

DISS. ETH NO. 19173

CRYSTALLIZATION AND NONCOHERENCE IN WIRELESS COMMUNICATION

A dissertation submitted to

ETH ZURICH

for the degree of

Doctor of Sciences

presented by

VENIAMIN I. MORGENSHTERN

Dipl. Math., Saint-Petersburg State University

born 23.06.1982

citizen of Russia

accepted on the recommendation of

Prof. Dr. Helmut Bölcskei, examiner

Prof. Dr. İ. Emre Telatar, coexaminer

Prof. Dr. Amos Lapidoth, coexaminer

2010

To my parents

Abstract

Channel capacity determines the largest amount of data that can be reliably transmitted per second over a communication channel. Finding the capacity of a *wireless* channel is a difficult problem because of the presence of two effects: the effect of *interference* and the effect of *fading*. When analyzing fading channels, it is common to assume that the receiver knows the channel state perfectly, i.e., a *genie* provides channel state information (CSI) to the receiver. This scenario is referred to as the *coherent* fading channel. However, in practical wireless systems such CSI must constantly be acquired and updated. To characterize the cost of acquiring CSI fundamentally, one needs to study the capacity of the fading channel in the setting when *no* CSI is available at the receiver; this scenario is referred to as the *noncoherent* fading channel.

Part I of this thesis is devoted to studying how relays can be used to improve spectral efficiency in *interference* wireless networks. Specifically, we analyze fading interference relay networks where M single-antenna source-destination terminal pairs communicate concurrently and in the same frequency band through a set of K single-antenna relays using half-duplex two-hop relaying.

First, we consider a *coherent-relaying* protocol, where the relays have CSI and perform matched-filtering. The destination terminals cannot cooperate, i.e., the network operates in a completely distributed fashion. It is shown that in the large- M limit, provided K grows fast enough as a function of M , the network “decouples”. This means that the capacities of all individual source-destination pair links are

strictly positive. The signal power in all individual source-destination pair links dominates the interference power in these links. Therefore, all users in the network can transmit concurrently and in the same frequency-band. The required (for the network to decouple) rate of growth of K as a function of M is found to be sufficient to also make the individual source-destination fading links converge to nonfading links. We say that the network “*crystallizes*” as it breaks up into a set of effectively isolated “*wires in the air*”. A large-deviations analysis is performed to characterize the “crystallization” rate, i.e., the rate (as a function of M, K) at which the decoupled links converge to nonfading links.

Second, we consider the case of relays that do not have CSI and perform simple amplify-and-forward (AF) relaying. The destination terminals have CSI, and are allowed to cooperate and perform joint decoding. Based on tools from large random-matrix theory, we compute the per source-destination terminal pair capacity of this network for $M, K \rightarrow \infty$, with $K/M \rightarrow \beta$ fixed. We also find that for $\beta \rightarrow \infty$, the AF relay network is turned into a point-to-point multiple-input multiple-output (MIMO) link. This result demonstrates that employing relays as “active scatterers” can recover spatial multiplexing gain in poor scattering environments.

In Part II of the thesis, we consider exclusively systems with a single transmitter and a single receiver. We study the impact of absence of CSI at the receiver on channel capacity in two different models. This is summarized next.

The noncoherent capacity of stationary discrete-time fading channels is known to be very sensitive to the fine details of the channel model. More specifically, the measure of the set of harmonics where the power spectral density of the fading process is nonzero determines if capacity grows logarithmically in signal-to-noise ratio (SNR) or slower than logarithmically. An engineering-relevant problem is to characterize the SNR value at which this sensitivity starts to matter. To address this problem we start from the general model of continuous-time Rayleigh-fading channels that satisfy the wide-sense stationary uncorrelated scattering (WSSUS) assumption. In addition,

we assume that the channel is *underspread*, i.e., its scattering function is highly concentrated around the origin of the Doppler-delay plane. This is a highly relevant assumption for most real-life wireless channels and it simplifies our analysis significantly. We show that the noncoherent capacity of the WSSUS underspread channel is close to the additive white Gaussian noise channel capacity for all SNR values of practical interest, independently of whether the scattering function is compactly supported or not. As a byproduct of our analysis, we obtain an information-theoretic pulse-design criterion for pulse-shaped orthogonal frequency division multiplexing (PS-OFDM) systems.

Finally, we analyze the impact of multiple antennas at the receiver side of noncoherent channels using the temporally correlated Rayleigh block-fading single-input multiple-output (SIMO) channel model. Our analysis is aimed at characterizing the capacity pre-log, i.e., the asymptotic ratio between the capacity and the logarithm of SNR, as SNR goes to infinity. We derive a lower bound on the channel capacity pre-log and show that the capacity pre-log in the SIMO case is generally *larger* than that in the single-input single-output (SISO) case. The result is surprising, because in the *coherent* case the capacity pre-log of a SIMO fading channel is *the same as* that of a SISO fading channel.

Kurzfassung

Die Kanalkapazität entspricht der maximal erreichbaren Datenmenge, die pro Sekunde fehlerfrei über einen Kommunikationskanal gesendet werden kann. Die Kapazität eines drahtlosen Kanals zu bestimmen, ist aufgrund zweier Effekte ein schwieriges Problem: zum einen Interferenz, zum anderen Kanalschwund. Bei der Untersuchung von Schwundkanälen ist es üblich anzunehmen, dass der Empfänger den Kanalzustand vollkommen kennt, d. h., ein Orakel stellt dem Empfänger die Kanalzustandsinformation (CSI) zur Verfügung. In realen Systemen muss das CSI jedoch laufend erworben und erneuert werden. Um die Kosten für den Erwerb des CSI grundlegend zu bestimmen, muss man die Kapazität des Schwundkanals für den Fall, für den der Empfänger kein CSI besitzt, bestimmen; diesen Fall bezeichnet man auch als den nicht kohärenten Schwundkanal.

Der erste Teil dieser Dissertation ist der Frage gewidmet, wie Relais verwendet werden können, die spektrale Effizienz in drahtlosen Interferenznetzwerken zu erhöhen. Insbesondere werden Relais-Interferenzschwundnetzwerke untersucht, in denen M Einantennen Sender-Empfängerpaare gleichzeitig und im selben Frequenzband mittels einer Anzahl von K Einantennenrelais unter Verwendung eines halb-duplex Protokolls über zwei Zeitschlitze miteinander kommunizieren.

Zuerst untersuchen wir ein kohärentes Relaisprotokoll, bei welchem die Relais über CSI verfügen und daran angepasst filtern. Die Empfänger können nicht kooperieren, d.h., das Netzwerk funktioniert in einer vollständig dezentralisierten Art. Es wird gezeigt, dass sich das Netzwerk

für den Fall, in dem M groß wird, und K als Funktion von M ebenfalls groß genug ist, “aufteilen”. Das bedeutet, dass die Kapazitäten aller individuellen Sender-Empfängerpaarverbindungen strikt positiv sind. Die Signalleistung in allen individuellen Sender-Empfängerpaarverbindungen dominiert die Interferenzleistung in diesen Verbindungen. Deshalb können alle Nutzer in dem Netzwerk gleichzeitig im gleichen Frequenzband senden. Die notwendige Wachstumsrate von K (damit das Netzwerk sich aufteilen kann) als Funktion von M ist hinreichend, um auch die individuellen Sender-Empfängerschwindkanäle zu konstanten Verbindungen konvergieren zu lassen. Da das Netzwerk sich effektiv als “Drahtverbindung in der Luft” auffassen lässt, sagen wir fortan, dass sich das Netzwerk “kristallisiert”. Eine Analyse der großen Abweichungen wird durchgeführt, um die Kristallisationsrate zu charakterisieren, d.h., jene Rate (als Funktion von M und K), bei welcher die aufgeteilten Verbindungen zu konstanten Verbindungen konvergieren.

Danach betrachten wir den Fall, in dem die Relais über kein CSI verfügen und das Signal einfach verstärken und weiterleiten (AF). Die Empfänger verfügen über CSI und dürfen kooperieren und gemeinsam dekodieren. Basierend auf Erkenntnissen aus der Theorie der großen Zufallsmatrizen berechnen wir die Kapazität pro Sender-Empfängerpaar in diesem Netzwerk wenn $M, K \rightarrow \infty$, wobei $K/M \rightarrow \beta$ konstant bleibt. Wir zeigen auch, dass sich das AF Relaisnetzwerk für $\beta \rightarrow \infty$ in eine Punkt-zu-Punkt Mehrantennenverbindung mit mehreren Antennen auf Sender- und Empfängerseite wandelt. Dieses Resultat demonstriert, dass die Verwendung von Relais als “aktive Streuer” dazu führen kann, dass man den räumlichen Multiplexgewinn in schlechten Streuungsumgebungen erzeugen kann.

Im zweiten Teil dieser Dissertation betrachten wir ausschließlich Systeme mit einem einzigen Sender und einem einzigen Empfänger. Wir studieren die Auswirkung des Nichtvorhandenseins von CSI auf Empfängerseite auf die Kanalkapazität in zwei verschiedenen Modellen. Dies ist im Folgenden zusammengefasst.

Die nichtkohärente Kapazität von stationären zeitdiskreten Schwundkanälen ist bekanntermaßen stark abhängig von kleinen Details des

Kanalmodells. Genauer gesagt bestimmt das Mass der Menge der harmonischen Schwingungen bei welchen die spektrale Leistungsdichte des Schwundprozesses ungleich Null ist, ob die Kapazität logarithmisch im Verhältnis von Signal- zu Rauschenergie (SNR) oder langsamer als logarithmisch wächst. Ein aus Ingenieursicht relevantes Problem ist es, den SNR Wert, bei welchem die Signifikanz dieser Abhängigkeit gross ist, zu bestimmen. Um dieses Problem zu lösen, betrachten wir zuerst das allgemeine Modell zeitkontinuierlicher Rayleigh-Schwundkanäle, die die Annahme der weitestgehenden Stationarität und unkorrelierten Streuung (WSSUS) erfüllen. Zusätzlich nehmen wir an, dass der Kanal untergespreizt ist, d.h., seine Streufunktion ist stark um den Ursprung der Doppler-Verzögerungsebene konzentriert. Diese Annahme ist für die meisten realen Drahtloskanäle vollkommen angebracht und vereinfacht unsere Analyse signifikant. Wir zeigen, dass die nichtkohärente Kapazität des WSSUS unterspreizten Kanals für alle SNR Werte, welche von praktischem Interesse sind, nahe an der Kapazität des Kanals unter additivem Gausschen Rauschen ist. Dies gilt unabhängig von der Kompaktheit der Streufunktion. Als ein Nebenprodukt unserer Analyse erhalten wir ein informationstheoretisches Pulsdesignkriterium für Systeme, welche ein pulsförmiges orthogonales Frequenzmultiplexverfahren verwenden.

Zuletzt analysieren wir die Auswirkung, die mehrere Antennen auf der Empfängerseite auf nichtkohärente Kanäle haben, indem wir das zeitlich korrelierte Rayleigh-Blockschwundkanalmodell mit einem einzelnen Eingang und mehreren Ausgängen (SIMO) verwenden. Unsere Analyse zielt darauf ab, den Kapazitäts-"Pre-Log" zu bestimmen, d.h., das asymptotische Verhältnis zwischen der Kapazität und des Logarithmus des SNR, wenn das SNR groß wird. Wir leiten eine untere Schranke für den Kanalkapazitäts-Pre-Log her und zeigen, dass dieser im SIMO Fall im Allgemeinen größer ist als im Fall von einem einzelnen Eingang und einem einzelnen Ausgang (SISO). Dieses Resultat ist überraschend, da im kohärenten Fall der Kapazitäts-Pre-Log eines SIMO Schwundkanals genau gleich wie jener eines SISO Schwundkanals ist.

Acknowledgments

Throughout my life I was extremely lucky to always have great teachers! I am deeply indebted to each of them for inspiring my interest in learning and for shaping me as a person. I feel it is the right moment to thank some of these people and to summarize how my interest in science was formed under their influence.

I must start by expressing my gratitude to Valentina Al'bertovna Pryadko, my elementary school teacher. Valentina Al'bertovna taught us with great joy and passion, as if we were her own kids. Day after day, for many years, she not only taught us the school subjects, but also how to be a better person. She could be extremely proud if she would realize how much better she made this world with immense amount of kindness that she managed to pass to her students over the years.

I am indebted to my mathematics teacher Yury Vladimirovich Berhman for his inspiring teaching. With his youthful attitude and courage to experiment he converted the standard elementary-school curriculum into the math that was fun! If it was not for this, I would probably never develop my interest in mathematics any further. Yury Vladimirovich also introduced me to alpine skiing. Once he took me to ski the little hills in Orehovo, near St. Petersburg. Back then, those little hills with outdated ski-lifts felt like real mountains. Since that day my interest in the alpine skiing grew into a strong passion, which played a role in bringing me to Switzerland.

Outside of school I was greatly influenced by Vladimir Balakirsky. Visiting my parents, he would often ask me to solve the simplified

ACKNOWLEDGMENTS

versions of the research problems in Information Theory he was working on. This was the first research I have ever done. And I loved it.

I was lucky to study in the most amazing high school I can imagine: The Laboratory of Continuous Mathematical Education in Anichkov Lyceum. The Laboratory was created and managed by Ilya Alexandrovich Chistiakov. His ambition was to create a high school with exquisite quality of education for kids interested in science. He fully succeeded in that. The system created by Ilya Alexandrovich has transformed me from a student with an interest in math into a mathematician. Never before and never after my high school years, was I motivated to work so hard. I was rewarded for my efforts: In short three years in Anichkov Lyceum I learned no less than half of mathematics and physics that I know now. Moreover, Ilya Alexandrovich taught me how to communicate my ideas clearly. The intense education in Anichkov Lyceum convinced me that I can accomplish anything I target. Anichkov Lyceum also gave me my closest friends.

In Anichkov Lyceum I met Yury Alexandrovich Antonov, my teacher of Physics. His lectures were captivating and inspiring. He always managed to demonstrate us with just a piece of chalk, fingers, and words (we never had an experimental lab) how beautiful and intriguing the laws of Nature are. Even more importantly, he showed us how amazing is the fact that human scientists were able to discover these laws. After learning from Antonov, physics became my religion. I learned to think about problems in a “physical” way. This helps me in my research immensely.

Yury Alexandrovich introduced me to lectures by Richard Feynman. Feynman’s joy about intellectual discovery became a model for me. I keep turning to Feynman’s legacy whenever I need inspiration.

I want to thank Edward A. Hirsch for the excellent lectures on Theoretical Computer Science I attended during my university years.

I must express my deepest gratitude to my advisor Prof. Helmut Bölcskei. Without him this work would have never been possible. I marvel Helmut’s style of supervising his students. He constantly generates a rich stream of research ideas, out of which I could always

ACKNOWLEDGMENTS

choose the one I liked to develop. Whenever I wanted to pursue my own idea, I always had a possibility to do so. When it came to writing down or presenting the results, Helmut worked closely with me to achieve the highest quality possible. I will never forget his ultra-detailed comments and questions to each sentence of my drafts, which would uncover mathematical errors, lead to new research directions, or simply improve my writing style. I will never be able to reproduce Helmut's level of care and attention to details when delivering talks and writing papers. If I manage to take ten percent of what he tried to teach me, my writing will be nearly perfect. Finally, I would like to thank Helmut for the unlimited freedom he provided me in shaping my Ph.D. Not only could I freely choose the topics I wanted to work on, but also I could work on those topics when and where I felt necessary. Due to Helmut's flexibility, during my Ph.D. I could allocate large amounts of time to learning advanced physics, studying German, and pursuing my mountaineering dreams; I enjoyed doing these things immensely.

I would like to thank Prof. P.R. Kumar for hosting me in UIUC for three months. The ideas initiated in the weekly discussions with him led me to the research that constitutes the second part of this thesis.

I would like to thank Prof. Amos Lapidot and Prof. İ. Emre Telatar for agreeing to read my work. Their comments on this thesis were most helpful. It was a great honor for me to have the scientist I greatly admire as my thesis referees.

I would like to thank the people, who made my stay in Zürich truly enjoyable, and who became my great friends over the years. Chris Peel for introducing me to rock-climbing. Dima Geshkenbein for his patience in teaching me alpinism. Vassily Samarin for sharing my passion for freeriding. Ateet Kapur for teaching me to enjoy life "the French way". Moritz Borgmann for spending countless late-night-shifts in the lab together, for giving me a very helpful feedback on my first (failed) attempt to deliver a research presentation, and for creating all the computer infrastructure I use every day. Cemal Akçaba for the great days we spent sharing our office and for the wonderful time in Cyprus. Patrick Kuppinger, Pedro Coronel, and Daniel Baum

ACKNOWLEDGMENTS

for teaching me how to play table-football. Jatin Thukral for the discussions during our dinners in the mensa. Ulrich Schuster for teaching me how to use L^AT_EX the right way, for parties in his student house, and for giving me a place to live during my long struggle to find accommodation in Zürich. Giuseppe Durisi and Erwin Riegler for collaborating on the hardest problems I ever managed to solve. Reinhard Heckel, Graeme Pope, Davide Cescato, Christoph Studer, Christoph Bunte, Michèle Wigger, and Toby Koch for a lot of fun time!

A very special “Thank You!” goes to my fiancée Marina. I’m grateful to her for each moment we spend together, for her unconditional love and support. I’m also indebted to her for the readiness to follow me in the long path towards my future research career; often at the expense of her own career success and plans for the future. I hope I will not disappoint her.

Finally, I want to thank my parents. There are no words that could express my gratitude to them.

Contents

Acknowledgments	xiii
1. Motivation and Overview	1
1.1. Wireless Channels	2
1.1.1. Interference	2
1.1.2. Fading	3
1.2. Outline and Contributions	5
1.2.1. Relaying in Wireless Networks	5
1.2.2. Noncoherent Wireless Point-to-Point Channels	6
I Relaying in Wireless Networks	9
2. “Crystallization” in Networks with Coherent Relaying	11
2.1. Contributions and Relation to Previous Work	12
2.2. Channel and Signal Model	16
2.2.1. General Assumptions	16
2.2.2. Channel and Signal Model	16
2.3. Protocols in the Coherent-Relaying Case	19
2.3.1. Protocol 1 (P1)	19
2.3.2. Protocol 2 (P2)	22
2.3.3. Large-Deviations Analysis of Signal-to-Interference- Plus-Noise Ratio	25
2.3.4. Concentration Results for P1 and P2	37

CONTENTS

2.4.	Ergodic Capacity and Cooperation at the Relay Level	39
2.4.1.	Ergodic Capacity of P1 and P2	39
2.4.2.	The “Crystallization” Phenomenon	48
2.4.3.	Cooperation at the Relay Level	54
2.5.	Summary of Results	62
3.	Networks with Noncoherent Amplify-and-Forward Relaying	65
3.1.	Contributions and Relation to Previous Work	65
3.2.	The Amplify-and-Forward Protocol	66
3.3.	Capacity of the Amplify-and-Forward Protocol	67
3.4.	Asymptotic Capacity Behavior	68
3.5.	Convergence to Point-to-Point MIMO Channel	76
3.6.	Summary of Results	78
II	Noncoherent Wireless Point-to-Point Channels	81
4.	Noncoherent WSSUS Channel	83
4.1.	System Model	85
4.1.1.	The Continuous-Time Input-Output Relation	85
4.1.2.	A Robust Definition of Underspread Channels	88
4.2.	The Information Capacity	89
4.2.1.	Outline of the Information-Theoretic Analysis	91
4.3.	A Lower Bound on Capacity	92
4.3.1.	Discretization of the Input-Output Relation	92
4.3.2.	Orthonormality, Completeness, and Localization	95
4.3.3.	Statistical Properties	96
4.3.4.	Input-Output Relation in Vector-Matrix Form	98
4.3.5.	Definition of the Capacity of the Discretized Channel Induced by (g, T, F)	99
4.3.6.	The Capacity Lower Bound	100
4.3.7.	Reduction to a Square Setting	102
4.3.8.	Pulse-Design Criterion and Approximation for m_g and M_g	103
4.3.9.	A Simple WH Set	105

CONTENTS

4.4. Finite-SNR Analysis of the Lower Bound	106
4.4.1. Trade-off between Interference Reduction and Maximization of Number of Degrees of Freedom	106
4.4.2. Sensitivity of Capacity to the Parameters $\Delta_{\mathbb{H}}$ and ϵ	108
4.5. Summary of Results	110
5. Noncoherent SIMO Channel	113
5.1. System Model	114
5.2. Intuitive Analysis	116
5.3. A Lower Bound on the Capacity Pre-Log	118
5.3.1. A Lower Bound on $h(\mathbf{y})$	120
5.4. Summary of Results	128
6. Conclusions and Outlook	131
A. Truncation of Random Variables and Large Deviations	135
B. Union Bounds	143
C. Technical Results from Chapter 2	147
C.1. Proof of Theorem 2.1	147
C.1.1. Analysis of $S^{(1)}$	147
C.1.2. Analysis of $S^{(2)}$	148
C.1.3. Analysis of $S^{(4)}$	149
C.2. Proof of Lower Bound in Theorem 2.3	153
C.3. Lower Bound on Channel Capacity with Imperfect Channel Knowledge	155
D. Some Essentials from Large Random-Matrix Theory	157
E. A calculation for Chapter 3	159
F. Technical Results from Chapter 4	163
F.1. AWGN-Capacity Upper Bound	163
F.2. The Transmit Signal in (4.16) Satisfies (4.4)	164

CONTENTS

F.3. Some Properties of the Ambiguity Function	168
F.4. Proof of Theorem 4.1	170
F.5. Proof of Corollary 4.2	176
F.5.1. The log det Term	176
F.5.2. Bounds on $r[0, 0]$ and on σ_I^2	179
F.6. Proof of Lemma 4.3	181
F.7. Proof of Lemma 4.4	182
F.7.1. Derivative-Based Approximation for m_g	182
F.7.2. Derivative-Based Approximation for M_g	183
G. A Result from Linear Algebra	187
H. Notation	189
H.1. Miscellaneous	189
H.2. Functions	190
H.3. Operators	190
H.4. Probability Theory	191
H.5. Linear Algebra	192
H.6. Communication and Information Theory	193
I. Acronyms	195
References	197
List of Publications	203
Curriculum Vitae	205

CHAPTER 1

Motivation and Overview

THE largest amount of information per unit of time that can be *reliably* transmitted over a communication channel is called the *achievable rate*. Can *noisy* communication channels support *strictly positive* achievable rates? The answer to this question was first found by Claude Shannon in 1948 and turned out to be “yes” (Shannon, 1948). Shannon characterized the channel through a probabilistic model, where the input and the output of the channel are represented by dependent random variables (RVs). The conditional distribution of the channel output given its input is referred to as the channel law. Given this law, Shannon calculated a certain number, called the *channel capacity*, which is equal to the supremum over all admissible input distributions of the mutual information between the input of the channel and the output of the channel. He then proved that using *coding* it is possible to communicate reliably at an arbitrary rate below the channel capacity, i.e., all rates below the capacity are achievable rates. A converse to this statement also exists: rates above the channel capacity are not achievable.

In his seminal paper, Shannon computed the capacity of the discrete memoryless channel (DMC) and the additive white Gaussian noise (AWGN) channel. These basic results provided general guidelines on how to design communication systems. It then took many years of extensive research and progress in physics, semiconductor technology, coding theory, and signal processing to convert this theoretical under-

standing into practice. As a result, nowadays we can enjoy lightning fast, reliable, and cheap communication systems.

1.1. WIRELESS CHANNELS

The main features of the DMC and the AWGN channel model describe the majority of the real-world *wireline* channels (communication over wires) very well. These and similar models can also be used to characterize *wireless* channels where a single transmitter communicates with a single receiver located in its line of sight (LOS) and both of them do not move too fast. In general, however, the situation in the wireless world is much richer and much more complicated. This is because two completely new effects come into play: the effect of *interference* and the effect of *fading*. We discuss these effects next.

1.1.1. Interference

All users willing to communicate “over the air” unavoidably share the same medium: the physical space with the electromagnetic waves in it. Therefore, each user’s transmitted signal interferes with the signals transmitted by the other users in the system. The use of *orthogonalization* in signal space can mitigate interference:

- Separate users in time, such that only one user at a time is allowed to transmit.
- Separate users in frequency, such that different frequency-bands are allocated to different users communicating concurrently.
- Separate users in space, such that different regions of space are allocated to different users communicating concurrently and in the same frequency-band; exploit the fact that interfering signals coming from far away get attenuated due to path loss and shadowing in the channel.

These three separation methods are used widely in modern wireless communication systems: i) the time division multiple access (TDMA) technique realizes separation in time; ii) the frequency division multi-

ple access (FDMA) technique achieves separation in frequency; iii) the interference avoidance pattern (Tse and Viswanath, 2005, Section 4.2) in the cell structure of the global system for mobile communication (GSM) provides spatial separation of the users: the users located in the adjacent cells are not allowed to communicate concurrently and in the same frequency-band to avoid interference, whereas the users located in the cells that are not adjacent are allowed communicate concurrently and in the same frequency-band, because path loss eliminates interference. These and similar techniques allow to mitigate multi-user interference and to reduce the problem of communicating over the shared wireless medium to the point-to-point communication problem (i.e., communication between a single transmitter and a single receiver), which is much better understood.

The downside of using orthogonalization techniques is that they do not allow for simultaneous use of the main resources of a wireless channel: time, frequency, and space. The division of the resources through orthogonalization is potentially highly suboptimal from an information-theoretic perspective. It leads, in general, to loss in spectral efficiency, which reduces system capacity. The consequences of this can be observed today, as the wide use of smartphones threatens to saturate the capacity of existing wireless networks.

It is, therefore, paramount to understand the impact of interference on the capacity of wireless networks where several users operate in the shared medium and orthogonalization in time, frequency, or space is not enforced by the system designer.

In Part I of this thesis, consisting of Chapters 2 and 3, we study one aspect of this general problem; specifically, we explore how *relays* can be used to improve spectral efficiency in wireless interference networks.

1.1.2. Fading

In wireline channels, communication takes place in an essentially static medium, to a large extent controlled by the system designer. In wireless channels, in contrast, the medium, in general, is dynamic

and hard to control. The state of the channel changes with time due to movements of objects in the environment. When the wireless transmitter and receiver are located in direct LOS of each other, the variations of the channel state over the time-interval of interest are insignificant and can be neglected. However, in situations where the transmitter and the receiver are not located in direct LOS of each other, which is almost always the case in urban environments, multi-path wave propagation can create rapid variations of the channel state. Significant variations may occur on the time scale of interest, even when the transmitter, the receiver, and the objects in the propagation environment move relatively slowly. Channels with such variations are called *fading channels*.

When analyzing fading channels, it is common to assume that the receiver knows the channel state perfectly, i.e., *a genie* provides channel state information (CSI) to the receiver. This scenario is referred to as the *coherent* fading channel. However, in practical wireless systems such CSI must constantly be acquired and updated. One way to do this is to regularly transmit so-called *pilot symbols*, i.e., signals, known a priori to both the transmitter and the receiver. From the received pilot symbols, CSI can then be extracted.

Pilot symbols occupy dimensions in signal-space, which could otherwise be used for transmission of payload data. Depending on the CSI requirements and on the rate of change of the channel state, sending pilots can become very costly in terms of the number of signal-space dimensions sacrificed. From an information-theoretic point of view, sending pilots is just a special form of coding that may or may not be optimal. To characterize the cost of acquiring CSI fundamentally, one needs to study the capacity of the fading channel in the setting when *no* CSI is available at the receiver; this scenario is referred to as the *noncoherent* fading channel. The price to pay for developing this more fundamental understanding is that the analysis in the noncoherent case is usually much harder than that in the coherent case.

In Part II of this thesis, consisting of Chapters 4 and 5, we study the capacity of point-to-point noncoherent fading channels.

1.2. OUTLINE AND CONTRIBUTIONS

1.2.1. Relaying in Wireless Networks

Part I of the thesis is devoted to studying how relays can be used to improve spectral efficiency in interference wireless networks. Specifically, we analyze fading interference relay networks where M single-antenna source-destination terminal pairs communicate concurrently and in the same frequency band through a set of K single-antenna relays using half-duplex two-hop relaying. In this analysis we make two simplifying assumptions. First, we consider specific communication protocols and study the network capacity induced by these protocols, not the capacity of the network itself. Second, we are exclusively interested in the asymptotic capacity behavior when the number of terminals in the network grows large.

A. “Crystallization” in Networks with Coherent Relaying

In Chapter 2, we analyze a *coherent-relaying* protocol, where the relays have CSI and perform matched-filtering. The destination terminals cannot cooperate, i.e., the network operates in a completely distributed fashion. It is shown that in the large- M limit, provided K grows fast enough as a function of M , the network “decouples”. This means that the capacities of all individual source-destination pair links are strictly positive. The signal power in all individual source-destination pair links dominates the interference power in these links. Therefore, all users in the network can transmit concurrently and in the same frequency-band. Orthogonalization in the network is achieved by using relays, without loss in spectral efficiency.

The required (for the network to decouple) rate of growth of K as a function of M is found to be sufficient to also make the individual source-destination fading links converge to nonfading links. We say that the network “*crystallizes*” as it breaks up into a set of effectively isolated “*wires in the air*”. A large-deviations analysis is performed to characterize the “crystallization” rate, i.e., the rate (as a function

of M, K) at which the decoupled links converge to nonfading links.

B. Networks with Noncoherent Amplify-and-Forward Relaying

In Chapter 3, we consider the case of relays that do not have CSI and perform simple amplify-and-forward (AF) relaying. The destination terminals have CSI, and, in contrast to Chapter 2, are allowed to cooperate and perform joint decoding. Based on tools from large random-matrix theory, we compute the per source-destination terminal pair capacity of this network for $M, K \rightarrow \infty$, with $K/M \rightarrow \beta$ fixed. We also find that for $\beta \rightarrow \infty$, the AF relay network is turned into a point-to-point multiple-input multiple-output (MIMO) link. This result demonstrates that employing relays as “active scatterers” can recover spatial multiplexing gain in poor scattering environments.

1.2.2. Noncoherent Wireless Point-to-Point Channels

In Part II of the thesis, we consider exclusively systems with a single transmitter and a single receiver. We study the impact of absence of CSI at the receiver on channel capacity in two different models.

A. Noncoherent WSSUS Channel

The noncoherent capacity of stationary discrete-time fading channels is known to be very sensitive to the fine details of the channel model. More specifically, the measure of the set of harmonics where the power spectral density of the fading process is nonzero determines if capacity grows logarithmically in signal-to-noise ratio (SNR) or slower than logarithmically (Lapidoth, 2005). An engineering-relevant problem is to characterize the SNR value at which this sensitivity starts to matter.

An attempt to resolve this problem was made by Etkin and Tse (2006), where, for a first-order Gauss-Markov fading process, the SNR beyond which capacity behaves slower than logarithmically is computed as a function of the innovation rate of the process. The main limitation of this result is that it is based on a very specific

channel model and that it is difficult to link the innovation rate to physical channel parameters.

In Chapter 4, we attempt to address the problem in more generality. Rather than focusing on a specific discretized channel model, we start from the general model of continuous-time Rayleigh-fading channels that satisfy the wide-sense stationary (WSS) and uncorrelated scattering (US) assumptions (Bello, 1963). In addition, we assume that the channel is *underspread*, i.e., its scattering function is highly concentrated around the origin of the Doppler-delay plane. This is a highly relevant assumption for most real-life wireless channels and it simplifies our analysis significantly.

We show that the noncoherent capacity of the wide-sense stationary uncorrelated scattering (WSSUS) underspread channel is close to the AWGN capacity for all SNR values of practical interest, independently of whether the scattering function is compactly supported or not. As a byproduct of our analysis, we obtain an information-theoretic pulse-design criterion for pulse-shaped orthogonal frequency division multiplexing (PS-OFDM) systems.

B. Noncoherent SIMO Channel

In Chapter 5, we study the impact of multiple antennas at the receiver side of noncoherent channels. We resort to a much simpler channel model than the one used in Chapter 4. Specifically, we investigate the temporally correlated Rayleigh block-fading single-input multiple-output (SIMO) channel. Our analysis is aimed at characterizing the capacity pre-log, i.e., the asymptotic ratio between the capacity and the logarithm of SNR, as SNR goes to infinity.

We derive a lower bound on the channel capacity pre-log. When the covariance matrix of the channel satisfies a certain technical condition related to the cardinality of its smallest set of linearly dependent rows, this lower bound reveals that the capacity pre-log in the SIMO case is *larger* than that in the single-input single-output (SISO) case. The result is surprising, because in the *coherent* case the capacity pre-log of a SIMO fading channel is *the same as* that of a SISO fading

1. MOTIVATION AND OVERVIEW

channel (Telatar, 1999).

PART I

RELAYING IN WIRELESS NETWORKS

CHAPTER 2

“Crystallization” in Networks with Coherent Relaying

Sparked by the work of Gupta and Kumar (2002), the analysis of the capacity scaling behavior of large wireless (relay) networks has emerged as an interesting tool in network information theory, which often allows to make stronger statements than a finite-number-of-nodes analysis. The general idea of this approach is the following. First, define a measure of the capacity of the network. For example, this could be the sum of the capacities of the individual links in the network, or the distance-weighted sum of the capacities of the individual links (Gupta and Kumar, 2003). Then, consider the wireless network operating under a certain protocol. The protocol induces a certain effective capacity of the network, which is analyzed as a function of the number of users, in particular, when the number of users grows large; this function characterizes the capacity scaling law. Different protocols can then be compared on the basis of their induced capacity scaling laws. Sometimes optimality results can be proven, i.e., it can be shown that a given protocol achieves the best possible scaling law under certain assumptions on the network (Lévêque and Telatar, 2005).

This area of research has seen remarkable activity during the last decade (Xue and Kumar, 2006; Xie and Kumar, 2004; Gast-

par and Vetterli, 2005; Grossglauser and Tse, 2002; Jovičić et al., 2004; Franceschetti et al., 2007; Franceschetti and Meester, 2007; Wang et al., 2006; Dana and Hassibi, 2006; Bölcskei et al., 2006; Özgür et al., 2007). For a more detailed discussion of some of the work mentioned above, we refer the interested reader to the monographs by Xue and Kumar (2006), Franceschetti and Meester (2007), and to the Ph.D. thesis of Özgür (2009).

In this thesis, we study capacity scaling laws in interference fading relay networks operating under specific protocols as described in detail next.

2.1. CONTRIBUTIONS AND RELATION TO PREVIOUS WORK

Part I of the thesis deals with interference fading relay networks where M single-antenna source-destination terminal pairs communicate concurrently and in the same frequency band through half-duplex two-hop relaying over a common set of K single-antenna relay terminals (see Figure 2.1). Two setups are considered. In the present chapter, we analyze the *coherent-relaying* case, where the relays have CSI, perform matched-filtering, and the destination terminals cannot cooperate, i.e., the network operates in a completely distributed fashion. In Chapter 3, we study the *noncoherent-relaying* case, where the relays do not have CSI, perform AF relaying, and the destination terminals are assumed to cooperate and perform joint decoding.

Our main contributions in this chapter can be summarized as follows.

- We consider two different protocols, P1 introduced (for the finite- M case) by Bölcskei et al. (2006) and P2 introduced by Dana and Hassibi (2006). P1 relies on the idea of relay partitioning (i.e., each relay is assigned to one source-destination terminal pair) and requires each relay terminal to know its assigned backward (source to relay) and forward (relay to destination) channel only. The relays perform matched-filtering with respect to (w.r.t.) their

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

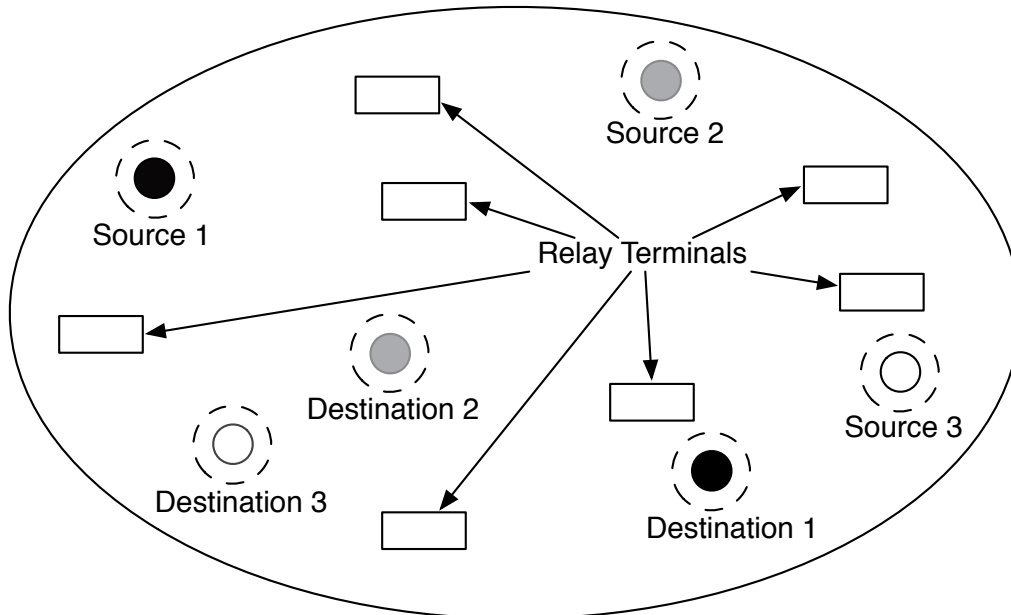


Fig. 2.1.: Dense wireless interference relay network with dead-zones around source and destination terminals. Each terminal employs one antenna.

assigned backward and forward channels. P2 does not use relay partitioning, requires each relay terminal to know all M backward and all M forward channels, and performs matched-filtering w.r.t. all M backward and M forward links, in a sense to be defined below.

Previous work for the coherent-relaying case has established the power efficiency scaling of P2 for $M \rightarrow \infty$ with $K = M^2$ (Dana and Hassibi, 2006). Bölcskei et al. (2006) have shown that for P1 with M fixed, in the $K \rightarrow \infty$ limit, network capacity¹ scales as $(M/2) \log(K) + \mathcal{O}(1)$. The results of Bölcskei et al. (2006) and the corresponding proof techniques, however, rely heavily on M being fixed when $K \rightarrow \infty$. When $M, K \rightarrow \infty$, the amount of interference (at each destination terminal) grows with M . Establishing the corresponding network capacity scaling behavior, therefore, requires fundamentally new techniques, which are de-

¹Throughout Part I of the thesis, when we talk about capacity, we mean the capacity induced by the considered protocols, not the capacity of the network itself. The term “network capacity” stands for the sum of the capacities of the individual channels between the source-destination terminal pairs.

veloped in this chapter. In particular, we derive the network (ergodic) capacity scaling behavior for $M, K \rightarrow \infty$ for P1 and P2 by computing a lower and an upper bound on the per source-destination terminal pair capacity², and by showing that the bounds exhibit the same scaling (in M, K) behavior. The technique used to establish the lower bound is based on a result found in a different context in the work of Médard (2000) and applied by Dana and Hassibi to derive the power efficiency scaling of P2. For our purposes, we need a slight generalization of the result in (Médard, 2000), which follows, in a straightforward fashion, from a result on nearest-neighbor decoding reported by Lapidoth and Shamai (Shitz) (2002). For the sake of completeness, we state, in Appendix C.3, the relevant inequality in the form needed in the context of this chapter. The corresponding upper bound on the per source-destination terminal pair capacity poses significantly more difficult technical challenges and is based on a large-deviations analysis of the individual link signal-to-interference-plus-noise ratio (SINR) RVs. In summary, we prove that in the large- M limit, provided the number of relay terminals K grows fast enough as a function of M , under both protocols P1 and P2 the network “decouples” in the sense that the individual source-destination terminal pair (ergodic) capacities are strictly positive. The corresponding minimum (for the network to decouple) rates of growth are $K \propto M^3$ for P1 and $K \propto M^2$ for P2, with the per source-destination terminal pair capacity scaling (for $M, K \rightarrow \infty$) given by $(1/2) \log(1 + \Theta(K/M^3))$ and $(1/2) \log(1 + \Theta(K/M^2))$, respectively. The protocols P1 and P2 thus trade off CSI at the relays for the required rate of growth of the number of relays. We hasten to add that an ergodic-capacity lower bound for P2 was previously established by Dana and Hassibi; this bound is restated (and reproved under slightly different assumptions) in this chapter for the sake of completeness. It

²The per source-destination terminal pair capacity is the network capacity divided by the number of source-destination pairs.

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

appears, however, that Dana and Hassibi do not establish the minimum rate of growth of the number of relays for the network to decouple.

- We analyze the network outage probability behavior induced by P1 and P2 using a large-deviations approach. More specifically, we show that the growth rates $K \propto M^3$ in P1 and $K \propto M^2$ in P2 are sufficient to not only make the network decouple, but also to make the individual source-destination fading links converge to nonfading links. We say that the *network “crystallizes”* as it breaks up into a set of *effectively isolated “wires in the air”*. Each of the decoupled links experiences distributed spatial diversity (or relay diversity), with the corresponding diversity order going to infinity as $M \rightarrow \infty$. Consequently, in the large- M limit, time diversity (achieved by coding over a sufficiently long time horizon) is not needed to achieve ergodic capacity. We obtain bounds on the outage probability of the individual source-destination links, which allow to characterize the “crystallization” rate (more precisely a guaranteed “crystallization” rate as we do not know whether our bounds are tight), i.e., the rate (as a function of M, K) at which the decoupled links converge to nonfading links. In the course of this analysis, we develop a new technique for characterizing the large-deviations behavior of certain sums of dependent RVs. This technique builds on the well-known truncation approach and is reported in Appendix A.
- For P1 and P2, we establish the impact of cooperation at the relay level on network (ergodic) capacity scaling. More specifically, it is shown that, asymptotically in M and K , cooperation (realized by vector matched filtering) in groups of L relays leads to an L -fold reduction in the total number of relays needed to achieve a given per source-destination terminal pair capacity.

2.2. CHANNEL AND SIGNAL MODEL

In this section, we present the channel and signal model and additional basic assumptions. We restrict ourselves to the aspects that apply to both coherent-relaying and noncoherent-relaying networks and to both protocols considered in the coherent-relaying case. Relevant specifics for the coherent-relaying case will be provided in Sections 2.3.1 and 2.3.2 and for the noncoherent-relaying case in Chapter 3.

2.2.1. General Assumptions

We consider an interference relay network (see Figures 2.1 and 2.2) consisting of $K + 2M$ single-antenna terminals with M designated source-destination terminal pairs $\{\mathcal{S}_m, \mathcal{D}_m\}$ ($m \in [1:M]$) and K relays \mathcal{R}_k ($k \in [1:K]$). We assume a “dead-zone” of non-zero radius, free of relays, around each of the source and destination terminals, no direct link between the individual source-destination terminal pairs (e.g., due to large separation), and a domain of fixed area (i.e., dense network assumption). Transmission takes place in half-duplex fashion (the terminals cannot transmit and receive simultaneously) in two hops (a.k.a. two-hop relaying) over two disjoint time slots. In the first time slot, the source terminals simultaneously broadcast their information to all the relay terminals (i.e., each relay terminal receives a superposition of all source signals). After processing the received signals, the relay terminals simultaneously broadcast the processed data to all the destination terminals during the second time slot. Our setup can be considered as an interference channel (Carleial, 1978) with dedicated relays, hence the terminology *interference relay network*.

2.2.2. Channel and Signal Model

Throughout Part I of the thesis, frequency-flat fading over the bandwidth of interest as well as perfectly synchronized transmission and reception between the terminals is assumed. The input-output (IO) re-

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

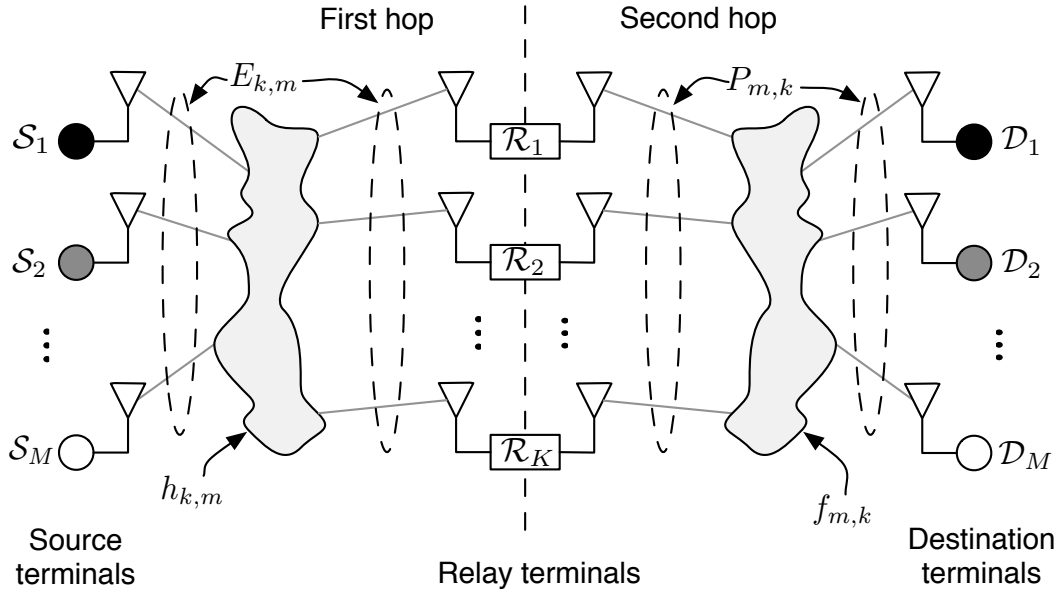


Fig. 2.2.: Two-hop wireless relay network setup.

lation for the link between the source terminals and the relay terminals during the first time slot is given by

$$\mathbf{r} = (\mathbf{E} \odot \mathbf{H}) \mathbf{s} + \mathbf{z} \quad (2.1)$$

where $\mathbf{r} = [r_1 \ r_2 \ \cdots \ r_K]^\top$ with r_k denoting the signal received at the k th relay terminal, $\mathbf{E} \in \mathbb{R}^{K \times M}$ with $[\mathbf{E}]_{k,m} = \sqrt{E_{k,m}}$ where $E_{k,m}$ denotes the average energy received at \mathcal{R}_k through the $\mathcal{S}_m \rightarrow \mathcal{R}_k$ link³ (accounting for path loss and shadowing in the $\mathcal{S}_m \rightarrow \mathcal{R}_k$ link), $\mathbf{H} \in \mathbb{C}^{K \times M}$ with $[\mathbf{H}]_{k,m} = h_{k,m}$ ($k \in [1:K]$, $m \in [1:M]$) where $h_{k,m} \sim \mathcal{CN}(0, 1)$ denotes the independent identically distributed (i.i.d.) complex-valued channel gains corresponding to the $\mathcal{S}_m \rightarrow \mathcal{R}_k$ links, $\mathbf{s} = [s_1 \ s_2 \ \cdots \ s_M]^\top$ where s_m is the zero-mean Gaussian signal transmitted by \mathcal{S}_m and the vector \mathbf{s} is i.i.d. temporally and spatially (across source terminals). Finally, $\mathbf{z} = [z_1 \ z_2 \ \cdots \ z_K]^\top$ where $z_k \sim \mathcal{CN}(0, \sigma^2)$ is temporally and spatially (across relay terminals) white noise. The k th relay terminal processes its received signal r_k to produce the output signal t_k . The collection of output signals t_k , organized in the vector $\mathbf{t} = [t_1 \ t_2 \ \cdots \ t_K]^\top$, is then broadcast to the destination terminals during the second time slot, while the source

³ $\mathcal{A} \rightarrow \mathcal{B}$ signifies communication from terminal \mathcal{A} to terminal \mathcal{B} .

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

terminals are silent. The m th destination terminal receives the signal y_m with $\mathbf{y} = [y_1 \ y_2 \ \cdots \ y_M]^\top$ given by

$$\mathbf{y} = (\mathbf{P} \odot \mathbf{F}) \mathbf{t} + \mathbf{w} \quad (2.2)$$

where $\mathbf{P} \in \mathbb{R}^{M \times K}$ with $[\mathbf{P}]_{m,k} = \sqrt{P_{m,k}}$ and $P_{m,k}$ denotes the average energy received at \mathcal{D}_m through the $\mathcal{R}_k \rightarrow \mathcal{D}_m$ link (having accounted for path loss and shadowing in the $\mathcal{R}_k \rightarrow \mathcal{D}_m$ link). Furthermore, $\mathbf{F} \in \mathbb{C}^{M \times K}$ with $[\mathbf{F}]_{m,k} = f_{m,k}$ ($m \in [1:M], k \in [1:K]$) where $f_{m,k} \sim \mathcal{CN}(0, 1)$ denotes the i.i.d. complex-valued channel gains corresponding to the $\mathcal{R}_k \rightarrow \mathcal{D}_m$ links, and $\mathbf{w} = [w_1 \ w_2 \ \cdots \ w_M]^\top$ with $w_m \sim \mathcal{CN}(0, \sigma^2)$ being temporally and spatially (across destination terminals) white noise. Throughout Part I of the thesis, we impose a per-source-terminal power constraint

$$\mathbb{E}[|s_m|^2] \leq 1/M, \quad m \in [1:M]$$

which results in the total transmit power trivially satisfying $\mathbb{E}[\|\mathbf{s}\|^2] \leq 1$. Furthermore, we impose a per-relay-terminal power constraint

$$\mathbb{E}[|t_k|^2] \leq P_{\text{rel}}/K, \quad k \in [1:K]$$

which results in the total power transmitted by the relay terminals satisfying $\mathbb{E}[\|\mathbf{t}\|^2] \leq P_{\text{rel}}$. As already mentioned above, path loss and shadowing are accounted for through the $E_{k,m}$ ($k \in [1:K], m \in [1:M]$) (for the first hop) and the $P_{m,k}$ ($m \in [1:M], k \in [1:K]$) (for the second hop). We assume that these parameters are deterministic, uniformly bounded from above (follows from the dead-zone assumption) and below (follows from considering a domain of fixed area) so that for all k, m

$$0 < \underline{E} \leq E_{k,m} \leq \bar{E} < \infty \quad 0 < \underline{P} \leq P_{m,k} \leq \bar{P} < \infty. \quad (2.3)$$

Throughout Part I of the thesis, we assume that the source terminals \mathcal{S}_m ($m \in [1:M]$) do not have CSI. The assumptions on CSI at the relays and the destination terminals depend on the setup (coherent-relaying or noncoherent-relaying case) and the protocol (in the coherent-relaying case) and will be made specific when needed.

A discussion of the motivation for the two scenarios analyzed in this chapter can be found in the work of Bölcskei et al. (2006).

2.3. PROTOCOLS IN THE COHERENT-RELAYING CASE

In this section, we describe the two protocols P1 and P2 and derive the corresponding SINR concentration results along with the resulting bounds on the individual source-destination link outage probability induced by P1 and P2. Note that the results in this section do not require ergodicity of \mathbf{H} and \mathbf{F} .

2.3.1. Protocol 1 (P1)

We shall next describe the specifics of P1. The K relay terminals are partitioned into M subsets \mathcal{M}_m ($m \in [1:M]$) with⁴ $|\mathcal{M}_m| = K/M$. The relays in \mathcal{M}_m are assumed to assist the m th source-destination terminal pair $\{\mathcal{S}_m, \mathcal{D}_m\}$. This assignment is succinctly described through the relay partitioning function $p : [1, K] \rightarrow [1, M]$ defined as

$$p(k) \triangleq m \Leftrightarrow \mathcal{R}_k \in \mathcal{M}_m.$$

We assume that the k th relay terminal has perfect knowledge of the phases $\arg(h_{k,p(k)})$ and $\arg(f_{p(k),k})$ of the SISO backward (from the perspective of the relay) channel $\mathcal{S}_{p(k)} \rightarrow \mathcal{R}_k$ and the corresponding forward channel $\mathcal{R}_k \rightarrow \mathcal{D}_{p(k)}$, respectively. We furthermore define $\tilde{h}_{k,p(k)} \triangleq \exp(i \arg(h_{k,p(k)}))$ and $\tilde{f}_{p(k),k} \triangleq \exp(i \arg(f_{p(k),k}))$. The signal r_k received at the k th relay terminal is first cophased w.r.t. the assigned backward channel followed by an energy normalization so that

$$u_k = d_{P1,k} \tilde{h}_{k,p(k)}^* r_k \quad (2.4)$$

⁴For simplicity, we assume that K is an integer multiple of M . Moreover, in the remainder of this chapter all results pertaining to P1 implicitly assume $K \geq M$.

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

where

$$d_{\text{P1},k} \triangleq \sqrt{P_{\text{rel}}} \left[\frac{K}{M} \sum_{m=1}^M E_{k,m} + K\sigma^2 \right]^{-1/2} \quad (2.5)$$

ensures that the per-relay power constraint $\mathbb{E}[|u_k|^2] = P_{\text{rel}}/K$ is met. The relay terminal \mathcal{R}_k then computes the transmit signal t_k by cophasing w.r.t. its assigned forward channel, i.e.,

$$t_k = \tilde{f}_{p(k),k}^* u_k \quad (2.6)$$

which, obviously, satisfies $\mathbb{E}[|t_k|^2] \leq P_{\text{rel}}/K$ with equality and hence meets the total power constraint (across relays)

$$\mathbb{E}[\|\mathbf{t}\|^2] = \sum_{k=1}^K \mathbb{E}[|t_k|^2] = P_{\text{rel}}.$$

In summary, P1 ensures that the relays $\mathcal{R}_k \in \mathcal{M}_m$ forward the signal intended for \mathcal{D}_m , namely, the signal transmitted by \mathcal{S}_m , in a “doubly coherent” (w.r.t. backward and forward channels) fashion, whereas the signals transmitted by the source terminals $\mathcal{S}_{\hat{m}}$ with $\hat{m} \neq m$ are forwarded to \mathcal{D}_m in a “noncoherent” fashion (i.e., phase incoherence occurs either on the backward or the forward link or on both links). The idea underlying P1 has originally been introduced by Bölcskei et al. (2006).

We shall next derive the IO relation for the SISO channels $\mathcal{S}_m \rightarrow \mathcal{D}_m$ ($m \in [1:M]$). The destination terminal \mathcal{D}_m receives doubly (backward and forward link) coherently combined contributions corresponding to the signal s_m , with interfering terms containing contributions from the signals $s_{\hat{m}}$ with $\hat{m} \neq m$ as well as noise, forwarded by the relays. Combining (2.1), (2.4), (2.6), and (2.2), it follows (after some straightforward algebra) that the signal received at \mathcal{D}_m ($m \in [1:M]$) is given by⁵

⁵The notation $\sum_{\hat{m} \neq m}$ stands for the summation over $\hat{m} \in [1:M]$ such that $\hat{m} \neq m$. If not specified, the upper limit of the summation is clear from the context.

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

$$\begin{aligned}
 y_m = s_m & \underbrace{\frac{1}{\sqrt{K}} \sum_{k=1}^K a_k^{m,m}}_{\text{effective channel gain}} \\
 & + \underbrace{\sum_{\hat{m} \neq m} s_{\hat{m}} \frac{1}{\sqrt{K}} \sum_{k=1}^K a_k^{m,\hat{m}}}_{\text{interference}} + \underbrace{\frac{1}{\sqrt{K}} \sum_{k=1}^K b_k^m z_k + w_m}_{\text{noise}} \quad (2.7)
 \end{aligned}$$

where

$$a_k^{m,\hat{m}} \triangleq C_{\text{P1},k}^{m,\hat{m}} \tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* h_{k,\hat{m}} \quad (2.8)$$

$$b_k^m \triangleq C_{\text{P1},k}^m \tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* \quad (2.9)$$

with

$$C_{\text{P1},k}^{m,\hat{m}} = \sqrt{K} d_{\text{P1},k} \sqrt{P_{m,k} E_{k,\hat{m}}} \quad (2.10)$$

$$C_{\text{P1},k}^m = \sqrt{K} d_{\text{P1},k} \sqrt{P_{m,k}}. \quad (2.11)$$

The normalization factor \sqrt{K} in (2.7), (2.10), and (2.11) is introduced for convenience of exposition. Using (2.3), it now follows that

$$\underline{C} \triangleq \sqrt{\frac{P \underline{E} P_{\text{rel}}}{\underline{E} + \sigma^2}} \leq C_{\text{P1},k}^{m,\hat{m}} \leq \sqrt{\frac{\overline{P} \overline{E} P_{\text{rel}}}{\underline{E} + \sigma^2}} \triangleq \overline{C} \quad (2.12)$$

$$\underline{c} \triangleq \sqrt{\frac{P P_{\text{rel}}}{\underline{E} + \sigma^2}} \leq C_{\text{P1},k}^m \leq \sqrt{\frac{\overline{P} P_{\text{rel}}}{\underline{E} + \sigma^2}} \triangleq \overline{c} \quad (2.13)$$

for all $k \in [1:K]$, $m \in [1:M]$, and $\hat{m} \in [1:M]$. In the following, it will be essential that the constants \underline{C} , \underline{c} , \overline{C} , and \overline{c} do not depend on M, K .

Since we assumed that the destination terminals \mathcal{D}_m ($m \in [1:M]$) cannot cooperate, the \mathcal{D}_m cannot perform joint decoding so that the network can be viewed as a collection of M SISO channels $\mathcal{S}_m \rightarrow \mathcal{D}_m$, i.e., as an interference channel with dedicated relays. We can see from (2.7) that each of these SISO channels consists of a fading effective channel, fading interference, caused by the source signals not intended

for a given destination terminal, and finally a noise term incorporating thermal noise forwarded by the relays and thermal noise added at the destination terminals. In the remainder of this chapter, we make the assumption that each of the destination terminals \mathcal{D}_m has perfect knowledge of the fading and path loss and shadowing coefficients in the entire network, i.e., \mathcal{D}_m ($m \in [1:M]$) knows \mathbf{H} , \mathbf{F} , \mathbf{E} and \mathbf{P} perfectly. An immediate consequence of this assumption is that \mathcal{D}_m ($m \in [1:M]$) has perfect knowledge of the effective channel gain $(1/\sqrt{K}) \sum_{k=1}^K a_k^{m,m}$, the interference channel gains $(1/\sqrt{K}) \sum_{k=1}^K a_k^{m,\hat{m}}$ ($\hat{m} \neq m$), and the quantity $(1/\sqrt{K}) \sum_{k=1}^K b_k^m$. Conditioned on \mathbf{H} and \mathbf{F} , both the interference and the noise term in (2.7) are Gaussian, so that the mutual information for the $\mathcal{S}_m \rightarrow \mathcal{D}_m$ link is given by

$$I(y_m; s_m | \mathbf{H}, \mathbf{F}) = \log(1 + \text{SINR}_m^{\text{P1}}) \quad (2.14)$$

where

$$\text{SINR}_m^{\text{P1}} \triangleq \frac{\left| \sum_{k=1}^K a_k^{m,m} \right|^2}{\sum_{\hat{m} \neq m} \left| \sum_{k=1}^K a_k^{m,\hat{m}} \right|^2 + \sigma^2 M \sum_{k=1}^K |b_k^m|^2 + KM\sigma^2} \quad (2.15)$$

is the effective SINR in the SISO channel $\mathcal{S}_m \rightarrow \mathcal{D}_m$.

We conclude by noting that the large-deviations results in Section 2.3.3 rely heavily on the assumption that \mathcal{D}_m ($m \in [1:M]$) knows \mathbf{H} , \mathbf{F} , \mathbf{E} , and \mathbf{P} perfectly. The ergodic capacity-scaling results in Section 2.4 will, however, be seen to require significantly less channel knowledge at the destination terminals.

2.3.2. Protocol 2 (P2)

The only difference between P1 and P2 is in the processing at the relays. Whereas in P1 the K relay terminals are partitioned into M clusters (of equal size) with each of these clusters assisting one particular source-destination terminal pair, in P2 each relay assists all source-destination terminal pairs so that relay partitioning is not

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

needed. In turn, P2 requires that each relay knows the phases of all its M backward and M forward channels, i.e., \mathcal{R}_k needs knowledge of $\tilde{h}_{k,m}$ and $\tilde{f}_{m,k}$, respectively, for $m \in [1:M]$. Consequently, P2 requires significantly more CSI at the relays than P1. The relay processing stage in P2 computes

$$t_k = d_{\text{P2},k} \left(\sum_{m=1}^M \tilde{h}_{k,m}^* \tilde{f}_{m,k}^* \right) r_k \quad (2.16)$$

where

$$d_{\text{P2},k} \triangleq \sqrt{P_{\text{rel}}} \left[K \sum_{m=1}^M E_{k,m} + MK\sigma^2 \right]^{-1/2}$$

ensures that the power constraint $\mathbb{E}[|t_k|^2] = P_{\text{rel}}/K$ and hence

$$\mathbb{E}[\|\mathbf{t}\|^2] = \sum_{k=1}^K \mathbb{E}[|t_k|^2] = P_{\text{rel}}$$

is met.

Again, we start by deriving the IO relation for the SISO channels $\mathcal{S}_m \rightarrow \mathcal{D}_m$ ($m \in [1:M]$). Like in P1, the destination terminal \mathcal{D}_m receives doubly (backward and forward link) coherently combined contributions corresponding to the signal s_m , interfering terms containing contributions from the signals $s_{\hat{m}}$ with $\hat{m} \neq m$, as well as noise forwarded by the relays. Combining (2.1), (2.16), and (2.2), it follows that the signal received at \mathcal{D}_m ($m \in [1:M]$) is given by

$$y_m = s_m \underbrace{\frac{1}{\sqrt{KM}} \sum_{k=1}^K \sum_{\tilde{m}=1}^M a_k^{m,m,\tilde{m}}}_{\text{effective channel gain}} + \underbrace{\sum_{\hat{m} \neq m} s_{\hat{m}} \frac{1}{\sqrt{KM}} \sum_{k=1}^K \sum_{\tilde{m}=1}^M a_k^{m,\hat{m},\tilde{m}}}_{\text{interference}}$$

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

$$+ \underbrace{\frac{1}{\sqrt{KM}} \sum_{k=1}^K \sum_{\tilde{m}=1}^M b_k^{m,\tilde{m}} z_k}_{\text{noise}} + w_m \quad (2.17)$$

where

$$\begin{aligned} a_k^{m,\hat{m},\tilde{m}} &\triangleq C_{\text{P2},k}^{m,\hat{m}} \tilde{f}_{\tilde{m},k}^* f_{m,k} \tilde{h}_{k,\tilde{m}}^* h_{k,\hat{m}} \\ b_k^{m,\tilde{m}} &\triangleq C_{\text{P2},k}^m \tilde{f}_{\tilde{m},k}^* f_{m,k} \tilde{h}_{k,\tilde{m}}^* \end{aligned}$$

with

$$C_{\text{P2},k}^{m,\hat{m}} \triangleq \sqrt{KM} d_{\text{P2},k} \sqrt{P_{m,k} E_{k,\hat{m}}} \quad (2.18)$$

$$C_{\text{P2},k}^m \triangleq \sqrt{KM} d_{\text{P2},k} \sqrt{P_{m,k}}. \quad (2.19)$$

Again, the normalization \sqrt{KM} in (2.17), (2.18) and (2.19) is introduced for convenience of exposition and

$$\underline{C} \leq C_{\text{P2},k}^{m,\hat{m}} \leq \bar{C}, \quad \underline{c} \leq C_{\text{P2},k}^m \leq \bar{c}$$

for all $k \in [1:K]$, $m \in [1:M]$, and $\hat{m} \in [1:M]$ with the constants \underline{C} , \underline{c} , \bar{C} , and \bar{c} not depending on M, K .

Recalling that we assume perfect knowledge of \mathbf{H} , \mathbf{F} , \mathbf{E} , and \mathbf{P} at each of the destination terminals, \mathcal{D}_m , the mutual information for the $\mathcal{S}_m \rightarrow \mathcal{D}_m$ link in P2 is given by

$$I(y_m; s_m | \mathbf{H}, \mathbf{F}) = \log(1 + \text{SINR}_m^{\text{P2}}) \quad (2.20)$$

where

$$\text{SINR}_m^{\text{P2}} \triangleq \frac{\left| \sum_{k=1}^K \sum_{\tilde{m}=1}^M a_k^{m,m,\tilde{m}} \right|^2}{\sum_{\hat{m} \neq m} \left| \sum_{k=1}^K \sum_{\tilde{m}=1}^M a_k^{m,\hat{m},\tilde{m}} \right|^2 + \sigma^2 M \sum_{k=1}^K \left| \sum_{\tilde{m}=1}^M b_k^{m,\tilde{m}} \right|^2 + KM^2 \sigma^2} \quad (2.21)$$

is the effective SINR in the SISO channel $\mathcal{S}_m \rightarrow \mathcal{D}_m$.

2.3.3. Large-Deviations Analysis of Signal-to-Interference-Plus-Noise Ratio

Our goal in this section is to prove that $\text{SINR}_m^{\text{P1}}$ and $\text{SINR}_m^{\text{P2}}$ for $m \in [1: M]$ (and, thus, the corresponding mutual information quantities (2.14) and (2.20)) lie within “narrow intervals” around their mean values with⁶ “high probability” when $M, K \rightarrow \infty$. The technique we use to prove these *concentration results* is based on a large-deviations analysis and can be summarized as follows:

- i. Consider each sum in the numerator and denominator of (2.15) and (2.21) separately.
- ii. Represent the considered sum as a sum of independent RVs or as a sum of dependent complex-valued RVs with independent phases.
- iii. Find the mean value of the considered sum.
- iv. Employ a large-deviations analysis to prove that the considered sum lies within a narrow interval around its mean with high probability, i.e., establish a concentration result.
- v. Combine the concentration results for the separate sums using the union bounds summarized in Appendix B to obtain concentration results for $\text{SINR}_m^{\text{P1}}$ and $\text{SINR}_m^{\text{P2}}$.

A. Chernoff bounds

Before embarking on a detailed discussion of the individual Steps i–v above, we note that a well-known technique to establish large-deviations results for sums of RVs (as required in Step iv above) is based on Chernoff bounds. This method, which yields the precise exponential behavior for the tails of the distributions under question, can, unfortunately, not be applied to all the sums in (2.15) and (2.21). To solve this problem, we develop a new technique, which allows

⁶The precise meaning of “narrow intervals” and “high probability” is explained in the formulation of Theorems 2.1 and 2.2 in Section 2.3.4.

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

to establish large-deviations results for certain sums of dependent complex-valued RVs with independent phases where the RVs occurring in the sum are such that their moment generating function (MGF) does not need to be known. The new technique is based on the well-known idea of truncation of RVs and will, therefore, be called truncation technique. Even though truncation of RVs is a standard concept in probability theory, and in particular in large-deviations analysis, we could not find the specific approach developed in this thesis in the literature. We therefore decided to present the truncation technique as a stand-alone concept and summarized the main results in Appendix A. Before proceeding, we note that even though the truncation technique has wider applicability than Chernoff bounds, it yields weaker exponents for the tails of the distributions under question.

Although the proofs of the main concentration results, Theorems 2.1 and 2.2 in Section 2.3.4, are entirely based on the truncation technique, we still discuss the results of the application of Chernoff bounds (without giving all the details) in the following, restricting our attention to P1, to motivate the development of the truncation technique and to provide a reference for the quality (in terms of tightness of the bounds) of the results in Theorems 2.1 and 2.2. Moreover, the developments below introduce some of the key elements of the proofs of Theorems 2.1 and 2.2.

Following the approach outlined in Steps i–v above, we start by writing $\text{SINR}_m^{\text{P1}}$ as

$$\text{SINR}_m^{\text{P1}} = \frac{|S^{(1)} + S^{(2)}|^2}{S^{(3)} + \sigma^2 M S^{(4)} + K M \sigma^2} \quad (2.22)$$

and establishing bounds on the probability of large deviations of

$$S^{(1)} \triangleq \sum_{k:p(k)=m} C_{\text{P1},k}^{m,m} |f_{m,k}| |h_{k,m}| \quad (2.23)$$

$$S^{(2)} \triangleq \sum_{k:p(k) \neq m} C_{\text{P1},k}^{m,m} \tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* h_{k,m} \quad (2.24)$$

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

$$S^{(3)} \triangleq \sum_{\hat{m} \neq m} \left| \sum_{k=1}^K C_{P1,k}^{m,\hat{m}} \tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* h_{k,\hat{m}} \right|^2 \quad (2.25)$$

$$S^{(4)} \triangleq \sum_{k=1}^K (C_{P1,k}^m)^2 |f_{m,k}|^2. \quad (2.26)$$

We shall see in the following that the probability density functions (PDFs) of the terms in $S^{(1)}$, $S^{(2)}$, and $S^{(4)}$ have a structure that is simple enough for Chernoff bounds to be applicable. We start with the analysis of the simplest term, namely $S^{(4)}$. To avoid unnecessary technical details and to simplify the exposition, we assume (only in the ensuing analysis of the large deviations behavior of $S^{(4)}$) that

$$C_{P1,k}^{m,\hat{m}} = C_{P1,k}^m = 1 \quad (2.27)$$

for all $m, \hat{m} \in [1:M]$, $k \in [1:K]$. Defining⁷ $X_k \triangleq |f_{m,k}|^2$, we have

$$S^{(4)} = \sum_{k=1}^K X_k$$

where the X_k are i.i.d. exponentially distributed⁸ with parameter $\lambda = 1$, and hence $\mathbb{E}[X_k] = 1$. For convenience, we centralize X_k and define $Z_k \triangleq X_k - 1$. The MGF of Z_k is given by

$$M_{Z_k}(s) = \int_0^{\infty} e^{s(x-1)} e^{-x} dx = \frac{e^{-s}}{1-s}, \quad \Re s \leq 1. \quad (2.28)$$

Since the RVs Z_k are independent, we obtain, using the standard Chernoff bound (Gallager, 1968, Section 5.4), for $x > 0$

$$\begin{aligned} \mathbb{P} \left\{ \sum_{k=1}^K Z_k \geq x \right\} &\leq \min_{0 \leq s \leq 1} (M_{Z_k}(s))^K e^{-sx} \\ &= \min_{0 \leq s \leq 1} e^{-Ks - K \ln(1-s) - sx}. \end{aligned} \quad (2.29)$$

⁷For notational convenience, we shall omit the index m in what follows.

⁸An exponentially distributed RV with parameter λ is a real-valued RV X with PDF given by $f_X(x) = \lambda \exp(-\lambda x) u(x)$.

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

Because $(M_{Z_k}(s))^K \exp(-sx)$ is convex in s (Gallager, 1968, Section 5.4), the minimum in (2.29) can easily be seen to be taken on for $s = x/(x + K)$, which gives

$$\mathbb{P}\left\{\sum_{k=1}^K Z_k \geq x\right\} \leq e^{K \ln(x+K) - K \ln(K) - x}. \quad (2.30)$$

The corresponding relation for negative deviations ($x < 0$) is

$$\mathbb{P}\left\{\sum_{k=1}^K Z_k \leq x\right\} \leq \begin{cases} e^{K \ln(x+K) - K \ln(K) - x}, & x > -K \\ 0, & x < -K. \end{cases} \quad (2.31)$$

Finally, setting $x = \sqrt{K}t$, we get the desired concentration result for the sum $S^{(4)}$ as

$$\mathbb{P}\left\{S^{(4)} - K \geq \sqrt{K}t\right\} \leq e^{K \ln(1+t/\sqrt{K}) - \sqrt{K}t}, \quad t \geq 0 \quad (2.32)$$

$$\mathbb{P}\left\{S^{(4)} - K \leq \sqrt{K}t\right\} \leq \begin{cases} e^{K \ln(1+t/\sqrt{K}) - \sqrt{K}t}, & -\sqrt{K} < t \leq 0 \\ 0, & t \leq -\sqrt{K}. \end{cases} \quad (2.33)$$

We now consider the case when K is large and $t = o(\sqrt{K})$ so that

$$\ln\left(1 + \frac{t}{\sqrt{K}}\right) = \frac{t}{\sqrt{K}} - \frac{t^2}{2K} + \mathcal{O}\left(\left(\frac{t}{\sqrt{K}}\right)^3\right). \quad (2.34)$$

If we omit higher (than second) order terms in (2.34), the bound in (2.32) and (2.33) can be compactly written as

$$\mathbb{P}\left\{\left|S^{(4)} - K\right| \geq \sqrt{K}t\right\} \leq 2e^{-t^2/2}. \quad (2.35)$$

We can, therefore, conclude that the probability of large deviations of $S^{(4)}$ decays exponentially.

Similar concentration results, using Chernoff bounds, can be established for $S^{(1)}$ and $S^{(2)}$. The derivation is somewhat involved (as it requires establishing upper bounds on the MGF), does not provide

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

insights into the problem and will, therefore, be omitted. Unfortunately, the simple technique used above to establish concentration results for $S^{(4)}$ (and applicable to $S^{(1)}$ and $S^{(2)}$) does not seem to be applicable to $S^{(3)}$. To see this, we start by noting that $S^{(3)}$ contains two classes of terms (in the sense of the properties of their PDF), i.e.,

$$S^{(3)} = S^{(31)} + S^{(32)} \quad (2.36)$$

with

$$S^{(31)} \triangleq \sum_{\hat{m} \neq m} \sum_{k=1}^K \left(C_{P1,k}^{m,\hat{m}} \right)^2 |f_{m,k}|^2 |h_{k,\hat{m}}|^2 \quad (2.37)$$

$$S^{(32)} \triangleq \sum_{\hat{m} \neq m} \sum_{k=1}^K \sum_{\hat{k} \neq k} C_{P1,k}^{m,\hat{m}} \tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* h_{k,\hat{m}} \\ \times C_{P1,\hat{k}}^{m,\hat{m}} \tilde{f}_{p(\hat{k}),\hat{k}}^* f_{m,\hat{k}} \tilde{h}_{\hat{k},p(\hat{k})}^* h_{\hat{k},\hat{m}}. \quad (2.38)$$

Now, there are two problems in applying the technique we have used so far to $S^{(3)}$: First, it seems very difficult to compute the MGFs for the individual terms in $S^{(31)}$ and $S^{(32)}$; second, the individual terms in $S^{(31)}$ and $S^{(32)}$ are not *jointly*⁹ independent across the summation indices. The first problem can probably be resolved using bounds on the exact MGFs (as can be done in the analysis of $S^{(1)}$ and $S^{(2)}$). The second problem, however, seems more fundamental. In particular, the individual terms in $S^{(31)}$ are independent across k but not across \hat{m} . In $S^{(32)}$, the individual terms are independent across k but not across \hat{k} and \hat{m} . Assuming that the problem of computing (or properly bounding) the MGFs is resolved, a natural way to overcome the second problem mentioned above would be to establish concentration results

⁹We write “jointly independent”, as opposed to “pairwise independent” here and in what follows to stress the fact that the joint PDF of the RVs under consideration can be factored into a product of the marginal PDFs. In several places throughout this chapter we will deal with sets of RVs that turn out to be pairwise independent, but not jointly independent.

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

for the sums over k , i.e., for

$$\hat{S}_{\hat{m}}^{(31)} \triangleq \sum_{k=1}^K \left(C_{P1,k}^{m,\hat{m}} \right)^2 |f_{m,k}|^2 |h_{k,\hat{m}}|^2 \quad (2.39)$$

$$\begin{aligned} \hat{S}_{\hat{m},\hat{k}}^{(32)} &\triangleq \sum_{k=1}^K C_{P1,k}^{m,\hat{m}} \tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* h_{k,\hat{m}} \\ &\quad \times C_{P1,\hat{k}}^{m,\hat{m}} \tilde{f}_{p(\hat{k}),\hat{k}}^* f_{m,\hat{k}} \tilde{h}_{\hat{k},p(\hat{k})}^* h_{\hat{k},\hat{m}} \end{aligned} \quad (2.40)$$

and to employ the union bound for sums (Lemmas B.1 and B.3 in Appendix B) to obtain concentration results for $S^{(31)}$ and $S^{(32)}$. Unfortunately, this method, although applicable, yields results that are very loose in the sense of not reflecting the correct “order-of-magnitude behavior” of the typical deviations. To understand why this is the case, we perform an order-of-magnitude analysis as follows. For simplicity, we again assume that the condition (2.27) is satisfied. Note that for every $\hat{k}, k \in [1:K]$ such that $\hat{k} \neq k$ and every $\hat{m} \in [1:M]$ such that $\hat{m} \neq m$, we have

$$\mathbb{E} \left[\tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* h_{k,\hat{m}} \tilde{f}_{p(\hat{k}),\hat{k}}^* f_{m,\hat{k}} \tilde{h}_{\hat{k},p(\hat{k})}^* h_{\hat{k},\hat{m}} \right] = 0.$$

Chernoff bounding $\hat{S}_{\hat{m},\hat{k}}^{(32)}$ would, therefore, yield that

$$\mathbb{P} \left\{ \left| \hat{S}_{\hat{m},\hat{k}}^{(32)} \right| \geq \sqrt{Kt} \right\}$$

decays exponentially¹⁰ in t . Then, applying the union bound for sums (Lemma B.1) to $S^{(32)} = \sum_{\hat{m} \neq m} \sum_{\hat{k} \neq k} \hat{S}_{\hat{m},\hat{k}}^{(32)}$, we would conclude that

$$\mathbb{P} \left\{ \left| S^{(32)} \right| \geq (M-1)(K-1)\sqrt{Kt} \right\} \quad (2.41)$$

decays exponentially in t . Even though the terms in $S^{(32)}$ are not completely independent across \hat{k} and \hat{m} , we will see in Section 2.3.3.B that

¹⁰We do not specify the exponent here.

there is still enough independence between them for the truncation technique to reveal that

$$\mathbb{P}\left\{\left|S^{(32)}\right| \geq \sqrt{(M-1)(K-1)Kt}\right\} \quad (2.42)$$

decays exponentially in t , which is a much stronger concentration result than (2.41). The importance of the difference between (2.42) and (2.41) becomes clear if we consider $S^{(31)}$. Since $\hat{S}_{\hat{m}}^{(31)}$ is a sum over K independent terms, each of which satisfies $\mathbb{E}\left[|f_{m,k}|^2|h_{k,\hat{m}}|^2\right] = 1$, Chernoff bounding would yield that

$$\mathbb{P}\left\{\left|\hat{S}_{\hat{m}}^{(31)} - K\right| \geq \sqrt{Kt}\right\}$$

decays exponentially in t . Applying the union bound to $S^{(31)} = \sum_{\hat{m} \neq m} \hat{S}_{\hat{m}}^{(31)}$, one can then show that

$$\mathbb{P}\left\{\left|S^{(31)} - K(M-1)\right| \geq (M-1)\sqrt{Kt}\right\} \quad (2.43)$$

decays exponentially in t . When M and K are large, we would now conclude from (2.41) and (2.43) that $S^{(3)} = S^{(31)} + S^{(32)}$ deviates around KM with a typical deviation of order $MK\sqrt{K}$. Since the typical deviations are larger (by a factor of \sqrt{K}) than the mean, the corresponding deviation result is useless. On the other hand, if we use the bound (2.42) combined with (2.43), again assuming that M and K are large, we can conclude that $S^{(3)}$ deviates around KM with a typical deviation of order $\sqrt{MK} + M\sqrt{K}$, which is an order of magnitude smaller than the mean. As already mentioned, the truncation technique allows us to establish useful concentration results for sums with dependent terms such as that in (2.40).

B. Application of the truncation technique

In this subsection, we demonstrate how the desired concentration results for $S^{(31)}$ and $S^{(32)}$, defined in (2.37) and (2.38), respectively, can be obtained by application of the truncation technique. The following results will be used in the proof of Theorem 2.1 and will, therefore, be formulated for general $C_{P1,k}^{m,\hat{m}}$ and $C_{P1,k}^m$.

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

Analysis of $S^{(31)}$: Consider $\hat{S}_{\hat{m}}^{(31)}$. The variables $X_k \triangleq |f_{m,k}|^2$ and $Y_{k,\hat{m}} \triangleq |h_{k,\hat{m}}|^2$ are exponentially distributed with parameter $\lambda = 1$. Therefore, we have

$$\mathbb{P}\{X_k \geq x\} = \mathbb{P}\{Y_{k,\hat{m}} \geq x\} \leq e^{-x}, \quad x \geq 0, \text{ for all } k, \hat{m}.$$

Define $Z_{k,\hat{m}} \triangleq X_k Y_{k,\hat{m}}$. From the union bound for products it follows that

$$\mathbb{P}\{Z_{k,\hat{m}} \geq x^2\} = \mathbb{P}\{X_k Y_{k,\hat{m}} \geq x^2\} \leq 2e^{-x}$$

which yields

$$\mathbb{P}\{Z_{k,\hat{m}} \geq x\} \leq 2e^{-\sqrt{x}}.$$

Next, using $\mathbb{E}[Z_{k,\hat{m}}] = 1$ and $\mathbb{E}[(Z_{k,\hat{m}})^2] = 4$ for all $k, \hat{m} \neq m$ and the independence of the RVs $Z_{k,\hat{m}}$ across $k \in [1:K]$, it follows from Corollary A.5, taking into account (2.12), that for $K \geq 2$

$$\mathbb{P}\left\{\left|\hat{S}_{\hat{m}}^{(31)} - \sum_{k=1}^K \left(C_{\text{P1},k}^{m,\hat{m}}\right)^2\right| \geq \sqrt{K}x\right\} \leq 6Ke^{-\Delta^{(31)}x^{2/5}}$$

where $\Delta^{(31)} \triangleq \min[1, (1/8)\bar{C}^{-4}]$. Applying the union bound for sums (see Lemma B.1) and using (2.12), we finally obtain the desired¹¹ concentration result for $S^{(31)}$ as

$$\mathbb{P}\left\{S^{(31)} \geq (M-1)K\bar{C}^2 + (M-1)\sqrt{K}x\right\} \leq 6(M-1)Ke^{-\Delta^{(31)}x^{2/5}} \quad (2.44)$$

and

$$\mathbb{P}\left\{S^{(31)} \leq (M-1)K\underline{C}^2 - (M-1)\sqrt{K}x\right\} \leq 6(M-1)Ke^{-\Delta^{(31)}x^{2/5}}. \quad (2.45)$$

Analysis of $S^{(32)}$: We start by rewriting (2.38) as

$$S^{(32)} = \sqrt{K-1} \sum_{\hat{m} \neq m} \sum_{k=1}^K C_{\text{P1},k}^{m,\hat{m}} \tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* h_{k,\hat{m}} T_{\hat{m},k}^{(32)} \quad (2.46)$$

¹¹We note that we do not avoid using the union bound on $S^{(31)}$. It is important, however, that we do not use it when analyzing $S^{(32)}$.

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

where $T_{\hat{m},k}^{(32)}$ is defined as

$$T_{\hat{m},k}^{(32)} \triangleq \frac{1}{\sqrt{K-1}} \sum_{\hat{k} \neq k} C_{P1,\hat{k}}^{m,\hat{m}} \tilde{f}_{p(\hat{k}),\hat{k}} \tilde{f}_{m,\hat{k}}^* \tilde{h}_{\hat{k},p(\hat{k})} h_{\hat{k},\hat{m}}^*.$$

The concentration result for $S^{(32)}$ (and other similar sums occurring in the proofs of Theorems 2.1 and 2.2) will be established by applying (one or multiple times) the following general steps:

- Establish a concentration result for $T_{\hat{m},k}^{(32)}$.
- Represent the terms on the right-hand side of (2.46) in the form $C_{P1,k}^{m,\hat{m}} Z_{\hat{m},k} \exp(i\hat{\phi}_{k,\hat{m}})$ where

$$Z_{\hat{m},k} \triangleq T_{\hat{m},k}^{(32)} |f_{m,k}| |h_{k,\hat{m}}|$$

and

$$\hat{\phi}_{k,\hat{m}} \triangleq \arg\left(\tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* h_{k,\hat{m}}\right)$$

so that the sum $S^{(32)}$ can be written as

$$S^{(32)} \triangleq \sqrt{K-1} \sum_{\hat{m} \neq m} \sum_{k=1}^K C_{P1,k}^{m,\hat{m}} Z_{\hat{m},k} e^{i\hat{\phi}_{k,\hat{m}}}.$$

- Use the concentration result for $T_{\hat{m},k}^{(32)}$ together with the union bound for products (see Lemma B.4) to establish bounds on the tail behavior of $Z_{\hat{m},k}$ and verify condition (A.1) in Theorem A.3.
- If needed, split up the sum $S^{(32)}$ into several sums, so that the phases $\exp(i\hat{\phi}_{k,\hat{m}})$ are jointly independent in each of these sums and Theorem A.3 can be applied (to each of these sums separately).
- Finally, apply Theorem A.3 to each of the sums resulting in the previous step separately and use the union bound for sums to establish the desired concentration result for $S^{(32)}$.

Following this procedure, we start by deriving a concentration result for $T_{\hat{m},k}^{(32)}$. Since $T_{\hat{m},k}^{(32)}$ is of the same nature as $S^{(2)}$, we could, in principle, use Chernoff bounds. This would, however, lead to an

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

exponent with a complicated dependence on t , which can be simplified only under certain assumptions on t , such as e.g., $t = o(\sqrt{K})$ in (2.34). What we need is a simple universal bound for $\mathbb{P}\{|T_{\hat{m},k}^{(32)}| \geq x\}$, which is valid for all x and allows to verify condition (A.1) in Theorem A.3 for $Z_{\hat{m},k}$. Such a bound can be obtained by applying the truncation technique to $T_{\hat{m},k}^{(32)}$ as follows. Define $X_{\hat{k}} \triangleq |f_{m,\hat{k}}|$, $Y_{\hat{k},\hat{m}} \triangleq |h_{\hat{k},\hat{m}}|$ and

$$\phi_{\hat{k},\hat{m}} \triangleq \arg\left(\tilde{f}_{p(\hat{k}),\hat{k}} f_{m,\hat{k}}^* \tilde{h}_{\hat{k},p(\hat{k})} h_{\hat{k},\hat{m}}^*\right)$$

so that

$$T_{\hat{m},k}^{(32)} = \frac{1}{\sqrt{K}-1} \sum_{\hat{k} \neq k} C_{P1,\hat{k}}^{m,\hat{m}} X_{\hat{k}} Y_{\hat{k},\hat{m}} e^{i\phi_{\hat{k},\hat{m}}}.$$

The RVs $X_{\hat{k}}$ and $Y_{\hat{k},\hat{m}}$ (for all \hat{k}, \hat{m}) are Rayleigh distributed¹² with parameter $\alpha^2 = 1/2$. Therefore, we have

$$\mathbb{P}\{X_{\hat{k}} \geq x\} = \mathbb{P}\{Y_{\hat{k},\hat{m}} \geq x\} \leq e^{-x^2}, \quad x \geq 0$$

and the union bound for products yields

$$\mathbb{P}\{X_{\hat{k}} Y_{\hat{k},\hat{m}} \geq x\} \leq 2e^{-x}, \quad x \geq 0 \quad (2.47)$$

which shows that condition (A.8) in Corollary A.4 is satisfied. Next, rewrite $\phi_{\hat{k},\hat{m}}$ as

$$\phi_{\hat{k},\hat{m}} = \arg\left(\tilde{f}_{p(\hat{k}),\hat{k}}\right) \oplus \arg\left(f_{m,\hat{k}}^*\right) \oplus \arg\left(\tilde{h}_{\hat{k},p(\hat{k})}\right) \oplus \arg\left(h_{\hat{k},\hat{m}}^*\right). \quad (2.48)$$

Because the f 's and the h 's in (2.48) are independent across $\hat{k} \in [1:K]$, it follows that the phases $\phi_{\hat{k},\hat{m}}$ are also independent across $\hat{k} \in [1:K]$, which is precisely what we need for the truncation technique to be applicable. Recalling that $m \neq \hat{m}$, and, therefore, either $p(\hat{k}) \neq m$ or $p(\hat{k}) \neq \hat{m}$, (2.48) implies that $\phi_{\hat{k},\hat{m}} \sim \mathcal{U}(-\pi, \pi)$ and hence, $\mathbb{E}[\exp(i\phi_{\hat{k},\hat{m}})] = 0$ for all \hat{k}, \hat{m} . Since $\phi_{\hat{k},\hat{m}}$ is independent of $X_{\hat{k}}$

¹²A Rayleigh-distributed RV with parameter α^2 is a real-valued RV X with PDF $f_X(x) = (x/\alpha^2) \exp(-x^2/(2\alpha^2))u(x)$.

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

and $Y_{\hat{k}, \hat{m}}$, we have $\mathbb{E}[\exp(i\phi_{\hat{k}, \hat{m}})X_{\hat{k}}Y_{\hat{k}, \hat{m}}] = 0$ for all \hat{k}, \hat{m} and hence, $\mathbb{E}[T_{\hat{m}, k}^{(32)}] = 0$ for all \hat{m}, k . Finally, applying Corollary A.4 to $T_{\hat{m}, k}^{(32)}$, taking into account (2.12), we get for $K \geq 2$ and $x \geq 0$ that

$$\mathbb{P}\left\{\left|T_{\hat{m}, k}^{(32)}\right| \geq x\right\} \leq 8(K-1)e^{-\Delta^{(T)}x^{2/3}} \quad (2.49)$$

with $\Delta^{(T)} \triangleq 2^{-1/3} \min\left[1, (1/2)\overline{C}^{-2}\right]$.

We are now ready to establish the concentration result for $S^{(32)}$. First, rewrite $\hat{\phi}_{k, \hat{m}}$ as

$$\hat{\phi}_{k, \hat{m}} \triangleq \arg\left(\tilde{f}_{p(k), k}^*\right) \oplus \arg(f_{m, k}) \oplus \arg\left(\tilde{h}_{k, p(k)}^*\right) \oplus \arg(h_{k, \hat{m}}). \quad (2.50)$$

Similar to $\phi_{\hat{k}, \hat{m}}$ in (2.48), because $\hat{m} \neq m$ we conclude that $\hat{\phi}_{k, \hat{m}} \sim \mathcal{U}(-\pi, \pi)$. Furthermore, because $\hat{k} \neq k$ the $\hat{\phi}_{k, \hat{m}}$ are independent of $T_{\hat{m}, k}^{(32)}$, and therefore also of $Z_{\hat{m}, k}$ (for all k, \hat{m}). To apply Corollary A.4 to $S^{(32)}$, the $\hat{\phi}_{k, \hat{m}}$ are required to be jointly independent across $\hat{m} \in [1 : M]$ for $\hat{m} \neq m$ and $k \in [1 : K]$. It can be verified that this is not the case. There is, however, a simple way to resolve this problem by considering the two disjoint index sets

$$\begin{aligned} \mathcal{I}_1 &\triangleq \left\{(\hat{m}, k) \mid \hat{m} \in [1 : M], \hat{m} \neq m, k \in [1 : K], p(k) \neq \hat{m}\right\} \\ \mathcal{I}_2 &\triangleq \left\{(\hat{m}, k) \mid \hat{m} \in [1 : M], \hat{m} \neq m, k \in [1 : K], p(k) = \hat{m}\right\}. \end{aligned}$$

It follows by inspection that within each of the sets $\{\hat{\phi}_{k, \hat{m}}\}_{(k, \hat{m}) \in \mathcal{I}_1}$ and $\{\hat{\phi}_{k, \hat{m}}\}_{(k, \hat{m}) \in \mathcal{I}_2}$ the phases are jointly independent. Separating $S^{(32)}$ into two sums corresponding to the group of indices \mathcal{I}_1 and \mathcal{I}_2 , we get

$$S^{(32)} = S^{(321)} + S^{(322)} \quad (2.51)$$

with

$$\begin{aligned} S^{(321)} &\triangleq \sqrt{K-1} \sum_{\hat{m} \neq m} \sum_{k: p(k) \neq \hat{m}} C_{P1, k}^{m, \hat{m}} Z_{\hat{m}, k} e^{i\hat{\phi}_{k, \hat{m}}} \\ S^{(322)} &\triangleq \sqrt{K-1} \sum_{\hat{m} \neq m} \sum_{k: p(k) = \hat{m}} C_{P1, k}^{m, \hat{m}} Z_{\hat{m}, k} e^{i\hat{\phi}_{k, \hat{m}}}. \end{aligned}$$

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

Applying the union bound for products first to $|f_{m,k}| |h_{k,\hat{m}}|$ as in (2.47), then to $Z_{\hat{m},k}$ using (2.49), and using the simple bound

$$2e^{-x} + 8(K-1)e^{-\Delta^{(T)}x^{1/3}} \leq 16(K-1)e^{-\Delta^{(T)}x^{1/3}}$$

which is valid for $x \geq 1$, we get

$$\mathbb{P}\{|Z_{\hat{m},k}| \geq x\} \leq 16(K-1)e^{-\Delta^{(T)}x^{1/3}}$$

for $K \geq 2$ and $x \geq 1$. Therefore, using $\mathbb{E}[Z_{\hat{m},k} \exp(i\hat{\phi}_{k,\hat{m}})] = 0$ for all $k, \hat{m} \neq m$, applying Corollary A.4 to $S^{(321)}$ (which consists of $K(M-1)^2/M$ terms) and to $S^{(322)}$ (which consists of $K(M-1)/M$ terms) separately, taking into account (2.12), we obtain that for $K \geq 2$, $M > 2$, and $x \geq 1$

$$\begin{aligned} \mathbb{P}\left\{|S^{(321)}| \geq \sqrt{\frac{(K-1)K(M-1)^2}{M}}x\right\} \\ \leq 64 \frac{(K-1)K(M-1)^2}{M} e^{-\Delta^{(32)}x^{2/7}} \end{aligned} \quad (2.52)$$

and

$$\begin{aligned} \mathbb{P}\left\{|S^{(322)}| \geq \sqrt{\frac{(K-1)K(M-1)}{M}}x\right\} \\ \leq 64 \frac{(K-1)K(M-1)}{M} e^{-\Delta^{(32)}x^{2/7}} \end{aligned} \quad (2.53)$$

where $\Delta^{(32)} = 2^{-10/21} \min[1, (1/2)\bar{C}^{-2}]$. Combining (2.35) (and similar bounds for $S^{(1)}$ and $S^{(2)}$), (2.52), (2.53), (2.51), (2.44), (2.45), and (2.36), we can now state the final concentration result for $\text{SINR}_m^{\text{P1}}$ by carrying out Step v in the summary presented in the first paragraph of Section 2.3.3. Recall, however, that we used the classical Chernoff-bounding technique to establish the large-deviations behavior of $S^{(1)}$, $S^{(2)}$, and $S^{(4)}$, whereas we employed the truncation technique to analyze the large-deviations behavior of $S^{(3)}$. Even though the Chernoff bounds are tighter than the bounds obtained through the truncation

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

technique, the tightness of the final bounds for the tail behavior of $\text{SINR}_m^{\text{P1}}$ and $\text{SINR}_m^{\text{P2}}$ is determined by the weakest exponent in the bounds for the individual terms $S^{(1)}$, $S^{(2)}$, $S^{(3)}$ and $S^{(4)}$. Therefore, employing Chernoff bounds for $S^{(1)}$, $S^{(2)}$, and $S^{(4)}$ and the truncation technique for $S^{(3)}$ will not lead to a significantly tighter final result, compared to the case where the truncation technique is used throughout. Motivated by this observation and for simplicity of exposition, we therefore decided to state the concentration results in Section 2.3.4 for $\text{SINR}_m^{\text{P1}}$ and $\text{SINR}_m^{\text{P2}}$ obtained by applying the truncation technique throughout.

2.3.4. Concentration Results for P1 and P2

In Section 2.3.3, we outlined how the large-deviations behavior of the SINR (for P1 and P2) can be established based on the truncation technique and on union bounds. The resulting key statement, made precise in Theorems 2.1 and 2.2 below, is that the probability of the SINR falling outside a narrow interval around its mean is “exponentially small”. We proceed with the formal statement of the results.

Theorem 2.1. *For every $K \geq 2$, $M \geq 2$, for every $x \geq 1$, the probability $P_{\text{P1}}(x)$ of the event*

$$\text{SINR}_m^{\text{P1}} \notin [L_{\text{P1}}(x), U_{\text{P1}}(x)], \quad m \in [1 : M]$$

where

$$L_{\text{P1}}(x) \triangleq \frac{\pi^2 \underline{C}^2}{16 \underline{C}_{\text{SN}}^2} \frac{K}{M^3} \frac{\max\left[0, 1 - \frac{8}{\underline{C}_\pi} \frac{M}{\sqrt{K}} x\right]^2}{\frac{\underline{C}^2}{\underline{C}_{\text{SN}}^2} + \frac{3}{\underline{C}_{\text{SN}}^2} \frac{x}{\sqrt{M}} + \frac{\sigma^2}{\underline{C}_{\text{SN}}^2} \left(\underline{c}^2 + \frac{x}{\sqrt{K}}\right) + \frac{\sigma^2}{\underline{C}_{\text{SN}}^2}} \quad (2.54)$$

and

$$U_{\text{P1}}(x) \triangleq \frac{\pi^2 \overline{C}^2}{16 \overline{C}_{\text{SN}}^2} \frac{K}{M^3} \times \frac{\left(1 + \frac{8}{\overline{C}_\pi} \frac{M}{\sqrt{K}} x\right)^2}{\max\left[0, \frac{\overline{C}^2}{\overline{C}_{\text{SN}}^2} \frac{M-1}{M} - \frac{3}{\overline{C}_{\text{SN}}^2} \frac{x}{\sqrt{M}}\right] + \max\left[0, \frac{\sigma^2}{\overline{C}_{\text{SN}}^2} \left(\overline{c}^2 - \frac{x}{\sqrt{K}}\right)\right] + \frac{\sigma^2}{\overline{C}_{\text{SN}}^2}},$$

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

with the constants \overline{C}_{SN} and $\underline{C}_{\text{SN}}$ given by

$$\overline{C}_{\text{SN}} \triangleq \sqrt{\overline{C}^2 + \sigma^2(\overline{c}^2 + 1)} \quad \underline{C}_{\text{SN}} \triangleq \sqrt{\underline{C}^2 + \sigma^2(\underline{c}^2 + 1)}$$

satisfies the following inequality

$$P_{\text{P1}}(x) \leq 302 K^2 M e^{-\Delta_{\text{P1}} x^{2/7}} \quad (2.55)$$

with $\Delta_{\text{P1}} \triangleq \min\left[2^{-\frac{10}{21}}, 1/(2^{\frac{31}{21}} \overline{C}^2), 1/(8 \overline{C}^4), 1/(4 \overline{c}^4)\right]$.

Proof. See Appendix C.1. □

Theorem 2.2. For every $K \geq 2$, $M \geq 2$, for every $x \geq 1$, the probability $P_{\text{P2}}(x)$ of the event

$$\text{SINR}_m^{\text{P2}} \notin [L_{\text{P2}}(x), U_{\text{P2}}(x)], \quad m \in [1 : M]$$

where

$$L_{\text{P2}}(x) \triangleq \frac{\pi^2}{16} \frac{\underline{C}^2}{\overline{C}_{\text{SN}}^2} \frac{K}{M^2} \frac{\max\left[0, 1 - \frac{8}{\underline{C}\pi} \sqrt{\frac{M}{K}} x\right]^2}{\frac{\overline{C}^2}{\overline{C}_{\text{SN}}^2} + \frac{4}{\overline{C}_{\text{SN}}^2} \frac{x}{\min[\sqrt{M}, \sqrt{K}]} + \frac{\sigma^2\left(\overline{c}^2 + \frac{2x}{\sqrt{K}}\right)}{\overline{C}_{\text{SN}}^2} + \frac{\sigma^2}{\overline{C}_{\text{SN}}^2}}$$

$$U_{\text{P2}}(x) \triangleq \frac{\pi^2}{16} \frac{\overline{C}^2}{\underline{C}_{\text{SN}}^2} \frac{K}{M^2}$$

$$\times \frac{\left(1 + \frac{8}{\underline{C}\pi} \sqrt{\frac{M}{K}} x\right)^2}{\max\left[0, \frac{\underline{C}^2}{\underline{C}_{\text{SN}}^2} \frac{M-1}{M} - \frac{4}{\underline{C}_{\text{SN}}^2} \frac{x}{\min[\sqrt{M}, \sqrt{K}]} \right] + \max\left[0, \frac{\sigma^2\left(\underline{c}^2 - \frac{2x}{\sqrt{K}}\right)}{\underline{C}_{\text{SN}}^2}\right] + \frac{\sigma^2}{\underline{C}_{\text{SN}}^2}}$$

satisfies the following inequality

$$P_{\text{P2}}(x) \leq 814 K^2 M^3 e^{-\Delta_{\text{P2}} x^{2/9}} \quad (2.56)$$

with $\Delta_{\text{P2}} \triangleq \min\left[2^{-\frac{11}{5}}, 1/(2^{\frac{61}{36}} \overline{C}^2), 1/(8 \overline{C}^4), 1/(4 \overline{c}^4)\right]$.

Proof. The proof idea is the same as that underlying the proof of Theorem 2.1 with large parts of the proof itself being very similar to the proof of Theorem 2.1. For the sake of brevity the details of the proof are therefore omitted. □

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

The concentration results in Theorems 2.1 and 2.2 form the basis for showing that, provided the rate of growth of K as a function of M is fast enough, the network “decouples” (see Theorems 2.3 and 2.4) and “crystallizes” (see Theorem 2.5). Moreover, as outlined in Theorem 2.5, the outage probability behavior of the $\mathcal{S}_m \rightarrow \mathcal{D}_m$ links can be inferred from (2.55) and (2.56).

2.4. ERGODIC CAPACITY AND COOPERATION AT THE RELAY LEVEL

The focus in the previous section was on establishing concentration results for the individual link SINRs for P1 and P2. Based on these results, in this section, we study the ergodic capacity realized by the two protocols and we establish the corresponding capacity scaling and outage probability behavior.

2.4.1. Ergodic Capacity of P1 and P2

Throughout this section, we assume that all channels in the network are ergodic. The two main results are summarized as follows.

Theorem 2.3 (Ergodic capacity of P1). *Suppose that destination terminal \mathcal{D}_m ($m \in [1:M]$) has perfect knowledge of the mean of the effective channel gain of the $\mathcal{S}_m \rightarrow \mathcal{D}_m$ link, given by $(\pi/(4\sqrt{K}))\sum_{k:p(k)=m} C_{P1,k}^{m,m}$. Then, the per source-destination terminal pair capacity achieved by P1 is given by*

$$C_{P1} = \frac{1}{2M} \sum_{m=1}^M I(y_m; s_m) \quad (2.57)$$

and for all $\epsilon, \delta > 0$ there exist $M_0, K_0 > 0$ such that for all $M \geq M_0$,

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

$K \geq K_0$, capacity C_{P1} satisfies

$$\begin{aligned} \frac{1}{2} \log \left(1 + \frac{\pi^2 \underline{C}^2}{16 \overline{C}_{\text{SN}}^2} \frac{K}{M^3} (1 - \epsilon) \right) &\leq C_{P1} \\ &\leq \frac{1}{2} \log \left(1 + \frac{\pi^2 \overline{C}^2}{16 \underline{C}_{\text{SN}}^2} \frac{\max[K, M^{2+\delta}]}{M^3} (1 + \epsilon) \right). \end{aligned} \quad (2.58)$$

Theorem 2.4 (Ergodic capacity of P2). *Suppose that destination terminal \mathcal{D}_m ($m \in [1:M]$) has perfect knowledge of the mean of the effective channel gain of the $\mathcal{S}_m \rightarrow \mathcal{D}_m$ link, given by $(\pi/(4\sqrt{KM})) \sum_{k=1}^K C_{P2,k}^{m,m}$. Then, the per source-destination terminal pair capacity achieved by P2 is given by*

$$C_{P2} = \frac{1}{2M} \sum_{m=1}^M I(y_m; s_m)$$

and for all $\epsilon, \delta > 0$ there exist $M_0, K_0 > 0$, such that for all $M \geq M_0$, $K \geq K_0$, capacity C_{P2} satisfies

$$\begin{aligned} \frac{1}{2} \log \left(1 + \frac{\pi^2 \underline{C}^2}{16 \overline{C}_{\text{SN}}^2} \frac{K}{M^2} (1 - \epsilon) \right) &\leq C_{P2} \\ &\leq \frac{1}{2} \log \left(1 + \frac{\pi^2 \overline{C}^2}{16 \underline{C}_{\text{SN}}^2} \frac{\max[K, M^{1+\delta}]}{M^2} (1 + \epsilon) \right). \end{aligned} \quad (2.59)$$

The proofs of Theorems 2.3 and 2.4 are very similar. Below we present the proof of Theorem 2.3 only. The proof of Theorem 2.4 is omitted.

Proof of Theorem 2.3. Formula (2.57) follows because the input distribution in the protocol P1 is fixed, therefore we should not optimize over it; the factor 1/2 in (2.57) results from the fact that data is transmitted over two time slots. Next, we focus on proving the inequalities in (2.58).

We start by establishing the lower bound in (2.58), the proof of which uses the result summarized in Appendix C.3. To apply

Lemma C.1 in Appendix C.3, we start from (2.7) and define

$$\begin{aligned}\bar{F}_m &\triangleq \frac{1}{\sqrt{K}} \sum_{k=1}^K \mathbb{E}[a_k^{m,m}] \\ \tilde{F}_m &\triangleq \frac{1}{\sqrt{K}} \sum_{k=1}^K (a_k^{m,m} - \mathbb{E}[a_k^{m,m}]) \\ W_m &\triangleq \sum_{\hat{m} \neq m} s_{\hat{m}} \frac{1}{\sqrt{K}} \sum_{k=1}^K a_k^{m,\hat{m}} + \frac{1}{\sqrt{K}} \sum_{k=1}^K b_k^m z_k + w_m.\end{aligned}$$

With these definitions, we can now rewrite (2.7) as

$$y_m = \left(\bar{F}_m + \tilde{F}_m \right) s_m + W_m.$$

Straightforward, but tedious, manipulations yield

$$\begin{aligned}\bar{F}_m &= \frac{\pi}{4} \frac{1}{\sqrt{K}} \sum_{k:p(k)=m} C_{P1,k}^{m,m} \\ \text{Var}[\tilde{F}_m] &= \frac{1}{K} \left(\sum_{k=1}^K \left(C_{P1,k}^{m,m} \right)^2 - \frac{\pi^2}{16} \sum_{k:p(k)=m} \left(C_{P1,k}^{m,m} \right)^2 \right) \\ \text{Var}[W_m] &= \frac{1}{KM} \sum_{\hat{m} \neq m} \sum_{k=1}^K \left(C_{P1,k}^{m,\hat{m}} \right)^2 + \frac{\sigma^2}{K} \sum_{k=1}^K \left(C_{P1,k}^m \right)^2 + \sigma^2.\end{aligned}$$

Next, we use (2.12) and (2.13) to lower-bound \bar{F}_m and upper-bound $\text{Var}[\tilde{F}_m]$ and $\text{Var}[W_m]$, substitute the resulting bounds into (C.22), and obtain¹³

$$I(y_m; s_m) \geq \log \left(1 + \frac{\pi^2}{16} \frac{\underline{C}^2}{(1/M)\bar{C}^2 + \bar{C}_{\text{SN}}^2} \frac{K}{M^3} \right). \quad (2.60)$$

Finally, fix $\epsilon > 0$ and set

$$M_0 = \frac{1 - \epsilon}{\epsilon} \frac{\bar{C}^2}{\bar{C}_{\text{SN}}^2}.$$

¹³We note that this bound is valid for arbitrary M and K and is, therefore, somewhat stronger than the asymptotic bound we are actually seeking.

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

It then follows that for every $M \geq M_0$, the inequality

$$\frac{\underline{C}^2}{(1/M)\underline{C}^2 + \underline{C}_{\text{SN}}^2} \geq \frac{\underline{C}^2}{\underline{C}_{\text{SN}}^2} (1 - \epsilon)$$

is satisfied, which together with (2.60) and (2.57) completes the proof of the lower bound.

Proving the upper bound on C_{P1} in (2.58) turns out to be significantly more challenging. The method we use to this end is based on the concentration result for $\text{SINR}_m^{\text{P1}}$ in Theorem 2.1. We start by noting that the per-stream mutual information can be upper-bounded by assuming that \mathcal{D}_m has perfect knowledge of \mathbf{H} and \mathbf{F} , i.e.,

$$\begin{aligned} I(y_m; s_m) &\leq \mathbb{E}_{\mathbf{H}, \mathbf{F}} [I(y_m; s_m \mid \mathbf{H}, \mathbf{F})] \\ &= \mathbb{E}_{\mathbf{H}, \mathbf{F}} [\log(1 + \text{SINR}_m^{\text{P1}})] \\ &\leq \log(1 + \mathbb{E}_{\mathbf{H}, \mathbf{F}} [\text{SINR}_m^{\text{P1}}]) \end{aligned}$$

where the last step follows from Jensen’s inequality.

Now fix $\epsilon > 0$. To prove the upper bound in (2.58), it suffices to show that there exist $M_0, K_0 > 0$ such that for all $M \geq M_0$ and $K \geq K_0$

$$\mathbb{E}_{\mathbf{H}, \mathbf{F}} [\text{SINR}_m^{\text{P1}}] \leq A \frac{\max[K, M^{2+\delta}]}{M^3} (1 + \epsilon)$$

where we define

$$A \triangleq \frac{\pi^2 \underline{C}^2}{16 \underline{C}_{\text{SN}}^2}.$$

To simplify the exposition, we define

$$g(M, K) \triangleq \frac{1}{A} \text{SINR}_m^{\text{P1}}(M, K) \frac{M^3}{\max[K, M^{2+\delta}]}.$$

Note that we make the dependence of $\text{SINR}_m^{\text{P1}}$ on M and K explicit by using the notation $\text{SINR}_m^{\text{P1}}(M, K)$. In the remainder of the proof, we show that

$$\mathbb{E}_{\mathbf{H}, \mathbf{F}} [g(M, K)] \leq 1 + \epsilon \tag{2.61}$$

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

for M and K large enough. Let $f_g(x)$ denote the PDF of $g(M, K)$. Then, the expectation $\mathbb{E}_{\mathbf{H}, \mathbf{F}}[g(M, K)]$ can be written as

$$\begin{aligned} \mathbb{E}_{\mathbf{H}, \mathbf{F}}[g(M, K)] &= \int_0^{\infty} t f_g(t) dt \\ &= \int_0^{1+\epsilon_1} t f_g(t) dt + \int_{1+\epsilon_1}^{\infty} t f_g(t) dt \end{aligned} \quad (2.62)$$

where $\epsilon_1 > 0$ is chosen such that

$$1 + \epsilon_1 < 1 + \epsilon/3.$$

Consequently, we have

$$\begin{aligned} \int_0^{1+\epsilon_1} t f_g(t) dt &\leq (1 + \epsilon_1) \int_0^{1+\epsilon_1} f_g(t) dt \\ &\leq 1 + \epsilon_1 < 1 + \epsilon/3. \end{aligned} \quad (2.63)$$

For bounding the second integral on the right-hand side of (2.62), it is convenient to write the upper bound in Theorem 2.1 in the following form: there exist $\Delta > 0$, $\delta_1 > 0$, $\delta_2 > 0$, and $A_1, A_2, A_3 > 0$ such that for all $x \geq 1$ and $M, K \geq 2$

$$\mathbb{P}\left\{g(M, K) \geq B(M, K, x)\right\} \leq A_3 M^{\delta_1} K^{\delta_2} e^{-\Delta x^{2/7}} \quad (2.64)$$

with

$$B(M, K, x) \triangleq \frac{B^N(M, K, x)}{B^D(M, K, x)}$$

where $B^N(M, K, x)$ and $B^D(M, K, x)$ are given by

$$B^N(M, K, x) \triangleq \frac{K}{\max[K, M^{2+\delta}]} \left(1 + A_1 \frac{M}{\sqrt{K}} x\right)^2 \quad (2.65)$$

$$B^D(M, K, x)$$

$$\triangleq \frac{\underline{C}^2 \max\left[0, \frac{M-1}{M} - \frac{A_2 x}{\underline{C}^2 \sqrt{M}}\right] + \underline{c}^2 \sigma^2 \max\left[0, 1 - \frac{x}{\underline{c}^2 \sqrt{K}}\right] + \sigma^2}{\underline{C}_{\text{SN}}^2}. \quad (2.66)$$

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

The second integral on the right-hand side of (2.62) will be shown, for M and K large enough, to be upper bounded by $2\epsilon/3$ by splitting it up and proving that

$$\int_{1+\epsilon_1}^{\lceil t_0 \rceil} t f_g(t) dt \leq \epsilon/3 \quad (2.67)$$

and

$$\int_{\lceil t_0 \rceil}^{\infty} t f_g(t) dt \leq \epsilon/3 \quad (2.68)$$

where the parameter $t_0 > 1 + \epsilon_1$, independent of M, K , will be chosen later. It will become clear later why we need to split up the second integral on the right-hand side of (2.62) according to (2.67) and (2.68). The integral in (2.67) can be bounded as follows

$$\begin{aligned} \int_{1+\epsilon_1}^{\lceil t_0 \rceil} t f_g(t) dt &\leq \lceil t_0 \rceil \int_{1+\epsilon_1}^{\lceil t_0 \rceil} f_g(t) dt \\ &\leq \lceil t_0 \rceil \mathbb{P}\left\{g(M, K) \geq 1 + \epsilon_1\right\}. \end{aligned}$$

Set $x(M) = (\min[\sqrt{M}, M^\delta])^{1/3}$. With this choice of $x(M)$, it is not difficult to show that

$$\lim_{M, K \rightarrow \infty} A_1 \frac{M x(M)}{\sqrt{\max[K, M^{2+\delta}]} } = 0$$

$$\lim_{M, K \rightarrow \infty} A_2 \frac{x(M)}{\underline{C}^2 \sqrt{M}} = 0$$

$$\lim_{M, K \rightarrow \infty} \frac{x}{\underline{c}^2 \sqrt{K}} = 0$$

which yields

$$\lim_{M, K \rightarrow \infty} B^N(M, K, x(M)) = \lim_{M, K \rightarrow \infty} \frac{K}{\max[K, M^{2+\delta}]} \leq 1. \quad (2.69)$$

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

Using $\underline{C}_{\text{SN}}^2 = \underline{C}^2 + \sigma^2(\underline{c}^2 + 1)$, we can furthermore conclude that

$$\lim_{M, K \rightarrow \infty} B^D(M, K, x(M)) = 1$$

which, together with (2.69), implies that

$$\lim_{M, K \rightarrow \infty} B(M, K, x(M)) \leq 1.$$

We can, therefore, conclude that there exist $M_0^{(11)}, K_0^{(11)} > 0$ such that for all $M \geq M_0^{(11)}$ and $K \geq K_0^{(11)}$

$$B(M, K, x(M)) \leq 1 + \epsilon_1. \quad (2.70)$$

Trivially, we have

$$\lim_{M, K \rightarrow \infty} M^{\delta_1} K^{\delta_2} e^{-\Delta(x(M))^{2/7}} = 0$$

and, therefore, there exist $M_0^{(12)}, K_0^{(12)} > 0$ such that for all $M \geq M_0^{(12)}$ and $K \geq K_0^{(12)}$

$$A_3 M^{\delta_1} K^{\delta_2} e^{-\Delta(x(M))^{2/7}} \leq \frac{\epsilon}{3 \lceil t_0 \rceil}. \quad (2.71)$$

Combining (2.70) and (2.71) and setting

$$M_0^{(1)} = \max[M_0^{(11)}, M_0^{(12)}], \quad K_0^{(1)} = \max[K_0^{(11)}, K_0^{(12)}]$$

we get that for all $M \geq M_0^{(1)}$ and $K \geq K_0^{(1)}$

$$\lceil t_0 \rceil \mathbb{P}\{g(M, K) \geq 1 + \epsilon_1\} \leq \epsilon/3 \quad (2.72)$$

which concludes the proof of (2.67).

To show (2.68), we note that

$$\int_{\lceil t_0 \rceil}^{\infty} t f_g(t) dt \leq \sum_{n=\lceil t_0 \rceil}^{\infty} (n+1) \mathbb{P}\{g(M, K) \geq n\} \triangleq S. \quad (2.73)$$

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

Expanding the square, upper-bounding x by x^2 in $B^N(M, K, x)$, and substituting the max terms in $B^D(M, K, x)$ by 0, we obtain the bound

$$B(M, K, x) \leq \underbrace{\frac{K}{\max[K, M^{2+\delta}]} \frac{C_{\text{SN}}^2}{\sigma^2} \left(1 + \left(2A_1 \frac{M}{\sqrt{K}} + A_1^2 \frac{M^2}{K} \right) x^2 \right)}_{\triangleq B_1(M, K, x^2)}. \quad (2.74)$$

Applying the change of variables $y = x^2$ in (2.74) and (2.64), we finally get

$$\begin{aligned} \mathbb{P}\left\{g(M, K) \geq B_1(M, K, \sqrt{y})\right\} &\leq \mathbb{P}\left\{g(M, K) \geq B(M, K, \sqrt{y})\right\} \\ &\leq A_3 M^{\delta_1} K^{\delta_2} e^{-\Delta y^{1/7}}. \end{aligned} \quad (2.75)$$

Equating $B_1(M, K, y)$ with n and solving for y , we find that

$$\mathbb{P}\left\{g(M, K) \geq n\right\} \leq A_3 M^{\delta_1} K^{\delta_2} e^{-\Delta(y_2(n, M, K))^{1/7}}$$

with

$$y_2(n, M, K) = \frac{\frac{\max[K, M^{2+\delta}]}{K} \left(\frac{\sigma^2}{C_{\text{SN}}^2} n - \frac{K}{\max[K, M^{2+\delta}]} \right)}{2A_1 \frac{M}{\sqrt{K}} + A_1^2 \frac{M^2}{K}}. \quad (2.76)$$

Now, S defined in (2.73) can be upper-bounded as

$$\begin{aligned} S &\leq 2 \sum_{n=\lceil t_0 \rceil}^{\infty} n \mathbb{P}\{g(M, K) \geq n\} \\ &\leq 2A_3 M^{\delta_1} K^{\delta_2} \sum_{n=\lceil t_0 \rceil}^{\infty} n e^{-\Delta(y_2(n, M, K))^{1/7}}. \end{aligned} \quad (2.77)$$

If n is such that $\sigma^2 n / C_{\text{SN}}^2 > 1$, then the expression in the parentheses in the numerator of (2.76) is strictly positive and it follows that

$$\lim_{M, K \rightarrow \infty} y_2(n, M, K) = \infty.$$

Therefore, if t_0 is chosen such that $\lceil t_0 \rceil > C_{\text{SN}}^2 / \sigma^2$, each term in the sum in (2.77) goes to zero exponentially fast in M, K . Note that

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

the split-up in (2.67) and (2.68) was needed to be able to choose t_0 large enough here. To simplify the exposition in the following, we set $t_0 = (2^7 + 1) \underline{C}_{\text{SN}}^2 / \sigma^2$, so that

$$\left(\frac{\sigma^2}{\underline{C}_{\text{SN}}^2} n - \frac{K}{\max[K, M^{2+\delta}]} \right)^{1/7} \geq 2$$

for $n \geq \lceil t_0 \rceil$. Next, we note that

$$\lim_{M, K \rightarrow \infty} \frac{\max[K, M^{2+\delta}]}{K} \frac{1}{2A_1 \frac{M}{\sqrt{K}} + A_1^2 \frac{M^2}{K}} = \infty$$

so that there exist $M_0^{(2)}, K_0^{(2)} > 0$ such that for all $M \geq M_0^{(2)}$ and $K \geq K_0^{(2)}$

$$\left(\frac{\max[K, M^{2+\delta}]}{K} \frac{1}{2A_1 \frac{M}{\sqrt{K}} + A_1^2 \frac{M^2}{K}} \right)^{1/7} \geq 2.$$

Now using that, trivially,

$$xy \geq x + y$$

for $x, y \geq 2$, we have for all $M \geq M_0^{(2)}, K \geq K_0^{(2)}$ and $n \geq \lceil t_0 \rceil$

$$\begin{aligned} (y_2(n, M, K))^{1/7} &\geq \left(\frac{\sigma^2}{\underline{C}_{\text{SN}}^2} n - \frac{K}{\max[K, M^{2+\delta}]} \right)^{1/7} \\ &\quad + \left(\frac{\max[K, M^{2+\delta}]}{K} \frac{1}{2A_1 \frac{M}{\sqrt{K}} + A_1^2 \frac{M^2}{K}} \right)^{1/7} \end{aligned}$$

which yields

$$S \leq 2A_3 M^{\delta_1} K^{\delta_2} e^{-\Delta \left(\frac{2A_1 M \sqrt{K} + A_1^2 M^2}{\max[K, M^{2+\delta}]} \right)^{-1/7}} \sum_{n=\lceil t_0 \rceil}^{\infty} \eta(n)$$

with

$$\eta(n) \triangleq n \exp \left(-\Delta \left(\frac{\sigma^2}{\underline{C}_{\text{SN}}^2} n - 1 \right)^{1/7} \right).$$

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

Clearly, $\eta(n)$ decays fast enough for $\sum_{n=\lceil t_0 \rceil}^{\infty} \eta(n)$ to converge to a finite limit, in other words, there exists a constant $c < \infty$ (independent of M, K) such that

$$\sum_{n=\lceil t_0 \rceil}^{\infty} \eta(n) \leq c. \quad (2.78)$$

Moreover, it is easily seen that

$$\lim_{M, K \rightarrow \infty} M^{\delta_1} K^{\delta_2} e^{-\Delta \left(\frac{2A_1 M \sqrt{K} + A_1^2 M^2}{\max[K, M^{2+\delta}]} \right)^{-1/7}} = 0$$

which, together with (2.78), shows that S can be made arbitrarily small by choosing M and K large enough. More specifically, there exist $M_0^{(3)}, K_0^{(3)} > 0$ such that for all $M \geq M_0^{(3)}$ and $K \geq K_0^{(3)}$

$$S \leq \epsilon/3. \quad (2.79)$$

Taking

$$\begin{aligned} M_0 &\triangleq \max[M_0^{(1)}, M_0^{(2)}, M_0^{(3)}] \\ K_0 &\triangleq \max[K_0^{(1)}, K_0^{(2)}, K_0^{(3)}] \end{aligned}$$

and combining (2.63), (2.72), and (2.79), we have shown (2.61), which completes the proof. \square

2.4.2. The “Crystallization” Phenomenon

As pointed out in the beginning of this chapter, the “crystallization” phenomenon occurs for $M, K \rightarrow \infty$, provided that K scales fast enough as a function of M , and manifests itself in two effects, namely, the *decoupling* of the individual $\mathcal{S}_m \rightarrow \mathcal{D}_m$ links and the *convergence of each of the resulting SISO links to a nonfading link*.

A. Decoupling of the network

Theorems 2.3 and 2.4 show that in the $M, K \rightarrow \infty$ limit, the per-source destination terminal pair capacity scales as

$$\begin{aligned} C_{\text{P1}} &= (1/2) \log(1 + \Theta(K/M^3)) \\ C_{\text{P2}} &= (1/2) \log(1 + \Theta(K/M^2)) \end{aligned}$$

in P1 and P2, respectively. We can, therefore, conclude that if $K \propto M^{3+\alpha}$ in P1 and $K \propto M^{2+\alpha}$ in P2 with $\alpha \geq 0$, apart from the factor $1/2$, which is due to the use of two time slots, P1 and P2 achieve full spatial multiplexing gain (Tse and Viswanath, 2005) (i.e., full sum-capacity pre-log) without any cooperation of the terminals in the network, not even the destination terminals. The corresponding distributed array gain (i.e., the factor inside the log) is given by M^α in both cases.

The fact that the per source-destination terminal pair capacity is strictly positive when K scales at least as fast as M^3 in P1 and at least as fast as M^2 in P2 shows that the individual $\mathcal{S}_m \rightarrow \mathcal{D}_m$ links in the network “decouple” in the sense that the SINR is strictly positive for each of the links. Note that this does not imply that the interference at the \mathcal{D}_m (created by $s_{\hat{m}}$ with $\hat{m} \neq m$) vanishes. Rather, if K scales fast enough, the signal power starts dominating the interference (plus noise) power. Since both upper and lower bounds in Theorems 2.3 and 2.4 exhibit the same scaling behavior, the $K \propto M^3$ and $K \propto M^2$, respectively, thresholds are fundamental in the sense of defining the critical scaling rate by delineating the regime where interference dominates over the signal and hence drives the per source-destination terminal pair capacity to zero from the regime where the signal dominates the interference and the per source-destination terminal pair capacity is strictly positive. Further inspection of the upper and lower bounds in (2.58) and (2.59) reveals that, for fixed $\epsilon > 0$, unless all path loss and shadowing coefficients $E_{k,m}$ and $P_{m,k}$ ($k \in [1:K], m \in [1:M]$) are equal and hence $\bar{C}^2 = \underline{C}^2$ and $\bar{C}_{\text{SN}}^2 = \underline{C}_{\text{SN}}^2$, there is a gap (apart from that due to $\epsilon > 0$) between the bounds.

The order-of-magnitude reduction in the threshold for critical scal-

ing in P2, when compared with P1, comes at the cost of each relay having to know all M backward and M forward channels. We can, therefore, conclude that P1 and P2 trade off the number of relay terminals for channel knowledge at the relays.

Finally, it is worthwhile to point out that in contrast to the finite- M results for P1 reported by Bölcskei et al. (2006), the destination terminals \mathcal{D}_m do not need knowledge of the fading coefficients $h_{k,m}$ and $f_{m,k}$. This can be seen by noting that the quantity $(\pi/(4\sqrt{K})) \sum_{k:p(k)=m} C_{P1,k}^{m,m}$, which has to be known at \mathcal{D}_m , depends on $E_{k,m}$, $P_{m,k}$, K , and M only. Moreover, the coefficient $(\pi/(4\sqrt{K})) \sum_{k:p(k)=m} C_{P1,k}^{m,m}$ can easily be acquired through training.

B. Convergence to nonfading links and “crystallization”

When the network decouples, it is interesting to ask how the decoupled SISO links behave (in terms of their fading statistics) when M and K grow large. The answer to this question follows from the concentration results in Theorems 2.1 and 2.2, which can be reformulated to establish upper bounds on the outage probability for the individual $\mathcal{S}_m \rightarrow \mathcal{D}_m$ links. For the sake of brevity, we focus on P1 in what follows. The goal is to arrive at a statement regarding

$$\begin{aligned} P_{\text{out},P1}(R) &\triangleq \mathbb{P}\left\{\frac{1}{2} \log(1 + \text{SINR}_m^{\text{P1}}) \leq R\right\} \\ &= \mathbb{P}\{\text{SINR}_m^{\text{P1}} \leq 2^{2R} - 1\}. \end{aligned}$$

The corresponding result is summarized as follows.

Theorem 2.5 (Outage probability for P1).

1. Assume that $K \geq 2$, $M \geq 2$, and $R \geq 0$ are such that

$$x(R) = \frac{1 - e_{P1}(M, K, R)}{\frac{16}{C\pi} \frac{M}{\sqrt{K}} + e_{P1}(M, K, R) \left(\frac{3}{C_{\text{SN}}^2} \frac{1}{\sqrt{M}} + \frac{\sigma^2}{C_{\text{SN}}^2} \frac{1}{\sqrt{K}} \right)} \geq 1 \quad (2.80)$$

where

$$e_{P1}(M, K, R) = \frac{16}{\pi^2} \frac{\overline{C}_{\text{SN}}^2}{C^2} \frac{M^3}{K} (2^{2R} - 1).$$

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

Then, the individual link outage probability is upper-bounded as

$$P_{\text{out},\text{P1}}(R) \leq 151 K^2 M e^{-\Delta_{\text{P1}} x(R)^{2/7}}. \quad (2.81)$$

2. Under the same conditions on K, M and R as in 1), for all $\epsilon, \delta > 0$, $K \geq M^{3+\delta}$, and

$$R \leq \frac{1}{2} \log \left(1 + \frac{\pi^2}{16} \frac{\underline{C}^2}{\overline{C}_{\text{SN}}^2} \frac{K}{M^3} (1 - \epsilon) \right), \quad (2.82)$$

we have

$$\lim_{M, K \rightarrow \infty} P_{\text{out},\text{P1}}(R) \leq \lim_{M, K \rightarrow \infty} 151 K^2 M e^{-\Delta_{\text{P1}} x(R)^{2/7}} = 0.$$

Proof. We start with the proof of statement 1). Recall that Theorem 2.1 provides us with a parametric upper bound on

$$\mathbb{P}\{\text{SINR}_m^{\text{P1}} \leq L_{\text{P1}}(x)\}$$

with $L_{\text{P1}}(x)$ defined in (2.54). Assuming that

$$x \leq \frac{C\pi\sqrt{K}}{16M} \quad (2.83)$$

and using $\overline{C}_{\text{SN}}^2 = \overline{C}^2 + \sigma^2(\overline{c}^2 + 1)$, we can lower-bound $L_{\text{P1}}(x)$ as

$$L_{\text{P1}}(x) \geq \frac{\pi^2}{16} \frac{\underline{C}^2}{\overline{C}_{\text{SN}}^2} \frac{K}{M^3} \frac{1 - \frac{16}{\underline{C}\pi} \frac{M}{\sqrt{K}} x}{1 + \frac{3}{\overline{C}_{\text{SN}}^2} \frac{x}{\sqrt{M}} + \frac{\sigma^2}{\overline{C}_{\text{SN}}^2} \frac{x}{\sqrt{K}}} \triangleq L'_{\text{P1}}(x).$$

Solving

$$2^{2R} - 1 = L'_{\text{P1}}(x) \quad (2.84)$$

for $x(R)$ yields (2.80), which, by assumption, satisfies $x(R) \geq 1$. With

$$\mathbb{P}\{\text{SINR}_m^{\text{P1}} \leq L'_{\text{P1}}(x)\} \leq \mathbb{P}\{\text{SINR}_m^{\text{P1}} \leq L_{\text{P1}}(x)\}$$

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

we can now apply¹⁴ Theorem 2.1 to obtain

$$P_{\text{out},\text{P1}}(R) \leq 151 K^2 M e^{-\Delta_{\text{P1}} x(R)^{2/7}}. \quad (2.85)$$

Finally, we note that $x(R)$ in (2.80) is trivially seen to satisfy (2.83). This concludes the proof of statement 1).

The proof of statement 2) is obtained by establishing a sufficient condition on $x(R)$, for every $R \geq 0$, to grow with increasing M (and by $K \geq M^{3+\delta}$ with increasing K). Using (2.80), it is easily verified that guaranteeing

$$0 \leq e_{\text{P1}}(M, K, R) \leq 1 - \epsilon$$

for some $0 < \epsilon < 1$ (independent of M, K) provides such a condition. The final result is now obtained by solving

$$e_{\text{P1}}(M, K, R) = \frac{16 \overline{C}_{\text{SN}}^2}{\pi^2} \frac{M^3}{\underline{C}^2} \frac{1}{K} (2^{2R} - 1) \leq 1 - \epsilon$$

for R . □

The implications of Theorem 2.5 are significant: For any transmission rate R less than the ergodic capacity (in the case $E_{k,m} = P_{m,k}$ for all k, m) or the ergodic capacity lower bound in Theorem 2.3 (in the case of general $E_{k,m}$ and $P_{m,k}$), the outage probability of each of the decoupled links goes to zero exponentially fast in the number of nodes in the network, provided K scales supercritically in M . We have thus shown that choosing the rate of growth of K fast enough for the network to decouple automatically guarantees that the decoupled SISO links converge to nonfading links. Equivalently, we can say that each of the decoupled links experiences a distributed spatial diversity (or, more precisely, relay diversity) order that goes to infinity as $M \rightarrow \infty$. Consequently, in the large- M limit time diversity (achieved by coding over a sufficiently long time horizon) is not needed to achieve ergodic capacity. We say that the network “crystallizes” as it breaks up into

¹⁴Strictly speaking, one needs to use the upper bounds on $\mathbb{P}\{\text{SINR}_m^{\text{P1}} \leq L_{\text{P1}}(x)\}$ derived in the last paragraph of Appendix C.1.

a set of effectively isolated “wires in the air”. From (2.81), we can furthermore infer the “crystallization” rate, i.e., the rate (as a function of M and K) at which the individual $\mathcal{S}_m \rightarrow \mathcal{D}_m$ links converge to nonfading links. We note, however, that the exponent $2/7$ (and $2/9$ for P2) is unlikely to be fundamental as it is probably a consequence of the application of the truncation technique. In this sense, we can only specify a guaranteed crystallization rate. We conclude by noting that the upper bound (2.85) (as well as the corresponding result for P2) tend to be rather loose. This is probably a consequence of the truncation technique and the use of union bounds to characterize the large-deviations behavior of the individual link SINR RVs.

Numerical results: We shall finally provide numerical results quantifying the outage behavior of P1 and P2. For simplicity, we set $E_{k,m} = P_{m,k} = 1$ for all m, k and $\sigma^2 = 0.01$ in both simulation examples. This choice for the path loss and shadowing parameters, although not representative of a real-world propagation scenario, isolates the dependence of our results on the network geometry. Moreover, it ensures that the distribution of the different SINR RVs for a given protocol is identical for all links so that it suffices to analyze the behavior of only one SINR RV for each of the two protocols. For $K = M^3$ in P1 and $K = M^2$ in P2, Figure 2.3 shows the cumulative distribution functions (CDFs) (obtained through Monte-Carlo simulation) of SINR^{P1} and SINR^{P2} , respectively, for different values of M . We observe that, for increasing M , the CDFs approach a step function at the corresponding mean values, i.e., the SINR RVs, indeed, converge to a deterministic quantity, and, consequently, the underlying fading channel converges to a nonfading channel. The limiting mean values are given by the lower and upper bounds (which coincide in the case $E_{k,m} = P_{m,k} = 1$ for all m, k) in (2.58) and (2.59) for P1 and P2, respectively. We can furthermore see that for fixed M the CDFs are very similar for P1 and P2 (recall, however, that $K = M^3$ in P1 and $K = M^2$ in P2), suggesting that the convergence behavior is similar for the two protocols. The difference in the theoretically predicted convergence exponents ($2/7$ for P1 and $2/9$ for P2) therefore does not

seem to be fundamental to the two protocols and may, indeed, be a consequence of our proof technique as already pointed out above.

2.4.3. Cooperation at the Relay Level

The analysis carried out so far was based on the assumption that the relays cannot cooperate. The purpose of this section is to investigate the impact of cooperation (in fact, a specific form of cooperation) at the relay level on the ergodic-capacity scaling behavior in the coherent-relaying case. Note that we continue to assume that the destination terminals cannot cooperate. Before proceeding, we would like to mention that concentration results and an outage analysis along the lines of the discussion in Sections 2.3 and 2.4.2 are possible, but will be omitted for brevity of exposition.

Cooperation at the relay level will be accounted for by grouping the K single-antenna relay terminals into Q groups

$$\mathcal{G}_q \triangleq \left\{ \mathcal{R}_{(q-1)L+1}, \mathcal{R}_{(q-1)L+2}, \dots, \mathcal{R}_{qL} \right\}, \quad q \in [1:Q]$$

with L relays in each group¹⁵ and by assuming that the relays in each group can fully cooperate, but cooperation across groups is not possible. In order to simplify the exposition, in the remainder of this section, we think of a group \mathcal{G}_q ($q \in [1:Q]$) as a single relay element with L antenna elements and use the term “vector-relay (v-relay)” terminal to address the L -antenna relays $\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_Q$. For $q \in [1:Q]$ and $m \in [1:M]$, the following notation will be used:

$$\begin{aligned} \mathbf{r}_q &\triangleq [r_{(q-1)L+1} \ r_{(q-1)L+2} \ \cdots \ r_{qL}]^\top \\ \mathbf{t}_q &\triangleq [t_{(q-1)L+1} \ t_{(q-1)L+2} \ \cdots \ t_{qL}]^\top \\ \mathbf{z}_q &\triangleq [z_{(q-1)L+1} \ z_{(q-1)L+2} \ \cdots \ z_{qL}]^\top \\ \mathbf{h}_{q,m} &\triangleq [h_{(q-1)L+1,m} \ h_{(q-1)L+2,m} \ \cdots \ h_{qL,m}]^\top \\ \mathbf{f}_{m,q} &\triangleq [f_{m,(q-1)L+1} \ f_{m,(q-1)L+2} \ \cdots \ f_{m,qL}]^\top \end{aligned}$$

¹⁵For simplicity, we assume that Q divides K so that $K = QL$.

2. "CRYSTALLIZATION" IN NETWORKS WITH COHERENT RELAYING

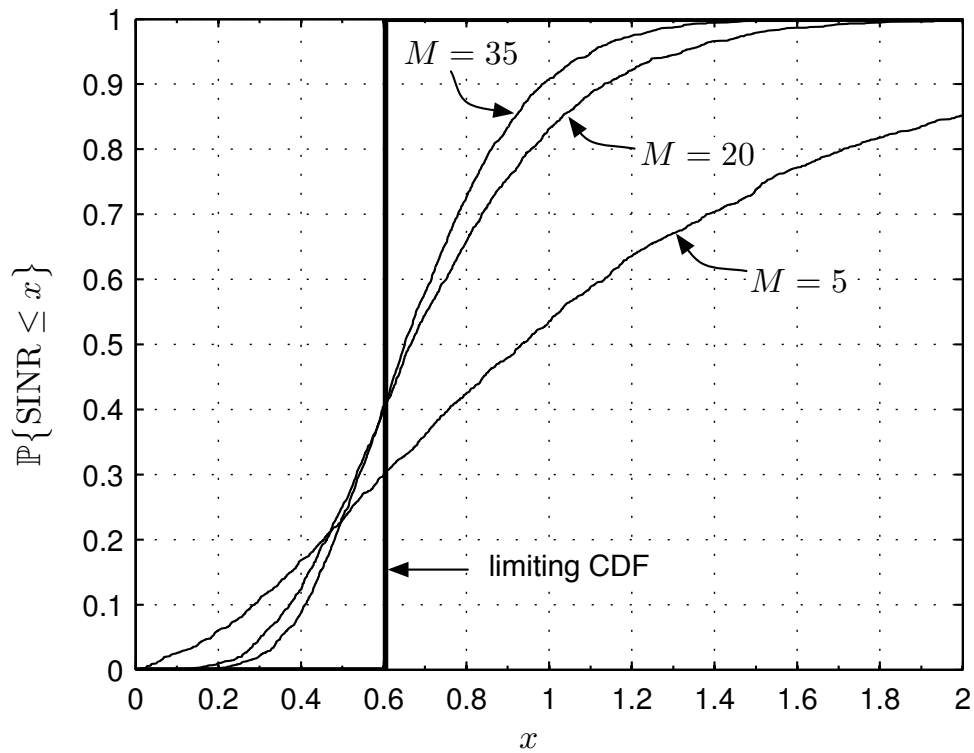
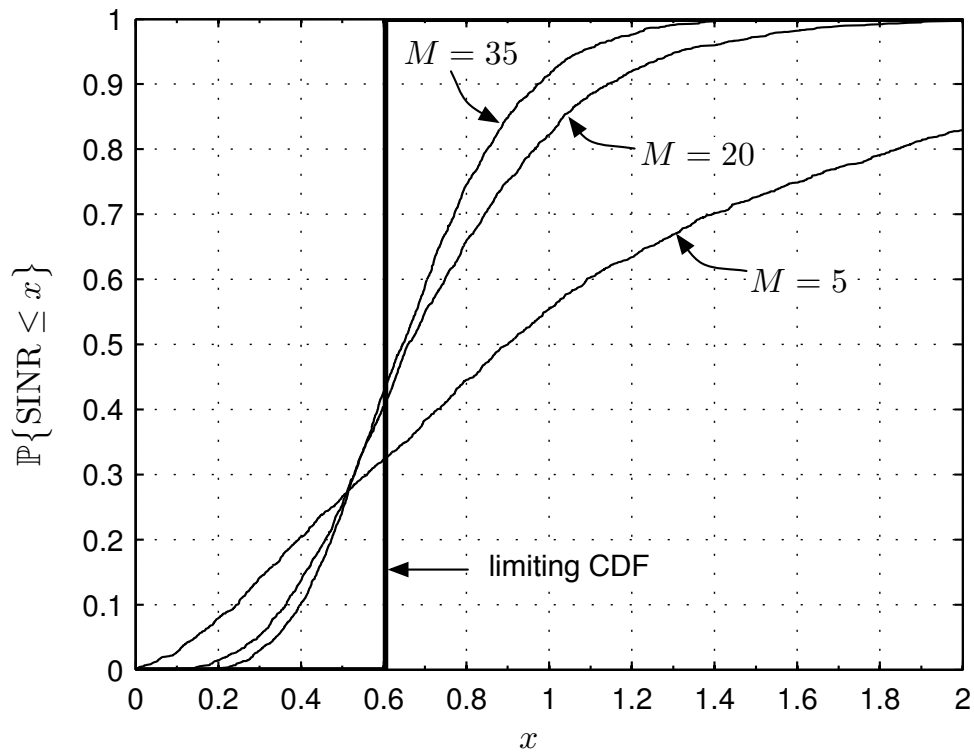


Fig. 2.3.: Simulated (Monte-Carlo) SINR CDFs for different values of M for (a) $K = M^3$ in P1 and (b) $K = M^2$ in P2.

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

where \mathbf{r}_q and \mathbf{t}_q are the (L -dimensional) vector-valued signals received and transmitted by the q th v-relay, respectively, \mathbf{z}_q is additive noise at the q th v-relay, $\mathbf{h}_{q,m}$ contains the channel gains for the $\mathcal{S}_m \rightarrow \mathcal{G}_q$ link, and $\mathbf{f}_{m,q}$ contains the channel gains for the $\mathcal{G}_q \rightarrow \mathcal{D}_m$ link. Additionally, for simplicity, we assume that relays belonging to a given group q are located close to each other so that

$$\begin{aligned}\hat{E}_{q,m} &\triangleq E_{(q-1)L+1,m} = E_{(q-1)L+2,m} = \cdots = E_{qL,m} \\ \hat{P}_{m,q} &\triangleq P_{m,(q-1)L+1} = P_{m,(q-1)L+2} = \cdots = P_{m,qL}\end{aligned}$$

for $q \in [1:Q]$ and $m \in [1:M]$. With this notation, the IO relations (2.1) and (2.2) for the $\mathcal{S}_m \rightarrow \mathcal{G}_q$ links and the $\mathcal{G}_q \rightarrow \mathcal{D}_m$ links can be written as

$$\mathbf{r}_q = \sum_{m=1}^M \hat{E}_{q,m} \mathbf{h}_{q,m} s_m + \mathbf{z}_q, \quad q \in [1:Q]$$

and

$$y_m = \sum_{q=1}^Q \hat{P}_{m,q} \mathbf{f}_{m,q}^\top \mathbf{t}_q + w_m, \quad m \in [1:M]$$

respectively. Next, we describe the generalization of the protocols P1 and P2 to the case of v-relays making the aspect of cooperation at the relay level explicit.

A. P1 for the Cooperative Case

Like in the case of single-antenna relays (described in Section 2.3.1), we partition the Q v-relay terminals into M subsets \mathcal{M}_m ($m \in [1:M]$) with¹⁶ $|\mathcal{M}_m| = Q/M$. The v-relays (each of which has L antenna elements) in \mathcal{M}_m are assumed to assist the m th source-destination terminal pair $\{\mathcal{S}_m, \mathcal{D}_m\}$, and the relay partitioning function $p : [1, Q] \rightarrow [1, M]$ is defined as

$$p(q) \triangleq m \Leftrightarrow \mathcal{G}_q \in \mathcal{M}_m.$$

¹⁶For simplicity, we assume that M divides Q .

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

We assume that the q th v-relay terminal has perfect knowledge of the phases of the SIMO backward channel $\mathcal{S}_{p(q)} \rightarrow \mathcal{G}_q$ and the phases of the corresponding multiple-input single-output (MISO) forward channel $\mathcal{G}_q \rightarrow \mathcal{D}_{p(q)}$. This implies that perfect knowledge of the vectors

$$\tilde{\mathbf{h}}_{q,p(q)} \triangleq \left[e^{i \arg([\mathbf{h}_{q,p(q)}]_1)} \quad e^{i \arg([\mathbf{h}_{q,p(q)}]_2)} \quad \dots \quad e^{i \arg([\mathbf{h}_{q,p(q)}]_L)} \right]^T$$

and

$$\tilde{\mathbf{f}}_{p(q),q} \triangleq \left[e^{i \arg([\mathbf{f}_{p(q),q}]_1)} \quad e^{i \arg([\mathbf{f}_{p(q),q}]_2)} \quad \dots \quad e^{i \arg([\mathbf{f}_{p(q),q}]_L)} \right]^T$$

is available at \mathcal{G}_q . The signal \mathbf{r}_q received at the q th v-relay terminal is phase-matched-filtered first w.r.t. the assigned backward channel $\mathcal{S}_{p(q)} \rightarrow \mathcal{G}_q$ and then w.r.t. the assigned forward channel $\mathcal{G}_q \rightarrow \mathcal{D}_{p(q)}$ followed by a normalization so that

$$\mathbf{t}_q = d_{P1,q} \tilde{\mathbf{f}}_{p(q),q}^* \left(\tilde{\mathbf{h}}_{q,p(q)}^H \mathbf{r}_q \right) \quad (2.86)$$

where¹⁷ the choice

$$d_{P1,q} \triangleq \frac{1}{L} \sqrt{P_{\text{rel}}} \left[\frac{Q}{M} \sum_{m=1}^M \hat{E}_{q,m} + \frac{\pi(L-1)Q}{4M} \hat{E}_{q,p(q)} + Q\sigma^2 \right]^{-1/2}$$

ensures that the per-v-relay power constraint

$$\mathbb{E}[\|\mathbf{t}_q\|^2] = P_{\text{rel}}/Q \quad q \in [1:Q]$$

and consequently the total (across v-relays) power constraint

$$\sum_{q=1}^Q \mathbb{E}[\|\mathbf{t}_q\|^2] = P_{\text{rel}}$$

¹⁷The quantity $d_{P1,q}$, used in this section is (for $L > 1$) different from $d_{P1,k}$ defined in (2.5). We use the same symbol for notational simplicity and employ the index q (instead of k) consistently, in order to resolve potential ambiguities. The same comment applies to other variables redefined in this section.

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

is met. As in the single-antenna relay (i.e., noncooperative) case, P1 ensures that the relays $\mathcal{G}_q \in \mathcal{M}_m$ forward the signal intended for \mathcal{D}_m in a “doubly coherent” (w.r.t. the assigned backward and forward channel) fashion whereas the signals transmitted by the source terminals $\mathcal{S}_{\hat{m}}$ with $\hat{m} \neq m$ are forwarded to \mathcal{D}_m in a “noncoherent” fashion (i.e., phase incoherence occurs either on the backward or the forward link or on both links). From (2.86), we can see that cooperation in groups of L single-antenna relays is realized by phase combining on the backward and forward links of each v-relay. More sophisticated forms of cooperation such as equalization on the backward link and precoding on the forward link are certainly possible, but are beyond the scope of this thesis.

B. P2 for the Cooperative Case

Like in the case of single-antenna relays (i.e., the noncooperative case), P2 requires that each relay, in fact here v-relay, knows the phases of all its M vector-valued backward and forward channels, i.e., \mathcal{G}_q needs knowledge of $\tilde{\mathbf{h}}_{q,m}$ and $\tilde{\mathbf{f}}_{m,q}$, respectively, for $m \in [1:M]$. The relay processing stage in P2 computes

$$\mathbf{t}_q = d_{\text{P2},q} \left(\sum_{m=1}^M \tilde{\mathbf{f}}_{m,q}^* \tilde{\mathbf{h}}_{q,m}^H \right) \mathbf{r}_q$$

where

$$d_{\text{P2},q} \triangleq \frac{1}{L} \sqrt{P_{\text{rel}}} \left[Q \sum_{m=1}^M \hat{E}_{q,m} + \frac{\pi(L-1)Q}{4M} \sum_{m=1}^M \hat{E}_{q,m} + MQ\sigma^2 \right]^{-1/2}$$

ensures that the per-v-relay power constraint

$$\mathbb{E}[\|\mathbf{t}_q\|^2] = P_{\text{rel}}/Q \quad q \in [1:Q]$$

and, consequently, the total (across relays) power constraint

$$\sum_{q=1}^Q \mathbb{E}[\|\mathbf{t}_q\|^2] = P_{\text{rel}}$$

is met.

C. Ergodic-Capacity Results

We are now ready to establish the impact of cooperation at the relay level on the ergodic capacity scaling laws for P1 and P2. Our results are summarized in Theorems 2.6 and 2.7 below.

Theorem 2.6 (Capacity of P1 with cooperation). *Suppose that destination terminal \mathcal{D}_m ($m \in [1:M]$) has perfect knowledge of the mean of the effective channel gain of the $\mathcal{S}_m \rightarrow \mathcal{D}_m$ link, given by $(\pi/4)L^2 \sum_{q:p(q)=m} d_{P1,q} \hat{P}_{m,q} \hat{E}_{q,m}$. Then, for all $\epsilon, \delta > 0$, there exist $M_0, Q_0 > 0$ such that for all $M \geq M_0$ and $Q \geq Q_0$ the per source-destination terminal pair ergodic capacity achieved by P1 satisfies¹⁸*

$$\begin{aligned} \frac{1}{2} \log \left(1 + \frac{\pi^2 Q L^2}{16 M^3} \frac{\underline{C}^2}{\underline{C}_{\text{SN}}^2} (1 - \epsilon) \right) &\leq C_{\text{P1}} \\ &\leq \frac{1}{2} \log \left(1 + \frac{\pi^2 \max[Q, M^{2+\delta}] L^2 \bar{C}^2}{16 M^3 \underline{C}_{\text{SN}}^2} (1 - \epsilon) \right). \end{aligned} \quad (2.87)$$

Theorem 2.7 (Capacity of P2 with cooperation). *Suppose that destination terminal \mathcal{D}_m ($m \in [1:M]$) has perfect knowledge of the mean of the effective channel gain of the $\mathcal{S}_m \rightarrow \mathcal{D}_m$ link, given by $(\pi/4)L^2 \sum_{q=1}^Q d_{P2,q} \hat{P}_{m,q} \hat{E}_{q,m}$. Then, for all $\epsilon, \delta > 0$, there exist $M_0, Q_0 > 0$ such that for all $M \geq M_0$, $Q \geq Q_0$ the per source-destination terminal pair ergodic capacity achieved by P2 satisfies*

$$\begin{aligned} \frac{1}{2} \log \left(1 + \frac{\pi^2 Q L^2}{16 M^2} \frac{\underline{C}^2}{\underline{C}_{\text{SN}}^2} (1 - \epsilon) \right) &\leq C_{\text{P2}} \\ &\leq \frac{1}{2} \log \left(1 + \frac{\pi^2 \max[Q, M^{1+\delta}] L^2 \bar{C}^2}{16 M^2 \underline{C}_{\text{SN}}^2} (1 - \epsilon) \right). \end{aligned} \quad (2.88)$$

Proof of Theorems 2.6 and 2.7. The upper bounds in (2.87) and (2.88) are again established based on a concentration result for the individual

¹⁸Note that the quantities \bar{C}_{SN} , \underline{C} , \bar{C} , and $\underline{C}_{\text{SN}}$ used in this section have been defined in Section 2.3.

link SINRs and the lower bounds build on the technique summarized in Appendix C.3. The proofs of Theorems 2.6 and 2.7 are almost identical to the proofs of Theorems 2.3 and 2.4, respectively, and do not require new techniques. There is, however, one important aspect in which Theorems 2.6 and 2.7 differ from Theorems 2.3 and 2.4, namely, the appearance of the factor L^2 in (2.87) and (2.88). To demonstrate where this factor comes from, we provide the proof of the ergodic capacity lower bound for P1 in Appendix C.2. The proofs of the remaining statements will be omitted for brevity of exposition. \square

Discussion of results: Just like in the noncooperative (i.e., single-antenna relay) case, we can conclude that asymptotically in M if $K \propto M^{3+\alpha}$ in P1 and $K \propto M^{2+\alpha}$ in P2 with $\alpha > 0$, the network decouples.

The effect of cooperation (through phase matched-filtering) at the relay level manifests itself in the presence of the factor L^2 inside the log in the bounds for C_{P1} and C_{P2} stated in Theorems 2.6 and 2.7, respectively. We can summarize the results of Theorems 2.6 and 2.7 as¹⁹

$$C_{P1} = \frac{1}{2} \log \left(1 + \Theta \left(\frac{QL^2}{M^3} \right) \right)$$

$$C_{P2} = \frac{1}{2} \log \left(1 + \Theta \left(\frac{QL^2}{M^2} \right) \right).$$

We can, therefore, conclude that the per-stream array gain A is given by $A_{P1} = QL^2/M^3$ for P1 and $A_{P2} = QL^2/M^2$ for P2. On a conceptual level, the array gain can be decomposed into a contribution due to distributed array gain, A_d , and a contribution due to cooperation at the relay level (realized by phase matching on backward and forward links), A_c , i.e., $A = A_d A_c$ with $A_{d,P1} = QL/M^3$, $A_{d,P2} = QL/M^2$, and $A_{c,P1} = A_{c,P2} = L$. To illustrate the impact of cooperation at

¹⁹Note that we use the $\Theta(\cdot)$ notation only to hide the dependence on \underline{E} , \underline{E} , \underline{P} , and \overline{P} . Strictly speaking, as L is finite it should also be hidden under the $\Theta(\cdot)$ notation. However, our goal is to exhibit the impact of cooperation at the relay level on C_{P1} and C_{P2} , which is the reason for making the dependence on L explicit.

the relay level, we compare a network with K noncooperating single-antenna relays to a network with a total of $K = QL$ single-antenna relays cooperating in groups of L single-antenna relays. In the case where there is no cooperation at the relay level, we have

$$C_{\text{P1}}^{(nc)} = \frac{1}{2} \log \left(1 + \Theta \left(\frac{K}{M^3} \right) \right)$$

whereas if the relays cooperate in groups of L single-antenna relays, we get

$$C_{\text{P1}}^{(c)} = \frac{1}{2} \log \left(1 + \Theta \left(\frac{KL}{M^3} \right) \right).$$

Cooperation at the relay level (realized by phase matched-filtering) in groups of L single-antenna relays therefore yields an L -fold increase in the effective per-stream SINR due to additional array gain given by $A_c = L$. Equivalently, the total number of single-antenna relays needed to achieve a given per source-destination terminal pair capacity is reduced by a factor of L through cooperation in groups of L single-antenna relay elements. The conclusions for P2 are identical.

As already pointed out above, the network decouples into effectively isolated source-destination pair links for every finite $L > 1$. Even though a concentration analysis along the lines of Theorems 2.1 and 2.2 was not reported (for the sake of brevity), it can be shown that for finite $L > 1$ the individual links converge to nonfading links as $M, Q \rightarrow \infty$, provided that Q scales supercritically as a function of M .

Numerical example: We conclude this section with a numerical example that demonstrates the impact of cooperation at the relay level, where we use the same parameters as in the simulation examples at the end of Section 2.4.2. Figure 2.4 shows the SINR CDF for P1 with $L = 4$ and $QL = M^3$ (the case $L = 1$ shown in Figure 2.3 is included for reference). We observe that, as pointed out above, for increasing M , we, indeed, get convergence of the fading link to a nonfading link. Moreover, we can also see that increasing L for fixed M results in higher per source-destination terminal pair capacity, but at

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

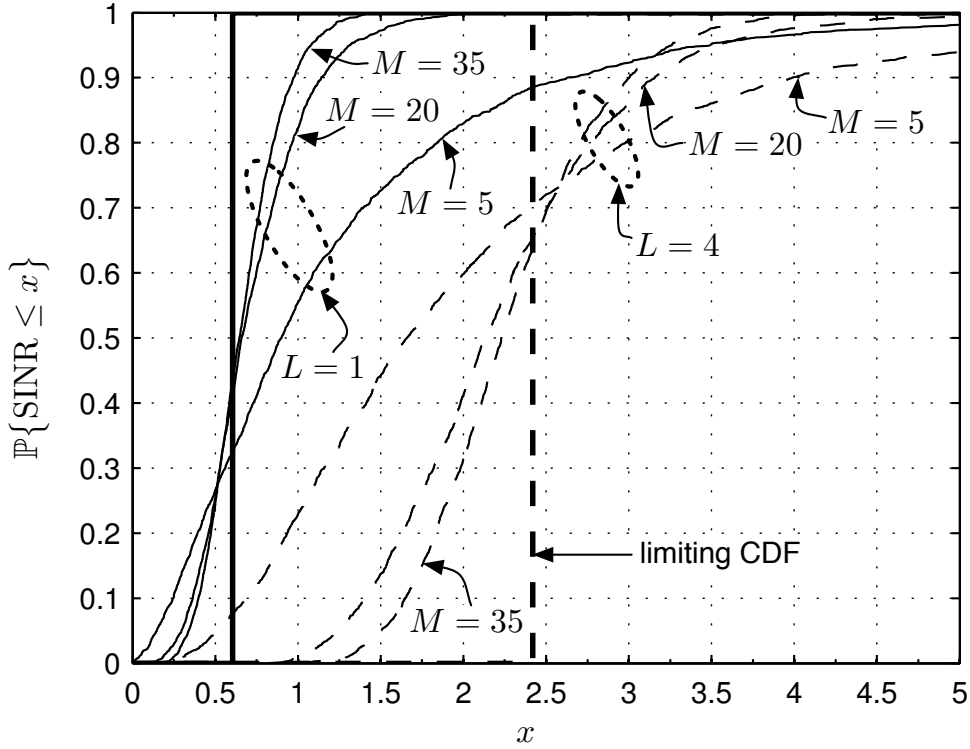


Fig. 2.4.: Simulated (Monte-Carlo) SINR CDFs for different values of M for $QL = M^3$ in P1 with $L = 1$ and $L = 4$.

the same time slows down convergence (w.r.t. M and hence also Q) of the link SINRs to their deterministic limits.

2.5. SUMMARY OF RESULTS

The minimum rate of growth of the number of relays K , as a function of the number of source-destination terminal pairs M , for fading interference networks with coherent relays to decouple was shown to be $K \propto M^3$ under protocol P1 and $K \propto M^2$ under protocol P2. P1 requires relay partitioning and the knowledge of one backward and one forward fading coefficient at each relay, whereas P2 does not need relay partitioning, but requires that each relay knows all its M backward and M forward fading coefficients. The protocols P1 and P2 are thus found to trade off CSI at the relays for the required (for the network to decouple) rate of growth of K as a function of M .

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

We found that cooperation at the relay level in groups of L relays, both for P1 and P2, results in an L -fold reduction of the total number of relays needed to achieve a given per source-destination terminal pair capacity. An interesting open question in this context is whether more sophisticated signal processing at the relays (such as equalization for the backward link and precoding for the forward link) could lead to improved capacity scaling behavior.

It was furthermore shown that the critical growth rates $K \propto M^3$ in P1 and $K \propto M^2$ in P2 are sufficient to not only make the network decouple, but to also make the individual source-destination fading links converge to nonfading links. We say that the *network “crystallizes”* as it breaks up into a set of *effectively isolated “wires in the air”*. More pictorially, the decoupled links experience increasing distributed spatial (or more specifically relay) diversity. Consequently, in the large- M limit time diversity (achieved by coding over a sufficiently long time horizon) is not needed to achieve ergodic capacity. We furthermore characterized the “crystallization” rate (more precisely a guaranteed “crystallization” rate as we do not know whether our bounds are tight), i.e., the rate (as a function of M, K) at which the decoupled links converge to nonfading links. In the course of our analysis, we developed a new technique for characterizing the large-deviations behavior of certain sums of dependent random variables.

The large-deviations analysis, along with the notion of decoupling of the network, as carried out in this thesis could serve as a general tool to assess the impact of protocols, processing at the relays, propagation conditions, routing, and scheduling on network outage and ergodic capacity performance. More specifically, an interesting question is under which conditions “crystallization” can happen in a general network and, if it occurs, what the corresponding “crystallization” rate would be. It has to be noted, however, that, in view of the technical difficulties posed by the basic case analyzed in this chapter, it is unclear whether this framework can yield substantial analytical insights into the above-mentioned questions.

Finally, we note that if we interpret our results in terms of per-node throughput, we find that P1 achieves $\mathcal{O}(1/n^{2/3})$ whereas P2

2. “CRYSTALLIZATION” IN NETWORKS WITH COHERENT RELAYING

realizes $\mathcal{O}(1/\sqrt{n})$, where $n = 2M + K$ is the total number of nodes in the network. The scaling law for P2 is exactly the same as the behavior established by Gupta and Kumar (2002) and the per-node throughput goes to zero. On the other hand, it is interesting to observe that we can get an $\mathcal{O}(1/\sqrt{n})$ throughput without imposing any assumptions on the path loss behavior. General conclusions on the impact of fading on the network-capacity scaling law cannot be drawn as we are considering a specific setup and specific protocols. Based on the work of Aeron and Saligrama (2007), Özgür et al. (2007) have recently shown, however, that under optimistic assumptions on CSI in the network (every user must know all the channels globally), $\mathcal{O}(1)$ per-node throughput can be achieved using hierarchical cooperation. For more details on this, see the Ph.D. thesis of Özgür (2009).

CHAPTER 3

Networks with Noncoherent Amplify-and-Forward Relaying

In Chapter 2, we have considered networks with coherent relaying, where each relay terminal is assumed to know its assigned backward and forward channels (P1) or all its backward and all its forward channels (P2) perfectly. In this chapter, we relax this assumption and study networks with no CSI at the relay terminals, i.e., networks with noncoherent relaying. In particular, we investigate a simple AF architecture where the relay terminals, in the second time slot, forward (without additional processing) a scaled version of the signal received in the first time slot. As already mentioned in Section 2.2, the source terminals do not have CSI. The destination terminals have CSI and, in contrast to Chapter 2, are allowed to cooperate and perform joint decoding.

3.1. CONTRIBUTIONS AND RELATION TO PREVIOUS WORK

Previous work of Bölcskei et al. (2006) for the noncoherent-relaying (AF) case demonstrated that for M fixed and $K \rightarrow \infty$, AF relaying turns the fading interference relay network into a fading point-to-point

MIMO link, showing that the use of relays as active scatterers can recover spatial multiplexing gain in poor scattering environments. Our main contributions for the noncoherent-relaying (AF) case are as follows:

- As in the coherent-relaying case, the proof techniques used by Bölcskei et al. (2006) for the noncoherent-relaying (AF) case rely heavily on M being finite. Building on results reported by Silverstein (1995), we compute the $M, K \rightarrow \infty$ (with $K/M \rightarrow \beta$ fixed) per source-destination terminal pair capacity using tools from large-random-matrix theory (Tulino and Verdú, 2004; Müller, 2003; Anderson et al., 2009). The limiting eigenvalue density function of the effective MIMO channel matrix between the source and destination terminals is characterized in terms of its Stieltjes transform as the unique solution of a fixed-point equation, which can be transformed into a fourth-order equation. Upon solving this fourth-order equation and applying the inverse Stieltjes transform, the remaining steps to computing the limiting eigenvalue density function, and based on that the asymptotic network capacity, need to be carried out numerically. We show that this can be accomplished in a straightforward fashion and provide a corresponding algorithm.
- We show that for $\beta \rightarrow \infty$, the fading AF relay network is turned into a fading point-to-point MIMO link (in a sense to be made precise in Section 3.5), thus establishing the large- M, K analog of the result found previously by Bölcskei et al. (2006) for the finite- M and $K \rightarrow \infty$ case.

3.2. THE AMPLIFY-AND-FORWARD PROTOCOL

Throughout this chapter, we use the basic setup introduced in Section 2.2. In addition, we assume that $E_{k,m} = P_{m,k} = 1$ for all $m \in [1 : M]$, $k \in [1 : K]$. This assumption is crucial as the technique used to derive the main result in this chapter does not seem to be applicable for general $E_{k,m}$ and $P_{m,k}$. On the other hand, the results

in this chapter do not require \mathbf{H} and \mathbf{F} to have Gaussian entries. Upon reception of r_k , the k th relay terminal simply scales the received signal to obtain $t_k = (d/\sqrt{K})r_k$. Choosing $d = \sqrt{P_{\text{rel}}/(1 + \sigma^2)}$ ensures that the per-relay power constraint $\mathbb{E}[|t_k|^2] \leq P_{\text{rel}}/K$ and hence the total power constraint $\mathbb{E}[\|\mathbf{t}\|^2] \leq P_{\text{rel}}$ is met.

With these assumptions, inserting (2.1) into (2.2), we get the following IO relation

$$\mathbf{y} = \frac{d}{\sqrt{K}}\mathbf{F}\mathbf{H}\mathbf{s} + \frac{d}{\sqrt{K}}\mathbf{F}\mathbf{z} + \mathbf{w}. \quad (3.1)$$

In the remainder of this chapter, we assume that the *jointly decoding destination terminals* have access to the realizations of \mathbf{H} and \mathbf{F} . In fact, as the analysis below shows, knowledge of $\mathbf{F}\mathbf{H}$ and \mathbf{F} is sufficient.

3.3. CAPACITY OF THE AMPLIFY-AND-FORWARD PROTOCOL

Based on the IO relation (3.1), we shall next study the behavior of $I(\mathbf{y}; \mathbf{s} | \mathbf{F}\mathbf{H}, \mathbf{F})$ when $M, K \rightarrow \infty$ with $K/M \rightarrow \beta$. We start by noting that

$$I(\mathbf{y}; \mathbf{s} | \mathbf{F}\mathbf{H}, \mathbf{F}) = \log \det \left(\mathbf{I} + \frac{d^2}{\sigma^2 MK} \mathbf{H}^H \mathbf{F}^H \left(\frac{d^2}{K} \mathbf{F}\mathbf{F}^H + \mathbf{I} \right)^{-1} \mathbf{F}\mathbf{H} \right).$$

Since the destination terminals perform joint decoding, the ergodic capacity per source-destination terminal pair is given by

$$C_{\text{AF}} = \frac{1}{2} \mathbb{E} \left[\frac{1}{M} \sum_{k=1}^K \log \left(1 + \frac{1}{\sigma^2} \lambda_k \left(\frac{1}{M} \mathbf{H}\mathbf{H}^H \mathbf{T} \right) \right) \right] \quad (3.2)$$

where

$$\mathbf{T} \triangleq \frac{d^2}{K} \mathbf{F}^H \left(\mathbf{I} + \frac{d^2}{K} \mathbf{F}\mathbf{F}^H \right)^{-1} \mathbf{F}$$

and the factor 1/2 in (3.2) results from the fact that data is transmitted over two time slots.

3.4. ASYMPTOTIC CAPACITY BEHAVIOR

To compute C_{AF} in the $M, K \rightarrow \infty$ limit with $K/M \rightarrow \beta$, we start by analyzing the corresponding asymptotic behavior of $\lambda_k((1/M)\mathbf{H}\mathbf{H}^H\mathbf{T})$. To this end, we define the empirical spectral distribution (ESD) of a matrix (random or deterministic).

Definition 3.1. Let $\mathbf{A} \in \mathbb{C}^{N \times N}$ be a Hermitian matrix. The ESD of \mathbf{A} is defined as

$$F_{\mathbf{A}}^N(x) \triangleq \frac{1}{N} \sum_{n=1}^N I[\lambda_n(\mathbf{A}) \leq x].$$

For random \mathbf{A} , the quantity $F_{\mathbf{A}}^N(x)$ is random as well, i.e., it is a RV for each x . In the following, our goal is to prove the convergence (in the sense defined below), when $M, K \rightarrow \infty$ with $K/M \rightarrow \beta$ and $\beta \in (0, \infty)$, of $F_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}^K(x)$ to a deterministic limit and to find the corresponding limiting eigenvalue distribution.

Definition 3.2. We say that the ESD $F_{\mathbf{A}}^N(x)$ of a random Hermitian matrix $\mathbf{A} \in \mathbb{C}^{N \times N}$ converges almost surely (a.s.) to a deterministic limiting function $F_{\mathbf{A}}(x)$, when $N \rightarrow \infty$, if for every $\epsilon > 0$ there exists an $N_0 > 0$ such that for all $N \geq N_0$ a.s.

$$\sup_{x \in \mathbb{R}} |F_{\mathbf{A}}^N(x) - F_{\mathbf{A}}(x)| \leq \epsilon.$$

To prove the convergence of $F_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}^K(x)$ to a deterministic limiting function, we start by analyzing $F_{\mathbf{T}}^K(x)$.

Lemma 3.1. For $M, K \rightarrow \infty$ with $K/M \rightarrow \beta$, the ESD $F_{\mathbf{T}}^K(x)$ converges a.s. to a nonrandom limiting distribution $F_{\mathbf{T}}(x)$ with corresponding density given by¹

$$f_{\mathbf{T}}(x) = \frac{\sqrt{(1 + \gamma_1)(1 + \gamma_2) \left(\frac{\gamma_2}{1 + \gamma_2} - x\right)^+ \left(x - \frac{\gamma_1}{1 + \gamma_1}\right)^+}}{2\pi d^2 x(1 - x)^2} + \left[1 - \frac{1}{\beta}\right]^+ \delta(x) \quad (3.3)$$

¹Note that (3.3) implies that $f_{\mathbf{T}}(x)$ is compactly supported in the interval $[\gamma_1/(1 + \gamma_1), \gamma_2/(1 + \gamma_2)]$.

where $\gamma_1 \triangleq d^2(1 - 1/\sqrt{\beta})^2$ and $\gamma_2 \triangleq d^2(1 + 1/\sqrt{\beta})^2$.

Proof. We start with the singular value decomposition

$$\frac{d}{\sqrt{K}}\mathbf{F} = \mathbf{U}\Sigma\mathbf{V}$$

where the columns of $\mathbf{U} \in \mathbb{C}^{M,M}$ are the eigenvectors of the matrix $(d^2/K)\mathbf{F}\mathbf{F}^H$, the columns of $\mathbf{V}^H \in \mathbb{C}^{K,K}$ are the eigenvectors of $(d^2/K)\mathbf{F}^H\mathbf{F}$, and the matrix $\Sigma \in \mathbb{R}^{M,K}$ contains $R = \min(M, K)$ nonzero entries $\Sigma_{11}, \Sigma_{22}, \dots, \Sigma_{RR}$, which are the positive square roots of the nonzero eigenvalues of the matrix $(d^2/K)\mathbf{F}\mathbf{F}^H$. Defining $\Lambda \triangleq \Sigma\Sigma^H \in \mathbb{R}^{M,M}$, we have

$$\mathbf{T} = \mathbf{V}^H\Sigma^H(\mathbf{I} + \Lambda)^{-1}\Sigma\mathbf{V}.$$

By inspection, it follows that

$$F_{\Sigma^H(\mathbf{I}+\Lambda)^{-1}\Sigma}^K(x) = \frac{M}{K}F_{\Lambda}^M\left(\frac{x}{1-x}\right) + \left(1 - \frac{M}{K}\right)u(x). \quad (3.4)$$

As $F_{\Lambda}^M(x) = F_{(d^2/K)\mathbf{F}\mathbf{F}^H}^M(x)$, by the Marčenko-Pastur law (see Theorem D.3 in Appendix D), we conclude that $F_{\Lambda}^M(x)$ converges a.s. to a limiting nonrandom distribution $F_{\Lambda}(x)$ with corresponding density

$$f_{\Lambda}(x) = \frac{\beta}{2\pi x d^2} \sqrt{(\gamma_2 - x)^+ (x - \gamma_1)^+} + [1 - \beta]^+ \delta(x). \quad (3.5)$$

From (3.4) we can, therefore, conclude that $F_{\Sigma^H(\mathbf{I}+\Lambda)^{-1}\Sigma}^K(x)$ converges a.s. to a nonrandom limit given by

$$F_{\Sigma^H(\mathbf{I}+\Lambda)^{-1}\Sigma}(x) = \frac{1}{\beta}F_{\Lambda}\left(\frac{x}{1-x}\right) + \left(1 - \frac{1}{\beta}\right)u(x). \quad (3.6)$$

Taking the derivative w.r.t. x on both sides of (3.6), the density corresponding to $F_{\Sigma^H(\mathbf{I}+\Lambda)^{-1}\Sigma}(x)$ is obtained as

$$f_{\Sigma^H(\mathbf{I}+\Lambda)^{-1}\Sigma}(x) = \frac{1}{\beta}f_{\Lambda}\left(\frac{x}{1-x}\right) \frac{1}{(1-x)^2} + \left(1 - \frac{1}{\beta}\right)\delta(x). \quad (3.7)$$

We obtain the final result in (3.3) now by noting that $f_{\mathbf{T}}(x) = f_{\Sigma^{\mathbf{H}}(\mathbf{I}+\Lambda)^{-1}\Sigma}(x)$ because of the unitarity of \mathbf{V} and by inserting (3.5) into (3.7) and carrying out straightforward algebraic manipulations. \square

Based on Lemma 3.1, we can now apply Theorem D.2 (Appendix D) to conclude that $F_{(1/M)\mathbf{H}\mathbf{H}^{\mathbf{H}}\mathbf{T}}^K(x)$ converges a.s. to a deterministic function $F_{(1/M)\mathbf{H}\mathbf{H}^{\mathbf{H}}\mathbf{T}}(x)$ as $M, K \rightarrow \infty$ with $K/M \rightarrow \beta$. The corresponding limiting density $f_{(1/M)\mathbf{H}\mathbf{H}^{\mathbf{H}}\mathbf{T}}(x)$ is obtained through the application of the Stieltjes inversion formula (D.1) to the solution of the fixed-point equation

$$G(z) = \underbrace{\int_{-\infty}^{\infty} \frac{f_{\mathbf{T}}(x)dx}{x(1 - \beta - \beta zG(z)) - z}}_I, \quad z \in \mathbb{C}^+ \quad (3.8)$$

in the set

$$\{G(z) \in \mathbb{C} \mid -(1 - \beta)/z + \beta G(z) \in \mathbb{C}^+\}, \quad z \in \mathbb{C}^+ \quad (3.9)$$

where we used the symbol $G(z)$ to denote the Stieltjes transform $G_{(1/M)\mathbf{H}\mathbf{H}^{\mathbf{H}}\mathbf{T}}(z)$. In the following, for brevity, we write G instead of $G(z)$. To solve (3.8), we first compute the integral I on the right-hand side of (3.8). We substitute $f_{\mathbf{T}}(x)$ from (3.3) into (3.8) and define

$$\eta_1 \triangleq \frac{\gamma_1}{1 + \gamma_1}, \quad \eta_2 \triangleq \frac{\gamma_2}{1 + \gamma_2}, \quad \rho \triangleq \frac{\sqrt{(1 + \gamma_1)(1 + \gamma_2)}}{2\pi d^2}$$

to obtain

$$I = -\frac{1}{z} \left[1 - \frac{1}{\beta}\right]^+ + \frac{1}{z} \underbrace{\int_{\eta_1}^{\eta_2} \frac{\rho \sqrt{(\eta_2 - x)(x - \eta_1)} dx}{x(1 - x)^2 \left(x \left(\frac{1 - \beta}{z} - \beta G\right) - 1\right)}}_{\hat{I}}. \quad (3.10)$$

The integral \hat{I} is computed in Appendix E. Employing the notation introduced in Appendix E, we can finally write the fixed point

equation (3.8) as

$$Gz = - \left[1 - \frac{1}{\beta} \right]^+ + \chi \left(A_1 \hat{I}_1 + A_2 \hat{I}_2 + A_3 \hat{I}_3 + A_4 \hat{I}_4 \right). \quad (3.11)$$

It is tedious, but straightforward, to show that for every $\beta > 0$

$$- \left[1 - \frac{1}{\beta} \right]^+ + \chi A_1 \hat{I}_1 = - \frac{\beta - 1}{2\beta}$$

so that (3.11) can be written as

$$Gz + \frac{\beta - 1}{2\beta} - \chi A_2 \hat{I}_2 - \chi A_3 \hat{I}_3 = \chi A_4 \hat{I}_4. \quad (3.12)$$

Next, multiplying (3.12) by $2d^2\beta(G\beta z + z + \beta - 1)^2$, squaring both sides, introducing the auxiliary variable

$$\hat{G} \triangleq - \frac{1 - \beta}{z} + \beta G$$

we obtain after straightforward, but tedious, manipulations that \hat{G} must satisfy the following quartic equation

$$\hat{G}^4 + a_3 \hat{G}^3 + a_2 \hat{G}^2 + a_1 \hat{G} + a_0 = 0 \quad (3.13)$$

with the coefficients

$$\begin{aligned} a_3 &= \frac{1}{z}(2z - \beta + 1) & a_2 &= \frac{1}{z} \left(z - \beta + 3 - \frac{\beta}{d^2} \right) \\ a_1 &= \frac{1}{z^2} \left(2z - \beta + 1 - \frac{\beta}{d^2} \right) & a_0 &= \frac{1}{z^2}. \end{aligned}$$

The quartic equation (3.13) can be solved analytically. The resulting expressions are, however, very lengthy, do not lead to interesting insights, and will therefore be omitted. It is important to note, however, that (3.13) has two pairs of complex conjugate roots. The solutions of (3.13) will henceforth be denoted as $\hat{G}_1, \hat{G}_1^*, \hat{G}_2$, and \hat{G}_2^* . We recall that our goal is to find the unique solution G of the fixed

point equation (3.8) such that $\hat{G} = -(1 - \beta)/z + \beta G \in \mathbb{C}^+$ for all $z \in \mathbb{C}^+$. Therefore, in each point $z \in \mathbb{C}^+$ we can immediately eliminate the two solutions (out of the four) that have a negative imaginary part. In practice, this can be done conveniently by constructing the functions $\hat{G}'_1 \triangleq \Re \hat{G}_1 + i|\Im \hat{G}_1|$ and $\hat{G}'_2 \triangleq \Re \hat{G}_2 + i|\Im \hat{G}_2|$, which can be computed analytically, satisfy (3.13), and are in \mathbb{C}^+ for every $z \in \mathbb{C}^+$. Next, note that (3.12) has a unique solution in the set (3.9), which is also the unique solution of (3.8). We can obtain this solution $G(z)$, $z \in \mathbb{C}^+$, by substituting

$$\begin{aligned} G_1 &= (1/\beta)(\hat{G}'_1 - (\beta - 1)/z) \quad \text{and} \\ G_2 &= (1/\beta)(\hat{G}'_2 - (\beta - 1)/z) \end{aligned}$$

into (3.12) and checking which of the two satisfies the equation. Unfortunately, it seems that this verification cannot be formalized in the sense of identifying the unique solution of (3.12) in analytic form. The primary reason for this is that to check *algebraically* if G_1 and G_2 satisfy (3.12), we have to perform a noninvertible transformation (squaring) of (3.12), which doubles the number of solutions of this equation, and results in G_1 and G_2 both satisfying the resulting formula. The second reason is that depending on the values of the parameters $\beta > 0, d > 0$, the correct solution is either G_1 or G_2 , and the dependence between G_1, G_2, β , and d has a complicated structure. Starting from the analytical expressions for G_1 and G_2 , we can identify, however, for all fixed $\beta > 0, d > 0$, the density function $f_{(1/M)\mathbf{HH}^{\mathbf{H}}\mathbf{T}}(x) = (1/\pi) \lim_{y \rightarrow 0^+} \Im[G(x + iy)]$ corresponding to the unique solution of (3.12) [and hence of (3.8)] numerically. This is accomplished as follows. We know that, for every x , $\lim_{y \rightarrow 0^+} \Im[G(x + iy)]$ is either equal to

$$L_1(x) \triangleq \lim_{y \rightarrow 0^+} \Im[G_1(x + iy)]$$

or

$$L_2(x) \triangleq \lim_{y \rightarrow 0^+} \Im[G_2(x + iy)].$$

Even though the functions $L_1(x)$ and $L_2(x)$ can be computed analytically (with the resulting expressions being very lengthy and involved), it seems that for every fixed $x > 0$ the correct choice between the values $L_1(x)$ and $L_2(x)$ can only be made numerically. The following algorithm constitutes one possibility to solve this problem.

Algorithm—Choice of the Limit

Input: $x > 0$

- 1) Choose a small enough $y > 0$
- 2) Substitute $G_1(x + iy)$ and $G_2(x + iy)$ into (3.12)
- 3) *If $G_1(x + iy)$ satisfies (3.12), then*
 return $L_1(x)$
 otherwise
 return $L_2(x)$

As every other numerical procedure, this algorithm includes a heuristic element. The following comments are therefore in order.

- In Step 1 of the algorithm, the choice of y cannot be formalized in the sense of giving an indication of how small it has to be as a function of β and d . On the one hand, y has to be strictly greater than zero, because (3.12) in general holds in \mathbb{C}^+ only and does not need to hold neither for $G_1(x + i0)$ nor for $G_2(x + i0)$. On the other hand, y should be small enough for $G_1(x + iy)$ to be close to $L_1(x)$ and $G_2(x + iy)$ to be close to $L_2(x)$. The correctness of the output of the algorithm is justified by the fact that $G(z)$ is analytic in \mathbb{C}^+ (see Definition D.1 in Appendix D).
- In Step 3, the check whether $G_1(x + iy)$ satisfies (3.12) is performed numerically. Therefore, rounding errors will arise. It turns out, however, that in practice, unless $|L_1(x) - L_2(x)|$ is very small (in this case it does not matter which of the two values we choose), the solution of (3.12) yields a clear indication of whether $G_1(x + iy)$ or $G_2(x + iy)$ is the correct choice.

3. NETWORKS WITH NONCOHERENT AMPLIFY-AND-FORWARD RELAYING

- To compute the density $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$ using the proposed algorithm, we need to run Steps 1–3 for every x . It will be proved later that $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$ is always compactly supported and bounds for its support will be given in analytic form (as a function of β and d). Since the algorithm consists of very basic arithmetic operations only, it is very fast and can easily be run on a dense grid inside the support region of $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$.

As an example, for $d = 1$ and $\beta = 1/2$, Figure 3.1(a) shows the density $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$ obtained by the algorithm formulated above along with the histogram of the same density obtained through Monte-Carlo simulation. We can see that the two curves match very closely and that our method allows to obtain a much more refined picture of the limiting density. Figure 3.1(b) shows the density $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$ for $\beta = 2, 1, 1/2$ obtained through our algorithm. We can see that the density function is always compactly supported.

The final step in computing the asymptotic capacity of the AF relay network is to take the limit $K, M \rightarrow \infty$ with $K/M \rightarrow \beta$ in (3.2) and to evaluate the resulting integral

$$C_{\text{AF}}^\beta \triangleq \frac{\beta}{2} \int_0^\infty \log\left(1 + \frac{x}{\sigma^2}\right) f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x) dx \quad (3.14)$$

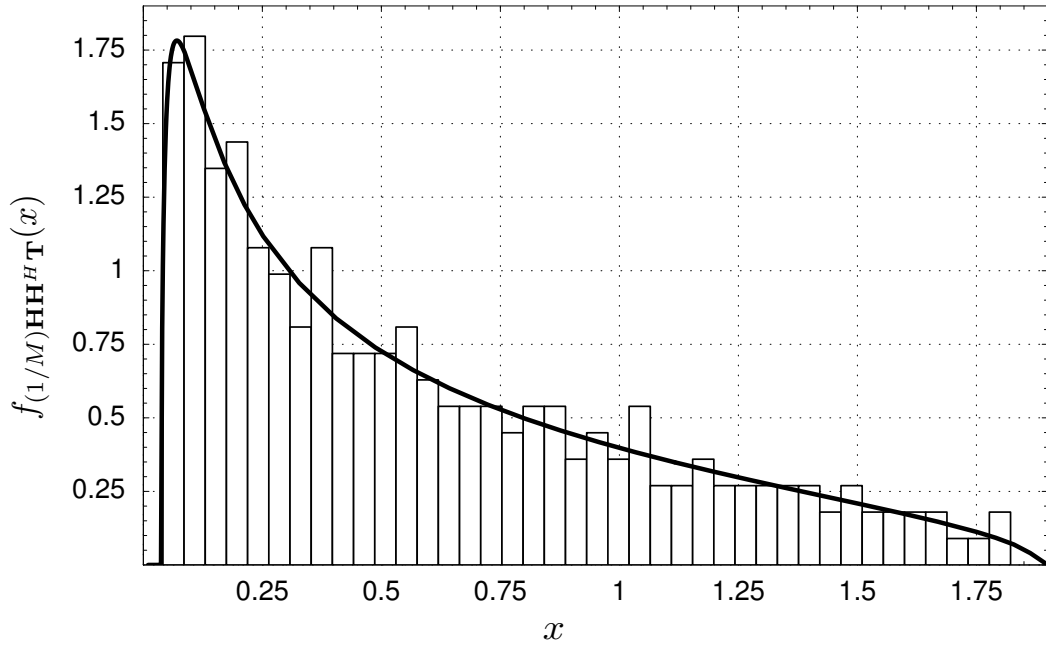
numerically. The evaluation of (3.14) is drastically simplified if we consider that $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$ is compactly supported. The corresponding interval boundaries (or, more specifically, bounds thereon) can be computed analytically as a function of β and d . We start by noting that the second part of Theorem D.3 in Appendix D implies that a.s.

$$\lim_{M \rightarrow \infty} \lambda_{\max}\left(\frac{1}{M}\mathbf{H}\mathbf{H}^H\right) = (1 + \sqrt{\beta})^2.$$

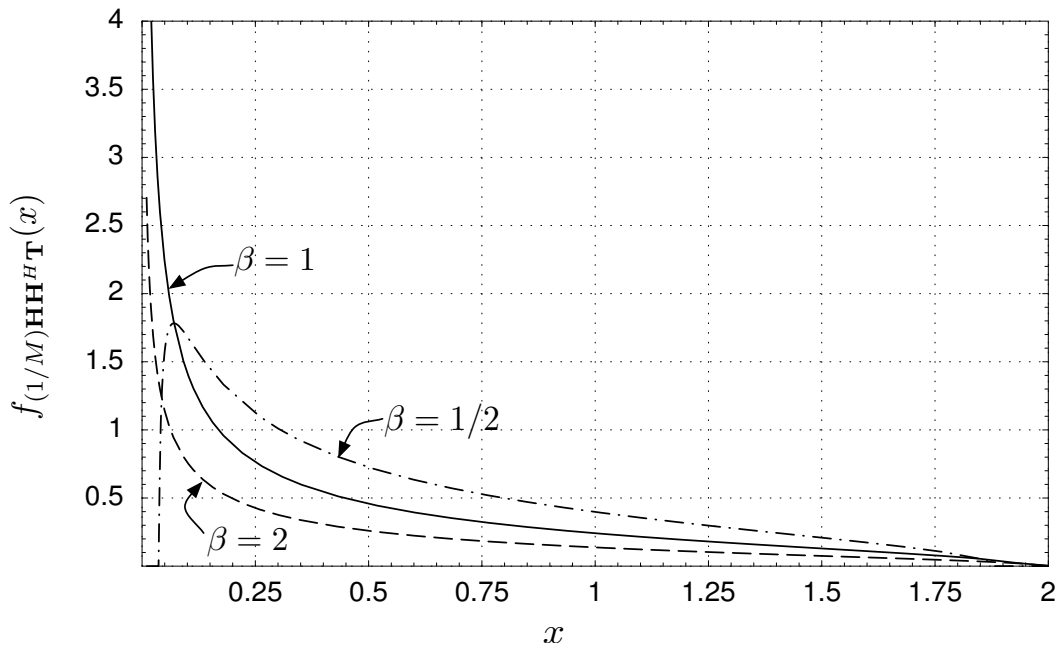
From (3.7) and Theorem D.3, it follows that a.s.

$$\lambda_{\max}(\mathbf{T}) = \frac{d^2(1 + \sqrt{\beta})^2}{\beta + d^2(1 + \sqrt{\beta})^2}.$$

3. NETWORKS WITH NONCOHERENT AMPLIFY-AND-FORWARD RELAYING



(a)



(b)

Fig. 3.1.: Limiting density $f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x)$ (a) for $\beta = 1/2$ and $d = 1$ along with its histogram (Monte-Carlo) and (b) for different values of $\beta = 2, 1, 1/2$ and $d = 1$.

For every realization of \mathbf{H} and \mathbf{T} and all M, K , by the submultiplica-

tivity of the spectral norm, we have

$$\lambda_{\max}\left(\frac{1}{M}\mathbf{H}\mathbf{H}^{\mathbf{H}}\mathbf{T}\right) \leq \lambda_{\max}\left(\frac{1}{M}\mathbf{H}\mathbf{H}^{\mathbf{H}}\right) \lambda_{\max}(\mathbf{T})$$

which implies that for $M, K \rightarrow \infty$ with $K/M \rightarrow \beta$ a.s.

$$\lambda_{\max}\left(\frac{1}{M}\mathbf{H}\mathbf{H}^{\mathbf{H}}\mathbf{T}\right) \leq \frac{d^2(1 + \sqrt{\beta})^4}{\beta + d^2(1 + \sqrt{\beta})^2} \triangleq x_{\max}.$$

We can thus conclude that $f_{(1/M)\mathbf{H}\mathbf{H}^{\mathbf{H}}\mathbf{T}}(x)$ is compactly supported on the interval² $[0, x_{\max}]$. Consequently, the integral in (3.14) becomes

$$C_{\text{AF}}^{\beta} = \frac{\beta}{2} \int_0^{x_{\max}} \log\left(1 + \frac{x}{\sigma^2}\right) f_{(1/M)\mathbf{H}\mathbf{H}^{\mathbf{H}}\mathbf{T}}(x) dx$$

which we can compute numerically, using any standard method for numerical integration and employing the algorithm described above to evaluate $f_{(1/M)\mathbf{H}\mathbf{H}^{\mathbf{H}}\mathbf{T}}(x)$ at the required grid points. Using this procedure, we computed C_{AF}^{β} as a function of β for $d = 1$ with the result depicted in Figure 3.2. We can see that for $\beta < 1$ (i.e., $K < M$), C_{AF}^{β} increases very quickly with β , which is because the corresponding effective MIMO channel matrix builds up rank and hence spatial multiplexing gain. For $\beta > 1$ (i.e., $K > M$), when the effective MIMO channel matrix is already full rank with high probability, the curve flattens out and for $\beta \rightarrow \infty$, the capacity C_{AF}^{β} seems to converge to a finite value. In the next section, we prove that C_{AF}^{β} indeed converges to a finite limit as $\beta \rightarrow \infty$. This result has an interesting interpretation as it allows to relate the AF relay network to a point-to-point MIMO channel.

3.5. CONVERGENCE TO POINT-TO-POINT MIMO CHANNEL

Bölcskei et al. (2006) have shown that for finite M , as $K \rightarrow \infty$, the two-hop AF relay network capacity converges to half the capacity

²The actual supporting interval of $f_{(1/M)\mathbf{H}\mathbf{H}^{\mathbf{H}}\mathbf{T}}(x)$ may, in fact, be smaller.

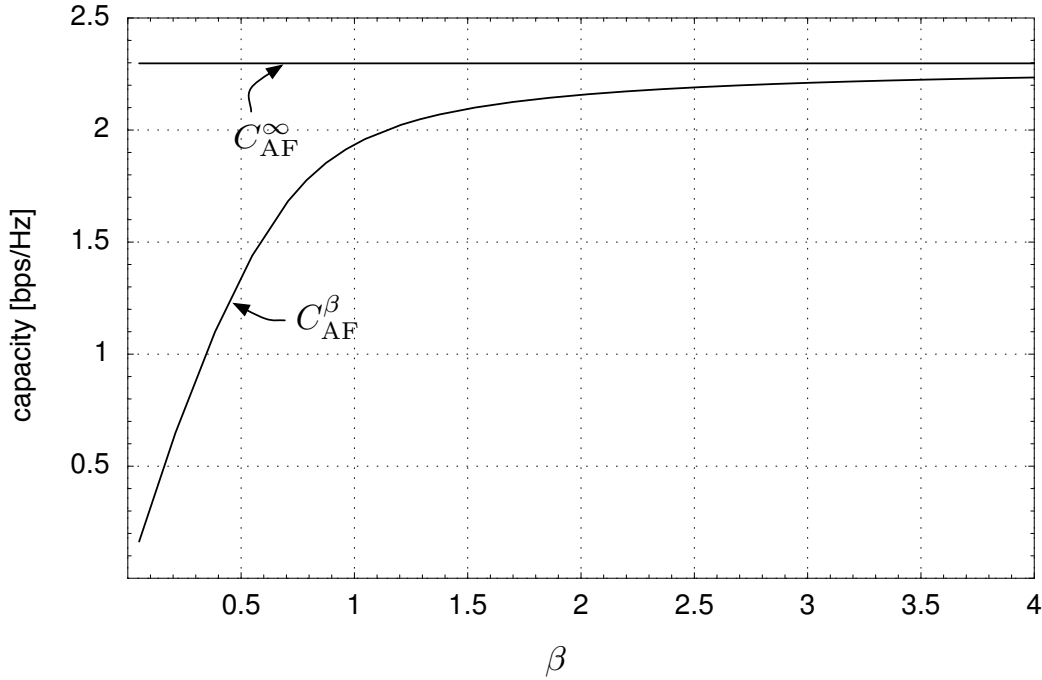


Fig. 3.2.: Capacity C_{AF}^β as a function of β for $d = 1$ and $\sigma^2 = 0.01$.

of a point-to-point MIMO link; the factor $1/2$ penalty comes from the fact that communication takes place over two time slots. In the following, we demonstrate that the result of Bölcskei et al. (2006) can be generalized to the $M, K \rightarrow \infty$ case. More specifically, we show that for $\beta \rightarrow \infty$ the asymptotic ($M, K \rightarrow \infty$) capacity of the two-hop AF relay network is equal to half the asymptotic ($M \rightarrow \infty$) capacity of a point-to-point MIMO channel with M transmit and M receive antennas. We start by dividing (3.13) by β and taking the limit³ $\beta \rightarrow \infty$, which yields the quadratic equation

$$z\hat{G}^2 + z\left(1 + \frac{1}{d^2}\right)\hat{G} + \left(1 + \frac{1}{d^2}\right) = 0. \quad (3.15)$$

The two solutions of (3.15) are given by

$$\hat{G}_{1,2}(z) = \frac{-z\left(1 + \frac{1}{d^2}\right) \pm \sqrt{z^2\left(1 + \frac{1}{d^2}\right)^2 - 4z\left(1 + \frac{1}{d^2}\right)}}{2z}. \quad (3.16)$$

³It is important that first we take the limit $M, K \rightarrow \infty$ with $K/M \rightarrow \beta$ and afterwards let $\beta \rightarrow \infty$.

Applying the Stieltjes inversion formula (D.1) to (3.16) and choosing the solution that yields a positive density function, we obtain

$$\begin{aligned}
 \beta f_{(1/M)\mathbf{H}\mathbf{H}^H\mathbf{T}}(x) &= \frac{1}{\pi} \lim_{y \rightarrow 0^+} \Im [\beta G(x + iy)] \\
 &= \frac{1}{\pi} \lim_{y \rightarrow 0^+} \Im [\hat{G}(x + iy)] \\
 &= \frac{1}{2\pi x} \sqrt{\left[4x \left(1 + \frac{1}{d^2} \right) - x^2 \left(1 + \frac{1}{d^2} \right)^2 \right]^+}.
 \end{aligned} \tag{3.17}$$

Inserting (3.17) into (3.14) and changing the integration variable according to $u \triangleq x(1 + 1/d^2)$, we find that $C_{\text{AF}}^\beta \xrightarrow{\beta \rightarrow \infty} C_{\text{AF}}^\infty$, where

$$C_{\text{AF}}^\infty \triangleq \frac{1}{4\pi} \int_0^4 \sqrt{\frac{4}{u} - 1} \log \left(1 + \frac{d^2}{(d^2 + 1)\sigma^2} u \right) du. \tag{3.18}$$

Comparing (3.18) with (Telatar, 1999, Equation (13)), it follows that for $\beta \rightarrow \infty$ the asymptotic $M, K \rightarrow \infty$ with $K/M \rightarrow \beta$ per source-destination terminal pair capacity in the two-hop AF relay network is equal to half the asymptotic ($M \rightarrow \infty$) per-antenna capacity in a point-to-point MIMO link with M transmit and M receive antennas, provided the SNR in the relay case is defined as $\text{SNR} \triangleq d^2 / ((d^2 + 1)\sigma^2)$. For M and K large, it is easy to verify that this choice corresponds to the SNR at each destination terminal in the AF relay network. In this sense, we can conclude that for $\beta \rightarrow \infty$ the AF relay network “converges” to a point-to-point MIMO link with the same received SNR.

3.6. SUMMARY OF RESULTS

For noncoherent fading interference relay networks with amplify-and-forward relaying and joint decoding at the cooperating destination terminals, we computed the asymptotic (in M and K with $K/M \rightarrow \beta$

fixed) network capacity using tools from large random-matrix theory. To the best of our knowledge, this is the first application of large random-matrix theory to characterize the capacity behavior of large fading networks. Yeh and L ev eque (2007) reported an elegant extension of this approach to the case of multiple layers of relays. We furthermore demonstrated that for $\beta \rightarrow \infty$ the relay network converges to a point-to-point MIMO link. This generalizes the finite- M result obtained by B olcskei et al. (2006) and shows that the use of relays as active scatterers can recover spatial multiplexing gain in poor scattering environments, even if the number of transmit and receive antennas grows large. More importantly, our result shows that linear increase in the number of relays as a function of transmit-receive antennas is sufficient for this to happen.

PART II

NONCOHERENT WIRELESS POINT-TO-POINT CHANNELS

CHAPTER 4

Noncoherent WSSUS Channel

The capacity of fading channels in the absence of CSI both at the transmitter and the receiver¹ is notoriously difficult to analyze even for simple channel models (Abou-Faycal et al., 2001). Most of the results available in the literature pertain to either low or high SNR asymptotics. While in the low-SNR regime the capacity behavior seems robust with respect to the underlying channel model (see, for example, Durisi et al., 2010, for a detailed review of low-SNR capacity results), this is not the case in the high-SNR regime, where capacity is very sensitive to the *fine details* of the channel model, as we are going to argue next.

Consider, as an example, a discrete-time stationary frequency-flat time-selective Rayleigh-fading channel subject to AWGN. Here, the channel law is fully specified by the power spectral density (PSD) $f(\theta)$, $\theta \in [-1/2, 1/2)$, of the fading process and by the noise variance. The high-SNR capacity of this channel depends on the measure μ of the set of harmonics θ where the PSD is nonzero. More specifically, if $\mu < 1$, capacity behaves as $(1 - \mu) \log \text{SNR}$, in the high-SNR regime (Lapidoth, 2005). If $\mu = 1$ and the fading process is regular, i.e., $\int_{-1/2}^{1/2} \log f(\theta) d\theta > -\infty$, then the high-SNR capacity behaves as $\log \log \text{SNR}$ Lapidoth (2005). As a consequence, two channels, one with PSD equal to $1/\Delta$ for $\theta \in [-\Delta/2, \Delta/2]$ and 0 else ($0 < \Delta < 1$),

¹This quantity, under the additional assumption that the transmitter and the receiver are aware of the channel law, is typically called *noncoherent* capacity; in the remainder of this chapter, it will be referred to simply as capacity.

and the other one with PSD equal to $(1 - \epsilon)/\Delta$ for $\theta \in [-\Delta/2, \Delta/2]$ and $\epsilon/(1 - \Delta)$ else ($0 < \epsilon < 1$), will have completely different high-SNR capacity behavior, no matter how small ϵ is. A result like this is clearly unsatisfactory from an engineering viewpoint, as the measure of the support of a PSD cannot be determined through channel measurements. Such a sensitive dependency of the capacity behavior on the fine details of the channel model (by fine details here, we mean details that, in the words of Slepian (Slepian, 1976), have “. . . no direct meaningful counterparts in the real world . . .”), should make one question the validity of the channel model itself.

An engineering-relevant problem is then to determine the SNR value at which capacity starts being sensitive to such fine details. An attempt to resolve this problem was recently made in (Etkin and Tse, 2006), where, for a first-order Gauss-Markov fading process, the SNR beyond which capacity behaves as $\log \log \text{SNR}$ is computed as a function of the innovation rate of the process. The main limitation of this result is that it is based on a very specific channel model and that it is difficult to link the innovation rate to physical channel parameters.

In this chapter, we attempt to address the problem in more generality. Rather than focusing on a specific discretized channel model, we start from the general class of continuous-time Rayleigh-fading linear time-variant (LTV) channels that satisfy the WSS and US assumptions (Bello, 1963) and that are, in addition, *underspread* (Kennedy, 1969). The Rayleigh-fading and the WSSUS assumptions imply that the statistics of the channel are fully characterized by its *scattering function* (Bello, 1963); the underspread assumption is satisfied if the scattering function is highly concentrated around the origin of the Doppler-delay plane. More concretely, we shall say that a WSSUS channel is underspread if its scattering function has only a fraction $\epsilon \ll 1$ of its volume outside a rectangle of area $\Delta_{\mathbb{H}} \ll 1$ (see Definition 4.1 in the next section). Our main result is the following: we provide a lower bound on the capacity of continuous-time WSSUS underspread Rayleigh-fading channels that is explicit in the parameters $\Delta_{\mathbb{H}}$ and ϵ . On the basis of this bound, we show that for all SNR values that

satisfy $\sqrt{\Delta_{\mathbb{H}}} \ll \text{SNR} \ll 1/(\Delta_{\mathbb{H}} + \epsilon)$, the fading-channel capacity is close to the capacity of a nonfading AWGN channel with the same SNR. Hence, the fading-channel capacity grows logarithmically in SNR up to (high) SNR values $\text{SNR} \ll 1/(\Delta_{\mathbb{H}} + \epsilon)$.

A crucial step in the derivation of our capacity lower bound is the discretization of the continuous-time channel IO relation, which is accomplished by transmitting and receiving on an *orthonormal* Weyl-Heisenberg (WH) set (Christensen, 2003, Chapter 8) of time-frequency shifts of a pulse $g(t)$. The resulting signaling scheme can be interpreted as PS-OFDM. This discretization technique is attractive in our setting, because due to the regular time-frequency structure of the WH set, the stationarity property of the continuous-time channel is inherited by the corresponding discretized channel, a fact that is essential for our analysis. Furthermore, by optimizing the parameters of the WH set, one can mitigate the intersymbol interference (ISI) and the intercarrier interference (ICI) that originate from the dispersive nature of the channel.

Durisi et al. (2010) used a similar discretization technique to characterize the capacity of WSSUS underspread fading channels in the low-SNR regime. In contrast to Durisi et al. (2010), in this chapter, we explicitly account for ISI and ICI terms in the discretized IO relation. This is crucial, as unlike in the low-SNR regime, these terms play a fundamental role at high SNR. Finally, as an interesting byproduct of our analysis, we obtain an information-theoretic pulse-design criterion for PS-OFDM systems that operate over WSSUS underspread fading channels.

4.1. SYSTEM MODEL

4.1.1. The Continuous-Time Input-Output Relation

In the following, we briefly summarize the continuous-time WSSUS underspread Rayleigh-fading channel model employed in this chapter. For a more complete description of this model, the interested reader is

referred to (Durisi et al., 2010). The IO relation of a continuous-time stochastic LTV channel \mathbb{H} can be written as

$$\begin{aligned} y(t) &= \underbrace{(\mathbb{H} s)(t)}_{\triangleq \hat{y}(t)} + w(t) \\ &= \int_{-\infty}^{\infty} h_{\mathbb{H}}(t, \tau) s(t - \tau) d\tau + w(t) \end{aligned} \quad (4.1)$$

where $y(t)$ is the received signal and $\hat{y}(t)$ is the received signal in the absence of additive noise. As in (Wyner, 1966, Model 2), the stochastic transmit signal $s(t)$

- i) satisfies the average-power constraint

$$(1/D) \mathbb{E} [\|s(t)\|^2] \leq P; \quad (4.2)$$

- ii) is strictly limited to a bandwidth of B Hz, i.e.,

$$S(f) = 0, \quad \text{for } |f| > B/2 \quad (4.3)$$

where $S(f) = \mathbb{F}[s(t)]$;

- iii) is approximately limited to a duration of D sec. More precisely, we shall assume that

$$\mathbb{E} [\|\mathbb{D}_D s(t)\|^2] / \mathbb{E} [\|s(t)\|^2] > 1 - \eta \quad (4.4)$$

where \mathbb{D}_D is the time-limiting operator (defined in Section H.3), and η is the average fraction of the energy of $s(t)$ lying outside $[-D/2, D/2]$.

The constraints (4.3) and (4.4) on the input signal capture the fact that the signals of interest in wireless communication are essentially time- and bandwidth-limited. Time limitation is important in an information-theoretic context, because it makes it possible to define the transmission rate at the encoder in a physically meaningful way (Wyner, 1966). Note that the strict band limitation (4.3) implies

that $s(t)$ can be limited in time in an approximate sense (Slepian, 1976) only, a consideration that justifies the constraint (4.4). Throughout the chapter, the parameter η is assumed to be a small positive constant.

The signal $w(t)$ is a zero-mean unit-variance proper AWGN process. As we assumed unit-variance noise, the SNR is given by $\text{SNR} = P/B$. Finally, the channel impulse response $h_{\mathbb{H}}(t, \tau)$ is a zero-mean jointly proper Gaussian (JPG) process in t and τ that satisfies the WSSUS assumption

$$\mathbb{E}[h_{\mathbb{H}}(t, \tau)h_{\mathbb{H}}^*(t', \tau')] = R_{\mathbb{H}}(t - t', \tau)\delta(\tau - \tau') \quad (4.5)$$

and is independent of $w(t)$ and $s(t)$. A more detailed review of the WSSUS channel model can be found for example in (Durisi et al., 2010). As a consequence of the JPG and the WSSUS assumption, the *time-delay correlation function* $R_{\mathbb{H}}(t, \tau)$ fully characterizes the channel statistics.

Often, it is convenient to describe \mathbb{H} in domains other than the time-delay domain. The *delay-Doppler spreading function* $S_{\mathbb{H}}(\nu, \tau) = \mathbb{F}_{t \rightarrow \nu}[h_{\mathbb{H}}(t, \tau)]$ can be used for this purpose. If we rewrite the IO relation (4.1) in terms of the spreading function as

$$y(t) = \underbrace{\iint_{\mathbb{R}^2} S_{\mathbb{H}}(\nu, \tau)s(t - \tau)e^{i2\pi\nu t}d\tau d\nu}_{=\hat{y}(t)} + w(t) \quad (4.6)$$

we obtain the following physical interpretation (Durisi et al., 2011): $\hat{y}(t) = (\mathbb{H}s)(t)$ is a weighted superposition of copies of the input signal $s(t)$ that are shifted in time by the delay τ and in frequency by the Doppler shift ν . It is worth noting that, because of the WSSUS assumption, the spreading function $S_{\mathbb{H}}(\nu, \tau)$ is uncorrelated in τ and ν :

$$\mathbb{E}[S_{\mathbb{H}}(\nu, \tau)S_{\mathbb{H}}^*(\nu', \tau')] = C_{\mathbb{H}}(\nu, \tau)\delta(\nu - \nu')\delta(\tau - \tau'). \quad (4.7)$$

The function $C_{\mathbb{H}}(\nu, \tau)$ is usually referred to as channel *scattering function*. In the remainder of the chapter, we let the scattering function

be normalized in volume according to

$$\iint_{\mathbb{R}^2} C_{\mathbb{H}}(\nu, \tau) d\tau d\nu = 1. \quad (4.8)$$

4.1.2. A Robust Definition of Underspread Channels

Qualitatively speaking, WSSUS underspread channels are WSSUS channels with a scattering function that is highly concentrated around the origin of the Doppler-delay plane (Bello, 1963). A mathematically precise definition of the underspread property is available for the case where $C_{\mathbb{H}}(\nu, \tau)$ is *compactly supported* within a rectangle. In this case, the channel is said to be underspread if the support area of $C_{\mathbb{H}}(\nu, \tau)$ is much smaller than 1 (see, for example, Kozek, 1997; Durisi et al., 2010). The compact-support assumption, albeit mathematically convenient, is a fine detail of the channel model in the terminology introduced in the previous section, because it is not possible to determine through channel measurements whether $C_{\mathbb{H}}(\nu, \tau)$ is indeed compactly supported or not. However, the results discussed in the previous section hint at a high sensitivity of capacity to this fine detail. To better understand and quantify this sensitivity, we need to take a more general approach. We replace the compact-support assumption by the following more robust and physically meaningful assumption: $C_{\mathbb{H}}(\nu, \tau)$ has a small fraction of its total volume outside a rectangle of an area that is much smaller than 1. More precisely, we have the following definition.

Definition 4.1. Let $\nu_0, \tau_0 \in \mathbb{R}^+, \epsilon \in [0, 1]$, and let $\mathcal{H}(\tau_0, \nu_0, \epsilon)$ be the set of all Rayleigh-fading WSSUS channels \mathbb{H} with scattering function $C_{\mathbb{H}}(\nu, \tau)$ satisfying

$$\int_{-\nu_0}^{\nu_0} \int_{-\tau_0}^{\tau_0} C_{\mathbb{H}}(\nu, \tau) d\tau d\nu \geq 1 - \epsilon. \quad (4.9)$$

We say that the channels in $\mathcal{H}(\tau_0, \nu_0, \epsilon)$ are *underspread* if $\Delta_{\mathbb{H}} = 4\tau_0\nu_0 \ll 1$ and $\epsilon \ll 1$.

Typical wireless channels are (highly) underspread, with most of the volume of $C_{\mathbb{H}}(\nu, \tau)$ supported over a rectangle of area $\Delta_{\mathbb{H}} \approx 10^{-3}$ for land-mobile channels, and $\Delta_{\mathbb{H}}$ as small as 10^{-7} for certain indoor channels with restricted terminal mobility. Note that $\epsilon = 0$ in Definition 4.1 yields the compact-support underspread definition of Kozek (1997) and Durisi et al. (2010).

4.2. THE INFORMATION CAPACITY

We are interested in this chapter in the characterization of the ultimate limit of reliable communication (i.e., of the capacity) of the continuous-time channel (4.1), under the assumption that both the transmitter and the receiver are not aware of the realization of \mathbb{H} , but are aware of its stochastic law. We will start by defining the information capacity and then discuss its achievability.

To define the information capacity of (4.1), we shall follow the (classic) approach (Gallager, 1968), and represent the complex signals at the input and output of \mathbb{H} in terms of a series expansion with respect to a complete orthonormal set for the underlying vector space. More specifically, let $\mathcal{L}^2(B) \subset \mathcal{L}^2(\mathbb{C})$ be the Hilbert space of $\mathcal{L}^2(\mathbb{C})$ -signals that have bandwidth no larger than B , and let $\{\phi_m(t)\}_{m=0}^{\infty}$ be a complete orthonormal set for $\mathcal{L}^2(B)$. We can write every $s(t) \in \mathcal{L}^2(B)$ as

$$s(t) = \sum_{m=0}^{\infty} \underbrace{\langle s(t), \phi_m(t) \rangle}_{\triangleq s_m} \phi_m(t). \quad (4.10)$$

Even though $s(t)$ has bandwidth no larger than B , the signal $\hat{y}(t) = (\mathbb{H}s)(t)$ does not satisfy in general a strict bandwidth constraint. However, if \mathbb{H} is underspread in the sense of Definition 4.1, most of the energy of $\hat{y}(t)$ will lie (on average) within a bandwidth of $(B + 2\nu_0)$ Hz. We assume that this fact is exploited at the receiver to reduce the impact of thermal noise. More precisely, we assume that $y(t)$ is passed through an ideal low-pass filter of bandwidth

$(B + 2\nu_0)$ Hz, yielding

$$y_f(t) = (\mathbb{B}_{B+2\nu_0}y)(t)$$

where $y_f(t)$ denotes the filtered received signal, and $\mathbb{B}_{B+2\nu_0}$ is the frequency-limiting operator (defined in Section H.3). Because of the low-pass filtering operation at the receiver, $y_f(t)$ is a $\mathcal{L}^2(B + 2\nu_0)$ -function, and, hence, it can be written as

$$y_f(t) = \sum_{m=0}^{\infty} \underbrace{\langle y(t), \phi'_m(t) \rangle}_{\triangleq y_m} \phi'_m(t) \quad (4.11)$$

where $\{\phi'_m(t)\}_{m=0}^{\infty}$ is a complete orthonormal set for $\mathcal{L}^2(B+2\nu_0)$. Note that (4.10) and (4.11) yield a discretization of the continuous-time IO relation (4.1).

Let now $\mathcal{Q}(B, D, \eta, P)$ be the set of probability measures on $s(t)$ that satisfy the average-power constraint (4.2) and the approximate time limitation (4.4). Every probability measure $\mathcal{Q}_s \in \mathcal{Q}(B, D, \eta, P)$ on $s(t)$ induces a probability measure on $\{s_m\}_{m=0}^{\infty}$. Hence, for a given $\mathcal{Q}_s \in \mathcal{Q}(B, D, \eta, P)$ we can define the mutual information between $s(t)$ and $y_f(t)$ as

$$I_D(s(t); y_f(t)) \triangleq \lim_{M \rightarrow \infty} I(\mathbf{s}_0^M; \mathbf{y}_0^M)$$

where $\mathbf{s}_0^M = [s_0 \ s_1 \ \cdots \ s_M]^\top$, and, similarly, $\mathbf{y}_0^M = [y_0 \ y_1 \ \cdots \ y_M]^\top$. The information capacity C_{WSSUS} of the channel (4.1) can now be defined as follows (Gallager, 1968, Eq. (8.1.55))

$$C_{\text{WSSUS}} \triangleq \liminf_{D \rightarrow \infty} \frac{1}{D} \sup_{\mathcal{Q}_s \in \mathcal{Q}(B, D, \eta, P)} I_D(s(t); y_f(t)). \quad (4.12)$$

Note that for the purpose of defining C_{WSSUS} , all pairs of orthonormal sets—one complete for $\mathcal{L}^2(B)$, and the other complete for $\mathcal{L}^2(B + 2\nu_0)$ —work equally well; however, complete orthonormal sets that are *matched* to the channel, i.e., that consist of channel singular functions (Durisi et al., 2010), are better suited to information-theoretic analysis (Gallager, 1968; Wyner, 1966). For example, the

characterization of the capacity of continuous-time bandlimited AWGN channels in (Wyner, 1966) is enabled by a discretization of the IO relation using prolate spheroidal wave functions (Landau and Pollak, 1960; Slepian and Pollak, 1961; Landau and Pollak, 1962). These functions are singular functions of the channel operator, which is in this case the frequency-limiting operator \mathbb{B}_B .

We conclude this section with some comments on the achievability of (4.12). By Fano's inequality (Cover and Thomas, 2006, Theorem 2.10.1) no rate above C_{WSSUS} is achievable (see Cover and Thomas, 2006, for a definition of achievable rate). For the channel of interest in this chapter, however, it is not clear whether the information capacity (4.12) coincides with the supremum of all achievable rates on the channel. Establishing whether this result holds is an interesting open problem, which goes beyond the scope of the thesis. In the remainder of the chapter, we shall refer to information capacity simply as capacity.

4.2.1. Outline of the Information-Theoretic Analysis

An analytic characterization of C_{WSSUS} for the setting of interest in this chapter (neither transmitter nor receiver are aware of the realization of \mathbb{H} but perfectly aware of its stochastic law) is difficult. Hence, our focus in this chapter will be on deriving tight bound on (4.12) for SNR values of practical interest.

For channels that are underspread according to Definition 4.1, a simple (yet tight as we shall see) upper bound on (4.12) can be obtained as follows: first note that

$$I_D(s(t); y_f(t)) \leq I_D(\hat{y}_f(t); y_f(t)) \quad (4.13)$$

where $\hat{y}_f(t) = (\mathbb{B}_{B+2\nu_0}\hat{y})(t)$. The right-hand side of (4.13) can be interpreted as the capacity of a continuous-time bandlimited AWGN channel. As most of the energy of $\hat{y}(t)$ lies in a time interval of $(D + 2\tau_0)$ sec and in a frequency interval of $(B + 2\nu_0)$ Hz, we can

use (Wyner, 1966, Theorem 2) (see Appendix F.1) and obtain

$$I_D(\hat{y}_f(t); y_f(t)) \leq (B + 2\nu_0) \log \left(1 + \frac{(1 - \eta)(1 - \epsilon)P}{B + 2\nu_0} \right) + (\eta + \epsilon - \eta\epsilon)P. \quad (4.14)$$

The right-hand side of (4.14) can be well-approximated by

$$C_{\text{AWGN}} = B \log(1 + \text{SNR}) \quad (4.15)$$

whenever ν_0 , η and ϵ are sufficiently small.

It is now appropriate to provide a preview of the nature of the results we are going to obtain on the basis of the novel underspread definition introduced in Section 4.1.2. We will show that, as long as $\Delta_{\mathbb{H}} \ll 1$ and $\epsilon \ll 1$, the capacity of all channels in $\mathcal{H}(\tau_0, \nu_0, \epsilon)$, independently of whether their scattering function is compactly supported or not, is close to the AWGN capacity C_{AWGN} for all SNR values typically encountered in practical wireless communication systems. To establish this result, we choose a specific transmit and receive scheme (detailed in the next section), which yields a capacity lower bound that is close to the upper bound C_{AWGN} .

4.3. A LOWER BOUND ON CAPACITY

4.3.1. Discretization of the Input-Output Relation

As discussed above, the starting point for an information-theoretic analysis of continuous-time channels is a discretization of the corresponding IO relation. The discretization is performed by means of two sets of orthonormal functions that satisfy the following conditions: i) the first set is complete for the input space of the channel operator, and the second set is complete for its output space; ii) both sets are made of channel singular functions (Wyner, 1966; Gallager, 1968). This approach is, however, not feasible in our setup. Assume that the functions in $\{\phi_m(t)\}_{m=0}^{\infty}$ and $\{\phi'_m(t)\}_{m=0}^{\infty}$ are the right singular functions and the left singular functions of the channel \mathbb{H} , respectively. To make use of the series expansions (4.10) and (4.11), and, hence,

discretize the channel, the set $\{\phi_m(t)\}_{m=0}^{\infty}$ needs to be known at the transmitter and $\{\phi'_m(t)\}_{m=0}^{\infty}$ at the receiver. But as \mathbb{H} is random, its singular functions are, in general, random as well, and consequently not known to transmitter and receiver in the noncoherent setting of interest in this chapter (see Durisi et al., 2011, for a more detailed discussion on this issue).

In contrast, if the singular functions of the random channel \mathbb{H} did not depend on the particular realization of \mathbb{H} , we could diagonalize \mathbb{H} without knowledge of the channel realizations. This is the case, for example, for linear time-invariant (LTI) channels, where (deterministic) complex sinusoids are always eigenfunctions, independently of the realization of the channel impulse response. This observation is crucial for the ensuing analysis. Specifically, our discretization strategy is based on the following fundamental property of underspread LTV channels (see Durisi et al., 2010, and references therein): the singular functions of a random underspread WSSUS channel can be well approximated by deterministic functions that are well localized in time and frequency. More specifically, we discretize the continuous-time IO relation (4.1) by transmitting and receiving on the highly structured WH set $(g, T, F) \triangleq \{g_{k,n}(t) = g(t - kT)e^{i2\pi nFt}\}_{k,n \in \mathbb{Z}}$ of time-frequency shifts of the pulse $g(t)$. We choose $g(t)$, T , and F such that the following properties are satisfied:

- i) $g(t)$ has unit energy, is strictly bandlimited, with bandwidth $F \leq B$, and satisfies $g(t) = \mathcal{O}(1/|t|^{1+\mu})$, $|t| \rightarrow \infty$, for some $\mu > 0$;
- ii) the signals in the WH set (g, T, F) are orthonormal.

An explicit construction of a family of WH sets (g, T, F) for which Properties i) and ii) are satisfied is provided in Section 4.3.9. These two properties allow us to construct a general class of signals that satisfy the constraints (4.2)–(4.4). More precisely, we shall proceed as follows. Fix η in (4.4) and the bandwidth B of $s(t)$. For a given approximate duration D of the transmit signal $s(t)$ [we will later take $D \rightarrow \infty$ according to (4.12)], the interval $[-D/2, D/2]$ is divided into

three parts: one interval $[-\tilde{K}T/2, \tilde{K}T/2]$ where most of the energy of the transmit signal will lie, and two *guard intervals* $[-D/2, -\tilde{K}T/2]$ and $[\tilde{K}T/2, D/2]$ of length $K'T = D/2 - \tilde{K}T$. We will let $\tilde{K} \rightarrow \infty$ as $D \rightarrow \infty$. On the contrary, K' does not depend on D , but depends on $g(t)$ and on η in (4.4) only. For simplicity of notation, we shall assume in the remainder of the chapter that K' is an integer and that \tilde{K} and $\tilde{N} \triangleq B/F$ are odd integers.

Let $g(t)$, T , and F be chosen so that Properties i) and ii) are satisfied and consider transmit signals of the form

$$s(t) = \sum_{k=-K}^K \sum_{n=-N}^N s[k, n] g_{k, n}(t) \quad (4.16)$$

where the data symbols $s[k, n] \in \mathbb{C}$ are chosen such that

$$\sum_{k=-K}^K \sum_{n=-N}^N \mathbb{E}[|s[k, n]|^2] \leq (2K + 1)TP. \quad (4.17)$$

Here and in (4.16), $N = (\tilde{N} - 1)/2$ and $K = (\tilde{K} - 1)/2$. The transmit signal (4.16) satisfies the average-power constraint (4.2) and the bandwidth constraint (4.3) by construction. Furthermore, as shown in Appendix F.2, one can choose K' (independently of D) such that $s(t)$ in (4.16) satisfies the approximate time-limitation constraint (4.4) as well. Because K' does not depend on D , the loss of degrees of freedom caused by the presence of the guard interval $2K'T$ vanishes when one computes the limit $D \rightarrow \infty$ in (4.12).

The received signal $y_f(t)$ is projected onto the signal set $\{g_{k, n}(t)\}$, $k \in [-K : K]$, $n \in [-N : N]$ to obtain

$$\begin{aligned} \underbrace{\langle y, g_{k, n} \rangle}_{\triangleq y[k, n]} &= \underbrace{\langle \mathbb{H} g_{k, n}, g_{k, n} \rangle}_{\triangleq h[k, n]} s[k, n] \\ &+ \sum_{\substack{l=-K \\ (l, m) \neq (k, n)}}^K \sum_{m=-N}^N \underbrace{\langle \mathbb{H} g_{l, m}, g_{k, n} \rangle}_{\triangleq p[l, m, k, n]} s[l, m] + \underbrace{\langle w, g_{k, n} \rangle}_{\triangleq w[k, n]} \end{aligned}$$

4. NONCOHERENT WSSUS CHANNEL

$$= h[k, n]s[k, n] + \sum_{\substack{l=-K \\ (l,m) \neq (k,n)}}^K \sum_{m=-N}^N p[l, m, k, n]s[l, m] + w[k, n] \quad (4.18)$$

for each *time-frequency* slot (k, n) . We refer to the channel with IO relation (4.18) as the discretized channel *induced* by the WH set (g, T, F) . Note that we do not require that the set (g, T, F) is complete for $\mathcal{L}^2(\mathbb{C})$. Hence, not all signals satisfying (4.2)–(4.4) can be represented in the form (4.16). As a consequence, the capacity of the discretized channel (4.18), defined in the next section, is a lower bound on the capacity of the underlying continuous-time channel (4.1) defined in (4.12).

The regular time-frequency structure of the WH-set implies that the channel gains $h[k, n]$ in (4.18) inherit the two-dimensional stationarity property of the underlying continuous-time channel (see Section 4.3.3), a fact that is crucial for the ensuing analysis.

The presence of the second term in (4.18), which corresponds to ISI and ICI, makes the derivation of capacity bounds involved. Fortunately, to establish our main result (see Section 4.3.6), it will be sufficient to treat the interference term as noise. The corresponding results are of practical interest, as receiver algorithms that take the structure of interference explicitly into account are, in general, computationally expensive. The variance of the interference term in (4.18) depends on the time-frequency localization properties of $g(t)$ as will be shown in Section 4.3.3. We finally note that the orthonormality of (g, T, F) implies that $w[k, n]$ in (4.18) is i.i.d. $\mathcal{CN}(0, 1)$.

4.3.2. Orthonormality, Completeness, and Localization

Orthonormality, completeness, and time-frequency localization are desirable properties of the WH set (g, T, F) . It is, therefore, sensible to ask whether complete orthonormal WH sets generated by a $g(t)$ with prescribed time-frequency localization exist. The answer is as follows:

- i) A necessary condition for the set (g, T, F) to be orthonormal is

$TF \geq 1$ (Gröchenig, 2001, Corollary 7.5.1, Corollary 7.3.2).

- ii) For $TF = 1$, it is possible to find orthonormal sets (g, T, F) that are complete in $\mathcal{L}^2(\mathbb{C})$ (Christensen, 2003, Theorem 8.3.1). These sets, however, do not exhibit good time-frequency localization, as a consequence of the Balian-Low Theorem (Christensen, 2003, Theorem 4.1.1), which states that if (g, T, F) is orthonormal and complete in $\mathcal{L}^2(\mathbb{C})$, then

$$\left(\int_{-\infty}^{\infty} |tg(t)|^2 dt \right) \left(\int_{-\infty}^{\infty} |fG(f)|^2 df \right) = \infty$$

where $G(f) = \mathbb{F}[g(t)]$.

- iii) For $TF > 1$, it is possible to have orthonormality and good time-frequency localization concurrently, but the resulting set (g, T, F) is necessarily incomplete in $\mathcal{L}^2(\mathbb{C})$. Lack of completeness entails a loss of degrees of freedom.
- iv) For $TF < 1$, it is possible to construct WH sets generated by a well-localized $g(t)$, which are also (over)complete in $\mathcal{L}^2(\mathbb{C})$. However, as a consequence of overcompleteness, the resulting input signal (4.16) cannot be recovered uniquely at the receiver, even in the absence of noise.

Our choice will be to privilege localization and orthonormality over completeness. The information-theoretic results in Section 4.4 will show that this choice is sound.

4.3.3. Statistical Properties

We derive in this section two useful properties of the statistics of $h[k, n]$ and $p[l, m, k, n]$. We will need these properties in the proof of the lower bound to be presented in Theorem 4.1 (see Section 4.3.6). The first property concerns the autocorrelation of $h[k, n]$. Let the

4. NONCOHERENT WSSUS CHANNEL

cross-ambiguity function of two signals $g(t)$ and $f(t)$ be defined as

$$A_{g,f}(\nu, \tau) \triangleq \int_{-\infty}^{\infty} g(t)f^*(t - \tau)e^{-i2\pi\nu t} dt \quad (4.19)$$

and let $A_g(\nu, \tau) \triangleq A_{g,g}(\nu, \tau)$. Some basic results about these two functions that are useful for our analysis are reviewed in Appendix F.3. We are now ready to compute the autocorrelation of $h[k, n]$, which is given by

$$\begin{aligned} \mathbb{E}[h[k, n]h^*[l, m]] &= \mathbb{E}[\langle \mathbb{H} g_{k,n}, g_{k,n} \rangle \langle \mathbb{H} g_{l,m}, g_{l,m} \rangle^*] \\ &\stackrel{(a)}{=} \iint_{\mathbb{R}^2} C_{\mathbb{H}}(\nu, \tau) A_{g_{k,n}}^*(\nu, \tau) A_{g_{l,m}}(\nu, \tau) d\tau d\nu \\ &\stackrel{(b)}{=} \iint_{\mathbb{R}^2} C_{\mathbb{H}}(\nu, \tau) |A_g(\nu, \tau)|^2 e^{i2\pi[(k-l)T\nu - (n-m)F\tau]} d\tau d\nu \\ &\triangleq r[k - l, n - m] \end{aligned} \quad (4.20)$$

where (a) follows from Property 4 in Appendix F.3 and because \mathbb{H} is WSSUS [see (4.7)], while (b) follows from Property 3 in Appendix F.3 [see in particular (F.3)]. As a consequence of (4.20), we have that $\{h[k, n]\}$ is stationary both in discrete time k and in discrete frequency n . Because of stationarity, we can associate to the discrete channel process $\{h[k, n]\}$ a two-dimensional power spectral density

$$c(\theta, \varphi) \triangleq \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} r[k, n] e^{-i2\pi(k\theta - n\varphi)}, \quad |\theta|, |\varphi| \leq 1/2. \quad (4.21)$$

The Poisson summation formula together with (4.20) implies that

$$\begin{aligned} c(\theta, \varphi) &= \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} e^{-i2\pi(k\theta - n\varphi)} \\ &\quad \times \iint_{\mathbb{R}^2} C_{\mathbb{H}}(\nu, \tau) |A_g(\nu, \tau)|^2 e^{i2\pi(kT\nu - nF\tau)} d\tau d\nu \end{aligned}$$

$$= \frac{1}{TF} \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_{\mathbb{H}}\left(\frac{\theta - k}{T}, \frac{\varphi - n}{F}\right) \left| A_g\left(\frac{\theta - k}{T}, \frac{\varphi - n}{F}\right) \right|^2. \quad (4.22)$$

Another property we shall often use is that

$$r[0, 0] = \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} c(\theta, \varphi) d\theta d\varphi = \iint_{\mathbb{R}^2} C_{\mathbb{H}}(\nu, \tau) |A_g(\nu, \tau)|^2 d\tau d\nu \leq 1 \quad (4.23)$$

where the last step follows from Property 1 in Appendix F.3, from the assumption that $g(t)$ has unit energy and from the normalization (4.8).

We next characterize the variance of $p[l, m, k, n]$. As $p[l, m, k, n]$ has zero mean,

$$\begin{aligned} & \mathbb{E} [|p[l, m, k, n]|^2] \\ &= \mathbb{E} [|\langle \mathbb{H} g_{l,m}, g_{k,n} \rangle|^2] \\ &= \iint_{\mathbb{R}^2} C_{\mathbb{H}}(\nu, \tau) |A_{g_{k,n}, g_{l,m}}(\nu, \tau)|^2 d\tau d\nu \\ &= \iint_{\mathbb{R}^2} C_{\mathbb{H}}(\nu, \tau) |A_g(\nu + (m - n)F, \tau + (l - k)T)|^2 d\tau d\nu \\ &\triangleq \sigma_p^2[k - l, n - m] \end{aligned} \quad (4.24)$$

where, again, we used first Property 4 in Appendix F.3 together with the WSSUS property of \mathbb{H} , and then Property 3 in Appendix F.3. As a consequence of (4.24), the variance of $p[l, m, k, n]$ depends only on the shift difference in the time-frequency plane between the two functions $g_{l,m}(t)$ and $g_{k,n}(t)$. Also this property will turn out to be crucial in the ensuing analysis.

4.3.4. Input-Output Relation in Vector-Matrix Form

For each $k \in [-K : K]$, we arrange the data symbols $s[k, n]$, the received signal samples $y[k, n]$, the channel coefficients $h[k, n]$, and

the noise samples $w[k, n]$ in corresponding \tilde{N} -dimensional vectors as follows:

$$\begin{aligned}\mathbf{s}[k] &\triangleq [s[k, -N] \ s[k, -N + 1] \ \cdots \ s[k, N]]^\top \\ \mathbf{y}[k] &\triangleq [y[k, -N] \ y[k, -N + 1] \ \cdots \ y[k, N]]^\top \\ \mathbf{h}[k] &\triangleq [h[k, -N] \ h[k, -N + 1] \ \cdots \ h[k, N]]^\top \\ \mathbf{w}[k] &\triangleq [w[k, -N] \ w[k, -N + 1] \ \cdots \ w[k, N]]^\top.\end{aligned}$$

To obtain a compact notation, we further stack \tilde{K} contiguous \tilde{N} -dimensional input, output, channel, and noise vectors,² into corresponding $\tilde{K}\tilde{N}$ -dimensional vectors \mathbf{s} , \mathbf{y} , \mathbf{h} , and \mathbf{w} , respectively, according to

$$\begin{aligned}\mathbf{s} &\triangleq [\mathbf{s}^\top[-K] \ \mathbf{s}^\top[-K + 1] \ \cdots \ \mathbf{s}^\top[K]]^\top \\ \mathbf{y} &\triangleq [\mathbf{y}^\top[-K] \ \mathbf{y}^\top[-K + 1] \ \cdots \ \mathbf{y}^\top[K]]^\top \\ \mathbf{h} &\triangleq [\mathbf{h}^\top[-K] \ \mathbf{h}^\top[-K + 1] \ \cdots \ \mathbf{h}^\top[K]]^\top \\ \mathbf{w} &\triangleq [\mathbf{w}^\top[-K] \ \mathbf{w}^\top[-K + 1] \ \cdots \ \mathbf{w}^\top[K]]^\top.\end{aligned}\tag{4.25}$$

Finally, we arrange ISI and ICI terms $\{p[l, m, k, n]\}$ in a $\tilde{K}\tilde{N} \times \tilde{K}\tilde{N}$ matrix \mathbf{P} with entries

$$[\mathbf{P}]_{n+k\tilde{N}, m+l\tilde{N}} = \begin{cases} p[l - K, m - N, k - K, n - N], & \text{if } (l, m) \neq (k, n) \\ 0, & \text{otherwise} \end{cases}$$

for $l, k \in [0 : \tilde{K} - 1]$ and $m, n \in [0 : \tilde{N} - 1]$. With these definitions, we can now compactly express the IO relation (4.18) as

$$\mathbf{y} = \mathbf{h} \odot \mathbf{s} + \mathbf{P}\mathbf{s} + \mathbf{w}.\tag{4.26}$$

4.3.5. Definition of the Capacity of the Discretized Channel Induced by (g, T, F)

For a given WH set (g, T, F) satisfying Properties i) and ii) in Section 4.3.1 and a given continuous-time channel \mathbb{H} , the capacity

²Recall that $\tilde{K} = 2K + 1$ and $\tilde{N} = 2N + 1$

$C_{\text{WSSUS-D}}$ of the induced discretized channel (4.18) is given by

$$C_{\text{WSSUS-D}} \triangleq \lim_{\tilde{K} \rightarrow \infty} \frac{1}{(\tilde{K} + 2K')T} \sup_{\mathcal{Q}} I(\mathbf{y}; \mathbf{s}). \quad (4.27)$$

Here, the supremum is taken over the set \mathcal{Q} of all distributions on \mathbf{s} that satisfy the average-power constraint (4.17). The value of K' in (4.27) is chosen so that (4.4) is satisfied (see Appendix F.2). We recall that, by construction, $C_{\text{WSSUS-D}}$ is a lower bound on the capacity C_{WSSUS} of the continuous-time channel, defined in (4.12).

4.3.6. The Capacity Lower Bound

We next derive a lower bound on $C_{\text{WSSUS-D}}$ by treating interference as noise. We then show that whenever the channel is underspread according to Definition 4.1, this lower bound, evaluated for an appropriately chosen WH set, is close to the AWGN-capacity upper bound C_{AWGN} in (4.15).

We shall first present a lower bound that is explicit in the power spectral density of the multivariate stationary channel process $\{\mathbf{h}[k]\}$

$$\mathbf{C}(\theta) \triangleq \sum_{k=-\infty}^{\infty} \mathbf{R}[k] e^{-i2\pi k\theta}, \quad |\theta| \leq \frac{1}{2} \quad (4.28)$$

where $\mathbf{R}[k' - k] \triangleq \mathbb{E}[\mathbf{h}[k']\mathbf{h}^H[k]]$.

Theorem 4.1. *Let (g, T, F) be a WH set satisfying Properties i) and ii) in Section 4.3.1 and consider an arbitrary Rayleigh-fading WSSUS channel with the scattering function $C_{\mathbb{H}}(\nu, \tau)$. Then, for a given SNR and a given bandwidth B , the capacity of the discretized channel induced by (g, T, F) is lower-bounded as:*

$$C_{\text{WSSUS-D}}(\text{SNR}) \geq \frac{B}{TF} \mathbb{E}_h \left[\log \left(1 + \frac{r[0,0]TF\text{SNR}|h|^2}{1 + TF\text{SNR}\sigma_I^2} \right) \right] \\ - \inf_{0 < \alpha < 1} \left\{ \frac{1}{T} \int_{-1/2}^{1/2} \log \det \left(\mathbf{I} + \frac{TF\text{SNR}}{\alpha} \mathbf{C}(\theta) \right) d\theta \right.$$

$$+ \frac{B}{TF} \log \left(1 + \frac{TF \text{SNR}}{1 - \alpha} \sigma_I^2 \right) \}. \quad (4.29)$$

Here,

$$\begin{aligned} h &\sim \mathcal{CN}(0, 1) \\ r[0, 0] &= \iint_{\mathbb{R}^2} C_{\mathbb{H}}(\nu, \tau) |A_g(\nu, \tau)|^2 d\tau d\nu \\ \sigma_I^2 &= \sum_{\substack{k=-\infty \\ (k,n) \neq (0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} \iint_{\mathbb{R}^2} C_{\mathbb{H}}(\nu, \tau) |A_g(\nu - nF, \tau - kT)|^2 d\tau d\nu \end{aligned} \quad (4.30)$$

and $\mathbf{C}(\theta)$, defined in (4.28), denotes the matrix-valued power spectral density of the discretized channel induced by (g, T, F) .

Proof. See Appendix F.4. □

The lower bound (4.29) we just obtained is difficult to analyze. In the corollary below we further lower-bound the right-hand side of (4.29) to get a less tight expression that is, however, explicit in the parameters ϵ and $\Delta_{\mathbb{H}}$ we introduced in Definition 4.1.

Corollary 4.2. *Let (g, T, F) be a WH set satisfying Properties i) and ii) in Section 4.3.1 and consider an arbitrary Rayleigh-fading WSSUS channel in the set $\mathcal{H}(\tau_0, \nu_0, \epsilon)$ with the scattering function $C_{\mathbb{H}}(\nu, \tau)$. Then, for a given SNR and a given bandwidth B , and under the technical condition³ $\tilde{\Delta}_{\mathbb{H}} \triangleq 2\nu_0 T < 1$, the capacity of the discretized channel induced by (g, T, F) is lower-bounded as*

$$C_{\text{WSSUS-D}}(\text{SNR}) \geq L(\text{SNR}, g, T, F, \tau_0, \nu_0, \epsilon)$$

³This technical condition is not restrictive for underspread channels if T and F are chosen so that $\nu_0 T = \tau_0 F$ (see Section 4.3.7). In this case, $2\nu_0 T = \sqrt{\Delta_{\mathbb{H}} TF} \ll 1$ for all values of TF of practical interest.

where

$$\begin{aligned}
 L(\text{SNR}, g, T, F, \tau_0, \nu_0, \epsilon) = & \frac{B}{TF} \left\{ \mathbb{E}_h \left[\log \left(1 + \frac{TF\text{SNR}(1-\epsilon)m_g|h|^2}{1+TF\text{SNR}(M_g+\epsilon)} \right) \right] \right. \\
 & - \inf_{0 < \alpha < 1} \left[\tilde{\Delta}_{\mathbb{H}} \log \left(1 + \frac{TF\text{SNR}}{\alpha \tilde{\Delta}_{\mathbb{H}}} \right) \right. \\
 & + (1 - \tilde{\Delta}_{\mathbb{H}}) \log \left(1 + \frac{TF\text{SNR} \epsilon}{\alpha(1 - \tilde{\Delta}_{\mathbb{H}})} \right) \\
 & \left. \left. + \log \left(1 + \frac{TF\text{SNR}}{1-\alpha} (M_g + \epsilon) \right) \right] \right\}. \tag{4.31}
 \end{aligned}$$

Here, $h \sim \mathcal{CN}(0, 1)$, $m_g \triangleq \min_{(\nu, \tau) \in \mathcal{D}} |A_g(\nu, \tau)|^2$,

$$M_g \triangleq \max_{(\nu, \tau) \in \mathcal{D}} \sum_{\substack{k=-\infty \\ (k, n) \neq (0, 0)}}^{\infty} \sum_{n=-\infty}^{\infty} |A_g(\nu - nF, \tau - kT)|^2$$

with $\mathcal{D} \triangleq [-\nu_0, \nu_0] \times [-\tau_0, \tau_0]$.

Proof. See Appendix F.5. □

The lower bound L in (4.31) is not useful in the asymptotic regimes $\text{SNR} \rightarrow 0$ and $\text{SNR} \rightarrow \infty$. In fact, the bound even turns negative when SNR is sufficiently small or sufficiently large. Nevertheless, as shown in Section 4.4, for underspread channels, L evaluated for particular WH sets is close to the capacity upper bound C_{AWGN} over all SNR values of practical interest. In the next two sections, we list some properties of L , which will be used in Section 4.4.

4.3.7. Reduction to a Square Setting

The lower bound $L(\text{SNR}, g, T, F, \tau_0, \nu_0, \epsilon)$ depends on seven parameters and is therefore difficult to analyze. We show next that if T and F

are chosen so that $\nu_0 T = \tau_0 F$, a condition often referred to as *grid matching rule* (Kozek, 1997, Equation (2.75)), two of these seven parameters can be dropped without loss of generality.

Lemma 4.3. *Let (g, T, F) be a WH set satisfying Properties i) and ii) in Section 4.3.1. Then, for every $\beta > 0$,*

$$L(\text{SNR}, g(t), T, F, \tau_0, \nu_0, \epsilon) = L\left(\text{SNR}, \sqrt{\beta}g(\beta t), \frac{T}{\beta}, \beta F, \frac{\tau_0}{\beta}, \beta\nu_0, \epsilon\right).$$

In particular, assume that $\nu_0 T = \tau_0 F$ and let $\beta = \sqrt{T/F} = \sqrt{\tau_0/\nu_0}$ and $\tilde{g}(t) = \sqrt{\beta}g(\beta t)$. Then,

$$\begin{aligned} L(\text{SNR}, g, T, F, \tau_0, \nu_0, \epsilon) \\ &= L\left(\text{SNR}, \tilde{g}, \sqrt{TF}, \sqrt{TF}, \sqrt{\Delta_{\mathbb{H}}}/2, \sqrt{\Delta_{\mathbb{H}}}/2, \epsilon\right) \\ &\triangleq L_s(\text{SNR}, \tilde{g}, TF, \Delta_{\mathbb{H}}, \epsilon). \end{aligned} \quad (4.32)$$

Proof. See Appendix F.6. □

In the remainder of the chapter, for the sake of simplicity of exposition, we will choose T and F so that the grid matching rule $\nu_0 T = \tau_0 F$ is satisfied. Then, as a consequence of Lemma 4.3, we can (and will) only consider, without loss of generality, WH sets of the form $(g, \sqrt{TF}, \sqrt{TF})$ and WSSUS channels in the class $\mathcal{H}(\sqrt{\Delta_{\mathbb{H}}}/2, \sqrt{\Delta_{\mathbb{H}}}/2, \epsilon)$.

4.3.8. Pulse-Design Criterion and Approximation for m_g and M_g

The lower bound in (4.31) can be tightened by maximizing it over all WH sets satisfying Properties i) and ii) in Section 4.3.1. This maximization implicitly provides an information-theoretic design criterion for $g(t)$, T , and F . Classic design rules for $g(t)$ available in the orthogonal frequency division multiplexing (OFDM) literature (see, for example, Matz et al., 2007, and references therein) are based on a maximization of the signal-to-interference ratio (SIR) in (4.18), for

a fixed value of TF (typically, $TF \approx 1.2$). The maximization of the lower bound (4.31) yields a more complete picture as it explicitly reveals the interplay between the product TF and the time-frequency localization properties of $g(t)$, reflected through the quantities m_g and M_g . Unfortunately, the maximization of L over (g, T, F) seems complicated, as the dependency of m_g and M_g on (g, T, F) is difficult to characterize analytically. This problem can be partially overcome when $\Delta_{\mathbb{H}} \ll 1$. In this case, a first-order Taylor-series expansion of m_g and M_g around $\Delta_{\mathbb{H}} = 0$ yields an accurate picture.

Lemma 4.4. *Let $(g, \sqrt{TF}, \sqrt{TF})$ be a WH set satisfying Properties i) and ii) in Section 4.3.1. Assume that $g(t)$ is real-valued and even, and that $A_g(\nu, \tau)$ is differentiable in the points $(n\sqrt{TF}, k\sqrt{TF})$ for all (n, k) and twice differentiable in $(0, 0)$; let $G(f) = \mathbb{F}[g(t)]$ and define $\tilde{\mathcal{D}} = [-\sqrt{\Delta_{\mathbb{H}}}/2, \sqrt{\Delta_{\mathbb{H}}}/2] \times [-\sqrt{\Delta_{\mathbb{H}}}/2, \sqrt{\Delta_{\mathbb{H}}}/2]$. For $\Delta_{\mathbb{H}} \ll 1$, we have*

$$m_g \approx \min_{(\nu, \tau) \in \tilde{\mathcal{D}}} |A_g(\nu, \tau)|^2 = 1 - c_m \Delta_{\mathbb{H}} \quad (4.33)$$

where $c_m = \pi^2(a_0^2 + b_0^2)$ with

$$a_0^2 = \int_{-\infty}^{\infty} t^2 |g(t)|^2 dt, \quad b_0^2 = \int_{-\infty}^{\infty} f^2 |G(f)|^2 df.$$

Moreover, still under the assumption that $\Delta_{\mathbb{H}} \ll 1$, we have

$$\begin{aligned} M_g &\approx \max_{(\nu, \tau) \in \tilde{\mathcal{D}}} \sum_{\substack{k=-\infty \\ (k,n) \neq (0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} |A_g(\nu - nF, \tau - kT)|^2 \\ &= c_M \Delta_{\mathbb{H}} \end{aligned} \quad (4.34)$$

where $c_M = \sum_{\substack{k=-\infty \\ (k,n) \neq (0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} [|a_{k,n}|^2 + |b_{k,n}|^2] / 4$, with

$$a_{k,n} = -i2\pi \int_{-\infty}^{\infty} tg(t)g(t + k\sqrt{TF})e^{i2\pi n\sqrt{TF}t} dt$$

$$b_{k,n} = i2\pi \int_{-\infty}^{\infty} fG(f - n\sqrt{TF})G(f)e^{-i2\pi k\sqrt{TF}f} df.$$

Proof. See Appendix F.7. □

4.3.9. A Simple WH Set

We next present an example of a family of WH sets $(g, \sqrt{TF}, \sqrt{TF})$ satisfying Properties i) and ii) in Section 4.3.1, and for which, in addition, $g(t)$ is real-valued and even. Take $1 < TF < 2$, let $\zeta = \sqrt{TF}$, $\delta = TF - 1$, and $G(f) = \mathbb{F}[g(t)]$. We choose $G(f)$ as the (positive) square root of a raised-cosine pulse:

$$G(f) = \begin{cases} \sqrt{\zeta}, & \text{if } |f| \leq \frac{1-\delta}{2\zeta} \\ \sqrt{\frac{\zeta}{2}(1 + S(f))}, & \text{if } \frac{1-\delta}{2\zeta} \leq |f| \leq \frac{1+\delta}{2\zeta} \\ 0, & \text{otherwise} \end{cases} \quad (4.35)$$

where $S(f) = \cos\left[\frac{\pi\zeta}{\delta}\left(|f| - \frac{1-\delta}{2\zeta}\right)\right]$. As $(1 + \delta)/(2\zeta) = \zeta/2$, the function $G(f)$ has compact support of length $\zeta = \sqrt{TF}$. Furthermore, $G(f)$ has unit energy, is real-valued and even, and satisfies

$$\sum_{n=-\infty}^{\infty} G(f - n/\zeta)G(f - n/\zeta - k\zeta) = \zeta\delta[k]. \quad (4.36)$$

By (Christensen, 2003, Theorem 8.7.2), we can, therefore, conclude that the WH set $(g(t), 1/\sqrt{TF}, 1/\sqrt{TF})$ is a tight WH frame for $\mathcal{L}^2(\mathbb{C})$, and, by duality, the WH set $(g(t), \sqrt{TF}, \sqrt{TF})$ is orthonormal. Finally, it can be shown that $g(t) = \mathcal{O}(1/|t|^2)$ whenever $TF > 1$. Note that, for $TF = 1$, the pulse $G(f)$ reduces to the rectangular pulse and, consequently, $g(t)$ reduces to a sinc function, which has poor time localization.

4.4. FINITE-SNR ANALYSIS OF THE LOWER BOUND

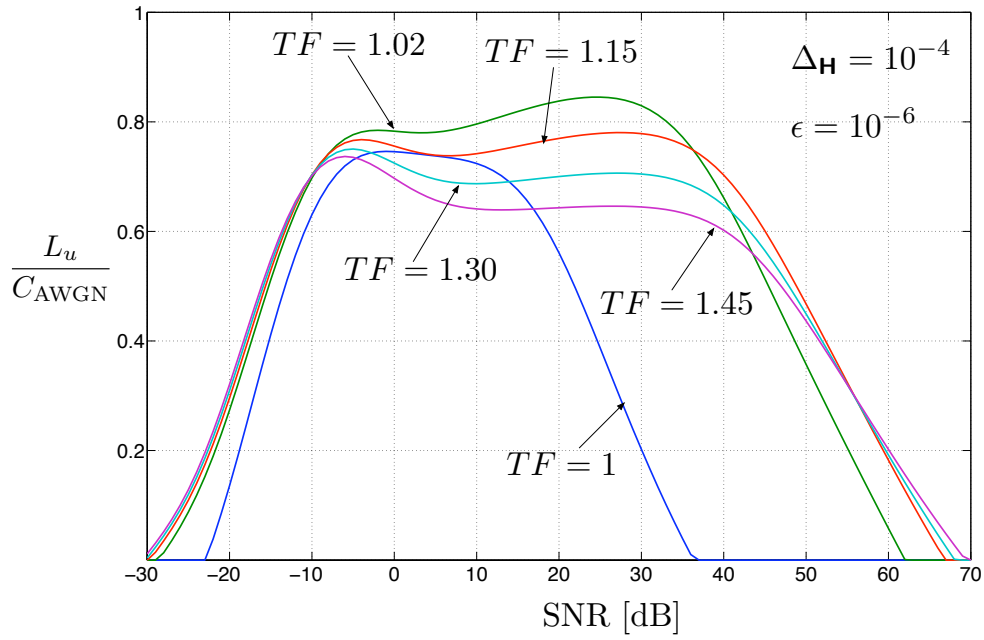
We now study the behavior of the lower bound L_s in (4.32) evaluated for the WH set constructed in the previous section, under the assumption that the underlying WSSUS channel is underspread according to Definition 4.1, i.e., $\Delta_{\mathbb{H}} \ll 1$ and $\epsilon \ll 1$. As $\Delta_{\mathbb{H}} \ll 1$, we can replace m_g and M_g in L_s by the first-order term of their Taylor-series expansions derived in Lemma 4.4 and obtain the following approximation:

$$\begin{aligned}
 & L_s(\text{SNR}, g, TF, \Delta_{\mathbb{H}}, \epsilon) \\
 & \approx \frac{B}{TF} \left\{ \mathbb{E}_h \left[\log \left(1 + \frac{TF\text{SNR}(1-\epsilon)(1-c_m\Delta_{\mathbb{H}})|h|^2}{1+TF\text{SNR}(c_M\Delta_{\mathbb{H}}+\epsilon)} \right) \right] \right. \\
 & \quad - \inf_{0 < \alpha < 1} \left\{ \sqrt{\Delta_{\mathbb{H}}TF} \log \left(1 + \frac{TF\text{SNR}}{\alpha\sqrt{\Delta_{\mathbb{H}}TF}} \right) \right. \\
 & \quad \left. + \left(1 - \sqrt{\Delta_{\mathbb{H}}TF} \right) \log \left(1 + \frac{TF\text{SNR}\epsilon}{\alpha(1-\sqrt{\Delta_{\mathbb{H}}TF})} \right) \right. \\
 & \quad \left. \left. + \log \left(1 + \frac{TF\text{SNR}}{1-\alpha} (c_M\Delta_{\mathbb{H}} + \epsilon) \right) \right\} \right\} \\
 & \triangleq L_u(\text{SNR}, g, TF, \Delta_{\mathbb{H}}, \epsilon).
 \end{aligned} \tag{4.37}$$

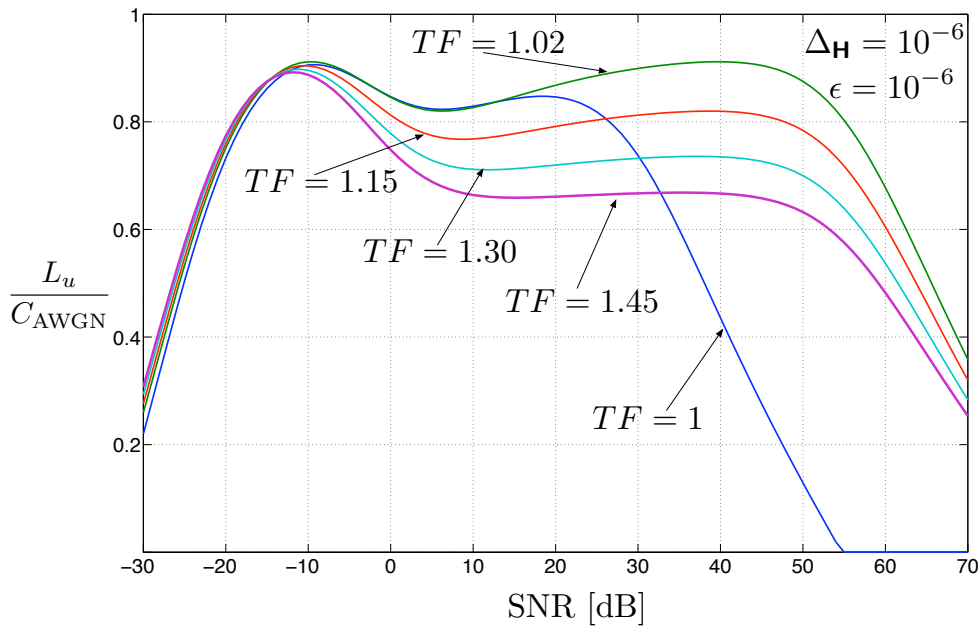
4.4.1. Trade-off between Interference Reduction and Maximization of Number of Degrees of Freedom

In Figure 4.1, we plot the ratio $L_u(\text{SNR}, g, TF, \Delta_{\mathbb{H}}, \epsilon)/C_{\text{AWGN}}$ for $\epsilon = 10^{-6}$, $\Delta_{\mathbb{H}} = 10^{-4}$ and for $\epsilon = 10^{-6}$, $\Delta_{\mathbb{H}} = 10^{-6}$. The different curves correspond to different values of TF . We can observe that the choice $TF = 1$ is highly suboptimal. The reason for this suboptimality is the significant reduction in SIR this choice entails. In fact (as already mentioned), when $TF = 1$, the pulse $g(t)$ reduces to a sinc function, which has poor time localization.

4. NONCOHERENT WSSUS CHANNEL



(a)



(b)

Fig. 4.1.: Lower bounds $L_u(\text{SNR}, g, TF, \Delta_{\mathbb{H}}, \epsilon)$ normalized with respect to the AWGN capacity. The bounds are computed for WH sets based on a root-raised-cosine pulse (see (4.35)), for different values of the grid-parameter product TF . The channel parameters $\Delta_{\mathbb{H}}$ and ϵ are set to be $\Delta_{\mathbb{H}} = 10^{-4}$, $\epsilon = 10^{-6}$ in (a) and $\Delta_{\mathbb{H}} = \epsilon = 10^{-6}$ in (b).

A value of TF slightly above 1 leads to a significant improvement in the time localization of $g(t)$ and to a corresponding increase in the lower bound $L_u(\text{SNR}, g, TF, \Delta_{\mathbb{H}}, \epsilon)$ for all SNR values of practical interest, despite the (small) loss of degrees of freedom. A further increase of the product TF seems to be detrimental for all but very high SNR values, where the ratio $L_u(\text{SNR}, g, TF, \Delta_{\mathbb{H}}, \epsilon)/C_{\text{AWGN}}$ is much smaller than 1: the rate loss due to the reduction of the number of degrees of freedom is more significant than the rate increase due to the resulting SIR improvement.

4.4.2. Sensitivity of Capacity to the Parameters $\Delta_{\mathbb{H}}$ and ϵ

The results presented in Figure 4.1 suggest that, for $TF = 1.02$, the lower bound $L_u(\text{SNR}, g, TF, \Delta_{\mathbb{H}}, \epsilon)$ is close to the AWGN-capacity upper bound $C_{\text{AWGN}}(\text{SNR})$ over a quite large range of SNR values. To make this statement precise, we compute the SNR interval $[\text{SNR}_{\min}, \text{SNR}_{\max}]$ over which

$$L_u(\text{SNR}, g, TF, \Delta_{\mathbb{H}}, \epsilon) \geq 0.75 C_{\text{AWGN}}(\text{SNR}). \quad (4.38)$$

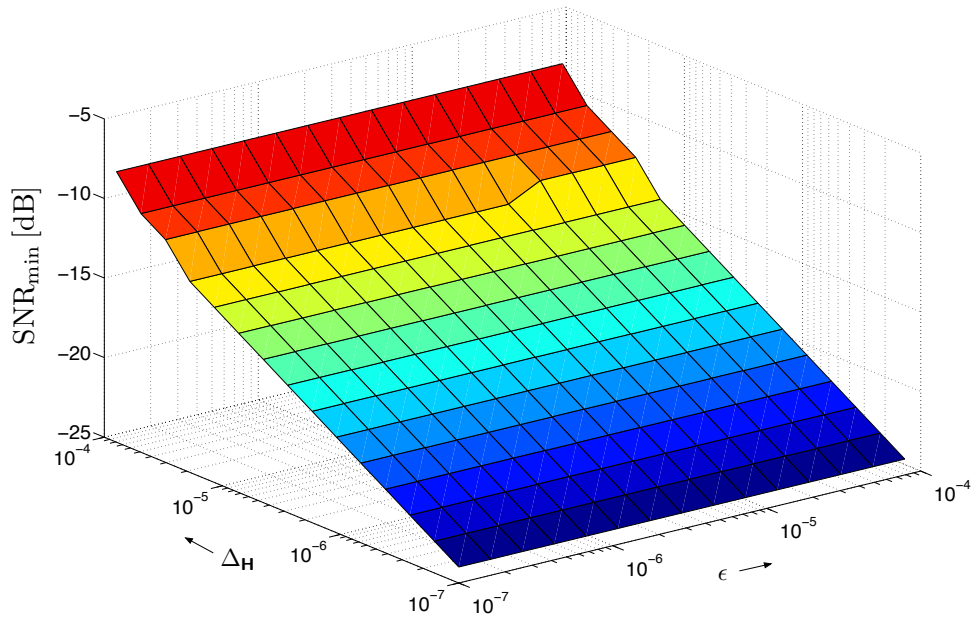
The interval end points SNR_{\min} and SNR_{\max} can easily be computed numerically; the corresponding values for SNR_{\min} and SNR_{\max} are illustrated in Figure 4.2 for different $(\Delta_{\mathbb{H}}, \epsilon)$ pairs. For the WH set and WSSUS underspread channels considered in this section, we have that $\text{SNR}_{\min} \in [-25 \text{ dB}, -7 \text{ dB}]$ and $\text{SNR}_{\max} \in [32 \text{ dB}, 68 \text{ dB}]$.

An analytic characterization of SNR_{\min} and SNR_{\max} is more difficult. Insights on how these two quantities are related to the channel parameters $\Delta_{\mathbb{H}}$ and ϵ can be obtained by the following crude “back-of-the-envelope” approximation to $L_u(\text{SNR}, g, TF, \Delta_{\mathbb{H}}, \epsilon)$.

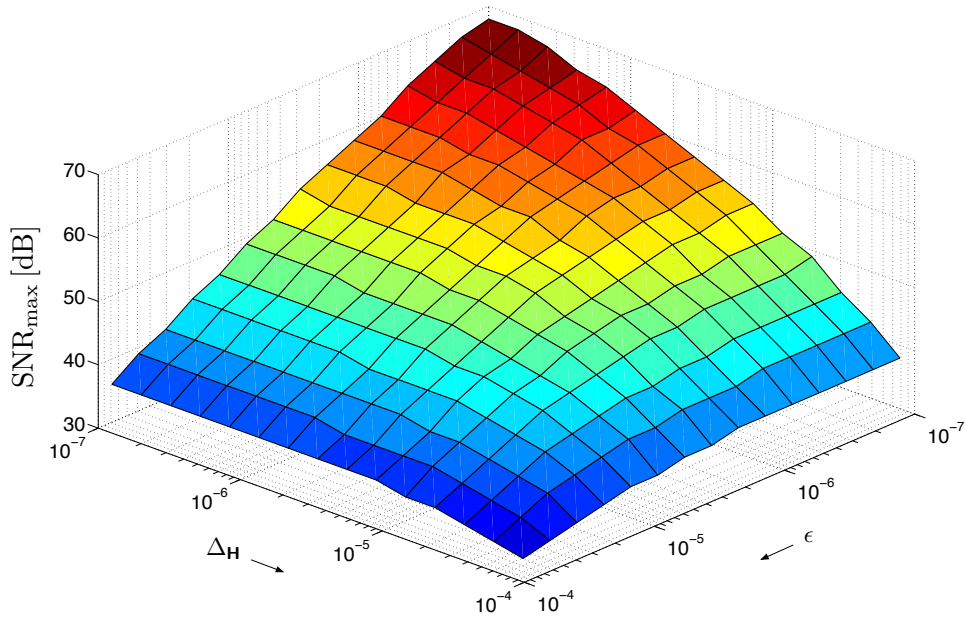
As a first step, we further lower-bound L_u by picking $\alpha = 1/2$ and dropping the last two (positive) terms in (4.37) to obtain

$$L_u \geq \frac{B}{TF} \left\{ \mathbb{E}_h \left[\log \left(1 + \frac{TF \text{SNR} (1 - \epsilon) (1 - c_m \Delta_{\mathbb{H}}) |h|^2}{1 + TF \text{SNR} (c_M \Delta_{\mathbb{H}} + \epsilon)} \right) \right] - \sqrt{\Delta_{\mathbb{H}} TF} \log \left(1 + \frac{2TF \text{SNR}}{\sqrt{\Delta_{\mathbb{H}} TF}} \right) \right\}. \quad (4.39)$$

4. NONCOHERENT WSSUS CHANNEL



(a)



(b)

Fig. 4.2.: Minimum SNR value SNR_{\min} in (a) and maximum SNR value SNR_{\max} in (b) for which (4.38) holds, as a function of $\Delta_{\mathbb{H}}$ and ϵ . The lower bound $L_s(\text{SNR}, g, TF, \Delta_{\mathbb{H}}, \epsilon)$ is computed for a WH set base on a root-raised-cosine pulse; furthermore, $TF = 1.02$.

Next, the choice $TF = 1.02$ results in $c_m \approx 25.87$ and $c_M \approx 0.77$, which allows us to use the approximation $(1 - c_m \Delta_{\mathbb{H}}) \approx 1$ and the bound $c_M \Delta_{\mathbb{H}} + \epsilon \leq \Delta_{\mathbb{H}} + \epsilon$ in the first term on the right-hand side of (4.39). Making further approximations $TF \approx 1$, $(1 - \epsilon) \approx 1$ we arrive at

$$L_u(\text{SNR}, g, TF, \Delta_{\mathbb{H}}, \epsilon) \gtrsim B \left\{ \mathbb{E}_h \left[\log \left(1 + \frac{\text{SNR} |h|^2}{1 + \text{SNR}(\Delta_{\mathbb{H}} + \epsilon)} \right) \right] - \sqrt{\Delta_{\mathbb{H}}} \log \left(1 + \frac{2\text{SNR}}{\sqrt{\Delta_{\mathbb{H}}}} \right) \right\}. \quad (4.40)$$

Now we see that as long as $\sqrt{\Delta_{\mathbb{H}}} \ll \text{SNR} \ll \Delta_{\mathbb{H}} + \epsilon$, the first term in (4.40) dominates the second term, and the right-hand side of (4.40) is close to $C_{\text{AWGN}}(\text{SNR})$.

The following rule of thumb then holds: the capacity of all WSSUS underspread channels with scattering function $C_{\mathbb{H}}(\nu, \tau)$ having no more than ϵ of its volume outside a rectangle (in the Doppler-delay plane) of area $\Delta_{\mathbb{H}}$, is close to $C_{\text{AWGN}}(\text{SNR})$ for all SNR that satisfy $\sqrt{\Delta_{\mathbb{H}}} \ll \text{SNR} \ll 1/(\Delta_{\mathbb{H}} + \epsilon)$, independently of whether $C_{\mathbb{H}}(\nu, \tau)$ is compactly supported or not, and of its shape. The condition $\sqrt{\Delta_{\mathbb{H}}} \ll \text{SNR} \ll 1/(\Delta_{\mathbb{H}} + \epsilon)$ holds for all channels and SNR values of practical interest.

4.5. SUMMARY OF RESULTS

In this chapter, we provided an information-theoretic characterization of Rayleigh-fading channels that satisfy the WSSUS and the underspread assumptions. The information-theoretic analysis is built upon a discretization of WSSUS underspread channels that takes the underspread property explicitly into account to minimize ISI and ICI in the discretized IO relation. The channel discretization is accomplished by transmitting and receiving on a WH set generated by a pulse that is well localized in time and frequency.

We derived a capacity lower bound that is obtained by treating interference as noise. This lower bound yields valuable insights into the

capacity of Rayleigh-fading underspread WSSUS channels over a large range of SNR values of practical interest. On the basis of this lower bound, we derived an information-theoretic criterion for the design of capacity-optimal WH sets. This criterion is more fundamental than criteria based on SIR maximization (see Matz et al., 2007), because it sheds light on the trade-off between number of degrees of freedom and time-frequency localization of the pulse $g(t)$. Unfortunately, the corresponding optimization problem is hard to solve. We simplified the problem by fixing $g(t)$ to be a root-raised-cosine pulse and performing an optimization over the grid parameters T and F . Our analysis shows that the optimal value of the grid-parameter product TF is close to 1 (but strictly larger than 1) for a large range of SNR values of practical interest. This result suggests that the maximization of the number of degrees of freedom should be privileged over the SIR maximization in the design of capacity-maximizing WH sets.

Even though our analysis was confined to a specific pulse shape (i.e., root-raised-cosine), we were able to show that for all Rayleigh-fading WSSUS channels that are underspread according to Definition 4.1, the corresponding capacity lower bound is close to the AWGN-capacity upper bound for all SNR values of practical interest, independently of whether the scattering function is compactly supported or not (a fine detail of the channel model). In other words, the capacity of Rayleigh-fading underspread WSSUS channels starts being sensitive to this fine detail of the channel model only for SNR values that lie outside the SNR range typically encountered in real-world systems. Hence, the Rayleigh-fading WSSUS underspread model is a robust model.

To conclude, an interesting open problem, the solution of which would strengthen our results, is to obtain an upper bound on the capacity of (4.1) based on perfect CSI at the receiver.

CHAPTER 5

Noncoherent SIMO Channel

It is well known that the *coherent-capacity* pre-log (i.e., the asymptotic ratio between capacity and the logarithm of SNR, as SNR goes to infinity) of a SIMO fading channel is equal to 1 and is, hence, the same as that of a SISO fading channel (Telatar, 1999). In the practically more relevant *noncoherent setting*, where neither transmitter nor receiver have CSI, but both are aware of the channel statistics, the effect of multiple antennas on the capacity¹ pre-log is understood only for a specific simple channel model, namely, the *constant block-fading* model. In this model, the channel is assumed to remain constant over a block of D symbols and to change in an independent fashion from block to block (Marzetta and Hochwald, 1999). For this model, the SIMO capacity pre-log is again equal to the SISO capacity pre-log, but, differently from the coherent case, is given by $1 - 1/D$ (Hochwald and Marzetta, 2000; Zheng and Tse, 2002).

A more general way of capturing channel variations in time is to assume that the fading process is *stationary*. In this case, the capacity pre-log is known only in the SISO (Lapidoth, 2005) and the MISO (Koch, 2009, Theorem 4.15) cases. The capacity bounds for the SIMO stationary-fading channel available in the literature (Koch, 2009,

¹In the remainder of this chapter, we consider the noncoherent setting only. Consequently, we will refer to capacity in the noncoherent setting simply as capacity. Furthermore, we shall assume Rayleigh fading throughout.

Theorem 4.13) do not allow one to determine whether the capacity pre-log in the SIMO case can be larger than that in the SISO case.

In this chapter, we focus on a channel model that can be seen as lying in between the general stationary-fading model considered by Lapidath (2005) and Koch (2009), and the simpler constant block-fading model analyzed by Marzetta and Hochwald (1999) and Zheng and Tse (2002). Specifically, we assume that the fading process is independent across blocks of length D and temporally correlated within blocks, with the rank of the corresponding $D \times D$ channel covariance matrix given by² $M < D$. For this channel model, referred to as the *correlated block-fading* model in the following, the SISO capacity pre-log is equal to $1 - M/D$ (Liang and Veeravalli, 2004).³ The SIMO and MIMO capacity pre-logs are not known in this case. A conjecture proposed by Liang and Veeravalli (2004) on the MIMO capacity pre-log implies that the capacity pre-log in the SIMO case would be the same as that in the SISO case. In this chapter, we disprove the Liang and Veeravalli (2004) conjecture by showing that in the SIMO case a capacity pre-log of $1 - 1/D$ can be obtained when the number of receive antennas is equal to M , and the channel covariance matrix satisfies a certain technical condition detailed in Theorem 5.1.

5.1. SYSTEM MODEL

We consider a SIMO channel with M receive antennas. The fading in each *component channel* follows the correlated block-fading model described in the previous section, namely, it is independent across blocks of length D , and correlated within blocks, with the rank of the corresponding channel covariance matrix given by $M < D$. Note that we assume the rank of the channel covariance matrix to be equal to the number of receive antennas. Our analysis relies heavily on this

²When $M = D$, capacity is known to grow double-logarithmically in SNR (Lapidath and Moser, 2003), and, hence, the capacity pre-log is zero.

³The constant block-fading model is obviously a special case ($M = 1$) of the correlated block-fading model.

assumption⁴. Across component channels, the fading is independent and identically distributed. The IO relation (within every block) for the m th component channel can be written as

$$\mathbf{y}_m = \sqrt{\text{SNR}} \text{diag}(\mathbf{h}_m) \mathbf{s} + \mathbf{w}_m, \quad m \in [1 : M]$$

where the vector $\mathbf{s} = [s_1 \cdots s_D]^\top \in \mathbb{C}^D$ contains the D -dimensional signal transmitted within the block, and the vectors $\mathbf{y}_m, \mathbf{w}_m \in \mathbb{C}^D$ contain the corresponding received signal and additive noise, respectively, at the m th antenna. Finally, $\mathbf{h}_m \in \mathbb{C}^D$ contains the channel coefficients between the transmit antenna and the m th receive antenna. We assume that $\mathbf{w}_m \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_D)$ and $\mathbf{h}_m \sim \mathcal{CN}(\mathbf{0}, \mathbf{P}\mathbf{P}^\text{H})$ are mutually independent (and independent across m) and that $\mathbf{P} = [p_{i,j}]_{i \in [1 : D], j \in [1 : M]} \in \mathbb{C}^{D \times M}$ (which is the same for all blocks) has rank $M < D$. It will turn out convenient to write the channel-coefficient vector in whitened form as $\mathbf{h}_m = \mathbf{P}\mathbf{g}_m$, where $\mathbf{g}_m \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$. Finally, we assume that \mathbf{g}_m and \mathbf{w}_m change in an independent fashion from block to block.

If we define

$$\begin{aligned} \mathbf{y}^\text{T} &\triangleq [\mathbf{y}_1^\text{T} \cdots \mathbf{y}_M^\text{T}] & \mathbf{g}^\text{T} &= [g_1 \cdots g_{M^2}] \triangleq [\mathbf{g}_1^\text{T} \cdots \mathbf{g}_M^\text{T}] \\ \mathbf{w}^\text{T} &\triangleq [\mathbf{w}_1^\text{T} \cdots \mathbf{w}_M^\text{T}] & \mathbf{S} &\triangleq \text{diag}(\mathbf{s}) \end{aligned} \quad (5.1)$$

we can write the channel IO relation in the following—more compact—form

$$\mathbf{y} = \sqrt{\text{SNR}} (\mathbf{I}_M \otimes \mathbf{S}\mathbf{P}) \mathbf{g} + \mathbf{w}. \quad (5.2)$$

The capacity of the channel (5.2) is defined as

$$C_{\text{SIMO}}(\text{SNR}) \triangleq \frac{1}{D} \sup_{f_{\mathbf{s}}(\cdot)} I(\mathbf{s}; \mathbf{y}) \quad (5.3)$$

where the supremum is taken over all input distributions $f_{\mathbf{s}}(\cdot)$ that satisfy the average-power constraint $\mathbb{E}[\|\mathbf{s}\|^2] \leq D$.

⁴The results in this thesis can be generalized to the case when the rank of the channel covariance matrix is not equal to the number of receive antennas. This generalization is based on Hironaka's theorem on resolution of singularities, a famous result from algebraic geometry. See Riegler et al. (2011) for details.

5.2. INTUITIVE ANALYSIS

In this section, we describe a “back-of-the-envelope” method for guessing the capacity pre-log. A formal justification of this procedure is provided in Section 5.3.

The capacity pre-log characterizes the asymptotic behavior of the fading-channel capacity at high SNR, i.e., in the regime where the additive noise can “effectively” be ignored. In order to guess the capacity pre-log, we therefore consider the problem of identifying the transmit symbols s_i , $i \in [1 : D]$, from the *noise-free* (and rescaled) observation

$$\hat{\mathbf{y}} = [\hat{y}_1 \ \cdots \ \hat{y}_{MD}]^T \triangleq (\mathbf{I}_M \otimes \mathbf{S}\mathbf{P}) \mathbf{g}. \quad (5.4)$$

Specifically, we shall ask the question: “How many symbols s_i can be identified uniquely from $\hat{\mathbf{y}}$ given that the channel coefficients \mathbf{g} are unknown but the statistics of the channel, i.e., the matrix \mathbf{P} , are known?” The claim we make is that the capacity pre-log is given by the number of these symbols divided by the block length D .

We start by noting that the unknown variables in (5.4) are \mathbf{g} and \mathbf{s} , which means that we have a quadratic system of equations. It turns out, however, that the simple change of variables $z_i \triangleq 1/s_i$, $i \in [1 : D]$, (we make the technical assumption $0 < |s_i| < \infty$, $i \in [1 : D]$, in the remainder of this section) transforms (5.4) into a system of equations that is linear in \mathbf{g} and z_i , $i \in [1 : D]$. Since the transformation $z_i \triangleq 1/s_i$ is invertible for $0 < s_i < \infty$, uniqueness of the solution of the linear system of equations in \mathbf{g} and z_i , $i \in [1 : D]$, is equivalent to uniqueness of the solution of the quadratic system of equations in \mathbf{g} and s_i , $i \in [1 : D]$. For simplicity of exposition and concreteness, we consider the special case $D = 3$ and $M = 2$. A direct computation

reveals that (5.4) is equivalent to

$$\begin{bmatrix} p_{11} & p_{12} & 0 & 0 & \hat{y}_1 & 0 & 0 \\ p_{21} & p_{22} & 0 & 0 & 0 & \hat{y}_2 & 0 \\ p_{31} & p_{32} & 0 & 0 & 0 & 0 & \hat{y}_3 \\ 0 & 0 & p_{11} & p_{12} & \hat{y}_4 & 0 & 0 \\ 0 & 0 & p_{21} & p_{22} & 0 & \hat{y}_5 & 0 \\ 0 & 0 & p_{31} & p_{32} & 0 & 0 & \hat{y}_6 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ -z_1 \\ -z_2 \\ -z_3 \end{bmatrix} = \mathbf{0}. \quad (5.5)$$

The solution of this linear system of equations is not unique, as we have 6 equations in 7 unknowns. The $s_i = 1/z_i$, $i \in [1:3]$, can, therefore, not be determined uniquely from $\hat{\mathbf{y}}$. However, if we transmit one pilot symbol and two data symbols, the system of equations becomes solvable. Take for example $s_1 = 1$ and let the receiver know the value of this (pilot) symbol. Then (5.5) reduces to the following inhomogeneous system of 6 equations in 6 unknowns

$$\underbrace{\begin{bmatrix} p_{11} & p_{12} & 0 & 0 & 0 & 0 \\ p_{21} & p_{22} & 0 & 0 & \hat{y}_2 & 0 \\ p_{31} & p_{32} & 0 & 0 & 0 & \hat{y}_3 \\ 0 & 0 & p_{11} & p_{12} & 0 & 0 \\ 0 & 0 & p_{21} & p_{22} & \hat{y}_5 & 0 \\ 0 & 0 & p_{31} & p_{32} & 0 & \hat{y}_6 \end{bmatrix}}_{\mathbf{B}} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ -z_2 \\ -z_3 \end{bmatrix} = \begin{bmatrix} \hat{y}_1 \\ 0 \\ 0 \\ \hat{y}_4 \\ 0 \\ 0 \end{bmatrix}. \quad (5.6)$$

This system of equations has a unique solution if $\det \mathbf{B} \neq 0$. We prove in Lemma 5.2 that under the technical condition on \mathbf{P} specified in Theorem 5.1 below, we, indeed, have that $\det \mathbf{B} \neq 0$ for almost all⁵ $\hat{y}_2, \hat{y}_3, \hat{y}_5, \hat{y}_6$. It, therefore, follows that for almost all $\hat{\mathbf{y}}$, the system of equations (5.6) has a unique solution. Consequently, we can recover z_2 and z_3 , and, hence, $s_2 = 1/z_2$ and $s_3 = 1/z_3$.

Summarizing our findings, we expect that the capacity pre-log of the channel (5.2), for the special case $D = 3$ and $M = 2$, is equal to $2/3$. This is larger than the capacity pre-log of the corresponding SISO

⁵Except for a set of measure zero.

channel (i.e., the capacity pre-log of one of the component channels), which is equal to $1 - M/D = 1/3$ (Liang and Veeravalli, 2004).

In general, we expect that under some technical conditions on \mathbf{P} the capacity pre-log of the SIMO channel as defined in Section 5.1 is equal to $(D - 1)/D = 1 - 1/D$. This is what we intend to show rigorously in the next section.

5.3. A LOWER BOUND ON THE CAPACITY PRE-LOG

The main result of this chapter is the following theorem.

Theorem 5.1. *Assume that there exists a subset of indices $\mathcal{I} \subset [1 : D]$ of cardinality $|\mathcal{I}| = M + 1$ such that the $[(M + 1) \times M]$ -dimensional submatrix⁶ $\tilde{\mathbf{P}} \triangleq [\mathbf{P}]_{\mathcal{I}, \diamond}$ of the matrix \mathbf{P} in (5.2) satisfies the following Property (A): Every set of M rows of $\tilde{\mathbf{P}}$ is linearly independent. Then, the capacity of the SIMO channel (5.2) can be lower-bounded as*

$$C_{\text{SIMO}}(\text{SNR}) \geq (1 - 1/D) \log(\text{SNR}) + \mathcal{O}(1), \quad \text{SNR} \rightarrow \infty. \quad (5.7)$$

Remark 5.1. For the special case $D = M + 1$, (5.7) yields a lower bound on the capacity pre-log that is tight. A matching upper bound can be obtained through steps similar to those in the proof of (Liang and Veeravalli, 2004, Proposition 4). Establishing tight upper bounds on the capacity pre-log for general values of D , however, seems to be an open problem. Other tools than those used by Liang and Veeravalli (2004) are probably needed.

Remark 5.2. When $M = 1$, the channel in (5.2) reduces to a SISO constant block-fading channel, and the lower bound (5.7) yields the correct capacity pre-log of Zheng and Tse (2002) and Hochwald and Marzetta (2000).

⁶Notations for subvectors and submatrices used in this chapter are define in Section H.5.

Remark 5.3. Property (A) is not very restrictive and is satisfied by many practically relevant matrices \mathbf{P} . For example, removing arbitrary set of $D - M$ columns from a $D \times D$ discrete Fourier transform (DFT) matrix, results in a matrix that satisfies Property (A) when D is prime (Tao, 2005). DFT covariance matrices occur naturally in basis-expansion models for time-varying channels (Liang and Veeravalli, 2004).

Proof. We choose an input distribution for which the entries s_i , $i \in [1 : D]$, of \mathbf{s} , are i.i.d., have zero mean and unit variance, and satisfy $\mathbb{E}[\log |s_i|] > -\infty$ and $h(s_i) > -\infty$. For example, we can take $s_i \sim \mathcal{CN}(0, 1)$. We then lower-bound $I(\mathbf{s}; \mathbf{y})$ in (5.3), evaluated for this input distribution. More precisely, we use $I(\mathbf{s}; \mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y} | \mathbf{s})$ and bound the two differential entropy terms separately. Note that the class of input distributions for which (5.7) holds is large. This does not come as a surprise, as we are interested in the capacity pre-log only.

As \mathbf{y} conditional on \mathbf{s} is JPG, the conditional differential entropy $h(\mathbf{y} | \mathbf{s})$ can be upper-bounded in a standard fashion as follows:

$$\begin{aligned}
 h(\mathbf{y} | \mathbf{s}) &= MD \log(\pi e) \\
 &+ \mathbb{E}_{\mathbf{s}} \left[\log \det(\mathbf{I}_{MD} + \text{SNR} (\mathbf{I}_M \otimes \mathbf{S}\mathbf{P}) \mathbb{E}_{\mathbf{g}}[\mathbf{g}\mathbf{g}^H] (\mathbf{I}_M \otimes \mathbf{P}^H \mathbf{S}^H)) \right] \\
 &= \mathbb{E}_{\mathbf{s}} \left[\log \det(\mathbf{I}_{MD} + \text{SNR} (\mathbf{I}_M \otimes \mathbf{S}\mathbf{P}) (\mathbf{I}_M \otimes \mathbf{P}^H \mathbf{S}^H)) \right] + c \\
 &= \mathbb{E}_{\mathbf{s}} \left[\log \det(\mathbf{I}_{MD} + \text{SNR} (\mathbf{I}_M \otimes \mathbf{S}\mathbf{P}\mathbf{P}^H \mathbf{S}^H)) \right] + c \\
 &= M \mathbb{E}_{\mathbf{s}} \left[\log \det(\mathbf{I}_D + \text{SNR} (\mathbf{S}\mathbf{P}\mathbf{P}^H \mathbf{S}^H)) \right] + c \\
 &= M \mathbb{E}_{\mathbf{s}} \left[\log \det(\mathbf{I}_M + \text{SNR} (\mathbf{P}^H \mathbf{S}^H \mathbf{S}\mathbf{P})) \right] + c \\
 &\stackrel{(a)}{\leq} M \log \det(\mathbf{I}_M + \text{SNR} (\mathbf{P}^H \mathbb{E}_{\mathbf{s}}[\mathbf{S}^H \mathbf{S}] \mathbf{P})) + c \\
 &= M \sum_{i=1}^M \log(1 + \text{SNR} \lambda_i(\mathbf{P}^H \mathbf{P})) + c \\
 &\stackrel{(b)}{\leq} M^2 \log(\text{SNR}) + \mathcal{O}(1), \quad \text{SNR} \rightarrow \infty, \tag{5.8}
 \end{aligned}$$

where c , here and in the remainder of this chapter, stands for a con-

stant⁷ that is independent of SNR; (a) follows from Jensen's inequality; (b) holds because \mathbf{P} has rank M .

Finding a tight lower bound on $h(\mathbf{y})$ is the main difficulty of the proof. In fact, the differential entropy of \mathbf{y} is often intractable even for simple input distributions. The main technical contribution of this chapter is presented in Section 5.3.1 below, where we show that if Property (A) is satisfied and if the input distribution satisfies the conditions specified at the beginning of this proof, we have

$$h(\mathbf{y}) \geq (D - 1 + M^2) \log(\text{SNR}) + c. \quad (5.9)$$

Combining (5.8) and (5.9) then yields the desired result. Note that in order to establish (5.8) it is sufficient to use that \mathbf{P} has rank M , whereas the more restrictive Property (A) is crucial to establish (5.9). \square

5.3.1. A Lower Bound on $h(\mathbf{y})$

The main idea of our approach is to relate $h(\mathbf{y})$ to $h(\mathbf{g}, \mathbf{s}) = h(\mathbf{g}) + h(\mathbf{s})$, which is generally much simpler to compute than $h(\mathbf{y})$. It is possible to relate the entropies of two random vectors in a simple way if the vectors are of the same dimension and are connected by a *deterministic one-to-one* (in the sense of (Rudin, 1987, p.7)) function (see Lemma 5.3 below).

The connection between \mathbf{y} and $[\mathbf{g}^T \mathbf{s}^T]^T$ is not deterministic, because of the presence of *noise* in the IO relation (5.2). Moreover, the vectors \mathbf{y} and $[\mathbf{g}^T \mathbf{s}^T]^T$ have, in general, *different dimensionality*. Therefore, it is hard to establish a direct relationship between $h(\mathbf{y})$ and $h(\mathbf{g}, \mathbf{s})$.

The difficulty can be resolved by the following observations.

- (i) It is possible to show that if s_1 is a fixed parameter, then there is a deterministic one-to-one function between $[\mathbf{g}^T \mathbf{s}_{[2:D]}^T]^T$ and a specific *subset* $\mathcal{J} \subset [1:DM]$ of components of the *noiseless* version $\hat{\mathbf{y}}$ of the output vector (see Lemma 5.2 below).

⁷The value of this constant can change at each appearance.

- (ii) The existence of this deterministic one-to-one function allows us to relate $h(\hat{\mathbf{y}}_{\mathcal{J}} | s_1)$ to $h(\mathbf{g}, \mathbf{s}_{[2:D]} | s_1) = h(\mathbf{g}, \mathbf{s}_{[2:D]})$ in a simple way (see Lemma 5.3 below).
- (iii) This relation will turn out to be sufficient for our purposes as using that conditioning reduces entropy we can link $h(\mathbf{y})$ to $h(\hat{\mathbf{y}}_{\mathcal{J}} | s_1)$ according to (5.25) below.

We now describe the details of the proof program outlined above.

Lemma 5.2. *Assume that the matrix \mathbf{P} satisfies the conditions of Theorem 5.1 and take the submatrix $\tilde{\mathbf{P}}$ defined in Theorem 5.1 to consist of the first $M + 1$ rows of \mathbf{P} for simplicity.⁸ Let*

$$\begin{aligned} \mathcal{J} \triangleq & [1 : D] \cup [D + 1 : D + M + 1] \\ & \cup [2D + 1 : 2D + M + 1] \cup \dots \\ & \cup [(M - 1)D + 1 : (M - 1)D + M + 1] \end{aligned} \quad (5.10)$$

where $|\mathcal{J}| = D - 1 + M^2$, and consider the vector-valued function $\hat{\mathbf{y}}_{\mathcal{J}} : \mathbb{C}^{D-1+M^2} \rightarrow \mathbb{C}^{D-1+M^2}$

$$\hat{\mathbf{y}}_{\mathcal{J}}(\mathbf{g}, \mathbf{s}_{[2:D]}) = ((\mathbf{I}_M \otimes \mathbf{S}\mathbf{P}) \mathbf{g})_{\mathcal{J}} \quad (5.11)$$

parametrized by $s_1 \neq 0$. To simplify the notation we will not indicate this parametrization explicitly. The function $\hat{\mathbf{y}}_{\mathcal{J}}(\cdot)$ is one-to-one almost everywhere (a.e.) on \mathbb{C}^{D-1+M^2} .

Proof. We need to show that the function $\hat{\mathbf{y}}_{\mathcal{J}}(\mathbf{g}, \mathbf{s}_{[2:D]})$ is one-to-one a.e. Hence, we can exclude sets of measure zero from its domain. In particular, we shall consider the restriction of the function $\hat{\mathbf{y}}_{\mathcal{J}}(\mathbf{g}, \mathbf{s}_{[2:D]})$ to the set of pairs $(\mathbf{g}, \mathbf{s}_{[2:D]})$, which satisfy

- (i) $0 < |s_i| < \infty$ for all $i \in [2:D]$;
- (ii) the matrix \mathbf{G} , defined by

$$\mathbf{G}^T = [\mathbf{g}_1 \ \dots \ \mathbf{g}_M] \quad (5.12)$$

is invertible [here $[\mathbf{g}_1 \ \dots \ \mathbf{g}_M]$ is related to \mathbf{g} according to (5.1)];

⁸This assumption will be made in the remainder of the chapter, without explicitly mentioning it again.

(iii) the sum $\sum_{m=1}^M g_{1m} p_{jm}$ is nonzero for all $j \in [M+2:D]$.

To show that this restriction of the function $\hat{\mathbf{y}}_{\mathcal{J}}(\cdot)$ [which, with slight abuse of notation we still call $\hat{\mathbf{y}}_{\mathcal{J}}(\cdot)$] is one-to-one, we take an element $\tilde{\mathbf{y}}$ from its range and prove that the equation

$$\hat{\mathbf{y}}_{\mathcal{J}}(\mathbf{g}', \mathbf{s}'_{[2:D]}) = \tilde{\mathbf{y}} \quad (5.13)$$

has a unique solution in the set of pairs $(\mathbf{g}', \mathbf{s}'_{[2:D]})$ satisfying the constraints (i)–(iii). We do this in two steps. First, we demonstrate that equation (5.13) is equivalent to a linear equation. Second, we show that this linear equation has a unique solution.

Step 1: The element $\tilde{\mathbf{y}}$ can be represented as

$$\tilde{\mathbf{y}} = ((\mathbf{I}_M \otimes \mathbf{S}\mathbf{P}) \mathbf{g})_{\mathcal{J}} \quad (5.14)$$

with

$$\mathbf{S} = \text{diag} \left(\begin{bmatrix} s_1 \\ \mathbf{s}_{[2:D]} \end{bmatrix} \right)$$

where $(\mathbf{g}, \mathbf{s}_{[2:D]})$ satisfies the constraints (i)–(iii). Hence, (5.13) can be rewritten in the following way

$$((\mathbf{I}_M \otimes \mathbf{S}'\mathbf{P}) \mathbf{g}')_{\mathcal{J}} = ((\mathbf{I}_M \otimes \mathbf{S}\mathbf{P}) \mathbf{g})_{\mathcal{J}} \quad (5.15)$$

with

$$\mathbf{S}' = \text{diag} \left(\begin{bmatrix} s_1 \\ \mathbf{s}'_{[2:D]} \end{bmatrix} \right).$$

To prove that (5.15) has a unique solution, we follow the approach described in Section 5.2 and convert (5.15) into a linear system of equations through a change of variables. In particular, thanks to constraint (i), we can multiply both sides of (5.15) by $[\mathbf{I}_M \otimes \mathbf{S}']_{\mathcal{J},\mathcal{J}}^{-1}$ and by $[\mathbf{I}_M \otimes \mathbf{S}]_{\mathcal{J},\mathcal{J}}^{-1}$ and perform the substitution $z'_i = 1/s'_i$, $i \in [2:D]$. We obtain that (5.15) is equivalent to

$$((\mathbf{I}_M \otimes \mathbf{S}^{-1}\mathbf{P}) \mathbf{g}')_{\mathcal{J}} = ((\mathbf{I}_M \otimes \mathbf{Z}'\mathbf{P}) \mathbf{g})_{\mathcal{J}} \quad (5.16)$$

where

$$\mathbf{Z}' = \text{diag} \left(\begin{bmatrix} 1/s'_1 \\ \mathbf{z}' \end{bmatrix} \right)$$

and $\mathbf{z}' = [z'_2 \cdots z'_D]^\top$. Using (Lütkepohl, 1996, Section 7.2, Equation 5), it can be easily shown that (5.16) holds if and only if

$$\left[[\mathbf{I}_M \otimes \mathbf{S}^{-1} \mathbf{P}]_{\mathcal{J}, \diamond} - [(\mathbf{G} \otimes \mathbf{I}_D) \mathbf{D}(\mathbf{P})]_{\mathcal{J}, \diamond} \right] \begin{bmatrix} \mathbf{g}' \\ 1/s'_1 \\ \mathbf{z}' \end{bmatrix} = \mathbf{0} \quad (5.17)$$

where the operator $\mathbf{D}(\cdot)$ is defined as

$$\mathbf{D}(\mathbf{P}) \triangleq \begin{bmatrix} \text{diag}([p_{11} \cdots p_{D1}]) \\ \vdots \\ \text{diag}([p_{1M} \cdots p_{DM}]) \end{bmatrix}. \quad (5.18)$$

Then, manipulate equation (5.17) such that all the unknowns are on one side of the equation and all the terms depending on the constant s'_1 are on the other side. This yield the following inhomogeneous linear system of equations

$$\mathbf{F} \begin{bmatrix} \mathbf{g}' \\ -\mathbf{z}' \end{bmatrix} = \frac{1}{s'_1} \mathbf{u} \quad (5.19)$$

where

$$\mathbf{F} = \left[[\mathbf{I}_M \otimes \mathbf{S}^{-1} \mathbf{P}]_{\mathcal{J}, \diamond} \quad [(\mathbf{G} \otimes \mathbf{I}_D) \mathbf{D}(\mathbf{P})]_{\mathcal{J}, [2:D]} \right] \quad (5.20)$$

$$\mathbf{u} = \left[\mathbf{p}_1^\top \mathbf{g}_1 \quad \underbrace{0 \cdots 0}_{D-1 \text{ times}} \quad \mathbf{p}_1^\top \mathbf{g}_2 \quad \underbrace{0 \cdots 0}_M \text{ times} \quad \cdots \quad \mathbf{p}_1^\top \mathbf{g}_M \quad \underbrace{0 \cdots 0}_M \text{ times} \right]^\top$$

and \mathbf{p}_1^\top denotes the first row of \mathbf{P} .

Step 2: The solution of (5.19) is unique if and only if $\det \mathbf{F} \neq 0$. Matrix \mathbf{F} depends on \mathbf{S} , \mathbf{G} , \mathbf{P} in a complicated way. The key to our analysis is to factorize \mathbf{F} into a product of terms, each of which can be easily shown to have a nonzero determinant.

As a consequence of the choice of \mathcal{J} in (5.10), each of the last $D - M - 1$ columns of \mathbf{F} has exactly one nonzero element. This allows

us to use the Laplace formula to expand the determinant along these columns iteratively to get

$$\begin{aligned}
 |\det \mathbf{F}| &= \left| \det \left[[\mathbf{I}_M \otimes \mathbf{S}^{-1} \mathbf{P}]_{\mathcal{J}, \diamond} \quad [(\mathbf{G} \otimes \mathbf{I}_D) \mathbf{D}(\mathbf{P})]_{\mathcal{J}, [2 : D]} \right] \right| \\
 &= \prod_{j \in [M+2 : D]} \left| \sum_{m=1}^M g_{1m} p_{jm} \right| |\det(\mathbf{E})| \tag{5.21}
 \end{aligned}$$

where

$$\begin{aligned}
 \mathbf{E} &= [\mathbf{I}_M \otimes \tilde{\mathbf{S}}^{-1} \tilde{\mathbf{P}} \quad (\mathbf{G} \otimes \mathbf{I}_{M+1}) \tilde{\mathbf{D}}] \\
 \tilde{\mathbf{D}} &= [\mathbf{D}(\tilde{\mathbf{P}})]_{\diamond, [2 : M+1]}, \quad \tilde{\mathbf{S}} = [\mathbf{S}]_{[1 : M+1], [1 : M+1]}.
 \end{aligned}$$

Next, we factorize \mathbf{E} into a product of simple terms and write

$$\begin{aligned}
 \mathbf{E} &= \left[(\mathbf{I}_M \otimes \tilde{\mathbf{S}}^{-1}) (\mathbf{I}_M \otimes \tilde{\mathbf{P}}) \quad (\mathbf{G} \otimes \mathbf{I}_{M+1}) \tilde{\mathbf{D}} \right] \\
 &= \underbrace{(\mathbf{I}_M \otimes \tilde{\mathbf{S}}^{-1})}_{\mathbf{M}_1(\mathbf{S})} \left[(\mathbf{I}_M \otimes \tilde{\mathbf{P}}) \quad (\mathbf{I}_M \otimes \tilde{\mathbf{S}}) (\mathbf{G} \otimes \mathbf{I}_{M+1}) \tilde{\mathbf{D}} \right] \\
 &= \mathbf{M}_1(\mathbf{S}) \left[(\mathbf{I}_M \otimes \tilde{\mathbf{P}}) \quad (\mathbf{G} \otimes \mathbf{I}_{M+1}) (\mathbf{I}_M \otimes \tilde{\mathbf{S}}) \tilde{\mathbf{D}} \right] \\
 &= \mathbf{M}_1(\mathbf{S}) \left[(\mathbf{I}_M \otimes \tilde{\mathbf{P}}) \quad (\mathbf{G} \otimes \mathbf{I}_{M+1}) \tilde{\mathbf{D}} \quad \underbrace{[\tilde{\mathbf{S}}]_{[2 : M+1], [2 : M+1]}} \right] \\
 &= \mathbf{M}_1(\mathbf{S}) \left[(\mathbf{I}_M \otimes \tilde{\mathbf{P}}) \quad (\mathbf{G} \otimes \mathbf{I}_{M+1}) \tilde{\mathbf{D}} \quad \underbrace{\begin{bmatrix} \mathbf{I}_{M^2} & \mathbf{0} \\ \mathbf{0} & [\tilde{\mathbf{S}}]_{[2 : M+1], [2 : M+1]} \end{bmatrix}}_{\mathbf{M}_2(\mathbf{S})} \right] \tag{5.22}
 \end{aligned}$$

where the first three equalities are consequences of simple properties of the Kronecker product and in the fourth equality we used the block-diagonal structure of $\tilde{\mathbf{D}}$. Next, again using simple properties of the Kronecker product, we can write

$$\left[(\mathbf{I}_M \otimes \tilde{\mathbf{P}}) \quad (\mathbf{G} \otimes \mathbf{I}_{M+1}) \tilde{\mathbf{D}} \right]$$

$$\begin{aligned}
 &= (\mathbf{G} \otimes \mathbf{I}_{M+1}) \left[(\mathbf{G}^{-1} \otimes \mathbf{I}_{M+1}) \begin{pmatrix} \mathbf{I}_M \otimes \tilde{\mathbf{P}} & \tilde{\mathbf{D}} \end{pmatrix} \right] \\
 &= (\mathbf{G} \otimes \mathbf{I}_{M+1}) \left[\begin{pmatrix} \mathbf{I}_M \otimes \tilde{\mathbf{P}} & \tilde{\mathbf{D}} \end{pmatrix} (\mathbf{G}^{-1} \otimes \mathbf{I}_M) \right] \\
 &= \underbrace{(\mathbf{G} \otimes \mathbf{I}_{M+1})}_{\mathbf{M}_3(\mathbf{G})} \underbrace{\begin{pmatrix} \mathbf{I}_M \otimes \tilde{\mathbf{P}} & \tilde{\mathbf{D}} \end{pmatrix}}_{\mathbf{M}_4(\mathbf{P})} \underbrace{\begin{bmatrix} (\mathbf{G}^{-1} \otimes \mathbf{I}_M) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_M \end{bmatrix}}_{\mathbf{M}_5(\mathbf{G})}. \quad (5.23)
 \end{aligned}$$

Inserting (5.23) into (5.22) and then inserting the result into (5.21) and using the multiplicativity of the determinant we obtain the desired factorization

$$\begin{aligned}
 |\det \mathbf{F}| &= \prod_{j \in [M+2 : D]} \left| \sum_{m=1}^M g_{1m} p_{jm} \right| |\det \mathbf{M}_1(\mathbf{S})| |\det \mathbf{M}_2(\mathbf{S})| \\
 &\quad \times |\det \mathbf{M}_3(\mathbf{G})| |\det \mathbf{M}_4(\mathbf{P})| |\det \mathbf{M}_5(\mathbf{G})|. \quad (5.24)
 \end{aligned}$$

The first factor in this product is nonzero by constraint (iii); the second and the third factors are nonzero by constraint (i); the fourth and the sixth factors are nonzero by constraint (ii). The fifth factor in the product depends only on \mathbf{P} , a parameter of the problem. It is curious to find out that the structure of $\mathbf{M}_4(\mathbf{P})$ is such that Property (A) guarantees that $|\det \mathbf{M}_4(\mathbf{P})| \neq 0$. This fact from linear algebra is proven in Appendix G. Therefore $|\det \mathbf{F}| \neq 0$ and the proof is completed. \square

The following comments on Lemma 5.2 are in order.

Remark 5.4. For $D = 3$ and $M = 2$ as in the simple example in Section 5.2, $\mathcal{J} = [1 : 6]$, so that $\hat{\mathbf{y}}_{\mathcal{J}} = \hat{\mathbf{y}}$. Therefore, the one-to-one correspondence established in this lemma simply means that (5.4) has a unique solution for fixed $s_1 \neq 0$. Note also, that the matrix \mathbf{B} defined in (5.6) is related to \mathbf{F} in (5.20) according to $\mathbf{B} = (\mathbf{I}_2 \otimes \mathbf{S})\mathbf{F}$, and therefore $\det \mathbf{B} \neq 0$ a.e., as claimed in Section 5.2.

Remark 5.5. For the proof of the lemma, it is crucial that $s_1 \neq 0$ is fixed. In fact, one can check that if none of the components of \mathbf{s} is fixed, the resulting equivalent of the function $\hat{\mathbf{y}}_{\mathcal{J}}(\cdot)$ cannot be

one-to-one, no matter how the set \mathcal{J} is chosen. Fixing $s_1 \neq 0$ in order to make the function $\hat{\mathbf{y}}_{\mathcal{J}}(\cdot)$ be one-to-one corresponds to transmitting a pilot, as done in the simple example in Section 5.2 by setting $s_1 = 1$. The cardinality of the set \mathcal{J} , which determines the lower bound on the capacity pre-log, as we shall see below, is dictated by the requirement that $\hat{\mathbf{y}}_{\mathcal{J}}$ and $[\mathbf{g}^{\top} \mathbf{s}_{[2:D]}^{\top}]^{\top}$ are of the same dimension, which implies that \mathcal{J} must contain $D - 1 + M^2$ elements. The specific choice of \mathcal{J} in (5.10) is arbitrary to a certain extent and is made this way to simplify the proof of the lemma as much as possible.

Lemma 5.2 can be used to relate the conditional differential entropy $h(\hat{\mathbf{y}}_{\mathcal{J}} | s_1)$ to $h(\mathbf{g}, \mathbf{s}_{[2:D]})$. Before doing so, we establish a simple lower bound on $h(\mathbf{y})$ that is explicit in $h(\hat{\mathbf{y}}_{\mathcal{J}} | s_1)$. Let \mathcal{N} be the complement of \mathcal{J} in $[1:MD]$. Then

$$\begin{aligned} h(\mathbf{y}) &= h(\mathbf{y}_{\mathcal{J}}, \mathbf{y}_{\mathcal{N}}) = h(\mathbf{y}_{\mathcal{J}}) + h(\mathbf{y}_{\mathcal{N}} | \mathbf{y}_{\mathcal{J}}) \\ &\geq h(\sqrt{\text{SNR}}\hat{\mathbf{y}}_{\mathcal{J}} + \mathbf{w}_{\mathcal{J}}) + h(\mathbf{y}_{\mathcal{N}} | \mathbf{g}, \mathbf{s}, \mathbf{y}_{\mathcal{J}}) \\ &\geq h(\sqrt{\text{SNR}}\hat{\mathbf{y}}_{\mathcal{J}} + \mathbf{w}_{\mathcal{J}} | \mathbf{w}_{\mathcal{J}}) + h(\mathbf{w}_{\mathcal{N}}) \\ &\geq |\mathcal{J}| \log(\text{SNR}) + h(\hat{\mathbf{y}}_{\mathcal{J}} | s_1) + c. \end{aligned} \tag{5.25}$$

Through this chain of inequalities, we got rid of the noise \mathbf{w} . This corresponds to considering the noise-free IO relation (5.4) in the intuitive explanation given in Section 5.2. Inserting $|\mathcal{J}| = D - 1 + M^2$ into (5.25), we obtain the desired result (5.9) provided that $h(\hat{\mathbf{y}}_{\mathcal{J}} | s_1)$ is finite, which will be proved by means of the following lemma.

Lemma 5.3 (Transformation of differential entropy). *Assume that $\mathbf{f} : \mathbb{C}^N \rightarrow \mathbb{C}^N$ is a continuous vector-valued function that is one-to-one a.e. on \mathbb{C}^N . Let $\mathbf{u} \in \mathbb{C}^N$ be a random vector and $\mathbf{v} = \mathbf{f}(\mathbf{u})$. Then*

$$h(\mathbf{v}) = h(\mathbf{u}) + 2 \mathbb{E}_{\mathbf{u}}[\log |\det(\partial \mathbf{f} / \partial \mathbf{u})|].$$

Proof. The proof follows from the change-of-variable theorem for integrals (Rudin, 1987, Theorem 7.26). \square

Let $f_{s_1}(x)$ denote the density of s_1 . Then

$$\begin{aligned} \mathbf{h}(\hat{\mathbf{y}}_{\mathcal{J}} | s_1) &= \int f_{s_1}(x) \mathbf{h}(\hat{\mathbf{y}}_{\mathcal{J}} | s_1 = x) dx \\ &= \mathbf{h}(\mathbf{g}, \mathbf{s}_{[2:D]} | s_1) + 2 \mathbb{E}_{\mathbf{g}, \mathbf{s}} \left[\log \left| \det \frac{\partial \hat{\mathbf{y}}_{\mathcal{J}}}{\partial (\mathbf{g}, \mathbf{s}_{[2:D]})} \right| \right] \end{aligned} \quad (5.26)$$

where in the second equality we applied Lemma 5.3 to $\mathbf{h}(\hat{\mathbf{y}}_{\mathcal{J}} | s_1 = x)$, using that the function $\hat{\mathbf{y}}_{\mathcal{J}}(\cdot)$ in (5.11) is continuous and is one-to-one a.e. as shown in Lemma 5.2. The first term on the right-hand side of (5.26) satisfies

$$\mathbf{h}(\mathbf{g}, \mathbf{s}_{[2:D]} | s_1) = \mathbf{h}(\mathbf{g}) + \mathbf{h}(\mathbf{s}_{[2:D]}) > -\infty$$

where the inequality follows because the $s_i, i \in [2:D]$, are i.i.d. and have finite differential entropy. It therefore remains to show that the second term on the right-hand side of (5.26) is finite as well. As the right-hand side of (5.11) is linear in \mathbf{g} , we have that

$$\frac{\partial \hat{\mathbf{y}}_{\mathcal{J}}}{\partial \mathbf{g}} = [\mathbf{I}_M \otimes \mathbf{S}\mathbf{P}]_{\mathcal{J}, \diamond}.$$

Furthermore, using (Lütkepohl, 1996, Section 7.2, Equation 5), the right-hand side of (5.11) can be rewritten as

$$\hat{\mathbf{y}}_{\mathcal{J}}(\mathbf{g}, \mathbf{s}_{[2:D]}) = ((\mathbf{G} \otimes \mathbf{I}_D) \mathbf{D}(\mathbf{P}) \mathbf{s})_{\mathcal{J}}. \quad (5.27)$$

Hence, we have that

$$\frac{\partial \hat{\mathbf{y}}_{\mathcal{J}}}{\partial \mathbf{s}_{[2:D]}} = [(\mathbf{G} \otimes \mathbf{I}_D) \mathbf{D}(\mathbf{P})]_{\mathcal{J}, [2:D]}.$$

To summarize, the Jacobian matrix in (5.26) is given by

$$\frac{\partial \hat{\mathbf{y}}_{\mathcal{J}}}{\partial (\mathbf{g}, \mathbf{s}_{[2:D]})} = \left[[\mathbf{I}_M \otimes \mathbf{S}\mathbf{P}]_{\mathcal{J}, \diamond} \quad [(\mathbf{G} \otimes \mathbf{I}_D) \mathbf{D}(\mathbf{P})]_{\mathcal{J}, [2:D]} \right]. \quad (5.28)$$

This Jacobian matrix has a structure similar to the matrix \mathbf{F} defined in (5.20) [the only difference is that \mathbf{S}^{-1} in \mathbf{F} is replaced by \mathbf{S} in (5.28)].

Therefore, we can factorize the determinant of this Jacobian matrix in exactly the same way as $\det \mathbf{F}$ and write

$$\begin{aligned} & \left| \det \left[[\mathbf{I}_M \otimes \mathbf{S}\mathbf{P}]_{\mathcal{J}, \diamond} \quad [(\mathbf{G} \otimes \mathbf{I}_D)\mathbf{D}(\mathbf{P})]_{\mathcal{J}, [2 : D]} \right] \right| \\ &= \prod_{j \in [M+2 : D]} \left| \sum_{m=1}^M g_{1m} p_{jm} \right| |\det \mathbf{M}_1(\mathbf{S}^{-1})| |\det \mathbf{M}_2(\mathbf{S}^{-1})| \\ & \quad \times |\det \mathbf{M}_3(\mathbf{G})| |\det \mathbf{M}_4(\mathbf{P})| |\det \mathbf{M}_5(\mathbf{G})|. \end{aligned} \quad (5.29)$$

Hence, we can rewrite the second term on the right-hand side of (5.26) as

$$\begin{aligned} & \mathbb{E}_{\mathbf{g}, \mathbf{s}} \left[\log \left| \det \frac{\partial \hat{\mathbf{y}}_{\mathcal{J}}}{\partial (\mathbf{g}, \mathbf{s}_{[2 : D]})} \right| \right] \\ &= \sum_{j \in [M+2 : D]} \mathbb{E}_{\mathbf{g}} \left[\log \left| \sum_{m=1}^M g_{1m} p_{jm} \right| \right] \\ & \quad + \mathbb{E}_{\mathbf{s}} [\log |\det \mathbf{M}_1(\mathbf{S}^{-1})|] + \mathbb{E}_{\mathbf{s}} [\log |\det \mathbf{M}_2(\mathbf{S}^{-1})|] \\ & \quad + \mathbb{E}_{\mathbf{g}} [\log |\det \mathbf{M}_3(\mathbf{G})|] + \log |\det \mathbf{M}_4(\mathbf{P})| \\ & \quad + \mathbb{E}_{\mathbf{g}} [\log |\det \mathbf{M}_5(\mathbf{G})|]. \end{aligned} \quad (5.30)$$

The first and the fourth term on the right-hand side of (5.30) are finite because \mathbf{G} has i.i.d. Gaussian components. The sixth term is finite for the same reason, because $\log |\det \mathbf{G}^{-1}| = -\log |\det \mathbf{G}|$. The second and the third term are finite because $\mathbb{E} [\log |s_i|] > -\infty, i \in [1 : D]$, by assumption. Finally, as we already mentioned, we show in Appendix G that the matrix $\mathbf{M}_3(\mathbf{P})$ has full rank if Property (A) is satisfied. This then implies that the fifth term on the right-hand side of (5.30) is also finite.

5.4. SUMMARY OF RESULTS

In this chapter, we analyzed the noncoherent-capacity pre-log of a temporally correlated block-fading channel. We showed that, surprisingly, the capacity pre-log in the SIMO case can be larger than that in

5. NONCOHERENT SIMO CHANNEL

the SISO case. This result was established for the special case of the number of receive antennas being equal to the rank of the channel covariance matrix. Interesting open issues include extending the lower bound in Theorem 5.1 to an arbitrary number of receive antennas and finding a tight upper bound on the capacity pre-log.

CHAPTER 6

Conclusions and Outlook

In this thesis, we explored how relays can be used to improve spectral efficiency in wireless interference networks and we investigated the impact of the absence of CSI on the capacity of wireless point-to-point fading channels.

The results in Chapter 2 demonstrate that it is possible to use relay terminals to make an interference network decouple and “crystallize”, provided there is CSI at the relays. “Crystallization” means that the network breaks up into a set of effectively isolated “wires in the air”, i.e., the links between all source-destination terminal pairs in the network become nonfading and signal power dominates interference power in every link. We found that the protocols P1 and P2 trade off the required (for the network to crystallize) rate of growth of the number of relay terminals for the amount of CSI at the relay level.

Interpreting our results in terms of capacity scaling, we found that in a network with n nodes, P1 achieves $\mathcal{O}(1/n^{2/3})$ per-node throughput whereas P2 realizes $\mathcal{O}(1/\sqrt{n})$. The scaling law for P2 is exactly the same as the behavior established by Gupta and Kumar (2002). However, it is interesting to observe that the interference management technique employed by P2 is completely different from the one used by Gupta and Kumar. Gupta and Kumar manage the interference in the network by carefully adjusting the transmit powers. They exploit the fact that if the signal attenuation is sufficiently high, then users situated far away do not interfere with each other and,

6. CONCLUSIONS AND OUTLOOK

therefore, can transmit (to the nearby users) simultaneously. Protocol P2 does not require any assumptions on the path loss behavior and handles interference using match-filtering and coherent combining at the relay level. This technique, however, requires each relay to know all its backward and all its forward channels.

Özgür et al. (2007) have shown that $\mathcal{O}(1)$ throughput can be achieved in an interference network using hierarchical cooperation. The protocol developed by Özgür et al. manages interference as follows. First, it carefully adjusts the power of local transmissions (i.e., the transmissions to the nearby users), just like in (Gupta and Kumar, 2002). Second, it uses the local transmissions to emulate collocated antenna arrays by exchanging information between the nearby users and to create large-scale virtual MIMO links; it then benefits from the spatial multiplexing gain provided by these links. In order to sustain the required number of local transmissions simultaneously, the signal attenuation should be sufficiently strong. The use of the large-scale virtual MIMO links requires enough CSI: each node must know all the channels in the network. Such global CSI is difficult to acquire and to maintain. An interesting problem, therefore, is to determine the network capacity scaling behavior in the noncoherent case.

As discussed above, to achieve the optimal capacity scaling behavior in the coherent case, Özgür et al. (2007) relied heavily on the concept of distributed large-scale virtual MIMO links. The only known upper bound in network information theory is the cut-set bound; applied to a wireless network this bound is equal to the capacity of the corresponding MIMO channel, where nodes on one side of the cut form the receiver and nodes on the other side of the cut form the transmitter. These two observations suggest that in order to make progress in understanding the capacity of noncoherent networks, one has to understand the capacity behavior of the noncoherent MIMO point-to-point channel first. Motivated by this idea, we started to look at the capacity of noncoherent point-to-point channels.

Our results in Chapter 4 demonstrate that in underspread WSSUS SISO channels (most wireless channels are highly underspread), the absence of CSI does not play a significant role for all SNR values of

practical interest, at least in terms of channel capacity. It, however, remains unclear what the situation in MIMO channels is. It seems difficult to answer this question using the general WSSUS model, because the gap between a lower bound based on the ideas from Chapter 4 and the corresponding AWGN upper bound widens as the number of antennas increases. This is due to the fact that the AWGN upper bound is insensitive to CSI uncertainty, which becomes important in multi-antenna channels even when the spread is small. To get first insights into the capacity behavior of point-to-point channels with multiple antennas, we decided to use the much simpler correlated block-fading model in Chapter 5.

The result reported in Chapter 5 is surprising: the capacity pre-log of the noncoherent SIMO channel is strictly larger than the capacity pre-log of the corresponding noncoherent SISO channel, an effect that does not exist in the coherent case. Many interesting problems related to the work in Chapter 5 remain:

1. Generalizing the lower bound found in Chapter 5 to the case where the number of receive antennas is not equal to the rank of the correlation matrix of the channel.
2. Finding the corresponding matching upper bound.
3. Generalizing the SIMO results to the MIMO case.

If the solution to these three problems is found, it will become possible to analyze the hierarchical cooperation protocol of Özgür et al. in the noncoherent setting. It will then be interesting to see, what the noncoherent capacity scaling behavior achieved by the protocol is, whether $\mathcal{O}(1)$ throughput is achievable, and how the capacity scaling behavior depends on the rank of the channel correlation matrix.

APPENDIX A

Truncation of Random Variables and Large Deviations

We start by recalling the famous Hoeffding inequality along with an important variation that will be central for our developments.

Theorem A.1 (Hoeffding, 1963). *Let X_1, X_2, \dots, X_N be independent real-valued RVs and $A_n \leq X_n \leq B_n$ for $n \in [1:N]$. Let $S_N = \sum_{n=1}^N X_n$. Then,*

$$\mathbb{P}\{S_N - \mathbb{E}[S_N] \geq Nx\} \leq \exp\left(-\frac{2N^2x^2}{\sum_{n=1}^N (B_n - A_n)^2}\right).$$

Theorem A.2 (Maurer, 2003). *Let X_1, X_2, \dots, X_N be independent real-valued RVs with $X_n \geq 0$ and $\mathbb{E}[X_n^2] < \infty$ for $n \in [1:N]$. Let $S_N = \sum_{n=1}^N X_n$. Then,*

$$\mathbb{P}\{S_N - \mathbb{E}[S_N] \leq -Nx\} \leq \exp\left(-\frac{N^2x^2}{2\sum_{n=1}^N \mathbb{E}[X_n^2]}\right).$$

The following theorem builds on the Hoeffding inequality (Theorem A.1) and constitutes the core of the truncation technique.

Theorem A.3. *Assume the following on a common probability space.*

- The real-valued RVs X_1, X_2, \dots, X_N (possibly dependent) have marginal distribution functions $F_{X_n}(x)$, $n \in [1:N]$. The tails of these distributions are exponentially decaying uniformly in n , i.e., there exist $B > 0$, $\alpha > 0$, $\beta > 0$, and $x_0 > 0$ such that for $x \geq x_0 > 0$ and $n \in [1:N]$

$$\mathbb{P}\{|X_n| \geq x\} = 1 - F_{X_n}(x) + F_{X_n}(-x) \leq Be^{-\alpha x^\beta}. \quad (A.1)$$

- The real-valued RVs $\phi_1, \phi_2, \dots, \phi_N$ are jointly independent and satisfy

$$-1 \leq \phi_n \leq 1, \quad \mathbb{E}[\phi_n] = 0, \quad n \in [1:N].$$

- The real-valued deterministic nonnegative coefficients A_1, A_2, \dots, A_N are uniformly bounded from above, i.e., there exists a constant A independent of n such that

$$0 \leq A_n \leq A, \quad n \in [1:N].$$

- The set of RVs $\{X_n\}_{n=1}^N$ is independent of the set $\{\phi_n\}_{n=1}^N$.

Let $S_N = \sum_{n=1}^N A_n X_n \phi_n$. Then, for all $N > 0$ and $x > 0$ such that $x \geq x_0^{(2+\beta)/2}$

$$\mathbb{P}\{|S_N| \geq \sqrt{N}x\} \leq 2 \max[2, NB] \exp\left(-\min\left[\frac{1}{2A^2}, \alpha\right] x^{\frac{2\beta}{2+\beta}}\right). \quad (A.2)$$

Proof. The proof is based on the idea of truncation of the RVs X_n . We start by fixing N and choosing t such that $(Nt^2)^\gamma \geq x_0$. The truncation parameter $0 < \gamma < 1$ will be chosen later. Next, we truncate the RVs X_n , $n \in [1:N]$, according to

$$\hat{X}_n \triangleq X_n I[|X_n| \leq (Nt^2)^\gamma].$$

Define $\hat{S}_N \triangleq \sum_{n=1}^N A_n \hat{X}_n \phi_n$. Note that the independence of $\{X_n\}_{n=1}^N$ and $\{\phi_n\}_{n=1}^N$ and the condition $\mathbb{E}[\phi_n] = 0$ ($n \in [1:N]$) implies that $\mathbb{E}[S_N] = \mathbb{E}[\hat{S}_N] = 0$. Let \mathcal{I}_n denote the event that X_n is equal to its truncated version, i.e., $\mathcal{I}_n \triangleq \{X_n = \hat{X}_n\}$ and, \mathcal{I}_n^c the event that $X_n \neq$

\hat{X}_n , i.e., $\mathcal{I}_n^c \triangleq \{X_n \neq \hat{X}_n\}$. With these definitions, distinguishing the events where either all X_n are equal to their truncated version, i.e., $\bigcap_{n=1}^N \mathcal{I}_n$ and where at least one of the X_n is not equal to its truncated version, i.e., $\bigcup_{n=1}^N \mathcal{I}_n^c$, we get

$$\begin{aligned}
\mathbb{P}\{|S_N| \geq Nt\} &= \mathbb{P}\left\{|S_N| \geq Nt \mid \bigcap_{n=1}^N \mathcal{I}_n\right\} \mathbb{P}\left\{\bigcap_{n=1}^N \mathcal{I}_n\right\} \\
&\quad + \mathbb{P}\left\{|S_N| \geq Nt \mid \bigcup_{n=1}^N \mathcal{I}_n^c\right\} \mathbb{P}\left\{\bigcup_{n=1}^N \mathcal{I}_n^c\right\} \\
&= \mathbb{P}\left\{|\hat{S}_N| \geq Nt\right\} \mathbb{P}\left\{\bigcap_{n=1}^N \mathcal{I}_n\right\} \\
&\quad + \mathbb{P}\left\{|S_N| \geq Nt \mid \bigcup_{n=1}^N \mathcal{I}_n^c\right\} \mathbb{P}\left\{\bigcup_{n=1}^N \mathcal{I}_n^c\right\} \\
&\leq \mathbb{P}\left\{|\hat{S}_N| \geq Nt\right\} + \sum_{n=1}^N \mathbb{P}\{\mathcal{I}_n^c\} \tag{A.3}
\end{aligned}$$

where the last step follows by using the trivial bounds

$$\mathbb{P}\left\{\bigcap_{n=1}^N \mathcal{I}_n\right\} \leq 1, \quad \mathbb{P}\left\{|S_N| \geq Nt \mid \bigcup_{n=1}^N \mathcal{I}_n^c\right\} \leq 1$$

and applying the union bound to $\mathbb{P}\{\bigcup_{n=1}^N \mathcal{I}_n^c\}$. Since $-1 \leq \phi_n \leq 1$, $n \in [1:N]$, we obtain the following bounds for the individual terms in \hat{S}_N

$$-A_n(Nt^2)^\gamma \leq A_n \hat{X}_n \phi_n \leq A_n(Nt^2)^\gamma, \quad n \in [1:N].$$

Moreover, owing to the independence of the ϕ_n , conditioned on the set $\mathcal{X} \triangleq \{\hat{X}_1, \hat{X}_2, \dots, \hat{X}_N\}$, the RVs $A_n \hat{X}_n \phi_n$ are independent. Therefore, using Bayes' rule and the Hoeffding inequality (Theorem A.1), noting that $\mathbb{E}[\hat{S}_N | \mathcal{X}] = 0$, we can conclude that

$$\mathbb{P}\left\{|\hat{S}_N| \geq Nt\right\} = \mathbb{E}_{\mathcal{X}} \left[\mathbb{P}\left\{|\hat{S}_N - \mathbb{E}[\hat{S}_N | \mathcal{X}]| \geq Nt \mid \mathcal{X}\right\} \right]$$

$$\begin{aligned}
&\leq 2 \exp\left(-\frac{N^2 t^2}{2 \sum_{n=1}^N A_n^2 (Nt^2)^{2\gamma}}\right) \\
&\leq 2 \exp\left(-\frac{(Nt^2)^{1-2\gamma}}{2A^2}\right). \tag{A.4}
\end{aligned}$$

Next, using (A.1), and assuming [this will be justified in (A.6)] that $(Nt^2)^\gamma \geq x_0$, we have

$$\begin{aligned}
\mathbb{P}\{\mathcal{I}_n^c\} &= \mathbb{P}\{X_n \neq \hat{X}_n\} \\
&= \mathbb{P}\{|X_n| \geq (Nt^2)^\gamma\} \leq B e^{-\alpha(Nt^2)^{\gamma\beta}}. \tag{A.5}
\end{aligned}$$

To get the fastest possible exponential decay in (A.3), we need to choose the free parameter γ such that it maximizes $\min[1 - 2\gamma, \gamma\beta]$, which is the solution that makes the exponents of t in (A.4) and (A.5) equal and is given by $\gamma = 1/(2 + \beta)$. Finally, setting $t = x/\sqrt{N}$ results in

$$(Nt^2)^\gamma = x^{2\gamma} = x^{2/(2+\beta)} \geq x_0 \tag{A.6}$$

as required. Combining (A.3), (A.4) and (A.5), we finally obtain

$$\mathbb{P}\{|S_N| \geq \sqrt{N}x\} \leq 2 \exp\left(-\frac{1}{2A^2} x^{\frac{2\beta}{2+\beta}}\right) + NB \exp\left(-\alpha x^{\frac{2\beta}{2+\beta}}\right). \tag{A.7}$$

The final result (A.2) is a trivial upper bound to (A.7). \square

The following corollary is the generalization of Theorem A.3 to the complex-valued case and will be used repeatedly in the proofs of Theorems 2.1 and 2.2.

Corollary A.4. *Assume the following on a common probability space.*

- *The absolute values of the complex-valued (possibly dependent) RVs X_1, X_2, \dots, X_N have marginal distribution functions $F_{X_n}(x)$, $n \in [1:N]$. The tails of these distributions are exponentially decaying uniformly in n , i.e., there exist $B > 0$, $\alpha > 0$, $\beta > 0$ and $x_0 > 0$ such that for $x \geq x_0 > 0$ and $n \in [1:N]$*

$$\mathbb{P}\{|X_n| \geq x\} = 1 - F_{X_n}(x) \leq B e^{-\alpha x^\beta}. \tag{A.8}$$

- The real-valued RVs $\phi_1, \phi_2, \dots, \phi_N$ are jointly independent and satisfy $\phi_n \sim \mathcal{U}(-\pi, \pi)$ and hence $\mathbb{E}[e^{i\phi_n}] = 0$ for all $n \in [1:N]$.
- The real-valued deterministic nonnegative coefficients A_1, A_2, \dots, A_N are uniformly bounded from above, i.e., there exists a constant A independent of n such that

$$0 \leq A_n \leq A, \quad n \in [1:N].$$

- The set of RVs $\{X_n\}_{n=1}^N$ is independent of the set $\{\phi_n\}_{n=1}^N$.

Let $S_N = \sum_{n=1}^N A_n X_n e^{i\phi_n}$. Then, for all $N > 0$ and $x > 0$ such that $x \geq x_0^{(2+\beta)/2}$

$$\mathbb{P}\left\{|S_N| \geq \sqrt{N}x\right\} \leq 4 \max[2, NB] \exp\left(-\min\left[\frac{1}{2A^2}, \alpha\right] 2^{-\frac{\beta}{\beta+2}} x^{\frac{2\beta}{\beta+2}}\right).$$

Proof. Apply Theorem A.3 to $\Re S_N$ and $\Im S_N$ separately and combine the two bounds using the Pythagorean union bound (Lemma B.2). \square

The following corollary is a modification of Theorem A.3 for the case of independent nonnegative RVs and will be used repeatedly in the proofs of Theorems 2.1 and 2.2.

Corollary A.5. *Assume the following on a common probability space.*

- The real-valued nonnegative RVs X_1, X_2, \dots, X_N are jointly independent and have marginal distribution functions $F_{X_n}(x)$, $n \in [1:N]$. The right tails of these distributions are exponentially decaying uniformly in n , i.e., there exist $B > 0$, $\alpha > 0$, $\beta > 0$ and $x_0 > 0$ such that for all $x \geq x_0 > 0$ and $n \in [1:N]$

$$\mathbb{P}\{X_n \geq x\} = 1 - F_{X_n}(x) \leq B e^{-\alpha x^\beta}. \quad (\text{A.9})$$

- The expectations $\mathbb{E}[X_n^2]$ are uniformly bounded from above, i.e., there exists a constant C independent of n such that

$$\mathbb{E}[X_n^2] \leq C, \quad n \in [1:N]. \quad (\text{A.10})$$

- The real-valued deterministic nonnegative coefficients A_1, A_2, \dots, A_N are uniformly bounded from above, i.e., there exists a constant A independent of n such that

$$0 \leq A_n \leq A, \quad n \in [1:N]. \quad (\text{A.11})$$

Let $S_N = \sum_{n=1}^N A_n X_n$. Then, for all $N > 0$ and $x > 0$ such that $x \geq x_0^{(2+\beta)/2}$

$$\begin{aligned} & \mathbb{P}\left\{|S_N - \mathbb{E}[S_N]| \geq \sqrt{N}x\right\} \\ & \leq 3 \max[1, NB] \exp\left(-\min\left[\frac{2}{A^2}, \alpha, \frac{1}{2A^2C}\right] x^{\frac{2\beta}{\beta+2}}\right). \end{aligned} \quad (\text{A.12})$$

Proof. The proof idea of this corollary is similar to that used in Theorem A.3. However, there are several technical details, which do not occur in the proof of Theorem A.3. We have, therefore, decided to present the full version of the proof of Corollary A.5.

Unlike in the proof of Theorem A.3, here we have $\mathbb{E}[S_N] \neq 0$. To obtain an upper bound on $\mathbb{P}\{|S_N - \mathbb{E}[S_N]| \geq \sqrt{N}x\}$, we establish an upper bound on $\mathbb{P}\{S_N \geq \mathbb{E}[S_N] + \sqrt{N}x\}$ and on $\mathbb{P}\{S_N \leq \mathbb{E}[S_N] - \sqrt{N}x\}$ and use the union bound to combine the results.

We start by deriving an upper bound on $\mathbb{P}\{S_N \geq \mathbb{E}[S_N] + \sqrt{N}x\}$. Following the same steps as in the proof of Theorem A.3, we define the truncation parameter $0 < \gamma < 1$, which will be chosen later. Fix N and choose t such that $(Nt^2)^\gamma \geq x_0$. We truncate the RVs X_n ($n \in [1:N]$) according to

$$\hat{X}_n \triangleq X_n I[X_n \leq (Nt^2)^\gamma]$$

and define $\hat{S}_N \triangleq \sum_{n=1}^N A_n \hat{X}_n$. It is easily seen that $\mathbb{E}[S_N] \geq \mathbb{E}[\hat{S}_N]$ and therefore

$$\mathbb{P}\{S_N \geq \mathbb{E}[S_N] + Nt\} \leq \mathbb{P}\{S_N \geq \mathbb{E}[\hat{S}_N] + Nt\}. \quad (\text{A.13})$$

Let \mathcal{I}_n denote the event that X_n is equal to its truncated version, i.e., $\mathcal{I}_n \triangleq \{X_n = \hat{X}_n\}$, and \mathcal{I}_n^c the event that $X_n \neq \hat{X}_n$, i.e., $\mathcal{I}_n^c \triangleq \{X_n \neq \hat{X}_n\}$. With these definitions, distinguishing the events where either all X_n are equal to their truncated version, i.e., $\bigcap_{n=1}^N \mathcal{I}_n$ and where at least one of the X_n is not equal to its truncated version, i.e.,

$\bigcup_{n=1}^N \mathcal{I}_n^c$, we get

$$\begin{aligned}
& \mathbb{P}\left\{S_N \geq \mathbb{E}\left[\hat{S}_N\right] + Nt\right\} \\
&= \mathbb{P}\left\{S_N \geq \mathbb{E}\left[\hat{S}_N\right] + Nt \mid \bigcap_{n=1}^N \mathcal{I}_n\right\} \mathbb{P}\left\{\bigcap_{n=1}^N \mathcal{I}_n\right\} \\
&\quad + \mathbb{P}\left\{S_N \geq \mathbb{E}\left[\hat{S}_N\right] + Nt \mid \bigcup_{n=1}^N \mathcal{I}_n^c\right\} \mathbb{P}\left\{\bigcup_{n=1}^N \mathcal{I}_n^c\right\} \\
&= \mathbb{P}\left\{\hat{S}_N \geq \mathbb{E}\left[\hat{S}_N\right] + Nt\right\} \mathbb{P}\left\{\bigcap_{n=1}^N \mathcal{I}_n\right\} \\
&\quad + \mathbb{P}\left\{S_N \geq \mathbb{E}\left[\hat{S}_N\right] + Nt \mid \bigcup_{n=1}^N \mathcal{I}_n^c\right\} \mathbb{P}\left\{\bigcup_{n=1}^N \mathcal{I}_n^c\right\} \\
&\leq \mathbb{P}\left\{\hat{S}_N \geq \mathbb{E}\left[\hat{S}_N\right] + Nt\right\} + \sum_{n=1}^N \mathbb{P}\{\mathcal{I}_n^c\} \tag{A.14}
\end{aligned}$$

where the last step is obtained by using the trivial bounds

$$\mathbb{P}\left\{\bigcap_{n=1}^N \mathcal{I}_n\right\} \leq 1, \quad \mathbb{P}\left\{S_N \geq \mathbb{E}\left[\hat{S}_N\right] + Nt \mid \bigcup_{n=1}^N \mathcal{I}_n^c\right\} \leq 1$$

and applying the union bound to $\mathbb{P}\{\bigcup_{n=1}^N \mathcal{I}_n^c\}$. The individual terms in \hat{S}_N are bounded according to

$$0 \leq A_n \hat{X}_n \leq A_n (Nt^2)^\gamma, \quad n \in [1:N].$$

Using Bayes' rule and the Hoeffding inequality (Theorem A.1), we can conclude that

$$\begin{aligned}
\mathbb{P}\left\{\hat{S}_N \geq \mathbb{E}\left[\hat{S}_N\right] + Nt\right\} &\leq \exp\left(-\frac{2N^2t^2}{\sum_{n=1}^N A_n^2 (Nt^2)^{2\gamma}}\right) \\
&\leq \exp\left(-\frac{2(Nt^2)^{1-2\gamma}}{A^2}\right). \tag{A.15}
\end{aligned}$$

Next, using (A.9), and assuming [this will be justified in (A.17)] that $(Nt^2)^\gamma \geq x_0$, we have

$$\begin{aligned} \mathbb{P}\{\mathcal{I}_n^c\} &= \mathbb{P}\{X_n \neq \hat{X}_n\} \\ &= \mathbb{P}\{X_n \geq (Nt^2)^\gamma\} \leq B e^{-\alpha(Nt^2)^{\gamma\beta}}. \end{aligned} \quad (\text{A.16})$$

To get the fastest possible exponential decay in (A.14), we need to choose the free parameter γ such that it maximizes $\min[1 - 2\gamma, \gamma\beta]$, which is the solution that makes the exponents of t in (A.15) and (A.16) equal and is given by $\gamma = 1/(2 + \beta)$. Finally, setting $t = x/\sqrt{N}$ results in

$$(Nt^2)^\gamma = x^{2\gamma} = x^{2/(2+\beta)} \geq x_0 \quad (\text{A.17})$$

as required. Combining (A.13)–(A.16), we obtain

$$\begin{aligned} \mathbb{P}\{S_N \geq \mathbb{E}[S_N] + \sqrt{N}x\} \\ \leq \exp\left(-\frac{2}{A^2}x^{\frac{2\beta}{2+\beta}}\right) + NB \exp\left(-\alpha x^{\frac{2\beta}{2+\beta}}\right). \end{aligned} \quad (\text{A.18})$$

It remains to establish an upper bound on $\mathbb{P}\{S_N \leq \mathbb{E}[S_N] - \sqrt{N}x\}$. From Theorem A.2 it follows that

$$\mathbb{P}\{S_N \leq \mathbb{E}[S_N] - \sqrt{N}x\} \leq \exp\left(-\frac{Nx^2}{2\sum_{n=1}^N \mathbb{E}[A_n^2 X_n^2]}\right)$$

which, using (A.10) and (A.11), can be further upper-bounded as

$$\mathbb{P}\{S_N \leq \mathbb{E}[S_N] - \sqrt{N}x\} \leq \exp\left(-\frac{x^2}{2A^2C}\right). \quad (\text{A.19})$$

Combining (A.18) and (A.19) and using the union bound, we obtain

$$\begin{aligned} \mathbb{P}\{|S_N - \mathbb{E}[S_N]| \geq \sqrt{N}x\} &\leq \exp\left(-\frac{2}{A^2}x^{\frac{2\beta}{2+\beta}}\right) \\ &\quad + NB \exp\left(-\alpha x^{\frac{2\beta}{2+\beta}}\right) + \exp\left(-\frac{x^2}{2A^2C}\right). \end{aligned} \quad (\text{A.20})$$

The final result (A.12) is a trivial upper bound to (A.20). \square

APPENDIX B

Union Bounds

In this appendix, as a reference, we present several variations of union bounds for probability that we use frequently throughout the Chapter 2.

Lemma B.1 (Union bound for sums). *Assume the complex-valued RVs X_1, X_2, \dots, X_N are such that*

$$\mathbb{P}\left\{|X_n| \geq C_n\right\} \leq P_n, \quad n \in [1:N]$$

where C_1, C_2, \dots, C_N and P_1, P_2, \dots, P_N are fixed positive constants. Then,

$$\mathbb{P}\left\{\left|\sum_{n=1}^N X_n\right| \geq \sum_{n=1}^N C_n\right\} \leq \sum_{n=1}^N P_n.$$

Proof. Let \mathcal{A}_n denote the event that $|X_n| \geq C_n$, $n \in [1:N]$. Let \mathcal{B} denote the event that $|\sum_{n=1}^N X_n| \geq \sum_{n=1}^N C_n$. By inspection, it follows that $\mathcal{B} \Rightarrow \bigcup_{n=1}^N \mathcal{A}_n$, which implies $\mathbb{P}\{\mathcal{B}\} \leq \sum_{n=1}^N \mathbb{P}\{\mathcal{A}_n\}$. \square

The proofs of the remaining union bounds follow exactly the same pattern as the proof of Lemma B.1 and will hence be omitted.

Lemma B.2 (Pythagorean union bound). *Assume the complex-valued RV X is such that*

$$\mathbb{P}\left\{|\Re X| \geq C_R\right\} \leq P_R \quad \text{and} \quad \mathbb{P}\left\{|\Im X| \geq C_I\right\} \leq P_I$$

B. UNION BOUNDS

where C_R, C_I, P_R , and P_I are fixed positive constants. Then,

$$\mathbb{P}\left\{|X| \geq \sqrt{C_R^2 + C_I^2}\right\} \leq P_R + P_I.$$

Lemma B.3 (Union bound for mixed sums). *Assume that the complex-valued RVs X_1, X_2, \dots, X_N are such that*

$$\mathbb{P}\left\{|X_n| \geq C_n\right\} \leq P_n, \quad n \in [1:N]$$

where C_1, C_2, \dots, C_N and P_1, P_2, \dots, P_N are fixed positive constants; then, the following statements hold.

1. *If the real-valued RVs $X'_1, X'_2, \dots, X'_{N'}$ are such that*

$$\mathbb{P}\left\{X'_n \leq C'_n\right\} \leq P'_n, \quad n \in [1:N]$$

where $C'_1, C'_2, \dots, C'_{N'}$ and $P'_1, P'_2, \dots, P'_{N'}$ are fixed positive constants, then

$$\begin{aligned} \mathbb{P}\left\{\left|\sum_{n=1}^N X_n + \sum_{n=1}^{N'} X'_n\right| \leq \max\left[0, \sum_{n=1}^{N'} C'_n - \sum_{n=1}^N C_n\right]\right\} \\ \leq \sum_{n=1}^N P_n + \sum_{n=1}^{N'} P'_n. \end{aligned}$$

2. *If the real-valued RVs $X'_1, X'_2, \dots, X'_{N'}$ are such that*

$$\mathbb{P}\left\{X'_n \geq C'_n\right\} \leq P'_n, \quad n \in [1:N]$$

then,

$$\mathbb{P}\left\{\left|\sum_{n=1}^N X_n + \sum_{n=1}^{N'} X'_n\right| \geq \sum_{n=1}^{N'} C'_n + \sum_{n=1}^N C_n\right\} \leq \sum_{n=1}^N P_n + \sum_{n=1}^{N'} P'_n.$$

Lemma B.4 (Union bound for products). *Assume the complex-valued RVs X_1, X_2, \dots, X_N are such that*

$$\mathbb{P}\left\{|X_n| \geq C_n\right\} \leq P_n, \quad n \in [1:N]$$

where C_1, C_2, \dots, C_N and P_1, P_2, \dots, P_N are fixed positive constants. Then,

$$\mathbb{P}\left\{\left|\prod_{n=1}^N X_n\right| \geq \prod_{n=1}^N C_n\right\} \leq \sum_{n=1}^N P_n.$$

Lemma B.5 (Union bound for fractions). *If for real-valued positive RVs X_1 and X_2 and positive constants C_1, C_2 and P_1, P_2*

$$\mathbb{P}\left\{X_1 \geq C_1\right\} \leq P_1 \quad \text{and} \quad \mathbb{P}\left\{X_2 \leq C_2\right\} \leq P_2$$

then

$$\mathbb{P}\left\{X_1/X_2 \geq C_1/C_2\right\} \leq P_1 + P_2.$$

If, in turn,

$$\mathbb{P}\left\{X_1 \leq C_1\right\} \leq P_1 \quad \text{and} \quad \mathbb{P}\left\{X_2 \geq C_2\right\} \leq P_2$$

then

$$\mathbb{P}\left\{X_1/X_2 \leq C_1/C_2\right\} \leq P_1 + P_2.$$

APPENDIX C

Technical Results from Chapter 2

C.1. PROOF OF THEOREM 2.1

We start by recalling that we want to establish a concentration result for $\text{SINR}_m^{\text{P1}}$, given by (2.22), using the truncation technique throughout. As already mentioned, this entails establishing the large-deviations behavior of $S^{(1)}$, $S^{(2)}$, $S^{(3)}$, and $S^{(4)}$. For $S^{(3)}$, this has already been done in Section 2.3.3.B. It remains to establish the corresponding (based on the truncation technique) concentration results for $S^{(1)}$, $S^{(2)}$, and $S^{(4)}$ defined by (2.23), (2.24), and (2.26), respectively.

C.1.1. Analysis of $S^{(1)}$

The sum $S^{(1)}$ can be written as

$$S^{(1)} = \sum_{k:p(k)=m} C_{\text{P1},k}^{m,m} Z_k^{(1)} \quad (\text{C.1})$$

with

$$Z_k^{(1)} \triangleq |f_{m,k}| |h_{k,m}|.$$

For every $k \in [1:K]$ such that $p(k) = m$, we have $\mathbb{E}[Z_k^{(1)}] = \pi/4$ and $\mathbb{E}[(Z_k^{(1)})^2] = 1$. Application of the union bound for products yields

$$\mathbb{P}\left\{Z_k^{(1)} \geq x\right\} \leq 2e^{-x}, \quad x \geq 0.$$

Noting that the sum $S^{(1)}$ contains K/M terms, which are jointly independent, taking into account (2.12), and using Corollary A.5, we get for $x \geq 0$ and $K/M \geq 1$

$$\mathbb{P} \left\{ \left| S^{(1)} - \frac{\pi}{4} \sum_{k:p(k)=m} C_{P1,k}^{m,m} \right| \geq \sqrt{\frac{K}{M}} x \right\} \leq 6 \frac{K}{M} e^{-\Delta^{(1)} x^{2/3}}$$

with $\Delta^{(1)} = \min[1, 1/(2\bar{C}^2)]$. Finally, using (2.12), it follows that

$$\mathbb{P} \left\{ S^{(1)} \geq \frac{\pi}{4} \bar{C} \frac{K}{M} + \sqrt{\frac{K}{M}} x \right\} \leq 6 \frac{K}{M} e^{-\Delta^{(1)} x^{2/3}} \quad (\text{C.2})$$

and

$$\mathbb{P} \left\{ S^{(1)} \leq \frac{\pi}{4} \underline{C} \frac{K}{M} - \sqrt{\frac{K}{M}} x \right\} \leq 6 \frac{K}{M} e^{-\Delta^{(1)} x^{2/3}} \quad (\text{C.3})$$

for all $x \geq 0$ and $K/M \geq 1$.

C.1.2. Analysis of $S^{(2)}$

The sum $S^{(2)}$ can be written as

$$S^{(2)} = \sum_{k:p(k) \neq m} C_{P1,k}^{m,m} Z_k^{(2)}$$

with

$$Z_k^{(2)} \triangleq \tilde{f}_{p(k),k}^* f_{m,k} \tilde{h}_{k,p(k)}^* h_{k,m}.$$

For every $k \in [1 : K]$ such that $p(k) \neq m$, we have $\mathbb{E}[Z_k^{(2)}] = 0$. Application of the union bound for products yields

$$\mathbb{P} \left\{ |Z_k^{(2)}| \geq x \right\} \leq 2e^{-x}, \quad x \geq 0.$$

Noting that the sum $S^{(2)}$ contains $K(M-1)/M$ terms, which are jointly independent, taking into account (2.12), and using Corollary A.4, we get for $x \geq 0$ and $K(M-1)/M \geq 1$

$$\mathbb{P} \left\{ |S^{(2)}| \geq \sqrt{\frac{K(M-1)}{M}} x \right\} \leq 8 \frac{K(M-1)}{M} e^{-\Delta^{(2)} x^{2/3}} \quad (\text{C.4})$$

with $\Delta^{(2)} = 2^{-\frac{1}{3}} \min[1, 1/(2\bar{C}^2)]$.

C.1.3. Analysis of $S^{(4)}$

The sum $S^{(4)}$ can be written as

$$S^{(4)} = \sum_{k=1}^K (C_{\text{P1},k}^m)^2 Z_k^{(4)}$$

with

$$Z_k^{(4)} = |f_{m,k}|^2.$$

Since $Z_k^{(4)}$ is exponentially distributed with parameter $\lambda = 1$, we have

$$\mathbb{P}\left\{Z_k^{(4)} \geq x\right\} \leq e^{-x}, \quad k \in [1:K], \quad x \geq 0.$$

Noting that the sum $S^{(4)}$ contains K jointly independent terms, taking into account (2.13) and using

$$\mathbb{E}\left[Z_k^{(4)}\right] = 1 \quad \text{and} \quad \mathbb{E}\left[(Z_k^{(4)})^2\right] = 2, \quad k \in [1:K]$$

we get for $x \geq 0$ and $K \geq 1$

$$\mathbb{P}\left\{\left|S^{(4)} - \sum_{k=1}^K (C_{\text{P1},k}^m)^2\right| \geq \sqrt{K}x\right\} \leq 3Ke^{-\Delta^{(4)}x^{2/3}}$$

with $\Delta^{(4)} = \min[1, 1/(4\bar{c}^4)]$. Therefore, using (2.13), it follows that

$$\mathbb{P}\left\{S^{(4)} \geq K\bar{c}^2 + \sqrt{K}x\right\} \leq 3Ke^{-\Delta^{(4)}x^{2/3}} \quad (\text{C.5})$$

and

$$\mathbb{P}\left\{S^{(4)} \leq K\underline{c}^2 - \sqrt{K}x\right\} \leq 3Ke^{-\Delta^{(4)}x^{2/3}}. \quad (\text{C.6})$$

We are now ready to carry out the final Step v of the program outlined in the first paragraph of Section 2.3.3. The concentration result for $\text{SINR}_m^{\text{P1}}$ is expressed in terms of upper bounds on $\mathbb{P}\{\text{SINR}_m^{\text{P1}} \geq$

\hat{U}_{P1} and $\mathbb{P}\{\text{SINR}_m^{P1} \leq \hat{L}_{P1}\}$, where the exact form of \hat{U}_{P1} and \hat{L}_{P1} is specified below.

To establish an upper bound on $\mathbb{P}\{\text{SINR}_m^{P1} \geq \hat{U}_{P1}\}$, we proceed as follows:

1. Apply Part 2 of Lemma B.3 to (C.2) and (C.4) to establish a stochastic upper bound¹ for $|S^{(1)} + S^{(2)}|$.
2. Apply Part 1 of Lemma B.3 to (2.45), (2.52), and (2.53) to establish a stochastic lower bound² for $|S^{(3)}|$.
3. Apply Part 1 of Lemma B.3 to the result from Step 2) and (C.6) to establish a stochastic lower bound for $S^{(3)} + \sigma^2 MS^{(4)} + KM\sigma^2$.
4. Apply the union bound for fractions (Lemma B.5) to the stochastic upper bound from Step 1 and to the stochastic lower bound from Step 3 to establish the final result:

$$\mathbb{P}\left\{\text{SINR}_m^{P1} \geq \hat{U}_{P1}\right\} \leq P_{P1}^U \quad (\text{C.7})$$

with

$$\hat{U}_{P1} \triangleq \frac{\pi^2 \bar{C}^2}{16 \underline{C}_{\text{SN}}^2} \frac{K}{M^3} \frac{\hat{U}_{P1}^N}{\hat{U}_{P1}^D} \quad (\text{C.8})$$

and P_{P1}^U , \hat{U}_{P1}^N , and \hat{U}_{P1}^D defined as

$$\begin{aligned} P_{P1}^U &\triangleq 6 \frac{K}{M} e^{-\Delta^{(1)} x_1^{2/3}} + 8 \frac{K(M-1)}{M} e^{-\Delta^{(2)} x_2^{2/3}} \\ &\quad + 6(M-1)K e^{-\Delta^{(31)} x_{31}^{2/5}} \\ &\quad + 64 \frac{(K-1)K(M-1)^2}{M} e^{-\Delta^{(32)} x_{321}^{2/7}} \\ &\quad + 64 \frac{(K-1)K(M-1)}{M} e^{-\Delta^{(32)} x_{322}^{2/7}} + 3K e^{-\Delta^{(4)} x_4^{2/3}} \end{aligned} \quad (\text{C.9})$$

¹For a RV X , a “stochastic upper bound” in this context means a bound of the form $\mathbb{P}\{X \geq A\} \leq P$.

²For a RV X , a “stochastic lower bound” in this context means a bound of the form $\mathbb{P}\{X \leq A\} \leq P$.

$$\hat{U}_{P1}^N \triangleq \left(1 + \frac{4}{\underline{C}_\pi} \sqrt{\frac{M}{K}} x_1 + \frac{4}{\underline{C}_\pi} \sqrt{\frac{M(M-1)}{K}} x_2 \right)^2 \quad (\text{C.10})$$

$$\begin{aligned} \hat{U}_{P1}^D \triangleq \max & \left[0, \frac{\underline{C}^2}{\underline{C}_{\text{SN}}^2} \frac{M-1}{M} - \frac{1}{\underline{C}_{\text{SN}}^2} \frac{M-1}{M\sqrt{K}} x_{31} \right. \\ & - \frac{1}{\underline{C}_{\text{SN}}^2} \sqrt{\frac{(K-1)(M-1)^2}{KM^3}} x_{321} \\ & \left. - \frac{1}{\underline{C}_{\text{SN}}^2} \sqrt{\frac{(K-1)(M-1)}{KM^3}} x_{322} \right] \\ & + \frac{\sigma^2}{\underline{C}_{\text{SN}}^2} \max \left[0, \underline{c}^2 - \frac{1}{\sqrt{K}} x_4 \right] + \frac{\sigma^2}{\underline{C}_{\text{SN}}^2} \end{aligned} \quad (\text{C.11})$$

An upper bound on $\mathbb{P}\{\text{SINR}_m^{\text{P1}} \leq \hat{L}_{P1}\}$ can be obtained as follows:

1. Apply Part 1 of Lemma B.3 to (C.3) and (C.4) to establish a stochastic lower bound for $|S^{(1)} + S^{(2)}|$.
2. Apply Part 2 of Lemma B.3 to (2.44), (2.52), and (2.53) to establish a stochastic upper bound for $|S^{(3)}|$.
3. Apply Part 2 of Lemma B.3 to the result from Step 2 and to (C.5) to establish a stochastic upper bound for $S^{(3)} + \sigma^2 MS^{(4)} + KM\sigma^2$.
4. Apply the union bound for fractions to the stochastic lower bound from Step 1 and to the stochastic upper bound from Step 3 to establish the final result:

$$\mathbb{P}\left\{\text{SINR}_m^{\text{P1}} \leq \hat{L}_{P1}\right\} \leq P_{P1}^U \quad (\text{C.12})$$

with

$$\hat{L}_{P1} \triangleq \frac{\pi^2}{16} \frac{\underline{C}^2}{\underline{C}_{\text{SN}}^2} \frac{K}{M^3} \frac{\hat{L}_{P1}^N}{\hat{L}_{P1}^D} \quad (\text{C.13})$$

and \hat{L}_{P1}^N and \hat{L}_{P1}^D defined as

$$\hat{L}_{P1}^N \triangleq \max \left[0, 1 - \frac{4}{\underline{C}\pi} \sqrt{\frac{M}{K}} x_1 - \frac{4}{\underline{C}\pi} \sqrt{\frac{M(M-1)}{K}} x_2 \right]^2 \quad (\text{C.14})$$

$$\begin{aligned} \hat{L}_{P1}^D \triangleq & \frac{\bar{C}^2}{\underline{C}_{\text{SN}}^2} \frac{M-1}{M} + \frac{1}{\underline{C}_{\text{SN}}^2} \frac{M-1}{M\sqrt{K}} x_{31} \\ & + \frac{1}{\underline{C}_{\text{SN}}^2} \sqrt{\frac{(K-1)(M-1)^2}{KM^3}} x_{321} \\ & + \frac{1}{\underline{C}_{\text{SN}}^2} \sqrt{\frac{(K-1)(M-1)}{KM^3}} x_{322} \\ & + \frac{\sigma^2}{\underline{C}_{\text{SN}}^2} \left(\bar{c}^2 + \frac{1}{\sqrt{K}} x_4 \right) + \frac{\sigma^2}{\underline{C}_{\text{SN}}^2}. \end{aligned} \quad (\text{C.15})$$

The result presented in Theorem 2.1 is a simpler and slightly weaker form of the bounds (C.7) and (C.12). To obtain this simplification we proceed as follows. Set

$$x_1 = x_2 = x_{31} = x_4 = x_{321} = x_{322} = x$$

in (C.9), (C.10), (C.11), (C.14) and (C.15). Note that in this case $L_{P1}(x) \leq \hat{L}_{P1}(x)$ and $U_{P1}(x) \geq \hat{U}_{P1}(x)$ and therefore

$$\mathbb{P}\left\{\text{SINR}_m^{\text{P1}} \geq U_{P1}\right\} \leq \mathbb{P}\left\{\text{SINR}_m^{\text{P1}} \geq \hat{U}_{P1}\right\} \leq P_{P1}^U \quad (\text{C.16})$$

$$\mathbb{P}\left\{\text{SINR}_m^{\text{P1}} \leq L_{P1}\right\} \leq \mathbb{P}\left\{\text{SINR}_m^{\text{P1}} \leq \hat{L}_{P1}\right\} \leq P_{P1}^U. \quad (\text{C.17})$$

Finally, combine the bounds (C.16) and (C.17) according to

$$\begin{aligned} & \mathbb{P}\left\{(\text{SINR}_m^{\text{P1}} \geq U_{P1}) \cup (\text{SINR}_m^{\text{P1}} \leq L_{P1})\right\} \\ & \leq \mathbb{P}\left\{\text{SINR}_m^{\text{P1}} \geq U_{P1}\right\} + \mathbb{P}\left\{\text{SINR}_m^{\text{P1}} \leq L_{P1}\right\} \leq 2P_{P1}^U \end{aligned}$$

and note that $2P_{P1}^U$ is upper bounded by the right-hand side of (2.55). ■

C.2. PROOF OF LOWER BOUND IN THEOREM 2.3

As already mentioned in Chapter 2, the proof of the lower bound in (2.87) is based on the technique summarized in Appendix C.3. After straightforward algebra, it follows that the IO relation of the SISO channel between the terminals \mathcal{S}_m and \mathcal{D}_m ($m \in [1:M]$) is given by

$$y_m = \left(\bar{F}_m + \tilde{F}_m \right) s_m + W_m$$

where

$$\bar{F}_m \triangleq \frac{1}{\sqrt{Q}} \sum_{q=1}^Q \mathbb{E}[a_q^{m,m}]$$

$$\tilde{F}_m \triangleq \frac{1}{\sqrt{Q}} \sum_{q=1}^Q (a_q^{m,m} - \mathbb{E}[a_q^{m,m}])$$

$$W_m \triangleq \sum_{\hat{m} \neq m} s_{\hat{m}} \frac{1}{\sqrt{Q}} \sum_{q=1}^Q a_q^{m,\hat{m}} + \frac{1}{\sqrt{Q}} \sum_{q=1}^Q b_q^m \tilde{\mathbf{h}}_{q,p(q)}^H \mathbf{z}_q + w_m$$

and

$$\begin{aligned} a_q^{m,\hat{m}} &\triangleq C_{\text{P1},q}^{m,\hat{m}} \left(\tilde{\mathbf{f}}_{p(q),q}^H \mathbf{f}_{m,q} \right) \left(\tilde{\mathbf{h}}_{q,p(q)}^H \mathbf{h}_{q,\hat{m}} \right) \\ b_q^m &\triangleq C_{\text{P1},q}^m \left(\tilde{\mathbf{f}}_{p(q),q}^H \mathbf{f}_{m,q} \right) \\ C_{\text{P1},q}^{m,\hat{m}} &\triangleq \sqrt{Q} d_{\text{P1},q} \hat{P}_{m,q} \hat{E}_{q,\hat{m}} \\ C_{\text{P1},q}^m &\triangleq \sqrt{Q} d_{\text{P1},q} \hat{P}_{m,q}. \end{aligned}$$

It is not difficult, but tedious, to verify that

$$\bar{F}_m = \frac{\pi}{4} \frac{L^2}{\sqrt{Q}} \sum_{q:p(q)=m} C_{\text{P1},q}^{m,m} \tag{C.18}$$

$$\text{Var}[\tilde{F}_m] = \frac{L^2}{Q} \sum_{q:p(q) \neq m} \left(C_{\text{P1},q}^{m,m} \right)^2$$

$$+ \frac{(L + (\pi/4)(L-1)L)^2 - (\pi^2/16)L^4}{Q} \sum_{q:p(q)=m} \left(C_{P1,q}^{m,m}\right)^2 \quad (\text{C.19})$$

$$\begin{aligned} \text{Var}[W_m] &= \frac{L^2 + (\pi/4)L^2(L-1)}{QM} \sum_{\hat{m} \neq m} \sum_{q:p(q)=m} \left(C_{P1,q}^{m,\hat{m}}\right)^2 \\ &+ \frac{L^2 + (\pi/4)L^2(L-1)}{QM} \sum_{\hat{m} \neq m} \sum_{q:p(q)=\hat{m}} \left(C_{P1,q}^{m,\hat{m}}\right)^2 \\ &+ \frac{L^2}{QM} \sum_{\hat{m} \neq m} \sum_{\substack{q:p(q) \neq m \\ p(q) \neq \hat{m}}} \left(C_{P1,q}^{m,\hat{m}}\right)^2 \\ &+ \frac{L^2 + (\pi/4)L^2(L-1)}{Q} \sigma^2 \sum_{q:p(q)=m} \left(C_{P1,q}^m\right)^2 \\ &+ \frac{L^2}{Q} \sigma^2 \sum_{q:p(q) \neq m} \left(C_{P1,q}^m\right)^2 + \sigma^2. \end{aligned} \quad (\text{C.20})$$

Using (2.3), we lower-bound \bar{F}_m and upper-bound $\text{Var}[\tilde{F}_m]$ and $\text{Var}[W_m]$, substitute the resulting bounds into (C.22), and obtain

$$I(y_m; s_m) \geq \log\left(1 + \frac{\pi^2}{16} \frac{Q}{M^3} \underline{f}(M, L)\right) \quad (\text{C.21})$$

where

$$\underline{f}(M, L) = \frac{\underline{P} \underline{E} P_{\text{rel}} L^2}{\left(\bar{E} + \frac{\pi(L-1)}{4M} \bar{E} + \sigma^2\right) \left(\epsilon(M, L) + \bar{C}^2 + \sigma^2 \bar{c}^2 + \sigma^2\right)}$$

with

$$\begin{aligned} \epsilon(M, L) &= \frac{\bar{C}^2}{M} + \frac{(1 + (\pi/4)(L-1))^2 \bar{C}^2}{M^2} \\ &+ \frac{(1 + (\pi/4)(L-1)) (2\bar{C}^2 + \sigma^2 \bar{c}^2)}{M}. \end{aligned}$$

Finally, since L is finite, it follows by inspection that

$$\lim_{M \rightarrow \infty} \epsilon(M, L) = 0$$

and, therefore,

$$\lim_{M \rightarrow \infty} \underline{f}(M, L) = \frac{L^2 \underline{C}^2}{\underline{C}_{\text{SN}}^2}$$

which, together with (C.21), concludes the proof. ■

C.3. LOWER BOUND ON CHANNEL CAPACITY WITH IMPERFECT CHANNEL KNOWLEDGE

The following Lemma is obtained by recognizing that the expression in (Lapidoth and Shamai (Shitz), 2002, Equation (66)) is trivially a lower bound to $I(X; Y)$ in (C.22) below. For completeness, we present the result in the form needed in this thesis. For the proof of the (general) statement the interested reader is referred to (Lapidoth and Shamai (Shitz), 2002).

Lemma C.1. *Consider a SISO channel with IO relation*

$$Y = FX + W$$

where $X \sim \mathcal{CN}(0, \sigma_X^2)$, W is zero-mean noise³ with variance σ_W^2 , F is the random channel gain with variance σ_F^2 , and Y is the output of the channel. Assume that F can be decomposed as

$$F = \bar{F} + \tilde{F}$$

where $\bar{F} = \mathbb{E}[F]$ is known at the receiver and \tilde{F} with $\mathbb{E}[\tilde{F}] = 0$ is not known at the receiver. Assume that X is statistically independent⁴ of both F and W . Then, the mutual information $I(X; Y)$ can be lower-bounded as follows:

$$I(X; Y) \geq \log \left(1 + \frac{\bar{F}^2 \sigma_X^2}{\sigma_F^2 \sigma_X^2 + \sigma_W^2} \right). \quad (\text{C.22})$$

³In contrast to (Médard, 2000, Section III), the noise is not necessarily Gaussian.

⁴In (Médard, 2000, Section III), it is assumed that X , F , and W are statistically independent. The condition required here is weaker: F and W need not be statistically independent.

APPENDIX D

Some Essentials from Large Random-Matrix Theory

In this appendix, we briefly summarize the basic definitions and results from large random-matrix theory used in Chapter 3. An excellent tutorial on this subject is (Tulino and Verdú, 2004).

Definition D.1 (Stieltjes transform). Let $F(x)$ be a distribution function with density $f(x)$. The analytic function

$$G_F(z) \triangleq \int \frac{f(x)}{x-z} dx, \quad z \in \mathbb{C}^+$$

is called the Stieltjes transform of $F(x)$.

Lemma D.1 (Inversion formula). *Let $G_F(z)$ be the Stieltjes transform of a distribution function $F(x)$. The corresponding density function can be obtained as*

$$f(x) = \frac{1}{\pi} \lim_{y \rightarrow 0^+} \Im [G_F(x + iy)]. \quad (D.1)$$

Theorem D.2 (Silverstein, 1995). *Define the following quantities on a common probability space.*

- *The random matrix $\mathbf{A} \in \mathbb{C}^{N \times N'}$ has i.i.d. zero-mean entries with variance one.*

- The random matrix $\mathbf{B} \in \mathbb{C}^{N \times N}$ is Hermitian nonnegative definite with $F_{\mathbf{B}}^N(x)$, for $N \rightarrow \infty$, converging on $[0, \infty)$ a.s. to a non-random distribution function $F_{\mathbf{B}}(x)$ with corresponding density $f_{\mathbf{B}}(x)$.

Assume that the matrices \mathbf{A} and \mathbf{B} are statistically independent. Then, for $N, N' \rightarrow \infty$ with $N/N' \rightarrow \beta$,

$$F_{(1/N')\mathbf{A}\mathbf{A}^H\mathbf{B}}^N(x) \xrightarrow{\text{a.s.}} F_{(1/N')\mathbf{A}\mathbf{A}^H\mathbf{B}}(x)$$

with its Stieltjes transform $G_{F_{(1/N')\mathbf{A}\mathbf{A}^H\mathbf{B}}}(z)$ satisfying

$$G_{F_{(1/N')\mathbf{A}\mathbf{A}^H\mathbf{B}}}(z) = \int_{-\infty}^{\infty} \frac{f_{\mathbf{B}}(x)dx}{x(1 - \beta - \beta z G_{F_{(1/N')\mathbf{A}\mathbf{A}^H\mathbf{B}}}(z)) - z}, \quad z \in \mathbb{C}^+.$$

The solution of this fixed-point equation is unique in the set

$$\left\{ G_{F_{(1/N')\mathbf{A}\mathbf{A}^H\mathbf{B}}}(z) \in \mathbb{C} \mid -\frac{1 - \beta}{z} + \beta G_{F_{(1/N')\mathbf{A}\mathbf{A}^H\mathbf{B}}}(z) \in \mathbb{C}^+ \right\}.$$

We shall furthermore use the Marčenko-Pastur law as stated by Bai (1999).

Theorem D.3 (Marčenko and Pastur, 1967). *Assume that the matrix $\mathbf{A} \in \mathbb{C}^{N \times N'}$ has i.i.d. zero-mean entries with variance d^2 . Then, for $N, N' \rightarrow \infty$ with $N'/N \rightarrow \beta$, the ESD of $(1/N')\mathbf{A}\mathbf{A}^H$ converges a.s. to a limiting distribution function with density*

$$f_{(1/N')\mathbf{A}\mathbf{A}^H}(x) = \frac{\beta}{2\pi x d^2} \sqrt{(\gamma_2 - x)^+ (x - \gamma_1)^+} + [1 - \beta]^+ \delta(x)$$

where $\gamma_1 = d^2(1 - 1/\sqrt{\beta})^2$ and $\gamma_2 = d^2(1 + 1/\sqrt{\beta})^2$.

Under the same assumptions as in the first statement, if, in addition, the entries of \mathbf{A} have finite fourth moments, then a.s.

$$\begin{aligned} \lim_{N' \rightarrow \infty} \lambda_{\min} \left(\frac{1}{N'} \mathbf{A}\mathbf{A}^H \right) &= \gamma_1 \\ \lim_{N' \rightarrow \infty} \lambda_{\max} \left(\frac{1}{N'} \mathbf{A}\mathbf{A}^H \right) &= \gamma_2. \end{aligned}$$

APPENDIX E

A calculation for Chapter 3

In the following, we detail the computation of the integral

$$\hat{I} \triangleq \rho \int_{\eta_1}^{\eta_2} \frac{\sqrt{(\eta_2 - x)(x - \eta_1)} dx}{x(1-x)^2 \left(x \left(\frac{1-\beta}{z} - \beta G \right) - 1 \right)} \quad (\text{E.1})$$

on the right-hand side of (3.10). With the change of variables

$$t = \sqrt{\frac{x - \eta_1}{\eta_2 - x}} \quad (\text{E.2})$$

and the notation

$$\begin{aligned} \mu_1 &\triangleq 1 - \eta_1 & \nu_1 &\triangleq \eta_1 \left(\frac{1-\beta}{z} - \beta G \right) - 1 \\ \mu_2 &\triangleq 1 - \eta_2 & \nu_2 &\triangleq \eta_2 \left(\frac{1-\beta}{z} - \beta G \right) - 1 \end{aligned}$$

the integral \hat{I} can be written as

$$\hat{I} = 2(\eta_2 - \eta_1)^2 \rho \int_0^\infty \frac{t^2(t^2 + 1)dt}{(\eta_2 t^2 + \eta_1)(\mu_2 t^2 + \mu_1)^2(\nu_2 t^2 + \nu_1)}.$$

To simplify further, we introduce the notation

$$\kappa_1 \triangleq -\frac{\eta_1}{\eta_2}, \quad \kappa_2 \triangleq -\frac{\mu_1}{\mu_2}, \quad \kappa_3 \triangleq -\frac{\nu_1}{\nu_2}, \quad \chi \triangleq \frac{2(\eta_2 - \eta_1)^2}{\eta_2 \mu_2^2 \nu_2} \rho$$

so that

$$\hat{I} = \chi \int_0^{\infty} \frac{t^2(t^2 + 1)dt}{(t^2 - \kappa_1)(t^2 - \kappa_2)^2(t^2 - \kappa_3)}. \quad (\text{E.3})$$

Upon partial fraction expansion of the integrand in (E.3), we obtain

$$\hat{I} = \chi(A_1\hat{I}_1 + A_2\hat{I}_2 + A_3\hat{I}_3 + A_4\hat{I}_4) \quad (\text{E.4})$$

where

$$\begin{aligned} \hat{I}_1 &\triangleq \int_0^{\infty} \frac{dt}{t^2 - \kappa_1} & \hat{I}_2 &\triangleq \int_0^{\infty} \frac{dt}{(t^2 - \kappa_2)^2} \\ \hat{I}_3 &\triangleq \int_0^{\infty} \frac{dt}{t^2 - \kappa_2} & \hat{I}_4 &\triangleq \int_0^{\infty} \frac{dt}{t^2 - \kappa_3} \end{aligned} \quad (\text{E.5})$$

with

$$A_1 = \frac{\kappa_1(\kappa_1 + 1)}{(\kappa_1 - \kappa_2)^2(\kappa_1 - \kappa_3)} \quad (\text{E.6})$$

$$A_2 = \frac{\kappa_2(\kappa_2 + 1)}{(\kappa_2 - \kappa_1)(\kappa_2 - \kappa_3)} \quad (\text{E.7})$$

$$A_3 = \frac{-\kappa_2^2 - \kappa_1\kappa_2^2 + \kappa_1\kappa_3 + 2\kappa_1\kappa_2\kappa_3 - \kappa_2^2\kappa_3}{(\kappa_2 - \kappa_1)^2(\kappa_2 - \kappa_3)^2} \quad (\text{E.8})$$

$$A_4 = \frac{\kappa_3(\kappa_3 + 1)}{(\kappa_3 - \kappa_1)(\kappa_3 - \kappa_2)^2}. \quad (\text{E.9})$$

The integrals in (E.5) can be evaluated resulting in

$$\hat{I}_1 = \frac{1}{\sqrt{-\kappa_1}} \arctan \frac{t}{\sqrt{-\kappa_1}} \Big|_0^{\infty} = \frac{\pi}{2\sqrt{-\kappa_1}} \quad (\text{E.10})$$

$$\begin{aligned} \hat{I}_2 &= -\frac{t}{2\kappa_2(t^2 - \kappa_2)} \Big|_0^{\infty} - \frac{1}{2\kappa_2\sqrt{-\kappa_2}} \arctan \frac{t}{\sqrt{-\kappa_2}} \Big|_0^{\infty} \\ &= -\frac{\pi}{4\kappa_2\sqrt{-\kappa_2}} \end{aligned} \quad (\text{E.11})$$

$$\hat{I}_3 = \frac{1}{\sqrt{-\kappa_2}} \arctan \frac{t}{\sqrt{-\kappa_2}} \Big|_0^{\infty} = \frac{\pi}{2\sqrt{-\kappa_2}} \quad (\text{E.12})$$

$$\hat{I}_4 = \frac{1}{\sqrt{-\kappa_3}} \arctan \frac{t}{\sqrt{-\kappa_3}} \Big|_0^\infty = \frac{\pi}{2\sqrt{-\kappa_3}}. \quad (\text{E.13})$$

The quantity κ_3 is complex-valued, and the arctan and square root in (E.12) are understood as the principal values of these functions in \mathbb{C} as defined by Abramowitz and Stegun (1972).

Finally, by inspection, combining (E.10)–(E.13) with (E.6)–(E.9) and resubstituting the values of the parameters κ_1 , κ_2 , κ_3 , χ , ρ , μ_1 , μ_2 , η_1 , η_2 , ν_1 , ν_2 , γ_1 , and γ_2 , after straightforward but tedious simplifications, we find

$$\begin{aligned} \chi A_1 \hat{I}_1 &= \frac{(\sqrt{\beta} + 1) |\sqrt{\beta} - 1|}{2\beta} \\ \chi A_2 \hat{I}_2 &= -\frac{z}{\sqrt{\beta}(G\beta z + z + \beta - 1)} \\ \chi A_3 \hat{I}_3 &= -\frac{z d^2 (\sqrt{\beta} - 1)^2 (G\beta z + z + \beta - 1)}{2d^2 \beta (G\beta z + z + \beta - 1)^2} \\ &\quad + \frac{z\beta(G\beta z + \beta - 1)}{2d^2 \beta (G\beta z + z + \beta - 1)^2} \\ \chi A_4 \hat{I}_4 &= -\frac{(G\beta z + \beta - 1)}{2d^2 \beta (G\beta z + z + \beta - 1)^2} \\ &\quad \times \sqrt{\frac{d^2(G\beta z + z + \beta - 1)(\sqrt{\beta} - 1)^2 + z\beta}{d^2(G\beta z + z + \beta - 1)(\sqrt{\beta} + 1)^2 + z\beta}} \\ &\quad \times \left(d^2(G\beta z + z + \beta - 1) (\sqrt{\beta} + 1)^2 + z\beta \right). \end{aligned}$$

APPENDIX F

Technical Results from Chapter 4

F.1. AWGN-CAPACITY UPPER BOUND

To prove (4.14), we shall make use of (Wyner, 1966, Theorem 2). For the application of the theorem (see, in particular, Wyner, 1966, Equation (47c)), we need a lower bound on the average energy of $\hat{y}_f(t)$, after $\hat{y}_f(t)$ is passed through the time-limiting operator \mathbb{D}_D . To account for the time-dispersive nature of the channel \mathbb{H} , we shall replace \mathbb{D}_D with $\mathbb{D}_{D+2\tau_0}$. As τ_0 does not depend on D , such replacement does not affect capacity. The signal $(\mathbb{D}_{D+2\tau_0}\hat{y}_f)(t)$ is given by

$$\begin{aligned} & (\mathbb{D}_{D+2\tau_0}\hat{y}_f)(t) \\ &= \begin{cases} \iint_{\mathbb{R}^2} S_{\mathbb{H}}(\nu, \tau) [\mathbb{B}_{B+2\nu_0}(s(t-\tau)e^{i2\pi t\nu})] d\tau d\nu, & \text{if } |t| \leq D/2 + \tau_0 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

so that its average energy can be computed as

$$\begin{aligned} & \mathbb{E} [\|(\mathbb{D}_{D+2\tau_0}\hat{y}_f)(t)\|^2] \\ &= \iint_{\mathbb{R}^2} C_{\mathbb{H}}(\nu, \tau) \mathbb{E} \left[\int_{-D/2-\tau_0}^{D/2+\tau_0} |\mathbb{B}_{B+2\nu_0}(s(t-\tau)e^{i2\pi t\nu})|^2 dt \right] d\tau d\nu \end{aligned}$$

$$\geq \int_{-\nu_0}^{\nu_0} \int_{-\tau_0}^{\tau_0} C_{\mathbb{H}}(\nu, \tau) \mathbb{E} \left[\int_{-D/2-\tau_0}^{D/2+\tau_0} |\mathbb{B}_{B+2\nu_0}(s(t-\tau)e^{i2\pi t\nu})|^2 dt \right] d\tau d\nu$$

where the equality follows from the WSSUS property of \mathbb{H} [see (4.7)], and the inequality follows from the non-negativity of the integrand. As a consequence of the bandwidth constraint (4.3) and of the approximate time constraint (4.4) the input signal $s(t)$ is subject to, we have that, for all $(\nu, \tau) \in [-\nu_0, \nu_0] \times [-\tau_0, \tau_0]$,

$$\mathbb{E} \left[\int_{-D/2-\tau_0}^{D/2+\tau_0} |\mathbb{B}_{B+2\nu_0}(s(t-\tau)e^{i2\pi t\nu})|^2 dt \right] \geq (1-\eta) \mathbb{E} [\|s(t)\|^2].$$

Hence,

$$\begin{aligned} \mathbb{E} [\|(\mathbb{D}_{D+2\tau_0} \hat{y}_f)(t)\|^2] &\geq (1-\eta) \mathbb{E} [\|s(t)\|^2] \int_{-\nu_0}^{\nu_0} \int_{-\tau_0}^{\tau_0} C_{\mathbb{H}}(\nu, \tau) d\tau d\nu \\ &\geq (1-\eta)(1-\epsilon) \mathbb{E} [\|s(t)\|^2] \end{aligned}$$

where the last step follows from the underspread definition (4.9). We can now apply (Wyner, 1966, Theorem 2) and get

$$I_D(\hat{y}_f(t); y_f(t)) \leq (B+2\nu_0) \log \left(1 + \frac{(1-\eta)(1-\epsilon)P}{B+2\nu_0} \right) + (\eta+\epsilon-\eta\epsilon)P$$

which concludes the proof.

F.2. THE TRANSMIT SIGNAL IN (4.16) SATISFIES (4.4)

Let \mathbf{s} be the vector of dimension $\tilde{K}\tilde{N}$ obtained by stacking the data symbols $s[k, n]$ as in (4.25). Furthermore, let

$$d[k, n, l, m] = \int_{|t|>D/2} g_{k,n}(t) g_{l,m}^*(t) dt$$

and let \mathbf{D} be a square matrix of dimension $\tilde{K}\tilde{N} \times \tilde{K}\tilde{N}$ with entries

$$[\mathbf{D}]_{\tilde{m}+\tilde{l}\tilde{N}, \tilde{n}+\tilde{k}\tilde{N}} = d[\tilde{k} - K, \tilde{n} - N, \tilde{l} - K, \tilde{m} - N]$$

for $\tilde{k}, \tilde{l} \in [0 : \tilde{K} - 1]$ and $\tilde{n}, \tilde{m} \in [0 : \tilde{N} - 1]$. Note that \mathbf{D} is a Hermitian matrix. The average amount of energy of $s(t)$ [defined in (4.16)] outside an interval of length D can be expressed as a function of \mathbf{s} and \mathbf{D} as follows

$$\mathbb{E}[\|(\mathbb{I} - \mathbb{D}_D)s(t)\|^2] = \mathbf{s}^H \mathbf{D} \mathbf{s} \leq \lambda_{\max}(\mathbf{D}) \mathbb{E}[\|\mathbf{s}\|^2]$$

where the last step follows from the Rayleigh-Ritz theorem (Horn and Johnson, 1985, Theorem 4.2.2). We next use Geršgorin's disc theorem (Horn and Johnson, 1985, Corollary 6.1.5) to obtain an upper bound on $\lambda_{\max}(\mathbf{D})$ that is explicit in the entries of \mathbf{D} :

$$\lambda_{\max}(\mathbf{D}) \leq \max_{\substack{k \in [-K : K] \\ n \in [-N : N]}} \left[d[k, n, k, n] + \sum_{\substack{l=-K \\ (l,m) \neq (k,n)}}^K \sum_{m=-N}^N |d[k, n, l, m]| \right]. \quad (\text{F.1})$$

Our goal is to show that the right-hand side of (F.1) can be made smaller than η for an appropriate choice of K' . To establish this result, we bound each term on the right-hand side of (F.1) separately. For the first term we have that

$$\begin{aligned} d[k, n, k, n] &= \int_{|t| > D/2} |g_{k,n}(t)|^2 dt \\ &= \int_{(K+K'+1/2)T}^{\infty} |g(t - kT)|^2 dt + \int_{-\infty}^{-(K+K'+1/2)T} |g(t - kT)|^2 dt \\ &= \int_{(K+K'-k+1/2)T}^{\infty} |g(t)|^2 dt + \int_{-\infty}^{-(K+K'+k+1/2)T} |g(t)|^2 dt \\ &\leq \int_{(K'+1/2)T}^{\infty} |g(t)|^2 dt + \int_{-\infty}^{-(K'+1/2)T} |g(t)|^2 dt \end{aligned}$$

$$= \int_{|t| > (K'+1/2)T} |g(t)|^2 dt.$$

As $g(t) = \mathcal{O}(1/|t|^{1+\mu})$ by assumption, there exists a constant $\gamma > 0$ such that $|g(t)| < \gamma/|t|^{1+\mu}$ for $|t| \geq \gamma$. Hence, if we choose K' such that $K'T > \gamma$, we get

$$d[k, n, k, n] \leq \int_{|t| > (K'+1/2)T} \frac{\gamma^2}{|t|^{2(1+\mu)}} dt = 2 \int_{(K'+1/2)T}^{\infty} \frac{\gamma^2}{t^{2(1+\mu)}} dt.$$

We next upper-bound each of the terms $|d[k, n, l, m]|$ in (F.1), still under the assumption that $K'T > \gamma$.

$$\begin{aligned} |d[k, n, l, m]| &= \left| \int_{|t| > (K+K'+1/2)T} g_{k,n}(t) g_{l,m}^*(t) dt \right| \\ &\leq \int_{|t| > (K+K'+1/2)T} |g_{k,n}(t) g_{l,m}^*(t)| dt \\ &= \int_{|t| > (K+K'+1/2)T} |g(t - kT) g^*(t - lT)| dt \\ &\leq \gamma^2 \int_{|t| > (K+K'+1/2)T} \frac{1}{|t - kT|^{1+\mu}} \frac{1}{|t - lT|^{1+\mu}} dt \\ &= \gamma^2 \int_{(K+K'+1/2)T}^{\infty} \frac{1}{|t - kT|^{1+\mu}} \frac{1}{|t - lT|^{1+\mu}} dt \\ &\quad + \gamma^2 \int_{-\infty}^{-(K+K'+1/2)T} \frac{1}{|t - kT|^{1+\mu}} \frac{1}{|t - lT|^{1+\mu}} dt \\ &\leq \gamma^2 \int_{(K+K'+1/2)T}^{\infty} \frac{1}{|t - KT|^{1+\mu}} \frac{1}{|t - lT|^{1+\mu}} dt \end{aligned}$$

$$\begin{aligned}
 & + \gamma^2 \int_{-\infty}^{-(K+K'+1/2)T} \frac{1}{|t+KT|^{1+\mu}} \frac{1}{|t-lT|^{1+\mu}} dt \\
 \leq & \gamma^2 \int_{(K'+1/2)T}^{\infty} \frac{1}{|t|^{1+\mu}} \frac{1}{|t-(l-K)T|^{1+\mu}} dt \\
 & + \gamma^2 \int_{-\infty}^{-(K'+1/2)T} \frac{1}{|t|^{1+\mu}} \frac{1}{|t-(l+K)T|^{1+\mu}} dt.
 \end{aligned}$$

Note that for $t \geq 0$

$$\begin{aligned}
 \sum_{l=-K}^K \frac{1}{|t-(l-K)T|^{1+\mu}} &= \sum_{l=0}^{2K} \frac{1}{|t+lT|^{1+\mu}} \\
 &\leq \sum_{l=0}^{2K} \frac{1}{(lT)^{1+\mu}} \leq \sum_{l=0}^{\infty} \frac{1}{(lT)^{1+\mu}} \triangleq \gamma' < \infty
 \end{aligned}$$

where in the last step we used that $\mu > 0$ and, hence, the series converges. Similarly, for $t < 0$

$$\begin{aligned}
 \sum_{l=-K}^K \frac{1}{|t-(l+K)T|^{1+\mu}} &= \sum_{l=0}^{2K} \frac{1}{|t-lT|^{1+\mu}} \\
 &\leq \sum_{l=0}^{2K} \frac{1}{(lT)^{1+\mu}} \leq \sum_{l=0}^{\infty} \frac{1}{(lT)^{1+\mu}} = \gamma'.
 \end{aligned}$$

Putting all pieces together, we get,

$$\begin{aligned}
 & \sum_{l=-K}^K \sum_{\substack{m=-N \\ (l,m) \neq (k,n)}}^N |d[k, n, l, m]| \\
 \leq & \sum_{\substack{l=-K \\ (l,m) \neq (k,n)}}^K \sum_{m=-N}^N \gamma^2 \left[\int_{(K'+1/2)T}^{\infty} \frac{1}{|t|^{1+\mu}} \frac{1}{|t-(l-K)T|^{1+\mu}} dt \right.
 \end{aligned}$$

$$\begin{aligned}
 & + \left. \int_{-\infty}^{-(K'+1/2)T} \frac{1}{|t|^{1+\mu}} \frac{1}{|t - (l + K)T|^{1+\mu}} dt \right] \\
 \leq & (2N + 1)\gamma^2 \left[\int_{(K'+1/2)T}^{\infty} \frac{1}{|t|^{1+\mu}} \sum_{l=-K}^K \frac{1}{|t - (l - K)T|^{1+\mu}} dt \right. \\
 & \left. + \int_{-\infty}^{-(K'+1/2)T} \frac{1}{|t|^{1+\mu}} \sum_{l=-K}^K \frac{1}{|t - (l + K)T|^{1+\mu}} dt \right] \\
 \leq & 2(2N + 1)\gamma^2\gamma' \int_{(K'+1/2)T}^{\infty} \frac{1}{t^{1+\mu}} dt.
 \end{aligned}$$

To summarize, we obtained the following upper bound on the right-hand side of (F.1):

$$\lambda_{\max}(\mathbf{D}) \leq 2\gamma^2 \left[2 \int_{(K'+1/2)T}^{\infty} \frac{\gamma}{t^{2(1+\mu)}} dt + (2N + 1)\gamma' \int_{(K'+1/2)T}^{\infty} \frac{1}{t^{1+\mu}} dt \right].$$

The right-hand side of the inequality can be made arbitrarily small (for sufficiently large D) by choosing K' sufficiently large. In other words, for sufficiently large D , we can find a finite K' for which the right-hand side of the inequality falls below η .

F.3. SOME PROPERTIES OF THE AMBIGUITY FUNCTION

In this appendix, we summarize for completeness some useful properties of the ambiguity function defined in (4.19).

Property 1. For every function $g(t) \in \mathcal{L}^2(\mathbb{C})$, the ambiguity surface attains its maximum magnitude at the origin: $|A_g(\nu, \tau)|^2 \leq [A_g(0, 0)]^2 = \|g(t)\|^4$, for all ν and τ . This property follows directly

from the Cauchy-Schwarz inequality as shown in (Gröchenig, 2001, Lemma 4.2.1).

Property 2. Let $g(t) \in \mathcal{L}^2(\mathbb{C})$ and $e(t) = \sqrt{\beta}g(\beta t)$. Then

$$\begin{aligned} A_e(\nu, \tau) &= \int_{-\infty}^{\infty} e(t)e^*(t - \tau)e^{-i2\pi\nu t} dt \\ &= \beta \int_{-\infty}^{\infty} g(\beta t)g^*(\beta(t - \tau))e^{-i2\pi\nu t} dt \\ &\stackrel{(a)}{=} \int_{-\infty}^{\infty} g(z)g^*(z - \beta\tau)e^{-i2\pi\nu z/\beta} dz = A_g\left(\frac{\nu}{\beta}, \beta\tau\right) \end{aligned}$$

where (a) follows from the change of variable $z = \beta t$.

Property 3. The cross-ambiguity function between the two time- and frequency-shifted versions $g_{(\alpha, \beta)}(t) \triangleq g(t - \alpha)e^{i2\pi\beta t}$ and $g_{(\alpha', \beta')}(t) \triangleq g(t - \alpha')e^{i2\pi\beta' t}$ of a function $g(t) \in \mathcal{L}^2(\mathbb{C})$ is given by

$$\begin{aligned} &A_{g_{(\alpha, \beta)}, g_{(\alpha', \beta')}}(\nu, \tau) \\ &= \int_{-\infty}^{\infty} g(t - \alpha)e^{i2\pi\beta t} g^*(t - \alpha' - \tau)e^{-i2\pi\beta'(t - \tau)} e^{-i2\pi\nu t} dt \\ &\stackrel{(a)}{=} e^{i2\pi\beta'\tau} e^{-i2\pi(\nu + \beta' - \beta)\alpha} \int_{-\infty}^{\infty} g(t')g^*(t' - (\alpha' - \alpha) - \tau)e^{-i2\pi(\nu + \beta' - \beta)t'} dt' \\ &= A_g(\nu + \beta' - \beta, \tau + \alpha' - \alpha)e^{-i2\pi(\nu\alpha - \tau\beta')} e^{-i2\pi(\beta' - \beta)\alpha} \end{aligned} \tag{F.2}$$

where (a) follows from the change of variables $t' = t - \alpha$. As a direct consequence of (F.2), we have

$$A_{g_{(\alpha, \beta)}}(\nu, \tau) = A_g(\nu, \tau)e^{-i2\pi(\nu\alpha - \tau\beta)}. \tag{F.3}$$

Property 4. Let $S_{\mathbb{H}}(\nu, \tau)$ be the delay-Doppler spreading function of the channel \mathbb{H} . Then, for all $g(t), f(t) \in \mathcal{L}^2(\mathbb{C})$,

$$\begin{aligned} \langle \mathbb{H} g, f \rangle &\stackrel{(a)}{=} \iiint_{\mathbb{R}^3} S_{\mathbb{H}}(\nu, \tau) g(t - \tau) e^{i2\pi t\nu} f^*(t) d\tau d\nu dt \\ &= \iint_{\mathbb{R}^2} S_{\mathbb{H}}(\nu, \tau) \left[\int_{-\infty}^{\infty} f(t) g^*(t - \tau) e^{-i2\pi t\nu} dt \right]^* d\tau d\nu \\ &= \iint_{\mathbb{R}^2} S_{\mathbb{H}}(\nu, \tau) A_{f,g}^*(\nu, \tau) d\tau d\nu = \langle S_{\mathbb{H}}, A_{f,g} \rangle \end{aligned}$$

where in (a) we used (4.6).

F.4. PROOF OF THEOREM 4.1

We obtain a capacity lower bound by computing the mutual information for a specific input distribution. In particular, we take $s[k, n]$ i.i.d. JPG with zero mean and variance $TF\text{SNR}$, so that the average power constraint (4.2) is satisfied. The corresponding input \mathbf{s} is independent of \mathbf{h} , \mathbf{P} , and \mathbf{w} , because no feedback from the receiver to the transmitter is assumed. We use the chain rule for mutual information and the fact that mutual information is nonnegative to obtain the following lower bound:

$$\begin{aligned} I(\mathbf{y}; \mathbf{s}) &= I(\mathbf{y}; \mathbf{s}, \mathbf{h}) - I(\mathbf{y}; \mathbf{h} | \mathbf{s}) \\ &= I(\mathbf{y}; \mathbf{h}) + I(\mathbf{y}; \mathbf{s} | \mathbf{h}) - I(\mathbf{y}; \mathbf{h} | \mathbf{s}) \\ &\geq I(\mathbf{y}; \mathbf{s} | \mathbf{h}) - I(\mathbf{y}; \mathbf{h} | \mathbf{s}). \end{aligned} \tag{F.4}$$

A. The ‘‘Coherent’’ Term

The first term can be further lower-bounded as follows

$$\begin{aligned} I(\mathbf{y}; \mathbf{s} | \mathbf{h}) &= h(\mathbf{s} | \mathbf{h}) - h(\mathbf{s} | \mathbf{h}, \mathbf{y}) \\ &\stackrel{(a)}{=} h(\mathbf{s}) - h(\mathbf{s} | \mathbf{h}, \mathbf{y}) \end{aligned}$$

$$\begin{aligned}
 &\stackrel{(b)}{=} \sum_{k=-K}^K \sum_{n=-N}^N \left[\mathbf{h}\left(s[k, n] \mid \mathbf{s}_{\text{pred}}^{(k, n)}\right) - \mathbf{h}\left(s[k, n] \mid \mathbf{h}, \mathbf{y}, \mathbf{s}_{\text{pred}}^{(k, n)}\right) \right] \\
 &\stackrel{(c)}{=} \sum_{k=-K}^K \sum_{n=-N}^N \left[\mathbf{h}(s[k, n]) - \mathbf{h}\left(s[k, n] \mid \mathbf{h}, \mathbf{y}, \mathbf{s}_{\text{pred}}^{(k, n)}\right) \right] \\
 &\stackrel{(d)}{\geq} \sum_{k=-K}^K \sum_{n=-N}^N [\mathbf{h}(s[k, n]) - \mathbf{h}(s[k, n] \mid h[k, n], y[k, n])] \\
 &= \sum_{k=-K}^K \sum_{n=-N}^N I(y[k, n]; s[k, n] \mid h[k, n]).
 \end{aligned}$$

Here, (a) follows because \mathbf{s} and \mathbf{h} are independent; in (b) we used chain rule for differential entropy [we denoted by $\mathbf{s}_{\text{pred}}^{(k, n)}$ a $[(k+K)\tilde{N} + n + N]$ -dimensional vector containing the first $(k+K)\tilde{N} + n + N$ entries of \mathbf{s}]. Next, (c) holds because \mathbf{s} has i.i.d. entries and (d) follows because conditioning reduces entropy.

We next seek a lower bound on $I(y[k, n]; s[k, n] \mid h[k, n])$ that does not depend on the time-frequency (TF) slot position $[k, n]$. Let $\tilde{w}[k, n]$ be the sum of the interference and noise term in $y[k, n]$ [see (4.18)], i.e.,

$$\tilde{w}[k, n] \triangleq \sum_{\substack{l=-K \\ (l, m) \neq (k, n)}}^K \sum_{m=-N}^N p[l, m, k, n] s[l, m] + w[k, n].$$

Furthermore, let $\tilde{w}_G[k, n]$ be a proper Gaussian random variable that has the same variance as $\tilde{w}[k, n]$. It follows from (Diggavi and Cover, 2001, Lemma II.2) that $I(y[k, n]; s[k, n] \mid h[k, n])$ does not increase if we replace $w[k, n]$ with $\tilde{w}_G[k, n]$. In other words, Gaussian noise is the worst noise for this setting. Hence,

$$\begin{aligned}
 I(y[k, n]; s[k, n] \mid h[k, n]) &= I(h[k, n]s[k, n] + \tilde{w}[k, n]; s[k, n] \mid h[k, n]) \\
 &\geq I(h[k, n]s[k, n] + \tilde{w}_G[k, n]; s[k, n] \mid h[k, n]) \\
 &\stackrel{(a)}{=} \mathbb{E}_{h[k, n]} \left[\log \left(1 + \frac{TF\text{SNR} |h[k, n]|^2}{\mathbb{E}[|\tilde{w}_G[k, n]|^2]} \right) \right]
 \end{aligned}$$

$$\stackrel{(b)}{=} \mathbb{E}_h \left[\log \left(1 + \frac{r[0,0]TF\text{SNR}|h|^2}{\mathbb{E}[|\tilde{w}_G[k,n]|^2]} \right) \right] \quad (\text{F.5})$$

where (a) follows because $s[k,n] \sim \mathcal{CN}(0, TF\text{SNR})$, and (b) follows because $h[k,n] \sim \mathcal{CN}(0, r[0,0])$ [see (4.20)], so that we can replace $h[k,n]$ with $r[0,0]h$, where $h \sim \mathcal{CN}(0, 1)$. As the input symbols $s[k,n]$ are independent across TF slots, and as $\mathbb{E}[|p[l,m,k,n]|^2] = \sigma_p^2[k-l, n-m]$ [see (4.24)], we have that

$$\begin{aligned} \mathbb{E}[|\tilde{w}_G[k,n]|^2] &= \mathbb{E}[|\tilde{w}[k,n]|^2] \\ &= 1 + TF\text{SNR} \sum_{\substack{l=-K \\ (l,m) \neq (k,n)}}^K \sum_{m=-N}^N \sigma_p^2[k-l, n-m]. \end{aligned} \quad (\text{F.6})$$

The nonnegativity of $\sigma_p^2[k,n]$ allows us to upper-bound (F.6) as follows

$$\begin{aligned} \mathbb{E}[|\tilde{w}_G[k,n]|^2] &\leq 1 + TF\text{SNR} \sum_{\substack{l=-\infty \\ (l,m) \neq (k,n)}}^{\infty} \sum_{m=-\infty}^{\infty} \sigma_p^2[k-l, n-m] \\ &= 1 + TF\text{SNR} \sum_{\substack{l=-\infty \\ (l,m) \neq (0,0)}}^{\infty} \sum_{m=-\infty}^{\infty} \sigma_p^2[l, m] = 1 + TF\text{SNR}\sigma_I^2 \end{aligned} \quad (\text{F.7})$$

where we set

$$\sigma_I^2 \triangleq \sum_{\substack{l=-\infty \\ (l,m) \neq (0,0)}}^{\infty} \sum_{m=-\infty}^{\infty} \sigma_p^2[l, m]. \quad (\text{F.8})$$

If we now substitute (F.7) into (F.5) we obtain

$$I(y[k,n]; s[k,n] | h[k,n]) \geq \mathbb{E}_h \left[\log \left(1 + \frac{r[0,0]TF\text{SNR}|h|^2}{1 + TF\text{SNR}\sigma_I^2} \right) \right]$$

and, consequently,

$$I(\mathbf{y}; \mathbf{s} | \mathbf{h}) \geq \tilde{K}\tilde{N} \mathbb{E}_h \left[\log \left(1 + \frac{r[0,0]TF\text{SNR}|h|^2}{1 + TF\text{SNR}\sigma_I^2} \right) \right]. \quad (\text{F.9})$$

B. The Penalty Term

We next seek an upper bound on the penalty term $I(\mathbf{y}; \mathbf{h} | \mathbf{s})$ in (F.4). Let $\mathbf{w}_1 \sim \mathcal{CN}(\mathbf{0}, \alpha \mathbf{I})$ and $\mathbf{w}_2 \sim \mathcal{CN}(\mathbf{0}, (1 - \alpha) \mathbf{I})$, where $0 < \alpha < 1$, be two $\tilde{K}\tilde{N}$ -dimensional independent JPG vectors. Then,

$$\begin{aligned} \mathbf{y} &= \mathbf{s} \odot \mathbf{h} + \mathbf{P}\mathbf{s} + \mathbf{w} \\ &= \underbrace{\mathbf{s} \odot \mathbf{h} + \mathbf{w}_1}_{\triangleq \mathbf{y}_1} + \underbrace{\mathbf{P}\mathbf{s} + \mathbf{w}_2}_{\triangleq \mathbf{y}_2} \end{aligned}$$

By the data processing inequality (Cover and Thomas, 2006, Theorem 2.8.1) and the chain rule for mutual information we have that

$$I(\mathbf{y}; \mathbf{h} | \mathbf{s}) \leq I(\mathbf{y}_1, \mathbf{y}_2; \mathbf{h} | \mathbf{s}) = I(\mathbf{y}_1; \mathbf{h} | \mathbf{s}) + I(\mathbf{y}_2; \mathbf{h} | \mathbf{s}, \mathbf{y}_1). \quad (\text{F.10})$$

As \mathbf{h} is JPG, the first term on the right-hand side of (F.10) admits a simple closed-form expression, which can be upper-bounded as follows:

$$\begin{aligned} I(\mathbf{y}_1; \mathbf{h} | \mathbf{s}) &= I(\mathbf{h} \odot \mathbf{s} + \mathbf{w}_1; \mathbf{h} | \mathbf{s}) \\ &= \mathbb{E}_{\mathbf{s}} \left[\log \det \left(\mathbf{I} + \frac{1}{\alpha} \text{diag}(\mathbf{s}) \mathbb{E}[\mathbf{h}\mathbf{h}^H] \text{diag}(\mathbf{s}^H) \right) \right] \\ &\stackrel{(a)}{=} \mathbb{E}_{\mathbf{s}} \left[\log \det \left(\mathbf{I} + \frac{1}{\alpha} \text{diag}(\mathbf{s}^H) \text{diag}(\mathbf{s}) \mathbb{E}[\mathbf{h}\mathbf{h}^H] \right) \right] \\ &\stackrel{(b)}{\leq} \log \det \left(\mathbf{I} + \frac{TF\text{SNR}}{\alpha} \mathbb{E}[\mathbf{h}\mathbf{h}^H] \right). \end{aligned} \quad (\text{F.11})$$

Here, (a) follows from the identity $\det(\mathbf{I} + \mathbf{A}\mathbf{B}^H) = \det(\mathbf{I} + \mathbf{B}^H\mathbf{A})$ for all \mathbf{A} and \mathbf{B} of appropriate dimension (Horn and Johnson, 1985, Theorem 1.3.20) and (b) from Jensen's inequality.

For the second term on the right-hand side of (F.10) we note that

$$\begin{aligned} I(\mathbf{y}_2; \mathbf{h} | \mathbf{s}, \mathbf{y}_1) &= h(\mathbf{y}_2 | \mathbf{s}, \mathbf{y}_1) - h(\mathbf{y}_2 | \mathbf{s}, \mathbf{y}_1, \mathbf{h}) \\ &\stackrel{(a)}{=} h(\mathbf{y}_2 | \mathbf{s}, \mathbf{y}_1) - h(\mathbf{y}_2 | \mathbf{s}, \mathbf{h}) \\ &\stackrel{(b)}{\leq} h(\mathbf{y}_2 | \mathbf{s}) - h(\mathbf{y}_2 | \mathbf{s}, \mathbf{h}, \mathbf{P}) \\ &\stackrel{(c)}{=} h(\mathbf{y}_2 | \mathbf{s}) - h(\mathbf{y}_2 | \mathbf{s}, \mathbf{P}) \\ &= I(\mathbf{y}_2; \mathbf{P} | \mathbf{s}). \end{aligned}$$

Here, (a) holds because \mathbf{y}_1 and \mathbf{y}_2 are conditionally independent given \mathbf{s} and \mathbf{h} , in (b) we used that conditioning reduces entropy (Cover and Thomas, 2006, Theorem 2.6.5), and (c) follows because \mathbf{y}_2 and \mathbf{h} are conditionally independent given \mathbf{P} .

Let $\mathbf{K}(\mathbf{s}) = \mathbb{E}_{\mathbf{P}} [\mathbf{P}\mathbf{s}\mathbf{s}^H\mathbf{P}^H]$ be the $\tilde{K}\tilde{N} \times \tilde{K}\tilde{N}$ covariance matrix of the vector $\mathbf{P}\mathbf{s}$. We next upper-bound $I(\mathbf{y}_2; \mathbf{P} | \mathbf{s})$ as follows:

$$\begin{aligned}
 I(\mathbf{y}_2; \mathbf{P} | \mathbf{s}) &= I(\mathbf{P}\mathbf{s} + \mathbf{w}_2; \mathbf{P} | \mathbf{s}) \\
 &\stackrel{(a)}{=} \mathbb{E}_{\mathbf{s}} \left[\log \det \left(\mathbf{I} + \frac{1}{1-\alpha} \mathbf{K}(\mathbf{s}) \right) \right] \\
 &\stackrel{(b)}{\leq} \sum_{\tilde{k}=0}^{\tilde{K}-1} \sum_{\tilde{n}=0}^{\tilde{N}-1} \mathbb{E}_{\mathbf{s}} \left[\log \left(1 + \frac{1}{1-\alpha} [\mathbf{K}(\mathbf{s})]_{(\tilde{n}+\tilde{k}\tilde{N}, \tilde{n}+\tilde{k}\tilde{N})} \right) \right] \\
 &\stackrel{(c)}{\leq} \sum_{\tilde{k}=0}^{\tilde{K}-1} \sum_{\tilde{n}=0}^{\tilde{N}-1} \log \left(1 + \frac{1}{1-\alpha} \mathbb{E}_{\mathbf{s}} \left[[\mathbf{K}(\mathbf{s})]_{(\tilde{n}+\tilde{k}\tilde{N}, \tilde{n}+\tilde{k}\tilde{N})} \right] \right)
 \end{aligned}$$

where (a) follows because \mathbf{P} is a JFG matrix, in (b) we used Hadamard's inequality, and (c) follows from Jensen's inequality. As the entries of \mathbf{s} are i.i.d. with zero mean, we have that

$$\begin{aligned}
 &\mathbb{E}_{\mathbf{s}} \left[[\mathbf{K}(\mathbf{s})]_{(\tilde{n}+\tilde{k}\tilde{N}, \tilde{n}+\tilde{k}\tilde{N})} \right] \\
 &= \mathbb{E}_{\mathbf{s}} \left[\mathbb{E}_{\mathbf{P}} \left[[\mathbf{P}\mathbf{s}\mathbf{s}^H\mathbf{P}^H]_{(\tilde{n}+\tilde{k}\tilde{N}, \tilde{n}+\tilde{k}\tilde{N})} \right] \right] \\
 &= TF \text{ SNR} \sum_{\substack{l=-K \\ (l,m) \neq (\tilde{k}-K, \tilde{n}-N)}}^K \sum_{m=-N}^N \sigma_p^2 [k-l-K, \tilde{n}-m-\tilde{N}] \\
 &\leq TF \text{ SNR} \sigma_I^2
 \end{aligned}$$

where σ_I^2 has been defined in (F.8). Hence,

$$I(\mathbf{y}_2; \mathbf{P} | \mathbf{s}) \leq \tilde{K}\tilde{N} \log \left(1 + \frac{TF \text{ SNR}}{1-\alpha} \sigma_I^2 \right). \quad (\text{F.12})$$

If we now substitute (F.11) and (F.12) in (F.10), we get that

$$I(\mathbf{y}; \mathbf{h} | \mathbf{s}) \leq \log \det \left(\mathbf{I} + \frac{TF\text{SNR}}{\alpha} \mathbb{E}[\mathbf{h}\mathbf{h}^H] \right) + \tilde{K}\tilde{N} \log \left(1 + \frac{TF\text{SNR}}{1-\alpha} \sigma_I^2 \right).$$

Furthermore, as the bound holds for all $\alpha \in (0, 1)$,

$$\begin{aligned} I(\mathbf{y}; \mathbf{P} | \mathbf{s}) \\ \leq \inf_{0 < \alpha < 1} \left\{ \log \det \left(\mathbf{I} + \frac{TF\text{SNR}}{\alpha} \mathbb{E}[\mathbf{h}\mathbf{h}^H] \right) + \tilde{K}\tilde{N} \log \left(1 + \frac{\text{SNR}}{1-\alpha} \sigma_I^2 \right) \right\}. \end{aligned} \quad (\text{F.13})$$

C. Putting Pieces Together

We substitute (F.9) and (F.13) in (F.4) and (F.4) in (4.27) to get the following lower bound on capacity:

$$\begin{aligned} C_{\text{WSSUS-D}}(\text{SNR}) &\geq \frac{\tilde{N}}{T} \mathbb{E}_h \left[\log \left(1 + \frac{r[0,0]TF\text{SNR}|h|^2}{1 + TF\text{SNR}\sigma_I^2} \right) \right] \\ &\quad - \inf_{0 < \alpha < 1} \left\{ \lim_{\tilde{K} \rightarrow \infty} \frac{1}{(\tilde{K} + 2K')T} \log \det \left(\mathbf{I} + \frac{TF\text{SNR}}{\alpha} \mathbb{E}[\mathbf{h}\mathbf{h}^H] \right) \right. \\ &\quad \left. + \frac{\tilde{N}}{T} \log \left(1 + \frac{TF\text{SNR}}{1-\alpha} \sigma_I^2 \right) \right\}. \end{aligned} \quad (\text{F.14})$$

By direct application of (Miranda and Tilli, 2000, Theorem 3.4), an extension of Szegö's theorem (on the asymptotic eigenvalue distribution of Toeplitz matrices) to two-level Toeplitz matrices, we obtain

$$\begin{aligned} \lim_{\tilde{K} \rightarrow \infty} \frac{1}{(\tilde{K} + 2K')T} \log \det \left(\mathbf{I} + \frac{TF\text{SNR}}{\alpha} \mathbb{E}[\mathbf{h}\mathbf{h}^H] \right) \\ = \frac{1}{T} \int_{-1/2}^{1/2} \log \det \left(\mathbf{I} + \frac{TF\text{SNR}}{\alpha} \mathbf{C}(\theta) \right) d\theta. \end{aligned}$$

Substituting this expression into (F.14) and noting that

$$\begin{aligned}
 \sigma_I^2 &= \sum_{\substack{k=-\infty \\ (k,n) \neq (0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} \sigma_p^2[k, n] \\
 &= \sum_{\substack{k=-\infty \\ (k,n) \neq (0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} \iint_{\mathbb{R}^2} C_{\mathbb{H}}(\nu, \tau) |A_g(\nu - nF, \tau - kT)|^2 d\tau d\nu
 \end{aligned} \tag{F.15}$$

we complete the proof.

F.5. PROOF OF COROLLARY 4.2

To prove the corollary we further lower bound each term in (4.29) separately.

F.5.1. The log det Term

We start with an upper bound on the log det term on the right-hand side of (4.29). The matrix $\mathbf{C}(\theta)$ is Toeplitz. Hence, the entries on the main diagonal of $\mathbf{C}(\theta)$ are all equal. Let $c_0(\theta)$ denote one such entry; then

$$\begin{aligned}
 c_0(\theta) &\stackrel{(a)}{=} \sum_{k=-\infty}^{\infty} r[k, 0] e^{-i2\pi k\theta} \\
 &\stackrel{(b)}{=} \int_{-1/2}^{1/2} c(\theta, \varphi) d\varphi, \quad |\theta| \leq 1/2.
 \end{aligned} \tag{F.16}$$

Here, (a) follows from (4.28) and (4.20); (b) follows from (4.21) and from Poisson summation formula. By Hadamard's inequality, we can upper-bound the log det term on the right-hand side of (4.29) as

follows:

$$\begin{aligned} \frac{1}{T} \int_{-1/2}^{1/2} \log \det \left(\mathbf{I} + \frac{TF\text{SNR}}{\alpha} \mathbf{C}(\theta) \right) d\theta \\ \leq \frac{\tilde{N}}{T} \int_{-1/2}^{1/2} \log \left(1 + \frac{TF\text{SNR}}{\alpha} c_0(\theta) \right) d\theta. \end{aligned} \quad (\text{F.17})$$

We next use that $\nu_0 T < 1/2$, by assumption, to split the integral in two parts; the use of Jensen's inequality on both terms yields

$$\begin{aligned} \frac{\tilde{N}}{T} \int_{-1/2}^{1/2} \log \left(1 + \frac{TF\text{SNR}}{\alpha} c_0(\theta) \right) d\theta \\ = \frac{\tilde{N}}{T} \int_{|\theta| < \nu_0 T} \log \left(1 + \frac{TF\text{SNR}}{\alpha} c_0(\theta) \right) d\theta \\ + \frac{\tilde{N}}{T} \int_{\nu_0 T < |\theta| < 1/2} \log \left(1 + \frac{TF\text{SNR}}{\alpha} c_0(\theta) \right) d\theta \\ \leq \frac{2\nu_0 T \tilde{N}}{T} \log \left(1 + \frac{TF\text{SNR}}{2\nu_0 T \alpha} \int_{|\theta| < \nu_0 T} c_0(\theta) d\theta \right) \\ + \frac{\tilde{N}}{T} (1 - 2\nu_0 T) \log \left(1 + \frac{TF\text{SNR}}{(1 - 2\nu_0 T) \alpha} \int_{\nu_0 T < |\theta| < 1/2} c_0(\theta) d\theta \right). \end{aligned} \quad (\text{F.18})$$

Now note that

$$\begin{aligned} \int_{|\theta| < \nu_0 T} c_0(\theta) d\theta \\ \stackrel{(a)}{=} \int_{-1/2}^{1/2} \int_{|\theta| < \nu_0 T} c(\theta, \varphi) d\theta d\varphi \end{aligned}$$

$$\begin{aligned}
 & \stackrel{(b)}{=} \int_{-1/2}^{1/2} \int_{|\theta| < \nu_0 T} \frac{1}{TF} \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_{\mathbb{H}}\left(\frac{\theta-k}{T}, \frac{\varphi-n}{F}\right) \\
 & \qquad \qquad \qquad \times \left| A_g\left(\frac{\theta-k}{T}, \frac{\varphi-n}{F}\right) \right|^2 d\theta d\varphi \\
 & \stackrel{(c)}{\leq} \frac{1}{TF} \int_{-1/2}^{1/2} \int_{|\theta| < \nu_0 T} \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_{\mathbb{H}}\left(\frac{\theta-k}{T}, \frac{\varphi-n}{F}\right) d\theta d\varphi \\
 & \leq \frac{1}{TF} \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_{\mathbb{H}}\left(\frac{\theta-k}{T}, \frac{\varphi-n}{F}\right) d\theta d\varphi \\
 & = \iint_{\mathbb{R}^2} C_{\mathbb{H}}(\nu, \tau) d\tau d\nu = 1 \tag{F.19}
 \end{aligned}$$

where (a) follows from (F.16), (b) follows from (4.22), and (c) follows from Property 1 in Appendix F.3. Furthermore, with similar steps, we have that

$$\begin{aligned}
 & \int_{\nu_0 T < |\theta| < 1/2} c_0(\theta) d\theta \\
 & \leq \frac{1}{TF} \int_{-1/2}^{1/2} \int_{\nu_0 T < |\theta| < 1/2} \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_{\mathbb{H}}\left(\frac{\theta-k}{T}, \frac{\varphi-n}{F}\right) d\theta d\varphi \\
 & \leq \int_{|\nu| \geq \nu_0} \int_{-\infty}^{\infty} C_{\mathbb{H}}(\nu, \tau) d\tau d\nu \leq \epsilon \tag{F.20}
 \end{aligned}$$

where the last step follows from (4.9). If we now substitute (F.19) and (F.20) into (F.18), insert the result into (F.17), set $\tilde{\Delta}_{\mathbb{H}} = 2\nu_0 T$

and use $B = \tilde{N}F$, we get

$$\begin{aligned}
 & \frac{1}{T} \int_{-1/2}^{1/2} \log \det \left(\mathbf{I} + \frac{TF\text{SNR}}{\alpha} \mathbf{C}(\theta) \right) d\theta \\
 & \leq \frac{B\tilde{\Delta}_{\mathbb{H}}}{TF} \log \left(1 + \frac{TF\text{SNR}}{\alpha\tilde{\Delta}_{\mathbb{H}}} \right) + \frac{B}{TF} (1 - \tilde{\Delta}_{\mathbb{H}}) \log \left(1 + \frac{TF\text{SNR}\epsilon}{\alpha(1 - \tilde{\Delta}_{\mathbb{H}})} \right).
 \end{aligned} \tag{F.21}$$

F.5.2. Bounds on $r[0, 0]$ and on σ_I^2

To further lower bound the right-hand side of (4.29), we next derive a lower bound on $r[0, 0]$ and an upper bound on σ_I^2 that are explicit in $\Delta_{\mathbb{H}}$ and ϵ . These bounds will be explicit in the ambiguity function of the prototype pulse $g(t)$.

Let $\mathcal{D} = \{(\nu, \tau) \in [-\nu_0, \nu_0] \times [-\tau_0, \tau_0]\}$ be the rectangular area in the delay doppler plane over which $C_{\mathbb{H}}(\nu, \tau)$ has at least $1 - \epsilon$ of its volume according to (4.9). The following chain of inequalities holds:

$$\begin{aligned}
 r[0, 0] &= \iint_{\mathbb{R}^2} C_{\mathbb{H}}(\nu, \tau) |A_g(\nu, \tau)|^2 d\tau d\nu \\
 &\geq \iint_{\mathcal{D}} C_{\mathbb{H}}(\nu, \tau) |A_g(\nu, \tau)|^2 d\tau d\nu \\
 &\geq \min_{(\nu, \tau) \in \mathcal{D}} |A_g(\nu, \tau)|^2 \iint_{\mathcal{D}} C_{\mathbb{H}}(\nu, \tau) d\tau d\nu \\
 &\geq \min_{(\nu, \tau) \in \mathcal{D}} |A_g(\nu, \tau)|^2 (1 - \epsilon) = m_g (1 - \epsilon).
 \end{aligned} \tag{F.22}$$

We now seek an upper bound on σ_I^2 . Let

$$M(\nu, \tau) = \sum_{\substack{k=-\infty \\ (k,n) \neq (0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} |A_g(\nu - nF, \tau - kT)|^2$$

and note that

$$\begin{aligned}
 M(\nu, \tau) &\leq \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} |A_g(\nu - nF, \tau - kT)|^2 \\
 &\leq \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} |\langle g(t + \tau)e^{i2\pi\nu t}, g_{k,n}(t) \rangle|^2 \\
 &\stackrel{(a)}{\leq} |g(t)|^4 = 1
 \end{aligned} \tag{F.23}$$

where (a) follows from Bessel's inequality (Kreyszig, 1989, Theorem 3.4-6). The following chain of inequalities holds:

$$\begin{aligned}
 \sigma_I^2 &= \iint_{\mathbb{R}^2} C_{\mathbb{H}}(\nu, \tau) M(\nu, \tau) d\tau d\nu \\
 &= \iint_{\mathcal{D}} C_{\mathbb{H}}(\nu, \tau) M(\nu, \tau) d\tau d\nu + \iint_{\mathbb{R}^2 \setminus \mathcal{D}} C_{\mathbb{H}}(\nu, \tau) M(\nu, \tau) d\tau d\nu \\
 &\leq \max_{(\nu, \tau) \in \mathcal{D}} M(\nu, \tau) \iint_{\mathcal{D}} C_{\mathbb{H}}(\nu, \tau) d\tau d\nu \\
 &\quad + \max_{(\tau, \nu) \in \mathbb{R}^2 \setminus \mathcal{D}} M(\nu, \tau) \iint_{\mathbb{R}^2 \setminus \mathcal{D}} C_{\mathbb{H}}(\nu, \tau) d\tau d\nu \\
 &\stackrel{(a)}{\leq} \max_{(\nu, \tau) \in \mathcal{D}} M(\nu, \tau) + \epsilon = M_g + \epsilon
 \end{aligned} \tag{F.24}$$

where (a) follows from (F.23) and (4.8), (4.9).

The proof is completed by substituting (F.21) into (4.29), and using (F.22) and (F.24).

F.6. PROOF OF LEMMA 4.3

To prove the lemma, we verify that after the substitutions

$$\begin{aligned} e(t) &= \sqrt{\beta}g(\beta t) \\ \tilde{T} &= T/\beta \\ \tilde{F} &= F\beta \\ \tilde{\tau}_0 &= \tau_0/\beta \\ \tilde{\nu}_0 &= \nu_0\beta \end{aligned}$$

the lower bound $L(\text{SNR}, g, T, F, \tau_0, \nu_0, \epsilon)$ in (4.31) does not change. Note first that $TF = \tilde{T}\tilde{F}$ and that $\nu_0 T = \tilde{\nu}_0 \tilde{T}$. Furthermore, $\|e(t)\| = \|g(t)\| = 1$ and, by Property 2 in Appendix F.3, the orthonormality of $\{g_{k,n}(t)\}$ implies the orthonormality of $\{e(t - k\tilde{T})e^{i2\pi n\tilde{F}t}\}$, for all k, n . Therefore, the triple $\{e(t), \tilde{T}, \tilde{F}\}$ is a valid WH set to be used in the lower bound. Let now $\mathcal{E} = [-\tilde{\nu}_0, \tilde{\nu}_0] \times [-\tilde{\tau}_0, \tilde{\tau}_0]$; we have that

$$\begin{aligned} m_g &= \min_{(\nu, \tau) \in \mathcal{D}} |A_g(\nu, \tau)|^2 \\ &= \min_{(\nu, \tau) \in \mathcal{D}} \left| A_e\left(\beta\nu, \frac{\tau}{\beta}\right) \right|^2 \\ &= \min_{(\nu, \tau) \in \mathcal{E}} |A_e(\nu, \tau)|^2. \end{aligned}$$

Similarly,

$$\begin{aligned} M_g &= \max_{(\nu, \tau) \in \mathcal{D}} \sum_{\substack{k=-\infty \\ (k,n) \neq (0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} |A_g(\nu - nF, \tau - kT)|^2 \\ &= \max_{(\nu, \tau) \in \mathcal{D}} \sum_{\substack{k=-\infty \\ (k,n) \neq (0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} \left| A_e\left(\beta(\nu - nF), \frac{\tau - kT}{\beta}\right) \right|^2 \\ &= \max_{(\nu, \tau) \in \mathcal{D}} \sum_{\substack{k=-\infty \\ (k,n) \neq (0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} \left| A_e\left(\beta\nu - n\tilde{F}, \frac{\tau}{\beta} - k\tilde{T}\right) \right|^2 \end{aligned}$$

$$= \max_{(\nu, \tau) \in \mathcal{E}} \sum_{\substack{k=-\infty \\ (k, n) \neq (0, 0)}}^{\infty} \sum_{n=-\infty}^{\infty} \left| A_e(\nu - n\tilde{F}, \tau - k\tilde{T}) \right|^2.$$

As a last step, we note that for $\beta = \sqrt{T/F}$ and under the assumption that $\nu_0 T = \tau_0 F$, we get $\tilde{T} = \tilde{F} = \sqrt{TF}$, and $\tilde{\tau}_0 = \tilde{\nu}_0 = \sqrt{\Delta_{\mathbb{H}}}/2$.

F.7. PROOF OF LEMMA 4.4

F.7.1. Derivative-Based Approximation for m_g

Since $|A_g(\nu, \tau)| \leq |A_g(0, 0)| = 1$, the gradient of $A_g(\nu, \tau)$ in $(\nu, \tau) = (0, 0)$ is zero. For small $\Delta_{\mathbb{H}}$, we can approximate $A_g(\nu, \tau)$ around $(0, 0)$ with its second order Taylor-series expansion, and obtain (Wilcox, 1991, Section 6)

$$|A_g(\nu, \tau)|^2 \approx 1 - 4\pi^2(a_0^2\nu^2 + b_0^2\tau^2) \quad (\text{F.25})$$

where

$$a_0^2 = \int_{-\infty}^{\infty} t^2 |g(t)|^2 dt$$

$$b_0^2 = \int_{-\infty}^{\infty} f^2 |G(f)|^2 df.$$

Consequently, for small $\Delta_{\mathbb{H}}$

$$\begin{aligned} m_g &\approx \min_{(\nu, \tau) \in \tilde{\mathcal{D}}} [1 - 4\pi^2(a_0^2\nu^2 + b_0^2\tau^2)] \\ &= 1 - \pi^2(a_0^2 + b_0^2)\Delta_{\mathbb{H}} \\ &= 1 - c_m\Delta_{\mathbb{H}}. \end{aligned}$$

F.7.2. Derivative-Based Approximation for M_g

We want to find a small- $\Delta_{\mathbb{H}}$ approximation for

$$M_g = \max_{(\nu, \tau) \in \tilde{\mathcal{D}}} \sum_{\substack{k=-\infty \\ (k,n) \neq (0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} \left| A_g\left(\nu - n\sqrt{TF}, \tau - k\sqrt{TF}\right) \right|^2. \quad (\text{F.26})$$

Under the assumption that $A_g(\nu, \tau)$ is differentiable in $(n\sqrt{TF}, k\sqrt{TF})$ for all (n, k) , we can approximate the function to maximize on the right-hand side of (F.26), with its first-order Taylor-series expansion around the point $(0, 0)$. Let $G(f) = \mathbb{F}[g(t)]$ (note that the assumption that $g(t)$ is real and even implies that $G(f)$ is real and even as well) and define

$$\begin{aligned} a_{k,n} &\triangleq \left. \frac{\partial A_g\left(\nu - n\sqrt{TF}, \tau - k\sqrt{TF}\right)}{\partial \nu} \right|_{(0,0)} \\ &= -i2\pi \int_t tg(t)g(t + k\sqrt{TF})e^{i2\pi n\sqrt{TF}t} dt \quad (\text{F.27}) \end{aligned}$$

$$\begin{aligned} b_{k,n} &\triangleq \left. \frac{\partial A_g\left(\nu - n\sqrt{TF}, \tau - k\sqrt{TF}\right)}{\partial \tau} \right|_{(0,0)} \\ &= i2\pi \int_{-\infty}^{\infty} fG(f - n\sqrt{TF})G(f)e^{-i2\pi k\sqrt{TF}f} df. \quad (\text{F.28}) \end{aligned}$$

Then, around the point $(\nu, \tau) = (0, 0)$ we have that

$$A_g\left(\nu - n\sqrt{TF}, \tau - k\sqrt{TF}\right) \approx a_{k,n}\nu + b_{k,n}\tau.$$

Consequently, for $\Delta_{\mathbb{H}}$ small, the right-hand side of (F.26) can be approximated as

$$\begin{aligned}
 & \sum_{\substack{k=-\infty \\ (k,n) \neq (0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} \left| A_g \left(\nu - n\sqrt{TF}, \tau - k\sqrt{TF} \right) \right|^2 \\
 & \approx \sum_{\substack{k=-\infty \\ (k,n) \neq (0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} |a_{k,n}\nu + b_{k,n}\tau|^2. \quad (\text{F.29})
 \end{aligned}$$

It follows from (F.27) and (F.28) by direct computation that

$$a_{k,-n} = -a_{k,n}^* \quad a_{-k,n} = a_{k,n}^* \quad a_{-k,-n} = -a_{k,n} \quad (\text{F.30})$$

$$b_{k,-n} = b_{k,n}^* \quad b_{-k,n} = -b_{k,n}^* \quad b_{-k,-n} = -b_{k,n}. \quad (\text{F.31})$$

But these six relations imply that the sum of the cross products on the right-hand side of (F.29) vanishes:

$$\begin{aligned}
 & \sum_{\substack{k=-\infty \\ (k,n) \neq (0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} \Re\{a_{k,n}b_{k,n}^*\} \\
 & = \sum_{k=1}^{\infty} \Re\{a_{k,0}b_{k,0}^* + a_{-k,0}b_{-k,0}^*\} + \sum_{n=1}^{\infty} \Re\{a_{0,n}b_{0,n}^* + a_{0,-n}b_{0,-n}^*\} \\
 & \quad + \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} \Re\{a_{k,n}b_{k,n}^* + a_{-k,n}b_{-k,n}^* + a_{k,-n}b_{k,-n}^* + a_{-k,-n}b_{-k,-n}^*\} \\
 & = \sum_{k=1}^{\infty} \Re\{a_{k,0}b_{k,0}^* - (a_{k,0}b_{k,0}^*)^*\} + \sum_{n=1}^{\infty} \Re\{a_{0,n}b_{0,n}^* - (a_{0,n}b_{0,n}^*)^*\} \\
 & \quad + 2 \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} \Re\{a_{k,n}b_{k,n}^* - (a_{k,n}b_{k,n}^*)^*\} \\
 & = 0
 \end{aligned}$$

where the last step follows because for all $z \in \mathbb{C}$, $\Re\{z - z^*\} = 0$. Hence,

$$M_g \approx \max_{(\nu,\tau) \in \tilde{\mathcal{D}}} \sum_{\substack{k=-\infty \\ (k,n) \neq (0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} |a_{k,n}\nu + b_{k,n}\tau|^2$$

F. TECHNICAL RESULTS FROM CHAPTER 4

$$\begin{aligned}
 &= \max_{(\nu, \tau) \in \tilde{\mathcal{D}}} \left\{ \nu^2 \sum_{\substack{k=-\infty \\ (k,n) \neq (0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} |a_{k,n}|^2 + \tau^2 \sum_{\substack{k=-\infty \\ (k,n) \neq (0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} |b_{k,n}|^2 \right\} \\
 &= \frac{\Delta_{\mathbb{H}}}{4} \sum_{\substack{k=-\infty \\ (k,n) \neq (0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} [|a_{k,n}|^2 + |b_{k,n}|^2] = c_M \Delta_{\mathbb{H}}.
 \end{aligned}$$

APPENDIX G

A Result from Linear Algebra

Lemma G.1. *Let \mathbf{A} be an $(N + 1) \times N$ matrix. If every set of N rows of \mathbf{A} is linearly independent, then the $N(N + 1) \times N(N + 1)$ matrix $\hat{\mathbf{A}}$ defined as $\hat{\mathbf{A}} = \begin{bmatrix} (\mathbf{I}_N \otimes \mathbf{A}) & [\mathbf{D}(\mathbf{A})]_{\diamond, [2 : N+1]} \end{bmatrix}$ has full rank.*

Proof. The proof is by contradiction. Assume that $\hat{\mathbf{A}}$ does not have full rank. Then there exists an $N(N + 1)$ -dimensional nonzero vector $\mathbf{u}^\top = [\mathbf{u}_1^\top \ \cdots \ \mathbf{u}_N^\top]$, where $\mathbf{u}_n \in \mathbb{C}^{N+1}$, such that $\mathbf{u}^\top \hat{\mathbf{A}} = \mathbf{0}$. Because $\hat{\mathbf{A}} = \begin{bmatrix} (\mathbf{I}_N \otimes \mathbf{A}) & [\mathbf{D}(\mathbf{A})]_{\diamond, [2 : N+1]} \end{bmatrix}$, we have in particular that (i) $\mathbf{u}^\top (\mathbf{I}_N \otimes \mathbf{A}) = \mathbf{0}$ and (ii) $\mathbf{u}^\top [\mathbf{D}(\mathbf{A})]_{\diamond, [2 : N+1]} = \mathbf{0}$. We next analyze these two equalities separately. Equality (i) can be restated as $\mathbf{u}_n^\top \mathbf{A} = \mathbf{0}$ for all $n \in [1 : N]$, which implies that all vectors \mathbf{u}_n lie in the kernel of the $N \times (N + 1)$ matrix \mathbf{A}^\top . Because \mathbf{A}^\top has rank N , its kernel must be of dimension 1. Hence, all vectors \mathbf{u}_n must be collinear, i.e., there exists a vector \mathbf{v} and a set of N constants c_n such that $\mathbf{u}_n = c_n \mathbf{v}$, for all $n \in [1 : N]$. The vector \mathbf{v} and at least one of the constants c_n must be nonzero because \mathbf{u} is nonzero. Furthermore, because $\mathbf{v}^\top \mathbf{A} = \mathbf{0}$, and because every set of N rows of \mathbf{A} is linearly independent by assumption, *all* components of \mathbf{v} must be nonzero.

We now use this property of \mathbf{v} to analyze equality (ii), which can be restated as

$$\mathbf{u}^\top [\mathbf{D}(\mathbf{A})]_{\diamond, [2 : N+1]} = [c_1 \mathbf{v}^\top \ \cdots \ c_N \mathbf{v}^\top] [\mathbf{D}(\mathbf{A})]_{\diamond, [2 : N+1]} = \mathbf{0}$$

or, after straightforward manipulations, as

$$[\text{diag}(\mathbf{v})]_{[2:N+1],[2:N+1]} [\mathbf{A}]_{[2:N+1],\diamond} \begin{bmatrix} c_1 \\ \vdots \\ c_N \end{bmatrix} = \mathbf{0}.$$

Because all the components of \mathbf{v} are nonzero, this last equality implies that

$$[\mathbf{A}]_{[2:N+1],\diamond} \begin{bmatrix} c_1 \\ \vdots \\ c_N \end{bmatrix} = \mathbf{0}.$$

However, this contradicts the assumption that every set of N rows of \mathbf{A} is linearly independent (recall that at least one of the constants c_n is nonzero). Hence, $\hat{\mathbf{A}}$ must have full rank. \square

APPENDIX H

Notation

H.1. MISCELLANEOUS

i	$\sqrt{-1}$
\triangleq	definition
$\mathcal{A}, \mathcal{B}, \dots$	sets
$ \mathcal{A} $	cardinality of the set \mathcal{A}
\mathcal{A}^c	complement of the set \mathcal{A} in some specified superset
$\mathbb{R}, \mathbb{C}, \mathbb{Z}, \mathbb{N}$	real line, complex plane, set of all integers, set of natural numbers, including zero
$\mathbb{R}^+, \mathbb{C}^+$	set of positive real numbers; set of complex numbers with positive imaginary part: $\{z \in \mathbb{C} \mid \Im z > 0\}$
$\mathcal{L}^2(\mathbb{C})$	Hilbert space of complex-valued finite-energy signals
\Re, \Im	real, imaginary part of a complex-valued quantity
$[n : m]$	set of integer numbers $\{n, n + 1, \dots, m\}$
$a \oplus b$	sum modulo 2π of scalars a and b
$\langle \mathbf{a}, \mathbf{b} \rangle$	inner product of the vectors \mathbf{a} and \mathbf{b}
$\langle f(\cdot), g(\cdot) \rangle$	inner product of the functions $f(\cdot)$ and $g(\cdot)$
$\ \mathbf{a}\ ^2; \ g(x)\ ^2$	squared ℓ^2 -norm of the vector \mathbf{a} : $\ \mathbf{a}\ ^2 \triangleq \sum_i [\mathbf{a}]_i ^2$; squared \mathcal{L}^2 -norm of the function $g(x)$: $\ g(x)\ ^2 \triangleq \int_{-\infty}^{\infty} g(x) ^2 dx$

H.2. FUNCTIONS

$\ln(x)$	logarithm of x to the base e (natural logarithm)
$\log(x)$	logarithm of x to the base 2
$u(x)$	unit step function; takes on value 0 for $x < 0$ and 1 for $x \geq 0$
$\arg(x)$	argument of $x \in \mathbb{C}$
$\delta[k]$	discrete Kronecker Delta; takes on value 1 if $k = 0$ and 0 else
$I[x] = 1$	indicator function; takes on value 1 if x is true and 0 if x is false
$\delta(x)$	Dirac delta distribution
$f(x) = \mathcal{O}(g(x))$	$ f(x)/g(x) $ remains bounded as $x \rightarrow \infty$
$g(x) = \Theta(f(x))$	$f(x) = \mathcal{O}(g(x))$ and $g(x) = \mathcal{O}(f(x))$
$f(x) = o(g(x))$	$ f(x)/g(x) \rightarrow 0$ as $x \rightarrow \infty$
$\lceil x \rceil$	smallest integer equal to or larger than x
$(x)^+$	positive part of x ; takes value x for $x > 0$ and 0 else

H.3. OPERATORS

\mathbb{I}	identity operator
\mathbb{D}_D	time-limiting projection operator; limits the duration of $s(t) \in \mathcal{L}^2(\mathbb{C})$ to D sec as follows: $(\mathbb{D}_D s)(t) \triangleq I[t \leq D/2] s(t)$
\mathbb{B}_B	frequency-limiting projection operator; limits the bandwidth of $s(t) \in \mathcal{L}^2(\mathbb{C})$ to B Hz as follows: $(\mathbb{B}_B s)(t) \triangleq \int_{-\infty}^{\infty} \frac{\sin[\pi B(t-t')]}{\pi(t-t')} s(t') dt'$
$\mathbb{F}[\cdot]$	Fourier transform: $\mathbb{F}_{t \rightarrow f}[g(t)] \triangleq \int_{-\infty}^{\infty} g(t) e^{-i2\pi ft} dt$
$\partial \mathbf{g} / \partial \mathbf{u}$	Jacobian matrix of the vector-valued function $\mathbf{g}(\mathbf{u})$, i.e., the matrix that contains the partial derivative $\partial g_i / \partial u_j$ in its i th row and j th column

H.4. PROBABILITY THEORY

$\mathbb{P}\{\mathcal{A}\}$	probability of event \mathcal{A}
$\mathbb{P}\{\mathcal{A} \mathcal{B}\}$	probability of event \mathcal{A} conditioned on \mathcal{B}
$F_X(x), f_X(x)$	CDF, PDF of the RV X
$M_X(s)$	moment-generating function of the RV X : $M_X(s) \triangleq \int_{-\infty}^{\infty} e^{sx} f_X(x) dx$
$G_F(z)$	Stieltjes transform of the CDF $F(\cdot)$: $G_F(z) \triangleq \int_{-\infty}^{\infty} \frac{f(x)}{x-z} dx, \quad z \in \mathbb{C}^+; \quad f(\cdot)$ is the PDF of $F(\cdot)$
$\mathbb{E}_X[\cdot]$	expectation operator (w.r.t. the random variable X); if X is omitted, then w.r.t. all random variables
$\text{Var}[\cdot]$	variance operator
$\mathcal{CN}(\mathbf{m}, \mathbf{C})$	multivariate jointly proper Gaussian (JPG) distribution with mean \mathbf{m} and covariance matrix \mathbf{C}
$\mathcal{U}(a, b)$	uniform distribution over the interval $[a, b]$
\sim	“distributed as”

H.5. LINEAR ALGEBRA

$a, A, b, B \dots$	scalars
$\mathbf{a}, \mathbf{b}, \dots$	vectors
$\mathbf{A}, \mathbf{B}, \dots$	matrices
$\mathbf{a}^\top, \mathbf{A}^\top$	transpose of the vector \mathbf{a} and the matrix \mathbf{A}
$a^*, \mathbf{a}^*, \mathbf{A}^*$	element-wise complex conjugate of the scalar a , the vector \mathbf{a} , and the matrix \mathbf{A}
$\mathbf{a}^H, \mathbf{A}^H$	Hermitian transpose of the vector \mathbf{a} and the matrix \mathbf{A}
$[\mathbf{a}]_i$	i th element of the vector \mathbf{a}
$[\mathbf{A}]_{i,j}$	element in the i th row and j th column of the matrix \mathbf{A}
$\mathbf{u}_{\mathcal{I}}$	$ \mathcal{I} $ -dimensional subvector of the vector \mathbf{u} containing the elements $[u_i]_{i \in \mathcal{I}}$ corresponding to the index set \mathcal{I}
$[\mathbf{A}]_{\mathcal{I}, \mathcal{J}}$	$ \mathcal{I} \times \mathcal{J} $ -dimensional submatrix of the matrix \mathbf{A} containing the elements $[a_{i,j}]_{i \in \mathcal{I}, j \in \mathcal{J}}$ corresponding to the index sets \mathcal{I} and \mathcal{J}
$[\mathbf{A}]_{\diamond, \mathcal{J}}, [\mathbf{A}]_{\mathcal{I}, \diamond}$	short notations: $[\mathbf{A}]_{\diamond, \mathcal{J}} \triangleq [\mathbf{A}]_{[1:M], \mathcal{J}}$ and $[\mathbf{A}]_{\mathcal{I}, \diamond} \triangleq [\mathbf{A}]_{\mathcal{I}, [1:N]}$ for an $M \times N$ matrix \mathbf{A}
$\mathbf{I}, \mathbf{0}$	identity and all zero matrices of appropriate sizes
$\text{diag}(\mathbf{a})$	diagonal matrix with vector \mathbf{a} on its main diagonal
$\ \mathbf{A}\ _F^2$	squared Frobenius norm of \mathbf{A} : $\ \mathbf{A}\ _F^2 \triangleq \sum_{i,j} [\mathbf{A}]_{i,j} ^2$
$\text{tr } \mathbf{A}, \det \mathbf{A}$	trace, determinant of the square matrix \mathbf{A}
$\text{rank } \mathbf{A}$	rank of the matrix \mathbf{A}
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of the matrices \mathbf{A} and \mathbf{B}
$\mathbf{A} \odot \mathbf{B}$	Schur-Hadamard product of the matrices \mathbf{A} and \mathbf{B}
$\mathbf{A}\mathbf{B} \otimes \mathbf{C}$	the standard matrix product takes precedence over
$\triangleq (\mathbf{A}\mathbf{B}) \odot \mathbf{C}$	Kronecker and Hadamard products
$\lambda(\mathbf{A})$	eigenvalue of matrix \mathbf{A}
$\lambda_n(\mathbf{A})$	n th eigenvalue of the $N \times N$ matrix \mathbf{A} , sorted in nonincreasing order and including multiplicities, i.e., $\lambda_1 \geq \dots \geq \lambda_N$
$\lambda_{\min}(\mathbf{A})$	the smallest eigenvalue of the matrix \mathbf{A}
$\lambda_{\max}(\mathbf{A})$	the largest eigenvalue of the matrix \mathbf{A}

H.6. COMMUNICATION AND INFORMATION THEORY

SNR	signal-to-noise ratio
SINR	signal-to-interference-plus-noise ratio
$h(\mathbf{x})$	differential entropy of the random vector \mathbf{x}
$I(\mathbf{x}; \mathbf{y})$	mutual information between the random vectors \mathbf{x} and \mathbf{y}
C_{P1}, C_{P2}, C_{AF}	per source-destination terminal pair capacity induced by the protocol P1, the protocol P2, and the AF protocol, respectively
C_{WSSUS}	information capacity of the WSSUS continuous-time channel
C_{AWGN}	capacity of the AWGN channel
$C_{WSSUS-D}$	capacity of the discretized channel induced by WH set
C_{SIMO}	capacity of the noncoherent correlated block-fading SIMO channel

APPENDIX I

Acronyms

AWGN	additive white Gaussian noise
AF	amplify-and-forward
CDF	cumulative distribution function
CSI	channel state information
GSM	global system for mobile communication
DFT	discrete Fourier transform
DMC	discrete memoryless channel
ESD	empirical spectral distribution
FDMA	frequency division multiple access
ICI	intercarrier interference
IO	input-output
ISI	intersymbol interference
JPG	jointly proper Gaussian
LOS	line of sight
LTI	linear time-invariant
LTV	linear time-variant
MIMO	multiple-input multiple-output
MISO	multiple-input single-output
MGF	moment generating function
OFDM	orthogonal frequency division multiplexing
PDF	probability density function
PSD	power spectral density
PS-OFDM	pulse-shaped orthogonal frequency division multiplexing

I. ACRONYMS

RV	random variable
SIMO	single-input multiple-output
SISO	single-input single-output
SNR	signal-to-noise ratio
SINR	signal-to-interference-plus-noise ratio
SIR	signal-to-interference ratio
TDMA	time division multiple access
TF	time-frequency
US	uncorrelated scattering
WH	Weyl-Heisenberg
WSS	wide-sense stationary
WSSUS	wide-sense stationary uncorrelated scattering

References

- 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.
- Abramowitz, M. and Stegun, I. A. (1972), *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, no. 55 in Applied Mathematics Series, Washington, D.C. 20402: U.S. Government Printing Office, 10th ed.
- Aeron, S. and Saligrama, V. (2007), “Wireless ad hoc networks: Strategies and scaling laws for the fixed SNR regime,” *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2044–2059.
- Anderson, G. W., Guionnet, A., and Zeitouni, O. (2009), *An Introduction to Random Matrices*, no. 118 in Cambridge Studies in Advanced Mathematics, Cambridge Univ. Press.
- Bai, Z. D. (1999), “Methodologies in spectral analysis of large dimensional random matrices,” *Statistica Sinica*, vol. 9, pp. 611–677.
- Bello, P. A. (1963), “Characterization of randomly time-variant linear channels,” *IEEE Trans. Commun.*, vol. 11, no. 4, pp. 360–393.
- Bölcskei, H., Nabar, R. U., Oyman, Ö., and Paulraj, A. J. (2006), “Capacity scaling laws in MIMO relay networks,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1433–1444.
- Carleial, A. (1978), “Interference channels,” *IEEE Trans. Inf. Theory*, vol. 24, no. 1, pp. 60–71.
- Christensen, O. (2003), *An Introduction to Frames and Riesz Bases*, Boston, MA, U.S.A.: Birkhäuser.

- Cover, T. M. and Thomas, J. A. (2006), *Elements of Information Theory*, New York, NY, U.S.A.: Wiley, 2nd ed.
- Dana, A. F. and Hassibi, B. (2006), “On the power efficiency of sensory and ad-hoc wireless networks,” *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 2890–2914.
- 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.
- Durisi, G., Morgenshtern, V. I., Bölcskei, H., Schuster, U. G., and Shamai (Shitz), S. (2011), “Information theory of underspread WSSUS channels,” in F. Hlawatsch and G. Matz (Eds.), *Wireless Communications over Rapidly Time-Varying Channels*, Academic Press.
- Durisi, G., Schuster, U. G., Bölcskei, H., and Shamai (Shitz), S. (2010), “Noncoherent capacity of underspread fading channels,” *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 367–395.
- Etkin, R. and Tse, D. N. C. (2006), “Degrees of freedom in some underspread MIMO fading channels,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1576–1608.
- Franceschetti, M., Dousse, O., Tse, D. N. C., and Thiran, P. (2007), “Closing the gap in the capacity of wireless networks via percolation theory,” *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 1009–1018.
- Franceschetti, M. and Meester, R. (2007), *Random Networks for Communication*, no. 24 in Cambridge Series in Statistical and Probabilistic Mathematics, New York: Cambridge Univ. Press.
- Gallager, R. G. (1968), *Information Theory and Reliable Communication*, New York, NY, U.S.A.: Wiley.
- Gastpar, M. and Vetterli, M. (2005), “On the capacity of large Gaussian relay networks,” *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 765–779.
- Gröchenig, K. (2001), *Foundations of Time-Frequency Analysis*, Boston, MA, U.S.A.: Birkhäuser.
- Grossglauser, M. and Tse, D. N. C. (2002), “Mobility increases the

- capacity of ad hoc wireless networks,” *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 477–486.
- Gupta, P. and Kumar, P. R. (2002), “The capacity of wireless networks,” *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404.
- (2003), “Towards an information theory of large networks: An achievable rate region,” *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1877–1894.
- Hochwald, B. M. and Marzetta, T. L. (2000), “Unitary space–time modulation for multiple-antenna communications in Rayleigh flat fading,” *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 543–564.
- Hoeffding, W. (1963), “Probability inequalities for sums of bounded random variables,” *J. Am. Statist. Assoc.*, pp. 13–30.
- Horn, R. A. and Johnson, C. R. (1985), *Matrix Analysis*, Cambridge, U.K.: Cambridge Univ. Press.
- Jovičić, A., Viswanath, P., and Kulkarni, S. R. (2004), “Upper bounds to transport capacity of wireless networks,” *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2555–2565.
- Kennedy, R. S. (1969), *Fading Dispersive Communication Channels*, New York, NY, U.S.A.: Wiley.
- Koch, T. (2009), *On heating up and fading in communication channels*, Ph.D. thesis, ETH Zurich, Konstanz, Germany.
- Kozek, W. (1997), *Matched Weyl-Heisenberg Expansions of Nonstationary Environments*, Ph.D. thesis, Vienna University of Technology, Department of Electrical Engineering, Vienna, Austria.
- Kreyszig, E. (1989), *Introductory Functional Analysis with Applications*, Wiley.
- Landau, H. J. and Pollak, H. O. (1960), “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.
- Lapidoth, A. (2005), “On the asymptotic capacity of stationary Gaus-

- sian 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.
- Lapidoth, A. and Shamai (Shitz), S. (2002), “Fading channels: How perfect need “perfect side information” be?” *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134.
- Lévêque, O. and Telatar, I. E. (2005), “Information-theoretic upper bounds on the capacity of large extended ad hoc wireless networks,” *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 858–865.
- Liang, Y. and Veeravalli, V. V. (2004), “Capacity of noncoherent time-selective Rayleigh-fading channels,” *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3095–3110.
- Lütkepohl, H. (1996), *Handbook of Matrices*, Chichester, U.K.: Wiley.
- Marčenko, V. A. and Pastur, L. A. (1967), “Distribution of some sets of random matrices,” *Math. USSR-Sb*, vol. 1, pp. 457–483.
- 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.
- Matz, G., Schafhuber, D., Hartmann, K. G. 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.
- Maurer, A. (2003), “A bound on the deviation probability for sums of non-negative random variables,” *J. Inequalities Pure Appl. Math*, vol. 4, no. 1.
- Médard, M. (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.
- 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.
- Müller, R. R. (2003), “Applications of large random matrices in

- communications engineering,” in *Proc. Int. Conf. on Advances Internet, Process., Syst., Interdisciplinary Research (IPSI)*, Sveti Stefan, Montenegro.
- Özgür, A. (2009), *Fundamental limits and optimal operation in large wireless networks*, Ph.D. thesis, EPFL.
- Özgür, A., Lévêque, O., and Tse, D. N. C. (2007), “Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks,” *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3549–3572.
- Riegler, E., Morgenshtern, V. I., Durisi, G., Lin, S., Sturmfels, B., and Bölcskei, H. (2011), “Noncoherent SIMO pre-log via resolution of singularities,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, St. Petersburg, Russia. To appear.
- Rudin, W. (1987), *Real and Complex Analysis*, New York, NY, U.S.A.: McGraw-Hill, 3rd ed.
- Shannon, C. (1948), “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656.
- Silverstein, J. W. (1995), “Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices,” *J. Multivariate Anal.*, vol. 55, pp. 331–339.
- Slepian, D. (1976), “On bandwidth,” *Proc. IEEE*, vol. 64, no. 3, pp. 292–300.
- Slepian, D. and Pollak, H. O. (1961), “Spheroidal wave functions, Fourier analysis, and uncertainty—I,” *Bell Syst. Tech. J.*, vol. 40, no. 1, pp. 43–63.
- Tao, T. (2005), “An uncertainty principle for cyclic groups of prime order,” *Math. Res. Lett.*, vol. 12, no. 1, pp. 121–127.
- Telatar, Í. E. (1999), “Capacity of multi-antenna Gaussian channels,” *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595.
- Tse, D. and Viswanath, P. (2005), *Fundamentals of Wireless Communication*, Cambridge Univ. Press.
- Tulino, A. M. and Verdú, S. (2004), *Random Matrix Theory and Wireless Communications, Foundations and Trends in Commun. and Inf. Theory*, vol. 1, NOW Publishers.
- Wang, B., Zhang, J., and Zheng, L. (2006), “Achievable rates and

REFERENCES

- scaling laws of power-constrained wireless sensory relay networks,” *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4084–4104.
- 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.
- Wyner, A. D. (1966), “The capacity of band-limited Gaussian channel,” *Bell Syst. Tech. J.*, vol. 45, no. 3, pp. 359–395.
- Xie, L.-L. and Kumar, P. R. (2004), “A network information theory for wireless communication: Scaling laws and optimal operation,” *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 748–767.
- Xue, F. and Kumar, P. R. (2006), *Scaling Laws for Ad Hoc Wireless Networks: An Information Theoretic Approach, Foundations and Trends in Networking*, vol. 1, NOW Publishers.
- Yeh, S.-p. and Lévêque, O. (2007), “Asymptotic capacity of multi-level amplify-and-forward relay networks,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1436–1440, Nice, France.
- 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.

List of Publications

- Durisi, G., Morgenshtern, V. I., and Bölcskei, H. (2009), “On the sensitivity of noncoherent capacity to the channel model,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2174–2178, Seoul, Korea.
- (2010), “Sensitivity of the noncoherent capacity of WSSUS fading channels,” To be submitted.
- Durisi, G., Morgenshtern, V. I., Bölcskei, H., Schuster, U. G., and Shamai (Shitz), S. (2011), “Information theory of underspread WSSUS channels,” in F. Hlawatsch and G. Matz (Eds.), *Wireless Communications over Rapidly Time-Varying Channels*, Academic Press.
- Morgenshtern, V. I. and Bölcskei, H. (2005), “On the value of cooperation in interference relay networks,” in *Proc. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL.
- (2006), “Random matrix analysis of large relay network,” in *Proc. Allerton Conf. Commun., Contr., Comput.*, pp. 106–112, Monticello, IL.
- (2007), “Crystallization in large wireless networks,” *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3319–3349.
- (2011), “A short course on frame theory,” in T. Chen, R. Dinesh, and E. Serpedin (Eds.), *Mathematical Foundations for Signal Processing, Communications and Networking*, CRC Press. To appear.
- Morgenshtern, V. I., Bölcskei, H., and Nabar, R. U. (2005), “Distributed orthogonalization in large interference relay networks,”

- in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1211–1215, Adelaide, Australia.
- Morgenshtern, V. I., Durisi, G., and Bölcskei, H. (2010), “The SIMO pre-log can be larger than the SISO pre-log,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 320–324, Austin, TX, U.S.A.
- Morgenshtern, V. I., Durisi, G., Riegler, E., Lin, S., Sturmfels, B., and Bölcskei, H. (2011), “Capacity pre-log of noncoherent SIMO correlated-block-fading channel,” To be submitted.
- Riegler, E., Morgenshtern, V. I., Durisi, G., Lin, S., Sturmfels, B., and Bölcskei, H. (2011), “Noncoherent SIMO pre-log via resolution of singularities,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, St. Petersburg, Russia. To appear.

Curriculum Vitae

Veniamin I. Morgenshtern
born 23 June 1982
Leningrad, Russia

Education

09/2004–07/2010 working towards Doctor of Sciences degree at
ETH Zurich
09/1999–06/2004 Studies in Mathematics and Software
Engineering, Saint-Petersburg State University,
Russia; Dipl. Math. with Honors

Research

07/2010–present Postdoctoral researcher, Communication
Technology Laboratory, ETH Zurich,
Switzerland
09/2004–07/2010 Research Assistant, Communication Technology
Laboratory, ETH Zurich, Switzerland
08/2007–11/2007 Graduate Student Research Collaborator,
University of Illinois at Urbana-Champaign,
U.S.A.
01/2004–06/2004 Student Researcher, Communication Technology
Laboratory, ETH Zurich, Switzerland