

WIRELESS COMMUNICATION OVER WIDEBAND CHANNELS

Ulrich G. Schuster

WIRELESS COMMUNICATION OVER WIDEBAND CHANNELS

Hartung-Gorre Verlag Konstanz
2009

Reprint of Diss. ETH No. 17565

SERIES IN COMMUNICATION THEORY VOLUME 4

edited by Helmut Bölcskei

Bibliographic Information published by Die Deutsche Bibliothek

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

Copyright © 2009 by Ulrich G. Schuster

First edition 2009

HARTUNG-GORRE VERLAG KONSTANZ

ISSN 1865-6765

ISBN-10 3-86628-245-1

ISBN-13 978-3-86628-245-2

To my parents

Abstract

The amount of data that can be reliably transmitted per second over a given channel, called the *channel capacity*, depends on the received power and on the number of degrees of freedom (DOFs) per second that the combination of transmitter, channel, and receiver allows for. An increase in either power or DOFs increases channel capacity, everything else being equal. Bandwidth and time are the main sources of DOFs in wireline as well as wireless communication systems; directional transmission and reception, i.e., the use of space, can offer additional DOFs in wireless systems. While the radiated power is strictly regulated for most applications of wireless communications, DOFs abound in so-called ultrawideband (UWB) channels of several gigahertz bandwidth, the license-free use of which has been permitted recently in the United States of America. Similar regulations for the use of UWB communications are expected for many other countries in the near future. Therefore, the focus of this thesis is on wireless channels with many DOFs and the performance of communication systems that operate over such channels.

Wireless channels change over time, space, and frequency in a seemingly random manner; therefore, each DOF in a wireless channel is commonly described by random coefficient. To communicate reliably, the receiver not only needs to resolve the uncertainty caused by the noise of all electronic components but also the uncertainty introduced by the random channel. We quantify the latter uncertainty through the number of degrees of uncertainty (DOUs)—effectively the number of DOFs with uncorrelated coefficients. Resolution of channel uncer-

tainty requires DOFs and power, which are then no longer available for communication. For example, we can send known pilot symbols over some DOFs to estimate the channel. It is not guaranteed that channel uncertainty can always be resolved. If the number of DOUs increases at the same rate as we add DOFs, e.g., by enlarging the bandwidth, adding more bandwidth might actually be detrimental after some point.

While the capacity under channel uncertainty is an information-theoretic problem, the relation between the number of DOFs and DOUs depends on the physical channel and its mathematical model. In the first part of this thesis, we review standard channel models and their physical foundation, all with special emphasis on channels of wide bandwidth. Of particular importance for information-theoretic analysis is a suitable stochastic channel model, i.e., a joint distribution for the time-variant channel impulse response that is accurate yet mathematically tractable.

As common modeling assumptions for channels of small bandwidth might no longer hold for UWB channels, we complement the theoretical modeling considerations in Part I with statistical analysis of measured wideband channels in Part II. We describe two channel measurement campaigns in the band from 2 GHz to 5 GHz conducted in a public space; in the first campaign we moved the transmit antenna on a regular grid and kept the environment static, and in the second campaign we fixed the antennas while people were moving about the environment. On the basis of the measured channel impulse responses, we select marginal amplitude distributions from the Rayleigh, Rice, Nakagami, lognormal, and Weibull families by means of information criteria and use tools from multivariate statistical analysis to obtain a stochastic channel description of second order. While the channel with moving terminals can be sensibly modeled as zero-mean jointly proper Gaussian (JPG) distributed, measurement data for the channel with static terminals does not seem to contain sufficient evidence to unequivocally select a single stochastic channel model. But physical considerations, like a strong mean component in every channel tap, and the need for a parsimonious mathematical model prompt us to

advocate the JPG distribution with nonzero mean to describe the latter type of channel. An analysis of channel correlation matrices for the second measurement campaign shows that their number of significant eigenvalues scales linearly with increasing bandwidth. We interpret this scaling behavior as an indication that the number of DOUs increases linearly with the number of DOFs over the measured frequency band.

The modeling considerations in Part I and Part II indicate that a JPG linear time-variant description might be adequate for channels with several gigahertz bandwidth. Hence, we use a discretized version of the standard proper Gaussian wide-sense stationary uncorrelated scattering (WSSUS) channel model for the information-theoretic analysis in Part III. We extend said model to the spatially correlated multiantenna setting and use it to derive bounds on channel capacity under a constraint on the transmit signal's peak power and under the assumption that neither the transmitter nor the receiver know the instantaneous channel realization but both know the channel law. The bounds are useful for a large range of bandwidth and allow to coarsely identify the capacity-optimal combination of bandwidth and number of transmit antennas. We also obtain a closed-form expression for the first-order Taylor series expansion of capacity in the limit of large bandwidth. From this expression, we conclude that in the wideband regime: (i) it is optimal to use only one transmit antenna when the channel is spatially white; (ii) one should transmit along the strongest channel eigenmode if the channel is spatially correlated; (iii) spatial correlation is beneficial.

Kurzfassung

Die Datenmenge welche pro Sekunde zuverlässig über einen gegebenen Kanal übertragen werden kann, *Kanalkapazität* genannt, hängt ab von der empfangenen Leistung sowie von der Anzahl der Freiheitsgrade pro Sekunde des effektiven Kanals, bestehend aus Sender, physikalischem Kanal und Empfänger. Eine Erhöhung der empfangenen Leistung oder der Anzahl Freiheitsgrade führt zu einer Kapazitätssteigerung, vorausgesetzt dass alle anderen Systemparameter konstant gehalten werden. Bandbreite und Zeit sind ursächlich für Freiheitsgrade in drahtgebundenen sowie drahtlosen Kommunikationssystemen; zusätzliche Freiheitsgrade in drahtlosen Systemen können durch richtungsgebundene Übertragung, also durch die Nutzung des Raumes, erzielt werden. Während die abgestrahlte Leistung in den meisten drahtlosen Kommunikationssystemen streng beschränkt ist, stehen Freiheitsgrade in so genannten ultra-breitband (UWB) Kanälen mit mehreren Gigahertz Bandbreite im Überfluss zur Verfügung. Die lizenzgebührenfreie Nutzung solcher Kanäle wurde vor kurzem in den Vereinigten Staaten von Amerika zugelassen; ähnliche Betriebsgenehmigungen sind in der nahen Zukunft in vielen anderen Ländern zu erwarten. Aus diesem Grund liegt das Schwerpunkt der vorliegenden Dissertation auf Funkkanälen mit vielen Freiheitsgraden sowie auf der Leistungsfähigkeit von Kommunikationssystemen die ebendiese Kanäle verwenden.

Funkkanäle ändern sich in scheinbar zufälliger Weise im Lauf der Zeit, mit der räumlichen Position der Antennen, sowie mit der Frequenz. Daher wird gewöhnlich jeder Freiheitsgrad durch einen Ko-

effizienten in form einer Zufallsvariable modelliert. Um zuverlässige Kommunikation zu ermöglichen muss der Empfänger nicht nur die Unsicherheit bezüglich der empfangenen Nachricht in Folge des additiven thermischen Rauschens beseitigen, sondern auch die Kanalunsicherheit. Wir beschreiben letzter durch die Anzahl statistischer Freiheitsgrade—die Anzahl der regulären Freiheitsgrade mit unkorrelierten Koeffizienten. Um die Kanalunsicherheit zu beseitigen, müssen Sendeleistung und Freiheitsgrade aufgewendet werden, die letztendlich nicht mehr für die Datenübertragung zur Verfügung stehen. Wir können zum Beispiel einige Freiheitsgrade reservieren um dem Empfänger bekannte Symbole zu senden, mit deren Hilfe dann eine Schätzung aller Freiheitsgrade möglich ist. Wächst nun die Anzahl statistischer Freiheitsgrade zu schnell mit steigender Anzahl regulärer Freiheitsgrade das Kanals, so besteht die Gefahr dass die Wahl einer größeren Bandbreite oder mehrerer Sende- und Empfangsantennen die zuverlässig erzielbare Datenrate reduziert anstatt sie zu erhöhen.

Die Frage nach der Kapazität unbekannter Schwundkanäle ist informationstheoretischer Natur; der Zusammenhang zwischen regulären und stochastischen Freiheitsgraden ist jedoch eine Funktion des physikalischen Funkkanals und des verwendeten mathematischen Modells. Im ersten Teil der vorliegenden Dissertation geben wir daher einen Überblick über oft benutzte Kanalmodelle und deren physikalische Grundlagen unter besonderer Beachtung von Kanälen großer Bandbreite. Für informationstheoretische Betrachtungen sind insbesondere stochastische Kanalmodelle von großer Bedeutung, d.h., die Beschreibung der zeitvarianten Kanalimpulsantwort durch eine möglichst getreue, gleichzeitig jedoch mathematisch gut handhabbare Zufallsverteilung.

Viele gebräuchliche Annahmen zur Modellierung von Funkkanälen mit einigen Megahertz Bandbreite erscheinen zweifelhaft für die Modellierung von UWB Kanälen. Daher ergänzen wir die theoretischen Betrachtungen in Teil I dieser Arbeit durch die statistische Analyse gemessener Breitbandkanäle in Teil II. Wir erläutern zwei Messkampagnen, welche wir in einem öffentlichen Raum durchgeführt haben. In der ersten Kampagne nahmen wir Kanalstichproben an verschiedenen

räumlich versetzt Messpunkten auf während keine Bewegung in der Umgebung zu verzeichnen war. In der zweiten Messkampagne sorgten sich bewegenden Menschen für eine Änderung des Kanals während wir Sende- und Empfangsantennen an Ort und Stelle beließen. Die so gewonnenen Kanalimpulsantworten benutzen wir um Randverteilungen der Kanalamplitude mittels so genannter Informationskriterien zu bestimmen und um eine stochastische Beschreibung zweiter Ordnung mittels konventioneller multivariater statistischer Methoden zu erhalten. Es zeigt sich, dass der Kanal mit räumlich variierender Antennenposition durch eine mittelwertfreie ordentliche multivariate Normalverteilung beschrieben werden kann, wohingegen unsere Messdaten nicht genügend Anhaltspunkte für eine eindeutige statistische Charakterisierung des Kanals mit stationären Antennen bieten. Auf Grund physikalischer Überlegungen, wie zum Beispiel eines starken Mittelwerts in jedem Wert der diskreten Kanalimpulsantwort, sowie der Notwendigkeit eines mathematisch handhabbaren Modells empfehlen wir auch für letzteren Typ von Funkkanal die Beschreibung mittels einer ordentlichen multivariaten Normalverteilung, jedoch mit positivem Mittelwert für jeden Kanalwert. Eine Analyse der Kanalkorrelationsmatrizen, berechnet aus den Messdaten der zweiten Messkampagne, ergibt dass deren Anzahl signifikanter Eigenwerte linear mit der Kanalbandbreite zunimmt. Wir interpretieren dieses Resultat dahingehend, dass die Anzahl stochastischer Freiheitsgrade des Kanals linear mit der Anzahl regulärer Freiheitsgrade wächst.

Die Überlegungen zur Kanalmodellierung in Teil I und II sind Indiz dafür dass eine ordentliche reguläre Normalverteilung UWB Kanäle in angemessener Weise beschreiben kann. Folglich greifen wir für unsere informationstheoretischen Studien in Teil III auf die diskretisierte Version eines weit verbreiteten Kanalmodells auf der Grundlage einer ordentlichen Normalverteilung mit stationärer Zeit- und Frequenzvariation zurück, das so genannte WSSUS Model, und erweitern dieses Model zur Beschreibung von Mehrantennensystemen mit räumlicher Korrelation. Für dieses erweiterte WSSUS Model entwickeln wir obere und untere Schranken auf die Kanalkapazität unter der Annahme, dass die maximale Amplitude des gesendeten Signals streng beschränkt

ist, und dass weder Sender noch Empfänger die augenblickliche Realisierung des Kanals, hingegen jedoch die stochastische Kanalbeschreibung kennen. Die Schranken sind nützlich für eine große Spanne von Bandbreiten; sie erlauben, in groben Zügen die kapazitätsoptimale Kombination aus Bandbreite und Anzahl der Sendeantennen zu bestimmen. Des weiteren berechnen wir einen Ausdruck für das Glied erster Ordnung der Reihenentwicklung nach Taylor der Kapazität für den Grenzübergang zu unendlicher Bandbreite. Besagter Ausdruck ermöglicht folgende Schlussfolgerungen für Kanäle mit sehr großer Bandbreite: (i) Im Falle eines räumlich unkorrelierten Kanals ist es optimal, nur eine einzige Sendeantenne zu verwenden; andernfalls sollte nur entlang der Eigenmode des Kanals übertragen werden welche zum stärksten Eigenwert gehört. (ii) Räumliche Korrelation erhöht die Kanalkapazität.

Acknowledgments

During the course of my graduate studies at UC Berkeley and at ETH Zurich, I had the opportunity to work, discuss, and cooperate with many brilliant people; their thoughts, insights, critique, and motivation greatly helped me to learn and understand the fundamental concepts of communication and information theory, and to appreciate the beauty of theory as much as the fascination of real-world applications. Above all, through their own examples, they showed me how to conduct the research that led to this thesis. In particular, I would like to thank my advisor, Prof. H. Bölcskei, and Dr. G. Durisi for the fruitful cooperation, the results of which are documented in the following chapters.

Contents

Acknowledgments	xv
1 Wideband Channels	1
1.1 Problem Formulation and Overview	2
1.2 Models for Wideband Channels	3
1.3 Wideband Channel Measurements	6
1.4 Capacity of Wideband Channels	8
I Models for Wideband Channels	13
2 Channel Modeling Foundations	15
2.1 Modeling Philosophy	15
2.2 Physical Channel Models	18
2.2.1 Electrodynamic Foundations	19
2.2.2 Channel Variation	22
2.3 Input-Output Description of Wireless Channels	25
2.3.1 Types of Input-Output Relations	26
2.3.2 Multipath Input-Output Relations	27
2.4 Mathematical Structure	34
2.4.1 Linear Time-Variant Systems	35
2.4.2 The Channel as a Linear Operator	36
2.4.3 System Functions	38
2.4.4 Discretization of the System Functions	39

CONTENTS

2.5	Stochastic Models	42
2.5.1	Structure of a Stochastic Model	43
2.5.2	Marginal Distributions	45
2.5.3	Second-Order Characterization and the WSSUS Assumption	47
3	Discretized Models	53
3.1	Block Fading Models	54
3.1.1	Modeling Assumptions	54
3.1.2	The Specular Block-Fading Model	56
3.1.3	The Model of Saleh and Valenzuela	57
3.1.4	The Standard Discrete-Time Block-Fading Model	58
3.2	A Model for Underspread Channels	60
3.2.1	Underspread Channels	60
3.2.2	Approximate Diagonalization of Underspread Channels	61
3.2.3	Approximate Eigenfunctions and Eigenvalues	65
3.2.4	Discretized Input-Output Relation	70
3.2.5	OFDM Interpretation of the Approximating Channel Model	74
3.2.6	OFDM Pulse Design for Minimum ISI and ICI	76
II	Empirical Characterization of Wideband Channels	81
4	Wideband Channel Measurements	83
4.1	Measurement Principles	83
4.1.1	General Considerations	83
4.1.2	Measurement Techniques	84
4.1.3	Time-Domain Measurements	86
4.1.4	Frequency-Domain Measurements	89
4.1.5	Calibration	90
4.2	A Survey of Wideband Channel Measurements and Models	91
4.2.1	Measurements	91

4.2.2	Models	94
4.3	The ETH Measurement Campaigns	98
4.3.1	Objectives	99
4.3.2	Measurements	101
4.3.3	Limitations	102
4.4	Measurement Setup and Equipment	104
4.4.1	Measurement Campaign I	104
4.4.2	Measurement Campaign II	112
4.4.3	Measurement Environment	116
4.5	Postprocessing	121
4.5.1	Measurement Campaign I	122
4.5.2	Measurement Campaign II	127
4.6	Data Exploration	134
4.6.1	Power-Delay Profiles	136
4.6.2	Mean Delay and Delay Spread	142
5	Marginal Tap Distributions	145
5.1	Preliminaries	145
5.2	The Statistical Modeling Problem	148
5.3	Model Selection	150
5.3.1	Candidate Families	151
5.3.2	Parameter Estimation	157
5.3.3	Goodness-of-Fit Tests	159
5.3.4	Discrepancy Minimization	161
5.4	Akaike's Information Criterion	165
5.4.1	Kullback-Leibler Discrepancy	166
5.4.2	Approximating the Trace Term	168
5.4.3	Akaike Weights	170
5.4.4	A Simulation Study	171
5.4.5	Comments	178
5.4.6	Other Model Selection Criteria	181
5.5	MCI Marginal Tap Distributions	182
5.5.1	Observations	183
5.5.2	Analysis and Interpretation	189
5.6	MCII Marginal Tap Distributions	192

CONTENTS

5.6.1	Observations	196
5.6.2	Analysis and Interpretation	197
5.7	Complex Tap Distributions	200
6	Second-Order Analysis	205
6.1	The Jointly Proper Gaussian Distribution	205
6.2	Intertap Correlation	208
6.3	Eigenvalues and Degrees of Uncertainty	213
6.3.1	Eigenvalue Estimation	214
6.3.2	Counting the Effective Degrees of Uncertainty	216
6.4	Capacity Estimates	221
6.4.1	Ergodic Capacity	222
6.4.2	Outage Capacity	223
6.5	Summary and Discussion	225
6.5.1	Marginal Tap Distributions	226
6.5.2	Joint Tap Distribution	230
III	Capacity of Multiantenna WSSUS Channels	233
7	Capacity Bounds	235
7.1	From Channel to System Model	236
7.1.1	Aspects of a Realistic System Model	236
7.1.2	Modeling Assumptions	238
7.2	Related Capacity Results	243
7.2.1	High-SNR Regime	244
7.2.2	Wideband Regime	246
7.3	System Model	248
7.3.1	Discretized IO Relation	249
7.3.2	Channel Correlation	251
7.3.3	Power Constraints	253
7.3.4	Spatially Decorrelated IO Relation	254
7.4	An Upper Bound on Capacity	255
7.4.1	Bounding Idea	256
7.4.2	The Upper Bound	258

7.4.3	Refinement for Channels of Practical Interest	265
7.4.4	Impact of Channel Characteristics	266
7.5	A Lower Bound on Capacity	269
7.5.1	The Lower Bound	269
7.5.2	Some Simplifications and Approximations	272
7.6	Numerical Examples	275
7.6.1	Setting	275
7.6.2	Evaluation Methodology	276
7.6.3	Numerical Results	277
7.6.4	Observations	280
7.7	Asymptotic Wideband Analysis	283
7.7.1	Wideband Asymptotes	283
7.7.2	Impact of Spatial Correlation	285
8	Technical Results	287
8.1	A Determinant Inequality	287
8.2	Majorization	288
8.2.1	Basic Definitions and Results	288
8.2.2	Majorization and Hadamard Products	290
8.2.3	Alternative Proof of Lemma 8.1	292
8.3	Mutual Information and MMSE	293
8.4	Optimization of the Upper Bound	298
8.5	Penalty Term of the Lower Bound	301
8.5.1	Upper Bound	301
8.5.2	Lower Bound	306
8.6	Taylor Series Expansion of Capacity	306
8.6.1	Upper Bound	306
8.6.2	Lower Bound	308
9	Conclusions	315
9.1	Modeling Consequences	316
9.2	Design Implications	317
9.3	Open Problems	319

CONTENTS

A	Notation	323
A.1	Miscellaneous	323
A.2	Linear Algebra	324
A.3	Probability and Statistics	325
A.4	Communication and Information Theory	326
A.5	Linear Time-Variant Systems	327
B	Acronyms	329
	References	333
	List of Publications	359
	Curriculum Vitae	361

CHAPTER 1

Wideband Channels

POWER and degrees of freedom are the two fundamental resources in communications engineering. The ratio of received power to the number of degrees of freedom (DOFs) is proportional to the signal-to-noise ratio (SNR) at the receiver, the single most important figure of merit in the design of a communication system. Bandwidth and time are the premier sources of DOFs; a channel of bandwidth B offers approximately $2B$ DOFs per second to communicate over. In the classical additive white Gaussian noise (AWGN) channel studied by Shannon (1948), receive power and the number of DOFs completely determine the maximum rate of reliable transmission, called the *capacity* of the channel. If the SNR is low, the channel is said to be operated in the *power-limited regime*, also called the *wideband regime*. Here, an increase in received power results in an almost linear increase of channel capacity, while the marginal return of each additional DOF approaches zero. Conversely, if the SNR is high—the channel is said to be operated in the *bandwidth-limited regime*—additional DOFs lead to an almost linear capacity increase while an increase in received power only increases capacity logarithmically.

An additional source of DOFs in wireless communications is *space*. Multiple antennas at the transmitter and the receiver enable directional transmission and reception, so that a multiple input multiple output (MIMO) system can sustain several spatially separated data

streams in a common frequency band.

Compared with the AWGN channel, the major difference of most wireless channels is channel variation. The attenuation and phase associated with each DOF varies as the transmitting and receiving antennas are displaced or if the propagation environment changes. As the propagation environment and its variation over time is unknown upon system design, the resulting propagation channel is best treated as random. Hence, the attenuation and phase of each DOF is best modeled as a random variable. We refer to a set of DOFs whose associated random coefficients are fully correlated as a degree of uncertainty (DOU). To ascertain the transmitted message, the receiver of a wireless communication system needs to jointly resolve the uncertainty introduced by the additive noise and by the DOUs.

1.1. PROBLEM FORMULATION AND OVERVIEW

A widely used simplification when designing a wireless communication system is to assume that the receiver does not suffer from channel uncertainty, i.e., that it knows the realization of each DOU perfectly. This so-called *coherent* setting is well understood, and, despite its obvious simplification, provides correct design guidelines, as evinced by the many wireless communication systems in use today. However, most of these systems operate over small or moderate bandwidth, from several kilohertz up to about 40 MHz, and the SNR at the receiver required for satisfactory operation is high. In the present work, we are interested in systems that operate over larger bandwidth, i.e., over channels with many DOFs and low effective SNR. If the number of DOUs grows if we increase the number of DOFs, i.e., increase channel bandwidth or use additional transmit and receive antennas, the coherent assumption might no longer be suitable because resolution of channel uncertainty becomes challenging, so much that the rate gain afforded by the additional DOFs might be completely negated. To study this effect, we need to understand how channel uncertainty is linked to physical channel parameters, and we need to devise a

model to analyze the behavior of channel capacity as a function of the DOFs and DOUs.

Models for wireless channels abound for all imaginable application scenarios and propagation environments. As our focus is on fundamental aspects of wideband channels, we need to use a model that captures those fundamental aspects but that is at the same time mathematically tractable. Part I of this thesis is devoted to modeling wideband wireless channels. It is mainly tutorial in nature and sets the stage for the core results in subsequent chapters.

Many of the standard modeling assumptions are questionable for channels of very wide bandwidth and, therefore, require empirical scrutiny. Although results of several measurement campaigns aimed at the characterization of so-called ultrawideband (UWB) channels are available, almost all of these measurement campaigns were designed to devise channel models suitable for simulation, but not to answer fundamental modeling questions, like “What is a suitable distribution to stochastically model UWB channels?”, or “How are DOFs and DOUs related?” Therefore, we conducted our own channel measurements, the results of which are reported in Part II.

In Part III, we combine the theoretical modeling considerations from Part I and the empirical findings from Part II and analyze the capacity of wideband systems under a set of realistic constraints. We obtain capacity bounds that depend on the number of DOFs provided by both bandwidth and space.

1.2. MODELS FOR WIDEBAND CHANNELS

All modeling is an approximation of reality; the key is to discern those aspects of reality that are fundamental for a given application, i.e., that significantly influence the predictions of the model. As we are mainly interested in communication-theoretic analysis of wideband channels—as opposed to cell planning and coverage analysis—we restrict our attention to channel variations over short time scales and small spatial displacements of transmit and receive antennas,

commonly referred to as *small-scale fading*.

Our modeling efforts in Chapter 2 follow the standard approach in wireless channel modeling to decompose the electrodynamic field emitted by the transmitting antenna into waves that propagate along a number of *paths* and to stochastically describe the resulting superposition of many individual propagation paths at the receiver. However, this approach becomes questionable if the bandwidth under study is very large, on the order of several gigahertz, as then objects in the environment might be much larger than the smallest wavelength under consideration but of similar extent as the largest wavelength. Furthermore, it becomes questionable if time variation of the channel can still be described by Doppler shifts or if the more complicated scale representation needs to be used. Because there does not seem to be much evidence to date that the more general but also less tractable time-scale model is indeed more accurate and leads to better predictions, we maintain the standard delay-Doppler formulation of channel variability.

Wireless channels are, in general, *time-selective*, i.e., they change in a correlated way as time progresses, and *frequency selective*—channel attenuation is different for different frequencies of the transmitted signal, where adjacent frequencies might be correlated. Time and frequency selectivity can be captured mathematically by modeling wireless channels as linear time-variant (LTV) systems. In a stochastic LTV channel model of the form we review in Chapter 2, a second-order description is quite complex, as all correlation functions depend on four parameters in general. A standard modeling assumption is to describe time variations of the channel as wide-sense stationary (WSS) and individual propagation paths as uncorrelated (uncorrelated scattering, US), which leads to the widely used wide-sense stationary uncorrelated scattering (WSSUS) channel model. Of particular importance is that a US model predicts a linear scaling of the number of DOUs with bandwidth. However, for channels of very wide bandwidth, the US assumption becomes questionable. An individual propagation path might experience frequency-selective propagation mechanisms that result in dispersion and subsequent temporal overlapping and correla-

tion of several paths. We attempt to empirically evaluate in Part II if the latter effect is strong enough to invalidate the US assumption for practical purposes.

All the models reviewed and devised in Chapter 2 are time continuous. But for channel measurements, statistical analysis, and information-theoretic studies, a discretized channel model is more suitable, i.e., a model that depends on a countable or even a finite set of parameters. We review several such discretized channel models in Chapter 3, like Turin's specular block-fading model, the clustered impulse response model by Saleh and Valenzuela, or the standard discrete-time block-fading model. As these models are of block-fading type, they are simplifications of the general stochastic LTV model and do not account for channel correlation between blocks. All mentioned models are widely used. In particular, the discrete-time block-fading model plays an important role in Part II, because we take our measured data to be discrete-time channel impulse responses of a block-fading channel.

As channel correlation in both time and frequency might be profitably used by a wireless system to combat channel uncertainty, we also review a discretization and approximate diagonalization of the more general WSSUS LTV channel model. This discretization, devised by Kozek, builds on the important observation that all wireless channels of interest are *underspread*, i.e., that their product of maximum delay and maximum Doppler shift is small. Underspread WSSUS channels have a highly structured set of approximate eigenfunctions that do not depend on the channel realization and are well concentrated in time and frequency. We can construct a canonical signaling scheme that closely resembles orthogonal frequency division multiplexing (OFDM) by projecting the transmitted signal onto these eigenfunctions. The received signal can be similarly decomposed to obtain a set of scalar input-output (IO) relations. The resulting channel description forms the basis for our capacity analysis in Part III.

1.3. WIDEBAND CHANNEL MEASUREMENTS

Many UWB channel measurement campaigns are devised to parameterize channel models used to simulate communication systems. As our focus is on channel models for theoretical analysis and on fundamental modeling aspects, we take a somewhat different approach to both measurements and statistical analysis of the resulting data than in most reported measurement campaigns.

One main problem with channel measurements is technical: how to obtain data of high fidelity in a form that is easy to handle and to interpret. We were fortunate to have measurement devices of high quality at our disposal, which allowed us to record channel samples at high SNR even over distances between transmitting and receiving antenna of about 27 m. A second problem is of more philosophical nature: the question what one actually measures and what assumptions are implicit in any statistical analysis of the measured data. This problem shows up when designing the measurement system, upon postprocessing the measured data, and during statistical analysis. For example, our measurement setup only allows to characterize block-fading models even if the channel was time variant.

We conducted two measurement campaigns in the entrance lobby of the main electrical engineering building at ETH Zurich, a typical public open space environment with large windows and a tiled floor. Both campaigns are described in detail in Chapter 4. In measurement campaign I (MCI), we fixed the position of the receive antenna, kept the environment static, and automatically positioned the transmit antenna at 90 different locations on a rectangular grid. For each location of the transmit antenna, we recorded the transfer function of the channel with a vector network analyzer (VNA). Such a measurement setup represents applications like wireless access point scenarios, where the terminals move with respect to the environment.

In wireless personal area networks and wireless sensor networks, e.g., for building automation or industrial applications, the terminals are typically static, and variations in the channel are mainly caused by moving scatterers, like moving persons. To characterize

the resulting channel, we fixed both the transmit and the receive antenna in measurement campaign II (MCII), sounded the channel with a pseudonoise sequence of chip rate 10 GHz, and recorded the received signal in real time with a digital sampling oscilloscope (DSO). In both MCI and MCII, we obtain, after postprocessing, channel impulse response vectors for the band from 2 GHz to 5 GHz, which we subsequently use for statistical analysis as described in Chapter 5 and Chapter 6.

Because the measured impulse responses are quite long, with over 1000 channel taps, a complete characterization of their joint distribution is impossible. Instead, we characterize the marginal tap distributions in Chapter 5 and attempt to obtain a second-order description in Chapter 6.

In Chapter 5, we set out to select, from a set of candidate distributions, a suitable stochastic description for the tap amplitudes of both MCI and MCII impulse responses. Our candidate set contains the parametric families of Rayleigh, Rice, Nakagami, lognormal, and Weibull distributions. The goal in statistical model selection is to choose a distribution that approximates as closely as possible the *operating model*, defined as the nearest representation of the true situation that can be constructed by means of a probability model. A good model should be based on *physical insight*, be *mathematically tractable*, it should be *general*, and lead to *consistent predictions*. A measure of approximation quality between a candidate distribution and the distribution of the operating model is called a *discrepancy*. For reasons detailed in Chapter 5, we do not follow the standard approach in the channel modeling literature to assess the approximation quality of a distribution family by means of goodness of fit (GOF) tests. Instead, we use Akaike's Information Criterion (AIC), which is based on the Kullback-Leibler (KL) relative entropy as a measure for the similarity between distributions. AIC explicitly takes into account the number of samples available for statistical analysis and the complexity of the candidate families. We find that the Rayleigh distribution is still adequate to model the amplitude of UWb channel taps if channel variations result from motion of the

antennas, i.e., in MCI; an exemption are taps that belong to clusters in the impulse response, most probably caused by strong reflections from walls and windows. The Ricean distribution provides the best fit for UWB channel tap amplitudes when transmitter and receiver are static and channel variations are caused by movements of objects in the environment, i.e., in MCII. From the Rayleigh, respectively Rice amplitude distribution it is only a small step to hypothesize that the corresponding complex-valued channel taps are proper Gaussian distributed, i.e., circularly symmetric Gaussian around a possibly nonzero mean component. A Kolmogorov-Smirnov test shows that for MCI the proper Gaussian tap distribution is indeed a viable stochastic model. The picture is not that clear for MCII data, where the same test is inconclusive when viewed globally for all taps of a given impulse response. As a strong mean component seems to be present in most taps, we nevertheless opt for a proper Gaussian model of the MCII taps.

A second-order characterization of the joint impulse response distribution is the main theme of Chapter 6. An attempt to estimate the complete channel correlation matrix of size 700×700 is hampered by a lack of samples. Even though we only use MCII data, where we have up to 2722 samples from a single measurement set, the resulting correlation matrices are difficult to interpret. Intertap correlation is low in general, yet the eigenvalues of the estimated channel correlation matrices seem to be significantly impacted by correlation when compared with a hypothetical uncorrelated channel. However, we find that the number of significant eigenvalues of the channel covariance matrices, which we take as a measure for the number of DOUs, scales approximately linearly with bandwidth.

1.4. CAPACITY OF WIDEBAND CHANNELS

Channel models are not an end in themselves but tools for analysis and design of communication systems. Contrary to claims in the literature, our empirical analysis in Part II does not provide sufficient

evidence to warrant a complete renunciation of the standard WSSUS model discussed in Part I. The finding of marginally Rayleigh, respectively Rice distributed tap amplitudes agrees with the jointly proper Gaussian (JPG) tap distribution in the WSSUS model, as does our finding of a linear increase of the number of DOUs with bandwidth. Some experimental findings, like the Ricean component in MCI or the specular reflections in MCI, are not well captured by the WSSUS model although they might be profitably exploited in system design. However, our goal is not to analyze particular channel features in isolation but the overall effect of fading channels with many DOFs and DOUs. Therefore, we take as our starting point Kozek’s discretized and approximately diagonalized WSSUS model from Chapter 3, well aware of the provision that our channel measurements do not allow for characterization of time variation, and that some of our empirical findings have no direct counterpart in this discretized model. Yet, we believe that this model provides a suitable compromise between mathematical tractability, accuracy, and generality, so that our capacity results, if parameterized appropriately, are useful for a wide range of application scenarios.

In accord with the main theme of this thesis—to analyze fundamental effects of practical importance—we consider the noncoherent setting where neither transmitter nor receiver know the channel realization but both are aware of the channel law. We impose a constraint on the average transmitted power to model aspects like limited battery resources, and we constrain the radiated peak power in a given time interval and frequency band. The latter constraint models typical regulatory limitations for UWB systems. A closed-form expression for the capacity of noncoherent channels is not known to date, not even for the simple flat fading memoryless channel model. Therefore, we resort to derive upper and lower bounds on capacity as a function of the DOFs, which in the WSSUS setting are linearly related to the DOUs. Although we were not able to measure MIMO UWB channels, we extend our analysis to the MIMO setting to include the impact of spatial DOFs. In particular, we consider a point-to-point MIMO link where the component channels between individual transmit and

receive antennas are all discretized in the same way. This extension is a formal one so far, as we do not back it up by physical modeling considerations; a detailed treatment of spatial wideband channel modeling remains as an interesting open problem. However, as we allow for spatial correlation, chances are that future findings about wideband multiantenna channels can be mapped to our generic results simply by choosing appropriate spatial correlation matrices.

Our upper and lower bounds hold for a wide range of bandwidths, not just in the limit for bandwidth approaching infinity. While the bounds are loose for very high SNR—and correspondingly low bandwidth—they are reasonably good for many bandwidths of practical interest and become very tight for large bandwidth. The bounds allow to coarsely determine the number of DOFs, bandwidth and the number of antennas above which capacity starts to decrease. Because our channel model is fairly general, we can study the impact of channel correlation in time, frequency, and space. In particular, the larger the channel’s spread, i.e., the less memory in the channel, the lower are our capacity bounds. For spatially white channels and at large bandwidth, using a single transmit antenna is optimal, while multiple receive antennas are always beneficial. Spatial correlation can both increase or decrease capacity, depending on the receive SNR per DOF. Differently from the coherent setting, both transmit and receive correlation are beneficial at low SNR.

Research on the capacity of wideband channels during the past decade revealed that for a channel whose number of DOUs increases linearly with the number of DOFs the capacity starts to diminish from a certain critical number of DOFs on if the channel input is subject to some kind of peak limitation. From numerical evaluation of our upper and lower bounds, we find that current wideband systems and even many envisioned UWB systems operate over channels whose number of DOUs is small enough to avoid the problem of diminishing capacity with increasing DOFs. In fact, the critical number of DOFs above which our bounds predict a decrease in channel capacity is so high that our channel model probably breaks down before this critical number is reached.

1.4 CAPACITY OF WIDEBAND CHANNELS

We evaluate the bounds only for very simple channel parameters; yet, the direct link to the underlying WSSUS model allows for a more fine-grained analysis of wireless channels, because the bounds are explicit in the channels scattering function.

PART I

MODELS FOR WIDEBAND CHANNELS

CHAPTER 2

Channel Modeling Foundations

THE FIRST STEP in the solution of any engineering problem is to suitably model it, to find a mathematical approximation of the physical world we want to manipulate, because any quantitative analysis can only be performed on the basis of a model. A mathematical model comprises

- a set of controls: all determinants of physical features that are important for the application at hand,
- a set of observables: all aspects that we can measure to learn about the state of the physical reality,
- a mathematical description of the relation between controls and observables.

All three aspects depend not only on the physical reality under study but at least to the same extent on the intended application of the model.

2.1. MODELING PHILOSOPHY

A model is always an approximation of the physical world; it has to satisfy a conflicting set of requirements:

- *Accuracy*: a model needs to faithfully describe, reproduce, and predict the interrelation of a comprehensive set of controls and observables.

- *Low complexity*: we would like the number of controls and observables of a model to be low and its mathematical description to be tractable.
- *Generality*: a model should be adaptable to many similar physical situations.

All three requirements are not absolute but depend on the intended application of the model.

Accuracy

Accuracy has two aspects that are equally important: predictions of a model should be close to observations of the physical reality, and the variation across multiple predictions should be small. In statistical language, a model should have small bias and low variance, respectively. Inherent in the latter formulation is the notion that perfect accuracy does not exist and all modeling efforts are ultimately stochastic. The *loss function*, or *error criterion*, according to which we judge model accuracy depends on the application at hand, and the ultimate assessment of modeling accuracy is empirical—we need to actually measure those aspects of reality that we want to model and compare the measurements with the prediction of the model by means of our loss function. Because measurements are never perfect, this empirical approach implies that increasing the model complexity does not always increase model accuracy: it is not sensible to model aspects that exceed measurement accuracy or that are not accessible to measurements at all, for we would mistake measurement noise as features of reality (Forster and Sober, 1994). Measurement aspects are treated in detail in Part II.

Complexity

The admissible complexity of a model greatly depends on its intended application. A simulation model can contain thousands of controls and observables, and its mathematical description can be implicit, e.g., the solution of a fixed-point problem. On the other hand, a model

for mathematical analysis needs to be tractable, which most often means that its controls and observables need to be highly structured and that the relation between them should be available in closed form. We are mainly interested in the latter type of model.

Generality

Aspects that are not modeled do not exist for the purpose of analysis and design, although these aspects may very well influence the operation of the resulting implementation. As a device designed on the basis of a model will be faced with the real world, not the model used to design it, Murphy's law (Spark, 2003) predicts that the aspects not modeled will influence the operation of the device in a negative way with probability 1 (w.p.1). This suggests to model many aspects of the physical world, i.e., to keep the model general. On the other hand, if too many aspects are modeled, chances are that the engineering problem cannot be solved at all because the model is too complex to be tractable. Hence, we need to find the *fundamental* set of controls and observables, those that have the largest influence for the intended application of the model, and we need to choose them *parsimoniously*. The later principle is known as *Occam's Razor*: "Entities should not be multiplied beyond necessity."

A Model for Information-Theoretic Analysis

The three requirements just stated are often conflicting. For a given accuracy requirement, large generality means that the model should be able to cover many special cases. This often comes at the price of large complexity because the model needs to comprise many controls and observables that are only relevant in certain special situations. For fixed complexity, the model can be made more general if we relax the accuracy requirement: such a model covers a larger number of cases, but with larger modeling error in most of them compared with the error of a comparably complex but less general model.

Our main interest in this thesis is to gain fundamental insights

into wideband communications, i.e., to identify ultimate performance limits and design criteria. Thus, our models need to be mathematically tractable, which sharply limits model complexity. The following sections attempt to identify the distinctive and fundamental features of wireless channels so as to obtain parsimonious models.

2.2. PHYSICAL CHANNEL MODELS

Our starting point is the propagation of electromagnetic waves as modeled by Maxwell's equations. This model is far too complex; hence, we use the intuitively appealing *multipath approximation* to describe the electromagnetic field at the receiver and the resulting received signal, analyze the effect of frequency-dependent propagation mechanisms and temporal variation of the channel, and further simplify the channel model by means of a stochastic characterization of its IO relation.

The arguments presented in the following are not novel; most of them can be found in the books by Plonsey and Collin (1961) and McNamara et al. (1990) for the main physical developments, and the books by Parsons (2000), Rappaport (2002), Vaughan and Bach Andersen (2003) and Tse and Viswanath (2005) for the engineering arguments. What might set our presentation apart from other discussions of wireless channels is the focus on parsimonious models of low complexity for information-theoretic analysis, like the ones we see in Part III. On the basis of the general physical and mathematical modeling considerations in the present chapter, we discuss several more specific models for wireless channels in Chapter 3. A second distinctive element of our modeling approach is the focus on channels of large bandwidth. We use the term UWB loosely to refer to wireless channels whose bandwidth is significantly wider than in most commercial wireless systems today, i.e., hundreds of megahertz up to several gigahertz instead of tens of megahertz. Many widely accepted modeling assumptions become questionable for UWB channels. The only way to figure out which assumptions are still valid and which

need to be modified or abandoned altogether is to measure wideband channels and to analyze the resulting data. This is the theme of Part II.

2.2.1. Electrodynamic Foundations

Maxwell's equations provide for a very general and extremely elegant model for the time-variant electromagnetic field in a given volume of space (Plonsey and Collin, 1961). If all materials in the environment are linear, then the corresponding solution of Maxwell's equations is linear as well. Nevertheless, to compute such a solution for a real-world environment, we would require an enormous amount of information to specify all boundary conditions. Thus, to obtain a tractable engineering model, we need to reduce model complexity at the expense of modeling accuracy.

A fairly intuitive and successful approach to simplify the description of the electrodynamic field in a given environment is the multipath approximation: the description of the electromagnetic field at all points in space and time is replaced by individual *partial waves*, each of which travels along a specific geometric *propagation path* from the transmitting antenna to the receiving antenna and undergoes one or several *propagation effects* along its way. A great many different effects can be distinguished, but this defies the goal to arrive at a simple model; hence, we restrict ourselves to the, in our view, most important ones. The following paragraphs summarize these effects on the basis of the texts by McNamara et al. (1990), Parsons (2000), and Rappaport (2002).

The theoretical foundation of the multipath approximation is the so-called *high-frequency approximation* of electromagnetic wave propagation (McNamara et al., 1990). In its simplest form, this approximation describes propagation in purely geometrical terms as *rays* emanating from a source that are reflected and refracted from surfaces in the environment. Only the rays that arrive at the receiving antenna need to be considered. The main reason why this so-called geometrical optics (GO) model often yields accurate results is that

many objects can be treated as *electrically large*, i.e., their extent is large compared with the wavelengths under study, and their electrical properties do not change appreciably over this extent. Frequencies of interest for wireless communications range from about 1 GHz up to 60 GHz, which corresponds to wavelengths from around 0.3 m down to 5 mm, respectively; hence, the GO approximation is less accurate than, say, for light waves, whose wavelengths measure several hundred nanometers. Nevertheless, the interaction of radio waves with large objects, like floors, ceilings, walls, and windows, can be very well modeled by GO techniques, as experimentally supported by Dersch and Zollinger (1994).

Free Space Propagation

The simplest propagation mechanism is the radiation of a monochromatic electromagnetic wave in free space. In the so-called *far field region*, or *Fraunhofer region*, the total field along a ray emanating from the source can be approximated by the radiated field alone; its strength decays proportional to the inverse radial distance from the transmitting antenna; the constant of proportionality depends on the frequency of the signal.* For typical indoor applications with small terminals and operating frequencies well above 2 GHz, the far field approximation is in general a good one.

Antenna Effects

An ideal point source that radiates isotropically in all directions is physically not possible. Even the simplest antenna geometry, the Hertzian dipole, has a radiation pattern that is not uniform in space. Although all practical antennas show some directivity, it is sensible to assume a uniform antenna pattern, as such a pattern best models our ignorance of the antenna placement and orientation in typical wireless application scenarios, where terminal placement is beyond the control of the designer.

* See, for example, Fig. 1.2 in the book by Parsons (2000).

Diffraction

The GO approximation is not valid at object corners, wedges, curvatures, and other surface irregularities, even if the object under consideration is electrically large. Similarly, creeping waves along curved surfaces cannot be treated in purely geometrical terms. Both problems are addressed by the geometrical theory of diffraction (GTD) and its extension, the uniform theory of diffraction (UTD) (McNamara et al., 1990). The important feature of these theories for our purposes is that we may continue to use the abstraction of propagation along rays, but the interaction of rays with objects needs to be described by diffraction coefficients that are, in general, frequency dependent.

Scattering

If a radio wave that propagates along the direction of a ray in the above geometric approximation interacts with an object whose dimension is comparable to the wavelength, the UTD approximation is no longer applicable. Such interactions are commonly called *scattering*.* The scattering object is taken as the source of a new elementary wave by Huygen's principle. Exact computation of the resulting field is virtually impossible, as it requires exact knowledge of the surface geometry and material properties. What can be computed, however, is the power of the resulting scattered field at different points in space. Thus, a sensible simplification is to model the scattered field as a random quantity and to relate the parameters of its distribution to physical quantities like the overall dimension of the scattering object. A simplification along these lines for objects that are *electrically small*, i.e., with physical dimensions that are much smaller than the wavelengths of interest, is to treat them as *point scatterers*. A point scatterer does not disturb the field of the impinging wave and generates another elementary wave that radiates isotropically in all

* In many texts, scattering is also used to mean any type of interaction of radio waves with objects in the propagation environment, and every such object is called a *scatterer*.

directions, with amplitude and phase taken as random according to a distribution that depends on the effective size and surface properties of the object to be modeled as point scatterer. These properties are commonly measured by a single parameter called the *radar cross section* of the scatterer.

Reflections from Rough Surfaces

Another important propagation effect for which the UTD is not applicable is the reflection of a radio wave from a rough surface. The object of which the surface is part might be electrically large, but if surface protuberances are large as well,* the reflected field cannot be described analytically without perfect knowledge of the exact surface geometry. A viable approximation is to treat the rough surface as an infinite collection of point scatterers and to describe the resulting field stochastically.

2.2.2. Channel Variation

The GO and UTD approximations we use to dissect wave propagation as the propagation of partial waves along individual paths and the interaction of these partial waves with objects in the environment are vital to get an intuitive understanding of the electromagnetic field at different points in space. In particular, the following picture emerges:

The transmitting antenna illuminates a certain area, determined by its radiation pattern. [...] The incident wave also interacts with surface irregularities via diffraction, scattering, reflection, and absorption, creating a continuum of scattered partial waves. The amplitudes and phases of these partial waves depend on the physical properties of the surface structure such as geometrical proportions and electromagnetic reflection properties. At every

* Surfaces are often classified as *rough* if the difference between maximum and minimum protuberance is larger than the *Rayleigh number* $\lambda/(8 \sin \theta)$, where λ denotes the wavelength and θ the incidence angle of the impinging wavefront with respect to the surface normal vector.

point in space, scattered partial waves interfere with each other and, possibly with the direct wave, building up an irregular electromagnetic field. (Braun and Dersch, 1991)

As the terminals of wireless systems are often mobile and objects in the environment may be mobile as well, it is important to understand how the interference pattern of all the reflected, diffracted, and scattered partial waves changes over space. Similarly, because some propagation mechanisms are frequency dependent, it is important to understand how the aggregate electromagnetic field depends on frequency.

A. Spatial Variation

The irregular pattern of constructively and destructively interfering partial waves is a spatial phenomenon. Displacements of either the transmitter or the receiver by half a wavelength changes the phases of the interfering partial waves by up to 180 degrees, which can result in a drastic amplitude change of the electromagnetic field. At operating frequencies of several gigahertz, half a wavelength corresponds to several centimeters; significant changes in the received field strength may thus happen over very short distances. Therefore, the resulting phenomenon is aptly called *small scale fading*.

If we average the received field strength over a local area, i.e., over several wavelengths, the resulting mean field strength will still vary over space, yet changes will be more gradual and occur over longer distances. Propagation effects such as the decay of the field strength with distance, shadowing of the transmitted field by large obstacles, or transmission through walls are the main reason for this so-called *large scale fading*.

Large scale fading is of major concern for the design of cellular systems, as the attenuation of the field with distance determines the spatial coverage a single cell is able to provide (Lee, 1993; Rappaport, 2002). For this reason, much research effort is spent to characterize the *path loss*, i.e., the average power attenuation from the source to the sink, as a function of distance and environment characteristics like the number of walls or the type of buildings in a given area. For many

applications where terminal placement is beyond the control of the system designer, like in a consumer wireless local area network (WLAN), a wireless personal area network (WPAN), or in ad-hoc networks, large scale fading needs to be accounted for in the link budget and mitigated by high-level power control. The fundamental issue at the physical layer is the small scale fading, which necessitates modulation and coding schemes, power allocation, channel estimation, and cooperative schemes that are much different from their counterparts in wireline communications. Therefore, we are mainly concerned with small-scale fading in this thesis; *fading* in the following always refers to small scale fading, unless explicitly noted otherwise.

B. Temporal Variation

The pattern of constructive and destructive interference is spatial in nature; hence, if transmitter, receiver, and the environment are static, the field distribution over space will not change in time. Enter mobility. Motion of the receiver can be visualized as successively probing the wave field at different points in space. Hence, the receiver experiences constructive and destructive interference in quick succession, so that the received field strength and phase vary rapidly—the more rapid the faster the receiver moves. By reciprocity, the same picture holds true if the transmitter is in motion. Alternatively, we can imagine both transmitter and receiver being fixed and objects in the environment moving. This leads to less severe channel variation in general, as typically not all of the environment is in motion relative to the terminals in such a scenario; e.g., walls and ceilings remain fixed, but people or machines might move about.

C. Variation with Frequency

Many of the approximations of radio wave propagation used so far are valid only if the considered wavelengths are either small or large compared with the dimension of objects in the environment. As wideband channels are the main focus in this thesis, the description

of propagation effects with respect to a single frequency needs to be expanded upon. Several physical effects may lead to frequency-dependence of individual propagation paths.

- The free space loss depends on the absorption of the atmosphere, which is frequency dependent. Attenuation in general increases with frequency, albeit slowly in the band below 10 GHz. Therefore, we neglect this effect.
- Specular reflections and refractions, as described by the GO approximation, do not depend on frequency as long as the material properties of the reflecting and refracting surfaces are not frequency dependent. However, the assumption that the surfaces are electrically large might not be satisfied at all frequencies: e.g., below a certain frequency, a specular reflection might very well be more suitably modeled as scattering.
- A surface that is considered as smooth at low frequencies, leading to a specular reflection, might no longer be smooth at high frequencies and corresponding short wavelength.
- Diffraction is a frequency dependent phenomenon. Hence, diffraction coefficients vary with frequency; the exact behavior depends on the geometry of the diffracting object (McNamara et al., 1990).

2.3. INPUT-OUTPUT DESCRIPTION OF WIRELESS CHANNELS

The model of the electromagnetic field as the superposition of partial waves that result from simplified propagation mechanisms is a global one; it allows to describe the electromagnetic field at almost every point in space, and it enables differentiation between large-scale and small-scale variations of the received field strength. However, the resulting model complexity is still only manageable in a simulation model, as evinced, e.g., by Seidel and Rappaport (1994).

The focus of this thesis is on communication- and information-theoretic analysis of a single point-to-point wireless link. Consequently,

we only need to model the transformation a signal undergoes between transmitter and receiver. In this sense, a model for wireless channels is an IO description, it relates output signals to corresponding input signals. This section deals with the question of how to translate the physical channel description into a mathematical IO relation.

2.3.1. Types of Input-Output Relations

Wireless transmitters and receivers consist of several functional blocks that transform the communication signal in various ways. Therefore, we are at liberty to consider different points in the signal path at the transmitter as channel input and at the receiver as channel output. To simplify the following discussion, we adapt a classification from Vaughan and Bach Andersen (2003):

- The *propagation channel* describes solely the effect of electromagnetic wave propagation between a transmitting point source and a receiving point sink.
- The *radio channel* takes into account the effect of transmitting and receiving antennas in addition to all effects of the propagation channel.
- The *radio frequency (RF) channel* consists of the radio channel and the effect of all filters, duplexers, amplifiers, and other analog hardware that operates at RF, i.e., up to the frequency conversion stages.
- The *baseband channel* takes as its input and output continuous-time complex-valued signals, i.e., the signal before frequency up-conversion and after down-conversion, respectively.

Virtually all contemporary communication theory uses the baseband channel IO relation, as then all signals and systems can be represented economically in complex baseband notation (Zhang and Miller, 1992; Wilson, 1996; Proakis, 2001).

As we are interested primarily in communications aspects, we neglect most impairments in the baseband channel, like power amplifier compression, I-Q imbalance, DC-offset (Razavi, 1998; Yee, 2001), and oscillator phase noise (Kouznetsov and Meyer, 2000; Demir, 2002),

and only retain the two most fundamental impairments:

- The receiver effectively limits the duration and the bandwidth of the received signal.
- All resistive elements in the receive circuitry introduce thermal noise.

2.3.2. Multipath Input-Output Relations

In Section 2.2, we presented physical approximations for the propagation of electromagnetic waves mostly on intuitive grounds. In the present subsection, we try to express these approximations mathematically.

The wavefront impinging at the receiving antenna can be modeled as the superposition of many partial waves; and these partial waves can in turn be characterized by the various propagation effects they experience on their way along the different paths from the transmitting antenna to the receiving antenna. In the following, we first analyze the signal received over a single propagation path and assume that the only propagation mechanism on this path is a reflection off a smooth surface. The overall received signal can then be obtained by linearity as the superposition of many such contributions; this is essentially the approach taken by Kennedy (1969). In a second step, we include frequency dependent propagation behavior along the individual paths.

A. Specular Multipath Models

The signal received via a propagation path that only undergoes a specular reflection can be characterized by an amplitude coefficient $\rho(t)$ that models the path loss and the reflection loss, a phase shift $\varphi(t)$ that results from the material properties of the reflector, and by the overall delay $\tau(t)$ between transmission and reception along the selected path. In general, all three parameters may depend on time.

The signal received over the *radio channel* can then be written as

$$\begin{aligned} \Re\{r(t)e^{i2\pi f_c t}\} \\ = \Re\left\{\rho(t)e^{i2\pi\varphi(t)}x(t-\tau(t))e^{i2\pi f_c(t-\tau(t))}\right\}, \end{aligned} \quad (2.1)$$

where $x(t)$ and $r(t)$ are the baseband input and output signals, respectively, and f_c denotes the carrier frequency. The variation in delay over time is caused by the motion of the transmitter, the receiver, and the reflecting object. In typical indoor environments, motion is slow and there are no abrupt changes; therefore, we approximate $\tau(t)$ by its first-order Taylor series, $\tau(t) \approx \tau + \dot{\tau}t$, where $\dot{\tau}$ stands for the time derivative of the arrival delay $\tau(t)$. Furthermore, we neglect the time variation of the complex-valued reflection coefficient, i.e., we set $\rho(t) \approx \rho$ and $\varphi(t) \approx \varphi$; these latter approximations are admissible as long as we are only interested in small-scale effects, because reflection and attenuation change significantly only upon large variations of the total propagation distance along the path under consideration, and the reflection phase shift varies only with large changes in the angle of arrival (AOA) and the angle of departure (AOD) of a path. Furthermore, we absorb the phase shift caused by the specular reflection into the delay coefficient, i.e., $\rho e^{i2\pi\varphi} \rightarrow \rho$. Consequently,

$$\Re\{r(t)e^{i2\pi f_c t}\} = \Re\left\{\underbrace{\rho x([1-\dot{\tau}]t-\tau)}_{r(t)} e^{-i2\pi f_c(\tau+\dot{\tau}t)} e^{i2\pi f_c t}\right\}. \quad (2.2)$$

The change in delay, which results from a change in path length, has two effects. First, the transmitted signal $x(t)$ is compressed in time by a factor $(1-\dot{\tau})$; this effect is called *time dilation* or *scaling* (Cohen, 1993). Second, the transmitted signal $x(t)$ is delayed by τ . Define $a \triangleq 1-\dot{\tau}$ and $b \triangleq \tau/(1-\dot{\tau})$; then, (2.2) can be rewritten as

$$\Re\{r(t)e^{i2\pi f_c t}\} = \Re\left\{\rho e^{i2\pi f_c a(t-b)}\sqrt{|a|x(a(t-b))}\right\}. \quad (2.3)$$

To obtain the corresponding IO relation of the wireless channel, we use the modeling assumption discussed previously: the total electro-

magnetic field at the receiver can be expressed as the linear superposition of many partial waves, each of which can be described by the propagation effects along a specific propagation path from the transmitter to the receiver. It follows from (2.3) that each path can be characterized by a corresponding parameter pair (a, b) . With high probability, no two paths have exactly the same parameter combination (a, b) , so that we can index the attenuation coefficient for a particular path (a, b) as $\rho(a, b)$. To simplify notation, we aggregate the path coefficients of (2.3) into a function $c_{\mathbb{S}}(a, b) \triangleq \rho(a, b)e^{i2\pi f_c a(t-b)}$, the so-called *wideband spreading function*.^{*} If there are many propagation paths, by linearity we can express the aggregated received signal in integral form as

$$\Re\{r(t)e^{i2\pi f_c t}\} = \Re\left\{\iint_{ab} c_{\mathbb{S}}(a, b)\sqrt{|a|}x(a(t-b))dbda\right\}, \quad (2.4)$$

i.e., the received signal is a superposition of delayed and scaled replicas of the transmitted signal, weighted by the wideband spreading function $c_{\mathbb{S}}(a, b)$.

If $x(t)$ varies slowly enough, we can simplify the IO relation (2.4). Let D denote the effective duration and B the effective bandwidth[†] of $x(t)$; then, we can approximate

$$x([1 - \dot{\tau}]t - \tau) \approx x(t - \tau) \quad (2.5)$$

as long as

$$DB \ll \frac{1}{|\dot{\tau}|}. \quad (2.6)$$

A more intuitive condition results if we relate the rate of change in delay $\dot{\tau}$ to the joint radial velocity v_r of transmitter, reflector, and

^{*} The subscript \mathbb{S} indicates that we consider a channel \mathbb{S} in which propagation is affected only by *specular* reflections.

[†] See Definition 3.2 to follow.

receiver as* $\dot{\tau} = -v_r/c$, where c stands for the speed of light in air. This allows us to define the *Doppler shift* $\nu \triangleq -f_c \dot{\tau} = f_c v_r/c$, and to express the condition (2.6) as

$$DB \ll \frac{f_c}{|\nu|}, \quad (2.7)$$

i.e., the approximation (2.5) is valid if the *time-bandwidth product* DB of the transmitted signal is appreciably less than the ratio $f_c/|\nu|$ between carrier frequency and Doppler shift, or—stated differently—if the *fractional bandwidth* B/f_c is significantly smaller than $1/(|\nu|D)$. This assumption is commonly referred to as the *narrowband assumption* (Weiss, 1996), although it should be more appropriately called the *small fractional bandwidth* assumption. Effectively, we approximate scaling of the received signal by Doppler shifts. Typical communication signals have duration D well below one second,[†] and the maximum Doppler shift encountered in indoor channels is often less than 10 Hz. Therefore, many communication signals are of small fractional bandwidth although their absolute bandwidth might be quite large. The narrowband assumption leads to the following simplification of the one-path IO relation (2.2) in baseband notation:

$$\begin{aligned} r(t) &= \rho x(t - \tau) e^{-i2\pi f_c(\tau + \dot{\tau}t)} \\ &= \rho e^{-i2\pi f_c \tau} x(t - \tau) e^{i2\pi \nu t}. \end{aligned} \quad (2.8)$$

Thus, under the assumption of small fractional bandwidth, the received signal along an individual propagation path subject to specular reflection is a time-delayed and frequency-shifted version of the transmitted signal.

The superposition of partial received signals via multiple propagation paths can be handled as before: We denote the attenuation coefficient of the path with Doppler shift ν and delay τ as $\rho(\nu, \tau)$,

* We choose the sign of the radial velocity to be negative merely for notational convenience.

† The duration here refers to the duration of the entire codeword, not just to the duration of a single transmitted symbol.

and aggregate the coefficients of all propagation paths as $s_S(\nu, \tau) \triangleq \rho(\nu, \tau)e^{-i2\pi f_c \tau}$. This function was termed the *delay-Doppler spreading function* by Bello (1963); we simply refer to it as the *spreading function*. The superposition of all received time- and frequency-shifted copies of the transmitted signal can now be expressed as

$$r(t) = \iint_{\nu \tau} s_S(\nu, \tau)x(t - \tau)e^{i2\pi\nu t} d\tau d\nu. \quad (2.9)$$

B. Delay-Scale or Delay-Doppler Modeling?

Which of the two IO relations (2.4) or (2.9) is more suited to model wireless channels? On the basis of the physical intuition developed so far, the delay-scale IO relation (2.4) is more accurate because it does not rely on the small fractional bandwidth assumption; but it is also mathematically more involved. Most conventional communication systems satisfy the small fractional bandwidth condition (2.7); consequently, analysis and empirical characterization of communication channels that follow the delay-scale IO relation did not receive much research attention in the past, except in the radar and sonar literature (Blahut et al., 1991). The possibility to commercially deploy UWB communication systems has led to novel research activities to characterize and analyze communication channels of large fractional bandwidth. Yet, there are still very few publications that do use the delay-scale channel characterization (2.4) for communication applications (Margetts and Schniter, 2004; Margetts et al., 2005; Jiang and Papandreou-Suppappola, 2006).

Hampering the development might also be a lack of empirical evidence that the more complicated delay-scale model (2.4) is indeed more appropriate than the delay-Doppler model (2.9) to describe wireless channels of large fractional bandwidth. It is not known starting at what fractional bandwidth a system designed on the basis of the delay-scale model might outperform a system designed on the basis of the delay-Doppler model, if at all. An empirical comparison of the delay-scale and the delay-Doppler approach requires time-variant

channel measurements of large enough bandwidth for the theoretical difference to become apparent. Such measurements are not yet technically feasible, as will be discussed in greater detail in Chapter 4.

The assumption of small fractional bandwidth is much more readily violated for underwater acoustic communication channels, as these channels are characterized by a much lower velocity of wave propagation compared with wireless channels (Stojanovic, 1996; Sibul et al., 1994). Yet, it seems that even in underwater acoustic communication, the delay-Doppler channel description (2.9) can be successfully used for system design (Eggen et al., 2000).

To conclude, the case for the delay-scale model (2.4) seems to be rather weak to date. We did not find compelling empirical evidence to suggest that the performance of a wireless system designed on the basis of a delay-Doppler model instead of a delay-scale model will carry a significant performance penalty, so that the mathematical difficulties analyzing (2.4) seem to outweigh its potential benefits. Therefore, we opt to use a delay-Doppler formulation along the lines of (2.9) in the following. This decision is not meant to devalue a delay-scale approach for the analysis of wideband wireless channels, though. On the contrary, it shows that research effort should be increased to provide empirical comparison of the two models and to translate the mathematical tools from harmonic analysis and group representation theory that are needed to analyze the delay-scale model (Auslander and Tolimieri, 1985; Heil and Walnut, 1989; Weiss, 1996; Gröchenig, 2001) into a form that is suitable for engineering analysis of delay-scale channels.

C. Path-Wise Frequency Dependence

We derived the preceding IO relations under the assumption that the only propagation mechanism along each path is a specular reflection. This is a gross simplification in the light of the discussion in Section 2.2. Propagation mechanisms such as diffraction and scattering are frequency dependent in general; furthermore, the type of propagation mechanism that is appropriate to model a certain interaction of a

partial wave with an object might not be the same for all frequencies.

Although the range of propagation mechanisms that may lead to frequency-dependent behavior of individual propagation paths is quite diverse, we make the simplifying assumption that each path acts as a linear time-invariant (LTI) filter whose impulse response $\gamma(\tau')$ captures the frequency dependence of the propagation mechanisms that affect this particular path. The LTI approximation can be motivated in that the frequency dependence of each path does not change significantly over time as long as we only consider small-scale fading. Our modeling approach is similar to the one used by Qiu and Lu (1996) to derive a wideband extension of the classical specular multipath model by Turin (1972). Cramer et al. (2002) provide empirical evidence for the frequency dependence of individual propagation paths.

For the reasons already discussed, we assume that the path delay varies linearly with time as $\tau(t) = \tau + \dot{\tau}t$, and we approximate $x([1 - \dot{\tau}]t - \tau)$ by $x(t - \tau)$. Then, the IO relation of a path with given delay τ and Doppler shift ν can be derived analogously to (2.8) and expressed in complex baseband notation as the convolution

$$r(t) = e^{i2\pi\nu t} e^{-i2\pi f_c \tau} \int_{\tau'} \gamma(\tau') \bar{x}(t - \tau') d\tau'. \quad (2.10)$$

As in the preceding subsection, we index each propagation path by its Doppler-delay pair (ν, τ) , i.e., set $\gamma(\tau') = \gamma(\nu, \tau, \tau')$, and express the received signal as a superposition of a large number of paths:

$$\begin{aligned} r(t) &= \iiint_{\nu \tau \tau'} \gamma(\nu, \tau, \tau') e^{-i2\pi f_c \tau} x(t - \tau - \tau') e^{i2\pi\nu t} d\tau' d\tau d\nu \\ &= \iiint_{\nu \tau \zeta} \gamma(\nu, \tau, \zeta - \tau) e^{-i2\pi f_c \tau} x(t - \zeta) e^{i2\pi\nu t} d\zeta d\tau d\nu, \end{aligned} \quad (2.11)$$

where the second equality follows from the substitution $\zeta \triangleq \tau + \tau'$. Next, we exchange the order of integration, define the *spreading*

function* as

$$s_{\mathbb{H}}(\nu, \zeta) \triangleq \int_{\tau} \gamma(\nu, \tau, \zeta - \tau) e^{-i2\pi f_c \tau} d\tau, \quad (2.12)$$

and perform another change of variables, this time purely cosmetic, to obtain the IO relation

$$r(t) = \iint_{\nu \tau} s_{\mathbb{H}}(\nu, \tau) x(t - \tau) e^{i2\pi \nu t} d\tau d\nu. \quad (2.13)$$

This IO relation has the same structure as (2.9): the channel output is a weighted superposition of time- and frequency-shifted copies of the input signal. The fundamental difference between the two delay-Doppler IO relations (2.9) and (2.13) is the physical interpretation of the delay variable. While in $s_{\mathbb{S}}(\nu, \tau)$ each τ signifies the contribution of a distinct propagation path, multiple paths can contribute to the value of $s_{\mathbb{H}}(\nu, \tau)$ for any given delay.

It is physically sensible to require that both $s_{\mathbb{S}}(\nu, \tau)$ and $s_{\mathbb{H}}(\nu, \tau)$ are energy limited, which means for $s_{\mathbb{H}}(\nu, \tau)$ that

$$\iint_{\nu \tau} |s_{\mathbb{H}}(\nu, \tau)|^2 d\tau d\nu < \infty, \quad (2.14)$$

i.e., the spreading functions need to be square integrable.

2.4. MATHEMATICAL STRUCTURE

At this point in the modeling process, we have reduced to the IO relation (2.13) the complicated initial model that was the solution of Maxwell's equations for a detailed description of all objects in the environment. In most modeling steps so far we put intuition above mathematical rigor; this approach seems well justified because an

* The subscript \mathbb{H} stands for the *channel operator*. We provide a precise definition in 2.4.2.

engineering model that critically depended on some mathematical pathology certainly would not be a very robust one. Yet, the main purpose of a mathematical model is to expose the fundamental *structure* of the physical reality under consideration, as it is this structure that can be exploited for analysis and design. Therefore, we now take a closer look at the IO relation (2.13) and try to discern its structural properties. In particular, we use two equivalent but complementary approaches to this end. The engineering approach is to view the channel as an LTV system that is characterized by its *system functions*, while the mathematical approach is to describe the channel as a linear operator on a Hilbert space.

2.4.1. Linear Time-Variant Systems

The spreading function $s_{\mathbb{H}}(\nu, \tau)$ can be interpreted as a specific system function of an LTV system. Equivalent descriptions of the same LTV system are given by Fourier transforms of $s_{\mathbb{H}}(\nu, \tau)$, as shown by Zadeh (1950, 1961), Kailath (1959), and Bello (1963). The most commonly used system function is the *time-variant impulse response* $h_{\mathbb{H}}(t, \tau)$, which is related to the spreading function by the Fourier transform

$$s_{\mathbb{H}}(\nu, \tau) \triangleq \int_t h_{\mathbb{H}}(t, \tau) e^{-i2\pi\nu t} dt. \quad (2.15)$$

The time-variant impulse response characterizes the response of the channel at time t to a Dirac impulse τ seconds earlier. The corresponding IO relation, which is equivalent to (2.13), is given by:

$$r(t) = \int_{\tau} h_{\mathbb{H}}(t, \tau) x(t - \tau) d\tau. \quad (2.16)$$

Another useful system function is the *time-variant transfer function*

$$l_{\mathbb{H}}(t, f) \triangleq \int_{\tau} h_{\mathbb{H}}(t, \tau) e^{-i2\pi f \tau} d\tau. \quad (2.17)$$

For completeness, we also mention the *bifrequency function*

$$b_{\mathbb{H}}(\nu, f) \triangleq \iint_{\tau t} h_{\mathbb{H}}(t, \tau) e^{-i2\pi(\nu t + f\tau)} dt d\tau, \quad (2.18)$$

although we do not need it in the thesis. The above Fourier relations are compactly summarized in the following diagram:

$$\begin{array}{ccc}
 & h_{\mathbb{H}}(t, \tau) & \\
 \mathbb{F}_{t \rightarrow \nu} \swarrow & & \searrow \mathbb{F}_{\tau \rightarrow f} \\
 s_{\mathbb{H}}(\nu, \tau) & & l_{\mathbb{H}}(t, f) \\
 \searrow \mathbb{F}_{\tau \rightarrow f} & & \swarrow \mathbb{F}_{t \rightarrow \nu} \\
 & b_{\mathbb{H}}(\nu, f) &
 \end{array} \quad (2.19)$$

As the Fourier transform is a unitary operation, the energy limitation on $s_{\mathbb{H}}(\nu, \tau)$ in (2.14) carries over to the other system functions: they are all square integrable.

2.4.2. The Channel as a Linear Operator

The key structural property of the transformation between input signal $x(t)$ and output signal $r(t)$ in (2.13) is *linearity*, as a consequence of the solution to Maxwell's equations being linear as long as all materials are linear. Linearity is extremely important as it allows us to abstract all input and output signals as vectors in a Hilbert space, the so-called *signal space* popularized by Wozencraft and Jacobs (1965) for communications engineering.

Physically realizable signals $s(t)$, expressed as a univariate function of the time parameter $t \in \mathcal{R}$, have finite energy content, i.e.,

$$\int_t |s(t)|^2 dt < \infty. \quad (2.20)$$

Thus, they can be taken as elements of the Hilbert space \mathcal{L}^2 of square-integrable functions over the real line \mathcal{R} (Naylor and Sell, 1982).

We use the term *signal* to refer to square-integrable functions of time $s(t) \in \mathcal{L}^2$ throughout.

The signal space is a powerful model because many important results from functional analysis can be used directly (Naylor and Sell, 1982; Kreyszig, 1989). In particular, we can model the effect of the channel as a linear operator $\mathbb{H} : \mathcal{L}^2 \rightarrow \mathcal{R}_{\mathbb{H}}$ that maps an input signal $x \in \mathcal{L}^2$ into an output signal $r \in \mathcal{R}_{\mathbb{H}}$, where $\mathcal{R}_{\mathbb{H}}$ denotes the range space of \mathbb{H} ; to aid intuition, we call $\mathcal{R}_{\mathbb{H}}$ the channel's *output space* in the following. The set of input signals $x(t)$ that result in a zero output, i.e., for which $\mathbb{H}x = 0$, span the *null-space* $\mathcal{N}_{\mathbb{H}}$ of \mathbb{H} . The signals in $\mathcal{N}_{\mathbb{H}}$ are useless for communication purposes. Conversely, we denote as the channel's *input space* $\mathcal{I}_{\mathbb{H}}$ the linear space of all those signals that do result in a nonzero output, i.e., $\mathcal{I}_{\mathbb{H}} \triangleq \overline{\text{span}}(\mathcal{L}^2 - \mathcal{N}_{\mathbb{H}})$.

Conservation of energy implies that the channel cannot have infinite gain. Thus, \mathbb{H} is a *bounded* linear operator, which means that it is continuous—a physically very reasonable assumption. The corresponding IO relation between a given input signal $x \in \mathcal{L}^2$ and the corresponding noise-free output signal $r \in \mathcal{R}_{\mathbb{H}}$ can now be written as

$$r = \mathbb{H}x. \quad (2.21)$$

The square integrability condition (2.14) on the system functions means that the channel operator is energy limited. From Section 2.3, this condition can be motivated in that the channel is nonzero only over a finite, albeit large, amount of time and over a finite frequency band. In mathematical terms, \mathbb{H} is a *Hilbert-Schmidt (HS)* operator: Let $\{e_i\}$ be an arbitrary orthonormal basis for \mathcal{L}^2 ; then, a HS operator \mathbb{H} is defined as a linear operator that satisfies (Dunford and Schwarz, 1963, Section XI.6)

$$\sum_{i=0}^{\infty} \|\mathbb{H}e_i\|^2 < \infty. \quad (2.22)$$

Mathematically, this assumption is rather strong: it implies that \mathbb{H} can be expressed as the limit* of an operator sequence where each operator

* Convergence is with respect to HS norm.

has a finite-dimensional range space. An important implication is that the IO relation (2.21) can be written in integral form as (Dunford and Schwarz, 1963, p. 1083)

$$r(t) = (\mathbb{H}x)(t) = \int_{t'} k_{\mathbb{H}}(t, t')x(t')dt', \quad (2.23)$$

where $k_{\mathbb{H}}(t, t')$ is called the *kernel* of the HS operator \mathbb{H} . The kernel of every HS operator is square integrable, i.e.,

$$\iint_{t' t} |k_{\mathbb{H}}(t, t')|^2 dt dt' < \infty. \quad (2.24)$$

2.4.3. System Functions

The channel operator \mathbb{H} with kernel $k_{\mathbb{H}}(t, t')$ and the system functions of an LTV system can be formally related through the identity*

$$k_{\mathbb{H}}(t, t') = h_{\mathbb{H}}(t, t - t'), \quad (2.25)$$

which means that every HS operator can be represented as an LTV system and vice versa.

Yet, it is not the formal equivalence of HS operators and LTV systems that is of importance here but the physical interpretation we can now give to the effect of the channel operator \mathbb{H} , on the basis of the developments in Sections 2.2 and 2.3. Two different system functions might correspond to the same HS operator, but the physical

* The identity (2.25) shows that the time-variant impulse response results from a particular coordinate transform between absolute time and delay. Many other transforms are possible, where each one leads to a different set of system functions. A generalized formulation that takes into account all possible ways to define the time-delay coordinate system was devised by Kozek (1992) on the basis of the generalized Wigner distribution (Hlawatsch and Boudreaux-Bartels, 1992). Artés et al. (1998) provide a comprehensive summary of these generalized system functions and their corresponding IO relations. We do not need the generalized system functions in the following.

interpretation of each system function might be completely different. For example, the spreading function $s_{\mathbb{H}}(\nu, \tau)$ and the wideband spreading function $c_{\mathbb{S}}(a, b)$ are mathematically equivalent, as recently shown by Rickard et al. (2005); i.e., they may correspond to the same HS operator \mathbb{H} . Yet, their physical interpretations differ.

Two important special cases of LTV systems are (i) LTI systems, characterized by an *impulse response* $h(\tau)$ that depends only on delay τ but not on absolute time t , so that $h_{\mathbb{H}}(t, \tau) = h(\tau)$, and (ii) linear frequency-invariant (LFI) systems, characterized by a *modulation function* $g(t)$ that depends only on absolute time but not on delay τ , so that $h_{\mathbb{H}}(t, \tau) = g(t)\delta(\tau)$. Both LTI and LFI system cannot be represented as a HS operator because the corresponding kernels are not square integrable. Yet, they can be seen as limiting cases (Naylor and Sell, 1982).

2.4.4. Discretization of the System Functions

We have not yet specified $s_{\mathbb{H}}(\nu, \tau)$, or any of the equivalent system functions. They must all be square integrable, and we have the physical intuition from Sections 2.2 and 2.3 to guide us in the choice of their exact shape. But the specification of a continuous function in two variables is still a formidable task. Especially if we want to use measured data to characterize the channel, we can at best hope to obtain a finite number of parameters. Therefore, it is advantageous to *discretize* the system functions, i.e., to describe them by a countable or, even better, a finite set of parameters.

One particular discretization follows directly from the HS property of \mathbb{H} : the representation of the kernel $k_{\mathbb{H}}(t, t')$ by means of a *singular value decomposition (SVD)*. For a given HS operator \mathbb{H} there exist an orthonormal sequence of signals $\{v_i\}$, called *right singular functions*, that span the channel's input space $\mathcal{I}_{\mathbb{H}}$, and an orthonormal sequence of signals $\{u_i\}$, called *left singular functions*, that span the channel's output space $\mathcal{R}_{\mathbb{H}}$. Furthermore, there exists a sequence of positive real numbers $\{\sigma_i\}$, called *singular values*, such that (Naylor and Sell,

1982, Theorem 6.14.1)

$$k_{\mathbb{H}}(t, t') = \sum_{i=0}^{\infty} \sigma_i u_i(t) v_i(t'). \quad (2.26)$$

Every signal x in the input space $\mathcal{I}_{\mathbb{H}}$ is uniquely characterized by its projection onto the right singular functions,

$$x = \sum_{i=0}^{\infty} \langle x, v_i \rangle v_i \quad (2.27)$$

and every signal r in the output space $\mathcal{R}_{\mathbb{H}}$ can similarly be expressed by its projection onto the left singular functions. Because the singular functions are orthonormal, we can write the projection of the output $r(t)$ in (2.21) onto a left singular function $u_j(t)$ for all $j = 0, 1, \dots$ as

$$\begin{aligned} \langle r, u_j \rangle &= \langle \mathbb{H} x, u_j \rangle \\ &\stackrel{(a)}{=} \sum_{i=0}^{\infty} \langle x, v_i \rangle \langle \mathbb{H} v_i, u_j \rangle \\ &\stackrel{(b)}{=} \sum_{i=0}^{\infty} \langle x, v_i \rangle \left\langle \sum_{n=0}^{\infty} \sigma_n u_n \langle v_n, v_i \rangle, u_j \right\rangle \\ &= \sum_{i=0}^{\infty} \sigma_i \langle x, v_i \rangle \langle u_i, u_j \rangle \\ &= \sigma_j \langle x, v_j \rangle. \end{aligned} \quad (2.28)$$

Here, (a) follows upon insertion of (2.27) and from the linearity of the inner product in its first operand (Kreyszig, 1989), equality (b) results from the integral IO relation (2.23) and the SVD of the kernel in (2.26), and the remaining equalities follow because the singular functions are orthonormal. Hence, the discretization (2.28) on the basis of the SVD (2.26) *diagonalizes* the IO relation (2.23), i.e., it decomposes it into a set of parallel scalar IO relations.

The set of parallel IO relations (2.28) constitutes an elegant and simple description of the channel \mathbb{H} ; it appears to be the perfect representation to analyze communication over LTV channels. Unfortunately,

there are some important practical problems with the diagonalization (2.28) that result because the left and right eigenfunctions of \mathbb{H} depend on the channel operator \mathbb{H} in general.

- As the SVD is a nonredundant representation of \mathbb{H} , the diagonalization (2.28) is in general not robust; i.e., a small change in \mathbb{H} might lead to a large change in the singular functions (Christensen, 2003; Kovačević and Chebira, 2007). Channel measurements, treated in Part II, always contain measurement errors, which might completely change the set of left and right singular functions. Hence, the diagonalized representation (2.28) is not suitable for the description of measured channels.
- The transmitter and receiver in any real world wireless communication system do not know \mathbb{H} . Therefore, they cannot use the right and left singular functions for transmission and reception, respectively.
- Estimation of the u_i and v_i is impossible while the system is operated because at any given point in time knowledge of \mathbb{H} implies anticausal knowledge of its singular functions, which may have infinite duration.

Despite these conceptual difficulties diagonalizing LTV channels by means of the SVD, an approximate diagonalization that is structurally similar to (2.28) is possible if \mathbb{H} is *underspread* (Kozek, 1997a). We discuss this approximate diagonalization in more detail later on, in Section 3.2.

The SVD of the HS kernel $k_{\mathbb{H}}(t, t')$ is a *matched* expansion, matched to the structure of the operator under study. In contrast, *mismatched* discretizations do not cater for the structure of a specific realization of \mathbb{H} , but rely on certain more general properties. The most widely used class of mismatched discretizations are sampling expansions; they are based on the assumption that a specific combination of the duration and the bandwidth of the input and the output signal is limited. Possible assumptions are an input band limitation in combination with an output time limitation, or an input time limitation in combination with an output band limitation. A joint limitation of the duration

and the bandwidth of either the input signal or the output signal is mathematically not possible; but these mathematical difficulties are not physically fundamental and can be dealt with (Slepian, 1976). The various types of discretized IO relations that can be obtained by sampling the input or output signals (or both), are described in detail by Kailath (1959), Bello (1963), and Artés et al. (1998). The resulting models are used, among others, by Médard (2000) and Goldsmith and Médard (2007).

Band limitations can also be placed directly on the system functions instead of on the input and output signals. Physically the most sensible assumption seems to be a joint limitation of the Doppler shift ν and the delay τ . As discussed in Section 2.3.2, the maximum Doppler shift is determined by the velocity of the transmitter, the receiver, and objects in the environment, all of which can readily be assumed finite. The maximum delay is determined by the time it takes for the transmitted energy of a single pulse to be absorbed by materials. Thus, we can limit the essential support of the spreading function $s_{\mathbb{H}}(\nu, \tau)$ to lie in the rectangle $[-\nu_0/2, \nu_0/2] \times [-\tau_0/2, \tau_0/2]$, where, without loss of generality, we removed an overall Doppler shift and introduced a negative delay to render the support rectangle symmetric around the origin. Because the support of $s_{\mathbb{H}}(\nu, \tau)$ is limited as above, the time-variant transfer function $l_{\mathbb{H}}(t, f)$ can be sampled in time t at rate ν_0 and in frequency f at rate τ_0 . The corresponding IO relation was derived by Kailath (1959, Section 3.13) and Bello (1963, Section IV.A.2).

2.5. STOCHASTIC MODELS

Most indoor wireless systems are commodity products, standardized and manufactured in large quantities for operation in a multitude of different environments. It is virtually impossible to characterize spreading functions for even a small subset of the different channels these systems might operate over, let alone to design a system for each specific spreading function. A pragmatic solution is to explicitly take

into account the ignorance about the channel a specific deployment of a wireless system will face; hence, a stochastic channel model describes our ignorance about the actual propagation environment a system will operate in, so that a design for the average case can be robust enough to operate under a wide range of conditions.

2.5.1. Structure of a Stochastic Model

In Section 2.2.2 we distinguished between the variation of the mean received signal strength and the fluctuation of the magnitude and phase of the received signal. This distinction between large-scale fading and small-scale fading is useful also to build stochastic models of wireless channels.

Large-scale variations in general happen over times that are much longer than the duration of a typical transmitted codeword. Therefore, the main purpose of a stochastic model for large-scale variations is to aid the design of the power and rate control algorithms, feedback schemes, admission control, and other functions in the upper layers of the standard network stack (Bertsekas and Gallager, 1996; Tanenbaum, 2003).

Small scale variations of the channel, on the other hand, may happen over the course of a couple of transmitted symbols during the transmission of a single codeword. Therefore, small-scale fading often cannot be easily mitigated by closed-loop control schemes, and is often best modeled as an integral part of the wireless channel. A stochastic description of small-scale fading is an approach well suited to model many types of fading channels in the sense that it correctly predicts the significant difference between the rates achievable over a fading channel in comparison with the capacity of a channel that is only affected by AWGN. From a mathematical perspective, a stochastic model for small-scale fading is appropriate as the models commonly used for both information sources and receiver noise are stochastic in nature. As already pointed out, we are mainly concerned with the mathematical communication problem in this thesis; therefore, we will exclusively use a stochastic small-scale fading model and exclude

from consideration all large-scale effects.

Formally, the change from the deterministic IO relation (2.13) to its stochastic counterpart amounts to replacing the deterministic spreading function $s_{\mathbb{H}}(\nu, \tau)$ with a random process $\mathbf{s}_{\mathbb{H}}(\nu, \tau)$. To complete the model, we also include the additive noise $\mathbf{w}(t)$ introduced by the receiver. At the frequencies of interest, this noise is mostly thermal in nature; therefore, it can be very well modeled as a white JPG process. For convenience, we normalize its power spectral density (PSD) to unity, i.e., $\mathbb{E}[\mathbf{w}(t)\overline{\mathbf{w}}(t')] = \delta(t - t')$. Furthermore, we incorporate into our model that information-bearing signals are best described as random processes as well. Hence, a stochastic model that relates the random output signal $\mathbf{y}(t)$ of a small-scale fading wireless channel \mathbb{H} with additive noise $\mathbf{w}(t)$ to the random input signal $\mathbf{x}(t)$ is

$$\mathbf{y}(t) = \iint_{\nu \tau} \mathbf{s}_{\mathbb{H}}(\nu, \tau) \mathbf{x}(t - \tau) e^{i2\pi\nu t} d\tau d\nu + \mathbf{w}(t). \quad (2.29)$$

Barring certain technical conditions on the nature of the stochastic process $\mathbf{s}_{\mathbb{H}}(\nu, \tau)$, all other system functions can be defined as in Section 2.4.1 by means of the appropriate Fourier transforms:

$$\begin{array}{ccc} & \mathbf{h}_{\mathbb{H}}(t, \tau) & \\ \mathbb{F}_{t \rightarrow \nu} \swarrow & & \searrow \mathbb{F}_{\tau \rightarrow f} \\ \mathbf{s}_{\mathbb{H}}(\nu, \tau) & \xrightarrow{\mathbb{F}_{\nu \rightarrow t}^{-1} \mathbb{F}_{\tau \rightarrow f}} & \mathbf{l}_{\mathbb{H}}(t, f) \end{array} \quad (2.30)$$

Hence, the IO relation on the basis of the random time-variant impulse response $\mathbf{h}_{\mathbb{H}}(t, \tau)$ is

$$\mathbf{y}(t) = \int_{\tau} \mathbf{h}_{\mathbb{H}}(t, \tau) \mathbf{x}(t - \tau) d\tau + \mathbf{w}(t), \quad (2.31)$$

and the kernel representation follows as before from a coordinate transform $\mathbf{k}_{\mathbb{H}}(t, t') = \mathbf{h}_{\mathbb{H}}(t, t - t')$. Consequently, the channel operator \mathbb{H} is now an operator-valued random variable.

A complete stochastic description of the channel requires specification of the distribution followed by the two-dimensional *channel process* $\mathbf{k}_{\mathbb{H}}(t, t')$, or equivalently by any of the stochastic processes $\mathbf{h}_{\mathbb{H}}(t, \tau)$, $\mathbf{l}_{\mathbb{H}}(t, f)$, or $\mathbf{s}_{\mathbb{H}}(\nu, \tau)$. Adapting terminology from Linhart and Zucchini (1986), we call this complete stochastic description the *operating distribution*—the closest possible stochastic representation of a given physical situation. The operating distribution is a conceptual modeling assumption, there is in general no “true” stochastic process according to which channel realizations are generated. As in the deterministic modeling steps in the preceding sections, we need to balance tractability with accuracy and approximate the operating distribution by a stochastic description with reasonable complexity. We also need to keep in mind that the transition between a deterministic, site-specific description and a generic, stochastic description of wireless channels is gradual—the more detailed the stochastic model, the more do we tie the model to a specific environment that fits the selected set of parameters. This is a manifestation of the trade-off between accuracy and generality discussed in Section 2.1.

A suitable compromise between accuracy, generality, and mathematical and experimental tractability is often to describe the process under study only up to second order, i.e., to specify its marginal distribution and its correlation function. We adopt this approach here and discuss marginal distributions and second-order characterization in the following two subsections.

2.5.2. Marginal Distributions

Many statistical models rely on the central limit theorem (CLT): if a quantity to be modeled stochastically results from the superposition of many unknown physical effects, chances are that the quantity can be very well modeled as a Gaussian random variable. However, if only a small number of unknown effects contribute to the quantity under study, this Gaussian approximation may not be a good one.

In a baseband equivalent channel description, it is common to model both the real and the imaginary part of the spreading func-

tion $\mathbf{s}_{\mathbb{H}}(\nu, \tau)$ as Gaussian for each (ν, τ) (Bello, 1963; Proakis, 2001; Tse and Viswanath, 2005). If the two parts are i.i.d., the complex random variable $\mathbf{s}_{\mathbb{H}}(\nu, \tau)$ is *proper Gaussian* distributed (Wooding, 1956; Miller, 1969; Neeser and Massey, 1993). A proper Gaussian random variable of zero mean has a uniformly distributed phase and its magnitude follows a Rayleigh distribution. If the mean is not zero, the magnitude of the corresponding random variable follows a Rice distribution (Rice, 1944).

Though, when we recapitulate the derivation of the nonrandom IO relations in Section 2.3 it seems questionable if the CLT can indeed be invoked to justify this Gaussian assumption. Only a single propagation path contributes to the value of the specular spreading function $\mathbf{s}_{\mathbb{S}}(\nu, \tau)$ for each (ν, τ) , so that in this particular case the CLT is certainly not applicable. It might be applicable in the more realistic propagation scenario with frequency dependence of individual propagation paths because of the aggregation of contributions from different paths for each fixed delay value according to (2.12). An even stronger argument in favor of the CLT and the resulting Gaussian marginal distribution of $\mathbf{s}_{\mathbb{H}}(\nu, \tau)$ is an indirect one: All communication systems use effectively band-limited signals, so that the effective channel they see is band limited as well. A band limitation reduces the temporal resolution. Therefore, propagation paths whose delay difference is less than the inverse bandwidth cannot be resolved at the receiver; consequently, only the aggregate effect of these paths needs to be modeled (Tse and Viswanath, 2005). And as a small change in delay can lead to a large change in the phase term in (2.8), a uniform distribution for the aggregated phase of these paths seems sensible as well. From this point of view, to model the not band-limited spreading function $\mathbf{s}_{\mathbb{H}}(\nu, \tau)$ as proper Gaussian is similar to the engineering model of white Gaussian noise—both models are physically questionable and mathematically ill defined, but a band-limited version of the noise, or in our case the channel’s spreading function, is often a sensible model.

But if $\mathbf{s}_{\mathbb{H}}(\nu, \tau)$ has operational significance only if band limited with respect to τ , it is important to ask how large the bandwidth of a communication system can be for the proper Gaussian model

of $\mathbf{s}_{\mathbb{H}}(\nu, \tau)$ to remain valid. For if the bandwidth is large and the corresponding temporal resolution is high, there might not be enough paths averaging to justify the Gaussian assumption via the CLT. Several researchers have questioned the validity of the Gaussian model on these grounds, especially for UWB channels with several gigahertz of bandwidth (Zhang et al., 2002; Cassioli et al., 2002; Molisch et al., 2003a). As an alternative for the Rayleigh or Rice amplitude distributions that correspond to a proper Gaussian random variable with zero or nonzero mean, respectively, Cassioli et al. (2002) proposed to use the Nakagami distribution (Nakagami, 1960), Foerster (2003), Buehrer et al. (2003), and Li and Wong (2003) advocate the lognormal distribution, and Chong and Yong (2005) use the Weibull distribution to model small-scale amplitude variations in UWB channels. All of these different amplitude distributions are proposed on the basis of UWB channel measurements. However, the measurement methodology and the statistical procedure used to assess the suitability of different distributions differs widely among the cited references, so that it is not clear if the variety of proposed amplitude distributions results from the different measured environments or if it is an artifact of the different statistical methods used. We take up this question in Part II, where we systematically compare the suitability of the above-listed distributions to model amplitude variations of measured UWB channels.

2.5.3. Second-Order Characterization and the WSSUS Assumption

A stochastic process is completely specified by its joint distribution at any finite number of epochs. Because the complexity of such a description is prohibitive, we restrict our attention to the moments of second order, i.e., we partially characterize the joint distribution of the random system functions in that we specify the corresponding correlation functions, following Bello (1963). The natural extension of the proper Gaussian marginal distribution to the entire stochastic processes is a jointly proper Gaussian distribution for any finite num-

ber of epochs (Neuser and Massey, 1993). Bello (1963) and Kennedy (1969) model $\mathbf{s}_{\mathbb{H}}(\nu, \tau)$ as a JPG process in both ν and τ . This implies that all other system functions as well as the kernel $\mathbf{k}_{\mathbb{H}}(t, t')$ are also two-dimensional JPG processes. For JPG processes, a second-order characterization is indeed a complete specification of the corresponding stochastic processes. Without loss of generality, we assume that all system functions are of zero mean. This assumption is not restrictive, as a mean component can always be added.

As the system functions are two-dimensional, their correlation functions depend on four parameters in general.

- For the time-variant impulse response $\mathbf{h}_{\mathbb{H}}(t, \tau)$:

$$r_{\mathbf{h}}(t, t', \tau, \tau') \triangleq \mathbb{E}[\mathbf{h}_{\mathbb{H}}(t, \tau)\bar{\mathbf{h}}_{\mathbb{H}}(t', \tau')]. \quad (2.32)$$

- For the spreading function $\mathbf{s}_{\mathbb{H}}(\nu, \tau)$:

$$r_{\mathbf{s}}(\nu, \nu', \tau, \tau') \triangleq \mathbb{E}[\mathbf{s}_{\mathbb{H}}(\nu, \tau)\bar{\mathbf{s}}_{\mathbb{H}}(\nu', \tau')]. \quad (2.33)$$

- For the time-variant transfer function $\mathbf{l}_{\mathbb{H}}(t, f)$:

$$r_{\mathbf{l}}(t, t', f, f') \triangleq \mathbb{E}[\mathbf{l}_{\mathbb{H}}(t, f)\bar{\mathbf{l}}_{\mathbb{H}}(t', f')]. \quad (2.34)$$

These four-dimensional correlation functions are still quite complex to work with. Therefore, Bello (1963) introduced two additional assumptions to simplify the correlation structure. Expressed in terms of the spreading function, these assumptions state that the channel is uncorrelated for different Doppler shifts and for different delays, i.e.,

$$\mathbb{E}[\mathbf{s}_{\mathbb{H}}(\nu, \tau)\bar{\mathbf{s}}_{\mathbb{H}}(\nu', \tau')] = c_{\mathbf{s}}(\nu, \tau)\delta(\nu - \nu')\delta(\tau - \tau'), \quad (2.35)$$

where $c_{\mathbf{s}}(\nu, \tau)$ is called the channel's *scattering function*. Expressed in terms of the time-variant transfer function, a symplectic Fourier transform shows that

$$\mathbb{E}[\mathbf{l}_{\mathbb{H}}(t, f)\bar{\mathbf{l}}_{\mathbb{H}}(t', f')] = c_{\mathbf{l}}(\underbrace{t - t'}_{\triangleq \Delta t}, \underbrace{f - f'}_{\triangleq \Delta f}). \quad (2.36)$$

Consequently, the spreading function being uncorrelated in both Doppler and delay implies that the time-variant transfer function is WSS in both time and frequency. Furthermore, the correlation function of the time-variant impulse response $\mathbf{h}_{\mathbb{H}}(t, \tau)$ is WSS in time and uncorrelated in delay:

$$\mathbb{E}[\mathbf{h}_{\mathbb{H}}(t, \tau)\overline{\mathbf{h}_{\mathbb{H}}(t', \tau')}] = c_{\mathbf{h}}(\Delta t, \tau)\delta(\tau - \tau'). \quad (2.37)$$

It is this last representation from which the assumption derives its name: the channel is assumed to be WSS in time and scatterers that correspond to paths of different delays are assumed to be uncorrelated; thus, the channel is said to be WSSUS. The Fourier transform relation between the three simplified correlation functions is summarized in the following diagram.

$$\begin{array}{ccc}
 & c_{\mathbf{h}}(\Delta t, \tau) & \\
 \mathbb{F}_{\tau \rightarrow \Delta f} \swarrow & & \searrow \mathbb{F}_{\Delta t \rightarrow \nu} \\
 c_{\mathbf{t}}(\Delta t, \Delta f) & \xrightarrow{\mathbb{F}_{\Delta t \rightarrow \nu}^{-1} \mathbb{F}_{\Delta f \rightarrow \tau}} & c_{\mathbf{s}}(\nu, \tau)
 \end{array} \quad (2.38)$$

As our focus is on small-scale fading, we can take care of the overall channel attenuation by appropriate scaling of the transmitted signal power; hence, we normalize the channel attenuation to unity throughout,

$$\iint_{\nu \tau} c_{\mathbf{s}}(\nu, \tau) d\tau d\nu = 1. \quad (2.39)$$

By Parseval's theorem, the other correlation functions are normalized accordingly. For convenience, we define a related quantity that is often used to characterize wireless channels: the power-delay profile (PDP) is

$$p_{\mathbf{s}}(\tau) \triangleq \int_{\nu} c_{\mathbf{s}}(\nu, \tau) d\nu. \quad (2.40)$$

That the channel evolves in a stationary way in time is a sensible assumption only with respect to the small-scale fading effects we consider here. Changes in path loss, for example, are certainly not stationary; thus, if we violate the small-scale fading assumption, we also violate the WSS assumption in t and vice versa.

There are two main problems with the physical motivation of the US part in the WSSUS assumption for the types of channel under consideration here. (i) The stochastic system functions have no operational relevance if not properly band limited in f , as discussed in the previous subsection. But the resulting function would no longer be WSS in frequency, i.e., its time-domain dual would no longer be uncorrelated in delay. Thus, the US assumption is an assumption on the physically meaningless infinite-bandwidth stochastic system functions. (ii) For channels with per-path frequency dependence, different delays do not necessarily correspond to different physical propagation paths; it follows from (2.12) that several paths can contribute to the channel output at the same delay. Hence, the motivation that US results because different uncorrelated physical scatterers contribute to the channel output only at different and distinct delays does not hold true. However, it is not clear a priori how strong the correlation between different paths is. Therefore, we analyze the correlation of measured channel impulse responses in Part II.

Finally, we need to point out a fundamental mathematical incompatibility between the WSSUS assumption and the channel description as a random HS operator. The HS property implies square integrability of each realization of the kernel $\mathbf{k}_{\mathbb{H}}(t, t')$, which physically means that the channel gain vanishes with increasing time and frequency. But the WSSUS assumption necessitates that the channel correlation functions depend only on the time and frequency difference but not on absolute time and frequency. Serious as this incompatibility may sound, it is not relevant for engineering applications, for which the model needs to hold only for a finite time duration and bandwidth. Physically, the WSSUS assumption only makes sense over a limited time and bandwidth, so that outside of this range we can safely assume the channel to decay so as to satisfy the square-integrability condition.

Hence, all we need is a quasi-WSSUS description, valid over long but finite time duration and bandwidth (Bello, 1963).

The WSSUS correlation functions $c_s(\nu, \tau)$, $c_h(\Delta t, \tau)$, and $c_l(\Delta t, \Delta f)$ describe the variation of the channel over time and over frequency; therefore, they are at the core of the stochastic model, and their specification is a delicate task. The two main approaches how to obtain correlation functions are by measurements and from theoretical considerations. A reliable estimate of, say, $c_s(\nu, \tau)$ requires a great many independent measurements to average over; this is challenging but can be done for channels of small bandwidth, as demonstrated by Cox (1972, 1973b,a), Bajwa and Parsons (1978), and others. Theoretical derivations of the correlation functions often start from a simplified geometry of the environment and some basic assumptions on the dynamics of the terminals and on the individual scattering parameters; these assumptions are then used in a multipath model along the lines of the derivation in Section 2.3. The most prominent such model is probably the one-ring scatterer model by Clarke (1968), with its modified formulation in terms of the fading process's PSD by Gans (1972) and Jakes (1974), the generalization of said model to three spatial dimensions by Aulin (1979), and a modification to include frequency-dependent behavior by Lauritzen and Bach Andersen (1990). A more generic geometry, namely a spatial Poisson field of scatterers is considered by Sadowsky and Kafedziski (1998). The advantage of a second-order description obtained from theoretical considerations is that the resulting correlation functions only depend on a small number of parameters, like the mobile velocity or the scatterer density. However, even such extremely simplified models might be too cumbersome to work with. Therefore, even simpler models are often used in theoretical analyses where mainly the qualitative behavior is of importance. One such simplification, discussed in the following chapter, is the well-known block-fading model, probably the simplest channel model to still yield nontrivial and insightful results.

CHAPTER 3

Discretized Models

IN CHAPTER 2, we set out to simplify the description of a wireless channel from a solution of Maxwell's equations to a description that is mathematically tractable, general enough to cover many different scenarios, but yet accurate in the sense that it captures the fundamental effects of the channel. The canonical stochastic IO relation (2.29) together with its second-order stochastic characterization satisfies these requirements. Yet, there is still one open issue we need to address before we can empirically characterize wideband channels in Part II and analyze their capacity in Part III: the stochastic system functions, as specified in Section 2.5, depend on an infinite number of parameters, and so does their stochastic characterization.

We already mentioned in Section 2.4.4 that a discretization of the channel's system functions is necessary if we are to characterize them by channel measurements. But an appropriate discretization is equally important for theoretical analysis of communication systems. A key abstraction in information theory is a *channel use*; virtually all coding theorems are stated for a countable set of channel uses (Cover and Thomas, 1991). The counterpart to one channel use in the continuous-time physical world is one dimension in signal space, one DOF. Although real-world signals cannot be strictly limited in bandwidth and in time simultaneously, Slepian and Pollak (1961) and Landau and Pollak (1961, 1962) showed that the space of signals with essential duration D and essential bandwidth B is of dimension

$2DB + 1$, i.e., there are $(2DB + 1)/D$ *real* DOFs per second available for communication.

In this chapter, we introduce several discretized models that are described by a countably infinite or even finite number of parameters. Some of these models are concisely described in standard textbooks and widely used, so that we only comment briefly on them, while another model, the matched discretization of underspread WSSUS channels on which we base our information theoretic analysis in Part III, is not yet as pervasively used, and the description of this model is distributed over several publications (Kozek, 1997b; Kozek and Molisch, 1997; Kozek, 1997a; Matz and Hlawatsch, 1998; Matz, 2000; Bölcskei et al., 2002b; Matz and Hlawatsch, 2003a,b; Matz et al., 2007). Therefore, we consolidate results from the mentioned sources and give a, hopefully complete, description of this channel model.

3.1. BLOCK FADING MODELS

A simple model for the time variation of a wireless channel is to take the channel as constant over a given time interval, also called a *block*, and as independently changing from one such block to the next. This model is arguably the simplest possible that still captures the essential features of wireless channels—namely, channel variability and randomness—and that at the same time leads to fundamental observations like the role of diversity and the dependence of capacity on the channel’s DOFs. Several variations of this model are widely used for analysis and design of wireless communication systems (Hashemi, 1993b; Biglieri et al., 1998; Tse and Viswanath, 2005).

3.1.1. Modeling Assumptions

Channel variation over time is caused by the relative motion of transmitter, receiver, and objects in the environment. As discussed in Chapter 2, a measure for how fast the channel changes over time is the *maximum Doppler shift* $\nu_0/2$. It depends on the relative speed of the

terminals and objects in the environment and on the carrier frequency. Typical Doppler shifts are rather small, in the range from a few hertz up to about 100 Hz; i.e., many channels change very slowly compared with the duration of one transmitted symbol. Therefore, the first simplifying assumption of the block-fading model is to approximate the channel as constant over a block of duration T_c , called the *coherence time* of the channel. Consequently, during one such block the channel can be modeled as a random LTI system $\mathbf{h}_{\mathbb{H}}(\tau)$ instead of as a random LTV system $\mathbf{h}_{\mathbb{H}}(t, \tau)$. The coherence time is often related to ν_0 as $T_c = \alpha/\nu_0$, where α varies between 1 and 50 among different authors. There is no agreement about the best value for α (Clarke, 1968; Lee, 1993; Rappaport, 2002; Tse and Viswanath, 2005).

The second assumption of the block-fading model concerns channel variation across blocks: the individual blocks are modeled as i.i.d.. This assumption is more difficult to justify than modeling the channel to be constant over the duration of a block; why should the channel change to a completely independent realization from one block to the next although it was assumed to not vary at all during a single block? Indeed, if the physical channel consists of a contiguous frequency band that is permanently used to transmit data from a single transmitter to a single receiver, the independence assumption is quite weak. But if we expand the notion of a channel model to include system aspects, the resulting *effective channel* might very well be modeled as independent across blocks. For example:

- If several users access the same channel by means of time division multiple access (TDMA), it might very well be that transmissions of a single user are separated far enough in time for the assumption of independent blocks to become viable.
- Successive point-to-point transmissions over changing frequency bands that are widely separated, e.g., in a frequency hopping (FH) system, can often be modeled as uses of independently fading blocks.
- If the data stream to be transmitted is heavily interleaved, or if the receiver only processes one block at a time, channel memory

between successive blocks is effectively useless, and there is no need to model this memory in the first place. Consequently, blocks can be treated as independent.

Thus, block-fading in time is difficult to justify for physical channels, but might be a good model for certain effective channels. We summarize the main features of three widely used block-fading models in the following.

3.1.2. The Specular Block-Fading Model

Analysis of electromagnetic wave propagation in terms of the propagation effects listed in Section 2.2 already entails a significant simplification of the physical reality. An even more basic description of radio wave propagation is to consider *only* specular reflections. If the reflection coefficients ρ_l and path delays τ_l for all significant specular paths $l = 0, 1, \dots, L-1$ are modeled as random, constant over blocks of length T_c and independently changing from block to block, the corresponding random channel impulse response can be expressed as

$$\mathbf{h}_{\mathbb{H}}(\tau) \triangleq \sum_{l=0}^{L-1} \rho_l \delta(\tau - \tau_l). \quad (3.1)$$

This type of block-fading model is often attributed to Turin (1972), who performed an extensive measurement campaign to characterize the probabilistic structure of the ρ_l and τ_l .

An apparent problem of the impulse response (3.1) is its infinite bandwidth. As it stands, this model is physically impossible and cannot be characterized from measurements. However, in combination with a transmitted signal $\mathbf{x}(t)$ of finite bandwidth, the resulting IO relation

$$\mathbf{y}(t) = \sum_{l=0}^{L-1} \rho_l \mathbf{x}(t - \tau_l) + \mathbf{w}(t) \quad (3.2)$$

is physically sensible. Similarly, the statistical characterization of the amplitude parameters $\{\rho_l\}$ is only viable for a lowpass-filtered

version of the impulse response (3.1), because only then can each ρ_l be interpreted as resulting from the superposition of many partial waves arriving in a given time interval. Unfortunately, this statistical characterization contradicts the original physical motivation that each Dirac impulse in (3.1) describes the contribution of a single propagation path.

Several models exist to describe the *excess delays* τ_l , i.e., the path arrival times measured with respect to the first path. The simplest model is to take the τ_l as nonrandom integer multiples of a basic delay quantum, most often the inverse bandwidth of the system under study. Although seemingly popular, there does not appear to be an advantage of such a description over a discrete-time block-fading model described below in Section 3.1.4. Stochastic descriptions for the delays are a Poisson process or a modified Poisson process (Turin, 1972). As virtually all wideband channel measurements indicate that the impulse response power decays with increasing delay, the distribution of the channel gains ρ_l need to depend on the corresponding delay value τ_l . Consequently, a suitable joint description of the channel gains and delays is by means of a marked Poisson process (Kingman, 1993).

3.1.3. The Model of Saleh and Valenzuela

A result of virtually all wideband measurement campaigns of sufficient time resolution is that individual impulse responses show a clustered structure, as illustrated in Figure 3.1. The clusters can often be attributed to specular reflections of large smooth objects in the environment. Saleh and Valenzuela (1987) proposed a model that describes the clustering phenomenon. It is an extension of Turin's specular continuous-time block-fading model (3.1) with impulse response

$$\mathbf{h}_{\mathbb{H}}(\tau) = \sum_{l=0}^{L-1} \sum_{i=0}^{N_l-1} \rho_{li} \delta(\tau - \tau_l - \kappa_{li}) \quad (3.3)$$

3 DISCRETIZED MODELS

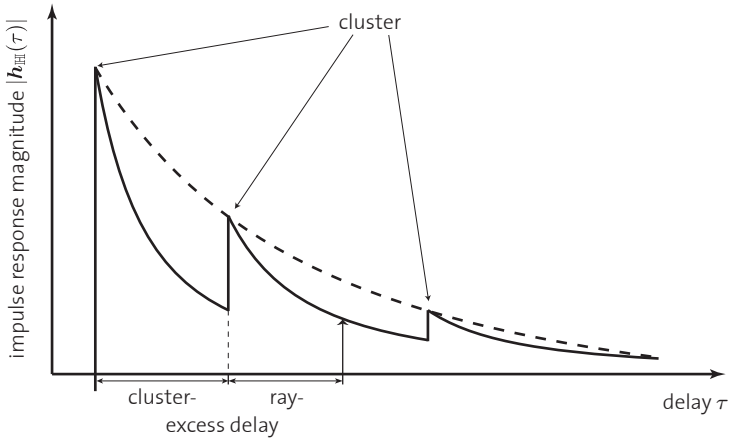


Fig. 3.1: Shape of the impulse response magnitude according to the SV model.

that consists of L clusters characterized by their *cluster excess delay* τ_l ; the l th cluster in turn is made up of N_l rays, each of which models a specular propagation path that arrives at a *ray excess delay* κ_{li} , measured relative to the corresponding cluster start. Cluster and ray excess delays are modeled as arrival process. The path coefficients ρ_{li} are assumed to follow a common parameterized distribution with parameters that depend on the corresponding cluster and ray arrival times. As a direct extension of the Turin model, the SV model is physically sensible only if appropriately band limited.

3.1.4. The Standard Discrete-Time Block-Fading Model

As discussed in Section 2.5.2, the random channel description only has an operational meaning when appropriately band limited. Thus, we assume that the communication system we wish to analyze uses signals that are strictly limited to a bandwidth of B hertz. Without loss of generality, we can now assume that $\mathbf{h}_{\text{HH}}(\tau)$ is band limited to B as well, so that we can discretize time at rate B . The resulting discrete-time

values* $\mathbf{h}[l] \triangleq \mathbf{h}_{\mathbb{H}}(l/B)$ of the impulse response $\mathbf{h}_{\mathbb{H}}(\tau)$ are commonly called *channel taps* because they can be interpreted as the values of a tapped delay line filter. We similarly discretize the input signal $\mathbf{x}(t)$, the output signal $\mathbf{y}(t)$, and the additive noise process $\mathbf{w}(t)$ to obtain the corresponding discrete-time input sequence $\{\mathbf{x}[k]\}$, output sequence $\{\mathbf{y}[k]\}$, and noise sequence $\{\mathbf{w}[k]\}$. Then, the discretized IO relation for all k is

$$\mathbf{y}[k] = \sum_{l=-\infty}^{\infty} \mathbf{h}[l]\mathbf{x}[k-l] + \mathbf{w}[k], \quad (3.4)$$

where the discretized noise process is white and of unit variance, i.e., $\mathbb{E}[\mathbf{w}[k]\overline{\mathbf{w}}[k']] = \delta[k-k']$.

The sequence of channel taps $\{\mathbf{h}[l]\}$ is of infinite extent in general; yet, it is convenient to truncate it to, say, L taps. If $L > \lceil \tau_0 B \rceil$, the approximation error will be small. We assemble the L taps in a random vector,

$$\mathbf{h} \triangleq [\mathbf{h}[0] \ \mathbf{h}[1] \ \dots \ \mathbf{h}[L-1]]^T \quad (3.5)$$

and define its covariance matrix

$$\mathbf{R}_{\mathbf{h}} \triangleq \mathbb{E}[\mathbf{h}\mathbf{h}^\dagger]. \quad (3.6)$$

Similarly to the continuous-time setting in (2.39), we normalize the channel attenuation according to

$$\text{tr } \mathbf{R}_{\mathbf{h}} = 1. \quad (3.7)$$

It is common in many theoretical studies of wireless channels to assume that the entries $\mathbf{h}[l]$ of the channel vector \mathbf{h} are uncorrelated so that $\mathbf{R}_{\mathbf{h}}$ is a diagonal matrix (Tse and Viswanath, 2005). We call this

* A word on notation: we use the letter \mathbf{h} in several flavors—with and without indices, with different types of arguments, and as a vector \mathbf{h} . The channel model to which it pertains is different in different parts of the thesis, i.e., we do not use a unique notation for channel quantities of different channel models, but hope that the model used is always clear from the context.

assumption the *discrete-time US hypothesis*. Though mathematically convenient, it is difficult to justify. Even if we model the infinite-bandwidth continuous-time impulse response $\mathbf{h}_{\mathbb{H}}(\tau)$ as US, the band-limited version $\mathbf{h}_{\mathbb{H}}(\tau)$ is correlated across τ and the taps $\mathbf{h}[l]$ are correlated as well. Whether this correlation is indeed significant is difficult to assess on theoretical grounds; we try to answer this question empirically in Part II.

3.2. A MODEL FOR UNDERSPREAD CHANNELS

The strength of the discrete-time block-fading model is its analytical simplicity. However, because long-term memory in the channel is neglected and short-term memory is modeled to be perfect, analyses of the long-term channel behavior might lead to overly optimistic results (Lapidoth, 2003). Therefore, we now present a more general discretized channel model, which enables us to model continuous channel variation in time and frequency.

3.2.1. Underspread Channels

Because the velocity of the transmitter, of the receiver, and of the objects in the environment is limited, so is the maximum Doppler shift $\nu_0/2$ experienced by the transmitted signal. We also assume that the maximum delay is strictly smaller than τ_0 . For simplicity and without loss of generality, we consider scattering functions that are centered at $\tau = 0$ and $\nu = 0$, i.e., we remove any overall fixed delay and Doppler shift. The assumptions of limited Doppler shift and delay then imply that the scattering function is supported on a rectangle of *spread* $\Delta_{\mathbb{H}} \triangleq \nu_0\tau_0$,

$$c_{\mathbf{s}}(\nu, \tau) = 0 \quad \text{for } (\nu, \tau) \notin [-\nu_0/2, \nu_0/2] \times [-\tau_0/2, \tau_0/2]. \quad (3.8)$$

This support condition in turn implies that the spreading function $\mathbf{s}_{\mathbb{H}}(\nu, \tau)$ is also supported on the same rectangle w.p.1. If $\Delta_{\mathbb{H}} < 1$, the channel is said to be *underspread* (Bello, 1963; Kennedy, 1969;

Kozek, 1997a). Virtually all wireless communication channels are highly underspread, with $\Delta_{\mathbb{H}} \approx 10^{-3}$ for typical land-mobile channels and as low as $\Delta_{\mathbb{H}} \approx 10^{-7}$ for some indoor channels with restricted mobility of the terminals (Hashemi, 1993b; Parsons, 2000; Rappaport, 2002).

The underspread property of typical wireless channels is very important for two reasons: (i) A realization of a random LTV channel can be identified if and only if it is underspread. More precisely, only for underspread channels is it possible to find an input signal that allows to perfectly reconstruct the channel realization from a noise-free observation of the channel's output signal. This identifiability property was discovered by Kailath (1963) for deterministic underspread channels with rectangular support of their spreading function, and later extended to channels with arbitrarily shaped support of the spreading function by Bello (1969) and Pfander and Walnut (2006). Identifiability in the latter case is guaranteed if the area of the support is less than one.* (ii) Underspread operators \mathbb{H} have much more structure than general HS operators; in particular, they have a well-structured set of approximate eigenfunctions that do not depend on the particular realization of \mathbb{H} (Kozek, 1997a). We use these approximate eigenfunctions to discretize and approximately diagonalize underspread channels in a way that circumvents the problems discussed in Section 2.4.4.

3.2.2. Approximate Diagonalization of Underspread Channels

The underspread property is of central importance, as it allows to construct a discretization of the kernel of any underspread operator \mathbb{H} in a way similar to the SVD in (2.26), but by means of a generic set of functions instead of particular singular functions that depend on the exact realization of \mathbb{H} (Kozek, 1997a). These generic functions used in the discretization of underspread operators need to be well localized in time and frequency, as measured by their time-bandwidth product.

* The area of the support set is measured by its Jordan content (Pfander and Walnut, 2006).

A. Time- and Frequency Localization

Let the Fourier transform of the unit-norm signal $g(t)$ be denoted as $G(f) \triangleq \mathbb{F}_{t \rightarrow f} g(t)$.

Definition 3.1. The time location of a unit-norm signal $g(t)$ is defined as

$$\check{t}_g \triangleq \int_t t |g(t)|^2 dt. \quad (3.9)$$

Similarly, the frequency location is

$$\check{f}_g \triangleq \int_f f |G(f)|^2 df. \quad (3.10)$$

Definition 3.2. The *effective duration* of a unit-norm signal $g(t)$ is defined as

$$d_g \triangleq \left(4\pi \int_t (t - \check{t}_g)^2 |g(t)|^2 dt \right)^{\frac{1}{2}}. \quad (3.11)$$

Similarly, the *effective bandwidth* is

$$b_g \triangleq \left(4\pi \int_f (f - \check{f}_g)^2 |G(f)|^2 df \right)^{\frac{1}{2}}. \quad (3.12)$$

Hence, d_g^2 is the second central moment of $|g(t)|^2$, and b_g^2 is the second central moment of $|G(f)|^2$. The *time-bandwidth product* of a unit-norm signal $g(t)$ is thus given as $d_g b_g$. As any unit-norm signal $g(t)$ satisfies the *Heisenberg uncertainty principle*, $d_g b_g \geq 1$ (Gröchenig, 2001, Theorem 2.2.1), we say that g is *well localized* in time and frequency if its time-bandwidth product is not much larger than 1.

B. Properties of Underspread Operators

An underspread random operator \mathbb{H} has the following properties (Kozek, 1997a):

- U1: All realizations of the underspread operator \mathbb{H} are approximately *normal*. This implies that for each realization $k_{\mathbb{H}}(t, t')$ of the random kernel $\mathbf{k}_{\mathbb{H}}(t, t')$, the SVD (2.26) can be replaced by a spectral decomposition.
- U2: Any deterministic unit-norm signal $g(t)$ that is well localized in time and frequency is an *approximate eigenfunction* of \mathbb{H} in the mean square sense: the mean squared error $\mathbb{E}[\|\langle \mathbb{H}g, g \rangle g - \mathbb{H}g\|^2]$ is small if \mathbb{H} is underspread. This error can be further reduced by an appropriate choice of $g(t)$, where the choice depends on the scattering function $c_s(\nu, \tau)$.
- U3: If $g(t)$ is an approximate eigenfunction, as just defined, then so is $g_{(\alpha, \beta)}(t) \triangleq g(t - \alpha)e^{i2\pi\beta t}$ for any time shift $\alpha \in \mathcal{R}$ and any frequency shift $\beta \in \mathcal{R}$.
- U4: For any (α, β) , the time-variant transfer function $\mathbf{l}_{\mathbb{H}}(\alpha, \beta)$ is an *approximate eigenvalue* of \mathbb{H} that corresponds to the approximate eigenfunction $g_{(\alpha, \beta)}(t)$ in the sense that the mean squared error $\mathbb{E}[\|\langle \mathbb{H}g_{(\alpha, \beta)}, g_{(\alpha, \beta)} \rangle - \mathbf{l}_{\mathbb{H}}(\alpha, \beta)\|^2]$ is small.

C. Approximate Diagonalization

We use the above properties of underspread operators to construct an approximation of the random channel \mathbb{H} that has a well-structured set of deterministic eigenfunctions and subsequently discuss the errors incurred by this approximation in Section 3.2.3. To synthesize the approximating operator, we use Property U1, the approximate normality of \mathbb{H} , together with Property U2 to construct an approximating kernel as

$$\sum_{l=-\infty}^{\infty} \lambda_l z_l(t) \overline{z_l}(t') \quad (3.13)$$

where, differently from (2.26), the λ_l are now random eigenvalues instead of random singular values, and the $z_l(t)$ constitute a set of

deterministic orthonormal eigenfunctions instead of random singular functions. Given that the approximating operator is normal, its input and output spaces are equivalent.

Property U2 means that we are at liberty to choose the approximate eigenfunctions $z_l(t)$ among all signals that are well localized in time and frequency. In particular, we would like the resulting approximating kernel to be convenient to work with; therefore, we choose the set of approximate eigenfunctions to be highly structured. By Property U3, it is possible to use time- and frequency-shifted versions of a single well-localized prototype function $g(t)$ as eigenfunctions. Furthermore, because the support of $\mathbf{s}_{\mathbb{H}}(\nu, \tau)$ is strictly limited in Doppler ν and delay τ , it follows from the Nyquist theorem and the Fourier relation (2.19) between the spreading function $\mathbf{s}_{\mathbb{H}}(\nu, \tau)$ and the time-variant transfer function $\mathbf{l}_{\mathbb{H}}(t, f)$ that the discrete-time discrete-frequency values $\mathbf{l}_{\mathbb{H}}(kT, nF)$, taken on a rectangular grid with

$$T \leq \frac{1}{\nu_0}, \quad (3.14a)$$

$$F \leq \frac{1}{\tau_0}, \quad (3.14b)$$

are sufficient to characterize the channel. The Nyquist condition (3.14) on the grid parameters T and F can be readily satisfied for all underspread channels. Hence, we take as our set of approximate eigenfunctions the so-called *Weyl-Heisenberg set* $\{g_{k,n}(t)\}$, where

$$g_{k,n}(t) \triangleq g(t - kT)e^{i2\pi nFt} \quad (3.15)$$

are orthonormal signals. The requirement that the $g_{k,n}(t)$ be orthonormal and at the same time well localized in time and frequency implies $TF > 1$ (Christensen, 2003), as a consequence of the Balian-Low Theorem (Gröchenig, 2001, Chapter 8). Large values of the product TF allow for better time-frequency (TF) localization of $g(t)$, but result in a loss of dimensions of the space spanned by the $\{g_{k,n}(t)\}$ with the case $TF = 1$, i.e., a loss of DOFs.

The values $\mathbf{l}_{\mathbb{H}}(kT, nF)$ are approximate eigenvalues of \mathbb{H} by Property U4; hence, our choice of approximate eigenfunctions results in

the following approximating spectral decomposition of the random kernel $\mathbf{k}_{\mathbb{H}}(t, t')$

$$\mathbf{k}_{\mathbb{H}}(t, t') \approx \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \mathbf{l}_{\mathbb{H}}(kT, nF) g_{k,n}(t) \overline{g_{k,n}(t')}. \quad (3.16)$$

For $TF > 1$, the Weyl-Heisenberg set $\{g_{k,n}(t)\}$ is not total in \mathcal{L}^2 (Christensen, 2003, Theorem 8.3.1). Therefore, the null-space of the approximating operator with kernel on the right-hand side of (3.16) is nonempty. This null-space might differ from $\mathcal{N}_{\mathbb{H}}$, the null space of the original channel operator \mathbb{H} , because the kernel only approximates $\mathbf{k}_{\mathbb{H}}(t, t')$. Similarly, the output space of the approximating operator in (3.16) might differ from the original output space $\mathcal{R}_{\mathbb{H}}$. The characterization of the difference between these spaces is an important open problem.

3.2.3. Approximate Eigenfunctions and Eigenvalues

The construction of the approximating channel operator in the preceding section relies on Property U1 through Property U4 of underspread WSSUS channels, in particular on the existence of approximate eigenfunctions and eigenvalues with small approximation error. In this subsection, we give bounds on the mean square approximation error for both approximate eigenfunctions and eigenvalues. The results we present here were first derived by Kozek (1997a) and Kozek and Molisch (1998) and later generalized by Matz and Hlawatsch (1998, 2003a). Our goal here is to provide a self-contained and unified exposition.

A. Ambiguity Function

The matching rules for $g(t)$ can be expressed in terms of the *ambiguity function* $A_g(\nu, \tau)$ of the prototype pulse $g(t)$, which is defined as

(Woodward, 1953)

$$A_g(\nu, \tau) \triangleq \int_t g(t)\bar{g}(t - \tau)e^{-i2\pi\nu t} dt. \quad (3.17)$$

Without loss of generality we can assume that $g(t)$ is normalized, so that $A_g(0, 0) = \|g\|^2 = 1$. For two signals $g(t)$ and $f(t)$, the *cross-ambiguity function* is defined as

$$A_{g,f}(\nu, \tau) \triangleq \int_t g(t)\bar{f}(t - \tau)e^{-i2\pi\nu t} dt \quad (3.18)$$

For the derivations below, we need the following properties of the ambiguity and cross-ambiguity functions (Wilcox, 1991):

A1: The volume under the so-called *ambiguity surface* $|A_g|^2$ is constant. In particular, if $g(t)$ is normalized such that $A_g(0, 0) = 1$, then

$$\iint_{\nu \tau} |A_g(\nu, \tau)|^2 d\tau d\nu = 1. \quad (3.19)$$

A2: The ambiguity surface attains its maximum magnitude at the origin: $|A_g(\nu, \tau)|^2 \leq |A_g(0, 0)|^2 = 1$ for all ν and τ .

A3: The cross-ambiguity function of the time- and frequency-shifted signals $g_{(\alpha, \beta)}(t) \triangleq g(t - \alpha)e^{i2\pi\beta t}$ and $g_{(\alpha', \beta')}(t) = g(t - \alpha')e^{i2\pi\beta' t}$ is given by

$$\begin{aligned} & A_{g_{(\alpha, \beta)}, g_{(\alpha', \beta')}}(\nu, \tau) \\ &= \int_t g(t - \alpha)e^{i2\pi\beta t}\bar{g}(t - \alpha' - \tau)e^{-i2\pi\beta'(t - \tau)}e^{-i2\pi\nu t} dt \\ &\stackrel{(a)}{=} e^{i2\pi\beta'\tau}e^{-i2\pi(\nu + \beta' - \beta)\alpha} \\ &\quad \times \int_{t'} g(t')\bar{g}(t' - (\alpha' - \alpha) - \tau)e^{-i2\pi(\nu + \beta' - \beta)t'} dt' \\ &= A_g(\nu + \beta' - \beta, \tau + \alpha' - \alpha)e^{-i2\pi(\nu\alpha - \tau\beta')}e^{-i2\pi(\beta' - \beta)\alpha} \end{aligned} \quad (3.20)$$

where (a) follows from the change of variables $t' = t - \alpha$. As a direct consequence of (3.20),

$$A_{g(\alpha,\beta)}(\nu, \tau) = A_g(\nu, \tau)e^{-i2\pi(\nu\alpha - \tau\beta)}. \quad (3.21)$$

A4: The second partial derivatives of the ambiguity function are related to the effective duration d_g and the effective bandwidth b_g of $g(t)$ as

$$\frac{\partial^2 A_g(\nu, \tau)}{\partial \nu^2} = -\pi(d_g^2 + 4\pi\check{t}_g^2), \quad (3.22)$$

$$\frac{\partial^2 A_g(\nu, \tau)}{\partial \tau^2} = -\pi(b_g^2 + 4\pi\check{f}_g^2). \quad (3.23)$$

A5: This last property pertains to both the ambiguity function $A_g(\nu, \tau)$ and any HS operator \mathbb{H} .

$$\begin{aligned} \langle \mathbb{H}g, f \rangle &\stackrel{(a)}{=} \iiint_{t \nu \tau} s_{\mathbb{H}}(\nu, \tau) g(t - \tau) e^{i2\pi t\nu} \bar{f}(t) d\tau d\nu dt \\ &= \iint_{\nu \tau} s_{\mathbb{H}}(\nu, \tau) \int_t \overline{f\bar{g}(t - \tau)} e^{i2\pi t\nu} dt d\tau d\nu \\ &= \iint_{\nu \tau} s_{\mathbb{H}}(\nu, \tau) \overline{A_{g,f}}(\nu, \tau) d\tau d\nu = \langle \mathbf{s}_{\mathbb{H}}, A_{g,f} \rangle, \end{aligned} \quad (3.24)$$

where (a) follows from the IO relation (2.13).

Property A1 and Property A2 constitute the so-called *radar uncertainty principle*, a manifestation of the classical Heisenberg uncertainty principle. These properties imply that it is not possible to find a signal $g(t)$ with a corresponding ambiguity function $A_g(\nu, \tau)$ that is arbitrarily well concentrated in ν and τ (Wilcox, 1991). When $g(t)$ has effective duration d_g and effective bandwidth b_g , the corresponding ambiguity function $A_g(\nu, \tau)$ is highly concentrated on a rectangle of area $d_g b_g / \pi$, but this area cannot be made arbitrarily small.

B. Approximate Eigenfunctions

Lemma 3.1. *Let \mathbb{H} be an underspread random HS operator with scattering function $c_s(\nu, \tau)$. Then, any unit-norm signal $g(t)$ that is well*

concentrated in time and frequency is an approximate eigenfunction of \mathbb{H} in mean square, and the approximation error is given by

$$\begin{aligned} \epsilon_1 &\triangleq \mathbb{E} [\|\langle \mathbb{H} g, g \rangle g - \mathbb{H} g\|^2] \\ &= \iint_{\nu \tau} c_s(\nu, \tau) \left(1 - |A_g(\nu, \tau)|^2\right) d\tau d\nu. \end{aligned} \quad (3.25)$$

Proof. As $g(t)$ has unit norm by assumption,

$$\begin{aligned} &\mathbb{E} [\|\langle \mathbb{H} g, g \rangle g - \mathbb{H} g\|^2] \\ &= \mathbb{E} [\|\langle \mathbb{H} g, g \rangle g\|^2] + \mathbb{E} [\|\mathbb{H} g\|^2] - 2 \mathbb{E} [|\langle \mathbb{H} g, g \rangle|^2] \\ &= \mathbb{E} [\|\mathbb{H} g\|^2] - \mathbb{E} [|\langle \mathbb{H} g, g \rangle|^2]. \end{aligned} \quad (3.26)$$

We now evaluate the two terms in (3.26) separately. The first one is

$$\begin{aligned} \mathbb{E} [\|\mathbb{H} g\|^2] &\stackrel{(a)}{=} \mathbb{E} \left[\int_t \left| \iint_{\nu \tau} \mathbf{s}_{\mathbb{H}}(\nu, \tau) g(t - \tau) e^{i2\pi t\nu} d\tau d\nu \right|^2 dt \right] \\ &\stackrel{(b)}{=} \iint_{\nu \tau} c_s(\nu, \tau) \int_t g(t - \tau) \overline{g}(t - \tau) dt d\tau d\nu \\ &\stackrel{(c)}{=} \iint_{\nu \tau} c_s(\nu, \tau) d\tau d\nu \end{aligned} \quad (3.27)$$

where (a) follows from the noise-free version of (2.29), (b) from the WSSUS property, and (c) from the normalization of $g(t)$. For the second term we have

$$\begin{aligned} \mathbb{E} [|\langle \mathbb{H} g, g \rangle|^2] &\stackrel{(a)}{=} \mathbb{E} [|\langle \mathbf{s}_{\mathbb{H}}, A_g \rangle|^2] \\ &= \mathbb{E} \left[\left| \iint_{\nu \tau} \mathbf{s}_{\mathbb{H}}(\nu, \tau) \overline{A_g(\nu, \tau)} d\tau d\nu \right|^2 \right] \\ &\stackrel{(b)}{=} \iint_{\nu \tau} c_s(\nu, \tau) |A_g(\nu, \tau)|^2 d\tau d\nu. \end{aligned} \quad (3.28)$$

Here, (a) follows from Property A5 and (b) follows from the WSSUS assumption. To conclude the derivation of (3.25), we substitute (3.27) and (3.28) into (3.26).

To minimize the error ϵ_1 in (3.25), we need to choose $g(t)$ such that $A_g(\nu, \tau) \approx A_g(0, 0) = 1$ over the support of the scattering function. If the channel is highly underspread, we can replace $A_g(\nu, \tau)$ on the right-hand side of (3.25) with its second-order Taylor series expansion around the point $(\nu, \tau) = (0, 0)$; Property A4 now shows that good time and frequency localization of $g(t)$ is *necessary* for ϵ_1 to be small. If $g(t)$ is taken to be real and even, the second-order Taylor series expansion of $A_g(\nu, \tau)$ around the point $(\nu, \tau) = (0, 0)$ takes on an especially simple form because (i) the first-order term is zero, (ii) we can approximate $A_g(\nu, \tau)$ around $(0, 0)$ as follows (Wilcox, 1991):

$$A_g(\nu, \tau) \approx 1 - [d_g^2 \nu^2 + b_g^2 \tau^2 - i\nu\tau] / 2. \quad (3.29)$$

Hence, when $g(t)$ is real and even, good time and frequency localization of $g(t)$ is also *sufficient* for ϵ_1 to be small. \square

The time- and frequency-shifted version $g_{(\alpha, \beta)}(t)$ of the prototype pulse is an approximate eigenfunction of \mathbb{H} for all $\alpha, \beta \in \mathcal{R}$ whenever $g(t)$ is an approximate eigenfunction because the ambiguity surface is invariant under time and frequency translates of $g(t)$ by Property A3.

C. Approximate Eigenvalues

Lemma 3.2. *Let \mathbb{H} be an underspread random HS operator with time-variant transfer function $\mathbf{l}_{\mathbb{H}}(t, f)$ and scattering function $c_{\mathbf{s}}(\nu, \tau)$, and let $g_{(\alpha, \beta)}(t)$ be any unit-norm approximate eigenfunction of \mathbb{H} . Then, $\mathbf{l}_{\mathbb{H}}(\alpha, \beta)$ is an approximate eigenvalue of \mathbb{H} associated with $g_{(\alpha, \beta)}(t)$, and the approximation error in mean square,*

$$\epsilon_2 \triangleq \mathbb{E} \left[\left| \langle \mathbb{H} g_{(\alpha, \beta)}(t), g_{(\alpha, \beta)} \rangle - \mathbf{l}_{\mathbb{H}}(\alpha, \beta) \right|^2 \right], \quad (3.30)$$

is given by

$$\epsilon_2 = \iint_{\nu \tau} c_{\mathbf{s}}(\nu, \tau) |1 - A_g(\nu, \tau)|^2 d\tau d\nu. \quad (3.31)$$

Proof. The proof follows from arguments by Matz and Hlawatsch (2003a, Section 9.5.3) and a theorem by the same authors (Matz and Hlawatsch, 1998, Theorem III.6): We use Property A5 and the Fourier relation (2.30) to write ϵ_2 as

$$\begin{aligned} \epsilon_2 &= \mathbb{E} \left[\left| \iint_{\nu \tau} \mathbf{s}_{\mathbb{H}}(\nu, \tau) \left[\overline{A_{g(\alpha, \beta)}(\nu, \tau)} - e^{i2\pi(\nu\alpha - \tau\beta)} \right] d\tau d\nu \right|^2 \right] \\ &\stackrel{(a)}{=} \mathbb{E} \left[\left| \iint_{\nu \tau} \mathbf{s}_{\mathbb{H}}(\nu, \tau) e^{i2\pi(\nu\alpha - \tau\beta)} \left[\overline{A_g(\nu, \tau)} - 1 \right] d\tau d\nu \right|^2 \right] \\ &\stackrel{(b)}{=} \iint_{\nu \tau} c_{\mathbf{s}}(\nu, \tau) |1 - A_g(\nu, \tau)|^2 d\tau d\nu. \end{aligned} \quad (3.32)$$

Here, (a) follows from (3.21) and (b) is a consequence of the WSSUS assumption. \square

As before, the error ϵ_2 is minimized if the ambiguity function satisfies $A_g(\nu, \tau) = A_g(0, 0) = 1$ over the support of the scattering function. This is facilitated if the support of $c_{\mathbf{s}}(\nu, \tau)$ is small, or, equivalently, if the channel is underspread. In this case, $g(t)$ can be chosen such that $A_g(\nu, \tau) \approx A_g(0, 0)$ for all $\nu \in [-\nu_0/2, \nu_0/2]$ and all $\tau \in [-\tau_0/2, \tau_0/2]$.

3.2.4. Discretized Input-Output Relation

The approximating random channel operator in (3.16) has a highly structured set of deterministic orthonormal eigenfunctions. Hence, we can diagonalize the corresponding IO relation without the need for perfect channel knowledge at both transmitter and receiver. Any

input signal $\mathbf{x}(t)$ that lies in the input space of the approximating operator is uniquely characterized by its projection onto the $g_{k,n}(t)$.

Although the signals in the input space of both the original and the approximating operators can be of infinite bandwidth and time duration, we need to model the fact that all physically realizable transmit signals are effectively band and time limited. Because the prototype function $g(t)$ is well concentrated in time and frequency by construction, this effective limitation can be modeled very elegantly by constraining the TF slots used by a given input signal to lie in a $K \times N$ -dimensional rectangle in the time-frequency plane. The resulting input signal

$$\mathbf{x}(t) = \sum_{k=0}^{K-1} \sum_{n=0}^{N-1} \underbrace{\langle \mathbf{x}, g_{k,n} \rangle}_{\triangleq \mathbf{x}[k,n]} g(t), \quad (3.33)$$

though not perfectly limited neither in time nor in frequency, has most of its energy concentrated in a finite TF rectangle of effective duration $D = KT$ and effective bandwidth $B = NF$. We call the coefficient $\mathbf{x}[k, n]$ the *transmit symbol* in the TF slot (k, n) .

The received signal can be expanded in the orthonormal basis $\{g_{k,n}(t)\}$. To compute the resulting projection coefficients, we substitute the approximating kernel (3.16) and the canonical input signal (3.33) into the integral IO relation (2.23), add white Gaussian noise $\mathbf{w}(t)$, and project the resulting noisy received signal $\mathbf{y}(t)$ onto the basis functions $g_{k,n}(t)$. In TF slot (k, n) , this yields the *received symbol*

$$\begin{aligned} \mathbf{y}[k, n] &\triangleq \langle \mathbf{y}, g_{k,n} \rangle \\ &= \langle \mathbb{H} \mathbf{x}, g_{k,n} \rangle + \langle \mathbf{w}, g_{k,n} \rangle \\ &= \sum_{k', n'} \mathbf{x}[k', n'] \langle \mathbb{H} g_{k', n'}, g_{k,n} \rangle + \mathbf{w}[k, n] \\ &= \underbrace{\mathbb{I}_{\mathbb{H}}(kT, nF)}_{\triangleq \mathbf{h}[k,n]} \mathbf{x}[k, n] + \mathbf{w}[k, n] \\ &= \mathbf{h}[k, n] \mathbf{x}[k, n] + \mathbf{w}[k, n], \end{aligned} \quad (3.34)$$

where we defined the *channel coefficients* $\mathbf{h}[k, n] \triangleq \mathbf{l}_{\mathbb{H}}(kT, nF)$. Orthonormality of the set $\{g_{k,n}(t)\}$ implies that the discrete-time noise signal $\mathbf{w}[k, n]$ is JPG, i.i.d. over time k and frequency n ; for convenience, we normalize the noise variance so that $\mathbf{w}[k, n] \sim \mathcal{CN}(0, 1)$ for all k and n . The diagonalized IO relation (3.34) is completely generic, i.e., it is not limited to a specific signaling scheme.

The discrete-time discrete-frequency channel coefficients $\mathbf{h}[k, n]$ constitute a two-dimensional discrete-parameter random process that is stationary in both k and n , and JPG with zero mean and correlation function

$$\begin{aligned} c_{\mathbf{l}}[\Delta k, \Delta n] &\triangleq \mathbb{E}[\mathbf{h}[k + \Delta k, n + \Delta n]\overline{\mathbf{h}}[k, n]] \\ &= \mathbb{E}[\mathbf{l}_{\mathbb{H}}((k + \Delta k)T, (n + \Delta n)F)\overline{\mathbf{l}_{\mathbb{H}}}(kT, nF)] \\ &= c_{\mathbf{l}}(\Delta k T, \Delta n F). \end{aligned} \quad (3.35)$$

We assume that its two-dimensional power spectral density is well-defined for all $|\varphi|, |\zeta| \leq 1/2$; it is given as

$$\psi(\varphi, \zeta) \triangleq \sum_{\Delta k=-\infty}^{\infty} \sum_{\Delta n=-\infty}^{\infty} c_{\mathbf{l}}[\Delta k, \Delta n] e^{-i2\pi(\Delta k\varphi - \Delta n\zeta)}. \quad (3.36)$$

The approximation (3.16), on which the discretization $\mathbf{h}[k, n]$ of the channel \mathbb{H} is based, allows us to express the spectrum $\psi(\varphi, \zeta)$ of $\mathbf{h}[k, n]$ in terms of the original channel's scattering function $c_{\mathbf{s}}(\nu, \tau)$. This is a very useful property of the discretization (3.34), because the scattering function of \mathbb{H} is operationally meaningful, it can be directly obtained from measurements.

$$\begin{aligned} \psi(\varphi, \zeta) &\stackrel{(a)}{=} \sum_{\Delta k=-\infty}^{\infty} \sum_{\Delta n=-\infty}^{\infty} e^{-i2\pi(\Delta k\varphi - \Delta n\zeta)} \\ &\quad \times \iint_{\nu \tau} c_{\mathbf{s}}(\nu, \tau) e^{i2\pi(\Delta k T \nu - \Delta n F \tau)} d\tau d\nu \end{aligned}$$

$$\begin{aligned}
 &= \iint_{\nu \tau} c_{\mathbf{s}}(\nu, \tau) \sum_{\Delta k=-\infty}^{\infty} e^{i2\pi\Delta k T(\nu - \frac{\varphi}{T})} \\
 &\quad \times \sum_{\Delta n=-\infty}^{\infty} e^{-i2\pi\Delta n F(\tau - \frac{\zeta}{F})} d\nu d\tau \\
 &\stackrel{(b)}{=} \frac{1}{TF} \iint_{\nu \tau} c_{\mathbf{s}}(\nu, \tau) \sum_{\Delta k=-\infty}^{\infty} \delta\left(\nu - \frac{\varphi + \Delta k}{T}\right) \\
 &\quad \times \sum_{\Delta n=-\infty}^{\infty} \delta\left(\tau - \frac{\zeta + \Delta n}{F}\right) d\nu d\tau \\
 &= \frac{1}{TF} \sum_{\Delta k=-\infty}^{\infty} \sum_{\Delta n=-\infty}^{\infty} c_{\mathbf{s}}\left(\frac{\varphi + \Delta k}{T}, \frac{\zeta + \Delta n}{F}\right) \tag{3.37}
 \end{aligned}$$

where (a) follows from the Fourier relation (2.38), and (b) results from Poisson's summation formula. For underspread channels and if the TF grid parameters T and F are chosen to satisfy the Nyquist condition (3.14), it follows that

$$\psi(\varphi, \zeta) = \frac{1}{TF} c_{\mathbf{s}}\left(\frac{\varphi}{T}, \frac{\zeta}{F}\right), \quad |\varphi|, |\zeta| \leq 1/2. \tag{3.38}$$

Under the same conditions, the variance of each channel coefficient is

$$\begin{aligned}
 \mathbb{E}\left[|\mathbf{h}[k, n]|^2\right] &= \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \psi(\varphi, \zeta) d\varphi d\zeta \\
 &\stackrel{(a)}{=} \frac{1}{TF} \sum_{\Delta k=-\infty}^{\infty} \sum_{\Delta n=-\infty}^{\infty} \\
 &\quad \times \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} c_{\mathbf{s}}\left(\frac{\varphi + \Delta k}{T}, \frac{\zeta + \Delta n}{F}\right) d\varphi d\zeta
 \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{TF} \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} c_{\mathbf{s}}\left(\frac{\varphi}{T}, \frac{\zeta}{F}\right) d\varphi d\zeta \\
&\stackrel{(b)}{=} \iint_{\nu \tau} c_{\mathbf{s}}(\nu, \tau) d\tau d\nu, \tag{3.39}
\end{aligned}$$

where (a) follows from (3.37) and (b) results from a change of variables: $\zeta/F \rightarrow \tau$ and $\varphi/T \rightarrow \nu$. Consistent with (2.39), we normalize the variance of all channel coefficients, i.e., we set

$$\mathbb{E} \left[|\mathbf{h}[k, n]|^2 \right] = \iint_{\nu \tau} c_{\mathbf{s}}(\nu, \tau) d\tau d\nu = 1. \tag{3.40}$$

3.2.5. OFDM Interpretation of the Approximating Channel Model

The canonical signaling scheme (3.33) and the corresponding diagonalized IO relation (3.34) can be interpreted in several different ways.

- The set of orthonormal basis functions $\{g_{k,n}(t)\}$ spans the input and output spaces of the approximating channel operator. Hence, every conceivable transmitted signal, independent of the modulation scheme used, can be described in our framework. Therefore, the diagonalized IO relation (3.34) is well suited to analyze the capacity of WSSUS underspread channels. Our capacity analysis in Chapter 7 builds on this interpretation.
- The IO relation (3.34) also describes a practical transmission system. The decomposition of the input signal (3.33) can be interpreted as pulse-shaped orthogonal frequency division multiplexing (PS-OFDM) (Kozek and Molisch, 1998), where discrete data symbols $\mathbf{x}[k, n]$ are modulated onto a set of orthogonal subcarriers, indexed by n , and OFDM symbols, indexed by k . From this point of view, the capacity results in Chapter 7 can be interpreted as results on the *achievable rates* of PS-OFDM over a random LTV channel \mathbb{H} (Matz et al., 2007).

The latter system perspective leads to an operational interpretation of the error incurred when approximating $\mathbf{k}_{\mathbb{H}}(t, t')$ as in (3.16). The time- and frequency-dispersive nature of LTV channels leads to intersymbol interference (ISI) and intercarrier interference (ICI) in the received PS-OFDM signal. This is apparent if we project the noiseless received signal $\mathbf{r}(t) \triangleq (\mathbb{H}\mathbf{x})(t)$ onto an arbitrary basis function $g_{k,n}(t)$:

$$\begin{aligned} \langle \mathbf{r}, g_{k,n} \rangle &= \langle \mathbb{H}\mathbf{x}, g_{k,n} \rangle = \sum_{k'=-\infty}^{\infty} \sum_{n'=0}^{N-1} \mathbf{x}[k', n'] \langle \mathbb{H} g_{k',n'}, g_{k,n} \rangle \\ &= \langle \mathbb{H} g_{k,n}, g_{k,n} \rangle \mathbf{x}[k, n] + \sum_{\substack{k'=-\infty \\ (k',n') \neq (k,n)}}^{\infty} \sum_{n'=0}^{N-1} \mathbf{x}[k', n'] \langle \mathbb{H} g_{k',n'}, g_{k,n} \rangle. \end{aligned} \quad (3.41)$$

The second term on the right-hand side of (3.41) corresponds to ISI and ICI, while the first term is the desired signal; we approximate it as $\mathbf{l}_{\mathbb{H}}(kT, nF)\mathbf{x}[k, n]$ by Property U4 in Section 3.2.2. Hence, the IO relation (3.34), which results from the approximation (3.16), can be interpreted as PS-OFDM transmission over the original channel \mathbb{H} if we neglect all ISI and ICI terms.

With proper design of the pulse shape $g(t)$ and choice of the grid parameters T and F , both ISI and ICI can be mitigated (Kozek and Molisch, 1998; Liu et al., 2004). The larger the TF grid product TF , the more effective the reduction in ISI and ICI, but the larger also the loss of dimensions of the input space. Heuristically, a useful compromise between loss of dimensions in signal space and reduction of the interference terms seems to result for $TF \approx 1.2$ (Kozek and Molisch, 1998; Matz et al., 2007). The cyclic prefix (CP) in a conventional CP-OFDM system, like an IEEE 802.11.a/g/n WLAN, incurs a similar dimension loss. We provide an upper bound on the mean squared value of the interference term in (3.41) and show how this upper bound can be minimized by careful choice of the pulse $g(t)$ and the grid parameters T and F in the next subsection.

To summarize, the conceptual difference between the interpretation of the diagonalized IO relation (3.34) as an approximation of

the original IO relation (2.31) on the one hand and PS-OFDM without ISI and ICI on the other hand is that the former allows us to make statements of an approximate nature about the capacity of the continuous-time LTV channel \mathbb{H} , while the latter yields results on the achievable rates of a specific signaling scheme, namely, PS-OFDM modulation. As the choice of the Weyl-Heisenberg set $\{g_{k,n}(t)\}$ as an approximate eigenbasis for \mathbb{H} is motivated by the fundamental Property U3 of underspread channels, we believe that PS-OFDM is a suitable modulation scheme for such channels.

3.2.6. OFDM Pulse Design for Minimum ISI and ICI

Let $\mathbf{r}(t) = (\mathbb{H} \mathbf{x})(t)$ denote the noise-free channel output signal when the input signal $\mathbf{x}(t)$ is PS-OFDM. For mathematical convenience, we consider the case of an infinite time and frequency horizon, assume that the input symbols $\mathbf{x}[k, n]$ are i.i.d., and further assume that they have zero mean and that their variance is upper-bounded as $\mathbb{E}[|\mathbf{x}[k, n]|^2] \leq 1$.

We want to quantify the error incurred if we assume that the projection of the received signal $\mathbf{r}(t)$ onto any one basis function $g_{k,n}(t)$ equals $\langle \mathbf{r}, g_{k,n} \rangle = \mathbf{x}[k, n] \mathbf{l}_{\mathbb{H}}(kT, nF)$, i.e., if we neglect the interference terms in (3.41) and use that the transmit signal is an approximate eigenfunction by Property U4. This error is given in the mean-square sense by

$$\epsilon_3 \triangleq \mathbb{E} \left[|\langle \mathbf{r}, g_{k,n} \rangle - \mathbf{x}[k, n] \mathbf{l}_{\mathbb{H}}(kT, nF)|^2 \right], \quad (3.42)$$

where the expectation is over the channel realizations and the input symbols. We bound ϵ_3 as follows:

$$\begin{aligned} \epsilon_3 = \mathbb{E} \left[& |\langle \mathbf{r}, g_{k,n} \rangle - \mathbf{x}[k, n] \langle \mathbb{H} g_{k,n}, g_{k,n} \rangle \\ & + \mathbf{x}[k, n] (\langle \mathbb{H} g_{k,n}, g_{k,n} \rangle - \mathbf{l}_{\mathbb{H}}(kT, nF))|^2 \right] \end{aligned}$$

$$\begin{aligned}
 &\stackrel{(a)}{\leq} 2 \underbrace{\mathbb{E} \left[\left| \langle \mathbf{r}, g_{k,n} \rangle - \mathbf{x}[k, n] \langle \mathbb{H} g_{k,n}, g_{k,n} \rangle \right|^2 \right]}_{\triangleq \epsilon_4} \\
 &\quad + 2 \mathbb{E} \left[\left| \mathbf{x}[k, n] (\langle \mathbb{H} g_{k,n}, g_{k,n} \rangle - \mathbf{l}_{\mathbb{H}}(kT, nF)) \right|^2 \right] \\
 &= 2\epsilon_4 + 2 \mathbb{E} \left[\left| \mathbf{x}[k, n] \right|^2 \right] \underbrace{\mathbb{E} \left[\left| \langle \mathbb{H} g_{k,n}, g_{k,n} \rangle - \mathbf{l}_{\mathbb{H}}(kT, nF) \right|^2 \right]}_{\epsilon_2} \\
 &\leq 2\epsilon_4 + 2\epsilon_2. \tag{3.43}
 \end{aligned}$$

where (a) holds because for any two complex numbers u and v we have that $|u + v|^2 \leq 2|v|^2 + 2|u|^2$. The error ϵ_2 is the same as the one computed in Lemma 3.2. It quantifies the approximation error that arises when we treat $g_{k,n}(t)$ as an eigenfunction of \mathbb{H} and $\mathbf{l}_{\mathbb{H}}(kT, nF)$ as the corresponding eigenvalue. The error ϵ_4 results because we neglect ISI and ICI. Under the assumption that the $\mathbf{x}[k, n]$ are i.i.d. and the variance is bounded as $\mathbb{E}[|\mathbf{x}[k, n]|^2] \leq 1$ for all k and n , the error ϵ_4 can be bounded as follows (Kozek, 1997a; Matz et al., 2007):

$$\begin{aligned}
 \epsilon_4 &= \mathbb{E} \left[\left| \langle \mathbf{r}, g_{k,n} \rangle \right|^2 \right] + \mathbb{E} \left[\left| \mathbf{x}[k, n] \right|^2 \right] \mathbb{E} \left[\left| \langle \mathbb{H} g_{k,n}, g_{k,n} \rangle \right|^2 \right] \\
 &\quad - 2\Re \left\{ \mathbb{E} \left[\mathbf{x}[k, n] \langle \mathbf{r}, g_{k,n} \rangle \overline{\langle \mathbb{H} g_{k,n}, g_{k,n} \rangle} \right] \right\} \\
 &\stackrel{(a)}{=} \sum_{\substack{k'=-\infty \\ (k', n') \neq (k, n)}}^{\infty} \sum_{n'=-\infty}^{\infty} \mathbb{E} \left[\left| \mathbf{x}[k', n'] \right|^2 \right] \mathbb{E} \left[\left| \langle \mathbb{H} g_{k', n'}, g_{k, n} \rangle \right|^2 \right] \\
 &\stackrel{(b)}{\leq} \sum_{k'=-\infty}^{\infty} \sum_{\substack{n'=-\infty \\ (k', n') \neq (k, n)}}^{\infty} \mathbb{E} \left[\left| \langle \mathbb{H} g_{k', n'}, g_{k, n} \rangle \right|^2 \right]. \tag{3.44}
 \end{aligned}$$

Here, (a) follows (3.41) and because $\mathbf{x}[k, n]$ are i.i.d. and zero mean, and (b) results because $\mathbb{E}[|\mathbf{x}[k, n]|^2] \leq 1$. Next, we evaluate the last

expectation on the right-hand side in (3.44):

$$\begin{aligned}
 & \mathbb{E} \left[\left| \langle \mathbb{H} g_{k',n'}, g_{k,n} \rangle \right|^2 \right] \\
 & \stackrel{(a)}{=} \mathbb{E} \left[\left| \langle \mathbf{s}_{\mathbb{H}}, A_{g_{k,n}, g_{k',n'}} \rangle \right|^2 \right] \\
 & \stackrel{(b)}{=} \iint_{\nu \tau} c_{\mathbf{s}}(\nu, \tau) \left| A_{g_{k,n}, g_{k',n'}}(\nu, \tau) \right|^2 d\tau d\nu \\
 & \stackrel{(c)}{=} \iint_{\nu \tau} c_{\mathbf{s}}(\nu, \tau) \left| A_g(\nu + (n' - n)F, \tau + (k' - k)T) \right|^2 d\tau d\nu \\
 & = \iint_{\nu \tau} c_{\mathbf{s}}(\nu - (n' - n)F, \tau - (k' - k)T) \left| A_g(\nu, \tau) \right|^2 d\tau d\nu. \quad (3.45)
 \end{aligned}$$

Equality in (a) follows from Property A5, (b) is again a consequence of the WSSUS assumption, and (c) results from Property A3. We finally substitute (3.45) into (3.44) and obtain the following bound on the approximation error ϵ_4 :

$$\begin{aligned}
 \epsilon_4 & \leq \\
 & \leq \sum_{\substack{k'=-\infty \\ (k',n') \neq (k,n)}}^{\infty} \sum_{\substack{n'=-\infty \\ (k',n') \neq (k,n)}}^{\infty} \iint_{\nu \tau} c_{\mathbf{s}}(\nu - (n' - n)F, \tau - (k' - k)T) \left| A_g(\nu, \tau) \right|^2 d\tau d\nu \\
 & = \sum_{\substack{k=-\infty \\ (k,n) \neq (0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} \iint_{\nu \tau} c_{\mathbf{s}}(\nu - nF, \tau - kT) \left| A_g(\nu, \tau) \right|^2 d\tau d\nu. \quad (3.46)
 \end{aligned}$$

Hence, the error ϵ_4 is small if the ambiguity surface $|A_g(\nu, \tau)|^2$ of $g(t)$ takes on small values on the periodically repeated rectangles $[-\nu_0/2 + nF, \nu_0/2 + nF] \times [-\tau_0/2 + kT, \tau_0/2 + kT]$, except for the rectangle centered at the origin, i.e., the hatched rectangles in Figure 3.2. This condition can be satisfied if the channel is highly underspread and if the grid parameters T and F are chosen such that the gray region in Figure 3.2 is large enough to allow the ambiguity surface $|A_g(\nu, \tau)|^2$ to decay. If $g(t)$ has effective duration d_g and effective bandwidth b_g , the

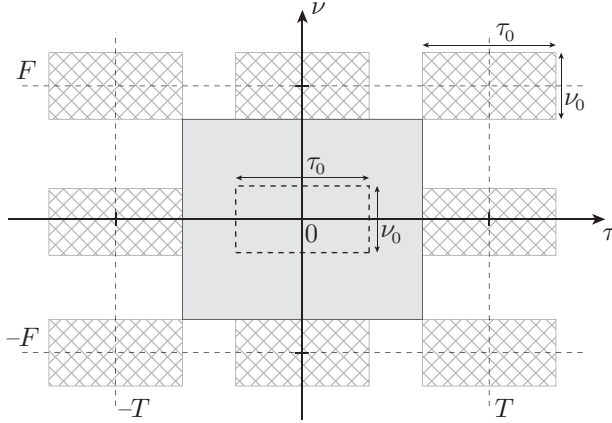


Fig. 3.2: The support set of the periodized scattering function in (3.46) are the hatched rectangles, while the area on which the ambiguity function $A_g(\nu, \tau)$ should be concentrated to minimize ϵ_4 is shaded in gray.

latter condition requires that $2T \geq \tau_0 + d_g/\sqrt{\pi}$ and $2F \geq \nu_0 + b_g/\sqrt{\pi}$. Given a constraint on the product TF , good localization of $g(t)$, both in time and frequency, is necessary for these two inequalities to hold.

The minimization of ϵ_4 in (3.46) over the set of all orthonormal Weyl-Heisenberg bases $\{g_{k,n}(t)\}$ does not admit a solution in closed form (Matz et al., 2007), but numerical methods to minimize ϵ_4 are available in the radar literature (Wilcox, 1991). However, we can provide some guidelines on how to choose the grid parameters T and F . For known maximum delay τ_0 and maximum Doppler shift ν_0 , and for a fixed product TF , the area $(2T - \tau_0)(2F - \nu_0)$ of the rectangle in gray in Figure 3.2 is maximized if (Kozek, 1997a; Kozek and Molisch, 1998)

$$\frac{T}{F} = \frac{\tau_0}{\nu_0}. \quad (3.47)$$

This choice of T and F is also appropriate to minimize ϵ_2 : the larger the gray rectangle in Figure 3.2, the simpler it is to find a pulse $g(t)$

3 DISCRETIZED MODELS

with ambiguity function almost equal to 1 on the support of the scattering function, i.e., the dashed rectangle centered at the origin in Figure 3.2.

PART II

EMPIRICAL CHARACTERIZATION OF WIDEBAND CHANNELS

CHAPTER 4

Wideband Channel Measurements

WE DEVELOPED the channel models in Part I on the basis of physical insight and engineering intuition, stressed the point that all models only *approximate* physical reality, and cautioned that measurements are needed to assess the approximation quality of a specific model. Consequently, this second part of the thesis is devoted to an empirical characterization of wideband channels. We present results of two measurement campaigns we conducted in 2004 at ETH Zurich (Schuster and Bölcskei, 2005a; Schuster et al., 2005; Schuster and Bölcskei, 2007) and discuss their implication for some of the modeling questions raised in Part I.

4.1. MEASUREMENT PRINCIPLES

4.1.1. General Considerations

Measuring the channel and devising a mathematical model for it are interdependent tasks, as

1. The measurement system itself needs to be designed on the basis of a model, and thus may not be able to correctly reproduce certain features not accounted for in its design.
2. The physical reality we are interested in is often only accessible through measurements, so that the measurement apparatus is

part of the reality we wish to model.

3. Measurements are never perfect but always contain errors, which might be mistaken as features of the measured reality and incorporated into the model.

It is in general not possible to disentangle these dependencies; the perfect measurement apparatus does not exist nor does the perfect model, as discussed in Chapter 2.

As we would like to statistically characterize wireless channels for communication purposes, the first point might not be as crucial as if we wanted to establish a deterministic channel model. Because the intention of a stochastic model is precisely to account for our ignorance about the physical reality under investigation, we may expect that channel characteristics we did not envision upon design of the measurement system will be reflected in the same way as in a communication system that is not designed to make use of these features. Hence, the model will most likely be robust, but a corresponding system design might not use the channel to its fullest potential.

As the measurement system is part of the measured channel, a sensible design choice is to render its impact similar to the impact an actual wireless transceiver chain exerts on the propagation channel, so that we directly measure the baseband channel.

Finally, the third point reflects our discussion in Section 2.1, that the complexity of a model needs to be adapted to the amount of data available to avoid fitting the model to measurement noise.

4.1.2. Measurement Techniques

Although we are interested in a stochastic characterization of wireless channels, we first need to measure a series of channel realizations, called channel *samples** in statistical terminology. These samples can

* We have an unfortunate collision of terms here. *Sampling* in the context of measurement and instrumentation commonly means to record the value of a continuous-time signal at regular time instants, while in statistics a *sample* means the random selection of a single entity out of a population. Because we

subsequently be used for statistical analysis. Measuring the channel means to send a deterministic probing signal $x(t)$ and record the resulting realization of the output signal $y(t)$, which in the scope of our LTV model from Chapter 2 is

$$y(t) = \int_{\tau} h_{\mathbb{H}}(t, \tau)x(t - \tau)d\tau + w(t). \quad (4.1)$$

From the output signal $y(t)$ we wish to infer the channel $h_{\mathbb{H}}(t, \tau)$. This is a standard system identification problem. Not all LTV systems can be identified, even in the absence of noise. Fortunately, the class of underspread LTV systems, i.e., LTV systems whose spreading function $s_{\mathbb{H}}(\nu, \tau)$ is compactly supported on a rectangle $[-\nu_0/2, \nu_0/2] \times [-\tau_0/2, \tau_0/2]$ of spread $\Delta_{\mathbb{H}} \triangleq \nu_0\tau_0 < 1$, are identifiable, as discussed in Section 3.2. Because the limitation of Doppler shift ν and delay τ is a physical one, we can assume that it holds reasonably well for all channel realizations.

As the time-variant impulse response $h_{\mathbb{H}}(t, \tau)$, the spreading function $s_{\mathbb{H}}(\nu, \tau)$, and the time-variant transfer function $l_{\mathbb{H}}(t, f)$ are related by unitary transforms, it is a matter of convenience upon which of these channel representations we base a measurement procedure.* To perform *time-domain measurements*, i.e., to directly measure $h_{\mathbb{H}}(t, \tau)$, the measurement apparatus transmits pulses and records the output signal $y(t)$ in the time domain. In contrast, a *frequency domain* measurement device sends sinusoidal signals and measures the corresponding spectrum of the received signal. We describe both measurement approaches in the following two subsections. No matter if measurements are taken in the frequency domain or in the time domain, any practical measurement apparatus transmits effectively

deal with both measurements and statistical analysis, we need to use both flavors of the term. Whenever possible, we attempt to use *sample* and *sampling* in its statistical context, e.g., a *channel sample* refers to a specific realization of a random channel impulse response. On some occasions, however, we are forced to use also the instrumentation and measurement meaning of the term for lack of a better alternative; we hope that the meaning is clear from the context.

* Note that we use a specific channel model to devise a measurement procedure. This illustrates the mentioned interlinking between model and measurement.

time- and band-limited signals and observes the output signal over an interval that is effectively limited in time and frequency; yet, for mathematical analysis it is often more convenient to assume either an infinite time and frequency horizon or a circulant sounding signal and channel response. We follow the later approach.

4.1.3. Time-Domain Measurements

The following is a summary of common time-domain measurement techniques. A more comprehensive account is given by Parsons et al. (1991), Fannin et al. (1991), Cullen et al. (1993), Lehne et al. (1998), and Parsons (2000). Conceptually the simplest way to measure the time-variant impulse response $h_{\mathbb{H}}(t, \tau)$ of an underspread LTV system with maximum Doppler shift $\nu_0/2$ and maximum delay τ_0 is to transmit a train of Dirac impulses,

$$x(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT). \quad (4.2)$$

In absence of noise and any other imperfections in the measurement system, the received signal is

$$\begin{aligned} r(t) &= \sum_{k=-\infty}^{\infty} \int_{\tau} h_{\mathbb{H}}(t, \tau) \delta(t - kT - \tau) d\tau \\ &= \sum_{k=-\infty}^{\infty} h_{\mathbb{H}}(t, t - kT). \end{aligned} \quad (4.3)$$

If the pulse repetition time T is chosen such that $T \geq \tau_0$, the output signal $r(t)$ consists of a train of nonoverlapping impulse responses. If furthermore $T \leq 1/\nu_0$, i.e., if the pulse repetition rate satisfies the Nyquist theorem with respect to the band limitation in the Doppler domain, the train of measured impulse responses (4.3) allows to perfectly reconstruct $h_{\mathbb{H}}(t, \tau)$. Taken together, the two conditions on T are $\tau_0 \leq T \leq 1/\nu_0$, which can only be satisfied if $\nu_0\tau_0 \leq 1$, i.e., if the channel is underspread.

It is challenging with current circuit techniques to generate an extremely well-defined sub-nanosecond pulse as required for channel measurements of several gigahertz bandwidth. Even more difficult is the design of wideband power amplifiers that offer high enough peak power to attain a sufficiently high SNR. The resulting SNR limitation is indeed the main reason that the direct pulse measurement technique is hardly ever used. However, this technique is only necessary if the channel's spread $\Delta_{\mathbb{H}} = \nu_0 \tau_0$ is close to 1, which is virtually never the case for wireless channels. If $\Delta_{\mathbb{H}} \ll 1$, the transmit signal can be spread over a longer time interval, so that more energy can be sent for a given peak-power constraint.

Measurement systems that spread the input signal in time and de-spread the received signal are commonly referred to as *pulse-compression* systems. Instead of an impulse train, as in (4.2), an ideal pulse compression system transmits a periodic repetition of a sounding signal $m(t)$ that is spread out in time and whose autocorrelation function is an impulse: $(m \star m)(t) = \delta(t)$. Practical sounding signals that approximate this ideal autocorrelation property are, for example, so-called pseudonoise (PN) sequences, which can be efficiently generated with maximum-length shift registers. The receiver correlates $r(t)$ with the sounding signal to obtain

$$\begin{aligned} (r \star m)(t) &= \left(\sum_{k=-\infty}^{\infty} \int_{\tau} h_{\mathbb{H}}(t, \tau) m(t - kT - \tau) d\tau \right) \star m(t) \\ &= \sum_{k=-\infty}^{\infty} \iint_{t', \tau} h_{\mathbb{H}}(t - t', \tau) \\ &\quad \times m(t - t' - kT - \tau) m(t') d\tau dt'. \end{aligned} \quad (4.4)$$

Because $h_{\mathbb{H}}(t, \tau)$ is time variant, correlation with the sounding signal at the receiver does not lead to a signal of the form (4.3); however, if $h_{\mathbb{H}}(t, \tau)$ varies slowly enough so that it can be approximated as constant over the duration of $m(t)$, the correlated received signal

becomes

$$\begin{aligned}
 (r \star m)(t) &\approx \sum_{k=-\infty}^{\infty} \iint_{t' \tau} h_{\mathbb{H}}(t, \tau) m(t - t' - kT - \tau) m(t') d\tau dt' \\
 &= \sum_{k=-\infty}^{\infty} \int_{\tau} h_{\mathbb{H}}(t, \tau) \delta(t - kT - \tau) d\tau \\
 &= \sum_{k=-\infty}^{\infty} h_{\mathbb{H}}(t, t - kT), \tag{4.5}
 \end{aligned}$$

which is the same as (4.3). The requirement that $h_{\mathbb{H}}(t, \tau)$ does not vary appreciably for the duration of $m(t)$ needs to be ensured by proper choice of the sounding signal $m(t)$.

The factor that limits the maximum measurable channel bandwidth in most measurement systems is the rate at which measured data can be processed and stored. Older measurement systems used a storage oscilloscope or a measurement tape recorder (Cox, 1972; Parsons and Bajwa, 1982), which offered only limited bandwidth. Offline processing of the recorded signals was severely limited, so that the important correlation step (4.5) was usually performed in real time by the measurement system. Two popular implementations were the swept-time delay crosscorrelator (Cox, 1972) and the convolution matched filter technique (Bajwa and Parsons, 1978). However, any analog implementation of the correlation operation is limited by the precision of the necessary filters and by synchronization errors. Therefore, most current time-domain channel sounders directly digitize the received signal, i.e., discretize it in time and quantize its amplitude. Subsequently, the measured data points need to be stored for further processing. As all processing, like the correlation with the PN sequence, is done offline, more sophisticated algorithms can be used (Cullen et al., 1993).

This fully digital approach is common practice for channel measurements of some hundred megahertz, but it is still quite challenging to measure channels with several gigahertz bandwidth. The Nyquist theorem requires the sampling rate to be at least twice the highest

frequency component in the signal to be digitized, but practical DSOs with filters of finite decay, sampling jitter, and other impairments should be operated at a sampling rate around four times the maximum desired frequency of the input signal. DSOs that provide the sampling rates required to measure channels of several gigahertz bandwidth became available only very recently. The technical limitation, though, has shifted from actually digitizing the received signal to storing the resulting data stream: to date, there does not exist a system that allows to continuously record the data stream of a multi-gigahertz analog-to-digital converter (ADC). The ultra high-speed acquisition memory used in current DSOs is fairly small; short measurement sequences can be recorded, but the transfer of these measurements to permanent storage takes quite long, up to a second on some devices. But a measurement of the channel impulse response every second is not sufficient to accurately characterize channel time variability; therefore, the data sets recorded in most of the published time-domain UWB channel measurement campaigns to date are suitable only to characterize a block-fading channel model, where the channel is assumed to take on independent realizations over successive blocks.

4.1.4. Frequency-Domain Measurements

To directly characterize the time-variant transfer function $l_{\text{H}}(t, f)$ of a channel, we need to start from an expression dual to (4.2): instead of impulses in time, the measurement system needs to transmit impulses in frequency, i.e., the transmit signal needs to consist of a number of pure sinusoids at preferably equidistant points across the band of interest. If only a discrete number of points in frequency are measured, the corresponding time-domain signal is periodic; thus, the number of frequency points needs to be chosen large enough to avoid aliasing in the time domain, i.e., the duration of a fundamental period needs to be larger than the largest expected delay of the channel. Fortunately, the number of frequency points needed does not exceed several thousand for typical indoor channels with delays up to several hundred nanoseconds; this is a relatively small number of points to be

stored compared with the billions of data points a multi-gigahertz DSO produces every second. Therefore, wideband channel measurements in the frequency domain were preferred over time-domain measurements in the past decade (see the summary in Section 4.2).

A VNA is typically used for frequency-domain measurements. Because a VNA houses transmitter and receiver in the same enclosure, a common clock is used for both of them so transmission and reception are always synchronized. Furthermore, the tight integration of transmitter and receiver enables calibration routines to be built into the system. On the negative side, co-location of transmitter and receiver necessitates long cables if transmit and receive antennas are to be separated in space. Because VNAs generate the stimulus sinusoids at different frequencies in succession rather than in parallel (Street et al., 2001), the most severe drawback of VNAs is the substantial amount of time necessary to measure a single transfer function. The two main reasons for the slow measurement speed is the necessity to tune and stabilize the internal oscillators for each frequency point, and the desire to transmit as pure a sinusoid as possible, which in turn requires the sounding signal to be of long enough duration so that the time limitation does not appreciably broaden its spectrum. The time necessary for a single transfer function measurement might easily be on the order of seconds, clearly too long to assume that $l_{\mathbb{H}}(t, f)$ remains constant. Hence, VNA measurements normally need to be conducted in a controlled environment where any motion has to be stopped during a measurement. Consequently, the resulting set of measured transfer functions does not allow to infer channel time variability, but can only be used to characterize a block-fading model.

4.1.5. Calibration

The above description of both time-domain and frequency-domain channel measurements assumes an ideal measurement system response. In reality, any measurement apparatus can only deal with effectively band-limited signals, so that the autocorrelation function of the sounding signal m in a time-domain sounding system can only approximate

4.2 A SURVEY OF WIDEBAND CHANNEL MEASUREMENTS AND MODELS

an ideal impulse $\delta(t)$, i.e., its spectrum is not flat in general. Similarly, the sinusoids used to measure the transfer function in the frequency-domain approach are time limited and, therefore, not mathematically perfect sinusoids; hence, the corresponding measured complex phasors are not ideal discrete-frequency values of the channel's transfer function. Furthermore, cables, transmit and receive filters, and antennas all further distort the channel transfer function. To some extent this distortion can be controlled for by means of *calibration*: The measurement system is connected back-to-back, i.e., without the wireless channel; the system's transfer function is measured, and this transfer function is used later on to equalize channel samples. Calibration is a standard procedure with most VNAs and requires little additional effort, while calibration of a time-domain measurement system is often more involved because it needs to be performed manually.

A measurement system that is operated in its linear regime has an almost flat calibrated transfer function, yet the noise level is raised at some frequencies. Nonlinear distortions cannot be removed by calibration and, therefore, need to be avoided as well as possible already during the design of the measurement system.

4.2. A SURVEY OF WIDEBAND CHANNEL MEASUREMENTS AND MODELS

4.2.1. Measurements

As the hardware necessary to perform wideband channel measurements became available only about ten years ago, most of the reported measurement results are fairly recent. The number of published results steadily increased over the last years, so that a significant number of results appeared in print only after we performed our measurement campaign in mid-2004. Table 4.1 on page 93 provides a fairly comprehensive overview of published wideband measurement campaigns as of fall 2007. When a single measurement campaign is the subject of several publications, i.e., a conference paper and a journal paper, we

only list the one with the most comprehensive description.

Saleh and Valenzuela (1987) conducted one of the first measurement campaigns to characterize indoor channels with large fractional bandwidth. Several years later, Scholtz (1993) and Win and Scholtz (1997, 1998a, 2000) conducted wideband measurement campaigns to complement their research on impulse radio (IR). This research illustrated the potential benefits of UWB communications and spurred increased research interest in wideband channels. Win et al. (1995) measured indoor and outdoor UWB channels with a bandwidth of about 1 GHz. First results were published in 1997 (Win et al., 1997a,b); several different channel models and various studies of specific communication techniques for wideband channels are based on their original set of indoor measurements (Win and Scholtz, 1998b; Scholtz et al., 1998; Cramer et al., 1999, 2002; Cassioli et al., 2002; Win and Scholtz, 2002). The results of several other measurement campaigns were reported at the first conference dedicated solely to UWB communications in 2002 (Kunisch and Pamp, 2002; Turin et al., 2002; Ghassemzadeh et al., 2002; Keignart and Daniele, 2002; Prettie et al., 2002; Hovinen et al., 2002). Most of the measured channel impulse responses in these references show the characteristic clustering structure, as illustrated in Figure 3.1, with either a single cluster for indoor non line of sight (NLOS) measurements or multiple clusters, depending on the measurement environment. Many other observations are not shared by all references, which resulted in a plurality of UWB channel models.

A particularly interesting observation by Kunisch and Pamp (2002) is that the channel attenuation over the measured band is frequency dependent. If the system bandwidth is large, the antennas may significantly shape the spectrum of the LTI radio channel. By the classical free-space propagation formula of Friis (1946), the amplitude of the received signal decays inversely proportional with frequency if both the transmitting and the receiving antenna have constant gain. If, however, the antenna gain changes with frequency, the frequency dependence of the free-space attenuation will be frequency-dependent as well.

4.2 A SURVEY OF WIDEBAND CHANNEL MEASUREMENTS AND MODELS

Table 4.1: Wideband Channel Measurements

Reference	Band [GHz]	Type	Environment
Saleh and Valenzuela (1987)	1.4–1.6	DSO	office
Win et al. (1997a)	0–1.3	DSO	outdoor
Win et al. (1997b)	0–1.3	DSO	office
Yano (2002)	1.25–2.75	DSO	office
Kunisch and Pamp (2002)	1–11	VNA	office
Prettie et al. (2002)	2–8	VNA	town houses
Hovinen et al. (2002)	2–8	VNA	office
Welch et al. (2002)	1–11	VNA	body
Poon and Ho (2003)	2–8	VNA	town houses
Keignart and Daniele (2003)	2–6	VNA	office
Zasowski et al. (2003)	3–6	VNA	body
Dabin et al. (2003)	2–6	VNA	classroom
Kovács et al. (2004)	3–6	VNA	body
Pagani and Pajusco (2004)	3.1–11	VNA	office
Ghassemzadeh et al. (2004)	4.37–5.63	VNA	town houses
Karedal et al. (2004b)	3.1–10.6	VNA	industrial
Menouni Hayar et al. (2005)	3–9	VNA	laboratory
Ciccognani et al. (2005)	3.6–6	DSO	office
Chong and Yong (2005)	3–10	VNA	apartments
Hentilä et al. (2005)	3.1–6	VNA	hospital
Kunisch and Pamp (2005)	3–11	VNA	industrial
Geng et al. (2005)	3–6.5	VNA	hall, office
Irahhauten et al. (2005)	3.1–10.6	DSO	office
Muqaibel et al. (2006)	0.1–12	DSO	office
Zasowski et al. (2006)	1.5–8	VNA	human head
Fort et al. (2006)	3–6	VNA	body
Jämsä et al. (2006)	1–11	VNA	classroom
Keignart et al. (2006)	3.1–4.5	DSO	laboratory
Kobayashi (2006)	3.1–10.6	VNA	car
Irahhauten et al. (2006)	3.1–10.6	DSO	ind., office
Geng and Vainikainen (2006)	3–10	VNA	office
Pagani and Pajusco (2006)	4–5	DSO	hallway
Chehri et al. (2006)	2–5	VNA	mine
Haneda et al. (2006)	3.1–10.6	VNA	wood house
Chuang et al. (2007)	3–10.6	VNA	aircraft
Schuster and Bölcskei (2007)	2–5	both	public space

The first applications of UWB communication were perceived to be WPANS; therefore, most of the early measurements were taken in offices and residential environments. Increased interest in other application scenarios lead to channel measurements of industrial environments (Karedal et al., 2004b; Kunisch and Pamp, 2005), UWB propagation at the human body (Zasowski et al., 2003; Fort et al., 2006), in cars (Kobayashi, 2006), and even aircraft (Chuang et al., 2007) and underground mines (Chehri et al., 2006).

A range of reported measurement campaigns are dedicated to specific aspects of UWB channels, like antenna directivity (Hoff et al., 2003; Dabin et al., 2003), frequency dependence of various channel parameters (Jämsä et al., 2006; Geng and Vainikainen, 2006), and material properties (Lao et al., 2003; Jaturatussanai et al., 2006).

Because of the high time resolution, spatial properties of UWB channels are markedly different from those of narrowband channels. The clustering observed in the impulse response directly corresponds to clustering in the angular domain, as the AOD and AOA of waves traveling via major scattering clusters can be resolved at least partially. Wideband channel measurements that specifically investigate spatial properties were reported by Prettie et al. (2002), Poon and Ho (2003), Keignart et al. (2006) and Haneda et al. (2006).

Most of the aforementioned measurements results are for time-invariant channels, for the reasons outlined in Sections 4.1.3 and 4.1.4. The first generation of channel measurement equipment that offers the continuously high sampling rates and data throughput necessary to measure time-variant channels became available recently (Kmec et al., 2005; Pagani and Pajusco, 2006), so that the characterization of truly time-variant UWB channels might become possible soon.

4.2.2. Models

A. Discrete-Time Block Fading

In several UWB measurement and modeling articles, sampled output sequences from a pulse-compression measurement system are directly

used to compute the *PDP* or parameters like the *path loss*, *delay spread*, and *mean delay* (Win et al., 1997a; Ghassemzadeh et al., 2004). These statistics are sufficient to describe a discrete-time US JPG block-fading channel, and can hence be considered as an implicit channel modeling step: the covariance matrix of such a channel is diagonal with the PDP on its main diagonal. Specification of delay spread and mean delay alone is not sufficient to completely specify a discrete-time block-fading model unless further assumptions on the shape of the PDP are made.

Another way to specify the covariance matrix of the channel vector, though by a smaller number of parameters, is to devise an autoregressive (AR) description of the channel frequency response (Howard and Pahlavan, 1992); implicit in such an approach is the JPG assumption on the channel vector. Ghassemzadeh et al. (2004) devised an AR model to characterize the frequency response of wideband indoor channels and found that two poles are sufficient to accurately specify the correlation structure.

Kunisch and Pamp (2002) modeled the frequency-dependent attenuation of the entire channel impulse response by an additional LTI filter with transfer function proportional to f^β , where β characterizes the effects of the combination of frequency-dependent antenna gain and all other large-scale frequency-dependent mechanisms in the channel, e.g., frequency-dependent absorption coefficients of materials.

B. Specular Block Fading

Instead of estimating the parameters of a discrete-time block-fading model from measured data, some researchers assume that the channel can be described by a block fading model with specular reflections, as proposed by Turin (1972) (see also the summary in Section 3.1.2). Because the Turin model is of infinite bandwidth, any analysis or simulation needs to band limit the corresponding channel impulse response, either implicitly by means of a band-limited input signal, or explicitly. Conversely, because any measurement system is band limited, some sort of deconvolution is necessary to estimate the model

parameters of the infinite-bandwidth Turin model from measured data of finite bandwidth. This is either done by inspection, i.e., visually identifying path arrivals in all measured discrete-time impulse responses, by direct estimation of the delay parameters (Win and Scholtz, 2002), or by means of superresolution algorithms. Cramer et al. (2002) adapted the CLEAN algorithm to separate the individual path arrivals in measured channel impulse responses under the assumption of a particular received pulse shape. The same method was later also used by Yano (2002) and Irahhtauten et al. (2006). A double-directional extension of the specular block-fading model to account for the AOD and AOA of an individual path was used by Poon and Ho (2003), in conjunction with a corresponding extension of the CLEAN algorithm to estimate the model parameters path delay and AOA. Other superresolution algorithms, like space alternating generalized expectation maximization (SAGE) (Fessler and Hero, 1994; Fleury et al., 1999) or multiple signal classification (MUSIC) rely on the assumption that the bandwidth of the signal under analysis is much smaller than its center frequency; therefore, it is not clear if these algorithms can be applied directly to UWB channel measurement data. Haneda et al. (2004, 2006) use an extension of the SAGE algorithm (Haneda and Takada, 2003) to estimate parameters of an extended directional model that also accounts for the AOA and frequency dependence of individual propagation paths.

C. Modifications of the Model by Saleh & Valenzuela

Measured UWB channel impulse responses in many different environments exhibit the characteristic clustered structure depicted in Figure 3.1. This observation suggests a model of the type proposed by Saleh and Valenzuela (1987), as described in Section 3.1.3. Indeed, the two influential UWB channel models that were originally developed to support the standardization activities in the IEEE 802.15.3a and IEEE 802.15.4a Task Groups* are derivatives of the original SV model.

* The goal of the IEEE 802.15.3a Task Group was to standardize an alternative physical layer for 802.15.3 WPANs. Because of political and technological

The IEEE 802.15.3a model is described in a report by Foerster (2003) and later detailed in a paper by Molisch et al. (2003a). It deviates from the original SV model in that it is a model for the radio channel with $f_c = 0$ so that the ray coefficients ρ_{li} in (3.3) are real-valued; furthermore, the coefficient magnitudes are lognormally instead of Gaussian distributed, with shape parameters that depend on both the cluster and the ray index, and the polarity of each coefficient is either positive or negative with equal probability. The UWB portion of the IEEE 802.15.4a channel model is specified at complex baseband, like the original SV model; however, the magnitude of the path weights ρ_{li} are modeled to follow a Nakagami distribution, while the phases are uniformly distributed. The ray arrival process is a Poisson process in the original SV model; in the IEEE 802.15.4a model, it is a mixture of two different Poisson processes, and the parameters of this mixture process for each cluster depend on the corresponding cluster arrival time. Two additional modifications are a weighting function for the entire impulse response that attenuates the early rays in NLOS settings, and a linear filter to model the frequency-dependence of the overall attenuation. Molisch et al. (2006b) describe the IEEE 802.15.4a channel model in detail. Many measurement campaigns listed in Table 4.1 on page 93 were conducted specifically to extract parameters for use in the definition of the above two models.

D. Space-Variant Models

One of the few models that takes into account temporal variation in the channel is the space-variant multipath model by Kunisch and Pamp (2003, 2006). This model is a derivative of the so-called

differences among group members, the task group never achieved this goal and disbanded itself. The IEEE 802.15.4a Task Group was chartered to standardize an alternative physical layer for industrial, sensing, monitoring, location-tracking, and safety applications on the basis of the IEEE 802.15.4 standard. The IEEE 802.15.4a standard was approved in March 2007. The role of channel models in both task groups was to aid in the selection of competing physical layer proposals for both standards; especially the IEEE 802.15.3a model was never envisioned as a comprehensive model for various types of scientific investigations. Nevertheless, it has been used to this end quite often.

geometry-based stochastic channel model (GSCM) that is used, for example, in the COST259 directional channel model described by Molisch et al. (2006a, 2003b). A GSCM describes propagation paths impinging at the receiver as emanating from a set of virtual sources, whose locations are random variables. This allows to adapt the model to specific environments. Furthermore, the virtual sources can move over time, which results in a truly time-variant channel instead of a block-fading description. The UWB extension by Kunisch and Pamp (2003) uses the GSCM approach to describe the arrival and the power of impulse response clusters, while individual channel taps are modeled stochastically, with uncorrelated proper Gaussian taps whose variance depends on the corresponding GSCM cluster energy. Furthermore, their model takes into account the frequency-dependent attenuation discussed previously. As the location of clusters in the generated impulse responses is determined by the location of virtual sources, and these locations can be made to vary in a physically sensible way over time, the variation in cluster location in the resulting impulse response is physically sensible as well. In fact, the source variation can be tuned to match a measured Doppler spectrum. The price for the high accuracy, however, is model complexity, so that the model is only suitable for simulation.

4.3. THE ETH MEASUREMENT CAMPAIGNS

The space-variant multipath model or the various extensions of the SV model are mainly developed to generate realizations of UWB channel impulse responses for use in simulators (Jeruchim et al., 2000); their utility in information- and communication-theoretic analysis of wireless communication systems is rather limited as their complexity is high. In contrast, the discrete-time block-fading model is suitable for theoretical studies, as evinced by its pervasive use (Tse and Viswanath, 2005; Biglieri, 2005). The virtue of such theoretical studies lies in the insight they provide into the fundamental mechanisms that determine the performance of wireless systems and the

high-level design guidelines they offer.

Powerful analytical results require powerful models, yet powerful results also depend on models that are simple enough to derive these results in the first place. Hence, the models need to retain those features of the real world that are of fundamental importance. The developments in Chapter 2 illustrated that a great many modeling approximations are necessary to arrive at a mathematical channel model that is analytically tractable. In particular, we found that some of the approximations made are questionable for channels of large bandwidth. Yet, most of the measurement campaigns and modeling efforts surveyed in the previous section do not address these important modeling questions. Therefore, we conducted a set of wideband channel measurements on the premises of ETH Zurich and assembled statistical tools to specifically scrutinize these modeling assumptions.

4.3.1. Objectives

Our main interest is the analysis of communication over wideband channels, in particular the impact of small scale fading. Therefore, our measurements were designed to provide empirical evidence about some of the basic assumptions that underly the stochastic channel models used in information- and communication-theoretic analysis of wideband wireless communication systems, as described in Chapter 2 and Chapter 3.

1. We argue in Section 2.5.2 that a stochastic channel model with JPG taps is only sensible for band-limited channels, as only in this case we can hope to have enough averaging for the JPG distribution to be justified by the CLT. If bandwidth becomes large, one could suspect that less and less paths contribute to each channel tap, which renders the assumption of proper Gaussian tap distributions questionable (Zhang et al., 2002; Cassioli et al., 2002; Molisch et al., 2003a). Therefore, one objective of the measurement campaign is to assess the validity of the assumption that channel taps are marginally proper Gaussian distributed.

2. The US assumption is particularly convenient for the analysis of wideband wireless channels, especially in the form of the discrete-time US assumption. As discussed in Section 3.1, this assumption is questionable because any band limitation necessarily introduces correlation. Furthermore, the frequency dependence of individual propagation paths discussed in Section 2.3.2 and evidenced in (2.12) also contradicts the US assumption.
3. The US assumption on the continuous-time channel predicts that the number of DOUs in a band-limited discrete-time representation of the channel scales linearly with bandwidth. Such a linear scaling is assumed in many theoretical analyses of wideband channels; it is fundamental in the sense that the achievable rates over such channels in the limit for infinite bandwidth and under a constraint on the peak transmitted power are limited to values less than the capacity of the corresponding AWGN channel (Viterbi, 1967; Abou-Faycal et al., 2001; Verdú, 2002); under certain types of peak constraints, the achievable rates may even vanish at large bandwidth (Telatar and Tse, 2000; Médard and Gallager, 2002; Subramanian and Hajek, 2002). As the continuous-time US model is conceptual only and always needs to be band limited for the model to have operational significance, it is important to ask over what range the linear relation between bandwidth and DOUs holds. A sublinear scaling would indicate either that the small fractional bandwidth assumption (2.7) becomes invalid starting at a certain bandwidth or that the number of uncorrelated scatterers in the environment is limited. In the latter case, we would also expect that the Gaussian assumption discussed previously does not hold because of limited averaging at large bandwidth.

In addition to these primary objectives, several secondary considerations also influenced the design of the measurement campaign. (i) As apparent in Table 4.1 on page 93, most measurement campaigns before 2005 focussed on residential and office environments, mainly because these environments were the envisioned deployment sites

for UWB systems conforming to the IEEE 802.15.3a standard then under discussion, i.e., for high-rate short-range WPANs like wireless universal serial bus (USB) or wireless media distribution. But UWB communications might turn out to be even better suited for other applications, notably in sensor networks for building management and automation like air conditioning and lightning, where extremely low power consumption and low cost are key factors (Rabaey et al., 2002). Thus, we decided to leave the standard office setting and perform measurements in a public space. (ii) At the time we took the measurements, all reported results were for distances between transmitting and receiving antennas of less than 20 m, often significantly less. We aimed at an antenna separation between 21 m and 27 m to provide a hopefully welcome addition to the body of available measurements.

4.3.2. Measurements

With the outlined objectives in mind, we conducted two different measurement campaigns, both in the frequency band from 2 GHz to 5 GHz in the same indoor public space, the lobby of the ETZ building at ETH Zurich.

- In measurement campaign I (MCI), the channel was static and we used a VNA to sample it spatially. This type of measurement setting is similar to, e.g., a wireless access point scenario or a peer-to-peer setting where one of the terminals moves with respect to the static environment. Hence, MCI allows us to assess the impact of the corresponding spatial variations of the channel.
- In measurement campaign II (MCII), the transmitting and receiving antennas were fixed, and channel variation was induced by people moving in the environment. This type of antenna placement might represent the situation in a WPAN or a wireless sensor network, where terminals are typically immobile, like for building automation. We excited the channel with a PN sounding signal and used a DSO to record the channel output.

For completeness, we briefly mention some external circumstances that might have influenced the two measurements and subsequent sta-

tistical analysis. We took actual measurements for MCI in June 2004. At this time, it was not foreseen to conduct an additional measurement campaign. Only during postprocessing and statistical evaluation it became apparent that the number of channel samples was not sufficient to compute certain statistics,* and that a sufficient number of samples could not be obtained by spatially sampling the channel. Temporal sampling seemed to be much better suited for this task; to this end, we devised MCII. Naturally, experience gained during MCI is reflected in MCII. By chance, we had access to a very advanced DSO for a brief period of time; this external circumstance determined the timing of MCII, which was conducted over three days in December 2004.

4.3.3. Limitations

The design of a channel measurement campaign is contingent on a channel model; conversely, the type of channel model that can be characterized from measurement data depends on what can actually be measured. We presuppose a linear channel with an IO relation that is identifiable; consequently, we use the measurement techniques described in Section 4.1.2. As already discussed in the same place, these measurement techniques are subject to technical constraints, which in turn limit the type of channel model we can characterize.

- The VNA we had available for our measurements required a significant amount of time to measure a single transfer function; this precluded any meaningful time variability in the channel. With such a setup, the large number of channel realizations that are necessary for a statistical channel characterization can only be obtained over space, i.e., from local displacements of either the transmitting or the receiving antenna, while the environment needs to be kept static. Consequently, only a model of the block-fading type can be characterized with such a setup.

* Ideally, such a problem should be detected *before* any measurements are taken. Yet, many of the statistical methods described in Chapter 5 and Chapter 6 have not been used before to analyze channel measurement data, so that there was no past experience on the basis of which to decide on certain parameters of MCI.

- The DSO used for the time-domain measurements allowed for time variability of the channel while measurements were in progress. Yet, its high-speed acquisition memory could only store up to four consecutive repetitions of the PN sequence used to probe the channel, and the transfer of the recorded output signal to secondary storage took about one second each time. But a one second delay is way too long for the channel to remain constant, so that consecutive recordings could not be used to characterize channel variation over time but only for statistical purposes. Again, the resulting channel model is confined to be block fading.
- Measurements taken with a VNA are discrete in frequency; hence, the corresponding impulse responses are periodic in time, with the period determined by the maximum number of points that can be measured in a given frequency band. This places a hard limit on the maximum delay in the channel, i.e., we have to make the a priori assumption that delay is limited. Similarly, the repetition rate of the transmitted PN sequence in the time-domain setup strictly limits the maximum delay that can be measured.
- Both measurement setups used a single transmit antenna and a single receive antenna. Although the spatial displacement of one antenna in the frequency-domain measurement system yields transfer functions that contain spatial information about the channel, they can either be used to characterize spatial properties or for statistical purposes, but not for both; i.e., our measurement setup does not allow us to obtain a statistical characterization of a multiantenna channel.

To summarize: our measurement setups, both for frequency-domain measurements and for time-domain measurements, only allow us to characterize a single input single output (SISO) block-fading channel model with a finite number of channel taps, i.e., a channel with IO relation

$$\mathbf{y}[k] = \sum_{l=0}^{L-1} \mathbf{h}[l]\mathbf{x}[k-l] + \mathbf{w}[k]. \quad (4.6)$$

4 WIDEBAND CHANNEL MEASUREMENTS

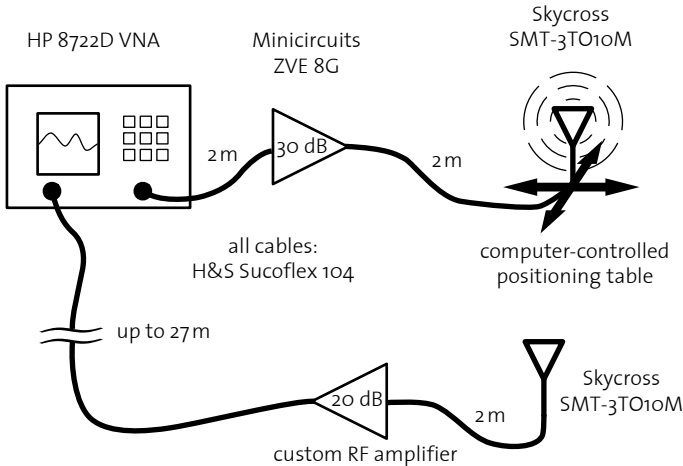


Fig. 4.1: The measurement system used in MCI.

Consequently, we are mainly concerned in the following chapters with measurement and statistical analysis of the L -dimensional channel vector

$$\mathbf{h} \triangleq [\mathbf{h}[0] \ \mathbf{h}[1] \ \dots \ \mathbf{h}[L-1]]^T. \quad (4.7)$$

4.4. MEASUREMENT SETUP AND EQUIPMENT

4.4.1. Measurement Campaign I

The frequency-domain measurement system used in MCI is shown schematically in Figure 4.1. The goals in the design of this measurement system were to achieve sufficient dynamic range up to the maximum antenna separation of about 27 m, to automate most of the measurement tasks, to guarantee reproducibility of the results, and to aim for the highest possible fidelity of the recorded channel

transfer functions.* Our measurements spanned the frequency band from 2 GHz to 8 GHz, limited by several components in the measurement setup. Unfortunately, the channel bandwidth that could be measured in the time-domain MCII turned out to be smaller, only 2 GHz–5 GHz. To allow a direct comparison of measurements from both campaigns, we decided to use MCI data only for the band from 2 GHz to 5 GHz. The following subsections provide a detailed description of the measurement setup in Figure 4.1.

A. Vector Network Analyzer

The VNA used in MCI was a HP 8722D with firmware version 6.04, dated 07/29/1996. This VNA is capable of measuring S-parameters from 50 MHz to 40 GHz. Of interest for our purposes is the parameter vector S_{12} , the transfer function of the channel at discrete frequency points. A maximum of 1601 points can be recorded in any given frequency band. The option 012, “Direct Sampler Access,” allows to bypass the input and output couplers to directly access the sampling circuits; this configuration significantly increases the dynamic range of the measurements compared with the standard setup. In this configuration, the sampling circuits are no longer protected from overvoltage and DC offset; hence, special care must be taken when handling RF components. As a protective measure, we inserted a 3 dB attenuator and a blocking capacitor at the return port, as shown in Figure 4.2. In the configuration for increased dynamic range, crosstalk from the internal reference channel of the VNA becomes a problem. Therefore, we attenuated this reference channel by 23 dB. All MCI measurements used the following configuration of the VNA:

- We measured the band from 2 GHz to 5 GHz and additionally the band from 5 GHz to 8 GHz. For each band we obtained 1601 equispaced points in frequency at a point spacing of 1.875 MHz.

* The measurement system was designed in cooperation with H.-R. Benedickter from the Laboratory for Electromagnetic Fields and Microwave Electronics at ETH Zurich. We greatly appreciate his expertise and experience with RF measurement and instrumentation.

4 WIDEBAND CHANNEL MEASUREMENTS

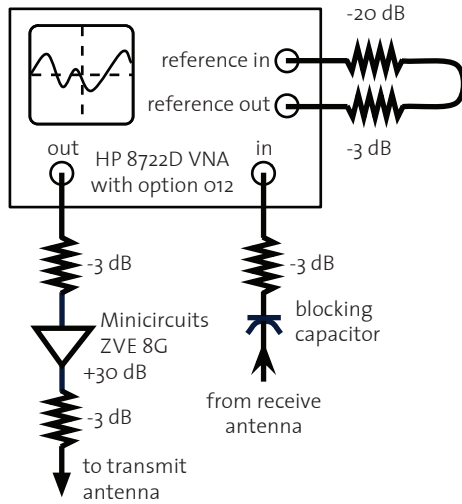


Fig. 4.2: Configuration of the HP 8722D VNA for direct sampler access.

However, we only used the lower band from 2 GHz to 5 GHz in all subsequent analyses.

- Output power was set to -5 dBm.
- All transfer functions were measured in stepped frequency mode. This means that the stimulus signal was synthesized separately for each discrete frequency point instead of being continuously swept over the whole band. Operation in stepped frequency mode was necessary to avoid phase errors that may result because of the long delay introduced by the propagation channel (Street et al., 2001).
- We set the intermediate frequency (IF) bandwidth to 300 Hz to minimize the noise floor while maintaining an acceptable measurement duration. The IF bandwidth determines the SNR—the smaller the IF bandwidth, the less noise is present in the measured signal. But a small IF bandwidth also means that the duration of the measurement signal needs to be long. We did

not use the averaging functionality of the VNA, only one transfer function was measured for every spatial location of the transmit antenna; every such individual measurement, comprising 1601 points in frequency, took 9.5 s.

To capture all significant multipath components in the channel, the frequency domain sampling rate should be set to at least twice the coherence bandwidth. Worst-case delay spreads previously reported for UWB indoor measurements were around 30 ns, so that the estimated coherence bandwidth of the corresponding channels is larger than 3.5 MHz. Hence, we deemed 1601 points with a resulting point spacing in frequency of 1.875 MHz sufficient for the lobby measurements.

The VNA has built-in calibration routines that require a set of calibration measurements. We performed these measurements approximately 45 minutes after powering up the whole measurement system to ensure that all components had reached thermal equilibrium. The calibration data was saved together with the instrument state. Calibration included all cables, amplifiers and antenna feeds. The effect of the antennas, however, was not calibrated because the antenna frequency response was not equal for all azimuth and elevation angles, as evident from Figure 4.4.

B. Virtual Antenna Array

The channel has to remain constant during measurement of each transfer function; but for meaningful statistical analysis of small-scale fading effects we need the channel to change—i.e., we need to measure several realizations of the small scale fading that can be regarded as independent. The fading phenomenon is a spatial effect, caused by constructive and destructive interference of the reflected, diffracted, and scattered partial waves; hence, different channel realizations can be obtained from *spatial sampling* of the corresponding wave field: we leave the receiving antenna fixed and measure the channel for different positions of the transmit antenna within a confined area. If the points are taken as the nodes of a rectangular grid, such a setup is often

referred to as a *virtual array*. The area of the array needs to be small enough so that changes in large-scale channel parameters like path loss are insignificant across the array. At the same time, the spacing of the individual grid points needs to be large enough for the resulting spatial samples to appear independent. This independence over space is not a property of the physical channel, it is a modeling assumption we make, like the stochastic description of wireless channels is a modeling assumption (see Section 2.5). By proper design and placement of the virtual array, we have to ensure that the assumption of spatially independent channel samples describes the measured data sufficiently well. As a rule of thumb, the different transfer functions should result from the superposition of waves whose phases differ by more than π between neighboring grid points. Hence, grid spacing should be at least as large as half a wavelength at the lowest measured frequency. We chose a grid spacing of 7 cm in both dimensions; although slightly smaller than half a wavelength at 2 GHz, this spacing allowed us to maximize the number of spatial samples, as described next.

After a transfer function is recorded, the transmitting antenna needs to be moved to the next grid point before another measurement can take place. This was done by a positioning table whose two jackscrews could independently position the antenna in the two horizontal dimensions. The same apparatus was already used in the ECHO 24 channel sounder (Truffer and Leuthold, 2001). Our custom control software achieved a worst-case positioning accuracy of about 2 mm independently in both directions, which is sufficient for our purposes. With a grid spacing of 7 cm, the positioning table allowed for a 9×5 grid, 45 points in total. For statistical analysis, the largest possible number of samples is always just good enough. Therefore, we measured two times 45 spatial positions for a given distance d between transmitter virtual array and receiver by manually displacing the positioning table by about 50 cm after the first 45 samples were recorded. We term the 90 channel samples recorded this way a *measurement set*. The exact location of the two virtual arrays that make up one measurement set are shown in Figure 4.7; we describe the measurement environment in more detail later on.



Fig. 4.3: Prototype of the Skycross SMT-3TO10M antenna. This antenna was used in all MCI and MCII measurements.

The cart could not be positioned with high accuracy, but as we are only interested in spatial samples for statistical analysis and not for estimating directional parameters, this does not incur any penalty. The transmitting antenna was attached to the positioning apparatus at a height of 1.9 m above the floor. To prevent mechanical oscillations of the antenna during repositioning, we enclosed the antenna feed in a stabilization and damping construction made of plastic and foam.

C. Antennas

Both transmitter and receiver were equipped with a prototype version of the Skycross SMT-3TO10M meander-line antenna shown in Figure 4.3. The SMT-3TO10M is a commercially available off-the-shelf antenna of small size and low cost, like it might be used in many envisioned UWB transceivers. We used this application-grade antenna instead of a high-precision measurement device because we are mainly interested in modeling the baseband channel, which includes the an-

4 WIDEBAND CHANNEL MEASUREMENTS

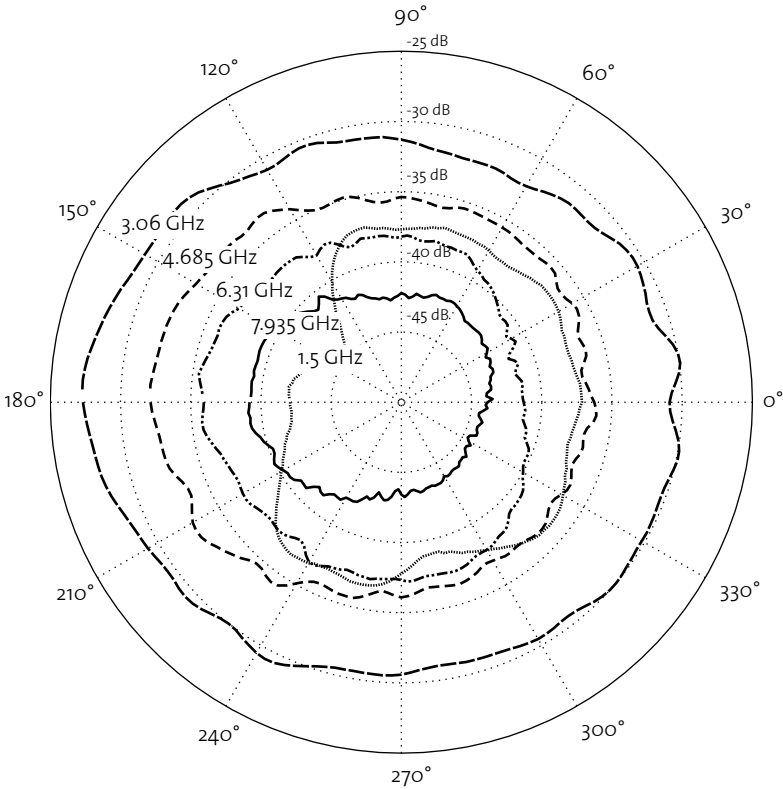


Fig. 4.4: Azimuth response of the Skycross SMT-3TO10M antenna prototype. Measurements courtesy of T. Zasowski and H.-R. Benedickter.

tenna effects. The radiation pattern* of the selected antenna in the horizontal plane is not perfectly omnidirectional, but still reasonably good over the band of interest, as can be seen in Figure 4.4. All antennas were fed through a custom-made 20 cm semirigid feed with

* All antenna measurements were performed by H.-R. Benedickter and T. Zasowski in the anechoic chamber of the Laboratory for Electromagnetic Fields and Microwave Electronics at ETH Zurich. The reference antenna was a Watkins Johnson 2 horn. We did not remove the effect of the reference antenna.

almost perfectly flat frequency response over the band of interest.

D. Amplifiers

The output power of the VNA was insufficient to obtain an SNR at the receiver that was high enough for sensible statistical analysis. Therefore, we used two different external amplifiers: (i) We placed a Minicircuits ZVE-8G wideband power amplifier that provides 30 dB gain in the band from 2 GHz to 8 GHz right before the transmit antenna to boost the power of the transmitted signal. (ii) At the receiver side, we followed the receive antenna with a custom RF amplifier with bandwidth of over 10 GHz and a noise figure of less than 6 dB over the whole band.

E. Cables and Connectors

All cables used in the measurement chain were H&S Sucoflex 104 assemblies in different length. The long cable from the receiving antenna back to the VNA was composed of several shorter pieces.* In particular, we used the following composition: 7 m + 3 m + 3 m + 3 m + 2 m + 3 m + 7 m.

Cables are a crucial part of every measurement setup because bending and torsion can lead to degradation of their frequency response. The H&S cables used are certified up to 24 GHz and have low loss. During the measurements, the cables experienced all sorts of mechanical stress while moving the antennas, especially during the calibration procedure where the far end of the long cable needed to be moved all the way back to the VNA for the back-to-back calibration measurement. Tests showed that this mechanical stress had some impact on the measured frequency responses, depending on the way the cables were bend and laid out on the floor. In addition, there was an observable effect due to surface waves traveling along the outer shielding of the cables. Hence, although great care was exercised to achieve

* This composition was necessary as we did not have longer single-piece high-quality assemblies in our laboratory stock.

the lowest possible distortion in the overall measurement setup, the cables constituted the weakest part of the presented channel sounding system.

F. Measurement Control and Automation

All measurement functions were controlled from a laptop computer, which was connected to the VNA via an IEEE 488.2 interface and to the positioning table via the IEEE 1284 parallel port. Control and data acquisition routines were written in Matlab, using the Data Acquisition and Instrument Control Toolboxes. The measured frequency responses were saved as three-dimensional Matlab arrays.

4.4.2. Measurement Campaign II

We adopted the pulse compression measurement principle described in Section 4.1.3; our sounding signal was a PN sequence. A DSO directly acquired the output signal, i.e., without prior downconversion of the received signal. A schematic of the measurement setup used in MCII is shown in Figure 4.5; we describe its core components in the following subsections.

Whereas the VNA in MCI handles synthesis of the transmit signal, amplification, sampling and digitization of the received signal, calibration, and various control tasks, several of these tasks have to be carried out by separate devices in MCII because no integrated time-domain measurement system for channels of several gigahertz bandwidth is yet available. A major challenge in MCII was to establish a common time reference at the transmitter and the receiver side. This problem did not arise in MCI because the VNA houses transmitter and receiver, and both use the same clock signal. To extract accurate timing information also in the time-domain measurement setup, we recorded the output signal of a dedicated wireline *reference channel* in addition to the output signals of the wireless channel under study. The reference channel conveyed the same signal used to sound the wireless channel, but over a cable instead of “over the air.”

4.4 MEASUREMENT SETUP AND EQUIPMENT

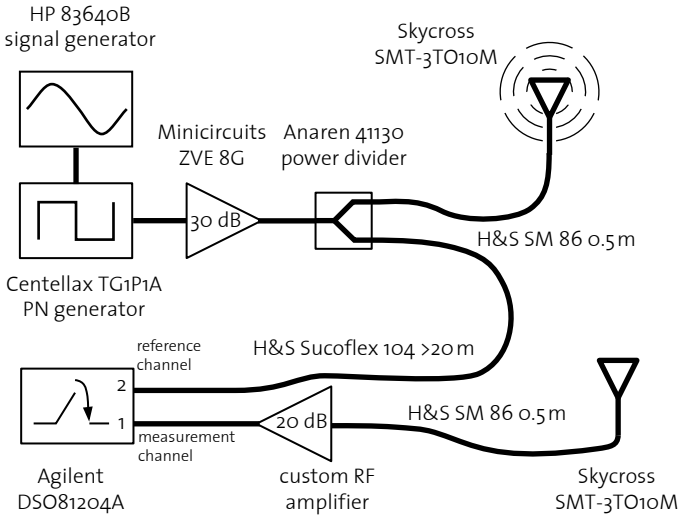


Fig. 4.5: The measurement system used in MC II.

A. Signal Generation

The ideal sounding signal in a pulse-compression measurement system has a Dirac impulse as autocorrelation function and a correspondingly flat spectrum. Any realizable signal can only approximate this ideal, though. We used a PN sequence as sounding signal, generated by a Centellax TG1P1A pseudorandom bit sequence generator. This device converts a clock input signal into a periodically repeated PN sequence, generated by a maximum-length shift register. Out of the choices of maximum-length shift register sequences offered by the device, we opted for the one of length $2^{15} - 1 = 32767$ with generator polynomial $x^{15} + x + 1$. The sounding signal was transmitted continuously at a clock rate of 10 GHz. This resulted in a repetition period of the PN sequence of $3.2767 \mu\text{s}$, equivalent to a repetition rate of 305 185 kHz. The duration of the chosen sequence is such that it can comfortably fit the available acquisition memory of the DSO, as described below. We used a HP 83640B signal generator as source for

4 WIDEBAND CHANNEL MEASUREMENTS

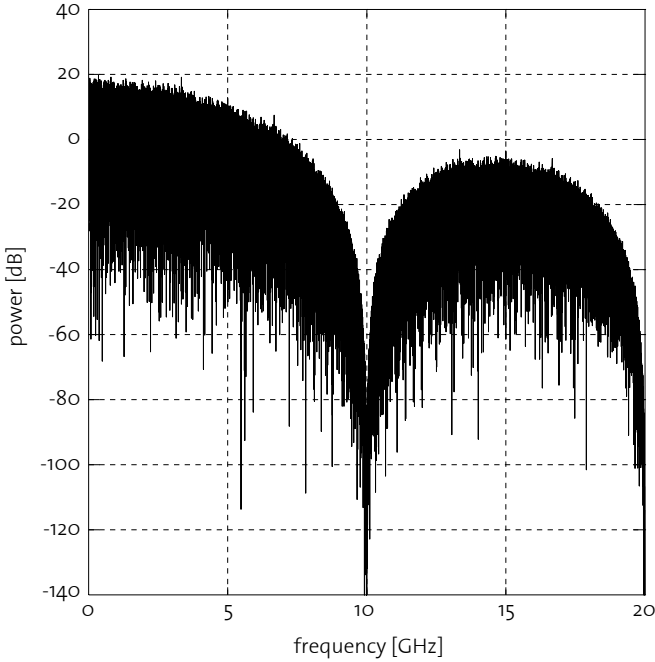


Fig. 4.6: DFT power spectrum estimate of the PN sequence with generator polynomial $x^{15} + x + 1$, clocked at 10 GHz.

the 10 GHz clock input; its clock signal is extremely stable and has very low phase noise. The PN sequence generator itself is specified to have root mean square (RMS) jitter of 0.9 ps at 10 GHz clock frequency. A discrete Fourier transform (DFT) estimate of the sequence's power spectrum is shown in Figure 4.6. Clearly visible is the deviation from an ideally flat spectrum. The drop in magnitude with increasing frequency prompted us to limit the analysis of all MCII measurement data to frequencies below 5 GHz.

B. Digital Sampling Oscilloscope

The core component of the receiver in MCII was an Agilent Technologies DSO81204A DSO.* This device can simultaneously acquire two channels of 12 GHz analog bandwidth at a sampling rate of 40 GHz and 8 bit resolution. The signal received “over the air” was fed to the first DSO channel, which we call the *measurement channel* in the following; the second channel of the DSO, termed the *reference channel*, was used to acquire samples of the reference signal for timing recovery; this signal was the transmitted PN sequence, but conducted over a long cable rather than received over the wireless channel.

Two million samples can be acquired and stored in real time in the DSO’s acquisition memory. After every successful acquisition, the data in the acquisition memory need to be transferred to the DSO’s internal hard drive before another acquisition can take place. Data transfer is rather slow, which severely limits the acquisition rate and prevents characterization of any meaningful time-variation in the channel. For a sustained acquisition rate of about one measurement and one reference snapshot per second, we had to limit acquisition to 400 000 samples per channel, i.e., to 10 μ s of output signal per acquisition. We call the 400 000 points of an individual measurement channel acquisition a *channel snapshot*, and the corresponding points from the reference channel a *reference snapshot*.

C. Other Components

Although the measurement methodology in MCII was different from the one in MCI, we used the same antennas and amplifiers as already described in the previous section. A modification to the antenna feeds was to enclose them in clamp-on ferrite beads; this prevents reflections from the antenna to travel along the cable and radiate.†

* We greatly acknowledge the support of Agilent Technologies to provide us with a DSO for MCII that, at the time of the measurement campaign, was probably one of a few DSOs worldwide with the stated capabilities.

† Unfortunately, we were too late to realize the problem of parasitic transmission from cables in MCI and used the ferrite beads only in MCII.

The cable used to conduct the reference signal from the signal source to the DSO was the same as the cable used in MCI. For accurate timing, we used phase-stable semirigid H&S Sucoform 86 cables to connect the antennas to the transmit and receive amplifiers. The DSO was controlled over an IEEE 488.2 interface from a laptop computer running Matlab, as in MCI.

4.4.3. Measurement Environment

We conducted both channel measurement campaigns in the lobby of the ETZ building at ETH Zurich. The building itself is a typical early-1980s multistory office building, with the lobby located on the ground floor of the building. A floor plan of the relevant part of this floor is shown in Figure 4.7. The lobby stretches from a conference room at the bottom in Figure 4.7 to an office in the top right corner. On the right, the whole length of the lobby is enclosed by large windows; on the left are four concrete structures. The bottom-most one houses service rooms; its walls facing the lobby are plated with metal sheets. The second structure, counted from the bottom of Figure 4.7, is a concrete staircase; it is partially enclosed behind a row of vertical metal bars. The third structure houses elevators pits and service rooms; its concrete walls are again plated with metal sheets. The topmost structure, next to the entrance doors, is an office enclosed behind concrete walls, with a counter facing downward. Further to the left of the four structures just described are hallways that lead to classrooms and offices. A row of reinforced concrete pillars runs through the whole lobby in parallel to the windows, about two thirds between windows and walls. The floor of the lobby is tiled with large natural stones, the ceiling is of poured concrete; a rectangular structure that carries neon lights is suspended from it. Furniture in the lobby comprises mainly tables and chairs, some vending machines, and decorative plants, as shown on the photograph in Figure 4.8.

4.4 MEASUREMENT SETUP AND EQUIPMENT

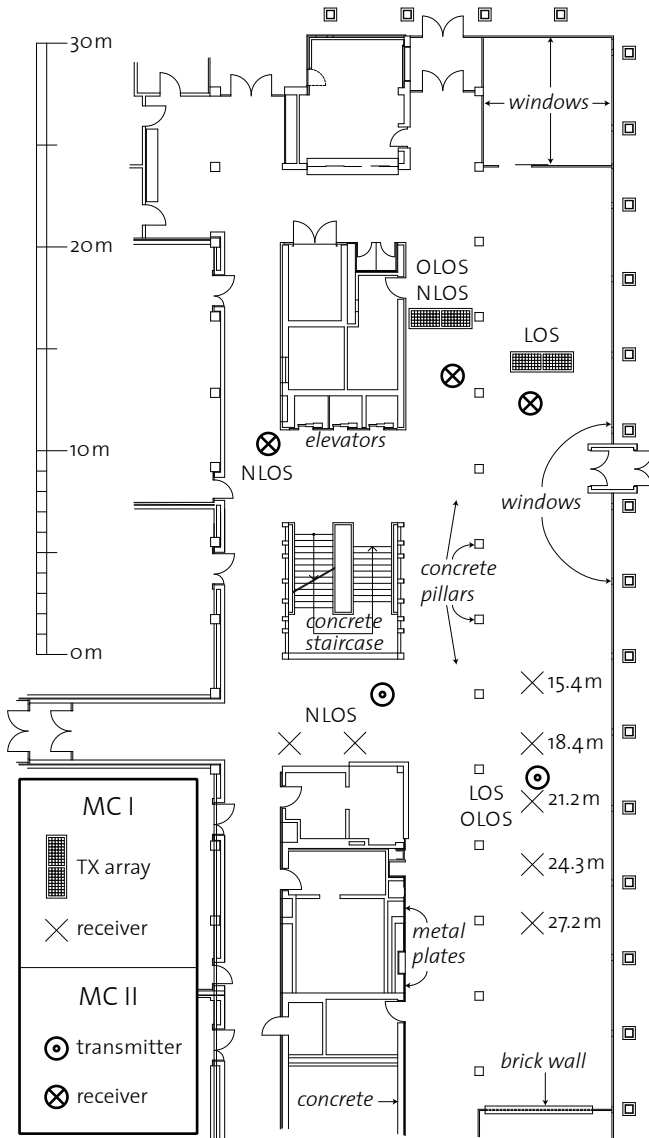


Fig. 4.7: Floor plan of the lobby in the ETZ building, ETH Zurich.

A. Measurement Campaign I Locations

Frequency-domain measurements are contingent on a static channel because of the time required by the VNA to record a single transfer function, as discussed in Section 4.1.4. To ensure that the channel did not change during acquisition, we performed all measurements at times when there were no people present in the lobby, i.e., during two successive nights on a weekend. Nevertheless, some nocturnal graduate students interrupted the measurement process at times, so that we had to repeat the affected measurements. Other sources of channel variation, like fluorescent lights (Truffer and Leuthold, 2001, Section V) and IEEE 802.11b base stations, were switched off.

We placed transmitter and receiver at different locations throughout the lobby. The transmitting antenna was mounted atop the positioning table to form a virtual array, as detailed in Section 4.4.1.B; the receiving antenna was placed on a tripod. Antenna heights above the floor were 1.9 m for the transmitter and 1.75 m for the receiver.

Conforming with well-established channel measurement terminology, we distinguish between three different *measurement settings*.

- In a *line of sight* setting, the line of sight (LOS) between transmitter and receiver is not blocked by any object. In the MCI LOS setting, we took several *measurement sets* of $N = 90$ transfer functions each, along the length of the lobby. Each measurement set is characterized by the distance d between the transmitter virtual array and the receiving antenna. The virtual array remained in the same location in the upper part of the lobby for all LOS measurement sets, while we placed the receiving antenna at $d = 15.4$ m, $d = 18.4$ m, $d = 21.2$ m, $d = 24.3$ m, and $d = 27.2$ m distance as indicated in Figure 4.7. Thus, we took five LOS measurement sets of $N = 90$ spatial channel samples each, for a total of $5 \times 90 = 450$ LOS transfer functions in MCI.
- *Obstructed line of sight (OLOS)* means that the LOS between transmitter and receiver is partially shadowed by objects, in our case by the row of concrete pillars in the lobby. These pillars are small enough for the transmitted waves to be diffract around

them instead of being shadowed. Hence, channel variation over the extent of the virtual array can be seen as small-scale fading. While we displaced the transmitting array with respect to its LOS position to create the OLOS setup, we used three of the five receiver positions from the LOS measurements also for the OLOS measurements; hence, we recorded three OLOS measurement sets of $N = 90$ spatial channel samples each. The resulting distances d between transmitter array and receiver antenna in the OLOS measurement setting are $d = 21.7$ m, $d = 24.6$ m, and $d = 27.5$ m.

- *Non-line of sight (NLOS)* characterizes measurement settings where the LOS between transmitting and receiving antenna is completely blocked by a large object. In MCI, this large object was the concrete staircase behind which we positioned the receiving antenna at two different spots for a total two measurement sets and correspondingly 180 measured transfer functions. We left the transmitter array in the same spot as for the OLOS measurements; the distances to the receiving antennas were approximately $d = 22.2$ m and $d = 22.7$ m. Unfortunately, we only realized upon postprocessing that not all NLOS impulse responses are alike, but that probably shadowing and diffraction around the corner of the elevator pit resulted in some quite strong and other rather weak impulse responses, depending on the location of the transmit antenna on the virtual array. Such large-scale variations of channel samples in a single measurement set are not suitable for statistical analysis of small-scale fading; therefore, we use MCI NLOS data only in Section 4.6 for exploratory purposes but not in the statistical analysis in Chapter 5 and Chapter 6.

B. Measurement Campaign II Locations

As we kept both the transmitting and the receiving antenna fixed during a sequence of channel acquisitions in MCII, variation in the environment was solely due to people moving about. The lobby connects classrooms on the left-hand side in Figure 4.7, a cafeteria reachable via the doors at the top, and a patio and further classrooms behind

the door to the right; because we conducted MCII at daytime during a normal week of lectures, many people passed through the lobby in the course of our measurements. A rather annoying source of channel variation was introduced by operational IEEE 802.11b base stations, which, unfortunately, we could not deactivate during the day. The impact of the resulting narrowband interference is discussed in detail in Section 4.5.2.

For ease of comparison, we chose similar measurement locations as in MCII, although we reduced the distance between transmitter and receiver to obtain a higher SNR in anticipation of the more complicated postprocessing steps necessary for the time-domain measurement data. Compared with MCI, we also had to interchange the position of transmitter and receiver, as indicated in Figure 4.7. Both antennas were mounted on tripods, approximately 1.6 m above the floor.

The main objective in MCII was to record a sufficient number of channel samples for sophisticated statistical analysis. As the acquisition rate was limited by the data transfer between the DSO's acquisition memory and its hard disk, we elected to reduce the number of measurement sets to only one set for each measurement setting, LOS, OLOS, and NLOS; at the same time, we greatly extend the number N of channel samples in each measurement set compared with MCI. Hence, in MCII a measurement set is synonymous with a measurement setting.

- In the LOS setting, we recorded $N = 1011$ snapshots. The distance between the transmitting and receiving antenna was $d \approx 20$ m.
- The distance between the antennas in the OLOS setting was also $d \approx 20$ m. In this setting, we recorded $N = 2722$ snapshots.
- Compared with MCI, we placed the receiver in a different spot in the NLOS setting, right behind the concrete staircase next to the elevators. The reason for this change was the expected intense pedestrian traffic between stairs and elevators, and thus high channel variation. The distance between the two antennas of $d \approx 13$ m was smaller than in the LOS and OLOS settings. We recorded $N = 1256$ channel snapshots.



Fig. 4.8: Impression of the lobby from the position of the MCII transmitter in the LOS and OLOS setting.

Figure 4.8 shows the lobby from the vantage point of the MCII transmitter in the LOS and OLOS setting. The line of vision is in the direction of the office in the top right corner in Figure 4.7.

4.5. POSTPROCESSING

The raw data generated by the measurement devices in MCI and MCII are not suited for direct analysis, they need to be processed first. This is a delicate task as it is contingent on several modeling assumptions. It is a subjective task as well, because the investigator can easily introduce significant bias without any intention to malevolently manipulate the data (Gould, 1981). As such experimenter bias seems to be unavoidable to a certain extent, we can only try to explain all postprocessing steps in detail, so that any omissions, additions, or erroneous manipulation

can be detected easily. Furthermore, we make the raw measurement data available for third parties to analyze and compare with our findings (Schuster and Bölcskei, 2005b).

Although the raw measurement data from MCI and MCII are quite different, our goal is to use the same statistical methods for both data sets, as only then a valid comparison between the two types of channels might eventually be possible. Thus, the goal of the postprocessing steps described in the following is to transform the measured transfer functions of MCI and the measured channel snapshots from MCII into a common format: a set of complex-valued baseband impulse response vectors as defined in (4.7), that describe the channel in the band from 2 GHz to 5 GHz.

4.5.1. Measurement Campaign I

Raw measurement data in MCI consists of transfer functions that are discrete in frequency. For each individual measurement, the VNA records a vector of 1601 complex-valued entries; the entry at position n represents the magnitude and the phase of the channel transfer function at frequency $2 \text{ GHz} + n \cdot 1.875 \text{ GHz}$, for $n = 0, 1, \dots, 1600$. The raw transfer functions are not completely “raw”, as the VNA already applies its calibration routines described in Section 4.1.5. Because of the two amplifiers in the signal path, we had to insert 50 dB of attenuation during the calibration measurements to compensate for the absent channel path loss. Starting from the calibrated frequency-domain vectors recorded by the VNA, we apply the signal processing steps described in the following.

A. Inverse Discrete Fourier Transform

The impulse response vectors we are interested in are time domain quantities, yet the measured transfer functions are frequency-domain quantities. Because these transfer functions are discrete in frequency and band limited, the corresponding impulse responses are periodic and discrete in time; both representations are linked by the DFT.

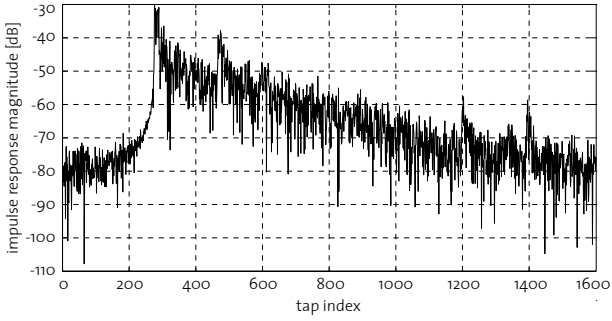
The original continuous-time channel impulse responses are not periodic; thus, discretizing a transfer function might introduce aliasing in the corresponding impulse responses. With a point spacing of $3 \text{ GHz}/1600 = 1.875 \text{ MHz}$, the impulse response vector obtained via the inverse discrete Fourier transform (IDFT) is 533 ns long, which corresponds to about 160 m propagation distance of electromagnetic waves in air. Contributions from propagation paths longer than this distance lead to aliasing in the corresponding impulse response vector. In view of the distances indicated in Figure 4.7, such long propagation paths can only result from multiple reflections or scattering; we can thus expect that they convey only a very small amount of the total received energy, at a level well below the measurement noise floor.

The impulse response vectors obtained via the IDFT are already in complex baseband form because the discrete-frequency transfer functions are band limited and centered at 3.5 GHz instead of occupying the whole band from 0 Hz up to 5 GHz. Figure 4.9 shows the magnitude of three impulse response vectors for a randomly chosen position on the virtual array in one of the LOS, OLOS, and NLOS measurement sets each. Displayed is the IDFT log-magnitude of the VNA output vector without any further processing.*

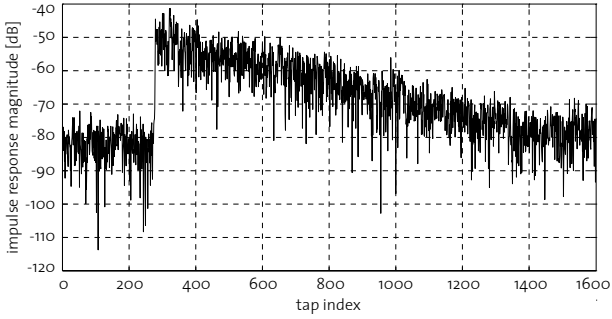
The measured channel impulse responses are band limited, i.e., they are impulse responses of the baseband channel defined in Section 2.3.1. If the goal of the channel measurement campaign is to parameterize a channel model that is inherently of infinite bandwidth, like Turin's specular block fading model discussed in Section 3.1.2, the measured transfer functions are commonly multiplied by a window function before computing the IDFT (Dabin et al., 2003; Chong and Yong, 2005; Jämsä et al., 2006). This windowing is supposed to mitigate the spectral broadening of the Dirac impulses in Turin's model that results from the strict band limitation of the measurement system. As we do not assume that the channel can be accurately modeled by

* Because we did not yet subtract the 50 dB attenuation used in the calibration process but show the raw measurement data in these plots, the scale on the ordinate does not represent the actual attenuation of the channel.

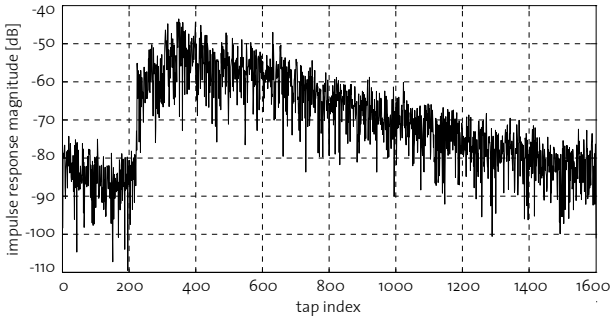
4 WIDEBAND CHANNEL MEASUREMENTS



(a) MCI LOS, $d = 27.2$ m.



(b) MCI OLOS, $d = 27.5$ m.



(c) MCI NLOS, $d = 22.1$ m.

Fig. 4.9: Measured MCI impulse response magnitudes.

a Turin-type specular model, we do not apply any window function and consider the band limitation of the measurement system as part of the channel instead.

B. Time Normalization

Displacement of the transmit antenna on the virtual array changes its distance d to the receiving antenna. This change can be sufficient to move a significant portion of the energy conveyed over a given specular path from one tap in the discrete-time impulse response to an adjacent earlier or later tap because the effective time resolution of a UWB signal is very high. For example, if the transmit antenna is displaced in LOS direction over the full length of the array, i.e., by $5 \times 7 \text{ cm} = 45 \text{ cm}$, the distance change equals a change in delay of 1.5 ns, a duration equivalent to 4.5 taps at a channel bandwidth of 3 GHz. This shifting of specular paths from one tap to another may create problems during statistical analysis. Most statistical methods rest on the assumption that the samples they operate on are independent and equally distributed, so that ensemble averages can be replaced with spatial averages in our case; i.e., we would like to treat the $N = 90$ spatial samples measured in a given measurement set as different realizations of the *same* random channel vector. But the transition of specular reflections from one tap to the next over the size of the array violates the modeling assumption of independent channel realizations. Yet, the problem of the specular paths transitioning from tap to tap is not a mathematical problem because there is no “true” underlying stochastic process according to which channel realizations are drawn (see also the discussion in Section 2.5). It is a modeling problem: we are faced with the question if the resulting channel impulse responses can be reasonably well modeled as spatially independent or not. By the present observation, this does not seem to be the case.

However, we can transform the data to alleviate this problem to some extent and make the modeling assumption of spatial independence at least partially viable. The channel we are ultimately interested in is not the propagation channel but the channel as experi-

enced by a communication system. But virtually every communication system synchronizes the receiver clock to the transmitter clock, i.e., it synchronizes the received signal with respect to the signal arrival time, which is in turn dictated by the arrival of the first path. Hence, a suitable transformation of our measured data is to align all impulse responses in a given measurement set with respect to the first arrival. This results in a set of channel samples as a receiver synchronized to the transmitter would see them. With this alignment, the problem of delay shifts of the LOS path disappears.*

Typical synchronizers align the receiver clock and the first path quite accurately, to within a fraction of the tap duration (Meyr et al., 1998), because operation at Nyquist rate is not sufficient to achieve good synchronization accuracy as the following example demonstrates. Consider time samples of a sine wave taken at Nyquist rate. The sample time that corresponds to the maximum sample value within one period of the sine wave might differ from the continuous-time instant at which the sine wave indeed reaches its maximum by up to a quarter period in both positive and negative time direction. Alignment of several taps with maximum magnitude might thus yield a phase error of the corresponding continuous-time signals of up to 180 degrees. To ameliorate this problem without excessive computation, we interpolate the measured impulse responses before the alignment step. In particular,

1. We transform the complex baseband impulse responses recorded by the VNA to passband, i.e., we shift each discrete-frequency transfer function to a center frequency of 3.5 GHz, which yields an analytic signal, and then add a conjugated and reversed copy of this frequency vector to obtain a Hermitian symmetric passband transfer function.
2. We zero-pad the measured transfer functions to eight times their original length and subsequently compute the IDFT of

* However, specular reflections different from the LOS path may still show up in several different impulse response taps, especially if they arrive at different angles than the LOS path. We discuss this effect in more detail in Section 5.5.

the zero-padded vectors. This amounts to eight-fold circulant interpolation.

3. We manually inspect each interpolated passband impulse response and note the index of the first tap whose magnitude significantly exceeds the noise floor.
4. We circularly shift all impulse responses to align the so identified taps.
5. We plot a superposition of all 90 interpolated and shifted impulse responses and, upon visual inspection, make some more adjustments as indicated, e.g., not just by the first maximum, but also by the rising edge of the first arrival.
6. We extract the discrete-time impulse responses in the band from 2 GHz to 5 GHz by a brick-wall filter and transform the resulting signals to complex baseband representation.
7. We cut off all but 20 taps before the identified start of the impulse response.
8. Finally, we decimate all aligned impulse responses by a factor of eight.

The resulting vectors are now in the form (4.7), as required for statistical analysis described in Chapter 5.

4.5.2. Measurement Campaign II

It takes more involved postprocessing steps to bring MCII snapshots into the desired form of complex baseband impulse response vectors (4.7) than for MCI transfer functions. First, the snapshots need to be correlated with the PN sequence to recover an estimate of the channel impulse response. Then, all impulse responses in a measurement set need to be aligned in time. The lack of a common clock reference at transmitter and receiver means that we need to extract a clock signal from the reference channel. Furthermore, all impulse responses need to be calibrated and converted to complex baseband form.

4 WIDEBAND CHANNEL MEASUREMENTS

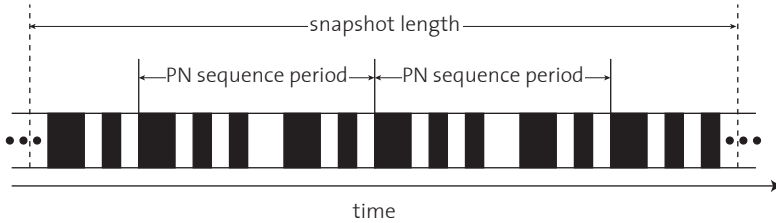


Fig. 4.10: Typical arrangement of the received PN sequences in a channel snapshot.

A. Correlation

Each channel snapshot contains the response of the channel to the periodically repeated PN sequence; hence, we need to correlate the snapshot with the PN sequence like in (4.5) to extract an estimate of the actual channel impulse response. The transmitter broadcast the PN sequence continuously at a repetition rate of over 300 kHz, a rate orders of magnitude higher than the worst-case Doppler frequency of the channel. Therefore, we can safely assume that the channel did not vary significantly over several periods of the PN sequence and the approximation (4.5) to hold.

Recorded at 40 gigasamples per second, each channel snapshot contains 400 000 points, or roughly three PN sequences. Because transmitter and receiver were not synchronized, in general only two PN sequences are recorded completely while the PN sequences at the beginning and the end of the snapshot are truncated, as illustrated in Figure 4.10. Linear correlation of the snapshot with one period of the PN sequence thus yields three distinct correlation peaks, of which only the two in the middle part of the snapshot result from a complete correlation over the entire length of the PN sequence.

B. Alignment in Time

As with the data from MCI, we assume for the purpose of statistical analysis that all snapshots in a given measurement set are realizations

of the same random channel vector. Therefore, the corresponding impulse responses need to be aligned in time. As in MCI, we interpolate all correlated snapshots for better alignment accuracy, i.e., we compute the DFT of each correlated channel and reference snapshot, pad the resulting vectors with zeros to eight times their original length, and subsequently take the IDFT. Hence, the effective sampling rate after interpolation is 320 GHz, for an oversampling ratio of 32 with respect to the maximum desired frequency component at 5 GHz; the resulting worst-case phase ambiguity after interpolation and alignment is $180^\circ/32 \approx 5.6^\circ$. We denote the correlated and interpolated channel snapshot a *raw channel sample*; similarly, we refer to the correlated and interpolated reference snapshot as a *reference sample*.

Different from MCI, there is no conceptual difficulty aligning the measured impulse responses because the transmitting and receiving antennas were fixed in each measurement set. In practice, though, alignment of all raw channel samples is a difficult task because it was not possible to synchronize transmitter and receiver. Visual alignment as used in MCI is infeasible given the large number of raw channel samples we need to process, and automatic extraction of a reliable time reference from the raw channel samples alone is not possible either, as we expect the channel to vary between different snapshots. But the reference channel—the PN sequence transmitted over a long cable—does not vary significantly during the whole sequence of measurements and can thus be used to extract a time reference. As illustrated in Figure 4.11, the first complete impulse response in a raw reference sample starts at some index a , and the first complete impulse response in the raw channel sample is offset from said reference impulse response by b taps. Both a and b are measured in impulse response taps at 320 GHz sampling rate. Because the propagation distance from transmit to receive antenna is fixed throughout a given measurement set and because the reference channel is static, the offset b stays constant for all samples in a given measurement set; we determine b visually from the first measured raw channel and reference samples. The index a , on the other hand, changes from sample to sample, because the exact start time of each channel acquisition could not be

4 WIDEBAND CHANNEL MEASUREMENTS

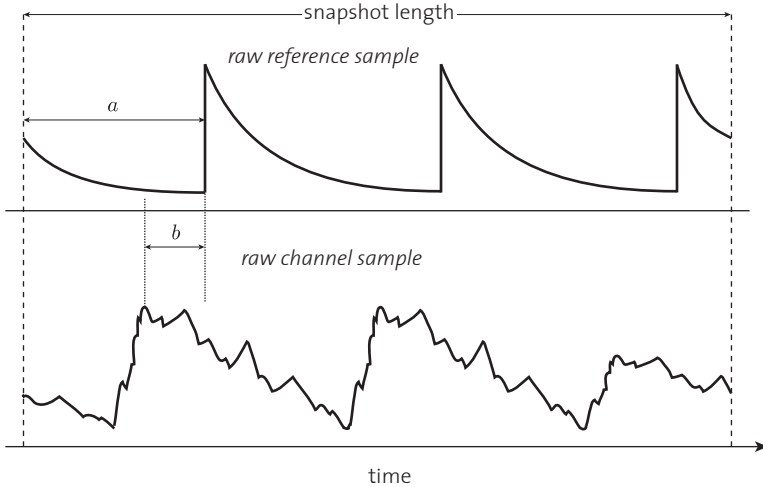


Fig. 4.11: Timing relation between raw channel and measurement samples.

synchronized to the PN sequence clock. This is where the reference channel comes in. Each impulse response in every raw reference sample is well accentuated and time invariant, because the reference channel is a wireline channel. Hence, a can be extracted automatically for each raw reference sample. With b known, we can now automatically extract and align all impulse responses in the raw measurement samples of a given measurement set.

That is, we could do so in theory. In practice, unfortunately, we are confronted with an additional impairment that we detected only during processing: a seemingly random selection of raw reference samples in each measurement set was not offset by b from the corresponding raw measurement samples, but by $b + 16$. With high probability, this extra offset is not caused by instabilities in the reference channel, as the offset does not appear in bursts of subsequent snapshots, nor does the offset error vary—it is either 16 taps (at 320 GHz) or not present at all. The only remotely plausible explanation seems to be an internal timing problem between the two channels of the DSO. Luckily,

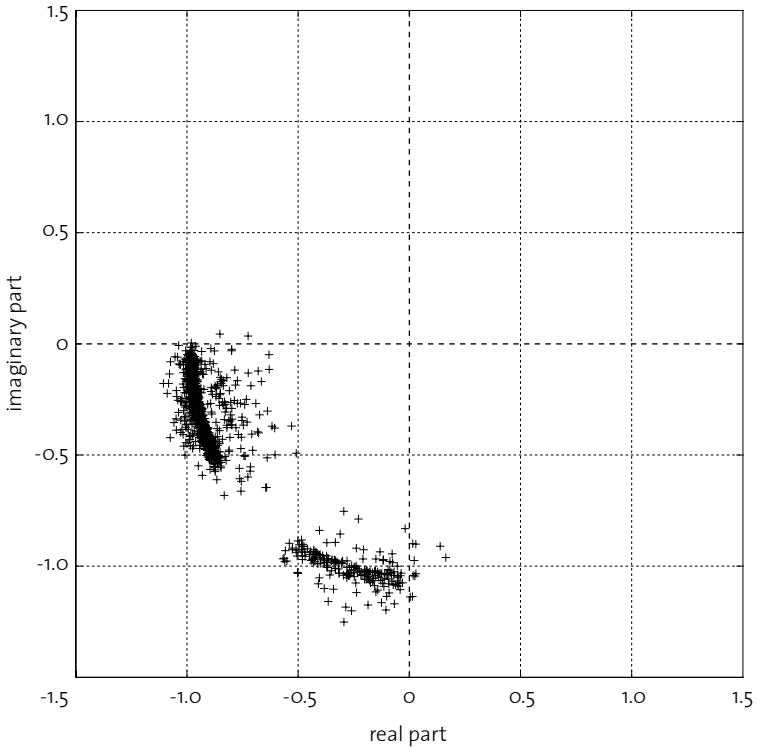


Fig. 4.12: All $N = 1011$ LOS measurements of the fourth impulse response tap in complex baseband form after alignment but without interpolation and offset correction. The lower cluster results from the timing offset error; the spread of each cluster shows the alignment error without interpolation. The average energy of the taps is normalized.

because the offset was a binary phenomenon—present or absent—we are able to correct for it automatically.

In summary, we automatically extract the index $a + b$ at which the first complete impulse response starts for each of the interpolated raw channel samples and use this index to align the interpolated impulse responses. The impact of the interpolation step and the offset

4 WIDEBAND CHANNEL MEASUREMENTS

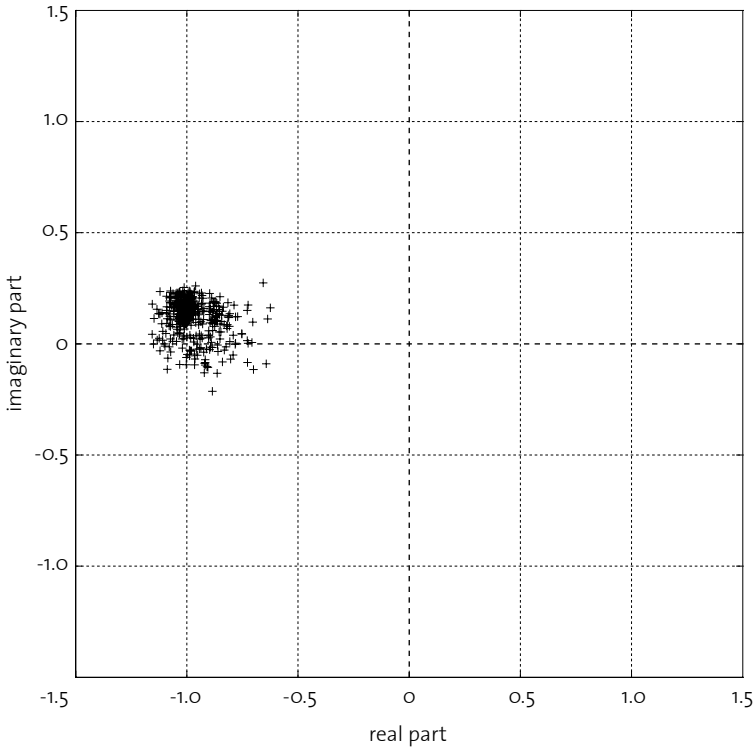


Fig. 4.13: All $N = 1011$ LOS measurements of the fourth impulse response tap in complex baseband form after alignment with interpolation and offset correction. The average energy of the taps is normalized.

correction on the alignment accuracy can be observed by comparing Figure 4.12 with Figure 4.13. Both figures show all $N = 1011$ measured realizations of the fourth tap in the LOS impulse response, after conversion to complex baseband. Alignment without interpolation is used in Figure 4.12, and the offset error just described is not compensated for, while eight-fold oversampling is used to increase alignment accuracy in Figure 4.13, and the offset error is corrected.

C. Averaging and Downsampling

Each raw channel sample contains two complete impulse responses in succession, as shown schematically in Figure 4.11. Because of the high repetition rate of the PN sequence, we can safely assume that the channel does not change significantly between these two sequences, so that we can average them to increase the measurement SNR.

Another benefit of the high PN sequence repetition rate is that we can regard the output signal as quasi-periodic, so that again the DFT is the appropriate transform between time and frequency domain for all processing steps. To obtain the desired complex baseband impulse responses, we convert each averaged interpolated impulse response to the frequency domain, extract the frequency points that correspond to the band from 2 GHz to 5 GHz, and convert the resulting frequency-domain vector back to the time domain. Although now in complex baseband representation, the resulting impulse response vectors are still sampled at 320 GHz; therefore, we decimate them to the required sampling rate of 3 GHz and truncate the length of each impulse response vector to 2001 taps.

D. Calibration

The measurement system itself has a transfer function that is not perfectly flat over the frequency band of interest. We estimate this transfer function from 100 snapshots of the measurement system connected back to back, including all amplifiers and cables but without the antennas, and use the estimate to equalize the measured data, i.e., to remove the linear distortion of the measurement system.

A second impairment we have to deal with did not result from the measurement system itself but from the measurement environment. We conducted MCII during daytime on a weekday; hence, the IEEE 802.11b base stations located at several places in the lobby could not be deactivated, and we could not prevent people in the lobby from using laptops that were equipped with IEEE 802.11b transmitters, either. The signals from these transmitters lie right in our

measurement band and result in significant interference. To suppress the effect of this interference on subsequent statistical analysis, we nulled all frequency points that correspond to the band from 2.4 GHz to 2.485 GHz.

At the end of all postprocessing steps, the resulting time-domain complex baseband impulse response vectors are now in the form (4.7). Three such vectors, randomly chosen from the LOS, OLOS, and NLOS measurement set, are shown in Figure 4.14. The absolute values on the ordinates are incommensurate with the actual received power levels,* but their relative scale shows the dynamic range of the MCII measurement setup.

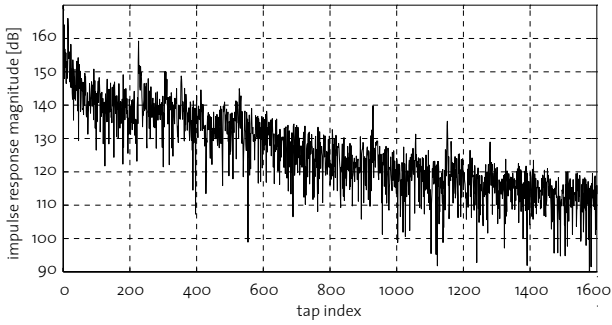
4.6. DATA EXPLORATION

Before proceeding to statistically analyze the measurement data, we take a phenomenological approach to obtain some feeling for the measurement results. This data exploration is important as the resulting understanding of the data sets provides a link between the physical channel and its abstract representation in the form of impulse response vectors.

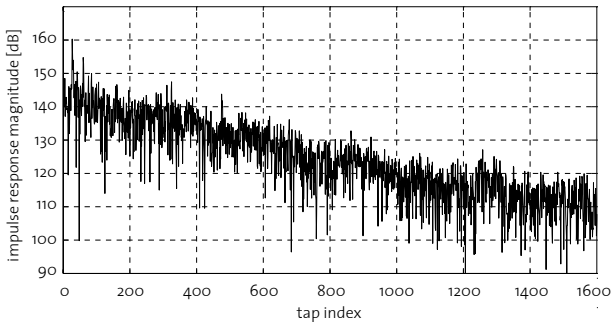
The first observation from the individual MCI impulse responses in Figure 4.9 and MCII impulse responses in Figure 4.14 is that both measurement systems provide for a sufficiently high SNR, between 30 dB and 40 dB in MCI for a separation of up to 27.5 m between transmitting and receiving antennas, and an SNR in the same range for MCII, though for smaller separation between transmitter and receiver of only up to 20 m. These numbers are rough estimates only; for more accurate SNR estimates we need to average over the individual impulse responses that correspond to a given measurement set.

* We did not normalize power in the postprocessing steps because the absolute magnitude is irrelevant for the following statistical analysis.

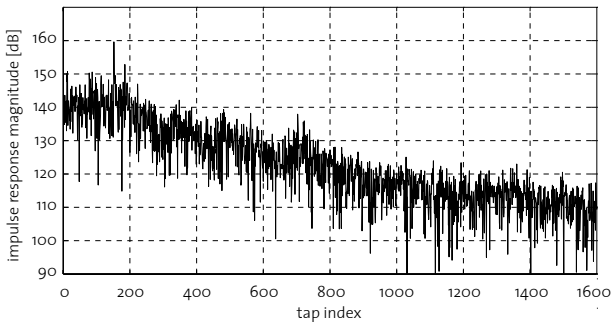
4.6 DATA EXPLORATION



(a) MC II LOS, $d = 20$ m.



(b) MC II OLOS, $d = 20$ m.



(c) MC II NLOS, $d = 13$ m.

Fig. 4.14: Measured MC II impulse response magnitudes.

4.6.1. Power-Delay Profiles

Let the n th measured complex baseband impulse response vector in a given measurement set be denoted $\mathbf{h}_n \triangleq [h_n[0] \ h_n[1] \ \cdots \ h_n[L-1]]^T$, where $h_n[l]$ is the l th tap of this complex baseband impulse response. The PDP for a given measurement set is defined as

$$\mathbf{p} \triangleq [p[0] \ p[1] \ \cdots \ p[L-1]]^T \quad (4.8a)$$

with

$$p[l] \triangleq \frac{1}{N} \sum_{n=0}^{N-1} |h_n[l]|^2. \quad (4.8b)$$

The PDP averaged over all $N = 90$ impulse response samples in the MCI LOS set for $d = 27.2$ m is shown in Figure 4.15.* For convenience, the abscissa is indexed in terms of propagation distance instead of delay, i.e., it reads measured delay divided by the speed of light. Figure 4.15 shows the distinct structure commonly encountered in PDPs from UWB measurements: several clusters protrude from the otherwise linearly (in dB) decaying PDP. The effect of the LOS path is clearly visible, and the propagation distance that corresponds to its arrival time is congruent with the physical distance d by which we separated transmitting array and receiving antenna in this measurement set.

Many clusters can be attributed to specular reflections. For a specular reflection to occur, the reflecting surface must be smooth and its dimension much larger than a wavelength. The propagation distance of 47 m that corresponds to the second cluster peak can be linked to a path that is reflected from the large window pane separating the office in the top right corner in Figure 4.7 from the lobby. The third cluster peak, with a propagation distance of 59.9 m, probably stems from a reflection at the outside window of the same office; the corresponding propagation path would have been refracted by the

* Compared with Figure 4.9, the ordinate is now correctly calibrated, i.e., the 50 dB line attenuation during the calibration measurements is taken into account.

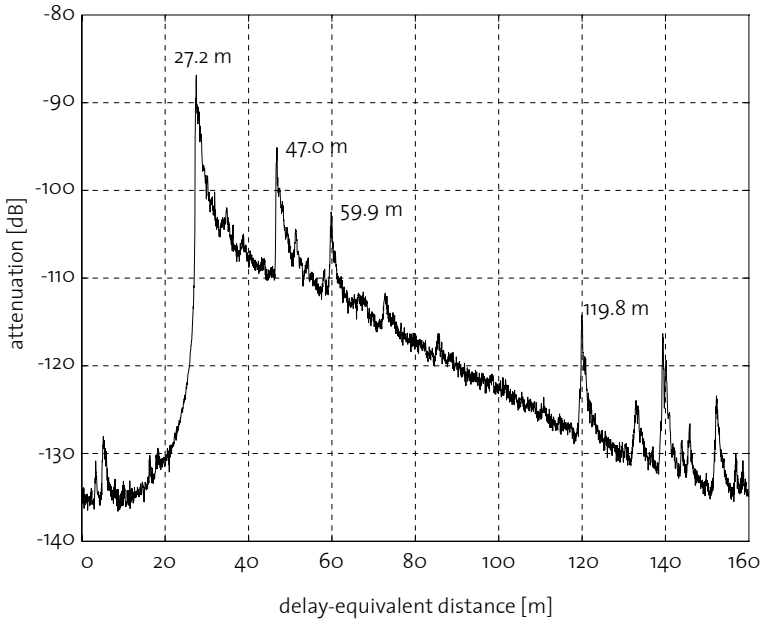


Fig. 4.15: PDP of the MCI LOS setting with measurement distance $d = 27.2$ m; averaged over $N = 90$ samples.

window between lobby and office, reflected by the outside window pane, and refracted again on its way back toward the receiver.

The later clusters in the tail of the PDP in Figure 4.15 are more difficult to associate with specular propagation paths in the floor plan because they probably result from multiple reflections, maybe from the wall at the bottom of the lobby in Figure 4.7 and from either the inside or the outside window of the conference room at the top. The propagation distances of these cluster peaks do match the physical lengths of the indicated paths. What is surprising, though, is the lack of a strong first reflection from said wall at the bottom in Figure 4.7. This casts some doubt on the interpretation of the late clusters to

4 WIDEBAND CHANNEL MEASUREMENTS

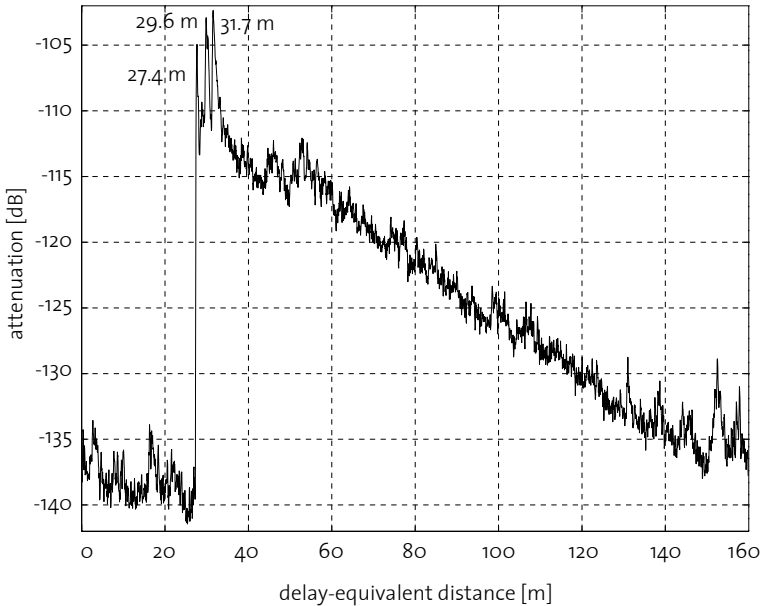


Fig. 4.16: PDP of the MCI OLOS setting with measurement distance $d = 27.5$ m; averaged over $N = 90$ samples.

be the result of multiple reflections. Another possible cause for these clusters could be crosstalk in the measurement system or the effect of surface waves traveling along the cables. Yet, both explanations are unlikely because no such strong late clusters do appear neither in the OLOS PDP shown in Figure 4.16 nor in the NLOS PDP shown in Figure 4.17, although all MCI measurements were taken with the same measurement system. On the other hand, similar late clusters can be observed in the LOS PDP computed from MCII data in Figure 4.19.

The PDP for the MCI OLOS measurement set at $d = 27.5$ m is shown in Figure 4.16. Different from the LOS PDP, there are three peaks at the start of the impulse response but no clearly distinguishable

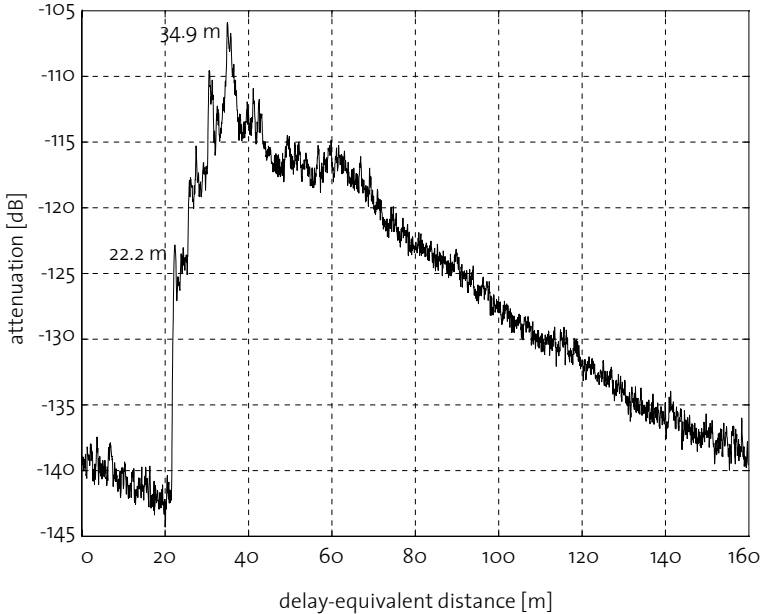


Fig. 4.17: PDP of the MCI NLOS setting with measurement distance $d = 22.2$ m; averaged over $N = 90$ samples.

clusters later on. The propagation distances that correspond to the three initial peaks can be linked to the LOS path and to reflections from the large window front on the right-hand side of the lobby in Figure 4.7. The LOS path does not seem to convey the most energy, probably because it was partially blocked by the row of concrete columns. This effect is even more pronounced in the PDP computed from NLOS data shown in Figure 4.17.

We would expect the PDPs from MCII to look more smooth than the corresponding PDPs from MCI because we have many more samples to average over in each MCII measurement set. Interestingly, though, all three PDPs, for the LOS, OLOS, and NLOS settings in MCII appear to

4 WIDEBAND CHANNEL MEASUREMENTS

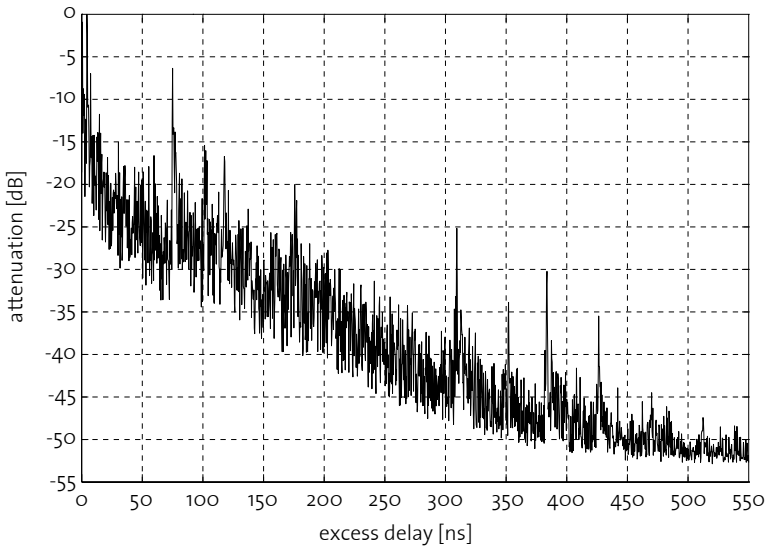


Fig. 4.18: PDP of the MCII LOS setting with measurement distance $d = 20$ m, computed from $N = 1011$ samples.

be much *less* smooth than the ones from MCI. Exemplarily, we plot in Figure 4.18 the PDP for the MCII LOS measurement set, with a distance between transmitter and receiver of $d = 20$ m and $N = 1011$ to average over. Because of the temporal alignment procedure described in the previous section, the precursor signal before the PDP onset is not shown in this figure. Without any absolute delay reference, it is not possible to translate the delay values into propagation distances. Therefore, we refer all time values to the arrival of the first path, i.e., we index the ordinate in terms of excess delays. The PDP in Figure 4.18 looks very noise-like, despite the large number of $N = 1011$ averaged impulse responses. In direct comparison with the corresponding PDP from MCI, we can notice the same clusters at similar excess delays, yet these clusters are partially hidden in the noise-like structure of the PDP in Figure 4.18.

Both antennas in MCII were fixed for the duration of the measure-

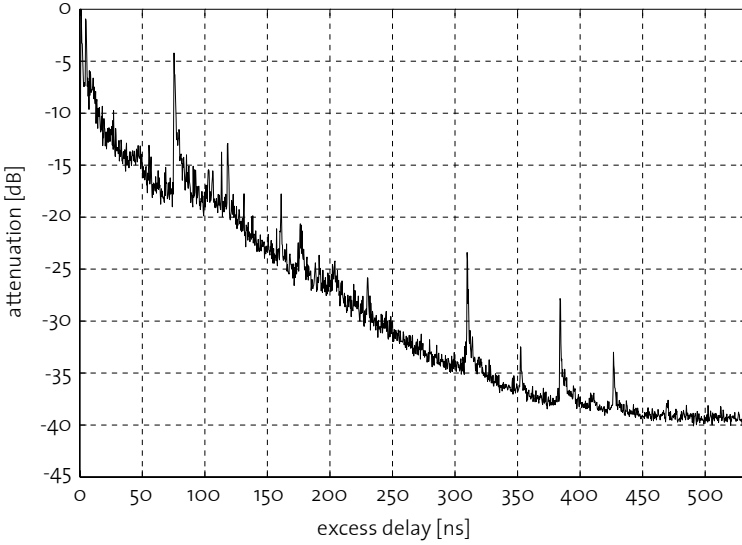


Fig. 4.19: PDP of the MCII LOS data, computed from $N = 1011$ samples after subtracting the mean of each tap.

ments; thus, one conclusion from Figure 4.18 might be that the PDP in such a setting is indeed much less smooth because we cannot average the channel realizations over space and hence see essentially a single spatial channel realization, only slightly perturbed by moving people in the environment. This hypothesis is supported by the following observation: Figure 4.19 shows again a PDP computed from $N = 1011$ MCII LOS samples, but for its computation we first subtract the empirical mean in each tap, i.e., Figure 4.19 shows a PDP vector with elements

$$p_0[l] \triangleq \frac{1}{N} \sum_{n=0}^{N-1} |h_n[l] - \hat{\mu}[l]|^2, \quad (4.9a)$$

Table 4.2: MCI LOS mean delay and delay spread

Distance d	Mean delay μ_τ [ns]	Delay spread σ_τ [ns]
27.2 m	26.2	41.5
24.3 m	26.3	41.1
21.2 m	30.1	43.2
18.4 m	29.2	42.5
15.4 m	26.9	40.5

where

$$\hat{\mu}[l] \triangleq \frac{1}{N} \sum_{n=0}^{N-1} h_n[l]. \quad (4.9b)$$

This zero-mean PDP is much more smooth than the one computed according to (4.8), which indicates that it might indeed be the mean value in each tap that leads to the ragged appearance of the PDP in Figure 4.18. Thus, there seems to be a fundamental difference between wireless channels whose variation is induced by displacement of the terminals in space, and channels where the main source of channel variation is the mobility of objects, in our particular case people, in the environment.

Another important observation are the late clusters in Figure 4.19, which indicate that similar findings in MCI are not merely measurement artifacts.

4.6.2. Mean Delay and Delay Spread

Two channel parameters quoted in every publication on channel measurements are the mean delay and the delay spread (Hashemi, 1993b). The mean delay is the first moment of the PDP $p[l]$ defined in (4.8), i.e.,

$$\mu_\tau \triangleq \frac{\sum_{l=0}^{L-1} \frac{l}{B} p[l]}{\sum_{l=0}^{L-1} p[l]}, \quad (4.10)$$

Table 4.3: MCI OLOS mean delay and delay spread

Distance d	Mean delay μ_τ [ns]	Delay spread σ_τ [ns]
27.5 m	47.0	52.5
24.6 m	44.0	38.6
21.7 m	42.8	40.0

and the delay spread is the square root of the PDP's second central moment,

$$\sigma_\tau \triangleq \left(\frac{\sum_{l=0}^{L-1} (l/B - \mu_\tau)^2 p[l]}{\sum_{l=0}^{L-1} p[l]} \right)^{1/2}. \quad (4.11)$$

Together with the channel bandwidth B , these parameters allow for a back-of-the-envelope estimation of the coherence bandwidth, and hence, the diversity order of the channel (Tse and Viswanath, 2005). They are easy to use, easy to measure, and widely available for many types of channels.

The mean delay and delay spread for the MCI LOS measurement sets is given in Table 4.2 on the preceding page. The numbers are in line with the ones reported in other UWB measurement campaigns in Table 4.1 on page 93. An interesting observation is that both mean delay and delay spread do not vary much across measurement locations. The walls and windows in the lobby reflect much of the overall transmitted power, which arrives mainly in the clusters already observed in Figure 4.15. Because of the geometry of the LOS measurement setup, moving the receiving antenna away from the transmitter increases the propagation distance of the direct path but reduces the propagation distance of reflected paths. The net effect is that delay spread is insensitive to variations in d .

The MCI OLOS measurements are different in this respect; mean delay and delay spread given in Table 4.3 do change significantly with a change in distance d between transmitting and receiving antenna. It seems as if the mean delay decreases with distance; but with only three OLOS measurement sets, any statistical analysis of this phenomenon is futile.

4 WIDEBAND CHANNEL MEASUREMENTS

Table 4.4: MCII mean delay and delay spread

Setting	Mean delay μ_τ [ns]	Delay spread σ_τ [ns]
LOS	28.4	46.7
OLOS	45.4	45.1
NLOS	61.6	48.0

The values for mean delay and delay spread for the three MCII measurement sets listed in Table 4.4 are consistent with the corresponding parameters from MCI. The mean delay seems to increase with a decreasing number of visible clusters in the corresponding PDPs.

CHAPTER 5

Marginal Tap Distributions

MEASUREMENTS are only useful if analyzed and interpreted. The exploration of the recorded impulse responses described in Section 4.6 provides a general feeling for the measured channels, but our goal in the present and the next chapter is somewhat more ambitious: to build a stochastic model for UWB channels that is accurate yet simple enough for theoretical analysis of wideband communication systems, and to answer some of the fundamental modeling questions raised in Section 2.5.

5.1. PRELIMINARIES

A stochastic model requires three main ingredients:

1. The overall structure of the model, as discussed in Chapter 2 and Chapter 3. Many aspects of the model structure are similar for deterministic and stochastic models.
2. The structure of the probability law that jointly describes all quantities that are modeled as random.
3. A parameterization of the parametric distribution families used.

Both the stochastic description and a suitable parameterization can either be derived from first principles and basic assumptions, e.g., like the fading model of Clarke (1968) and Gans (1972), or directly obtained from measurements. We combine both approaches and rely

on our measurement data and on physical principles to select a suitable probabilistic model structure and to characterize it statistically.

The available number of measurement sets and the number of samples per set as well as the technical limitations of our measurement systems outlined in Section 4.3.3 restrict the type of models we can statistically characterize to block-fading models discussed in Section 3.1. These models have a discrete-time IO relation

$$\mathbf{y}[k] = \sum_{l=0}^{L-1} \mathbf{h}[l] \mathbf{x}[k-l] + \mathbf{w}[k] \quad (5.1)$$

with random channel vector

$$\mathbf{h} \triangleq [\mathbf{h}[0] \ \mathbf{h}[1] \ \dots \ \mathbf{h}[L-1]]^T. \quad (5.2)$$

The noise vector $\mathbf{w} = [\mathbf{w}[0] \ \mathbf{w}[1] \ \dots \ \mathbf{w}[L-1]]^T$ describes the thermal noise introduced by all resistive components in the signal path. Thermal noise dominates other noise sources at the frequencies of interest (Rappaport, 2002); it can be very well modeled as i.i.d. JPG. Hence, statistical modeling of the IO relation (5.1) chiefly means to find a stochastic model for the random channel vector \mathbf{h} on the basis of our measurements. The most general such model is a joint distribution for \mathbf{h} .

Most statistical methods rely on the concept of a *random sample*, where individual measurements are treated as independent random variables that all follow the same distribution. Each small-scale measurement set $\{\mathbf{h}_n\}_{n=0}^{N-1}$ of a given measurement campaign (MCI or MCII), large-scale setting (LOS, OLOS, or NLOS), and distance d between transmitting and receiving antenna comprises a fixed number N of complex baseband channel impulse response vectors \mathbf{h}_n . To use the data from a specific measurement set as a random sample, we need to assume that all individual measurements in this set are i.i.d. random vectors of the same distribution. We make this assumption explicit and say that the channel vector \mathbf{h} for a given measurement set is distributed according to $F(\mathbf{h})$, the *operating distribution* introduced in Section 2.5. The operating distribution is the closest approximation

of the physical situation by means of a stochastic description (Linhart and Zucchini, 1986). Both the assumption of a random sample and the operating distribution itself are modeling assumptions and not a trait of the physical reality we want to analyze, in the same way as the statistical modeling approach in general is a modeling assumption rather than reflecting physical reality.

So far we used the terms “statistical model” and “distribution” rather carelessly. For the following analysis, we need to be more precise. Unfortunately, researchers in wireless channel modeling and various branches of statistics do not always use a common terminology when discussing statistical models. We deem the following terms most suitable for the tasks at hand.

- A univariate *cumulative distribution function (CDF)* $F(x)$ of a random variable \mathbf{x} , also called *distribution function*, or simply a *distribution*, is a right-continuous nondecreasing function with $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = 1$ (Feller, 1971). Multivariate distributions are defined analogously.
- We denote the set of all univariate distributions by \mathcal{M} .
- If it exists, we denote by $f(x)$ the probability density function (PDF) that corresponds to a given distribution $F(x)$.
- A *parametric family of distributions*, or simply a *family*, is a set $\mathcal{F} \triangleq \{F_{\mathbf{q}} : \mathbf{q} \in \mathcal{T}\}$ of distributions $F_{\mathbf{q}}$ where each distribution is parameterized by a D -dimensional parameter vector $\mathbf{q} \triangleq [q_1 \ q_2 \ \dots \ q_D]^T$ that takes values in some set $\mathcal{T} \subset \mathcal{R}^D$.
- A distribution $F_{\hat{\mathbf{q}}}(x)$ that is parameterized by a parameter vector $\hat{\mathbf{q}}$ estimated from measurements on the basis of said distribution’s family is called a *fitted distribution*.

In the model selection literature, a different terminology is commonly used. The term (*fitted*) *model* is substituted for fitted distribution, a family of distributions is referred to as a *family of models* (Linhart and Zucchini, 1986). In the present work, a “model” has a broader meaning that also encompasses the overall structure of the channel IO relation. Therefore, we prefer the terms distribution and family when discussing stochastic channel modeling aspects.

5.2. THE STATISTICAL MODELING PROBLEM

The operating distribution $F(h)$ is the closest possible stochastic representation of the given discrete-time block-fading channel. It is only conceptual, though; a complete specification of $F(h)$ is not possible in general as it would be way too complex to be tractable. Therefore, statistical modeling means to find a CDF that approximates $F(h)$. Following the discussion in Section 2.1, such an approximation should be accurate, tractable, and general. Compared with Section 2.1, we may make these requirements more specific, now that we established a model structure.

- Accuracy of a stochastic model means that, on the one hand, it should fit our measurement data. On the other hand, though, it also needs to lead to *consistent predictions*: the model distribution needs to be consistent with the empirical distribution of future channel realizations.
- Tractability means that the mathematical complexity of the joint CDF should be low; but for a stochastic model it also means that the choice of a CDF is not completely abstract, but that we can retain a link to the physical reality we set out to model; such a link often helps to guide intuition.

Several stochastic modeling procedures exist that take into account some but not all of the above requirements. For example, a density or a CDF can be estimated on the basis of the measured impulse responses alone, without recourse to any physical description of the channel (Devroye and Györfi, 1985; Devroye, 1987). The resulting densities would certainly show good agreement with measured data, but might fall short of providing design intuition and generalizing easily to other environments and application scenarios. A converse approach is to use the principle of maximum entropy, proposed for wireless channel modeling by Debbah and Müller (2005); here, modeling complexity is reduced and physical knowledge about the environment can be included, but consistency with measurements is difficult to assess.

With our approach presented in the following sections, we try to strike a balance between physical intuition, mathematical rigor, and, hopefully, common sense. We do not follow a single modeling paradigm, but use the methods we deem most suitable for a given specific modeling question. However, such an approach is vulnerable to significant experimenter bias: we might easily retain only the results that are in agreement with our preconceived expectations and stereotypes or choose those statistical methods that most likely lead to the desired results. To minimize the likelihood of this happening, we often double check results by different methods. Furthermore, we list findings in the present and the next chapter even if they are inconclusive or contradictory.

The number of samples necessary to completely characterize the joint distribution of the channel impulse response vector cannot sensibly be measured. For an impulse response vector of 700 taps, even the relatively simple JPG PDF has over 245 000 complex-valued parameters. The number of samples necessary for any sensible statistical method to yield reliable results would easily exceed several million, a number impossible to obtain with the maximum acquisition rate of about one channel snapshot per second in MCII. But at the same time, the search for a mathematically tractable joint CDF for \mathbf{h} is not much of a search at all because there do not seem to exist mathematically tractable multivariate distributions for random vectors of dimension $L > 2$ besides the multivariate Gaussian (Fang et al., 1990; Anderson, 2003; Muirhead, 1982). However, these issues do not mean that we should abandon all modeling efforts and simply use what is available. Instead, we simplify the problem in that we first model the marginal distribution of the individual channel taps in the present chapter and characterize the joint distribution of \mathbf{h} up to second order in Chapter 6. Yet, even a second-order analysis, which requires estimation of the mean and covariance of \mathbf{h} , is difficult to perform reliably given the relatively small number of samples available.

5.3. MODEL SELECTION

As a compromise between the purely empirical approach of directly estimating the marginal tap distributions and the theoretical approach of constructing such distributions by the principle of maximum entropy, we specify a set of *candidate families* on the basis of physical insight and published measurement results, and then use statistical tools to determine which distribution in this candidate set provides the best approximation to the marginal operating distribution of each individual channel tap.

As already outlined, we focus on the marginal distribution of individual channel taps $\mathbf{h}[l]$. Specific aspects of the multivariate channel distribution are treated in Chapter 6. Our approach is one of *model selection*, i.e., our goal is to choose from a set of fitted distributions the one that best approximates the marginal operating distribution for a given channel tap l . Even this simplified univariate model selection problem is difficult to solve because each tap $\mathbf{h}[l]$ is a complex random variable and is thus characterized by the multivariate distribution of its real and imaginary part. Hence, we further simplify the modeling problem and select a distribution for the tap magnitudes first. As the following developments are not contingent on a specific tap index l , we simplify notation and use the generic random variable $\mathbf{a} = |\mathbf{h}[l]|$ for all tap magnitudes, with the univariate CDF $F(a)$ as operating model. We have N samples in each measurement set; correspondingly, we denote the n th realization of \mathbf{a} by a_n .

Let \mathcal{M} stand for the set of all univariate CDFs. A parametric *candidate family*

$$\mathcal{G}^j = \{G_{\mathbf{q}^j}^j \mid \mathbf{q}^j \in \mathcal{T}^j\} \quad (5.3)$$

is a subset of \mathcal{M} , where individual CDFs $G_{\mathbf{q}^j}^j(a)$ are identified by the D -dimensional parameter vector $\mathbf{q}^j \triangleq [q_1 \ q_2 \ \dots \ q_D]^T \in \mathcal{T}^j$, with $\mathcal{T}^j \subset \mathcal{R}^D$. For notational convenience, we take $G_{\mathbf{q}^j}^j(a)$ to mean $G_{\mathbf{q}^j}^j(a)$

in the following, i.e., we drop the redundant superscript. The set

$$\mathcal{J} = \bigcup_{j=1}^J \mathcal{G}^j \quad (5.4)$$

of J candidate families constitutes the *candidate set*.

5.3.1. Candidate Families

Before we can analyze statistical procedures to select a distribution that minimizes the approximation error with respect to $F(a)$, the marginal operating distribution corresponding to a given tap amplitude, we need to specify the candidate set. Ideally, candidate families are physically motivated, e.g., they are derived from physical principles like the superposition of partial waves that result from certain propagation mechanisms. Yet, such a derivation is often elusive, except for highly simplified propagation scenarios (Vaughan and Bach Andersen, 2003). Therefore, we also take into account distributions that do not have a clear physical explanation but whose suitability has been demonstrated in a number of measurement campaigns. In accord with the overall goal to find a stochastic model that is a low-complexity approximation of the operating model, we require candidate families to be reasonably simple; e.g., we rule out families like the one proposed by Zhang et al. (2002). We furthermore exclude families that are specifically derived to jointly model both small- and large-scale fading, like the family proposed by Suzuki (1977) or combinations such as Rice-Lognormal or Nakagami-Lognormal (Yacoub, 2000). In accord with these criteria, and after an extensive literature study, we chose to include the families described in the following subsections. Hashemi (1993b) provides a tutorial review that covers all of the listed families and contains further references.

A. Rayleigh

If a complex random variable is proper Gaussian distributed, i.e., if its real and imaginary parts are i.i.d. Gaussian, then said random vari-

able's amplitude follows a Rayleigh distribution. We already discussed the proper Gaussian assumption for the taps of \mathbf{h} in Section 2.5.2; if many partial waves that result from independent propagation mechanisms arrive in a range of delays that is not resolvable at the receiver, the resulting superposition of partial waves in any given tap will, by the CLT, be JPC distributed (Vaughan and Bach Andersen, 2003). Consequently, the corresponding channel tap has a Rayleigh-distributed amplitude and a uniformly distributed phase. In particular, if $\mathbf{h}[l] \sim \mathcal{CN}(0, \sigma^2)$, then the PDF of the tap amplitude $\mathbf{a} = |\mathbf{h}[l]|$ is

$$g_{\mathbf{q}}(a) = \frac{2a}{\sigma^2} e^{-\frac{a^2}{\sigma^2}}, \quad a \geq 0, \quad (5.5)$$

where the parameter vector \mathbf{q} contains only a single entry, the variance σ^2 . The Rayleigh family is almost universally used to model amplitude fading of narrowband wireless channels (Rappaport, 2002; Vaughan and Bach Andersen, 2003; Tse and Viswanath, 2005). However, if the bandwidth of the baseband channel grows, the time resolution becomes finer and less and less partial waves might aggregate in every particular tap of the discrete-time channel impulse response. For this reason, several researchers question the validity of the Rayleigh family to accurately model amplitude fading of UWB channels (Zhang et al., 2002; Cassioli et al., 2002; Molisch et al., 2003a; Irahauten et al., 2006; Muqaibel et al., 2006). So far, though, this hypothesis is not universally agreed on, as several other measurement campaigns indicated that a Rayleigh distribution is indeed suitable to model the tap amplitudes of UWB channels (Saleh and Valenzuela, 1987; Ghassemzadeh et al., 2004; Karedal et al., 2004a)

B. Rice

If a strong partial wave of fixed delay* arrives at the receiver in addition to many partial waves received over time-variant paths, the tap of the complex baseband discrete-time impulse response that

* The delay has to be fixed with respect to the first arrival, similar to the discussion about time normalization in Section 4.5.1.B.

corresponds to the delay of the fixed path can often be modeled as JPG distributed around a nonzero mean. The physical motivation for the Gaussian distribution is the same as just discussed, whereas the mean component might result from a LOS path of which the receiver is able to track the time variation, or from a strong reflection of a large, smooth surface in the environment if the delay of the corresponding path does not change over time (Rappaport, 2002; Vaughan and Bach Andersen, 2003; Tse and Viswanath, 2005). The family of amplitude distributions that corresponds to a nonzero mean JPG random variable was derived by Rice (1944). It can be expressed in two different forms. In the first representation, the relation with the underlying JPG distribution is apparent. Let $\mathbf{h}[l] \sim \mathcal{CN}(\mu, \sigma^2)$; then, the PDF of $\mathbf{a} = |\mathbf{h}[l]|$ is

$$g_{\mathbf{q}}(a) = \frac{2a}{\sigma^2} \exp\left(-\frac{a^2 + |\mu|^2}{\sigma^2}\right) I_0\left(2a\frac{|\mu|}{\sigma^2}\right) \quad a \geq 0, \quad (5.6)$$

where $I_0(a)$ is the modified Bessel function of the first kind and order zero, and where the parameter vector is $\mathbf{q} = [\mu, \sigma^2]^T$. The second representation is parameterized in terms of the total power

$$\Omega \triangleq \sigma^2 + |\mu|^2 \quad (5.7a)$$

and the *Ricean K-factor*

$$K \triangleq |\mu|^2 / \sigma^2, \quad (5.7b)$$

the ratio of the power in the mean component and the random component. The Ricean PDF in terms of the parameter vector $\mathbf{q} = [\Omega, K]^T$ is then

$$g_{\mathbf{q}}(a) = 2a \frac{1+K}{\Omega} \exp\left(-a^2 \frac{1+K}{\Omega} - K\right) \times I_0\left(2a\sqrt{\frac{K(1+K)}{\Omega}}\right) \quad a \geq 0. \quad (5.8)$$

For $K = 0$, we recover the Rayleigh PDF. The Rice family is subject to the same criticism as the Rayleigh family as a model for small-scale fading in wireless channels of very wide bandwidth. Though, again the evidence against the Rice family is not conclusive so far, as there are some publications that support it (Kunisch and Pamp, 2002; Hovinen et al., 2002; Laine et al., 2004; Pagani and Pajusco, 2006).

C. Nakagami

Nakagami proposed the family carrying his name to fit data of an extensive measurement campaign of the tropospheric scattering channel (Nakagami, 1960). The family is characterized by a two-parameter PDF with parameter vector $\mathbf{q} = [\Omega, m]^T$,

$$g_{\mathbf{a}}(a) = \frac{2}{\Gamma(m)} \left(\frac{m}{\Omega}\right)^m a^{2m-1} \exp\left\{-\frac{ma^2}{\Omega}\right\}, \quad a \geq 0, \quad (5.9)$$

where $\Gamma(m)$ is the gamma function. As in the Rice family, the parameter Ω denotes the total power of $\mathbf{h}[l]$, i.e., $\Omega = \mathbb{E}[|\mathbf{h}[l]|^2] = \mathbb{E}[\mathbf{a}]$. The *Nakagami parameter* m is defined as the ratio of moments,

$$m \triangleq \frac{\Omega^2}{\mathbb{E}[(\mathbf{a}^2 - \Omega)^2]} \geq \frac{1}{2}, \quad (5.10)$$

and is sometimes called the *fading figure*. Although two parameters specify the Nakagami PDF, only the m parameter changes its shape. For $m = 1$, the Nakagami PDF reduces to the Rayleigh PDF; for $m > 1$ the tails of the Nakagami PDF are less accentuated than in the Rayleigh case, which indicates less severe fading; for $m < 1$, the tails are heavier, which means that fading is stronger.

The physical motivation for the Nakagami family as a model for small-scale amplitude fading arises from the analysis of statistical scattering from random rough surfaces and the resulting superposition of partial waves. In particular, the Nakagami PDF is an approximation of the PDF that results if one allows for correlated real and imaginary parts of the superposing partial waves (Braun and Dersch, 1991). The Nakagami parameter m in the corresponding derivation is linked to

the surface height parameter of the random rough surface, which in turn determines if a specular reflection results or if the returned wave field is the superposition of many scattered partial waves (Nakagami, 1960; Braun and Dersch, 1991; Parsons, 2000).

The effect of the Nakagami parameter is comparable to the Ricean K -factor (5.7b) in that it determines the severity of the fading. Indeed, the two densities approximate each other by setting (see, e.g., equation (59) in the paper by Nakagami, 1960, or Appendix C.12 in the book by Vaughan and Bach Andersen, 2003)

$$K = m - 1 + \sqrt{m^2 - m}. \quad (5.11)$$

Although originally derived and fitted for data obtained from measurements of tropospheric scattering channels, Nakagami distributions have been found to provide a good fit also to outdoor urban, suburban and rural measurements (Suzuki, 1977; Braun and Dersch, 1991; Lacroix et al., 1997) and to indoor measurements (Hashemi, 1993a; Handforth et al., 1993; Abouraddy and Elnoubi, 2000). Recently, Nakagami distributions were found suitable to model small-scale fading in UWB channels (Cassoli et al., 2002; Balakrishnan et al., 2004; Hentilä et al., 2005); the Nakagami family is used to describe the amplitude coefficients in the IEEE 802.15.4a SV block-fading impulse response (Molisch et al., 2006b). However, some researchers dispute the utility of the Nakagami family to stochastically model wideband fading altogether (Stein, 1987).

D. Lognormal

The lognormal family (Limpert et al., 2001) is often used to model large-scale fading, also known as *shadowing*, that results if the strength of the received signal varies because of slowly changing attenuation from hills, buildings, or other structures. If a random variable \mathbf{x} is normally distributed, i.e., $\mathbf{x} \sim \mathcal{N}(\mu, \sigma^2)$, then $\mathbf{a} = e^{\mathbf{x}}$ is lognormal distributed with PDF

$$g_{\mathbf{q}}(a) = \frac{1}{a\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\ln a - \mu)^2}{2\sigma^2}\right\} \quad a \geq 0. \quad (5.12)$$

The parameter vector $\mathbf{q} = [\mu, \sigma^2]^T$ contains the parameters of the corresponding underlying Gaussian distribution. While the Rayleigh, Rice, and Nakagami amplitude families arise from the summation of many independent random quantities, the lognormal family, by the CLT, results if many independent quantities are multiplied. While this multiplicative effect can be understood in the context of large-scale fading, where successive attenuation of a given propagation path acts multiplicatively, it is difficult to physically motivate the lognormal family as a model for small-scale fading. Hashemi (1993b) conjectures that small-scale lognormal fading might result from multiple reflections of partial waves, which would lead to a multiplicative effect.* Although physical motivation is scarce, some researchers report a good fit of lognormal amplitude distributions for their channel measurement data for conventional narrowband and wideband outdoor (Lacroix et al., 1997) and indoor channels (Nikookar and Hashemi, 1993). Most importantly however, some researchers found that lognormal distributions provide a good fit to measured data for UWB channels (Foerster and Li, 2002a,b; Keignart and Daniele, 2003; Li and Wong, 2003; Hentilä et al., 2005), so that the lognormal family was adopted to model the path amplitudes in the IEEE 802.15.3a reference model (Foerster, 2003; Molisch et al., 2003a).

E. Weibull

The Weibull family was originally proposed to model failure rates over time (Johnson et al., 1995); it is a two-parameter family with parameter vector $\mathbf{q} = [b, d]^T$. Its PDF is given as

$$g_{\mathbf{q}}(a) = db^{-d}a^{d-1}e^{-(a/b)^d} \quad a \geq 0. \quad (5.13)$$

As in the lognormal case, a physical explanation why small-scale amplitude fading should be Weibull distributed is hard to come by—except for $d = 2$, in which case the Weibull PDF reduces to the Rayleigh PDF.

* This reasoning originates in a paper by Suzuki (1977), but it is disputed, e.g., by Abbas and Sheikh (1997).

Nevertheless, there are several published measurement campaigns where researchers report a good fit of Weibull distributions for several environments, carrier frequencies and bandwidths (Nikookar and Hashemi, 1993; Hashemi et al., 1994; Healey et al., 2000; Tzeremes and Christodoulou, 2002; Babich and Lombardi, 2000). In an extensive report, the IEEE Vehicular Technology Society Committee on Radio Propagation (1988) also recommended the use of either the Nakagami or the Weibull family for models in the 800 MHz to 900 MHz range. Pagani and Pajusco (2004) and Chong and Yong (2005) successfully fitted Weibull distributions to UWB measurement results.

5.3.2. Parameter Estimation

The families just presented are parametric. The Rayleigh family is completely specified by a single parameter, while all other candidate families depend on two parameters. The first step in any model selection method is to fit the candidate families with respect to the measured data, i.e., to estimate the parameters of each family. Although parameter estimation is treated in virtually every statistics textbook, we briefly outline the specific estimators we use for the five candidate families so as to guarantee reproducibility of our results.

A. Rayleigh

The sole parameter of the Rayleigh family is equal to the variance of the underlying JPG family. Consequently, we estimate it as

$$\hat{\sigma}^2 = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} a_n^2}, \quad (5.14)$$

where the average is taken over the N channel samples in a given measurement set.

B. Rice

The maximum likelihood (ML) estimate of the total power Ω is the sample second moment (Taludkar and Lawing, 1991), as for the Rayleigh family. However, estimation of the K -factor is tricky: Tepedelenlioğlu et al. (2003) numerically computed the Cramér-Rao lower bound on the estimation error variance and showed that it diverges for $K \rightarrow 0$, i.e., when the Rice distribution approaches the Rayleigh distribution. A closed-form ML estimate of Ω and K , or $|\mu|$ and σ^2 , does not exist; hence, we need to resort to numerical methods. An implicit formula for the ML estimate is given by Taludkar and Lawing (1991), and an expectation-maximization algorithm by Marzetta (1995). We choose yet another approach and numerically maximize the log-likelihood function by standard optimization methods. In particular, we use the Nelder-Mead algorithm as implemented in Matlab's `fminsearch` function, which is in turn called by the `ricefit` function.* A method of moments estimate serves as a starting value for the numerical maximum search (Rastogi and Holt, 1981; Tepedelenlioğlu et al., 2003). Yet for small values of K , the numerical minimum search does not converge, manifesting the theoretical impossibility to accurately estimate small K -factors. In this case, though, the difference between the fitted Rice and Rayleigh distributions is very small for K close to 0; hence, we simply set $K = 0$.

C. Nakagami

A closed-form ML estimator of the Nakagami parameters Ω and m does not exist, but several approximations of the ML solution are available, like the estimators recently proposed by Cheng and Beaulieu (2001) and Ko and Alouini (2003), as well as the classical approximation of the ML solution by Greenwood and Durand (1960). The latter ML approximation is actually an estimator for the parameters of the gamma distribution; because the square of a Nakagami-distributed random variable is gamma distributed, the estimates are equivalent.

* We had to modify this function as it contained a bug.

For reasons detailed later, we need ML parameter estimates. Therefore, we square the channel taps to estimate the parameters of the corresponding gamma distribution, but instead of approximating the ML estimate by the method of Greenwood and Durand (1960), we numerically maximize the gamma likelihood function by means of the `gamfit` function in Matlab's Statistics Toolbox.

D. Lognormal

We obtain ML parameter estimates of the lognormal family from ML estimates of the corresponding normal family after a logarithmic transformation of the data.

E. Weibull

The Weibull family is related to the extreme value family (Johnson et al., 1995) through a logarithmic transformation of the random variable; thus, we use a numerical ML estimate for the parameters of the extreme value family, as implemented in Matlab's `evfit` function.

5.3.3. Goodness-of-Fit Tests

With the candidate set in place, our main task is now to evaluate which family with which parameter values provides the best approximation for the operating distribution of our measurements. Several methods exist in the literature on wireless channel modeling and in the statistics literature. Each approach has its strengths and weaknesses in terms of mathematical rigor, ease of use, intuition, and generality.

Probably the most well known and most commonly used method to determine small-scale fading distributions is by means of GOF tests. We do not believe that GOF tests are well suited to select small-scale fading distributions; hence, we do not use GOF tests in this work. But their dominance in much of the literature on empirical modeling of wireless channels obliges us to briefly review the main concepts and to justify why we take a different approach.

A GOF tests is a specific type of hypothesis test (Papoulis and Pillai, 2002; Bartoszyński and Niewiadomska-Bugaj, 1996; Dixon and Massey, Jr., 1983). In the framework of Neyman and Pearson (Spanos, 1999), the *challenging hypothesis** \mathcal{H}_c is that a given parameterized candidate CDF $G_q(a)$ equals the operating CDF, i.e., $\mathcal{H}_c : G_q = F$. The *incumbent hypothesis* \mathcal{H}_i is the complementary event, $\mathcal{H}_i : G_q \neq F$. The general procedure to test for the hypothesis \mathcal{H}_c is to partition the sample space into an acceptance region and a rejection region, where these regions are defined through a *test statistic* $t(\mathbf{a})$ computed from the data vector $\mathbf{a} \triangleq [a_0 \ a_1 \ \dots \ a_{N-1}]^T$. The rejection region is the set $\{a : t(a) < c\}$ for some c . The event that \mathcal{H}_c is rejected although it was true is called a *type I error*. We choose c so as to attain a small probability of type I error, called *significance level* α of the test: if we consider the data as a random sample \mathbf{a} , so that the test statistic $t(\mathbf{a})$ is a random variable, the significance level is $\alpha \triangleq \mathbb{P}(t(\mathbf{a}) < c | \mathcal{H}_c)$. To compute this probability, the test statistic needs to be such that its sample distribution, given that \mathcal{H}_c is true, is known or can be approximated by its limiting distribution for $N \rightarrow \infty$. A *type II error* is the event that \mathcal{H}_c is false but not rejected. A good GOF test should be *powerful*, i.e., it should minimize the probability of type II error for a given significance level.

GOF tests are not well suited for model selection for two main reasons. (i) Most GOF tests are applicable for a specific distribution as challenging hypothesis, not for a complete parametric family of distributions. If the distribution parameters are first estimated from measured data, the corresponding parameterized distribution cannot be subsequently used as challenging hypothesis in a GOF test on the basis of the *same* data because the test would be biased: always the best-fitting distribution for the given data would be used to compute the test statistic, so that the passing probability of the test would be inflated. (ii) GOF tests are meant to check if a given distribution

* As explained by Guttman (1977), the term “challenging hypothesis” is preferable over the more commonly used “null hypothesis” because the challenging hypothesis should represent the unexpected outcome.

provides an adequate fit for a set of data or not; a GOF test does not provide a measure of how good the fit actually is. Hence, GOF tests cannot be used to compare the fit of different distributions. But model selection requires to estimate parameters from measured data and to select one family among a whole class of candidate families.

Some researchers apply GOF tests separately for each fitted distribution in the candidate set and for different measurement sets, e.g., for LOS measurements taken at different distances d between transmitter and receiver. They then use as a measure of fit the percentage with which each fitted distribution in the candidate set passes the test. In such an approach, the problem how to consistently estimate parameters and compute test statistics from the same data persists; in addition, such an approach presupposes that the test used is equally *powerful* for all distributions, i.e., the type II error is the same for all challenging hypothesis. This is not the case in general.

5.3.4. Discrepancy Minimization

A measure of how good one distribution approximates another distribution is called a *discrepancy* (Linhart and Zucchini, 1986). The main problem with GOF tests for model selection is that they do not provide a well-defined discrepancy between the distribution of the data and the distribution of the challenging hypothesis on the basis of which different distributions can be compared. The binary result of a standard GOF test cannot be used as a discrepancy, it does not allow to compare different distributions. A composite selection criteria like the passing percentage of successive GOF tests cannot be used as discrepancy, either, for the reasons pointed out in the last subsection.

The selection of a suitable approximating distribution out of a given candidate set requires computation of a true discrepancy for each candidate distribution with respect to the operating distribution, and the selection of the distribution with the lowest discrepancy. To compute such a discrepancy we need to know the operating distribution. But as the operating distribution is unknown, we need to find a way to consistently estimate discrepancies. Any such estimator is

plagued by a similar problem as already encountered in the previous subsection: parameter estimation and discrepancy computation from the same set of data may lead to biased results. The problems of discrepancy estimation and bias correction were solved by Akaike (1973) for the specific class of linear* models. His approach can be generalized to other candidate families, as we are going to outline now. The following summary is based on the book by Linhart and Zucchini (1986) and a tutorial paper by Zucchini (2000).

A *discrepancy* is a functional $\Delta : \mathcal{M} \times \mathcal{M} \rightarrow \mathcal{R}$ that satisfies $\Delta(G, F) \geq \Delta(F, F)$ for all $G, F \in \mathcal{M}$. Every candidate family \mathcal{G}^j induces an *optimal parameter vector* q^* with the property $\Delta(G_{q^*}^j, F) \leq \Delta(G_q^j, F)$ for all $q^j \in \mathcal{T}^j$. This optimal parameter vector exists even if the operating distribution does not belong to the candidate set. A consistent estimator for the discrepancy $\Delta(G_q^j, F)$ on the basis of N samples $\mathbf{a} \triangleq [a_1 \ a_2 \ \dots \ a_N]^T$ of \mathbf{a} is called an *empirical discrepancy*; we denote it by $\Delta_N(G_q^j, F)$. For a given candidate family \mathcal{G}^j , the empirical discrepancy induces the *minimum discrepancy estimator*

$$\hat{q}^j(\mathbf{a}) \triangleq \arg \min_{q \in \mathcal{T}^j} \Delta_N(G_q^j, F). \quad (5.15)$$

When the dependence on the data vector \mathbf{a} is clear from the context, we simply write \hat{q}^j instead of $\hat{q}^j(\mathbf{a})$ in the following.

Our goal is to choose the distribution that minimizes the discrepancy among all members of the candidate set. Yet, although the empirical discrepancy is a consistent estimator, the resulting estimate

$$\hat{G} \triangleq \arg \min_{G \in \mathcal{J}} \Delta_N(G, F) \quad (5.16)$$

is not the best estimate if the number of samples N is finite: because we use the same data to estimate the distribution parameters and to select the model family, the estimate (5.16) is biased. The bias results because families of increasing complexity, as measured by the

* A *linear* model in the statistical literature refers to a linear relation between controls and observables, corrupted by additive noise (Burnham and Anderson, 2002). Prime examples are regressive and autoregressive models.

number U of independent parameters, tend to provide more flexibility to fit the data; thus, the chance of selecting a very complex distribution to best fit a given set of data is high, an effect called *overfitting*. The trade-off between complexity and estimation error can be seen if we split the overall discrepancy of any candidate distribution with estimated parameter vector G_q^j into two distinct contributions:

1. The *approximation discrepancy* is the discrepancy $\Delta(G_{q^*}^j, F)$ that results if we select a family \mathcal{G}^j different from the operating family, even if we parameterized the selected family by its discrepancy-minimizing parameter vector q^* .
2. The *estimation discrepancy* is the discrepancy $\Delta(G_q^j, G_{q^*}^j)$ that results if we estimate the parameter vector q from a finite number of samples.

A complex CDF with many free parameters U has, in general, a lower approximation discrepancy at the cost of a larger estimation discrepancy. The goal of a sensible selection procedure is thus to balance both discrepancies with respect to the number of samples available.

As a simple example for this observation, consider a least squares fit of a polynomial $q[k] = \alpha_0 + \alpha_1 k + \alpha_2 k^2 + \dots$ of arbitrary order to a set of equidistant points $\mathbf{y}[k] = k + \mathbf{w}[k]$, where $\mathbf{w}[k]$ is white Gaussian noise and $k = 0, 1, \dots, K - 1$. Although the operating model is a first order polynomial, the best fit is achieved, in general, by the polynomial of order at least $K - 1$ that goes through all K points $\mathbf{y}[k]$ because the squared error is zero in this case. However, if we draw a new set of points according to the same distribution as before, the polynomial just constructed would most certainly result in a large discrepancy.*

* This example is taken from a paper by Forster and Sober (1994); a nice graphical illustration for a similar case—polynomial regression with the relative Kullback-Leibler discrepancy—is depicted in the book by Sakamoto et al. (1986). The present example illustrates that model selection is not just a statistical problem but a problem of *epistemology*. The criteria a good model has to satisfy, like parsimony and mathematical tractability, are *outside* the realm of rigorous mathematical reasoning. Forster and Sober (1994) give a nontechnical introduc-

The above example illustrates what we have called the predictive property of a good model: it should fit available measurements equally well as hypothetical future measurements. Hence, a suitable distribution should not necessarily minimize the empirical discrepancy for a given set of measurements, but it should minimize it on average, over all possible measurement sets. Let $\mathbf{a} \triangleq [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_N]^T$ denote a random vector with i.i.d. entries distributed according to the operating CDF $F(a)$. For each candidate family \mathcal{G}^j , the minimum discrepancy estimator (5.15) on the basis of \mathbf{a} is a random vector $\hat{\mathbf{q}}^j(\mathbf{a})$. The preceding discussion suggests that we select the candidate family \mathcal{G}^j that minimizes the *expected discrepancy*

$$\min_j \mathbb{E}_F \left[\Delta(G_{\hat{\mathbf{q}}}^j, F) \right] \quad (5.17)$$

so as to minimize the discrepancy over all possible realizations of \mathbf{a} , not just over the measured data \mathbf{a} .

The problem with the expected discrepancy is that the above expectation cannot be computed because it depends on the unknown operating distribution $F(a)$. However, it can be estimated. An estimator of the expected discrepancy is called a *model selection criterion*. If the operating family, but not the exact operating distribution, is known, it is sometimes possible to derive criteria for finite sample sizes. In our case, with an unknown operating family, we have to resort to asymptotic methods, i.e., we have to contend ourselves with criteria that hold if the number of samples approaches infinity. The idea is to approximate, for a given candidate family \mathcal{G}^j , the expected discrepancy (5.17) by the first two terms of its Taylor series expansion in $\hat{\mathbf{q}}$ around the unknown optimal parameter vector \mathbf{q}^* , and then compute its expectation with respect to $\hat{\mathbf{q}}$. Linhart and Zucchini (1986, A.1) show that the resulting approximation is

$$\mathbb{E}_F \left[\Delta(G_{\hat{\mathbf{q}}}^j, F) \right] \approx \Delta(G_{\mathbf{q}^*}^j, F) + \frac{\text{tr}(\mathbf{\Omega}_j^{-1} \mathbf{\Sigma}_j)}{2N}, \quad (5.18)$$

tion to model selection by discrepancy minimization and discuss the relation between the principle of parsimony and elementary empiricism as well as related epistemological questions.

where Ω_j is the Hessian of $\Delta(G_{\mathbf{q}}^j, F)$ with respect to \mathbf{q} and evaluated at $\mathbf{q} = \mathbf{q}^*$; furthermore, Σ_j is the asymptotic ($N \rightarrow \infty$) covariance matrix of the vector whose j th entry is

$$\sqrt{N} \frac{\partial \Delta_N(G_{\mathbf{q}}^j, F)}{\partial q_j} \Bigg|_{\mathbf{q}=\mathbf{q}^*}. \quad (5.19)$$

In a further step, the approximation discrepancy $\Delta(G_{\mathbf{q}^*}^j, F)$ in (5.18) is replaced by the expected empirical discrepancy plus half the trace term of (5.18), i.e., (Linhart and Zucchini, 1986, Proposition 2)

$$\mathbb{E}_F \left[\Delta(G_{\hat{\mathbf{q}}}^j, F) \right] \approx \mathbb{E}_F \left[\Delta_N(G_{\hat{\mathbf{q}}}^j, F) \right] + \frac{\text{tr}(\Omega_j^{-1} \Sigma_j)}{N}, \quad (5.20)$$

where the expectation on the right-hand side is with respect to the random parameter vector $\hat{\mathbf{q}} = \hat{\mathbf{q}}(\mathbf{a})$. The proof of both approximations relies on a set of regularity conditions (Linhart and Zucchini, 1986, A.1.1.) and the rather strong assumption that the operating family is a subset of the candidate set. The latter is unlikely in our case; however, Linhart and Zucchini (1986) claim that the resulting criterion is still useful if the candidate distributions are not grossly different from the operating distribution. We further comment on this point in Section 5.4.5.

The criterion (5.20) is not yet in a form that can be used in practice. First of all, we still need to specify what discrepancy functional to use. Second, the expected empirical discrepancy still depends on the operating distribution; hence, we need to find an estimator for it. Third, the trace term is rather complicated to compute and might incur a large estimation variance for finite N . We address all three issues in the following section.

5.4. AKAIKE'S INFORMATION CRITERION

A widely used discrepancy is based on the *KL distance*, also called *relative entropy* (Cover and Thomas, 1991). The resulting model

selection criterion is known as Akaike's Information Criterion (AIC). It can be derived in several different ways: Burnham and Anderson (2002) directly manipulate the relative entropy functional and the resulting log-likelihood expressions. A very intuitive geometrical derivation can be found in the book by Sakamoto et al. (1986) and the papers by Clergeot (1984) and Matsuoka and Ulrych (1986); however, the geometrical interpretation only works for linear operating models. We continue along the lines of Linhart and Zucchini (1986), whose derivation is the most general we are aware of.

5.4.1. Kullback-Leibler Discrepancy

For simplicity, we assume that all distributions have corresponding PDFs. This assumption is certainly satisfied for our candidate set from Section 5.3.1. The relative entropy between the operating distribution $F(a)$ with PDF $f(a)$ and a candidate distribution $G_q(a)$ with PDF $g_q(a)$ is the functional

$$D(f||g_q) = \mathbb{E}_F[\log f(\mathbf{a})] - \mathbb{E}_F[\log g_q(\mathbf{a})], \quad (5.21)$$

where the random variable \mathbf{a} is distributed according to $F(a)$. Relative entropy $D(f||g_q)$ is nonnegative and equals zero only if $f(a) \equiv g_q(a)$. It is not a true mathematical distance measure, yet it quantifies how dissimilar two distributions are. In particular, relative entropy measures the inefficiency if we substitute a distribution $G_q(a)$ for the true distribution $F(a)$: a code for the distribution $G_q(a)$ used to encode independent source letters of distribution $F(a)$ has an average description length that is $D(f||g_q)$ nat longer per source symbol than necessary (Cover and Thomas, 1991). If used as a discrepancy, relative entropy ranks a distribution the higher the more probability this distribution assigns on average to actual observations (Zucchini, 2000).

The first term on the right-hand side of (5.21) depends on the operating distribution only; it contributes equally to the discrepancy between the operating distribution and any candidate distribution.

Thus, it suffices to consider the second term, which is called *KL discrepancy*,

$$\Delta(G_{\mathbf{q}}^j, F) \triangleq -\mathbb{E}_F[\log g_{\mathbf{q}}^j(\mathbf{a})]. \quad (5.22)$$

Consequently, the index j of the candidate family with lowest expected KL discrepancy is, by (5.17), given as (Burnham and Anderson, 2002)

$$\max_j \mathbb{E}_{\tilde{\mathbf{a}}} \left[\mathbb{E}_{\mathbf{a}} \left[\log g_{\hat{\mathbf{q}}(\tilde{\mathbf{a}})}^j(\mathbf{a}) \right] \right], \quad (5.23)$$

where \mathbf{a} and the entries of $\tilde{\mathbf{a}}$ are independent random variables that all follow the operating distribution $F(a)$. It is (5.23) we want to estimate along the lines of (5.20).

The empirical discrepancy that corresponds to the KL discrepancy in (5.22) is the empirical log-likelihood,

$$\Delta_N(G_{\mathbf{q}}^j, F) \triangleq -\frac{1}{N} \sum_{n=1}^N \log g_{\mathbf{q}}^j(a_n); \quad (5.24)$$

it is obtained from (5.22) upon substitution of the empirical operating CDF, i.e., the cumulative relative frequency of a given data set $\{a_n\}_{n=0}^{N-1}$, for the operating CDF $F(a)$. The minimum discrepancy estimator is the ML estimator,

$$\hat{\mathbf{q}} \triangleq \arg \max_{\mathbf{q} \in \mathcal{T}} \frac{1}{N} \sum_{n=1}^N \log g_{\mathbf{q}}(a_n). \quad (5.25)$$

In hindsight, this justifies our insistence on ML estimators in Section 5.3.2. Finally, Burnham and Anderson (2002) show that the expected empirical log-likelihood on the right-hand side in (5.20) can be estimated simply by the empirical log-likelihood itself,

$$\mathbb{E}_F \left[\Delta_N(G_{\hat{\mathbf{q}}}^j, F) \right] \approx \Delta_N(G_{\hat{\mathbf{q}}}^j, F) = \frac{1}{N} \sum_{n=1}^N \log g_{\hat{\mathbf{q}}}^j(a_n). \quad (5.26)$$

5.4.2. Approximating the Trace Term

The trace term $\text{tr}(\Omega_j^{-1}\Sigma_j)$ in (5.20) depends on the operating distribution because the matrices Σ_j and Ω_j are expectations with respect to the operating model. For the KL discrepancy, we have

$$[\Omega_j]_{rs} = -\mathbb{E}_F \left[\frac{\partial^2 \log g_q^j(\mathbf{a})}{\partial q_r \partial q_s} \right]_{q=q^*}, \quad (5.27)$$

$$[\Sigma_j]_{rs} = \mathbb{E}_F \left[\left(\frac{\partial \log g_q^j(\mathbf{a})}{\partial q_r} \right) \left(\frac{\partial \log g_q^j(\mathbf{a})}{\partial q_s} \right) \right]_{q=q^*}, \quad (5.28)$$

where $r, s = 1, 2, \dots, U$.

One possibility to remove the dependence on $F(\mathbf{a})$ is to estimate both matrices from the data. Each candidate family \mathcal{G}^j requires a different estimator, e.g., the ones provided for some common families by Linhart and Zucchini (1986). Unfortunately, closed-form estimators for the two matrices are only available for some families in our candidate set. Another drawback in estimating Ω_j and Σ_j is that a large number of samples might be necessary to obtain a reliable estimate.

An alternative approach is to approximate the trace term by the value it would attain if the operating distribution was part of the candidate family. Indeed, if there is no misspecification, i.e., $F \equiv G_{q^*}^j$ for some j , then $\Omega_j = \Sigma_j$. But even if the operating distribution is not part of the candidate family, like in most cases of practical interest, we can approximate $\Omega_j \approx \Sigma_j$ if the distributions in the candidate set are judiciously selected in advance to provide a reasonable fit. Then, the resulting approximation of the trace term is extremely simple,

$$\text{tr}(\Omega_j^{-1}\Sigma_j) \approx U_j. \quad (5.29)$$

Linhart and Zucchini (1986) and Burnham and Anderson (2002) report that the resulting simplified criterion often performs better than if the trace term was estimated from the data, under the condition that the candidate family does not contain families whose best-fitting

distribution is vastly different in relative entropy from the operating family.

We can now substitute the estimate (5.26) of the expected discrepancy and the approximation for the trace term (5.29) into the Taylor series approximation (5.20) of the expected KL discrepancy to obtain the model selection criterion

$$\arg \min_j -\frac{1}{N} \sum_{n=1}^N \log g_q^j(a_n) + \frac{U_j}{N}. \quad (5.30)$$

This criterion, known as AIC, was first derived by Akaike (1973). It is commonly stated in a slightly different form,

$$\text{AIC}_j \triangleq -2 \sum_{n=1}^N \log g_q^j(a_n) + 2U_j. \quad (5.31)$$

AIC is an approximately unbiased estimator of the expected discrepancy (5.23); hence, its estimation error depends on the number N of samples used to compute it. Sakamoto et al. (1986) recommend to limit the number U of free parameters in all candidate families to less than $2\sqrt{N}$ if only N independent samples are available. The approximation of the trace term (5.29) is an asymptotically unbiased estimator if the candidate set contains the operating model; but for small N , it might be significantly biased. Hurvich and Tsai (1989) developed a small-sample corrected version of AIC, called AICC, to alleviate this problem,

$$\text{AICC}_j \triangleq \text{AIC}_j + \frac{2U_j(U_j + 1)}{N - U_j - 1}. \quad (5.32)$$

Although the derivation of AICC relies on the assumption of linear models with Gaussian residuals, e.g., regressive and autoregressive models, Burnham and Anderson (2002) recommend using AICC throughout when the number of samples per parameter, N/U is less than 40. Their recommendation is based on a large simulation study of the performance of AIC and AICC. For larger sample sizes or fewer parameters, the influence of the extra correction term tends to decrease.

Instead of approximating the trace term as we did, it is possible to directly estimate the expected KL discrepancy (5.23) using the bootstrap (Linhart and Zucchini, 1986; Efron and Tibshirani, 1993). We tried this approach, but did not succeed because of the immense computing requirements of the bootstrap algorithm.

5.4.3. Akaike Weights

AIC estimates the approximation quality of different candidate distributions with respect to the operating distribution. Hence, it can be used to rank the candidate distributions: the minimum AIC value indicates the best fit. Yet, the raw AIC (and AICC) values are difficult to interpret directly because we discard the point of reference, the expected log-likelihood of the operating distribution in (5.22). To conveniently compare the relative fit of each distribution within the candidate set, we define the *AIC differences* (Burnham and Anderson, 2002)

$$\phi_j = \text{AIC}_j - \min_i \text{AIC}_i, \quad (5.33)$$

where $\min_i \text{AIC}_i$ denotes the minimum AIC value over all J candidate families, and compute the *Akaike weights* (Akaike, 1978)

$$w_j \triangleq \frac{e^{-\frac{1}{2}\phi_j}}{\sum_{i=1}^J e^{-\frac{1}{2}\phi_i}}. \quad (5.34)$$

The Akaike weights are normalized, i.e., they satisfy $\sum_{j=1}^J w_j = 1$. So far, this definition is a formal one only, meant to simplify comparison of different candidate distributions. However, the Akaike weights can be given a Bayesian interpretation as the *likelihood of a model* (Akaike, 1978; Burnham and Anderson, 2002; Wagenmakers and Farrell, 2004), i.e., the Akaike weight w_j can be interpreted as the conditional probability that the candidate family \mathcal{G}^j contains the model of lowest KL discrepancy in the candidate set \mathcal{J} , given the data. Consequently, the Akaike weights allow us not only to select the best distribution in the candidate set, they also provide information

about the relative approximation quality of each family and about the reliability of selecting the AIC-best model.

One way to assess the reliability of a given selection is to compute the *confidence set*, the equivalent to a confidence interval for a given significance level in classical model-based parameter estimation. A confidence set is a set of families so that the family with the best-fitting distribution is in the confidence set with high probability. The simplest way to obtain such a confidence set is to include those families in the set whose best-fitting distributions have cumulative Akaike weights just larger than a given confidence level; e.g., if the fitted distributions in the set $\{1, 5\}$ have total Akaike weight $w_1 + w_5 = 0.81$, the ones in the set $\{1, 2, 5\}$ have total Akaike weight 0.967, but including the next best distribution yields $w_1 + w_2 + w_3 + w_5 = 0.98$, then the 95% confidence set will be $\{1, 2, 5\}$. Thus, the Akaike weights can be used to assess the reliability of any selection. If several distributions have similar Akaike weights, the evidence provided by the data is not sufficient to allow for a better differentiation (Burnham and Anderson, 2002, Chapter 4.5).

5.4.4. A Simulation Study

Apparently, AIC (or AICC) has never been used before to select suitable distributions for small scale amplitude fading in wireless channels. The only publication we are aware of in which model selection techniques are employed to determine fading distributions is the one by Taneda et al. (2001); the authors use the criterion of minimum description length (MDL) to select a fading distribution among the Rayleigh, Rice, Nakagami, Nakagami-q, Weibull, and generalized Nakagami families on the basis of narrowband measurements of outdoor urban and suburban channels. Unfortunately, they only report the percentage of each family to provide the best fit but do not provide a more detailed description of their results. Hence, we can only draw on very limited previous experience with information criteria for wireless channel modeling in general, and apparently on no such experience at all in the particular case of AIC.

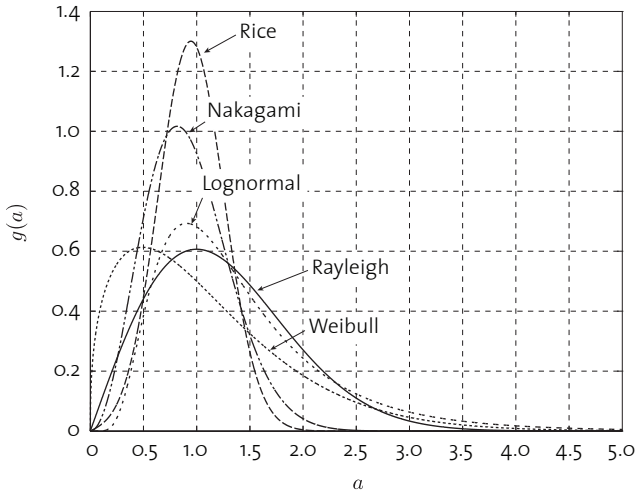


Fig. 5.1: Plot of the PDFs used in the simulation experiment.

Hence, to obtain at least some experience and intuition about the operation and limitation of AICC before actually applying it to our measurement data, we perform a Monte-Carlo study. In particular, we use Matlab's random number generator to generate sample vectors according to five specific distributions from the five families listed in Section 5.3.1. For each family, the parameters below are similar to those extracted from measurements reported in the literature.

- Rayleigh: $\sigma^2 = 2$
- Rice: $K = 4, \Omega = 1$
- Nakagami: $m = 1.5, \Omega = 1$
- lognormal: $\mu = 0.2, \sigma^2 = 0.3$
- Weibull: $b = 1.2, d = 1.4$

The corresponding PDFs $g_q^j(a)$ are shown in Figure 5.1. We generate 90 i.i.d. random vectors for every listed distribution, and compute AICC values for all five families; the whole process is repeated 10 000 times. To compare the fit of each family, we use the Akaike weights presented in Section 5.4.3.

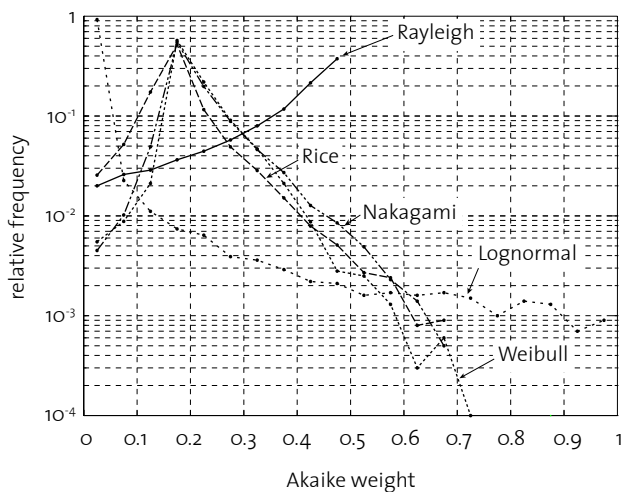


Fig. 5.2: Histogram of Akaike weights; generating distribution is Rayleigh.

The simulation study is not in complete agreement with the assumptions on the basis of which we want to apply AIC or AICC to our measurement data. The five operating distributions used in the simulation experiment are very simple, with only one or two parameters, while we assume the operating distribution of the actual channel taps to be of high complexity. Furthermore, the five operating distributions used here are part of the candidate family by construction, while this is most certainly not the case for the channel operating distribution. Hence, the results of our simulation study might not necessarily be representative for the performance of AICC on the basis of our measurement data. Nevertheless, we can hope to get some understanding of the general behavior of AICC and the Akaike weights.

A. Rayleigh

For the Rayleigh operating distribution with $\sigma^2 = 2$, Figure 5.2 shows a normalized histogram for the 10 000 Akaike weights com-

5 MARGINAL TAP DISTRIBUTIONS

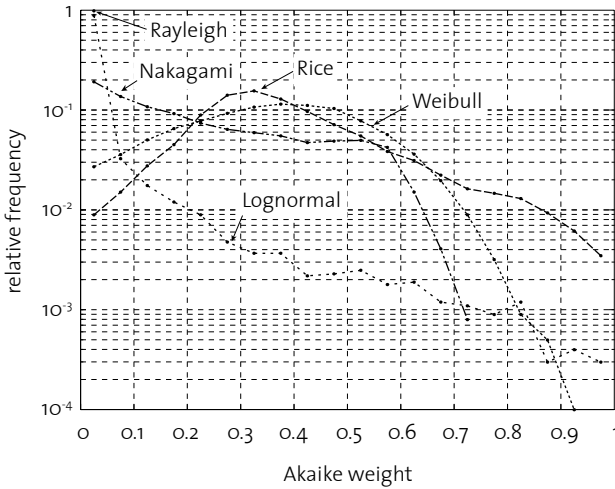


Fig. 5.3: Histogram of Akaike weights; generating distribution is Ricean. The Rayleigh candidate family results in a single mass point at an Akaike weight of 0.

puted for each candidate family. Two important observations are that (i) Rayleigh provides indeed the best fit, the corresponding weight is largest with high relative frequency, and (ii) the variance of the weights is quite high. The Nakagami, Rice, and Weibull distributions, which contain the Rayleigh distribution as a special case, come in second. In comparison with the Rayleigh family, we may ascribe the difference to the second parameter in each of these families, which penalizes the corresponding AICC values. Indeed, we find the maximum AICC difference between the Rayleigh distribution on the one hand and the Rice, Nakagami, and Weibull distributions on the other hand to be around -2 , the penalty incurred in (5.32) by the extra parameter. The best lognormal distribution fits significantly worse than all the other families; this might result because the lognormal family does not reduce to the Rayleigh family for any parameter settings.

B. Rice

If the generating distribution is Ricean with the above-listed parameters, it is difficult to decide which of the Weibull, Nakagami, or Rice distributions provides the best fit, as can be observed in the histogram in Figure 5.3. Over a reasonable range of parameters, said families can be parameterized in a way that they closely approximate each other, as discussed in Section 5.3.1. Thus, the large variance of the corresponding Akaike weights can be interpreted within the model selection framework as lack of evidence—the available data does not allow a better distinction between these three distributions in the simulated scenario. On the other hand, even the best-fitting lognormal distribution does provide a consistently bad fit. The Rayleigh distribution is even worse—it is indeed so bad that it has a zero weight with relative frequency one, as there is only a single data point in the histogram.

C. Nakagami

The weight histogram in Figure 5.4, which results from the Nakagami operating distribution with $m = 1.5$ and $\Omega = 1$, is very similar to the one in Figure 5.3 for the Ricean operating distribution. This observation agrees with the earlier remarks that the Rice, Nakagami, and Weibull distributions approximate each other for a certain range of parameters. It seems as if the approximation is not as good for the chosen parameter setting than in the previous example, but the Rice, Nakagami, and Weibull distributions still provide a much better fit than the Rayleigh and lognormal distributions.

D. Lognormal

We already observed in the previous examples that the lognormal family seems to be different from the rest as it does not fit data generated according to either a Rayleigh, a Rice, or a Nakagami distribution. Figure 5.5 shows that the converse also seems to be true: data generated according to the lognormal distribution with

5 MARGINAL TAP DISTRIBUTIONS

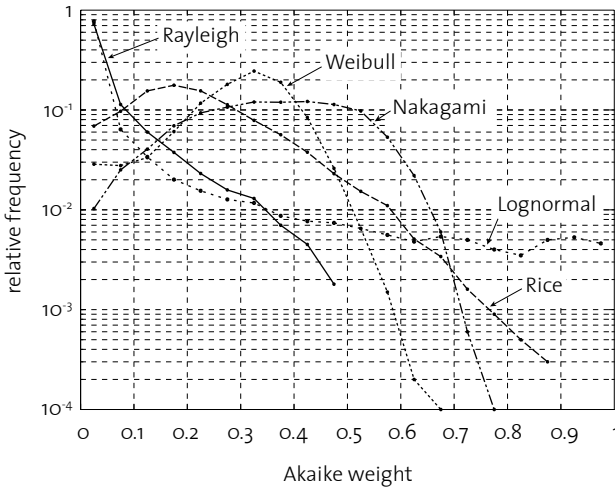


Fig. 5.4: Histogram of Akaike weights; generating distribution is Nakagami.

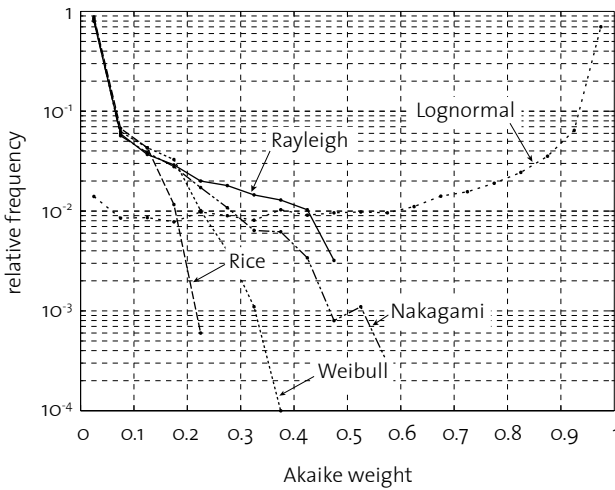


Fig. 5.5: Histogram of Akaike weights; generating distribution is lognormal.

5.4 AKAIKE'S INFORMATION CRITERION

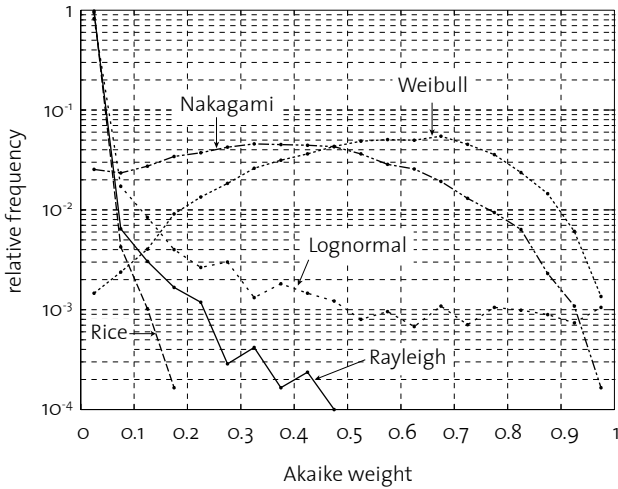


Fig. 5.6: Histogram of Akaike weights; generating distribution is Weibull.

parameters as above seems to be difficult to fit with any of the four other candidate families, while the lognormal family itself shows an excellent fit. This clear-cut picture is important insofar as we can expect to obtain a clear-cut answer to the question if the lognormal distribution is indeed appropriate to model small-scale amplitude fading of wideband channels.

E. Weibull

The last example uses the Weibull distribution with parameters $b = 1.2$ and $d = 1.4$ as operating distribution. This distribution seems to be rather different than anything the Rayleigh, Rice, or lognormal family can approximate. Only the Nakagami family provides a good approximation, as can be seen in Figure 5.6. As already observed several times, the corresponding weights show large variability, which indicates that there is not enough evidence in the data to warrant a better discrimination of the two distributions.

5.4.5. Comments

As the use of AIC to select suitable amplitude fading distributions on the basis of measured channel impulse responses seems to be novel, we cannot rely on past experience to guide us in its application. The method still has a number of weak spots, both in the underlying theory as well as in its application. We discuss the—in our opinion—most important ones below.

A. Fundamental Assumptions

The derivation of AIC, as just outlined, rests on two crucial assumptions: the operating distribution is part of the candidate family, and the asymptotic criterion is a sensible estimator also for finite sample sizes.

We have no reason to believe that the operating distribution of the channel tap amplitudes is contained in the candidate set of the rather simple families listed in Section 5.3.1—in fact, the main reason why we resort to a stochastic channel model is to reduce model complexity. No rigorous derivation of a model selection criterion seems to exist for the case when the operating distribution is not part of the candidate set. The only justification to nevertheless use AIC, and in our opinion a strong justification at present, is that the approach of model selection by discrepancy minimization as such is theoretically sound, much more so than any other statistical method currently used to determine suitable small-scale fading amplitude distributions. In particular, the necessity to judge how much information is actually contained in a finite number of channel samples and to avoid overfitting seems to be overlooked far too often in the context of wireless channel modeling. The simulation examples by Burnham and Anderson (2002) and our own simulation study in the preceding subsection support the claim by Zucchini (2000) and others that AIC provides a sensible approach to model selection as long as the best-fitting distributions in each candidate family are not too far off.

How many samples are necessary for AIC or AICC to yield sensible

results is difficult to answer. In both MCI and MCII we have enough samples N per measurement set to satisfy the rule of thumb by Sakamoto et al. (1986) that the number U of free parameters should be smaller than $2\sqrt{N}$. However, our simulation studies showed that the variance of the Akaike weights might be quite large across a number of independent experiments. Yet, this variability was most pronounced in the cases where several families were similarly well suited to approximate the operating distribution. In such a situation, we should not try to extract more information than provided by the data and expect to obtain an unequivocal answer. Instead, if several distributions provide an adequate fit, we are at liberty to choose the one that is most convenient to work with, i.e., that best satisfies our criterion of mathematical tractability.

B. Do We Need AIC?

All our candidate families from Section 5.3.1 are completely characterized by either one or two parameters; and out of the five families under consideration, four have two parameters. This implies that the second term in AIC and AICC is ineffective for the comparison of four out of five families. Hence, is AIC actually necessary to select a small-scale fading distribution? In fact, we could as well just compute the expected log-likelihood for each fitted distribution, make a comparison of distributions on the basis of this measure, and, if in doubt, select the simplest distribution. Essentially, this is all that AIC amounts to in our case. Nevertheless, we believe that the approach we took is more comprehensive than evinced by the two formulas (5.31) and (5.32) that we finally use to compute statistics.

- The expected log-likelihood that is the first term in AIC and AICC results from a well-defined discrepancy, the KL discrepancy. This already is an important improvement over many techniques commonly used in wireless channel modeling, which are all to some extent *ad-hoc*.
- The fact that the available data limits the complexity of a suitable distribution is often disregarded. AIC directly shows when this

issue is of importance, even if only for a single family in the candidate set.

- The interpretation of the Akaike weights as the likelihood of corresponding fitted distributions is contingent on the approach taken and yields valuable additional insight because it allows us to directly compare the approximation quality of different distributions.

In summary, we believe that the model selection approach as such is sensible for the application at hand, even though we do not use it to its full potential.

C. Measurement Noise

We base our approach to model selection on the assumption that our data are channel impulse responses. Real life measurement data is contaminated with noise picked up by the receiver and the different amplification stages, and by interference. A Gaussian noise model is a very good assumption in most cases, since noise is created by many independent sources, such that the CLT is valid. Hence, it would be possible to include the contribution of the noise into the model selection process. Since the effect of the noise is almost always additive and independent of the signal, the resulting overall density of the output would be the convolution of the candidate PDF with the noise PDF. In consequences, all resulting models would have one more free parameter, the noise variance. However, including the noise effect into the model has several drawbacks: the models might become analytically intractable, often it would not be possible to compute the resulting PDF in the first place; parameter estimation would be even more difficult. Most important, since the noise is well below the signal level, we would attempt to include effects in our model that are too weak to be quantified using the small number of samples available in MCI. Hence, we would counteract the parsimonious approach taken in selecting the candidate models and built into the model selection criterion. For these reasons we opt against inclusion of the noise effects into the model and attempted instead to obtain measurements

with the best possible SNR, ensuring that the noise effects are indeed negligible.

5.4.6. Other Model Selection Criteria

AIC has proven very popular for linear models; e.g., for the regressive and autoregressive models ubiquitous in most empirical sciences, AIC is used to select the model order. Specialized applications are discussed in the book by Sakamoto et al. (1986) and in a tutorial paper by Stoica and Selén (2004). Another common application is selecting the number of sources in array processing along the lines of Wax and Kailath (1985). Because the class of linear models is highly structured, a more detailed analysis of AIC is possible, and the approximations in the derivation of the general case, e.g., in (5.29), often hold with equality.

Following Akaike's work, a wealth of other model selection criteria have been proposed. Some of them are just modifications of Akaike's original result, others are derived taking a completely different approach. The most popular ones are the criterion of minimum description length (MDL), proposed by Rissanen, and the Bayes information criterion (BIC), derived by Schwarz.

MDL is based on the theory of stochastic complexity; the approximation quality of a distribution is measured through the length of the shortest description for the data at hand, i.e., the shortest encoding. In this sense, MDL is related to relative entropy, the basis of AIC. The simplest derivation of MDL uses two-stage coding, where in a first step, for each model family parameters are estimated and quantized in steps of size $1/\sqrt{N}$. In a second step, the parameters are encoded under the assumption of a uniform distribution while the data itself is encoded following the fitted distribution. The total description length is thus the sum of the code length for the data and the code length for the parameters, which is for i.i.d. data

$$\sum_{n=1}^N \log g_{\hat{q}}(a_n) + \frac{U}{2} \log N. \quad (5.35)$$

An introduction to MDL can be found in a tutorial paper by Hansen and Yu (2001).

BIC is asymptotically equivalent to MDL, although the theoretical starting point of BIC is to attach a prior probability to each family and then to select the fitted distribution with the highest posterior probability of being correct.

Although AIC and MDL/BIC differ only in their respective second term, the theoretical implications are far-reaching. For linear models, MDL/BIC is consistent if the operating distribution is of finite dimension, i.e., the correct model order is selected in the large-sample limit. On the other hand, if the operating model is of infinite order, AIC selects the distribution with minimal squared error, i.e., it is efficient. These two characteristics are mutually exclusive—a consistent model selection procedure is not efficient if the operating model is of infinite dimension and vice versa. (Shibata, 1989; Yang, 2005). Heuristically extrapolating these observations, AIC should be preferred in our case, as we expect the operating model to be highly complex. In practice, though, the difference in the second term between AIC and MDL/BIC is of no importance for the set of candidate families we consider.

5.5. MCI MARGINAL TAP DISTRIBUTIONS

In the present and the next section, we use AIC and AICC to assess which of the proposed amplitude fading distributions listed in Section 5.3.1 is best suited to statistically describe the small-scale fading phenomenon in the two different types of wideband channels we measured in MCI and MCII. This analysis only concerns the marginal distribution of each tap in the complex baseband channel impulse responses because a complete characterization of the joint distribution is infeasible for lack of sufficient channel samples. In accordance with the approach outlined in Section 5.3, we attempt to characterize the joint distribution of the discrete-time wideband block-fading channel model up to second order in Chapter 6.

To assess the fit of the various candidate families listed in Sec-

tion 5.3.1, we compute the AICC values according to (5.31) and (5.32), the AICC differences as in (5.33), and the Akaike weights following (5.34), for all measurement sets in the MCI LOS and OLOS settings. The parameter estimates for each family that are needed in the computation of AICC are detailed in Section 5.3.2. Each AICC value is computed from $N = 90$ spatial samples. The impulse responses and PDPs of the different measurement sets consist of a large number of taps. Figure 5.7 and Figure 5.8 are our attempts to visualize the Akaike weights for all taps of the channel vector in a given measurement set in the LOS and OLOS setting, respectively. The PDP of the corresponding measurement set is shown in the bottom-most graph for reference. Similar plots for other measurement sets, i.e., taken at different distances d between transmitter and receiver, look very similar to the ones shown; hence, we refrain from presenting them here.

5.5.1. Observations

The plots in Figure 5.7 and Figure 5.8 show the probability of best fit of each fitted candidate model for each channel tap; in addition, they also show interesting relations between different taps.

- Viewed globally, the fitted Rayleigh amplitude distribution provides the best fit for most channel taps in both the LOS and the OLOS measurement set shown. The Akaike weight of the Rayleigh distribution never exceeds 0.5, though, while significant downward variability across different taps exists.
- Distributions from the Rice, Nakagami, and Weibull families provide a very similar fit in total across all taps of a given channel vector. Their Akaike weights are of similar magnitude, on average around 0.2. However, these weights do vary significantly around this mean value across taps.
- The lognormal family does not contain distributions that provide a fit comparable to the Rayleigh distribution; for most of the taps, the lognormal fit is also far less than that of the Rice, Nakagami, and Weibull families; Akaike weights for lognormal distributions

5 MARGINAL TAP DISTRIBUTIONS

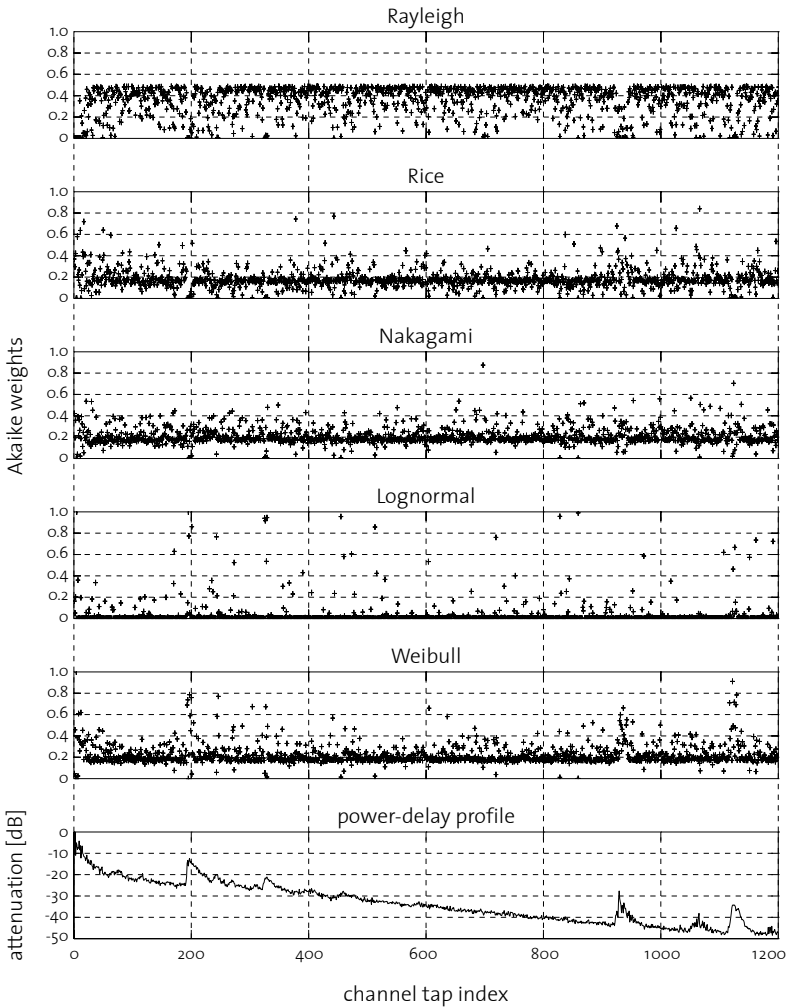


Fig. 5.7: MCI: Akaike weights and PDP, LOS setting, $d = 27.2$ m.

5.5 MCI MARGINAL TAP DISTRIBUTIONS

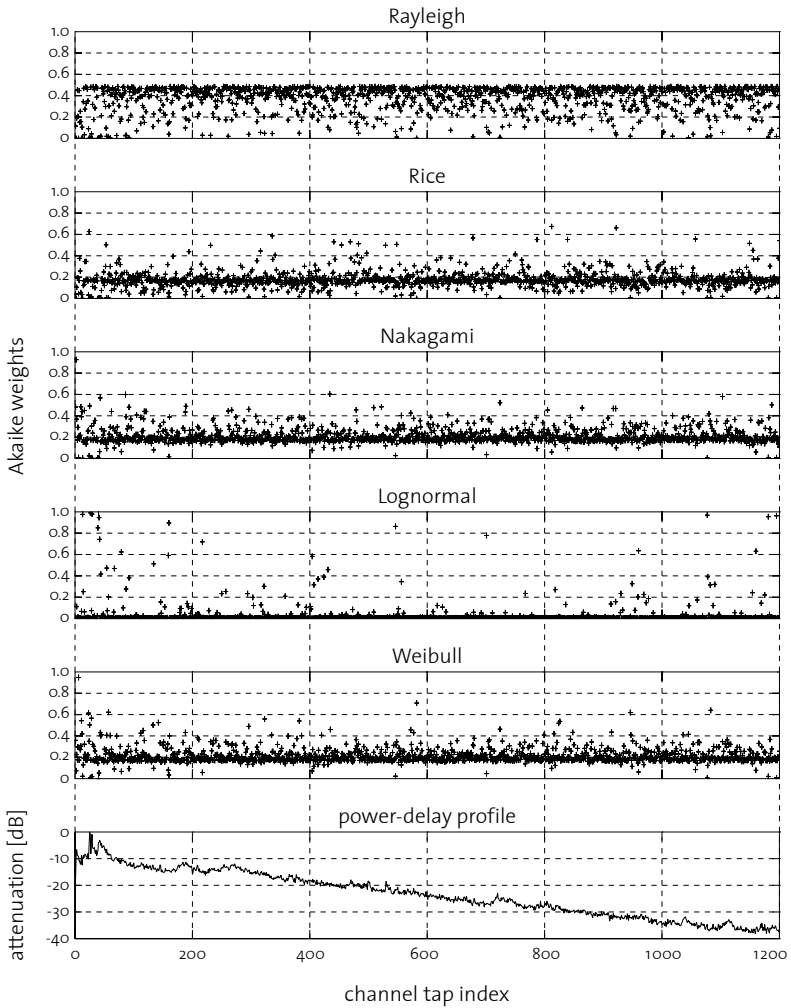


Fig. 5.8: MCI: Akaike weights and PDP, OLOS setting, $d = 27.4$ m.

are often close to 0.

- Especially in the LOS measurement set, but also at the beginning of the OLOS impulse response, there exist short sequences of taps where the lognormal family provides the best fit, with Akaike weights close to 1. Similarly, for some consecutive taps the Rice, Nakagami, and Weibull families show an improved quality of fit compared with most other taps. These short sequences of taps seem to correspond to the peaks of clusters in the corresponding PDP. More such short runs of good lognormal fit or good fit of the Rice, Nakagami, and Weibull fit appear in the LOS plot in Figure 5.7 than in the OLOS plot in Figure 5.8.

Some of these observations are surprising as they seem to contradict results of other measurement campaigns. Before analyzing these observations, though, we complement them by some related statistical quantities.

Figure 5.9 shows, on a logarithmic scale, the estimates of the Ricean K -factor and the Nakagami- m parameter for the LOS measurement set with $d = 27.2$ m. Both parameters are usually taken as an indicator of how severe the fading is in a given tap; smaller values indicate more severe fading. Most of the estimates are small, in the Ricean case close to or below 0 dB, and in the Nakagami case around 0 dB. For these parameter settings, both distributions are very close to the Rayleigh distribution. But for some groups of taps, especially the ones around tap indices 936 and 1065, the Rice and Nakagami parameter estimates are significantly larger, while around the tap indices 197 and 1122 they are smaller than average. These taps correspond to specific cluster peaks in the PDP of Figure 5.7. The parameter estimates for OLOS measurement sets not shown here are even closer to the values that reduce the Rice respectively Nakagami distribution to the Rayleigh distribution. Only for the first few taps does the Ricean K -factor significantly exceed 5 dB and the Nakagami- m parameter 2 dB.

A special behavior of the channel taps at the beginning of the impulse response and at the cluster peaks in the LOS measurement

5.5 MCI MARGINAL TAP DISTRIBUTIONS

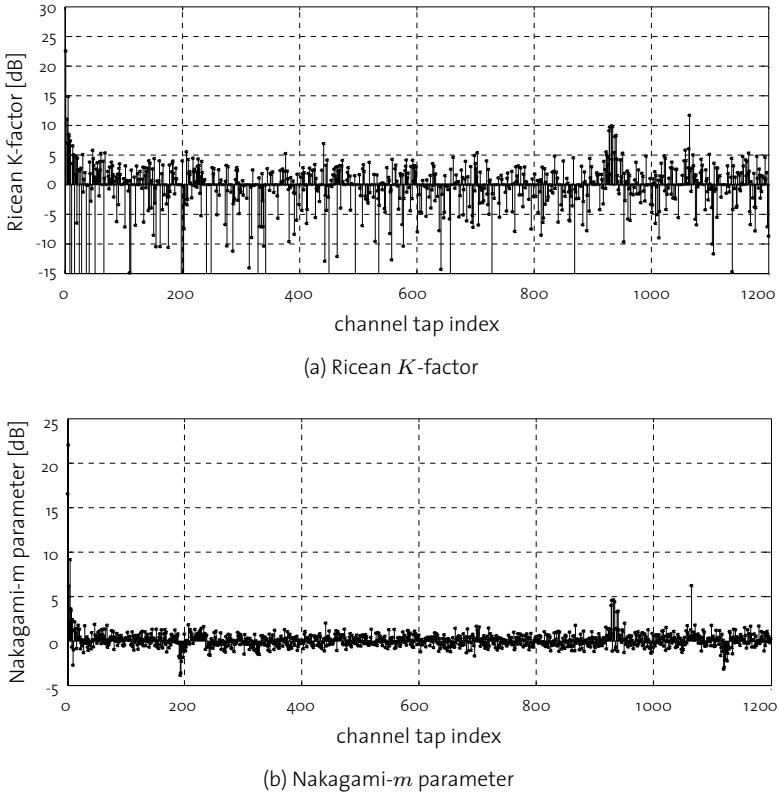
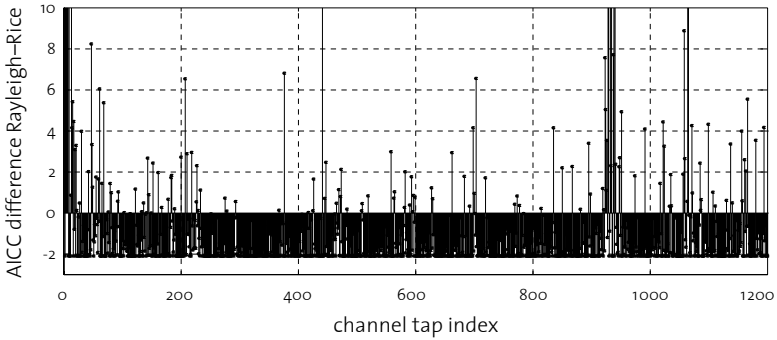


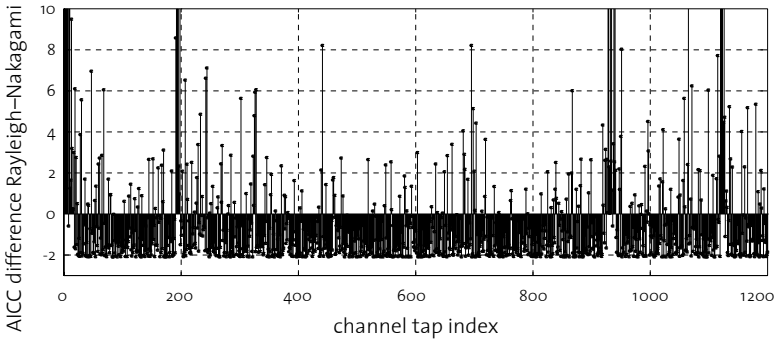
Fig. 5.9: Parameter estimates for the MCI LOS measurement set with $d = 27.2$ m.

set around tap indices 197, 936, 1065 and 1122 can also be observed directly in the AICC differences ϕ_j defined in (5.33). For the same measurement set as in the previous plots, i.e., MCI LOS at $d = 27.2$ m, Figure 5.10 shows the difference between the Rayleigh AICC values and the Rice, respectively Nakagami AICC values. While for the vast majority of taps the difference is approximately -2, the taps just mentioned show a positive difference, even a very large one at times—we had to clip some AICC differences in Figure 5.10 because the corresponding values were too large to be plotted, up to $\phi \approx 400$. The

5 MARGINAL TAP DISTRIBUTIONS



(a) Difference between the Rayleigh and the Rice AICC values.



(b) Difference between the Rayleigh and the Nakagami AICC values.

Fig. 5.10: Difference between AICC values, MC I LOS, $d = 27.2$ m.

AICC differences between the fitted Rayleigh and Weibull distributions are similar to the differences just discussed; therefore, we do not plot the Weibull differences here.

Except for taps that correspond to cluster peaks, neither the Akaike weights and the corresponding AICC differences nor the parameter estimates show a trend over the length of the impulse response; all listed quantities vary considerably from one tap to the next, but these changes appear to be purely random.

5.5.2. Analysis and Interpretation

Especially in the LOS setting, but also in our OLOS measurement sets at the onset of the impulse response, there is a clear distinction between the many *ordinary taps*, and the few *cluster taps*, i.e., taps that correspond to a cluster peak. As the difference between ordinary taps and cluster taps is striking in Figure 5.7 and Figure 5.8, we discuss the two separately.

A. Ordinary Taps

Overall, the Rayleigh distribution seems to provide the best fit, except for cluster peaks in the LOS impulse responses and at the beginning of the OLOS impulse responses. However, the variation of the Akaike weights across taps is high.

In most taps where the Rayleigh distribution has the largest Akaike weight, the weight differences to the fitted Rice, Nakagami, and Weibull distributions are small, which means that all these distributions provide an adequate fit with nonnegligible probability. Even more, Figure 5.10 shows that the AICC differences between the Rayleigh distribution and the Rice, Nakagami, and Weibull distributions is around -2 in most of these taps. But -2 is exactly the difference that result from the term $2U$ in (5.31) that penalizes more complex families, because the Rice, Nakagami, and Weibull families have one more parameter than the Rayleigh family. Together with the estimated parameters of the Rice and Nakagami families in Figure 5.9, which are often such that they reduce the fitted distributions to the Rayleigh distribution, there is ample evidence that many individual taps are best modeled by the simplest family available, the Rayleigh family.

But what about the ordinary taps where the Rayleigh distribution does not fit best? If we strictly follow the theory outlined in Section 5.3, we cannot say more than the Akaike weights tell us—that in those taps another distribution has a higher probability of best fit. It is now a modeling decision to weight the evidence for another distribution

against the simplicity that results from the use of a single family to describe all marginal tap amplitude distributions.

Additional information beyond what is contained in the AICC differences and Akaike weights can be used to help making the aforementioned modeling decision. The large variability of the Akaike weights between adjacent taps cannot be quantified from within the theory at present. But we discussed in Section 5.3.2 that some of our parameter estimates have a large error variance, especially the estimate of the Ricean K -factor. Furthermore, the simulation study in Section 5.4.4 indicated that a large variation in Akaike weights can be observed even when the operating distribution is known, simple, and part of the candidate set. In fact, the histogram in Figure 5.2 for the case of a Rayleigh operating distribution is very similar to histograms for Akaike weights computed across the channel taps of OLOS impulse responses, and also for LOS impulse responses if we exclude the taps around cluster peaks. Because said channel taps cannot be taken as i.i.d. random variables—already their variances differ, as seen in any PDP—this observation does not show that the variability of Akaike weights solely manifests the inherent estimator variance; but we can take it as an indication that modeling all taps as Rayleigh distributed is not a bad modeling assumption. This modeling assumption is also supported by two more arguments. Physically, there does not seem to be a reason for a significant change over the duration of the impulse response in the composition of the arriving partial waves. If at all, we expect that more partial waves average in later taps than in earlier taps because of multiple reflections, diffraction, and scattering of paths. From an estimation-theoretic viewpoint, the variation in Akaike weights over taps indicates that we do not have enough data to obtain low estimator variance. Hence, a parsimonious modeling choice is to use the same family for all taps; everything else would amount to overfitting in the same way as a family with two many parameters provides a better fit at the cost of increased predictive power of the resulting model.

B. Cluster Taps

The cluster taps are clearly distinct from the ordinary taps in that the Akaike weights for those taps do not vary in the same way. For all cluster taps, the Rayleigh fit is rather bad, e.g., for the LOS measurement set in Figure 5.7 with Akaike weights close to 0 for the first taps and for the clusters around tap indices 197, 325, 936, and 1122. The Weibull family shows a good fit for the taps that correspond to the LOS path and the clusters at tap indices 197, 936, and 1122, the Rice family seems suitable for cluster taps around indices 936 and 1065, and the lognormal family provides a reasonable fit for the taps around indices 197, 325, and 1122. One possible and plausible explanation for this striking difference of the cluster taps compared with the ordinary taps is the variation of the arrival times of strong specular paths as the transmit antenna is moved over the virtual array, in combination with the temporal alignment of the first arrival described in Section 4.5.1.

- The LOS path is the one aligned in time. If the paths around tap indices 936 and 1065 in the LOS measurement are caused by a double reflection off the window at the top and the brick wall at the bottom right-hand side in the floor plan of Figure 4.7, they will be similarly aligned as these paths are collinear with the LOS path so that their path length differential upon moving the transmit antenna will be equal to the path length differential of the LOS path. The Weibull and Rice distributions provide the best fit for the taps around indices 936 and 1065, as indicated by their high Akaike weight in Figure 5.7. A look at the corresponding parameter estimates reveals that both fitted distributions for these cluster taps have a very light right tail; the corresponding Weibull parameter b is on the order 10^{-2} in all LOS measurement sets while the parameter d is large, between 6 and 9. Similarly, the Ricean K -factor for the taps around indices 936 and 1065 is large, as shown in Figure 5.9. Both observations indicate little fading. This finding can be physically expected for aligned specular arrivals.

- For the cluster peaks around tap indices 197, 325, and 1122, both the lognormal and the Weibull family provide a good fit. But for the corresponding distribution parameters estimated from our data, the lognormal and Weibull PDFs have a heavy right tail for these taps. A possible explanation is that a nonaligned specular reflection does not always contribute to the same tap for all 90 measured impulse response vectors in a given measurement set but appears as an outlier in many samples. Because the $N = 90$ impulse responses of a given measurement set are recorded at different locations of the transmitting antenna on the virtual array, the corresponding delay of a specular reflection changes from one measured impulse response to the next. We aligned the LOS arrival, as discussed in Section 4.5.1.B, but this alignment step does not apply to specular paths that are incident under a different AOA than the LOS path. Hence, a strong specular reflection may appear as an outlier in several adjacent taps. Strong outliers increase the weight of the tail of a PDF fitted to the measurements of the affected taps, which might explain the good fit of the lognormal and Weibull distributions.

5.6. MC II MARGINAL TAP DISTRIBUTIONS

We compute the Akaike weights for the three MCII measurement sets in the same way as for MCI data. The only difference is that we used AIC instead of AICC, because the number of samples in each MCII measurement set is much larger than in the MCI measurement sets. Indeed, we could observe hardly any difference between AIC and AICC values computed from MCII data.

The AIC values are shown in Figure 5.11 for the LOS measurement set with $N = 1011$ samples, in Figure 5.12 for the OLOS measurement set with $N = 2722$ samples, and in Figure 5.13 for the NLOS measurement set with $N = 1256$ samples. As before, we also include in each figure the corresponding PDP for reference.

5.6 MC II MARGINAL TAP DISTRIBUTIONS

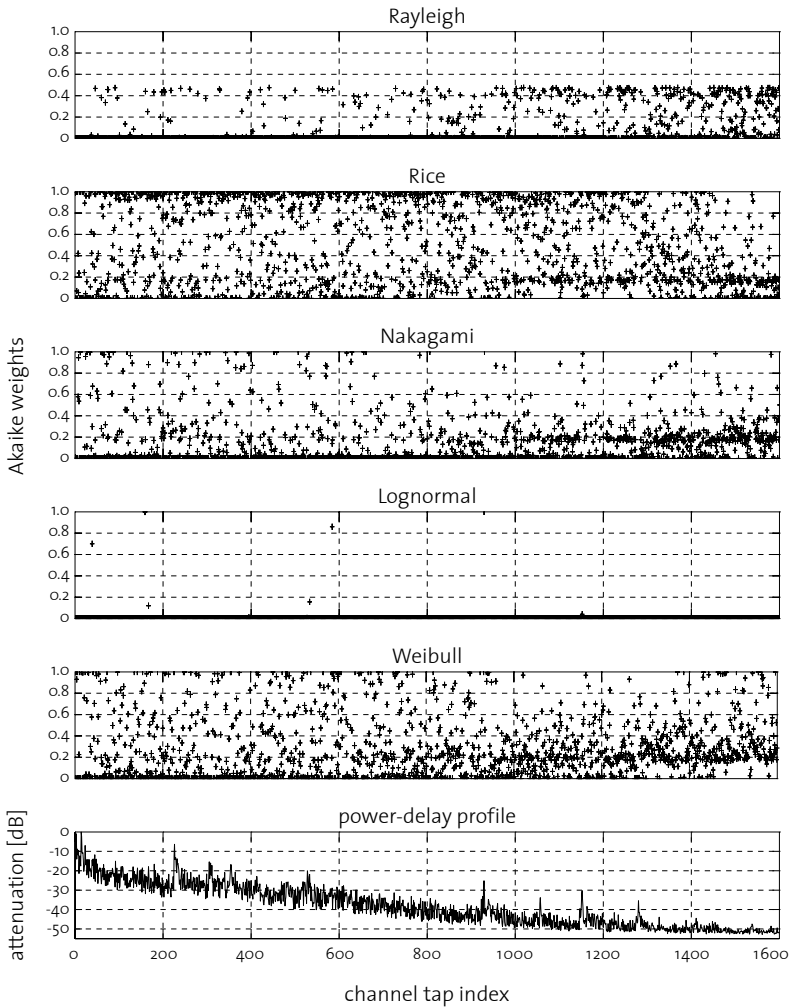


Fig. 5.11: MC II: Akaike weights and PDP, LOS setting, $d = 20$ m.

5 MARGINAL TAP DISTRIBUTIONS

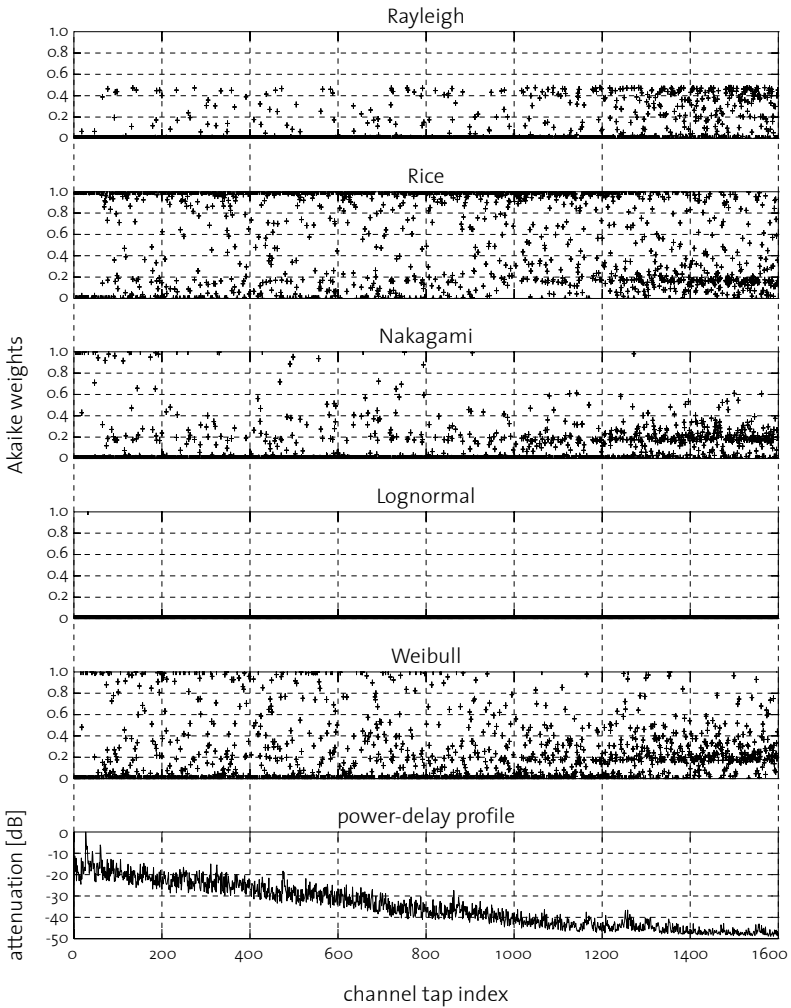


Fig. 5.12: MC II: Akaike weights and PDP, OLOS setting, $d = 20$ m.

5.6 MC II MARGINAL TAP DISTRIBUTIONS

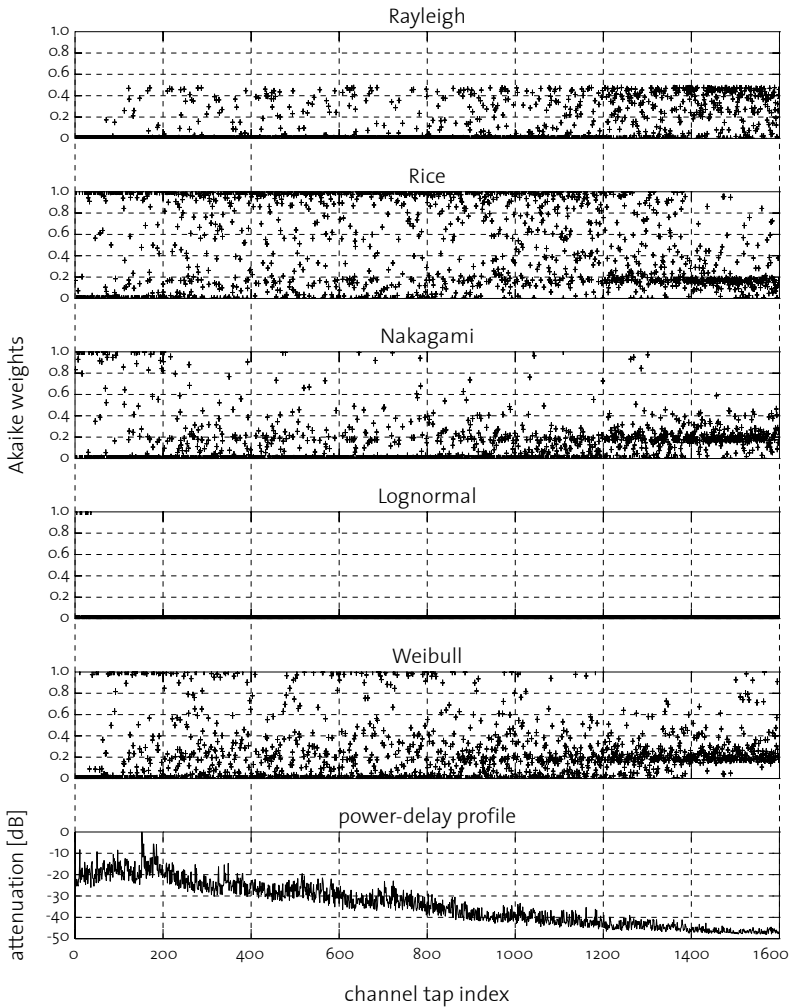
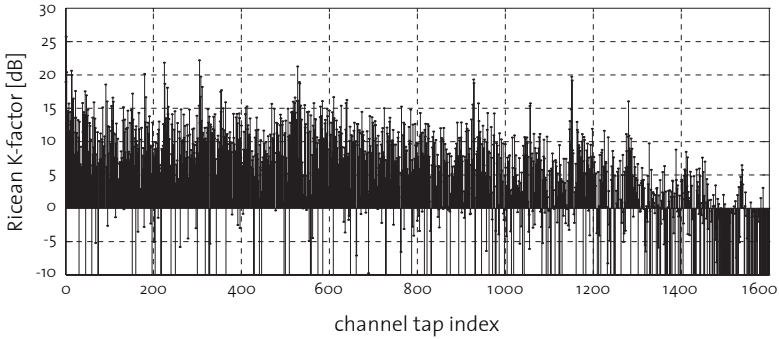
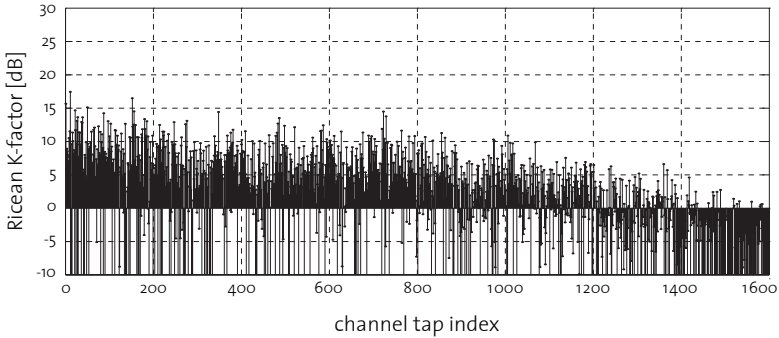


Fig. 5.13: MC II: Akaike weights and PDP, NLOS setting, $d = 13$ m.

5 MARGINAL TAP DISTRIBUTIONS



(a) MC II LOS, $d = 20$ m



(b) MC II NLOS, $d = 13$ m

Fig. 5.14: Ricean K -factors estimated from MC II measurements.

5.6.1. Observations

All three figures are quite different compared with the ones computed from MCI data.

- Rice distributions provide the best fit for most taps up to around tap index 1200 in all three measurement sets. The Weibull family also has a high probability of fit in some of the taps in this range. The Nakagami family seems to be suitable for taps at the beginning of the impulse responses in all three settings, and only

sporadically useful for later taps.

- The Rayleigh family provides a good fit for later taps, starting around index 1200 in all three settings. For these later taps, the Akaike weight plots very much resemble the plots for the ordinary taps in all MCI LOS and MCI OLOS measurement sets.
- The lognormal family is almost never adequate to model small scale amplitude fading in all three measurement sets. Indeed, the Akaike weights for the lognormal family are often so small that they are indistinguishable from zero in all three plots.
- Although there do exist clusters in the LOS PDP, and to a lesser extent also in the OLOS and NLOS PDPs, there is no apparent change in the Akaike weights of the taps that correspond to cluster peaks.

The estimated Ricean K -factor for the LOS and NLOS setting in Figure 5.14 is quite high in both cases, compared with the K -factor from MCI shown in Figure 5.9a. Only for later taps, starting around tap index 1200, does the MCII K -factor become as small as in MCI. The sporadic low values result from convergence problems of the parameter estimation algorithm, as described in Section 5.3.2.

5.6.2. Analysis and Interpretation

Ricean fading is often attributed to a strong mean component in an impulse response tap (Vaughan and Bach Andersen, 2003), which renders the affected tap less variable compared with Rayleigh fading. Most statistical models for wireless channels that make use of the Rice family do so only for the first tap of the impulse response vector, because a strong LOS path is usually taken as physical justification. The underlying modeling assumption is that the first tap is proper Gaussian distributed with nonzero mean.

The Akaike weights for all three MCII measurement sets indicate a good fit of the Rice family not just for the first tap but for many taps of the impulse response vector. We would like to know if this fit indeed results from a strong mean component in every tap, because a Ricean envelope can also result if the underlying complex-valued channel tap

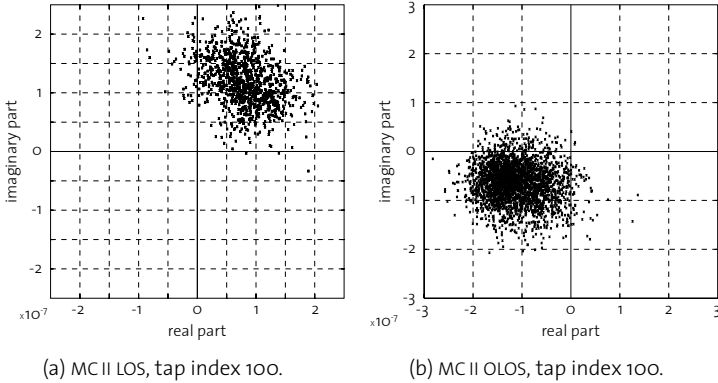


Fig. 5.15: Scatter plots of a randomly selected MCII tap.

has zero mean; e.g., from the multiplication of a nonzero-mean JPG random variable with a random variable of constant modulus and uniformly distributed phase. To exclude the possibility of a zero-mean tap distribution giving rise to the Rice amplitude distribution, we inspect some scatter plots of individual impulse response taps, as shown in Figure 5.15. Both plots indicate that all samples within one measurement set cluster in one quadrant of the complex plane. We cannot inspect every single tap in this way, but the random taps we analyzed all show the same clustering. Hence, it is very likely that the Rice amplitude distribution indeed fits because there is a strong mean component in every tap. Further evidence for this hypothesis follows from our analysis of MCII PDPs in Section 4.6.1. The complete PDP in Figure 4.18 looks rather random, while the PDP computed from the same data but with the estimated mean subtracted in Figure 4.19 looks much less erratic. This behavior indicates that it is the mean component in every tap that is responsible for the ragged appearance of the PDP in Figure 4.18.

A physical explanation for the presence of a strong mean component in every single tap might be as follows: While in MCI, most scattering and reflecting objects move relative to the position of the

antennas, the major reflectors and scatterers in MCII are static. Only persons did move about in the lobby during the measurements; the energy scattered by them is much less compared to the amount of energy transferred by the static environment. The channel variations induced by moving persons lead to small fluctuations around the mean components from specular reflections. Similar observations were already reported by Hashemi et al. (1994) on the basis of narrowband measurements and by Bultitude et al. (1989) for indoor channels of 80 MHz bandwidth.

Because transmit and receive antennas were mounted approximately 1.6 m above the floor, people moving in the environment sometimes blocked the LOS or propagation paths of some of the dominant reflections. Consequently, it is not possible to perfectly separate small-scale fading and shadowing in MCII, as we will discuss in more detail in Section 6.4.2. Maybe it is this mix of shadowing and small-scale fading that leads to a good fit of the Weibull family for the corresponding shadowed taps. In a recent study of temporal variation in a UWB channel with fixed antennas, Pagani and Pajusco (2004) also find a good fit of Weibull distributions. However, their measurement methodology is different from MCII.

Approximately from tap 1200 on, the Rayleigh family exhibits a better fit than the Rice family for all three MCII measurement sets. Indeed, the Akaike weights for the affected taps are very similar to the ones obtained from MCI data for all impulse response taps. Therefore, we hypothesize that all those taps are best modeled to be Rayleigh distributed. The Rayleigh family arises if the underlying complex-valued channel taps are zero-mean proper Gaussian, the family most often used to model thermal noise. Hence, we believe that the good fit of the Rayleigh family from tap 1200 onward results because the measurement SNR of these taps is so low that we are essentially fitting Gaussian-distributed noise.

5.7. COMPLEX TAP DISTRIBUTIONS

We use AIC to select a suitable amplitude distribution for the channel taps, but did not look at the tap phases so far. In this section we attempt to characterize the phase distribution of the individual channel taps; however, there are several issues that might lead to results less strong than for the amplitude distributions.

- First and foremost, we lack physically motivated candidate distributions for the tap phases. If the taps are proper Gaussian distributed, the phase of each tap around a potentially present mean component is uniform. Hence, the Rayleigh and Rice amplitude distributions come with a corresponding phase distribution that derives from the physical explanation of the underlying complex tap distribution. But there do not seem to exist such companion phase distributions for the other families we investigated, i.e., phase distributions that complement the Nakagami, lognormal, and Weibull amplitude distributions.
- We invested much effort to devise a postprocessing methodology that preserves phase information as well as possible. Nevertheless, the quality of the phase information in our complex baseband impulse responses is of much lower quality than the amplitude information, especially for MCII where alignment of the first path is difficult.

Because we lack an alternative to the uniform phase distribution, because this distribution is the physically sensible companion to the Rice and Rayleigh families that we found to provide the best fit for many taps in MCI and MCII impulse responses, respectively, and because the uniform distribution is the maximum entropy distribution barring any further information, all we can do in this section is to evaluate if the phases of individual channel taps can be sensibly modeled as uniform around a potentially present mean or not. We are thus faced with a binary decision problem instead of a model selection problem; a GOF test is a suitable statistical tool in this case.

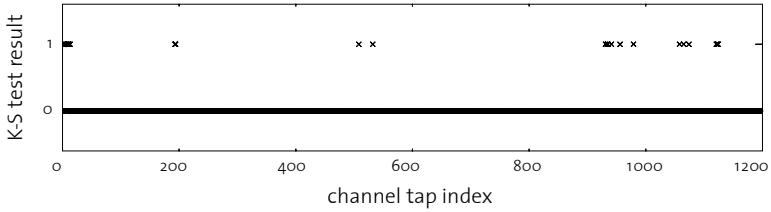
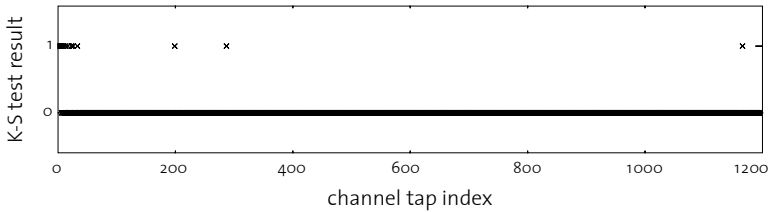
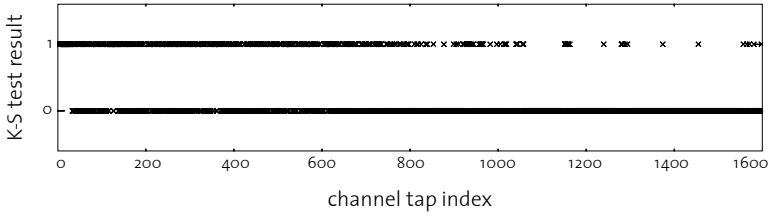
(a) MCI LOS, $d = 27.2$ m.(b) MCI OLOS, $d = 27.4$ m.

Fig. 5.16: Results of the Kolmogorov-Smirnov GOF test for MCI data under the challenging hypothesis of a uniform phase distribution.

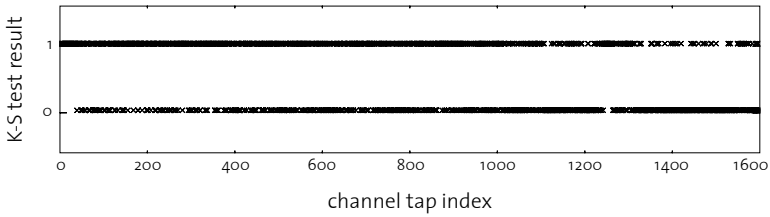
A. Measurement Campaign I

Figure 5.16 shows the results of a series of Kolmogorov-Smirnov GOF test results (Bartoszyński and Niewiadomska-Bugaj, 1996; Zey et al., 2006). We perform the test for each tap in a given MCI measurement set under the challenging hypothesis of a uniform phase distribution for each tap and with a significance level of 5%. Hence, each test is computed on the basis of $N = 90$ samples. A negative test result, indicated by 0 in Figure 5.16, means that the challenging hypothesis of a uniform phase distribution cannot be rejected for most ordinary taps. Only cluster taps fail the test, which is consistent with our interpretation of these taps in Section 5.5.2.B.

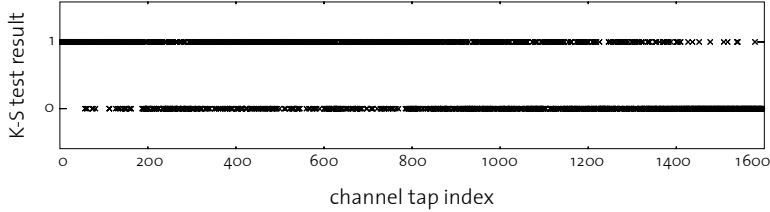
5 MARGINAL TAP DISTRIBUTIONS



(a) MC II LOS, $d = 20$ m.



(b) MC II OLOS, $d = 20$ m.



(c) MC II NLOS, $d = 13$ m.

Fig. 5.17: Results of the Kolmogorov-Smirnov GOF test for MC II data under the challenging hypothesis of a uniform phase distribution.

B. Measurement Campaign II

The picture changes drastically if we apply the same Kolmogorov-Smirnov GOF test just described to MC II data. The corresponding test results are shown for all three MC II measurement sets in Figure 5.17. According to the test, the uniform phase hypothesis must be rejected for about half of all channel taps in a given MC II measurement set.

C. Discussion

The test results for MCI seem to be unequivocal—the uniform phase hypothesis does not need to be rejected, except for cluster taps. For MCII, the situation is not as clear and the corresponding modeling decision a difficult one. A parsimonious model would ideally use the same phase distribution in all taps, but if we follow the test results, we need to reject the uniform phase hypothesis for around half the channel taps in each measurement set.

Any statistical test can only make use of the information contained in the data and used in the design of the test; yet, we do know more, about the measurement environment, about the measurement system, about the postprocessing steps. Unfortunately, this additional knowledge does not help to make our modeling decision easier, as the following considerations show.

- The MCI GOF test statistics are computed on the basis of $N = 90$ samples, while for MCII, we have 1011 samples in the LOS set, 2722 samples in the OLOS set, and 1256 samples in the NLOS set available. Hence, our confidence in the test results for MCII should be higher.
- There does not seem to be a compelling physical reason for the observed variation in test results from one tap to the next.
- MCII phase variation is either induced by people moving in the lobby or by inaccurate measurements and postprocessing. The large impact of our postprocessing on the phase of each tap is best illustrated by comparing Figure 4.12 with Figure 4.13, where the difference in phase without and with interpolation-based tap alignment is shown. Combined with the higher sensitivity of the test because of the large number of samples, measurement and postprocessing inaccuracies in MCII might be a possible reason for the inconclusive result.
- The two taps for which a scatter plot of all measured values is shown in Figure 5.15 both belong to the set of taps for which the uniform phase hypothesis needs to be rejected according to the test result in Figure 5.17a and Figure 5.17b, respectively. The

deviation from uniformity does not seem to follow a common pattern, an observation bolstered by similar inspections of other taps.

To summarize: For MCI, the almost unequivocal test results suggest to model the corresponding tap phases as uniform, while the same test for MCII does not support a single distribution for the phase of all impulse response taps; the latter finding might very well result from imperfections of the measurement system and the postprocessing steps applied to the raw MCII measurement data. For lack of an alternative, and appealing to the modeling principles of parsimony and mathematical tractability, we decide, nevertheless, to model the phase of all MCII impulse response taps as uniformly distributed around their means. The uniform distribution is also the maximum entropy distribution, i.e., the distribution that requires the least number of assumptions about the physical reality. The combination of a Rayleigh or Rice amplitude distribution with the uniformly distributed phase of the zero-mean component results in a proper Gaussian distribution for each individual tap.

CHAPTER 6

Second-Order Analysis

A COMPLETE stochastic description of the block-fading model (5.1) requires specification of a joint distribution for the random channel vector \mathbf{h} . As it is impossible to collect enough independent channel samples to use the model selection techniques presented in the previous chapter as a means of selecting a suitable joint distribution for \mathbf{h} , we need to contend ourselves with a second-order analysis of this joint distribution.

6.1. THE JOINTLY PROPER GAUSSIAN DISTRIBUTION

The second-order properties of a complex-valued random vector \mathbf{h} are completely specified by its mean $\boldsymbol{\mu}_{\mathbf{h}} \triangleq \mathbb{E}[\mathbf{h}]$, covariance matrix

$$\mathbf{R}_{\mathbf{h}} \triangleq \mathbb{E}[(\mathbf{h} - \boldsymbol{\mu}_{\mathbf{h}})(\mathbf{h} - \boldsymbol{\mu}_{\mathbf{h}})^\dagger], \quad (6.1)$$

and *pseudocovariance matrix*

$$\mathbf{J}_{\mathbf{h}} \triangleq \mathbb{E}[(\mathbf{h} - \boldsymbol{\mu}_{\mathbf{h}})(\mathbf{h} - \boldsymbol{\mu}_{\mathbf{h}})^T]. \quad (6.2)$$

The term pseudocovariance matrix seems to originate in a paper by Neeser and Massey (1993); other commonly used names are *complementary covariance matrix* (Schreier and Scharf, 2003) or *relation matrix* (Picinbono, 1996).

A further simplification in the second-order description of complex-valued random vectors arises if the pseudocovariance vanishes. Such random vectors are called *proper* (Neuser and Massey, 1993). If a complex-valued random vector \mathbf{h} is proper, the real and imaginary parts of a given element of this vector are uncorrelated and of the same variance; furthermore, the covariance matrix of $\Re\{\mathbf{h}\}$ equals the covariance matrix of $\Im\{\mathbf{h}\}$. If a complex-valued random vector \mathbf{h} is JPG, it is completely described by its mean $\mu_{\mathbf{h}}$ and covariance matrix $R_{\mathbf{h}}$. JPG vectors are especially amenable to mathematical analysis because affine transformations of JPG vectors are again JPG. Therefore, many relations that result from the linear structure of the space of JPG vectors, like linear estimators, carry over verbatim to the case of JPG vectors (Miller, 1973). Densities and generating functions can also be defined for JPG vectors analogously to their real-valued counterparts (Wooding, 1956; Miller, 1969), and many fundamental results from real multivariate statistical theory have a JPG counterpart (Goodman, 1963).

The complex baseband impulse response of a wireless channel is often taken as a proper random process, or as a proper random vector in discrete-time, because it seems physically sensible to model the phase as uniformly distributed for all delays, as discussed in Section 2.5. This assumption implies that, for a given delay, real and imaginary part of the impulse response are uncorrelated. A further common assumption is that the real and imaginary parts of the impulse response have the same covariance. This assumption can be justified by the definition of a baseband equivalent signal: the choice of center frequency is arbitrary, and hence is the choice of real and imaginary part of the impulse response, so that there does not seem to be any reason to distinguish the real part from the imaginary part and break the symmetry. The two conditions just discussed are necessary for a vector or process to be proper, but they are not sufficient—the crosscorrelation between imaginary and real parts needs to be Hermitian symmetric in addition. This latter condition is difficult to justify on physical grounds, though, and virtually impossible to check empirically.

Properness is a sensible modeling assumption for many complex baseband signals, as it can be shown that a complex baseband representation of a real-valued WSS process is proper (Schreier and Scharf, 2003). In our case, though, two complications arise: First, we are dealing with discrete-time and discrete-frequency signals, for which the relation between properness and stationarity is more involved (Picinbono, 1994).^{*} Second, our LTI channel impulse responses are indexed by a delay variable, but wide-sense stationarity in delay is not a sensible modeling assumption. Nevertheless, virtually all families used to model the complex baseband channel are proper, e.g., in the WSSUS model introduced in Section 2.5.3, or for most block-fading models with tap correlation (Tse and Viswanath, 2005). Similarly, it is common to specify spatial covariance matrices for multiantenna channels while the corresponding pseudocovariances are implicitly assumed to vanish (Paulraj et al., 2003; Goldsmith et al., 2003).

We would like to know if we can sensibly model the measured UWB channels as proper Gaussian distributed. The analysis in Chapter 5 shows that the individual channel taps of MCI impulse responses can be described as zero-mean proper Gaussian. The evidence for proper Gaussian marginal taps is not as compelling for MCII impulse responses, but a proper Gaussian distribution, in this cases with nonzero mean component, still seems to be a reasonable choice. Marginal properness does not imply joint properness, and a marginal Gaussian distribution does not imply that the overall vector is jointly Gaussian distributed. Unfortunately, it seems to be extremely difficult to make a sound judgment on the validity of the JPG assumption, because we do not have enough data relative to the size of the covariance and pseudocovariance matrices at hand to reliably estimate them. Schreier et al. (2006) propose a generalized likelihood ratio test for properness whose test statistic depends on the covariance and pseudocovariance matrix of the channel vector. We did not succeed

^{*} Picinbono (1994) discusses the relation between stationarity and circularity, not properness. However, circularity and zero-mean properness are equivalent in the Gaussian case (Neeser and Massey, 1993).

computing this test statistic from MCII data because of numerical instability. The test fails because the JPG property is a global one. We need to consider the entire covariance matrix and the entire pseudo-covariance matrix, both of which are described by L^2 real variables,* where L denotes the length of the impulse response vector. Reliable estimation of that many parameters even from MCII data is difficult—not for an isolated entry of the covariance matrix, but in order to infer structural properties of the entire matrix.

The only possible way to proceed is to indirectly characterize the joint distribution via low-dimensional quantities that can be reliably estimated. One such low-dimensional test is to analyze the marginal distribution of transformed channel vectors. If some random vector \mathbf{x} is JPG distributed, the DFT, or any other linear transformation of \mathbf{x} , will result in another JPG random vector. Therefore, as a simple heuristic test, we compute the DFTs of all MCII impulse response vectors \mathbf{h} and verify through AIC that the tap amplitudes of the DFT of \mathbf{h} are Rayleigh, respectively Rice, distributed. This simple test does, of course, not solve the problem to determine if the JPG assumption on \mathbf{h} is sensible, but with no evidence against it so far, we take it as a working hypothesis in the following.

6.2. INTERTAP CORRELATION

Any baseband channel includes the effect of band-limiting filters; hence, the corresponding discrete-time complex baseband channel impulse response cannot have uncorrelated taps, even if the continuous-time propagation channel was US—a highly idealized modeling assumption, as discussed in Section 2.5.3. Furthermore, the effect of frequency-dependent propagation paths and a violation of the small fractional bandwidth assumption from Section 2.3.2 also manifest themselves in correlation of the complex baseband channel vector. Because we only have access to measurements of the discrete-time

* Any $L \times L$ Hermitian matrix is parameterized by L^2 free real variables.

baseband channel impulse response, we cannot separate the three mentioned sources of tap correlation; but we can analyze if correlation is significant enough to question the commonly used discrete-time US assumption.

An unbiased estimate for the channel mean vector $\mu_{\mathbf{h}}$ is the empirical mean

$$\hat{\mu}_{\mathbf{h}} \triangleq \frac{1}{N-1} \sum_{n=1}^N \mathbf{h}_n, \quad (6.3)$$

where the \mathbf{h}_n are the individual measured channel impulse response vectors in a given measurement set. The channel covariance $\mathbf{R}_{\mathbf{h}}$ can be estimated unbiasedly as

$$\hat{\mathbf{R}}_{\mathbf{h}} \triangleq \frac{1}{N} \sum_{n=1}^N (\mathbf{h}_n - \hat{\mu}_{\mathbf{h}})(\mathbf{h}_n - \hat{\mu}_{\mathbf{h}})^\dagger, \quad (6.4)$$

which, if the underlying distribution is JCG, is the ML estimate (Goodman, 1963). If we normalize all entries of this matrix, we obtain the normalized *sample correlation matrix*, whose entries are the normalized *sample correlation coefficients*

$$\rho_{ij} \triangleq \frac{[\hat{\mathbf{R}}_{\mathbf{h}}]_{ij}}{\sqrt{[\hat{\mathbf{R}}_{\mathbf{h}}]_{ii} [\hat{\mathbf{R}}_{\mathbf{h}}]_{jj}}}, \quad i, j = 0, 1, \dots, L-1. \quad (6.5)$$

Every individual correlation coefficient can be estimated quite accurately from MCII data because we have a sufficient number of samples available; it is the overall *structure* of the correlation matrix, i.e., the relation between individual entries, that cannot be estimated reliably.

A plot of the normalized correlation coefficient magnitudes $|\rho_{ij}|$ is shown in Figure 6.1 for the MCII LOS measurement set. For ease of presentation, we compute the corresponding matrix from truncated channel impulse response vectors of $L = 700$ taps. The lighter the shading at a given index pair (i, j) in Figure 6.1, the less correlated are the i th and the j th tap. Because of the normalization in (6.5), the

6 SECOND-ORDER ANALYSIS

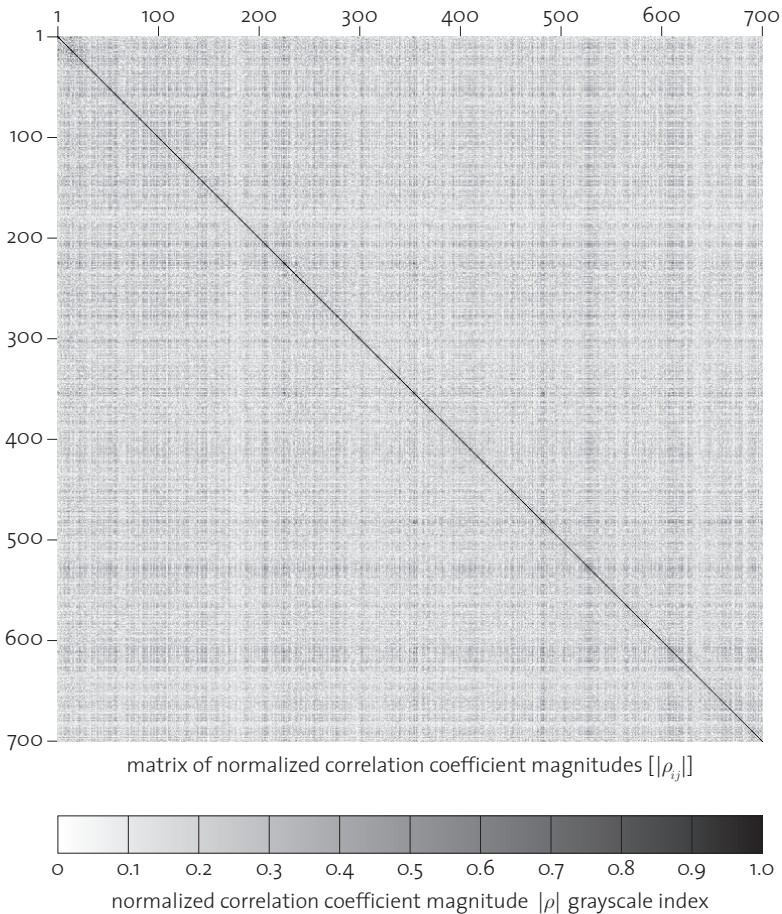


Fig. 6.1: Matrix of normalized correlation coefficients ρ_{ij} for the MCII LOS measurement setting with $d = 20$ m, computed from $N = 1011$ sample impulse response vectors of length $L = 700$ each.

autocorrelation coefficients on the main diagonal satisfy $\rho_{ii} = 1$ by definition for all i . Correlation matrices for the MCII OLOS and NLOS settings look very similar to the LOS correlation matrix shown, with

6.2 INTERTAP CORRELATION

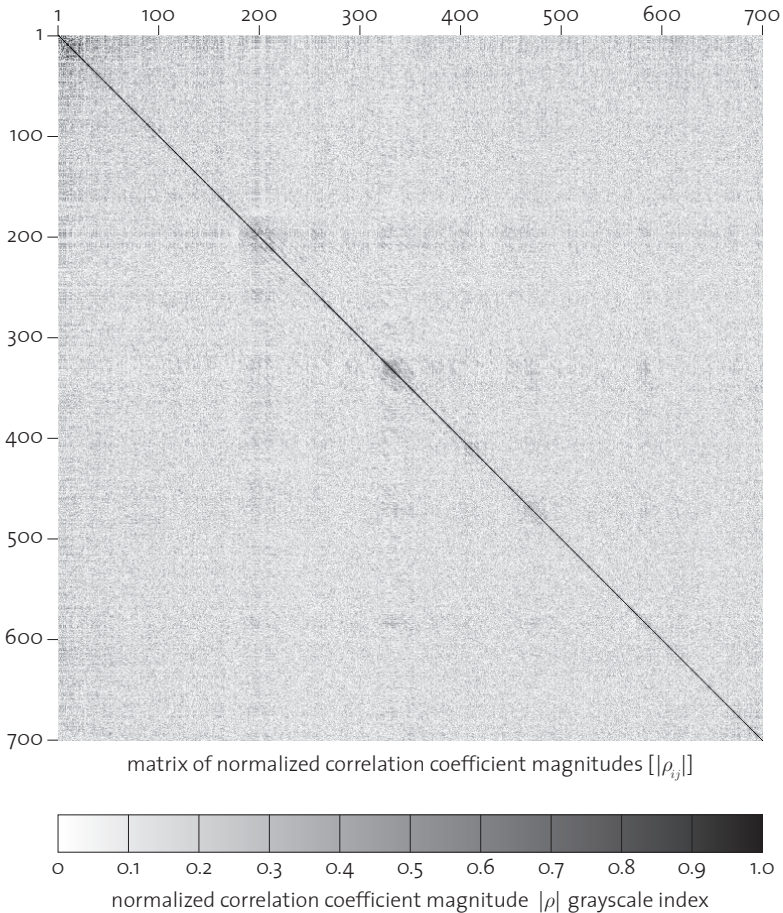


Fig. 6.2: Matrix of normalized correlation coefficients ρ_{ij} for the MCI LOS measurement set with $d = 27.2$ m, computed from $N = 90$ sample impulse response vectors of length $L = 700$ each.

correlation in the OLOS setting a little lower and in the NLOS setting a little higher than in the LOS setting. Figure 6.2 presents a similar plot, but computed from MCI LOS data.

The two plots from MCI and MCII data show some striking similarities but also some curious differences.

- The magnitude of the normalized cross-correlation coefficients ρ_{ij} for $i \neq j$ are small, around 0.15 on average in both cases, apparently somewhat smaller in MCI than in MCII. Yet it is impossible to determine if the values are small enough to be considered negligible.
- Especially in the MCI plot in Figure 6.2, there are small areas of high correlation around some points on the main diagonal. A comparison with the PDP for the same measurement set in Figure 4.15 indicates that the corresponding tap indices belong to cluster peaks. But for those “cluster taps” we argued in Section 5.5.2.B that strong specular reflections appear in several neighboring taps because of their delay variation over the virtual array. Consistent with this interpretation, we find a stronger clustering effect in Figure 6.2, computed from MCI data, than in Figure 6.1, computed from MCII data.
- Probably the most curious difference between the MCI and MCII sample correlation matrices is the hatched structure in Figure 6.1, which is almost nonexistent in Figure 6.2. Unfortunately, we do not have an unequivocal explanation for this observation. It might be related to the fewer number of samples in MCI and thus simply lost in the noise, but it might equally well result from the not always perfectly uniform phase of MCII taps, as discussed in Section 5.7.

In summary, correlation between most taps seems to be low, lower in MCI than in MCII. However, while individual correlation coefficients can be estimated quite accurately, we cannot say much about the overall correlation structure of the channel impulse response vectors for a lack of sufficient data. Thus, we cannot determine if effects like band limitation and frequency-dependent propagation paths, which might lead to correlation of impulse response taps, have a significant impact on the joint distribution of \mathbf{h} .

6.3. EIGENVALUES AND DEGREES OF UNCERTAINTY

The receiver of a wireless communication system generally does not know the instantaneous realizations of the channel coefficients; hence, it needs to jointly resolve the uncertainty in the transmitted signal and in the channel. This channel uncertainty can be expected to increase with increasing bandwidth. For certain US channel models, it is well known that too much bandwidth can be detrimental in that the achievable rates over such channels under a peak power constraint at the transmitter start to decrease if bandwidth is increased beyond a certain *critical bandwidth* (Telatar and Tse, 2000; Médard and Gallager, 2002; Subramanian and Hajek, 2002; Verdú, 2002). This effect is important for the optimal choice of bandwidth in future UWB communication systems. To empirically characterize channel uncertainty, though, we first need to determine how to best measure and devise an estimator for it.

From a communication-theoretic perspective, a sensible measure of channel uncertainty is the number of *uncorrelated* fading dimensions in signal space. This number is commonly referred to as the *diversity order* of a wireless channel; it determines the slope of the high-SNR error probability curve for coded transmission (Tse and Viswanath, 2005). We prefer to use the term *degrees of uncertainty (DOU)* here, because the realization of a diversity gain is contingent on the receiver. The number of DOUs in a block-fading channel increases linearly over blocks of time because these blocks fade independently. Thus, we are interested in the number of DOUs within a single block. If the channel under study satisfies the discrete-time US assumption, i.e., if the taps of the impulse response are uncorrelated, the number of DOUs is simply given by the number of nonzero taps. However, the correlation analysis in the preceding section was inconclusive as to whether significant correlation exists in the channels measured in MCI and MCII; hence, we have to take a more cautious approach. The *Karhunen-Loève expansion*, also known as *principal component*

analysis (Anderson, 2003), linearly decomposes a random vector into uncorrelated components. Mathematically, the Karunen-Loève expansion is simply a change of coordinates, where the new coordinate system results from the spectral decomposition of the covariance matrix $\mathbf{R}_{\mathbf{h}}$ of the random vector \mathbf{h} under study. Therefore, the number of uncorrelated components, and hence the number of DOUs is equal to the number of nonzero eigenvalues of $\mathbf{R}_{\mathbf{h}}$. We denote its l th eigenvalue by λ_l , where $l = 0, 1, \dots, L - 1$, and we assume that the eigenvalues are arranged in nonincreasing order, i.e., $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_{L-1}$. As $\mathbf{R}_{\mathbf{h}}$ is nonnegative definite, all eigenvalues are real-valued and nonnegative.

Because we do not know $\mathbf{R}_{\mathbf{h}}$, we also do not know its eigenvalues λ_l . Hence, to analyze the scaling behavior of the DOUs with bandwidth, we need a suitable eigenvalue estimator. As estimated eigenvalues are never exactly equal to zero, we also need to devise a method to estimate the *effective* number of eigenvalues.

6.3.1. Eigenvalue Estimation

Most of the multivariate statistical theory, as presented, e.g., in the book by Anderson (2003), deals with real-valued sample vectors and corresponding real covariance matrices. To analyze complex random vectors $\mathbf{z} \triangleq \mathbf{x} + i\mathbf{y}$, where \mathbf{x} and \mathbf{y} are jointly Gaussian, we can always resort to real statistical analysis on the basis of the stacked random vector $\tilde{\mathbf{z}} \triangleq [\mathbf{x} \ \mathbf{y}]^T$. However, if \mathbf{z} is JPC distributed, many results can be expressed directly in terms of the complex random vector \mathbf{z} instead of the stacked real vector $\tilde{\mathbf{z}}$ (Goodman, 1963). Of particular importance for the problem at hand is the fact that for a JPC vector \mathbf{z} , each eigenvalue of the covariance matrix $\mathbb{E}[\mathbf{z}\mathbf{z}^\dagger]$ appears twice as an eigenvalue of $\mathbb{E}[\tilde{\mathbf{z}}\tilde{\mathbf{z}}^T]$ (Schreier and Scharf, 2003); hence, all results about eigenvalue estimation in the real case also apply to the JPC case.

The most straightforward estimator for the eigenvalues of $\mathbf{R}_{\mathbf{h}}$ are the eigenvalues of $\hat{\mathbf{R}}_{\mathbf{h}}$, which we denote as $\hat{\lambda}_l$, with $l = 0, 1, \dots, L - 1$. As before, we assume that these eigenvalues are arranged in nonin-

creasing order, i.e., $\hat{\lambda}_0 \geq \hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_{L-1}$. Although $\hat{\mathbf{R}}_{\mathbf{h}}$, defined in (6.4), is an unbiased estimator for the covariance matrix $\mathbf{R}_{\mathbf{h}}$, the eigenvalues $\hat{\lambda}_l$ of $\hat{\mathbf{R}}_{\mathbf{h}}$ are biased estimates for the eigenvalues λ_l of $\mathbf{R}_{\mathbf{h}}$ if the number of samples used in the computation of $\mathbf{R}_{\mathbf{h}}$ is finite. In particular, the vector of estimated eigenvalues $\hat{\boldsymbol{\lambda}} \triangleq [\hat{\lambda}_0 \ \hat{\lambda}_1 \ \dots \ \hat{\lambda}_{L-1}]^T$ is more dispersed than the vector of eigenvalues $\boldsymbol{\lambda} \triangleq [\lambda_0 \ \lambda_1 \ \dots \ \lambda_{L-1}]^T$, i.e., estimates of small eigenvalues are biased down while estimates of large eigenvalues are biased up (Muirhead, 1987). This property is apparent in the extreme case where the dimension L of the covariance matrix $\mathbf{R}_{\mathbf{h}}$ is larger than the number N of samples available for its estimation; then, the corresponding estimate $\hat{\mathbf{R}}_{\mathbf{h}}$ is rank deficient and the smallest $L - N$ eigenvalues equal zero. For this reason we need to confine our analysis to MCII data. But even then the number of samples is not much larger than the number of parameters we wish to estimate. For MCII channel impulse responses truncated to $L = 700$ taps, we have $N/L \approx 1.44$ in the LOS setting, $N/L \approx 3.89$ in the OLOS setting, and $N/L \approx 1.79$ in the NLOS setting. Clearly, we cannot satisfy the assumption of $N/L \gg 1$, under which most estimators are derived. Improved eigenvalue estimators were proposed by Anderson (1965), Stein,* and Haff (1980). However, these improved estimators are still derived for a number of samples that is large compared with the size of the covariance matrix.

An alternative approach is to devise estimators for the covariance matrix under *generalized asymptotes*, i.e., for $N, L \rightarrow \infty$ with fixed N/L . Ledoit and Wolf (2004) provide such a covariance matrix estimator, which is well conditioned and invertible for small sample sizes. This estimate induces an estimate for the eigenvalues of $\mathbf{R}_{\mathbf{h}}$, given by

$$\hat{\lambda}_l^* \triangleq \hat{\zeta}_1 + \hat{\zeta}_2 \hat{\lambda}_l, \quad l = 0, 1, \dots, L - 1, \quad (6.6)$$

where $\hat{\zeta}_1$ and $\hat{\zeta}_2$ need to be estimated from the data. For typical parameters $\hat{\zeta}_1$ and $\hat{\zeta}_2$, the eigenvalue estimate $\hat{\lambda}_l^*$ corrects the biased

* Cited following Muirhead (1987), as the original references seem to be inaccessible.

sample estimate $\hat{\lambda}_l$ toward 1, i.e., it *shrinks* the dispersion of the eigenvalue vector. We compute the estimate (6.6) for all L eigenvalues from MCH data and find that $\hat{\zeta}_1 \approx 0$ and $\hat{\zeta}_2 \approx 1$, i.e., hardly any correction is applied compared with the sample eigenvalue $\hat{\lambda}_l$. This suggests that N might be large enough to reliably estimate the eigenvalues $\hat{\lambda}_l$ of \hat{R}_h in our case.

6.3.2. Counting the Effective Degrees of Uncertainty

Heuristically, the more a channel tap varies, the more uncertain the receiver is about this tap. In the coordinate system defined by the Karunen-Loève transform, the eigenvalues indicate the variance of their corresponding dimension; hence, a sensible measure for the number of effective DOUs is how many eigenvalues contribute a fixed fraction, say 90%, of the total channel variance. A similar criterion has already been used by Patenaude et al. (1999) to measure the number of effective diversity branches for channels at 950 MHz and 40 GHz center frequency. More precisely, we normalize the total estimated variance according to $\sum_{l=0}^{L-1} \hat{\lambda}_l = 1$ and declare all eigenvalues $\hat{\lambda}_l$ with index $l \leq L_s$ to be significant, where L_s is the largest integer that satisfies $\sum_{l=0}^{L_s-1} \hat{\lambda}_l \leq s$, with $0 \leq s \leq 1$. This criterion essentially measures the number of diversity branches with an effective branch receive SNR above an implicitly defined threshold.

Our main interest is in the scaling behavior of the effective DOUs with bandwidth. Therefore, we compute the just described measure L_s for various bandwidths in the following way:

1. We truncate all measured impulse responses to a length of 700 taps.
2. We use the DFT to transform all N measured impulse responses to the frequency domain and compute the empirical covariance matrix (6.4) from these frequency-domain samples. As the DFT is a unitary transform, it does not change the eigenvalues of the covariance matrix.
3. We subdivide the overall bandwidth of 3 GHz into 30 intervals of 100 MHz each and map this division to the 700 frequency points.

6.3 EIGENVALUES AND DEGREES OF UNCERTAINTY

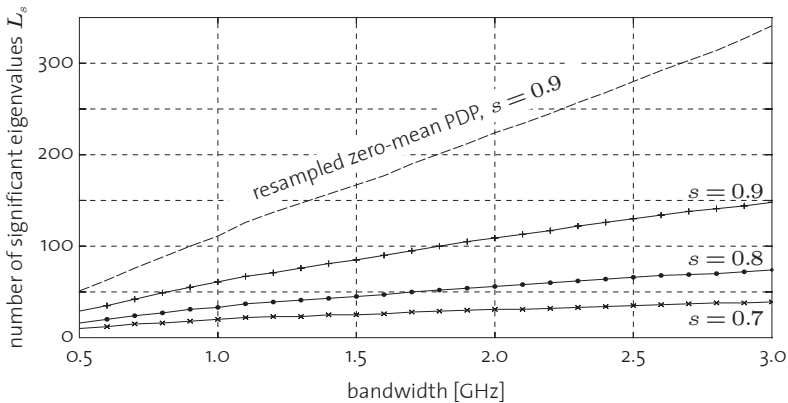


Fig. 6.3: Number L_s of largest estimated eigenvalues that account for a fraction s of the total channel variance. Computed from $N = 1011$ MCII olos samples.

One interval thus corresponds to either 23 or 24 frequency points. Step by step we reduce the bandwidth from the upper band edge by pruning the covariance matrix: for each bandwidth reduction by 100 MHz we remove the last 23 respectively 24 rows of the frequency-domain covariance matrix.

4. We compute the eigenvalues for each pruned covariance matrix.

If the channel taps were uncorrelated, the channel covariance matrix for each measurement set would be diagonal, and the diagonal would consist of the PDP after removing the channel mean, like in (4.9). To test whether the discrete-time US assumption is viable, we treat this zero-mean PDP as an eigenvalue spectrum and compute its scaling behavior as just outlined.

The resulting scaling of the number of effective eigenvalues as a function of bandwidth is shown in Figure 6.3 for the MCII LOS setting, in Figure 6.4 for the MCII OLOS setting, and in Figure 6.5 for the MCII NLOS setting. In all three measurement sets, the number of effective eigenvalues L_s seems to scale approximately linearly with bandwidth. Physically, this hints at the possibility that the amount

6 SECOND-ORDER ANALYSIS

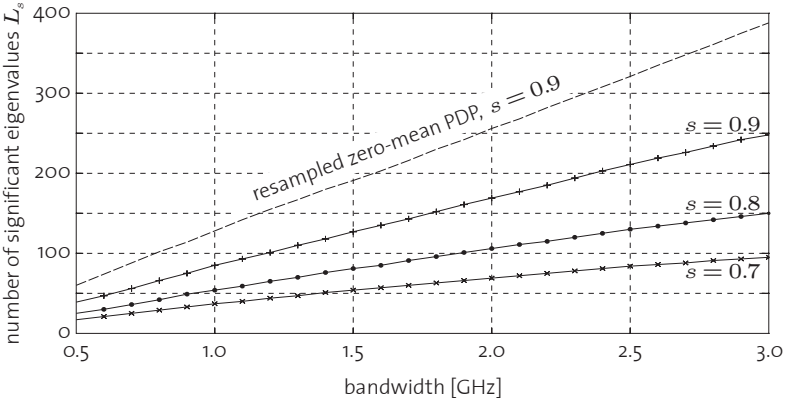


Fig. 6.4: Number L_s of largest estimated eigenvalues that account for a fraction s of the total channel variance. Computed from $N = 2722$ MCII OLS samples.

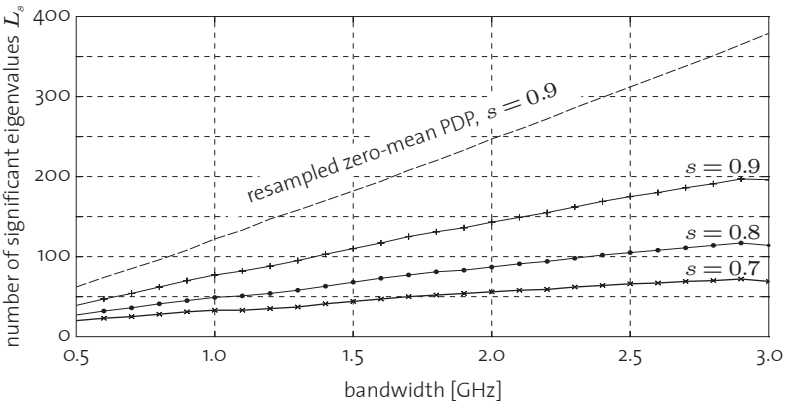


Fig. 6.5: Number L_s of largest estimated eigenvalues that account for a fraction s of the total channel variance. Computed from $N = 1256$ MCII NLOS samples.

of scattering in the environment is not limited over the measured frequency band. Such a finding lends support to the proper Gaussian tap distribution in Chapter 5, which can be explained by rich scattering and correspondingly many partial waves arriving over the duration of each tap. However, we can also observe that the slope of the eigenvalue vs. bandwidth curve of the hypothetical uncorrelated channel is higher than the empirical eigenvalue curves in all three measurement sets. This hints at the existence of intertap correlation; though, in view of the discussion in Section 6.2 about the tap correlation coefficients, the source of the observed correlation cannot be unequivocally determined. Another curious observation for which we do not have a sensible explanation is the decrease of L_s at very large bandwidth in the NLOS setting shown in Figure 6.5.

To double-check our results, we also use AIC and MDL in the form described by Wax and Kailath (1985). Both criteria are often used to determine the number of signals impinging on a linear array. We adapt both criteria following Zhao et al. (1986) to take into account the noise color introduced through the calibration process. The application of information criteria to determine the number of signals is structurally equivalent to our problem of discerning the number of nonzero eigenvalues (Stoica and Selén, 2004). Both AIC and MDL show an approximately linear scaling behavior of the number of eigenvalues as a function of bandwidth, consistent with the findings reported here.

We mention in Section 6.3.1 that the eigenvalues of the sample covariance matrix \hat{R}_h are biased estimates for the eigenvalues of the channel covariance matrix R_h itself and discuss several improved eigenvalue estimators. To bolster the claim that our eigenvalue estimates are accurate enough to indicate the correct scaling of L_s , we recompute the number of eigenvalues as a function of bandwidth for the improved eigenvalue estimators by Anderson (1965), Stein (Muirhead, 1987), and Haff (1980). The result for the MCII OLOS setting with $s = 0.9$ in Figure 6.6 shows that the scaling behavior is indeed unchanged, although a small correction of the slope is apparent.

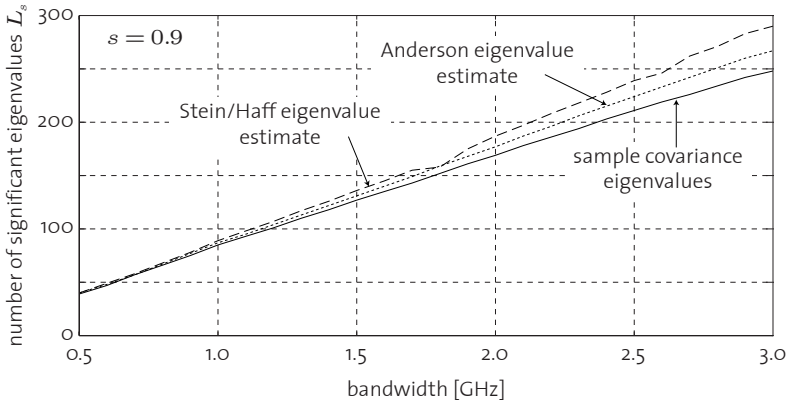


Fig. 6.6: Comparison of different eigenvalue estimates on the scaling behavior of the number of significant eigenvalues that account for 90% of the total variance in the channel. Estimated from 2722 MCII OLOS samples.

Contradicting Results

Our finding on the scaling of the number of DOUs with bandwidth seems to be consistent with other observations from our measurements, yet it is inconsistent with a similar analysis by Menouni Hayar et al. (2005),* who found a sublinear scaling of the DOUs. As often in experimental work, we can think of several explanations for this difference, but we cannot be certain which one is correct.

- Our measurement system might have introduced artifacts. One prominent impairment that might lead to an inflation of the number of DOUs is sampling jitter (Balakrishnan, 1962; Brown, 1963; Liu and Stanley, 1965). To check if the different scaling behavior indeed results from sampling jitter introduced by the DSO, we used data from Menouni Hayar et al. (2005) and artificially added sampling jitter. We found that jitter with standard deviation of 10 ps, almost ten times more than given in our DSO

* See also the papers by Saadane et al. (2004, 2005).

specification, is too small to have any noticeable effect upon the number of significant eigenvalues.

- The way how the DOUs are measured is different in our analysis compared with the analysis by Menouni Hayar et al. (2005).
- The measurement setup of Menouni Hayar et al. (2005) is different from our MCII, more similar to MCI, with a VNA as measurement device and a virtual array to induce channel variation.

If indeed the different sources of channel variation—motion of the entire propagation environment relative to the antennas compared with the motion within a fixed environment of people only—lead to the different scaling behavior, the design implications for UWB communication system could be quite far-reaching. However, we were not able to test this hypothesis.

6.4. CAPACITY ESTIMATES

Statistical channel models are typically used to predict the performance of communication systems. Because the most fundamental performance measure is the channel capacity C , the goal of this section is to assess the approximation quality of the discrete-time block-fading channel model with JPG taps of nonzero mean with respect to the capacity directly estimated from MCII data. To this end we use the Monte-Carlo method (MCM) to synthesize JPG-distributed impulse responses with either uncorrelated or correlated taps; mean and covariance in both cases are ML-estimated from MCII data, as described in Section 5.3.2. Throughout, we assume no instantaneous channel state information (CSI) at the transmitter and perfect CSI at the receiver, i.e., we consider the *coherent* setting, mainly because no closed-form capacity expression is known for the noncoherent case. A detailed study of the noncoherent capacity of wideband channels, albeit on the basis of the discretized WSSUS model from Section 3.2.4, is the theme of Part III.

The coherent mutual information of a block-fading channel with random transfer function $\mathbf{H}(f)$ and additive white Gaussian noise is

a random variable \mathbf{I} , given as (Biglieri et al., 1998)

$$\mathbf{I} = \int_{-B/2}^{B/2} \log(1 + \text{SNR} |\mathbf{H}(f)|^2) df \quad [\text{nat/s}]. \quad (6.7)$$

6.4.1. Ergodic Capacity

An estimate* of the ergodic capacity $C = \mathbb{E}[\mathbf{I}]$ on the basis of N samples of the discrete-time channel impulse response \mathbf{h} can be obtained as (Hirt and Massey, 1988; Bölcskei et al., 2002a)

$$\hat{C} = \frac{1}{N} \sum_{n=1}^N \sum_{l=0}^{L-1} \log\left(1 + \frac{P}{LN_0} |H_n[l]|^2\right) \quad [\text{nat/s}] \quad (6.8)$$

where $[H_n[0] H_n[1] \dots H_n[L-1]]^T$ is the length- L DFT of the n th channel vector sample \mathbf{h}_n . The average transmit power is constrained to P , allocated uniformly over all L parallel channels, and N_0 is the noise variance in each parallel channel. This estimate does not assume a specific distribution of the taps, nor does it impose any correlation structure between taps. We compute the capacity estimate (6.8) from the measured and the synthesized impulse responses as a function of bandwidth B . All impulse responses are normalized to unit average power and truncated after $L = 701$ taps. The result for the MCII NLOS setting, along with the capacity of the AWGN channel with the same receive SNR, is shown in Figure 6.7. For this figure, we set $P/N_0 = 10$ dB and use $L = 5608$ parallel channels at $B = 3$ GHz. The channel synthesized according to the uncorrelated Ricean model predicts the ergodic capacity of the measured channel very accurately, with a relative mean squared error at $P/N_0 = 10$ dB of less than 0.3% in the NLOS setting, less than 0.2% in the OLOS setting, and less than 0.07% in the LOS setting. The corresponding plots for the latter two

* The estimator (6.8) is a so-called *plug-in* estimate (Efron and Tibshirani, 1993); effectively, we use the empirical CDF in lieu of the operating CDF.

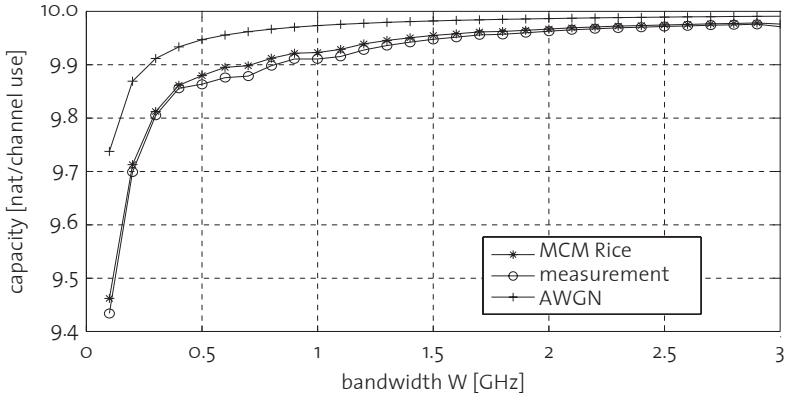


Fig. 6.7: Estimated capacity \hat{C} according to (6.8) from 1265 MCM NLOS samples, Monte-Carlo capacity estimate of the corresponding synthetic channel, and AWGN channel capacity. $P/N_0 = 10$ dB.

measurement sets are not shown here because they are very similar to Figure 6.7.

6.4.2. Outage Capacity

For slow-fading channels (Ozarow et al., 1994; Tse and Viswanath, 2005), the ϵ -outage capacity C_ϵ , defined as $\mathbb{P}(\mathbf{I} < C_\epsilon) \leq \epsilon$, is a more sensible performance measure than the ergodic capacity (Biglieri et al., 1998). Figure 6.8 shows empirical CDFs of \mathbf{I} for the measured LOS channel and two synthetic channels. One of the synthetic channels takes into account intertap correlation according to $\hat{\mathbf{R}}_{\mathbf{h}}$, i.e., the individual MCM samples are drawn according to a JPG distribution with covariance matrix $\hat{\mathbf{R}}_{\mathbf{h}}$. Measured and synthetic channels behave quite differently at low outage probabilities. For the measured channel, the large change in mutual information from 8.8 nat to 9.3 nat at nearly constant outage probability of 0.12 indicates a reduction in total received power for some channel realizations.

In fact, a simulation experiment shows that it suffices to heavily attenuate the cluster taps of the strongest cluster in the synthesized

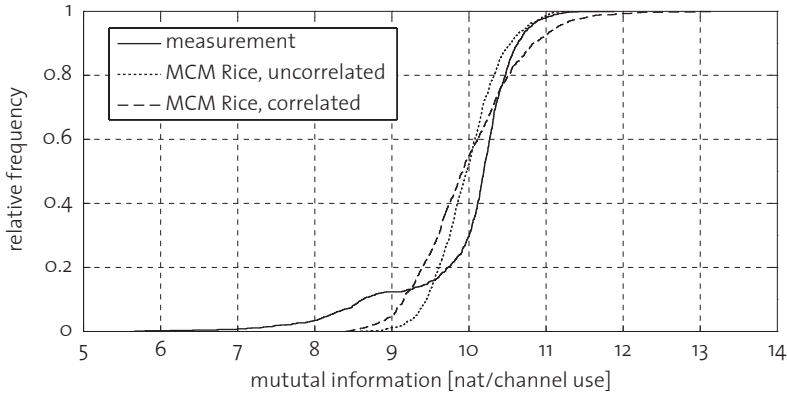


Fig. 6.8: Empirical CDF of the instantaneous mutual information for MCII LOS data, and Monte-Carlo estimate of the CDF for the corresponding synthetic channels. $P/N_0 = 10$ dB.

LOS impulse responses in 10% of all realizations to observe an even stronger increase in outage probability than in Figure 6.8. The empirical CDF of the resulting simulated impulse responses is shown in Figure 6.9. A possible physical explanation for the drop in received power in some channel realizations is that persons moving in the lobby might have sometimes blocked the LOS or some other propagation paths that convey a significant fraction of the total energy. Indeed, the LOS between transmitting and receiving antenna was frequently blocked in the MCII LOS setting because both antennas were positioned at a height of approximately 1.6 m above the floor.

Although it is common practice in channel modeling to separate shadowing and small-scale fading, this separation is not an intrinsic property of the channel but a modeling assumption. Figure 6.8 indicates that this assumption might not always be valid for the measured channels of MCII. The distribution of the multiplicative shadowing random variable that leads to the best agreement between simulated CDFs and the CDF estimated from MCII LOS data is bimodal, to model presence or absence of a person blocking a dominant propagation

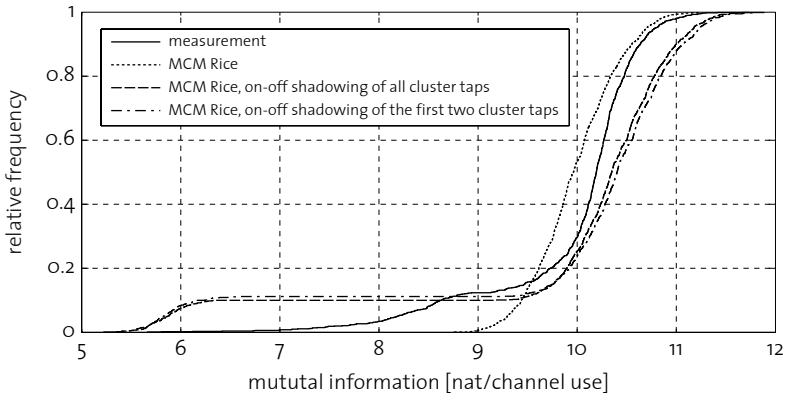


Fig. 6.9: Empirical CDF of the instantaneous mutual information for MCII LOS data, and Monte-Carlo estimate of the CDF for the corresponding synthetic channels with cluster taps shadowed in 10% of all channel realizations. $P/N_0 = 10$ dB.

path; this is in stark contrast to the commonly used lognormal distribution for shadow fading. Scatter plots for individual taps support the finding that only a small number of taps are affected by shadowing. The AIC-best distribution for these taps is most often the Weibull distribution, consistent with our findings in Section 5.6. Shadowing has an even larger impact on the CDF of the mutual information in the NLOS setting, because the overall received SNR is lower, while the impact in the OLOS setting is smaller for the LOS setting depicted in Fig. 6.8.

6.5. SUMMARY AND DISCUSSION

The stated goal in this part of the thesis was to empirically assess the validity of standard modeling assumptions for channels of very wide bandwidth, in particular the proper Gaussian distribution of channel taps, the US assumption, and the scaling of the number of DOUs with bandwidth, as detailed in Section 4.3.1. To this end, we devised

two UWB measurement campaigns and presented statistical methods to answer the three modeling questions on the basis of measured data. Despite the recent advances in measurement and instrumentation technology, channel sounding with several gigahertz bandwidth is still challenging; consequently, it is sometimes difficult to discern measurement artifacts from channel effects, as evinced in the analysis of the intertap correlation in Section 6.2. The single most important hurdle to obtain statistically sound results is the lack of an adequate number of measurements that can be reasonably well considered as a random sample. For this reason, we sometimes have to resort to extrinsic modeling criteria like mathematical tractability or modeling convention to bridge empirical gaps. A common argument of ours if empirical results are inconclusive can be best summarized as “if it ain’t broke don’t fix it,” i.e., we use a well established tried and proven standard model as long as the data does not clearly falsify said model. This is an engineering approach and not scientifically and epistemologically completely sound; nevertheless, as models of wireless channels ultimately are tools to solve engineering problems, we believe that our results can be valuable to this end; they provide design guidelines for wireless systems that operate over channels of large bandwidth. On the one hand, our results dispel some apparently common beliefs that basic communication theory for UWB channels is fundamentally different from the theory used to design systems of several megahertz bandwidth; on the other hand, we found some interesting aspects of wideband channels not accounted for in the present theory that might be beneficially used in future wideband systems.

6.5.1. Marginal Tap Distributions

The model selection criterion AIC allows us to compare the suitability of different fitted amplitude distributions. Both the analysis of individual taps as well as the variation of the Akaike weights across taps yield interesting results.

A. Individual Taps

Especially surprising is the bad fit of the lognormal family for both MCI and MCII data, as this family of amplitude distributions was found to provide a good fit in some early UWB measurement campaigns and was consequently adopted in the channel model of the IEEE 802.15.3a Task Group (Foerster, 2003). The differentiation between the other candidate families is not as clear cut, which hints at the fact that there is not enough evidence in our measurement data to unequivocally select a single family on the basis of the data *alone*; the Akaike weights reveal that the differences between the Rayleigh family and the Rice, Nakagami, and Weibull families in MCI and between the Rice family and the Weibull family in MCII are often small. One possible resolution of this ambiguous statistical result is multimodel inference (Burnham and Anderson, 2002), i.e., to simultaneously use different amplitude distributions and average the results of whatever analysis we wish to perform on the basis of these models. The Akaike weights can be used to weight the different results in a final averaging step. Because of its inherent complexity, such a multimodel approach appears to be suitable only in the simulation stage of a design, not for manual analysis.

For initial analysis and design, the small differences between different candidate families mandate the use of the simplest possible model that is still good enough (see the modeling principles outlined in Section 2.1). The AIC results from Chapter 5 favor the Rayleigh and Rice amplitude distributions, although not always by a strong margin; hence, we believe that for most analysis purposes, these two families are well suited to model small scale fading even of UWB channels. Because the differences between the analyzed amplitude distributions are often small, analyses performed on the basis of either family should yield similar results. Widely differing performance predictions for a specific modulation and coding technique and a specific type of receiver between, say, Nakagami or Ricean fading should not be taken as an indication of huge possible gains to be had if only the system was designed to take advantage of the specific features of either the

Nakagami or the Rice distribution for some parameter setting but rather as a warning to rethink ones modeling assumptions. To state it more bluntly, there appears to be no need for analysis and design of systems for fading different than Rayleigh and Ricean.

Most of the appeal of the Rayleigh and Rice amplitude distribution results from the link to the proper Gaussian distribution. Our analysis in Section 5.7 shows that a zero-mean proper Gaussian distribution for each individual channel tap is a sensible model for MCI, while the tests performed on the basis of MCII data do not provide a clear answer to the question if all MCII taps can sensibly be modeled as proper Gaussian with nonzero mean. On the other hand, we also did not find evidence that clearly contradicts the proper Gaussian assumption. Hence, without a clear alternative, and referring again to the criterion of parsimony and tractability of a suitable model, we believe that the complex Gaussian distribution is a suitable working hypothesis for channels similar to the one measured in MCII.

B. Cluster Taps in MCI

Not all channel taps in MCI can be sensibly modeled as complex Gaussian, the *cluster taps*, as we called them, are distinct. The straightforward consequence would be to model cluster tap amplitudes as Weibull or lognormal distributed.* However, we do not believe that such a modeling approach is sensible. Certainly not to keep the model tractable; but more importantly, we do not believe a system designed on the basis of such a model would offer any performance advantage. Instead, the physical cause for the distinct distribution of cluster taps needs to be taken into account and exploited by a viable design.

We argued in Section 5.5.2.B that the cluster tap amplitudes follow a distribution distinct from the ordinary taps because a single strong specular reflection contributes to several adjacent taps. But such a distinct arrival that moves between taps as the transmitter or the

* We might even presume that it is the distinction of the cluster taps that has lead to the adoption of the lognormal distribution in the IEEE 802.15.3a standard model.

receiver is displaced violates the small-scale fading assumption, i.e., the assumption that the channel impulse response at different points on the virtual array can be treated as independent realizations of the same random vector.* Thus, the difference in Akaike weights for the cluster taps is a manifestation of a large-scale effect, and modeling it in a small-scale fashion, by means of a small-scale amplitude distribution, is not a sensible approach. Instead, a system should exploit the clusters and track them over time, as the corresponding taps convey more energy than other taps in the impulse response, and they are less affected by fading.† For example, a receiver that can track not just the first arrival but also subsequent specular reflections can use a selective rake combiner with less fingers than a conventional rake receiver of comparable performance that simply weights all taps according to the channel's PDP (Win and Scholtz, 2002). Yet to design such a system, the channel model needs to be time variant and correctly reproduce the time variation of strong specular reflections. An SV-type model is not suitable here, because the cluster excess delays τ_l in (3.3) change in a completely random fashion between individual channel realizations. Much more appropriate is a GSCM, as discussed briefly in Section 4.2.2.D.

C. Ricean Component in MCH

Our finding in Section 5.6 that virtually all taps in an MCH impulse response have a strong mean component around which the channel fluctuates is not novel, Bultitude et al. (1989) and Hashemi et al. (1994) reported similar observations. Apparently not much attention has been paid to these results almost 20 years ago, when mobile telephony was the main driver of wireless technology. Nowadays, many wireless systems operate with predominantly static terminals—they are to serve as cable replacements in WPANs, in sensor networks, or as links between stationary laptop computers and access points.

* For details, refer to the design of the virtual array described in Section 4.4.1.B.

† The lognormal amplitude distribution with its heavy right tail predicts exactly the opposite, that cluster taps are more severely faded.

In all those settings we may expect the channel to feature a strong mean component in *every* impulse response tap because the major features of the environment do not change relative to transmitter and receiver. Such a mean component in each tap greatly decreases the channel uncertainty the receiver needs to resolve; in fact, there might exist DOFs that are not affected by fading at all (Nabar et al., 2005; Gärtner and Bölcskei, 2007).

6.5.2. Joint Tap Distribution

Statistical analysis of the joint distribution of an entire channel impulse response is practically infeasible because of the huge number of samples that would be required for most statistical methods. Our working hypothesis of proper complex Gaussian taps suggests to use a JPC distribution for the entire channel vector. This choice is mathematically appealing because it is virtually the only tractable multivariate distribution. Yet this lack of alternatives puts us into a methodological dilemma: An unbiased statistical analysis is supposed to weight all options only on the basis of the available data. But we know already in advance that there is only a single viable option; hence, there is great danger to interpret all findings in a way to fit this anticipated result. Indeed, this is to some extent what we did in the preceding sections. Most of the multivariate analysis we performed had the goal to identify if the data openly violated some of the key consequences of a JPC distribution. In statistical language, we performed a hypothesis test where the incumbent hypothesis was the JPC distribution of the channel vector and the challenging hypothesis was its converse. On the basis of the data alone, incumbent and challenging hypothesis should be exchanged, but on physical grounds and from the vast experience of working wireless communication systems designed on the basis of a JPC channel model, this methodical mistake may be acceptable, although it cannot be fully justified.

A. Tap Correlation

Taken individually, most taps in MCII impulse responses are only weakly correlated, with magnitude of the normalized correlation coefficient around 0.15 on average. For such low intertap correlation, Intarapanich et al. (2004) and Xiao and Zheng (2004) observe only a small decrease in coherent ergodic capacity compared with the uncorrelated case in their numerical analyses. However, both results are for channels with few taps only and assume perfect CSI at the receiver; it is not clear if small intertap correlation is equally unimportant for the performance of noncoherent wireless receivers, in which we are mainly interested in the thesis. While intertap correlation is detrimental in the coherent setting, it might actually be beneficial in the noncoherent setting as it can help to reduce channel uncertainty at the receiver. Indeed, we find in Section 7.7 that correlation between antennas can boost noncoherent capacity. As multiple antennas at transmitter and receiver are a source of DOFs just as bandwidth, the hypothesis that tap correlation may be beneficial seems reasonable.

B. Degrees of Uncertainty

The accuracy of multivariate statistical methods increases with shrinking dimension of the random vector under study. For this reason the eigenvalue estimation and the scaling analysis of the number of DOUs in Section 6.3 are relatively robust, as indicated by the similar results of the different eigenvalue estimates in Figure 6.6. The finding of a linear scaling of the number of DOUs with bandwidth is consistent with the physical justification of the proper Gaussian tap distributions even for 3 GHz channel bandwidth, yet it is not consistent with findings by Menouni Hayar et al. (2005). We conjecture in Section 6.3.2 that the difference might result because we estimate the eigenvalues of the channel's covariance matrix from MCII data, where all channel variation results from moving people in the lobby, whereas the eigenvalue scaling result by Menouni Hayar et al. (2005) derives from measurements similar to our MCI, where the antennas move with

respect to the fixed environment. In the latter setting, all major reflectors and scatterers maintain their geometrical configuration relative to each other, whereas the relative location of people as time-variant scattering objects in our MCII varies. This difference between the two measurement campaigns might indeed be the reason for the different scaling behavior, but so far we are not aware of any physical explanation of this phenomenon. Furthermore, with only two independent measurement campaigns that result in two different findings, neither of the two results should be given too much significance until further independent measurement campaigns either substantiate the two different scaling behaviors of the two different types of channels or falsify at least one of them.

PART III

CAPACITY OF MULTIAN TENNA WSSUS CHANNELS

CHAPTER 7

Capacity Bounds

THE THEME of the previous two parts of this thesis was channel modeling, both from a theoretical and an empirical perspective. Yet, a channel model is not an end in itself but—in our case—a tool for analysis and design of wireless communication systems that operate not over the channel as modeled but over real-world channels of which the model is only a more or less accurate approximation. In this third part of the thesis, we make use of the modeling efforts just presented to analyze the performance attainable over wideband wireless channels. In doing so, we continue the overall theme of this thesis to analyze *fundamental* effects rather than specifics of particular communication systems. Therefore, the natural setting is information theoretic, and we are interested in the *capacity* of wireless channels as the ultimate performance limit on information transmission (Gallager, 1968; Cover and Thomas, 1991).

Operational capacity of a given channel—the supremum of all rates achievable with arbitrarily small probability of error—is a benchmark for the design of any practical wireless system. But the most important asset of a capacity expression besides its numerical value is the insight it offers into which channel mechanisms, which system parameters, and which constraints are fundamental, i.e., significantly change capacity. These insights, often a result of the techniques used to compute or bound channel capacity, provide guidelines for the design of practical systems, e.g., how to best utilize the resources DOFs and power, and

how to design efficient modulation and coding schemes (Biglieri et al., 1998, Section III.3).

The main results of the present chapter are upper and lower bounds on the capacity of wideband multiantenna channels that are selective in time and frequency. The material presented in the current and the next chapter builds on many contributions of Dr. G. Durisi; in fact, all the results are the outcome of a fruitful cooperation (Durisi et al., 2007, 2008; Schuster et al., 2009).

7.1. FROM CHANNEL TO SYSTEM MODEL

Capacity analysis of fading channels requires a *system model*, not just a channel model. Although both terms are often used synonymously in the information-theoretic literature, we prefer to keep them separate in the light of the specific channel models developed in the previous chapters.

7.1.1. Aspects of a Realistic System Model

A system model is more comprehensive than a channel model because it specifies power constraints, the amount of channel knowledge available at the terminals, possible processing steps like interleaving or power control, the user configuration in a multiuser system, feedback links, and any further application and implementation constraints, all in addition to the basic channel model. Our goal here is to analyze the capacity of a system model that is of practical importance. Like in all our modeling efforts discussed so far, we need to strike a balance between complexity and tractability of the model. Therefore, we restrict our attention to aspects that are both of *practical interest* as well as *fundamental*. In particular:

1. *Neither transmitter nor receiver know the realization of the channel:* As the received signal is always corrupted by thermal noise and other impairments, even the best estimate of the channel coefficients cannot be perfectly accurate. Without perfect chan-

nel knowledge at the receiver, perfect channel knowledge at the transmitter is impossible as well, even in the rather unrealistic case of an error-free instantaneous or even anticausal feedback link. The amount of channel knowledge is a fundamental aspect: lack of perfect CSI drastically alters the low-SNR capacity behavior (Verdú, 2002); channel capacity with perfect channel knowledge at the receiver is always larger than the capacity without channel knowledge (Médard, 2000), and the signaling schemes necessary to achieve capacity are also very different in the two cases (Biglieri et al., 1998).

2. *Channel fading is selective in time and frequency:* i.e., the channel exhibits memory in frequency and time. Frequency selectivity is realistic for channels of very wide bandwidth, even in indoor environments with small delay spread. Time selectivity is an equally realistic modeling assumption, as either the terminals or the propagation environment change over time in virtually all application scenarios. Channel selectivity over time and frequency is fundamental as it significantly impacts capacity behavior. For example, without instantaneous channel knowledge, the capacity of a block-fading channel scales differently with increasing SNR than the capacity of certain types of channels with stationary time variation, as shown by Lapidoth and Moser (2003) and Lapidoth (2005); furthermore, different types of frequency selectivity may result in different capacity behavior at low SNR and large bandwidth (Porrat et al., 2007).
3. *The peak power of the transmitted signal is limited.* This assumption is of practical importance as any power amplifier has finite gain and every mobile transmitter has limited battery resources. In addition, regulatory bodies often constrain the admissible radiated power. However, it is more common in information-theoretic analysis to constrain the average power of the transmitted signal. The type of power constraint—peak vs. average—is fundamental: while capacity increases with bandwidth and the infinite-bandwidth AWGN capacity can be attained asymptotically if a constraint on the average transmitted power only is

in place (Pierce, 1966; Kennedy, 1969; Gallager, 1968; Verdú, 2002; Durisi et al., 2006), certain types of peak constraints can lead to vanishing capacity in the large bandwidth limit (Telatar and Tse, 2000; Médard and Gallager, 2002; Subramanian and Hajek, 2002).

The attainable SNR in all practical receivers is limited because of limited resolution of the ADC, limited operating range of the amplification stages, phase noise, and other impairments. For the systems of large bandwidth we are interested in, this means that the SNR per DOF is often very low. Therefore, we primarily analyze capacity at large bandwidth and correspondingly low SNR.

7.1.2. Modeling Assumptions

With the requirements established, we can now review the modeling results of Part I and the empirical evidence of Part II to determine a suitable channel description that, in combination with the mentioned system aspects, allows us to analyze the capacity of wideband wireless systems of practical interest.

A. Single-Antenna Channel

The requirements just outlined in combination with the empirical results in Part II suggest* to use one specific channel model discussed in Part I, namely, the WSSUS model in Section 2.5.3, chiefly because this model captures channel selectivity in time and in frequency. However, the WSSUS model entails several assumptions that need to be scrutinized.

Tap distribution: Our empirical analysis suggests to model the individual tap amplitudes as Rayleigh, respectively Rice distributed. Although the analysis of the tap phases and of the joint tap distribu-

* Although the requirements may strongly suggest to analyze the underspread WSSUS model in Section 2.5.3, these requirements themselves are not set in stone but result from our personal and subjective preferences and experience.

tion did not yield equally strong results in favor of the JPG distribution as a suitable joint distribution of the channel impulse response, we did not find a compelling reason to abandon this mathematically tractable model, i.e., we did not find strong evidence to discount the standard WSSUS model's JPG-distributed two-dimensional channel process.

Uncorrelated scattering: The US assumption is difficult to analyze empirically. Our correlation analysis in Section 6.2 is inconclusive: correlation between taps is low in general, its impact on the coherent mutual information, discussed in Section 6.4.2, is weak, yet we find in Section 6.3.2 that tap correlation has an observable impact on the number of significant eigenvalues of the channel covariance matrix. On physical grounds, the channel cannot be US as it cannot be WSS in frequency: the effective baseband channel is band limited, and channel attenuation increases with increasing frequency. But for capacity analysis of real-world systems, it suffices if the US assumption is a sensible model for bandwidths of practical interest. The fundamental implication of channel correlation is the scaling behavior of the DOUs with bandwidth. The US assumption entails linear scaling, which is also what we find in our empirical study in Section 6.3.2. However, it is difficult to reconcile the US assumption with the strong mean component we found in all MCII impulse response taps. For the channel to be US, it needs to be WSS in frequency, so that a nonzero mean in all time-domain channel taps would need to result in an equal mean across all frequency points. As we do not find such a constant nonzero mean over frequency in MCII data, the analysis in the following sections pertains only to settings where at least one of the terminals is mobile and all channel taps can be assumed to have zero mean, as in the setting of MCI. Those settings are exactly the ones that lead to the type of time- and frequency-selective channels we are interested in, and it is for these settings that we believe the US assumption to be reasonable over frequency bands of practical interest.

Stationarity: We did not assess the WSS assumption in Part II because we could not record time-variant channels with the measurement equipment we had available. Some findings, like the PDP clusters in the LOS setting that change only slowly over time and can be related to major reflectors in the environment, may render the WSS assumption questionable. In the OLOS and LOS settings, however, the impact of specular reflections seems to be greatly diminished. Yet, without first-hand time-variant data, it is difficult to decide if the WSS assumption needs to be discarded altogether, or if it is sufficient to take special care of specular reflections in the channel impulse response. Most importantly, the WSS assumption in time needs to hold reasonably well only over time intervals of interest, i.e., for the duration of a transmitted packet, frame, or codeword, whichever constitutes the longest entity to be jointly processed at the receiver.

The standard WSSUS model is not well suited for information-theoretic studies because it is continuous in time and frequency. But the approximate diagonalization presented in Section 3.2 is; indeed, we believe that it is better suited for the specific setting at hand than most other discretizations because (i) for a suitable choice of the prototype pulse $g(t)$ and of the time-frequency shift parameters T and F , the error incurred in the approximation of the kernel $\mathbf{k}_{\mathbb{H}}(t, t')$ in (3.16) is minimized. As the optimal choice of $g(t)$, T , and F depends on the channel's scattering function $c_s(\nu, \tau)$ as discussed in Section 3.2.3, we say that the approximate diagonalization (3.16) is a *matched expansion* of the LTV channel \mathbb{H} (Kozek, 1997a); (ii) the approximate diagonalization explicitly uses that wireless channels are highly underspread; thus, it exploits an important property of wireless channels that is not made use of in other types of channel decompositions, cf. the basis expansion model of Giannakis and Tepedelenlioglu (1998) or the sampling decomposition of Médard (1995); (iii) the approximate eigenvectors of the matched Weyl-Heisenberg expansion (3.16) are deterministic, so that the diagonalization (3.34) does not require instantaneous channel knowledge at the transmitter and at the receiver, but knowledge of the channel's scattering function

only; (iv) the prototype pulse $g(t)$ of the expansion can be chosen to be well TF localized, so that the canonical input signal (3.33) suitably models the fact that real-world transmitted signals are always effectively time and band limited; (v) the link between the resulting discretized channel model and the continuous-time WSSUS model allows to express results directly in terms of the channel's scattering function, which can be estimated from measurements* (Gaarder, 1968; Artés et al., 2004); (vi) the canonical IO relation (3.34) on the basis of the Weyl-Heisenberg transmission set $\{g_{k,n}(t)\}$ can be interpreted as PS-OFDM, which is a practical transmission scheme (see the discussion in Section 3.2.5); furthermore, the matching rule (3.47) for T and F and the conditions on $g(t)$ reviewed in Section 3.2 give design guidelines to jointly mitigate ISI and ICI. With hindsight, these benefits explain the prominent exposition of the discretized underspread WSSUS model in Section 3.2.

B. Multiantenna Extension

Although practical wireless systems are in general designed to serve multiple users, we limit our analysis to a single point-to-point link, because capacity in the noncoherent setting is unknown even for this simple case. We do, however, take a very small first step that might at some future time enable an extension of our results to the multiuser setting: we augment the SISO channel model of Section 3.2 to a MIMO channel model. As will become clear in Section 7.3, this extension is primarily a *formal* one, i.e., without physical modeling support. We directly extend the standard narrowband multiantenna channel description established in the information-theoretic multiantenna literature (Telatar, 1999; Foschini and Gans, 1998; Paulraj et al., 2003; Goldsmith et al., 2003) to the case that the *component channels* between individual transmit and receive antennas are described by the discretized and diagonalized WSSUS channel model from Section 3.2. While we spent much effort in the previous chapters to analyze the

* As discussed in Section 4.3.3 time variant measurements of UWB channels are still extremely challenging to date.

small-scale fading model on which we base the capacity analysis for a point-to-point link in the present chapter, we did not concern ourselves with theoretical and experimental assessment of the spatial channel properties required for a physically sound description of multiantenna channels (Steinbauer et al., 2001; Vaughan and Bach Andersen, 2003; Poon et al., 2004). The standard approach to model the multiantenna channel by means of matrix-valued channel coefficients is sensible for small bandwidth, but becomes questionable as bandwidth increases. Antenna patterns start to play an important role, as they often show great variation with frequency (e.g., see the pattern of the measurement antenna in Figure 4.4), and the narrowband array assumption (van Trees, 2002; Tse and Viswanath, 2005), which lies at the heart of the standard MIMO channel model, breaks down. The question what constitutes a sensible wideband multiantenna channel model is certainly an important and interesting one, but goes far beyond the scope of the present work.

Many spatial channel properties result in spatial correlation of the *component channels* between individual transmit and receive antennas. We provision for such correlation in our multiantenna extension, though we do not describe its physical basis nor do we provide empirical assessment of spatial correlation. But we may hope that the generality of our MIMO extension covers some cases of practical interest, so that future measurement results may be mapped to the model to be presented in Section 7.3.

C. System

The amount of channel knowledge at the transmitter and the receiver can be modeled in different gradations. We use the common assumption that neither transmitter nor receiver know the realization of the channel process, but both know the channel law. This assumption is commonly referred to as the *noncoherent setting*, as opposed to the *coherent setting*, where the receiver has perfect CSI while the transmitter knows the channel law only. Perfect knowledge of the channel law is still a rather strong assumption, yet a much more reasonable

one than perfect knowledge of the channel realization. For the WSSUS model, knowledge of the channel law amounts to knowledge of the scattering function, which the receiver can measure with reasonable accuracy even during data transmission (Artés et al., 2004).

A constraint on the average power of the transmitted signal is suitable to model effects such as a limited battery capacity of mobile devices and limited cooling of integrated circuits. But such an average-power constraint alone is not realistic, as already outlined. All physically realizable power amplifiers can only provide limited output power (Gray et al., 2001), and regulatory bodies often constrain the maximum radiated power at a specified distance from the transmitter. Therefore, we also limit the peak power of the transmitted signal. Various possibilities exist to model such a peak limitation: per antenna or summed over all antennas (Sethuraman et al., 2007); as a constraint on the peak signal value or on higher-order moments (Médard and Gallager, 2002; Subramanian and Hajek, 2002); as a fixed value or in relation to the constraint on the average power (Hajek and Subramanian, 2002); in a given frequency band or accumulated over the entire bandwidth. In consideration of current UWB regulations, we limit the peak radiated power in a given frequency band, i.e., we place a hard limit on the squared magnitude of the transmit signal in a given frequency band and summed over all transmit antennas.

As already discussed, we do not analyze notions of multiuser capacity; neither do we explicitly take into account interference from other users. Instead, we simply aggregate all impairments besides fading in an additive noise term.

7.2. RELATED CAPACITY RESULTS

In the following, we give a brief overview of capacity results for fading channels to put our work into perspective. We limit our discussion to the noncoherent single-user setting, as channel capacity in the coherent setting is well understood and widely documented, e.g., in the extensive survey by Biglieri et al. (1998), or in standard textbooks

like the one by Tse and Viswanath (2005).

A multitude of possibilities exist to model power constraints and channel variation in noncoherent communication systems; equally diverse are the available results on the capacity of those systems. A common trait of all these results, however, is that a closed-form capacity expression is not known, even for the seemingly simple memoryless and frequency flat Rayleigh-fading channel (Abou-Faycal et al., 2001). Nevertheless, several asymptotic results exist for both $\text{SNR} \rightarrow \infty$ and $\text{SNR} \rightarrow 0$.

7.2.1. High-SNR Regime

The performance of wireless systems at high SNR has been the focus of much research in the past in terms of channel capacity, capacity versus outage, and error probability, mainly because many widely used communication system that operate over a relatively narrow frequency band, say, up to several megahertz bandwidth, also operate at high SNR. Because the bandwidth of such systems is small, the high total receive SNR of 20 dB and more in practice implies a high effective SNR, i.e., high SNR per DOF.

As no closed-form expressions for capacity are known in the noncoherent setting, one resorts to characterizing the scaling behavior of capacity as SNR grows without bound. This scaling behavior is highly sensitive to the channel model used. In the AWGN channel and in coherent fading channels, both subject to an average-power constraint, capacity scales logarithmically with increasing power (Telatar, 1999; Tse and Viswanath, 2005); logarithmic capacity scaling with SNR also holds for noncoherent SISO and MIMO block-fading channels with block length larger than the minimum of M_T and M_R , the number of transmit and receive antennas, respectively (Marzetta and Hochwald, 1999; Zheng and Tse, 2002; Liang and Veeravalli, 2004). In all three types of channels, the multiplicative constant in front of the logarithmic term in the high-SNR capacity expansion, referred to as *pre-log* or *multiplexing gain*, can be interpreted as the number of DOFs these channels provide for communication; thus, capacity increases linearly

in the number of DOFs at high SNR (Tse and Viswanath, 2005) for the mentioned channel models.

If the fading process is stationary—this includes memoryless fading and, therefore, the block-fading model of block length one—capacity does not necessarily scale logarithmically with SNR. Lapidath and Moser (2003) show that the noncoherent capacity of general flat fading SISO and MIMO channels scales doubly logarithmic at high SNR if the channel process is of finite differential entropy rate, irrespective of the transmitted average power or the peak power being constrained. This double-logarithmic capacity increase is dominant at very high SNR values, while for smaller, but still large SNR, the *fading number*, which increases linearly in the number of degrees of freedom, is the dominant term in the high-SNR capacity expansion (Lapidath and Moser, 2003; Moser, 2007; Koch and Lapidath, 2005).

Channel processes whose differential entropy rate is unbounded seem to be difficult to analyze in general. An important exception are stationary JPG processes, whose scaling behavior can be characterized in detail in the case of both finite and infinite differential entropy rate (Lapidath, 2003, 2005). If the channel is flat fading, so that the channel process is univariate stationary JPG, its differential entropy rate depends only on the one-step infinite-horizon causal prediction error (Kolmogorov, 1992b,a), i.e., the error in predicting the current channel realization from noiseless observations of its infinite past. Channels with unbounded differential entropy rate have a one-step prediction error of zero and are referred to as *deterministic* because their future can be predicted perfectly given the infinite past of the process. Finite differential entropy rate implies a nonzero one-step prediction error, and the corresponding processes are, therefore, called *nondeterministic*.^{*} The capacity scaling behavior at high SNR is quite diverse for deterministic stationary JPG channel processes (Lapidath, 2003, 2005); in particular, if the spectrum of the channel

^{*} We follow Wiener and Masani (1957) as their terminology appears to be the most descriptive. Kolmogorov (1992b,a) and Rozanov (1967) use the terms *singular* and *nonsingular* instead of deterministic and nondeterministic, respectively.

process vanishes on a set of positive Lebesgue measure, capacity scales logarithmically at high SNR, and the pre-log is determined by said measure. Chen and Veeravalli (2007) extended the latter result to the block-stationary fading channel introduced by Liang and Veeravalli (2004).

7.2.2. Wideband Regime

The terms *low-power regime*, *power-limited regime*, and *wideband regime* are often used interchangeably to mean low effective SNR. While the high-SNR capacity scaling of nondeterministic channels does not seem to depend strongly on the nature of the power constraint—average or peak—the type of power constraint is crucially important at low SNR. Conversely, capacity results seem to be insensitive to whether the channel is deterministic or nondeterministic in the wideband regime.

A. Asymptotic Results

If only a constraint on the average transmitted power is imposed, the infinite bandwidth AWGN capacity can be achieved also in the presence of fading. This result is quite robust, as it holds for a wide variety of channel models (Pierce, 1958; Jacobs, 1963; Gallager, 1968; Kennedy, 1969; Telatar and Tse, 2000; Verdú, 2002; Durisi et al., 2006). Verdú (2002) showed that flash signaling, which implies unbounded peak power of the input symbols, is necessary and sufficient to achieve the infinite-bandwidth AWGN capacity on block-memoryless channels; a form of flash signaling is also infinite-bandwidth optimal for the more general time- and frequency-selective channel model that we use in the present work (Durisi et al., 2006). If the peakiness of the input symbols is restricted, the resulting capacity limit for infinite bandwidth SISO channels is less than the corresponding AWGN capacity (Viterbi, 1967; Gallager, 1968; Telatar and Tse, 2000; Médard and Gallager, 2002; Subramanian and Hajek, 2002). For channels that are selective in time, frequency, or both, the exact capacity behavior depends on

the type of input peak constraint. The analysis by Viterbi (1967) provides a result that can be interpreted as a lower bound on the infinite-bandwidth SISO capacity of a time- and frequency-selective channel if a peak constraint is imposed only in time but not across frequency. For SISO channels that are time-selective but frequency-flat, structurally similar expressions exist for the infinite-bandwidth capacity (Sethuraman and Hajek, 2006; Zhang and Laneman, 2007b) and for the capacity per unit energy (Sethuraman and Hajek, 2005). Durisi et al. (2008) obtain Viterbi's lower bound for the channel model we use in this thesis, and also provide an upper bound that coincides with the lower bound for a specific class of channels.

If we allow for multiple antennas at the transmitter and the receiver, an additional determinant of capacity behavior is spatial correlation. For the *separable* (Kronecker) spatial correlation model (Chuah et al., 2002; Kermaol et al., 2002), Jafar and Goldsmith (2005) proved that transmit correlation increases the capacity of a memoryless fading channel. Moreover, in the low-SNR regime, the rates achievable with on-off keying on memoryless fading channels (Zhang and Laneman, 2007a) and with finite-cardinality constellations on block-fading channels (Srinivasan and Varanasi, 2007) increase in the presence of spatial correlation at the transmitter, the receiver, or both.

Sethuraman et al. (2009) analyzed the capacity of peak-constrained MIMO Rayleigh-fading channels that are frequency flat, time selective, and spatially uncorrelated and derived an upper bound and a low-SNR lower bound that allow to characterize the second-order Taylor series expansion of capacity around the point $\text{SNR} = 0$. In particular, they proved that in the low-SNR regime it is optimal to use only a single transmit antenna, while additional receive antennas are always beneficial. These low-SNR results also apply to a wideband channel with fixed total transmit power and increasing bandwidth if the wideband channel can be decomposed into a set of i.i.d. parallel subchannels in frequency.

B. Nonasymptotic Results

If a peak constraint is imposed in both time and frequency, like we have chosen to do in the present work, a closed-form capacity expression valid for all bandwidth is not available. However, it is known that the infinite-bandwidth capacity is zero for various channel models (Telatar and Tse, 2000; Médard and Gallager, 2002; Subramanian and Hajek, 2002). This asymptotic capacity behavior is often used to discuss the suitability of different signaling schemes in the large bandwidth regime. Yet, even more useful for this purpose would be capacity bounds for finite bandwidth. For frequency flat time-selective SISO channels, such an upper bound was derived by Sethuraman et al. (2005) and subsequently extended by Sethuraman et al. (2009) to frequency-flat, time-selective, and spatially uncorrelated Rayleigh-fading MIMO channels. For the more general time- and frequency-selective case treated here, upper bounds seem to exist only on the rates achievable with specific signaling schemes, namely, for constant-modulus OFDM over underspread SISO WSSUS channels, as derived by Schafhuber et al. (2004), and for unitarily space-frequency coded MIMO OFDM over frequency-selective block-fading channels, derived by Borgmann and Bölcskei (2005).

7.3. SYSTEM MODEL

In the present section, we translate the requirements and modeling assumptions discussed in Section 7.1 into mathematical language. We start from the discretized and diagonalized IO relation (3.34) for underspread WSSUS channels, extend it formally to multiple transmit and receive antennas, specify how the resulting time- and frequency-selective MIMO channel can be spatially correlated, and state the power constraints at the transmitter.

7.3.1. Discretized IO Relation

The IO relation (3.34) describes a channel that is selective in time and frequency. To supplement the spatial dimension, we consider a point-to-point MIMO system with M_T transmit antennas, indexed by q , and M_R receive antennas, indexed by r , and assume that all component channels are WSSUS and identically distributed, though not necessarily independent. Hence, all component channels are characterized by the same scattering function $c_s(\nu, \tau)$ so that they are diagonalized by the same Weyl-Heisenberg set $\{g_{k,n}(t)\}$. For each TF slot (k, n) and component channel (r, q) , the resulting scalar channel coefficient is denoted as $\mathbf{h}_{r,q}[k, n]$. We arrange the coefficients for a given TF slot (k, n) in an $M_R \times M_T$ matrix $\mathbf{H}[k, n]$ with entries $[\mathbf{H}[k, n]]_{r,q} \triangleq \mathbf{h}_{r,q}[k, n]$. Similarly, the symbol transmitted from the q th antenna in the TF slot (k, n) is denoted $\mathbf{x}_q[k, n]$, the symbol received at the r th antenna in the same TF slot is $\mathbf{y}_r[k, n]$, and the corresponding noise random variable is $\mathbf{w}_r[k, n]$, which we assume to be proper Gaussian of unit variance. For each TF slot (k, n) , we define the input, output, and noise vectors respectively as*

$$\mathbf{x}[k, n] \triangleq [\mathbf{x}_0[k, n] \ \mathbf{x}_1[k, n] \ \cdots \ \mathbf{x}_{M_T-1}[k, n]]^T, \quad (7.1a)$$

$$\mathbf{y}[k, n] \triangleq [\mathbf{y}_0[k, n] \ \mathbf{y}_1[k, n] \ \cdots \ \mathbf{y}_{M_T-1}[k, n]]^T, \quad (7.1b)$$

$$\mathbf{w}[k, n] \sim \mathcal{CN}(0, \mathbf{I}_{M_R}). \quad (7.1c)$$

The diagonalized IO relation of the multiantenna channel is then given by a countable set of standard MIMO IO relations of the form

$$\mathbf{y}[k, n] = \mathbf{H}[k, n]\mathbf{x}[k, n] + \mathbf{w}[k, n]. \quad (7.2)$$

Transmitted signals are always effectively limited in time duration and bandwidth. Therefore, we limit the signal of each transmit antenna to a rectangle of $K \times N$ TF slots, i.e., we describe each transmit signal

* To distinguish quantities that pertain to the MIMO IO relation for an individual TF slot (k, n) from the corresponding quantities of the joint time-frequency-space IO relation (7.5) to be introduced momentarily, we use a sans-serif font for the former quantities.

in the canonical form (3.33) defined in Section 3.2.4. For simplicity, we refer to the above rectangle of $K \times N$ TF slots as one *channel use*.

With four parameters to index time, frequency, transmit antennas, and receive antennas, notation can become very cumbersome quickly. Therefore, we devise a compact multivariate notation in the following. We stack the symbols $\{\mathbf{x}_q[k, n]\}$ transmitted from all M_T transmit antennas during one channel use into a large $M_T KN$ -dimensional vector \mathbf{x} , the corresponding output $\{\mathbf{y}_r[k, n]\}$ for all M_R receive antennas into an $M_R KN$ -dimensional vector \mathbf{y} , and likewise the noise $\{\mathbf{w}_r[k, n]\}$ into an $M_R KN$ -dimensional vector \mathbf{w} . Stacking proceeds first along frequency, then along time, and finally along space, as shown exemplarily for the input vector \mathbf{x} :

$$\mathbf{x}_q[k] = [\mathbf{x}_q[k, 0] \ \mathbf{x}_q[k, 1] \ \cdots \ \mathbf{x}_q[k, N-1]]^T, \quad (7.3a)$$

$$\mathbf{x}_q = [\mathbf{x}_q^T[0] \ \mathbf{x}_q^T[1] \ \cdots \ \mathbf{x}_q^T[K-1]]^T, \quad (7.3b)$$

$$\mathbf{x} = [\mathbf{x}_0^T \ \mathbf{x}_1^T \ \cdots \ \mathbf{x}_{M_T-1}^T]^T. \quad (7.3c)$$

In the same way, we stack the channel coefficients first in frequency to obtain the vectors $\mathbf{h}_{r,q}[k]$, and then in time to obtain a vector $\mathbf{h}_{r,q}$ for each component channel (r, q) ; further stacking of these vectors along transmit antennas q and then along receive antennas r results in the $M_R M_T KN$ -dimensional vector \mathbf{h} . In mathematical notation:

$$\mathbf{h}_{r,q}[k] = [\mathbf{h}_{r,q}[k, 0] \ \mathbf{h}_{r,q}[k, 1] \ \cdots \ \mathbf{h}_{r,q}[k, N-1]]^T, \quad (7.4a)$$

$$\mathbf{h}_{r,q} = [\mathbf{h}_{r,q}^T[0] \ \mathbf{h}_{r,q}^T[1] \ \cdots \ \mathbf{h}_{r,q}^T[K-1]]^T, \quad (7.4b)$$

$$\mathbf{h}_r = [\mathbf{h}_{r,0}^T \ \mathbf{h}_{r,1}^T \ \cdots \ \mathbf{h}_{r,M_T-1}^T]^T, \quad (7.4c)$$

$$\mathbf{h} = [\mathbf{h}_0^T \ \mathbf{h}_1^T \ \cdots \ \mathbf{h}_{M_R-1}^T]^T. \quad (7.4d)$$

Here, each component channel process $\{\mathbf{h}_{r,q}[k]\}_k$ is multivariate stationary. Let $\mathbf{X}_q \triangleq \text{diag } \mathbf{x}_q$ and $\mathbf{X} \triangleq [\mathbf{X}_0 \ \mathbf{X}_1 \ \cdots \ \mathbf{X}_{M_T-1}]$, where the vectors \mathbf{x}_q are defined in (7.3b). With this notation, the IO relation of the WSSUS MIMO channel for one channel use can be conveniently expressed as

$$\mathbf{y} = (\mathbf{I}_{M_R} \otimes \mathbf{X})\mathbf{h} + \mathbf{w}. \quad (7.5)$$

7.3.2. Channel Correlation

As we assume that all continuous-time component channels are stochastically modeled by the same scattering function, their discretized counterparts all have the same correlation function, which is matrix valued and given by

$$\begin{aligned} \mathbf{R}_{\mathbf{h}}[\Delta k] &\triangleq \mathbb{E} [\mathbf{h}_{r,q}[\Delta k + k] \mathbf{h}_{r,q}^\dagger[k]] \\ &= \begin{bmatrix} c_l[\Delta k, 0] & \bar{c}_l[\Delta k, 1] & \dots & \bar{c}_l[\Delta k, N-1] \\ c_l[\Delta k, 1] & c_l[\Delta k, 0] & \dots & \bar{c}_l[\Delta k, N-2] \\ \vdots & \vdots & \ddots & \vdots \\ c_l[\Delta k, N-1] & c_l[\Delta k, N-2] & \dots & c_l[\Delta k, 0] \end{bmatrix}. \end{aligned} \quad (7.6)$$

During one channel use, the corresponding completely stacked channel vector \mathbf{h} is characterized by its correlation matrix

$$\begin{aligned} \mathbf{R}_{\mathbf{h}} &= \mathbb{E} [\mathbf{h}_{r,q} \mathbf{h}_{r,q}^\dagger] \\ &= \begin{bmatrix} \mathbf{R}_{\mathbf{h}}[0] & \mathbf{R}_{\mathbf{h}}^\dagger[1] & \dots & \mathbf{R}_{\mathbf{h}}^\dagger[K-1] \\ \mathbf{R}_{\mathbf{h}}[1] & \mathbf{R}_{\mathbf{h}}[0] & \dots & \mathbf{R}_{\mathbf{h}}^\dagger[K-2] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{\mathbf{h}}[K-1] & \mathbf{R}_{\mathbf{h}}[K-2] & \dots & \mathbf{R}_{\mathbf{h}}[0] \end{bmatrix}. \end{aligned} \quad (7.7)$$

This correlation matrix does not depend on the transmit and receive antenna indices either, as we assumed all component channels to have the same scattering function. Because each component channel process $\{\mathbf{h}_{r,q}[k, n]\}_{k,n}$ is stationary in time and in frequency, $\mathbf{R}_{\mathbf{h}}$ is a two-level Hermitian Toeplitz matrix (Tyrtysnikov and Zamarashkin, 1998), i.e., $\mathbf{R}_{\mathbf{h}}$ is block-Toeplitz with Toeplitz blocks.

In addition to correlation over time and frequency as in (7.7), which in turn is a consequences of (3.35), we allow for spatial correlation, albeit in a rather restrictive form to keep the model tractable. In particular, we assume that:

- Spatial correlation does not change over time and frequency. This is a direct extension of the WSSUS assumption for each component channel.

- Spatial correlation is *separable*, i.e., we use the so-called *Kronecker* correlation model of Chuah et al. (2002) and Kermoal et al. (2002). Separability means that spatial correlation is local to the transmitter and the receiver: For any given TF slot, correlation between a component channel (r, q) and a component channel (r', q') can be expressed in terms of a *transmit correlation* coefficient $a_{qq'}$ and a *receive correlation* coefficient $b_{rr'}$.

In combination with the use of a single scattering function for all component channels, the above assumptions on spatial correlation imply that

$$\mathbb{E}[\mathbf{h}_{r,q}[\Delta k + k, \Delta n + n] \bar{\mathbf{h}}_{r',q'}[k, n]] = b_{rr'} a_{qq'} c_l[\Delta k, \Delta n]. \quad (7.8)$$

The $M_T \times M_T$ matrix \mathbf{A} with entries $[\mathbf{A}]_{q,q'} = a_{qq'}$ is called the *transmit correlation matrix*, and the $M_R \times M_R$ matrix \mathbf{B} , with entries $[\mathbf{B}]_{r,r'} = b_{rr'}$, is the *receive correlation matrix*. Both spatial correlation matrices are normalized according to

$$\text{tr } \mathbf{A} = M_T, \quad (7.9a)$$

$$\text{tr } \mathbf{B} = M_R, \quad (7.9b)$$

i.e., the overall receive SNR remains constant irrespectively of the correlation structure. For all (k, n) , let $\mathbf{H}_w[k, n]$ be an $M_R \times M_T$ matrix with i.i.d. JPG entries of zero mean and unit variance. Then, the channel matrix in each TF slot can be written as

$$\mathbf{H}[k, n] = \mathbf{B}^{1/2} \mathbf{H}_w[k, n] (\mathbf{A}^{1/2})^T. \quad (7.10)$$

It now follows from (7.8) that the distribution of the channel in one channel use is completely characterized by the $M_R M_T K N \times M_R M_T K N$ correlation matrix

$$\mathbb{E}[\mathbf{h}\mathbf{h}^\dagger] = \mathbf{B} \otimes \mathbf{A} \otimes \mathbf{R}_h. \quad (7.11)$$

We assume that the three matrices \mathbf{A} , \mathbf{B} , and \mathbf{R}_h are known to the transmitter and the receiver.

7.3.3. Power Constraints

As already discussed in Section 7.1.1, we impose a constraint on the peak transmitted power, as such a constraint appropriately models real-world implementation constraints. In addition, we also constrain the average transmitted power.

A. Average Power

The most commonly used input constraint in information-theoretic analysis is on the average transmitted power. We follow this practice and impose the limitation

$$\frac{1}{T} \mathbb{E} [\|\mathbf{x}\|^2] = \sum_q \sum_{k,n} \mathbb{E} [|\mathbf{x}_q[k, n]|^2] \leq KP. \quad (7.12)$$

An average-power constraint models, for example, the limited battery resources of typical mobile devices and heat dissipation limits of integrated circuits used in the transmitter and the receiver.

B. Peak Power

We opt for a constraint on the peak power summed over all antennas, and relate it to the average power according to

$$\frac{1}{T} \|\mathbf{x}[k, n]\|^2 = \frac{1}{T} \sum_{q=0}^{M_T-1} |\mathbf{x}_q[k, n]|^2 \leq \frac{\beta P}{N} \quad \text{w.p.1} \quad (7.13)$$

for each TF slot. This constraint models, e.g., a limitation of the total radiated peak power in a given frequency band. The parameter β stands for the *nominal* peak- to average-power ratio (PAPR), which is commonly used in circuit design.

The peak constraint (7.13) is imposed on the input symbols $\mathbf{x}[k, n]$, i.e., in the eigenspace of the approximating channel operator. This limitation is mathematically convenient; however, the peak value of the corresponding transmitted continuous time signal also depends on the prototype pulse $g(t)$, so that a limit on $\mathbf{x}[k, n]$ does not imply in general that the transmitted signal is peak limited to the same value.

7.3.4. Spatially Decorrelated IO Relation

Before proceeding to analyze the capacity of the channel just introduced, we make one more cosmetic change to the IO relation (7.5), which simplifies the exposition of our results considerably. For each TF slot, we express the input and output vectors in the coordinate systems defined by the eigendecomposition of the transmit and receive correlation matrices, respectively. A similar transformation is used by Jafar and Goldsmith (2005) and Srinivasan and Varanasi (2007) for a frequency-flat block-fading spatially correlated MIMO channel. Let the eigendecomposition of the spatial correlation matrices be

$$\mathbf{A} = \mathbf{U}_A \mathbf{\Sigma} \mathbf{U}_A^\dagger, \quad (7.14a)$$

$$\mathbf{B} = \mathbf{U}_B \mathbf{\Lambda} \mathbf{U}_B^\dagger. \quad (7.14b)$$

where $\mathbf{\Sigma} = \text{diag}[\sigma_0 \sigma_1 \cdots \sigma_{M_T-1}]^T$ contains the eigenvalues $\{\sigma_q\}$ of \mathbf{A} , ordered according to $\sigma_0 \geq \sigma_1 \geq \cdots \geq \sigma_{M_T-1}$ and, similarly, $\mathbf{\Lambda} = \text{diag}[\lambda_0 \lambda_1 \cdots \lambda_{M_R-1}]^T$ contains the eigenvalues $\{\lambda_r\}$ of \mathbf{B} , ordered according to $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{M_R-1}$. The columns of \mathbf{U}_A are called the *transmit eigenmodes* and the columns of \mathbf{U}_B are the *receive eigenmodes*.

Instead of the vectors $\mathbf{x}[k, n]$ and $\mathbf{y}[k, n]$, we use the rotated vectors $\mathbf{U}_A^T \mathbf{x}[k, n]$ and $\mathbf{U}_B^\dagger \mathbf{y}[k, n]$, respectively, to obtain the following *spatially decorrelated IO relation* in each TF slot (k, n) :

$$\begin{aligned} \mathbf{U}_B^\dagger \mathbf{y}[k, n] &= \mathbf{U}_B^\dagger \mathbf{H}[k, n] \mathbf{x}[k, n] + \mathbf{U}_B^\dagger \mathbf{w}[k, n] \\ &\stackrel{(a)}{=} \mathbf{U}_B^\dagger (\mathbf{U}_B \mathbf{\Lambda}^{1/2} \mathbf{U}_B^\dagger) \mathbf{H}_w[k, n] (\mathbf{U}_A \mathbf{\Sigma}^{1/2} \mathbf{U}_A^\dagger)^T \mathbf{x}[k, n] \\ &\quad + \mathbf{U}_B^\dagger \mathbf{w}[k, n] \\ &= \mathbf{\Lambda}^{1/2} \mathbf{U}_B^\dagger \mathbf{H}_w[k, n] \overline{\mathbf{U}_A} \mathbf{\Sigma}^{1/2} \mathbf{U}_A^T \mathbf{x}[k, n] + \mathbf{U}_B^\dagger \mathbf{w}[k, n], \end{aligned} \quad (7.15)$$

where (a) follows from (7.10). Rotations are unitary operations; therefore, $\mathbf{U}_B^\dagger \mathbf{H}_w[k, n] \overline{\mathbf{U}_A} \sim \mathbf{H}_w[k, n]$ and $\mathbf{U}_B^\dagger \mathbf{w}[k, n] \sim \mathbf{w}[k, n]$. Furthermore, rotations preserve norms, so that the rotated input vector $\mathbf{U}_A^T \mathbf{x}[k, n]$ satisfies the same power constraints as the unrotated input vector $\mathbf{x}[k, n]$. Finally, $\mathbf{U}_B^\dagger \mathbf{y}[k, n]$ is a sufficient statistic for the output

vector $\mathbf{y}[k, n]$. These three properties imply that the capacity of the channel with input $\mathbf{x}[k, n]$ and output $\mathbf{y}[k, n]$ in (7.2) is the same as the capacity of the spatially decorrelated channel $\Lambda^{1/2}\mathbf{H}_w[k, n]\Sigma^{1/2}$ in (7.15) with input $U_A^T\mathbf{x}[k, n]$ and output $U_B^\dagger\mathbf{y}[k, n]$. In the new coordinate system, q indexes transmit eigenmodes instead of transmit antennas, and r indexes receive eigenmodes instead of receive antennas.

It is now tedious but straightforward to similarly rotate the stacked IO relation (7.5). To keep notation simple, we chose not to introduce new symbols for the rotated input and output and for the spatially decorrelated channel; from here on, all inputs and outputs are with respect to the rotated coordinate systems, and the channel vector \mathbf{h} now stands for the spatially decorrelated stacked channel with correlation matrix

$$\mathbb{E}[\mathbf{h}\mathbf{h}^\dagger] = \Lambda \otimes \Sigma \otimes R_{\mathbf{h}}, \quad (7.16)$$

This correlation matrix is block diagonal, and hence of much simpler structure than (7.11).

7.4. AN UPPER BOUND ON CAPACITY

In the present and the next sections, we analyze the capacity of the diagonalized underspread WSSUS multiantenna channel in (7.5) subject to the constraint on the average transmitted power (7.12) and to the peak constraint (7.13). The link established in Section 3.2 between the diagonalized component channels (3.34) and the continuous-time WSSUS channel model \mathbb{H} then allows us to express the resulting bounds in terms of the channel's scattering function $c_s(\nu, \tau)$.

Operational channel capacity is the ultimate rate at which transmission with arbitrarily small error probability is possible. If the channel varies ergodically over time, a coding theorem exists, so that operational capacity is equivalent to information-theoretic capacity. Ergodicity is a modeling assumption, similar to the WSSUS assumption in Section 2.5.3 or the assumption on spatial independence in

Section 4.4.1.B. Over time intervals of practical interest, i.e., time intervals over which the small-scale fading assumption is reasonable, we can sensibly assume that $\{\mathbf{h}_{r,q}[k, n]\}$ is ergodic in k . In fact, we implicitly made this assumption already when we required the process $\{\mathbf{h}_{r,q}[k, n]\}$ to be JPC and to have a spectral density (3.36) for all component channels (r, q) , as these conditions guarantee ergodicity (Maruyama, 1949). Hence, operational and information-theoretic capacity coincide, and we henceforth simply refer to both as capacity. It is given by (Gray, 2007, Chapter 12)

$$C(B) = \lim_{K \rightarrow \infty} \frac{1}{KT} \sup_{\mathcal{P}} I(\mathbf{x}; \mathbf{y}), \quad (7.17)$$

for any fixed bandwidth $B = NF$. The supremum is taken over the set \mathcal{P} of all input distributions that satisfy the constraints on the average transmitted power (7.12) and on the peak power (7.13) in Section 7.3.3.

7.4.1. Bounding Idea

Our derivation of an upper bound on (7.17) is rather involved; not because the bounding idea is intricate, but simply because the system model in Section 7.3 is quite general, so that we need to take care of many technical details that may obscure the main thoughts. To expose the key bounding steps without the clutter that comes with the complicated model, we illustrate them here for a very simple setting.

We briefly deviate from the notation introduced in Section 7.3, but do so only in the present subsection. Let $\mathbf{h} \sim \mathcal{CN}(0, 1)$ denote the random gain of a memoryless flat-fading SISO channel with input \mathbf{x} , output \mathbf{y} , and additive noise $\mathbf{w} \sim \mathcal{CN}(0, 1)$, i.e., with IO relation $\mathbf{y} = \mathbf{h}\mathbf{x} + \mathbf{w}$. Let \mathcal{P} denote the set of all distributions on \mathbf{x} that satisfy the average-power constraint $\mathbb{E}[|\mathbf{x}|^2] \leq P$ and the peak constraint $|\mathbf{x}|^2 \leq \beta P$ w.p.1. We obtain an upper bound on the capacity $C = \sup_{\mathcal{P}} I(\mathbf{x}; \mathbf{y})$

per channel use along the lines of Sethuraman and Hajek (2006):

$$\begin{aligned}
C &\stackrel{(a)}{=} \sup_{\mathcal{P}} \{I(\mathbf{x}, \mathbf{h}; \mathbf{y}) - I(\mathbf{h}; \mathbf{y} | \mathbf{x})\} \\
&\stackrel{(b)}{\leq} \sup_{\mathcal{P}} \left\{ \log(1 + \mathbb{E}[|\mathbf{x}|^2]) - \mathbb{E} \left[\log(1 + |\mathbf{x}|^2) \right] \right\} \\
&= \sup_{0 \leq \alpha \leq 1} \sup_{\substack{\mathcal{P} \\ \mathbb{E}[|\mathbf{x}|^2] = \alpha P}} \left\{ \log(1 + \mathbb{E}[|\mathbf{x}|^2]) - \mathbb{E} \left[\log(1 + |\mathbf{x}|^2) \right] \right\} \\
&\leq \sup_{0 \leq \alpha \leq 1} \left\{ \log(1 + \alpha P) - \inf_{\substack{\mathcal{P} \\ \mathbb{E}[|\mathbf{x}|^2] = \alpha P}} \mathbb{E} \left[\log(1 + |\mathbf{x}|^2) \right] \right\} \\
&\stackrel{(c)}{\leq} \sup_{0 \leq \alpha \leq 1} \left\{ \log(1 + \alpha P) \right. \\
&\quad \left. - \inf_{\substack{\mathcal{P} \\ \mathbb{E}[|\mathbf{x}|^2] = \alpha P}} \mathbb{E} \left[\inf_{|x|^2 \leq \beta P} \frac{\log(1 + |x|^2)}{|x|^2} |\mathbf{x}|^2 \right] \right\} \\
&\stackrel{(d)}{=} \sup_{0 \leq \alpha \leq 1} \left\{ \log(1 + \alpha P) - \frac{\alpha}{\beta} \log(1 + \beta P) \right\}. \tag{7.18}
\end{aligned}$$

Here, (a) follows from the chain rule for mutual information (Cover and Thomas, 1991); the inequality (b) results because we take $\mathbf{h}\mathbf{x}$ as JPG with variance $\mathbb{E}[|\mathbf{h}\mathbf{x}|^2] = \mathbb{E}[|\mathbf{x}|^2]$. To obtain the inequality (c), we multiply and divide the second term inside the braces by $|\mathbf{x}|^2$ and lower-bound the first factor inside the expectation by its infimum over all inputs x that satisfy the peak constraint. Finally, (d) results because $\log(1 + |x|^2)/|x|^2$ is convex and monotonically decreasing, so that its infimum is achieved for $|x|^2 = \beta P$. If the supremum in (7.18) is achieved for $\alpha = 1$, the upper bound simplifies to

$$C \leq \log(1 + P) - \frac{1}{\beta} \log(1 + \beta P). \tag{7.19}$$

This bound can be interpreted as the capacity of an AWGN channel with SNR P minus a penalty term that quantifies the capacity loss because of channel uncertainty. The higher the allowed peakiness of the input, as measured by the PAPR β , the smaller the penalty.

With $\log(1+x) = x + x^2/2 + o(x^2)$, the upper bound (7.19) can be expanded into a Taylor series around $P = 0$ as $P^2(\beta - 1)/2 + o(P^2)$. This expansion matches the Taylor series derived by Wang (2006) for more general time-selective fading channels. Hence, the simple bound (7.19) has the correct low-SNR behavior. However, its asymptotic behavior for increasing SNR is far too optimistic: the bound (7.19) scales logarithmically in P , while the correct high-SNR scaling is doubly logarithmic for memoryless channels (Taricco and Elia, 1997; Lapidoth and Moser, 2003).

7.4.2. The Upper Bound

In this section, we expand the simple bounding idea just presented to the full-fledged system model from Section 7.3. The resulting upper bound on capacity $C(B)$ is given in the following theorem.

Theorem 7.1. *The capacity (7.17) of the underspread WSSUS MIMO channel with spatial correlation described in Section 7.3, whose average input power is constrained as (7.12) and whose peak power follows (7.13) is upper-bounded as $C(B) \leq U_1(B)$, where*

$$U_1(B) \triangleq \sup_{0 \leq \alpha \leq \sigma_0} \sum_{r=0}^{M_R-1} \left\{ \frac{B}{TF} \log \left(1 + \alpha \lambda_r \frac{PTF}{B} \right) - \alpha \underbrace{\frac{B}{\sigma_0 \beta} \iint_{\nu \tau} \log \left(1 + \frac{\sigma_0 \lambda_r \beta P}{B} c_s(\nu, \tau) \right) d\tau d\nu}_{\triangleq G_r(B)} \right\}, \quad (7.20)$$

and σ_0 denotes the maximum eigenvalue of the transmit correlation matrix \mathbf{A} .

Proof. We initially fix K . The first step to upper-bound (7.17) is to relax the constraint on the average power (7.12). Let \mathcal{Q} denote the set of input distributions that satisfy

$$\frac{1}{T} \sum_{q=0}^{M_T-1} \sigma_q \mathbb{E} [\|\mathbf{x}_q\|^2] \leq \sigma_0 KP \quad (7.21)$$

and the peak constraint (7.13). As

$$\sum_{q=0}^{M_T-1} \sigma_q \mathbb{E}[\|\mathbf{x}_q\|^2] \leq \sigma_0 \sum_{q=0}^{M_T-1} \mathbb{E}[\|\mathbf{x}_q\|^2] = \sigma_0 \mathbb{E}[\|\mathbf{x}\|^2], \quad (7.22)$$

any input distribution that satisfies the average-power constraint $\mathbb{E}[\|\mathbf{x}\|^2]/T \leq KP$ in (7.12) also satisfies (7.21), so that $\mathcal{P} \subset \mathcal{Q}$. To upper-bound $C(B)$, we replace the supremum over \mathcal{P} in the capacity formula (7.17) with a supremum over \mathcal{Q} , use the chain rule to decompose the mutual information as

$$\sup_{\mathcal{Q}} I(\mathbf{x}; \mathbf{y}) = \sup_{\mathcal{Q}} \{I(\mathbf{x}, \mathbf{h}; \mathbf{y}) - I(\mathbf{h}; \mathbf{y} | \mathbf{x})\} \quad (7.23)$$

and then split the supremum over \mathcal{Q} in two parts:

1. One supremum over a restricted set $\mathcal{Q}|_{\alpha}$ of input distributions that satisfy the peak constraint (7.13) and attain the relaxed constraint on the average power, i.e., satisfy (7.21) with equality for some fixed parameter $\alpha \in [0, \sigma_0]$:

$$\frac{1}{T} \sum_{q=0}^{M_T-1} \sigma_q \mathbb{E}[\|\mathbf{x}_q\|^2] = \alpha KP. \quad (7.24)$$

2. Another supremum over the parameter α .

Consequently,

$$\begin{aligned} \sup_{\mathcal{P}} I(\mathbf{x}; \mathbf{y}) &\leq \sup_{\mathcal{Q}} I(\mathbf{x}; \mathbf{y}) \\ &= \sup_{\mathcal{Q}} \{I(\mathbf{x}, \mathbf{h}; \mathbf{y}) - I(\mathbf{h}; \mathbf{y} | \mathbf{x})\} \\ &= \sup_{0 \leq \alpha \leq \sigma_0} \sup_{\mathcal{Q}|_{\alpha}} \{I(\mathbf{x}, \mathbf{h}; \mathbf{y}) - I(\mathbf{h}; \mathbf{y} | \mathbf{x})\} \\ &\leq \sup_{0 \leq \alpha \leq \sigma_0} \left\{ \sup_{\mathcal{Q}|_{\alpha}} I(\mathbf{x}, \mathbf{h}; \mathbf{y}) - \inf_{\mathcal{Q}|_{\alpha}} I(\mathbf{h}; \mathbf{y} | \mathbf{x}) \right\}. \end{aligned} \quad (7.25)$$

Next, we bound the two terms inside the braces individually. While standard steps suffice for the bound on the first term, the second

term requires some auxiliary derivations that are rather technical. We focus on the main steps here and relegate detailed derivations of the auxiliary results to Chapter 8.

A. Upper Bound on the First Term

The output vector \mathbf{y} depends on the input vector \mathbf{x} and the channel vector \mathbf{h} only through $\mathbf{s} \triangleq (\mathbf{I}_{M_R} \otimes \mathbf{X})\mathbf{h}$, so that $I(\mathbf{x}, \mathbf{h}; \mathbf{y}) = I(\mathbf{s}; \mathbf{y})$. We drop the peak constraint on the input distributions in $\mathcal{Q}|_\alpha$ but retain the relaxed constraint on the average power (7.24). Let $\mathcal{R}|_\alpha$ denote the resulting set of distributions. We obtain an upper bound on $I(\mathbf{s}; \mathbf{y})$ if we choose \mathbf{x} such that \mathbf{s} is JCG distributed with zero mean and covariance matrix

$$\begin{aligned}
 \mathbb{E}[\mathbf{s}\mathbf{s}^\dagger] &= \mathbb{E}[(\mathbf{I}_{M_R} \otimes \mathbf{X})\mathbf{h}\mathbf{h}^\dagger(\mathbf{I}_{M_R} \otimes \mathbf{X})^\dagger] \\
 &= \mathbb{E}[(\mathbf{I}_{M_R} \otimes \mathbf{X})(\Lambda \otimes \Sigma \otimes \mathbf{R}_h)(\mathbf{I}_{M_R} \otimes \mathbf{X})^\dagger] \\
 &= \mathbb{E}[(\Lambda \otimes \mathbf{X}(\Sigma \otimes \mathbf{R}_h))(\mathbf{I}_{M_R} \otimes \mathbf{X})^\dagger] \\
 &= \Lambda \otimes \mathbb{E}[\mathbf{X}(\Sigma \otimes \mathbf{R}_h)\mathbf{X}^\dagger] \\
 &\stackrel{(a)}{=} \Lambda \otimes \sum_{q=0}^{M_T-1} \sigma_q \mathbb{E}[\mathbf{x}_q \mathbf{x}_q^\dagger] \odot \mathbf{R}_h. \tag{7.26}
 \end{aligned}$$

In the above derivation, we used several times that for matrices A, B, C, and D of appropriate dimension (Horn and Johnson, 1991, Lemma 4.2.10)

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}. \tag{7.27}$$

Furthermore, the equation (a) in (7.26) results because

$$\mathbf{X}(\Sigma \otimes \mathbf{R}_h)\mathbf{X}^\dagger = \sum_{q=0}^{M_T-1} \sigma_q \mathbf{X}_q \mathbf{R}_h \mathbf{X}_q^\dagger = \sum_{q=0}^{M_T-1} \sigma_q \mathbf{x}_q \mathbf{x}_q^\dagger \odot \mathbf{R}_h. \tag{7.28}$$

As $\mathcal{Q}|_\alpha \subset \mathcal{R}|_\alpha$, we can now upper-bound the first term inside the

braces in (7.25) as follows:

$$\begin{aligned}
 \sup_{\mathcal{Q}|\alpha} I(\mathbf{x}; \mathbf{h}, \mathbf{y}) &= \sup_{\mathcal{Q}|\alpha} I(\mathbf{s}; \mathbf{y}) \leq \sup_{\mathcal{R}|\alpha} I(\mathbf{s}; \mathbf{y}) \\
 &\stackrel{(a)}{\leq} \sup_{\mathcal{R}|\alpha} \sum_{r=0}^{M_R-1} \log \det \left(\mathbf{I}_{KN} + \lambda_r \sum_{q=0}^{M_T-1} \sigma_q \mathbb{E} [\mathbf{x}_q \mathbf{x}_q^\dagger] \odot \mathbf{R}_h \right) \\
 &\stackrel{(b)}{\leq} \sup_{\mathcal{R}|\alpha} \sum_{r=0}^{M_R-1} \sum_{n=0}^{N-1} \sum_{k=0}^{K-1} \log \left(1 + \lambda_r \sum_{q=0}^{M_T-1} \sigma_q \mathbb{E} [|\mathbf{x}_q[k, n]|^2] \right) \\
 &\stackrel{(c)}{\leq} KN \sum_{r=0}^{M_R-1} \log \left(1 + \frac{\alpha \lambda_r PT}{N} \right) \\
 &\stackrel{(d)}{=} \frac{KB}{F} \sum_{r=0}^{M_R-1} \log \left(1 + \frac{\alpha \lambda_r PTF}{B} \right). \tag{7.29}
 \end{aligned}$$

Here, (a) follows from the entropy-maximizing property of JPG random vectors (Neeser and Massey, 1993) and from (7.26); we can sum over the receive eigenmodes because the covariance matrix (7.26) of the JPG input vector \mathbf{s} is block diagonal. As the channel taps are normalized so that $[\mathbf{R}_h]_{i,i} = 1$ for all i , Hadamard's inequality gives (b). Next, (c) follows from Jensen's inequality and the constraint on the average power (7.24). Finally, we obtain (d) upon substitution of the physical channel parameter B for NF .

B. Lower Bound on the Second Term

We use the fact that the channel \mathbf{h} is JPG, so that

$$\begin{aligned}
 \inf_{\mathcal{Q}|\alpha} I(\mathbf{h}; \mathbf{y} | \mathbf{x}) &= \mathbb{E} \left[\log \det (\mathbf{I}_{M_R KN} + \Lambda \otimes \mathbf{X} (\Sigma \otimes \mathbf{R}_h) \mathbf{X}^\dagger) \right] \\
 &= \sum_{r=0}^{M_R-1} \mathbb{E} \left[\log \det \left(\mathbf{I}_{KN} + \lambda_r \sum_{q=0}^{M_T-1} \sigma_q \mathbf{x}_q \mathbf{x}_q^\dagger \odot \mathbf{R}_h \right) \right], \tag{7.30}
 \end{aligned}$$

where the second equality once more follows because $\Lambda \otimes \mathbf{X} (\Sigma \otimes \mathbf{R}_h) \mathbf{X}^\dagger$ is block diagonal and from (7.28). Then, we expand the expectation

operator and multiply and divide by $\sum_{q=0}^{M_T-1} \sigma_q \|\mathbf{x}_q\|^2$:

$$\begin{aligned}
& \inf_{\mathcal{Q}|\alpha} I(\mathbf{h}; \mathbf{y} | \mathbf{x}) \\
&= \inf_{\mathcal{Q}|\alpha} \sum_{r=0}^{M_R-1} \mathbb{E} \left[\log \det \left(\mathbf{I}_{KN} + \lambda_r \sum_{q=0}^{M_T-1} \sigma_q \mathbf{x}_q \mathbf{x}_q^\dagger \odot \mathbf{R}_h \right) \right] \\
&= \inf_{G_{\mathbf{x}} \in \mathcal{Q}|\alpha} \sum_{r=0}^{M_R-1} \\
&\quad \int_{\mathbf{x}} \frac{\log \det \left(\mathbf{I}_{KN} + \lambda_r \sum_{q=0}^{M_T-1} \sigma_q \mathbf{x}_q \mathbf{x}_q^\dagger \odot \mathbf{R}_h \right)}{\sum_{q=0}^{M_T-1} \sigma_q \|\mathbf{x}_q\|^2} \\
&\quad \quad \quad \times \left(\sum_{q=0}^{M_T-1} \sigma_q \|\mathbf{x}_q\|^2 \right) dG_{\mathbf{x}}. \quad (7.31)
\end{aligned}$$

Both factors inside the integral are nonnegative;* hence, we obtain a lower bound if we replace the first factor by its infimum over all deterministic input vectors \mathbf{x} that satisfy the peak constraint (7.13); furthermore, we exchange the order of infimum and summation. Then,

$$\begin{aligned}
& \inf_{\mathcal{Q}|\alpha} I(\mathbf{h}; \mathbf{y} | \mathbf{x}) \\
&\geq \sum_{r=0}^{M_R-1} \inf_{G_{\mathbf{x}} \in \mathcal{Q}|\alpha} \\
&\quad \int_{\mathbf{x}} \frac{\log \det \left(\mathbf{I}_{KN} + \lambda_r \sum_{q=0}^{M_T-1} \sigma_q \mathbf{x}_q \mathbf{x}_q^\dagger \odot \mathbf{R}_h \right)}{\sum_{q=0}^{M_T-1} \sigma_q \|\mathbf{x}_q\|^2} \\
&\quad \quad \quad \times \left(\sum_{q=0}^{M_T-1} \sigma_q \|\mathbf{x}_q\|^2 \right) dG_{\mathbf{x}}
\end{aligned}$$

* The point $\mathbf{x} = 0$ can be excluded from the integration interval as it is a zero of the integrand.

$$\begin{aligned}
 &\geq \sum_{r=0}^{M_R-1} \inf_{\mathbf{x}} \frac{\log \det \left(\mathbf{I}_{KN} + \lambda_r \sum_{q=0}^{M_T-1} \sigma_q \mathbf{x}_q \mathbf{x}_q^\dagger \odot \mathbf{R}_{\mathbf{h}} \right)}{\sum_{q=0}^{M_T-1} \sigma_q \|\mathbf{x}_q\|^2} \\
 &\quad \times \underbrace{\inf_{\mathcal{Q}|\alpha} \mathbb{E} \left[\sum_{q=0}^{M_T-1} \sigma_q \|\mathbf{x}_q\|^2 \right]}_{\alpha KPT} \\
 &= \alpha KPT \sum_{r=0}^{M_R-1} \inf_{\mathbf{x}} \frac{\log \det \left(\mathbf{I}_{KN} + \lambda_r \sum_{q=0}^{M_T-1} \sigma_q \mathbf{x}_q \mathbf{x}_q^\dagger \odot \mathbf{R}_{\mathbf{h}} \right)}{\sum_{q=0}^{M_T-1} \sigma_q \|\mathbf{x}_q\|^2} \\
 &\stackrel{(a)}{\geq} \alpha KPT \sum_{r=0}^{M_R-1} \inf_{\mathbf{x}} \frac{\log \det \left(\mathbf{I}_{KN} + \lambda_r \sum_{q=0}^{M_T-1} \sigma_q \mathbf{X}_q^\dagger \mathbf{X}_q \mathbf{R}_{\mathbf{h}} \right)}{\sum_{q=0}^{M_T-1} \sigma_q \|\mathbf{x}_q\|^2}. \tag{7.32}
 \end{aligned}$$

Inequality (a) follows from Lemma 8.1, which states that for any two nonnegative definite matrices \mathbf{A} and \mathbf{B} of equal dimension, $\det(\mathbf{I} + \mathbf{A} \odot \mathbf{B}) \geq \det(\mathbf{I} + (\mathbf{I} \odot \mathbf{A})\mathbf{B})$. We prove this lemma in Section 8.1. As the matrix $\mathbf{R}_{\mathbf{h}}$ is nonnegative definite, the infimum on the right-hand side in (7.32) is achieved on the boundary of the admissible set, as shown by Sethuraman and Hajek (2005, Section VI.A). In other words, the vector \mathbf{x} that minimizes the right-hand side of (7.32) has entries that satisfy

$$\sum_{q=0}^{M_T-1} |x_q[k, n]|^2 \in \{0, \beta PT/N\} \tag{7.33}$$

for all TF slots (k, n) . We use this fact and the relation between mutual information and the minimum mean squared error (MMSE), recently discovered by Guo et al. (2005) to further lower-bound the

infimum on the right-hand side of (7.32) as

$$\begin{aligned} \inf_{\mathbf{x}} & \frac{\log \det \left(\mathbf{I}_{KN} + \lambda_r \sum_{q=0}^{M_T-1} \sigma_q \mathbf{X}_q^\dagger \mathbf{X}_q \mathbf{R}_h \right)}{\sum_{q=0}^{M_T-1} \sigma_q \|\mathbf{x}_q\|^2} \\ & \geq \frac{N}{\sigma_0 \beta PT} \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \log \left(1 + \frac{\sigma_0 \lambda_r \beta PT}{N} \psi(\varphi, \zeta) \right) d\varphi d\zeta \quad (7.34) \end{aligned}$$

where $\psi(\varphi, \zeta)$ is the two-dimensional spectral density of the multivariate component channel processes $\{\mathbf{h}_{r,q}[k, n]\}_{k,n}$ defined in (3.36). The derivation of the bound (7.34) is detailed in Section 8.3, where we also show that the bound can be achieved for $B \rightarrow \infty$.

Finally, we substitute (7.34) into (7.32) and relate $\psi(\varphi, \zeta)$ to the scattering function $c_s(\nu, \tau)$ by means of (3.37). Hence,

$$\begin{aligned} & \inf_{\mathcal{Q}_\alpha} I(\mathbf{h}; \mathbf{y} | \mathbf{x}) \\ & \geq \frac{\alpha KN}{\sigma_0 \beta} \sum_{r=0}^{M_R-1} \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \log \left[1 + \frac{\sigma_0 \lambda_r \beta PT}{N} \right. \\ & \quad \left. \times \sum_{\Delta k=-\infty}^{\infty} \sum_{\Delta n=-\infty}^{\infty} c_s \left(\frac{\varphi + \Delta k}{T}, \frac{\zeta + \Delta n}{F} \right) \right] d\varphi d\zeta \\ & = \frac{\alpha KN}{\sigma_0 \beta} \sum_{r=0}^{M_R-1} \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \log \left[1 + \frac{\sigma_0 \lambda_r \beta PT}{N} c_s \left(\frac{\varphi}{T}, \frac{\zeta}{F} \right) \right] d\varphi d\zeta \\ & = \frac{\alpha KTNF}{\sigma_0 \beta} \sum_{r=0}^{M_R-1} \iint_{\nu \tau} \log \left(1 + \frac{\sigma_0 \lambda_r \beta P}{NF} c_s(\nu, \tau) \right) d\tau d\nu \\ & = \frac{\alpha KTB}{\sigma_0 \beta} \sum_{r=0}^{M_R-1} \iint_{\nu \tau} \log \left(1 + \frac{\sigma_0 \lambda_r \beta P}{B} c_s(\nu, \tau) \right) d\tau d\nu, \quad (7.35) \end{aligned}$$

where we follow the derivation in (3.39), i.e., we use the underspread property to eliminate the double summation, followed by a change of

variables. As in the upper bound on the first term (7.29), we substitute the physical channel parameter B for NF to obtain the last equality.

C. Completing the Proof

The bounds in (7.29) and (7.35) are valid for any K , in particular for $K \rightarrow \infty$. Consequently, we insert (7.29) and (7.35) in (7.25), divide by KT , and set $B = NF$ to obtain the upper bound (7.20) on capacity (7.17). \square

7.4.3. Refinement for Channels of Practical Interest

As the value of α that achieves the supremum in (7.20) depends on B in general, the upper bound $U_1(B)$ is difficult to interpret. However, for the special case that the supremum is attained for $\alpha = \sigma_0$ independently of B , the upper bound can be interpreted as the capacity of a set of M_R parallel AWGN channels with received power $\sigma_0 \lambda_r P$ and $B/(TF)$ DOFs per second, minus a penalty term that quantifies the capacity loss because of channel uncertainty. Fortunately, it turns out that only the latter case is of practical interest. Let $\tilde{\Delta}_{\mathbb{H}} \triangleq TF\Delta_{\mathbb{H}}/\beta$ denote the *effective spread* and recall that the maximum eigenvalue of the receive correlation matrix B is λ_0 . As we show in Section 8.4, a sufficient condition for the supremum in (7.20) to be achieved for $\alpha = \sigma_0$ is

$$\tilde{\Delta}_{\mathbb{H}} \leq 1/3 \tag{7.36a}$$

and

$$0 \leq \frac{PTF}{B} < \frac{\tilde{\Delta}_{\mathbb{H}}}{\sigma_0 \lambda_0} \left[\exp\left(\frac{1}{2\tilde{\Delta}_{\mathbb{H}}}\right) - 1 \right]. \tag{7.36b}$$

As virtually all wireless channels are highly underspread, as $\beta \geq 1$ and, typically, $TF \approx 1.25$, condition (7.36a) is always satisfied, so that the only relevant condition is (7.36b); but even for large channel

spread, this condition holds for all SNR values P/B of practical interest. As an example, consider a system with $\beta = 1$, and $M_T = M_R = 4$ that operates over a channel with spread $\Delta_{\mathbb{H}} = 10^{-2}$. If we use the upper bound $\sigma_0 \lambda_0 \leq M_R M_T$, which follows from the normalization (7.9), we find from (7.36) that $P/B < 141$ dB is sufficient for the supremum in (7.20) to be achieved for $\alpha = \sigma_0$. The conditions (7.36) are analytically tractable but rather loose. A numerical analysis with the parameters of the above example indicates that $\alpha = \sigma_0$ even for P/B up to 315 dB. These values are far in excess of the receive SNR in practical systems. Therefore, we exclusively consider the case $\alpha = \sigma_0$ in the following.

7.4.4. Impact of Channel Characteristics

The number of transmit and receive antennas, the spatial correlation structure, the spread $\Delta_{\mathbb{H}}$, and the shape of the scattering function $c_{\mathbf{s}}(\nu, \tau)$ are important characteristics of wireless channels. As the upper bound (7.20) is explicit in the scattering function, we can analyze the impact of $c_{\mathbf{s}}(\nu, \tau)$ on $U_1(B)$. As just discussed, we restrict our discussion to the practically relevant case $\alpha = \sigma_0$.

A. Spread

For a fixed shape of the scattering function, the upper bound $U_1(B)$ decreases for increasing channel spread $\Delta_{\mathbb{H}}$. To see this, we define a normalized scattering function $\tilde{c}_{\mathbf{s}}(\nu, \tau)$ with unit spread, so that $c_{\mathbf{s}}(\nu, \tau) = \tilde{c}_{\mathbf{s}}(\nu/\nu_0, \tau/\tau_0)/\Delta_{\mathbb{H}}$. By a change of variables, the penalty term $G_r(B)$ in (7.20) can be written as

$$\begin{aligned} G_r(B) &= \frac{B}{\sigma_0 \beta} \sum_{r=0}^{M_R-1} \iint_{\nu \tau} \log \left(1 + \frac{\sigma_0 \lambda_r \beta P}{B} c_{\mathbf{s}}(\nu, \tau) \right) d\tau d\nu \\ &= \frac{B \Delta_{\mathbb{H}}}{\sigma_0 \beta} \sum_{r=0}^{M_R-1} \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \log \left(1 + \frac{\sigma_0 \lambda_r \beta P}{B \Delta_{\mathbb{H}}} \tilde{c}_{\mathbf{s}}(\nu, \tau) \right) d\tilde{\tau} d\tilde{\nu}. \quad (7.37) \end{aligned}$$

Because $\Delta_{\mathbb{H}} \log(1 + \rho/\Delta_{\mathbb{H}})$ is monotonically increasing in $\Delta_{\mathbb{H}}$ for any positive constant $\rho > 0$, the penalty term $G_r(B)$ increases with increasing spread. On the other hand, the first term on the right-hand side of (7.20) does not depend on the spread; consequently, the upper bound $U_1(B)$ decreases with increasing spread. The larger the spread, the more rapidly does the channel change either in time, in frequency, or in both domains. Intuitively, a large spread means little memory in the channel that the receiver can exploit to reduce channel uncertainty. Consequently, we expect the capacity to decrease with increasing spread—an intuition that is consistent with the above finding.

The penalty term (7.37) expressed in terms of the normalized scattering function $\tilde{c}_s(\nu, \tau)$ depends on $\Delta_{\mathbb{H}}$ and β only through the ratio $\beta/\Delta_{\mathbb{H}}$. Hence, increasing the peakiness of the input signal, as measured by β , has the same effect as transmitting over a channel with lower spread.

B. Shape of the Scattering Function

For fixed τ_0 and ν_0 , and hence fixed spread $\Delta_{\mathbb{H}}$, the scattering function that results in the lowest upper bound $U_1(B)$ is the “brick-shaped” scattering function:

$$c_s(\nu, \tau) = 1/\Delta_{\mathbb{H}}, \quad (\nu, \tau) \in [-\nu_0/2, \nu_0/2] \times [-\tau_0/2, \tau_0/2]. \quad (7.38)$$

We prove this claim in two steps. First, we apply Jensen’s inequality to the penalty term $G_r(B)$ in (7.20),

$$\begin{aligned} & \iint_{\nu \tau} \log \left(1 + \frac{\sigma_0 \lambda_r \beta P}{B} c_s(\nu, \tau) \right) d\tau d\nu \\ & \leq \Delta_{\mathbb{H}} \log \left(1 + \frac{\sigma_0 \lambda_r \beta P}{B \Delta_{\mathbb{H}}} \iint_{\nu \tau} c_s(\nu, \tau) d\tau d\nu \right) \\ & = \Delta_{\mathbb{H}} \log \left(1 + \frac{\sigma_0 \lambda_r \beta P}{\Delta_{\mathbb{H}} B} \right). \end{aligned} \quad (7.39)$$

Second, we note that a brick-shaped scattering function achieves this upper bound. The worst-case upper bound is thus

$$C \leq \frac{B}{TF} \sum_{r=0}^{M_R-1} \left\{ \log \left(1 + \sigma_0 \lambda_r \frac{PTF}{B} \right) - \tilde{\Delta}_{\mathbb{H}} \log \left(1 + \sigma_0 \lambda_r \frac{PTF}{\tilde{\Delta}_{\mathbb{H}} B} \right) \right\}, \quad (7.40)$$

where we substituted the effective spread $\tilde{\Delta}_{\mathbb{H}} = \Delta_{\mathbb{H}} TF / \beta$.

The representation of the bound in terms of the effective spread allows for an interesting comparison with the simple bound (7.19) for the memoryless frequency-flat SISO setting derived in Section 7.4.1, which has the same structure as (7.40): the complicated upper bound (7.20) for an underspread WSSUS MIMO channel can be interpreted as an upper bound on the capacity per unit time of $M_R B / F$ parallel flat-fading memoryless SISO channels with effective PAPR $\tilde{\Delta}_{\mathbb{H}}$, where the r th channel has an SNR of $\sigma_0 \lambda_r PTF / B$.

The observation that a brick-shaped scattering function minimizes the upper bound $U_1(B)$ sheds some light on the common practice to use ν_0 and τ_0 in the initial design stages of a communication system. A design on the basis of ν_0 and τ_0 alone, without further recourse to $c_s(\nu, \tau)$ is implicitly targeted at a channel with brick-shaped scattering function. It is thus a min-max design: one optimizes the system for the worst-case channel.

C. Spatial Correlation and Number of Antennas

The upper bound $U_1(B)$ depends on the transmit correlation matrix A only through its maximum eigenvalue σ_0 , which plays the role of a power gain. This observation shows that rank-one statistical beamforming along any eigenvector of A corresponding to σ_0 is optimal whenever $U_1(B)$ is tight. At high P/B and correspondingly small bandwidth, $U_1(B)$ increases linearly in the number of nonzero eigenvalues of the receive correlation matrix, that is, in $\text{rank}(B)$. As the capacity in the coherent setting, which is a simple upper

bound on $C(B)$, increases at high P/B linearly only in the minimum of $\text{rank}(A)$ and $\text{rank}(B)$ (Tulino et al., 2005, Proposition 4), we conclude that $U_1(B)$ is not tight at high P/B . However, for large bandwidth and corresponding small P/B , we show in Section 7.7 that $U_1(B)$ is tight and that rank-one statistical beamforming is indeed optimal in the wideband regime.

7.5. A LOWER BOUND ON CAPACITY

Lower bounds on capacity are, in general, easier to derive than upper bounds, as the mutual information of any specific signaling scheme may serve as a lower bound. The key point is thus to choose a signaling scheme that is tractable yet good enough to furnish a tight bound. A popular choice is a JCG-distributed input vector because it is easy to handle and known to be optimal in the coherent setting. Unfortunately, a Gaussian distribution is incompatible with the peak constraint (7.13). Instead, we use input symbols of constant amplitude and rotation-symmetric phase, e.g., a phase-shift keying (PSK) constellation.

7.5.1. The Lower Bound

To state our lower bound on capacity (7.17), we require the following definitions.:

- Let $\Psi(\varphi)$ denote the $N \times N$ matrix-valued spectral density of an arbitrary component channel process* $\{\mathbf{h}[k]\}$, i.e.,

$$\Psi(\varphi) \triangleq \sum_{k=-\infty}^{\infty} \mathbb{E}[\mathbf{h}[\Delta k + k]\mathbf{h}^\dagger[k]] e^{-i2\pi\Delta k\varphi} \quad |\varphi| \leq \frac{1}{2}. \quad (7.41)$$

- Let \mathbf{s} denote an M_T -dimensional random vector whose first Q elements are i.i.d. and of *constant modulus*, i.e., they have zero mean and satisfy $|\mathbf{s}_q|^2 = PT/(NQ)$, and let the remaining $M_T - Q$ elements of \mathbf{s} be zero.

* The vector processes $\mathbf{h}[k]$ of all component channels (r, q) have the same spectral density by assumption; therefore, we drop the subscripts r and q .

- Let \mathbf{H}_w be an $M_R \times M_T$ matrix and let \mathbf{w} be an M_R -dimensional vector, both with i.i.d. JPG entries of zero mean and unit variance.

With these definitions, the coherent mutual information of the fading MIMO channel with IO relation $\mathbf{y} = \Lambda^{1/2} \mathbf{H}_w \Sigma^{1/2} \mathbf{s} + \mathbf{w}$ is given by $I(\mathbf{s}; \mathbf{y} | \mathbf{H}_w)$, a quantity we need momentarily.

Theorem 7.2. *The capacity (7.17) of the underspread WSSUS MIMO channel with spatial correlation described in Section 7.3, whose average input power is constrained as (7.12) and whose peak power follows (7.13) is lower-bounded as $C(B) \geq \max_{1 \leq Q \leq M_T} L_1(B)$, where*

$$L_1(B) \triangleq \max_{1 \leq \gamma \leq \beta} \left\{ \frac{B}{\gamma T F} I(\sqrt{\gamma} \mathbf{s}; \mathbf{y} | \mathbf{H}_w) - \frac{1}{\gamma T} \sum_{q=0}^{Q-1} \sum_{r=0}^{M_R-1} \int_{-1/2}^{1/2} \log \det \left(\mathbf{I}_N + \sigma_q \lambda_r \frac{\gamma P T F}{Q B} \Psi(\varphi) \right) d\varphi \right\}. \quad (7.42)$$

Proof. Any specific input distribution leads to a lower bound on capacity; in particular, we choose to transmit constant modulus symbols $\mathbf{x}_q[k, n] = \mathbf{s}_q[k, n]$ that are i.i.d. over time, frequency, and eigenmodes, and that satisfy $|\mathbf{s}_q[k, n]|^2 = PT/(QN)$ w.p.1 for all (k, n) and for $q = 0, 1, \dots, Q-1$. The remaining $M_T - Q$ eigenmodes are not used to transmit information. We stack the symbols $\mathbf{s}_q[k, n]$ as in (7.3) and define the $KN \times M_T KN$ matrix

$$\mathbf{S} \triangleq [\mathbf{S}_0 \quad \mathbf{S}_1 \quad \cdots \quad \mathbf{S}_{Q-1} \quad 0_{KN} \quad \cdots \quad 0_{KN}] \quad (7.43)$$

with $\mathbf{S}_q \triangleq \text{diag } \mathbf{s}_q$, and where the last $M_T - Q$ entries are all-zero matrices of dimension KN . By the chain rule for mutual information and the fact that mutual information is nonnegative,

$$\begin{aligned} I(\mathbf{S}; \mathbf{y}) &= I(\mathbf{S}, \mathbf{h}; \mathbf{y}) - I(\mathbf{h}; \mathbf{y} | \mathbf{S}) \\ &= I(\mathbf{h}; \mathbf{y}) + I(\mathbf{S}; \mathbf{y} | \mathbf{h}) - I(\mathbf{h}; \mathbf{y} | \mathbf{S}) \\ &\geq I(\mathbf{S}; \mathbf{y} | \mathbf{h}) - I(\mathbf{h}; \mathbf{y} | \mathbf{S}). \end{aligned} \quad (7.44)$$

Next, we analyze the two terms on the right-hand side separately.

Because we chose the input symbols to be i.i.d., and because all channel coefficients have the same marginal distribution,

$$I(\mathbf{S}; \mathbf{y} | \mathbf{h}) = KN I(\mathbf{s}; \mathbf{y} | \mathbf{H}_w). \quad (7.45)$$

Given the input \mathbf{S} , the corresponding output vector \mathbf{y} is JPG distributed. Therefore, the second term on the right-hand side in (7.44) can be evaluated as

$$\begin{aligned} I(\mathbf{h}; \mathbf{y} | \mathbf{S}) &= \sum_{r=0}^{M_R-1} \mathbb{E} [\log \det(\mathbf{I}_{KN} + \lambda_r \mathbf{S}(\Sigma \otimes \mathbf{R}_h) \mathbf{S}^\dagger)] \\ &\stackrel{(a)}{\leq} \sum_{r=0}^{M_R-1} \log \det(\mathbf{I}_{KN} + \lambda_r \mathbb{E}[\mathbf{S}^\dagger \mathbf{S}] (\Sigma \otimes \mathbf{R}_h)) \\ &\stackrel{(b)}{=} \sum_{q=0}^{Q-1} \sum_{r=0}^{M_R-1} \log \det \left(\mathbf{I}_{KN} + \sigma_q \lambda_r \frac{PT}{QN} \mathbf{R}_h \right). \end{aligned} \quad (7.46)$$

Here, (a) follows from the determinant identity $\det(\mathbf{I} + \mathbf{A}\mathbf{B}^\dagger) = \det(\mathbf{I} + \mathbf{B}^\dagger\mathbf{A})$ for any two square matrices \mathbf{A} and \mathbf{B} of appropriate dimension (Horn and Johnson, 1985, Theorem 1.3.20) and from Jensen's inequality; the latter is applicable because the log-determinant expression is concave in $\mathbf{S}^\dagger \mathbf{S}$ (Diggavi and Cover, 2001). Subsequently, (b) follows because the $\{\mathbf{s}_q[k, n]\}$ are i.i.d. with zero mean and constant modulus.

We combine the two terms (7.45) and (7.46), set $B = NF$, divide by KT , and take the limit $K \rightarrow \infty$ to obtain the lower bound

$$\begin{aligned} C(B) &\geq \lim_{K \rightarrow \infty} \frac{1}{KT} I(\mathbf{S}; \mathbf{y}) \\ &\geq \frac{B}{TF} I(\mathbf{s}; \mathbf{y} | \mathbf{H}_w) \\ &\quad - \lim_{K \rightarrow \infty} \frac{1}{KT} \sum_{q=0}^{Q-1} \sum_{r=0}^{M_R-1} \log \det \left(\mathbf{I}_{KN} + \sigma_q \lambda_r \frac{PTF}{QB} \mathbf{R}_h \right). \end{aligned} \quad (7.47)$$

The correlation matrix \mathbf{R}_h of the stacked channel vector \mathbf{h} is two-level Toeplitz with blocks that are $N \times N$ correlation matrices $\mathbf{R}_h[\Delta k]$,

as shown in (7.7) and (7.6), respectively. Therefore, we can evaluate the limit on the right-hand side in (7.47) and express it in terms of an integral over the matrix-valued spectral density $\Psi(\varphi)$ of the multivariate channel process $\mathbf{h}[k]$ by means of an extension of Szegő's Theorem to the case of block-Toeplitz matrices generated by matrix-valued spectra (Miranda and Tilli, 2000, Theorem 3.4):

$$\begin{aligned} \lim_{K \rightarrow \infty} \frac{1}{KT} \sum_{q=0}^{Q-1} \sum_{r=0}^{M_R-1} \log \det \left(\mathbf{I}_{KN} + \sigma_q \lambda_r \frac{PTF}{QB} \mathbf{R}_h \right) \\ = \frac{1}{T} \sum_{q=0}^{Q-1} \sum_{r=0}^{M_R-1} \int_{-1/2}^{1/2} \log \det \left(\mathbf{I}_N + \sigma_q \lambda_r \frac{PTF}{QB} \Psi(\varphi) \right) d\varphi. \end{aligned} \quad (7.48)$$

The lower bound that results by substituting (7.48) in (7.47) can be improved upon by time-sharing: Let $1 \leq \gamma \leq \beta$. We transmit $\sqrt{\gamma} \mathbf{s}$ during a fraction $1/\gamma$ of the transmission time and let the transmitter be silent otherwise, so that the constraint on the average power is satisfied. \square

7.5.2. Some Simplifications and Approximations

Evaluation of the lower bound (7.42) is complicated by two facts.

1. The mutual information $I(\sqrt{\gamma} \mathbf{s}; \mathbf{y} | \mathbf{H}_w)$ in the first term on the right-hand side of (7.42) needs to be evaluated for constant modulus input vectors.
2. In general, the eigenvalues of $\Psi(\varphi)$ in the second term on the right-hand side of (7.42) cannot be derived in closed form.

While the coherent mutual information $I(\sqrt{\gamma} \mathbf{s}; \mathbf{y} | \mathbf{H}_w)$ can be evaluated numerically for constant modulus inputs at reasonable expense, computing the eigenvalues of $\Psi(\varphi)$ is challenging for channels of very wide bandwidth because the matrix might be very large. Therefore, we derive a lower bound on $L_1(B)$ that can be efficiently evaluated. Furthermore, we present a closed-form approximation of (7.42) that

can be used for back-of-the-envelope calculations. Both the bound and the approximation are tight for large bandwidth.

We first analyze the second term of $L_1(B)$, which can be again interpreted as a penalty term that quantifies the lack of channel knowledge.

Lemma 7.3. *Let*

$$d_j \triangleq \Re \left\{ \frac{2}{N} \sum_{\Delta n=0}^{N-1} (N - \Delta n) c_l[0, \Delta n] e^{-i2\pi \frac{j\Delta n}{N}} \right\} - 1. \quad (7.49)$$

Then, the (q, r) th addend of the penalty term in (7.48) can be bounded as

$$\begin{aligned} & \nu_0 \sum_{j=0}^{N-1} \log \left(1 + \sigma_q \lambda_r \frac{PF}{\nu_0 QB} d_j \right) \\ & \geq \frac{1}{T} \int_{-1/2}^{1/2} \log \det \left(\mathbf{I}_N + \sigma_q \lambda_r \frac{PTF}{QB} \Psi(\varphi) \right) d\varphi \\ & \geq B \iint_{\nu \tau} \log \left(1 + \sigma_q \lambda_r \frac{P}{QB} c_s(\nu, \tau) \right) d\tau d\nu. \end{aligned} \quad (7.50)$$

Furthermore, the following asymptotic results hold:

- *The difference between the (q, r) th addend of the penalty term in (7.48) and its lower bound in (7.50) vanishes as B grows large.*
- *For scattering functions that are flat in the Doppler domain, i.e., that satisfy*

$$c_s(\nu, \tau) = \frac{1}{\nu_0} p_s(\tau) \quad (7.51)$$

*for $(\nu, \tau) \in [-\nu_0/2, \nu_0/2] \times [-\tau_0/2, \tau_0/2]$, the difference between the upper bound and the lower bound in (7.50) vanishes as B grows large.**

* The normalization of the scattering function in (2.39) implies that the PDP is normalized as well, i.e., $\int_{\tau} p_s(\tau) d\tau = 1$.

Proof. We present the proof in Section 8.5. The key step to upper-bound the penalty term tightly for large bandwidth is to replace $\Psi(\varphi)$ with a circulant matrix that is, in N , asymptotically equivalent to $\Psi(\varphi)$ (Gray, 2005), while the lower bound follows from Lemma 8.5 in Section 8.3. \square

A. Further Lower Bound for Efficient Evaluation

If we replace the penalty term in (7.42) with the upper bound on each of its addends in (7.50), we obtain the lower bound on capacity $L_2(B) \leq C(B)$, where

$$L_2(B) \triangleq \max_{1 \leq \gamma \leq \beta} \left\{ \frac{B}{\gamma TF} I(\sqrt{\gamma} \mathbf{s}; \mathbf{y} | \mathbf{H}_w) - \frac{\nu_0}{\gamma} \sum_{q=0}^{Q-1} \sum_{r=0}^{M_R-1} \sum_{j=0}^{N-1} \log \left(1 + \sigma_q \lambda_r \frac{\gamma PF}{\nu_0 QB} d_j \right) \right\}. \quad (7.52)$$

This lower bound is less tight than $L_1(B)$; however, it can be evaluated numerically in a much more efficient way than $L_1(B)$ because the coefficient sequence $\{d_j\}_{j=0}^{N-1}$ can be obtained from the values $\{(N - \Delta n)c_l[0, \Delta n]\}_{\Delta n=0}^{N-1}$ through the DFT.

B. Closed-Form Approximation of the Lower Bound

For large enough bandwidth, the difference between each addend of the penalty term (7.48) and the corresponding lower bound (7.50) is so small that we can replace the individual addends in the second term of $L_1(B)$ by their corresponding lower bounds to obtain a good approximation of $L_1(B)$:

$$L_1(B) \approx L_a(B) \triangleq \max_{1 \leq \gamma \leq \beta} \left\{ \frac{B}{\gamma TF} I(\sqrt{\gamma} \mathbf{s}; \mathbf{y} | \mathbf{H}_w) - \frac{B}{\gamma} \sum_{q=0}^{Q-1} \sum_{r=0}^{M_R-1} \iint_{\nu \tau} \log \left(1 + \sigma_q \lambda_r \frac{\gamma P}{QB} c_s(\nu, \tau) \right) d\tau d\nu \right\}. \quad (7.53)$$

Furthermore, we can replace the first term of $L_a(B)$ in (7.53) by its Taylor series up to first order, as given by Prelov and Verdú (2004, Theorem 3). This series expansion requires the evaluation of several expected trace terms that involve the channel matrix; Lemmas 3 and 4 by Lozano et al. (2003) provide these expectations. The resulting coarser approximation of $L_1(B)$ is

$$L_{aa}(B) \triangleq \max_{1 \leq \gamma \leq \beta} \left\{ \frac{M_R P}{Q} \sum_{q=0}^{Q-1} \sigma_q - \gamma P^2 \frac{TF}{B} \frac{\left(\sum_{q=0}^{Q-1} \sigma_q \right)^2 \sum_{r=0}^{M_R-1} \lambda_r^2 + M_R^2 \sum_{q=0}^{Q-1} \sigma_q^2}{2Q^2} - \frac{B}{\gamma} \sum_{q=0}^{Q-1} \sum_{r=0}^{M_R-1} \iint_{\nu \tau} \log \left(1 + \sigma_q \lambda_r \frac{\gamma P}{QB} c_s(\nu, \tau) \right) d\tau d\nu \right\}. \quad (7.54)$$

Both $L_a(B)$ and $L_{aa}(B)$ are tight for $B \rightarrow \infty$.

7.6. NUMERICAL EXAMPLES

7.6.1. Setting

The relation between the WSSUS channel model and our discrete approximation (3.16) allows us to plot upper and lower bounds for realistic channel and system parameters. In particular, we use the following parameter set for the numerical evaluation of our bounds:

- A brick-shaped scattering function with maximum delay $\tau_0 = 10 \mu\text{s}$, maximum positive or negative Doppler shift $\nu_0/2 = 50 \text{ Hz}$, and corresponding spread $\Delta_{\mathbb{H}} = \tau_0 \nu_0 = 10^{-3}$.
- Grid parameters* $T = 353 \mu\text{s}$ and $F = 3.53 \text{ kHz}$, so that $TF \approx 1.25$ and $T/F = \tau_0/\nu_0$, as suggested by the matching rule for the

* Note that these parameters deviate significantly from corresponding parameters of commercially deployed OFDM systems: e.g., $T = 4 \mu\text{s}$ and $F = 325 \text{ kHz}$ in the IEEE 802.11a/g/n standard, which means that the standard is optimized for channels with a much smaller ratio τ_0/ν_0 than the one considered here, i.e.,

TF discretization grid (3.47).

- Receive power normalized with respect to the noise spectral density and the pathloss of

$$\frac{P}{1 \text{ W/Hz}} = 2.42 \cdot 10^8 \text{ s}^{-1}.$$

These parameters are representative for several different types of systems under the assumption of a thermal noise level at the receiver of -174 dBm/Hz. For example:

1. An IEEE 802.11a system with transmitted power of 200 mW, pathloss of 108 dB, and receiver noise figure (Razavi, 1998) of 5 dB; the pathloss is rather pessimistic for typical indoor link distances and includes the attenuation of the signal by, e.g., a concrete wall.
2. A UWB system with transmitted power of 0.5 mW, pathloss of 67 dB, and receiver noise figure of 20 dB. The latter also accounts for ADC back-off.

The large maximum delay and maximum Doppler shift are certainly not representative for typical home environments, but may be encountered, for example, in a factory with moving machinery.

7.6.2. Evaluation Methodology

The upper bound $U_1(B)$ in Theorem 7.1 can be evaluated in closed form for a brick-shaped scattering function, but the lower bound $L_1(B)$ in Theorem 7.2 is difficult to evaluate numerically. Therefore, we turn to the simplifications discussed in Section 7.5.2. The lower bound $L_2(B)$ in (7.52) can be evaluated much more efficient than $L_1(B)$, yet the memory requirement to compute the d_j for bandwidths in excess of several hundred megahertz is still prohibitive for the parameters

for channels that are much more time variant. This difference results because the *effective* channel of real-world communication systems also accounts for effects like carrier frequency offset and phase noise, which significantly increase channel variation over time.

specified above. Instead, we use the approximation $L_a(B)$, as its second term can be evaluated in closed form for a brick-shaped scattering function. However, $L_a(B)$ is an upper bound on the lower bound $L_1(B)$ instead of a true lower bound. Fortunately, the match between $L_a(B)$ and the true lower bound $L_2(B)$ is extremely good already for the range of bandwidth for which we can actually compute $L_2(B)$ —so good that the resulting curves are virtually indistinguishable. Thus, *de facto* the lower bound $L_1(B)$ in (7.42), which lies between $L_2(B)$ and $L_a(B)$, is fully characterized. Therefore, we label the corresponding plots as both L_1 and L_a .

To compute the first term of $L_a(B)$, i.e., the mutual information $I(\mathbf{s}; \mathbf{y} | \mathbf{H}_w)$, we use Monte-Carlo integration and input symbols with quaternary phase-shift keying (QPSK) modulation. The loss because of the finite-dimensional QPSK constellation compared with continuous-valued input symbols is negligible at low SNR. In addition to the two mentioned bounds, we also compute the second-order approximation $L_{aa}(B)$ in (7.54), which turns out to be suitable for very large bandwidth.

7.6.3. Numerical Results

For a 3×3 MIMO system, we show in this section plots of the upper bound $U_1(B)$ of Theorem 7.1, and—for Q between 1 and 3—plots of the approximation $L_a(B)$ in (7.53) that is indistinguishable from the true lower bound $L_1(B)$ of Theorem 7.2. In addition, we also show plots of the corresponding bounds of a SISO system that operates over a channel with spread $\Delta_{\text{H}} = 10^{-5}$. In all figures, we also plot the large-bandwidth approximation $L_{aa}(B)$ of the lower bound in (7.54).

A. Spatially Uncorrelated Channel

Figure 7.1 shows the bounds and approximations just mentioned for the spatially uncorrelated case $\Sigma = \Lambda = \mathbf{I}_3$. For comparison, we also plot a standard capacity upper bound $U_c(B)$, obtained for the coherent setting and with input subject to an average-power

7 CAPACITY BOUNDS

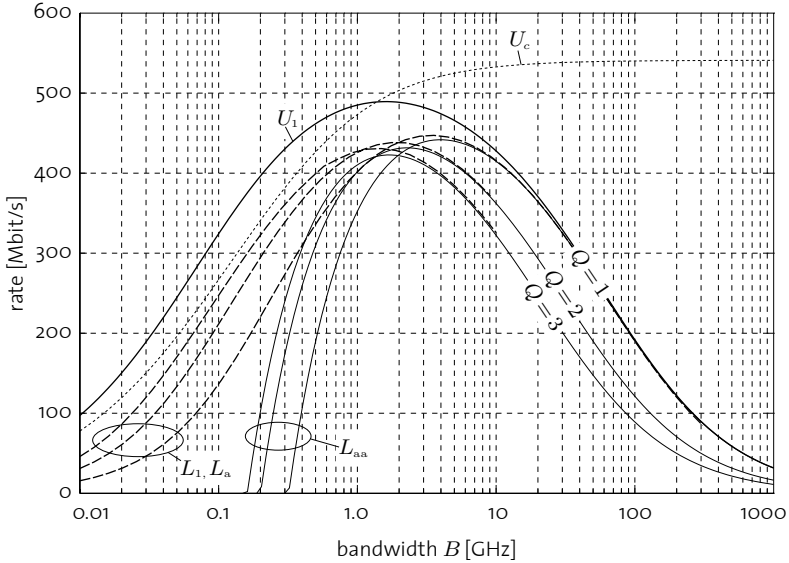


Fig. 7.1: Upper and lower bounds on the capacity of a spatially uncorrelated underspread WSSUS MIMO channel with $\Sigma = \Lambda = \mathbf{I}_3$, $M_T = M_R = 3$, $\beta = 1$, and $\Delta_{\mathbb{H}} = 10^{-3}$.

constraint only. We can observe that $U_c(B)$ is tighter than $U_1(B)$ for small bandwidth B ; this holds true in general as for small B the penalty term in (7.20) can be neglected and $U_1(B)$ in the spatially uncorrelated case reduces to

$$U_1(B) \approx \frac{M_R B}{TF} \log\left(1 + \frac{PTF}{B}\right) \quad (7.55)$$

which is the Jensen upper bound on the capacity upper bound $U_c(B)$ in the coherent setting. For small and medium bandwidth, the lower bound $L_1(B)$ increases with Q and comes surprisingly close to the coherent capacity upper bound $U_c(B)$ for $Q = 3$.

As can be expected in the light of results, e.g., by Médard and Gallager (2002) and Subramanian and Hajek (2002), when bandwidth

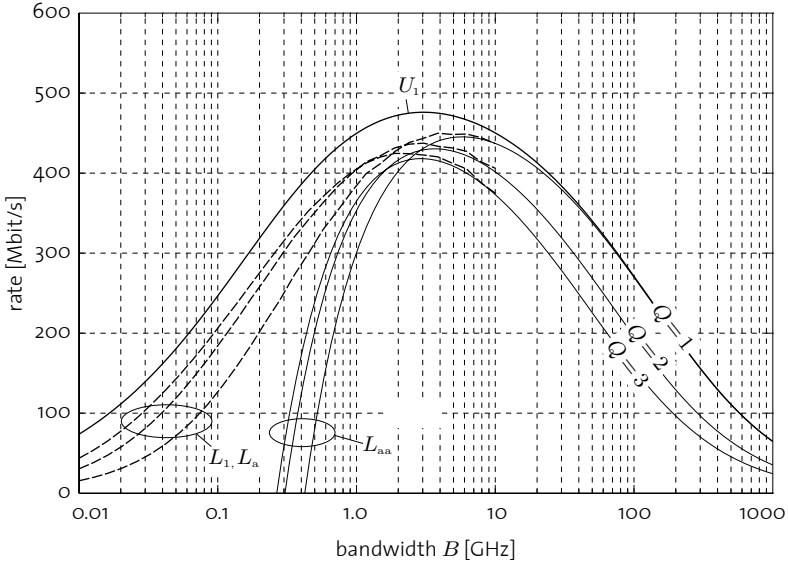


Fig. 7.2: Upper and lower bounds on the capacity of an underspread wssus MIMO channel that is spatially uncorrelated at the transmitter, $\Sigma = \mathbf{I}_3$, but correlated at the receiver with $\Lambda = \text{diag}[2.6 \ 0.3 \ 0.1]^T$; $M_T = M_R = 3$, $\beta = 1$, and $\Delta_{\text{H}} = 10^{-3}$.

increases above a certain *critical bandwidth*, both $U_1(B)$ and $L_1(B)$ start to decrease; in this regime, the rate gain resulting from the additional DOFs is offset by the resources required to resolve channel uncertainty. The same argument seems to hold in the wideband regime for the DOFs by multiple transmit and receive antennas: $U_1(B)$ appears to match $L_1(B)$ for $Q = 1$; hence, using a single transmit antenna seems optimal in the wideband regime.

B. Impact of Receive Correlation

Figure 7.2 shows the same bounds as before, but evaluated with spatial correlation $\Lambda = \text{diag}[2.6 \ 0.3 \ 0.1]^T$ at the receiver and a spatially

uncorrelated channel at the transmitter, i.e., $\Sigma = \mathbf{I}_3$. The curves in Figure 7.2 are very similar to the ones shown in Figure 7.1 for the spatially uncorrelated case, yet they are shifted toward higher bandwidth while the maximum rate is lower. Hence, at least for the example at hand, receive correlation decreases capacity at small bandwidth but it is beneficial at large bandwidth.

C. Impact of Transmit Correlation

We evaluate the same bounds once more, but this time for spatial correlation $\Sigma = \text{diag}[1.7 \ 1.0 \ 0.3]^T$ at the transmitter and a spatially uncorrelated channel at the receiver, i.e., $\Lambda = \mathbf{I}_3$. The corresponding curves are shown in Figure 7.3. Here, transmit correlation increases the capacity at large bandwidth, while its impact at small bandwidth is more difficult to judge because the distance between upper and lower bound increases compared to the spatially uncorrelated case.

D. Impact of the Spread

A spread of $\Delta_{\text{H}} = 10^{-3}$ is rather large for typical indoor channels, where the maximum delay τ_0 is small and mobility is low. Therefore, Figure 7.4 shows the same capacity bounds as before but for a channel of spread $\Delta_{\text{H}} = 10^{-5}$, where we again assumed a brick-shaped scattering function, but this time with maximum delay $\tau_0 = 1 \mu\text{s}$ and maximum positive or negative Doppler shift $\nu_0/2 = 5 \text{ Hz}$, as well as a received-power to noise-density ratio of $P/(1 \text{ W}/Hz) = 2.42 \cdot 10^8 \text{ s}^{-1}$. For simplicity, we plot the bounds for a SISO system only. We can observe that the maximum of both upper and lower bounds broaden and extend to higher bandwidths. Specifically, the wideband approximation $L_{\text{aa}}(B)$ appears useful down to modest bandwidths.

7.6.4. Observations

All four figures show that especially for large bandwidth the approximation $L_{\text{aa}}(B)$ of $L_1(B)$ is quite accurate. An observation of significant practical importance is that the bounds $U_1(B)$ and $L_1(B)$

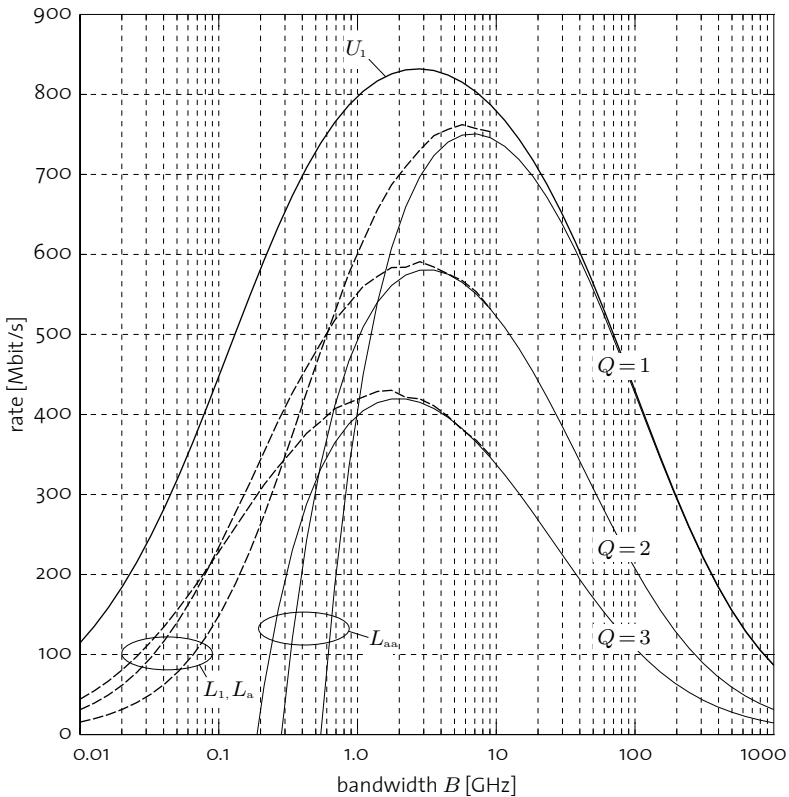


Fig. 7.3: Upper and lower bounds on the capacity of an underspread WSSUS MIMO channel that is correlated at the transmitter with $\Sigma = \text{diag}[1.7 \ 1.0 \ 0.3]^T$ and spatially uncorrelated at the receiver, $\Lambda = \mathbf{I}_3$; $M_T = M_R = 3$, $\beta = 1$, and $\Delta_{\mathbb{H}} = 10^{-3}$.

are quite flat over a wide range of bandwidth around their maxima, and that these maxima broaden with decreasing spread. Further numerical results not shown here point at an increase in the gap between upper and lower bounds as β increases.

The single most important caveat with all four figures is that much of the bandwidth shown certainly exceeds the range of bandwidths for

7 CAPACITY BOUNDS

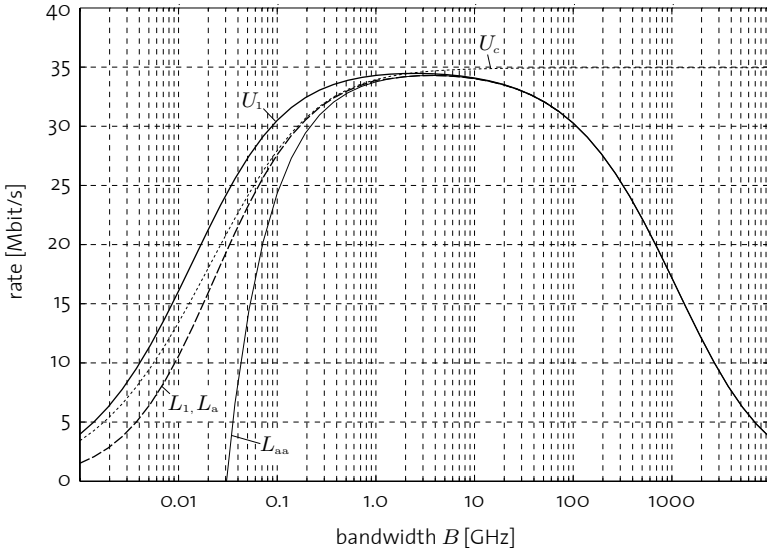


Fig. 7.4: Upper and lower bounds on the capacity of an underspread WSSUS SISO channel, $\beta = 1$, and $\Delta_{\mathbb{H}} = 10^{-5}$.

which our channel model is a reasonable approximation of physical channels. Even the critical bandwidth, somewhere between 1 GHz and 10 GHz in the plots, is already a stretch for the model. This does not mean that the bounds of Theorem 7.1 and Theorem 7.2 are useless, though; on the contrary, we can draw the following important conclusions:

- Even for rather pessimistic assumptions about the physical channel does capacity increase with an increase in bandwidth for virtually all practically relevant frequency ranges over which our model makes sense. The effect of vanishing wideband capacity is not practically relevant in the scenarios we analyzed, and probably not relevant either for smaller values of the spread $\Delta_{\mathbb{H}}$.
- Our bounds allow to judge how much bandwidth can be benefi-

cially used for a given set of system parameters as long as the underlying channel model is sensible. The bounds themselves are reasonably accurate for the parameters considered here.

Finally, we can see from all four figures that virtually all wireless systems currently deployed operate well below the critical bandwidth, at least for the settings considered.

7.7. ASYMPTOTIC WIDEBAND ANALYSIS

The plots in Section 7.6 suggest that in the wideband regime (i) using a single transmit antenna is optimal when the channel is spatially uncorrelated at the transmitter side, (ii) it is optimal to signal on the maximum transmit eigenmode if transmit correlation is present, and (iii) both transmit and receive correlation are beneficial. To substantiate these observations, we compute the first-order Taylor series expansion of $C(B)$ around $1/B = 0$.

7.7.1. Wideband Asymptotes

We would like to find a closed-form expression for the first-order Taylor expansion

$$C(B) = \frac{\phi}{B} + o\left(\frac{1}{B}\right). \quad (7.56)$$

To describe this expansion in a compact way, we need the following definitions:

$$\kappa_{\mathbb{H}} \triangleq \iint_{\nu \tau} c_{\mathbf{s}}^2(\nu, \tau) d\tau d\nu, \quad (7.57)$$

$$\vartheta \triangleq \sum_{r=0}^{M_{\mathbb{R}}-1} \lambda_r^2. \quad (7.58)$$

Theorem 7.4. *The capacity $C(B)$ of an underspread WSSUS channel that satisfies the constraint on the average power (7.12) and the peak constraint (7.13) has the following first-order coefficient ϕ in the Taylor series expansion (7.56) around the point $1/B = 0$:*

$$\begin{aligned} \phi &= \lim_{B \rightarrow \infty} B U_1(B) \\ &= \begin{cases} \frac{\vartheta(\sigma_0 P)^2}{2} (\beta \kappa_{\mathbb{H}} - T F), & \text{if } \beta > \frac{2TF}{\kappa_{\mathbb{H}}} \\ \frac{\vartheta(\sigma_0 \kappa_{\mathbb{H}} \beta P)^2}{8TF}, & \text{if } \beta \leq \frac{2TF}{\kappa_{\mathbb{H}}}. \end{cases} \end{aligned} \quad (7.59)$$

Proof. A seemingly direct approach to prove Theorem 7.4 is to expand both the upper bound $U_1(B)$ in Theorem 7.1 and the lower bound $L_1(B)$ in Theorem 7.2 into a Taylor series around the point $1/B = 0$ and show that both expansions match up to first order. Unfortunately they don't. The culprit is the lower bound, where the signaling scheme we use to derive it in Section 7.5.1 is not optimal at low SNR—and correspondingly large bandwidth—because information is not encoded in the signal amplitude. Thus, we derive a lower bound on the infinite bandwidth capacity in Section 8.6.2 that has the correct asymptotic first-order behavior.

There are two key elements in the derivation of this new lower bound: an extension of the block-constant signaling scheme used by Sethuraman et al. (2009) to prove asymptotic capacity results for flat-fading time-selective channels, and transmission only over the strongest transmit eigenmode, as we observed in Section 7.6 that signaling over a single eigenmode seems to be optimal for large bandwidth. Such a signaling scheme is often referred to as *rank-one beamforming*. Hence, we transmit on the strongest eigenmode a signal with uniformly distributed phase whose magnitude is toggled on and off at random with a prescribed probability, so that information is encoded jointly in the amplitude and in the phase. In comparison, the signaling scheme used to obtain $L_1(B)$ transmits a signal of constant amplitude in all time-frequency slots. The so constructed asymptotic lower bound matches the first-order Taylor expansion of $U_1(B)$. We

present the derivation of both the Taylor series expansion of $U_1(B)$ and of the asymptotically matching lower bound in Section 8.6 as they are rather technical. \square

Similar to the capacity behavior of a discrete-time stationary flat-fading channel for vanishing SNR (Wang, 2006), the first-order Taylor expansion coefficient ϕ in (7.59) can take on two different forms as a function of the channel parameters. However, the link in (3.38) between the discretized channel and the WSSUS channel \mathbb{H} allows us to conclude that $\beta > 2TF/\kappa_{\mathbb{H}}$ and thus $\phi = P^2(\beta\kappa_{\mathbb{H}} - TF)/2$ for virtually all wireless channels of practical interest: In fact, $\kappa_{\mathbb{H}} \geq \Delta_{\mathbb{H}}^{-1}$ by Jensen's inequality, so that $2TF\Delta_{\mathbb{H}} \geq 2TF/\kappa_{\mathbb{H}}$; consequently, a sufficient condition for $\beta > 2TF/\kappa_{\mathbb{H}}$ is $\beta > 2TF\Delta_{\mathbb{H}}$. But for typical values of TF , e.g., $TF \approx 1.25$, and virtually all sensible values of $\Delta_{\mathbb{H}}$, say, $\Delta_{\mathbb{H}} < 10^{-2}$, this latter condition is satisfied for every admissible β .

7.7.2. Impact of Spatial Correlation

Rank-one statistical beamforming along any eigenvector of \mathbf{A} associated with σ_0 is optimal to attain the wideband asymptotes of Theorem 7.4. For channels that are spatially uncorrelated at the transmitter, this result implies that using only one transmit antenna is optimal, as previously shown by Sethuraman et al. (2009) for the frequency-flat time-selective case.

To further assess the impact of correlation on capacity, we follow Chuah et al. (2002), Jafar and Goldsmith (2005), Jorswieck and Boche (2006) and define a partial ordering of correlation matrices through *majorization* (Marshall and Olkin, 1979). We say that a correlation matrix \mathbf{K} entails more correlation than a correlation matrix \mathbf{C} if the vector of eigenvalues $\lambda(\mathbf{K})$ majorizes $\lambda(\mathbf{C})$. A brief review of majorization theory is provided in Section 8.2 for convenience.

In the *coherent setting*, capacity is Schur concave in $\lambda(\mathbf{B}) = \lambda(\mathbf{A}) = [\lambda_0 \lambda_1 \cdots \lambda_{M_{\mathbb{R}}-1}]$, the eigenvalue vector of the receive correlation matrix while—for sufficiently large bandwidth—it is Schur convex in $\lambda(\mathbf{A}) = \lambda(\mathbf{\Sigma}) = [\sigma_0 \sigma_1 \cdots \sigma_{M_{\mathbb{T}}-1}]$, the eigenvalue vector of the

transmit correlation matrix (Tulino et al., 2005; Jorswieck and Boche, 2006). Hence, in the coherent setting receive correlation is detrimental at any bandwidth while transmit correlation is beneficial at large bandwidth. The intuition is that transmit correlation allows to focus the transmit power into the maximum transmit eigenmode, and the corresponding power gain offsets the reduction in effective transmit signal space dimensions in the power-limited regime, i.e., at large bandwidth. On the other hand, receive correlation is detrimental at any bandwidth because it reduces the effective dimensionality of the receive signal space without any power gain, as shown by Lozano et al. (2006).

On the basis of Theorem 7.4, we conclude that the picture is fundamentally different in the *noncoherent setting*. The coefficient ϕ in (7.59) is a Schur-convex function in both the eigenvalue vector $\lambda(\mathbf{B})$ of the transmit correlation matrix and the eigenvalue vector $\lambda(\mathbf{B})$ of the receive correlation matrix because σ_0 and ϑ are continuous convex functions of the corresponding eigenvalue vectors (Marshall and Proschan, 1965). Hence, both receive and transmit correlation are beneficial for sufficiently large bandwidth. This observation agrees with the results for memoryless and block-fading channels reported by Jafar and Goldsmith (2005), Zhang and Laneman (2007a), and Srinivasan and Varanasi (2007). In the wideband regime, while transmit correlation is beneficial in both the coherent and the noncoherent setting because it allows for power focusing,* receive correlation is beneficial rather than detrimental in the noncoherent setting for the following reason: for fixed M_T and M_R , the rate gain obtained from additional bandwidth is offset in the wideband regime by the corresponding increase in channel uncertainty (see Figures 7.1, 7.2, and 7.3); yet, for fixed but large bandwidth, channel uncertainty decreases in the presence of receive correlation so that capacity increases.

* Differently from the coherent setting treated by Tulino et al. (2005, Proposition 3) though, the multiplicity of the maximum eigenvalue does not need to be taken into account in the design of capacity-achieving signaling schemes at low SNR.

CHAPTER 8

Technical Results

IN THE present chapter, we compile several technical results that are needed for proofs and derivations in Chapter 7. Some of these results, like the determinant inequality of Lemma 8.1, the majorization relation of Lemma 8.3, or the bound in Lemma 8.5 on the basis of the relation between mutual information and the MMSE might be of independent interest, while other results are probably not directly applicable elsewhere.

8.1. A DETERMINANT INEQUALITY

Lemma 8.1. *Let A and B be two $N \times N$ nonnegative definite Hermitian matrices. Then,*

$$\det(\mathbf{I}_N + A \odot B) \geq \det(\mathbf{I}_N + (\mathbf{I}_N \odot A)B).$$

Proof. Assume for now that A does not have zeros on its main diagonal and define $\tilde{A} \triangleq (\mathbf{I}_N \odot A)^{-1}$. Then,

$$\begin{aligned} \det(\mathbf{I}_N + A \odot B) &= \det(A \odot (\tilde{A} + B)) \\ &\stackrel{(a)}{\geq} \det(\mathbf{I}_N \odot A) \det(\tilde{A} + B) \\ &= \det((\mathbf{I}_N \odot A)\tilde{A} + (\mathbf{I}_N \odot A)B) \\ &= \det(\mathbf{I}_N + (\mathbf{I}_N \odot A)B) \end{aligned} \tag{8.1}$$

where (a) is a direct consequence of Oppenheim's inequality (Horn and Johnson, 1985, Theorem 7.8.6). To conclude the proof, we remove the restriction that A has only nonzero diagonal entries. Because A is nonnegative definite, its i th row and its i th column are zero if $[A]_{ii} = 0$ (Horn and Johnson, 1985, Section 7.1), so that, by the definition of the Hadamard product, the i th row and the i th column of $A \odot B$ are zero as well. Let \mathcal{I} be the set that contains all indices i for which $[A]_{ii} = 0$, assume without loss of generality that there are L such indices, and let $A_{\mathcal{I}}$ and $B_{\mathcal{I}}$ denote the submatrices of A and B , respectively, with all rows and columns corresponding to \mathcal{I} removed. An expansion by minors of $\det(I_N + A \odot B)$ now shows that

$$\det(I_N + A \odot B) = \det(I_L + A_{\mathcal{I}} \odot B_{\mathcal{I}}). \quad (8.2)$$

Hence, it suffices to apply the inequality (8.1) to the right-hand side of (8.2). \square

8.2. MAJORIZATION

Majorization theory is a general tool to derive inequalities between vector- and matrix-valued quantities (Marshall and Olkin, 1979). We need it to compare the impact of spatial correlation on the wideband capacity of MIMO channels in Section 7.7, and we also use majorization theory to provide an alternative proof of Lemma 8.1.

8.2.1. Basic Definitions and Results

Definition 8.1 (Marshall and Olkin 1979, Definition 1.A.1). Let \mathbf{a} and \mathbf{b} be two real-valued vectors of length N , and denote by $\boldsymbol{\alpha} \triangleq [\alpha_0 \ \alpha_1 \ \cdots \ \alpha_{N-1}]^T$ and $\boldsymbol{\beta} \triangleq [\beta_0 \ \beta_1 \ \cdots \ \beta_{N-1}]^T$ two permutations of \mathbf{a} and \mathbf{b} , respectively, so that $\alpha_0 \geq \alpha_1 \geq \cdots \geq \alpha_{N-1}$ and $\beta_0 \geq \beta_1 \geq \cdots \geq \beta_{N-1}$, i.e., $\boldsymbol{\alpha}$ contains the elements of \mathbf{a} and $\boldsymbol{\beta}$ contains the elements of \mathbf{b} in nonincreasing order. Then, the vector \mathbf{a} is said

to *majorize* the vector \mathbf{b} , written as $\mathbf{a} \succ \mathbf{b}$, if

$$\sum_{j=0}^{k-1} \alpha_j \geq \sum_{j=0}^{k-1} \beta_j, \quad k = 1, 2, \dots, N \quad (8.3)$$

and

$$\sum_{j=0}^{N-1} \alpha_j = \sum_{j=0}^{N-1} \beta_j. \quad (8.4)$$

Majorization provides a partial ordering on real Euclidean space \mathcal{R}^N . Intuitively, $\mathbf{a} \succ \mathbf{b}$ means that \mathbf{a} is more spread out than \mathbf{b} .

Definition 8.2 (Marshall and Olkin 1979, Definition 3.A.1). A real-valued function f defined on $\mathcal{A} \subset \mathcal{R}^N$ is said to be *Schur convex* on \mathcal{A} if for $\mathbf{a}, \mathbf{b} \in \mathcal{A}$,

$$\mathbf{a} \succ \mathbf{b} \quad \implies \quad f(\mathbf{a}) \geq f(\mathbf{b}). \quad (8.5)$$

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

$$\mathbf{a} \succ \mathbf{b} \quad \implies \quad f(\mathbf{a}) \leq f(\mathbf{b}). \quad (8.6)$$

Lemma 8.2 to follow, due to Hardy, Littlewood, and Pólya, is of particular importance to verify majorization relations. The lemma relates majorization to a linear relation between two vectors by means of a doubly stochastic matrix.

Definition 8.3 (Marshall and Olkin 1979, Definition 2.A.1). A square matrix S of dimension $N \times N$ is called *doubly stochastic* if all its entries are nonnegative, i.e.,

$$s_{i,j} \triangleq [S]_{i,j} \geq 0 \quad \text{for } i, j = 0, 1, \dots, N-1, \quad (8.7)$$

and if its row and column sums equal 1:

$$\sum_{i=0}^{N-1} s_{i,j} = 1 \quad \text{for } j = 0, 1, \dots, N-1, \quad (8.8a)$$

$$\sum_{j=0}^{N-1} s_{i,j} = 1 \quad \text{for } i = 0, 1, \dots, N-1. \quad (8.8b)$$

Lemma 8.2 (Marshall and Olkin 1979, Theorem 2.B.2). *Take \mathbf{a} and \mathbf{b} as in Definition 8.1. A necessary and sufficient condition for $\mathbf{a} \succ \mathbf{b}$ is that there exists a doubly stochastic matrix \mathbf{S} such that $\mathbf{b} = \mathbf{S}\mathbf{a}$.*

8.2.2. Majorization and Hadamard Products

We can use majorization theory to provide an alternative proof of Lemma 8.1. Although more involved than the proof given in Section 8.1, the next lemma, on which this alternative proof is based may be interesting in its own right.

Lemma 8.3. *Let \mathbf{A} and \mathbf{B} be two $N \times N$ Hermitian matrices, and let \mathbf{d} be the vector that contains the square-root of the diagonal elements of \mathbf{A} , i.e., $[d]_i = \sqrt{[\mathbf{A}]_{i,i}}$. Then, the eigenvalue vector $\lambda(\mathbf{d}\mathbf{d}^\dagger \odot \mathbf{B})$ majorizes the vector of eigenvalues $\lambda(\mathbf{A} \odot \mathbf{B})$:*

$$\lambda(\mathbf{d}\mathbf{d}^\dagger \odot \mathbf{B}) \succ \lambda(\mathbf{A} \odot \mathbf{B}). \quad (8.9)$$

To simplify presentation of the corresponding proof, we first state a lemma by Bapat and Sunder (1985).

Lemma 8.4 (Bapat and Sunder 1985, Lemma 1). *Let \mathbf{P} and \mathbf{Q} be complex-valued $N \times N$ matrices and let $\mathbf{a} \in \mathbb{C}^N$. Then, the diagonal of the matrix $\mathbf{P} \text{diag}(\mathbf{a}) \mathbf{Q}$, arranged as a column vector, equals the vector $(\mathbf{P} \odot \mathbf{Q}^T)\mathbf{a}$.*

Proof of Lemma 8.3. The proof of Lemma 8.3 follows along the lines of Bapat and Sunder (1985). To simplify notation, let $\mathbf{X} \triangleq \mathbf{A} \odot \mathbf{B}$ and $\mathbf{Y} \triangleq \mathbf{d}\mathbf{d}^\dagger \odot \mathbf{B}$. By Lemma 8.2, the majorization relation in Lemma 8.3 holds if we can find a doubly stochastic matrix \mathbf{S} such that $\lambda(\mathbf{X}) = \mathbf{S}\lambda(\mathbf{Y})$.

Assume for now that \mathbf{A} does not have zero entries on its main diagonal. Any Hermitian matrix \mathbf{A} can be written as a sum of outer products, $\mathbf{A} = \sum_{i=0}^{N-1} \mathbf{a}_i \mathbf{a}_i^\dagger$ (Horn and Johnson, 1985). Let $\mathbf{A}_i = \text{diag } \mathbf{a}_i$ and define for all i the normalized diagonal matrices $\mathbf{C}_i \triangleq \mathbf{A}_i \mathbf{D}^{-1}$,

where $D = \text{diag } d$. Then,

$$\begin{aligned} \sum_{i=0}^{N-1} C_i C_i^\dagger &= \sum_{i=0}^{N-1} A_i D^{-1} D^{-1} A_i^\dagger \\ &= D^{-1} D^{-1} \sum_{i=0}^{N-1} A_i A_i^\dagger \stackrel{(a)}{=} D D D^{-1} D^{-1} = I, \end{aligned} \quad (8.10)$$

where (a) follows from the definition of D . Define the spectral decompositions

$$X = U_X \Lambda_X U_X^\dagger, \quad (8.11a)$$

$$Y = U_Y \Lambda_Y U_Y^\dagger. \quad (8.11b)$$

where $\Lambda_X \triangleq \text{diag } \lambda(X)$ and $\Lambda_Y \triangleq \text{diag } \lambda(Y)$, and where the columns of U_X and U_Y are the corresponding eigenvectors. Then,

$$\begin{aligned} \Lambda_X &= U_X^\dagger \left(\sum_{i=0}^{N-1} a_i a_i^\dagger \odot B \right) U_X = U_X^\dagger \left(\sum_{i=0}^{N-1} A_i B A_i^\dagger \right) U_X \\ &= \sum_{i=0}^{N-1} U_X^\dagger C_i D B D^\dagger C_i^\dagger U_X \\ &= \sum_{i=0}^{N-1} U_X^\dagger C_i \underbrace{(d d^\dagger \odot B)}_Y C_i^\dagger U_X \\ &= \sum_{i=0}^{N-1} \underbrace{U_X^\dagger C_i U_Y}_\tilde{C}_i \Lambda_Y \underbrace{U_Y^\dagger C_i^\dagger U_X}_{\tilde{C}_i^\dagger}. \end{aligned} \quad (8.12)$$

It then follows by Lemma 8.4 that

$$\lambda(X) = \underbrace{\left(\sum_{i=0}^{N-1} \tilde{C}_i \odot \tilde{C}_i^\dagger \right)}_S \lambda(Y). \quad (8.13)$$

The matrix S has nonnegative entries by definition. Its row sums can be computed as Sa and its column sums as $S^T a = S^\dagger a$, with a taken

as the all-1 vector. To show that row and column sums are all unity, it suffices to prove that $\sum_{i=0}^{N-1} \tilde{C}_i \tilde{C}_i^\dagger$, respectively $\sum_{i=0}^{N-1} \tilde{C}_i^\dagger \tilde{C}_i$, has an all-1 diagonal, by Lemma 8.4. This follows because

$$\sum_{i=0}^{N-1} \tilde{C}_i \tilde{C}_i^\dagger = \sum_{i=0}^{N-1} U_X^\dagger C_i U_Y U_Y^\dagger C_i^\dagger U_X = I = \sum_{i=0}^{N-1} \tilde{C}_i^\dagger \tilde{C}_i. \quad (8.14)$$

Hence, S is indeed doubly stochastic.

To conclude the proof, we remove the restriction that A has only nonzero diagonal entries in a similar same way as in the first proof of Lemma 8.1: Because A is nonnegative definite, it follows that if a diagonal entry is zero, the entire row and the entire column that contain this entry vanish. As zero eigenvalues do not alter majorization, it suffices to apply the steps just described to the remaining nonzero submatrix to obtain the desired majorization relation in its general form. \square

8.2.3. Alternative Proof of Lemma 8.1

Proof. For any two nonnegative definite matrices A and B , the majorization $\lambda(B) \succ \lambda(A)$ implies $\det A \geq \det B$ (Marshall and Olkin, 1979, Proposition 3.F.1.a). Hence, by Lemma 8.3,

$$\det(I + A \odot B) \geq \det(I + dd^\dagger \odot B). \quad (8.15)$$

Furthermore, as $dd^\dagger \odot B = DBD^\dagger$, where $D = \text{diag } d$, it follows that

$$\det(I + dd^\dagger \odot B) = \det(I + DBD^\dagger) = \det(I + D^\dagger DB), \quad (8.16)$$

where the last equality follows from Theorem 1.3.20 by Horn and Johnson (1985). Now, $D^\dagger D = I \odot A$ by definition. \square

8.3. MUTUAL INFORMATION AND MMSE

Lemma 8.5. *Let $\{\mathbf{h}[k]\}$ be a stationary random process with auto-correlation $R_{\mathbf{h}}[\Delta k] \triangleq \mathbb{E}[\mathbf{h}[\Delta k + k]\bar{\mathbf{h}}[k]]$ and spectral density*

$$\psi_{\mathbf{h}}(\varphi) \triangleq \sum_{\Delta k=-\infty}^{\infty} R_{\mathbf{h}}[\Delta k] e^{-i2\pi\Delta k\varphi}, \quad |\varphi| \leq 1/2. \quad (8.17)$$

Furthermore, let $\mathbf{h} \triangleq [\mathbf{h}[0] \ \mathbf{h}[1] \ \dots \ \mathbf{h}[K-1]]^T$ be a K -dimensional random vector constructed from $\{\mathbf{h}[k]\}_{k=0}^{K-1}$, and denote the $K \times K$ covariance matrix of \mathbf{h} by $\mathbf{R}_{\mathbf{h}} \triangleq \mathbb{E}[\mathbf{h}\mathbf{h}^\dagger]$. This covariance matrix is Hermitian Toeplitz with entries $[\mathbf{R}_{\mathbf{h}}]_{i,j} = R_{\mathbf{h}}[i-j]$. Then, for any vector \mathbf{x} with binary entries $\{0, 1\}$ and for any $\rho > 0$, the following inequality holds for any finite K :

$$\begin{aligned} \inf_{\mathbf{x}} \frac{1}{\|\mathbf{x}\|^2} \log \det(\mathbf{I}_K + \rho \mathbf{x}\mathbf{x}^\dagger \odot \mathbf{R}_{\mathbf{h}}) \\ \geq \int_{-1/2}^{1/2} \log(1 + \rho \psi_{\mathbf{h}}(\varphi)) d\varphi. \end{aligned} \quad (8.18)$$

Furthermore, in the limit for $K \rightarrow \infty$, the above inequality is achieved by a vector \mathbf{x} whose entries are all equal to 1.

Remark 8.1. The second part of Lemma 8.5, that the infimum can be achieved by an all-1 vector in the limit for $K \rightarrow \infty$, was already proved by Sethuraman and Hajek (2005, Section VI.B). Their proof relies on technical set-theoretic and limiting arguments, so that it is not easy to see how the fundamental property—the stationarity of the process $\{\mathbf{h}[k]\}$ and the corresponding Toeplitz structure of $\mathbf{R}_{\mathbf{h}}$ —come into play. Therefore, it is also difficult to extend their proof to multidimensional stationary processes like we need them for the time and the frequency dimension in Chapter 7. We give an alternative proof that is much shorter, explicitly uses the stationarity property, and which even allows us to derive the seemingly new lower bound (8.18). The proof can be directly generalized to the case of multidimensional stationary processes, as we show in Corollary 8.7 below.

A key idea of the proof is already contained in a derivation by Sanderovich et al. (2007); Dr. S. Shamai (Shitz) suggested its use for the problem at hand. In particular, we employ the relation between mutual information and the MMSE that was discovered recently by Guo et al. (2005). We need a modification of their Theorem 2 to the case of JPG random vectors, which we state here as a lemma for convenience.

Lemma 8.6 (Guo et al., 2005). *Let \mathbf{h} be a random vector that satisfies $\mathbb{E}[\|\mathbf{h}\|^2] < \infty$, and let \mathbf{w} be a zero-mean JPG vector, $\mathbf{w} \sim \mathcal{CN}(0, \mathbf{I})$, independent of \mathbf{h} . Then, for any fixed vector \mathbf{x} of dimension KN ,*

$$\begin{aligned} \frac{d}{d\gamma} I(\mathbf{h}; \sqrt{\gamma}\mathbf{x} \odot \mathbf{h} + \mathbf{w}) \\ = \mathbb{E}[\|\mathbf{x} \odot \mathbf{h} - \mathbf{x} \odot \mathbb{E}[\mathbf{h} | \sqrt{\gamma}\mathbf{x} \odot \mathbf{h} + \mathbf{w}]\|^2]. \end{aligned} \quad (8.19)$$

The expression on the right-hand side in (8.19) is the MMSE if $\mathbf{x} \odot \mathbf{h}$ is estimated from the noisy observation $\sqrt{\gamma}\mathbf{x} \odot \mathbf{h} + \mathbf{w}$.

Proof of Lemma 8.5. We first derive the lower bound (8.18) and then show achievability in the limit for $K \rightarrow \infty$ in a second step.

To apply Lemma 8.6, we rewrite the left-hand side of (8.18) as a mutual information,

$$\frac{1}{\|\mathbf{x}\|^2} \log \det(\mathbf{I}_K + \rho \mathbf{x} \mathbf{x}^\dagger \odot \mathbf{R}_{\mathbf{h}}) = \frac{1}{\|\mathbf{x}\|^2} I(\mathbf{h}; \sqrt{\rho}\mathbf{x} \odot \mathbf{h} + \mathbf{w}), \quad (8.20)$$

where $\mathbf{w} \sim \mathcal{CN}(0, \mathbf{I}_K)$ is a JPG random vector. Without loss of generality, we assume that the vector \mathbf{x} has exactly M nonzero elements with indices in a given index set \mathcal{I} . Then,

$$\begin{aligned} \frac{1}{\|\mathbf{x}\|^2} I(\mathbf{h}; \sqrt{\rho}\mathbf{x} \odot \mathbf{h} + \mathbf{w}) = \\ \stackrel{(a)}{=} \frac{1}{\|\mathbf{x}\|^2} \int_0^\rho \mathbb{E}[\|\mathbf{x} \odot \mathbf{h} - \mathbb{E}[\mathbf{x} \odot \mathbf{h} | \sqrt{\gamma}\mathbf{x} \odot \mathbf{h} + \mathbf{w}]\|^2] d\gamma \end{aligned}$$

$$\begin{aligned}
 &\stackrel{(b)}{=} \frac{1}{M} \int_0^\rho \sum_{m \in \mathcal{I}} \mathbb{E} \left[\left| \mathbf{h}[m] - \mathbb{E} \left[\mathbf{h}[m] \mid \{\sqrt{\gamma} \mathbf{h}[k] + \mathbf{w}[k]\}_{k \in \mathcal{I}} \right] \right|^2 \right] d\gamma \\
 &\stackrel{(c)}{\geq} \frac{1}{M} \int_0^\rho \sum_{m \in \mathcal{I}} \mathbb{E} \left[\left| \mathbf{h}[m] - \mathbb{E} \left[\mathbf{h}[m] \mid \{\sqrt{\gamma} \mathbf{h}[k] + \mathbf{w}[k]\}_{k \in \mathcal{Z}} \right] \right|^2 \right] d\gamma \\
 &\stackrel{(d)}{=} \int_0^\rho \mathbb{E} \left[\left| \mathbf{h}[0] - \mathbb{E} \left[\mathbf{h}[0] \mid \{\sqrt{\gamma} \mathbf{h}[k] + \mathbf{w}[k]\}_{k=-\infty}^\infty \right] \right|^2 \right] d\gamma.
 \end{aligned} \tag{8.21}$$

Here, (a) follows from the relation between mutual information and the MMSE in Lemma 8.6. Equality (b) holds because \mathbf{x} has exactly M nonzero entries whose indices are in \mathcal{I} , and the components of the observation that contain only noise do not influence the estimation error. To obtain inequality (c), we estimate each $\mathbf{h}[m]$ not just from a finite set of noisy observations of the random process $\{\mathbf{h}[k]\}$ but also from noisy observations of the process's infinite past and future. This is the so-called *infinite horizon noncausal* MMSE estimate. Finally, we obtain (d) because the process $\{\mathbf{h}[k]\}$ is stationary and its infinite horizon noncausal MMSE is, therefore, the same for all indices $m \in \mathcal{I}$.

The noncausal MMSE can be expressed in terms of the spectral density of the process $\{\mathbf{h}[k]\}$ (Poor, 1994, Equation (V.D.28)):

$$\begin{aligned}
 &\mathbb{E} \left[\left| \mathbf{h}[0] - \mathbb{E} \left[\mathbf{h}[0] \mid \{\sqrt{\gamma} \mathbf{h}[k] + \mathbf{w}[k]\}_{k=-\infty}^\infty \right] \right|^2 \right] \\
 &= \int_{-1/2}^{1/2} \frac{\psi_{\mathbf{h}}(\varphi)}{1 + \gamma \psi_{\mathbf{h}}(\varphi)} d\varphi. \tag{8.22}
 \end{aligned}$$

To obtain the desired inequality (8.18), we substitute (8.22) in (8.21) and note that the lower bound (8.21) does not depend on \mathbf{x} . Therefore, it is also a lower bound on the left-hand side of (8.18). We finally

integrate over γ to obtain

$$\begin{aligned} & \inf_{\mathbf{x}} \frac{1}{\|\mathbf{x}\|^2} \log \det(\mathbf{I}_K + \rho \mathbf{x} \mathbf{x}^\dagger \odot \mathbf{R}_{\mathbf{h}}) \\ & \geq \int_{-1/2}^{1/2} \int_0^\rho \frac{\psi_{\mathbf{h}}(\varphi)}{1 + \gamma \psi_{\mathbf{h}}(\varphi)} d\gamma d\varphi = \int_{-1/2}^{1/2} \log(1 + \rho \psi_{\mathbf{h}}(\varphi)) d\varphi. \end{aligned} \quad (8.23)$$

To prove the second part of Lemma 8.5, we choose \mathbf{x} to be the all-1 vector for any dimension K and evaluate the limit for $K \rightarrow \infty$ by means of Szegő's Theorem on the eigenvalue distribution of a Toeplitz matrix (Grenander and Szegő, 1984; Gray, 2005):

$$\lim_{K \rightarrow \infty} \frac{1}{K} \log \det(\mathbf{I}_K + \rho \mathbf{R}_{\mathbf{h}}) = \int_{-1/2}^{1/2} \log(1 + \rho \psi_{\mathbf{h}}(\varphi)) d\varphi. \quad (8.24)$$

This shows that the lower bound in (8.18) can indeed be achieved in the limit for $K \rightarrow \infty$. \square

We already mentioned that our proof allows for a simple generalization of Lemma 8.5 to multidimensional stationary processes. In the following corollary, we state this result for the case of a two-dimensional stationary process.

Corollary 8.7. *Let $\{\mathbf{h}[k, n]\}$ be a random process that is stationary in discrete time k and discrete frequency n with two-dimensional autocorrelation function $R_{\mathbf{h}}[\Delta k, \Delta n] \triangleq \mathbb{E}[\mathbf{h}[\Delta k + k, \Delta n + n] \bar{\mathbf{h}}[k, n]]$ and two-dimensional spectral density*

$$\psi_{\mathbf{h}}(\varphi, \zeta) \triangleq \sum_{\Delta k=-\infty}^{\infty} \sum_{\Delta n=-\infty}^{\infty} R_{\mathbf{h}}[\Delta k, \Delta n] e^{-i2\pi(\Delta k\varphi - \Delta n\zeta)} \quad (8.25)$$

for all $|\varphi|, |\zeta| \leq 1/2$. Furthermore, let

$$\mathbf{h}[k] \triangleq [\mathbf{h}[k, 0] \ \mathbf{h}[k, 1] \ \dots \ \mathbf{h}[k, N-1]]^T \quad (8.26)$$

8.3 MUTUAL INFORMATION AND MMSE

be an N -dimensional random vector constructed from $\{\mathbf{h}[k, n]\}_{n=0}^{N-1}$, define the stacked vector

$$\mathbf{h} \triangleq [\mathbf{h}^T[0] \ \mathbf{h}^T[1] \ \cdots \ \mathbf{h}^T[K-1]]^T, \quad (8.27)$$

and denote the $KN \times KN$ covariance matrix of \mathbf{h} by $\mathbf{R}_{\mathbf{h}} \triangleq \mathbb{E}[\mathbf{h}\mathbf{h}^\dagger]$. Then, for any vector \mathbf{x} with binary entries $\{0, 1\}$ and for any $\rho > 0$, the following inequality holds for any finite K and N :

$$\begin{aligned} \inf_{\mathbf{x}} \frac{1}{\|\mathbf{x}\|^2} \log \det(\mathbf{I}_{KN} + \rho \mathbf{x} \mathbf{x}^\dagger \odot \mathbf{R}_{\mathbf{h}}) \\ \geq \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \log(1 + \rho \psi_{\mathbf{h}}(\varphi, \zeta)) d\varphi d\zeta. \end{aligned} \quad (8.28)$$

Furthermore, in the limit for $K, N \rightarrow \infty$, the above inequality can be achieved by a vector \mathbf{x} whose entries are all equal to 1.

Proof. Without loss of generality, we assume that the vector \mathbf{x} has exactly M nonzero elements with indices in a given index set \mathcal{I} . The arguments used in (8.21) directly apply, and we obtain

$$\begin{aligned} \frac{1}{\|\mathbf{x}\|^2} \log \det(\mathbf{I}_{KN} + \rho \mathbf{x} \mathbf{x}^\dagger \odot \mathbf{R}_{\mathbf{h}}) \geq \\ \int_0^\rho \mathbb{E} \left[\left| \mathbf{h}[0, 0] - \mathbb{E} \left[\mathbf{h}[0, 0] \mid \{\sqrt{\gamma} \mathbf{h}[k, n] + \mathbf{w}[k, n]\}_{k, n=-\infty}^\infty \right] \right|^2 \right] d\gamma. \end{aligned} \quad (8.29)$$

To complete the proof, we use the two-dimensional counterpart of (8.22)—the closed-form expression for the two-dimensional non-causal MMSE (Helstrom, 1967, Equation (2.6)), and we compute the two-dimensional equivalent of (8.24) by an extension of Szegő's Theorem to two-level Toeplitz matrices (Sakrison, 1969; Tyrtshnikov, 1996). \square

8.4. OPTIMIZATION OF THE UPPER BOUND

In this section, we analyze the remaining optimization in the upper bound of Theorem 7.1, restated here for convenience. A closed-form expression of the parameter α that maximizes

$$U_\alpha \triangleq \sum_{r=0}^{M_R-1} \left(\frac{B}{TF} \log \left(1 + \alpha \lambda_r \frac{PTF}{B} \right) - \alpha \underbrace{\frac{B}{\sigma_0 \beta} \iint_{\nu \tau} \log \left(1 + \frac{\sigma_0 \lambda_r \beta P}{B} c_s(\nu, \tau) \right) d\tau d\nu}_{G_r(B)} \right) \quad (8.30)$$

over all $\alpha \in [0, \sigma_0]$ does not exist for $M_R > 4$, and even for $M_R = 1$, the upper bound (7.20) becomes intractable for all parameter values except $\alpha = \sigma_0$. Therefore, we analyze here if the case that $\alpha = \sigma_0$ achieves the supremum in (7.20) is operationally relevant. Our goal is to find a sufficient condition for $\alpha = \sigma_0$ to maximize U_α and to assess the range of the physical parameters power, spread, and bandwidth under which this condition is satisfied.

The function U_α to be optimized is continuous and its first and second derivatives with respect to the optimization parameter α are

$$\frac{\partial}{\partial \alpha} U_\alpha = \sum_{r=0}^{M_R-1} \left(\frac{\lambda_r P}{1 + \alpha \lambda_r PTF/B} - G_r(B) \right), \quad (8.31a)$$

$$\frac{\partial^2}{\partial \alpha^2} U_\alpha = - \sum_{r=0}^{M_R-1} \frac{\lambda_r^2 TFB P^2}{(B + \alpha \lambda_r TFP)^2}. \quad (8.31b)$$

It follows that U_α is concave in α ; hence, the optimizing parameter α is unique. We bound the penalty term $G_r(B)$ by Jensen's inequality as

$$G_r(B) \leq \frac{B \Delta_{\mathbb{H}}}{\sigma_0 \beta} \log \left(1 + \frac{\sigma_0 \lambda_r \beta P}{\Delta_{\mathbb{H}} B} \right) \leq \lambda_r P \quad (8.32)$$

to conclude that

$$\frac{\partial}{\partial \alpha} U_\alpha \Big|_{\alpha=0} \geq 0. \quad (8.33)$$

Furthermore, we can see from (8.30) that $U_\alpha = 0$ for $\alpha = 0$. Therefore, a necessary and sufficient condition for $\alpha = \sigma_0$ to maximize U_α is that the zero of the first derivative (8.31a) occurs at a point larger or equal to σ_0 , or, equivalently, that (8.31a) is positive for $\alpha \in (0, \sigma_0)$. Then, concavity guarantees that the maximum is attained on the boundary of the admissible set, at $\alpha = \sigma_0$.

Finding the zeros of (8.31a) is difficult for rank $B > 1$. However, for our purposes it suffices to bound these zeros away from $(0, \sigma_0)$, i.e., it suffices if each term on the right-hand side of (8.31a) satisfies

$$\frac{\lambda_r P}{1 + \alpha \lambda_r PTF/B} > G_r(B). \quad (8.34)$$

Assume now that each of the following implications holds:

$$\begin{aligned} \xrightarrow{(a)} \quad & \frac{\lambda_r P}{1 + \alpha \lambda_r PTF/B} > \frac{B \Delta_{\mathbb{H}}}{\sigma_0 \beta} \log \left(1 + \frac{\sigma_0 \lambda_r \beta P}{\Delta_{\mathbb{H}} B} \right) \\ \xrightarrow{(b)} \quad & \frac{\lambda_r P}{1 + \sigma_0 \lambda_r PTF/B} > \frac{B \Delta_{\mathbb{H}}}{\sigma_0 \beta} \log \left(1 + \frac{\sigma_0 \lambda_r \beta P}{\Delta_{\mathbb{H}} B} \right) \\ \Leftrightarrow^{(c)} \quad & \left(\frac{1}{\lambda_r \rho} + \sigma_0 TF \right)^{-1} > \frac{\Delta_{\mathbb{H}}}{\sigma_0 \beta} \log \left(1 + \frac{\sigma_0 \lambda_r \beta \rho}{\Delta_{\mathbb{H}}} \right). \end{aligned} \quad (8.35)$$

For (a), we use Jensen's inequality, as in (8.32); if we upper-bound α by σ_0 , (b) results, and (c) follows from a change of variables: we define the SNR $\rho \triangleq P/B$. Reversing the argument, the last condition, i.e., condition (8.35), is sufficient for (8.34) to hold, and consequently for the r th addend of the first derivative (8.31a) to be positive in $(0, \sigma_0)$. To render the condition (8.35) more tractable, we replace it by two complementary conditions that are easier to evaluate:

- We replace the left-hand side of the sufficient condition (8.35) by a lower bound.
- We replace the right-hand side in (8.35) by an upper bound.

A. First Condition

If $\rho \geq 1/(\sigma_0\lambda_rTF)$, we lower-bound the left-hand side in (8.35) so that

$$\frac{1}{2\sigma_0TF} > \frac{\Delta_{\mathbb{H}}}{\sigma_0\beta} \log\left(1 + \frac{\sigma_0\lambda_r\beta\rho}{\Delta_{\mathbb{H}}}\right) \quad \text{and} \quad \rho \geq 1/(\sigma_0\lambda_rTF), \quad (8.36)$$

which can be expressed in terms of ρ as

$$\frac{\Delta_{\mathbb{H}}}{\sigma_0\lambda_r\beta} \left[\exp\left(\frac{\beta}{2\Delta_{\mathbb{H}}TF}\right) - 1 \right] > \rho \geq \frac{1}{\sigma_0\lambda_rTF} \quad (8.37)$$

B. Second Condition

An upper bound on the right-hand side of (8.35) results from the inequality*

$$\frac{1}{x} \log(1+x) \leq \frac{1}{\sqrt{1+x}}, \quad \text{for all } x \geq 0. \quad (8.38)$$

Hence, we obtain a second sufficient condition:

$$\begin{aligned} \left(\frac{1}{\lambda_r\rho} + \sigma_0TF\right)^{-1} &> \lambda_r\rho \left(1 + \frac{\sigma_0\lambda_r\beta\rho}{\Delta_{\mathbb{H}}}\right)^{-1/2} \\ \iff \frac{\beta/(\Delta_{\mathbb{H}}TF) - 2}{\sigma_0\lambda_rTF} &> \rho > 0, \end{aligned} \quad (8.39)$$

where equivalency holds because on physical grounds we can rule out all solutions that require $\rho \leq 0$.

C. Joint Condition

For $\rho \leq 1/(\sigma_0\lambda_rTF)$, we only need to consider the second condition (8.39); otherwise, we can use the condition (8.37). Hence, to obtain a joint sufficient condition, we need to ensure that the admissible sets for ρ in (8.37) and (8.39) overlap, i.e.,

$$\frac{\beta/(\Delta_{\mathbb{H}}TF) - 2}{\sigma_0\lambda_rTF} \geq \frac{1}{\sigma_0\lambda_rTF}. \quad (8.40)$$

* C. Aćkaba suggested this inequality.

But this latter condition is equivalent to

$$\underbrace{\frac{\Delta_{\mathbb{H}}TF}{\beta}}_{\tilde{\Delta}_{\mathbb{H}}} \leq \frac{1}{3}. \quad (8.41)$$

Therefore, we combine (8.37) and (8.39) to find the following set of sufficient conditions for $\alpha = \sigma_0$:

$$0 < \rho TF < \frac{\tilde{\Delta}_{\mathbb{H}}}{\sigma_0 \lambda_r} \left[\exp\left(\frac{1}{2\tilde{\Delta}_{\mathbb{H}}}\right) - 1 \right] \quad \text{and} \quad \tilde{\Delta}_{\mathbb{H}} \leq \frac{1}{3}, \quad (8.42)$$

where $\tilde{\Delta}_{\mathbb{H}} \triangleq \Delta_{\mathbb{H}}TF/\beta$ is the effective spread of the channel.

8.5. PENALTY TERM OF THE LOWER BOUND

To proof Lemma 7.3, we need to upper- and lower-bound the penalty term addend

$$\frac{1}{T} \int_{-1/2}^{1/2} \log \det \left(\mathbf{I}_N + \sigma_q \lambda_r \frac{PTF}{QB} \Psi(\varphi) \right) d\varphi. \quad (8.43)$$

An upper bound follows from the asymptotic equivalence between Toeplitz and circulant matrices, a lower bound from Lemma 8.5.

8.5.1. Upper Bound

We would like to upper-bound (8.43) by an expression that can be evaluated efficiently even for large bandwidth and that is asymptotically tight. To obtain such a bound, we need to solve two problems. (i) The eigenvalues of the Toeplitz matrix $\Psi(\varphi)$ are difficult to compute. Therefore, we replace it with an asymptotically equivalent circulant matrix whose eigenvalues can be obtained efficiently by the DFT. The asymptotic equivalence guarantees that the resulting bound is asymptotically tight. For an extensive discussion of asymptotic equivalence between Toeplitz and circulant matrices, see the

tutorial article by Gray (2005). (ii) The determinant expression in the penalty term (8.43) needs to be evaluated for all values of the continuous parameter φ . To remove this dependency, we use Jensen's inequality.

We need the following result on the asymptotic equivalence between Toeplitz and circulant matrices.

Lemma 8.8 (Pearl, 1973). *Let \mathbf{T} be a $N \times N$ Toeplitz matrix. Furthermore, let \mathbf{F} be the DFT matrix, i.e., the matrix*

$$\mathbf{F} \triangleq [\mathbf{f}_0 \ \mathbf{f}_1 \ \cdots \ \mathbf{f}_{N-1}] \quad (8.44)$$

whose columns

$$\mathbf{f}_n \triangleq [\beta^{0n} \ \beta^{1n} \ \cdots \ \beta^{(N-1)n}]^T / \sqrt{N} \quad (8.45)$$

contain increasing powers of the N th root of unity, $\beta = e^{i2\pi/N}$. Construct from the matrix $\mathbf{F}^\dagger \mathbf{T} \mathbf{F}$ the diagonal matrix \mathbf{D} so that the entries on the main diagonal of \mathbf{D} and on the main diagonal of $\mathbf{F}^\dagger \mathbf{T} \mathbf{F}$ are the same. Then, \mathbf{T} and the circulant matrix $\mathbf{F} \mathbf{D} \mathbf{F}^\dagger$ are asymptotically equivalent.

The asymptotically equivalent matrix $\mathbf{F} \mathbf{D} \mathbf{F}^\dagger$ is the best circulant approximation of \mathbf{T} in Frobenius norm for any finite dimension N (Chan et al., 1991).

We want to upper-bound a function of the form $\log \det(\mathbf{I}_N + \mathbf{T}/N)$. Because \mathbf{F} is unitary, and by Hadamard's inequality,

$$\begin{aligned} \log \det \left(\mathbf{I}_N + \frac{1}{N} \mathbf{T} \right) &= \log \det \left(\mathbf{I}_N + \frac{1}{N} \mathbf{F}^\dagger \mathbf{T} \mathbf{F} \right) \\ &\leq \log \det \left(\mathbf{I}_N + \frac{1}{N} \mathbf{D} \right) \\ &= \log \det \left(\mathbf{I}_N + \frac{1}{N} \mathbf{F} \mathbf{D} \mathbf{F}^\dagger \right). \end{aligned} \quad (8.46)$$

As \mathbf{T} and $\mathbf{F} \mathbf{D} \mathbf{F}^\dagger$ are asymptotically equivalent by Lemma 8.8, the difference between $\log \det(\mathbf{I}_N + \mathbf{D}/N)$ and $\log \det(\mathbf{I}_N + \mathbf{T}/N)$ vanishes

as N grows large (Gray, 2005, Lemma 4.3), i.e., the bound is asymptotically tight.

To apply the bound (8.46) to our problem of upper-bounding (8.43), we need to compute the diagonal entries of $\mathbf{F}^\dagger \boldsymbol{\Psi}(\varphi) \mathbf{F}$. We denote the entries of the Toeplitz covariance matrix $\boldsymbol{\Psi}(\varphi)$ as $\{\psi_{\Delta n}(\varphi)\}_{\Delta n=-(N-1)}^{N-1}$, so that

$$\boldsymbol{\Psi}(\varphi) = \begin{bmatrix} \psi_0(\varphi) & \psi_{-1}(\varphi) & \cdots & \psi_{-(N-1)}(\varphi) \\ \psi_1(\varphi) & \psi_0(\varphi) & \cdots & \psi_{-(N-2)}(\varphi) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{N-1}(\varphi) & \psi_{N-2}(\varphi) & \cdots & \psi_0(\varphi) \end{bmatrix}. \quad (8.47)$$

By (3.36), each entry $\psi_{\Delta n}(\varphi)$ is related to the discrete-time discrete-frequency correlation function $c_{\mathcal{I}}[\Delta k, \Delta n]$ by the following Fourier transform:

$$\psi_{\Delta n}(\varphi) = \sum_{\Delta k=-\infty}^{\infty} c_{\mathcal{I}}[\Delta k, \Delta n] e^{-i2\pi \Delta k \varphi} \quad (8.48)$$

$$= \frac{1}{T} \sum_{\Delta k=-\infty}^{\infty} \int_{-\tau_0/2}^{\tau_0/2} c_s\left(\frac{\varphi + \Delta k}{T}, \tau\right) e^{-i2\pi \tau \Delta n F} d\tau. \quad (8.49)$$

Consequently, the j th element on the main diagonal of $\mathbf{F}^\dagger \boldsymbol{\Psi}(\varphi) \mathbf{F}$, which we denote as $d_j(\varphi)$, can be expressed as a function of the entries of $\boldsymbol{\Psi}(\varphi)$, using the substitution $\Delta n = q - p$.

$$\begin{aligned} d_j(\varphi) &= \frac{1}{N} \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} \beta^{-jq} \psi_{q-p}(\varphi) \beta^{jp} \\ &= \frac{1}{N} \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} \psi_{q-p}(\varphi) \beta^{-j(q-p)} \\ &= \frac{1}{N} \sum_{\Delta n=-(N-1)}^{N-1} (N - |\Delta n|) \psi_{\Delta n}(\varphi) e^{-i2\pi \frac{j\Delta n}{N}} \\ &= \Re \left\{ \frac{2}{N} \sum_{\Delta n=0}^{N-1} (N - \Delta n) \psi_{\Delta n}(\varphi) e^{-i2\pi \frac{j\Delta n}{N}} \right\} - \psi_0(\varphi). \quad (8.50) \end{aligned}$$

8 TECHNICAL RESULTS

We are now in a position to use the upper bound (8.46) for the penalty term addend (8.43):

$$\begin{aligned}
 & \frac{1}{T} \int_{-1/2}^{1/2} \log \det \left(\mathbf{I}_N + \sigma_q \lambda_r \frac{PTF}{QB} \Psi(\varphi) \right) d\varphi \\
 &= \frac{1}{T} \int_{-1/2}^{1/2} \log \det \left(\mathbf{I}_N + \sigma_q \lambda_r \frac{PTF}{QB} \mathbf{F}^\dagger \Psi(\varphi) \mathbf{F} \right) d\varphi \\
 &\leq \frac{1}{T} \int_{-1/2}^{1/2} \sum_{j=0}^{N-1} \log \left(1 + \sigma_q \lambda_r \frac{PTF}{QB} d_j(\varphi) \right) d\varphi \\
 &= \int_{-\nu_0/2}^{\nu_0/2} \sum_{j=0}^{N-1} \log \left(1 + \sigma_q \lambda_r \frac{PTF}{QB} d_j(\nu T) \right) d\nu, \quad (8.51)
 \end{aligned}$$

where the last equality follows from a change of variables $\nu = \varphi/T$.

We proceed to remove the dependence on φ . To this end, we further upper-bound (8.51) by means of Jensen's inequality and obtain the desired upper bound in (7.50). Define

$$d_j \triangleq T \int_{-\nu_0/2}^{\nu_0/2} d_j(\nu T) d\nu. \quad (8.52)$$

Then,

$$\begin{aligned}
 & \int_{-\nu_0/2}^{\nu_0/2} \sum_{j=0}^{N-1} \log \left(1 + \sigma_q \lambda_r \frac{PTF}{QB} d_j(\nu T) \right) d\nu \\
 &\leq \nu_0 \sum_{j=0}^{N-1} \log \left(1 + \sigma_q \lambda_r \frac{PTF}{\nu_0 QB} \int_{-\nu_0/2}^{\nu_0/2} d_j(\nu T) d\nu \right) \\
 &= \nu_0 \sum_{j=0}^{N-1} \log \left(1 + \sigma_q \lambda_r \frac{PF}{\nu_0 QB} d_j \right). \quad (8.53)
 \end{aligned}$$

Because

$$\begin{aligned}
 T \int_{-\nu_0/2}^{\nu_0/2} \psi_{\Delta n}(\nu T) d\nu &= \sum_{k=-\infty}^{\infty} \int_{-\tau_0/2}^{\tau_0/2} \int_{-\nu_0/2}^{\nu_0/2} c_{\mathbf{s}}\left(\nu + \frac{k}{T}, \tau\right) e^{-i2\pi\tau\Delta n F} d\tau d\nu \\
 &= \int_{-\tau_0/2}^{\tau_0/2} \int_{-\nu_0/2}^{\nu_0/2} c_{\mathbf{s}}(\nu, \tau) e^{-i2\pi\tau\Delta n F} d\tau d\nu \\
 &= c_{\mathbf{t}}[0, \Delta n], \tag{8.54}
 \end{aligned}$$

it follows from (8.50) that

$$d_j = \Re \left\{ \frac{2}{N} \sum_{\Delta n=0}^{N-1} (N - \Delta n) c_{\mathbf{t}}[0, \Delta n] e^{-i2\pi \frac{j\Delta n}{N}} \right\} - 1 \tag{8.55}$$

as defined in (7.49). Because we obtained the upper bound (8.53) from the tighter upper bound (8.51) by Jensen's inequality, the bound (8.53) is asymptotically tight only if the inequality (8.51) reduces to an equality. This happens for scattering functions that are flat in the Doppler domain, or, equivalently, that satisfy (7.51).

8.5.2. Lower Bound

To lower-bound the penalty term addend (8.43), we use Lemma 8.5 and the relation between the spectral density of the underlying discrete-time discrete-frequency component channel and the scattering function of the continuous-time component channel in (3.37) to conclude that

$$\begin{aligned} \frac{1}{T} \int_{-1/2}^{1/2} \log \det \left(\mathbf{I}_N + \sigma_q \lambda_r \frac{PTF}{QB} \Psi(\varphi) \right) d\varphi \\ \geq B \iint_{\nu \tau} \log \left(1 + \sigma_q \lambda_r \frac{P}{QB} c_s(\nu, \tau) \right) d\tau d\nu. \end{aligned} \quad (8.56)$$

Furthermore, by Szegő's Theorem, the difference between the right-hand side and the left-hand side of the above inequality vanishes as $B \rightarrow \infty$. We rigorously prove this last result in Section 8.6.2, where we analyze the first order Taylor expansion of the lower bound $L_1(B)$ in (7.42).

8.6. TAYLOR SERIES EXPANSION OF CAPACITY

8.6.1. Upper Bound

We would like to compute $\lim_{B \rightarrow \infty} BU_1(B)$. This task is complicated by the supremum over α in (7.20). However, the argument of the supremum is a measurable function in α , so that by Fatou's Lemma (Rudin, 1976),

$$\begin{aligned} \lim_{B \rightarrow \infty} BU_1(B) \leq \sup_{0 \leq \alpha \leq \sigma_0} \lim_{B \rightarrow \infty} \sum_{r=0}^{M_R-1} \left(\frac{B^2}{TF} \log \left(1 + \alpha \lambda_r \frac{PTF}{B} \right) \right. \\ \left. - \frac{\alpha B^2}{\sigma_0 \beta} \iint_{\nu \tau} \log \left(1 + \frac{\sigma_0 \lambda_r \beta P}{B} c_s(\nu, \tau) \right) d\tau d\nu \right). \end{aligned} \quad (8.57)$$

Hence, we can compute the limit first. To do so, we take a look at the two terms inside the summation separately. We expand the first term

in a Taylor series according to $\log(1+x) = x - x^2/2 + o(x^2)$, i.e.,

$$\begin{aligned} \frac{B^2}{TF} \log \left(1 + \alpha \lambda_r \frac{PTF}{B} \right) \\ = \alpha \lambda_r BP - \frac{\alpha^2 \lambda_r^2 P^2 T F}{2} + o \left(\frac{1}{B^2} \right). \end{aligned} \quad (8.58)$$

As $c_s(\nu, \tau)$ does not depend on B , the Monotone Convergence Theorem (Rudin, 1976) allows to similarly expand the logarithm inside the penalty term integral,

$$\begin{aligned} \frac{\alpha B^2}{\sigma_0 \beta} \iint_{\nu \tau} \log \left(1 + \frac{\sigma_0 \lambda_r \beta P}{B} c_s(\nu, \tau) \right) d\tau d\nu \\ = \alpha \lambda_r BP \iint_{\nu \tau} c_s(\nu, \tau) d\tau d\nu \\ - \frac{\alpha \sigma_0 \lambda_r^2 \beta P^2}{2} \underbrace{\iint_{\nu \tau} c_s^2(\nu, \tau) d\tau d\nu}_{\kappa_{\mathbb{H}}} + o \left(\frac{1}{B^2} \right) \\ = \alpha \lambda_r BP - \frac{\alpha \sigma_0 \lambda_r^2 \kappa_{\mathbb{H}} \beta P^2}{2} + o \left(\frac{1}{B^2} \right) \end{aligned} \quad (8.59)$$

We can now combine the two terms and compute the limit in (8.57):

$$\begin{aligned} \lim_{B \rightarrow \infty} B U_1(B) &\leq \sup_{0 \leq \alpha \leq \sigma_0} \lim_{B \rightarrow \infty} \sum_{r=0}^{M_{\mathbb{R}}-1} \left(+ \frac{\alpha \sigma_0 \lambda_r^2 \kappa_{\mathbb{H}} \beta P^2}{2} \right. \\ &\quad \left. - \frac{\alpha^2 \lambda_r^2 P^2 T F}{2} - o \left(\frac{1}{B^2} \right) \right) \\ &= \sup_{0 \leq \alpha \leq \sigma_0} \left\{ \frac{\alpha P^2}{2} (\sigma_0 \kappa_{\mathbb{H}} \beta - \alpha T F) \underbrace{\sum_{r=0}^{M_{\mathbb{R}}-1} \lambda_r}_{\vartheta} \right\}. \end{aligned} \quad (8.60)$$

Elementary optimization now yields the maximizing parameter α , so that

$$\lim_{B \rightarrow \infty} BU_1(B) \leq \begin{cases} \frac{\vartheta(\sigma_0 P)^2}{2} (\beta \kappa_{\mathbb{H}} - TF), & \text{if } \beta > \frac{2TF}{\kappa_{\mathbb{H}}} \\ \frac{\vartheta(\sigma_0 \kappa_{\mathbb{H}} \beta P)^2}{8TF}, & \text{if } \beta \leq \frac{2TF}{\kappa_{\mathbb{H}}}. \end{cases} \quad (8.61)$$

8.6.2. Lower Bound

To prove Theorem 7.4, we need to find a lower bound on $C(B)$ whose first-order Taylor expansion matches (8.61). To this end, we compute the mutual information for a specific signaling scheme that generalizes the scheme used by Sethuraman et al. (2009).

We observe in Section 7.6 that signaling over a single transmit eigenmode seems to be optimal for very wide bandwidth; hence, it is sensible to use a signaling scheme that uses only the strongest transmit eigenmode in each TF slot for the construction of the asymptotic lower bound. Such a signaling scheme transmits over all available antennas in general; only if there is no spatial correlation at the transmitter can antennas be physically switched off. The effective IO relation induced by this signaling scheme has a single input and M_{R} outputs in each TF slot. On the strongest eigenmode, we let the transmitter modulate data onto blocks of K' consecutive slots in time and N' consecutive slots in frequency, where $K' \leq K$ and $N' \leq N$ are fixed. For now, we also assume that the total available bandwidth N and the total available time per channel use K are integer multiples of N' and K' , respectively, so that the transmitter can send $N/N' \times K/K'$ blocks in one channel use. The signal in each TF slot in a given block is independent and identically distributed with unit amplitude and zero mean, i.e., we use constant modulus signaling within each block. We arrange the $K' \times N'$ block of these constant-modulus signals in a vector \mathbf{d} in the same way as in (7.3), i.e., we stack first in frequency and then in time. The transmitter does not directly send \mathbf{d} along the strongest eigenmode but $\tilde{\mathbf{x}}_0 = b \mathbf{d}$, where b is a binary random variable

that takes independent values across blocks and has distribution

$$b = \begin{cases} \sqrt{\beta PT/N} & \text{with probability } \zeta, \\ 0 & \text{with probability } 1 - \zeta. \end{cases} \quad (8.62)$$

This means that each block of i.i.d. constant modulus signals undergoes on-off modulation with *duty cycle* ζ . The above signaling scheme satisfies the peak constraint (7.13) by construction. The covariance matrix of the input vector $\tilde{\mathbf{x}}_0$ is given by

$$\mathbb{E} [\tilde{\mathbf{x}}_0 \tilde{\mathbf{x}}_0^\dagger] = \mathbb{E}_b [\mathbb{E}_{\tilde{\mathbf{x}}_0} [\tilde{\mathbf{x}}_0 \tilde{\mathbf{x}}_0^\dagger | b]] = \zeta \frac{\beta PT}{N} \mathbf{I}_{K'N'}, \quad (8.63)$$

from which we can infer that $\tilde{\mathbf{x}}_0$ also satisfies the constraint on the average power if we chose $\zeta \leq 1/\beta$. Because the input vector is nonzero only on the strongest eigenmode, the MIMO channel reduces to a single input multiple output (SIMO) channel that consists of M_R component channels. The r th component channel, truncated to K' time slots and N' frequency slots, is denoted as $\tilde{\mathbf{h}}_r$, and we stack all M_R component channel vectors according to

$$\tilde{\mathbf{h}} \triangleq [\tilde{\mathbf{h}}_0^T \ \tilde{\mathbf{h}}_1^T \ \cdots \ \tilde{\mathbf{h}}_{M_R-1}^T]^T. \quad (8.64)$$

The covariance matrices of the time-frequency truncated component channels are all equal; we denote them as

$$\mathbf{R}_{\tilde{\mathbf{h}}} = \mathbb{E} [\tilde{\mathbf{h}}_r \tilde{\mathbf{h}}_r^\dagger]. \quad (8.65)$$

Consequently, the covariance of the stacked channel is $\mathbb{E} [\tilde{\mathbf{h}} \tilde{\mathbf{h}}^\dagger] = \sigma_0 \Lambda^{1/2} \otimes \mathbf{R}_{\tilde{\mathbf{h}}}$. We denote the TF-truncated and stacked output vector as $\tilde{\mathbf{y}}$, where stacking proceeds as in (7.3), and define $\tilde{\mathbf{X}}_0 \triangleq \text{diag } \tilde{\mathbf{x}}_0$. Then, the IO relation (7.5) for the transmission of one block reduces to

$$\tilde{\mathbf{y}} = (\mathbf{I}_{M_R} \otimes \tilde{\mathbf{X}}_0) \tilde{\mathbf{h}} + \tilde{\mathbf{w}}. \quad (8.66)$$

As the transmit signals are i.i.d. across blocks, we can use the chain rule for mutual information to lower-bound the rates achievable by

this signaling scheme:

$$C(B) = \lim_{K \rightarrow \infty} \sup_{\mathcal{P}} \frac{1}{KT} I(\mathbf{x}; \mathbf{y}) \geq \frac{N}{K'N'T} I(\tilde{\mathbf{x}}_0; \tilde{\mathbf{y}}). \quad (8.67)$$

The right-hand side of (8.67) is the rate achievable over a generalized SIMO block fading channel with blocks of size $K' \times N'$, independent fading across blocks, and correlation within each block equal to the correlation of the original channel. Such a channel model is also analyzed by Liang and Veeravalli (2004).

Because we are only interested in the asymptotic behavior of the above bound, it suffices to analyze the second-order expansion of the mutual information $I(\tilde{\mathbf{x}}_0; \tilde{\mathbf{y}})$ on the right-hand side in (8.67). For the channel model at hand, this second-order expansion was derived by Prelov and Verdú (2004, Corollary 1) for fixed K' and N' :

$$\begin{aligned} I(\tilde{\mathbf{x}}_0; \tilde{\mathbf{y}}) &= \frac{1}{2} \operatorname{tr} \left\{ \mathbb{E}_{\tilde{\mathbf{x}}_0} \left[\left(\mathbb{E}_{\tilde{\mathbf{h}}} \left[(\mathbf{I}_{M_R} \otimes \tilde{\mathbf{X}}_0) \tilde{\mathbf{h}} \tilde{\mathbf{h}}^\dagger (\mathbf{I}_{M_R} \otimes \tilde{\mathbf{X}}_0)^\dagger \right] \right)^2 \right] \right\} \\ &\quad - \frac{1}{2} \operatorname{tr} \left\{ \left(\mathbb{E}_{\tilde{\mathbf{h}}, \tilde{\mathbf{x}}_0} \left[(\mathbf{I}_{M_R} \otimes \tilde{\mathbf{X}}_0) \tilde{\mathbf{h}} \tilde{\mathbf{h}}^\dagger (\mathbf{I}_{M_R} \otimes \tilde{\mathbf{X}}_0)^\dagger \right] \right)^2 \right\} \\ &\quad + o\left(\frac{1}{N^2}\right). \end{aligned} \quad (8.68)$$

In the following, we use that

$$\begin{aligned} \mathbb{E}_{\tilde{\mathbf{h}}} \left[(\mathbf{I}_{M_R} \otimes \tilde{\mathbf{X}}_0) \tilde{\mathbf{h}} \tilde{\mathbf{h}}^\dagger (\mathbf{I}_{M_R} \otimes \tilde{\mathbf{X}}_0)^\dagger \right] &= \sigma_0 \Lambda \otimes (\tilde{\mathbf{X}}_0 \mathbf{R}_{\tilde{\mathbf{h}}} \tilde{\mathbf{X}}_0^\dagger) \\ &= \sigma_0 \Lambda \otimes (\mathbf{R}_{\tilde{\mathbf{h}}} \odot \tilde{\mathbf{x}}_0 \tilde{\mathbf{x}}_0^\dagger) \end{aligned} \quad (8.69)$$

and analyze the two trace terms in (8.68) separately.

The first trace term in (8.68) is

$$\begin{aligned}
 & \text{tr} \left\{ \mathbb{E} \left[\left(\sigma_0 \Lambda \otimes (\mathbf{R}_{\tilde{\mathbf{h}}} \odot \tilde{\mathbf{x}}_0 \tilde{\mathbf{x}}_0^\dagger) \right)^2 \right] \right\} \\
 &= \sigma_0^2 \text{tr} \left\{ \Lambda^2 \otimes \mathbb{E} \left[(\mathbf{R}_{\tilde{\mathbf{h}}} \odot \tilde{\mathbf{x}}_0 \tilde{\mathbf{x}}_0^\dagger)^2 \right] \right\} \\
 &\stackrel{(a)}{=} \sigma_0^2 \underbrace{\left(\sum_{r=0}^{M_{\mathbf{R}}-1} \lambda_r^2 \right)}_{\vartheta} \text{tr} \left\{ \mathbb{E} \left[(\mathbf{R}_{\tilde{\mathbf{h}}} \odot \tilde{\mathbf{x}}_0 \tilde{\mathbf{x}}_0^\dagger)^2 \right] \right\} \\
 &\stackrel{(b)}{=} \sigma_0^2 \vartheta \text{tr} \left\{ \mathbb{E} \left[\left(\mathbf{R}_{\tilde{\mathbf{h}}} \odot \tilde{\mathbf{x}}_0 \tilde{\mathbf{x}}_0^\dagger \right)^\dagger \left(\mathbf{R}_{\tilde{\mathbf{h}}} \odot \tilde{\mathbf{x}}_0 \tilde{\mathbf{x}}_0^\dagger \right) \right] \right\} \\
 &\stackrel{(c)}{=} \sigma_0^2 \vartheta \text{tr} \left\{ \mathbb{E} \left[\mathbf{R}_{\tilde{\mathbf{h}}}^\dagger \left(\overline{\tilde{\mathbf{x}}_0 \tilde{\mathbf{x}}_0^T} \odot \mathbf{R}_{\tilde{\mathbf{h}}} \odot \tilde{\mathbf{x}}_0 \tilde{\mathbf{x}}_0^\dagger \right) \right] \right\} \\
 &\stackrel{(d)}{=} \sigma_0^2 \vartheta \zeta \text{tr} \left\{ \mathbf{R}_{\tilde{\mathbf{h}}}^\dagger \left(\mathbf{R}_{\tilde{\mathbf{h}}} \odot \mathbb{E} \left[\overline{\tilde{\mathbf{x}}_0 \tilde{\mathbf{x}}_0^T} \odot \tilde{\mathbf{x}}_0 \tilde{\mathbf{x}}_0^\dagger \mid b = \sqrt{\frac{\beta PT}{N}} \right] \right) \right\} \\
 &\stackrel{(e)}{=} \sigma_0^2 \vartheta \zeta \left(\frac{\beta PT}{N} \right)^2 \text{tr} \left\{ \mathbf{R}_{\tilde{\mathbf{h}}}^\dagger \mathbf{R}_{\tilde{\mathbf{h}}} \right\}. \tag{8.70}
 \end{aligned}$$

Here, (a) is a consequence of Theorem 4.2.12 by Horn and Johnson (1991); (b) results because the matrices involved are all Hermitian, and (c) follows from the Hermitian product identity (Lütkepohl, 1996, p. 42)

$$\text{tr} \{ (\mathbf{A}^T \odot \mathbf{B}^T) \mathbf{C} \} = \text{tr} \{ \mathbf{A}^T (\mathbf{B} \odot \mathbf{C}) \}. \tag{8.71}$$

To obtain (d), we partially evaluate the expectation and use that the Hadamard product is commutative; finally, (e) holds because the entries of the matrix $\overline{\tilde{\mathbf{x}}_0 \tilde{\mathbf{x}}_0^T} \odot \tilde{\mathbf{x}}_0 \tilde{\mathbf{x}}_0^\dagger$ are all equal to $(\beta PT)^2 / N^2$ w.p.1 given that $b = \sqrt{\beta PT / N}$.

8 TECHNICAL RESULTS

The second trace term in (8.68) is

$$\begin{aligned}
 & \operatorname{tr} \left\{ \mathbb{E} \left[\sigma_0 \Lambda \otimes (\mathbf{R}_{\tilde{\mathbf{h}}} \odot \tilde{\mathbf{x}}_0 \tilde{\mathbf{x}}_0^\dagger) \right]^2 \right\} \\
 &= \sigma_0^2 \operatorname{tr} \left\{ \left(\Lambda \otimes \left(\mathbf{R}_{\tilde{\mathbf{h}}} \odot \mathbb{E} \left[\tilde{\mathbf{x}}_0 \tilde{\mathbf{x}}_0^\dagger \right] \right) \right)^2 \right\} \\
 &= \sigma_0^2 \operatorname{tr} \left\{ \left(\Lambda \otimes \left(\mathbf{R}_{\tilde{\mathbf{h}}} \odot \frac{\zeta \beta PT}{N} \mathbf{I}_{K'N'} \right) \right)^2 \right\} \\
 &\stackrel{(a)}{=} \sigma_0^2 \operatorname{tr} \left\{ \Lambda^2 \otimes \left(\frac{\zeta \beta PT}{N} \right)^2 \mathbf{I}_{K'N'} \right\} \\
 &= \sigma_0^2 \vartheta \left(\frac{\zeta \beta PT}{N} \right)^2 K'N', \tag{8.72}
 \end{aligned}$$

where (a) follows from (7.27).

We have thus shown that the asymptotic mutual information expression (8.68) simplifies to

$$\begin{aligned}
 I(\tilde{\mathbf{x}}_0; \tilde{\mathbf{y}}) &= \frac{1}{2} \vartheta \zeta \left(\frac{\sigma_0 \beta PT}{N} \right)^2 \\
 &\quad \times \left(\operatorname{tr} \{ \mathbf{R}_{\tilde{\mathbf{h}}}^\dagger \mathbf{R}_{\tilde{\mathbf{h}}} \} - \zeta K'N' \right) + o \left(\frac{1}{N^2} \right) \tag{8.73}
 \end{aligned}$$

which we can now substitute in (8.67) to compute a lower bound

on $\lim_{B \rightarrow \infty} BC(B)$. This lower bound is valid for any fixed K' and N' :

$$\begin{aligned}
 \lim_{B \rightarrow \infty} BC(B) &\geq \lim_{B \rightarrow \infty} \frac{B^2}{K'N'TF} I(\tilde{\mathbf{x}}_0; \tilde{\mathbf{y}}) \\
 &= \lim_{B \rightarrow \infty} \frac{B^2}{K'N'TF} \left[\frac{1}{2} \vartheta \zeta \left(\frac{\sigma_0 \beta PTF}{B} \right)^2 \right. \\
 &\quad \left. \left(\text{tr}\{R_{\mathbf{h}}^\dagger R_{\tilde{\mathbf{h}}}\} - \zeta K'N' \right) + o\left(\frac{1}{B^2}\right) \right] \\
 &= \frac{\vartheta}{2} TF (\sigma_0 \zeta \beta P)^2 \left(\frac{\text{tr}\{R_{\mathbf{h}}^\dagger R_{\tilde{\mathbf{h}}}\}}{\zeta K'N'} - 1 \right) \quad (8.74)
 \end{aligned}$$

To obtain a closed-form expression for (8.74), we take its limit for $K' \rightarrow \infty$ and $N' \rightarrow \infty$. Although this limit is no longer a valid lower bound itself because we require that K' and N' are fixed to apply the asymptotic expansion of mutual information by Prelov and Verdú (2004, Corollary 1), we can always choose K' and N' large enough but fixed to come arbitrarily close to this limit. To identify its value, we use a result by Tyrtyshnikov (1994, Section 4) and Capizzano (2001) about the asymptotic eigenvalue spectrum of the product of two-level Toeplitz matrices. This result generalizes the relation between the product of two Hermitian Toeplitz matrices and the product of corresponding asymptotically equivalent circulant matrices by Grenander and Szegő (1984, Chapter 7) (see also the simplified derivation by Gray 2005, Theorem 5.3).

$$\begin{aligned}
 \lim_{K', N' \rightarrow \infty} \frac{1}{K'N'} \text{tr}\{R_{\mathbf{h}}^\dagger R_{\tilde{\mathbf{h}}}\} &= \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} (\psi(\varphi, \zeta))^2 d\varphi d\zeta \\
 &= \frac{1}{TF} \underbrace{\iint_{\nu \tau} c_s^2(\nu, \tau) d\nu d\tau}_{\kappa_{\mathbb{H}}}. \quad (8.75)
 \end{aligned}$$

Thus, if we choose K' and N' large enough, we can come arbitrarily

8 TECHNICAL RESULTS

close to

$$\frac{\vartheta(\sigma_0\zeta\beta P)^2}{2} \left(\frac{\kappa_{\mathbb{H}}}{\zeta} - TF \right). \quad (8.76)$$

But for $\zeta = 1/\beta$ and $\zeta = \kappa_{\mathbb{H}}/(2TF)$, this limit equals the upper bound on the first-order Taylor series coefficient ϕ of U_1 in (8.61) for the two cases $\beta > 2TF/\kappa_{\mathbb{H}}$ and $\beta \leq 2TF/\kappa_{\mathbb{H}}$. Hence, the first order Taylor expansion of the lower bound (8.74) can be made to match the upper bound (8.61) as closely as desired.

CHAPTER 9

Conclusions

THE GAP between model and reality, how large is it? If we consider all the assumptions we made to get from the description of the wireless channel by means of Maxwell's equations to the discretized and approximately diagonalized underspread WSSUS model in Section 3.2, or even to the discrete-time block-fading model in Section 3.1.4, and if we also take into account the rather inconclusive empirical results on the joint tap distribution from Section 6.3, which in turn rest on measurements that required complicated postprocessing, we might easily conclude that the models used for communication-theoretic analysis are disconnected from reality. And in an absolute sense, if we just care about modeling accuracy, we might actually be right with this conclusion.

But maybe we are just asking the wrong question. "How important is the gap?" or "How fundamental is the gap?" are probably much more relevant questions to ask, because in the end the utility of a model is determined by the designs enabled by it. And to this end we hope to have contributed some aspects in this thesis, even if they are only minor aspects.

9.1. MODELING CONSEQUENCES

Motivated by our considerations in Chapter 2, we stated three important modeling questions in Section 4.3.1, about marginal tap distributions, tap correlation, and DOU scaling. From these three questions, we could empirically answer the one about marginal tap distributions in a satisfactory way, the one about the number of DOUs as a function of bandwidth positively only contingent on our measurements alone, but inconclusively in view of other reported results, and the question about tap correlation only to a certain degree—correlation is low, but seems to play a role nevertheless.

Maybe of equal importance is what we did not find: clear evidence *against* most of the fundamental features of the standard discrete-time block-fading model commonly used for the analysis of wideband wireless channels (Tse and Viswanath, 2005). With the advent of UWB communications, several researchers claimed that completely different models would be necessary, that the conventional design principles no longer held. The answers to these supposed challenges have been mainly specular block-fading models with various tap amplitude distributions and various cluster and ray arrival processes. While the findings that are undisputed in the literature, the exponentially decaying clustered PDP, the large number of channel taps, the reported delay spread and mean delay values for indoor scenarios, are also reflected in MCI and MCII, our results seem to imply that Rayleigh and Rice amplitude distributions are still adequate to model small-scale fading of UWB channels and, more importantly, that the JPG distribution for the entire channel impulse response is empirically at least equally well supported than other models. On the basis of these observations, we are confident that the conventional discrete-time wideband block-fading channel model reviewed in Section 3.1.4 is much better suited for mathematical analysis of ultrawideband communication system than the proposed alternatives, and that a time variant extension of this model like the GSCM by Kunisch and Pamp (2003), which adequately describes the evolution of specular reflections, is better suited for simulations than derivatives of the

SV model, like the IEEE 802.15.3a and IEEE 802.15.4a models, for the reasons detailed in Section 6.5.1.

Though we do not advance a radically different approach to wideband channel modeling, we find several interesting features that might be beneficially targeted in the design of future UWB systems, namely, the evolution of specular paths over space in application scenarios with mobile terminals and the strong mean component in every impulse response tap for channels between static terminals.

The basic empirical findings about tap distributions and DOU scaling hint at the possibility that Kozek's more general time- and frequency-selective fading model presented in Section 3.2 captures several fundamental features of wideband channels, so that our capacity analysis in Chapter 7 may be of practical relevance. Because of the connection between the discrete-time and continuous-time models, it is possible to evaluate our capacity bounds for scattering functions that are obtained from channel measurements and to use the results to optimize a communication system for specific channel characteristics.

9.2. DESIGN IMPLICATIONS

A model is only as good as the designs it enables. We argue above that our results support the conventional discrete-time JPC channel model for mathematical analysis, and that the more comprehensive discretized WSSUS model is also in line with our empirical findings. The upper and lower bounds from Chapter 7 may, therefore, provide important quantitative design guidelines for a channel with known scattering function, e.g., obtained by measurements. But even if the scattering function is not known completely and the channel is only characterized coarsely by its maximum delay τ_0 and maximum Doppler shift $\nu_0/2$, the bounds may serve as an efficient design tool. Maximum delay and maximum Doppler shift are sufficient to characterize a brick-shaped scattering function, and for such a scattering function the upper bound (7.20) and the approximate lower bound (7.53) are especially easy to evaluate. Furthermore, a brick-shaped scattering

function results in the worst-case capacity upper bound for a given τ_0 and ν_0 . Therefore, the widely accepted practice to characterize wireless channels simply by means of these two parameters leads to a robust design.

The numerical evaluation in Section 7.6 indicates that our bounds are surprisingly good over a large range of bandwidths for a set of practically relevant parameters. In particular, our lower bound is close to the coherent upper bound in Figure 7.1 for bandwidths up to 1 GHz. It is exactly this regime that is of most interest for current wideband and ultrawideband communication systems, and also the regime where we can expect the modeling assumptions in Chapter 2 to be most accurate. While our bounds are general, the numerical examples pertain only to certain parameter choices and do not allow for general statements. However, we use rather conservative parameters in the plots of Section 7.6; many systems might operate over channels with smaller spread and at higher SNR. And even for very different parameters, the bounds are appropriate to check if a given bandwidth still lies in the benign regime where capacity increases with DOFs and where the lower bound is close to the coherent upper bound.

The small difference between noncoherent lower bound and coherent upper bound for practically relevant bandwidth hints at the possibility that the standard coherent design methodology for coding and modulation schemes, in which perfect CSI is assumed at the receiver, might still be useful for UWB systems, and that well established receiver architectures, codes, training schemes, etc. might be scaled to larger bandwidth. To prove this claim rigorously, though, some additional work is required.

In addition to standard design principles, our results give several hints on how to improve a design to operate over wideband channels that are possibly time-variant.

- If both terminals can be expected to move little, the Ricean component in all channel taps should be made use of.
- If terminals are mobile, it might be worthwhile to track several

strong specular path.

- The numerical evaluation of our lower bound for QPSK modulation indicates that low-order modulation schemes may be close to optimal for a large range of bandwidth.
- The number of DOU increases with bandwidth and limits the performance if bandwidth gets too large. The most efficient way to combat this problem seems to make use of channel memory in space, time, and frequency. Many current wireless systems do not exploit channel memory over more than a single packet. But many indoor channels are highly underspread, i.e., have lots of memory, and it would be wasteful not to use it, even if only very coarsely to initialize a channel estimation algorithm or adaptive equalizer at the start of a new packet.

Many of these design guidelines have one important drawback, though. They require substantial computation. While the tremendous advances in silicon integration have enabled many computationally intensive receiver algorithms to be implemented for wideband systems of, say, 40 MHz bandwidth, they are still extremely challenging to build for bandwidths that are one or two orders of magnitude larger. But then, nobody envisioned in 1958 that the then refrigerator-sized rake receiver (Price and Green, 1958) is part of most wireless devices some 50 years later.

9.3. OPEN PROBLEMS

In this thesis, we tried to analyze the problem of wireless communications over wideband fading channels with a focus on fundamental yet practically relevant aspects. Our results shed light on some interesting questions, but they are by no means comprehensive. We briefly summarize some important, as we believe, open problems we came across.

The delay-scale representation: To derive the LTV IO relation used throughout the thesis, we require the small fractional bandwidth as-

sumption (2.7). Without it, we need to resort to the more complicated *delay-scale* IO representation (2.4). Two important problems are to analyze if and when the latter model can be beneficially used, and to develop mathematical tools with which the delay-scale model can be as efficiently handled as the delay-Doppler model.

Time-variant and multiantenna measurements: Our multiantenna extension of the basic SISO WSSUS channel model in Section 7.3 is not backed up empirically, mainly because truly time-variant multiantenna channel measurements are still difficult to perform.

Scaling of the number of degrees of uncertainty in different types of channels: We were not able to fundamentally explain the different scaling behavior of the DOUs in our MCH and the measurements by Menouni Hayar et al. (2005). As this question is of fundamental importance for our capacity results in Chapter 7, further empirical and fundamental theoretical investigations would be valuable.

Channel models for the design of synchronization algorithms: We aligned individual impulse response samples in Section 4.5 to obtain impulse responses as a receiver synchronized to the transmitter would perceive them. This implies that our model is not suited to design synchronization algorithms, as the synchronization step is already implicitly contained in the model. To the best of our knowledge, no dedicated channel models are available for the design of synchronizers.

Capacity results for different channel models: Our capacity analysis in Chapter 7 fundamentally requires that the channel is *underspread* and *stationary*. How does capacity behave if either assumption is violated? Matz (2003, 2005) provides some first promising results for non-WSSUS channels. Overspread channels can be found mainly in underwater acoustic communications. A promising approach to analyze their asymptotic capacity behavior might follow along the lines of Koch et al. (2007a,b).

Capacity results for a constraint on the transmitted PSD: Current UWB regulations do not constrain the average transmitted power as we do in Section 7.3.3 but limit the transmitted signal's PSD. Therefore, the total transmitted power is allowed to increase with increasing bandwidth, which results in a scaling behavior different from our results in Chapter 7. The analysis for a PSD constraint only follows along the same lines as in Chapter 7; however, the more interesting and more realistic case is to include the frequency dependent attenuation as observed by Kunisch and Pamp (2002), so that the received power does not scale with bandwidth in the same way than the transmit power. The corresponding channel model would no longer be US, though.

Signaling schemes that take advantage of the Ricean component: An important finding in Chapter 5 is the strong Ricean component in every impulse response tap in MCII. As already discussed in Section 6.5, a strong mean component in every tap might significantly increase capacity compared to purely Rayleigh fading. However, our capacity analysis in Chapter 7 requires a zero-mean JPC channel process. A constant mean can probably be incorporated as well, but a more realistic mean component that is frequency dependent breaks the US assumption; hence, different analysis techniques might be necessary.

Assessment of pilot-aided transmission: We hypothesized in Section 7.6 that the small gap between the noncoherent lower bound L_1 and the coherent upper bound U_c observed in Figure 7.1 for bandwidths up to 1 GHz means that not knowing the channel incurs only a small capacity penalty compared with the coherent setting, a statement theoretically verified by Médard (2000). A question of practical interest is if the commonly used pilot-aided transmission schemes in combination with codes designed for the coherent setting achieve rates comparable to our lower bound L_1 , or if a different design paradigm is necessary.

APPENDIX A

Notation

A.1. MISCELLANEOUS

$\mathcal{A}, \mathcal{B}, \dots$	sets
$ \mathcal{A} $	cardinality of the set \mathcal{A}
\mathcal{A}^c	complement of the set \mathcal{A} in some specified superset
$\mathcal{A} - \mathcal{B}$	set difference
\mathcal{R}, \mathcal{C}	real line, complex plane
\mathcal{Z}	set of all integers
\mathbb{F}	Fourier transform: $(\mathbb{F}_{t \rightarrow f} f)(f) \triangleq \int f(t)e^{-i2\pi ft} dt$
\mathbb{F}^{-1}	inverse Fourier transform: $(\mathbb{F}_{f \rightarrow t}^{-1} g)(t) \triangleq \int g(f)e^{i2\pi ft} df$
$\log x$	logarithm of x to the base e (natural logarithm)
$\delta(t)$	Dirac distribution
$\delta_{i,j}$	discrete Kronecker delta; takes on the value 1 if $i = j$ and the value 0 otherwise
\star	convolution: $(x \star y)(t) \triangleq \int_{\tau} x(\tau)\overline{y}(t - \tau)d\tau$
\Re, \Im	real, imaginary part of a complex-valued quantity
$\Gamma(x)$	gamma function
$I_0(x)$	modified Bessel function of the first kind and order 0
$f(x) = o(x)$	$f(x)/x \rightarrow 0$ for $x \rightarrow 0$
$\{x_n\}, \{x_n\}_{n=0}^{N-1}$	generic sequence, finite sequence with N elements
$\lceil x \rceil$	smallest integer equal to or larger than x

A.2. LINEAR ALGEBRA

$a, A, b, B \dots$	deterministic scalars
$\mathbf{a}, \mathbf{b}, \dots$	deterministic vectors
A, B, \dots	deterministic matrices
\mathbf{a}^T, A^T	transpose of the vector \mathbf{a} and the matrix A
$\bar{a}, \bar{\mathbf{a}}, \bar{A}$	element-wise complex conjugate of the scalar a , the vector \mathbf{a} , and the matrix A
$\mathbf{a}^\dagger, A^\dagger$	Hermitian transpose of the vector \mathbf{a} and the matrix A
$[A]_{i,j}$	element in the i th row and j th column of the matrix A
$I_N, 0_N$	$N \times N$ identity, $N \times N$ all-zero matrix
$\text{diag}(\mathbf{a})$	diagonal matrix with \mathbf{a} on its main diagonal
$\ A\ _F^2$	squared Frobenius norm of A : $\ A\ _F^2 = \sum_{i,j} [A]_{i,j} ^2$
$\text{tr } A, \det A$	trace, determinant of the square matrix A
$\text{rank } A$	rank of the matrix A
$A \otimes B$	Kronecker product of the matrices A and B
$A \odot B$	Schur-Hadamard product of the matrices A and B
$(\mathbf{a}\mathbf{a}^\dagger) \otimes B$	the standard matrix product takes precedence over
$= \mathbf{a}\mathbf{a}^\dagger \otimes B$	Kronecker and Hadamard products
$\langle \mathbf{a}, \mathbf{b} \rangle$	inner product of the vectors \mathbf{a} and \mathbf{b}
$\mathbf{a} \succ \mathbf{b}$	the vector \mathbf{a} majorizes the vector \mathbf{b}
λ, σ	eigenvalue, singular value
$\lambda_n(A)$	n th eigenvalue of the $N \times N$ matrix A , sorted in nonincreasing order and including multiplicities, i.e., $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_{N-1}$
$\lambda(A)$	vector of eigenvalues of A : $\lambda(A) = [\lambda_0(A) \ \lambda_1(A) \ \dots \ \lambda_{N-1}(A)]$
F_K	$K \times K$ discrete Fourier transform (DFT) matrix: $[F_K]_{m,n} = (1/\sqrt{K}) e^{-j2\pi mn/K}$, for $m = 0, 1, \dots, K-1$ and $n = 0, 1, \dots, K-1$
$\overline{\text{span}} \mathcal{A}$	closed linear span of the vectors in \mathcal{A}

A.3. PROBABILITY AND STATISTICS

$\mathbb{P}(\mathcal{A})$	probability of the event \mathcal{A}
$\mathbb{P}(\mathcal{A} \mathcal{B})$	conditional probability of the event \mathcal{A} given the event \mathcal{B}
$\mathbf{a}, \mathbf{b}, \dots$	random variables
$\mathbf{a}, \mathbf{b}, \dots$	random vectors
$\mathbf{A}, \mathbf{B}, \dots$	random matrices
$F_{\mathbf{q}}(a), f_{\mathbf{q}}(a)$	CDF, PDF of the random variable \mathbf{a} , parametrized by the parameter vector \mathbf{q}
$F(a b), f(a b)$	conditional CDF, PDF of \mathbf{a} given the realization $\mathbf{b} = b$
\mathcal{M}	set of all univariate CDFs
\mathcal{G}	candidate family of CDFs
\mathcal{J}	candidate set
$\mathbb{E}_F[\cdot]$	expectation operator (with respect to the CDF F)
$\boldsymbol{\mu}$	generic mean vector
\mathbf{R}, \mathbf{J}	generic covariance, pseudocovariance matrix
$\psi(\varphi, \zeta)$	bivariate spectral density
$\boldsymbol{\Psi}(\varphi)$	matrix-valued spectral density
ρ_{ij}	normalized correlation coefficient
$\mathcal{CN}(\boldsymbol{\mu}, \mathbf{R})$	multivariate jointly proper Gaussian (JPG) distribution with mean $\boldsymbol{\mu}$ and covariance matrix \mathbf{R}
$\hat{\mathbf{q}}$	estimate of the parameter vector \mathbf{q}
$\mathcal{H}_i, \mathcal{H}_c$	incumbent, challenging hypothesis
$\Delta(\cdot, \cdot)$	discrepancy between two probability distributions
$\Delta_N(\cdot, \cdot)$	empirical discrepancy between two probability distributions, computed from N samples
AIC	value of Akaike's Information Criterion (AIC)
AICC	value of the small-sample corrected version of AIC
ϕ_j	difference between the AIC-best distribution and the fitted distribution j
w_j	Akaike weight of the fitted distribution j

A.4. COMMUNICATION AND INFORMATION THEORY

N_0	spectral density of the thermal noise
P	average power
β	peak- to average-power ratio (PAPR)
E_b	energy per bit
SNR	signal to noise ratio
f_c	carrier frequency
K	Ricean K-factor
M_T, M_R	number of transmit, receive antennas
A, B	transmit, receive spatial correlation matrix
$\mathbf{x}, \mathbf{x}, \mathbf{x}, \mathbf{X}$	(random) channel input
\mathbf{r}	noise-free (random) channel output
$\mathbf{y}, \mathbf{y}, \mathbf{y}$	noisy channel output
$\mathbf{h}, \mathbf{h}, \mathbf{H}$	(random) channel
$\mathbf{w}, \mathbf{w}, \mathbf{w}$	white noise
D	duration
B	bandwidth
ρ	(random) path gain
μ_τ, σ_τ	mean delay, delay spread
$h(\mathbf{x})$	differential entropy of the random vector \mathbf{x}
$I(\mathbf{x}; \mathbf{y})$	mutual information between the random vectors \mathbf{x} and \mathbf{y}
\mathcal{P}, \mathcal{Q}	sets of input distributions
$D(f g)$	relative entropy between the probability densities f and g , also called Kullback-Leibler (KL) distance or divergence
C, C_ϵ	channel capacity, ϵ -outage capacity
$L_1, L_2,$	lower bounds on capacity
U_1, U_c	upper bound on capacity, coherent capacity

A.5. LINEAR TIME-VARIANT SYSTEMS

\mathcal{L}^2	Hilbert space of square-integrable functions
$t, \Delta t, k$	continuous time, time difference, discrete time
τ, l	continuous, discrete delay
τ_0	maximum delay
$f, \Delta f, n$	continuous frequency, frequency difference, discrete frequency
ν, m	continuous, discrete Doppler frequency
ν_0	maximum Doppler shift
T, F	time, frequency grid spacing
$\Delta_{\mathbb{H}}, \tilde{\Delta}_{\mathbb{H}}$	spread, effective spread
\mathbb{H}	Hilbert-Schmidt (HS) channel operator
$\mathcal{I}_{\mathbb{H}}, \mathcal{R}_{\mathbb{H}}, \mathcal{N}_{\mathbb{H}}$	channel input, output, null-space
$\mathbf{k}_{\mathbb{H}}(t, t')$	(random) kernel of the HS operator \mathbb{H}
$\mathbf{h}_{\mathbb{H}}(t, \tau)$	(random) time-variant impulse response
$\mathbf{l}_{\mathbb{H}}(t, f)$	(random) time-variant transfer function
$\mathbf{s}_{\mathbb{H}}(\nu, \tau)$	(random) spreading function
$\mathbf{b}_{\mathbb{H}}(\nu, f)$	(random) bi-frequency function
$\mathbf{s}_{\mathbb{S}}(\nu, \tau)$	specular spreading function
$c_{\mathbb{S}}(a, b)$	wideband spreading function
$c_{\mathbf{h}}(\Delta t, \tau)$	WSSUS impulse response correlation function
$c_{\mathbf{l}}(\Delta t, \Delta f)$	WSSUS time-frequency correlation function
$c_{\mathbf{s}}(\nu, \tau)$	WSSUS scattering function
$p_{\mathbf{s}}(\tau)$	power-delay profile (PDP)
$u_i(t), v_i(t')$	left and right singular functions of \mathbb{H}
$g(t)$	Weyl-Heisenberg (WH) prototype pulse
$A_g(\nu, \tau)$	ambiguity function of $g(t)$

APPENDIX B

Acronyms

ADC	analog-to-digital converter
AIC	Akaike's Information Criterion
AICC	small-sample corrected AIC
AOA	angle of arrival
AOD	angle of departure
AR	autoregressive
AWGN	additive white Gaussian noise
BIC	Bayes information criterion
CLT	central limit theorem
CDF	cumulative distribution function
CP	cyclic prefix
CSI	instantaneous channel state information
DFT	discrete Fourier transform
DOF	degrees of freedom
DOU	degrees of uncertainty
DSO	digital sampling oscilloscope
FH	frequency hopping
GO	geometrical optics
GOF	goodness of fit
GSCM	geometry-based stochastic channel model
GTD	geometrical theory of diffraction
HS	Hilbert-Schmidt
ICI	intercarrier interference

B ACRONYMS

IDFT	inverse discrete Fourier transform
IEEE	Institute of Electrical and Electronics Engineers
IF	intermediate frequency
IO	input-output
IR	impulse radio
ISI	intersymbol interference
JPG	jointly proper Gaussian
KL	Kullback-Leibler
LFI	linear frequency-invariant
LOS	line of sight
LTI	linear time-invariant
LTV	linear time-variant
MCI	measurement campaign I
MCII	measurement campaign II
MCM	Monte-Carlo method
MDL	criterion of minimum description length
MIMO	multiple input multiple output
ML	maximum likelihood
MMSE	minimum mean squared error
MUSIC	multiple signal classification
NLOS	non line of sight
OFDM	orthogonal frequency division multiplexing
OLOS	obstructed line of sight
PAPR	peak- to average-power ratio
PDF	probability density function
PDP	power-delay profile
PN	pseudonoise
PSD	power spectral density
PSK	phase-shift keying
PS-OFDM	pulse-shaped orthogonal frequency division multiplexing
QPSK	quaternary phase-shift keying
RF	radio frequency
RMS	root mean square
SAGE	space alternating generalized expectation maximization
SIMO	single input multiple output

SISO	single input single output
SNR	signal-to-noise ratio
SV	Saleh-Valenzuela
SVD	singular value decomposition
TDMA	time division multiple access
TF	time-frequency
US	uncorrelated scattering
USB	universal serial bus
UTD	uniform theory of diffraction
UWB	ultrawideband
VNA	vector network analyzer
WH	Weyl-Heisenberg
WLAN	wireless local area network
WPAN	wireless personal area network
w.p.1	with probability 1
WSS	wide-sense stationary
WSSUS	wide-sense stationary uncorrelated scattering

References

- Abbas, S. A. and Sheikh, A. U. (1997), “On understanding the nature of slow fading in LOS microcellular channels,” in *Proc. IEEE Veh. Technol. Conf. (VTC)*, vol. 2, pp. 662–666, Phoenix, AZ, U.S.A.
- Abou-Faycal, I. C., Trott, M. D., and Shamai (Shitz), S. (2001), “The capacity of discrete-time memoryless Rayleigh-fading channels,” *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1301.
- Abouraddy, A. F. and Elnoubi, S. M. (2000), “Statistical modeling of the indoor radio channel at 10 GHz through propagation measurements—Part I: Narrow-band measurements and modeling,” *IEEE Trans. Veh. Technol.*, vol. 49, no. 5, pp. 1491–1507.
- Akaike, H. (1973), “Information theory and an extension of the maximum likelihood principle,” in B. N. Petrov and F. Csaki (Eds.), *Proc. Int. Symp. Inf. Theory*, pp. 267–281, Budapest, Hungary: Akademiai Kiado. Reproduced in *Breakthroughs in Statistics* vol. 1, S. Kotz and N. L. Johnson, eds., New York, NY, U.S.A.: Springer, 1992.
- (1978), “On the likelihood of a time series model,” *The Statistician*, vol. 27, no. 3/4, pp. 217–235.
- Anderson, G. A. (1965), “An asymptotic expansion for the distribution of the latent roots of the estimated covariance matrix,” *Ann. Math. Statist.*, vol. 36, no. 4, pp. 1153–1173.
- Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis*, Hoboken, NJ, U.S.A.: Wiley, 3rd ed.
- Artés, H., Matz, G., and Hlawatsch, F. (1998), “Linear time-varying channels,” Tech. Rep. 98-06, Department of Communications and Radio-Frequency Engineering, Vienna University of Technology, Vienna, Austria.
- (2004), “Unbiased scattering function estimators for underspread channels and extension to data-driven operation,” *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1387–1402.

REFERENCES

- Aulin, T. (1979), "A modified model for the fading signal at a mobile radio channel," *IEEE Trans. Veh. Technol.*, vol. 28, no. 3, pp. 182–203.
- Auslander, L. and Tolimieri, R. (1985), "Radar ambiguity functions and group theory," *SIAM J. Math. Anal.*, vol. 16, no. 3, pp. 577–601.
- Babich, F. and Lombardi, G. (2000), "Statistical analysis and characterization of the indoor propagation channel," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 455–464.
- Bajwa, A. S. and Parsons, J. D. (1978), "Time-delay/Doppler scattering function for suburban mobile radio propagation at 436 MHz," *Electron. Lett.*, vol. 14, pp. 423–425.
- Balakrishnan, A. V. (1962), "On the problem of time jitter in sampling," *IEEE Trans. Inf. Theory*, vol. 8, no. 3, pp. 2226–236.
- Balakrishnan, K., Wee, K. C., Xu, S., Chuan, C. L., Chin, F., Huat, C. Y., Choy, C. C., Thiang, T. T., Xiaoming, P., Ong, M., and Krishnan, S. (2004), "Characterization of ultra-wideband channels: Small-scale parameters for indoor & outdoor office environments," Presentation IEEE P802.15-04-0342-00-004a, Institute for Infocomm Research, Singapore. IEEE P802.15 SG4a contribution.
- Bapat, R. B. and Sunder, V. S. (1985), "On majorization and Schur products," *Lin. Algeb. Appl.*, vol. 72, pp. 107–117.
- Bartoszyński, R. and Niewiadomska-Bugaj, M. (1996), *Probability and Statistical Inference*, New York, NY, U.S.A.: Wiley.
- Bello, P. A. (1963), "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun.*, vol. 11, no. 4, pp. 360–393.
- (1969), "Measurement of random time-variant linear channels," *IEEE Trans. Inf. Theory*, vol. 15, no. 4, pp. 469–475.
- Bertsekas, D. and Gallager, R. G. (1996), *Data Networks*, New Delhi, India: Prentice Hall, 2nd ed.
- Biglieri, E. (2005), *Coding for Wireless Channels*, New York, NY, U.S.A.: Springer.
- Biglieri, E., Proakis, J., and Shamai (Shitz), S. (1998), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692.
- Blahut, R. E., Miller, Jr., W., and Wilcox, C. H. (1991), *Radar and Sonar*, vol. 1, New York, NY, U.S.A.: Springer.
- Bölcskei, H., Gesbert, D., and Paulraj, A. J. (2002a), "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 225–234.
- Bölcskei, H., Koetter, R., and Mallik, S. (2002b), "Coding and modulation for underspread fading channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, p. 358, Lausanne, Switzerland.

- Borgmann, M. and Bölcskei, H. (2005), “On the capacity of noncoherent wideband MIMO-OFDM systems,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 651–655, Adelaide, Australia.
- Braun, W. R. and Dersch, U. (1991), “A physical mobile radio channel model,” *IEEE Trans. Veh. Technol.*, vol. 40, no. 2, pp. 472–482.
- Brown, W. M. (1963), “Sampling with random jitter,” *J. Soc. Industrial Appl. Math.*, vol. 11, no. 2, pp. 460–473.
- Buehrer, R. M., Davis, W. A., Safaai-Jazi, A., and Sweeney, D. (2003), “Characterization of the ultra-wideband channel,” in *IEEE Conf. Ultra Wideband Syst. Technol. (UWBST) Dig. Tech. Papers*, pp. 26–31, Reston, VA, U.S.A.
- Bultitude, R. J. C., Mahmoud, S. A., and Sullivan, W. A. (1989), “A comparison of indoor radio propagation characteristics at 910 MHz and 1.75 GHz,” *IEEE J. Sel. Areas Commun.*, vol. 7, no. 1, pp. 20–30.
- Burnham, K. P. and Anderson, D. R. (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, New York, NY, U.S.A.: Springer, 2nd ed.
- Capizzano, S. S. (2001), “Distribution results on the algebra generated by Toeplitz sequences: A finite-dimensional approach,” *Lin. Algeb. Appl.*, vol. 328, no. 1–3, pp. 121–130.
- Cassoli, D., Win, M. Z., and Molisch, A. F. (2002), “The ultra-wide bandwidth indoor channel: From statistical models to simulations,” *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1247–1257.
- Chan, R. H., Jin, X.-Q., and Yeung, M.-C. (1991), “The circulant operator in the Banach algebra of matrices,” *Lin. Algeb. Appl.*, vol. 149, pp. 41–53.
- Chehri, A., Fortier, P., and Tardif, P.-M. (2006), “Measurements and modeling of line-of-sight UWB channel in underground mines,” in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, pp. 1–5, San Francisco, CA, U.S.A.
- Chen, J. and Veeravalli, V. V. (2007), “Capacity results for block-stationary Gaussian fading channels with a peak power constraint,” *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4498–4520.
- Cheng, J. and Beaulieu, N. C. (2001), “Maximum-likelihood based estimation of the Nakagami m parameter,” *IEEE Commun. Lett.*, vol. 5, no. 3, pp. 101–103.
- Chong, C.-C. and Yong, S. K. (2005), “A generic statistical-based UWB model for high-rise apartments,” *IEEE Trans. Antennas Propag.*, vol. 53, no. 8, pp. 2389–2399.
- Christensen, O. (2003), *An Introduction to Frames and Riesz Bases*, Boston, MA, U.S.A.: Birkhäuser.

REFERENCES

- Chuah, C.-N., Tse, D. N. C., Kahn, J. M., and Valenzuela, R. A. (2002), "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 637–650.
- Chuang, J., Xin, N., Huang, H., Chiu, S., and Michelson, D. G. (2007), "UWB radiowave propagation within the passenger cabin of a Boeing 737-200 aircraft," in *Proc. IEEE Veh. Technol. Conf. Spring (VTC-Spring)*, pp. 496–500, Dublin, Ireland.
- Ciccognani, W., Durantini, A., and Cassioli, D. (2005), "Time domain propagation measurements of the UWB indoor channel using PN-sequence in the FCC-compliant band 3.6–6 GHz," *IEEE Trans. Antennas Propag.*, vol. 53, no. 4, pp. 1542–1549.
- Clarke, R. H. (1968), "A statistical theory of mobile-radio reception," *Bell Syst. Tech. J.*, vol. 47, pp. 957–1000.
- Clergeot, H. (1984), "Filter-order selection in adaptive maximum likelihood estimation," *IEEE Trans. Inf. Theory*, vol. 30, no. 2, pp. 199–210.
- Cohen, L. (1993), "The scale representation," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3275–3292.
- Cover, T. M. and Thomas, J. A. (1991), *Elements of Information Theory*, New York, NY, U.S.A.: Wiley.
- Cox, D. C. (1972), "Delay Doppler characterization of multipath propagation at 910 MHz in a suburban radio environment," *IEEE Trans. Antennas Propag.*, vol. 20, no. 5, pp. 625–635.
- (1973a), "910 MHz urban mobile radio propagation: Multipath characteristics in New York City," *IEEE Trans. Commun.*, vol. 21, no. 11, pp. 1188–1194.
- (1973b), "A measured delay-Doppler scattering function for multipath propagation at 910 MHz in an urban mobile radio environment," *Proc. IEEE*, vol. 61, no. 4, pp. 479–480.
- Cramer, R. J.-M., Scholtz, R. A., and Win, M. Z. (1999), "Spatio-temporal diversity in ultra-wideband radio," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, vol. 2, pp. 888–892, Los Angeles, CA, U.S.A.
- (2002), "Evaluation of an ultra-wide-band propagation channel," *IEEE Trans. Antennas Propag.*, vol. 50, no. 5, pp. 561–570.
- Cullen, P. J., Fannin, P. C., and Molina, A. (1993), "Wideband measurement and analysis techniques for the mobile radio channel," *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 589–603.
- Dabin, J. A., Ni, N., Haimovich, A. M., Niver, E., and Grebel, H. (2003), "The effects of antenna directivity on path loss and multipath propagation in UWB indoor wireless channels," in *IEEE Conf. Ultra Wideband Syst. Technol. (UWBST) Dig. Tech. Papers*, pp. 305–309, Reston, VA, U.S.A.

- Debbah, M. and Müller, R. R. (2005), “MIMO channel modeling and the principle of maximum entropy,” *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1667–1690.
- Demir, A. (2002), “Phase noise and timing jitter in oscillators with colored-noise sources,” *IEEE Trans. Circuits Syst. I*, vol. 49, no. 12, pp. 1782–1791.
- Dersch, U. and Zollinger, E. (1994), “Propagation mechanisms in microcell and indoor environments,” *IEEE Trans. Veh. Technol.*, vol. 43, no. 4, pp. 1058–1066.
- Devroye, L. (1987), *A Course in Density Estimation*, Boston, MA, U.S.A.: Birkhäuser.
- Devroye, L. and Györfi, L. (1985), *Nonparametric Density Estimation*, New York, NY, U.S.A.: Wiley.
- Diggavi, S. N. and Cover, T. M. (2001), “The worst additive noise under a covariance constraint,” *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 3072–3081.
- Dixon, W. J. and Massey, Jr., F. J. (1983), *Introduction to Statistical Analysis*, New York, NY, U.S.A.: McGraw-Hill, 4th ed.
- Dunford, N. and Schwarz, J. T. (1963), *Linear Operators*, vol. 2, New York, NY, U.S.A.: Wiley.
- Durisi, G., Bölcskei, H., and Shamai (Shitz), S. (2006), “Capacity of underspread WSSUS fading channels in the wideband regime,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1500–1504, Seattle, WA, U.S.A.
- (2007), “Capacity of underspread WSSUS fading channels under peak signal constraints,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 156–160, Nice, France.
- Durisi, G., Schuster, U. G., Bölcskei, H., and Shamai (Shitz), S. (2008), “Capacity of underspread WSSUS fading channels in the wideband regime under peak constraints,” *IEEE Trans. Inf. Theory*, Available: <http://arxiv.org/abs/0804.1748>. Submitted for review.
- Efron, B. and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, New York, NY, U.S.A.: Chapman & Hall.
- Eggen, T. H., Baggerier, A. B., and Preisig, J. C. (2000), “Communication over Doppler spread channels—part I: Channel and receiver presentation,” *IEEE J. Ocean. Eng.*, vol. 25, no. 1, pp. 62–71.
- Fang, K.-T., Kotz, S., and Ng, K. W. (1990), *Symmetric Multivariate and Related Distributions*, London, U.K.: Chapman & Hall.
- Fannin, P. C., Molina, A., Swords, S. S., and Cullen, P. J. (1991), “Digital signal processing techniques applied to mobile radio channel sounding,”

REFERENCES

- IEEE Proc. F, Commun., Radar, Signal Proc.*, vol. 138, no. 5, pp. 502–508.
- Feller, W. (1971), *An Introduction to Probability Theory and its Applications*, vol. 2, New York, NY, U.S.A.: Wiley, 2nd ed.
- Fessler, J. A. and Hero, A. O. (1994), “Space-alternating generalized expectation-maximization algorithm,” *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2664–2677.
- Fleury, B. H., Tschudin, M., Heddergott, R., Dahlhaus, D., and Pedersen, K. I. (1999), “Channel parameter estimation in mobile radio environments using the SAGE algorithm,” *IEEE J. Sel. Areas Commun.*, vol. 17, no. 3, pp. 434–450.
- Foerster, J. R. (2003), “Channel modeling sub-committee report final,” Tech. Rep. P802.15 02/490r1, IEEE 802.15 SG3a.
- Foerster, J. R. and Li, Q. (2002a), “UWB channel modeling contribution from Intel,” Tech. Rep. P802.15 02/279SG3a, Intel Corporation, Hillsboro, OR, U.S.A. IEEE P802.15 SG3a contribution.
- (2002b), “UWB channel modeling contribution from Intel,” Tech. Rep. P802.15 02/287SG3a, Intel Corporation, Hillsboro, OR, U.S.A. IEEE P802.15 SG3a contribution.
- Forster, M. and Sober, E. (1994), “How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions,” *British J. for the Philosophy of Science*, vol. 45, no. 1, pp. 1–35.
- Fort, A., Ryckaert, J., Desset, C., De Doncker, P., Wambacq, P., and Van Biesen, L. (2006), “Ultra-wideband channel model for communication around the human body,” *IEEE J. Sel. Areas Commun.*, vol. 24, no. 4, pp. 927–933.
- Foschini, G. J. and Gans, M. J. (1998), “On limits of wireless communications in a fading environment when using multiple antennas,” *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335.
- Friis, H. T. (1946), “A note on a simple transmission formula,” *Proc. IRE*, vol. 34, no. 5, pp. 254–256.
- Gaarder, N. T. (1968), “Scattering function estimation,” *IEEE Trans. Inf. Theory*, vol. 14, no. 5, pp. 684–693.
- Gallager, R. G. (1968), *Information Theory and Reliable Communication*, New York, NY, U.S.A.: Wiley.
- Gans, M. J. (1972), “A power-spectral theory of propagation in the mobile-radio environment,” *IEEE Trans. Veh. Technol.*, vol. 21, no. 1, pp. 27–38.
- Gärtner, M. E. and Bölcskei, H. (2007), “On the ‘critical rate’ in Ricean MIMO channels,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 526–530, Nice, France.

- Geng, S., Ranvier, S., Zhao, X., Kivinen, J., and Vainikainen, P. (2005), "Multipath propagation characterization of ultra-wide band indoor radio channels," in *Proc. IEEE Int. Conf. Ultra-Wideband (ICU)*, pp. 11–15, Zurich, Switzerland.
- Geng, S. and Vainikainen, P. (2006), "Frequency and bandwidth dependency of UWB propagation channels," in *Proc. IEEE Int. Symp. Personal, Indoor, Mobile Radio Commun. (PIMRC)*, pp. 1–5, Helsinki, Finland.
- Ghassemzadeh, S. S., Jana, R., Rice, C. W., Turin, W., and Tarokh, V. (2002), "A statistical path loss model for in-home UWB channels," in *IEEE Conf. Ultra Wideband Syst. Technol. (UWBST) Dig. Tech. Papers*, pp. 59–64, Baltimore, MD, U.S.A.
- (2004), "Measurement and modeling of an ultra-wide bandwidth indoor channel," *IEEE Trans. Commun.*, vol. 52, no. 10, pp. 1786–1796.
- Giannakis, G. B. and Tepedelenlioglu, C. (1998), "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. IEEE*, vol. 86, no. 10, pp. 1969–1986.
- Goldsmith, A., Jafar, S. A., Jindal, N., and Vishwanath, S. (2003), "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702.
- Goldsmith, A. and Médard, M. (2007), "Capacity of time-varying channels with causal channel side information," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 881–889.
- Goodman, N. R. (1963), "Statistical analysis on a certain multivariate complex Gaussian distribution," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 152–177.
- Gould, S. J. (1981), *The Mismeasure of Man*, New York, NY, U.S.A.: Norton.
- Gray, P. R., Hurst, P. J., Lewis, S. H., and Meyer, R. G. (2001), *Analysis and Design of Analog Integrated Circuits*, New York, NY, U.S.A.: Wiley, 4th ed.
- Gray, R. M. (2005), "Toeplitz and circulant matrices: A review," in *Foundations and Trends in Communications and Information Theory*, vol. 2, Delft, The Netherlands: now Publishers.
- (2007), *Entropy and Information Theory*, New York, NY, U.S.A.: Springer, revised ed., Available: <http://ee.stanford.edu/~gray/it.pdf>.
- Greenwood, J. A. and Durand, D. (1960), "Aids for fitting the gamma distribution by maximum likelihood," *Technometrics*, vol. 2, no. 1, pp. 55–66.
- Grenander, U. and Szegő, G. (1984), *Toeplitz Forms and Their Applications*, New York, NY, U.S.A.: Chelsea Publishing.

REFERENCES

- Gröchenig, K. (2001), *Foundations of Time-Frequency Analysis*, Boston, MA, U.S.A.: Birkhäuser.
- Guo, D., Shamai (Shitz), S., and Verdú, S. (2005), “Mutual information and minimum mean-square error in Gaussian channels,” *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282.
- Guttman, L. (1977), “What is not what in statistics,” *The Statistician*, vol. 26, no. 2, pp. 81–107.
- Haff, L. R. (1980), “Empirical Bayes estimation of the multivariate normal covariance matrix,” *Ann. Statist.*, vol. 8, no. 3, pp. 586–897.
- Hajek, B. and Subramanian, V. G. (2002), “Capacity and reliability function for small peak signal constraints,” *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 828–839.
- Handforth, M., Sheikh, A. U., and Abdi, M. (1993), “A statistical characterization of the indoor mobile radio channel at 946 MHz,” in *Conf. Record Int. Conf. Univ. Personal Commun.*, vol. 2, pp. 955–959, Ottawa, Ont., Canada.
- Haneda, K. and Takada, J. (2003), “An application of SAGE algorithm for UWB propagation channel estimation,” in *IEEE Conf. Ultra Wideband Syst. Technol. (UWBST) Dig. Tech. Papers*, pp. 483–487, Reston, VA, U.S.A.
- Haneda, K., Takada, J., and Kobayashi, T. (2004), “Experimental evaluation of a SAGE algorithm for ultra wideband channel sounding in an anechoic chamber,” in *Proc. Int. Workshop Ultra-Wideband Syst. (IWUWBS)*, pp. 66–70, Kyoto, Japan.
- (2006), “Cluster properties investigated from a series of ultrawideband double directional propagation measurements in home environments,” *IEEE Trans. Antennas Propag.*, vol. 54, no. 12, pp. 3778–3788.
- Hansen, M. H. and Yu, B. (2001), “Model selection and the principle of minimum description length,” *J. Am. Statist. Assoc.*, vol. 96, no. 454, pp. 746–774.
- Hashemi, H. (1993a), “Impulse response modeling of indoor radio propagation channels,” *IEEE J. Sel. Areas Commun.*, vol. 11, no. 7, pp. 967–978.
- (1993b), “The indoor radio propagation channel,” *Proc. IEEE*, vol. 81, no. 7, pp. 943–968.
- Hashemi, H., McGuire, M., Vlasschaert, T., and Tholl, D. (1994), “Measurement and modeling of temporal variations of the indoor radio propagation channel,” *IEEE Trans. Veh. Technol.*, vol. 43, no. 3, pp. 733–737.
- Healey, A., Bianchi, C. H., and Sivaprasad, K. (2000), “Wideband outdoor

- channel sounding at 2.4 GHz,” in *Proc. Conf. Antennas and Propagat. for Wireless Commun.*, pp. 95–98, Waltham, MA, U.S.A.
- Heil, C. E. and Walnut, D. F. (1989), “Continuous and discrete wavelet transforms,” *SIAM Rev.*, vol. 31, no. 4, pp. 628–666.
- Helstrom, C. W. (1967), “Image restoration by the method of least squares,” *J. Opt. Soc. Am.*, vol. 57, pp. 297–303.
- Hentilä, L., Taparungssanagorn, A., Viittala, H., and Hämäläinen, M. (2005), “Measurement and modelling of an UWB channel at hospital,” in *Proc. IEEE Int. Conf. Ultra-Wideband (ICU)*, pp. 113–117, Zurich, Switzerland.
- Hirt, W. and Massey, J. L. (1988), “Capacity of the discrete-time Gaussian channel with intersymbol interference,” *IEEE Trans. Inf. Theory*, vol. 34, no. 3, pp. 380–388.
- Hlawatsch, F. and Boudreaux-Bartels, G. F. (1992), “Linear and quadratic time-frequency signal representations,” *IEEE Signal Process. Mag.*, vol. 9, no. 2, pp. 21–67.
- Hoff, H., Eggers, P. C. F., and Kovács, I. Z. (2003), “Directional indoor ultra wideband propagation mechanisms,” in *Proc. IEEE Veh. Technol. Conf. Fall (VTC-Fall)*, vol. 1, pp. 188–192, Orlando, FL, U.S.A.
- Horn, R. A. and Johnson, C. R. (1985), *Matrix Analysis*, Cambridge, U.K.: Cambridge Univ. Press.
- (1991), *Topics in Matrix Analysis*, Cambridge, U.K.: Cambridge Univ. Press.
- Hovinen, V., Hämäläinen, M., and Pätsi, T. (2002), “Ultra wideband indoor radio channel models: Preliminary results,” in *IEEE Conf. Ultra Wideband Syst. Technol. (UWBST) Dig. Tech. Papers*, pp. 75–79, Baltimore, MD, U.S.A.
- Howard, S. J. and Pahlavan, K. (1992), “Autoregressive modeling of wideband indoor radio propagation,” *IEEE Trans. Commun.*, vol. 40, no. 9, pp. 1540–1552.
- Hurvich, C. M. and Tsai, C.-L. (1989), “Regression and time series model selection in small samples,” *Biometrika*, vol. 76, no. 2, pp. 297–307.
- IEEE Vehicular Technology Society Committee on Radio Propagation (1988), “Coverage prediction for mobile radio systems operating in the 800/900 MHz frequency range,” *IEEE Trans. Veh. Technol.*, vol. 37, no. 1, pp. 3–72.
- Intarapanich, A., Kafle, P. L., Davies, R. J., and Sesay, A. B. (2004), “Effect of tap gain correlation on capacity of OFDM MIMO systems,” *Electron. Lett.*, vol. 40, no. 1, pp. 86–88.
- Irahauten, Z., Dacuña, J., Janssen, G. J. M., and Nikoogar, H. (2005), “UWB channel measurements and results for wireless personal area net-

REFERENCES

- works applications,” in *Proc. Europ. Conf. Wireless Technol. (ECWT)*, pp. 189–192, Paris, France.
- Irahauten, Z., Janssen, G. J. M., Nikookar, H., Yarovoy, A., and Lighthart, L. P. (2006), “UWB channel measurements and results for office and industrial environments,” in *Proc. IEEE Int. Conf. Ultra-Wideband (ICU)*, pp. 225–230, Waltham, MA, U.S.A.
- Jacobs, I. M. (1963), “The asymptotic behavior of incoherent M-ary communication systems,” *Proc. IEEE*, vol. 51, no. 1, pp. 251–252.
- Jafar, S. A. and Goldsmith, A. (2005), “Multiple-antenna capacity in correlated Rayleigh fading with channel covariance information,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 990–997.
- Jakes, W. C. (Ed.) (1974), *Microwave Mobile Communications*, New York, NY, U.S.A.: Wiley.
- Jämsä, T., Hovinen, V., Karjalainen, A., and Iinatti, J. H. J. (2006), “Frequency dependency of delay spread and path loss in indoor ultra-wideband channels,” in *Proc. IET Sem. Ultra Wideband Syst. Technol. Appl.*, pp. 254–258, London, U.K.
- Jaturatussanai, P., Chamchoy, M., and Promwong, S. (2006), “Characteristics of UWB radio propagation with building materials,” in *Proc. Int. Symp. Commun. Inform. Technol.*, pp. 1225–1228, Bangkok, Thailand.
- Jeruchim, M. C., Balaban, P., and Shanmugan, K. S. (2000), *Simulation of Communication Systems*, Kluwer, 2nd ed.
- Jiang, Y. and Papandreou-Suppappola, A. (2006), “Discrete time-scale characterization of wideband time-varying systems,” *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1364–1375.
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995), *Continuous Univariate Distributions*, vol. 2, New York, NY, U.S.A.: Wiley, 2nd ed.
- Jorswieck, E. A. and Boche, H. (2006), “Performance analysis of MIMO systems in spatially correlated fading using matrix-monotone functions,” *IEICE Trans. Fund. Elec. Commun. Comp. Sc.*, vol. E89-A, no. 5, pp. 1454–1482.
- Kailath, T. (1959), “Sampling models for linear time-variant filters,” Tech. Rep. 352, Massachusetts Institute of Technology, Research Laboratory of Electronics, Cambridge, MA, U.S.A.
- (1963), “Time-variant communication channels,” *IEEE Trans. Inf. Theory*, vol. 9, no. 4, pp. 233–237.
- Karedal, J., Wyne, S., Almers, P., Tufvesson, F., and Molisch, A. F. (2004a), “Statistical analysis of the UWB channel in an industrial environment,” in *Proc. IEEE Veh. Technol. Conf. Fall (VTC-Fall)*, vol. 1, pp. 81–85, Los Angeles, CA, U.S.A.
- (2004b), “UWB channel measurements in an industrial environment,”

- in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, vol. 6, pp. 3511–3516, San Francisco, CA, U.S.A.
- Keignart, J., Abou-Rjeily, C., Delaveaud, C., and Daniele, N. (2006), “UWB SIMO channel measurements and simulations,” *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 4, pp. 1812–1819.
- Keignart, J. and Daniele, N. (2002), “Subnanosecond UWB channel sounding in frequency and temporal domain,” in *IEEE Conf. Ultra Wideband Syst. Technol. (UWBST) Dig. Tech. Papers*, pp. 25–30, Baltimore, MD, U.S.A.
- (2003), “Channel sounding and modelling for indoor UWB communications,” in *Proc. Int. Workshop Ultra-Wideband Syst. (IWUWBS)*, Oulu, Finland.
- Kennedy, R. S. (1969), *Fading Dispersive Communication Channels*, New York, NY, U.S.A.: Wiley.
- Kermoal, J. P., Schumacher, L., Pedersen, K. I., Mogensen, P. E., and Frederiksen, F. (2002), “A stochastic MIMO radio channel model with experimental validation,” *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226.
- Kingman, J. F. C. (1993), *Poisson Processes*, Oxford, U.K.: Oxford Univ. Press.
- Kmec, M., Sachs, J., Peyerl, P., Rauschenbach, P., Thomä, R., and Zetik, R. (2005), “A novel ultra-wideband real-time MIMO channel sounder architecture,” in *Proc. General Assembly Int. Union of Radio Science (URSI)*, New Delhi, India.
- Ko, Y.-C. and Alouini, M.-S. (2003), “Estimation of Nakagami- m fading channel parameters with applications to optimized transmitter diversity systems,” *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 250–260.
- Kobayashi, T. (2006), “Measurements and characterization of ultra wideband propagation channels in a passenger-car compartment,” in *Proc. IEEE Int. Symp. Spread Spectrum Tech. Applicat. (ISSTA)*, pp. 228–232, Manaus, Brazil.
- Koch, T. and Lapidoth, A. (2005), “The fading number and degrees of freedom in non-coherent MIMO fading channels: A peace pipe,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 661–665, Adelaide, Australia.
- Koch, T., Lapidoth, A., and Sotiriadis, P. P. (2007a), “A channel that heats up,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 906–910, Nice, France.
- (2007b), “A hot channel,” in *Proc. IEEE Inf. Theory Workshop (ITW)*, pp. 511–516, Tahoe City, CA, U.S.A.

REFERENCES

- Kolmogorov, A. N. (1992a), "Interpolation and extrapolation of stationary random sequences," in A. N. Shiriyayev (Ed.), *Selected Works of A. N. Kolmogorov*, vol. 2, chap. 28, pp. 272–280, Dordrecht, The Netherlands: Kluwer. Translated from the Russian by G. Lindquist; originally published in *Akad. Nauk. SSSR*, Ser. Mat. 2, no. 6, pp. 3–14, 1941.
- (1992b), "Stationary sequences in Hilbert space," in A. N. Shiriyayev (Ed.), *Selected Works of A. N. Kolmogorov*, vol. 2, chap. 27, pp. 228–271, Dordrecht, The Netherlands: Kluwer. Translated from the Russian by G. Lindquist; originally published in *Byull. Moskov. Gos. Univ. Mat.*, vol. 2, no. 6, pp. 1–40, 1940.
- Kouznetsov, K. A. and Meyer, R. G. (2000), "Phase noise in LC oscillators," *IEEE J. Solid-State Circuits*, vol. 35, no. 8, pp. 1244–1248.
- Kovačević, J. and Chebira, A. (2007), "Life beyond bases: The advent of frames (part I)," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 86–104.
- Kovács, I. Z., Pedersen, G. F., Eggers, P. C. F., and Olesen, K. (2004), "Ultra wideband radio propagation in body area network scenarios," in *Proc. IEEE Int. Symp. Spread Spectrum Tech. Applicat. (ISSA)*, pp. 102–106, Sydney, Australia.
- Kozek, W. (1992), "On the generalized Weyl correspondence and its application to time-frequency analysis of linear time-varying systems," in *Proc. IEEE Int. Symp. Time-Freq. Time-Scale Anal.*, pp. 167–170, Victoria, BC, Canada.
- (1997a), *Matched Weyl-Heisenberg Expansions of Nonstationary Environments*, Ph.D. thesis, Vienna University of Technology, Department of Electrical Engineering, Vienna, Austria.
- (1997b), "On the transfer function calculus of underspread LTV channels," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 219–223.
- Kozek, W. and Molisch, A. F. (1997), "On the eigenstructure of underspread WSSUS channels," in *Proc. IEEE Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, pp. 325–328, Paris, France.
- (1998), "Nonorthogonal pulseshapes for multicarrier communications in doubly dispersive channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1579–1589.
- Kreyszig, E. (1989), *Introduction to Functional Analysis with Applications*, New York, NY, U.S.A.: Wiley.
- Kunisch, J. and Pamp, J. (2002), "Measurement results and modeling aspects for the UWB radio channel," in *IEEE Conf. Ultra Wideband Syst. Technol. (UWBST) Dig. Tech. Papers*, pp. 19–23, Baltimore, MD, U.S.A.
- (2003), "An ultra-wideband space-variant multipath indoor ra-

- dio channel model,” in *IEEE Conf. Ultra Wideband Syst. Technol. (UWBST) Dig. Tech. Papers*, pp. 290–294, Reston, VA, U.S.A.
- (2005), “UWB radio channel measurements for sensor communication in an industrial environment,” in *Proc. Int. Conf. Electromagnetics in Adv. Applicat. and Eur. Electromagnetics Structures Conf.*, Torino, Italy.
- (2006), “Locally coherent ultra-wideband radio channel model for sensor networks in industrial environment,” in *Proc. IEEE Int. Conf. Ultra-Wideband (ICU)*, pp. 363–368, Waltham, MA, U.S.A.
- Lacroix, D., Despins, C. L., Delisle, G. Y., Mariner, P., and Luneau, P. (1997), “Experimental characterization of outdoor microcellular quasi-static channels in the UHF and SHF bands,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 1, pp. 41–45, Montreal, Canada.
- Laine, N., Hämäläinen, M., Hentilä, L., Iinatti, J. H. J., Nenonen, M., and Karhu, S. (2004), “Outdoor ultra wideband propagation channel modeling at VHF/UHF band,” in *Proc. Nordic Radio Symp.*, Oulu, Finland.
- Landau, H. J. and Pollak, H. O. (1961), “Prolate spheroidal wave functions, Fourier analysis and uncertainty—II,” *Bell Syst. Tech. J.*, vol. 40, no. 1, pp. 65–84.
- (1962), “Prolate spheroidal wave functions, Fourier analysis and uncertainty—III: The dimension of the space of essentially time- and band-limited signals,” *Bell Syst. Tech. J.*, vol. 41, pp. 1295–1336.
- Lao, R.-R., Tarng, J.-H., and Hsiao, C. (2003), “Transmission coefficient measurement of building materials for UWB systems in 3-10 GHz,” in *Proc. IEEE Veh. Technol. Conf. Spring (VTC-Spring)*, vol. 1, pp. 11–14, Jeju, Korea.
- Lapidoth, A. (2003), “On the high SNR capacity of stationary Gaussian fading channels,” in *Proc. Allerton Conf. Commun., Contr., Comput.*, pp. 410–419, Monticello, IL, U.S.A.
- (2005), “On the asymptotic capacity of stationary Gaussian fading channels,” *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 437–446.
- Lapidoth, A. and Moser, S. M. (2003), “Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels,” *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467.
- Lauritzen, S. L. and Bach Andersen, J. (1990), “A stochastic model in mobile communications,” *Stoch. Proc. Appl.*, vol. 36, no. 1, pp. 165–172.
- Ledoit, O. and Wolf, M. (2004), “A well-conditioned estimator for large-dimensional covariance matrices,” *J. Multivariate Anal.*, vol. 88, no. 2, pp. 365–411.

REFERENCES

- Lee, W. C. Y. (1993), *Mobile Communications Design Fundamentals*, New York, NY, U.S.A.: Wiley, 2nd ed.
- Lehne, P. H., Aanvik, F., Bic, J.-C., Pajusco, P., Grigat, M., Gaspard, I., Martin, U., and Steinbauer, M. (1998), "Review of existing channel sounder measurement setups and applied calibration methods," Deliverable META/D1-TR/D1/1/b1, Measurement, Testing, and Calibration of Advanced Mobile Radio-Channel Test Equipment Project (METAMORP).
- Li, Q. and Wong, S. (2003), "Measurement and analysis of the indoor UWB channel," in *Proc. IEEE Veh. Technol. Conf. Fall (VTC-Fall)*, vol. 1, pp. 1–5, Orlando, FL, U.S.A.
- Liang, Y. and Veeravalli, V. V. (2004), "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3096–3110.
- Limpert, E., Stahel, W. A., and Abbt, M. (2001), "Log-normal distributions across the sciences: Keys and clues," *Bioscience*, vol. 51, no. 5, pp. 341–352.
- Linhart, H. and Zucchini, W. (1986), *Model Selection*, New York, NY, U.S.A.: Wiley.
- Liu, B. and Stanley, T. P. (1965), "Error bounds for jittered sampling," *IEEE Trans. Autom. Control*, vol. 10, no. 4, pp. 449–454.
- Liu, K., Kadous, T., and Sayeed, A. (2004), "Orthogonal time-frequency signaling over doubly dispersive channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2583–2603.
- Lozano, A., Tulino, A. M., and Verdú, S. (2003), "Multiple-antenna capacity in the low-power regime," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2527–2544.
- (2006), "Multiantenna capacity: Myths and realities," in H. Bölcskei, D. Gesbert, C. B. Papadias, and A.-J. van der Veen (Eds.), *Space-Time Wireless Systems—From Array Processing to MIMO Communications*, chap. 5, pp. 87–107, Cambridge, U.K.: Cambridge Univ. Press.
- Lütkepohl, H. (1996), *Handbook of Matrices*, Chichester, U.K.: Wiley.
- Margetts, A. R. and Schniter, P. (2004), "Joint scale-lag diversity in mobile ultra-wideband systems," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, vol. 2, pp. 1496–1500, Pacific Grove, CA, U.S.A.
- Margetts, A. R., Schniter, P., and Swami, A. (2005), "Scale-lag diversity reception in mobile wideband channels," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 3, pp. 321–324, Philadelphia, PA, U.S.A.
- Marshall, A. W. and Olkin, I. (1979), *Inequalities: Theory of Majorization and Its Applications*, New York, NY, U.S.A.: Academic Press.

- Marshall, A. W. and Proshan, F. (1965), “An inequality for convex functions involving majorization,” *J. Math. Anal. Appl.*, vol. 12, no. 1, pp. 87–90.
- Maruyama, G. (1949), “The harmonic analysis of stationary stochastic processes,” *Memoirs of the Faculty of Science, Kyūshū University, Ser. A*, vol. 4, no. 1, pp. 45–106.
- Marzetta, T. L. (1995), “EM algorithm for estimating the parameters of a multivariate complex Rician density for polarimetric SAR,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 5, pp. 3651–3654, Detroit, MI, U.S.A.
- Marzetta, T. L. and Hochwald, B. M. (1999), “Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading,” *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157.
- Matsuoka, T. and Ulich, T. J. (1986), “Information theory measures with application to model identification,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 3, pp. 511–517.
- Matz, G. (2000), *A Time-Frequency Calculus for Time-Varying Systems and Nonstationary Processes with Applications*, Ph.D. thesis, Vienna University of Technology, Vienna, Austria.
- (2003), “Characterization of non-WSSUS fading dispersive channels,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 2480–2484, Anchorage, AK, U.S.A.
- (2005), “On non-WSSUS wireless fading channels,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2465–2478.
- Matz, G. and Hlawatsch, F. (1998), “Time-frequency transfer function calculus (symbolic calculus) of linear time-varying systems (linear operators) based on a generalized underspread theory,” *J. Math. Phys.*, vol. 39, no. 8, pp. 4041–4070.
- (2003a), “Time-frequency characterization of randomly time-varying channels,” in B. Boashash (Ed.), *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, chap. 9.5, pp. 410–419, Oxford, U.K.: Elsevier.
- (2003b), “Time-frequency transfer function calculus of linear time-varying systems,” in B. Boashash (Ed.), *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, chap. 4.7, pp. 135–144, Oxford, U.K.: Elsevier.
- Matz, G., Schafhuber, D., Gröchenig, K., Hartmann, M., and Hlawatsch, F. (2007), “Analysis, optimization, and implementation of low-interference wireless multicarrier systems,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1921–1931.
- McNamara, D. A., Pistorius, C. W. I., and Malherbe, J. A. G. (1990),

REFERENCES

- Introduction to the Uniform Geometrical Theory of Diffraction*, Boston, MA, U.S.A.: Artech House.
- Médard, M. (1995), *The Capacity of Time Varying Multiple User Channels in Wireless Communications*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, U.S.A.
- (2000), “The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel,” *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946.
- Médard, M. and Gallager, R. G. (2002), “Bandwidth scaling for fading multipath channels,” *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 840–852.
- Menouni Hayar, A., Knopp, R., and Saadane, R. (2005), “Subspace analysis of indoor UWB channels,” *EURASIP J. Appl. Signal Process.*, vol. 2005, no. 3, pp. 287–295.
- Meyr, H., Moeneclaey, M., and Fechtel, S. A. (1998), *Digital Communication Receivers*, New York, NY, U.S.A.: Wiley.
- Miller, K. S. (1969), “Complex Gaussian processes,” *SIAM Rev.*, vol. 11, no. 4, pp. 544–567.
- (1973), “Complex linear least squares,” *SIAM Rev.*, vol. 15, no. 4, pp. 706–726.
- Miranda, M. and Tilli, P. (2000), “Asymptotic spectra of Hermitian block Toeplitz matrices and preconditioning results,” *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 3, pp. 867–881.
- Molisch, A. F., Asplund, H., Heddergott, R., Steinbauer, M., and Zwick, T. (2006a), “The COST259 directional channel model—part I: Overview and methodology,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3421–3433.
- Molisch, A. F., Cassioli, D., Chong, C.-C., Emami, S., Fort, A., Balakrishnan, K., Karedal, J., Kunisch, J., Schanz, H. G., Siwiak, K., and Win, M. Z. (2006b), “A comprehensive standardized model for ultrawideband propagation channels,” *IEEE Trans. Antennas Propag.*, vol. 54, no. 11, pp. 3151–3166.
- Molisch, A. F., Foerster, J. R., and Pendergrass, M. (2003a), “Channel models for ultrawideband personal area networks,” *IEEE Wireless Commun. Mag.*, vol. 10, no. 6, pp. 14–21.
- Molisch, A. F., Kuchar, A., Laurila, J., Hugl, K., and Schmalenberger, R. (2003b), “Geometry-based directional model for mobile radio channels—principles and implementation,” *Eur. Trans. Telecommun.*, vol. 14, no. 4, pp. 351–359.
- Moser, S. M. (2007), “The fading number of multiple-input multiple-output

- fading channels with memory,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 521–525, Nice, France.
- Muirhead, R. J. (1982), *Aspects of Multivariate Statistical Theory*, New York, NY, U.S.A.: Wiley.
- (1987), “Developments in eigenvalue estimation,” in A. K. Gupta (Ed.), *Advances in Multivariate Statistical Analysis*, pp. 277–288, Dordrecht, The Netherlands: D. Reidel.
- Muqaibel, A. H., Safaai-Jazi, A., Attiya, A. M., Woerner, B., and Riad, S. M. (2006), “Path-loss and time dispersion parameters for indoor UWB propagation,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 3, pp. 550–559.
- Nabar, R. U., Bölcskei, H., and Paulraj, A. J. (2005), “Diversity and outage performance in space-time block coded Ricean MIMO channels,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2519–2532.
- Nakagami, M. (1960), “The m -distribution—a general formula of intensity distribution of rapid fading,” in W. C. Hoffman (Ed.), *Statistical Methods in Radio Wave Propagation*, pp. 3–36, New York, NY, U.S.A.: Pergamon.
- Naylor, A. W. and Sell, G. R. (1982), *Linear Operator Theory in Engineering and Science*, New York, NY, U.S.A.: Springer.
- Neeser, F. D. and Massey, J. L. (1993), “Proper complex random processes with applications to information theory,” *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302.
- Nikookar, H. and Hashemi, H. (1993), “Statistical modeling of signal amplitude fading of indoor radio propagation channels,” in *Proc. Conf. Record Int. Univ. Personal Commun. Conf.*, vol. 1, pp. 84–88.
- Ozarow, L. H., Shamai (Shitz), S., and Wyner, A. D. (1994), “Information theoretic considerations for cellular mobile radio,” *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 359–378.
- Pagani, P. and Pajusco, P. (2004), “Experimental assessment of the UWB channel variability in a dynamic indoor environment,” in *Proc. IEEE Int. Symp. Personal, Indoor, Mobile Radio Commun. (PIMRC)*, vol. 4, pp. 2973–2977, Barcelona, Spain.
- (2006), “Characterization and modeling of temporal variations on an ultrawideband link,” *IEEE Trans. Antennas Propag.*, vol. 54, no. 11, pp. 3198–3206.
- Papoulis, A. and Pillai, S. U. (2002), *Probability, Random Variables and Stochastic Processes*, Boston, MA, U.S.A.: McGraw-Hill, 4th ed.
- Parsons, J. D. (2000), *The Mobile Radio Propagation Channel*, Chichester, U.K.: Wiley, 2nd ed.
- Parsons, J. D. and Bajwa, A. S. (1982), “Wideband characterization of

REFERENCES

- fading mobile radio channels,” *IEE Proc. F, Commun., Radar, Signal Proc.*, vol. 129, no. 2, pp. 95–101.
- Parsons, J. D., Demery, D. A., and Turkmani, A. M. D. (1991), “Sounding techniques for wideband mobile radio channels: A review,” *IEE Proc. I, Commun., Speech, Vision*, vol. 138, no. 5, pp. 437–446.
- Patenaude, F., Lodge, J., and Chouinard, J.-Y. (1999), “Eigen analysis of wide-band fading channel impulse responses,” *IEEE Trans. Veh. Technol.*, vol. 48, no. 2, pp. 593–606.
- Paulraj, A., Nabar, R., and Gore, D. (2003), *Introduction to Space-Time Wireless Communications*, Cambridge, U.K.: Cambridge Univ. Press.
- Pearl, J. (1973), “On coding and filtering stationary signals by discrete Fourier transforms,” *IEEE Trans. Inf. Theory*, vol. 19, no. 2, pp. 229–232.
- Pfander, G. E. and Walnut, D. F. (2006), “Measurement of time-variant linear channels,” *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 4808–4820.
- Picinbono, B. (1994), “On circularity,” *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3473–3482.
- (1996), “Second-order complex random vectors and normal distributions,” *IEEE Trans. Signal Process.*, vol. 44, no. 10, pp. 2637–2640.
- Pierce, J. R. (1958), “Theoretical diversity improvement in frequency-shift keying,” *Proc. IRE*, vol. 46, pp. 903–910.
- (1966), “Ultimate performance of M -ary transmission on fading channels,” *IEEE Trans. Inf. Theory*, vol. 12, no. 1, pp. 2–5.
- Plonsey, R. and Collin, R. E. (1961), *Principles and Applications of Electromagnetic Fields*, New York, NY, U.S.A.: McGraw-Hill.
- Poon, A. S. Y., Brodersen, R. W., and Tse, D. N. C. (2004), “A spatial channel model for multiple-antenna systems,” in *IEEE Antennas and Propag. Soc. Int. Symp.*, vol. 4, pp. 3665–3668, Monterey, CA, U.S.A.
- Poon, A. S. Y. and Ho, M. (2003), “Indoor multiple-antenna channel characterization from 2 to 8 GHz,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 5, pp. 3519–3523, Anchorage, AK, U.S.A.
- Poor, H. V. (1994), *An Introduction to Signal Detection and Estimation*, New York, NY, U.S.A.: Springer, 2nd ed.
- Porrat, D., Tse, D. N. C., and Nacu, Șerban. (2007), “Channel uncertainty in ultra-wideband communication systems,” *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 194–208.
- Prelov, V. V. and Verdú, S. (2004), “Second-order asymptotics of mutual information,” *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1567–1580.
- Prettie, C., Cheung, D., Rusch, L., and Ho, M. (2002), “Spatial correlation of UWB signals in a home environment,” in *IEEE Conf. Ultra Wideband*

- Syst. Technol. (UWBST) Dig. Tech. Papers*, pp. 65–70, Baltimore, MD, U.S.A.
- Price, R. and Green, P. E. (1958), “A communication technique for multipath channels,” *Proc. IRE*, vol. 46, no. 3, pp. 555–570.
- Proakis, J. G. (2001), *Digital Communications*, New York, NY, U.S.A.: McGraw-Hill, 4th ed.
- Qiu, R. C. and Lu, I.-T. (1996), “Wideband wireless channel modeling with path frequency dependence,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 1, pp. 277–281, Dallas, TX, U.S.A.
- Rabaey, J. M., Ammer, J., Karalar, T., Li, S., Otis, B., Sheets, M., and Tuan, T. (2002), “Picoradios for wireless sensor networks—the next challenge in ultra-low power design,” in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, pp. 200–201, San Francisco, CA, U.S.A.
- Rappaport, T. S. (2002), *Wireless Communications: Principles and Practice*, Upper Saddle River, NJ, U.S.A.: Prentice Hall, 2nd ed.
- Rastogi, P. K. and Holt, O. (1981), “On detecting reflections in presence of scattering from amplitude statistics with applications to D region partial reflections,” *Radio Sci.*, vol. 16, no. 6, pp. 1431–1443.
- Razavi, B. (1998), *RF Microelectronics*, Upper Saddle River, NJ, U.S.A.: Prentice Hall.
- Rice, S. O. (1944), “Mathematical analysis of random noise,” *Bell Syst. Tech. J.*, vol. 23, pp. 282–332.
- Rickard, S. T., Balan, R. V., Poor, H. V., and Verdú, S. (2005), “Canonical time-frequency, time-scale, and frequency-scale representations of time-varying channels,” *Commun. Inf. Syst.*, vol. 5, no. 2, pp. 197–226.
- Rozanov, Y. A. (1967), *Stationary Random Processes*, San Francisco, CA, U.S.A.: Holden-Day.
- Rudin, W. (1976), *Principles of Mathematical Analysis*, New York, NY, U.S.A.: McGraw-Hill, 3rd ed.
- Saadane, R., Aboutajdine, D., Hayar, A. M., and Knopp, R. (2005), “On the estimation of the degrees of freedom of in-door UWB channels,” in *Proc. IEEE Veh. Technol. Conf. Spring (VTC-Spring)*, vol. 5, pp. 3147–3151, Stockholm, Sweden.
- Saadane, R., Menouni Hayar, A., Knopp, R., and Aboutajdine, D. (2004), “Empirical eigenanalysis of indoor UWB propagation channels,” in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, vol. 5, pp. 3215–3219, Dallas, TX, U.S.A.
- Sadowsky, J. S. and Kafedziski, V. (1998), “On the correlation and scattering functions of the WSSUS channel for mobile communications,” *IEEE Trans. Veh. Technol.*, vol. 47, no. 1, pp. 270–282.

REFERENCES

- Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986), *Akaike Information Criterion Statistics*, Tokyo, Japan: KTK Scientific Publishers.
- Sakrison, D. J. (1969), "An extension of the theorem of Kač, Murdock and Szegö to N dimensions," *IEEE Trans. Inf. Theory*, vol. 15, no. 5, pp. 608–610.
- Saleh, A. A. M. and Valenzuela, R. A. (1987), "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 128–137.
- Sanderovich, A., Somekh, O., and Shamai (Shitz), S. (2007), "Uplink macro diversity with limited backhaul capacity," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 11–15, Nice, France.
- Schafhuber, D., Bölcskei, H., and Matz, G. (2004), "System capacity of wideband OFDM communications over fading channels without channel knowledge," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, p. 391, Chicago, IL, U.S.A., Available: <http://www.nari.ee.ethz.ch/commth/pubs/p/ofdm04>. Corrected version online.
- Scholtz, R. A. (1993), "Multiple access with time-hopping impulse modulation," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, vol. 2, pp. 447–450, Boston, MA, U.S.A.
- Scholtz, R. A., Cramer, R. J.-M., and Win, M. Z. (1998), "Evaluation of the propagation characteristics of ultra-wideband communication channels," in *Proc. IEEE Antennas and Propagat. Soc. Int. Symp.*, vol. 2, pp. 626–630, Montreal, Canada.
- Schreier, P. J. and Scharf, L. L. (2003), "Second-order analysis of improper complex random vectors and processes," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 714–725.
- Schreier, P. J., Scharf, L. L., and Hanssen, A. (2006), "A generalized likelihood ratio test for impropriety of complex signals," *IEEE Signal Process. Lett.*, vol. 13, no. 7, pp. 433–436.
- Schuster, U. G. and Bölcskei, H. (2005a), "How different are UWB channels from conventional wideband channels?" in *Proc. Int. Workshop Convergent Technol. (IWCT)*, Oulu, Finland.
- (2005b), "UWB measurement database," Available: www.nari.ee.ethz.ch/commth/uwb-measurements.
- (2007), "Ultrawideband channel modeling on the basis of information-theoretic criteria," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2464–2475.
- Schuster, U. G., Bölcskei, H., and Durisi, G. (2005), "Ultra-wideband channel modeling on the basis of information-theoretic criteria," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 97–101, Adelaide, Australia.

- Schuster, U. G., Durisi, G., Bölcskei, H., and Poor, H. V. (2009), “Capacity bounds for peak-constrained multiantenna wideband channels,” *IEEE Trans. Commun.*, Available: <http://arxiv.org/abs/0801.1002>. To appear.
- Seidel, S. Y. and Rappaport, T. S. (1994), “Site-specific propagation prediction for wireless in-building personal communication system design,” *IEEE Trans. Veh. Technol.*, vol. 43, no. 4, pp. 879–891.
- Sethuraman, V. and Hajek, B. (2005), “Capacity per unit energy of fading channels with peak constraint,” *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3102–3120.
- (2006), “Low SNR capacity of fading channels with peak and average power constraints,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 689–693, Seattle, WA, U.S.A.
- Sethuraman, V., Hajek, B., Lapidath, A., and Wang, L. (2007), “Low SNR capacity of fading channels—MIMO and multipath,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 516–520, Nice, France.
- Sethuraman, V., Hajek, B., and Narayanan, K. (2005), “Capacity bounds for noncoherent fading channels with a peak constraint,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 515–519, Adelaide, Australia.
- Sethuraman, V., Wang, L., Hajek, B., and Lapidath, A. (2009), “Low SNR capacity of noncoherent fading channels,” *IEEE Trans. Inf. Theory*, Available: <http://arxiv.org/abs/0712.2872>. To appear.
- Shannon, C. E. (1948), “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656.
- Shibata, R. (1989), “Statistical aspects of model selection,” in J. C. Willems (Ed.), *From Data to Model*, pp. 215–240, Berlin, Germany: Springer.
- Sibul, L. H., Weiss, L. G., and Dixon, T. L. (1994), “Characterization of stochastic propagation and scattering via Gabor and wavelet transforms,” *J. Comput. Acoust.*, vol. 2, no. 3, pp. 345–369.
- Slepian, D. (1976), “On bandwidth,” *Proc. IEEE*, vol. 64, no. 3, pp. 292–300.
- Slepian, D. and Pollak, H. O. (1961), “Prolate spheroidal wave functions, Fourier analysis, and uncertainty—I,” *Bell Syst. Tech. J.*, vol. 40, no. 1, pp. 43–63.
- Spanos, A. (1999), *Probability Theory and Statistical Inference*, Cambridge, U.K.: Cambridge Univ. Press.
- Spark, N. T. (2003), “The fastest man on earth,” *Ann. Improb. Research*, vol. 9, no. 5, pp. 4–26, Available: <http://www.improb.com/airchives/paperair/volume9/v9i5/murphy/murphy0.html>.
- Srinivasan, S. G. and Varanasi, M. K. (2007), “Optimal spatial correlations for the noncoherent MIMO Rayleigh fading channel,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 10, pp. 3760–3769.

REFERENCES

- Stein, S. (1987), "Fading channel issues in system engineering," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 68–89.
- Steinbauer, M., Molisch, A. F., and Bonek, E. (2001), "The double-directional radio channel," *IEEE Antennas Propag. Mag.*, vol. 43, no. 4, pp. 51–63.
- Stoica, P. and Selén, Y. (2004), "Model-order selection—a review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47.
- Stojanovic, M. (1996), "Recent advances in high-speed underwater acoustic communications," *IEEE J. Ocean. Eng.*, vol. 21, no. 2, pp. 125–136.
- Street, A. M., Lukama, L., and Edwards, D. J. (2001), "Use of VNAs for wideband propagation measurements," *IEE Proc. Commun.*, vol. 148, no. 6, pp. 411–415.
- Subramanian, V. G. and Hajek, B. (2002), "Broad-band fading channels: Signal burstiness and capacity," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 809–827.
- Suzuki, H. (1977), "A statistical model for urban radio propagation," *IEEE Trans. Commun.*, vol. 25, no. 7, pp. 673–680.
- Taludkar, K. K. and Lawing, W. D. (1991), "Estimation of the parameters of the Rice distribution," *J. Acoust. Soc. Am.*, vol. 89, no. 3, pp. 1193–1197.
- Taneda, M. A., Takada, J., and Araki, K. (2001), "The problem of the fading model selection," *IEICE Trans. Commun.*, vol. E84-B, no. 3, pp. 660–666.
- Tanenbaum, A. S. (2003), *Computer Networks*, Upper Saddle River, NJ, U.S.A.: Prentice Hall, 4th ed.
- Taricco, G. and Elia, M. (1997), "Capacity of fading channel with no side information," *Electron. Lett.*, vol. 33, no. 16, pp. 1368–1370.
- Telatar, I. E. (1999), "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595.
- Telatar, I. E. and Tse, D. N. C. (2000), "Capacity and mutual information of wideband multipath fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1384–1400.
- Tepedelenlioğlu, C., Abdi, A., and Giannakis, G. B. (2003), "The Ricean K factor: Estimation and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 799–810.
- Truffer, P. and Leuthold, P. E. (2001), "Wide-band channel sounding at 24 GHz based on a novel fiber-optic synchronization concept," *IEEE Trans. Microw. Theory Tech.*, vol. 49, no. 4, pp. 692–700.
- Tse, D. N. C. and Viswanath, P. (2005), *Fundamentals of Wireless Communication*, Cambridge, U.K.: Cambridge Univ. Press.

- Tulino, A. M., Lozano, A., and Verdú, S. (2005), “Impact of antenna correlation on the capacity of multi-antenna channels,” *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491–2509.
- Turin, G. L. (1972), “A statistical model of urban multipath propagation,” *IEEE Trans. Veh. Technol.*, vol. 21, no. 1, pp. 1–8.
- Turin, W., Jana, R., Ghassemzadeh, S. S., Rice, C. W., and Tarokh, V. (2002), “Autoregressive modeling of an indoor UWB channel,” in *IEEE Conf. Ultra Wideband Syst. Technol. (UWBST) Dig. Tech. Papers*, pp. 71–74, Baltimore, MD, U.S.A.
- Tyrtshnikov, E. E. (1994), “Influence of matrix operations on the distribution of eigenvalues and singular values of Toeplitz matrices,” *Lin. Algeb. Appl.*, vol. 207, pp. 225–249.
- (1996), “A unifying approach to some old and new theorems on distribution and clustering,” *Lin. Algeb. Appl.*, vol. 232, pp. 1–43.
- Tyrtshnikov, E. E. and Zamarashkin, N. L. (1998), “Spectra of multilevel Toeplitz matrices: Advanced theory via simple matrix relations,” *Lin. Algeb. Appl.*, vol. 270, pp. 15–27.
- Tzeremes, G. and Christodoulou, C. G. (2002), “Use of Weibull distribution for describing outdoor multipath fading,” in *IEEE Antennas and Propagat. Soc. Int. Symp.*, vol. 1, pp. 232–235, San Antonio, TX, U.S.A.
- van Trees, H. L. (2002), *Detection, Estimation, and Modulation Theory—Optimum Array Processing*, vol. 4, New York, NY, U.S.A.: Wiley.
- Vaughan, R. and Bach Andersen, J. (2003), *Channels, Propagation and Antennas for Mobile Communications*, London, U.K.: The Institution of Electrical Engineers.
- Verdú, S. (2002), “Spectral efficiency in the wideband regime,” *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343.
- Viterbi, A. J. (1967), “Performance of an M-ary orthogonal communication system using stationary stochastic signals,” *IEEE Trans. Inf. Theory*, vol. 13, no. 3, pp. 414–422.
- Wagenmakers, E.-J. and Farrell, S. (2004), “AIC model selection using Akaike weights,” *Psychonomic Bulletin & Rev.*, vol. 11, no. 1, pp. 192–196.
- Wang, L. (2006), *On Non-Coherent Fading Channels at Low SNR*, Master’s thesis, Signal and Inform. Proc. Lab., ETH Zurich, Zurich, Switzerland. See also arXiv:cs/0604031.
- Wax, M. and Kailath, T. (1985), “Detection of signals by information theoretic criteria,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392.

REFERENCES

- Weiss, L. G. (1996), "Time-varying system characterization for wideband input signals," *EURASIP J. Signal Process.*, vol. 55, no. 3, pp. 295–304.
- Welch, T. B., Musselman, R. L., Emessiene, B. A., Gift, P. D., Choudhury, D. K., Cassadine, D. N., and Yano, S. M. (2002), "The effects of the human body on UWB signal propagation in an indoor environment," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 9, pp. 1778–1782.
- Wiener, N. and Masani, P. (1957), "The prediction theory of multivariate stochastic processes, I. The regularity condition," *Acta Math.*, vol. 98, pp. 111–150.
- Wilcox, C. H. (1991), "The synthesis problem for radar ambiguity functions," in R. E. Blahut, W. Miller, Jr., and C. H. Wilcox (Eds.), *Radar and Sonar*, vol. 1, pp. 229–260, New York, NY, U.S.A.: Springer.
- Wilson, S. G. (1996), *Digital Modulation and Coding*, Upper Saddle River, NJ, U.S.A.: Prentice Hall.
- Win, M. Z., Ramírez-Mireles, F., and Scholtz, R. A. (1997a), "Ultra-wide bandwidth (UWB) signal propagation for outdoor wireless communications," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, vol. 1, pp. 251–255, Phoenix, AZ, U.S.A.
- Win, M. Z. and Scholtz, R. A. (1997), "Comparison of analog and digital impulse radio for wireless multiple-access communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 91–95.
- (1998a), "Impulse radio: How it works," *IEEE Commun. Lett.*, vol. 2, no. 2, pp. 36–38.
- (1998b), "On the robustness of ultra-wide bandwidth signals in dense multipath environments," *IEEE Commun. Lett.*, vol. 2, no. 2, pp. 10–12.
- (2000), "Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple-access communications," *IEEE Trans. Commun.*, vol. 48, no. 4, pp. 679–691.
- (2002), "Characterization of ultra-wide bandwidth wireless indoor channels: A communication-theoretic view," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 9, pp. 1613–1627.
- Win, M. Z., Scholtz, R. A., and Barnes, M. A. (1995), "Time Domain Corporation UWB database," Available: http://ultra.usc.edu/uwb_database/.
- (1997b), "Ultra-wide bandwidth signal propagation for indoor wireless communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 1, pp. 56–60, Montreal, Canada.
- Wooding, R. A. (1956), "The multivariate distribution of complex normal variables," *Biometrika*, vol. 43, no. 1/2, pp. 212–215.

- Woodward, P. M. (1953), *Probability and Information Theory, with Applications to Radar*, London, U.K.: Pergamon Press.
- Wozencraft, J. M. and Jacobs, I. M. (1965), *Principles of Communication Engineering*, New York, NY, U.S.A.: Wiley.
- Xiao, C. and Zheng, Y. R. (2004), “Ergodic capacity of doubly selective Rayleigh fading MIMO channels,” in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, vol. 1, pp. 345–350, Atlanta, GA, U.S.A.
- Yacoub, M. D. (2000), “Fading distributions and co-channel interference in wireless systems,” *IEEE Antennas Propag. Mag.*, vol. 42, no. 1, pp. 150–158.
- Yang, Y. (2005), “Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation,” *Biometrika*, vol. 92, no. 4, pp. 937–950.
- Yano, S. M. (2002), “Investigating the ultra-wideband indoor wireless channel,” in *Proc. IEEE Veh. Technol. Conf. Spring (VTC-Spring)*, vol. 3, pp. 1200–1204, Birmingham, AL, U.S.A.
- Yee, D. G.-W. (2001), *A Design Methodology for Highly-Integrated Low-Power Receivers for Wireless Communications*, Ph.D. thesis, University of California at Berkeley, Berkeley, CA, U.S.A.
- Zadeh, L. A. (1950), “Frequency analysis of variable networks,” *Proc. IRE*, vol. 38, no. 3, pp. 291–299.
- (1961), “Time-varying networks—I,” *Proc. IRE*, vol. 49, no. 10, pp. 1488–1503.
- Zasowski, T., Althaus, F., Stäger, M., Wittneben, A., and Tröster, G. (2003), “UWB for noninvasive wireless body area networks: Channel measurements and results,” in *IEEE Conf. Ultra Wideband Syst. Technol. (UWBST) Dig. Tech. Papers*, pp. 285–289, Reston, VA, U.S.A.
- Zasowski, T., Meyer, G., Althaus, F., and Wittneben, A. (2006), “UWB signal propagation at the human head,” *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 4, pp. 1836–1844.
- Zey, C., Filliben, J. J., Heckert, A., Croarkin, C., Hembree, B., Guthrie, W., Trutna, L., Spagon, P., del Castillo, E., Moore, T., Hartley, S., Hurwitz, A., Prins, J., McCormack, D., Michelson, D., Horrell, K., and Tobias, P. (2006), *NIST/SEMATECH e-Handbook of Statistical Methods*, National Institute of Standards and Technology, Gaithersburg, MD, U.S.A., Available: <http://www.itl.nist.gov/div898/handbook/>. Last visited Oct. 2007.
- Zhang, H., Udagawa, T., Arita, T., and Nakagawa, M. (2002), “A statistical model for the small-scale multipath fading characteristics of ultra-wideband indoor channel,” in *IEEE Conf. Ultra Wideband Syst.*

REFERENCES

- Technol. (UWBST) Dig. Tech. Papers*, pp. 81–85, Baltimore, MD, U.S.A.
- Zhang, W. and Laneman, J. N. (2007a), “Benefits of spatial correlation for multi-antenna non-coherent communication over fading channels at low SNR,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 887–896.
- (2007b), “How good is PSK for peak-limited fading channels in the low-SNR regime?” *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 236–251.
- Zhang, W. and Miller, M. J. (1992), “Baseband equivalents in digital communication system simulation,” *IEEE Trans. Educ.*, vol. 35, no. 4, pp. 376–382.
- Zhao, L. C., Krishnaiah, P. R., and Bai, Z. D. (1986), “On detection of the number of signals when the noise covariance matrix is arbitrary,” *J. Multivariate Anal.*, vol. 20, no. 1, pp. 26–49.
- Zheng, L. and Tse, D. N. C. (2002), “Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel,” *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383.
- Zucchini, W. (2000), “An introduction to model selection,” *J. Math. Psychology*, vol. 44, no. 1, pp. 41–61.

List of Publications

- Althaus, F., Chin, F., Kunisch, J., Pamp, J., Radunović, B., Schuster, U. G., Trösch, F., Truong, H. L., Weisenhorn, M., and Zetik, R. (2006), “A comprehensive investigation of UWB sensor networks for industrial applications,” in *Proc. IST Mobile Summit*, Myconos, Greece.
- Durisi, G., Schuster, U. G., Bölcskei, H., and Shamai (Shitz), S. (2008), “Capacity of underspread WSSUS fading channels in the wideband regime under peak constraints,” *IEEE Trans. Inf. Theory*, Available: <http://arxiv.org/abs/0804.1748>. Submitted for review.
- Schuster, U. G. and Bölcskei, H. (2005), “How different are UWB channels from conventional wideband channels?” in *Proc. Int. Workshop Convergent Technol. (IWCT)*, Oulu, Finland.
- (2007), “Ultrawideband channel modeling on the basis of information-theoretic criteria,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2464–2475.
- Schuster, U. G., Bölcskei, H., and Durisi, G. (2005), “Ultra-wideband channel modeling on the basis of information-theoretic criteria,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 97–101, Adelaide, Australia.
- Schuster, U. G., Borgmann, M., and Bölcskei, H. (2004), “Semicoherent PPM for wideband communications,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, p. 383, Chicago, IL, U.S.A.
- Schuster, U. G., Durisi, G., Bölcskei, H., and Poor, H. V. (2008), “Capacity bounds for peak-constrained multiantenna wideband channels,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1582–1586, Toronto, ON, Canada.
- (2009), “Capacity bounds for peak-constrained multiantenna wideband channels,” *IEEE Trans. Commun.*, Available: <http://arxiv.org/abs/0801.1002>. To appear.

Curriculum Vitae

Ulrich G. Schuster
born 2 September 1977

Education

12/2007	Dr. sc. ETH Zurich, Switzerland
12/2002	M.S. in Electrical Engineering and Computer Sciences, University of California at Berkeley, U.S.A.
10/1997–07/2001	Studies in Electrical Engineering, RWTH Aachen University, Germany

Research

04/2003–02/2008	Research Assistant, Communication Technology Laboratory, ETH Zurich, Switzerland
07/2006–09/2006	Graduate Student Research Collaborator, Princeton University, NJ, U.S.A.
07/2004–08/2004	Visiting Researcher, Centre for Wireless Communications, University of Oulu, Finland
08/2001–12/2002	Graduate Student Researcher, Berkeley Wireless Research Center, University of California at Berkeley, U.S.A.
11/1998–06/2000	Student Researcher, Institute for Electrical Machines, RWTH Aachen University, Germany

Series in Communication Theory

Edited by Helmut Bölcskei

- Vol. 1 Markus E. Gärtner, *Space-Time Coding and Multiple Access in MIMO Fading Channels*. ISBN 3-86628-181-1
- Vol. 2 Moritz Borgmann, *Noncoherent MIMO Wideband Communications*. ISBN 3-86628-246-X
- Vol. 3 Daniel S. Baum, *Information-Theoretic Analysis of a Class of MIMO Channel Measurement Devices*. ISBN 3-86628-247-8
- Vol. 4 Ulrich G. Schuster, *Wireless Communication over Wideband Channels*. ISBN 3-86628-245-1

More information at <http://www.Hartung-Gorre.de>
Hartung-Gorre Verlag Konstanz

