Diss. ETH No. 18339

On Heating Up and Fading in Communication Channels

A dissertation submitted to the
ETH Zurich
for the degree of
Doctor of Sciences ETH Zurich

presented by

Tobias Koch

M.Sc. ETH Zurich
born on December 22, 1978
citizen of Büttikon AG

accepted on the recommendation of
Prof. Dr. Amos Lapidoth, examiner
Prof. Dr. Helmut Bölcskei, co-examiner
Prof. Dr. Emre Telatar, co-examiner

Hartung-Gorre Verlag, Konstanz, May 2009

To my dad

I may not have gone where I intended to go,
but I think I have ended up where I intended to be.

Douglas Adams (1952–2001)

Acknowledgments

It is common to start a thesis with a quote. My two favorite quotes were one by my advisor Prof. Amos Lapidoth ("you will never graduate") and one by Douglas Adams, best known as the author of *The Hitchhiker's Guide to the Galaxy* ("I may not have gone where I intended to go, but I think I have ended up where I intended to be"). As you can see, I decided for the latter one. It describes accurately the process of writing a thesis: doing a Ph.D. is more than just solving problems and writing them up in a book. Before I was able to find a problem that I could finally solve, I had to fail miserably at a number of others. I talk myself into believing that this has nothing to do with me, but is part of doing research.

Amos was a great help during my four and a half years at the ISI and is largely responsible for the (positive) outcome of this thesis. I profited a lot from his good ideas, and I'm deeply grateful for the effort he made to teach me how to write papers, to give talks, and to do research in general. But most of all, I valued that I was allowed to disagree with him—even though he was mostly right in the end. I know that this is not something that can be taken for granted. Thank you!

A huge "thank you!" goes to Prof. Gerhard Kramer who was mentor during my internship at Bell Labs in New Jersey, USA. He helped me to find accommodation, was patient enough to spend hours with me discussing my project, and kindly took me along when he went out for dinner. Thanks to him, these ten weeks in the US were not only highly instructive, but also very enjoyable.

I would like to say "thank you!" to Prof. Andi Loeliger for the nice discussions we had and for the kind feedback he always gave me after my talks. I further would like to say "thank you!" to Prof. Helmut Bölcskei and Prof. Emre Telatar who kindly accepted to be co-examiners at my Ph.D. examination. I truly value their effort to read my thesis—their comments were a big help. A big "thank you!" goes also to Prof. Paul Sotiriadis who suggested the problem of *Channels that Heat Up*. His enthusiastic e-mails about our results were a huge morale booster. Furthermore, I would like to say "thank you!" to Prof. Ezio Biglieri and Prof. Angel Lozano for receiving me in Barcelona. I always felt welcome.

During my time at the ISI, I shared an office with Michèle. She had to put up with my good moods and my bad moods—and most of the time I was driving her crazy by playing with my pen. I hope she misses me a bit, too... A huge "thank you!" goes to the former and current members of the "IT gang": Dani, Ligong, Michèle, Natalia, Stefan, and Stephan. The discussions about our research problems, but also about more general problems in information theory, helped me a great deal. But most of all, I valued the great spirit we had in our group. I will keep the conferences we visited and the vacations we enjoyed together in good memory. I would further like to thank all the members of the ISI. Even though I'm too lazy to list all of them by name, I well remember our "scientific" meetings at the fussball table; the regular hours at the gym; the coffee breaks in the afternoon; the ski and hiking days; as well as our discussions about such diverse topics as computers, music, sports, women (yes, some engineers are also interested in women), culture in general, or chaos. Thank you! A special "thank you!" goes to Mehdi, without whom this thesis wouldn't have a title.

Sitting in front of a sheet of paper and trying to evaluate differential entropies (i.e., doing research) is fun, but it can make you forget about the "real life" that is happening around you. I was fortunate to have good friends with whom I spent my time with, and with whom I discussed problems that cannot be described with math and whose solutions are not of a scientific nature. I won't list your names here, but I'm sure you know that I'm referring to *You*. I can't thank you enough for your support!

And last but not least, I would like to thank my brothers, Florian and Nicolas, and my mom, without whom nothing would have been possible.

I dedicate this thesis to my dad, who regrettably passed away shortly after I started my Ph.D. I would give anything to show him the outcome of my work. The last thing he said to me was: "Du machsch das scho guet." I hope he was right.

Abstract

This dissertation studies two phenomena that affect the transmission of data: *heating up* and *fading*. In particular, the effect of these phenomena on channel capacity, which is the largest rate at which data transmission with arbitrarily lower error probability is possible, is investigated.

Heating up is relevant in on-chip communication, where multiple terminals that are located on the same microchip wish to communicate with each other. It accounts for thermal coupling of data and noise. Indeed, the data to be transmitted are corrupted by thermal noise, whose variance depends on the local temperature of the chip. Furthermore, the transmission of data is associated with dissipation of energy into heat and raises therefore the local temperature of the chip. This gives rise to a channel model where the variance of the additive noise is data-dependent. The capacity of this channel is studied at low and at high transmit powers. At low transmit powers, the slope of the capacity-vs-power curve at zero is computed, and it is shown that the heating-up effect is beneficial. At high transmit powers, it is demonstrated that the heating-up effect is detrimental. In fact, if the heat dissipates slowly then the capacity is bounded in the transmit power, i.e., the capacity does not tend to infinity as the allowed average power tends to infinity. A sufficient condition and a necessary condition for the capacity to be bounded is derived.

The results of the above analyses suggest that at low transmit powers heat sinks are not only unnecessary, but they even reduce the capacity by dissipating heat, which contains information about the transmitted signal. The results further accentuate the importance of an efficient heat sink at large transmit powers.

Fading occurs in wireless communication channels. In such channels the transmitted signal is not only corrupted by additive noise, but also by multiplicative noise, which accounts for the variation of the signal's attenuation. This multiplicative noise is referred to as *fading.* In contrast to many other information-theoretic studies, where it is assumed that the receiver has perfect knowledge of the fading, in this dissertation it is assumed that the transmitter and the receiver only know the statistics of the fading but not its realization.

First, the capacity of multiple-input multiple-output (MIMO) Gaussian flat-fading channels with memory is considered. Nonasymptotic upper and lower bounds on the capacity are derived, and their asymptotic behavior is analyzed in the limit as the signal-to-noise ratio (SNR) tends to infinity. In particular, upper bounds on the fading number (which is defined as the second-order term in the high-SNR expansion of capacity) and on the capacity pre-log (which is defined as the limiting ratio of capacity to $\log \text{SNR}$ as SNR tends to infinity) are computed. Furthermore, an approach to derive lower bounds on the fading number is proposed. This lower bound is applied to derive a lower bound on the fading number of spatially IID, zero-mean, MIMO Gaussian fading channels with memory. The derived upper and lower bounds on the fading number demonstrate that when the number of receive antennas does not exceed the number of transmit antennas, the fading number of spatially IID, zero-mean, slowly-varying, Gaussian fading channels is proportional to the number of degrees of freedom, i.e., to the minimum of the number of transmit and receive antennas.

Second, the capacity pre-log of single-input single-output (SISO) flat-fading channels with memory is studied. It is shown that, among all stationary and ergodic fading processes of a given spectral distribution function and whose law has no mass point at zero, the Gaussian process gives rise to the smallest pre-log. It is further demonstrated that the assumption that the fading law has no mass point at zero is essential in the sense that there exist stationary and ergodic fading processes of some spectral distribution function (and whose law has a mass point at zero) that give rise to a smaller pre-log than the Gaussian process of equal spectral distribution function. These results are then extended to multiple-input single-output (MISO) fading channels with memory.

Finally, the capacity of multipath (frequency-selective) fading channels is studied. It is shown that if the delay spread is large in the sense that the variances of the path gains decay exponentially or slower, then the capacity is bounded in the SNR. Thus, in this case the capacity does not grow to infinity as the SNR tends to infinity. In contrast, if the variances of the path gains decay faster than exponentially, then the capacity is unbounded in the SNR. It is further demonstrated that if the number of paths is finite, then the capacity pre-loglog, which is defined as the limiting ratio of capacity to $\log\log \mathrm{SNR}$ as SNR tends to infinity, is 1, irrespective of the number of paths.

The conclusions that can be drawn from the above described analyses of fading channels are manifold. First, the presence of multiple antennas at the transmitter and receiver is very beneficial, even if the receiver does not know the realization of the fading. Second, the Gaussian fading assumption in the analysis of fading channels at high SNR is conservative in the sense that for a large class of fading processes the Gaussian process gives rise to the smallest capacity pre-log. Third, at high SNR multipath fading channels with an infinite number of paths should not be approximated by multipath fading channels with a finite number of paths, since these channels possess completely different high-SNR capacity behaviors. And last but not least, the high-SNR asymptotic behavior of the capacity of fading channels is very sensitive to the employed channel model. Thus, in the information-theoretic analysis of fading channels at high SNR and in the evaluation of the results thereof, one should attach great importance to the channel model.

Keywords: Information theory, channel capacity, capacity per unit cost, channels with memory, high signal-to-noise ratio, on-chip communication, wireless communication, flat-fading channels, multipath fading channels.

Kurzfassung

Diese Dissertation behandelt zwei Phänomene, welche die Übertragung von Daten beeinträchtigen: Erhitzung und Schwund. Insbesondere wird untersucht, inwiefern diese Phänomene die Kanalkapazität, die definiert ist als die grösste Datenrate mit welcher Daten mit beliebig kleiner Fehlerwahrscheinlichkeit übertragen werden können, beeinflussen.

Das Problem der Erhitzung ist in der On-Chip Kommunikation von Bedeutung, wo mehrere Datenstationen, welche sich auf dem selben Mikrochip befinden, miteinander kommunizieren. Die übertragenen Daten werden üblicherweise durch thermisches Rauschen gestört, wobei die Varianz dieses Rauschens von der Temperatur des Mikrochips abhängt. Da bei der Datenübertragung Energie in Wärme umgewandelt wird, welche dann den Chip erhitzt, hängt die Varianz des thermischen Rauschens von der Energie der bereits übertragenen Daten ab. In dieser Dissertation wird die Kapazität dieses Kanals bei geringer Signalleistung und bei grosser Signalleistung untersucht. Im ersten Fall (d.h. bei geringer Signalleistung) wird die Steigung der Funktion, welche die Signalleistung auf die Kapazität abbildet, im Nullpunkt ermittelt, und es wird gezeigt dass in diesem Fall Erhitzung von Vorteil ist. Im zweiten Fall (d.h. bei grosser Signalleistung) wird aufgezeigt, dass sich Erhitzung nachteilig auf die Kapazität auswirkt: Wenn die Wärme nicht genug schnell abgeführt werden kann, dann strebt die Kapazität mit steigender Signalleistung nicht gegen Unendlich. Des Weiteren werden eine hinreichende und eine notwendige Bedingung dafür dass die Kapazität nicht gegen Unendlich strebt hergeleitet.

Die Resultate der oben beschriebenen Analyse deuten darauf hin, dass bei geringer Signalleistung eine Wärmesenke nicht nur unnötig ist, sondern sogar die Kapazität verringert, da sie Wärme abführt, die Information über die gesendeten Daten enthält. Des Weiteren heben die Resultate die Wichtigkeit einer effizienten Wärmesenke bei grosser Si-

gnalleistung hervor.

Schwund tritt in der drahtloser Kommunikation auf. Drahtlose Übertragung wird häufig durch ein Kanalmodell beschrieben in welchem das übertragene Signal nicht nur durch additives, sondern auch durch multiplikatives Rauschen gestört wird. Dieses multiplikative Rauschen wird *Fading* genannt. Im Gegensatz zu vielen informationstheoretischen Arbeiten, wo angenommen wird, dass der Empfänger das Fading perfekt kennt, wird in dieser Dissertation angenommen, dass sowohl Sender als auch Empfänger lediglich die Wahrscheinlichkeitsverteilung des Fadings kennen, aber nicht seine Realisation.

Zuerst wird die Kanalkapazität von Gauss'schen MIMO Flat-Fading Kanälen mit Gedächtnis untersucht (MIMO ist neudeutsch für Mehrfachantennen). Es werden nichtasymptotische untere und obere Schranken für die Kapazität hergeleitet sowie dessen asymptotisches Verhalten untersucht wenn der Störabstand gegen Unendlich wächst. Insbesondere werden obere Schranken für die Fading Number (welche definiert ist als der Term zweiter Ordnung der asymptotischen Reihenentwicklung der Kanalkapazität bei hohem Störabstand) und den Pre-Log (welcher definiert ist als das asymptotische Verhältnis der Kapazität zum Logarithmus des Störabstandes wenn der Störabstand gegen Unendlich strebt) berechnet. Des Weiteren wird eine Methode zur Berechnung von unteren Schranken für die Fading Number eingeführt. Diese Methode wird angewendet, um eine untere Schranke für die Fading Number von räumlich-IID, mittelwertfreien, Gauss'schen MIMO Fading Kanälen herzuleiten. Die hergeleiteten oberen und unteren Schranken für die Fading Number zeigen, dass wenn die Anzahl Antennen am Empfänger nicht grösser ist als die Anzahl Antennen am Sender, dann ist die Fading Number von räumlich-IID, mittelwertfreien, langsam-variierenden, Gauss'schen Fading Kanälen proportional zur Anzahl Freiheitsgrade des Systems, d.h. zum Minimum der Anzahl Sende- und der Anzahl Empfangsantennen.

In einem zweiten Schritt wird der Pre-Log von (nicht notwendigerweise Gauss'schen) Flat-Fading Kanälen mit einer einzelnen Antenne am Sender und am Empfänger untersucht. Es wird gezeigt, dass von allen stationären und ergodischen Fading Prozessen mit einer gegebenen spektralen Verteilungsfunktion und einer (kumulative) Verteilungsfunktion die stetig im Nullpunkt ist, der Gauss'sche Prozess den kleinsten Pre-Log ergibt. Weiter wird aufgezeigt, dass die Annahme einer im Nullpunkt stetigen Verteilungsfunktion notwendig ist, sprich dass es stationäre und ergodische Fading Prozesse gibt mit einer bestimmten spektralen Verteilungsfunktion und einer (kumulativen) Verteilungsfunktion die nicht stetig im Nullpunkt ist, welche einen kleineren Pre-Log ergeben als der Gauss'sche Prozess mit derselben spektralen Verteilungsfunktion. Schliesslich wird die obige Aussage für Fading Kanäle mit mehreren Antennen am Sender und einer einzelnen Antenne am Empfänger erweitert.

Zum Schluss wird die Kanalkapazität von Fading Kanälen mit Mehrfachausbreitung untersucht. Es wird gezeigt, dass wenn der Delay Spread gross ist, sprich wenn die Varianzen der einzelnen Ausbreitungspfade exponentiell oder langsamer abfallen, dann strebt die Kanalkapazität mit steigendem Störabstand nicht gegen Unendlich. Andererseits, wenn die Varianzen der einzelnen Ausbreitungspfade schneller als exponentiell abfallen, dann strebt die Kanalkapazität mit steigendem Störabstand gegen Unendlich. Des Weiteren wird gezeigt, dass wenn die Anzahl Ausbreitungspfade endlich ist, dann ist der Pre-Loglog (welcher definiert ist als das asymptotische Verhältnis der Kapazität zum Logarithmus des Logarithmus des Störabstandes wenn der Störabstand gegen Unendlich strebt), unabhängig von der Anzahl Pfade, immer 1.

Die Resultate der oben beschriebenen Analysen von Fading Kanälen lassen folgende Schlüsse zu: Erstens, die Verwendung von mehreren Antennen am Sender und am Empfänger ist sehr nutzbringend, selbst wenn der Empfänger das Fading nicht perfekt kennt. Zweitens, die in informationstheoretischen Arbeiten häufig getroffene Annahme dass das Fading gaussverteilt ist, ist insofern konservativ, als in den meisten Fällen Gauss'sches Fading den kleinsten Pre-Log ergibt. Drittens, man sollte bei grossem Störabstand Mehrfachausbreitungskanäle mit einer unendlichen Anzahl Ausbreitungspfade nicht durch Mehrfachausbreitungskanäle mit einer endlich Anzahl Ausbreitungspfade annähern, da die Kapazitäten dieser beiden Kanäle völlig unterschiedliche aysmptotische Verhalten aufweisen. Und drittens, das asymptotische Verhalten der Kanalkapazität (wenn der Störabstand gegen Unendlich strebt) hängt stark vom gewählten Kanalmodell ab. Man sollte deshalb bei der informationstheoretischen Analyse von Fading Kanälen und bei der Auswertung deren Resultate dem Kanalmodell grosse Beachtung schenken.

Stichworte: Informationstheorie, Kanalkapazität, Capacity per Unit Cost, Kanäle mit Gedächtnis, high SNR, On-Chip Kommunikation, drahtlose Kommunikation, Fading Kanäle, Mehrfachausbreitung.

Contents

Chapter 1

Introduction

1.1 Motivation

In this dissertation, we study two phenomena that affect the transmission of data: *heating up* in on-chip communication and *fading* in wireless communication. We aim to characterize the impact of these phenomena on the information-theoretic limits of the above communication scenarios, i.e., we wish to characterize how these phenomena affect channel capacity.

On-chip communication names the scenario where multiple terminals that are located on the same microchip wish to communicate with each other. The data that are transmitted from one terminal to another are corrupted by thermal noise, and the variance of this noise depends on the local temperature of the chip. Since the transmission of data is associated with dissipation of energy into heat, it follows that the local temperature of the chip depends on the energy of the transmitted signal's history, which in turn implies that the variance of the thermal noise is data-dependent. We thus model on-chip communication by an additive noise channel where the noise variance depends on the power of the past channel inputs. This thermal coupling of noise and data is expected to become a bottleneck in future technology; the more so as it is the trend of modern microelectronics technology to pack more and faster operations within the smallest possible physical area.

Wireless communication names the scenario where multiple terminals communicate with each other without the aid of wires. *Inter alia*, the absence of wires is beneficial because it raises the mobility of each terminal and allows the operator of a wireless communication system more flexibility in operating his system. The other side of the coin is that wireless links are more sensitive to the nature of its surroundings and to disturbances thereof than wired links. This is modeled by a channel where the channel input is not only corrupted by additive (thermal) noise, but also by multiplicative noise. The multiplicative noise accounts for the variation of the signal's attenuation. We refer to it as *fading*, and we refer to the corresponding channel as a *fading channel*. To simplify analysis, information-theoretic studies often assume that the receiver has perfect knowledge of the fading. This is usually justified by arguing that the fading varies typically slowly over time, and one can therefore estimate it by transmitting a training-sequence at the beginning of each transmission. However, this simplification yields overly optimistic results, since the resources needed to estimate the fading are not taken into consideration. In this dissertation, we dispose of the above assumption. Specifically, we study the information-theoretic limits of fading channels when the receiver only knows the statistics of the fading, but not its realization. We shall refer to models where the realization of the fading is unknown to the transmitter and the receiver as *noncoherent* channel models.

While on-chip communication and wireless communication seem to have little in common, their channel models are very similar. Indeed, the channel that models on-chip communication is an additive noise channel whose noise variance depends on the channel inputs. Likewise, when the realization of the fading is unknown at the receiver, a fading channel is statistically equivalent to an additive noise channel (without fading), where the law of the noise depends on the channel inputs. Thus, both scenarios can be modeled by an additive noise channel with input-dependent noise.

The dependence of the noise on the channel inputs makes it all but impossible to obtain analytic results for the whole capacity-vs-transmit-power curve. We therefore resort to asymptotic analyses of capacity in the limit as the transmit power tends to zero and to infinity. These analyses provide an indication of how capacity grows with the allowed transmit power.

1.2 Outline and Contributions

This dissertation is organized as follows. Chapter 2 introduces the notions of achievable rate and channel capacity. Chapter 3 addresses the capacity of on-chip communication channels. Chapter 4 studies the capacity of multiple-input multiple-output (MIMO) Gaussian flat-fading channels. Chapter 5 demonstrates that for single-input single-output (SISO) flat-fading channels, Gaussian fading is (typically) the worst fading. Chapter 6 investigates the capacity of multipath (frequency-selective) fading channels. And Chapter 7 concludes with a summary and discussion of our results.

In the following we summarize the main contributions of this dissertation.

Channels that heat up (Chapter 3) We study a model for on-chip communication and analyze its capacity. At low transmit powers, we compute the capacity per unit cost, which describes the slope of the capacity-vs-power curve at zero. At large transmit powers, we exhibit a necessary and a sufficient condition under which the capacity is bounded in the transmit power, i.e., under which the capacity does not tend to infinity as the available transmit power tends to infinity. Our results accentuate the importance of heat sinks at high transmit powers, while they suggest that at low transmit powers heat sinks are not only unnecessary, but that they even reduce channel capacity.

Degrees of freedom and the fading number (Chapter 4) We study the capacity of noncoherent MIMO Gaussian flat-fading channels with memory. We propose nonasymptotic upper and lower bounds on the capacity. The upper bounds are then used to analyze the asymptotic behavior of capacity at high signal-to-noise ratio (SNR): for the cases where the fading process is of finite entropy rate we compute upper bounds on the fading number (which is defined as the second-order term in the high-SNR expansion of capacity); for the cases where the entropy rate of the fading process is infinite, we compute upper bounds on the capacity pre-log (which is defined as the limiting ratio of capacity to log SNR as SNR tends to infinity). We further propose an approach to derive lower bounds on the fading number of MIMO fading channels. This approach is applied to derive a lower bound on the fading number of spatially IID, zero-mean, MIMO Gaussian fading channels with memory. Our upper and lower bounds on the fading number demonstrate that when the number of receive antennas does not exceed the number of transmit antennas, the fading number of spatially IID, zero-mean, slowly-varying, MIMO Gaussian fading channels is proportional to the number of degrees of freedom, i.e., to the minimum number of transmit and receive antennas.

Gaussian fading is the worst fading (Chapter 5) We study a noncoherent SISO flat-fading channel with memory. We show that, among all stationary and ergodic fading processes of a given spectral distribution function and whose law has no mass point at zero, the Gaussian process gives rise to the smallest capacity pre-log. We further demonstrate that the assumption that the fading has no mass point is zero is essential in the sense that there exist stationary and ergodic processes of some spectral distribution function (and whose law has a mass point at zero) that give rise to a smaller pre-log than the Gaussian process of equal spectral distribution function. We then extend these results to multiple-input single-output (MISO) fading channels with memory.

Multipath fading channels (Chapter 6) We study a noncoherent multipath (frequency-selective) fading channel. We demonstrate that if the delay spread is large in the sense that the variances of the path gains decay exponentially or slower, then the capacity is bounded in the SNR. In contrast, if the variances of the path gains decay faster than exponentially, then the capacity is unbounded in the SNR. We further show that if the number of paths is finite, then at high SNR capacity grows double-logarithmically with the SNR, and the capacity pre-loglog (which is defined as the limiting ratio of the capacity to log log SNR as SNR tends to infinity) is 1, irrespective of the number of paths.

1.3 Notation

In the following we introduce the notation that is used throughout this dissertation. Notation that is specific to a chapter will be introduced there.

Unless otherwise stated, we use upper case letters from random quantities and lower case letters for their realizations. We denote vectors by boldface letters, and matrices by upper case letters of a special font, e.g., A for a deterministic matrix and $\mathbb{A}$ for a random matrix. The (j,ℓ)-th entry of a random matrix $\mathbb{A}$ will be denoted by $A(j,\ell)$, and its realization will be denoted by $a(j,\ell)$, where we implicitly assume that the indices j and ℓ are within the range of the matrix. Likewise, we denote the j-th entry of a random vector $\mathbf{A}$ by $A(j)$, and its realization by $a(j)$. We use $\|\cdot\|$ to denote the Euclidean norm of vectors or the Euclidean operator norm of matrices, i.e.,

$$\|\mathbf{a}\| = \sqrt{\sum_{\ell} |a(\ell)|^2},$$

$$\|\mathsf{A}\| = \max_{\|\hat{\mathbf{x}}\|=1} \|\mathsf{A}\hat{\mathbf{x}}\|.$$

Thus $\|\mathsf{A}\|$ is the maximal singular value of A.

We shall use $\det(\cdot)$ to denote the determinant and $\operatorname{tr}(\cdot)$ to denote the trace of a matrix. We shall further use $(\cdot)^*$ to denote conjugation, $(\cdot)^{\mathsf{T}}$ to denote the transpose of a matrix, and $(\cdot)^\dagger$ to denote Hermitian conjugation, i.e., $A^\dagger = \left(\mathsf{A}^*\right)^{\mathsf{T}}$.

The Frobenius norm of matrices is denoted by $\|\cdot\|_{\mathrm{F}}$; it is given by

$$\|\mathsf{A}\|_{\mathrm{F}} = \sqrt{\operatorname{tr}(\mathsf{A}^\dagger\mathsf{A})}.$$

Note that for any matrix A

$$\|\mathsf{A}\| \leq \|\mathsf{A}\|_{\mathrm{F}}.$$

We denote the $n \times n$ identity matrix by I_n.

We shall denote the indicator function by $\mathrm{I}\{\text{statement}\}$. It is 1 if the statement is true and 0 if it is false.

The set $\mathbb{R}$ denotes the set of real numbers, $\mathbb{C}$ denotes the set of complex numbers, $\mathbb{Z}$ denotes the set of integers, $\mathbb{N}$ denotes the set of positive integers, and $\mathbb{N}_0$ denotes the set of nonnegative integers. We shall denote the square-root of -1 by i, i.e., $\mathsf{i} = \sqrt{-1}$.

We denote the floor function by $\lfloor \cdot \rfloor$, and the ceiling function by $\lceil \cdot \rceil$. Thus $\lfloor a \rfloor$ denotes the largest integer that is less than or equal to a, and $\lceil a \rceil$ denotes the smallest integer that is greater than or equal to a. We shall use a combination of superscripts and subscripts to address sequences, e.g., A_m^n denotes the sequence $A_m, \ldots, A_n$. We shall denote the *limit superior* by $\overline{\lim}$ and the *limit inferior* by $\underline{\lim}$.

All rates specified in this dissertation are in nats per channel use. We use $\log(\cdot)$ to denote the natural logarithm function.

Chapter 2

Channel Capacity

2.1 Achievable Rates and Channel Capacity

This section provides the information-theoretic framework for studying communication channels. *Inter alia*, we introduce the notions of an achievable rate and of channel capacity.

We consider the communication system depicted in Figure 2.1. The message M represents the data we wish to transmit. It could be, for example, a file that we want to send to a friend, or music that we would like to store on a harddrive. We assume that M is uniformly distributed over the set $\mathcal{M} = \{1, \ldots, |\mathcal{M}|\}$, where $|\mathcal{M}|$ is some positive integer. The encoder maps the message to the length-n sequence $X_1, \ldots, X_n$, where n is called the *blocklength*, and where the sequence takes place in $\mathcal{X}$ (in this dissertation, $\mathcal{X}$ will be either the set of real numbers or the set of complex numbers). In the absence of feedback, the sequence X_1^n is a function of the message M, i.e., $X_1^n = \phi_n(M)$ for some mapping $\phi_n : \mathcal{M} \to \mathcal{X}^n$. If there is a feedback link, then X_k, $k = 1, \ldots, n$ is not only a function of the message M but also of the past channel output symbols Y_1^{k-1}, which take value in $\mathcal{Y}$, i.e., $X_k = \varphi_n^{(k)}(M, Y_1^{k-1})$ for some mapping $\varphi_n^{(k)} : \mathcal{M} \times \mathcal{Y}^{k-1} \to \mathbb{R}$. (Again, in this dissertation, $\mathcal{Y}$ will be either the set of real numbers or the set of complex numbers.) The receiver guesses the transmitted message M based on the n channel output symbols Y_1^n, i.e., $\hat{M} = \psi_n(Y_1^n)$ for some mapping $\psi_n : \mathcal{Y}^n \to \mathcal{M}$.

We describe the channel by the *channel law* W, where $W\bigl(\cdot \bigm| x_1^k, y_1^{k-1}\bigr)$ is the probability distribution of Y_k corresponding to the present and past channel inputs x_1^k and past channel outputs y_1^{k-1}. Roughly speaking, W specifies with what probability we observe y_k at the receiver

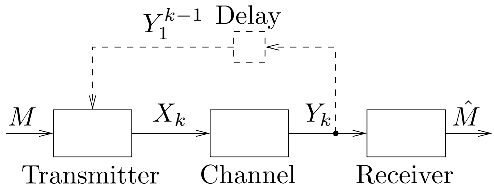

Figure 2.1: A schema of the communication system.

when the transmitter emitted the sequence x_1^k and the receiver observed so far the sequence y_1^{k-1}. The channel law defines the problem we are studying, i.e, it determines whether we study, say, communication in electronic circuits or communication in fading channels. We shall specify the considered channel law in each chapter.

Often the transmitter is permitted to transmit only sequences X_1^n whose energy is smaller than a certain number. The practical reason behind this constraint could be that the transmitter is driven by a battery and one wishes to communicate in an efficient way so as to ensure that the battery does not discharge too soon or, as in the example of wireless communication, that one would like to keep electromagnetic emissions small in order to reduce electromagnetic pollution. When $\mathcal{X}$ is the set of real or of complex numbers, common constraints are the *average-power constraint*

$$\frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\left[|X_k|^2\right] \leq \mathsf{P} \tag{2.1}$$

(where the average is over all realizations of X_1^n and Y_1^n) or the *peak-power constraint*

$$|X_k|^2 \leq \mathsf{A}^2, \quad k = 1, \ldots, n \qquad \text{with probability one.} \tag{2.2}$$

We shall specify in each chapter which power constraint we impose.

A *rate* R (in nats per channel use) is said to be *achievable* if for every $\delta > 0$ there exist sequences of mappings $\{\phi_n,\, n \in \mathbb{N}\}$ (without feedback) or $\left\{\left(\varphi_n^{(1)}, \ldots, \varphi_n^{(n)}\right),\, n \in \mathbb{N}\right\}$ (with feedback) and $\{\psi_n,\, n \in \mathbb{N}\}$

such that for each $n \in \mathbb{N}$

$$\frac{\log |\mathcal{M}|}{n} > R - \delta,$$

and such that the error probability $\Pr\big(\hat{M} \neq M\big)$ tends to zero as n goes to infinity. The *capacity* is defined as the supremum of all achievable rates. We shall denote it by C when there is no feedback, and we add the subscript "FB" to indicate that there is a feedback link. Clearly

$$C \leq C_{\text{FB}} \tag{2.3}$$

as we can always ignore the feedback.

In the absence of feedback, the *information capacity* is defined as [5, Ch. 8]

$$C_{\text{Info}} \triangleq \varliminf_{n\to\infty} \frac{1}{n} \sup I(X_1^n; Y_1^n), \tag{2.4}$$

where the supremum is over all joint distributions on $X_1, \ldots, X_n$ (possibly satisfying some power constraint). When there is a feedback link, then we define the information capacity as

$$C_{\text{Info,FB}} \triangleq \varliminf_{n\to\infty} \frac{1}{n} \sup I(M; Y_1^n), \tag{2.5}$$

where the supremum is over all mappings $\varphi_n^{(1)}, \ldots, \varphi_n^{(n)}$ (possibly satisfying some power constraint). By Fano's inequality [5, Thm. 2.11.1] no rate above C_{Info} and $C_{\text{Info,FB}}$ is achievable, i.e.,

$$C \leq C_{\text{Info}} \qquad \text{and} \qquad C_{\text{FB}} \leq C_{\text{Info,FB}}. \tag{2.6}$$

We say that a channel is memoryless if for any $k \in \mathbb{N}$ and for any Borel set $\mathcal{B} \subseteq \mathcal{Y}$

$$W\big(Y_k \in \mathcal{B} \,\big|\, x_1^k, y_1^{k-1}\big) = W\big(Y_k \in \mathcal{B} \,\big|\, x_k\big), \quad \big(x_1^k \in \mathcal{X}^k,\, y_1^{k-1} \in \mathcal{Y}^{k-1}\big).$$

For memoryless channels, it was shown by Shannon [40] that, in the absence of feedback, the information capacity is equal to the capacity, i.e.,

$$C = \lim_{n\to\infty} \frac{1}{n} \sup I(X_1^n; Y_1^n) = \sup I(X_1; Y_1).$$

Shannon further showed that in this case feedback does not help, i.e., that $C_{\text{FB}} = C$, see also [5, Sec. 8.12].

The channels we study in this thesis are not memoryless and it is therefore *prima facie* not clear whether the information capacity is equal to the capacity. In the absence of feedback, this claim was shown to hold for a class of channels that are called *information stable*; see [19, Thm. 4.5.1] (Ihara refers to such channels as *stable* [19, Def. 4.5.2]). Regrettably, the definition of this class of channels can be roughly paraphrased by saying that a channel is information stable if the capacity is equal to the information capacity. Thus, proving that a channel is information stable is usually as hard as proving that C_{Info} is achievable.

In this dissertation we shall not resort to the notion of information stability. We either derive achievable rates (as in Chapter 3), or we present references for the achievability of the information capacity and proceed by analyzing C_{Info} (as in Chapters 4 and 5), or we study the information capacity and indicate under what conditions it is equal to the capacity (as in Chapter 6). Since the information capacity constitutes an upper bound on the capacity, it provides a good estimate on what rates are supported by a channel.

2.2 A General Upper Bound on Mutual Information

Computing the information capacity is a difficult task: in the absence of feedback one needs to maximize

$$\frac{1}{n} I(X_1^n; Y_1^n)$$

for every n over all joint distributions on $X_1, \ldots, X_n$, so as to compute the limit of this maximum as n tends to infinity; when there is a feedback link, one needs to maximize

$$\frac{1}{n} I(M; Y_1^n)$$

for every n over all mappings $\varphi_n^{(1)}, \ldots, \varphi_n^{(n)}$, which is not likely to be the easier task. Information theorists typically attack this problem by deriving a lower bound and an upper bound on the information capacity, and they are happy if these bounds coincide.

Since any particular choice of distribution on $X_1, \ldots, X_n$ (without feedback) or $\varphi_n^{(1)}, \ldots, \varphi_n^{(n)}$ (with feedback) yields a lower bound, deriving lower bounds is, in general, easier than deriving upper bounds. In the following, we present a general upper bound on mutual information that is useful in deriving upper bounds on the information capacity:

Theorem 2.1. *Let the random variables X and Y take value in the sets $\mathcal{X}$ and $\mathcal{Y}$, let X be of law Q, and let the conditional law of Y, conditioned on X, be given by V. Assume that $\mathcal{X}$ and $\mathcal{Y}$ are separable metric spaces, and assume that for any Borel set $\mathcal{B} \subseteq \mathcal{Y}$ the mapping $x \mapsto V(\mathcal{B}|x)$ from $\mathcal{X}$ to $[0,1]$ is Borel measurable. Then*

$$I(X;Y) \leq \int D\big(V(\cdot|x) \,\big\|\, R(\cdot)\big) \, \mathrm{d}Q(x), \tag{2.7}$$

where $D(\cdot\|\cdot)$ denotes relative entropy, i.e.,

$$D(P_1\|P_0) = \begin{cases} \displaystyle\int \log \frac{\mathrm{d}P_1}{\mathrm{d}P_0} \, \mathrm{d}P_1 & \text{if } P_1 \ll P_0 \\ +\infty & \text{otherwise,} \end{cases} \tag{2.8}$$

and where R is any distribution on $\mathcal{Y}$.

Proof. For general alphabets $\mathcal{X}$ and $\mathcal{Y}$, this inequality was derived by Lapidoth and Moser [28, Thm. 5.1]. When $\mathcal{X}$ and $\mathcal{Y}$ are finite, it follows by Topsøe's identity [44]; see also [6, Thm. 3.4]. □

Theorem 2.1 demonstrates that, for any choice of R, the right-hand side of (2.7) yields an upper bound on the mutual information $I(X;Y)$. Thus, while choosing a distribution for X yields a lower bound on the information capacity, an upper bound can be found by choosing a distribution on Y. We use this technique in Chapters 3 and 6.

Chapter 3

Channels that Heat Up

3.1 Introduction

Heating in electronics is strongly related to performance limitation, aging, and reliability issues. High performance-density and small physical size make heating important and challenging to address. This is reinforced by the trend of modern (micro-)electronics technology to pack more and faster operations within the smallest possible physical area in order to increase performance, reduce cost and size, and therefore expand the potential applications of the product and make it more profitable.

Electrical power dissipation into heat raises the local temperature of the circuit, so the temperature depends on the circuit activity. The raised temperature results in higher intrinsic noise in the circuit which in turn reduces its effective communication and computation capacity. This "negative" performance feedback is expected to become an important issue in the years to come [1, 21, 46].

We aim to add this dimension to our understanding of the coupling mechanism between communication and computation performance and heating. To this end, a class of communication channels, whose noise power depends dynamically on their activity, is introduced and studied.

To motivate the mathematical development of this new class of channels we first discuss the underlying physical mechanism that connects circuit activity with power consumption and heating. Heating is unavoidable in electronic circuits since they convert part of the power they draw from the power supply network (and other circuits they are connected to) into heat.

Every circuit is a three dimensional object embedded inside the substrate and the surrounding packaging material. It generates heat, in a distributed manner, that is diffused according to the *heat diffusion equation*

$$\mathsf{C}_{\mathrm{hv}} \frac{\partial \mathsf{T}}{\partial t} = \nabla \cdot \left(\mathsf{k}\, \nabla \mathsf{T} \right) + \dot{q}. \tag{3.1}$$

Here C_{hv} is the volumetric heat capacity of the material, T is the point temperature, k is the thermal conductivity, and q is the heat flux generated by the distributed conversion of electrical power into heat [15,31]. (If other heat sources exist in the volume of the circuit, they should be included in the heat diffusion equation as well.)

In many cases, (3.1) can be simplified to the corresponding *ordinary differential equation* (3.2) providing a lumped model of the thermal dynamics

$$\mathsf{C}_{\mathrm{h}} \frac{\mathrm{d}\mathsf{T}}{\mathrm{d}t} = \frac{\mathsf{T}_{\mathrm{e}} - \mathsf{T}}{\mathsf{R}_{\mathrm{th}}} + \mathsf{P}_{\mathrm{th}}. \tag{3.2}$$

Here C_{h} is the lumped heat capacity of the circuit (partially including the substrate and packaging), R_{th} is the thermal resistance between the circuit and the external heat-sinking environment (e.g., the air) whose temperature is T_{e}, and P_{th} is the instantaneous electrical power in the circuit that is converted into heat.

Assuming that the environmental temperature T_{e} is fixed and that $\mathsf{T}(0) = \mathsf{T}_e$, the solution of (3.2) is given by

$$\mathsf{T}(t) = \mathsf{T}_{\mathrm{e}} + \frac{1}{\mathsf{C}_{\mathrm{h}}} \int_0^t e^{\frac{\xi - t}{\mathsf{R}_{\mathrm{th}} \mathsf{C}_{\mathrm{h}}}}\, \mathsf{P}_{\mathrm{th}}(\xi)\, \mathrm{d}\xi, \quad t > 0. \tag{3.3}$$

Now suppose that our circuit operates according to a reference clock of period τ, i.e., it transmits an output value $x_k \in \mathbb{R}$ at the beginning of every clock period $t_k = k\tau$, $k \in \mathbb{N}$. Further assume that the part of the electrical energy converted into heat due to the transmission of x_k is (proportional to) $x_k^2\tau$—a typical case in circuits when x_k is voltage or current. Then (3.3) can be approximated by its discrete version

$$\mathsf{T}_k = \mathsf{T}_{\mathrm{e}} + \frac{1}{\mathsf{C}_{\mathrm{h}}} \sum_{\ell=1}^{k-1} e^{-\frac{\tau}{\mathsf{R}_{\mathrm{th}} \mathsf{C}_{\mathrm{h}}}(k-\ell)}\, \tau\, x_\ell^2, \quad k \in \mathbb{N}. \tag{3.4}$$

By defining

$$a_\ell \triangleq \frac{\mathcal{T}}{\mathsf{C}_\mathrm{h}} e^{-\frac{\mathcal{T}}{\mathsf{R}_\mathrm{th}\mathsf{C}_\mathrm{h}}\ell}, \quad \ell \in \mathbb{N},$$

(3.4) becomes

$$\mathsf{T}_k = \mathsf{T}_\mathrm{e} + \sum_{\ell=1}^{k-1} a_{k-\ell}\, x_\ell^2, \quad k \in \mathbb{N}. \tag{3.5}$$

Equation (3.5) describes the relation between the local temperature of the electronic circuit and the circuit activity. Note that (3.5), being a general discrete-time convolution, also captures discretized versions of higher-order lumped approximations of the diffusion equation (3.1). It therefore represents a general model of the circuit-heating process, despite the simplifying assumptions used in its derivation.

Every electronic circuit has some intrinsically generated noise, which is added to the received signal and degrades its quality. In wideband circuits, the dominant type of noise is typically thermal noise [10,37,45]. Thermal noise is stationary Gaussian, and in most applications it can be considered white within the bandwidth of interest. The variance of the thermal noise θ^2 follows the Johnson-Nyquist formula

$$\theta^2 = \eta\, \mathsf{T} \mathsf{W}, \tag{3.6}$$

where W is the circuit's bandwidth, T is the absolute temperature of the circuit, and η is a proportionality constant.

Applying (3.5) to (3.6), and assuming that the intrinsic noise is only additive, yields a channel model where the variance θ^2 of the additive noise is determined by the history of the power of the transmitted signal, i.e.,

$$\theta^2(x_1, \ldots, x_{k-1}) = \sigma^2 + \sum_{\ell=1}^{k-1} \alpha_{k-\ell} x_\ell^2, \quad k \in \mathbb{N}, \tag{3.7}$$

where σ^2 and $\{\alpha_\ell\}$ are discussed in more detail in Section 3.2 (proportionality constants like η are incorporated into the parameters σ and $\{\alpha_\ell\}$).

While in today's microelectronics technology the increase in thermal noise due to data transmission is often marginal compared to the signal power and can therefore be neglected, there are scenarios where the thermal coupling of data and noise becomes significant. For example, consider a communication system where the transmission of data is assisted by a repeater, which receives the transmitted signal, amplifies it, and retransmits it. The signal at the repeater's input is typically corrupted by thermal noise, which is then amplified together with the signal. When the repeater is a monolithic circuit, the temperature of the repeater's receiving end (input)—and hence also the variance of the thermal noise—depends on the power of the signal sent out by the repeater's transmitting end (output), which in turn depends on the power of the signal sent out by the transmitter. Since the signal power at the repeater's output is much larger than that at the repeater's input, the increase in thermal noise due to retransmission of data can be significant compared to the repeater's input-signal power.

We also expect that the above channel model will be relevant to the next generation of nanoscale electronic technologies based on silicon or biological substrates [21, 32], as well as to the interface between nanocircuits and conventional microelectronics [53].

The rest of this chapter is organized as follows. Section 3.2 describes the channel model in more detail. Section 3.3 discusses channel capacity and lists some important properties thereof. Section 3.4 presents our main results. Sections 3.5 and 3.6 provide the proofs of these results. And Section 3.7 concludes with a summary and a discussion of our results.

3.2 Channel Model

We consider the communication system described in Section 2.1. Conditional on $(X_1, \ldots, X_k) = (x_1, \ldots, x_k) \in \mathbb{R}^k$, the time-$k$ channel output $Y_k \in \mathbb{R}$ is given by

$$Y_k = x_k + \sqrt{\left(\sigma^2 + \sum_{\ell=1}^{k-1} \alpha_{k-\ell} x_\ell^2\right)} \cdot U_k, \quad k \in \mathbb{N}, \tag{3.8}$$

where $\{U_k,\, k \in \mathbb{Z}\}$ is a zero-mean, unit-variance, stationary, weakly-mixing random process, drawn independently of M, and being of finite fourth moment and of finite differential entropy rate, i.e.,

$$\mathsf{E}\big[U_k^4\big] < \infty \quad \text{and} \quad h\big(\{U_k\}\big) \triangleq \lim_{n\to\infty} \frac{1}{n} h(U_1,\dots,U_n) > -\infty. \tag{3.9}$$

See [35] for a definition of weak mixing. For example, $\{U_k,\, k \in \mathbb{Z}\}$ could be a stationary ergodic Gaussian process [33] (see also [39, Sec. II]). In particular, the case of most interest is when $\{U_k,\, k \in \mathbb{Z}\}$ are independent and identically distributed (IID), zero-mean, unit-variance Gaussian random variables, and the reader is encouraged to focus on this case.

The parameter σ^2 is assumed to be positive. It accounts for the temperature of the device when the transmitter is silent. The coefficients α_ℓ, $\ell \in \mathbb{N}$ are nonnegative and bounded, i.e.,

$$\alpha_\ell \geq 0, \quad \ell \in \mathbb{N} \qquad \text{and} \qquad \sup_{\ell\in\mathbb{N}} \alpha_\ell < \infty. \tag{3.10}$$

They characterize the dissipation of the heat produced by transmitting message M. (It seems reasonable to assume that the sequence $\{\alpha_\ell\}$ is monotonically nonincreasing, i.e., $\alpha_\ell \geq \alpha_{\ell'}$ for $\ell \leq \ell'$. This assumption is, however, not required for the results that are derived in this chapter.)

An example of a heat dissipation profile that satisfies (3.10) is the *geometric* heat dissipation profile where $\{\alpha_\ell\}$ is a geometric sequence, i.e.,

$$\alpha_\ell = \rho^\ell, \quad \ell \in \mathbb{N} \tag{3.11}$$

for some $0 < \rho < 1$.

The heat dissipation depends *inter alia* on the efficiency of the heat sink that is employed in order to absorb the produced heat. In the above example (3.11), the heat sink's efficiency is described by the parameter ρ: the smaller ρ, the more efficient the heat sink. In general, an efficient heat sink is modeled by a heat dissipation profile for which the sequence $\{\alpha_\ell\}$ decays fast.

We study the above channel under an average-power constraint on the inputs, i.e., averaged over the message M and channel outputs Y_1^n, the sequence X_1^n satisfies

$$\frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\left[X_k^2\right] \leq \mathsf{P}, \tag{3.12}$$

and we define the signal-to-noise ratio (SNR) as

$$\text{SNR} \triangleq \frac{\mathsf{P}}{\sigma^2}. \tag{3.13}$$

Note 3.1. *The results presented in this chapter do not change when* (3.12) *is replaced by a* per-message *average-power constraint, i.e., when for each message* $m \in \mathcal{M}$ *and for any given sequence of output symbols* $Y_1^n = y_1^n$*, the sequence* x_1^n *satisfies*

$$\frac{1}{n}\sum_{k=1}^{n} x_k^2 \leq \mathsf{P}. \tag{3.14}$$

Indeed, all achievability results (which are based on schemes that ignore the feedback) are derived under (3.14)*, whereas all converse results are derived under* (3.12)*. Since all mappings* ϕ_n *and* $\varphi_n^{(1)}, \ldots, \varphi_n^{(n)}$ *that satisfy* (3.14) *also fulfill* (3.12)*, this implies that the achievability results as well as the converse results derived in this chapter hold irrespective of whether constraint* (3.12) *or* (3.14) *is imposed.*

The channel (3.8) is reminiscent of a multipath fading channel, when the transmitter and the receiver are not aware of the realization of the fading but only of its statistics. In fact, some of the techniques used this chapter are extended in Chapter 6 to study the high-SNR asymptotic behavior of the capacity of such channels.

3.3 Channel Capacity

The *capacity* C was defined in Section 2.1 as the supremum of all achievable rates. We denote by $C(\text{SNR})$ the capacity under the input constraint (3.12) when there is no feedback, and we add the subscript "FB" to indicate that there is a feedback link. Clearly

$$C(\text{SNR}) \leq C_{\text{FB}}(\text{SNR}) \tag{3.15}$$

as we can always ignore the feedback.

In the absence of feedback, the *information capacity* $C_{\text{Info}}(\text{SNR})$ is defined as (2.4)

$$C_{\text{Info}}(\text{SNR}) \triangleq \varliminf_{n\to\infty} \frac{1}{n} \sup I\big(X_1^n; Y_1^n\big), \tag{3.16}$$

where the supremum is over all joint distributions on $X_1, \ldots, X_n$ satisfying (3.12). When there is a feedback link, the information capacity is defined as (2.5)

$$C_{\text{Info,FB}}(\text{SNR}) \triangleq \varliminf_{n\to\infty} \frac{1}{n} \sup I\big(M; Y_1^n\big), \tag{3.17}$$

where the supremum is over all mappings $\varphi_n^{(1)}, \ldots, \varphi_n^{(n)}$ satisfying (3.12) (cf. Section 2.1). By Fano's inequality we have (2.6)

$$C(\text{SNR}) \leq C_{\text{Info}}(\text{SNR}) \tag{3.18}$$

and

$$C_{\text{FB}}(\text{SNR}) \leq C_{\text{Info,FB}}(\text{SNR}). \tag{3.19}$$

See [49] for conditions that guarantee that $C_{\text{Info}}(\text{SNR})$ is achievable. Note that the channel (3.8) is not stationary[1] since the variance of the additive noise depends on the time-index k. It is therefore *prima facie* not clear whether the inequalities in (3.18) and (3.19) hold with equality.

In this paper, we shall investigate the capacities $C(\text{SNR})$ and $C_{\text{FB}}(\text{SNR})$ at low SNR and at high SNR. To study capacity at low SNR, we compute the *capacities per unit cost* defined as [47]

$$\dot{C}(0) \triangleq \sup_{\text{SNR}>0} \frac{C(\text{SNR})}{\text{SNR}} \tag{3.20}$$

and

$$\dot{C}_{\text{FB}}(0) \triangleq \sup_{\text{SNR}>0} \frac{C_{\text{FB}}(\text{SNR})}{\text{SNR}}. \tag{3.21}$$

[1] By a *stationary channel* we mean a channel where for any stationary sequence of channel inputs $\{X_k,\, k \in \mathbb{Z}\}$ and corresponding channel outputs $\{Y_k,\, k \in \mathbb{Z}\}$ the pair $\{(X_k, Y_k),\, k \in \mathbb{Z}\}$ is jointly stationary.

It will become apparent later that the suprema in (3.20) and (3.21) are attained when SNR tends to zero. Note that (3.15) implies

$$\dot{C}(0) \leq \dot{C}_{\mathrm{FB}}(0). \tag{3.22}$$

At high SNR, we study conditions under which capacity is unbounded in the SNR. Notice that when the allowed transmit power is large, there is a trade-off between optimizing the present transmission and minimizing the interference to future transmissions. Indeed, increasing the transmission power may help to overcome the present ambient noise, but it also heats up the chip and thus increases the noise variance in future receptions. We shall see that, as we increase the allowed transmit power, the capacity does not necessarily tend to infinity.

3.4 Main Results

Our main results are presented in the following two sections. Section 3.4.1 focuses on capacity at low SNR and presents our results on the capacity per unit cost. Section 3.4.2 provides a sufficient condition and a necessary condition on $\{\alpha_\ell\}$ under which capacity is bounded in the SNR.

3.4.1 Capacity per Unit Cost

The results presented in this section hold under the additional assumptions that $\{U_k,\, k \in \mathbb{Z}\}$ is IID and that

$$\sum_{\ell=1}^{\infty} \alpha_\ell < \infty. \tag{3.23}$$

To shorten notation we denote this sum by

$$\alpha \triangleq \sum_{\ell=1}^{\infty} \alpha_\ell. \tag{3.24}$$

Proposition 3.1. *Consider the above channel model, and assume additionally that the sequence* $\{\alpha_\ell\}$ *satisfies* (3.23) *and that* $\{U_k,\, k \in \mathbb{Z}\}$ *is IID. Then*

$$\sup_{\mathrm{SNR}>0} \frac{C_{\mathrm{Info}}(\mathrm{SNR})}{\mathrm{SNR}} \geq \sup_{\mathrm{SNR}>0} \frac{C_{\alpha=0}(\mathrm{SNR})}{\mathrm{SNR}}, \tag{3.25}$$

where $C_{\alpha=0}(\text{SNR})$ *denotes the capacity of the channel*

$$Y_k = x_k + \sigma\, U_k,$$

which is a special case of (3.8) *for* $\alpha = 0$.

Proof. See Appendix A.1. □

For $\alpha = 0$, Equation (3.8) describes a channel with an ideal heat sink or, equivalently, a channel that does not heat up. Proposition 3.1 thus demonstrates that the heating up can only increase the information capacity per unit cost. In other words, at low SNR the heating-up effect is unharmful.

For *Gaussian* noise, i.e., when $\{U_k,\, k \in \mathbb{Z}\}$ is a sequence of IID, zero-mean, unit-variance *Gaussian* random variables, the heating-up effect is beneficial.

Theorem 3.2. *Consider the above channel model. Assume additionally that the sequence* $\{\alpha_\ell\}$ *satisfies* (3.23) *and that* $\{U_k,\, k \in \mathbb{Z}\}$ *is a sequence of IID, zero-mean, unit-variance, Gaussian random variables. Then, irrespective of whether feedback is available or not,*

$$\dot{C}_{\text{FB}}(0) = \dot{C}(0) = \lim_{\text{SNR}\downarrow 0} \frac{C(\text{SNR})}{\text{SNR}} = \frac{1}{2}\left(1 + \sum_{\ell=1}^{\infty} \alpha_\ell\right). \tag{3.26}$$

Proof. See Section 3.5. □

For example, for the geometric heat dissipation profile (3.11) we obtain from Theorem 3.2

$$\dot{C}_{\text{FB}}(0) = \dot{C}(0) = \frac{1}{2}\frac{1}{1-\rho}, \quad 0 < \rho < 1. \tag{3.27}$$

Thus the capacity per unit cost is monotonically *decreasing* in ρ.

The above result might be counterintuitive, because it suggests not to use heat sinks at low SNR. Nevertheless it can be heuristically explained by noting that the heating-up effect increases the *channel gain*[2]. Indeed, if we split up the channel output

$$Y_k = X_k + \sqrt{\left(\sigma^2 + \sum_{\ell=1}^{k-1} \alpha_{k-\ell} X_\ell^2\right)} \cdot U_k$$

into a data-dependent part

$$\tilde{X}_k = X_k + \sqrt{\left(\sum_{\ell=1}^{k-1} \alpha_{k-\ell} X_\ell^2\right)} \cdot U_k$$

and a data-independent part Z_k (with $\{Z_k,\ k \in \mathbb{Z}\}$ being a sequence of IID, zero-mean, variance-σ^2, Gaussian random variables drawn independently of $\{(U_k, X_k),\ k \in \mathbb{Z}\}$), then the channel gain G for (3.8) is given by

$$\mathsf{G} \triangleq \lim_{n\to\infty} \sup \frac{\sum_{k=1}^{n} \mathsf{E}\left[\tilde{X}_k^2\right]}{\sum_{k=1}^{n} \mathsf{E}[X_k^2]} = 1 + \sum_{\ell=1}^{\infty} \alpha_\ell, \tag{3.28}$$

where the supremum is over all joint distributions on $X_1, \ldots, X_n$ satisfying (3.12). Thus, in view of (3.28), Theorem 3.2 demonstrates that the capacity per unit cost is determined by the channel gain G. This result is not specific to (3.8) but has also been observed for other channel models. For example, the same is true for fading channels whenever the additive noise is Gaussian [30, 48].

3.4.2 Conditions for Bounded Capacity

While at low SNR the heating-up effect is beneficial, at high SNR it is detrimental. In fact, it turns out that the capacity can be bounded in the SNR, i.e., the capacity does not necessarily tend to infinity as the SNR tends to infinity. The following theorem provides a sufficient condition and a necessary condition on $\{\alpha_\ell\}$ for the capacity to be bounded. Note that the results presented in this section do not require

[2]The channel gain is given by the ratio of the "desired" power at the channel output to the "desired" power at the channel input.

the additional assumptions made in Section 3.4.1: we neither assume that the sequence $\{\alpha_\ell\}$ satisfies (3.23) nor that $\{U_k,\, k \in \mathbb{Z}\}$ is IID.

Theorem 3.3. *Consider the channel model described in Section 3.2. Then*

$$(i)\quad \left(\varliminf_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} > 0\right) \quad\Longrightarrow\quad \left(\sup_{\mathrm{SNR}>0} C_{\mathrm{FB}}(\mathrm{SNR}) < \infty\right) \tag{3.29}$$

$$(ii)\quad \left(\lim_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} = 0\right) \quad\Longrightarrow\quad \left(\sup_{\mathrm{SNR}>0} C(\mathrm{SNR}) = \infty\right), \tag{3.30}$$

where we define $a/0 \triangleq \infty$ for every $a > 0$ and $0/0 \triangleq 0$.

Proof. See Section 3.6. □

For example, for the geometric heat dissipation (3.11) we have

$$\lim_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} = \rho, \quad 0 < \rho < 1$$

and it follows from Theorem 3.3 that the capacity is bounded. On the other hand, for a supergeometric heat dissipation, i.e., when

$$\alpha_\ell = \rho^{\ell^\kappa}, \quad \ell \in \mathbb{N}$$

for some $0 < \rho < 1$ and $\kappa > 1$, we obtain

$$\lim_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} = \lim_{\ell\to\infty} \rho^{(\ell+1)^\kappa - \ell^\kappa} = 0$$

and Theorem 3.3 implies that the capacity is unbounded. Roughly speaking, we can say that when the sequence of coefficients $\{\alpha_\ell\}$ decays *not faster than geometrically*, the capacity is *bounded* in the SNR, and when the sequence of coefficients $\{\alpha_\ell\}$ decays *faster than geometrically*, the capacity is *unbounded* in the SNR.

Note 3.2. *For Part (i) of Theorem 3.3, the assumptions that the process $\{U_k,\, k \in \mathbb{Z}\}$ is weakly-mixing and that it has a finite fourth moment are not needed. These assumptions are only needed for Lemma 3.5 in* *the proof of Part (ii). In Part (ii) of Theorem 3.3, the condition on the left-hand side (LHS) of* (3.30) *can be replaced by*

$$\lim_{\ell\to\infty} \frac{1}{\ell} \log \frac{1}{\alpha_\ell} = \infty. \tag{3.31}$$

This condition (3.31) *is weaker than the original condition* (3.30) *because*

$$\left(\lim_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} = 0\right) \quad \Longrightarrow \quad \left(\lim_{\ell\to\infty} \frac{1}{\ell} \log \frac{1}{\alpha_\ell} = \infty\right).$$

If neither the LHS of (3.29) nor the LHS of (3.30) holds, i.e.,

$$\varlimsup_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} > 0 \qquad \text{and} \qquad \varliminf_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} = 0, \tag{3.32}$$

then the capacity can be bounded or unbounded. Example 3.1 exhibits a sequence $\{\alpha_\ell\}$ satisfying (3.32) for which the capacity is bounded, and Example 3.2 provides a sequence $\{\alpha_\ell\}$ satisfying (3.32) for which the capacity is unbounded. (The provided sequences $\{\alpha_\ell\}$ are not monotonically decreasing in ℓ. Consequently, Examples 3.1 and 3.2 are rather of mathematical than of practical interest. Nevertheless they show that when neither condition of Theorem 3.3 is satisfied, then one can construct simple examples yielding a bounded capacity or an unbounded capacity, thus demonstrating the difficulty of finding conditions that are necessary *and* sufficient for the capacity to be bounded.)

Example 3.1. *Consider the sequence $\{\alpha_\ell\}$ where all coefficients with an even index are equal to* 1*, and where all coefficients with an odd index are* 0*. It satisfies* (3.32) *because $\varlimsup_{\ell\to\infty} \alpha_{\ell+1}/\alpha_\ell = \infty$ and $\varliminf_{\ell\to\infty} \alpha_{\ell+1}/\alpha_\ell = 0$. Then the time-$k$ channel output Y_k corresponding to the channel inputs $(x_1, \ldots, x_k)$ is given by*

$$Y_k = x_k + \sqrt{\left(\sigma^2 + \sum_{\ell=1}^{\lfloor (k-1)/2 \rfloor} x_{k-2\ell}^2\right)} \cdot U_k, \quad k \in \mathbb{N}.$$

Thus at even times the output Y_{2k}, $k \in \mathbb{N}$ only depends on the "even" inputs $(X_2, X_4, \ldots, X_{2k})$, while at odd times the output Y_{2k+1}, $k \in \mathbb{N}_0$ only depends on the "odd" inputs $(X_1, X_3, \ldots, X_{2k+1})$. By proceeding along the lines of the proof of Part (i) of Theorem 3.3 while choosing *in* (3.62) $\beta = 1/y_{k-2}^2$, *it can be shown that the capacity of this channel is bounded.*[3]

Example 3.2. *Consider the sequence* $\{\alpha_\ell\}$ *where all coefficients with an even positive index are* 0, *and where all other coefficients are* 1. *(Again, we have* $\overline{\lim}_{\ell\to\infty} \alpha_{\ell+1}/\alpha_\ell = \infty$ *and* $\underline{\lim}_{\ell\to\infty} \alpha_{\ell+1}/\alpha_\ell = 0$.*) In this case the time-*k *channel output* Y_k *corresponding to* $(x_1, \ldots, x_k)$ *is given by*

$$Y_k = x_k + \sqrt{\left(\sigma^2 + \sum_{\ell=1}^{\lfloor k/2 \rfloor} x_{k-2\ell+1}^2\right)} \cdot U_k, \quad k \in \mathbb{N}.$$

Using Gaussian inputs of power 2P *at even times while setting the inputs to be zero at odd times, and measuring the channel outputs only at even times, reduces the channel to a memoryless additive noise channel and demonstrates (using the result of* [23]*) the achievability of*

$$R = \frac{1}{4}\log(1 + 2\,\mathrm{SNR}),$$

which is unbounded in the SNR.

The two seemingly-similar examples thus lead to completely different capacity results. The crucial difference between Example 3.1 and Example 3.2 is that in the former example at even times the interference is caused by the past channel inputs at *even* times, whereas in the latter example at even times the interference is caused by the past channel inputs at *odd* times. Thus, in Example 3.2, setting all "odd" inputs to zero cancels (at even times) the interference from past channel inputs and hence transforms the channel into an additive noise channel whose capacity is unbounded. Evidently, this approach does not work for Example 3.1.

3.5 Proof of Theorem 3.2

In Section 3.5.1 we derive an upper bound on the feedback capacity $C_{\mathrm{FB}}(\mathrm{SNR})$, and in Section 3.5.2 we derive a lower bound on the ca-

[3] Intuitively, with this choice of $\{\alpha_\ell\}$ the channel can be divided into two parallel channels, one connecting the inputs and outputs at even times, and the other connecting the inputs and outputs at odd times. As both channels have the coefficients $\tilde{\alpha}_0 = \tilde{\alpha}_1 = \ldots = 1$, it follows from Theorem 3.3 that the capacity of each parallel channel is bounded, so the capacity of the original channel is bounded as well.

pacity $C(\mathrm{SNR})$ in the absence of feedback. These bounds are used in Section 3.5.3 to derive an upper bound on $\dot{C}_{\mathrm{FB}}(0)$ and a lower bound on $\dot{C}(0)$, which are then both shown to be equal to $1/2\,(1+\alpha)$. Together with (3.22) this proves Theorem 3.2.

3.5.1 Converse

The upper bound on $C_{\mathrm{FB}}(\mathrm{SNR})$ is based on (3.19) and on an upper bound on $\frac{1}{n}I(M;Y_1^n)$, which for our channel can be expressed as

$$\begin{aligned}
&\frac{1}{n}I\big(M;Y_1^n\big)\\
&\quad=\frac{1}{n}\sum_{k=1}^{n}\Big(h\big(Y_k\mid Y_1^{k-1}\big)-h\big(Y_k\mid Y_1^{k-1},M\big)\Big)\\
&\quad=\frac{1}{n}\sum_{k=1}^{n}\Big(h\big(Y_k\mid Y_1^{k-1}\big)-h\big(Y_k\mid Y_1^{k-1},M,X_1^k\big)\Big)\\
&\quad=\frac{1}{n}\sum_{k=1}^{n}\left(h\big(Y_k\mid Y_1^{k-1}\big)-h(U_k)-\frac{1}{2}\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2\right)\right]\right),
\end{aligned} \tag{3.33}$$

where the first step follows from the chain rule for mutual information [5, Thm. 2.5.2]; the second step follows because X_1^k is a function of M and Y_1^{k-1}; and the last step follows from the behavior of differential entropy under translation and scaling [5, Thms. 9.6.3 & 9.6.4], and because U_k is independent of $\big(Y_1^{k-1},M,X_1^k\big)$.

Evaluating the differential entropy $h(U_k)$ of a Gaussian random variable, and using the trivial lower bound

$$\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2\right)\right]\geq\log\sigma^2,$$

we obtain the final upper bound

$$\frac{1}{n}I\big(M;Y_1^n\big)\leq\frac{1}{n}\sum_{k=1}^{n}\left(h\big(Y_k\mid Y_1^{k-1}\big)-\frac{1}{2}\log\big(2\pi e\sigma^2\big)\right)$$

$$
\begin{aligned}
&\leq \frac{1}{n}\sum_{k=1}^{n}\frac{1}{2}\log\left(1+\sum_{\ell=1}^{k}\alpha_{k-\ell}\mathsf{E}\left[X_\ell^2\right]/\sigma^2\right)\\
&\leq \frac{1}{2}\log\left(1+\frac{1}{n}\sum_{k=1}^{n}\sum_{\ell=1}^{k}\alpha_{k-\ell}\mathsf{E}\left[X_\ell^2\right]/\sigma^2\right)\\
&= \frac{1}{2}\log\left(1+\frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\left[X_k^2\right]/\sigma^2\sum_{\ell=0}^{n-k}\alpha_\ell\right)\\
&\leq \frac{1}{2}\log\left(1+(1+\alpha)\,\frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\left[X_k^2\right]/\sigma^2\right)\\
&\leq \frac{1}{2}\log\bigl(1+(1+\alpha)\,\mathrm{SNR}\bigr),
\end{aligned}
\tag{3.34}
$$

where we define $\alpha_0 \triangleq 1$. Here the second step follows because conditioning cannot increase entropy and from the entropy maximizing property of Gaussian random variables [5, Thm. 9.6.5]; the third step follows from Jensen's inequality; the fourth step by rewriting the double sum; the fifth step follows because the coefficients are nonnegative which implies that $\sum_{\ell=0}^{n-k}\alpha_\ell \leq \sum_{\ell=0}^{\infty}\alpha_\ell = 1+\alpha$; and the last step follows from the power constraint (3.12).

3.5.2 Direct Part

As aforementioned, the above channel (3.8) is not stationary, and it is therefore *prima facie* not clear whether $C_{\text{Info}}(\mathrm{SNR})$ is achievable. We shall sidestep this problem by studying the capacity of a different channel whose time-k channel output $\tilde{Y}_k \in \mathbb{R}$ is, conditional on the sequence $X_k = x_k$, $X_{k-1} = x_{k-1}, \ldots$, given by

$$
\tilde{Y}_k = x_k + \sqrt{\left(\sigma^2 + \sum_{\ell=-\infty}^{k-1}\alpha_{k-\ell}x_\ell^2\right)}\cdot U_k, \quad k\in\mathbb{N}, \tag{3.35}
$$

where $\{U_k,\, k\in\mathbb{Z}\}$ and $\{\alpha_\ell\}$ are defined in Section 3.2. This channel has the advantage that it is stationary and ergodic in the sense that when $\{X_k,\, k\in\mathbb{Z}\}$ is a stationary ergodic process, the pair $\{(X_k,\tilde{Y}_k),\, k\in\mathbb{Z}\}$ is jointly stationary ergodic. It follows that if the sequences $\{X_k\,,\, k=0,-1,\ldots\}$ and $\{X_k\,,\, k=1,2,\ldots\}$ are independent of each other, and if the random variables X_k, $k=0,-1,\ldots$ are bounded, then any rate that can be achieved over this new channel is also achievable over the original channel. Indeed, the original channel (3.8) can be converted into (3.35) by adding

$$S_k = \sqrt{\left(\sum_{\ell=-\infty}^{0} \alpha_{k-\ell} X_\ell^2\right)} \cdot V_k, \quad k \in \mathbb{N}$$

to the channel output Y_k (where $\{V_k,\, k \in \mathbb{Z}\}$ is a sequence of IID, zero-mean, unit-variance, Gaussian random variables drawn independently of $\{(U_k, X_k),\, k \in \mathbb{Z}\}$),[4] and since the independence of $\{X_k\,,\, k = 0, -1, \ldots\}$ and $\{X_k\,,\, k = 1, 2, \ldots\}$ ensures that the sequence $\{S_k\,,\, k \in \mathbb{N}\}$ is independent of the message M, it follows that any rate achievable over (3.35) can be achieved over (3.8) by using a receiver that generates $\{S_k,\, k \in \mathbb{N}\}$ and then guesses M based on $(Y_1 + S_1, \ldots, Y_n + S_n)$.[5]

We shall consider channel inputs $\{X_k,\, k \in \mathbb{Z}\}$ that are blockwise IID in blocks of L symbols (for some $L \in \mathbb{N}$). Thus, denoting $\mathbf{X}_b = (X_{bL+1}, \ldots, X_{(b+1)L})^{\mathsf{T}}$, we have that $\{\mathbf{X}_b,\, b \in \mathbb{Z}\}$ is a sequence of IID random length-L vectors with $\mathbf{X}_b$ taking on the value $(\xi, 0, \ldots, 0)^{\mathsf{T}}$ with probability δ and $(0, \ldots, 0)^{\mathsf{T}}$ with probability $1-\delta$, for some $\xi \in \mathbb{R}$. Note that to satisfy the average-power constraint (3.12) we shall choose ξ and δ so that

$$\frac{\xi^2}{\sigma^2}\delta = L\,\mathsf{SNR}. \tag{3.36}$$

Let $\tilde{\mathbf{Y}}_b = (\tilde{Y}_{bL+1}, \ldots, \tilde{Y}_{(b+1)L})^{\mathsf{T}}$. Noting that the pair $\{(\mathbf{X}_b, \tilde{\mathbf{Y}}_b),\, b \in \mathbb{Z}\}$ is jointly stationary ergodic, it follows from [49] that the rate

$$\lim_{n\to\infty} \frac{1}{n} I\left(\mathbf{X}_0^{\lfloor n/L \rfloor - 1}; \tilde{\mathbf{Y}}_0^{\lfloor n/L \rfloor - 1}\right)$$

[4]The boundedness of the random variables X_k, $k = 0, -1, \ldots$ guarantees that the quantity $\sum_{\ell=-\infty}^{0} \alpha_{k-\ell} x_\ell^2$ is finite for any realization of $\{X_k\,,\, k = 0, -1, \ldots\}$.

[5]This approach is specific to the case where $\{U_k,\, k \in \mathbb{Z}\}$ is a Gaussian process. Indeed, it relies heavily on the fact that, given $\ldots, X_{-1} = x_{-1}, X_0 = x_0, X_1 = x_1, \ldots$, the additive noise term on the right-hand side of (3.35) can be written as the sum of two independent random variables, of which one only depends on $\{X_k\,,\, k = 0, -1, \ldots\}$ and the other only on $\{X_k\,,\, k = 1, 2, \ldots\}$. This certainly holds for Gaussian random variables, but it does not necessarily hold for other distributions on $\{U_k,\, k \in \mathbb{Z}\}$.

is achievable over the new channel (3.35) and thus yields a lower bound on the capacity $C(\mathrm{SNR})$ of the original channel (3.8). We have

$$\begin{aligned}
&\frac{1}{n} I\Big(\mathbf{X}_0^{\lfloor n/L\rfloor-1}; \tilde{\mathbf{Y}}_0^{\lfloor n/L\rfloor-1}\Big) \\
&= \frac{1}{n} \sum_{b=0}^{\lfloor n/L\rfloor-1} I\big(\mathbf{X}_b; \tilde{\mathbf{Y}}_0^{\lfloor n/L\rfloor-1} \bigm| \mathbf{X}_0^{b-1}\big) \\
&\geq \frac{1}{n} \sum_{b=0}^{\lfloor n/L\rfloor-1} I\big(\mathbf{X}_b; \tilde{\mathbf{Y}}_b \bigm| \mathbf{X}_0^{b-1}\big) \\
&\geq \frac{1}{n} \sum_{b=0}^{\lfloor n/L\rfloor-1} \Big(I\big(\mathbf{X}_b; \tilde{\mathbf{Y}}_b \bigm| \mathbf{X}_{-\infty}^{b-1}\big) - I\big(\mathbf{X}_{-\infty}^{-1}; \tilde{\mathbf{Y}}_b \bigm| \mathbf{X}_0^{b}\big)\Big), \qquad (3.37)
\end{aligned}$$

where we use the chain rule and the nonnegativity of mutual information. It is shown in Appendix A.2 that

$$\lim_{b\to\infty} I\big(\mathbf{X}_{-\infty}^{-1}; \tilde{\mathbf{Y}}_b \bigm| \mathbf{X}_0^{b}\big) = 0. \qquad (3.38)$$

This together with a Cesáro-type theorem [5, Thm. 4.2.3] yields

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} I\Big(\mathbf{X}_0^{\lfloor n/L\rfloor-1}; \tilde{\mathbf{Y}}_0^{\lfloor n/L\rfloor-1}\Big) \\
&\geq \frac{1}{L} I\big(\mathbf{X}_0; \tilde{\mathbf{Y}}_0 \bigm| \mathbf{X}_{-\infty}^{-1}\big) \\
&\quad - \frac{1}{L} \lim_{n\to\infty} \frac{1}{\lfloor n/L\rfloor} \sum_{b=0}^{\lfloor n/L\rfloor-1} I\big(\mathbf{X}_{-\infty}^{-1}; \tilde{\mathbf{Y}}_b \bigm| \mathbf{X}_0^{b}\big) \\
&= \frac{1}{L} I\big(\mathbf{X}_0; \tilde{\mathbf{Y}}_0 \bigm| \mathbf{X}_{-\infty}^{-1}\big), \qquad (3.39)
\end{aligned}$$

where the first step follows by the stationarity of $\{(\mathbf{X}_b, \tilde{\mathbf{Y}}_b),\, b \in \mathbb{Z}\}$, which implies that the mutual information $I(\mathbf{X}_b; \tilde{\mathbf{Y}}_b|\mathbf{X}_{-\infty}^{b-1})$ does not depend on b, and by noting that

$$\lim_{n\to\infty} \frac{\lfloor n/L\rfloor}{n} = 1/L.$$

We proceed to analyze $I(\mathbf{X}_0; \tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1})$ for a given sequence $\mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}$. Making use of the canonical decomposition of mutual information (e.g., [47, Eq. (10)]), we have

$$\begin{aligned}
&I\big(\mathbf{X}_0; \tilde{\mathbf{Y}}_0 \mid \mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}\big) \\
&\quad = I\big(X_1; \tilde{\mathbf{Y}}_0 \mid \mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}\big) \\
&\quad = \int D\Big(P_{\tilde{\mathbf{Y}}_0|X_1=x,\mathbf{x}_{-\infty}^{-1}} \Big\| P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big)\,\mathrm{d}P_{X_1}(x) \\
&\qquad - D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \Big\| P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big) \\
&\quad = \delta D\Big(P_{\tilde{\mathbf{Y}}_0|X_1=\xi,\mathbf{x}_{-\infty}^{-1}} \Big\| P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big) \\
&\qquad - D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \Big\| P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big),
\end{aligned} \tag{3.40}$$

where the first step follows because, for our choice of input distribution, $X_2 = \ldots = X_L = 0$ and hence X_1 conveys as much information about $\tilde{\mathbf{Y}}_0$ as $\mathbf{X}_0$. Here $D(\cdot\|\cdot)$ denotes relative entropy (2.8), P_{X_1} denotes the distribution of X_1, and

$$P_{\tilde{\mathbf{Y}}_0|X_1=\xi,\mathbf{x}_{-\infty}^{-1}}, \quad P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}, \quad \text{and} \quad P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}}$$

denote the distributions of $\tilde{\mathbf{Y}}_0$ conditional on $(X_1 = \xi, \mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1})$, $(X_1 = 0, \mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1})$, and $\mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}$, respectively. Thus $P_{\tilde{\mathbf{Y}}_0|X_1=\xi,\mathbf{x}_{-\infty}^{-1}}$ is the law of an L-variate Gaussian random vector of mean $(\xi, 0, \ldots, 0)^{\mathsf{T}}$ and of diagonal covariance matrix $\mathsf{K}_{\mathbf{x}_{-\infty}^{-1}}^{(\xi)}$ with diagonal entries

$$\begin{aligned}
\mathsf{K}_{\mathbf{x}_{-\infty}^{-1}}^{(\xi)}(1,1) &= \sigma^2 + \sum_{\ell=-\infty}^{-1} \alpha_{-\ell L} x_{\ell L+1}^2 \\
\mathsf{K}_{\mathbf{x}_{-\infty}^{-1}}^{(\xi)}(i,i) &= \sigma^2 + \alpha_{i-1}\xi^2 + \sum_{\ell=-\infty}^{-1} \alpha_{-\ell L+i-1} x_{\ell L+1}^2, \qquad i = 2, \ldots, L;
\end{aligned}$$

$P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}$ is the law of an L-variate, zero-mean, Gaussian random vector of diagonal covariance matrix $\mathsf{K}_{\mathbf{x}_{-\infty}^{-1}}^{(0)}$ with diagonal entries

$$\mathsf{K}_{\mathbf{x}_{-\infty}^{-1}}^{(0)}(i,i) = \sigma^2 + \sum_{\ell=-\infty}^{-1} \alpha_{-\ell L+i-1} x_{\ell L+1}^2, \qquad i = 1, \ldots, L;$$

and $P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}}$ is given by

$$P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} = \delta P_{\tilde{\mathbf{Y}}_0|X_1=\xi,\mathbf{x}_{-\infty}^{-1}} + (1-\delta) P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}.$$

In order to evaluate the first term on the right-hand side (RHS) of (3.40) we note that the relative entropy of two real, L-variate, Gaussian random vectors of means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and of covariance matrices K_1 and K_2 is given by

$$\begin{aligned} &D\big(\mathcal{N}(\boldsymbol{\mu}_1,\mathsf{K}_1) \,\big\|\, \mathcal{N}(\boldsymbol{\mu}_2,\mathsf{K}_2)\big) \\ &\quad= \frac{1}{2}\log\det\mathsf{K}_2 - \frac{1}{2}\log\det\mathsf{K}_1 + \frac{1}{2}\operatorname{tr}\left(\mathsf{K}_1\mathsf{K}_2^{-1} - \mathsf{I}_L\right) \\ &\qquad + \frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^{\mathsf{T}}\mathsf{K}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2). \end{aligned} \tag{3.41}$$

The second term on the RHS of (3.40) is analyzed in the next subsection.

Let $\mathsf{E}\Big[D\big(P_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}} \,\|\, P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\big)\Big]$ denote the second term on the RHS of (3.40) averaged over $\mathbf{X}_{-\infty}^{-1}$, i.e.,

$$\begin{aligned} &\mathsf{E}\Big[D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}} \,\Big\|\, P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\Big)\Big] \\ &\qquad\qquad = \mathsf{E}_{\mathbf{X}_{-\infty}^{-1}}\Big[D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \,\Big\|\, P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big)\Big], \end{aligned}$$

where $\mathsf{E}_{X_{-\infty}^{-1}}$ denotes expectation with respect to $X_{-\infty}^{-1}$. Then, using (3.41) & (3.40) and taking expectations over $\mathbf{X}_{-\infty}^{-1}$, we obtain, again defining $\alpha_0 \triangleq 1$,

$$\begin{aligned} &\frac{1}{L} I\big(\mathbf{X}_0; \tilde{\mathbf{Y}}_0 \,\big|\, \mathbf{X}_{-\infty}^{-1}\big) \\ &\quad= \frac{\delta}{L}\frac{\xi^2}{\sigma^2}\frac{1}{2}\sum_{i=1}^{L}\mathsf{E}\left[\frac{\alpha_{i-1}}{1+\sum_{\ell=-\infty}^{-1}\alpha_{-\ell L+i-1}X_{\ell L+1}^2/\sigma^2}\right] \\ &\qquad - \frac{\delta}{L}\frac{1}{2}\sum_{i=2}^{L}\mathsf{E}\left[\log\left(1+\frac{\alpha_{i-1}\xi^2}{\sigma^2+\sum_{\ell=-\infty}^{-1}\alpha_{-\ell L+i-1}X_{\ell L+1}^2}\right)\right] \\ &\qquad - \frac{1}{L}\mathsf{E}\Big[D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}} \,\Big\|\, P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\Big)\Big] \end{aligned}$$

$$
\begin{aligned}
&\geq \frac{\delta}{L}\frac{\xi^2}{\sigma^2}\frac{1}{2}\sum_{i=1}^{L}\frac{\alpha_{i-1}}{1+\sum_{\ell=-\infty}^{-1}\alpha_{-\ell L+i-1}\mathsf{E}\left[X_{\ell L+1}^2\right]/\sigma^2}\\
&\quad -\frac{\delta}{L}\frac{1}{2}\sum_{i=2}^{L}\log\bigl(1+\alpha_{i-1}\xi^2/\sigma^2\bigr)\\
&\quad -\frac{1}{L}\mathsf{E}\left[D\Bigl(P_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}}\Bigm\| P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\Bigr)\right]\\
&\geq \frac{1}{2}\mathrm{SNR}\sum_{i=1}^{L}\frac{\alpha_{i-1}}{1+\alpha\,L\,\mathrm{SNR}}\\
&\quad -\frac{1}{2}\mathrm{SNR}\sum_{i=2}^{L}\frac{\log\bigl(1+\alpha_{i-1}\xi^2/\sigma^2\bigr)}{\xi^2/\sigma^2}\\
&\quad -\frac{1}{L}\mathsf{E}\left[D\Bigl(P_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}}\Bigm\| P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\Bigr)\right],
\end{aligned}
\tag{3.42}
$$

where the second step follows by applying Jensen's inequality to the convex function $f(x) = 1/(1+x)$, $x > 0$, and from the upper bound

$$
\mathsf{E}\left[\log\left(1+\frac{\alpha_{i-1}\xi^2}{\sigma^2+\sum_{\ell=-\infty}^{-1}\alpha_{-\ell L+i-1}X_{\ell L+1}^2}\right)\right] \leq \log\bigl(1+\alpha_{i-1}\xi^2/\sigma^2\bigr);
$$

and the third step follows from (3.36) and by upper bounding

$$
\sum_{\ell=-\infty}^{-1}\alpha_{-\ell L+i-1} \leq \sum_{\ell=1}^{\infty}\alpha_\ell = \alpha.
$$

The final lower bound follows now by (3.42) and (3.39)

$$
\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}I\Bigl(\mathbf{X}_0^{\lfloor n/L\rfloor-1};\tilde{\mathbf{Y}}_0^{\lfloor n/L\rfloor-1}\Bigr)\\
&\quad\geq \frac{1}{2}\mathrm{SNR}\sum_{i=1}^{L}\frac{\alpha_{i-1}}{1+\alpha\,L\,\mathrm{SNR}}\\
&\qquad -\frac{1}{2}\mathrm{SNR}\sum_{i=2}^{L}\frac{\log\bigl(1+\alpha_{i-1}\xi^2/\sigma^2\bigr)}{\xi^2/\sigma^2}\\
&\qquad -\frac{1}{L}\mathsf{E}\left[D\Bigl(P_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}}\Bigm\| P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\Bigr)\right]
\end{aligned}
\tag{3.43}
$$

and by recalling that

$$C(\text{SNR}) \geq \lim_{n\to\infty} \frac{1}{n} I\Big(\mathbf{X}_0^{\lfloor n/L\rfloor -1}; \tilde{\mathbf{Y}}_0^{\lfloor n/L\rfloor -1}\Big). \tag{3.44}$$

3.5.3 Asymptotic Analysis

We start with analyzing the upper bound (3.34). Using that

$$\log(1+x) \leq x, \quad x > -1$$

we have

$$\frac{C_{\text{FB}}(\text{SNR})}{\text{SNR}} \leq \frac{\frac{1}{2}\log(1+(1+\alpha)\,\text{SNR})}{\text{SNR}} \leq \frac{1}{2}(1+\alpha), \tag{3.45}$$

and we thus obtain

$$\dot{C}_{\text{FB}}(0) \triangleq \sup_{\text{SNR}>0} \frac{C_{\text{FB}}(\text{SNR})}{\text{SNR}} \leq \frac{1}{2}(1+\alpha). \tag{3.46}$$

In order to derive a lower bound on $\dot{C}(0)$ we first note that

$$\dot{C}(0) \triangleq \sup_{\text{SNR}>0} \frac{C(\text{SNR})}{\text{SNR}} \geq \lim_{\text{SNR}\downarrow 0} \frac{C(\text{SNR})}{\text{SNR}} \tag{3.47}$$

and proceed by analyzing the limiting ratio of the lower bound (3.43) to SNR as SNR tends to zero. To this end we first shall show that

$$\lim_{\text{SNR}\downarrow 0} \frac{\mathsf{E}\Big[D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}} \Big\| P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\Big)\Big]}{\text{SNR}} = 0. \tag{3.48}$$

We recall that for any pair of distributions P_0 and P_1 satisfying $P_1 \ll P_0$ [47, p. 1023]

$$\lim_{\beta\downarrow 0} \frac{D\left(\beta P_1 + (1-\beta)P_0 \| P_0\right)}{\beta} = 0. \tag{3.49}$$

Thus, for any $\mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}$, (3.49) together with $\delta = \text{SNR}\, L\, \sigma^2/\xi^2$ implies that

$$\lim_{\text{SNR}\downarrow 0} \frac{D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \Big\| P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big)}{\text{SNR}} = 0. \tag{3.50}$$

In order to show that this also holds when

$$D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \,\Big\|\, P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big)$$

is averaged over $\mathbf{X}_{-\infty}^{-1}$, we derive in the following the uniform upper bound

$$\sup_{\mathbf{x}_{-\infty}^{-1}} D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \,\Big\|\, P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big) = D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \,\Big\|\, P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big)\Big|_{\mathbf{x}_{-\infty}^{-1}=0}. \tag{3.51}$$

The claim (3.48) follows then by upper bounding

$$\mathsf{E}\Big[D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}} \,\Big\|\, P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\Big)\Big] \le D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \,\Big\|\, P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big)\Big|_{\mathbf{x}_{-\infty}^{-1}=0}$$

and by (3.50).

To prove (3.51) we use that every Gaussian random vector can be expressed as the sum of two independent Gaussian random vectors to write the channel output $\tilde{\mathbf{Y}}_0$ as

$$\tilde{\mathbf{Y}}_0 = \mathbf{X}_0 + \mathbf{V} + \mathbf{W}, \tag{3.52}$$

where, conditional on $\mathbf{X}_{-\infty}^{0} = \mathbf{x}_{-\infty}^{0}$, $\mathbf{V}$ and $\mathbf{W}$ are L-variate, zero-mean, Gaussian random vectors, drawn independently of each other and having the respective diagonal covariance matrices $\mathsf{K}_{\mathbf{V}|\mathbf{x}_0}$ and $\mathsf{K}_{\mathbf{W}|\mathbf{x}_{-\infty}^{-1}}$ whose diagonal entries are given by

$$\begin{aligned} \mathsf{K}_{\mathbf{V}|\mathbf{x}_0}(1,1) &= \sigma^2 \\ \mathsf{K}_{\mathbf{V}|\mathbf{x}_0}(i,i) &= \sigma^2 + \alpha_{i-1}x_1^2, \qquad i = 2,\dots,L, \end{aligned}$$

and

$$\mathsf{K}_{\mathbf{W}|\mathbf{x}_{-\infty}^{-1}}(i,i) = \sum_{\ell=-\infty}^{-1} \alpha_{-\ell L+i-1} x_{\ell L+1}^2, \qquad i = 1,\dots,L.$$

Thus $\mathbf{W}$ is the portion of the noise due to $\mathbf{X}_{-\infty}^{-1}$, and $\mathbf{V}$ is the portion of the noise that remains after subtracting $\mathbf{W}$. Note that $\mathbf{X}_0 + \mathbf{V}$ and $\mathbf{W}$ are independent of each other because $\mathbf{X}_0$ is, by construction, independent of $\mathbf{X}_{-\infty}^{-1}$. The upper bound (3.51) follows now by

$$\begin{aligned} & D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \,\Big\|\, P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big) \\ & \quad = D\Big(P_{\mathbf{X}_0+\mathbf{V}+\mathbf{W}|\mathbf{x}_{-\infty}^{-1}} \,\Big\|\, P_{\mathbf{X}_0+\mathbf{V}+\mathbf{W}|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big) \\ & \quad \leq D\big(P_{\mathbf{X}_0+\mathbf{V}} \,\big\|\, P_{\mathbf{X}_0+\mathbf{V}|X_1=0}\big) \\ & \quad = D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \,\Big\|\, P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big)\Big|_{\mathbf{x}_{-\infty}^{-1}=0}, \end{aligned} \tag{3.53}$$

where $P_{\mathbf{X}_0+\mathbf{V}+\mathbf{W}|\mathbf{x}_{-\infty}^{-1}}$ and $P_{\mathbf{X}_0+\mathbf{V}+\mathbf{W}|X_1=0,\mathbf{x}_{-\infty}^{-1}}$ denote the distributions of $\mathbf{X}_0 + \mathbf{V} + \mathbf{W}$ conditional on $\mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}$ and on $(X_1 = 0, \mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1})$; $P_{\mathbf{X}_0+\mathbf{V}}$ denotes the unconditional distribution of $\mathbf{X}_0 + \mathbf{V}$; and $P_{\mathbf{X}_0+\mathbf{V}|X_1=0}$ denotes the distribution of $\mathbf{X}_0 + \mathbf{V}$ conditional on $X_1 = 0$. Here the inequality follows from the data processing inequality [5, Sec. 2.9] and because $\mathbf{X}_0 + \mathbf{V}$ is independent of $\mathbf{X}_{-\infty}^{-1}$.

Returning to the analysis of (3.43), we obtain from (3.47) and (3.48)

$$\begin{aligned} \dot{C}(0) &\geq \lim_{\mathrm{SNR}\downarrow 0} \frac{C(\mathrm{SNR})}{\mathrm{SNR}} \\ &\geq \lim_{\mathrm{SNR}\downarrow 0} \frac{1}{2}\sum_{i=1}^{L} \frac{\alpha_{i-1}}{1+\alpha\, L\, \mathrm{SNR}} - \frac{1}{2}\sum_{i=2}^{L} \frac{\log\left(1+\alpha_{i-1}\xi^2/\sigma^2\right)}{\xi^2/\sigma^2} \\ &= \frac{1}{2}\sum_{i=1}^{L} \alpha_{i-1} - \frac{1}{2}\sum_{i=2}^{L} \frac{\log\left(1+\alpha_{i-1}\xi^2/\sigma^2\right)}{\xi^2/\sigma^2}. \end{aligned} \tag{3.54}$$

By letting first ξ^2 go to infinity while holding L fixed, and by letting then L go to infinity, we obtain the desired lower bound

$$\dot{C}(0) \geq \lim_{\mathrm{SNR}\downarrow 0} \frac{C(\mathrm{SNR})}{\mathrm{SNR}} \geq \frac{1}{2}(1+\alpha). \tag{3.55}$$

Thus (3.55), (3.22), and (3.46) yield

$$\frac{1}{2}(1+\alpha) \leq \lim_{\mathrm{SNR}\downarrow 0} \frac{C(\mathrm{SNR})}{\mathrm{SNR}} \leq \dot{C}(0) \leq \dot{C}_{\mathrm{FB}}(0) \leq \frac{1}{2}(1+\alpha), \tag{3.56}$$

which proves Theorem 3.2.

3.6 Proof of Theorem 3.3

3.6.1 Part (i)

In order to show that

$$\underline{\lim}_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} > 0 \tag{3.57}$$

implies that the feedback capacity $C_{\text{FB}}(\text{SNR})$ is bounded, we derive an upper bound on the capacity that is based on (3.19) and on an upper bound on $\frac{1}{n}I(M;Y_1^n)$. Again we define $\alpha_0 \triangleq 1$.

We first note that, according to (3.57), we can find an $\ell_0 \in \mathbb{N}$ and a $0 < \rho < 1$ such that

$$\alpha_{\ell_0} > 0 \qquad \text{and} \qquad \frac{\alpha_{\ell+1}}{\alpha_\ell} \geq \rho, \quad \ell \geq \ell_0. \tag{3.58}$$

We continue with the chain rule for mutual information

$$\begin{aligned}\frac{1}{n}I(M;Y_1^n) &= \frac{1}{n}\sum_{k=1}^{\ell_0} I\bigl(M;Y_k \bigm| Y_1^{k-1}\bigr)\\ &\quad + \frac{1}{n}\sum_{k=\ell_0+1}^{n} I\bigl(M;Y_k \bigm| Y_1^{k-1}\bigr).\end{aligned} \tag{3.59}$$

Each summand in the first sum on the RHS of (3.59) is upper bounded by

$$\begin{aligned}&I\bigl(M;Y_k \bigm| Y_1^{k-1}\bigr)\\ &\quad\leq h(Y_k) - h\bigl(Y_k \bigm| Y_1^{k-1}, M\bigr)\\ &\quad= h(Y_k) - \frac{1}{2}\mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2\right)\right] - h\bigl(U_k \bigm| U_1^{k-1}\bigr)\\ &\quad\leq \frac{1}{2}\log\left(2\pi e\left(1 + \sum_{\ell=1}^{k}\alpha_{k-\ell}\frac{\mathsf{E}\bigl[X_\ell^2\bigr]}{\sigma^2}\right)\right) - h\bigl(U_k \bigm| U_1^{k-1}\bigr)\\ &\quad\leq \frac{1}{2}\log\left(2\pi e\left(1 + \sup_{\ell'\in\mathbb{N}_0}\alpha_{\ell'}\sum_{\ell=1}^{k}\frac{\mathsf{E}\bigl[X_\ell^2\bigr]}{\sigma^2}\right)\right) - h\bigl(U_k \bigm| U_1^{k-1}\bigr)\\ &\quad\leq \frac{1}{2}\log\left(2\pi e\left(1 + \sup_{\ell'\in\mathbb{N}_0}\alpha_{\ell'}\, n\,\text{SNR}\right)\right) - h\bigl(U_k \bigm| U_1^{k-1}\bigr)\end{aligned}$$

$$\leq \frac{1}{2}\log\left(2\pi e\left(1+\sup_{\ell'\in\mathbb{N}_0}\alpha_{\ell'}\,n\,\mathrm{SNR}\right)\right)-h\big(\{U_k\}\big). \tag{3.60}$$

Recall that $\sup_{\ell'\in\mathbb{N}_0}\alpha_{\ell'}$ and $h(\{U_k\})$ are finite (3.10) & (3.9). Here the first step follows because conditioning cannot increase entropy; the second step follows because $\big(X_1^k, U_1^{k-1}\big)$ is a function of $\big(M, Y_1^{k-1}\big)$, from the behavior of entropy under translation and scaling [5, Thms. 9.6.3 & 9.6.4], and from the fact that, conditional on U_1^{k-1}, the random variable U_k is independent of $\big(X_1^k, M, Y_1^{k-1}\big)$; the third step follows from the entropy maximizing property of Gaussian random variables and by lower bounding

$$\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2\right)\right]\geq \log\sigma^2;$$

the fourth step by upper bounding each coefficient α_ℓ by the supremum of α_ℓ, $\ell\in\mathbb{N}_0$; the fifth step follows from the power constraint (3.12); and the last step follows because conditioning cannot increase entropy and because, by the stationarity of $\{U_k,\, k\in\mathbb{Z}\}$, we have $h(U_k|U_{-\infty}^{k-1}) = h(\{U_k\})$ [5, Thm. 4.2.1].

The summands in the second sum on the RHS of (3.59) are upper bounded using the general upper bound for mutual information (Theorem 2.1)

$$I(X;Y)\leq \int D\big(W(\cdot|x)\;\|\;R(\cdot)\big)\,\mathrm{d}Q(x). \tag{3.61}$$

For each $k=\ell_0+1,\ldots,n$, we upper bound $I\big(M;Y_k \bigm| Y_1^{k-1}=y_1^{k-1}\big)$ for a given $Y_1^{k-1}=y_1^{k-1}$ by choosing $R(\cdot)$ to be a Cauchy distribution whose density is given by

$$\frac{\sqrt{\beta}}{\pi}\frac{1}{1+\beta y_k^2},\quad y_k\in\mathbb{R}, \tag{3.62}$$

where we choose $\beta = 1/\big(\tilde{\beta}\,y_{k-\ell_0}^2\big)$ with

$$\tilde{\beta}=\min\left\{\rho^{\ell_0-1}\,\frac{\alpha_{\ell_0}}{\max_{0\leq\ell'<\ell_0}\alpha_{\ell'}},\rho^{\ell_0}\right\},$$

where $0 < \rho < 1$ and $\ell_0 \in \mathbb{N}$ are given by (3.58).[6] Note that (3.58) together with (3.10) implies that

$$0 < \tilde{\beta} < 1 \qquad \text{and} \qquad \tilde{\beta}\alpha_\ell \leq \alpha_{\ell+\ell_0}, \quad \ell \in \mathbb{N}_0. \tag{3.63}$$

Applying (3.62) to (3.61) yields

$$\begin{aligned} I\big(M; Y_k \bigm| Y_1^{k-1} = y_1^{k-1}\big) \leq \mathsf{E}\left[\log\left(1 + \frac{Y_k^2}{\tilde{\beta}\, Y_{k-\ell_0}^2}\right) \middle| \, Y_1^{k-1} = y_1^{k-1}\right] \\ + \frac{1}{2}\log\big(\tilde{\beta}\, y_{k-\ell_0}^2\big) + \log\pi \\ - h\big(Y_k \bigm| M, Y_1^{k-1} = y_1^{k-1}\big), \end{aligned} \tag{3.64}$$

and we thus obtain, averaging over Y_1^{k-1},

$$\begin{aligned} I\big(M; Y_k \bigm| Y_1^{k-1}\big) \leq \log\pi - h\big(Y_k \bigm| Y_1^{k-1}, M\big) \\ + \frac{1}{2}\mathsf{E}\Big[\log\big(\tilde{\beta}\, Y_{k-\ell_0}^2\big)\Big] \\ + \mathsf{E}\Big[\log\big(\tilde{\beta}\, Y_{k-\ell_0}^2 + Y_k^2\big)\Big] \\ - \mathsf{E}\big[\log\big(Y_{k-\ell_0}^2\big)\big] - \log\tilde{\beta}. \end{aligned} \tag{3.65}$$

We evaluate the terms on the RHS of (3.65) individually. We begin with

$$h\big(Y_k \big| Y_1^{k-1}, M\big) \geq \frac{1}{2}\mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2\right)\right] + h\big(\{U_k\}\big), \tag{3.66}$$

where we use the same arguments as in the second step in (3.60). The next term is upper bounded by

$$\begin{aligned} &\mathsf{E}\Big[\log\big(\tilde{\beta}Y_{k-\ell_0}^2\big)\Big] \\ &\quad = \mathsf{E}\Big[\mathsf{E}\Big[\log\Big(\tilde{\beta}\big(X_{k-\ell_0} + \theta\big(X_1^{k-\ell_0-1}\big)U_{k-\ell_0}\big)^2\Big) \Bigm| X_1^{k-\ell_0}\Big]\Big] \\ &\quad \leq \mathsf{E}\Big[\log\Big(\tilde{\beta}\,\mathsf{E}\Big[\big(X_{k-\ell_0} + \theta\big(X_1^{k-\ell_0-1}\big)U_{k-\ell_0}\big)^2 \Bigm| X_1^{k-\ell_0}\Big]\Big)\Big] \end{aligned}$$

[6] When $y_{k-\ell_0} = 0$ then with this choice of β the density of the Cauchy distribution (3.62) is undefined. However, this event is of zero probability and has therefore no impact on the mutual information $I\big(M; Y_k \bigm| Y_1^{k-1}\big)$.

$$
\begin{aligned}
&= \mathsf{E}\left[\log\left(\tilde{\beta}\, X_{k-\ell_0}^2 + \tilde{\beta}\,\sigma^2 + \tilde{\beta} \sum_{\ell=1}^{k-\ell_0-1} \alpha_{k-\ell_0-\ell}\, X_\ell^2\right)\right] \\
&\leq \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k-\ell_0} \alpha_{k-\ell}\, X_\ell^2\right)\right],
\end{aligned} \tag{3.67}
$$

where we define, for a given $X_1^{k-1} = x_1^{k-1}$,

$$
\theta\big(x_1^{k-1}\big) \triangleq \sqrt{\sigma^2 + \sum_{\ell=1}^{k-1} \alpha_{k-\ell}\, x_\ell^2}. \tag{3.68}
$$

Here the second step in (3.67) follows from Jensen's inequality, and the last step follows from (3.63). Similarly we use Jensen's inequality along with (3.63) to upper bound

$$
\begin{aligned}
&\mathsf{E}\left[\log\left(\tilde{\beta} Y_{k-\ell_0}^2 + Y_k^2\right)\right] \\
&\quad \leq \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k-\ell_0} \alpha_{k-\ell}\, X_\ell^2 + \sigma^2 + \sum_{\ell=1}^{k} \alpha_{k-\ell}\, X_\ell^2\right)\right] \\
&\quad \leq \log 2 + \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k} \alpha_{k-\ell}\, X_\ell^2\right)\right].
\end{aligned} \tag{3.69}
$$

In order to lower bound $\mathsf{E}\big[\log\big(Y_{k-\ell_0}^2\big)\big]$ we need the following lemma:

Lemma 3.4. *Let X be a random variable of density $f_X(x)$, $x \in \mathbb{R}$. Then for any $0 < \delta \leq 1$ and $0 < \eta < 1$*

$$
\sup_{c\in\mathbb{R}} \mathsf{E}\left[\log|X+c|^{-1}\, \mathrm{I}\{|X+c| \leq \delta\}\right] \leq \epsilon(\delta,\eta) + \frac{1}{\eta} h^-(X), \tag{3.70}
$$

where $h^-(X)$ is defined as

$$
h^-(X) \triangleq \int_{\{x\in\mathbb{R}: f_X(x) > 1\}} f_X(x) \log f_X(x)\, \mathrm{d}x; \tag{3.71}
$$

and where $\epsilon(\delta,\eta) > 0$ tends to zero as $\delta \downarrow 0$.

Proof. See [28, Lemma 6.7]. □

We write the expectation as

$$\mathsf{E}\left[\log\left(Y_{k-\ell_0}^2\right)\right] = \mathsf{E}\left[\mathsf{E}\left[\log\left(X_{k-\ell_0} + \theta\left(X_1^{k-\ell_0-1}\right)U_{k-\ell_0}\right)^2 \,\middle|\, X_1^{k-\ell_0}\right]\right]$$

and lower bound the conditional expectation for a given $X_1^{k-\ell_0} = x_1^{k-\ell_0}$

$$\begin{aligned}
&\mathsf{E}\left[\log\left(X_{k-\ell_0} + \theta\left(X_1^{k-\ell_0-1}\right)U_{k-\ell_0}\right)^2 \,\middle|\, X_1^{k-\ell_0} = x_1^{k-\ell_0}\right] \\
&\quad = \log\theta^2\left(x_1^{k-\ell_0-1}\right) \\
&\qquad - 2\,\mathsf{E}\left[\log\left|\frac{X_{k-\ell_0}}{\theta\left(X_1^{k-\ell_0-1}\right)} + U_{k-\ell_0}\right|^{-1} \,\middle|\, X_1^{k-\ell_0} = x_1^{k-\ell_0}\right] \\
&\quad \geq \log\theta^2\left(x_1^{k-\ell_0-1}\right) - 2\epsilon(\delta,\eta) - \frac{2}{\eta}h^-\left(U_{k-\ell_0}\right) + \log\delta^2
\end{aligned} \tag{3.72}$$

for some $0 < \delta \leq 1$ and $0 < \eta < 1$. Here the inequality follows by splitting the conditional expectation into the two expectations

$$\begin{aligned}
&\mathsf{E}\left[\log\left|\frac{x_{k-\ell_0}}{\theta\left(x_1^{k-\ell_0-1}\right)} + U_{k-\ell_0}\right|^{-1}\right] \\
&\quad = \mathsf{E}\left[\log\left|\frac{x_{k-\ell_0}}{\theta\left(x_1^{k-\ell_0-1}\right)} + U_{k-\ell_0}\right|^{-1} \mathrm{I}\left\{\left|\frac{x_{k-\ell_0}}{\theta\left(x_1^{k-\ell_0-1}\right)} + U_{k-\ell_0}\right| \leq \delta\right\}\right] \\
&\qquad + \mathsf{E}\left[\log\left|\frac{x_{k-\ell_0}}{\theta\left(x_1^{k-\ell_0-1}\right)} + U_{k-\ell_0}\right|^{-1} \mathrm{I}\left\{\left|\frac{x_{k-\ell_0}}{\theta\left(x_1^{k-\ell_0-1}\right)} + U_{k-\ell_0}\right| > \delta\right\}\right]
\end{aligned}$$

and by upper bounding then the first expectation using Lemma 3.4 and the second expectation by $-\log\delta$. Averaging (3.72) over $X_1^{k-\ell_0}$ yields

$$\begin{aligned}
\mathsf{E}\left[\log\left(Y_{k-\ell_0}^2\right)\right] \geq{}& \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k-\ell_0-1} \alpha_{k-\ell_0-\ell}\, X_\ell^2\right)\right] \\
&- 2\epsilon(\delta,\eta) - \frac{2}{\eta}h^-\left(U_{k-\ell_0}\right) + \log\delta^2.
\end{aligned} \tag{3.73}$$

Note that the fact that $U_{k-\ell_0}$ is of unit variance together with [28, Lemma 6.4] implies that $h^-(U_{k-\ell_0})$ is finite.

Turning back to the upper bound (3.65), we obtain from (3.66), (3.67), (3.69), and (3.73)

$$\begin{aligned} I\big(M;Y_k \mid Y_1^{k-1}\big) \leq \log\pi - \frac{1}{2}\mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2\right)\right] - h\big(\{U_k\}\big) \\ + \frac{1}{2}\mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k-\ell_0}\alpha_{k-\ell}\,X_\ell^2\right)\right] \\ + \log 2 + \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k}\alpha_{k-\ell}\,X_\ell^2\right)\right] \\ - \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k-\ell_0-1}\alpha_{k-\ell_0-\ell}\,X_\ell^2\right)\right] \\ + 2\,\epsilon(\delta,\eta) + \frac{2}{\eta}h^-(U_{k-\ell_0}) - \log\delta^2 - \log\tilde{\beta} \\ \leq \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k}\alpha_{k-\ell}\,X_\ell^2\right)\right] \\ - \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k-\ell_0-1}\alpha_{k-\ell_0-\ell}\,X_\ell^2\right)\right] + \mathsf{K}, \quad (3.74) \end{aligned}$$

where

$$\mathsf{K} \triangleq \log\frac{2\pi}{\tilde{\beta}\,\delta^2} - h\big(\{U_k\}\big) + \frac{2}{\eta}h^-(U_{k-\ell_0}) + 2\,\epsilon(\delta,\eta)$$

is a finite constant, and where the last step in (3.74) follows because we have with probability one

$$\sum_{\ell=1}^{k-\ell_0}\alpha_{k-\ell}\,X_\ell^2 \leq \sum_{\ell=1}^{k-1}\alpha_{k-\ell}\,X_\ell^2.$$

Note that K does not depend on k since the process $\{U_k,\, k\in\mathbb{Z}\}$ is stationary.

Turning back to the evaluation of the second sum on the RHS of (3.59), we use that, for any sequences $\{a_k\}$ and $\{b_k\}$,

$$\sum_{k=\ell_0+1}^{n}\big(a_k - b_k\big) = \sum_{k=n-2\ell_0+1}^{n}\big(a_k - b_{k-n+3\ell_0}\big)$$

$$+\sum_{k=\ell_0+1}^{n-2\ell_0}\big(a_k - b_{k+2\ell_0}\big). \tag{3.75}$$

Defining

$$a_k \triangleq \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k}\alpha_{k-\ell}\,X_\ell^2\right)\right] \tag{3.76}$$

and

$$b_k \triangleq \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k-\ell_0-1}\alpha_{k-\ell_0-\ell}\,X_\ell^2\right)\right] \tag{3.77}$$

we have for the first sum on the RHS of (3.75)

$$\begin{aligned}
&\sum_{k=n-2\ell_0+1}^{n}\big(a_k - b_{k-n+3\ell_0}\big)\\
&\quad= \sum_{k=n-2\ell_0+1}^{n}\mathsf{E}\left[\log\left(\frac{\sigma^2+\sum_{\ell=1}^{k}\alpha_{k-\ell}\,X_\ell^2}{\sigma^2+\sum_{\ell=1}^{k-n+2\ell_0-1}\alpha_{k-n+2\ell_0-\ell}\,X_\ell^2}\right)\right]\\
&\quad\le 2\,\ell_0\,\log\Big(1+\sup_{\ell\in\mathbb{N}_0}\alpha_\ell\,n\,\mathrm{SNR}\Big),
\end{aligned} \tag{3.78}$$

which follows by lower bounding the denominator by σ^2, and by using then Jensen's inequality together with the last two steps in (3.60). For the second sum on the RHS of (3.75) we have

$$\begin{aligned}
\sum_{k=\ell_0+1}^{n-2\ell_0}\big(a_k - b_{k+2\ell_0}\big) &= \sum_{k=\ell_0+1}^{n-2\ell_0}\mathsf{E}\left[\log\left(\frac{\sigma^2+\sum_{\ell=1}^{k}\alpha_{k-\ell}\,X_\ell^2}{\sigma^2+\sum_{\ell=1}^{k+\ell_0-1}\alpha_{k+\ell_0-\ell}\,X_\ell^2}\right)\right]\\
&\le \sum_{k=\ell_0+1}^{n-2\ell_0}\mathsf{E}\left[\log\left(\frac{\sigma^2+\sum_{\ell=1}^{k}\alpha_{k+\ell_0-\ell}\,X_\ell^2}{\sigma^2+\sum_{\ell=1}^{k+\ell_0-1}\alpha_{k+\ell_0-\ell}\,X_\ell^2}\right)\right]\\
&\qquad-(n-3\ell_0)\,\log\tilde{\beta}\\
&\le -(n-3\ell_0)\log\tilde{\beta},
\end{aligned} \tag{3.79}$$

where the second step follows by adding $\log\tilde{\beta}$ to the expectation and by upper bounding then $\tilde{\beta}\,\sigma^2 \le \sigma^2$ and $\tilde{\beta}\,\alpha_\ell \le \alpha_{\ell+\ell_0}$ (3.63); and the third step follows because we have with probability one

$$\sum_{\ell=1}^{k}\alpha_{k+\ell_0-\ell}\,X_\ell^2 \le \sum_{\ell=1}^{k+\ell_0-1}\alpha_{k+\ell_0-\ell}\,X_\ell^2.$$

We combine now (3.74), (3.75), (3.78), and (3.79) to upper bound

$$\frac{1}{n}\sum_{\ell=\ell_0+1}^{n} I\bigl(M;Y_k \bigm| Y_1^{k-1}\bigr) \leq \frac{n-\ell_0}{n}\mathsf{K} + \frac{2\ell_0}{n}\log\Bigl(1+\sup_{\ell\in\mathbb{N}_0}\alpha_\ell\, n\,\mathrm{SNR}\Bigr) - \frac{n-3\ell_0}{n}\log\tilde{\beta}, \tag{3.80}$$

which together with (3.59) and (3.60) yields

$$\frac{1}{n}I\bigl(M;Y_1^n\bigr) \leq \frac{n-\ell_0}{n}\mathsf{K} - \frac{n-3\ell_0}{n}\log\tilde{\beta} + \frac{\ell_0}{2n}\log(2\pi e) + \frac{\ell_0}{n}\frac{5}{2}\log\Bigl(1+\sup_{\ell\in\mathbb{N}_0}\alpha_\ell\, n\,\mathrm{SNR}\Bigr) - \frac{\ell_0}{n}h\bigl(\{U_k\}\bigr). \tag{3.81}$$

This converges to $\mathsf{K}-\log\tilde{\beta} < \infty$ as we let n tend to infinity, thus proving that $\underline{\lim}_{\ell\to\infty}\alpha_{\ell+1}/\alpha_\ell > 0$ implies that the capacity $C_{\mathrm{FB}}(\mathrm{SNR})$ is bounded in the SNR.

3.6.2 Part (ii)

We show that

$$\lim_{\ell\to\infty}\frac{1}{\ell}\log\frac{1}{\alpha_\ell} = \infty \tag{3.82}$$

implies that the capacity $C(\mathrm{SNR})$ is unbounded in the SNR. Part (ii) of Theorem 3.3 follows then by noting that

$$\overline{\lim_{\ell\to\infty}}\frac{\alpha_{\ell+1}}{\alpha_\ell} = 0 \quad\Longrightarrow\quad \lim_{\ell\to\infty}\frac{1}{\ell}\log\frac{1}{\alpha_\ell} = \infty.$$

We prove the claim by proposing a coding scheme that achieves an unbounded rate. We first note that (3.82) implies that for any $0<\varrho<1$ we can find an $\ell_0\in\mathbb{N}$ such that

$$\alpha_\ell < \varrho^\ell, \quad \ell \geq \ell_0. \tag{3.83}$$

If there exists an $\ell_0\in\mathbb{N}$ such that $\alpha_\ell = 0$, $\ell\geq\ell_0$, then we can achieve the (unbounded) rate

$$R = \frac{1}{2L}\log(1+L\,\mathrm{SNR}), \quad L\geq\ell_0$$

by a coding scheme where $\{X_{kL+1},\, k \in \mathbb{N}_0\}$ is a sequence of IID, zero-mean, Gaussian random variables of variance $L\mathsf{P}$, and where the other inputs are deterministically zero. Indeed, by waiting L time-steps, the chip's temperature cools down to the ambient one, so the noise variance is independent of the previous channel inputs and we can achieve—after appropriate normalization—the capacity of the additive white Gaussian noise (AWGN) channel [23].

For the more general case (3.83) we propose the following encoding and decoding scheme. Let $x_1^n(m)$, $m \in \mathcal{M}$ denote the codeword sent out by the transmitter that corresponds to the message $M = m$. We choose some $L \geq \ell_0$ and generate the components $x_{kL+1}(m)$, $m \in \mathcal{M}$, $k = 0, \ldots, \lfloor n/L \rfloor - 1$ independently of each other according to a zero-mean Gaussian law of variance P. The other components are set to zero. (It follows from the weak law of large numbers that $\frac{1}{n}\sum_{k=1}^{n} x_k^2(m)$ converges to P/L in probability as n tends to infinity. This guarantees that the probability that a codeword does not satisfy the per-message power constraint (3.14)—and hence also the average-power constraint (3.12)—vanishes as n tends to infinity.)

The receiver uses a *nearest neighbor decoder* in order to guess M based on the received sequence of channel outputs y_1^n. Thus it computes $\|\mathbf{y} - \mathbf{x}(m')\|^2$ for each $m' \in \mathcal{M}$ and decides on the message that satisfies

$$\hat{M} = \arg\min_{m' \in \mathcal{M}} \|\mathbf{y} - \mathbf{x}(m')\|^2, \tag{3.84}$$

where ties are resolved with a fair coin flip. Here

$$\begin{aligned}
\mathbf{y} \quad &\triangleq \left(y_1, y_{L+1}, \ldots, y_{(\lfloor n/L \rfloor - 1)L+1}\right)^{\mathsf{T}} \\
\mathbf{x}(m') &\triangleq \left(x_1(m'), x_{L+1}(m'), \ldots, x_{(\lfloor n/L \rfloor - 1)L+1}(m')\right)^{\mathsf{T}}.
\end{aligned}$$

We are interested in the average probability of error $\Pr\big(\hat{M} \neq M\big)$, averaged over all codewords in the codebook, and averaged over all codebooks. By the symmetry of the codebook construction, the probability of error corresponding to the m-th message $\Pr\big(\hat{M} \neq M \bigm| M = m\big)$ does not depend on m, and we thus conclude that

$$\Pr\big(\hat{M} \neq M\big) = \Pr\big(\hat{M} \neq M \bigm| M = 1\big).$$

We further note that

$$\Pr\bigl(\hat{M} \neq M \bigm| M = 1\bigr) \leq \Pr\left(\bigcup_{m'=2}^{|\mathcal{M}|} \|\mathbf{Y} - \mathbf{X}(m')\|^2 \leq \|\mathbf{Z}\|^2 \,\middle|\, M = 1\right), \tag{3.85}$$

where

$$\mathbf{Z} = \begin{pmatrix} \theta\bigl(X_1(1)\bigr)U_1 \\ \theta\bigl(X_1^L(1)\bigr)U_{L+1} \\ \vdots \\ \theta\Bigl(X_1^{(\lfloor n/L\rfloor-1)L}(1)\Bigr)U_{(\lfloor n/L\rfloor-1)L+1} \end{pmatrix},$$

which is, conditional on $M = 1$, equal to $\mathbf{Y} - \mathbf{X}(1)$. In order to analyze (3.85) we need the following lemma.

Lemma 3.5. *Consider the channel described in Section 3.2, and assume that* $\{\alpha_\ell\}$ *satisfies* (3.82). *Further assume that* $\{X_{kL+1}\,,\, k \in \mathbb{N}_0\}$ *is a sequence of IID, zero-mean, Gaussian random variables of variance* P, *and that* $X_k = 0$ *for* $k \mod L \neq 1$. *(Here* $k \mod L$ *denotes the remainder upon dividing* k *by* L*). Let the set* $\mathcal{D}_\epsilon$ *be defined as*

$$\mathcal{D}_\epsilon \triangleq \Biggl\{(\mathbf{y},\mathbf{z}) \in \mathbb{R}^{\lfloor n/L\rfloor} \times \mathbb{R}^{\lfloor n/L\rfloor} : \left|\frac{1}{\lfloor n/L\rfloor}\|\mathbf{y}\|^2 - \bigl(\sigma^2 + \mathsf{P} + \alpha^{(L)}\,\mathsf{P}\bigr)\right| \leq \epsilon, \left|\frac{1}{\lfloor n/L\rfloor}\|\mathbf{z}\|^2 - \bigl(\sigma^2 + \alpha^{(L)}\,\mathsf{P}\bigr)\right| \leq \epsilon \Biggr\}, \tag{3.86}$$

where we define

$$\alpha^{(L)} \triangleq \sum_{\ell=1}^{\infty} \alpha_{\ell L}. \tag{3.87}$$

Then

$$\lim_{n\to\infty} \Pr\bigl((\mathbf{Y},\mathbf{Z}) \in \mathcal{D}_\epsilon\bigr) = 1 \tag{3.88}$$

for any $\epsilon > 0$.

Proof. See Appendix A.3. □

In order to upper bound the RHS of (3.85) we proceed along the lines of [23, 30]. Using that, by the symmetry of the codebook construction, the law of $(\mathbf{Y}, \mathbf{Z})$ does not depend on m, and using that the codewords are independent so, conditional on $M = 1$, the distribution of $\big(\mathbf{X}(2), \ldots, \mathbf{X}(|\mathcal{M}|)\big)$ does not depend on $(\mathbf{y}, \mathbf{z})$, we obtain

$$\Pr\left(\bigcup_{m'=2}^{|\mathcal{M}|} \|\mathbf{Y} - \mathbf{X}(m')\|^2 \le \|\mathbf{Z}\|^2 \,\middle|\, M = 1\right) \le \Pr\big((\mathbf{Y}, \mathbf{Z}) \notin \mathcal{D}_\epsilon\big) + \int_{\mathcal{D}_\epsilon} \Pr\left(\bigcup_{m'=2}^{|\mathcal{M}|} \|\mathbf{y} - \mathbf{X}(m')\|^2 \le \|\mathbf{z}\|^2\right) \mathrm{d}P(\mathbf{y}, \mathbf{z}), \quad (3.89)$$

where $P(\mathbf{y}, \mathbf{z})$ denotes the distribution of $(\mathbf{Y}, \mathbf{Z})$. It follows from Lemma 3.5 that the first term on the RHS of (3.89) vanishes as n tends to infinity. To evaluate the second term on the RHS of (3.89), we note that by the union of events bound

$$\Pr\left(\bigcup_{m'=2}^{|\mathcal{M}|} \|\mathbf{y} - \mathbf{X}(m')\|^2 \le \|\mathbf{z}\|^2\right) \le \sum_{m'=2}^{|\mathcal{M}|} \Pr\big(\|\mathbf{y} - \mathbf{X}(m')\|^2 \le \|\mathbf{z}\|^2\big). \quad (3.90)$$

By upper bounding

$$\|\mathbf{z}\|^2 \le \lfloor n/L \rfloor \big(\sigma^2 + \alpha^{(L)}\,\mathsf{P} + \epsilon\big), \quad (\mathbf{y}, \mathbf{z}) \in \mathcal{D}_\epsilon,$$

by lower bounding

$$\|\mathbf{y}\|^2 \ge \lfloor n/L \rfloor \big(\sigma^2 + \mathsf{P} + \alpha^{(L)}\,\mathsf{P} - \epsilon\big), \quad (\mathbf{y}, \mathbf{z}) \in \mathcal{D}_\epsilon,$$

and by applying Chernoff's bound [14, Sec. 5.4], we obtain for each $m' = 2, \ldots, |\mathcal{M}|$ and for any $s < 0$

$$\Pr\big(\|\mathbf{y} - \mathbf{X}(m')\|^2 \le \|\mathbf{z}\|^2\big) \le \exp\Big(-s\lfloor n/L \rfloor \big(\sigma^2 + \alpha^{(L)}\,\mathsf{P} + \epsilon\big)\Big)$$

$$\times \exp\left(s \frac{\lfloor n/L \rfloor \big(\sigma^2 + \mathsf{P} + \alpha^{(L)}\,\mathsf{P} - \epsilon\big)}{1 - 2s\mathsf{P}} \right)$$
$$\times \exp\Big(-\frac{1}{2} \lfloor n/L \rfloor \log(1 - 2\,s\,\mathsf{P}) \Big), \quad (\mathbf{y}, \mathbf{z}) \in \mathcal{D}_\epsilon. \qquad (3.91)$$

Applying (3.90) and (3.91) to (3.89), it follows that

$$\Pr\left(\bigcup_{m'=2}^{|\mathcal{M}|} \|\mathbf{Y} - \mathbf{X}(m')\|^2 \le \|\mathbf{Z}\|^2 \,\middle|\, M = 1 \right)$$

tends to zero as n tends to infinity if for some $s < 0$ the rate R satisfies

$$R < \frac{s}{L}\big(\sigma^2 + \alpha^{(L)}\,\mathsf{P} + \epsilon\big) + \frac{1}{2L}\log(1 - 2s\mathsf{P})$$
$$- \frac{s}{L}\frac{\sigma^2 + \mathsf{P} + \alpha^{(L)}\,\mathsf{P} - \epsilon}{1 - 2s\mathsf{P}}. \qquad (3.92)$$

Hence, by choosing $s = -\frac{1}{2}\frac{1}{1+\alpha^{(L)}\,\mathsf{P}}$, it follows that any rate below

$$-\frac{1}{2L}\frac{\sigma^2 + \alpha^{(L)}\,\mathsf{P} + \epsilon}{1 + \alpha^{(L)}\,\mathsf{P}} + \frac{1}{2L}\log\left(1 + \frac{\mathsf{P}}{1 + \alpha^{(L)}\,\mathsf{P}}\right)$$
$$+ \frac{1}{2L}\frac{\sigma^2 + \mathsf{P} + \alpha^{(L)}\,\mathsf{P} - \epsilon}{1 + \alpha^{(L)}\,\mathsf{P}}\frac{1}{1 + \mathsf{P}/\big(1 + \alpha^{(L)}\,\mathsf{P}\big)}$$

is achievable. As P tends to infinity this converges to

$$\frac{1}{2L}\log\left(1 + \frac{1}{\alpha^{(L)}}\right) > \frac{1}{2L}\log\frac{1}{\alpha^{(L)}}. \qquad (3.93)$$

It remains to show that given (3.83) we can make $-\frac{1}{2\,L}\log\alpha^{(L)}$ arbitrarily large. Indeed, (3.83) implies that

$$\alpha^{(L)} = \sum_{\ell=1}^{\infty}\alpha_{\ell L} < \sum_{\ell=1}^{\infty}\varrho^{\ell L} = \frac{\varrho^L}{1 - \varrho^L},$$

and the RHS of (3.93) can therefore be further lower bounded by

$$\frac{1}{2L}\log\big(1 - \varrho^L\big) + \frac{1}{2}\log\frac{1}{\varrho}.$$

Letting L tend to infinity yields then that we can achieve any rate below

$$\frac{1}{2}\log\frac{1}{\varrho}.$$

Since this can be made arbitrarily large by choosing ϱ sufficiently small, we conclude that $\lim_{\ell\to\infty}\frac{1}{\ell}\log\frac{1}{\alpha_\ell}=\infty$ implies that the capacity is unbounded.

3.7 Conclusion

We studied a model for on-chip communication with nonideal heat sinks. To account for the heating-up effect we proposed a channel model where the variance of the additive noise depends on a weighted sum of the past channel input powers. The weights are related to the efficiency of the heat sink.

To study the capacity of this channel at low SNR, we computed the capacity per unit cost. We showed that, irrespective of the distribution on the (IID) noise sequence $\{U_k,\, k\in\mathbb{Z}\}$, the heating-up effect is unharmful in the sense that the capacity per unit cost cannot be smaller than the capacity per unit cost of the channel with an ideal sink (i.e., for $\alpha=0$). We further showed that if the noise $\{U_k,\, k\in\mathbb{Z}\}$ is IID Gaussian, then the heating-up effect is even beneficial in the sense that the capacity per unit cost is *larger* than the capacity per unit cost of the channel with an ideal heat sink. This suggests that at low SNR no heat sinks should be used. (Of course, there may be other reasons to use heat sinks.)

Studying capacity at high SNR, we derived a sufficient condition and a necessary condition for the capacity to be bounded in the SNR. We showed that when $\{\alpha_\ell\}$ decays not faster than geometrically, then the capacity is bounded in the SNR. On the other hand, if $\{\alpha_\ell\}$ decays faster than geometrically, then the capacity is unbounded in the SNR. This result demonstrates the importance of an efficient heat sink at high SNR.

Chapter 4

Gaussian Flat-Fading Channels

4.1 Introduction

The subject of this chapter is the capacity of multiple-input multiple-output (MIMO), discrete-time, stationary and ergodic, Gaussian flat-fading channels with memory. We study a noncoherent channel model where transmitter and receiver are not aware of the realization of the fading but only of its statistics. We focus on the capacity of this channel at high signal-to-noise ratio (SNR).

The high-SNR asymptotic behavior of this channel's capacity depends highly on the ability of predicting the present fading from its past. We shall refer to the case where the present fading cannot be predicted perfectly from its past as *regular fading* and to the case where the present fading can be predicted perfectly from its past as *non-regular fading*. Regular fading channels were studied by Lapidoth and Moser [28]. They showed that in this case the capacity increases double-logarithmically in the SNR, and they defined the *fading number* as the second-order term in the high-SNR asymptotic expansion of capacity in order to characterize the capacity at high SNR more accurately. Furthermore, they computed the fading number for various scenarios: their results include the fading numbers of single-input single-output (SISO), single-input multiple-output (SIMO), and multiple-input single-output (MISO) channels when the fading is memoryless, and the fading number of SISO fading channels with memory. The fading number of SIMO fading channels with memory was later derived in [29].

Nonregular fading channels were studied by Lapidoth in the single-antenna case. He showed that, at high SNR, capacity can increase in various ways with the SNR. He further derived an expression for the capacity pre-log, i.e., the limiting ratio of capacity to the logarithm of the SNR as the SNR tends to infinity, when capacity grows logarithmically with the SNR.

To the best of our knowledge, there exist only bounds on the fading number and on the capacity pre-log of MIMO fading channels with memory. To better understand the high-SNR capacity of such channels, we present firm (nonasymptotic) upper bounds on channel capacity and study their high-SNR asymptotic behavior to obtain bounds on the fading number (for regular fading) and on the capacity pre-log (for nonregular fading). We shall see that the notions of the fading number and the number of degrees of freedom are closely tied.

Another commonly used model for flat-fading channels is the block-constant fading model [34], see for example [34, 52] for results on the capacity of this channel at high SNR. Note however that the block-constant fading model is, in contrast to our channel model, not a stationary channel model. Thus the block-constant fading model and the stationary fading model may give rise completely different capacity results.

Some of the results presented in this chapter were obtained in [22]. For the sake of completeness, we repeat the proofs of these results in Appendix B.

4.2 Channel Model and Channel Capacity

We begin with a description of the channel model and with some definitions. We envision a channel with n_T transmit antennas and n_R receive antennas. Its time-k ($k \in \mathbb{Z}$) output $\mathbf{Y}_k \in \mathbb{C}^{n_\mathrm{R}}$ corresponding to the time-k channel input $\mathbf{x}_k \in \mathbb{C}^{n_\mathrm{T}}$ is an n_R-dimensional complex random vector that is given by

$$\mathbf{Y}_k = \mathbb{H}_k \mathbf{x}_k + \mathbf{Z}_k, \quad k \in \mathbb{Z}, \tag{4.1}$$

where the random $n_\mathrm{R} \times n_\mathrm{T}$ complex matrix $\mathbb{H}_k \in \mathbb{C}^{n_\mathrm{R} \times n_\mathrm{T}}$ denotes the time-k fading matrix and the random vector $\mathbf{Z}_k \in \mathbb{C}^{n_\mathrm{R}}$ denotes the additive noise. We assume throughout that the vectors $\{\mathbf{Z}_k,\, k \in \mathbb{Z}\}$ are independent and identically distributed (IID) according to a circularly-symmetric, complex multi-variate Gaussian law of zero mean and of covariance matrix $\sigma^2 \, \mathsf{I}_{n_\mathrm{R}}$ with $\sigma > 0$.

The matrix valued fading process $\{\mathbb{H}_k,\ k \in \mathbb{Z}\}$ is assumed to be stationary ergodic Gaussian and to satisfy the finite expected squared Frobenius norm condition

$$\mathsf{E}\left[\|\mathbb{H}_k\|_\mathrm{F}^2\right] < \infty. \tag{4.2}$$

The fading process $\{\mathbb{H}_k,\ k \in \mathbb{Z}\}$ and the additive noise process $\{\mathbf{Z}_k,\ k \in \mathbb{Z}\}$ are independent of each other and of a joint law that does not depend on the channel inputs $\{\mathbf{x}_k\}$.

The above conditions will be assumed throughout. Some theorems will require additional assumptions. These are defined next. We shall say that the fading process $\{\mathbb{H}_k,\ k \in \mathbb{Z}\}$ is *regular* if it has a finite differential entropy rate, i.e., if

$$h(\{\mathbb{H}_k\}) \triangleq \lim_{n\to\infty} \frac{1}{n} h(\mathbb{H}_1, \ldots, \mathbb{H}_n) > -\infty. \tag{4.3}$$

Otherwise, we shall say that it is *nonregular*. We say that the *fading is spatially independent* if the $n_\mathrm{R}\, n_\mathrm{T}$ processes $\{H_k(r,t),\ k \in \mathbb{Z}\}$ are independent. We say that the *fading is spatially IID* if they are additionally of the same law.

The capacity of this channel under an average-power constraint on the inputs is given by (see, e.g., [20, Thm. 2] or [39, Sec. II])

$$C_\mathrm{Avg}(\mathrm{SNR}) = \lim_{n\to\infty} \frac{1}{n} \sup I\big(\mathbf{X}_1^n; \mathbf{Y}_1^n\big), \tag{4.4}$$

where the maximization is over all joint distributions on $\mathbf{X}_1, \ldots, \mathbf{X}_n$ that satisfy

$$\frac{1}{n} \sum_{k=1}^{n} \mathsf{E}\left[\|\mathbf{X}_k\|^2\right] \le \mathsf{P}. \tag{4.5}$$

Here the subscript "Avg" indicates that an average-power constraint is imposed. For the peak-power A constrained capacity $C_\mathrm{PP}(\mathrm{SNR})$ the maximization is over all joint distributions under which with probability one

$$\|\mathbf{X}_k\|^2 \le \mathsf{A}^2, \quad k = 1, \ldots, n, \tag{4.6}$$

where the subscript "PP" indicates that a peak-power constraint is imposed. We shall follow this convention throughout this chapter: we add a subscript "Avg" to indicate that an average-power constraint is imposed, and we add a subscript "PP" to indicate that a peak-power constraint is imposed. We omit the subscript if the distinction is immaterial. Note that

$$C_{\mathrm{PP}}(\mathrm{SNR}) \leq C_{\mathrm{Avg}}(\mathrm{SNR}) \tag{4.7}$$

since any distribution on $\mathbf{X}_1, \ldots, \mathbf{X}_n$ satisfying the peak-power constraint satisfies also the average-power constraint. The signal-to-noise ratio (SNR) is defined depending on whether an average- or a peak-power constraint is imposed:

$$\mathrm{SNR} \triangleq \frac{\mathsf{P}}{\sigma^2}, \quad \text{for an average-power constraint} \tag{4.8}$$

and

$$\mathrm{SNR} \triangleq \frac{\mathsf{A}^2}{\sigma^2}, \quad \text{for a peak-power constraint.} \tag{4.9}$$

For *regular* fading capacity grows double-logarithmically with the SNR [28, Thm. 4.2], i.e.,

$$\varlimsup_{\mathrm{SNR}\to\infty} \bigl\{C(\mathrm{SNR}) - \log\log \mathrm{SNR}\bigr\} < \infty. \tag{4.10}$$

In this case we define the *fading number* as

$$\chi(\{\mathbb{H}_k\}) \triangleq \varlimsup_{\mathrm{SNR}\to\infty} \{C(\mathrm{SNR}) - \log\log \mathrm{SNR}\}. \tag{4.11}$$

Note that, by (4.7), $\chi_{\mathrm{PP}}(\{\mathbb{H}_k\}) \leq \chi_{\mathrm{Avg}}(\{\mathbb{H}_k\})$.

The exact calculation of the fading number for general (regular) fading channels with memory is a difficult task. An exact expression for the fading number for the SISO case ($n_{\mathrm{R}} = n_{\mathrm{T}} = 1$) is given in [28]

$$\chi(\{H_k\}) = \log \pi + \mathsf{E}\bigl[\log |H_1|^2\bigr] - h(\{H_k\}). \tag{4.12}$$

The SIMO case ($n_{\mathrm{T}} = 1$) was recently solved in [29]. Here we shall present results for the MISO case ($n_{\mathrm{R}} = 1$) when the fading is spatially independent. Specifically, Corollary 4.10 treats the case where the fading is spatially independent Gaussian with a *zero mean-vector*, and Corollary 4.11 treats the case where $\{\mathbf{H}_k - \mathsf{E}[\mathbf{H}_k]\,,\, k \in \mathbb{Z}\}$ is *spatially IID* Gaussian. For MIMO channels we shall present lower bounds (see Proposition 4.5 and Theorem 4.6) and upper bounds on the fading number (see Corollaries 4.7 and 4.8).

For *nonregular fading*, capacity can grow with the SNR in various ways [25, 26]. When it grows logarithmically in the SNR, the *pre-log* under a peak-power constraint is defined as [26]

$$\Pi_{\text{PP}} = \varlimsup_{\text{SNR}\to\infty} \frac{C_{\text{PP}}(\text{SNR})}{\log \text{SNR}} \tag{4.13}$$

with an analogous definition for the pre-log Π_{Avg} under an average-power constraint.

The pre-log for general MIMO fading channels is unknown. It was computed under a peak-power constraint for the SISO Gaussian case in [25, 26] where it was shown that Π_{PP} is given by the Lebesgue measure of the set of harmonics in the interval $[-1/2, 1/2]$ where the derivative of the spectral distribution function is zero:

$$\Pi_{\text{PP}} = \mu\left(\{\lambda : F'(\lambda) = 0\}\right), \tag{4.14}$$

where μ denotes the Lebesgue measure on the interval $[-1/2, 1/2]$ and $F'(\cdot)$ is the derivative of the spectral distribution function characterizing the memory of the fading process, see Section 4.3.

Here we shall present an upper bound on Π_{PP} for MIMO Gaussian fading in Theorem 4.13 and Corollary 4.14 and an exact expression for Π_{PP} for spatially independent MISO Gaussian fading in Corollary 4.15.

The rest of this chapter is organized as follows. Section 4.3 discusses the linear prediction problem. Section 4.4 presents firm upper and lower bounds on the capacity. Section 4.5 addresses the fading number: new lower bounds, new upper bounds, and the MISO cases where these bounds yield the exact fading number are presented. Section 4.6 deals with nonregular fading and the capacity pre-log: it includes upper bounds on the pre-log for nonregular MIMO fading as well as the expression for the pre-log of spatially independent MISO fading channels. Section 4.7 addresses the capacity pre-loglog, which is defined as the limiting ratio of the capacity to $\log\log \text{SNR}$ as SNR tends to infinity. Section 4.8 provides the main proofs of this chapter. Section 4.9 specializes our results to slowly-varying Gauss-Markov MIMO fading channels. And Section 4.10 concludes with a discussion of the relationship between the fading number and degrees of freedom.

4.3 Linear Prediction

In this section we recall some facts about the prediction problem for Gaussian processes. We shall focus on the vector-valued case, i.e., we consider the prediction of the η-variate, zero-mean, stationary, circularly-symmetric, complex Gaussian process $\{\mathbf{A}_k,\ k \in \mathbb{Z}\}$ of finite variance. The matrix-valued case can be reduced to the vector-valued case by stacking the column vectors in one big vector.

The (matrix-valued) autocovariance function of $\{\mathbf{A}_k,\ k \in \mathbb{Z}\}$ can be described by the matrix-valued spectral distribution function $\mathsf{F}(\cdot)$. Thus for each $\lambda \in [-1/2, 1/2]$ the matrix $\mathsf{F}(\lambda)$ is nonnegative definite and Hermitian (i.e., $\mathsf{F}(\lambda)^\dagger = \mathsf{F}(\lambda)$), and the matrix-valued function $\lambda \mapsto \mathsf{F}(\lambda)$ is monotonically nondecreasing on $[-1/2, 1/2]$ and satisfies [50, Sec. 7]

$$\mathsf{E}\left[\mathbf{A}_{k+m}\mathbf{A}_k^\dagger\right] = \int_{-1/2}^{1/2} e^{\mathrm{i}2\pi m\lambda}\, \mathrm{d}\mathsf{F}(\lambda), \quad \bigl(k \in \mathbb{Z},\ m \in \mathbb{Z}\bigr). \tag{4.15}$$

(A matrix-valued function $\mathsf{F}(\cdot)$ is said to be monotonically nondecreasing if for $\lambda \le \lambda'$ the difference $\mathsf{F}(\lambda') - \mathsf{F}(\lambda)$ is nonnegative definite.) We note here that $\mathsf{F}(\cdot)$ has a derivative almost everywhere, which we denote by $\mathsf{F}'(\cdot)$ and which is for almost every $\lambda \in [-1/2, 1/2]$ nonnegative definite and Hermitian [50, Sec. 7, (7.5)].

Our goal is to predict $\mathbf{A}_0$ from $\mathbf{A}_{-1}, \ldots, \mathbf{A}_{-k}$, i.e., we are aiming for a predictor $\overline{\mathbf{A}}_0^{(k)}$ (which is a function of $\mathbf{A}_{-k}^{-1}$) that minimizes

$$\mathsf{E}\left[\left(\mathbf{A}_0 - \overline{\mathbf{A}}_0^{(k)}\right)^\dagger \left(\mathbf{A}_0 - \overline{\mathbf{A}}_0^{(k)}\right)\right].$$

(In the univariate case, the above expectation reduces to the mean-square error.) It is well known that this predictor is given by

$$\overline{\mathbf{A}}_0^{(k)} = \mathsf{E}\left[\mathbf{A}_0 \;\middle|\; \mathbf{A}_{-k}^{-1}\right].$$

Note that, since $\{\mathbf{A}_k,\, k \in \mathbb{Z}\}$ is a Gaussian process, $\overline{\mathbf{A}}_0^{(k)}$ is of the form

$$\overline{\mathbf{A}}_0^{(k)} = \sum_{\ell=1}^{k} \mathsf{C}_\ell \mathbf{A}_{-\ell},$$

hence the section title "Linear Prediction". Further note that $\overline{\mathbf{A}}_0^{(k)}$ is a Gaussian vector that is independent of $\mathbf{A}_0 - \overline{\mathbf{A}}_0^{(k)}$ [50, Lemma 5.8]. This implies that, conditional on $\mathbf{A}_{-1}, \ldots, \mathbf{A}_{-k}$, the prediction error

$$\widetilde{\mathbf{A}}_0^{(k)} = \mathbf{A}_0 - \overline{\mathbf{A}}_0^{(k)}$$

is a zero-mean Gaussian vector whose conditional covariance matrix, conditional on $\mathbf{A}_{-k}^{-1} = \mathbf{a}_{-k}^{-1}$, does not depend on $\mathbf{a}_{k-1}^{-1}$, i.e.,

$$\mathsf{E}\left[\widetilde{\mathbf{A}}_0^{(k)} \left(\widetilde{\mathbf{A}}_0^{(k)}\right)^\dagger \;\middle|\; \mathbf{A}_{-k}^{-1} = \mathbf{a}_{-k}^{-1}\right] = \mathsf{E}\left[\widetilde{\mathbf{A}}_0^{(k)} \left(\widetilde{\mathbf{A}}_0^{(k)}\right)^\dagger\right]. \tag{4.16}$$

By [50, Lemmas 5.7(b) & 5.10(c)] the prediction error in predicting $\mathbf{A}_0$ from $\mathbf{A}_{-1}, \ldots, \mathbf{A}_{-k}$ converges to the prediction error in predicting $\mathbf{A}_0$ from $\mathbf{A}_{-1}, \mathbf{A}_{-2}, \ldots$, i.e.,

$$\lim_{k\to\infty} \mathsf{E}\left[\widetilde{\mathbf{A}}_0^{(k)} \left(\widetilde{\mathbf{A}}_0^{(k)}\right)^\dagger\right] = \mathsf{E}\left[\widetilde{\mathbf{A}}_0 \widetilde{\mathbf{A}}_0^\dagger\right], \tag{4.17}$$

where $\widetilde{\mathbf{A}}_0$ denotes the prediction error in predicting $\mathbf{A}_0$ from $\mathbf{A}_{-1}, \mathbf{A}_{-2}, \ldots$, i.e.,

$$\widetilde{\mathbf{A}}_0 = \mathbf{A}_0 - \mathsf{E}\left[\mathbf{A}_0 \;\middle|\; \mathbf{A}_{-\infty}^{-1}\right].$$

Let Σ denote the covariance matrix of $\widetilde{\mathbf{A}}_0$. It was shown by Wiener and Masani [50, Main Thm. I, p. 145] that if the spectral distribution function satisfies

$$\int_{-1/2}^{1/2} \log\det \mathsf{F}'(\lambda)\, \mathrm{d}\lambda > -\infty, \tag{4.18}$$

then Σ is given by

$$\det \Sigma = \exp\left(\int_{-1/2}^{1/2} \log \det \mathsf{F}'(\lambda) \, \mathrm{d}\lambda \right). \tag{4.19}$$

We shall refer to Σ as the prediction error covariance matrix or, in short, the *prediction error*. In the univariate case ($\eta = 1$) we shall denote the prediction error by ϵ^2, i.e.,

$$\epsilon^2 = \exp\left(\int_{-1/2}^{1/2} \log F'(\lambda) \, \mathrm{d}\lambda \right). \tag{4.20}$$

Equation (4.19) implies that all processes satisfying (4.18) yield a positive determinant $\det \Sigma$ (and hence also a positive entropy rate $h\big(\{\mathbf{A}_k\}\big)$) and are thus regular. To study nonregular processes we shall analyze the *noisy* prediction problem: Let $\{\mathbf{W}_k,\, k \in \mathbb{Z}\}$ be a sequence of IID, zero-mean, circularly-symmetric, complex Gaussian random vectors of covariance matrix I_η. The noisy prediction problem is to predict $\mathbf{A}_0$ from a noisy observation of its past $\mathbf{A}_{-1} + \delta\,\mathbf{W}_{-1},\, \mathbf{A}_{-2} + \delta\,\mathbf{W}_{-2}, \ldots$ Let $\Sigma(\cdot)$ denote the covariance matrix of the prediction error of the noisy prediction problem. Then, by extending the derivations in [26, Sec. III] to the multivariate case, we have

$$\det\big(\Sigma(\delta) + \delta\,\mathsf{I}_\eta\big) = \exp\left(\int_{-1/2}^{1/2} \log \det\big(\mathsf{F}'(\lambda) + \delta\,\mathsf{I}_\eta\big) \, \mathrm{d}\lambda \right), \quad \delta \geq 0. \tag{4.21}$$

We shall refer to $\Sigma(\cdot)$ as the *noisy prediction error*. In the univariate case, we shall denote the noisy prediction error by $\epsilon^2(\cdot)$, i.e.,

$$\epsilon^2(\delta) = \exp\left(\int_{-1/2}^{1/2} \log\big(F'(\lambda) + \delta\big) \, \mathrm{d}\lambda \right) - \delta, \quad \delta \geq 0. \tag{4.22}$$

We finally note that $\delta \mapsto \log \det\big(\Sigma(\delta) + \delta\big)$ is monotonically decreasing in δ and that for all $\mathsf{F}(\cdot)$ satisfying (4.18) we have

$$\lim_{\delta \downarrow 0} \log \det\big(\Sigma(\delta) + \delta\big) = \log \det \Sigma.$$

4.4 Nonasymptotic Bounds

4.4.1 SISO Fading Channels

The asymptotic capacity of SISO fading channels with memory is well understood. For instance, the fading number of a (regular) mean-d, unit-variance, SISO Gaussian fading channel with spectral distribution function $F(\cdot)$ is given by [28, Cor. 4.42]

$$\chi(\{H_k\}) = \log|d|^2 - \mathrm{Ei}\big(-|d|^2\big) - 1 + \log\frac{1}{\epsilon^2}, \tag{4.23}$$

where $\mathrm{Ei}(\cdot)$ denotes the exponential integral function, i.e.,

$$\mathrm{Ei}(-x) = -\int_x^\infty \frac{e^{-t}}{t}\,\mathrm{d}t, \quad x > 0, \tag{4.24}$$

and where ϵ^2 is given in (4.20). If we view the fading number as an indication of the maximal rate at which power-efficient communication is achievable (cf. [28, Sec. IV-C]), then it follows from (4.23) that this maximal rate only depends on the memory of the channel through the (noiseless) prediction error.

Nonasymptotic upper and lower bounds on the capacity of SISO Gaussian fading channels with memory under a peak-power constraint on the inputs were given in [26]. It was shown that the capacity is upper bounded by

$$C_{\mathrm{PP}}(\mathrm{SNR}) \le C_{\mathrm{PP}}^{(\mathrm{IID})}(\mathrm{SNR}) + \log\frac{1}{\epsilon^2(1/\mathrm{SNR})}, \tag{4.25}$$

where $C_{\mathrm{PP}}^{(\mathrm{IID})}(\mathrm{SNR})$ denotes the capacity in the memoryless fading case, and where $\epsilon^2(\cdot)$ is given in (4.22). The capacity is lower bounded by

$$\begin{aligned} C_{\mathrm{PP}}(\mathrm{SNR}) \ge \log\frac{1}{\epsilon^2(4/\mathrm{SNR}) + \frac{8}{5\,\mathrm{SNR}}} + \log|d|^2 \\ - \mathrm{Ei}\left(-\frac{|d|^2}{1-\epsilon^2(4/\mathrm{SNR})}\right) - \log\frac{5e}{6}. \end{aligned} \tag{4.26}$$

By evaluating the second term on the right-hand side (RHS) of [26, Eq. (28)] instead of further upper bounding it (as it is done in [26]), the upper bound (4.25) can be tightened to

$$C_{\mathrm{PP}}(\mathrm{SNR}) \leq C_{\mathrm{PP}}^{(\mathrm{IID})}(\mathrm{SNR}) + \log \frac{1 + 1/\mathrm{SNR}}{\epsilon^2(1/\mathrm{SNR}) + 1/\mathrm{SNR}}. \tag{4.27}$$

The capacity of the memoryless SISO Rayleigh-fading channel (i.e., zero-mean Gaussian fading) can be upper bounded by [28, Eq. (141)]

$$C_{\mathrm{PP}}^{(\mathrm{IID})}(\mathrm{SNR}) \leq \inf_{\alpha,\beta'>0} \inf_{\delta'\geq 0} \Bigg\{ -1 + \alpha \log \beta' + \log \Gamma\big(\alpha, \delta'/\beta'\big) + \frac{1+\mathrm{SNR}}{\beta'} + \frac{\delta'}{\beta'} + (1-\alpha)\Big(\log \delta' - e^{\delta'} \operatorname{Ei}(-\delta')\Big) \Bigg\}. \tag{4.28}$$

The lower bound (4.26) can be improved by considering inputs that are IID and uniformly distributed over the set $\{z \in \mathbb{C} : \alpha \mathsf{A} \leq |z| \leq \mathsf{A}\}$ with $0 < \alpha < 1$ (instead of considering inputs that are uniformly distributed over the set $\{z \in \mathbb{C} : \mathsf{A}/2 \leq |z| \leq \mathsf{A}\}$ as in [26]), and by maximizing over α:

$$C_{\mathrm{PP}}(\mathrm{SNR}) \geq \sup_{0<\alpha<1} \Bigg\{ -\log \frac{e(1+\alpha^2)}{2(1-\alpha^2)} - \exp\left(\frac{\epsilon^2(\xi) + \frac{2}{\mathrm{SNR}(1+\alpha^2)}}{1-\epsilon^2(\xi)}\right) \operatorname{Ei}\left(-\frac{\epsilon^2(\xi) + \frac{2}{\mathrm{SNR}(1+\alpha^2)}}{1-\epsilon^2(\xi)}\right) \Bigg\}, \tag{4.29}$$

where $\xi = \frac{1}{\alpha^2\,\mathrm{SNR}}$. However, the gap between (4.27) and (4.29) is still substantial. Moreover, for regular fading the lower bound (4.29) does not tend to infinity as the SNR tends to infinity. The following proposition presents a lower bound on the capacity that behaves comparably to (4.29) at low and moderate SNR, and that achieves the correct asymptotic behavior at high SNR.

Proposition 4.1. *Consider a mean-d, SISO Gaussian fading channel where the fading process is such that $\{H_k - d,\, k \in \mathbb{Z}\}$ is a zero-mean, unit-variance, stationary and ergodic, circularly-symmetric, complex*

Gaussian process of spectral distribution function $F(\cdot)$. Then

$$C_{\mathrm{PP}}(\mathrm{SNR}) \geq \sup_{0<\alpha<1} \left\{ \log \frac{1}{\epsilon^2(\xi)+\xi} + \log \frac{1}{\alpha} + \log \frac{1}{\mathrm{SNR}} \right. \\ \left. - \exp\left(\frac{e}{\log(1/\alpha^2)\,\alpha\,\mathrm{SNR}} \right) \mathrm{Ei}\left(-\frac{e}{\log(1/\alpha^2)\,\alpha\,\mathrm{SNR}} \right) \right\}, \tag{4.30}$$

where $\xi = \frac{1}{\alpha^2\,\mathrm{SNR}}$.

Proof. See Section 4.8.1. □

This lower bound will be used later to compute the capacity pre-loglog Λ_{PP} (i.e., the limiting ratio of capacity to $\log\log\mathrm{SNR}$ as SNR tends to infinity), when the fading is such that capacity increases double-logarithmically in the SNR (Theorem 4.16). It can also be used to compute the capacity pre-log of SISO Gaussian fading channels (4.14) as well as the fading number (4.23) for zero-mean Gaussian fading.

A better lower bound can be found numerically as follows. Let

$$C_{\mathrm{SI@Rx}}\left(\frac{x_{\min}^2}{\sigma^2}, \frac{x_{\max}^2}{\sigma^2}, \xi^2 \right) \triangleq \sup I(X; (D+\tilde{H})X + Z | D), \tag{4.31}$$

where the maximization is over all input distributions on X (independent of D) under which with probability one

$$x_{\min} \leq |X| \leq x_{\max}, \tag{4.32}$$

and where D, $\tilde{H}$, and Z are zero-mean, circularly-symmetric, complex Gaussian random variables, independent of each other and of X and of variances $1-\xi^2$, ξ^2, and σ^2, respectively. The capacity of a zero-mean, unit-variance, SISO Gaussian fading channel is lower bounded by

$$C_{\mathrm{PP}}(\mathrm{SNR}) \geq \sup_{x_{\min}^2 \leq \mathsf{A}^2} C_{\mathrm{SI@Rx}}\left(\frac{x_{\min}^2}{\sigma^2}, \mathrm{SNR}, \epsilon^2\left(\frac{\sigma^2}{x_{\min}^2} \right) \right). \tag{4.33}$$

This lower bound (4.33) can be computed numerically [51].

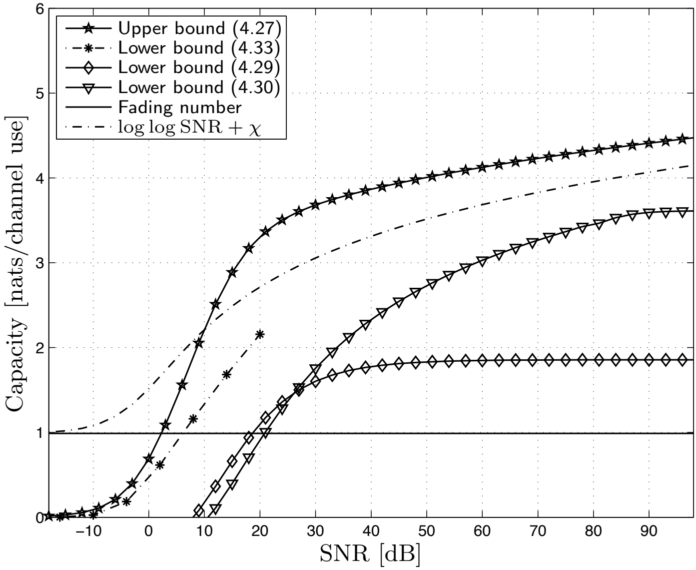

Figure 4.1: Bounds on the capacity of a zero-mean, unit-variance, SISO Gaussian fading channel with memory. Depicted are the capacity upper bound (4.27); the lower bound (4.29); the lower bound (4.30); the numerical lower bound (4.33); the asymptotic expansion $\log\bigl(1+\log(1+\text{SNR})\bigr)+\chi(\{H_k\})$; and the fading number $\chi(\{H_k\})$.

Figure 4.1 depicts the above analytic bounds (4.27), (4.29), and (4.30) together with the numerical lower bound (4.33) for a fading process whose spectral distribution function is of the form

$$F'(\lambda) = \begin{cases} \gamma_1, & \text{for } |\lambda| \leq \gamma_3 \\ \gamma_2, & \text{for } \gamma_3 < |\lambda| \leq \frac{1}{2}, \end{cases} \tag{4.34}$$

where γ_1 and γ_2 are positive real numbers, and where γ_3 is a positive real number satisfying $0 < \gamma_3 < 1/2$. Additionally, the asymptotic expansion

$$\log\bigl(1 + \log(1 + \text{SNR})\bigr) + \chi(\{H_k\})$$

and the fading number

$$\chi(\{H_k\}) = -1 - \gamma + \log\frac{1}{\epsilon^2}$$

(where $\gamma \approx 0.577$ denotes Euler's constant) are plotted in order to illustrate at which SNR the approximation

$$C(\text{SNR}) \approx \log\bigl(1 + \log(1 + \text{SNR})\bigr) + \chi(\{H_k\})$$

is reasonable.

In the following we consider two spectral distribution functions $F_1(\cdot)$ and $F_2(\cdot)$ of the form (4.34), where the parameters γ_1, γ_2, and γ_3 are chosen so that

$$\int_{-1/2}^{1/2} F'_\ell(\lambda)\,\mathrm{d}\lambda = 1, \quad \ell = 1, 2 \tag{4.35}$$

and

$$\epsilon_\ell^2 = \exp\left\{\int_{-1/2}^{1/2} \log F'_\ell(\lambda)\,\mathrm{d}\lambda\right\} = \frac{1}{100}, \quad \ell = 1, 2. \tag{4.36}$$

Thus we consider two unit-variance Gaussian fading processes with the same (noiseless) prediction error (of 1/100), but with different spectral densities. By controlling γ_1 while maintaining (4.35) and (4.36), we can control the sensitivity of the noisy prediction error with respect to the variance of the noise corrupting the observations.

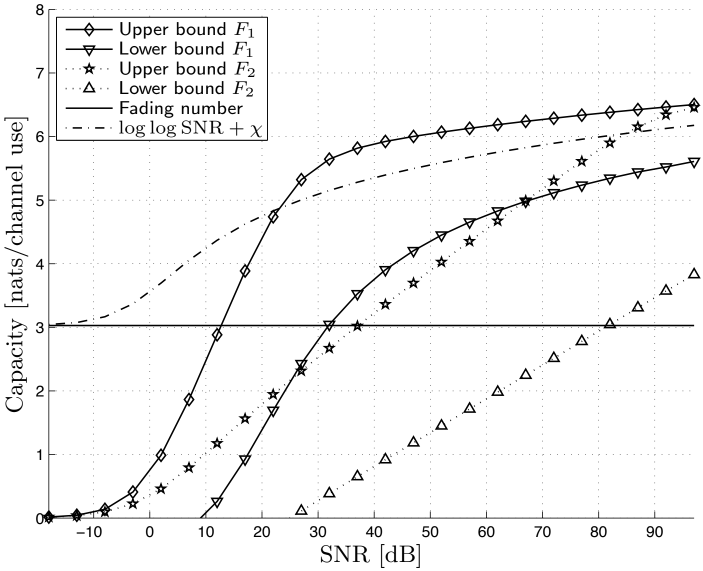

Figure 4.2: Bounds on the capacity of a zero-mean, unit-variance, SISO Gaussian fading channel with memory. Depicted are the upper bound (4.27) and the lower bound (4.30) for the spectral distribution functions $F_1(\cdot)$ and $F_2(\cdot)$; the asymptotic expansion $\log\bigl(1+\log(1+\mathrm{SNR})\bigr)+\chi(\{H_k\})$; and the fading number $\chi(\{H_k\})$.

Figure 4.2 depicts upper and lower bounds on the capacity of zero-mean, unit-variance, SISO Gaussian fading channels with memory for the two spectral distribution functions $F_1(\cdot)$ and $F_2(\cdot)$. The high-SNR asymptotic capacities corresponding to the two fading laws are identical, because these asymptotics are determined by the (noiseless) prediction error, which is identical for the two. At lower SNR the behaviors are, however, very different. The capacity lower bound corresponding to $F_1(\cdot)$ enters the power-inefficient regime at about 45 dB whereas the capacity upper bound corresponding to $F_2(\cdot)$ achieves this regime at far higher SNR, namely at about 90 dB. Thus, while the high-SNR expansion of channel capacity depends on the fading memory only via its (noiseless) prediction error, the moderate-SNR behavior depends on the memory more finely, namely, via the functional dependence of the noisy prediction error on the variance of the noise corrupting the observations.

This demonstrates that while the (noiseless) prediction error can sometimes indicate the *rates* above which communication becomes power-inefficient, it cannot indicate the corresponding SNR. For the latter one needs the functional dependence of the noisy prediction error on the variance of the corrupting noise.

4.4.2 MIMO Fading Channels

In the following we present nonasymptotic upper bounds on the capacity of MIMO fading channels.

Theorem 4.2. *Consider a mean-*D*, spatially independent, MIMO Gaussian fading channel where the fading process is such that* $\{H_k(r,t) - d(r,t),\ k \in \mathbb{Z}\}$ *is a zero-mean, unit-variance, stationary and ergodic, circularly-symmetric, complex Gaussian process of spectral distribution function* $F_{r,t}(\cdot)$*. Then*

$$\begin{aligned} &C_{\mathrm{PP}}(\mathrm{SNR}) \\ &\quad \leq C_{\mathrm{PP}}^{(\mathrm{IID})}(\mathrm{SNR}) + \max_{\|\hat{\mathbf{x}}\|=1} \sum_{r=1}^{n_{\mathrm{R}}} \log \frac{1 + \frac{1}{\mathrm{SNR}}}{\sum_{t=1}^{n_{\mathrm{T}}} |\hat{x}(t)|^2\, \epsilon_{r,t}^2\!\left(\frac{1}{\mathrm{SNR}}\right) + \frac{1}{\mathrm{SNR}}}, \end{aligned} \quad (4.37)$$

where $C_{\mathrm{PP}}^{(\mathrm{IID})}(\mathrm{SNR})$ *denotes the capacity in the memoryless fading case,* *and where $\epsilon^2_{r,t}(\cdot)$ denotes the error in predicting the (r,t)-th component of the fading matrix from a noisy observation of its past* (4.22)*.*

Proof. See [22, Thm. 5.14]. A proof can be found in Appendix B.1. □

When the fading process $\{\mathbb{H}_k - \mathsf{D},\ k \in \mathbb{Z}\}$ is spatially IID, (4.37) reduces to the following upper bound:

Corollary 4.3. *Consider a mean-*D*, MIMO Gaussian fading channel where the fading process is such that $\{\mathbb{H}_k - \mathsf{D},\ k \in \mathbb{Z}\}$ is spatially IID with each component being a zero-mean, unit-variance, stationary and ergodic, circularly-symmetric, complex Gaussian process of spectral distribution function $F(\cdot)$. Then, denoting $\epsilon^2_{r,t}(\cdot)$ by $\epsilon^2(\cdot)$,*

$$C_{\mathrm{PP}}(\mathrm{SNR}) \le C^{(\mathrm{IID})}_{\mathrm{PP}}(\mathrm{SNR}) + n_{\mathrm{R}} \log \frac{1 + 1/\mathrm{SNR}}{\epsilon^2(1/\mathrm{SNR}) + 1/\mathrm{SNR}}. \qquad (4.38)$$

Proof. Follows directly from Theorem 4.2 by noting that, since the process $\{\mathbb{H}_k - \mathsf{D},\ k \in \mathbb{Z}\}$ is spatially IID, $\epsilon^2_{r,t}(\cdot)$ does not depend on (r,t). □

An upper bound on the capacity of spatially independent MISO fading channels can be found by using Theorem 4.2 with $n_{\mathrm{R}} = 1$. The following theorem generalizes this bound to channels where the fading is not spatially independent.

For convenience we shall write the MISO fading process as a column vector and not as a row vector. The time-k channel output Y_k is given by

$$Y_k = \mathbf{H}_k^{\mathsf{T}} \mathbf{x}_k + Z_k, \quad k \in \mathbb{Z}. \qquad (4.39)$$

Theorem 4.4. *Consider a mean-*$\mathbf{d}$*, MISO Gaussian fading channel where $\{\mathbf{H}_k - \mathbf{d},\ k \in \mathbb{Z}\}$ is a zero-mean, stationary and ergodic, circularly-symmetric, complex Gaussian process of matrix-valued spectral distribution function $\mathsf{F}(\cdot)$. Further assume that the covariance matrix*

$$\mathsf{K} = \mathsf{E}\big[(\mathbf{H}_k - \mathbf{d})(\mathbf{H}_k - \mathbf{d})^{\dagger}\big]$$

is nonsingular. Then

$$C_{\mathrm{PP}}(\mathrm{SNR}) \leq C_{\mathrm{PP}}^{(\mathrm{IID})}(\mathrm{SNR}) + \log \frac{\|\mathsf{K}\| + 1/\mathrm{SNR}}{\lambda_{\min}(1/\mathrm{SNR}) + 1/\mathrm{SNR}}, \tag{4.40}$$

where $\lambda_{\min}(\cdot)$ *denotes the smallest eigenvalue of the (matrix-valued) noisy prediction error* $\Sigma(\cdot)$.

Proof. See [22, Thm. 5.12]. A proof can be found in Appendix B.2. □

4.5 The Fading Number

In this section we present lower and upper bounds on the fading number (4.11) of MIMO fading channels. We will assume throughout this section that the fading satisfies the finite differential entropy rate condition (4.3), i.e., that the fading is regular.

4.5.1 Lower Bounds

Proposition 4.5. *Consider a stationary and ergodic fading process* $\{\mathbb{H}_k,\ k \in \mathbb{Z}\}$ *that satisfies* (4.2) *and* (4.3)*. For any* n_{T}*-variate stationary and ergodic process* $\{\mathbf{V}_k,\ k \in \mathbb{Z}\}$ *that is independent of* $\{\mathbb{H}_k,\ k \in \mathbb{Z}\}$ *and that satisfies*

$$\lim_{k\to\infty} I\bigl(\mathbf{V}_k; \mathbf{V}_1^{k-1}\bigr) < \infty \tag{4.41}$$

and

$$\mathsf{E}\bigl[\|\mathbf{V}_k\|^2\bigr] < \infty \qquad \text{and} \qquad \mathsf{E}\bigl[\log \|\mathbf{V}_k\|^2\bigr] > -\infty, \tag{4.42}$$

the fading number $\chi_{\mathrm{Avg}}(\{\mathbb{H}_k\})$ *is lower bounded by*

$$\chi_{\mathrm{Avg}}(\{\mathbb{H}_k\}) \geq \chi(\{\mathbb{H}_k\mathbf{V}_k\}) + \lim_{n\to\infty} \frac{1}{n} I\bigl(\mathbf{V}_1^n; \{\mathbb{H}_k\mathbf{V}_k\}_{k=1}^n\bigr), \tag{4.43}$$

where $\chi(\{\mathbb{H}_k\mathbf{V}_k\})$ *is the fading number of a SIMO channel with fading* $\{\mathbb{H}_k\mathbf{V}_k,\ k \in \mathbb{Z}\}$*. (For SIMO fading, peak-power and average-power constraints yield the same fading number* [29]*. Note that the limit in* (4.41) *exists because* $\{\mathbf{V}_k,\ k \in \mathbb{Z}\}$ *is stationary.)*

Moreover, if $\{\mathbf{V}_k,\ k \in \mathbb{Z}\}$ *additionally satisfies*

$$\Pr(\|\mathbf{V}_k\| > \Upsilon) = 0, \quad \textit{for some } \Upsilon > 0, \tag{4.44}$$

then the lower bound holds also for the fading number $\chi_{\text{PP}}(\{\mathbb{H}_k\})$ *under a peak-power constraint.*

Proof. See Section 4.8.2. □

Note 4.1. *The assumption that the fading is Gaussian is not necessary. Proposition 4.5 holds for any stationary and ergodic fading process that satisfies* (4.2) *and* (4.3)*.*

An exact expression for the fading number $\chi(\{\mathbb{H}_k \mathbf{V}_k\})$ of the SIMO fading $\{\mathbb{H}_k \mathbf{V}_k,\ k \in \mathbb{Z}\}$ is given in [29]. However, this expression is not easy to evaluate. It can always be lower bounded by considering linear combining at the receiver, which reduces the SIMO channel to a SISO channel for which the fading number is easier to compute (4.12)

$$\chi(\{\mathbb{H}_k \mathbf{V}_k\}) \geq \chi(\{\boldsymbol{\alpha}^\mathsf{T} \mathbb{H}_k \mathbf{V}_k\}), \quad \boldsymbol{\alpha} \in \mathbb{C}^{n_\text{R}} \text{ deterministic}, \tag{4.45}$$

or by ignoring the memory in $\{\mathbb{H}_k \mathbf{V}_k,\ k \in \mathbb{Z}\}$

$$\chi(\{\mathbb{H}_k \mathbf{V}_k\}) \geq \chi^{(\text{IID})}(\mathbb{H}_1 \mathbf{V}_1), \tag{4.46}$$

or by applying both reductions

$$\chi(\{\mathbb{H}_k \mathbf{V}_k\}) \geq \chi^{(\text{IID})}(\boldsymbol{\alpha}^\mathsf{T} \mathbb{H}_1 \mathbf{V}_1), \quad \boldsymbol{\alpha} \in \mathbb{C}^{n_\text{R}} \text{ deterministic}, \tag{4.47}$$

where $\chi^{(\text{IID})}$ denotes the fading number in the memoryless case with equal marginals. The advantage of (4.46) and (4.47) is that it only depends on the marginal law of $\{\mathbb{H}_k,\ k \in \mathbb{Z}\}$.

Note 4.2. *For SIMO fading the lower bound* (4.43) *is tight, i.e., for any stationary and ergodic process* $\{\mathbf{V}_k,\ k \in \mathbb{Z}\}$ *that is independent of* $\{\mathbb{H}_k,\ k \in \mathbb{Z}\}$ *and that satisfies* (4.41) *and* (4.42)*, the RHS of* (4.43) *is equal to the SIMO fading number* $\chi(\{\mathbf{H}_k\})$*.*

Proof. See Section 4.8.3. □

We can use Theorem 4.5 to establish the following result on slowly-varying Gaussian fading channels:

Theorem 4.6. *Let the MIMO fading* $\{\mathbb{H}_k,\, k \in \mathbb{Z}\}$ *be spatially IID with each component of* $\{\mathbb{H}_k,\, k \in \mathbb{Z}\}$ *being a zero-mean, unit-variance, stationary and ergodic, circularly-symmetric, complex Gaussian process of temporal autocorrelation function*

$$\mathsf{K}(\nu) = \mathsf{E}\left[H_{k+\nu}(r,t)H_k^*(r,t)\right], \quad \nu \in \mathbb{Z}.$$

Let $n_{\min} = \min\{n_{\mathrm{R}}, n_{\mathrm{T}}\}$*, and let*

$$\begin{aligned}\epsilon_{\max}^2 &\triangleq \max_{1\le\nu\le n_{\min}+2} \mathsf{E}\left[\left|H_\nu(r,t) - H_0(r,t)\right|^2\right]\\ &= 2\max_{1\le\nu\le n_{\min}+2}\Big(1 - \mathrm{Re}\big(\mathsf{K}(\nu)\big)\Big),\end{aligned}$$

where $\mathrm{Re}\big(\mathsf{K}(\nu)\big)$ *denotes the real part of* $\mathsf{K}(\nu)$*. (Note that* K *and* $\epsilon_{\max}^2$ *do not depend on* (r,t) *because the fading is spatially IID.) Then*

$$\chi_{\mathrm{PP}}(\{\mathbb{H}_k\}) \ge n_{\min}\log\frac{1}{\epsilon_{\max}^2} + \Delta(n_{\min}), \tag{4.48}$$

where the correction term $\Delta(n_{\min})$ *depends only on* $n_{\min}$ *and not on the temporal autocorrelation function* K.

Proof. See Section 4.8.4. □

4.5.2 Upper Bounds

Corollary 4.7. *Let the mean-*D*, spatially independent, MIMO Gaussian fading channel be such that* $\{H_k(r,t) - d(r,t),\, k \in \mathbb{Z}\}$ *is a zero-mean, unit-variance, stationary and ergodic, circularly-symmetric, complex Gaussian process of spectral distribution function* $F_{r,t}(\cdot)$*. Then*

$$\chi_{\mathrm{Avg}}(\{\mathbb{H}_k\}) \le \chi_{\mathrm{Avg}}^{(\mathrm{IID})}(\mathbb{H}_1) + \max_{\|\hat{\mathbf{x}}\|=1}\sum_{r=1}^{n_{\mathrm{R}}}\log\frac{1}{\sum_{t=1}^{n_{\mathrm{T}}}|\hat{x}(t)|^2\epsilon_{r,t}^2}, \tag{4.49}$$

where $\epsilon_{r,t}^2 = \epsilon_{r,t}^2(0)$ *in (4.37).*

Proof. Follows directly from Theorem 4.2 by computing the difference between the RHS of (4.37) and $\log\log \mathrm{SNR}$ in the limit as SNR tends to infinity. □

When the fading process $\{\mathbb{H}_k - \mathsf{D},\ k \in \mathbb{Z}\}$ is spatially IID, (4.49) reduces to the following upper bound:

Corollary 4.8. *Let the mean-*D*, MIMO Gaussian fading channel be such that the process* $\{\mathbb{H}_k - \mathsf{D},\ k \in \mathbb{Z}\}$ *is spatially IID with each component being a zero-mean, unit-variance, stationary and ergodic, circularly-symmetric, complex Gaussian process of spectral distribution function* $F(\cdot)$*. Then, denoting* $\epsilon^2_{r,t}$ *by* ϵ^2*,*

$$\chi(\{\mathbb{H}_k\}) \leq \chi^{(\mathrm{IID})}_{\mathrm{Avg}}(\mathbb{H}_1) + n_{\mathrm{R}} \log \frac{1}{\epsilon^2}. \tag{4.50}$$

Proof. Follows directly from Corollary 4.7 by noting that, since the fading process $\{\mathbb{H} - \mathsf{D},\ k \in \mathbb{Z}\}$ is spatially IID, $\epsilon^2_{r,t}$ does not depend on (r, t). □

Note that $\chi^{(\mathrm{IID})}_{\mathrm{Avg}}(\mathbb{H}_1)$ is unknown for general fading matrices $\mathbb{H}_1$. However, it is known in the case where $\mathbb{H}_1$ is rotation commutative. General and specific upper bounds on $\chi^{(\mathrm{IID})}_{\mathrm{Avg}}(\mathbb{H}_1)$ for the case where the $n_{\mathrm{R}} \times n_{\mathrm{T}}$ matrix $\mathbb{H}_1$ is of the form $\mathbb{H}_1 = \mathsf{D} + \tilde{\mathbb{H}}_1$ (where D is deterministic and $\tilde{\mathbb{H}}_1$ is spatially IID with each component of $\tilde{\mathbb{H}}_1$ being a zero-mean, unit-variance, circularly-symmetric, complex Gaussian random variable) are given in [28, Sec. IV-D].

4.5.3 MISO Fading

An upper bound on the fading number of spatially independent MISO Gaussian channels follows from Corollary 4.7 by recalling that [28, Cor. 4.28]

$$\chi^{(\mathrm{IID})}(\mathbf{H}_1) = -1 + \log d_\star^2 - \mathrm{Ei}\bigl(-d_\star^2\bigr), \tag{4.51}$$

where

$$d_\star = \max_{\|\hat{\mathbf{x}}\|=1} \frac{\left|\mathsf{E}[\mathbf{H}_k^{\mathsf{T}}]\,\hat{\mathbf{x}}\right|}{\sqrt{\mathsf{Var}(\mathbf{H}_k^{\mathsf{T}}\hat{\mathbf{x}})}}. \tag{4.52}$$

Corollary 4.9. *Let the mean-*$\mathbf{d}$*, spatially independent, MISO Gaussian fading channel be such that* $\{H(t) - d(t),\, k \in \mathbb{Z}\}$ *is a zero-mean, unit-variance, stationary and ergodic, circularly-symmetric, complex Gaussian process of spectral distribution function* $F_t(\cdot)$*. Then*

$$\chi(\{\mathbf{H}_k\}) \leq -1 + \log d_\star^2 - \mathrm{Ei}\big(-d_\star^2\big) + \log \frac{1}{\epsilon_{\min}^2}, \tag{4.53}$$

where

$$\epsilon_{\min}^2 = \min_{1 \leq t \leq n_\mathrm{T}} \exp\bigg(\int_{-1/2}^{1/2} \log F_t'(\lambda)\,\mathrm{d}\lambda\bigg). \tag{4.54}$$

Proof. Follows immediately from Corollary 4.7 and (4.51) by noting that

$$\max_{\|\hat{\mathbf{x}}\|=1} \log \frac{1}{\sum_{t=1}^{n_\mathrm{T}} |\hat{x}(t)|^2\, \epsilon_t^2} = \log \frac{1}{\epsilon_{\min}^2}.$$ □

For some special cases this upper bound is tight:

Corollary 4.10. *Consider a zero-mean, spatially independent, MISO Gaussian fading channel where the fading process is such that* $\{H(t),\, k \in \mathbb{Z}\}$ *is a zero-mean, unit-variance, stationary and ergodic, circularly-symmetric, complex Gaussian process of spectral distribution* $F_t(\cdot)$*. Then*

$$\chi(\{\mathbf{H}_k\}) = -1 - \gamma + \log \frac{1}{\epsilon_{\min}^2}. \tag{4.55}$$

Moreover, this fading number can be achieved by transmitting from the antenna that yields the smallest prediction error, while keeping the other antennas silent.

Proof. The upper bound follows from Corollary 4.9 by noting that [28, Eqs. (210)–(213)]

$$\log d_\star^2 - \mathrm{Ei}\big(-d_\star^2\big) = -\gamma.$$

To derive a lower on the fading number, we define

$$t_\star \triangleq \arg \min_{1 \leq t \leq n_\mathrm{T}} \exp\bigg(\int_{-1/2}^{1/2} \log F_t'(\lambda)\,\mathrm{d}\lambda\bigg).$$

By transmitting only from antenna $t_\star$ while keeping the other antennas silent, we obtain the fading number of a zero-mean, SISO Gaussian fading channel of spectral distribution function $F_{t_\star}(\cdot)$, which is given by (4.23)

$$-1-\gamma+\log\frac{1}{\epsilon_{t_\star}^2}. \qquad \square$$

Corollary 4.11. *Consider a mean-$\mathbf{d}$, MISO Gaussian fading channel where the fading process is such that $\{\mathbf{H}_k - \mathbf{d},\, k \in \mathbb{Z}\}$ is spatially IID with each of its components being a zero-mean, unit-variance, stationary and ergodic, circularly-symmetric, complex Gaussian process of spectral distribution function $F(\cdot)$. Then*

$$\chi(\{\mathbf{H}_k\}) = -1 + \log\|\mathbf{d}\|^2 - \mathrm{Ei}\big(-\|\mathbf{d}\|^2\big) + \log\frac{1}{\epsilon^2}. \tag{4.56}$$

Moreover, this fading number is achievable with beam forming, i.e., by transmitting signals of the form

$$\mathbf{X}_k = \mathbf{d}^* X_k,$$

where the process $\{X_k,\, k \in \mathbb{Z}\}$ takes value in $\mathbb{C}$.

Proof. We first note that if the fading process is of covariance matrix $\mathsf{I}_{n_\mathrm{T}}$, then $d_\star^2 = \|\mathbf{d}\|^2$. We further note that if $\{\mathbf{H}_k - \mathbf{d},\, k \in \mathbb{Z}\}$ is spatially IID, then we have for any $\|\hat{\mathbf{x}}\| = 1$

$$\mathsf{E}\big[(\mathbf{H}_{k+m}-\mathbf{d})^\mathsf{T}\hat{\mathbf{x}}\,\hat{\mathbf{x}}^\dagger(\mathbf{H}_k-\mathbf{d})^*\big] = \mathsf{E}[(H_{k+m}(1)-\mathbf{d}^\mathsf{T}\hat{\mathbf{x}})(H_k(1)-\mathbf{d}^\mathsf{T}\hat{\mathbf{x}})^*]$$

and the process $\{\mathbf{H}_k^\mathsf{T}\hat{\mathbf{x}},\, k \in \mathbb{Z}\}$ is thus of spectral distribution function $F(\cdot)$. By transmitting signals of the form $\mathbf{d}^* X_k$ we therefore achieve the fading number of a mean-$\|\mathbf{d}\|$, Gaussian SISO fading channel of spectral distribution function $F(\cdot)$, which is given by (4.23). $\square$

An upper bound on the fading number of spatially independent MISO fading channels can be found using Corollary 4.7 with $n_\mathrm{R} = 1$. The following corollary generalizes this bound to channels where the fading is not spatially independent.

Corollary 4.12. *Let the mean-*$\mathbf{d}$*, MISO Gaussian fading channel be such that the process* $\{\mathbf{H}_k - \mathbf{d},\ k \in \mathbb{Z}\}$ *is a zero-mean, stationary and ergodic, circularly-symmetric, complex Gaussian process of covariance matrix* K *and of (matrix-valued) spectral distribution function* $\mathsf{F}(\cdot)$*. Then*

$$\chi_{\text{Avg}}(\{\mathbf{H}_k\}) \leq -1 + \log d_\star^2 - \text{Ei}\big(-d_\star^2\big) + \log \frac{\|\mathsf{K}\|}{\lambda_{\min}}, \tag{4.57}$$

where $\lambda_{\min}$ *denotes the smallest eigenvalue of* Σ*. (Note that the assumption that the fading is regular implies that* Σ *is nonsingular.)*

Proof. Follows directly from Theorem 4.4 by computing the difference between the RHS of (4.40) and $\log\log \text{SNR}$ in the limit as SNR tends to infinity. □

4.6 The Pre-Log

In this section we extend the results on the capacity pre-log of SISO fading channels [25, 26] to the multi-antenna case. In particular, we present upper bounds on the capacity pre-log of spatially independent MIMO fading channels, as well as the expression for the pre-log of spatially independent MISO fading channels.

Theorem 4.13. *Consider a mean-*D*, spatially independent, MIMO Gaussian fading channel where the fading process is such that* $\{H_k(r,t) - d(r,t),\ k \in \mathbb{Z}\}$ *is a zero-mean, unit-variance, stationary and ergodic, circularly-symmetric, complex Gaussian process of spectral distribution function* $F_{r,t}(\cdot)$*. Then*

$$\Pi_{\text{PP}} \leq \max_{1 \leq t \leq n_{\text{T}}} \sum_{r=1}^{n_{\text{R}}} \mu\left(\left\{\lambda : F'_{r,t}(\lambda) = 0\right\}\right). \tag{4.58}$$

Proof. See [22, Cor. 5.15]. A proof can be found in Appendix B.3. □

Corollary 4.14. *Consider a mean-*D*, MIMO Gaussian fading channel where the fading process is such that* $\{\mathbb{H}_k - \mathsf{D},\ k \in \mathbb{Z}\}$ *is spatially IID with each component being a zero-mean, unit-variance, stationary and* *ergodic, circularly-symmetric, complex Gaussian process of spectral distribution function $F(\cdot)$. Then*

$$\Pi_{\mathrm{PP}} \leq n_{\mathrm{R}}\, \mu\left(\{\lambda : F'(\lambda) = 0\}\right). \tag{4.59}$$

Proof. Follows directly from Theorem 4.13 by noting that if the fading process $\{\mathbb{H}_k - \mathsf{D},\, k \in \mathbb{Z}\}$ is spatially IID, then $F_{r,t}(\cdot)$ does not depend on (r,t). □

For spatially independent MISO Gaussian fading channels, the upper bound provided in Theorem 4.13 is tight.

Corollary 4.15. *Let the mean-$\mathbf{d}$, spatially independent, MISO Gaussian fading channel be such that $\{H_k(t) - d(t),\, k \in \mathbb{Z}\}$ is a zero-mean, unit-variance, stationary and ergodic, circularly-symmetric, complex Gaussian process of spectral distribution function $F_t(\cdot)$. Then*

$$\Pi_{\mathrm{PP}} = \max_{1 \leq t \leq n_{\mathrm{T}}} \mu\left(\{\lambda : F_t'(\lambda) = 0\}\right). \tag{4.60}$$

Moreover, this pre-log can be achieved by transmitting from the antenna that achieves the largest pre-log, while keeping the other antennas silent.

Proof. The upper bound follows directly from Theorem 4.13. To derive a lower bound on the pre-log, we define

$$t_\star \triangleq \arg \max_{1 \leq t \leq n_{\mathrm{T}}} \mu\left(\{\lambda : F_t'(\lambda) = 0\}\right).$$

By transmitting from antenna $t_\star$ while keeping the other antennas silent, we achieve the pre-log of a Gaussian SISO fading channel of spectral distribution function $F_{t_\star}(\cdot)$, which is given by (4.14)

$$\mu\left(\left\{\lambda : F_{t_\star}'(\lambda) = 0\right\}\right).$$

□

4.7 The Pre-LogLog

Lapidoth showed that the capacity grows double-logarithmically with the SNR if, and only if, [26, Sec. VII]

$$\varlimsup_{\delta \downarrow 0} \frac{-\int_{1/2}^{1/2} \log\bigl(F'(\lambda) + \delta\bigr)\, \mathrm{d}\lambda}{\log \log \frac{1}{\delta}} < \infty.$$

An expression for the capacity pre-loglog, which is defined as

$$\Lambda_{\text{PP}} \triangleq \varlimsup_{\text{SNR}\to\infty} \frac{C(\text{SNR})}{\log\log \text{SNR}},$$

was later derived in [22, Cor. 5.11].

Theorem 4.16. *Consider a mean-d, SISO Gaussian fading channel where the fading process is such that $\{H_k - d,\, k \in \mathbb{Z}\}$ is a zero-mean, unit-variance, stationary and ergodic, circularly-symmetric, complex Gaussian process of spectral distribution function $F(\cdot)$. Then*

$$\Lambda_{\text{PP}} = 1 + \varlimsup_{\delta\downarrow 0} \frac{-\int_{-1/2}^{1/2} \log\bigl(F'(\lambda) + \delta\bigr)\, \mathrm{d}\lambda}{\log\log \frac{1}{\delta}}. \tag{4.61}$$

Proof. An upper bound on Λ_{PP} follows by computing the limiting ratio of the upper bound (4.27) to $\log\log \text{SNR}$ as SNR tends to infinity. A lower bound on Λ_{PP} follows from Proposition 4.1 by evaluating the RHS of (4.30) for

$$\alpha^2 = \text{SNR}^{-(1-\beta)} \quad \text{for some } 0 < \beta < 1,$$

and by computing then the limiting ratio of this lower bound to $\log\log \text{SNR}$ as SNR tends to infinity. The details are carried out in Appendix B.4. □

4.8 Proofs

This section exhibits the main proofs of this chapter. Proposition 4.1 is proven in Section 4.8.1, Proposition 4.5 is proven in Section 4.8.2, Note 4.2 is proven in Section 4.8.3, and Theorem 4.6 is proven in Section 4.8.4.

4.8.1 Proof of Proposition 4.1

To derive a lower bound on $C_{\text{PP}}(\text{SNR})$, we evaluate $I(X_1^n; Y_1^n)$ for $\{X_k,\, k \in \mathbb{Z}\}$ being a sequence of IID, zero-mean, circularly-symmetric, complex random variables with $\log |X_k|^2$ uniformly distributed over

$[2\,\log(\alpha\mathsf{A}), 2\,\log\mathsf{A}]$. To this end we use the chain rule and a Cesàro-type theorem [5, Thm. 4.2.3] to lower bound

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}I\big(X_1^n;Y_1^n\big) &= \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}I\big(X_k;Y_1^n \mid X_1^{k-1}\big)\\
&\geq \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}I\big(X_k;Y_1^k \mid X_1^{k-1}\big)\\
&\geq \varliminf_{k\to\infty} I\big(X_k;Y_1^k \mid X_1^{k-1}\big). \qquad (4.62)
\end{aligned}$$

Let $\{W_k,\,k\in\mathbb{Z}\}$ be a sequence of IID, zero-mean, unit-variance, circularly-symmetric, complex Gaussian random variables, drawn independently of $\{(X_k,H_k,Z_k),\,k\in\mathbb{Z}\}$. Using that for our choice of input distribution $|X_k|\geq\alpha\mathsf{A}$, we obtain

$$\begin{aligned}
&I\big(X_k;Y_1^k \mid X_1^{k-1}\big)\\
&= I\left(X_k;Y_k,\left\{\frac{Y_\ell}{X_\ell}-d\right\}_{\ell=1}^{k-1}\,\middle|\, X_1^{k-1}\right)\\
&= I\left(X_k;Y_k,\left\{H_\ell+\frac{Z_\ell}{X_\ell}\right\}_{\ell=1}^{k-1}\,\middle|\, X_1^{k-1}\right)\\
&= I\left(X_k;Y_k,\left\{H_\ell+\frac{Z_\ell}{|X_\ell|}\right\}_{\ell=1}^{k-1}\,\middle|\, X_1^{k-1}\right)\\
&= I\left(X_k;Y_k,\left\{H_\ell+\frac{Z_\ell}{|X_\ell|}+\sigma\frac{\sqrt{|X_\ell|^2-(\alpha\mathsf{A})^2}}{|X_\ell|\,\alpha\mathsf{A}}W_\ell\right\}_{\ell=1}^{k-1}, W_1^{k-1}\,\middle|\, X_1^{k-1}\right)\\
&\geq I\left(X_k;Y_k,\left\{H_\ell+\frac{Z_\ell}{|X_\ell|}+\sigma\frac{\sqrt{|X_\ell|^2-(\alpha\mathsf{A})^2}}{|X_\ell|\,\alpha\mathsf{A}}W_\ell\right\}_{\ell=1}^{k-1}\,\middle|\, X_1^{k-1}\right)\\
&= I\left(X_k;Y_k,\left\{H_\ell+\frac{\sigma}{\alpha\mathsf{A}}W_\ell\right\}_{\ell=1}^{k-1}\,\middle|\, X_1^{k-1}\right)\\
&= I\left(X_k;Y_k\,\middle|\,\left\{H_\ell+\frac{\sigma}{\alpha\mathsf{A}}W_\ell\right\}_{\ell=1}^{k-1}\right), \qquad (4.63)
\end{aligned}$$

where the third step follows because Z_ℓ is circularly-symmetric, which implies that, conditional on X_ℓ, the random variable $Z_\ell/|X_\ell|$ has the same law as Z_ℓ/X_ℓ; the fifth step follows because reducing the number of observables cannot increase mutual information; the sixth step follows because $\{Z_k,\, k \in \mathbb{Z}\}$ and $\{W_k,\, k \in \mathbb{Z}\}$ are both IID and Gaussian, and because the sum of two Gaussian random variables is itself a Gaussian random variable; and the last step follows because $\{X_k,\, k \in \mathbb{Z}\}$ is IID and drawn independently of $\{(H_k, W_k),\, k \in \mathbb{Z}\}$.

We shall express the present fading as

$$H_k = \overline{H}_k + \widetilde{H}_k, \tag{4.64}$$

where

$$\overline{H}_k = \mathsf{E}\left[H_k \,\middle|\, \left\{H_\ell + \frac{\sigma}{\alpha\mathsf{A}} W_\ell\right\}_{\ell=1}^{k-1}\right]. \tag{4.65}$$

We have

$$\begin{aligned}
&I\left(X_k; Y_k \,\middle|\, \left\{H_\ell + \frac{\sigma}{\alpha\mathsf{A}} W_\ell\right\}_{\ell=1}^{k-1}\right) \\
&= h\left(H_k X_k + Z_k \,\middle|\, \left\{H_\ell + \frac{\sigma}{\alpha\mathsf{A}} W_\ell\right\}_{\ell=1}^{k-1}\right) \\
&\qquad - h\left((\overline{H}_k + \widetilde{H}_k) X_k + Z_k \,\middle|\, X_k, \left\{H_\ell + \frac{\sigma}{\alpha\mathsf{A}} W_\ell\right\}_{\ell=1}^{k-1}\right) \\
&= h\left(H_k X_k + Z_k \,\middle|\, \left\{H_\ell + \frac{\sigma}{\alpha\mathsf{A}} W_\ell\right\}_{\ell=1}^{k-1}\right) \\
&\qquad - h\left(\widetilde{H}_k X_k + Z_k \,\middle|\, X_k, \left\{H_\ell + \frac{\sigma}{\alpha\mathsf{A}} W_\ell\right\}_{\ell=1}^{k-1}\right) \\
&\geq h\left(H_k X_k + Z_k \,\middle|\, \left\{H_\ell + \frac{\sigma}{\alpha\mathsf{A}} W_\ell\right\}_{\ell=1}^{k-1}\right) - h\big(\widetilde{H}_k X_k + Z_k \bigm| X_k\big),
\end{aligned} \tag{4.66}$$

where the second step follows because differential entropy is invariant under deterministic translation; and the last step follows because conditioning cannot increase entropy.

The second entropy on the RHS of (4.66) can be evaluated using the scaling property of differential entropy

$$\begin{aligned}
&h\big(\widetilde{H}_k X_k + Z_k \bigm| X_k\big) \\
&\quad = \mathsf{E}\big[\log |X_k|^2\big] + h\left(\widetilde{H}_k + \frac{Z_k}{X_k} \,\middle|\, X_k\right)
\end{aligned}$$

$$
\begin{aligned}
&= \mathsf{E}\big[\log|X_k|^2\big] + \log(\pi e) + \mathsf{E}\left[\log\left(\epsilon_k^2(\xi) + \frac{\sigma^2}{|X_k|^2}\right)\right] \\
&\leq \mathsf{E}\big[\log|X_k|^2\big] + \log(\pi e) + \log\left(\epsilon_k^2(\xi) + \xi\right) \\
&= \log\big(\alpha \mathsf{A}^2\big) + \log(\pi e) + \log\left(\epsilon_k^2(\xi) + \xi\right), \quad \xi = \sigma^2/(\alpha\mathsf{A})^2, \quad (4.67)
\end{aligned}
$$

where $\epsilon_k^2(\xi)$ denotes the mean-square error in predicting H_k from $H_{k-1} + \sqrt{\xi}\,W_{k-1}, \ldots, H_1 + \sqrt{\xi}\,W_1$ (4.22). Here the second step follows by noting that, conditional on X_k, the random variable $\widetilde{H}_k X_k + Z_k$ is Gaussian, and by evaluating the entropy of a Gaussian random variable; the third step follows by lower bounding $|X_k|$ by $\alpha\mathsf{A}$; and the last step follows by computing the expected value of a random variable that is uniformly distributed on $[2\,\log(\alpha\mathsf{A}), 2\,\log\mathsf{A}]$.

The first entropy on the RHS of (4.66) can be lower bounded as

$$
\begin{aligned}
h\bigg(H_k X_k + Z_k \,\bigg|\, &\Big\{H_\ell + \frac{\sigma}{\alpha\mathsf{A}} W_\ell\Big\}_{\ell=1}^{k-1}\bigg) \\
&\geq h\bigg(H_k X_k + Z_k \,\bigg|\, H_k, \Big\{H_\ell + \frac{\sigma}{\alpha\mathsf{A}} W_\ell\Big\}_{\ell=1}^{k-1}\bigg) \\
&= h\big(H_k X_k + Z_k \mid H_k\big), \qquad (4.68)
\end{aligned}
$$

where the first step follows because conditioning cannot increase entropy; and where the second step follows because

$$
\Big\{H_\ell + \frac{\sigma}{\alpha\mathsf{A}} W_\ell\Big\}_{\ell=1}^{k-1} \multimap H_k \multimap H_k X_k + Z_k
$$

forms a Markov chain. We lower bound the RHS of (4.68) by applying, for each given $H_k = h_k$, the entropy power inequality [5, Thm. 16.6.3], and by averaging then over H_k

$$
\begin{aligned}
h\big(H_k X_k + Z_k \mid H_k\big) &\geq \mathsf{E}_{H_k}\left[\log\left(e^{h(H_k X_k | H_k = h_k)} + e^{h(Z_k)}\right)\right] \\
&= \mathsf{E}\left[\log\left(e^{\log|H_k|^2 + h(X_k)} + e^{h(Z_k)}\right)\right] \\
&= \mathsf{E}\left[\log\left(e^{\log|H_k|^2 + h(X_k)} + \pi e \sigma^2\right)\right], \quad (4.69)
\end{aligned}
$$

where E_{H_k} denotes expectation with respect to H_k. Here the second step follows from the scaling property of differential entropy and because X_k is independent of H_k; and the last step follows by evaluating the entropy of the Gaussian random variable Z_k.

Evaluating the differential entropy $h(X_k)$ for our choice of input distribution yields

$$\begin{aligned} h(X_k) &= h\big(\log |X_k|^2\big) + \mathsf{E}\big[\log |X_k|^2\big] + \log \pi \\ &= \log\log \frac{1}{\alpha^2} + \mathsf{E}\big[\log |X_k|^2\big] + \log \pi \\ &= \log\log \frac{1}{\alpha^2} + \log\big(\alpha \mathsf{A}^2\big) + \log \pi \\ &= \log\Big(\log\big(1/\alpha^2\big)\, \alpha \mathsf{A}^2 \pi \Big), \end{aligned} \tag{4.70}$$

where the first step follows from the circular symmetry of X_k [28, Lemmas 6.15 & 6.16]; and the second step follows by evaluating the expected value of a random variable that is uniformly distributed over $[2\,\log(\alpha\mathsf{A}), 2\,\log\mathsf{A}]$.

By applying (4.70) to (4.69), we obtain

$$\begin{aligned} &\mathsf{E}\left[\log\left(e^{\log |H_k|^2 + h(X_k)} + \pi e \sigma^2\right)\right] \\ &\quad = \mathsf{E}\left[\log\left(e^{\log |H_k|^2 + \log\left(\log(1/\alpha^2)\,\alpha \mathsf{A}^2 \pi\right)} + \pi e \sigma^2\right)\right] \\ &\quad = \mathsf{E}\left[\log\left(|H_k|^2 \log\big(1/\alpha^2\big)\, \alpha \mathsf{A}^2 \pi + \pi e \sigma^2\right)\right] \\ &\quad = \log\Big(\log\big(1/\alpha^2\big)\, \alpha \mathsf{A}^2 \pi\Big) + \mathsf{E}\left[\log\left(|H_k|^2 + \frac{e\sigma^2}{\log\big(1/\alpha^2\big)\, \alpha \mathsf{A}^2}\right)\right]. \end{aligned} \tag{4.71}$$

By noting that $|H_k|^2$ is stochastically larger than $|H_k - d|^2$ [28, Lemma 6.2b)], and by noting that, for every $\alpha > 0$, the function $f(x) = \log(\alpha + x)$, $x > 0$ is monotonically increasing, the second term on the RHS of (4.71) can be lower bounded by

$$\begin{aligned} &\mathsf{E}\left[\log\left(|H_k|^2 + \frac{e\sigma^2}{\log\big(\frac{1}{\alpha^2}\big)\, \alpha \mathsf{A}^2}\right)\right] \\ &\qquad\qquad \geq \mathsf{E}\left[\log\left(|H_k - d|^2 + \frac{e\sigma^2}{\log\big(\frac{1}{\alpha^2}\big)\, \alpha \mathsf{A}^2}\right)\right]; \end{aligned} \tag{4.72}$$

see [28, Sec. VI-B] on stochastic ordering.

We note that $|H_k - d|^2$ has an exponential distribution of mean 1, so the expectation on the RHS of (4.72) can be evaluated as [16, p. 568, Sec. 4.337]

$$\mathsf{E}\left[\log\left(|H_k - d|^2 + \frac{e\sigma^2}{\log\left(1/\alpha^2\right)\alpha\mathsf{A}^2}\right)\right] = \log\frac{e\sigma^2}{\log\left(1/\alpha^2\right)\alpha\mathsf{A}^2} - \exp\left(\frac{e\sigma^2}{\log\left(1/\alpha^2\right)\alpha\mathsf{A}^2}\right)\mathrm{Ei}\left(-\frac{e\sigma^2}{\log\left(1/\alpha^2\right)\alpha\mathsf{A}^2}\right). \quad (4.73)$$

Combining (4.68)–(4.73) yields

$$\begin{aligned} h&\left(H_kX_k + Z_k \,\middle|\, \left\{H_\ell + \frac{\sigma}{\alpha\mathsf{A}^2}W_\ell\right\}_{\ell=1}^{k-1}\right) \\ &\geq \log\left(\log\left(1/\alpha^2\right)\alpha\mathsf{A}^2\pi\right) + \log\frac{e\sigma^2}{\log\left(1/\alpha^2\right)\alpha\mathsf{A}^2} \\ &\quad - \exp\left(\frac{e\sigma^2}{\log\left(1/\alpha^2\right)\alpha\mathsf{A}^2}\right)\mathrm{Ei}\left(-\frac{e\sigma^2}{\log\left(1/\alpha^2\right)\alpha\mathsf{A}^2}\right) \\ &= \log\left(\pi e\sigma^2\right) - \exp\left(\frac{e\sigma^2}{\log\left(1/\alpha^2\right)\alpha\mathsf{A}^2}\right)\mathrm{Ei}\left(-\frac{e\sigma^2}{\log\left(1/\alpha^2\right)\alpha\mathsf{A}^2}\right), \end{aligned} \quad (4.74)$$

which in turn, along with (4.66) and (4.67), yields

$$\begin{aligned} I&\left(X_k; Y_k \,\middle|\, \left\{H_\ell + \frac{\sigma}{\alpha\mathsf{A}}W_\ell\right\}_{\ell=1}^{k-1}\right) \\ &\geq \log\left(\pi e\sigma^2\right) - \exp\left(\frac{e\sigma^2}{\log\left(1/\alpha^2\right)\alpha\mathsf{A}^2}\right)\mathrm{Ei}\left(-\frac{e\sigma^2}{\log\left(1/\alpha^2\right)\alpha\mathsf{A}^2}\right) \\ &\quad - \log\left(\alpha\mathsf{A}^2\right) - \log(\pi e) - \log\left(\epsilon_k^2(\xi) + \xi\right) \\ &= \log\frac{1}{\epsilon_k^2(\xi) + \xi} + \log\frac{1}{\alpha} + \log\frac{1}{\mathrm{SNR}} \\ &\quad - \exp\left(\frac{e}{\log\left(1/\alpha^2\right)\alpha\mathrm{SNR}}\right)\mathrm{Ei}\left(-\frac{e}{\log\left(1/\alpha^2\right)\alpha\mathrm{SNR}}\right), \end{aligned} \quad (4.75)$$

where $\xi = \frac{1}{\alpha^2\,\mathrm{SNR}}$.

We finally obtain from (4.4), (4.62), and (4.75)

$$
\begin{aligned}
& C_{\mathrm{PP}}(\mathrm{SNR}) \\
&\quad \geq \lim_{n\to\infty} \frac{1}{n} I\big(X_1^n; Y_1^n\big) \\
&\quad \geq \varliminf_{k\to\infty} I\left(X_k; Y_k \,\middle|\, \left\{H_\ell + \frac{\sigma}{\alpha \mathsf{A}} W_\ell\right\}_{\ell=1}^{k-1}\right) \\
&\quad \geq \varliminf_{k\to\infty} \log \frac{1}{\epsilon_k^2(\xi)+\xi} + \log\frac{1}{\alpha} + \log\frac{1}{\mathrm{SNR}} \\
&\qquad\quad - \exp\left(\frac{e}{\log\big(1/\alpha^2\big)\,\alpha\mathrm{SNR}}\right) \mathrm{Ei}\left(-\frac{e}{\log\big(1/\alpha^2\big)\,\alpha\mathrm{SNR}}\right) \\
&\quad = \log \frac{1}{\epsilon^2(\xi)+\xi} + \log\frac{1}{\alpha} + \log\frac{1}{\mathrm{SNR}} \\
&\qquad\quad - \exp\left(\frac{e}{\log\big(1/\alpha^2\big)\,\alpha\mathrm{SNR}}\right) \mathrm{Ei}\left(-\frac{e}{\log\big(1/\alpha^2\big)\,\alpha\mathrm{SNR}}\right),
\end{aligned} \tag{4.76}
$$

where $\xi = \frac{1}{\alpha^2\,\mathrm{SNR}}$. Here the last step follows by extending (4.17) to the noisy prediction problem.

Proposition 4.1 follows now by maximizing the RHS of (4.76) over α.

4.8.2 Proof of Proposition 4.5

In order to derive a lower bound on the fading number, we consider channel inputs of the form

$$
\mathbf{X}_k = \mathbf{V}_k\, R_k, \quad k \in \mathbb{Z},
$$

where $\{\mathbf{V}_k,\, k \in \mathbb{Z}\}$ is stationary ergodic and satisfies (4.41) and (4.42); where $\{R_k,\, k\in\mathbb{Z}\}$ is a sequence of IID, zero-mean, circularly-symmetric, complex random variables with $\log|R_k|^2$ being uniformly distributed on $\big[\log r_{\min}^2, \log\big(\mathsf{P}/\mathsf{E}\left[\|\mathbf{V}_1\|^2\right]\big)\big]$, i.e,

$$
\log|R_k|^2 \sim \mathcal{U}\left(\left[\log r_{\min}^2, \log\frac{\mathsf{P}}{\mathsf{E}[\|\mathbf{V}_1\|^2]}\right]\right), \quad k\in\mathbb{Z} \tag{4.77}
$$

for some positive and real $r_{\min}$; and where $\{\mathbf{V}_k,\, k \in \mathbb{Z}\}$ and $\{R_k,\, k\in\mathbb{Z}\}$ are independent of each other. Note that with this choice

$\{\mathbf{X}_k,\, k \in \mathbb{Z}\}$ satisfies the average-power constraint (4.5). If (4.44) holds, then $\{\mathbf{X}_k,\, k \in \mathbb{Z}\}$ can be chosen such that the peak-power constraint is satisfied by choosing $\log|R_k|^2$ to be uniformly distributed on $\bigl[\log r_{\min}^2, 2\,\log\bigl(\mathsf{A}/\Upsilon\bigr)\bigr]$, i.e,

$$\log|R_k|^2 \sim \mathcal{U}\left(\left[\log r_{\min}^2, \log\frac{\mathsf{A}^2}{\Upsilon^2}\right]\right), \quad k \in \mathbb{Z}.$$

The rest of the proof does not depend on whether an average-power or a peak-power constraint is imposed on the channels inputs.

Recall that the fading number is defined as (4.11)

$$\chi(\{\mathbb{H}_k\}) = \varlimsup_{\mathrm{SNR}\to\infty}\bigl\{C(\mathrm{SNR}) - \log\log\mathrm{SNR}\bigr\}. \tag{4.78}$$

To lower bound the fading number, we derive a lower bound on $C(\mathrm{SNR})$ by evaluating $\frac{1}{n}I(\mathbf{X}_1^n;\mathbf{Y}_1^n)$ for $\{\mathbf{X}_k,\, k \in \mathbb{Z}\}$ being distributed as described above. We have

$$\begin{aligned}\frac{1}{n}I\bigl(\mathbf{X}_1^n;\mathbf{Y}_1^n\bigr) &\geq \frac{1}{n}I\bigl(R_1^n, \mathbf{V}_1^n;\mathbf{Y}_1^n\bigr)\\ &= \frac{1}{n}I\bigl(R_1^n;\mathbf{Y}_1^n\bigr) + \frac{1}{n}I\bigl(\mathbf{V}_1^n;\mathbf{Y}_1^n \bigm| R_1^n\bigr)\\ &= \frac{1}{n}I\bigl(R_1^n;\{\mathbb{H}_\ell\mathbf{V}_\ell\,R_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n\bigr)\\ &\quad + \frac{1}{n}I\bigl(\mathbf{V}_1^n;\{\mathbb{H}_\ell\mathbf{V}_\ell\,R_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \bigm| R_1^n\bigr),\end{aligned} \tag{4.79}$$

where the first step follows from the data processing inequality [5, Thm. 2.8.1]; and where the second step follows from the chain rule for mutual information.

We continue by lower bounding the second term on the RHS of (4.79)

$$\begin{aligned}&\frac{1}{n}I\bigl(\mathbf{V}_1^n;\{\mathbb{H}_\ell\mathbf{V}_\ell\,R_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \bigm| R_1^n\bigr)\\ &\quad = \frac{1}{n}I\left(\mathbf{V}_1^n;\left\{\mathbb{H}_\ell\mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{R_\ell}\right\}_{\ell=1}^n \,\middle|\, R_1^n\right)\\ &\quad = \frac{1}{n}I\left(\mathbf{V}_1^n;\left\{\mathbb{H}_\ell\mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{|R_\ell|}\right\}_{\ell=1}^n \,\middle|\, R_1^n\right)\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n} I\left(\mathbf{V}_1^n; \left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{|R_\ell|} + \sigma \frac{\sqrt{|R_\ell|^2 - r_{\min}^2}}{|R_\ell|\, r_{\min}} \mathbf{W}_\ell\right\}_{\ell=1}^n, \mathbf{W}_1^n \,\middle|\, R_1^n\right) \\
&\geq \frac{1}{n} I\left(\mathbf{V}_1^n; \left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{|R_\ell|} + \sigma \frac{\sqrt{|R_\ell|^2 - r_{\min}^2}}{|R_\ell|\, r_{\min}} \mathbf{W}_\ell\right\}_{\ell=1}^n \,\middle|\, R_1^n\right) \\
&= \frac{1}{n} I\left(\mathbf{V}_1^n; \left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^n\right),
\end{aligned}
\tag{4.80}
$$

where $\{\mathbf{W}_k,\, k \in \mathbb{Z}\}$ is a sequence of IID, zero-mean, unit-variance, circularly-symmetric, complex Gaussian n_R-variate vectors, drawn independently of $\{(\mathbf{V}_k, \mathbb{H}_k, \mathbf{Z}_k),\, k \in \mathbb{Z}\}$. Here the first step follows by dividing $\mathbf{Y}_k$ by R_k; the second step follows because $\{\mathbf{Z}_k,\, k \in \mathbb{Z}\}$ is IID and circularly-symmetric, which implies that, conditional on R_1^n, the sequence $\mathbf{Z}_1/R_1, \ldots, \mathbf{Z}_n/R_n$ has the same joint law as $\mathbf{Z}_1/|R_1|, \ldots, \mathbf{Z}_n/|R_n|$; the third step follows by adding the noise $\mathbf{W}_\ell$, which is known and does therefore not change the mutual information; the fourth step follows because reducing observations cannot increase mutual information; and the last step follows because $\{\mathbf{Z}_k,\, k \in \mathbb{Z}\}$ and $\{\mathbf{W}_k,\, k \in \mathbb{Z}\}$ are both IID Gaussian, and because the sum of two Gaussian random vectors is again a Gaussian random vector.

We fix some positive integer κ. The chain rule for mutual information yields

$$
\begin{aligned}
&\frac{1}{n} I\left(\mathbf{V}_1^n; \left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^n\right) \\
&\quad = \frac{1}{n} \sum_{k=1}^n I\left(\mathbf{V}_k; \left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^n \,\middle|\, \mathbf{V}_1^{k-1}\right) \\
&\quad = \frac{1}{n} \sum_{k=1}^n I\left(\mathbf{V}_k; \left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^n, \mathbf{V}_1^{k-1}\right) - \frac{1}{n} \sum_{k=1}^n I\bigl(\mathbf{V}_k; \mathbf{V}_1^{k-1}\bigr) \\
&\quad \geq \frac{1}{n} \sum_{k=\kappa+1}^{n-\kappa} I\left(\mathbf{V}_k; \left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^n, \mathbf{V}_1^{k-1}\right) - \frac{1}{n} \sum_{k=1}^n I\bigl(\mathbf{V}_k; \mathbf{V}_1^{k-1}\bigr) \\
&\quad \geq \frac{1}{n} \sum_{k=\kappa+1}^{n-\kappa} I\left(\mathbf{V}_k; \left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=k-\kappa}^{k+\kappa}, \mathbf{V}_{k-\kappa}^{k-1}\right)
\end{aligned}
$$

$$
\begin{aligned}
&\quad - \frac{1}{n}\sum_{k=1}^{n} I\big(\mathbf{V}_k; \mathbf{V}_1^{k-1}\big) \\
&= \frac{1}{n}\sum_{k=\kappa+1}^{n-\kappa} \left(I\left(\mathbf{V}_k; \left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=k-\kappa}^{k+\kappa} \,\middle|\, \mathbf{V}_{k-\kappa}^{k-1}\right) + I\big(\mathbf{V}_k; \mathbf{V}_{k-\kappa}^{k-1}\big)\right) \\
&\quad - \frac{1}{n}\sum_{k=1}^{n} I\big(\mathbf{V}_k; \mathbf{V}_1^{k-1}\big) \\
&= \left(1 - \frac{2\kappa}{n}\right) I\left(\mathbf{V}_{\kappa+1}; \left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa}\right) \\
&\quad + \left(1 - \frac{2\kappa}{n}\right) I\big(\mathbf{V}_{\kappa+1}; \mathbf{V}_1^{\kappa}\big) - \frac{1}{n}\sum_{k=1}^{n} I\big(\mathbf{V}_k; \mathbf{V}_1^{k-1}\big),
\end{aligned} \tag{4.81}
$$

where the second step follows from the chain rule for mutual information; the third step follows from the nonnegativity of mutual information; the fourth step follows by reducing the number of observables; the fifth step follows again from the chain rule of mutual information; and the last step follows from the stationarity of the channel.

Writing mutual information in terms of differential entropies, we obtain for the first term on the RHS of (4.81)

$$
\begin{aligned}
&\left(1 - \frac{2\kappa}{n}\right) I\left(\mathbf{V}_{\kappa+1}; \left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa}\right) \\
&\quad = \left(1 - \frac{2\kappa}{n}\right)\left[h\left(\left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa}\right)\right. \\
&\qquad\qquad \left. - h\left(\left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa+1}\right)\right] \\
&\quad \geq \left(1 - \frac{2\kappa}{n}\right)\left[h\left(\left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa}, \mathbf{Z}_1^{2\kappa+1}\right)\right. \\
&\qquad\qquad \left. - h\left(\left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa+1}\right)\right] \\
&\quad = \left(1 - \frac{2\kappa}{n}\right)\left[h\left(\left\{\mathbb{H}_\ell \mathbf{V}_\ell\right\}_{\ell=1}^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa}\right) - h\left(\left\{\mathbb{H}_\ell \mathbf{V}_\ell\right\}_{\ell=1}^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa+1}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&- \left(1 - \frac{2\kappa}{n}\right) \left[h\left(\left\{ \mathbf{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}} \right\}_{\ell=1}^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa+1} \right) \right. \\
&\qquad\qquad\qquad\qquad \left. - h\left(\left\{ \mathbb{H}_\ell \mathbf{V}_\ell \right\}_{\ell=1}^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa+1} \right) \right] \\
&= \left(1 - \frac{2\kappa}{n}\right) I\left(\mathbf{V}_{\kappa+1}; \left\{ \mathbb{H}_\ell \mathbf{V}_\ell \right\}_{\ell=1}^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa} \right) \\
&\quad - \left(1 - \frac{2\kappa}{n}\right) I\left(\left\{ \mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}} \right\}_{\ell=1}^{2\kappa+1}; \mathbf{Z}_1^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa+1} \right). \qquad (4.82)
\end{aligned}
$$

Here the second step follows because conditioning cannot increase differential entropy; and the third step follows from the invariance of differential entropy under deterministic translation.

We continue by lower bounding the first mutual information on the RHS of (4.82). We have

$$
\begin{aligned}
&I\left(\mathbf{V}_{\kappa+1}; \left\{ \mathbb{H}_\ell \mathbf{V}_\ell \right\}_{\ell=1}^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa} \right) \\
&= I\left(\mathbf{V}_{\kappa+1}; \left\{ \mathbb{H}_\ell \mathbf{V}_\ell \right\}_{\ell=1}^{2\kappa+1}, \mathbf{V}_1^{\kappa} \right) - I\big(\mathbf{V}_{\kappa+1}; \mathbf{V}_1^{\kappa} \big) \\
&= \frac{1}{\kappa} \sum_{k=1}^{\kappa} I\left(\mathbf{V}_{\kappa+1}; \left\{ \mathbb{H}_\ell \mathbf{V}_\ell \right\}_{\ell=1}^{2\kappa+1}, \mathbf{V}_1^{\kappa} \right) - I\big(\mathbf{V}_{\kappa+1}; \mathbf{V}_1^{\kappa} \big) \\
&= \frac{1}{\kappa} \sum_{k=1}^{\kappa} I\left(\mathbf{V}_{k}; \left\{ \mathbb{H}_\ell \mathbf{V}_\ell \right\}_{\ell=k-\kappa}^{k+\kappa}, \mathbf{V}_{k-\kappa}^{k-1} \right) - I\big(\mathbf{V}_{\kappa+1}; \mathbf{V}_1^{\kappa} \big) \\
&\geq \frac{1}{\kappa} \sum_{k=1}^{\kappa} I\big(\mathbf{V}_{k}; \{ \mathbb{H}_\ell \mathbf{V}_\ell \}_{\ell=1}^{\kappa}, \mathbf{V}_1^{k-1} \big) - I\big(\mathbf{V}_{\kappa+1}; \mathbf{V}_1^{\kappa} \big) \\
&= \frac{1}{\kappa} \sum_{k=1}^{\kappa} \Big(I\big(\mathbf{V}_{k}; \{ \mathbb{H}_\ell \mathbf{V}_\ell \}_{\ell=1}^{\kappa} \mid \mathbf{V}_1^{k-1} \big) + I\big(\mathbf{V}_{k}; \mathbf{V}_1^{k-1} \big) \Big) - I\big(\mathbf{V}_{\kappa+1}; \mathbf{V}_1^{\kappa} \big) \\
&= \frac{1}{\kappa} I\big(\mathbf{V}_1^{\kappa}; \{ \mathbb{H}_\ell \mathbf{V}_\ell \}_{\ell=1}^{\kappa} \big) + \frac{1}{\kappa} \sum_{k=1}^{\kappa} I\big(\mathbf{V}_{k}; \mathbf{V}_1^{k-1} \big) - I\big(\mathbf{V}_{\kappa+1}; \mathbf{V}_1^{\kappa} \big), \quad (4.83)
\end{aligned}
$$

where the first step follows from the chain rule for mutual information; the third step follows from the stationarity of the channel; the fourth step follows by reducing the number of observables; and the fifth and the sixth step follow again from the chain rule for mutual information.

We next show that the second term on the RHS of (4.82) tends to zero as $r_{\min}$ tends to infinity, i.e.,

$$\lim_{r_{\min}\to\infty} I\left(\left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^{2\kappa+1}; \mathbf{Z}_1^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa+1}\right) = 0. \tag{4.84}$$

To this end, we first note that

$$I\left(\left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^{2\kappa+1}; \mathbf{Z}_1^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa+1}\right) \geq 0. \tag{4.85}$$

We further have

$$\begin{aligned}
I&\left(\left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^{2\kappa+1}; \mathbf{Z}_1^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa+1}\right) \\
&= I\left(\left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^{2\kappa+1}, \mathbf{V}_1^{k+1}; \mathbf{Z}_1^{2\kappa+1}\right) \\
&\leq I\left(\left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^{2\kappa+1}, \mathbf{V}_1^{2\kappa+1}; \mathbf{Z}_1^{2\kappa+1}\right) \\
&= I\left(\left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^{2\kappa+1}; \mathbf{Z}_1^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{2\kappa+1}\right) \\
&= I\left(\left\{\mathbb{H}_\ell \hat{\mathbf{V}}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}\,\|\mathbf{V}_\ell\|}\right\}_{\ell=1}^{2\kappa+1}; \mathbf{Z}_1^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{2\kappa+1}\right),
\end{aligned} \tag{4.86}$$

where $\hat{\mathbf{V}}_\ell = \mathbf{V}_\ell/\|\mathbf{V}_\ell\|$. Here the first step follows because $\mathbf{Z}_1^{2\kappa+1}$ and $\mathbf{V}_1^{\kappa+1}$ are independent; the second step follows by adding observables; the third step follows because $\mathbf{Z}_1^{2\kappa+1}$ and $\mathbf{V}_1^{2\kappa+1}$ are independent; and the last step follows by dividing $\mathbb{H}_\ell \mathbf{V}_\ell + \mathbf{Z}_\ell/r_{\min}$ by $\|\mathbf{V}_\ell\|$. (If we evaluate the mutual information in the third step of (4.86) for a particular $\mathbf{v}_1^{2\kappa+1}$, and if $\mathbf{v}_\ell = 0$ for at least one $\ell = 1, \ldots, 2\kappa+1$, then this mutual information is infinite. However, by (4.42), this event has probability zero and does therefore not influence the mutual information when averaged over $\mathbf{V}_1^{2\kappa+1}$. Without loss of generality, we can thus assume that $\mathbf{v}_\ell \neq 0$, $\ell = 1, \ldots, 2\kappa+1$, which also ensures that we do not divide by zero in the last step of (4.86).)

To show that the RHS of (4.86) tends to zero as SNR tends to infinity, we need to following lemma.

Lemma 4.17. *Let* $\mathbf{T}$ *be a random* n_{R}*-dimensional vector that satisfies*

$$\mathsf{E}\left[\|\mathbf{T}\|^2\right] < \infty \qquad \textit{and} \qquad h(\mathbf{T}) > -\infty. \tag{4.87}$$

Let $\mathbf{Z}$ *be a zero-mean, circularly-symmetric, complex Gaussian vector of dimension* n_{R} *and of covariance matrix* $\mathsf{I}_{n_{\mathrm{R}}}$*, drawn independently of* $\mathbf{T}$*. Then*

$$\lim_{\sigma \downarrow 0} I\big(\mathbf{T} + \sigma\,\mathbf{Z}; \mathbf{Z}\big) = 0. \tag{4.88}$$

Proof. See [28, Lemma 6.11a)]. □

We apply Lemma 4.17 to show that for a given $\mathbf{v}_1^{2\kappa+1}$ ($\mathbf{v}_\ell \neq 0$, $\ell = 1, \ldots, 2\kappa+1$)

$$\lim_{r_{\min}\to\infty} I\left(\left\{\mathbb{H}_\ell \hat{\mathbf{V}}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}\,\|\mathbf{V}_\ell\|}\right\}_{\ell=1}^{2\kappa+1}; \mathbf{Z}_1^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{2\kappa+1} = \mathbf{v}_1^{2\kappa+1}\right) = 0. \tag{4.89}$$

To this end, we have to demonstrate that the vector

$$\mathbf{T} = \begin{pmatrix} \mathbb{H}_1 \hat{\mathbf{v}}_1 \\ \vdots \\ \mathbb{H}_{2\kappa+1} \hat{\mathbf{v}}_{2\kappa+1} \end{pmatrix}$$

satisfies (4.87). Its second moment is upper bounded by

$$\mathsf{E}\left[\|\mathbf{T}\|^2\right] = \sum_{\ell=1}^{2\kappa+1} \mathsf{E}\left[\|\mathbb{H}_\ell \hat{\mathbf{v}}_\ell\|^2\right] \leq \sum_{\ell=1}^{2\kappa+1} \mathsf{E}\left[\|\mathbb{H}_\ell\|_{\mathrm{F}}^2\right] \underbrace{\|\hat{\mathbf{v}}_\ell\|^2}_{=1} < \infty, \tag{4.90}$$

where the second step follows from the Cauchy-Schwarz inequality; and the last step follows from (4.2). To show that $\mathbf{T}$ is of finite differential entropy, we use the chain rule for mutual information

$$\begin{aligned} h(\mathbf{T}) &= h(\mathbb{H}_1\hat{\mathbf{v}}_1, \ldots, \mathbb{H}_{2\kappa+1}\hat{\mathbf{v}}_{2\kappa+1}) \\ &= \sum_{k=1}^{2\kappa+1} h\big(\mathbb{H}_k\hat{\mathbf{v}}_k \bigm| \{\mathbb{H}_\ell\hat{\mathbf{v}}_\ell\}_{\ell=1}^{k-1}\big) \\ &\geq \sum_{k=1}^{2\kappa+1} h\big(\mathbb{H}_k\hat{\mathbf{v}}_k \bigm| \{\mathbb{H}_\ell\hat{\mathbf{v}}_\ell\}_{\ell=1}^{k-1}, \mathbb{H}_1^{k-1}\big) \\ &= \sum_{k=1}^{2\kappa+1} h\big(\mathbb{H}_k\hat{\mathbf{v}}_k \bigm| \mathbb{H}_1^{k-1}\big), \end{aligned} \tag{4.91}$$

where the third step follows because conditioning cannot increase entropy; and the last step follows because

$$\{\mathbb{H}_\ell\hat{\mathbf{v}}_\ell\}_{\ell=1}^{k-1} \multimap\!\!- \mathbb{H}_1^{k-1} \multimap\!\!- \mathbb{H}_k\hat{\mathbf{v}}_k$$

forms a Markov chain. The claim follows then by noting that

$$h\big(\mathbb{H}_k \bigm| \mathbb{H}_1^{k-1}\big) \geq h(\{\mathbb{H}_k\}) > \infty,$$

and from a conditional version of Lemma 6.6 in [28]. Here the first inequality follows from [5, Thms. 4.2.1 & 4.2.2], and the second inequality follows because, by the propositions assumption, the fading is regular.

In order to show that (4.89) also holds when averaged over $\hat{\mathbf{V}}_1^{2\kappa+1}$, we shall use the Monotone Convergence Theorem [38, Thm. 1.26]. To this end, we show that for every $\mathbf{v}_1^{2\kappa+1}$ ($\mathbf{v}_\ell \neq 0$, $\ell = 1, \ldots, 2\kappa+1$) the mutual information

$$I\left(\left\{\mathbb{H}_\ell\hat{\mathbf{v}}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}\|\mathbf{v}_\ell\|}\right\}_{\ell=1}^{2\kappa+1}; \mathbf{Z}_1^{2\kappa+1}\right)$$

is monotonically nonincreasing in $r_{\min}$. Indeed, we have for every $r_{\min}$ and $\alpha > 0$

$$\begin{aligned} I&\left(\left\{\mathbb{H}_\ell\hat{\mathbf{v}}_\ell + \frac{\mathbf{Z}_\ell}{(r_{\min}+\alpha)\|\mathbf{v}_\ell\|}\right\}_{\ell=1}^{2\kappa+1}; \mathbf{Z}_1^{2\kappa+1}\right) \\ &= h\left(\left\{\mathbb{H}_\ell\hat{\mathbf{v}}_\ell + \frac{\mathbf{Z}_\ell}{(r_{\min}+\alpha)\|\mathbf{v}_\ell\|}\right\}_{\ell=1}^{2\kappa+1}\right) - h\Big(\big\{\mathbb{H}_\ell\hat{\mathbf{v}}_\ell\big\}_{\ell=1}^{2\kappa+1}\Big) \end{aligned}$$

$$
\begin{aligned}
&= h\left(\left\{\mathbb{H}_\ell \hat{\mathbf{v}}_\ell + \frac{\mathbf{Z}_\ell}{(r_{\min}+\alpha)\,\|\mathbf{v}_\ell\|} + \varsigma\,\mathbf{W}_\ell\right\}_{\ell=1}^{2\kappa+1} \,\middle|\, \mathbf{W}_1^{2\kappa+1}\right) \\
&\quad - h\Big(\big\{\mathbb{H}_\ell \hat{\mathbf{v}}_\ell\big\}_{\ell=1}^{2\kappa+1}\Big) \\
&\le h\left(\left\{\mathbb{H}_\ell \hat{\mathbf{v}}_\ell + \frac{\mathbf{Z}_\ell}{(r_{\min}+\alpha)\,\|\mathbf{v}_\ell\|} + \varsigma\,\mathbf{W}_\ell\right\}_{\ell=1}^{2\kappa+1}\right) - h\Big(\big\{\mathbb{H}_\ell \hat{\mathbf{v}}_\ell\big\}_{\ell=1}^{2\kappa+1}\Big) \\
&= h\left(\left\{\mathbb{H}_\ell \hat{\mathbf{v}}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}\,\|\mathbf{v}_\ell\|}\right\}_{\ell=1}^{2\kappa+1}\right) - h\Big(\big\{\mathbb{H}_\ell \hat{\mathbf{v}}_\ell\big\}_{\ell=1}^{2\kappa+1}\Big) \\
&= I\left(\left\{\mathbb{H}_\ell \hat{\mathbf{v}}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}\,\|\mathbf{v}_\ell\|}\right\}_{\ell=1}^{2\kappa+1}; \mathbf{Z}_1^{2\kappa+1}\right), \qquad (4.92)
\end{aligned}
$$

where

$$
\varsigma = \sigma \frac{\sqrt{(r_{\min}+\alpha)^2 - r_{\min}^2}}{r_{\min}(r_{\min}+\alpha)\,\|\mathbf{v}_\ell\|}
$$

and where $\{\mathbf{W}_k,\ k \in \mathbb{Z}\}$ is as in (4.80). Here the second step follows because differential entropy is invariant under deterministic translation; the third step follows because conditioning cannot increase entropy; and the fourth step follows because $\{\mathbf{Z}_k,\ k \in \mathbb{Z}\}$ and $\{\mathbf{W}_k,\ k \in \mathbb{Z}\}$ are both IID Gaussian, and because the sum of two Gaussian vectors is again a Gaussian vector.

Equations (4.85)–(4.92) combine to prove (4.84). We return to the analysis of the fading number (4.78). We summarize the main steps of the proof to obtain the final lower bound

$$
\begin{aligned}
&\chi(\{\mathbb{H}_k\}) \\
&\quad \triangleq \varlimsup_{\mathrm{SNR}\to\infty} \big\{ C(\mathrm{SNR}) - \log\log \mathrm{SNR} \big\} \\
&\quad \ge \varlimsup_{\mathrm{SNR}\to\infty} \left\{ \lim_{n\to\infty} \frac{1}{n} I\big(R_1^n; \{\mathbb{H}_\ell \mathbf{V}_\ell R_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n\big) - \log\log \mathrm{SNR} \right\} \\
&\qquad + \varlimsup_{\mathrm{SNR}\to\infty} \lim_{n\to\infty} \frac{1}{n} I\big(\mathbf{V}_1^n; \mathbf{Y}_1^n \,\big|\, R_1^n\big) \\
&\quad = \chi(\{\mathbb{H}_k \mathbf{V}_k\}) + \varlimsup_{\mathrm{SNR}\to\infty} \lim_{n\to\infty} \frac{1}{n} I\big(\mathbf{V}_1^n; \mathbf{Y}_1^n \,\big|\, R_1^n\big) \\
&\quad \ge \chi(\{\mathbb{H}_k \mathbf{V}_k\}) \\
&\qquad + \lim_{n\to\infty} \left(1 - \frac{2\kappa}{n}\right) \left(I\Big(\mathbf{V}_{\kappa+1}; \big\{\mathbb{H}_\ell \mathbf{V}_\ell\big\}_{\ell=1}^{2\kappa+1} \,\Big|\, \mathbf{V}_1^{\kappa}\Big) + I\big(\mathbf{V}_{\kappa+1}; \mathbf{V}_1^{\kappa}\big) \right)
\end{aligned}
$$

$$
\begin{aligned}
&\quad - \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} I\bigl(\mathbf{V}_k; \mathbf{V}_1^{k-1}\bigr) \\
&\quad - \lim_{n\to\infty} \left(1 - \frac{2\kappa}{n}\right) I\left(\left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^{2\kappa+1}; \mathbf{Z}_1^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa+1}\right) \\
&= \chi(\{\mathbb{H}_k \mathbf{V}_k\}) + I\Bigl(\mathbf{V}_{\kappa+1}; \bigl\{\mathbb{H}_\ell \mathbf{V}_\ell\bigr\}_{\ell=1}^{2\kappa+1}\Bigr) + I\bigl(\mathbf{V}_{\kappa+1}; \mathbf{V}_1^{\kappa}\bigr) \\
&\quad - \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} I\bigl(\mathbf{V}_k; \mathbf{V}_1^{k-1}\bigr) \\
&\quad - I\left(\left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^{2\kappa+1}; \mathbf{Z}_1^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa+1}\right) \\
&\ge \chi(\{\mathbb{H}_k \mathbf{V}_k\}) + \frac{1}{\kappa} I\bigl(\mathbf{V}_1^{\kappa}; \{\mathbb{H}_\ell \mathbf{V}_\ell\}_{\ell=1}^{\kappa}\bigr) + \frac{1}{\kappa} \sum_{k=1}^{\kappa} I\bigl(\mathbf{V}_k; \mathbf{V}_1^{k-1}\bigr) \\
&\quad - I\bigl(\mathbf{V}_{\kappa+1}; \mathbf{V}_1^{\kappa}\bigr) + I\bigl(\mathbf{V}_{\kappa+1}; \mathbf{V}_1^{\kappa}\bigr) - \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} I\bigl(\mathbf{V}_k; \mathbf{V}_1^{k-1}\bigr) \\
&\quad - I\left(\left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^{2\kappa+1}; \mathbf{Z}_1^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa+1}\right) \\
&= \chi(\{\mathbb{H}_k \mathbf{V}_k\}) + \frac{1}{\kappa} I\bigl(\mathbf{V}_1^{\kappa}; \{\mathbb{H}_\ell \mathbf{V}_\ell\}_{\ell=1}^{\kappa}\bigr) \\
&\quad + \frac{1}{\kappa} \sum_{k=1}^{\kappa} I\bigl(\mathbf{V}_k; \mathbf{V}_1^{k-1}\bigr) - \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} I\bigl(\mathbf{V}_k; \mathbf{V}_1^{k-1}\bigr) \\
&\quad - I\left(\left\{\mathbb{H}_\ell \mathbf{V}_\ell + \frac{\mathbf{Z}_\ell}{r_{\min}}\right\}_{\ell=1}^{2\kappa+1}; \mathbf{Z}_1^{2\kappa+1} \,\middle|\, \mathbf{V}_1^{\kappa+1}\right),
\end{aligned} \tag{4.93}
$$

where the second step follows from (4.79); the third step by noting that the distribution of R_1^n achieves the fading number of the SIMO fading channel with fading $\{\mathbb{H}_k \mathbf{V}_k,\, k \in \mathbb{Z}\}$; the fourth step follows from (4.80)–(4.82) and by noting that the mutual informations on the RHS of (4.80)–(4.82) do not depend on the SNR; and the sixth step follows from (4.83). (It follows from (4.41) and Cesáro's mean [5, Thm. 4.2.3] that the limit of $\frac{1}{n}\sum_{k=1}^{n} I\bigl(\mathbf{V}_k; \mathbf{V}_1^{k-1}\bigr)$ exists and is finite.)

Proposition 4.5 follows now from (4.84) by letting first $r_{\min}$ tend to infinity while holding κ fixed, and by letting then κ tend to infinity.

4.8.3 Proof of Note 4.2

The fading number of SIMO fading channels with memory was computed by Lapidoth and Moser [29]. It is given by

$$\chi(\{\mathbf{H}_k\}) = h_\lambda\bigl(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Phi_0}\bigr) - h(\mathbf{H}_0) + n_\mathrm{R}\mathsf{E}\bigl[\log\|\mathbf{H}_0\|^2\bigr] - \log 2 \\ + I\bigl(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1}\bigr) - I\Bigl(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Phi_0};\bigl\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Phi_\ell}\bigr\}_{\ell=-\infty}^{-1}\Bigr), \tag{4.94}$$

where $\hat{\mathbf{H}}_\ell = \frac{\mathbf{H}_\ell}{\|\mathbf{H}_\ell\|}$; where $\{\Phi_k,\, k\in\mathbb{Z}\}$ is a sequence of IID random variables with Φ_k uniformly distributed over the interval $(-\pi,\pi]$ and drawn independently of $\{\mathbf{H}_k,\, k\in\mathbb{Z}\}$; and where $h_\lambda(\cdot)$ is defined in [28, Eq. (323)].

In the following we show that

$$\chi(\{\mathbf{H}_kV_k\}) + \lim_{n\to\infty}\frac{1}{n}I\bigl(V_1^n;\{\mathbf{H}_\ell V_\ell\}_{\ell=1}^n\bigr) \tag{4.95}$$

yields the same result. Here $\{V_k,\, k\in\mathbb{Z}\}$ is any stationary and ergodic process that is independent of $\{\mathbf{H}_k,\, k\in\mathbb{Z}\}$ and that satisfies (4.41) and (4.42).

We start by using the chain rule for mutual information to evaluate the second term in (4.95) by

$$\begin{aligned}
&\frac{1}{n}I\bigl(V_1^n;\{\mathbf{H}_\ell V_\ell\}_{\ell=1}^n\bigr)\\
&= \frac{1}{n}\sum_{k=1}^n I\bigl(V_1^n;\mathbf{H}_kV_k \bigm| \{\mathbf{H}_\ell V_\ell\}_{\ell=1}^{k-1}\bigr)\\
&= \frac{1}{n}\sum_{k=1}^n \Bigl(I\bigl(V_1^n,\{\mathbf{H}_\ell V_\ell\}_{\ell=1}^{k-1};\mathbf{H}_kV_k\bigr) - I\bigl(\mathbf{H}_kV_k;\{\mathbf{H}_\ell V_\ell\}_{\ell=1}^{k-1}\bigr)\Bigr)\\
&= \frac{1}{n}\sum_{k=1}^n \Bigl(I\bigl(V_1^k,\{\mathbf{H}_\ell V_\ell\}_{\ell=1}^{k-1};\mathbf{H}_kV_k\bigr) - I\bigl(\mathbf{H}_kV_k;\{\mathbf{H}_\ell V_\ell\}_{\ell=1}^{k-1}\bigr)\Bigr)\\
&= \frac{1}{n}\sum_{k=1}^n \Bigl(I\bigl(V_1^k,\{\mathbf{H}_\ell V_\ell\}_{\ell=1}^{k-1},\mathbf{H}_1^{k-1};\mathbf{H}_kV_k\bigr) - I\bigl(\mathbf{H}_kV_k;\{\mathbf{H}_\ell V_\ell\}_{\ell=1}^{k-1}\bigr)\Bigr)\\
&= \frac{1}{n}\sum_{k=1}^n \Bigl(I\bigl(\mathbf{H}_1^{k-1},V_k;\mathbf{H}_kV_k\bigr) - I\bigl(\mathbf{H}_kV_k;\{\mathbf{H}_\ell V_\ell\}_{\ell=1}^{k-1}\bigr)\Bigr)
\end{aligned}$$

$$= \frac{1}{n}\sum_{k=1}^{n}\Big(I\big(V_k;\mathbf{H}_kV_k\big) + I\big(\mathbf{H}_1^{k-1};\mathbf{H}_kV_k \bigm| V_k\big) - I\big(\mathbf{H}_kV_k;\{\mathbf{H}_\ell V_\ell\}_{\ell=1}^{k-1}\big)\Big)$$

$$= \frac{1}{n}\sum_{k=1}^{n}\Big(I\big(V_k;\mathbf{H}_kV_k\big) + I\big(\mathbf{H}_1^{k-1};\mathbf{H}_k\big) - I\big(\mathbf{H}_kV_k;\{\mathbf{H}_\ell V_\ell\}_{\ell=1}^{k-1}\big)\Big), \tag{4.96}$$

where the third step follows because the channel has no feedback in which case

$$V_{k+1}^n \multimap \big(\{\mathbf{H}_\ell V_\ell\}_{\ell-1}^{k-1}, V_1^k\big) \multimap \mathbf{H}_kV_k$$

forms a Markov chain; the fourth step follows by dividing $\mathbf{H}_\ell V_\ell$ by V_ℓ; the fifth step follows because

$$\big(\{\mathbf{H}_\ell V_\ell\}_{\ell=1}^{k-1}, V_1^{k-1}\big) \multimap \big(\mathbf{H}_1^{k-1}, V_k\big) \multimap \mathbf{H}_kV_k$$

forms a Markov chain; and the last step follows by dividing $\mathbf{H}_kV_k$ by V_k.

Note that $\{\mathbf{H}_k,\, k \in \mathbb{Z}\}$ (by (4.2) and (4.3)) and $\{\mathbf{H}_kV_k,\, k \in \mathbb{Z}\}$ are finite-variance, stationary random processes of finite entropy rate, so the limit

$$\lim_{k\to\infty}\Big\{I\big(V_k;\mathbf{H}_kV_k\big) + I\big(\mathbf{H}_1^{k-1};\mathbf{H}_k\big) - I\big(\mathbf{H}_kV_k;\{\mathbf{H}_\ell V_\ell\}_{\ell=1}^{k-1}\big)\Big\} \tag{4.97}$$

exists and is finite. Indeed, $\{\mathbf{H}_kV_k,\, k \in \mathbb{Z}\}$ is stationary because any (time-invariant) function of stationary processes is again a stationary process. Furthermore, it is of finite variance because

$$\mathsf{E}\left[\|\mathbf{H}_kV_k\|^2\right] = \mathsf{E}\left[\|\mathbf{H}_k\|^2\right]\mathsf{E}\left[|V_k|^2\right],$$

which is, by (4.2) and (4.42), finite; and it is of finite entropy rate because

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}h\big(\mathbf{H}_1V_1,\ldots,\mathbf{H}_nV_n\big) &= \lim_{n\to\infty} h\big(\mathbf{H}_nV_n \bigm| \{\mathbf{H}_\ell V_\ell\}_{\ell=1}^{n-1}\big)\\ &\geq \lim_{n\to\infty} h\big(\mathbf{H}_nV_n \bigm| \mathbf{H}_1^{n-1}, V_n\big)\\ &= n_{\mathrm{R}}\mathsf{E}\left[\log|V_1|^2\right] + \lim_{n\to\infty} h\big(\mathbf{H}_n \bigm| \mathbf{H}_1^{n-1}\big),\end{aligned}$$

which is, by (4.42) and (4.3), finite. Here the first step follows because $\{\mathbf{H}_kV_k,\, k \in \mathbb{Z}\}$ is stationary [5, Thm. 4.2.1]; the second step follows by conditioning the entropy additionally on $\big(\mathbf{H}_1^{n-1}, V_n\big)$, and by noting that

$$\{\mathbf{H}_\ell V_\ell\}_{\ell=1}^{n-1} \multimap \big(\mathbf{H}_1^{n-1}, V_n\big) \multimap \mathbf{H}_n V_n$$

forms a Markov chain; and the last step follows from the behavior of differential entropy under scaling and from the stationarity of $\{V_k,\, k \in \mathbb{Z}\}$.

Since the limit in (4.97) exists, we can make use of Cesáro's mean [5, Thm. 4.2.3] to obtain

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} I\big(V_1^n; \{\mathbf{H}_\ell V_\ell\}_{\ell=1}^n\big) \\
&\quad = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n \Big(I\big(V_k; \mathbf{H}_k V_k\big) + I\big(\mathbf{H}_1^{k-1}; \mathbf{H}_k\big) - I\big(\mathbf{H}_k V_k; \mathbf{H}_\ell V_\ell\}_{\ell=1}^{k-1}\big)\Big) \\
&\quad = I\big(V_0; \mathbf{H}_0 V_0\big) + I\big(\mathbf{H}_{-\infty}^{-1}; \mathbf{H}_0\big) - I\big(\mathbf{H}_0 V_0; \{\mathbf{H}_\ell V_\ell\}_{\ell=-\infty}^{-1}\big).
\end{aligned} \tag{4.98}$$

The first term in (4.95) can be computed from (4.94)

$$\begin{aligned}
\chi(\{\mathbf{H}_k V_k\}) = {} & h_\lambda\big(\hat{\mathbf{H}}_0 \hat{V}_0\, e^{\mathrm{i}\Phi_0}\big) - h\big(\mathbf{H}_0 V_0\big) + n_\mathrm{R} \mathsf{E}\big[\log \|\mathbf{H}_0 V_0\|^2\big] \\
& - \log 2 + I\big(\mathbf{H}_0 V_0; \{\mathbf{H}_\ell V_\ell\}_{\ell=-\infty}^{-1}\big) \\
& - I\big(\hat{\mathbf{H}}_0 \hat{V}_0\, e^{\mathrm{i}\Phi_0}; \{\hat{\mathbf{H}}_\ell \hat{V}_\ell\, e^{\mathrm{i}\Phi_\ell}\}_{\ell=-\infty}^{-1}\big),
\end{aligned} \tag{4.99}$$

where $\hat{V}_\ell = V_\ell/|V_\ell|$. Combining (4.98) and (4.99) yields

$$\begin{aligned}
&\chi(\{\mathbf{H}_k V_k\}) + \lim_{n\to\infty} \frac{1}{n} I\big(V_1^n; \{\mathbf{H}_k V_k\}_{k=1}^n\big) \\
&\quad = h_\lambda\big(\hat{\mathbf{H}}_0 \hat{V}_0\, e^{\mathrm{i}\Phi_0}\big) - h(\mathbf{H}_0 V_0) + n_\mathrm{R} \mathsf{E}\big[\log \|\mathbf{H}_0 V_0\|^2\big] - \log 2 \\
&\qquad + I\big(\mathbf{H}_0 V_0; \{\mathbf{H}_\ell V_\ell\}_{\ell=-\infty}^{-1}\big) - I\big(\hat{\mathbf{H}}_0 \hat{V}_0\, e^{\mathrm{i}\Phi_0}; \{\hat{\mathbf{H}}_\ell \hat{V}_\ell\, e^{\mathrm{i}\Phi_\ell}\}_{\ell=-\infty}^{-1}\big) \\
&\qquad + I\big(V_0; \mathbf{H}_0 V_0\big) + I\big(\mathbf{H}_{-\infty}^{-1}; \mathbf{H}_0\big) - I\big(\mathbf{H}_0 V_0; \{\mathbf{H}_\ell V_\ell\}_{\ell=-\infty}^{-1}\big) \\
&\quad = h_\lambda\big(\hat{\mathbf{H}}_0 \hat{V}_0\, e^{\mathrm{i}\Phi_0}\big) - h\big(\mathbf{H}_0 V_0\big) + n_\mathrm{R} \mathsf{E}\big[\log \|\mathbf{H}_0 V_0\|^2\big] - \log 2 \\
&\qquad - I\big(\hat{\mathbf{H}}_0 \hat{V}_0\, e^{\mathrm{i}\Phi_0}; \{\hat{\mathbf{H}}_\ell \hat{V}_\ell\, e^{\mathrm{i}\Phi_\ell}\}_{\ell=-\infty}^{-1}\big) + I\big(V_0; \mathbf{H}_0 V_0\big) + I\big(\mathbf{H}_{-\infty}^{-1}; \mathbf{H}_0\big) \\
&\quad = h_\lambda\big(\hat{\mathbf{H}}_0 \hat{V}_0\, e^{\mathrm{i}\Phi_0}\big) - h\big(\mathbf{H}_0 V_0\big) + n_\mathrm{R} \mathsf{E}\big[\log \|\mathbf{H}_0 V_0\|^2\big] - \log 2 \\
&\qquad - I\big(\hat{\mathbf{H}}_0 \hat{V}_0\, e^{\mathrm{i}\Phi_0}; \{\hat{\mathbf{H}}_\ell \hat{V}_\ell\, e^{\mathrm{i}\Phi_\ell}\}_{\ell=-\infty}^{-1}\big) + h\big(\mathbf{H}_0 V_0\big) - h\big(\mathbf{H}_0 V_0 \bigm| V_0\big) \\
&\qquad + I\big(\mathbf{H}_{-\infty}^{-1}; \mathbf{H}_0\big) \\
&\quad = h_\lambda\big(\hat{\mathbf{H}}_0 \hat{V}_0\, e^{\mathrm{i}\Phi_0}\big) + n_\mathrm{R} \mathsf{E}\big[\log \|\mathbf{H}_0 V_0\|^2\big] - \log 2 \\
&\qquad - I\big(\hat{\mathbf{H}}_0 \hat{V}_0\, e^{\mathrm{i}\Phi_0}; \{\hat{\mathbf{H}}_\ell \hat{V}_\ell\, e^{\mathrm{i}\Phi_\ell}\}_{\ell=-\infty}^{-1}\big) - h\big(\mathbf{H}_0 V_0 \bigm| V_0\big) + I\big(\mathbf{H}_{-\infty}^{-1}; \mathbf{H}_0\big)
\end{aligned}$$

$$
\begin{aligned}
&= h_\lambda\big(\hat{\mathbf{H}}_0\hat{V}_0\, e^{\mathrm{i}\Phi_0}\big) + n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big] + n_{\mathrm{R}}\mathsf{E}\big[\log|V_0|^2\big] - \log 2 \\
&\quad - I\big(\hat{\mathbf{H}}_0\hat{V}_0\, e^{\mathrm{i}\Phi_0}; \{\hat{\mathbf{H}}_\ell\hat{V}_\ell\, e^{\mathrm{i}\Phi_\ell}\}_{\ell=-\infty}^{-1}\big) - h\big(\mathbf{H}_0\big) - n_{\mathrm{R}}\mathsf{E}\big[\log|V_0|^2\big] \\
&\quad + I\big(\mathbf{H}_{-\infty}^{-1}; \mathbf{H}_0\big) \\
&= h_\lambda\big(\hat{H}_0\hat{V}_0\, e^{\mathrm{i}\Phi_0}\big) - h\big(\mathbf{H}_0\big) + n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big] - \log 2 \\
&\quad + I\big(\mathbf{H}_{-\infty}^{-1}; \mathbf{H}_0\big) - I\big(\hat{\mathbf{H}}_0\hat{V}_0\, e^{\mathrm{i}\Phi_0}; \{\hat{\mathbf{H}}_\ell\hat{V}_\ell\, e^{\mathrm{i}\Phi_\ell}\}_{\ell=-\infty}^{-1}\big) \\
&= h_\lambda\big(\hat{H}_0\, e^{\mathrm{i}\Phi_0}\big) - h\big(\mathbf{H}_0\big) + n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big] - \log 2 \\
&\quad + I\big(\mathbf{H}_{-\infty}^{-1}; \mathbf{H}_0\big) - I\big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Phi_0}; \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Phi_\ell}\}_{\ell=-\infty}^{-1}\big) \\
&= \chi(\{\mathbf{H}_k\}),
\end{aligned} \tag{4.100}
$$

where the fifth step follows from the behavior of differential entropy under scaling; and where the seventh step follows because $\{\Phi_k,\, k \in \mathbb{Z}\}$ is a sequence of IID random variables that are uniformly distributed over the interval $(\pi, \pi]$, drawn independently of $\{\hat{V}_k,\, k \in \mathbb{Z}\}$, which implies that $\{\hat{V}_k\, e^{\mathrm{i}\Phi_k},\, k \in \mathbb{Z}\}$ has the same law as $\{\, e^{\mathrm{i}\Phi_k},\, k \in \mathbb{Z}\}$.

This concludes the proof of Note 4.2.

4.8.4 Proof of Theorem 4.6

Without loss of generality, we assume that $n_{\min} = n_{\mathrm{R}} = n_{\mathrm{T}}$. (When $n_{\mathrm{R}} > n_{\mathrm{T}}$, we can ignore $n_{\mathrm{R}} - n_{\mathrm{T}}$ receive antennas, and when $n_{\mathrm{T}} > n_{\mathrm{R}}$, we can transmit only from n_{R} transmit antennas. This yields in both cases a lower bound.)

To derive the lower bound (4.48), we use the general lower bound on the fading number (4.43), namely

$$
\chi_{\mathrm{PP}}(\{\mathbb{H}_k\}) \geq \chi(\{\mathbb{H}_k\mathbf{V}_k\}) + \lim_{n\to\infty}\frac{1}{n} I\big(\mathbf{V}_1^n; \{\mathbb{H}_\ell\mathbf{V}_\ell\}_{\ell=1}^n\big), \tag{4.101}
$$

where $\{\mathbf{V}_k,\, k \in \mathbb{Z}\}$ is a stationary ergodic process that is independent of $\{\mathbb{H}_k,\, k \in \mathbb{Z}\}$, and that satisfies

$$
\lim_{k\to\infty} I\big(\mathbf{V}_k; \mathbf{V}_1^{k-1}\big) < \infty, \tag{4.102}
$$

$$
\mathsf{E}\big[\|\mathbf{V}_k\|^2\big] < \infty \qquad \text{and} \qquad \mathsf{E}\big[\log\|\mathbf{V}_k\|^2\big] > -\infty, \tag{4.103}
$$

and

$$
\Pr(\|\mathbf{V}_k\| > \Upsilon) = 0, \quad \text{for some } \Upsilon > 0. \tag{4.104}
$$

The assumption (4.104) ensures that the peak-power constraint is satisfied. (The limit in (4.103) exists because $\{\mathbf{V}_k,\, k \in \mathbb{Z}\}$ is stationary.)

We shall evaluate the RHS of (4.101) for $\{\mathbf{V}_k,\, k \in \mathbb{Z}\}$ being a sequence of IID random vectors satisfying (4.103) and (4.104). (This choice certainly satisfies (4.102) because $I\big(\mathbf{V}_k; \mathbf{V}_1^{k-1}\big) = 0$, $k \in \mathbb{N}$.)

We first lower bound $\chi(\{\mathbb{H}_k \mathbf{V}_k\})$. An exact expression for the fading number of SIMO fading with memory can be found in [29], but this expression is not easy to evaluate. We therefore lower bound it by considering a linear combining at the receiver, and by ignoring the memory in $\{\mathbb{H}_k \mathbf{V}_k,\, k \in \mathbb{Z}\}$, i.e.,

$$\chi(\{\mathbb{H}_k \mathbf{V}_k\}) \geq \chi^{(\mathrm{IID})}(\boldsymbol{\alpha}^{\mathsf{T}} \mathbb{H}_1 \mathbf{V}_1), \quad \boldsymbol{\alpha} \in \mathbb{C}^{n_{\mathrm{T}}} \text{ deterministic.} \tag{4.105}$$

We note that the fading number for IID SISO fading is given by [28, Thm. 4.16]

$$\chi(H_1) = \log \pi + \mathsf{E}\big[\log |H_1|^2\big] - h(H_1). \tag{4.106}$$

We thus have

$$\begin{aligned}
&\chi(\{\mathbb{H}_k \mathbf{V}_k\}) \\
&\quad \geq \log \pi + \mathsf{E}\Big[\log\big|\boldsymbol{\alpha}^{\mathsf{T}} \mathbb{H}_1 \mathbf{V}_1\big|^2\Big] - h\big(\boldsymbol{\alpha}^{\mathsf{T}} \mathbb{H}_1 \mathbf{V}_1\big) \\
&\quad \geq \log \pi + \mathsf{E}\Big[\log\big|\boldsymbol{\alpha}^{\mathsf{T}} \mathbb{H}_1 \mathbf{V}_1\big|^2\Big] - \log\Big(\pi e\, \mathsf{E}\Big[\big|\boldsymbol{\alpha}^{\mathsf{T}} \mathbb{H}_1 \mathbf{V}_1\big|^2\Big]\Big),
\end{aligned} \tag{4.107}$$

where the second step follows from the entropy maximizing property of Gaussian random variables. Using that $\mathbb{H}_k$ is spatially IID, it follows that, conditional on $\mathbf{V}_k = \mathbf{v}_k$, the random variable $\boldsymbol{\alpha}^{\mathsf{T}} \mathbb{H}_k \mathbf{v}_k$ is a zero-mean, circularly-symmetric, complex Gaussian random variable of variance $\|\boldsymbol{\alpha}\|^2 \, \|\mathbf{v}_k\|^2$. We thus have

$$\mathsf{E}\Big[\big|\boldsymbol{\alpha}^{\mathsf{T}} \mathbb{H}_1 \mathbf{V}_1\big|^2\Big] = \|\boldsymbol{\alpha}\|^2\, \mathsf{E}\big[\|\mathbf{V}_1\|^2\big]$$

and

$$\begin{aligned}
\mathsf{E}\Big[\log\big|\boldsymbol{\alpha}^{\mathsf{T}} \mathbb{H}_1 \mathbf{V}_1\big|^2\Big] &= \mathsf{E}\Big[\mathsf{E}\Big[\log\big|\boldsymbol{\alpha}^{\mathsf{T}} \mathbb{H}_1 \mathbf{V}_1\big|^2 \,\Big|\, \mathbf{V}_1 = \mathbf{v}_1\Big]\Big] \\
&= \mathsf{E}\big[\log\big(\|\boldsymbol{\alpha}\|^2 \, \|\mathbf{V}_1\|^2\big)\big] - \gamma,
\end{aligned}$$

where the second step follows from [16, p. 567, Sec. 4.331]. The fading number $\chi(\{\mathbb{H}_k \mathbf{V}_k\})$ is thus lower bounded by

$$\chi(\{\mathbb{H}_k \mathbf{V}_k\}) \geq \mathsf{E}\big[\log \|\mathbf{V}_1\|^2\big] - \log \mathsf{E}\big[\|\mathbf{V}_1\|^2\big] - \gamma - 1. \tag{4.108}$$

It follows from (4.103) that the RHS of (4.108) is finite. Note that it only depends on the statistics of $\mathbf{V}_1$ and not on the memory of the fading process.

We turn now to the second term on the RHS of (4.101). We use the chain rule for mutual information to obtain

$$\begin{aligned}\frac{1}{n} I\big(\mathbf{V}_1^n; \{\mathbb{H}_\ell \mathbf{V}_\ell\}_{\ell=1}^n\big) &= \frac{1}{n}\sum_{k=1}^{n} I\big(\mathbf{V}_k; \{\mathbb{H}_\ell \mathbf{V}_\ell\}_{\ell=1}^n \bigm| \mathbf{V}_1^{k-1}\big) \\ &\geq \frac{1}{n}\sum_{k=1}^{n} I\big(\mathbf{V}_k; \{\mathbb{H}_\ell \mathbf{V}_\ell\}_{\ell=1}^k \bigm| \mathbf{V}_1^{k-1}\big) \\ &= \frac{1}{n}\sum_{k=1}^{n} I\big(\mathbf{V}_k; \{\mathbb{H}_\ell \mathbf{V}_\ell\}_{\ell=1}^k, \mathbf{V}_1^{k-1}\big), \qquad (4.109)\end{aligned}$$

where the last step follows because $\{\mathbf{V}_k,\, k \in \mathbb{Z}\}$ is IID. Using a Cesáro-type theorem, we obtain

$$\begin{aligned}&\varliminf_{n\to\infty} \frac{1}{n} I\big(\mathbf{V}_1^n; \{\mathbb{H}_\ell \mathbf{V}_\ell\}_{\ell=1}^n\big) \\ &\quad\geq \varliminf_{k\to\infty} I\big(\mathbf{V}_k; \{\mathbb{H}_\ell \mathbf{V}_\ell\}_{\ell=1}^k, \mathbf{V}_1^{k-1}\big) \\ &\quad\geq \varliminf_{k\to\infty} I\big(\mathbf{V}_k; \{\mathbb{H}_\ell \mathbf{V}_\ell\}_{\ell=k-n_{\min}-2}^k, \mathbf{V}_{k-n_{\min}-2}^{k-1}\big) \\ &\quad= I\big(\mathbf{V}_0; \{\mathbb{H}_\ell \mathbf{V}_\ell\}_{\ell=-n_{\min}-2}^0, \mathbf{V}_{-n_{\min}-2}^{-1}\big), \qquad (4.110)\end{aligned}$$

where the second step follows by reducing the number of observables, and where the last step follows because the processes $\{\mathbf{V}_k,\, k \in \mathbb{Z}\}$ and $\{\mathbb{H}_k,\, k \in \mathbb{Z}\}$ are stationary.

Since Gaussian random variables are comparably easy to handle, we wish to analyze the RHS of (4.110) for $\mathbf{V}_{-n_{\min}-2}, \ldots, \mathbf{V}_0$ being a sequence of IID, zero-mean, circularly-symmetric, complex Gaussian vectors of covariance matrix $\mathsf{I}_{n_{\min}}$. Unfortunately, Gaussian vectors do not satisfy the boundedness condition (4.104), and we have to dig into our bag of mathematical tricks to sidestep this problem: instead of Gaussian vectors we shall consider vectors of a truncated Gaussian distribution, whose density is a normalized version of the Gaussian density for $\|\mathbf{v}_k\| \leq \Upsilon$, and whose density is zero for $\|\mathbf{v}_k\| > \Upsilon$. We then show that, for sufficiently large Υ, the fading number that can be achieved by vectors of a truncated Gaussian law cannot be much smaller than the fading number that can be achieved by vectors of a Gaussian law. In other words, by evaluating the RHS of (4.110) for Gaussian vectors, we obtain a valid lower bound on the fading number, even though condition (4.104) is not satisfied.

In the following we make the above argument rigorous. Note that condition (4.104) is only necessary to ensure that the channel inputs satisfy the peak-power constraint. So those readers who are only interested in the fading number under an average-power constraint might want to skip this part and continue with Equation (4.119).

Let $\mathbf{V}_{-n_{\min}-2}, \ldots, \mathbf{V}_0$ be a sequence of IID complex random vectors of a truncated zero-mean, circularly-symmetric, complex Gaussian distribution of covariance matrix $\mathsf{I}_{n_{\min}}$, whose density is given by

$$f_{\mathbf{V}_k}(\mathbf{v}_k) = \frac{\exp(-\|\mathbf{v}_k\|^2)}{\pi^{n_{\min}}\,\zeta(\Upsilon)}\,\mathrm{I}\left\{\|\mathbf{v}_k\| \leq \Upsilon\right\}, \quad \mathbf{v}_k \in \mathbb{C}^{n_{\min}} \tag{4.111}$$

for some $\Upsilon > 0$, where

$$\zeta(\Upsilon) = \int_{\|\boldsymbol{\xi}\|\leq\Upsilon} \frac{1}{\pi^{n_{\min}}} \exp\bigl(-\|\boldsymbol{\xi}\|^2\bigr)\,\mathrm{d}\boldsymbol{\xi}.$$

The random vectors $\mathbf{V}_{-n_{\min}-2}, \ldots, \mathbf{V}_0$ satisfy (4.104). In the following we demonstrate that they also satisfy (4.103). This is a direct consequence of Lemma 4.18 ahead. Lemma 4.19 shows then that, for sufficiently large Υ, the RHS of (4.110) evaluated for $\mathbf{V}_\ell$, $\ell = -n_{\min} - 2, \ldots, 0$ being of the truncated Gaussian distribution (4.111) cannot be much smaller than the same term evaluated for $\mathbf{V}_\ell$, $\ell = -n_{\min} - 2, \ldots, 0$ being of a zero-mean, circularly-symmetric, complex Gaussian distribution of covariance matrix $\mathsf{I}_{n_{\min}}$.

Lemma 4.18. *Let* $\mathbf{V}$ *be distributed according to the truncated Gaussian distribution given in* (4.111)*, and let* $\mathbf{U}$ *be distributed according to a* $n_{\min}$*-variate, zero-mean, circularly-symmetric, complex Gaussian distribution of covariance matrix* $\mathsf{I}_{n_{\min}}$*. Then*

$$\lim_{\Upsilon\to\infty} \mathsf{E}\left[\log \|\mathbf{V}\|^2\right] = \mathsf{E}\left[\log \|\mathbf{U}\|^2\right] \tag{4.112}$$

and

$$\lim_{\Upsilon\to\infty} \mathsf{E}\left[\|\mathbf{V}\|^2\right] = \mathsf{E}\left[\|\mathbf{U}\|^2\right]. \tag{4.113}$$

Proof. We first note that

$$\lim_{\Upsilon\to\infty}\zeta(\Upsilon)=\lim_{\Upsilon\to\infty}\int_{\|\boldsymbol{\xi}\|\le\Upsilon}\frac{1}{\pi^{n_{\min}}}\exp\bigl(-\|\boldsymbol{\xi}\|^2\bigr)\,\mathrm{d}\boldsymbol{\xi}=1, \tag{4.114}$$

which follows from the Monotone Convergence Theorem [38, Thm. 1.26]. To prove (4.112), i.e.,

$$\lim_{\Upsilon\to\infty}\frac{1}{\pi^{n_{\min}}\zeta(\Upsilon)}\int_{\|\mathbf{v}\|\le\Upsilon}\exp\bigl(-\|\mathbf{v}\|^2\bigr)\log\|\mathbf{v}\|^2\,\mathrm{d}\mathbf{v} = \int\exp\bigl(-\|\mathbf{v}\|^2\bigr)\log\|\mathbf{v}\|^2\,\mathrm{d}\mathbf{v},$$

we further note that, in view of (4.114), it suffices to show that

$$\lim_{\Upsilon\to\infty}\int_{\|\mathbf{v}\|\le\Upsilon}\exp\bigl(-\|\mathbf{v}\|^2\bigr)\log\|\mathbf{v}\|^2\,\mathrm{d}\mathbf{v}=\int\exp\bigl(-\|\mathbf{v}\|^2\bigr)\log\|\mathbf{v}\|^2\,\mathrm{d}\mathbf{v}.$$

But this follows from

$$\begin{aligned}
\lim_{\Upsilon\to\infty}&\int_{\|\mathbf{v}\|\le\Upsilon}\exp\bigl(-\|\mathbf{v}\|^2\bigr)\log\|\mathbf{v}\|^2\,\mathrm{d}\mathbf{v}\\
&=\int_{\|\mathbf{v}\|<1}\exp\bigl(-\|\mathbf{v}\|^2\bigr)\log\|\mathbf{v}\|^2\,\mathrm{d}\mathbf{v}\\
&\qquad+\lim_{\Upsilon\to\infty}\int_{1\le\|\mathbf{v}\|\le\Upsilon}\exp\bigl(-\|\mathbf{v}\|^2\bigr)\log\|\mathbf{v}\|^2\,\mathrm{d}\mathbf{v}\\
&=\int_{\|\mathbf{v}\|<1}\exp\bigl(-\|\mathbf{v}\|^2\bigr)\log\|\mathbf{v}\|^2\,\mathrm{d}\mathbf{v}+\int_{\|\mathbf{v}\|\ge1}\exp\bigl(-\|\mathbf{v}\|^2\bigr)\log\|\mathbf{v}\|^2\,\mathrm{d}\mathbf{v}\\
&=\int\exp\bigl(-\|\mathbf{v}\|^2\bigr)\log\|\mathbf{v}\|^2\,\mathrm{d}\mathbf{v},
\end{aligned}$$

where the second step follows from the Monotone Convergence Theorem. (Note that $\int_{\|\mathbf{v}\|<1}\exp\bigl(-\|\mathbf{v}\|^2\bigr)\log\|\mathbf{v}\|^2\,\mathrm{d}\mathbf{v}$ and $\int_{\|\mathbf{v}\|\ge1}\exp\bigl(-\|\mathbf{v}\|^2\bigr)\log\|\mathbf{v}\|^2\,\mathrm{d}\mathbf{v}$ are both finite.)

To prove (4.113), i.e.,

$$\lim_{\Upsilon\to\infty}\frac{1}{\pi^{n_{\min}}\,\zeta(\Upsilon)}\int_{\|\mathbf{v}\|\le\Upsilon}\|\mathbf{v}\|^2\exp\bigl(-\|\mathbf{v}\|^2\bigr)\,\mathrm{d}\mathbf{v} = \frac{1}{\pi^{n_{\min}}}\int\|\mathbf{v}\|^2\exp\bigl(-\|\mathbf{v}\|^2\bigr)\,\mathrm{d}\mathbf{v},$$

it suffices, by (4.114), to show that

$$\lim_{\Upsilon\to\infty}\int_{\|\mathbf{v}\|\leq\Upsilon}\|\mathbf{v}\|^2\exp\bigl(-\|\mathbf{v}\|^2\bigr)\,\mathrm{d}\mathbf{v}=\int\|\mathbf{v}\|^2\exp\bigl(-\|\mathbf{v}\|^2\bigr)\,\mathrm{d}\mathbf{v}. \quad (4.115)$$

But this follows again from the Monotone Convergence Theorem. □

Thus, since a circularly-symmetric, complex Gaussian vector of zero-mean and covariance matrix $\mathsf{I}_{n_{\min}}$ satisfies condition (4.103), it follows from Lemma 4.18 that, for sufficiently large Υ, a vector that is distributed according to the truncated Gaussian density given in (4.111) satisfies (4.103), too.

The next lemma considers the mutual information term on the RHS of (4.110).

Lemma 4.19. *Let $\mathbf{V}_{-n_{\min}-2},\ldots,\mathbf{V}_0$ be a sequence of IID complex random variables distributed according to the truncated Gaussian density given in* (4.111)*, and let $\mathbf{U}_{-n_{\min}-2},\ldots,\mathbf{U}_0$ be a sequence of IID, zero-mean, circularly-symmetric, complex Gaussian random vectors of covariance matrix $\mathsf{I}_{n_{\min}}$. Then*

$$\begin{aligned}\varliminf_{\Upsilon\to\infty} I\bigl(\mathbf{V}_0;\{\mathbb{H}_\ell\mathbf{V}_\ell\}_{\ell=-n_{\min}-2}^{0},\mathbf{V}_{-n_{\min}-2}^{-1}\bigr)&\\ \geq I\bigl(\mathbf{U}_0;\{\mathbb{H}_\ell\mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{0},\mathbf{U}_{n_{\min}-2}^{-1}\bigr).& \quad (4.116)\end{aligned}$$

Proof. Let $P_\Upsilon^{(x,y)}$ denote the law of $\bigl(\mathbf{V}_{-n_{\min}-2}^{0},\{\mathbb{H}_\ell\mathbf{V}_\ell\}_{\ell=-n_{\min}-2}^{0}\bigr)$, let $P_\Upsilon^{(x)}$ denote the law of $\mathbf{V}_0$, and let $P_\Upsilon^{(y)}$ denote the law of $\bigl(\mathbf{V}_{-n_{\min}-2}^{-1},\{\mathbb{H}_\ell\mathbf{V}_\ell\}_{\ell=-n_{\min}-2}^{0}\bigr)$. Similarly, let $Q^{(x,y)}$ denote the law of $\bigl(\mathbf{U}_{-n_{\min}-2}^{0},\{\mathbb{H}_\ell\mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{0}\bigr)$, let $Q^{(x)}$ denote the law of $\mathbf{U}_0$, and let $Q^{(y)}$ denote the law of $\bigl(\mathbf{U}_{-n_{\min}-2}^{-1},\{\mathbb{H}_\ell\mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{0}\bigr)$. We note that, as Υ tends to infinity,

(i) $P_\Upsilon^{(x,y)}$ converges weakly to $Q^{(x,y)}$;

(ii) $P_\Upsilon^{(x)}$ converges weakly to $Q^{(x)}$;

(iii) and $P_\Upsilon^{(y)}$ converges weakly to $Q^{(y)}$.

(See, e.g., [36, Sec. II] for a definition of weak convergence.) Expressing mutual information in terms of relative entropy [5, Sec. 9.5], it follows that (4.116) is equivalent to

$$\varliminf_{\Upsilon\to\infty} D\Big(P_{\Upsilon}^{(x,y)}\,\Big\|\,P_{\Upsilon}^{(x)}\times P_{\Upsilon}^{(y)}\Big) \geq D\Big(Q^{(x,y)}\,\Big\|\,Q^{(x)}\times Q^{(y)}\Big).$$

But this follows immediately from the lower semicontinuity of relative entropy [36, Thm. 1]. □

We return to the analysis of (4.110). Let $\mathbf{V}_{-n_{\min}-2},\ldots,\mathbf{V}_0$ and $\mathbf{U}_{-n_{\min}-2},\ldots,\mathbf{U}_0$ be as in Lemma 4.19. By Lemmas 4.18 and 4.19, there exists for every $\varepsilon>0$ an $\Upsilon>0$ such that

$$\begin{aligned} \Big|\mathsf{E}\big[\log\|\mathbf{V}_\ell\|^2\big]-\mathsf{E}\big[\log\|\mathbf{U}_\ell\|^2\big]\Big| &\leq \varepsilon, \\ \Big|\log\mathsf{E}\big[\|\mathbf{V}_\ell\|^2\big]-\log\mathsf{E}\big[\|\mathbf{U}_\ell\|^2\big]\Big| &\leq \varepsilon, \end{aligned} \qquad \ell=-n_{\min}-2,\ldots,0 \quad (4.117)$$

and

$$\begin{aligned} I\big(\mathbf{V}_0;\{\mathbb{H}_\ell\mathbf{V}_\ell\}_{\ell=-n_{\min}-2}^{0},\mathbf{V}_{-n_{\min}-2}^{-1}\big) \\ \geq I\big(\mathbf{U}_0;\{\mathbb{H}_\ell\mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{0},\mathbf{U}_{n_{\min}-2}^{-1}\big)-\varepsilon. \end{aligned} \qquad (4.118)$$

We lower bound the mutual information on the RHS of (4.118) as follows

$$\begin{aligned} &I\big(\mathbf{U}_0;\{\mathbb{H}_\ell\mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{0},\mathbf{U}_{n_{\min}-2}^{-1}\big) \\ &\quad\geq I\big(\mathbf{U}_0;\mathbb{H}_0\mathbf{U}_0\,\big|\,\{\mathbb{H}_\ell\mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{-1},\mathbf{U}_{n_{\min}-2}^{-1}\big) \\ &\quad= h\big(\mathbb{H}_0\mathbf{U}_0\,\big|\,\{\mathbb{H}_\ell\mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{-1},\mathbf{U}_{n_{\min}-2}^{-1}\big) \\ &\qquad - h\big(\mathbb{H}_0\mathbf{U}_0\,\big|\,\{\mathbb{H}_\ell\mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{-1},\mathbf{U}_{n_{\min}-2}^{0}\big), \end{aligned} \qquad (4.119)$$

where the first step follows by using the chain rule to expand the mutual information as a sum of an unconditional mutual information and a conditional mutual information, and by using then the nonnegativity of mutual information to lower bound the unconditional mutual information by zero.

The first entropy on the RHS of (4.119) can be lower bounded by

$$\begin{aligned} &h\Big(\mathbb{H}_0\mathbf{U}_0 \;\Big|\; \{\mathbb{H}_\ell\mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{-1}, \mathbf{U}_{-n_{\min}-2}^{-1}\Big) \\ &\quad \geq h\Big(\mathbb{H}_0\mathbf{U}_0 \;\Big|\; \mathbb{H}_0, \{\mathbb{H}_\ell\mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{-1}, \mathbf{U}_{-n_{\min}-2}^{-1}\Big) \\ &\quad = h(\mathbb{H}_0\mathbf{U}_0|\mathbb{H}_0), \end{aligned} \tag{4.120}$$

where the inequality follows because conditioning cannot increase entropy; and the equality follows because

$$\Big(\mathbf{U}_{-n_{\min}-2}^{-1}, \{\mathbb{H}_\ell\mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{-1}\Big) \multimap\!\!- \mathbb{H}_0 \multimap\!\!- \mathbb{H}_0\mathbf{U}_0$$

forms a Markov chain.

Since, conditional on $\mathbb{H}_0 = \mathsf{H}_0$, the vector $\mathbb{H}_0\mathbf{U}_0$ has a circularly-symmetric, complex Gaussian distribution of mean zero and covariance matrix

$$\mathsf{H}_0 \underbrace{\mathsf{E}\Big[\mathbf{U}_0\mathbf{U}_0^\dagger\Big]}_{=\mathsf{I}_{n_{\min}}} \mathsf{H}_0^\dagger = \mathsf{H}_0\mathsf{H}_0^\dagger,$$

we obtain

$$\begin{aligned} h(\mathbb{H}_0\mathbf{U}_0|\mathbb{H}_0) &= n_{\min}\,\log(\pi e) + \mathsf{E}\Big[\log\det \mathbb{H}_0\mathbb{H}_0^\dagger\Big] \\ &= n_{\min}\,\log(\pi e) + \sum_{i=0}^{n_{\min}-1} \psi(n_{\min}-i), \end{aligned} \tag{4.121}$$

where $\psi(\cdot)$ denotes Euler's psi-function. Here the last step follows from [17, Lemma A.2].

We next upper bound the second entropy on the RHS of (4.119). Consider a specific realization of

$$\Big(\{\mathbb{H}_\ell\mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{-1}, \mathbf{U}_{n_{\min}-2}^{0}\Big),$$

which we denote by

$$\Big(\{\mathsf{H}_\ell\mathbf{u}_\ell\}_{\ell=-n_{\min}-2}^{-1}, \mathbf{u}_{n_{\min}-2}^{0}\Big).$$

Let $\hat{\mathbf{u}}_0 \triangleq \mathbf{u}_0/\|\mathbf{u}_0\|$. With probability one, $\hat{\mathbf{u}}_0$ is in the span of $\mathbf{u}_{-n_{\min}-2}, \ldots, \mathbf{u}_{-1}$. Let $\alpha_1, \ldots, \alpha_{n_{\min}+2}$ be the set of coefficients of least L_2-norm such that

$$\hat{\mathbf{u}}_0 = \sum_{\ell=1}^{n_{\min}+2} \alpha_\ell \, \mathbf{u}_{-\ell}, \tag{4.122}$$

which can be written in matrix notation as

$$\underbrace{\begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \mathbf{u}_{-1} & \mathbf{u}_{-2} & \cdots & \mathbf{u}_{-n_{\min}-2} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}}_{\triangleq \mathsf{U}} \underbrace{\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{n_{\min}+2} \end{pmatrix}}_{\triangleq \boldsymbol{\alpha}} = \begin{pmatrix} \uparrow \\ \hat{\mathbf{u}}_0 \\ \downarrow \end{pmatrix}. \tag{4.123}$$

The problem of finding $\boldsymbol{\alpha} \triangleq (\alpha_1, \ldots, \alpha_{n_{\min}+2})^\mathsf{T}$ can be solved using the singular value decomposition (SVD):

$$\mathsf{U} = \mathsf{F}\mathsf{S}\mathsf{G}^\dagger, \tag{4.124}$$

where $\mathsf{F} \in \mathbb{C}^{n_{\min} \times n_{\min}}$ and $\mathsf{G} \in \mathbb{C}^{(n_{\min}+2) \times (n_{\min}+2)}$ are unitary matrices, and where the matrix $\mathsf{S} \in \mathbb{C}^{n_{\min} \times (n_{\min}+2)}$ is given by

$$\mathsf{S} = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \sigma_{n_{\min}} & 0 & \cdots & 0 \end{pmatrix}.$$

Here $\sigma_1, \ldots, \sigma_{n_{\min}}$ denote the singular values of U. We thus obtain from (4.123) and (4.124)

$$\mathsf{S}\mathsf{G}^\dagger \boldsymbol{\alpha} = \mathsf{F}^\dagger \hat{\mathbf{u}}_0, \tag{4.125}$$

which becomes, upon defining $\boldsymbol{\beta} \triangleq G^\dagger \boldsymbol{\alpha}$,

$$\mathsf{S}\boldsymbol{\beta} = F^\dagger \hat{\mathbf{u}}_0. \tag{4.126}$$

The solution for $\boldsymbol{\beta}$ with the least Euclidean norm satisfying (4.126) is

$$\beta_i = \begin{cases} \dfrac{\left(\mathsf{F}^\dagger \hat{\mathbf{u}}_0\right)_i}{\sigma_i}, & \text{for } i = 1, \ldots, n_{\min} \\ 0, & \text{otherwise,} \end{cases} \tag{4.127}$$

where β_i and $\left(\mathsf{F}^\dagger \hat{\mathbf{u}}_0\right)_i$ denote the i-th component of the vectors $\boldsymbol{\beta}$ and $\mathsf{F}^\dagger \hat{\mathbf{u}}_0$.

Later in the proof, we will need to upper bound $\|\boldsymbol{\alpha}\|^2$. Since G is unitary, it follows that $\|\boldsymbol{\alpha}\|^2 = \|\boldsymbol{\beta}\|^2$, which in turn can be computed from (4.127):

$$\begin{aligned}
\|\boldsymbol{\alpha}\|^2 &= \|\boldsymbol{\beta}\|^2 \\
&= \sum_{i=1}^{n_{\min}} \frac{\left|\left(\mathsf{F}^\dagger \hat{\mathbf{u}}_0\right)_i\right|^2}{\sigma_i^2} \\
&\le \frac{\sum_{i=1}^{n_{\min}} \left|\left(\mathsf{F}^\dagger \hat{\mathbf{u}}_0\right)_i\right|^2}{\sigma_{\min}^2} \\
&= \frac{\|\mathsf{F}^\dagger \hat{\mathbf{u}}_0\|^2}{\sigma_{\min}^2} \\
&= \frac{1}{\sigma_{\min}^2},
\end{aligned} \tag{4.128}$$

where $\sigma_{\min} = \min_{i=1,\ldots,n_{\min}} \sigma_i$. Here the last step follows because $\|\mathsf{F}^\dagger \hat{\mathbf{u}}_0\| = \|\hat{\mathbf{u}}_0\| = 1$. Note that $\sigma_{\min}^2$ is equal to the smallest eigenvalue (which we will denote by $\vartheta_{\min}$) of the matrix $\mathsf{U}\mathsf{U}^\dagger$.

If we do not condition on a specific realization of $\mathbf{U}_{n_{\min}-2}, \ldots, \mathbf{U}_0$, then $\mathbb{U}\mathbb{U}^\dagger$ is a Wishart matrix, and its smallest eigenvalue is a random variable. For future reference, we state its density in the next lemma.

Lemma 4.20. *Let the $\left(n_{\min} \times (n_{\min}+2)\right)$-dimensional random matrix $\mathbb{U}$ have IID, zero-mean, unit-variance, circularly-symmetric, complex Gaussian entries. Then the matrix $\mathbb{U}\mathbb{U}^\dagger$ is a Wishart matrix and the density of its smallest eigenvalue is given by*

$$f(\vartheta_{\min}) = c_{n_{\min},n_{\min}+2}\,\vartheta_{\min}^2 \exp\!\left(-\vartheta_{\min}\frac{n_{\min}}{2}\right) P_{n_{\min},n_{\min}+2}(\vartheta_{\min}), \qquad \vartheta_{\min} \ge 0, \tag{4.129}$$

where $c_{n_{\min},n_{\min}+2}$ is a constant and $P_{n_{\min},n_{\min}+2}(\vartheta_{\min})$ is a polynomial of degree $2\,(n_{\min}+1)$.

Proof. See [9, Thm. 5.4]. □

We return to the analysis of the second term on the RHS of (4.119). We have

$$\begin{aligned}
&h\big(\mathbb{H}_0\mathbf{U}_0 \bigm| \{\mathbb{H}_\ell\mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{-1}, \mathbf{U}_{n_{\min}-2}^{0}\big)\\
&\quad= n_{\min}\mathsf{E}\big[\log\|\mathbf{U}_0\|^2\big] + h\big(\mathbb{H}_0\hat{\mathbf{U}}_0 \bigm| \{\mathbb{H}_\ell\mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{-1}, \mathbf{U}_{n_{\min}-2}^{0}\big)\\
&\quad= n_{\min}\mathsf{E}\big[\log\|\mathbf{U}_0\|^2\big]\\
&\qquad+ h\left(\mathbb{H}_0\hat{\mathbf{U}}_0 - \sum_{\ell=1}^{n_{\min}+2}\alpha_\ell\mathbb{H}_{-\ell}\mathbf{U}_{-\ell} \,\middle|\, \{\mathbb{H}_\ell\mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{-1}, \mathbf{U}_{n_{\min}-2}^{0}\right)\\
&\quad\le n_{\min}\mathsf{E}\big[\log\|\mathbf{U}_0\|^2\big] + h\left(\mathbb{H}_0\hat{\mathbf{U}}_0 - \sum_{\ell=1}^{n_{\min}+2}\alpha_\ell\mathbb{H}_{-\ell}\mathbf{U}_{-\ell}\right)\\
&\quad= n_{\min}\mathsf{E}\big[\log\|\mathbf{U}_0\|^2\big] + h\left(\sum_{\ell=1}^{n_{\min}+2}\alpha_\ell(\mathbb{H}_0-\mathbb{H}_{-\ell})\mathbf{U}_{-\ell}\right),
\end{aligned}\tag{4.130}$$

where the first step follows from the scaling property of differential entropy; the second step follows because differential entropy is invariant under deterministic translation; the third step follows because conditioning cannot increase entropy; and the last step follows from (4.122).

Since the fading is zero-mean, we have

$$\mathsf{E}\left[\sum_{\ell=1}^{n_{\min}+2}\alpha_\ell(\mathbb{H}_0-\mathbb{H}_{-\ell})\mathbf{U}_{-\ell}\right] = 0.$$

Let $\mathbf{H}_0^{(r)}$ denote the r-th row of $\mathbb{H}_0$. Then the r-th component of $\sum_{\ell=1}^{n_{\min}+2}\alpha_\ell(\mathbb{H}_0-\mathbb{H}_{-\ell})\mathbf{U}_{-\ell}$ is given by $\sum_{\ell=1}^{n_{\min}+2}\alpha_\ell\Big(\mathbf{H}_0^{(r)}-\mathbf{H}_{-\ell}^{(r)}\Big)\mathbf{U}_{-\ell}$ and its variance can be upper bounded as follows:

$$\begin{aligned}
&\mathsf{E}\left[\left|\sum_{\ell=1}^{n_{\min}+2}\alpha_\ell\Big(\mathbf{H}_0^{(r)}-\mathbf{H}_{-\ell}^{(r)}\Big)\mathbf{U}_{-\ell}\right|^2\right]\\
&\quad\le \mathsf{E}\left[\sum_{\nu=1}^{n_{\min}+2}|\alpha_\nu|^2\sum_{\ell=1}^{n_{\min}+2}\left|\Big(\mathbf{H}_0^{(r)}-\mathbf{H}_{-\ell}^{(r)}\Big)\mathbf{U}_{-\ell}\right|^2\right]\\
&\quad= \mathsf{E}\left[\|\boldsymbol{\alpha}\|^2\sum_{\ell=1}^{n_{\min}+2}\left|\sum_{t=1}^{n_{\min}}\big(H_0(r,t)-H_{-\ell}(r,t)\big)U_{-\ell}(t)\right|^2\right]
\end{aligned}$$

$$
\begin{aligned}
&\leq \mathsf{E}\left[\|\boldsymbol{\alpha}\|^2 \sum_{\ell=1}^{n_{\min}+2} \sum_{t=1}^{n_{\min}} \left|H_0(r,t) - H_{-\ell}(r,t)\right|^2 \sum_{t'=1}^{n_{\min}} |U_{-\ell}(t')|^2\right] \\
&= \mathsf{E}\left[\|\boldsymbol{\alpha}\|^2 \sum_{\ell=1}^{n_{\min}+2} \sum_{t=1}^{n_{\min}} \mathsf{E}\left[\left|H_0(r,t) - H_{-\ell}(r,t)\right|^2\right] \|\mathbf{U}_{-\ell}\|^2\right] \\
&= \mathsf{E}\left[\|\boldsymbol{\alpha}\|^2 \sum_{\ell=1}^{n_{\min}+2} n_{\min}\, \epsilon^2(\ell)\, \|\mathbf{U}_{-\ell}\|^2\right], \qquad (4.131)
\end{aligned}
$$

where we define

$$
\epsilon^2(\ell) \triangleq \mathsf{E}\left[\left|H_0(r,t) - H_{-\ell}(r,t)\right|^2\right].
$$

(Note that $\epsilon^2(\ell)$ does not depend on (r,t) because the fading is spatially IID.) Here the first step follows from the Cauchy-Schwarz inequality; the second step follows by writing out the scalar product of two vectors; the third step follows again from the Cauchy-Schwarz inequality; the fourth step follows because $\{\mathbb{H}_k,\, k \in \mathbb{Z}\}$ and $\{\mathbf{U}_k,\, k \in \mathbb{Z}\}$ are independent; and the last step follows by defining $\epsilon^2(\ell)$ and by noting that, since the fading is spatially IID, it does not depend on (r,t).

Defining

$$
\begin{aligned}
\epsilon^2_{\max} &\triangleq \max_{\ell=1,\ldots,n_{\min}+2} \epsilon^2(\ell) \\
&= \max_{\ell=1,\ldots,n_{\min}+2} \mathsf{E}\left[\left|H_0(r,t) - H_{-\ell}(r,t)\right|^2\right],
\end{aligned}
$$

we can further upper bound (4.131) for each $r = 1, \ldots, n_{\min}$

$$
\begin{aligned}
&\mathsf{E}\left[\left|\sum_{\ell=1}^{n_{\min}+2} \alpha_\ell \left(\mathbf{H}_0^{(r)} - \mathbf{H}_{-\ell}^{(r)}\right)\mathbf{U}_{-\ell}\right|^2\right] \\
&\qquad \leq \mathsf{E}\left[\|\boldsymbol{\alpha}\|^2 \sum_{\ell=1}^{n_{\min}+2} n_{\min}\, \epsilon^2(\ell)\, \|\mathbf{U}_{-\ell}\|^2\right] \\
&\qquad \leq n_{\min}\, \epsilon^2_{\max}\, \mathsf{E}\left[\|\boldsymbol{\alpha}\|^2 \sum_{\ell=1}^{n_{\min}+2} \|\mathbf{U}_{-\ell}\|^2\right]. \qquad (4.132)
\end{aligned}
$$

Using again the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
&\mathsf{E}\left[\left|\sum_{\ell=1}^{n_{\min}+2}\alpha_\ell\Big(\mathbf{H}_0^{(r)}-\mathbf{H}_{-\ell}^{(r)}\Big)\mathbf{U}_{-\ell}\right|^2\right]\\
&\quad\le n_{\min}\,\epsilon_{\max}^2\sqrt{\mathsf{E}[\|\boldsymbol{\alpha}\|^4]}\sqrt{\mathsf{E}\left[\left(\sum_{\ell=1}^{n_{\min}+2}\|\mathbf{U}_{-\ell}\|^2\right)^2\right]}\\
&\quad\le n_{\min}\,\epsilon_{\max}^2\sqrt{\mathsf{E}\left[\frac{1}{\Theta_{\min}^2}\right]}\sqrt{\mathsf{E}\left[\left(\sum_{\ell=1}^{n_{\min}+2}\|\mathbf{U}_{-\ell}\|^2\right)^2\right]},
\end{aligned}\tag{4.133}$$

where $\Theta_{\min}$ denotes the smallest eigenvalue of the Wishart matrix $\mathbb{U}\mathbb{U}^\dagger$; see Lemma 4.20. Here we use (4.128) to upper bound $\|\boldsymbol{\alpha}\|^4\le 1/\Theta_{\min}^2$.

To evaluate the RHS of (4.133), we first analyze

$$\mathsf{E}\left[\left(\sum_{\ell=1}^{n_{\min}+2}\|\mathbf{U}_{-\ell}\|^2\right)^2\right].$$

To this end, we note that $\sum_{\ell=1}^{n_{\min}+2}\|\mathbf{U}_{-\ell}\|^2$ is of a central chi-square distribution with $2\,(n_{\min}+2)\,n_{\min}$ degrees of freedom, for which the second moment is given by [42, p. 14, Eq. (2.33)]

$$\mathsf{E}\left[\left(\sum_{\ell=1}^{n_{\min}+2}\|\mathbf{U}_{-\ell}\|^2\right)^2\right]=\Big(n_{\min}\,(n_{\min}+2)+1\Big)\,(n_{\min}+2)\,n_{\min}.\tag{4.134}$$

As for $\mathsf{E}\big[1/\Theta_{\min}^2\big]$, we recall that the density of $\Theta_{\min}$ is given by (Lemma 4.20)

$$f(\vartheta_{\min})=c_{n_{\min},n_{\min}+2}\vartheta_{\min}^2\exp\Big(-\vartheta_{\min}\frac{n_{\min}}{2}\Big)P_{n_{\min},n_{\min}+2}(\vartheta_{\min}),$$

so

$$\begin{aligned}
&\mathsf{E}\left[\frac{1}{\Theta_{\min}^2}\right]\\
&=c_{n_{\min},n_{\min}+2}\int_0^\infty\exp\Big(-\vartheta_{\min}\frac{n_{\min}}{2}\Big)P_{n_{\min},n_{\min}+2}(\vartheta_{\min})\,\mathrm{d}\vartheta_{\min}.
\end{aligned}\tag{4.135}$$

Note that (4.135) does only depend on $n_{\min}$ and not on the memory of the fading process. Further note that, since $P_{n_{\min},n_{\min}+2}(\vartheta_{\min})$ is a polynomial, the integral on the RHS of (4.135) is finite.

We return to the analysis of (4.130). Using that the entropy of a random vector of covariance matrix A is upper bounded by the entropy of a Gaussian vector of diagonal covariance matrix with diagonal entries $A(1,1),\ldots,A(n_{\min},n_{\min})$, we obtain

$$
\begin{aligned}
&h\big(\mathbb{H}_0\mathbf{U}_0 \bigm| \{\mathbb{H}_\ell\mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{-1}, \mathbf{U}_{n_{\min}-2}^{0}\big)\\
&\quad\le n_{\min}\mathsf{E}\big[\log\|\mathbf{U}_0\|^2\big] + h\left(\sum_{\ell=1}^{n_{\min}+2}\alpha_\ell(\mathbb{H}_0-\mathbb{H}_{-\ell})\mathbf{U}_{-\ell}\right)\\
&\quad\le n_{\min}\mathsf{E}\big[\log\|\mathbf{U}_0\|^2\big] + n_{\min}\,\log(\pi e)\\
&\qquad+\sum_{r=1}^{n_{\min}}\log\left(\mathsf{E}\left[\left|\sum_{\ell=1}^{n_{\min}+2}\alpha_\ell\Big(\mathbf{H}_0^{(r)}-\mathbf{H}_{-\ell}^{(r)}\Big)\mathbf{U}_{-\ell}\right|^2\right]\right)\\
&\quad\le n_{\min}\mathsf{E}\big[\log\|\mathbf{U}_0\|^2\big] + n_{\min}\,\log(\pi e) + n_{\min}\,\log\epsilon_{\max}^2\\
&\qquad+n_{\min}\,\log\left(n_{\min}\sqrt{\mathsf{E}\left[\frac{1}{\Theta_{\min}^2}\right]}\sqrt{\mathsf{E}\left[\left(\sum_{\ell=1}^{n_{\min}+2}\|\mathbf{U}_{-\ell}\|^2\right)^2\right]}\right), \qquad (4.136)
\end{aligned}
$$

where the last step follows from (4.133).

We summarize the main steps of our proof to obtain the final lower bound on $\chi_{\mathrm{PP}}(\{\mathbb{H}_k\})$

$$
\begin{aligned}
&\chi_{\mathrm{PP}}(\{\mathbb{H}_k\})\\
&\quad\ge \mathsf{E}\big[\log\|\mathbf{V}_0\|^2\big]-\log\big(\mathsf{E}\big[\|\mathbf{V}_0\|^2\big]\big)-\gamma-1\\
&\qquad+\lim_{n\to\infty}\frac{1}{n}I\big(\mathbf{V}_1^n;\{\mathbb{H}_\ell\mathbf{V}_\ell\}_{\ell=1}^n\big)\\
&\quad\ge \mathsf{E}\big[\log\|\mathbf{V}_0\|^2\big]-\log\big(\mathsf{E}\big[\|\mathbf{V}_0\|^2\big]\big)-\gamma-1\\
&\qquad+I\big(\mathbf{V}_0;\{\mathbb{H}_\ell\mathbf{V}_\ell\}_{\ell=-n_{\min}-2}^{0},\mathbf{V}_{n_{\min}-2}^{-1}\big)\\
&\quad\ge \mathsf{E}\big[\log\|\mathbf{U}_0\|^2\big]-\log\big(\mathsf{E}\big[\|\mathbf{U}_0\|^2\big]\big)-\gamma-1\\
&\qquad+I\big(\mathbf{U}_0;\{\mathbb{H}_\ell\mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{0},\mathbf{U}_{n_{\min}-2}^{-1}\big)-3\varepsilon\\
&\quad\ge \mathsf{E}\big[\log\|\mathbf{U}_0\|^2\big]-\log\big(\mathsf{E}\big[\|\mathbf{U}_0\|^2\big]\big)-\gamma-1-3\varepsilon\\
&\qquad+h\big(\mathbb{H}_0\mathbf{U}_0 \bigm| \{\mathbb{H}_\ell\mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{-1},\mathbf{U}_{n_{\min}-2}^{-1}\big)
\end{aligned}
$$

$$
\begin{aligned}
&\quad - h\big(\mathbb{H}_0 \mathbf{U}_0 \bigm| \{\mathbb{H}_\ell \mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{-1}, \mathbf{U}_{n_{\min}-2}^{0}\big) \\
&\ge \mathsf{E}\big[\log \|\mathbf{U}_0\|^2\big] - \log\big(\mathsf{E}\big[\|\mathbf{U}_0\|^2\big]\big) - \gamma - 1 - 3\varepsilon \\
&\quad + n_{\min} \log(\pi e) + \sum_{i=0}^{n_{\min}-1} \psi(n_{\min} - i) \\
&\quad - h\big(\mathbb{H}_0 \mathbf{U}_0 \bigm| \{\mathbb{H}_\ell \mathbf{U}_\ell\}_{\ell=-n_{\min}-2}^{-1}, \mathbf{U}_{n_{\min}-2}^{0}\big) \\
&\ge \mathsf{E}\big[\log \|\mathbf{U}_0\|^2\big] - \log\big(\mathsf{E}\big[\|\mathbf{U}_0\|^2\big]\big) - \gamma - 1 - 3\varepsilon \\
&\quad + n_{\min} \log(\pi e) + \sum_{i=0}^{n_{\min}-1} \psi(n_{\min} - i) \\
&\quad - n_{\min} \mathsf{E}\big[\log \|\mathbf{U}_0\|^2\big] - n_{\min} \log(\pi e) - n_{\min} \log \epsilon_{\max}^2 \\
&\quad - n_{\min} \log \left(n_{\min} \sqrt{\mathsf{E}\left[\frac{1}{\Theta_{\min}^2}\right]} \sqrt{\mathsf{E}\left[\left(\sum_{\ell=1}^{n_{\min}+2} \|\mathbf{U}_{-\ell}\|^2\right)^2\right]} \right) \\
&= n_{\min} \log \frac{1}{\epsilon_{\max}^2} + \Delta(n_{\min}),
\end{aligned}
\tag{4.137}
$$

where we define

$$
\begin{aligned}
\Delta(n_{\min}) \triangleq{}& -(n_{\min} - 1)\mathsf{E}\big[\log \|\mathbf{U}_0\|^2\big] - \log\big(\mathsf{E}\big[\|\mathbf{U}_0\|^2\big]\big) - \gamma - 1 - 3\varepsilon \\
& + \sum_{i=0}^{n_{\min}-1} \psi(n_{\min} - i) \\
& - n_{\min} \log \left(n_{\min} \sqrt{\mathsf{E}\left[\frac{1}{\Theta_{\min}^2}\right]} \sqrt{\mathsf{E}\left[\left(\sum_{\ell=1}^{n_{\min}+2} \|\mathbf{U}_{-\ell}\|^2\right)^2\right]} \right).
\end{aligned}
$$

Here the first step follows from (4.101) and (4.108); the second step follows from (4.110); the third step follows from (4.117) and (4.118); the fourth step follows from (4.119); the fifth step follows from (4.121); and the sixth step follows from (4.136).

This concludes the proof of Theorem 4.6.

4.9 A Gauss-Markov Fading Process

A very simple model for a slowly-varying channel is the Gauss-Markov fading model (see for example [4] or [13]). Here $\{\mathbb{H}_k,\, k \in \mathbb{Z}\}$ is a zero-mean, spatially IID, Gaussian process with

$$\mathbb{H}_k = \sqrt{1-\epsilon^2}\mathbb{H}_{k-1} + \epsilon \mathbb{W}_k, \quad k \in \mathbb{Z}, \tag{4.138}$$

where $\{\mathbb{W}_k,\, k \in \mathbb{Z}\}$ is spatially IID with $\{W_k(r,t),\, k \in \mathbb{Z}\}$ consisting of IID, zero-mean, unit-variance, circularly-symmetric, complex Gaussian random variables. In the above, ϵ^2 is the prediction error in predicting $H_0(r,t)$ from its past.

In the following, we consider a MIMO Gauss-Markov fading channel with $n = n_\mathrm{R} = n_\mathrm{T}$ transmit and receive antennas. Corollary 4.8 yields

$$\chi(\{\mathbb{H}_k\}) \leq n \log \frac{1}{\epsilon^2} + \Psi(n), \tag{4.139}$$

where the correction term $\Psi(n)$ is only a function of n and not of ϵ^2. Theorem 4.6 yields

$$\chi(\{\mathbb{H}_k\}) \geq n \log \frac{1}{\epsilon^2} + \Delta(n) + o(\epsilon^2), \tag{4.140}$$

where the $o(\epsilon^2)$-term vanishes as ϵ^2 tends to zero.

4.10 The Fading Number and Degrees of Freedom

The "number of degrees of freedom" n_min of a system employing n_T transmit antennas and n_R receive antennas is defined by

$$n_\mathrm{min} \triangleq \min\{n_\mathrm{T}, n_\mathrm{R}\}.$$

It plays an important role in the high-SNR asymptotic analysis of coherent MIMO fading channels [43] as well as in the asymptotic analysis of the block-constant fading model [34, 52].

The role of degrees of freedom in noncoherent MIMO fading channels is more subtle. Indeed, if the limit in (4.11) exists, then the asymptotic expansion

$$C(\mathrm{SNR}) \approx \log\bigl(1 + \log(1+\mathrm{SNR})\bigr) + \chi(\{\mathbb{H}_k\}) \tag{4.141}$$

indicates that at very high SNR, when the $\log\log\mathrm{SNR}$ term dominates the fading number χ, capacity grows double-logarithmically in the SNR and the number of transmit and receive antennas hardly influences capacity.

Great care, however, must be exercised when applying this argument. For this argument to demonstrate the irrelevance of the degrees of freedom in determining channel capacity, the SNR must not only be large enough so (4.141) is a good approximation, but it must also be large enough so the $\log\log \mathrm{SNR}$-term dominates the fading number $\chi(\{\mathbb{H}_k\})$. While, as we shall argue, the approximation (4.141) begins to hold at reasonable SNR, for the $\log\log \mathrm{SNR}$-term to dominate the fading number $\chi(\{\mathbb{H}_k\})$, the SNR must be larger than the double exponentiation of the fading number. When the fading number is large, as in slowly-varying channels, this latter condition only begins to hold at extremely high SNR.

What then is the role of degrees of freedom in slowly varying noncoherent communication? For slowly varying channels, degrees of freedom play a key role in determining the fading number. Indeed—at least when $n_\mathrm{R} \leq n_\mathrm{T}$—Theorem 4.6 and Corollary 4.8 combine to prove that for very slowly-varying fading channels the fading number is roughly proportional to $n_{\min}$.

The picture that emerges is thus the following. While the approximation (4.141) is quite reasonable as of relatively moderate SNR, for slowly-varying channels the $\log\log \mathrm{SNR}$ dominates the fading number only at extremely high SNR. At these extremely high SNR, degrees of freedom, indeed, hardly influence capacity. However, increasing the number of degrees of freedom increases the fading number $\chi(\{\mathbb{H}_k\})$ and hence pushes this extremely high-SNR regime further and further away. If we think of the fading number as an indication of the maximal rate at which power efficient communication is achievable on the channel [28], then we can say that for slowly-varying spatially independent Gaussian fading channels this rate is roughly proportional to the number of degrees of freedom. Thus, increasing the number of degrees of freedom increases the practical limit on power-efficient communication over the channel.

The capacity of noncoherent MIMO Gauss-Markov fading channels was also studied by Etkin and Tse [12, 13]. Their results are, in fact, in agreement with the above picture.

Note that our results on the fading number and degrees of freedom are not specific to Gauss-Markov fading. It suffices that the autocorrelation decay slowly and that the difference between the present fading and any fading in the past $n_{\min} + 2$ symbols is of expected squared error that is not much larger than the prediction error based on the infinite past. That is, we require that

$$\log \frac{\epsilon^2_{\max}}{\epsilon^2}$$

be roughly a constant. (Here $\epsilon^2_{\max}$ is defined in Theorem 4.6 and ϵ^2 is the prediction error in predicting the present value of the process $H_k(r,t)$ from its infinite past.) This is certainly the case for Gauss-Markov processes.

Chapter 5

Gaussian Fading is the Worst Fading

5.1 Introduction

We study the capacity of peak-power limited, single-antenna, discrete-time, flat-fading channels with memory. A noncoherent channel model is considered where the transmitter and the receiver are both aware of the law of the fading process, but not of its realization. Our focus is on the capacity at high signal-to-noise ratio (SNR). Specifically, we study the capacity pre-log, which is defined as the limiting ratio of channel capacity to the logarithm of the SNR, as the SNR tends to infinity.

The capacity pre-log of *Gaussian* fading channels was derived in [26] (see also [25]). It was shown that the pre-log is given by the Lebesgue measure of the set of harmonics where the derivative of the spectral distribution function that characterizes the memory of the fading process is zero. To the best of our knowledge, the capacity pre-log of *non-Gaussian* fading channels is unknown.

In this work, we demonstrate that the Gaussian assumption in the analysis of fading channels at high SNR is conservative in the sense that for a large class of fading processes the Gaussian process is the worst. More precisely, we show that among all stationary and ergodic fading processes of a given spectral distribution function and whose law has no mass point at zero, the Gaussian process gives rise to the smallest pre-log.

The rest of this chapter is organized as follows. Section 5.2 describes the channel model. Section 5.3 defines channel capacity and the capacity pre-log. Section 5.4 presents our main results. Section 5.5 provides the proofs of these results. Section 5.6 discusses the extension of our results to multiple-input single-output (MISO) fading channels with memory. And Section 5.7 concludes the chapter with a summary and a discussion of our results.

5.2 Channel Model

We consider a single-antenna flat-fading channel with memory where the time-k ($k \in \mathbb{Z}$) channel output $Y_k \in \mathbb{C}$ corresponding to the time-k channel input $x_k \in \mathbb{C}$ is given by

$$Y_k = H_k x_k + Z_k, \quad k \in \mathbb{Z}. \tag{5.1}$$

Here the random processes $\{Z_k,\, k \in \mathbb{Z}\}$ and $\{H_k,\, k \in \mathbb{Z}\}$ take value in $\mathbb{C}$ and model the additive and multiplicative noises, respectively. It is assumed that these processes are statistically independent and of a joint law that does not depend on the input sequence $\{x_k\}$.

The additive noise $\{Z_k,\, k \in \mathbb{Z}\}$ is a sequence of independent and identically distributed (IID) zero-mean, variance-σ^2, circularly-symmetric, complex Gaussian random variables. The multiplicative noise ("fading") $\{H_k,\, k \in \mathbb{Z}\}$ is a mean-d, unit-variance, stationary and ergodic stochastic process of spectral distribution function $F(\cdot)$, i.e., $\lambda \mapsto F(\lambda)$ is a bounded and nondecreasing function on $[-1/2, 1/2]$ satisfying [8, p. 474, Thm. 3.2]

$$\mathsf{E}\left[(H_{k+m} - d)(H_k - d)^*\right] = \int_{-1/2}^{1/2} e^{\mathrm{i}2\pi m\lambda}\, \mathrm{d}F(\lambda), \quad \big(k \in \mathbb{Z},\, m \in \mathbb{Z}\big). \tag{5.2}$$

Since $F(\cdot)$ is monotonic, it is almost everywhere differentiable, and we denote its derivative by $F'(\cdot)$. (At the discontinuity points of $F(\cdot)$ the derivative $F'(\cdot)$ is undefined.) For example, if the fading process $\{H_k,\, k \in \mathbb{Z}\}$ is IID, then

$$F'(\lambda) = 1, \quad -\frac{1}{2} \le \lambda \le \frac{1}{2}.$$

5.3 Channel Capacity and Capacity Pre-Log

Channel capacity was defined in Section 2.1 as the supremum of all achievable rates. In this chapter we study the capacity under a peak-power constraint A^2 on the inputs, which for our channel (5.1) is given by (see, e.g., [20, Thm. 2] or [39, Sec. II])

$$C(\mathrm{SNR}) = \lim_{n\to\infty} \frac{1}{n} \sup I(X_1^n; Y_1^n), \tag{5.3}$$

where SNR is defined as

$$\mathrm{SNR} \triangleq \frac{\mathsf{A}^2}{\sigma^2}, \tag{5.4}$$

and where the maximization is over all joint distributions on $X_1, \ldots, X_n$ satisfying with probability one

$$|X_k|^2 \leq \mathsf{A}^2, \quad k = 1, \ldots, n. \tag{5.5}$$

The capacity pre-log is defined as [26]

$$\Pi \triangleq \varlimsup_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\log \mathrm{SNR}}. \tag{5.6}$$

For *Gaussian fading*, i.e., when $\{H_k - d, \, k \in \mathbb{Z}\}$ is a circularly-symmetric, complex Gaussian process, the pre-log Π_{G} is given by the Lebesgue measure of the set of harmonics where the derivative of the spectral distribution function is zero, i.e.,

$$\Pi_{\mathrm{G}} = \mu\left(\{\lambda \colon F'(\lambda) = 0\}\right), \tag{5.7}$$

where $\mu(\cdot)$ denotes the Lebesgue measure on the interval $[-1/2, 1/2]$; see [25, 26]. (Here the subscript "G" stands for "Gaussian".)

This result indicates that if the fading process is Gaussian and satisfies

$$\mu\left(\{\lambda \colon F'(\lambda) = 0\}\right) > 0,$$

then the corresponding capacity grows logarithmically in the SNR. Note that otherwise the capacity can increase with the SNR in various ways. For instance, in [28] fading channels are studied that result in a capacity which increases double-logarithmically with the SNR (see also Chapter 4), and in [26] spectral distribution functions are presented for which capacity grows as a fractional power of the logarithm of the SNR.

5.4 Main Result

We show that, among all stationary and ergodic fading processes of a given spectral distribution function and whose law has no mass point at zero, the Gaussian process gives rise to the smallest pre-log. This is made precise in the following theorem.

Theorem 5.1. *Consider a mean-d, unit-variance, stationary, ergodic fading process $\{H_k, k \in \mathbb{Z}\}$ whose spectral distribution function is given by $F(\cdot)$ and whose law satisfies*

$$\Pr(H_k = 0) = 0, \quad k \in \mathbb{Z}.$$

Then the corresponding capacity pre-log Π is lower bounded by

$$\Pi \geq \mu\left(\{\lambda \colon F'(\lambda) = 0\}\right). \tag{5.8}$$

Proof. See Section 5.5.1. □

Note 5.1. *Theorem 5.1 continues to hold if $\{Z_k, k \in \mathbb{Z}\}$ is a sequence of IID, variance-σ^2, complex (not necessarily Gaussian) random variables of finite differential entropy. Thus, among all pairs of fading processes (satisfying the conditions of Theorem 5.1) and additive noise processes, the pair where both processes are Gaussian gives rise to the smallest pre-log.*

The assumption that the law of the fading process has no mass point at zero is essential in the following sense.

Note 5.2. *There exists a mean-d, unit-variance, stationary and ergodic fading process $\{H_k, k \in \mathbb{Z}\}$ of some spectral distribution function $F(\cdot)$ such that*

$$\Pi < \mu\left(\{\lambda \colon F'(\lambda) = 0\}\right). \tag{5.9}$$

By Theorem 5.1, this process must satisfy

$$\Pr(H_k = 0) > 0, \quad k \in \mathbb{Z}.$$

Proof. See Section 5.5.2. □

Note 5.3. *The inequality in* (5.8) *can be strict. For example, consider the* phase-noise channel *with memoryless phase noise. This channel can be viewed as a fading channel where the fading process* $\{H_k,\, k \in \mathbb{Z}\}$ *is given by*

$$H_k = e^{\mathrm{i}\Theta_k}, \quad k \in \mathbb{Z},$$

where $\{\Theta_k,\, k \in \mathbb{Z}\}$ *is IID with* Θ_k *being uniformly distributed over* $[-\pi, \pi)$*. This process gives rise to a pre-log* $\Pi = 1/2$*, whereas the Gaussian fading of equal spectral distribution function yields* $\Pi_{\mathrm{G}} = 0$*.*

Proof. For a derivation of the capacity pre-log of the phase-noise channel see Section 5.5.3. □

5.5 Proofs

This section provides the proofs of our main results. For a proof of Theorem 5.1 see Section 5.5.1, for a proof of Note 5.2 see Section 5.5.2, and for a proof of Note 5.3 see Section 5.5.3.

5.5.1 Proof of Theorem 5.1

To prove Theorem 5.1, we first derive a lower bound on the capacity, and proceed then to analyze its asymptotic growth as the SNR tends to infinity.

Capacity Lower Bound

To derive a lower bound on the capacity we consider inputs $\{X_k,\, k \in \mathbb{Z}\}$ that are IID, zero-mean, and circularly-symmetric, and for which $|X_k|^2$ is uniformly distributed over the interval $[0, \mathrm{A}^2]$. Our derivation is based on the lower bound

$$\frac{1}{n} I(X_1^n; Y_1^n) \geq \frac{1}{n} I(X_1^n; Y_1^n | H_1^n) - \frac{1}{n} I(H_1^n; Y_1^n | X_1^n), \tag{5.10}$$

which follows from the chain rule

$$\begin{aligned} I(X_1^n; Y_1^n) &= I(X_1^n, H_1^n; Y_1^n) - I(H_1^n; Y_1^n | X_1^n) \\ &= I(H_1^n; Y_1^n) + I(X_1^n; Y_1^n | H_1^n) - I(H_1^n; Y_1^n | X_1^n) \end{aligned}$$

and the nonnegativity of mutual information.

We first study the first term on the right-hand side (RHS) of (5.10). Making use of the stationarity of the channel and of the fact that the inputs are IID we have

$$\frac{1}{n} I(X_1^n; Y_1^n | H_1^n) = I(X_1; Y_1 | H_1). \tag{5.11}$$

We lower bound the RHS of (5.11) as follows. For any fixed $\Upsilon > 0$

$$\begin{aligned}
I(X_1; Y_1 | H_1) \\
&= h(H_1 X_1 + Z_1 | H_1) - h(Z_1) \\
&= \int_{|h_1| \geq \Upsilon} h(H_1 X_1 + Z_1 | H_1 = h_1) \, \mathrm{d}P_{H_1}(h_1) \\
&\quad + \int_{|h_1| < \Upsilon} h(H_1 X_1 + Z_1 | H_1 = h_1) \, \mathrm{d}P_{H_1}(h_1) - h(Z_1) \\
&\geq \int_{|h_1| \geq \Upsilon} h(H_1 X_1 + Z_1 | H_1 = h_1) \, \mathrm{d}P_{H_1}(h_1) \\
&\quad + \Pr(|H_1| < \Upsilon) \, h(Z_1) - h(Z_1) \\
&\geq \int_{|h_1| \geq \Upsilon} \bigl(\log |h_1|^2 + h(X_1) \bigr) \, \mathrm{d}P_{H_1}(h_1) \\
&\quad + \Pr(|H_1| < \Upsilon) \, h(Z_1) - h(Z_1) \\
&\geq \Pr(|H_1| \geq \Upsilon) \bigl(\log \Upsilon^2 + h(X_1) \bigr) + \Pr(|H_1| < \Upsilon) \, h(Z_1) - h(Z_1) \\
&= \Pr(|H_1| \geq \Upsilon) \bigl(\log \Upsilon^2 + \log \pi + h(|X_1|^2) \bigr) \\
&\quad + \Pr(|H_1| < \Upsilon) \, h(Z_1) - h(Z_1) \\
&= \Pr(|H_1| \geq \Upsilon) \log \mathrm{A}^2 + \Pr(|H_1| \geq \Upsilon) \log \left(\pi \Upsilon^2 \right) \\
&\quad + \Pr(|H_1| < \Upsilon) \, h(Z_1) - h(Z_1) \\
&= \Pr(|H_1| \geq \Upsilon) \log \mathrm{A}^2 + \Pr(|H_1| \geq \Upsilon) \log \left(\pi \Upsilon^2 \right) \\
&\quad + \bigl(\Pr(|H_1| < \Upsilon) - 1 \bigr) \log(\pi e \sigma^2) \\
&= \Pr(|H_1| \geq \Upsilon) \log \mathrm{SNR} - \Pr(|H_1| \geq \Upsilon) \left(1 - \log \Upsilon^2 \right),
\end{aligned} \tag{5.12}$$

where $P_{H_1}(\cdot)$ denotes the distribution of the fading H_1. Here the third step follows by conditioning the entropy in the second integral on X_1; the fourth step follows by conditioning the entropy in the first integral on Z_1 and by the behavior of differential entropy under scaling [5, Thm. 9.6.4]; the fifth step follows because over the range of integration $|h_1| \geq \Upsilon$ we have $\log |h_1|^2 \geq \log \Upsilon^2$; the sixth step follows because X_1 is circularly-symmetric [28, Lemma 6.16]; the seventh step follows by computing the entropy of a random variable that is uniformly distributed over the interval $\left[0, \mathrm{A}^2\right]$; the eighth step follows by evaluating the entropy of a zero-mean, variance-σ^2, circularly-symmetric, complex Gaussian random variable $h(Z_k) = \log(\pi e \sigma^2)$; and the last step follows from $\Pr(|H_1| \geq \Upsilon) = 1 - \Pr(|H_1| < \Upsilon)$.

We next turn to the second term on the RHS of (5.10). In order to upper bound it we proceed along the lines of [7], but for non-Gaussian fading. Let $\mathbf{Y}$, $\mathbf{H}$, and $\mathbf{Z}$ be the random vectors $(Y_1, \ldots, Y_n)^{\mathsf{T}}$, $(H_1, \ldots, H_n)^{\mathsf{T}}$, and $(Z_1, \ldots, Z_n)^{\mathsf{T}}$, and let X be a diagonal matrix with diagonal entries $x_1, \ldots, x_n$. It follows from (5.1) that

$$\mathbf{Y} = \mathsf{X}\mathbf{H} + \mathbf{Z}. \tag{5.13}$$

The conditional covariance matrix of $\mathbf{Y}$, conditional on $x_1, \ldots, x_n$, is given by

$$\mathsf{E}\left[(\mathbf{Y} - \mathsf{E}[\mathbf{Y}])\,(\mathbf{Y} - \mathsf{E}[\mathbf{Y}])^{\dagger} \,\Big|\, X_1^n = x_1^n \right] = \mathsf{X}\mathsf{K}_{\mathbf{HH}}\mathsf{X}^{\dagger} + \sigma^2 \mathsf{I}_n, \tag{5.14}$$

where

$$\mathsf{K}_{\mathbf{HH}} \triangleq \mathsf{E}\left[(\mathbf{H} - \mathsf{E}[\mathbf{H}])(\mathbf{H} - \mathsf{E}[\mathbf{H}])^{\dagger}\right]. \tag{5.15}$$

Using the entropy maximizing property of circularly-symmetric Gaussian vectors [5, Thm. 9.6.5], we have

$$\begin{aligned}
\frac{1}{n} I(H_1^n; Y_1^n | X_1^n) &= \frac{1}{n} h(Y_1^n | X_1^n) - \frac{1}{n} h(Z_1^n) \\
&\leq \frac{1}{n} \mathsf{E}\left[\log\det\left(\mathsf{I}_n + \frac{1}{\sigma^2} \mathbb{X} \mathsf{K}_{\mathbf{HH}} \mathbb{X}^{\dagger} \right) \right] \\
&= \frac{1}{n} \mathsf{E}\left[\log\det\left(\mathsf{I}_n + \frac{1}{\sigma^2} \mathsf{K}_{\mathbf{HH}} \mathbb{X}^{\dagger} \mathbb{X} \right) \right] \\
&\leq \frac{1}{n} \log\det\left(\mathsf{I}_n + \frac{\mathsf{E}\left[|X_1|^2\right]}{\sigma^2} \mathsf{K}_{\mathbf{HH}} \right) \\
&= \frac{1}{n} \sum_{k=1}^{n} \log\left(1 + \frac{\mathsf{E}\left[|X_1|^2\right]}{\sigma^2} \lambda_k \right) \\
&\leq \frac{1}{n} \sum_{k=1}^{n} \log(1 + \mathrm{SNR}\, \lambda_k),
\end{aligned} \tag{5.16}$$

where $\mathbb{X}$ is a random diagonal matrix with diagonal entries $X_1, \ldots, X_n$, and where $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of $\mathsf{K}_{\mathbf{HH}}$. Here the third step follows from the identity $\det(\mathsf{I}_n + \mathsf{AB}) = \det(\mathsf{I}_n + \mathsf{BA})$; the fourth step follows from Jensen's inequality and by noting that $\{X_k,\, k \in \mathbb{Z}\}$ is IID, so $\mathsf{E}\big[\mathbb{X}^\dagger \mathbb{X}\big] = \mathsf{E}\big[|X_1|^2\big]\,\mathsf{I}_n$; the fifth step follows because the determinant of a matrix is given by the product of its eigenvalues; and the last step follows because, by (5.5), we have $\mathsf{E}\big[|X_1|^2\big] \le \mathsf{A}^2$.

To evaluate the RHS of (5.16) in the limit as n tends to infinity, we apply Szegö's Theorem on the asymptotic behavior of the eigenvalues of Hermitian Toeplitz matrices [18] (see also [41, Thm. 2.7.13]). We obtain

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} I(H_1^n; Y_1^n | X_1^n) &\le \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \log(1 + \mathrm{SNR}\,\lambda_k) \\ &= \int_{-1/2}^{1/2} \log\big(1 + \mathrm{SNR}\, F'(\lambda)\big)\, \mathrm{d}\lambda. \qquad (5.17)\end{aligned}$$

Combining (5.10), (5.11), (5.12), and (5.17) yields the final lower bound

$$\begin{aligned}C(\mathrm{SNR}) \ge\ &\Pr(|H_1| \ge \Upsilon) \log \mathrm{SNR} - \Pr(|H_1| \ge \Upsilon)\left(1 - \log \Upsilon^2\right) \\ &- \int_{-1/2}^{1/2} \log\left(1 + \mathrm{SNR}\, F'(\lambda)\right) \mathrm{d}\lambda, \qquad (5.18)\end{aligned}$$

which holds for any fixed $\Upsilon > 0$. Note that this lower bound applies to all mean-d, unit-variance, stationary and ergodic fading processes $\{H_k,\, k \in \mathbb{Z}\}$ with spectral distribution function $F(\cdot)$, irrespective of whether $\Pr(H_k = 0)$ is zero or not.

Asymptotic Analysis

In the following we prove Theorem 5.1 by computing the limiting ratio of the lower bound (5.18) to $\log \mathrm{SNR}$ as SNR tends to infinity.

We first show that

$$\lim_{\mathrm{SNR}\to\infty} \int_{-1/2}^{1/2} \frac{\log\big(1 + \mathrm{SNR}\, F'(\lambda)\big)}{\log \mathrm{SNR}}\, \mathrm{d}\lambda = \mu\left(\{\lambda \colon F'(\lambda) > 0\}\right). \qquad (5.19)$$

To this end, we divide the integral into three parts, depending on whether λ takes part in the set $\mathcal{S}_1$, $\mathcal{S}_2$, or $\mathcal{S}_3$, where

$$\mathcal{S}_1 \triangleq \{\lambda \in [-1/2, 1/2] \colon F'(\lambda) = 0\} \tag{5.20}$$
$$\mathcal{S}_2 \triangleq \{\lambda \in [-1/2, 1/2] \colon F'(\lambda) \geq 1\} \tag{5.21}$$
$$\mathcal{S}_3 \triangleq \{\lambda \in [-1/2, 1/2] \colon 0 < F'(\lambda) < 1\}. \tag{5.22}$$

For $\lambda \in \mathcal{S}_1$ the integrand is zero and hence

$$\lim_{\text{SNR}\to\infty} \int_{\mathcal{S}_1} \frac{\log\bigl(1 + \text{SNR}\, F'(\lambda)\bigr)}{\log \text{SNR}} \, \mathrm{d}\lambda = 0. \tag{5.23}$$

For $\lambda \in \mathcal{S}_2$, i.e., when $F'(\lambda) \geq 1$, we note that, for sufficiently large SNR, the function

$$\text{SNR} \mapsto \frac{\log\bigl(1 + \text{SNR}\, F'(\lambda)\bigr)}{\log \text{SNR}}$$

is monotonically decreasing in SNR. Therefore, applying the Monotone Convergence Theorem [38, Thm. 1.26], we have

$$\begin{aligned}
&\lim_{\text{SNR}\to\infty} \int_{\mathcal{S}_2} \frac{\log\bigl(1 + \text{SNR}\, F'(\lambda)\bigr)}{\log \text{SNR}} \, \mathrm{d}\lambda \\
&\quad = \int_{\mathcal{S}_2} \lim_{\text{SNR}\to\infty} \frac{\log\bigl(1 + \text{SNR}\, F'(\lambda)\bigr)}{\log \text{SNR}} \, \mathrm{d}\lambda \\
&\quad = \mu\left(\mathcal{S}_2\right) \\
&\quad = \mu\left(\{\lambda \colon F'(\lambda) \geq 1\}\right).
\end{aligned} \tag{5.24}$$

For $\lambda \in \mathcal{S}_3$, i.e., when $0 < F'(\lambda) < 1$, we have

$$0 < \frac{\log\bigl(1 + \text{SNR}\, F'(\lambda)\bigr)}{\log \text{SNR}} < \frac{\log(1 + \text{SNR})}{\log \text{SNR}} \leq \log(1 + e), \quad \text{SNR} \geq e, \tag{5.25}$$

where the last inequality follows because, for sufficiently large SNR, the function

$$\text{SNR} \mapsto \frac{\log(1 + \text{SNR})}{\log \text{SNR}}$$

is monotonically decreasing in SNR. Since $\log(1 + e)$ is integrable over $\mathcal{S}_3$, we can apply the Dominated Convergence Theorem [38, Thm. 1.34] to obtain

$$\begin{aligned}
&\lim_{\mathrm{SNR}\to\infty}\int_{\mathcal{S}_3}\frac{\log\bigl(1+\mathrm{SNR}\,F'(\lambda)\bigr)}{\log\mathrm{SNR}}\,\mathrm{d}\lambda\\
&=\int_{\mathcal{S}_3}\lim_{\mathrm{SNR}\to\infty}\frac{\log\bigl(1+\mathrm{SNR}\,F'(\lambda)\bigr)}{\log\mathrm{SNR}}\,\mathrm{d}\lambda\\
&=\mu\left(\mathcal{S}_3\right)\\
&=\mu\left(\{\lambda\colon 0<F'(\lambda)<1\}\right).
\end{aligned} \tag{5.26}$$

Adding (5.23), (5.24), and (5.26) yields (5.19).

To continue with the asymptotic analysis of (5.18) we note that by (5.19)

$$\begin{aligned}
\Pi &\triangleq \varlimsup_{\mathrm{SNR}\to\infty}\frac{C(\mathrm{SNR})}{\log\mathrm{SNR}}\\
&\geq \Pr(|H_1|\geq\Upsilon)-\mu\left(\{\lambda: F'(\lambda)>0\}\right)\\
&=\mu\left(\{\lambda\colon F'(\lambda)=0\}\right)-\Pr(|H_1|<\Upsilon), \quad \Upsilon>0.
\end{aligned} \tag{5.27}$$

If the law of the fading process has no mass point at zero, then

$$\lim_{\Upsilon\downarrow 0}\Pr(|H_1|<\Upsilon)=0, \tag{5.28}$$

and Theorem 5.1 therefore follows from (5.27) by letting Υ tend to zero from above.

5.5.2 Proof of Note 5.2

We prove Note 5.2 by demonstrating that there exists a stationary and ergodic fading process of some spectral distribution function $F(\cdot)$ for which

$$\Pi<\mu\left(\{\lambda\colon F'(\lambda)=0\}\right).$$

By Theorem 5.1, the law of such a process must have a mass point at zero, i.e.,

$$\Pr(H_k=0)>0, \quad k\in\mathbb{Z}.$$

We first show that the capacity pre-log is upper bounded by

$$\Pi\leq \Pr(|H_1|>0). \tag{5.29}$$

Indeed, the capacity $C(\mathrm{SNR})$ does not decrease when the receiver additionally knows the realization of $\{H_k,\, k \in \mathbb{Z}\}$ and when the inputs have to satisfy an average-power constraint rather than a peak-power constraint, i.e.,

$$C(\mathrm{SNR}) \leq \lim_{n\to\infty} \frac{1}{n} \sup I(X_1^n; Y_1^n | H_1^n), \tag{5.30}$$

where the maximization is over all input distributions on $X_1, \ldots, X_n$ satisfying the average-power constraint

$$\frac{1}{n}\sum_{k=1}^{n} \frac{\mathsf{E}\big[|X_k|^2\big]}{\sigma^2} \leq \mathrm{SNR}. \tag{5.31}$$

(This follows because the availability of additional information cannot decrease the capacity, and because any distribution on the inputs satisfying the peak-power constraint (5.5) satisfies also (5.31).) It is well known that the expression on the RHS of (5.30) is equal to

$$\lim_{n\to\infty} \frac{1}{n} \sup I(X_1^n; Y_1^n | H_1^n) = \mathsf{E}\big[\log(1+|H_1|^2\,\mathrm{SNR})\big] \tag{5.32}$$

(see, e.g., [2, Eq. (3.3.10)]), which can be further upper bounded by

$$\begin{aligned}
&\mathsf{E}\big[\log(1+|H_1|^2\,\mathrm{SNR})\big] \\
&\quad = \Pr(|H_1|>0)\ \mathsf{E}\big[\log(1+|H_1|^2\,\mathrm{SNR})\,\big|\,|H_1|>0\big] \\
&\quad \leq \Pr(|H_1|>0)\ \log\big(1+\mathsf{E}\big[|H_1|^2\,\big|\,|H_1|>0\big]\ \mathrm{SNR}\big) \\
&\quad = \Pr(|H_1|>0)\ \log\left(1+\frac{\mathrm{SNR}}{\Pr(|H_1|>0)}\right).
\end{aligned} \tag{5.33}$$

Here the first step follows by writing the expectation as

$$\begin{aligned}
&\mathsf{E}\big[\log(1+|H_1|^2\,\mathrm{SNR})\big] \\
&\quad = \Pr(|H_1|=0)\ \mathsf{E}\big[\log(1+|H_1|^2\,\mathrm{SNR})\,\big|\,|H_1|=0\big] \\
&\qquad + \Pr(|H_1|>0)\ \mathsf{E}\big[\log(1+|H_1|^2\,\mathrm{SNR})\,\big|\,|H_1|>0\big]
\end{aligned}$$

and by noting that $\mathsf{E}\big[\log(1+|H_1|^2\,\mathrm{SNR})\,\big|\,|H_1|=0\big] = 0$; the second step follows from Jensen's inequality; and the last step follows because $\mathsf{E}\big[|H_1|^2\big]=1$, which implies that

$$\mathsf{E}\big[|H_1|^2\,\big|\,|H_1|>0\big] = \frac{1}{\Pr(|H_1|>0)}.$$

Dividing the RHS of (5.33) by $\log \mathrm{SNR}$ and computing the limit as SNR tends to infinity yields (5.29).

In view of (5.29), it suffices to demonstrate that there exists a fading process of some spectral distribution function $F(\cdot)$ that satisfies

$$\Pr(|H_1| > 0) < \mu\left(\{\lambda : F'(\lambda) = 0\}\right). \tag{5.34}$$

A first attempt of defining such a process (which, alas, does not work) is

$$\ldots, H_{-1}, H_0, H_1, \ldots = \begin{cases} \ldots, 0, 0, 0, \ldots & \text{with prob. } \delta \\ \ldots, B_{-1}, B_0, B_1, \ldots & \text{with prob. } 1-\delta, \end{cases} \tag{5.35}$$

where $\{B_k,\, k \in \mathbb{Z}\}$ is a zero-mean, circularly-symmetric, stationary and ergodic, complex Gaussian process of variance $1/(1-\delta)$ and of spectral distribution function $G(\cdot)$; and where δ and $G(\cdot)$ are chosen so that

$$1 - \delta < \mu\left(\{\lambda : G'(\lambda) = 0\}\right). \tag{5.36}$$

This process satisfies (5.34) because $\Pr(|H_1| > 0) = 1-\delta$, and because

$$\begin{aligned} &\mathsf{E}\left[(H_{k+m} - d)(H_k - d)^*\right] \\ &\qquad = (1-\delta)\,\mathsf{E}\left[B_{k+m} B_k^*\right], \quad \big(k \in \mathbb{Z},\, m \in \mathbb{Z}\big) \end{aligned} \tag{5.37}$$

which implies that $F(\lambda) = (1-\delta)G(\lambda)$ almost everywhere, so

$$\mu\left(\{\lambda : F'(\lambda) = 0\}\right) = \mu\left(\{\lambda : G'(\lambda) = 0\}\right). \tag{5.38}$$

Alas, the above fading process is stationary but not ergodic.

In the following, we exhibit a fading process that is stationary and ergodic and that satisfies (5.34). Let

$$\ldots, A_{-1}, A_0, A_1, A_2, \ldots = \begin{cases} \ldots, 0, 1, 0, 1, \ldots & \text{with prob. } \frac{1}{2} \\ \ldots, 1, 0, 1, 0, \ldots & \text{with prob. } \frac{1}{2}, \end{cases} \tag{5.39}$$

and let $\{B_k,\, k \in \mathbb{Z}\}$ be a zero-mean, variance-2, circularly-symmetric, stationary and ergodic, complex Gaussian process of spectral distribution function $G(\cdot)$. Furthermore let $\{A_k,\, k \in \mathbb{Z}\}$ and $\{B_k,\, k \in \mathbb{Z}\}$ be independent of each other. Let the fading process be given by

$$H_k = A_k \, B_k, \quad k \in \mathbb{Z}. \tag{5.40}$$

Note that $\{H_k, \, k \in \mathbb{Z}\}$ is of zero mean, and its law satisfies

$$\Pr(|H_k| > 0) = \Pr(A_k = 1) = \frac{1}{2}, \quad k \in \mathbb{Z}. \tag{5.41}$$

We first argue that $\{H_k, \, k \in \mathbb{Z}\}$ is stationary and ergodic. Indeed, $\{A_k, \, k \in \mathbb{Z}\}$ is stationary and ergodic. And since a Gaussian process is ergodic if, and only if, it is weakly-mixing (see, e.g., [39, Sec. II]), we have that $\{B_k, \, k \in \mathbb{Z}\}$ is stationary and weakly-mixing. (See [35, Sec. 2.6] for a definition of weakly-mixing stochastic processes.) It thus follows from [3, Prop. 1.6] that the process $\{(A_k, B_k), \, k \in \mathbb{Z}\}$ is jointly stationary and ergodic, which implies that

$$\{H_k, \, k \in \mathbb{Z}\} = \{A_k \cdot B_k, \, k \in \mathbb{Z}\},$$

is stationary and ergodic.

We next demonstrate that $G(\cdot)$ can be chosen so that $\{H_k, \, k \in \mathbb{Z}\}$ satisfies (5.34). We choose

$$G'(\lambda) = \begin{cases} \dfrac{1}{\mathsf{W}}, & \text{if } |\lambda| \le \mathsf{W} \\ 0, & \text{otherwise} \end{cases} \tag{5.42}$$

for some $\mathsf{W} \in (0, 1/8)$, which corresponds to the autocovariance function

$$\mathsf{E}[B_{k+m} B_k^*] = 2 \operatorname{sinc}(2\mathsf{W}m), \quad \big(k \in \mathbb{Z}, \, m \in \mathbb{Z}\big).$$

Here $\operatorname{sinc}(\cdot)$ denotes the sinc-function, i.e., $\operatorname{sinc}(x) = \sin(\pi x)/(\pi x)$ for $|x| > 0$ and $\operatorname{sinc}(0) = 1$. Using that

$$\mathsf{E}[A_{k+m} A_k^*] = \frac{1}{2} \, \mathrm{I}\{m \text{ is even}\}, \quad \big(k \in \mathbb{Z}, \, m \in \mathbb{Z}\big),$$

we have for the autocovariance function of $\{H_k, \, k \in \mathbb{Z}\}$

$$\begin{aligned} \mathsf{E}[H_{k+m} H_k^*] &= \mathsf{E}[A_{k+m} B_{k+m} A_k^* B_k^*] \\ &= \mathsf{E}[A_{k+m} A_k^*] \, \mathsf{E}[B_{k+m} B_k^*] \\ &= \mathrm{I}\{m \text{ is even}\} \operatorname{sinc}(2\mathsf{W}m), \quad \big(k \in \mathbb{Z}, \, m \in \mathbb{Z}\big), \end{aligned} \tag{5.43}$$

and the corresponding spectrum is given by

$$F'(\lambda) = \begin{cases} \frac{1}{4W}, & \text{if } |\lambda| \le W \text{ or } \frac{1}{2} - W \le |\lambda| \le \frac{1}{2} \\ 0, & \text{otherwise.} \end{cases} \tag{5.44}$$

We thus have

$$\mu\left(\{\lambda\colon F'(\lambda) = 0\}\right) = 1 - 4W, \tag{5.45}$$

and it follows from (5.41) and (5.45) that

$$\Pr(|H_k| > 0) < \mu\left(\{\lambda\colon F'(\lambda) = 0\}\right), \quad \text{for } W < \frac{1}{8}.$$

Thus there exist stationary and ergodic fading processes of some spectral distribution function that give rise to a capacity pre-log that is strictly smaller than the pre-log of a Gaussian fading channel of equal spectral distribution function.

5.5.3 Proof of Note 5.3

To prove Note 5.3, we first notice that, since the phase noise is memoryless, the derivative of the spectral distribution function is

$$F'(\lambda) = 1, \quad -\frac{1}{2} \le \lambda \le \frac{1}{2}.$$

Hence the capacity pre-log of the Gaussian fading channel of spectral distribution function $F(\cdot)$ equals

$$\Pi_{\mathrm{G}} = \mu\left(\{\lambda\colon F'(\lambda) = 0\}\right) = 0. \tag{5.46}$$

It remains to show that the pre-log of the phase-noise channel with memoryless phase noise is equal to

$$\Pi = \frac{1}{2}. \tag{5.47}$$

In [24] it was shown that at high SNR the capacity of the phase-noise channel under an average-power constraint on the inputs is given by

$$C_{\mathrm{Avg}}(\mathrm{SNR}) = \frac{1}{2}\log\left(1 + \frac{\mathrm{SNR}}{2}\right) + o(1), \tag{5.48}$$

where $o(1)$ tends to zero as SNR tends to zero. (The subscript "Avg" indicates that the inputs satisfy an average-power constraint and not a peak-power constraint.) Since any distribution on the inputs satisfying the peak-power constraint (5.5) satisfies also the average-power constraint, it follows that $C(\text{SNR}) \leq C_{\text{Avg}}(\text{SNR})$ and hence

$$\Pi \leq \frac{1}{2}. \tag{5.49}$$

To prove (5.47) it thus suffices to show that $\Pi \geq \frac{1}{2}$. To this end, we first note that, since the phase noise is memoryless, we have

$$C(\text{SNR}) = \sup I(X_1; Y_1), \tag{5.50}$$

where the maximization is over all distributions on X_1 satisfying with probability one

$$|X_1| \leq \mathsf{A}.$$

We derive a lower bound on $C(\text{SNR})$ by evaluating the RHS of (5.50) for X_1 being a zero-mean, circularly-symmetric, complex random variable with $|X_1|^2$ uniformly distributed over the interval $\left[0, \mathsf{A}^2\right]$. We have

$$\begin{aligned} I(X_1; Y_1) &\geq I\big(X_1; |Y_1|^2\big) \\ &= h\big(|Y_1|^2\big) - h\big(|Y_1|^2 \bigm| X_1\big) \\ &\geq h\big(|X_1|^2\big) - h\big(|Y_1|^2 \bigm| X_1\big), \end{aligned} \tag{5.51}$$

where the first step follows from the data processing inequality [5, Thm. 2.8.1]; and the last step follows by the circular symmetry of X_1 [24, p. 3, after Eq. (20)].

Computing the differential entropy of a uniformly distributed random variable, the first term on the RHS of (5.51) becomes

$$h\big(|X_1|^2\big) = \log \mathsf{A}^2. \tag{5.52}$$

As to the second term, we note that, for a given $X_1 = x_1$, the random variable $2/\sigma^2\,|Y_1|^2$ has a noncentral chi-square distribution with noncentrality parameter $2/\sigma^2\,|x_1|^2$ and two degrees of freedom. Its differential entropy can be upper bounded by [24, Eq. (8)]

$$\begin{aligned} h\big(|Y_1|^2 \mid X_1\big) &\le \frac{1}{2}\mathsf{E}\left[\log\left(4\pi e\Big(2+2\frac{2}{\sigma^2}|X_1|^2\Big)\right)\right] - \log\frac{2}{\sigma^2} \\ &\le \frac{1}{2}\log\left(4\pi e\Big(2+2\frac{2}{\sigma^2}\mathsf{A}^2\Big)\right) - \log\frac{2}{\sigma^2}, \end{aligned} \tag{5.53}$$

where the last step follows because $|X_1| \le \mathsf{A}$ with probability one. Combining (5.52) and (5.53) with (5.51) yields thus

$$I(X_1;Y_1) \ge \frac{1}{2}\log\mathrm{SNR} + o(\log\mathrm{SNR}), \tag{5.54}$$

where

$$\lim_{\mathrm{SNR}\to\infty}\frac{o(\log\mathrm{SNR})}{\log\mathrm{SNR}} = 0.$$

We finally obtain the lower bound

$$\Pi \ge \frac{1}{2}$$

upon dividing the RHS of (5.54) by $\log\mathrm{SNR}$ and letting SNR tend to infinity.

5.6 Extension to MISO Fading Channels

Theorem 5.1 can be extended to MISO fading channels with memory, when the fading processes corresponding to the different transmit antennas are independent. For such channels, the channel output $Y_k \in \mathbb{C}$ at time k corresponding to the channel input $\mathbf{x}_k \in \mathbb{C}^{n_\mathrm{T}}$ is given by

$$Y_k = \mathbf{H}_k^{\mathsf{T}}\mathbf{x}_k + Z_k, \quad k \in \mathbb{Z}, \tag{5.55}$$

where $\mathbf{H}_k = \big(H_k(1),\dots,H_k(n_\mathrm{T})\big)^{\mathsf{T}}$, and where the processes

$$\big\{H_k(1),\, k\in\mathbb{Z}\big\}, \big\{H_k(2),\, k\in\mathbb{Z}\big\} \dots, \big\{H_k(n_\mathrm{T}),\, k\in\mathbb{Z}\big\}$$

are jointly stationary ergodic and independent. We assume that for each $t = 1,\dots,n_\mathrm{T}$ the process $\big\{H_k(t),\, k\in\mathbb{Z}\big\}$ is of mean d_t, of unit variance, and of spectral distribution function $F_t(\cdot)$. We further assume that

$$\Pr(H_k(1) = 0) = \Pr(H_k(2) = 0) = \ldots = \Pr(H_k(n_\mathrm{T}) = 0) = 0, \ k \in \mathbb{Z}.$$

The additive noise $\{Z_k, \, k \in \mathbb{Z}\}$ is defined as in Section 5.2.

The capacity of this channel is given by (5.3), but with X_1^n replaced by $\mathbf{X}_1^n$, and with the peak-power constraint (5.5) altered accordingly:

$$\|\mathbf{X}_k\|^2 \leq \mathsf{A}^2 \quad \text{with probability one}, \quad k \in \mathbb{Z}. \tag{5.56}$$

The pre-log of MISO fading channels is defined as in the single-antenna case (5.6). For Gaussian fading, i.e., when the n_T processes $\{H(t) - d_t, \, k \in \mathbb{Z}\}$ are circularly-symmetric complex Gaussian, the pre-log is given by (Corollary 4.15)

$$\Pi_\mathrm{G} = \max_{1 \leq t \leq n_\mathrm{T}} \mu\left(\{\lambda \colon F_t'(\lambda) = 0\}\right). \tag{5.57}$$

Proving that the capacity pre-log Π of MISO fading channels is lower bounded by the pre-log of the MISO Gaussian fading channel of equal spectral distribution functions—namely $F_1(\cdot), \ldots, F_{n_\mathrm{T}}(\cdot)$—is straightforward. Let Π_t, $1 \leq t \leq n_\mathrm{T}$ denote the capacity pre-log of a single-antenna fading channel with fading process $\left\{H_k(t), \, k \in \mathbb{Z}\right\}$, and let

$$t_\star = \arg \max_{1 \leq t \leq n_\mathrm{T}} \Pi_t.$$

By signaling from antenna $t_\star$ while keeping the other antennas silent, we can achieve the pre-log $\Pi_{t_\star}$, so

$$\Pi \geq \max_{1 \leq t \leq n_\mathrm{T}} \Pi_t. \tag{5.58}$$

Theorem 5.1 yields then that

$$\Pi_t \geq \mu\left(\{\lambda \colon F_t'(\lambda) = 0\}\right), \quad 1 \leq t \leq n_\mathrm{T}, \tag{5.59}$$

which together with (5.58) proves the claim

$$\Pi \geq \max_{1 \leq t \leq n_\mathrm{T}} \mu\left(\{\lambda \colon F_t'(\lambda) = 0\}\right). \tag{5.60}$$

5.7 Conclusion

We showed that, among all stationary and ergodic fading processes of a given spectral distribution function and whose law has no mass point at zero, the Gaussian process gives rise to the smallest capacity pre-log. We further showed that if the fading law is allowed to have a mass point at zero, then the above statement is not necessarily true anymore. Roughly speaking, we can say that for a large class of fading processes the Gaussian process is the worst. This demonstrates the robustness of the Gaussian assumption in the analysis of fading channels at high SNR.

To give an intuition why Gaussian processes give rise to the smallest pre-log, we recall that for Gaussian fading [26, Eqs. (33) & (47)]

$$C(\mathrm{SNR}) = \log \frac{1}{\epsilon^2_{\mathrm{pred}}(1/\mathrm{SNR})} + o(\log \mathrm{SNR}),$$

where $\epsilon^2_{\mathrm{pred}}(\cdot)$ denotes the mean-square error in predicting the present fading H_0 from a noisy observation of its past (see Section 4.3). Thus for Gaussian fading the capacity pre-log is determined by $\epsilon^2_{\mathrm{pred}}(1/\mathrm{SNR})$, and it is plausible that also the pre-log of non-Gaussian fading channels is connected with the ability of predicting the present fading from a noisy observation of its past. Since, among all stationary and ergodic processes of a given spectral distribution function, the Gaussian process is hardest to predict, it is therefore plausible that the Gaussian process gives rise to the smallest pre-log.

Chapter 6

Multipath Fading Channels

6.1 Introduction

In this chapter, we study the capacity of discrete-time *multipath fading channels*. In multipath fading channels, the transmitted signal propagates along a multitude of paths, and the gains and delays of these paths vary over time. In general, the path delays differ from each other, and the receiver thus observes a weighted sum of delayed replicas of the transmitted signal, where the weights are random. We shall slightly abuse nomenclature and refer to each summand in the received signal as a path, and to the corresponding weight as its path gain, even if it is in fact composed of a multitude of paths. We consider a *noncoherent* channel model, where transmitter and receiver are cognizant of the statistics of the path gains, but are ignorant of their realization.

Multipath fading channels arise in wireless communication, where obstacles in the surroundings reflect the transmitted signal and force it to propagate along multiple paths, and where relative movements of transmitter, receiver, and obstacles lead to time-variations of the path gains and delays. Examples of wireless communication scenarios where the receiver observes typically more than one path include *radio communication* (particularly if the transmitted signal is of large bandwidth as, for example, in *Ultra-Wideband* or in *CDMA*) and *underwater acoustic communication*.

The capacity of noncoherent multipath fading channels has been investigated extensively in the wideband regime, where the signal-to-noise ratio (SNR) is typically small. It was shown by Kennedy that, in the limit as the available bandwidth tends to infinity, the capacity of the fading channel is the same as the capacity of the additive white Gaussian noise (AWGN) channel of equal received power; see [14, Sec. 8.6] and references therein.

To the best of our knowledge, not much is known about the capacity of noncoherent multipath fading channels at high SNR. For the special case of noncoherent *frequency-flat* fading channels (where we only have *one* path), it was shown by Lapidoth and Moser [28] that if the fading process is of finite entropy rate, then at high SNR capacity grows double-logarithmically in the SNR (see also Chapter 4). This is much slower than the logarithmic growth of the AWGN capacity [40].

We study the high-SNR behavior of the capacity of noncoherent *multipath* fading channels (where the number of paths is typically greater than one). We demonstrate that the capacity of such channels does not merely grow slower with the SNR than the capacity of the AWGN channel, but may be even *bounded* in the SNR. In other words, for such channels the capacity does not necessarily tend to infinity as the SNR tends to infinity.

We derive a necessary and a sufficient condition for the capacity to be bounded in the SNR. We show that if the variances of the path gains decay *exponentially or slower*, then the capacity is bounded in the SNR. In contrast, if the variances of the path gains decay *faster than exponentially*, then the capacity is unbounded in the SNR. We further show that if the number of paths is finite, then at high SNR capacity increases double-logarithmically with the SNR, and the capacity pre-loglog, which is defined as the limiting ratio of the capacity to $\log\log \mathrm{SNR}$ as SNR tends to infinity, is 1, irrespective of the number of paths.

The rest of this chapter is organized as follows. Section 6.2 describes the channel model. Section 6.3 is devoted to channel capacity. Section 6.4 summarizes our main results. Sections 6.5 and 6.6 derive the upper bounds and the lower bounds on the capacity, respectively, that are used to prove these results. Section 6.7 concludes with a brief summary and a discussion of our results.

6.2 Channel Model

We consider a discrete-time multipath fading channel whose channel output $Y_k \in \mathbb{C}$ at time $k \in \mathbb{N}$ corresponding to the time-1 through time-k channel inputs $x_1, \ldots, x_k \in \mathbb{C}$ is given by

$$Y_k = \sum_{\ell=0}^{k-1} H_k^{(\ell)} x_{k-\ell} + Z_k, \quad k \in \mathbb{N}. \tag{6.1}$$

Here $\{Z_k,\, k \in \mathbb{Z}\}$ models additive noise, and $H_k^{(\ell)}$ denotes the time-k gain of the ℓ-th path. We assume that $\{Z_k,\, k \in \mathbb{Z}\}$ is a sequence of independent and identically distributed (IID), zero-mean, variance-σ^2, circularly-symmetric, complex Gaussian random variables. For each path $\ell \in \mathbb{N}_0$, we assume that $\{H_k^{(\ell)},\, k \in \mathbb{Z}\}$ is a zero-mean complex stationary process. We denote its variance and its differential entropy rate by

$$\alpha_\ell \triangleq \mathsf{E}\left[\left|H_k^{(\ell)}\right|^2\right], \quad \ell \in \mathbb{N}_0 \tag{6.2}$$

and

$$h_\ell \triangleq \lim_{n\to\infty} \frac{1}{n} h\Big(H_1^{(\ell)}, \ldots, H_n^{(\ell)}\Big), \quad \ell \in \mathbb{N}_0. \tag{6.3}$$

We shall say that the channel has a *finite number of paths*, if for some finite integer $\mathsf{L} \in \mathbb{N}_0$

$$H_k^{(\ell)} = 0, \quad \big(\ell > \mathsf{L},\ k \in \mathbb{N}\big). \tag{6.4}$$

We assume that $\alpha_0 > 0$. We further assume that

$$\sup_{\ell \in \mathbb{N}_0} \alpha_\ell < \infty \tag{6.5}$$

and

$$\inf_{\ell \in \mathcal{L}} h_\ell > -\infty, \tag{6.6}$$

where the set $\mathcal{L}$ is defined as $\mathcal{L} \triangleq \{\ell \in \mathbb{N}_0 : \alpha_\ell > 0\}$. (In Chapter 4 we referred to processes that satisfy condition (6.6) as regular. When the path gains are Gaussian, then this condition is equivalent to saying that the mean-square error in predicting the present path gain from its past is strictly positive, i.e., that the present path gain cannot be predicted perfectly from its past.) We finally assume that the processes

$$\{H_k^{(0)},\, k \in \mathbb{Z}\}, \{H_k^{(1)},\, k \in \mathbb{Z}\}, \ldots$$

are independent ("uncorrelated scattering"); that they are jointly independent of $\{Z_k,\, k \in \mathbb{Z}\}$; and that the joint law of

$$\Big(\{Z_k,\, k \in \mathbb{Z}\}, \{H_k^{(0)},\, k \in \mathbb{Z}\}, \{H_k^{(1)},\, k \in \mathbb{Z}\}, \ldots\Big)$$

does not depend on the input sequence $\{x_k\}$. We consider a noncoherent channel model where the transmitter and the receiver are aware of the statistics of $\{H_k^{(\ell)},\, k \in \mathbb{Z}\}$, $\ell \in \mathbb{N}_0$, but not of their realizations. We do not assume that the path gains are Gaussian.

The channel (6.1) is akin to the channel (3.8) studied in Chapter 3. Indeed, in the Gaussian case, i.e., when the path gains in (6.1) as well as the noise $\{U_k,\, k \in \mathbb{Z}\}$ in (3.8) are Gaussian processes, the heating-up channel (3.8) can be viewed as a real-valued version of the multipath channel (6.1)—except that in (3.8) the gain of the shortest path is deterministic rather than a stochastic process.

6.3 Channel Capacity

We study the information capacity of the above channel (6.1) under an average-power constraint on the inputs, which is defined as (2.4)

$$C_{\text{Info}}(\text{SNR}) \triangleq \varliminf_{n\to\infty} \frac{1}{n} \sup I\big(X_1^n; Y_1^n\big), \tag{6.7}$$

where the supremum is over all joint distributions on $X_1, \ldots, X_n$ satisfying the power constraint

$$\frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\left[|X_k|^2\right] \le \mathsf{P}, \tag{6.8}$$

and where SNR is defined as

$$\text{SNR} \triangleq \frac{\mathsf{P}}{\sigma^2}. \tag{6.9}$$

Recall that, by Fano's inequality, no rate above $C_{\text{Info}}(\text{SNR})$ is achievable, i.e., we have $C(\text{SNR}) \leq C_{\text{Info}}(\text{SNR})$, where $C(\text{SNR})$ denotes the capacity under the input constraint (6.8). (Recall that the capacity was defined in Section 2.1 as the supremum of all achievable rates.) The information capacity $C_{\text{Info}}(\text{SNR})$ is achievable, for example, if the number of paths is finite, and if the processes corresponding to these paths $\{H_k^{(0)}, k \in \mathbb{Z}\}, \ldots, \{H_k^{(\text{L})}, k \in \mathbb{Z}\}$ are jointly ergodic [20, Thm. 2].

Note 6.1. *The results in this chapter do not change if* (6.8) *is replaced by the peak-power constraint* (2.2)*. Indeed, all upper bounds are derived under the average-power constraint* (6.8)*, while all lower bounds are derived under the peak-power constraint* (2.2)*. Since any distribution on the inputs satisfying* (2.2) *satisfies also* (6.8)*, it follows that all bounds derived in this chapter hold irrespective of whether an average-power constraint or a peak-power constraint is imposed.*

The special case of noncoherent frequency-flat fading channels (where we have only one path) was studied by Lapidoth and Moser [28] (see also (4.12)). They showed that if the fading process $\{H_k^{(0)}, k \in \mathbb{Z}\}$ is ergodic, then the capacity satisfies [28, Thm. 4.41]

$$\lim_{\text{SNR}\to\infty} \{C(\text{SNR}) - \log\log \text{SNR}\} = \log \pi + \mathsf{E}\left[\log\left|H_1^{(0)}\right|^2\right] - h_0. \quad (6.10)$$

Thus, at high SNR, the capacity of noncoherent frequency-flat fading channels grows double-logarithmically with the SNR. Lapidoth and Moser concluded that communicating over noncoherent frequency-flat fading channels at high SNR is extremely power-inefficient, as one should expect to square the SNR for every additional bit per channel use.[1]

In this chapter, we show *inter alia* that communicating over noncoherent multipath fading channels at high SNR is not merely power-inefficient, but may be even worse: if the delay spread is large in the sense that the sequence $\{\alpha_\ell\}$ (which describes the variances of the path gains) decays exponentially or slower, then the capacity is bounded in

[1] Note that the capacity of coherent fading channels (where the fading realization is known to the receiver) grows logarithmically with the SNR [11]. Thus in the coherent case it suffices to double the SNR for every additional bit per channel use.

the SNR. For such channels, capacity does not tend to infinity as the SNR tends to infinity. The main results of this chapter are presented in the following section.

6.4 Main Results

Our main results are a sufficient and a necessary condition on $\{\alpha_\ell\}$ for $C_{\text{Info}}(\text{SNR})$ to be bounded in SNR, as well as a characterization of the capacity pre-loglog when the number of paths is finite. Since $C(\text{SNR}) \leq C_{\text{Info}}(\text{SNR})$, it follows that any condition that implies that $C_{\text{Info}}(\text{SNR})$ is bounded implies also that $C(\text{SNR})$ is bounded.

Theorem 6.1. *Consider the above channel model. Then*

$$(i) \quad \left(\varliminf_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} > 0 \right) \quad \Longrightarrow \quad \left(\sup_{\text{SNR}>0} C(\text{SNR}) < \infty \right) \tag{6.11}$$

$$(ii) \quad \left(\lim_{\ell\to\infty} \frac{1}{\ell} \log \frac{1}{\alpha_\ell} = \infty \right) \quad \Longrightarrow \quad \left(\sup_{\text{SNR}>0} C_{\text{Info}}(\text{SNR}) = \infty \right), \tag{6.12}$$

where we define $a/0 \triangleq \infty$ for every $a > 0$ and $0/0 \triangleq 0$.

Proof. Part (i) is proven in Section 6.5.1, and Part (ii) is proven in Sections 6.6.1 and 6.6.2. □

By noting that

$$\left(\lim_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} = 0 \right) \quad \Longrightarrow \quad \left(\lim_{\ell\to\infty} \frac{1}{\ell} \log \frac{1}{\alpha_\ell} = 0 \right)$$

we obtain from Theorem 6.1 the immediate corollary:

Corollary 6.2. *Consider the above channel model. Then*

$$(i) \quad \left(\varliminf_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} > 0 \right) \quad \Longrightarrow \quad \left(\sup_{\text{SNR}>0} C(\text{SNR}) < \infty \right) \tag{6.13}$$

$$(ii) \quad \left(\lim_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} = 0 \right) \quad \Longrightarrow \quad \left(\sup_{\text{SNR}>0} C_{\text{Info}}(\text{SNR}) = \infty \right), \tag{6.14}$$

where we define $a/0 \triangleq \infty$ for every $a > 0$ and $0/0 \triangleq 0$.

For example, if

$$\alpha_\ell = e^{-\ell}, \quad \ell \in \mathbb{N}_0, \tag{6.15}$$

then

$$\lim_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} = \frac{1}{e} > 0 \tag{6.16}$$

and it follows from Part (i) of Corollary 6.2 that the capacity is bounded in the SNR. On the other hand, if

$$\alpha_\ell = \exp\bigl(-\ell^\kappa\bigr), \quad \ell \in \mathbb{N}_0, \quad \text{for some } \kappa > 1, \tag{6.17}$$

then

$$\lim_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} = \lim_{\ell\to\infty} \exp\bigl(\ell^\kappa - (\ell+1)^\kappa\bigr) = 0 \tag{6.18}$$

and it follows from Part (ii) of Corollary 6.2 that the (information) capacity is unbounded in the SNR. Roughly speaking, we can say that when $\{\alpha_\ell\}$ decays *exponentially or slower*, $C_{\text{Info}}(\text{SNR})$ (and hence also $C(\text{SNR})$) is bounded in SNR, and when $\{\alpha_\ell\}$ decays *faster than exponentially*, $C_{\text{Info}}(\text{SNR})$ is unbounded in SNR.

The condition on the left-hand side (LHS) of (6.14) is certainly satisfied if the channel has a finite number of paths, as in this case

$$H_k^{(\ell)} = 0, \quad \bigl(\ell > \mathsf{L},\ k \in \mathbb{N}\bigr),$$

which implies

$$\alpha_\ell = 0, \quad \ell > \mathsf{L} \qquad \text{and} \qquad \frac{\alpha_{\ell+1}}{\alpha_\ell} = \frac{0}{0} \triangleq 0, \quad \ell > \mathsf{L}.$$

Moreover, it was shown that in this case the information capacity is achievable, i.e., $C(\text{SNR}) = C_{\text{Info}}(\text{SNR})$ [20, Thm. 2]. Consequently, it follows from Corollary 6.2 that if the number of paths is finite, then $C(\text{SNR})$ is unbounded in SNR. However, for this case the high-SNR behavior of the capacity can be characterized more accurately: Theorem 6.3 ahead shows that if the number of paths is finite, then the capacity pre-loglog, which is defined as

$$\Lambda \triangleq \varlimsup_{\text{SNR}\to\infty} \frac{C(\text{SNR})}{\log\log \text{SNR}}, \tag{6.19}$$

is 1, irrespective of the number of paths. The pre-loglog in this case is thus the same as for frequency-flat fading.

Theorem 6.3. *Consider the above channel model. Further assume that the number of paths is finite. Then, irrespective of the number of paths, the capacity pre-loglog is given by*

$$\Lambda = \lim_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\log\log \mathrm{SNR}} = 1. \tag{6.20}$$

Proof. See Section 6.5.2 for the converse and Sections 6.6.1 and 6.6.3 for the direct part. □

When studying multipath fading channels at low or at moderate SNR, it is often assumed that the channel has a finite number of paths, even if the number of paths is in reality infinite. This assumption is commonly justified by saying that only the first $(\mathsf{L}+1)$ paths are relevant, since the variances of the remaining paths are typically small and hence the influence of these paths on the capacity is marginal. As we see from Theorems 6.1 and 6.3, this argument is not valid anymore when studying multipath fading channels at high SNR. In fact, when for example the sequence of variances $\{\alpha_\ell\}$ decays exponentially, then according to Part (i) of Theorem 6.1 the capacity is bounded in the SNR. However, if we consider only the first $(\mathsf{L}+1)$ paths and set the other paths to zero, then it follows from Theorem 6.3 that, irrespective of L, the capacity increases double-logarithmically with the SNR. Thus, even though the variances of the remaining paths α_ℓ, $\ell > \mathsf{L}$ can be made arbitrarily small by choosing L sufficiently large, these paths have a significant influence on the capacity behavior at high SNR.

The reason why paths with a small variance can affect the capacity behavior is that the capacity depends on the variance of the product between the path gains and the transmitted signal, and not on the variance of the path gains only. Since at high SNR the variance of $\sum_{\ell=\mathsf{L}+1}^{\infty} H_k^{(\ell)} X_{k-\ell}$ might be huge even if the variance of $\sum_{\ell=\mathsf{L}+1}^{\infty} H_k^{(\ell)}$ is small, the relevance of a path is determined not only by its own variance, but also by the power available at the transmitter. The number of paths that are needed to approximate a multipath channel typically depends on the SNR and may grow to infinity as the SNR tends to infinity.

In order to prove the above results, we derive upper and lower bounds on the capacity. Since these bounds may also be of independent interest, we summarize them in the following propositions.

Proposition 6.4 (Upper Bounds).

(i) Consider the above channel model. Further assume that for some $0 < \rho < 1$ *and some* $\ell_0 \in \mathbb{N}$

$$\alpha_{\ell_0} > 0 \qquad \text{and} \qquad \frac{\alpha_{\ell+1}}{\alpha_\ell} \geq \rho, \quad \ell \geq \ell_0.$$

Then the capacity $C(\text{SNR})$ *is upper bounded by*

$$C(\text{SNR}) \leq \log \frac{2\pi^2}{\sqrt{\tilde{\rho}}} - \inf_{\ell \in \mathcal{L}} (h_\ell - \log \alpha_\ell), \tag{6.21}$$

where

$$\tilde{\rho} = \min\left\{\rho^{\ell_0 - 1} \frac{\alpha_{\ell_0}}{\max_{0 \leq \ell' < \ell_0} \alpha_{\ell'}}, \rho^{\ell_0}\right\}. \tag{6.22}$$

(ii) Consider the above channel model. Further assume that

$$\sum_{\ell=0}^{\infty} \alpha_\ell < \infty. \tag{6.23}$$

Then

$$\varlimsup_{\text{SNR} \to \infty} \left\{C(\text{SNR}) - \log \log \text{SNR}\right\} \leq 1 + \log \pi - \inf_{\ell \in \mathcal{L}} (h_\ell - \log \alpha_\ell). \tag{6.24}$$

Proof. Part (i) is proven in Section 6.5.1 and Part (ii) is proven in Section 6.5.2. □

For example, if $\{\alpha_\ell\}$ is a geometric sequence, i.e.,

$$\alpha_\ell = \rho^\ell, \quad \ell \in \mathbb{N}_0, \qquad \text{for some } 0 < \rho < 1,$$

and if the path gains are Gaussian and memoryless, so

$$h_\ell = \log(\pi e \alpha_\ell), \quad \ell \in \mathbb{N}_0,$$

then Part (i) of Proposition 6.4 yields

$$C(\mathrm{SNR}) \leq \log \frac{2\pi}{\sqrt{\rho}} - 1. \tag{6.25}$$

Part (ii) of Proposition 6.4 combines with (6.10) to show that the pre-loglog of a multipath fading channel can never be larger than the pre-loglog of a frequency-flat fading channel. This result is consistent with the intuition that at high SNR the multipath behavior is detrimental.

Our last result is a lower bound on the capacity. This bound is the basis for the proof of Part (ii) of Theorem 6.1 and for the direct part of Theorem 6.3.

Proposition 6.5 (Lower Bound). *Consider the above channel model. Further assume that*

$$\sum_{\ell=0}^{\infty} \alpha_\ell < \infty. \tag{6.26}$$

Let $\mathsf{L}(\mathsf{P}) \in \mathbb{N}$ *be some positive integer that satisfies*

$$\sum_{\ell=\mathsf{L}(\mathsf{P})+1}^{\infty} \alpha_\ell \, \mathsf{P} \leq \sigma^2 \tag{6.27}$$

(typically $\mathsf{L}(\mathsf{P})$ *depends on* P*), and let* $\tau \in \mathbb{N}$ *be some arbitrary positive integer that is allowed to depend on* $\mathsf{L}(\mathsf{P})$*. Then the information capacity is lower bounded by*

$$C_{\mathrm{Info}}(\mathrm{SNR}) \geq \frac{\tau}{\mathsf{L}(\mathsf{P})+\tau} \log\log \mathsf{P}^{1/\tau} + \frac{\tau}{\mathsf{L}(\mathsf{P})+\tau} \Upsilon, \quad \mathsf{P} > 1, \tag{6.28}$$

where

$$\Upsilon \triangleq \mathsf{E}\left[\log\left|H_1^{(0)}\right|^2\right] - 1 - 2\log\left(\sqrt{\alpha_0} + \sqrt{\alpha + 2\sigma^2}\right). \tag{6.29}$$

Proof. See Section 6.6.1. □

6.5 Proofs of the Upper Bounds

In this section, we establish a proof of Proposition 6.4, which in turn will be used to prove Part (i) of Theorem 6.1 and the converse to Theorem 6.3.

Section 6.5.1 proves Part (i) of Proposition 6.4 and demonstrates that Part (i) of Theorem 6.1 follows immediately from this result. Section 6.5.2 proves Part (ii) of Proposition 6.4. This part provides an upper bound on the capacity pre-loglog and will be used later, together with a capacity lower bound that is derived in Section 6.6, to establish Theorem 6.3.

6.5.1 Bounded Capacity

We provide a proof of Part (i) of Proposition 6.4 by deriving an upper bound on channel capacity that holds under the assumption that for some $0 < \rho < 1$ and some $\ell_0 \in \mathbb{N}_0$

$$\alpha_{\ell_0} > 0 \qquad \text{and} \qquad \frac{\alpha_{\ell+1}}{\alpha_\ell} \geq \rho, \quad \ell \geq \ell_0. \tag{6.30}$$

As this bound is finite for SNR ≥ 0, Part (i) of Theorem 6.1 follows immediately from Part (i) of Proposition 6.4 by noting that if

$$\varliminf_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} > 0,$$

then we can find a $0 < \rho < 1$ and an $\ell_0 \in \mathbb{N}$ satisfying (6.30).

The proof of the desired upper bound is akin to the proof of the upper bound derived in Section 3.6.1. (However, in Chapter 3 we studied a channel with real-valued inputs and outputs, while here we study a channel with complex-valued inputs and outputs.) It is based on (2.6), (6.7), and on an upper bound on $\frac{1}{n}I(X_1^n; Y_1^n)$. To this end, we begin with the chain rule for mutual information [5, Thm. 2.5.2]

$$\begin{aligned}\frac{1}{n}I(X_1^n; Y_1^n) = {} & \frac{1}{n}\sum_{k=1}^{\ell_0} I\bigl(X_1^n; Y_k \bigm| Y_1^{k-1}\bigr) \\ & + \frac{1}{n}\sum_{k=\ell_0+1}^{n} I\bigl(X_1^n; Y_k \bigm| Y_1^{k-1}\bigr).\end{aligned} \tag{6.31}$$

Each term in the first sum on the right-hand side (RHS) of (6.31) is upper bounded by

$$\begin{aligned} I\big(X_1^n; Y_k \mid Y_1^{k-1}\big) &\le h(Y_k) - h\Big(Y_k \Bigm| Y_1^{k-1}, X_1^n, H_k^{(0)}, \dots, H_k^{(k-1)}\Big) \\ &\le \log\left(\pi e\left(\sigma^2 + \sum_{\ell=0}^{k-1} \alpha_\ell \mathsf{E}\left[|X_{k-\ell}|^2\right]\right)\right) - \log\left(\pi e \sigma^2\right) \\ &\le \log\Big(1 + \sup_{\ell \in \mathbb{N}_0} \alpha_\ell \, n \, \mathrm{SNR}\Big), \end{aligned} \tag{6.32}$$

where the first inequality follows because conditioning cannot increase differential entropy; the second inequality follows from the entropy maximizing property of Gaussian random variables [5, Thm. 9.6.5]; and the last inequality follows by upper bounding

$$\alpha_\ell \le \sup_{\ell' \in \mathbb{N}_0} \alpha_{\ell'}, \quad \ell = 0, \dots, k-1$$

and from the power constraint (6.8).

For $k = \ell_0 + 1, \dots, n$, we upper bound $I(X_1^n; Y_k | Y_1^{k-1})$ using the general upper bound for mutual information (Theorem 2.1)

$$I(X;Y) \le \int D\big(W(\cdot|x) \,\|\, R(\cdot)\big) \,\mathrm{d}Q(x). \tag{6.33}$$

For any given $Y_1^{k-1} = y_1^{k-1}$, we choose the output distribution $R(\cdot)$ to be of density

$$\frac{\sqrt{\beta}}{\pi^2 |y_k|} \frac{1}{1 + \beta |y_k|^2}, \quad y_k \in \mathbb{C}, \tag{6.34}$$

with $\beta = 1/(\tilde{\rho}|y_{k-\ell_0}|^2)$ and[2]

$$\tilde{\rho} = \min\left\{\rho^{\ell_0 - 1} \frac{\alpha_{\ell_0}}{\max_{0 \le \ell' < \ell_0} \alpha_{\ell'}}, \rho^{\ell_0}\right\}. \tag{6.35}$$

With this choice

$$0 < \tilde{\rho} < 1 \qquad \text{and} \qquad \tilde{\rho}\, \alpha_\ell \le \alpha_{\ell + \ell_0}, \quad \ell \in \mathbb{N}_0. \tag{6.36}$$

[2] When $y_{k-\ell_0} = 0$, the density (6.34) is undefined. However, this event is of zero probability and has therefore no impact on the mutual information $I\big(X_1^n; Y_k \mid Y_1^{k-1}\big)$.

Using (6.34) in (6.33), and averaging over Y_1^{k-1}, we obtain

$$\begin{aligned}
&I\big(X_1^n; Y_k \,\big|\, Y_1^{k-1}\big) \\
&\quad\le \frac{1}{2}\mathsf{E}\big[\log|Y_k|^2\big] + \frac{1}{2}\mathsf{E}\big[\log\big(\tilde{\rho}|Y_{k-\ell_0}|^2\big)\big] + \mathsf{E}\left[\log\left(1+\frac{|Y_k|^2}{\tilde{\rho}|Y_{k-\ell_0}|^2}\right)\right] \\
&\qquad - h\big(Y_k \,\big|\, X_1^n, Y_1^{k-1}\big) + \log\pi^2 \\
&\quad= \frac{1}{2}\mathsf{E}\big[\log|Y_k|^2\big] - \frac{1}{2}\mathsf{E}\big[\log|Y_{k-\ell_0}|^2\big] + \mathsf{E}\big[\log\big(\tilde{\rho}|Y_{k-\ell_0}|^2 + |Y_k|^2\big)\big] \\
&\qquad - h\big(Y_k \,\big|\, X_1^n, Y_1^{k-1}\big) + \log\frac{\pi^2}{\sqrt{\tilde{\rho}}}.
\end{aligned} \tag{6.37}$$

We bound the third and the fourth term in (6.37) individually. We begin with

$$\begin{aligned}
&\mathsf{E}\big[\log\big(\tilde{\rho}|Y_{k-\ell_0}|^2 + |Y_k|^2\big)\big] \\
&\quad= \mathsf{E}\Big[\mathsf{E}\big[\log\big(\tilde{\rho}|Y_{k-\ell_0}|^2 + |Y_k|^2\big) \,\big|\, X_1^k\big]\Big] \\
&\quad\le \mathsf{E}\Big[\log\Big(\tilde{\rho}\mathsf{E}\big[|Y_{k-\ell_0}|^2 \,\big|\, X_1^k\big] + \mathsf{E}\big[|Y_k|^2 \,\big|\, X_1^k\big]\Big)\Big] \\
&\quad= \mathsf{E}\left[\log\left((1+\tilde{\rho})\sigma^2 + \sum_{\ell=0}^{k-\ell_0-1}\tilde{\rho}\,\alpha_\ell|X_{k-\ell_0-\ell}|^2 + \sum_{\ell=0}^{k-1}\alpha_\ell|X_{k-\ell}|^2\right)\right] \\
&\quad\le \mathsf{E}\left[\log\left(2\sigma^2 + \sum_{\ell=0}^{k-\ell_0-1}\alpha_{\ell+\ell_0}|X_{k-\ell_0-\ell}|^2 + \sum_{\ell=0}^{k-1}\alpha_\ell|X_{k-\ell}|^2\right)\right] \\
&\quad= \mathsf{E}\left[\log\left(2\sigma^2 + \sum_{\ell'=\ell_0}^{k-1}\alpha_{\ell'}|X_{k-\ell'}|^2 + \sum_{\ell=0}^{k-1}\alpha_\ell|X_{k-\ell}|^2\right)\right] \\
&\quad\le \log 2 + \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=0}^{k-1}\alpha_\ell|X_{k-\ell}|^2\right)\right],
\end{aligned} \tag{6.38}$$

where the second step follows from Jensen's inequality; the third step follows by evaluating the conditional expectations; the fourth step follows from (6.36); the fifth step follows by substituting $\ell' = \ell + \ell_0$; and the sixth step follows because with probability one

$$\sum_{\ell=\ell_0}^{k-1}\alpha_\ell|X_{k-\ell}|^2 \le \sum_{\ell=0}^{k-1}\alpha_\ell|X_{k-\ell}|^2.$$

Next we derive a lower bound on $h\big(Y_k \big| X_1^n, Y_1^{k-1}\big)$. Let

$$\Big\{H_{k'}^{(\ell)}\Big\}_{k'=1}^{k-1} = \Big(H_1^{(\ell)}, \ldots, H_{k-1}^{(\ell)}\Big), \quad \ell \in \mathbb{N}_0, \tag{6.39}$$

and let

$$\mathbf{H}_1^{k-1} \triangleq \left(\Big\{H_{k'}^{(0)}\Big\}_{k'=1}^{k-1}, \ldots, \Big\{H_{k'}^{(k-1)}\Big\}_{k'=1}^{k-1}\right). \tag{6.40}$$

We have

$$\begin{aligned} h\big(Y_k \bigm| X_1^n, Y_1^{k-1}\big) &\geq h\big(Y_k \bigm| X_1^n, Y_1^{k-1}, \mathbf{H}_1^{k-1}\big) \\ &= h\big(Y_k \bigm| X_1^n, \mathbf{H}_1^{k-1}\big), \end{aligned} \tag{6.41}$$

where the inequality follows because conditioning cannot increase differential entropy; and where the equality follows because, conditional on $\big(X_1^n, \mathbf{H}_1^{k-1}\big)$, Y_k is independent of Y_1^{k-1}. Let $\mathcal{S}_k$ be defined as

$$\mathcal{S}_k \triangleq \{0 \leq \ell < k : |x_{k-\ell}|^2\, \alpha_\ell > 0\}. \tag{6.42}$$

Using the entropy power inequality [5, Thm. 16.6.3], and using that the processes

$$\big\{H_k^{(0)},\, k \in \mathbb{Z}\big\}, \big\{H_k^{(1)},\, k \in \mathbb{Z}\big\}, \ldots$$

are independent and jointly independent of X_1^n, it is shown in Appendix C that for any given $X_1^n = x_1^n$

$$\begin{aligned} &h\left(\sum_{\ell=0}^{k-1} H_k^{(\ell)} X_{k-\ell} + Z_k \;\middle|\; X_1^n = x_1^n, \mathbf{H}_1^{k-1}\right) \\ &\quad \geq \log\left(\sum_{\ell\in\mathcal{S}_k} e^{h\left(H_k^{(\ell)} X_{k-\ell} \middle| X_{k-\ell}=x_{k-\ell}, \left\{H_{k'}^{(\ell)}\right\}_{k'=1}^{k-1}\right)} + e^{h(Z_k)}\right). \end{aligned} \tag{6.43}$$

We lower bound the differential entropies on the RHS of (6.43) as follows. The differential entropies in the sum are lower bounded by

$$\begin{aligned} &h\left(H_k^{(\ell)} X_{k-\ell} \;\middle|\; X_{k-\ell} = x_{k-\ell}, \Big\{H_{k'}^{(\ell)}\Big\}_{k'=1}^{k-1}\right) \\ &\quad = \log\big(\alpha_\ell |x_{k-\ell}|^2\big) + h\left(H_k^{(\ell)} \;\middle|\; \Big\{H_{k'}^{(\ell)}\Big\}_{k'=1}^{k-1}\right) - \log\alpha_\ell \\ &\quad \geq \log\big(\alpha_\ell |x_{k-\ell}|^2\big) + \inf_{\ell\in\mathcal{L}} \big(h_\ell - \log\alpha_\ell\big), \quad \ell \in \mathcal{S}_k, \end{aligned} \tag{6.44}$$

where the equality follows from the behavior of differential entropy under scaling [5, Thm. 9.6.4]; and where the inequality follows from the stationarity of the process $\{H_k^{(\ell)}, k \in \mathbb{Z}\}$, which implies that the conditional differential entropy

$$h\Bigl(H_k^{(\ell)} \Bigm| \bigl\{H_{k'}^{(\ell)}\bigr\}_{k'=1}^{k-1}\Bigr), \quad \ell \in \mathcal{S}_k$$

cannot be smaller than the differential entropy rate h_ℓ [5, Thms. 4.2.1 & 4.2.2], and by lower bounding $(h_\ell - \log \alpha_\ell)$ by $\inf_{\ell\in\mathcal{L}}(h_\ell - \log \alpha_\ell)$ (which holds for each $\ell \in \mathcal{S}_k$ because $\mathcal{S}_k \subseteq \mathcal{L}$). The last differential entropy on the RHS of (6.43) is lower bounded by

$$h(Z_k) = \log(\pi e \sigma^2) \geq \inf_{\ell\in\mathcal{L}}\bigl(h_\ell - \log \alpha_\ell\bigr) + \log \sigma^2, \tag{6.45}$$

which follows because conditioning cannot increase differential entropy, and because Gaussian random variables maximize differential entropy:

$$\begin{aligned}\inf_{\ell\in\mathcal{L}}\bigl(h_\ell - \log \alpha_\ell\bigr) &\leq \inf_{\ell\in\mathcal{L}}\Bigl(h\Bigl(H_k^{(\ell)}\Bigr) - \log \alpha_\ell\Bigr) \\ &\leq \inf_{\ell\in\mathcal{L}}\bigl(\log(\pi e \alpha_\ell) - \log \alpha_\ell\bigr) \\ &= \log(\pi e). \end{aligned} \tag{6.46}$$

Applying (6.44) and (6.45) to (6.43), and averaging over X_1^n, yields

$$\begin{aligned} &h\bigl(Y_k \bigm| X_1^n, Y_1^{k-1}\bigr) \\ &\quad\geq \mathsf{E}\left[\log\left(\sum_{\ell\in\mathcal{S}_k} \alpha_\ell |X_{k-\ell}|^2 e^{\inf_{\ell\in\mathcal{L}}(h_\ell - \log \alpha_\ell)} + \sigma^2 e^{\inf_{\ell\in\mathcal{L}}(h_\ell - \log \alpha_\ell)}\right)\right] \\ &\quad= \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=0}^{k-1} \alpha_\ell |X_{k-\ell}|^2\right)\right] + \inf_{\ell\in\mathcal{L}}\bigl(h_\ell - \log \alpha_\ell\bigr). \end{aligned} \tag{6.47}$$

Returning to the analysis of (6.37), we obtain from (6.38) and (6.47)

$$\begin{aligned} &I\bigl(X_1^n; Y_k \bigm| Y_1^{k-1}\bigr) \\ &\quad\leq \frac{1}{2}\mathsf{E}\bigl[\log |Y_k|^2\bigr] - \frac{1}{2}\mathsf{E}\bigl[\log |Y_{k-\ell_0}|^2\bigr] \\ &\qquad\quad + \log 2 + \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=0}^{k-1} \alpha_\ell |X_{k-\ell}|^2\right)\right] \end{aligned}$$

$$- \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=0}^{k-1} \alpha_\ell |X_{k-\ell}|^2\right)\right] - \inf_{\ell \in \mathcal{L}} \big(h_\ell - \log \alpha_\ell\big) + \log \frac{\pi^2}{\sqrt{\tilde{\rho}}}$$
$$= \frac{1}{2}\mathsf{E}\big[\log |Y_k|^2\big] - \frac{1}{2}\mathsf{E}\big[\log |Y_{k-\ell_0}|^2\big] + \mathsf{K}, \tag{6.48}$$

where K is defined as

$$\mathsf{K} \triangleq \log \frac{2\pi^2}{\sqrt{\tilde{\rho}}} - \inf_{\ell \in \mathcal{L}} \big(h_\ell - \log \alpha_\ell\big). \tag{6.49}$$

Applying (6.48) and (6.32) to (6.31), we have

$$\begin{aligned} &\frac{1}{n} I(X_1^n; Y_1^n) \\ &\quad \le \frac{1}{n} \sum_{k=1}^{\ell_0} \log\Big(1 + \sup_{\ell \in \mathbb{N}_0} \alpha_\ell \, n \, \mathrm{SNR}\Big) \\ &\qquad + \frac{1}{n} \sum_{k=\ell_0+1}^{n} \left(\frac{1}{2}\mathsf{E}\big[\log |Y_k|^2\big] - \frac{1}{2}\mathsf{E}\big[\log |Y_{k-\ell_0}|^2\big] + \mathsf{K}\right) \\ &\quad = \frac{\ell_0}{n} \log\Big(1 + \sup_{\ell \in \mathbb{N}_0} \alpha_\ell \, n \, \mathrm{SNR}\Big) + \frac{n-\ell_0}{n} \mathsf{K} \\ &\qquad + \frac{1}{2n} \sum_{k=\ell_0+1}^{n} \Big(\mathsf{E}\big[\log |Y_k|^2\big] - \mathsf{E}\big[\log |Y_{k-\ell_0}|^2\big]\Big). \end{aligned} \tag{6.50}$$

To show that the RHS of (6.50) is bounded in the SNR, we use that, for any sequences $\{a_k\}$ and $\{b_k\}$,

$$\sum_{k=\ell_0+1}^{n} (a_k - b_k) = \sum_{k=n-\ell_0+1}^{n} (a_k - b_{k-n+2\ell_0}) + \sum_{k=\ell_0+1}^{n-\ell_0} (a_k - b_{k+\ell_0}). \tag{6.51}$$

Defining

$$a_k \triangleq \mathsf{E}\big[\log |Y_k|^2\big] \tag{6.52}$$

and

$$b_k \triangleq \mathsf{E}\big[\log |Y_{k-\ell_0}|^2\big] \tag{6.53}$$

we have for the first sum on the RHS of (6.51)

$$\begin{aligned}
&\sum_{k=n-\ell_0+1}^{n} (a_k - b_{k-n+2\ell_0}) \\
&\quad = \sum_{k=n-\ell_0+1}^{n} \Big(\mathsf{E}\big[\log|Y_k|^2\big] - \mathsf{E}\big[\log|Y_{k-n+\ell_0}|^2\big]\Big) \\
&\quad \le \sum_{k=n-\ell_0+1}^{n} \Big(\log \mathsf{E}\big[|Y_k|^2\big] - \mathsf{E}\big[\log|Y_{k-n+\ell_0}|^2\big]\Big) \\
&\quad \le \sum_{k=n-\ell_0+1}^{n} \left(\log\Big(\sigma^2 + \sup_{\ell\in\mathbb{N}_0} \alpha_\ell\, n\, \mathsf{P}\Big) - \mathsf{E}\big[\log|Y_{k-n+\ell_0}|^2\big]\right) \\
&\quad \le \sum_{k=n-\ell_0+1}^{n} \left(\log\Big(\sigma^2 + \sup_{\ell\in\mathbb{N}_0} \alpha_\ell\, n\, \mathsf{P}\Big) - \mathsf{E}\big[\log|Z_{k-n+\ell_0}|^2\big]\right) \\
&\quad = \ell_0 \log\Big(1 + \sup_{\ell\in\mathbb{N}_0} \alpha_\ell\, n\, \mathrm{SNR}\Big) + \ell_0\gamma, \qquad (6.54)
\end{aligned}$$

where $\gamma \approx 0.577$ denotes Euler's constant. Here the second step follows from Jensen's inequality; the third step follows by upper bounding

$$\mathsf{E}\big[|Y_k|^2\big] = \sigma^2 + \sum_{\ell=0}^{k-1} \alpha_\ell \mathsf{E}\big[|X_{k-\ell}|^2\big] \le \sigma^2 + \sup_{\ell\in\mathbb{N}_0} \alpha_\ell\, n\, \mathsf{P};$$

the fourth step follows by noting that, conditional on $\sum_{\ell=0}^{k-n+\ell_0-1} H_{k-n+\ell_0}^{(\ell)} X_{k-n+\ell_0-\ell} = \zeta$, the random variable $|Y_{k-n+\ell_0}|^2$ is of a Gaussian law of mean ζ and variance σ^2, so $|Y_{k-n+\ell_0}|^2$ is stochastically larger than $|Z_{k-n+\ell_0}|^2$ [28, Lemma 6.2b)] and hence

$$\begin{aligned}
&\mathsf{E}\left[\log|Y_{k-n+\ell_0}|^2 \,\middle|\, \sum_{\ell=0}^{k-n+\ell_0-1} H_{k-n+\ell_0}^{(\ell)} X_{k-n+\ell_0-\ell} = \zeta\right] \\
&\qquad \ge \mathsf{E}\left[\log|Z_{k-n+\ell_0}|^2 \,\middle|\, \sum_{\ell=0}^{k-n+\ell_0-1} H_{k-n+\ell_0}^{(\ell)} X_{k-n+\ell_0-\ell} = \zeta\right]
\end{aligned}$$

from which we obtain the lower bound

$$\mathsf{E}\big[\log|Y_{k-n+\ell_0}|^2\big] \ge \mathsf{E}\big[\log|Z_{k-n+\ell_0}|^2\big]$$

upon averaging over $\sum_{\ell=0}^{k-n+\ell_0-1} H_{k-n+\ell_0}^{(\ell)} X_{k-n+\ell_0-\ell}$ (see [28, Sec. VI–B] on stochastic ordering); and the last step follows by evaluating the expected logarithm of an exponentially distributed random variable of mean σ^2, i.e., $\mathsf{E}\big[\log|Z_{k-n+\ell_0}|^2\big] = \log\sigma^2 - \gamma$.

For the second sum on the RHS of (6.51) we have

$$\sum_{k=\ell_0+1}^{n-\ell_0} (a_k - b_{k+\ell_0}) = \sum_{k=\ell_0+1}^{n-\ell_0} \Big(\mathsf{E}\big[\log|Y_k|^2\big] - \mathsf{E}\big[\log|Y_k|^2\big]\Big) = 0. \quad (6.55)$$

Thus, applying (6.51)–(6.55) to (6.50), yields

$$\frac{1}{n} I(X_1^n; Y_1^n) \leq \frac{3\ell_0}{2n} \log\Big(1 + \sup_{\ell\in\mathbb{N}_0} \alpha_\ell \, n \, \mathrm{SNR}\Big) + \frac{n-\ell_0}{n}\mathsf{K} + \frac{\ell_0}{2n}\gamma, \quad (6.56)$$

which tends to

$$\mathsf{K} = \log\frac{2\pi^2}{\sqrt{\tilde{\rho}}} - \inf_{\ell\in\mathcal{L}}\big(h_\ell - \log\alpha_\ell\big)$$

as n tends to infinity. This proves Part (i) of Proposition 6.4.

6.5.2 Unbounded Capacity

We prove Part (ii) of Proposition 6.4 by deriving an upper bound on the capacity that holds under the assumption (6.26), namely,

$$\sum_{\ell=0}^{\infty} \alpha_\ell < \infty.$$

From this upper bound follows that

$$\varlimsup_{\mathrm{SNR}\to\infty} \big\{C(\mathrm{SNR}) - \log\log\mathrm{SNR}\big\} < \infty, \quad (6.57)$$

which in turn shows that the capacity pre-loglog is upper bounded by

$$\Lambda \triangleq \varlimsup_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\log\log\mathrm{SNR}} \leq 1. \quad (6.58)$$

This yields the converse to Theorem 6.3.

As in Section 6.5.1, the desired upper bound is based on (2.6), (6.7), and on an upper bound on $\frac{1}{n}I(X_1^n;Y_1^n)$. To this end, we begin with the chain rule for mutual information

$$I\bigl(X_1^n;Y_1^n\bigr) = \sum_{k=1}^{n} I\bigl(X_1^n;Y_k \bigm| Y_1^{k-1}\bigr) \tag{6.59}$$

and upper bound each summand on the RHS of (6.59) using [28, Eq. (27)]

$$\begin{aligned}
&I\bigl(X_1^n;Y_k \bigm| Y_1^{k-1}\bigr)\\
&\quad\le \mathsf{E}\bigl[\log|Y_k|^2\bigr] - h\bigl(Y_k \bigm| X_1^n, Y_1^{k-1}\bigr)\\
&\qquad + \xi\bigl(1+\log\mathsf{E}\bigl[|Y_k|^2\bigr] - \mathsf{E}\bigl[\log|Y_k|^2\bigr]\bigr) + \log\Gamma(\xi) - \xi\log\xi + \log\pi\\
&\quad= (1-\xi)\mathsf{E}\bigl[\log|Y_k|^2\bigr] - h\bigl(Y_k \bigm| X_1^n, Y_1^{k-1}\bigr)\\
&\qquad + \xi\bigl(1+\log\mathsf{E}\bigl[|Y_k|^2\bigr]\bigr) + \log\Gamma(\xi) - \xi\log\xi + \log\pi,
\end{aligned} \tag{6.60}$$

for any $\xi > 0$. Here $\Gamma(\cdot)$ denotes the Gamma function.

We evaluate the terms on the RHS of (6.60) individually. We upper bound the first term using Jensen's inequality

$$\begin{aligned}
\mathsf{E}\bigl[\log|Y_k|^2\bigr] &= \mathsf{E}\bigl[\mathsf{E}\bigl[\log|Y_k|^2 \bigm| X_1^k\bigr]\bigr]\\
&\le \mathsf{E}\bigl[\log\mathsf{E}\bigl[|Y_k|^2 \bigm| X_1^k\bigr]\bigr]\\
&= \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=0}^{k-1}\alpha_\ell|X_{k-\ell}|^2\right)\right].
\end{aligned} \tag{6.61}$$

The second term was already evaluated in (6.47)

$$\begin{aligned}
&h\bigl(Y_k \bigm| X_1^n, Y_1^{k-1}\bigr)\\
&\qquad\ge \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=0}^{k-1}\alpha_\ell|X_{k-\ell}|^2\right)\right] + \inf_{\ell\in\mathcal{L}}\bigl(h_\ell - \alpha_\ell\bigr),
\end{aligned} \tag{6.62}$$

and the next term is readily evaluated as

$$\log\mathsf{E}\bigl[|Y_k|^2\bigr] = \log\left(\sigma^2 + \sum_{\ell=0}^{k-1}\alpha_\ell\mathsf{E}\bigl[|X_{k-\ell}|^2\bigr]\right). \tag{6.63}$$

Our choice of ξ will satisfy $\xi < 1$ (see (6.65) ahead). We therefore obtain, upon substituting (6.61)–(6.63) in (6.60),

$$\begin{aligned}
& I\bigl(X_1^n; Y_k \big| Y_1^{k-1}\bigr) \\
&\quad \le (1-\xi)\mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=0}^{k-1} \alpha_\ell |X_{k-\ell}|^2\right)\right] \\
&\qquad - \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=0}^{k-1} \alpha_\ell |X_{k-\ell}|^2\right)\right] - \inf_{\ell\in\mathcal{L}} \bigl(h_\ell - \alpha_\ell\bigr) \\
&\qquad + \xi\left(1 + \log\left(\sigma^2 + \sum_{\ell=0}^{k-1} \alpha_\ell \mathsf{E}\left[|X_{k-\ell}|^2\right]\right)\right) \\
&\qquad + \log\Gamma(\xi) - \xi\log\xi + \log\pi \\
&\quad = -\inf_{\ell\in\mathcal{L}} \bigl(h_\ell - \alpha_\ell\bigr) + \log\Gamma(\xi) - \xi\log\xi + \log\pi \\
&\qquad + \xi\left(1 + \log\left(\sigma^2 + \sum_{\ell=0}^{k-1} \alpha_\ell \mathsf{E}\left[|X_{k-\ell}|^2\right]\right)\right. \\
&\qquad\qquad\qquad \left. - \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=0}^{k-1} \alpha_\ell |X_{k-\ell}|^2\right)\right]\right) \\
&\quad \le -\inf_{\ell\in\mathcal{L}} \bigl(h_\ell - \alpha_\ell\bigr) + \log\Gamma(\xi) - \xi\log\xi + \log\pi \\
&\qquad + \xi\left(1 + \log\left(1 + \sum_{\ell=0}^{k-1} \alpha_\ell \mathsf{E}\left[|X_{k-\ell}|^2\right]/\sigma^2\right)\right),
\end{aligned} \tag{6.64}$$

where the last inequality follows by lower bounding

$$\mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=0}^{k-1} \alpha_\ell |X_{k-\ell}|^2\right)\right] \ge \log\sigma^2.$$

We choose

$$\xi = \frac{1}{1 + \log\bigl(1 + \alpha\,\mathrm{SNR}\bigr)}, \tag{6.65}$$

where

$$\alpha \triangleq \sum_{\ell=0}^{\infty} \alpha_\ell.$$

Defining

$$\Psi(\mathrm{SNR}) \triangleq \left[\log\Gamma(\xi) - \log\frac{1}{\xi} - \xi\log\xi\right|_{\xi=\left(1+\log(1+\alpha\,\mathrm{SNR})\right)^{-1}}, \tag{6.66}$$

we obtain

$$\begin{aligned} I\big(X_1^n; Y_k \,\big|\, Y_1^{k-1}\big) &\\ \leq -\inf_{\ell\in\mathcal{L}}\big(h_\ell - \alpha_\ell\big) + \log\big(1 + \log(1+\alpha\,\mathrm{SNR})\big) + \Psi(\mathrm{SNR}) + \log\pi &\\ + \frac{1+\log\big(1+\sum_{\ell=0}^{k-1}\alpha_\ell\mathsf{E}\left[|X_{k-\ell}|^2\right]/\sigma^2\big)}{1+\log(1+\alpha\,\mathrm{SNR})}. \end{aligned} \tag{6.67}$$

Using (6.67) in (6.59) yields then

$$\begin{aligned} \frac{1}{n}I\big(X_1^n; Y_1^n\big) &\\ \leq -\inf_{\ell\in\mathcal{L}}\big(h_\ell - \alpha_\ell\big) + \log\big(1 + \log(1+\alpha\,\mathrm{SNR})\big) + \Psi(\mathrm{SNR}) + \log\pi &\\ + \frac{1+\frac{1}{n}\sum_{k=1}^{n}\log\big(1+\sum_{\ell=0}^{k-1}\alpha_\ell\mathsf{E}\left[|X_{k-\ell}|^2\right]/\sigma^2\big)}{1+\log(1+\alpha\,\mathrm{SNR})}. \end{aligned} \tag{6.68}$$

By Jensen's inequality we have

$$\begin{aligned} \frac{1}{n}\sum_{k=1}^{n}\log\left(1+\sum_{\ell=0}^{k-1}\alpha_\ell\mathsf{E}\left[|X_{k-\ell}|^2\right]/\sigma^2\right) &\leq \log\left(1+\frac{1}{n}\sum_{k=1}^{n}\sum_{\ell=0}^{k-1}\alpha_\ell\mathsf{E}\left[|X_{k-\ell}|^2\right]/\sigma^2\right) \\ &\leq \log(1+\alpha\,\mathrm{SNR}), \end{aligned} \tag{6.69}$$

where the last inequality follows by rewriting the double sum as

$$\frac{1}{n}\sum_{k=1}^{n}\frac{\mathsf{E}\left[|X_k|^2\right]}{\sigma^2}\sum_{\ell=0}^{n-k}\alpha_\ell$$

and by upper bounding then $\sum_{\ell=0}^{k-n}\alpha_\ell \leq \alpha$ and using the power constraint (6.8).

Combining (6.69) and (6.68) with (6.7) and (2.6), we obtain the upper bound

$$C(\mathrm{SNR}) \leq -\inf_{\ell\in\mathcal{L}}\bigl(h_\ell-\alpha_\ell\bigr)+\log\bigl(1+\log(1+\alpha\,\mathrm{SNR})\bigr) + \Psi(\mathrm{SNR})+\log\pi+1. \tag{6.70}$$

It follows from [28, Eq. (337)] that

$$\lim_{\mathrm{SNR}\to\infty}\Psi(\mathrm{SNR})=\lim_{\xi\downarrow 0}\left\{\log\Gamma(\xi)-\log\frac{1}{\xi}-\xi\log\xi\right\}=0. \tag{6.71}$$

Noting that

$$\lim_{\mathrm{SNR}\to\infty}\left\{\log\bigl(1+\log(1+\alpha\,\mathrm{SNR})\bigr)-\log\log\mathrm{SNR}\right\}=0,$$

we obtain from (6.70) and (6.71) the desired result

$$\varlimsup_{\mathrm{SNR}\to\infty}\bigl\{C(\mathrm{SNR})-\log\log\mathrm{SNR}\bigr\}\leq 1+\log\pi-\inf_{\ell\in\mathcal{L}}\bigl(h_\ell-\alpha_\ell\bigr). \tag{6.72}$$

6.6 Proofs of the Lower Bounds

In Section 6.6.1, we derive the lower bound on channel capacity that is presented in Proposition 6.5. This lower bound will be used in Sections 6.6.2 and 6.6.3 to prove Part (ii) of Theorem 6.1 and to prove the direct part of Theorem 6.3, respectively.

6.6.1 Capacity Lower Bound

To derive the desired lower bound on the (information) capacity, we evaluate $\frac{1}{n}I(X_1^n;Y_1^n)$ for the following distribution on the inputs $\{X_k,\, k\in\mathbb{Z}\}$:

Let $\mathsf{L}(\mathsf{P})$ be such that

$$\sum_{\ell=\mathsf{L}(\mathsf{P})+1}^{\infty}\alpha_\ell\,\mathsf{P}\leq\sigma^2. \tag{6.73}$$

To shorten notation, we shall write in the following L instead of $\mathsf{L}(\mathsf{P})$. Let $\tau\in\mathbb{N}$ be some positive integer that possibly depends on L, and let

$\mathbf{X}_b = (X_{b(\mathsf{L}+\tau)+1}, \ldots, X_{(b+1)(\mathsf{L}+\tau)})^\mathsf{T}$. We choose the sequence of vectors $\{\mathbf{X}_b,\, b \in \mathbb{Z}\}$ to be IID with

$$\mathbf{X}_b = \bigl(\underbrace{0, \ldots, 0}_{\mathsf{L}}, \tilde{X}_{b\tau+1}, \ldots, \tilde{X}_{(b+1)\tau}\bigr)^\mathsf{T},$$

where $\tilde{X}_{b\tau+1}, \ldots, \tilde{X}_{(b+1)\tau}$ is a sequence of independent, zero-mean, circularly-symmetric, complex random variables with $\log|\tilde{X}_{b\tau+\nu}|^2$ being uniformly distributed over the interval $\bigl[\log \mathsf{P}^{(\nu-1)/\tau}, \log \mathsf{P}^{\nu/\tau}\bigr]$, i.e., for each $\nu = 1, \ldots, \tau$

$$\log|\tilde{X}_{b\tau+\nu}|^2 \sim \mathcal{U}\Bigl(\bigl[\log \mathsf{P}^{(\nu-1)/\tau}, \log \mathsf{P}^{\nu/\tau}\bigr]\Bigr).$$

(Here and throughout this proof we assume that $\mathsf{P} > 1$.)

Let $\kappa \triangleq \bigl\lfloor \frac{n}{\mathsf{L}+\tau} \bigr\rfloor$, and let $\mathbf{Y}_b = (Y_{b(L+\tau)+1}, \ldots, Y_{(b+1)(\mathsf{L}+\tau)})^\mathsf{T}$. By the chain rule for mutual information we have

$$\begin{aligned} I\bigl(X_1^n; Y_1^n\bigr) &\geq I\bigl(\mathbf{X}_0^{\kappa-1}; \mathbf{Y}_0^{\kappa-1}\bigr) \\ &= \sum_{b=0}^{\kappa-1} I\bigl(\mathbf{X}_b; \mathbf{Y}_0^{\kappa-1} \bigm| \mathbf{X}_0^{b-1}\bigr) \\ &\geq \sum_{b=0}^{\kappa-1} I(\mathbf{X}_b; \mathbf{Y}_b), \end{aligned} \tag{6.74}$$

where the first step follows by restricting the number of observables; and where the last step follows by restricting the number of observables and by noting that $\{\mathbf{X}_b,\, b \in \mathbb{Z}\}$ is IID.

We continue by lower bounding each summand on the RHS of (6.74). We use again the chain rule and that reducing observations cannot increase mutual information to obtain

$$\begin{aligned} I(\mathbf{X}_b; \mathbf{Y}_b) &= \sum_{\nu=1}^{\tau} I\bigl(\tilde{X}_{b\tau+\nu}; \mathbf{Y}_b \bigm| \tilde{X}_{b\tau+1}^{b\tau+\nu-1}\bigr) \\ &\geq \sum_{\nu=1}^{\tau} I\bigl(\tilde{X}_{b\tau+\nu}; Y_{b(\mathsf{L}+\tau)+\mathsf{L}+\nu} \bigm| \tilde{X}_{b\tau+1}^{b\tau+\nu-1}\bigr) \\ &\geq \sum_{\nu=1}^{\tau} I\bigl(\tilde{X}_{b\tau+\nu}; Y_{b(\mathsf{L}+\tau)+\mathsf{L}+\nu}\bigr), \end{aligned} \tag{6.75}$$

where we have additionally used in the last step that $\tilde{X}_{b\tau+1}, \ldots, \tilde{X}_{(b+1)\tau}$ are independent.

Defining

$$W_{b\tau+\nu} \triangleq \sum_{\ell=1}^{b(\mathsf{L}+\tau)+\mathsf{L}+\nu-1} H_{b(\mathsf{L}+\tau)+\mathsf{L}+\nu}^{(\ell)} X_{b(\mathsf{L}+\tau)+\mathsf{L}+\nu-\ell} + Z_{b(\mathsf{L}+\tau)+\mathsf{L}+\nu}, \tag{6.76}$$

each summand on the RHS of (6.75) can be written as

$$I\big(\tilde{X}_{b\tau+\nu}; Y_{b(\mathsf{L}+\tau)+\mathsf{L}+\nu}\big) = I\big(\tilde{X}_{b\tau+\nu}; H_{b(\mathsf{L}+\tau)+\mathsf{L}+\nu}^{(0)} \tilde{X}_{b\tau+\nu} + W_{b\tau+\nu}\big). \tag{6.77}$$

A lower bound on (6.77) follows from the following lemma.

Lemma 6.6. *Let the random variables X, H, and W have finite second moments. Assume that both X and H are of finite differential entropy. Further assume that X is independent of H; that X is independent of W; and that X—∘—H—∘—W forms a Markov chain. Then*

$$I(X; HX + W) \geq h(X) - \mathsf{E}\big[\log|X|^2\big] + \mathsf{E}\big[\log|H|^2\big] - \mathsf{E}\left[\log\left(\pi e\left(\sigma_H + \frac{\sigma_W}{|X|}\right)^2\right)\right], \tag{6.78}$$

where $\sigma_H^2 \geq 0$ and $\sigma_H^2 > 0$ denote the variances of W and H. (Note that the assumptions that X and H have finite second moments and are of finite differential entropy guarantee that $\mathsf{E}\big[\log|X|^2\big]$ and $\mathsf{E}\big[\log|H|^2\big]$ are finite, see [28, Lemma 6.7e)]*.)*

Proof. See [27, Lemma 4]. □

It can be easily verified that for the channel model given in Section 6.2 and for the above coding scheme the lemma's conditions are satisfied. We therefore obtain from Lemma 6.6

$$\begin{aligned} I\big(\tilde{X}_{b\tau+\nu}; H_{b(\mathsf{L}+\tau)+\mathsf{L}+\nu}^{(0)} \tilde{X}_{b\tau+\nu} + W_{b\tau+\nu}\big) &\geq h\big(\tilde{X}_{b\tau+\nu}\big) - \mathsf{E}\Big[\log|\tilde{X}_{b\tau+\nu}|^2\Big] + \mathsf{E}\Big[\log\big|H_{b(\mathsf{L}+\tau)+\mathsf{L}+\nu}^{(0)}\big|^2\Big] \\ &\quad - \mathsf{E}\left[\log\left(\pi e\left(\sqrt{\alpha_0} + \frac{\sqrt{\mathsf{E}[|W_{b\tau+\nu}|^2]}}{|\tilde{X}_{b\tau+\nu}|}\right)^2\right)\right]. \end{aligned} \tag{6.79}$$

Using that the differential entropy of a circularly-symmetric random variable is given by [28, Eqs. (320) & (316)]

$$h\big(\tilde{X}_{b\tau+\nu}\big) = \mathsf{E}\Big[\log|\tilde{X}_{b\tau+\nu}|^2\Big] + h\big(\log|\tilde{X}_{b\tau+\nu}|^2\big) + \log\pi, \tag{6.80}$$

and evaluating $h(\log|\tilde{X}_{b\tau+\nu}|^2)$ for our choice of $\tilde{X}_{b\tau+\nu}$, yields for the first two terms on the RHS of (6.79)

$$h\big(\tilde{X}_{b\tau+\nu}\big) - \mathsf{E}\Big[\log|\tilde{X}_{b\tau+\nu}|^2\Big] = \log\log\mathsf{P}^{1/\tau} + \log\pi. \tag{6.81}$$

We next upper bound

$$\begin{aligned}\frac{\mathsf{E}\big[|W_{b\tau+\nu}|^2\big]}{|\tilde{X}_{b\tau+\nu}|^2} &= \sum_{\ell=1}^{\mathsf{L}} \alpha_\ell \frac{\mathsf{E}\big[|X_{b(\mathsf{L}+\tau)+\mathsf{L}+\nu-\ell}|^2\big]}{|\tilde{X}_{b\tau+\nu}|^2} \\ &+ \sum_{\ell=\mathsf{L}+1}^{b(\mathsf{L}+\tau)+\mathsf{L}+\nu-1} \alpha_\ell \frac{\mathsf{E}\big[|X_{b(\mathsf{L}+\tau)+\mathsf{L}+\nu-\ell}|^2\big]}{|\tilde{X}_{b\tau+\nu}|^2} + \frac{\sigma^2}{|\tilde{X}_{b\tau+\nu}|^2}.\end{aligned} \tag{6.82}$$

To this end, we note that for our choice of $\{X_k,\, k\in\mathbb{Z}\}$ and by the assumption that $\mathsf{P}>1$, we have

$$\mathsf{E}\big[|X_\ell|^2\big] \le \mathsf{P}, \qquad \ell\in\mathbb{N}, \tag{6.83}$$

$$\mathsf{E}\big[|X_{b(\mathsf{L}+\tau)+\mathsf{L}+\nu-\ell}|^2\big] \le \mathsf{P}^{(\nu-\ell)/\tau}, \quad \ell=1,\dots,\mathsf{L}, \tag{6.84}$$

and

$$|\tilde{X}_{b\tau+\nu}|^2 \ge \mathsf{P}^{(\nu-1)/\tau} \ge 1, \tag{6.85}$$

from which we obtain

$$\frac{\mathsf{E}\big[|X_{b(\mathsf{L}+\tau)+\mathsf{L}+\nu-\ell}|^2\big]}{|\tilde{X}_{b\tau+\nu}|^2} \le \frac{\mathsf{P}^{(\nu-\ell)/\tau}}{\mathsf{P}^{(\nu-1)/\tau}} \le 1, \quad \ell=1,\dots,\mathsf{L} \tag{6.86}$$

and

$$\frac{\mathsf{E}\big[|X_{b(\mathsf{L}+\tau)+\mathsf{L}+\nu-\ell}|^2\big]}{|\tilde{X}_{b\tau+\nu}|^2} \le \mathsf{P}, \quad \mathsf{L}+1\le\ell<b(\mathsf{L}+\tau)+\mathsf{L}+\nu. \tag{6.87}$$

Applying (6.85)–(6.87) to (6.82) yields

$$\begin{aligned}\frac{\mathsf{E}\left[|W_{b\tau+\nu}|^2\right]}{|\tilde{X}_{b\tau+\nu}|^2} &\le \sum_{\ell=1}^{\mathsf{L}} \alpha_\ell + \sum_{\ell=\mathsf{L}+1}^{b(\mathsf{L}+\tau)+\mathsf{L}+\nu-1} \alpha_\ell\,\mathsf{P} + \sigma^2 \\ &\le \alpha + \sum_{\ell=\mathsf{L}+1}^{\infty} \alpha_\ell\,\mathsf{P} + \sigma^2 \\ &\le \alpha + 2\sigma^2, \end{aligned} \tag{6.88}$$

where

$$\alpha \triangleq \sum_{\ell=0}^{\infty} \alpha_\ell.$$

Here the second step follows because α_ℓ, $\ell \in \mathbb{N}_0$ and P are nonnegative, and the last step follows from (6.73).

By combining (6.79) with (6.81) and (6.88), and by noting that by the stationarity of $\{H_k^{(0)},\, k \in \mathbb{Z}\}$

$$\mathsf{E}\left[\log\left|H^{(0)}_{b(\mathsf{L}+\tau)+\mathsf{L}+\nu}\right|^2\right] = \mathsf{E}\left[\log\left|H^{(0)}_1\right|^2\right],$$

we obtain the lower bound

$$\begin{aligned} &I\big(\tilde{X}_{b\tau+\nu}; H^{(0)}_{b(\mathsf{L}+\tau)+\mathsf{L}+\nu}\tilde{X}_{b\tau+\nu} + W_{b\tau+\nu}\big) \\ &\quad \ge \log\log \mathsf{P}^{1/\tau} + \mathsf{E}\left[\log\left|H^{(0)}_1\right|^2\right] - 1 - 2\log\big(\sqrt{\alpha_0} + \sqrt{\alpha+2\sigma^2}\big). \end{aligned} \tag{6.89}$$

Note that the RHS of (6.89) neither depends on ν nor on b. We therefore obtain from (6.89), (6.75), and (6.74)

$$I\big(X_1^n; Y_1^n\big) \ge \kappa\tau \log\log \mathsf{P}^{1/\tau} + \kappa\tau\Upsilon, \tag{6.90}$$

where we define Υ as

$$\Upsilon \triangleq \mathsf{E}\left[\log\left|H^{(0)}_1\right|^2\right] - 1 - 2\log\big(\sqrt{\alpha_0} + \sqrt{\alpha+2\sigma^2}\big).$$

Dividing the RHS of (6.90) by n, and computing the limit as n tends to infinity, yields the lower bound

$$C_{\text{Info}}(\text{SNR}) \ge \frac{\tau}{\mathsf{L}+\tau}\log\log \mathsf{P}^{1/\tau} + \frac{\tau}{\mathsf{L}+\tau}\Upsilon, \quad \mathsf{P} > 1, \tag{6.91}$$

where we have used that $\lim_{n\to\infty} \kappa/n = 1/(\mathsf{L}+\tau)$. This proves Proposition 6.5.

6.6.2 Condition for Unbounded Capacity

We use Proposition 6.5 to prove Part (ii) of Theorem 6.1. In particular, we show that if

$$\lim_{\ell\to\infty}\frac{1}{\ell}\log\frac{1}{\alpha_\ell}=\infty, \tag{6.92}$$

then, by cleverly choosing $\mathsf{L}(\mathsf{P})$ and τ, the lower bound (6.28), namely

$$C_{\text{Info}}(\text{SNR})\geq\frac{\tau}{\mathsf{L}(\mathsf{P})+\tau}\log\log\mathsf{P}^{1/\tau}+\frac{\tau}{\mathsf{L}(\mathsf{P})+\tau}\Upsilon,\quad \mathsf{P}>1,$$

can be made arbitrarily large as SNR tends to infinity. To this end, we first note that (6.92) implies that for every $0<\varrho<1$ we can find an $\ell_0\in\mathbb{N}$ such that

$$\alpha_\ell<\varrho^\ell,\quad \ell\geq\ell_0. \tag{6.93}$$

By choosing

$$\mathsf{L}(\mathsf{P})=\left\lceil\frac{\log\bigl(\mathsf{P}/\sigma^2\,\varrho/(1-\varrho)\bigr)}{\log(1/\varrho)}\right\rceil\qquad\text{and}\qquad \tau=\mathsf{L}(\mathsf{P}), \tag{6.94}$$

we obtain from (6.28) the lower bound

$$C_{\text{Info}}(\text{SNR})\geq\frac{1}{2}\log\frac{\log\mathsf{P}}{\left\lceil\frac{\log\left(\mathsf{P}/\sigma^2\,\varrho/(1-\varrho)\right)}{\log(1/\varrho)}\right\rceil}+\frac{1}{2}\Upsilon,\quad \mathsf{P}>1. \tag{6.95}$$

Taking the limit as SNR (and hence also $\mathsf{P}=\sigma^2\text{SNR}$) tends to infinity, yields

$$\lim_{\text{SNR}\to\infty}C_{\text{Info}}(\text{SNR})\geq\frac{1}{2}\log\log\frac{1}{\varrho}+\frac{1}{2}\Upsilon. \tag{6.96}$$

Since this holds for every $0<\varrho<1$, we have

$$\sup_{\text{SNR}>0}C_{\text{Info}}(\text{SNR})=\infty. \tag{6.97}$$

It remains to show that $\{\alpha_\ell\}$ and our choice of $\mathsf{L}(\mathsf{P})$ (6.94) satisfy the conditions (6.26) and (6.27) of Proposition 6.5, namely

$$\sum_{\ell=0}^{\infty}\alpha_\ell<\infty\qquad\text{and}\qquad\sum_{\ell=\mathsf{L}(\mathsf{P})+1}^{\infty}\alpha_\ell\,\mathsf{P}\leq\sigma^2.$$

It follows immediately from (6.5) and (6.93) that $\{\alpha_\ell\}$ satisfies the first condition (6.26):

$$\begin{aligned}\sum_{\ell=0}^{\infty}\alpha_\ell &= \sum_{\ell=0}^{\ell_0-1}\alpha_\ell+\sum_{\ell=\ell_0}^{\infty}\alpha_\ell\\ &< \ell_0\sup_{\ell\in\mathbb{N}_0}\alpha_\ell+\sum_{\ell=\ell_0}^{\infty}\varrho^\ell\\ &= \ell_0\sup_{\ell\in\mathbb{N}_0}\alpha_\ell+\frac{\varrho^{\ell_0}}{1-\varrho}\\ &< \infty.\end{aligned} \tag{6.98}$$

In order to show that $\mathsf{L}(\mathsf{P})$ satisfies the second condition (6.27), we first note that by (6.93)

$$\sum_{\ell=\ell'+1}^{\infty}\alpha_\ell < \sum_{\ell=\ell'+1}^{\infty}\varrho^\ell = \varrho^{\ell'}\frac{\varrho}{1-\varrho}, \quad \ell'\geq \ell_0-1. \tag{6.99}$$

Since $\mathsf{L}(\mathsf{P})$ tends to infinity as $\mathsf{P}\to\infty$ (6.94), it follows that $\mathsf{L}(\mathsf{P})$ is greater than (ℓ_0-1) for sufficiently large P. Furthermore, (6.94) implies

$$\varrho^{\mathsf{L}(\mathsf{P})}\frac{\varrho}{1-\varrho}\,\mathsf{P}\leq\sigma^2. \tag{6.100}$$

We therefore obtain from (6.99) and (6.100)

$$\sum_{\ell=\mathsf{L}(\mathsf{P})+1}^{\infty}\alpha_\ell\,\mathsf{P} < \varrho^{\mathsf{L}(\mathsf{P})}\frac{\varrho}{1-\varrho}\,\mathsf{P}\leq\sigma^2, \tag{6.101}$$

thus demonstrating that $\mathsf{L}(\mathsf{P})$ satisfies (6.27).

6.6.3 The Pre-LogLog

We use Proposition 6.5 to prove Theorem 6.3. To this end, we first note that because the number of paths is finite, we have for some $\mathsf{L}\in\mathbb{N}_0$

$$\alpha_\ell = 0, \quad \ell > \mathsf{L}, \tag{6.102}$$

which implies that

$$\sum_{\ell=0}^{\infty} \alpha_\ell = \sum_{\ell=0}^{\mathsf{L}} \alpha_\ell \le (\mathsf{L}+1) \sup_{\ell \in \mathbb{N}_0} \alpha_\ell < \infty \tag{6.103}$$

and

$$\sum_{\ell=\mathsf{L}+1}^{\infty} \alpha_\ell \mathsf{P} = 0 \le \sigma^2. \tag{6.104}$$

We further note that $C(\mathrm{SNR}) = C_{\mathrm{Info}}(\mathrm{SNR})$ [20, Thm. 2]. Consequently, it follows from (6.28) of Proposition 6.5 that the capacity is lower bounded by

$$C(\mathrm{SNR}) \ge \frac{\tau}{\mathsf{L}+\tau} \log\log \mathsf{P}^{1/\tau} + \frac{\tau}{\mathsf{L}+\tau}\Upsilon, \quad \mathsf{P} > 1. \tag{6.105}$$

Dividing the RHS of (6.105) by $\log\log\mathrm{SNR}$, and computing the limit as $\mathrm{SNR} \to \infty$, yields

$$\varliminf_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\log\log\mathrm{SNR}} \ge \frac{\tau}{\mathsf{L}+\tau}, \tag{6.106}$$

where we have used that for any fixed τ

$$\lim_{\mathrm{SNR}\to\infty} \frac{\log\log \mathsf{P}^{1/\tau}}{\log\log\mathrm{SNR}} = 1.$$

The lower bound on the capacity pre-loglog

$$\Lambda \triangleq \varlimsup_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\log\log\mathrm{SNR}} \ge \varliminf_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\log\log\mathrm{SNR}} \ge 1 \tag{6.107}$$

follows then by letting τ tend to infinity. Together with the upper bound $\Lambda \le 1$, which was derived in Section 6.5.2, this proves Theorem 6.3.

6.7 Conclusion

We studied the high-SNR behavior of the capacity of noncoherent multipath fading channels. We demonstrated that, depending on the decay rate of the sequence $\{\alpha_\ell\}$, the capacity may be bounded or unbounded in the SNR. We further showed that if the number of paths is finite, then at high SNR capacity grows double-logarithmically with the SNR, and the capacity pre-loglog does not depend on the number of paths. The picture that emerges is as follows:

- If the sequence of variances $\{\alpha_\ell\}$ decays exponentially or slower, then the capacity is bounded in the SNR.
- If the sequence of variances $\{\alpha_\ell\}$ decays faster than exponentially, then the capacity is unbounded in the SNR.
- If the number of paths is finite, then the capacity pre-loglog is equal to 1, irrespective of the number of paths.

The conclusions that can be drawn from these results are twofold. First, multipath channels with an infinite number of paths and multipath channels with a finite number of paths have in general completely different capacity behaviors at high SNR. Indeed, at high SNR, if the number of paths is finite, then capacity grows double-logarithmically with the SNR, whereas if the number of paths is infinite, then the capacity may even be bounded in the SNR. Thus, while for low and moderate SNR it might be reasonable to approximate a multipath channel with infinitely many paths by a multipath channel with only a finite number paths, this is not reasonable when the SNR tends to infinity. The number of paths that are needed to approximate a multipath channel typically depends on the SNR and may grow to infinity as the SNR tends to infinity.

Second, the above results indicate that the high-SNR behavior of the capacity of multipath fading channels depends critically on the assumed channel model. Thus when studying such channels at high SNR, the channel modeling is crucial, since slight changes in the channel model might lead to completely different capacity results.

Chapter 7

Summary and Conclusion

In this dissertation we studied the effect of heating up and fading in communication channels. In particular, we investigated the impact of these phenomena on channel capacity.

The heating-up effect was studied in **Chapter 3**. We proposed a channel model where the variance of the additive noise depends on a weighted sum of the past channel input powers. To study the capacity of this channel at low transmit powers, we computed the capacity per unit cost. We showed that, irrespective of the distribution of the additive noise, the heating-up effect is unharmful in the sense that the capacity per unit cost cannot be smaller than the capacity per unit cost of the channel with an ideal heat sink. We further showed that if the noise is Gaussian, then the heating-up effect is even beneficial in the sense that the capacity per unit cost is larger than the capacity per unit cost of the channel with an ideal heat sinks. This suggests that at low transmit powers no heat sinks should be used. To study capacity at large transmit powers, we derived a sufficient condition and a necessary condition on the weights describing the dependence of the noise variance on the past channel input powers for the capacity to be unbounded in the transmit power. We showed that when the sequence of weights decays not faster than geometrically, the capacity is bounded in the transmit power, and when the sequence of weights decays faster than geometrically, the capacity is unbounded in the transmit power. This demonstrates the importance of an efficient heat sink at large transmit powers. The main conclusions of this chapter are thus as follows:

- When the transmit power is low, heat sinks are not only unnecessary, but they may even reduce the capacity by dissipating heat which contains information about the transmitted signal.

- When the available transmit power is large, heat sinks can significantly increase the capacity. In fact, if the heat is dissipated slowly, then the capacity is bounded in the available transmit power, i.e., the capacity does not tend to infinity as the available power tends to infinity.

The remaining chapters were devoted to fading. While these chapters studied different versions of fading channels, they shared the assumption that the transmitter and the receiver only know the statistics of the fading, but not its realization, i.e., they all studied a noncoherent channel model.

In **Chapter 4**, we studied multiple-input multiple-output (MIMO) Gaussian flat-fading channels with memory. We first derived nonasymptotic upper and lower bounds on the capacity. The asymptotic behavior of these bounds was then analyzed in the limit as the signal-to-noise ratio (SNR) tends to infinity. The upper bounds were used to derive upper bounds on the fading number of regular Gaussian fading channels and on the capacity pre-log of nonregular Gaussian fading channels. The lower bounds were used to derive an expression for the capacity pre-loglog of nonregular single-input single-output (SISO) Gaussian fading channels. We further proposed a new approach to derive lower bounds on the fading number of MIMO fading channels. With this approach, we derived a lower bound on the fading number of spatially IID, zero-mean, Gaussian fading channels with memory. Our bounds demonstrate that when the number of receive antennas does not exceed the number of transmit antennas, the fading number of zero-mean, spatially IID, slowly-varying, MIMO Gaussian fading channels is proportional to the number of degrees of freedom, i.e., to the minimum number of transmit and receive antennas. The main conclusions of this chapter are as follows:

- While the high-SNR asymptotic capacity depends on the fading memory only via the (noiseless) prediction error, the capacity-vs-SNR curve depends on the memory of the fading process more finely, namely via the functional dependence of the noisy prediction error on the variance of the noise corrupting the observations. In fact, we demonstrated that two fading channels that have the same (noiseless) prediction error and therefore the same high-SNR asymptotic capacity can have completely different behaviors at low and moderate SNR.

- The fading number of MIMO Gaussian fading channels is proportional to the number of degrees of freedom. Thus, like in the high-SNR asymptotic analysis of coherent MIMO fading channels [43] (where the receiver is cognizant of the realization of the fading), the number of degrees of freedom plays also an important role in the high-SNR asymptotic analysis of noncoherent fading channels.

In **Chapter 5**, we investigated the robustness of the Gaussian fading assumption in the analysis of fading channels at high SNR. For SISO flat-fading channels, we showed that, among all stationary and ergodic fading processes of a given spectral distribution function and whose law has no mass point at zero, the Gaussian process gives rise to the smallest capacity pre-log. We further demonstrated that the assumption that the fading law has no mass point at zero is essential in the sense that there exist stationary and ergodic fading processes of some spectral distribution function (and whose law has a mass point at zero) that give rise to a smaller pre-log than the Gaussian process of equal spectral distribution function. An extension of our result to MISO fading channels with memory was also presented. The main conclusion of this chapter is:

- For a large class of fading processes the Gaussian process is the worst. The Gaussian fading assumption in the analysis of fading channels at high SNR is thus conservative.

Finally, in **Chapter 6** we studied the capacity of multipath (frequency-selective) fading channels. We showed that if the variances of the path gains decay exponentially or slower, then the capacity is bounded in the SNR. In contrast, if the variances of the path gains decay faster than exponentially, then the capacity is unbounded in the SNR. We further showed that, for multipath fading channels with a finite number of paths, the capacity is not merely unbounded in the SNR, but its high-SNR asymptotic behavior is also independent of the number of paths. The main conclusions of this chapter are the following:

- Multipath channels with an infinite number of paths and multipath channels with a finite number of paths possess usually completely different asymptotic capacity behaviors in the limit as the SNR tends to infinity. At high SNR it is thus not reasonable to approximate a multipath channel with an infinite number of paths by a multipath channel with a finite number of paths.
- In general, the high-SNR asymptotic capacity of multipath fading channels depends critically on the assumed channel model.

For noncoherent fading channels, particularly the high-SNR asymptotic behavior of channel capacity depends critically on the assumed channel model. Indeed, for flat-fading channels, the memory of the fading process determines whether the high-SNR asymptotic growth of the capacity with the SNR is double-logarithmic, logarithmic, or even something in between. For multipath (frequency-selective) fading channels, the asymptotic growth depends additionally on the decay rate of the sequence of the path-gains' variances. Thus, in the analysis of fading channels at high SNR, one has to attach great importance to the channel model, since slight changes in the model might lead to completely different capacity results.

Karma police, arrest this man, he talks in maths,
he buzzes like a fridge, he's like a detuned radio.

Radiohead, "Karma Police"

Appendix A

Appendix to Chapter 3

A.1 Proof of Proposition 3.1

We first note that by the expression of the capacity per unit cost of a memoryless channel [47] we have

$$\sup_{\mathrm{SNR}>0} \frac{C_{\alpha=0}(\mathrm{SNR})}{\mathrm{SNR}} = \sup_{\zeta>0} \frac{D\big(W_{\alpha=0}(\cdot|\zeta) \,\big\|\, W_{\alpha=0}(\cdot|0)\big)}{\zeta^2/\sigma^2}, \tag{A.1}$$

where $W_{\alpha=0}(\cdot|\cdot)$ denotes the channel law of the channel

$$Y_k = x_k + \sigma\, U_k. \tag{A.2}$$

Thus, to prove Proposition 3.1 it suffices to show that

$$\sup_{\mathrm{SNR}>0} \frac{C_{\mathrm{Info}}(\mathrm{SNR})}{\mathrm{SNR}} \geq \sup_{\zeta^2>0} \frac{D\big(W_{\alpha=0}(\cdot|\zeta) \,\big\|\, W_{\alpha=0}(\cdot|0)\big)}{\zeta^2/\sigma^2}.$$

We shall obtain this result by deriving a lower bound on $C_{\mathrm{Info}}(\mathrm{SNR})$ and by computing then its limiting ratio to SNR as SNR tends to zero.

In order to lower bound $C_{\mathrm{Info}}(\mathrm{SNR})$, which we defined in (3.16) as

$$C_{\mathrm{Info}}(\mathrm{SNR}) = \varliminf_{n\to\infty} \frac{1}{n} \sup I(X_1^n; Y_1^n),$$

we evaluate $\frac{1}{n} I(X_1^n; Y_1^n)$ for inputs $\{X_k,\, k \in \mathbb{Z}\}$ that are blockwise IID in blocks of L symbols (for some $L \in \mathbb{N}$). Thus $\{(X_{bL+1}, \ldots, X_{(b+1)L}), b \in \mathbb{N}_0\}$ is a sequence of IID random length-L vectors with $(X_{bL+1}, \ldots, X_{(b+1)L})$ taking on the value $(\xi, 0, \ldots, 0)$

with probability δ and $(0,\dots,0)$ with probability $1-\delta$, for some $\xi \in \mathbb{R}$. To satisfy the power constraint (3.12) we shall choose ξ and δ such that

$$\frac{\xi^2}{\sigma^2}\delta = L\,\mathrm{SNR}. \tag{A.3}$$

We use the chain rule for mutual information to write

$$\begin{aligned}\frac{1}{n}I\big(X_1^n;Y_1^n\big) &= \frac{1}{n}\sum_{b=0}^{\lfloor n/L\rfloor-1} I\big(X_{bL+1};Y_1^n \bigm| X_1^{bL}\big)\\ &\geq \frac{1}{n}\sum_{b=0}^{\lfloor n/L\rfloor-1} I\big(X_{bL+1};Y_{bL+1} \bigm| X_1^{bL}\big),\end{aligned} \tag{A.4}$$

where the inequality follows because reducing observations cannot increase mutual information.

Let $R^{(\xi)}_{\text{on-off}}(\mathsf{snr})$ denote the maximum rate achievable on (A.2) using on-off keying with on-symbol ξ and with its corresponding probability $\wp$ chosen in order to satisfy the power constraint snr, i.e.,

$$R^{(\xi)}_{\text{on-off}}(\mathsf{snr}) \triangleq \sup_{\substack{P_X(\xi)=1-P_X(0)=\wp,\\ \xi^2/\sigma^2\wp\leq\mathsf{snr}}} I(X;X+\sigma U_k). \tag{A.5}$$

Notice that $\mathsf{snr} \mapsto R^{(\xi)}_{\text{on-off}}(\mathsf{snr})$ is a nonnegative, monotonically nondecreasing function with $R^{(\xi)}_{\text{on-off}}(0) = 0$. From the strict concavity of mutual information it follows that

$$R^{(\xi)}_{\text{on-off}}(\mathsf{snr}) > 0, \qquad \mathsf{snr} > 0.$$

Also, for a fixed ξ, the function $\mathsf{snr} \mapsto R^{(\xi)}_{\text{on-off}}(\mathsf{snr})$ is concave in snr. Consequently, for some $\mathsf{snr}_0 > 0$, we have that $\mathsf{snr} \mapsto R^{(\xi)}_{\text{on-off}}(\mathsf{snr})$ is strictly monotonic in $\mathsf{snr} \in [0,\mathsf{snr}_0]$ and hence the supremum on the RHS of (A.5) is attained for

$$\wp = \mathsf{snr}\,\sigma^2/\xi^2, \qquad 0 \leq \mathsf{snr} \leq \mathsf{snr}_0.$$

By expressing $I\big(X_{bL+1};Y_{bL+1} \bigm| X_1^{bL}=x_1^{bL}\big)$ for a given $X_1^{bL}=x_1^{bL}$ as

$$\begin{aligned}I\big(X_{bL+1};Y_{bL+1} \bigm| X_1^{bL}=x_1^{bL}\big) &= I\Big(X_{bL+1};X_{bL+1}+\theta\big(x_1^{bL}\big)U_{bL+1}\Big)\\ &= I\left(X_{bL+1};\frac{\sigma}{\theta\big(x_1^{bL}\big)}X_{bL+1}+\sigma\,U_{bL+1}\right)\end{aligned}$$

(where $\theta\big(x_1^{bL}\big)$ is defined in (3.68)), and by using that for $0 \leq \mathsf{snr} \leq \mathsf{snr}_0$ the supremum on the RHS of (A.5) is attained for $\wp = \mathsf{snr}\,\sigma^2/\xi^2$, we obtain

$$\begin{aligned} I\big(X_{bL+1}; Y_{bL+1} \bigm| X_1^{bL} = x_1^{bL}\big) \\ = R_{\text{on-off}}^{(\xi)}\left(\frac{L\,\mathrm{SNR}}{1+\sum_{\ell=0}^{b-1}\alpha_{(b-\ell)L}\,x_{\ell L+1}^2/\sigma^2}\right), \end{aligned} \tag{A.6}$$

for $0 \leq \mathrm{SNR} \leq \mathsf{snr}_0/L$. Averaging (A.6) over X_1^{bL}, and applying the result to (A.4), yields

$$\begin{aligned} \frac{1}{n} I\big(X_1^n; Y_1^n\big) &\geq \frac{1}{n}\sum_{b=0}^{\lfloor n/L\rfloor-1} \mathsf{E}\left[R_{\text{on-off}}^{(\xi)}\left(\frac{L\,\mathrm{SNR}}{1+\sum_{\ell=0}^{b-1}\alpha_{(b-\ell)L}\,X_{\ell L+1}^2/\sigma^2}\right)\right] \\ &\geq \frac{\lfloor n/L\rfloor}{n} R_{\text{on-off}}^{(\xi)}\left(\frac{L\,\mathrm{SNR}}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\,\xi^2/\sigma^2}\right), \end{aligned} \tag{A.7}$$

where the second inequality follows by the monotonicity of $R_{\text{on-off}}^{(\xi)}(\cdot)$ and because we have with probability one

$$\sum_{\ell=0}^{b-1}\alpha_{(b-\ell)L}\,X_{\ell L+1}^2/\sigma^2 \leq \sum_{\ell=1}^{\infty}\alpha_{\ell L}\,\xi^2/\sigma^2.$$

The lower bound on $C_{\text{Info}}(\mathrm{SNR})$ follows by letting n tend to infinity

$$\begin{aligned} &C_{\text{Info}}(\mathrm{SNR}) \\ &\quad\geq \varliminf_{n\to\infty}\frac{1}{n}I\big(X_1^n;Y_1^n\big) \\ &\quad\geq \frac{1}{L}R_{\text{on-off}}^{(\xi)}\left(\frac{L\,\mathrm{SNR}}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\,\xi^2/\sigma^2}\right), \quad 0 \leq \mathrm{SNR} \leq \mathsf{snr}_0/L. \end{aligned} \tag{A.8}$$

We continue by lower bounding the information capacity per unit cost as

$$\begin{aligned} &\sup_{\mathrm{SNR}>0}\frac{C_{\text{Info}}(\mathrm{SNR})}{\mathrm{SNR}} \\ &\quad\geq \varliminf_{\mathrm{SNR}\downarrow 0}\frac{C_{\text{Info}}(\mathrm{SNR})}{\mathrm{SNR}} \end{aligned}$$

$$\geq \lim_{\mathrm{SNR}\downarrow 0} \frac{1}{L} \frac{R_{\text{on-off}}^{(\xi)}\left(\frac{L\,\mathrm{SNR}}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\,\xi^2/\sigma^2}\right)}{\mathrm{SNR}}$$
$$= \lim_{\mathrm{SNR}\downarrow 0} \frac{R_{\text{on-off}}^{(\xi)}\left(\frac{L\,\mathrm{SNR}}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\,\xi^2/\sigma^2}\right)}{\frac{L\,\mathrm{SNR}}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\,\xi^2/\sigma^2}} \frac{1}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\,\xi^2/\sigma^2}$$
$$= \lim_{\mathrm{SNR}'\downarrow 0} \frac{R_{\text{on-off}}^{(\xi)}(\mathrm{SNR}')}{\mathrm{SNR}'} \frac{1}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\,\xi^2/\sigma^2}, \tag{A.9}$$

where in the last step we substitute $\mathrm{SNR}' = \frac{L\,\mathrm{SNR}}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\xi^2/\sigma^2}$.

Proceeding along the lines of the proof of [47, Thm. 3], one can show that

$$\lim_{\mathrm{SNR}'\downarrow 0} \frac{R_{\text{on-off}}^{(\xi)}(\mathrm{SNR}')}{\mathrm{SNR}'} = \frac{D\big(W_{\alpha=0}(\cdot|\xi)\,\big\|\,W_{\alpha=0}(\cdot|0)\big)}{\xi^2/\sigma^2} \tag{A.10}$$

and therefore

$$\sup_{\mathrm{SNR}>0} \frac{C_{\text{Info}}(\mathrm{SNR})}{\mathrm{SNR}}$$
$$\geq \frac{D\big(W_{\alpha=0}(\cdot|\xi)\,\big\|\,W_{\alpha=0}(\cdot|0)\big)}{\xi^2/\sigma^2} \frac{1}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\,\xi^2/\sigma^2}. \tag{A.11}$$

Noting that (3.10) and (3.23) imply

$$0 \leq \lim_{L\to\infty} \sum_{\ell=1}^{\infty} \alpha_{\ell L} \leq \lim_{L\to\infty} \sum_{\ell=L}^{\infty} \alpha_{\ell} = 0 \tag{A.12}$$

we obtain, upon letting L tend to infinity,

$$\sup_{\mathrm{SNR}>0} \frac{C_{\text{Info}}(\mathrm{SNR})}{\mathrm{SNR}} \geq \frac{D\big(W_{\alpha=0}(\cdot|\xi)\,\big\|\,W_{\alpha=0}(\cdot|0)\big)}{\xi^2/\sigma^2}. \tag{A.13}$$

Maximizing (A.13) over ξ^2 yields then

$$\sup_{\mathrm{SNR}>0} \frac{C_{\text{Info}}(\mathrm{SNR})}{\mathrm{SNR}} \geq \sup_{\xi>0} \frac{D\big(W_{\alpha=0}(\cdot|\xi)\,\big\|\,W_{\alpha=0}(\cdot|0)\big)}{\xi^2/\sigma^2}, \tag{A.14}$$

which, in view of (A.1), proves Proposition 3.1.

A.2 Appendix to Section 3.5.2

We shall prove that

$$\lim_{b\to\infty} I\big(\mathbf{X}_{-\infty}^{-1}; \tilde{\mathbf{Y}}_b \bigm| \mathbf{X}_0^b\big) = 0. \tag{A.15}$$

Let $\alpha_b^{(i)}$ be defined as

$$\alpha_0^{(1)} \triangleq 0 \tag{A.16}$$
$$\alpha_b^{(i)} \triangleq \alpha_{bL+i-1}, \qquad (b,i) \in \mathbb{N}_0 \times \mathbb{N} \setminus \{(0,1)\}. \tag{A.17}$$

We have

$$\begin{aligned}
& I\big(\mathbf{X}_{-\infty}^{-1}; \tilde{\mathbf{Y}}_b \bigm| \mathbf{X}_0^b\big) \\
&= \sum_{i=1}^{L} I\big(\mathbf{X}_{-\infty}^{-1}; \tilde{Y}_{bL+i} \bigm| \mathbf{X}_0^b, \tilde{Y}_{bL+1}^{bL+i-1}\big) \\
&\leq \sum_{i=1}^{L} \Big(h\big(\tilde{Y}_{bL+i} \bigm| \mathbf{X}_0^b\big) - h\big(\tilde{Y}_{bL+i} \bigm| \mathbf{X}_{-\infty}^b\big)\Big) \\
&\leq \frac{1}{2}\sum_{i=1}^{L} \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=0}^{b} \alpha_{b-\ell}^{(i)} X_{\ell L+1}^2 + \mathsf{P}\, L \sum_{\ell=b+1}^{\infty} \alpha_\ell^{(i)}\right)\right] \\
&\quad - \frac{1}{2}\sum_{i=1}^{L} \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=-\infty}^{b} \alpha_{b-\ell}^{(i)} X_{\ell L+1}^2\right)\right] \\
&\leq \frac{1}{2}\sum_{i=1}^{L} \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=0}^{b} \alpha_{b-\ell}^{(i)} X_{\ell L+1}^2 + \mathsf{P}\, L \sum_{\ell=b+1}^{\infty} \alpha_\ell^{(i)}\right)\right] \\
&\quad - \frac{1}{2}\sum_{i=1}^{L} \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=0}^{b} \alpha_{b-\ell}^{(i)} X_{\ell L+1}^2\right)\right] \\
&= \frac{1}{2}\sum_{i=1}^{L} \mathsf{E}\left[\log\left(1 + \frac{\mathsf{P}\, L \sum_{\ell=b+1}^{\infty} \alpha_\ell^{(i)}}{\sigma^2 + \sum_{\ell=0}^{b} \alpha_{b-\ell}^{(i)} X_{\ell L+1}^2}\right)\right] \\
&\leq \frac{1}{2}\sum_{i=1}^{L} \log\left(1 + L\,\mathrm{SNR} \sum_{\ell=b+1}^{\infty} \alpha_\ell^{(i)}\right),
\end{aligned} \tag{A.18}$$

where the second step follows because conditioning cannot increase entropy and because, conditional on $\mathbf{X}_{-\infty}^b$, the random variable $\tilde{Y}_{bL+i}$

is independent of $\tilde{Y}_{bL+1}^{bL+i-1}$; the third step follows from the entropy maximizing property of Gaussian random variables and because, conditional on $\mathbf{X}_{-\infty}^{b}$, the random variable $\tilde{Y}_{bL+1}^{bL+i-1}$ is Gaussian; the fourth step follows because with probability one

$$\sum_{\ell=-\infty}^{-1} \alpha_{b-\ell}^{(i)} X_{\ell L+1}^2 \geq 0, \quad i = 1, \ldots, L;$$

and the last step follows because with probability one

$$\sum_{\ell=0}^{b} \alpha_{b-\ell}^{(i)} X_{\ell L+1}^2 \geq 0, \quad i = 1, \ldots, L.$$

By upper bounding

$$\sum_{\ell=b+1}^{\infty} \alpha_{\ell}^{(i)} \leq \sum_{\ell=b+1}^{\infty} \alpha_{\ell}, \quad i = 1, \ldots, L$$

we obtain

$$I\big(\mathbf{X}_{-\infty}^{-1}; \tilde{\mathbf{Y}}_b \mid \mathbf{X}_0^b\big) \leq \frac{L}{2} \log\left(1 + L\,\mathrm{SNR} \sum_{\ell=b+1}^{\infty} \alpha_\ell\right), \tag{A.19}$$

and (A.15) follows then by noting that (3.23) implies

$$\lim_{b\to\infty} \sum_{\ell=b+1}^{\infty} \alpha_i = 0.$$

A.3 Proof of Lemma 3.5

We show that for any $\epsilon > 0$

$$\lim_{n\to\infty} \Pr\left(\left|\frac{1}{\lfloor n/L \rfloor}\|\mathbf{Y}\|^2 - \big(\sigma^2 + \mathsf{P} + \alpha^{(L)}\,\mathsf{P}\big)\right| > \epsilon\right) = 0 \tag{A.20}$$

and

$$\lim_{n\to\infty} \Pr\left(\left|\frac{1}{\lfloor n/L \rfloor}\|\mathbf{Z}\|^2 - \big(\sigma^2 + \alpha^{(L)}\,\mathsf{P}\big)\right| > \epsilon\right) = 0. \tag{A.21}$$

Lemma 3.5 follows then by the union of events bound.

In order to prove (A.20) and (A.21), we first note that

$$\frac{1}{\lfloor n/L \rfloor} \mathsf{E}\left[\|\mathbf{Y}\|^2\right] = \sigma^2 + \mathsf{P} + \frac{\mathsf{P}}{\lfloor n/L \rfloor} \sum_{k=1}^{\lfloor n/L \rfloor - 1} \sum_{\ell=1}^{k} \alpha_{\ell L} \tag{A.22}$$

$$\frac{1}{\lfloor n/L \rfloor} \mathsf{E}\left[\|\mathbf{Z}\|^2\right] = \sigma^2 + \frac{\mathsf{P}}{\lfloor n/L \rfloor} \sum_{k=1}^{\lfloor n/L \rfloor - 1} \sum_{\ell=1}^{k} \alpha_{\ell L} \tag{A.23}$$

and hence, by Cesáro's mean [5, Thm. 4.2.3],

$$\lim_{n\to\infty} \frac{1}{\lfloor n/L \rfloor} \mathsf{E}\left[\|\mathbf{Y}\|^2\right] = \sigma^2 + \mathsf{P} + \alpha^{(L)}\,\mathsf{P} \tag{A.24}$$

$$\lim_{n\to\infty} \frac{1}{\lfloor n/L \rfloor} \mathsf{E}\left[\|\mathbf{Z}\|^2\right] = \sigma^2 + \alpha^{(L)}\,\mathsf{P}, \tag{A.25}$$

where $\alpha^{(L)}$ was defined in (3.87) as

$$\alpha^{(L)} = \sum_{\ell=1}^{\infty} \alpha_{\ell L}.$$

Thus for any $\epsilon > 0$ and $0 < \varepsilon < \epsilon$ there exists an n_0 such that for all $n \geq n_0$

$$\left| \frac{1}{\lfloor n/L \rfloor} \mathsf{E}\left[\|\mathbf{Y}\|^2\right] - (\sigma^2 + \mathsf{P} + \alpha^{(L)}\,\mathsf{P}) \right| \leq \varepsilon \tag{A.26}$$

$$\left| \frac{1}{\lfloor n/L \rfloor} \mathsf{E}\left[\|\mathbf{Z}\|^2\right] - (\sigma^2 + \alpha^{(L)}\,\mathsf{P}) \right| \leq \varepsilon, \tag{A.27}$$

and it follows from the triangle inequality that

$$\left| \frac{1}{\lfloor n/L \rfloor} \|\mathbf{Y}\|^2 - (\sigma^2 + \mathsf{P} + \alpha^{(L)}\,\mathsf{P}) \right| \leq \left| \frac{1}{\lfloor n/L \rfloor} \|\mathbf{Y}\|^2 - \frac{1}{\lfloor n/L \rfloor} \mathsf{E}\left[\|\mathbf{Y}\|^2\right] \right| + \varepsilon \tag{A.28}$$

and

$$\left| \frac{1}{\lfloor n/L \rfloor} \|\mathbf{Z}\|^2 - (\sigma^2 + \alpha^{(L)}\,\mathsf{P}) \right| \leq \left| \frac{1}{\lfloor n/L \rfloor} \|\mathbf{Z}\|^2 - \frac{1}{\lfloor n/L \rfloor} \mathsf{E}\left[\|\mathbf{Z}\|^2\right] \right| + \varepsilon. \tag{A.29}$$

This yields

$$\begin{aligned}\Pr\left(\left|\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Y}\|^2-(\sigma^2+\mathsf{P}+\alpha^{(L)}\,\mathsf{P})\right|>\epsilon\right)\\ \leq \Pr\left(\left|\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Y}\|^2-\frac{1}{\lfloor n/L\rfloor}\mathsf{E}\left[\|\mathbf{Y}\|^2\right]\right|>\epsilon-\varepsilon\right)\\ \leq \frac{\mathsf{Var}\left(\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Y}\|^2\right)}{(\epsilon-\varepsilon)^2}\end{aligned} \tag{A.30}$$

and

$$\begin{aligned}\Pr\left(\left|\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Z}\|^2-(\sigma^2+\alpha^{(L)}\,\mathsf{P})\right|>\epsilon\right)\\ \leq \Pr\left(\left|\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Z}\|^2-\frac{1}{\lfloor n/L\rfloor}\mathsf{E}\left[\|\mathbf{Z}\|^2\right]\right|>\epsilon-\varepsilon\right)\\ \leq \frac{\mathsf{Var}\left(\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Z}\|^2\right)}{(\epsilon-\varepsilon)^2},\end{aligned} \tag{A.31}$$

where $\mathsf{Var}(A)=\mathsf{E}\left[(A-\mathsf{E}[A])^2\right]$ denotes the variance of A. Here the last inequalities in (A.30) and (A.31) follow from Chebyshev's inequality [14, Sec. 5.4].

It remains to show that

$$\lim_{n\to\infty}\mathsf{Var}\left(\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Y}\|^2\right)=\lim_{n\to\infty}\mathsf{Var}\left(\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Z}\|^2\right)=0. \tag{A.32}$$

We shall prove (A.32) for $\mathbf{Y}$. The proof for $\mathbf{Z}$ follows along the same lines. We begin by writing $\mathsf{Var}\left(\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Y}\|^2\right)$ as

$$\begin{aligned}\mathsf{Var}\left(\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Y}\|^2\right)\\ =\frac{1}{\lfloor n/L\rfloor^2}\mathsf{Var}\left(\sum_{k=0}^{\lfloor n/L\rfloor-1}Y_{kL+1}^2\right)\\ =\frac{1}{\lfloor n/L\rfloor^2}\sum_{k=0}^{\lfloor n/L\rfloor-1}\mathsf{Var}\left(Y_{kL+1}^2\right)\end{aligned}$$

$$+ \frac{2}{\lfloor n/L \rfloor^2} \sum_{k>j} \mathsf{Cov}\big(Y_{kL+1}^2, Y_{jL+1}^2\big), \qquad \text{(A.33)}$$

where $\mathsf{Cov}(A, B) = \mathsf{E}[(A - \mathsf{E}[A])(B - \mathsf{E}[B])]$ denotes the covariance between A and B. We evaluate both terms on the RHS of (A.33) separately. For the sake of clarity, we shall omit the details and show only the main steps. Unless otherwise stated, these steps can be derived in a straightforward way using that

(i) $\{X_{kL+1}, k \in \mathbb{N}_0\}$ is a sequence of IID, zero-mean, variance-P, Gaussian random variables;

(ii) the fourth moment of a zero-mean, variance-P, Gaussian random variable is given by $3\mathsf{P}$, and all odd moments are zero;

(iii) $X_k = 0$ for $k \mod L \neq 1$;

(iv) $\{U_k, k \in \mathbb{Z}\}$ (and hence also $\{U_{kL+1}, k \in \mathbb{N}_0\}$) is a zero-mean, unit-variance, stationary, weakly-mixing random process;

(v) $\{X_k, k \in \mathbb{Z}\}$ and $\{U_k, k \in \mathbb{Z}\}$ are independent of each other.

For the first sum on the RHS of (A.33) it suffices to show that $\mathsf{Var}\big(Y_{kL+1}^2\big) < \infty$, $k \in \mathbb{N}_0$. Indeed, this sum contains only $\lfloor n/L \rfloor$ summands and hence, if $\mathsf{Var}\big(Y_{kL+1}^2\big) < \infty$, then its ratio to $\lfloor n/L \rfloor^2$ vanishes as n tends to infinity. We have

$$\begin{aligned}
&\mathsf{Var}\big(Y_{kL+1}^2\big) \\
&= \mathsf{E}\big[Y_{kL+1}^4\big] - \big(\mathsf{E}\big[Y_{kL+1}^2\big]\big)^2 \\
&\leq \mathsf{E}\big[Y_{kL+1}^4\big] \\
&= \mathsf{E}\bigg[\Big(X_{kL+1} + \theta\big(X_1^{kL}\big)U_{kL+1}\Big)^4\bigg] \\
&= 3\mathsf{P}^2 + 6\mathsf{P}\bigg(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{k} \alpha_{\ell L}\bigg) \\
&\quad + \left(\sigma^4 + 2\sigma^2\mathsf{P}\sum_{\ell=1}^{k} \alpha_{\ell L} + 2\mathsf{P}^2 \sum_{\ell=1}^{k} \alpha_{\ell L}^2 + \mathsf{P}^2\bigg(\sum_{\ell=1}^{k} \alpha_{kL}\bigg)^2\right) \mathsf{E}\big[U_{kL+1}^4\big]
\end{aligned}$$

$$\leq 3\mathsf{P}^2 + 6\mathsf{P}\Big(\sigma^2 + \mathsf{P}\,\alpha^{(L)}\Big)$$
$$+ \left(\sigma^4 + 2\sigma^2\mathsf{P}\,\alpha^{(L)} + 2\mathsf{P}^2\sum_{\ell=1}^{\infty}\alpha_{\ell L}^2 + \mathsf{P}^2\Big(\alpha^{(L)}\Big)^2\right)\mathsf{E}\left[U_{kL+1}^4\right], \tag{A.34}$$

where the last step follows by upper bounding $\sum_{\ell=1}^{k}\alpha_{\ell L}$ by $\alpha^{(L)}$ and $\sum_{\ell=1}^{k}\alpha_{\ell L}^2$ by $\sum_{\ell=1}^{\infty}\alpha_{\ell L}^2$. Note that (3.82) implies that

$$\alpha^{(L)} < \infty \quad \text{and} \quad \sum_{\ell=1}^{\infty}\alpha_{\ell L}^2 < \infty.$$

By additionally noting that U_{kL+1} has a finite fourth moment (3.9), it follows that (for any finite P)

$$\mathsf{Var}\big(Y_{kL+1}^2\big) < \infty, \quad k \in \mathbb{N}_0. \tag{A.35}$$

In order to show that the second term on the RHS of (A.33) vanishes as n tends to infinity, we evaluate

$$\mathsf{Cov}\big(Y_{kL+1}^2, Y_{jL+1}^2\big) = \mathsf{E}\left[Y_{kL+1}^2 Y_{jL+1}^2\right] - \mathsf{E}\left[Y_{kL+1}^2\right]\mathsf{E}\left[Y_{jL+1}^2\right], \quad k > j \tag{A.36}$$

We have

$$\mathsf{E}\left[Y_{kL+1}^2 Y_{jL+1}^2\right]$$
$$= \mathsf{P}^2 + \mathsf{P}\left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{j}\alpha_{\ell L}\right) + \mathsf{P}\left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}\right) + 2\mathsf{P}^2\,\alpha_{(k-j)L}$$
$$+ \left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}\right)\left(\sigma^2 + \mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\right)\mathsf{E}\left[U_{kL+1}^2 U_{jL+1}^2\right]$$
$$+ 2\mathsf{P}^2\sum_{\ell=1}^{j}\alpha_{\ell L}\,\alpha_{(\ell+k-j)L}\,\mathsf{E}\left[U_{kL+1}^2 U_{jL+1}^2\right] \tag{A.37}$$

and

$$\mathsf{E}\left[Y_{kL+1}^2\right]\mathsf{E}\left[Y_{jL+1}^2\right] = \mathsf{P}^2 + \mathsf{P}\left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{j}\alpha_{\ell L}\right) + \mathsf{P}\left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}\right)$$
$$+ \left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}\right)\left(\sigma^2 + \mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\right). \tag{A.38}$$

Equations (A.36), (A.37), and (A.38) thus yield

$$\begin{aligned}
&\mathsf{Cov}\big(Y_{kL+1}^2, Y_{jL+1}^2\big)\\
&= 2\mathsf{P}^2\,\alpha_{(k-j)L} + 2\mathsf{P}^2\sum_{\ell=1}^{j}\alpha_{\ell L}\alpha_{(\ell+k-j)L}\,\mathsf{E}\big[U_{kL+1}^2 U_{jL+1}^2\big]\\
&\quad + \Bigg(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}\Bigg)\Bigg(\sigma^2 + \mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\Bigg)\Big(\mathsf{E}\big[U_{kL+1}^2 U_{jL+1}^2\big] - 1\Big).
\end{aligned} \tag{A.39}$$

We continue by summing $\mathsf{Cov}\big(Y_{kL+1}^2, Y_{jL+1}^2\big)$ over (k,j)

$$\begin{aligned}
&\sum_{k>j}\mathsf{Cov}\big(Y_{kL+1}^2, Y_{jL+1}^2\big)\\
&= \sum_{k>j} 2\mathsf{P}^2\,\alpha_{(k-j)L} + \sum_{k>j} 2\mathsf{P}^2\sum_{\ell=1}^{j}\alpha_{\ell L}\alpha_{(\ell+k-j)L}\,\mathsf{E}\big[U_{kL+1}^2 U_{jL+1}^2\big]\\
&\quad + \sum_{k>j}\Bigg(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}\Bigg)\Bigg(\sigma^2 + \mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\Bigg)\Big(\mathsf{E}\big[U_{kL+1}^2 U_{jL+1}^2\big] - 1\Big)\\
&= \sum_{j=0}^{\lfloor n/L\rfloor-2}\ \sum_{\nu=1}^{\lfloor n/L\rfloor-1-j} 2\mathsf{P}^2\,\alpha_{\nu L}\\
&\quad + \sum_{j=0}^{\lfloor n/L\rfloor-2}\ \sum_{\nu=1}^{\lfloor n/L\rfloor-1-j} 2\mathsf{P}^2\sum_{\ell=1}^{j}\alpha_{\ell L}\alpha_{(\ell+\nu)L}\,\mathsf{E}\big[U_{\nu L+1}^2 U_1^2\big]\\
&\quad + \sum_{j=0}^{\lfloor n/L\rfloor-2}\ \sum_{\nu=1}^{\lfloor n/L\rfloor-1-j}\Bigg(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{j+\nu}\alpha_{\ell L}\Bigg)\Bigg(\sigma^2 + \mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\Bigg)\\
&\qquad\qquad \times\Big(\mathsf{E}\big[U_{\nu L+1}^2 U_1^2\big] - 1\Big),
\end{aligned} \tag{A.40}$$

where the second step follows by substituting $\nu = k - j$ and from the stationarity of $\{U_k,\, k\in\mathbb{Z}\}$.

The first two terms on the RHS of (A.40) can be upper bounded using (3.83), namely

$$\alpha_\ell < \varrho^\ell, \quad \big(0<\varrho<1,\, \ell\geq \ell_0\big).$$

Indeed, by noting that $L \geq \ell_0$, this yields

$$\sum_{\nu=1}^{\lfloor n/L \rfloor - 1 - j} \alpha_{\nu L} < \sum_{\nu=1}^{\lfloor n/L \rfloor - 1 - j} \varrho^{\nu L} < \sum_{\nu=1}^{\lfloor n/L \rfloor} \varrho^{\nu L} \tag{A.41}$$

and

$$\begin{aligned} \sum_{\nu=1}^{\lfloor n/L \rfloor - 1 - j} \sum_{\ell=1}^{j} \alpha_{\ell L} \alpha_{(\ell+\nu)L} &< \sum_{\nu=1}^{\lfloor n/L \rfloor - 1 - j} \sum_{\ell=1}^{j} \left(\varrho^{2L}\right)^{\ell} \varrho^{\nu L} \\ &< \sum_{\nu=1}^{\lfloor n/L \rfloor} \sum_{\ell=1}^{\infty} \left(\varrho^{2L}\right)^{\ell} \varrho^{\nu L} \\ &= \frac{\varrho^{2L}}{1-\varrho^{2L}} \sum_{\nu=1}^{\lfloor n/L \rfloor} \varrho^{\nu L}. \end{aligned} \tag{A.42}$$

Applying (A.41) we can upper bound the first term on the RHS of (A.40) by

$$\begin{aligned} &\frac{2}{\lfloor n/L \rfloor^2} \sum_{j=0}^{\lfloor n/L \rfloor - 2} \sum_{\nu=1}^{\lfloor n/L \rfloor - 1 - j} 2\mathsf{P}^2 \alpha_{\nu L} \\ &\qquad < \frac{4\mathsf{P}^2}{\lfloor n/L \rfloor^2} \sum_{j=0}^{\lfloor n/L \rfloor - 2} \sum_{\nu=1}^{\lfloor n/L \rfloor} \varrho^{\nu L} \\ &\qquad = 4\mathsf{P}^2 \frac{\lfloor n/L \rfloor - 1}{\lfloor n/L \rfloor} \frac{1}{\lfloor n/L \rfloor} \sum_{\nu=1}^{\lfloor n/L \rfloor} \varrho^{\nu L}. \end{aligned} \tag{A.43}$$

Likewise, applying (A.42) we can upper bound the second term on the RHS of (A.40) by

$$\begin{aligned} &\frac{2}{\lfloor n/L \rfloor^2} \sum_{j=0}^{\lfloor n/L \rfloor - 2} \sum_{\nu=1}^{\lfloor n/L \rfloor - 1 - j} 2\mathsf{P}^2 \sum_{\ell=1}^{j} \alpha_{\ell L} \alpha_{(\ell+\nu)L} \mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] \\ &\leq \frac{4\mathsf{P}^2}{\lfloor n/L \rfloor^2} \sum_{j=0}^{\lfloor n/L \rfloor - 2} \sum_{\nu=1}^{\lfloor n/L \rfloor - 1 - j} \sum_{\ell=1}^{j} \alpha_{\ell L} \alpha_{(\ell+\nu)L} \mathsf{E}\left[U_1^4\right] \\ &< 4\mathsf{P}^2 \frac{\varrho^{2L}}{1-\varrho^{2L}} \mathsf{E}\left[U_1^4\right] \frac{\lfloor n/L \rfloor - 1}{\lfloor n/L \rfloor} \frac{1}{\lfloor n/L \rfloor} \sum_{\nu=1}^{\lfloor n/L \rfloor} \varrho^{\nu L}, \end{aligned} \tag{A.44}$$

where the first inequality follows from the Cauchy-Schwarz inequality.

As for the last term on the RHS of (A.40), we upper bound each summand by

$$\begin{aligned}
&\left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{j+\nu}\alpha_{\ell L}\right)\left(\sigma^2 + \mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\right)\left(\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - 1\right) \\
&\quad \leq \left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{j+\nu}\alpha_{\ell L}\right)\left(\sigma^2 + \mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\right)\left|\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - 1\right| \\
&\quad \leq \left(\sigma^2 + \mathsf{P}\,\alpha^{(L)}\right)^2\left|\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - 1\right|, \qquad \text{(A.45)}
\end{aligned}$$

where the first inequality follows by upper bounding

$$\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - 1 \leq \left|\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - 1\right|;$$

and the second inequality follows by upper bounding

$$\sum_{\ell=1}^{j}\alpha_{\ell L} \leq \sum_{\ell=1}^{j+\nu}\alpha_{\ell L} \leq \sum_{\ell=1}^{\infty}\alpha_{\ell L} = \alpha^{(L)}.$$

Applying (A.43), (A.44), and (A.45) to (A.40) yields

$$\begin{aligned}
&\frac{2}{\lfloor n/L\rfloor^2}\sum_{k>j}\mathsf{Cov}\left(Y_{kL+1}^2, Y_{jL+1}^2\right) \\
&< 4\mathsf{P}^2\frac{\lfloor n/L\rfloor - 1}{\lfloor n/L\rfloor}\frac{1}{\lfloor n/L\rfloor}\sum_{\nu=1}^{\lfloor n/L\rfloor}\varrho^{\nu L} \\
&\quad + 4\mathsf{P}^2\frac{\varrho^{2L}}{1-\varrho^{2L}}\mathsf{E}\left[U_1^4\right]\frac{\lfloor n/L\rfloor - 1}{\lfloor n/L\rfloor}\frac{1}{\lfloor n/L\rfloor}\sum_{\nu=1}^{\lfloor n/L\rfloor}\varrho^{\nu L} \\
&\quad + \frac{2}{(\lfloor n/L\rfloor)^2}\sum_{j=0}^{\lfloor n/L\rfloor-2}\sum_{\nu=1}^{\lfloor n/L\rfloor-1-j}\left(\sigma^2 + \mathsf{P}\,\alpha^{(L)}\right)^2\left|\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - 1\right| \\
&\leq 4\mathsf{P}^2\frac{\lfloor n/L\rfloor - 1}{\lfloor n/L\rfloor}\frac{1}{\lfloor n/L\rfloor}\sum_{\nu=1}^{\lfloor n/L\rfloor}\varrho^{\nu L} \\
&\quad + 4\mathsf{P}^2\frac{\varrho^{2L}}{1-\varrho^{2L}}\mathsf{E}\left[U_1^4\right]\frac{\lfloor n/L\rfloor - 1}{\lfloor n/L\rfloor}\frac{1}{\lfloor n/L\rfloor}\sum_{\nu=1}^{\lfloor n/L\rfloor}\varrho^{\nu L}
\end{aligned}$$

$$+\,2\Big(\sigma^2+\mathsf{P}\,\alpha^{(L)}\Big)^2\frac{\lfloor n/L\rfloor-1}{\lfloor n/L\rfloor}\frac{1}{\lfloor n/L\rfloor}\sum_{\nu=1}^{\lfloor n/L\rfloor}\Big|\mathsf{E}\big[U_{\nu L+1}^2U_1^2\big]-1\Big|.\text{(A.46)}$$

Here the second step follows by upper bounding

$$\sum_{\nu=1}^{\lfloor n/L\rfloor-1-j}\Big(\sigma^2+\mathsf{P}\,\alpha^{(L)}\Big)^2\Big|\mathsf{E}\big[U_{\nu L+1}^2U_1^2\big]-1\Big|$$
$$\leq\sum_{\nu=1}^{\lfloor n/L\rfloor}\Big(\sigma^2+\mathsf{P}\,\alpha^{(L)}\Big)^2\Big|\mathsf{E}\big[U_{\nu L+1}^2U_1^2\big]-1\Big|.$$

By Cesáro's mean the first two terms on the RHS of (A.46) tend to zero as n tends to infinity, and by the weakly-mixing property of $\{U_k,\,k\in\mathbb{Z}\}$ the third term on the RHS of (A.46) tends to zero as n tends to infinity [35, Thm. 6.1]. It thus follows from (A.33), (A.35), and (A.46) that

$$\lim_{n\to\infty}\mathsf{Var}\left(\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Y}\|^2\right)=0.$$

Together with (A.30) this proves (A.20). The proof of (A.21) follows along the same lines.

Appendix B

Appendix to Chapter 4

The proofs given in this appendix were originally derived in [22]. We repeat them here for the sake of completeness.

B.1 Proof of Theorem 4.2

To derive an upper bound on the capacity of peak-power limited MIMO fading channels, we begin by using the chain rule

$$I\big(\mathbf{X}_1^n;\mathbf{Y}_1^n\big)=\sum_{k=1}^{n} I\big(\mathbf{Y}_k;\mathbf{X}_1^n \,\big|\, \mathbf{Y}_1^{k-1}\big) \tag{B.1}$$

and upper bound each of the terms in the sum by

$$\begin{aligned}
I\big(\mathbf{Y}_k;\mathbf{X}_1^n \,\big|\, \mathbf{Y}_1^{k-1}\big) &= I\big(\mathbf{Y}_k;\mathbf{X}_1^n,\mathbf{Y}_1^{k-1}\big)-I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big)\\
&\le I\big(\mathbf{Y}_k;\mathbf{X}_1^n,\mathbf{Y}_1^{k-1}\big)\\
&= I\big(\mathbf{Y}_k;\mathbf{X}_1^k,\mathbf{Y}_1^{k-1}\big)\\
&= I\big(\mathbf{Y}_k;\mathbf{X}_k\big)+I\big(\mathbf{Y}_k;\mathbf{X}_1^{k-1},\mathbf{Y}_1^{k-1}\,\big|\,\mathbf{X}_k\big)\\
&\le \sup I\big(\mathbf{Y}_k;\mathbf{X}_k\big)+I\big(\mathbf{Y}_k;\mathbf{X}_1^{k-1},\mathbf{Y}_1^{k-1}\,\big|\,\mathbf{X}_k\big),
\end{aligned} \tag{B.2}$$

where the maximization on the RHS of (B.2) is over all distributions on $\mathbf{X}_k$ satisfying with probability one $\|\mathbf{X}_k\|\le A$. Here the first step follows from the chain rule; the second step follows from the nonnegativity of mutual information; the third step follows from the absence of feedback, which implies that

$$\mathbf{X}_{k+1}^n \multimap\!\!- \big(\mathbf{X}_1^k,\mathbf{Y}_1^{k-1}\big) \multimap\!\!- \mathbf{Y}_k$$

forms a Markov chain; the fourth step follows from the chain rule; and the last step follows by maximizing the first term on the RHS of (B.2).

The first term on the RHS of (B.2) is the capacity of a memoryless fading channel

$$\sup I\big(\mathbf{Y}_k;\mathbf{X}_k\big) = C_{\mathrm{PP}}^{(\mathrm{IID})}(\mathrm{SNR}). \tag{B.3}$$

In the following we upper bound the second term on the RHS of (B.2).

Let the $(n_\mathrm{R}\times n_\mathrm{T})$-valued process $\{\mathbb{W}_k,\ k\in\mathbb{Z}\}$ be spatially IID, where each process $\{W_k(r,t),\ k\in\mathbb{Z}\}$ is a sequence of IID, zero-mean, unit-variance, circularly-symmetric, complex Gaussian random variables, drawn independently of $\{(\mathbf{X}_k,\mathbb{H}_k,\mathbf{Z}_k),\ k\in\mathbb{Z}\}$. We have

$$\begin{aligned}
& I\big(\mathbf{Y}_k;\mathbf{X}_1^{k-1},\mathbf{Y}_1^{k-1}\bigm|\mathbf{X}_k\big)\\
&\quad = I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\bigm|\mathbf{X}_1^{k}\big)\\
&\quad \leq I\left(\mathbf{Y}_k;\left\{\mathbb{H}_\ell+\frac{\sigma}{\mathsf{A}}\mathbb{W}_\ell\right\}_{\ell=1}^{k-1}\,\middle|\,\mathbf{X}_k\right)\\
&\quad \leq \sup_{\|\mathbf{x}_0\|\leq\mathsf{A}} I\left(\mathbf{Y}_0;\left\{\mathbb{H}_\ell+\frac{\sigma}{\mathsf{A}}\mathbb{W}_\ell\right\}_{\ell=-\infty}^{-1}\,\middle|\,\mathbf{X}_0=\mathbf{x}_0\right),
\end{aligned} \tag{B.4}$$

where the first step follows because, conditional on $\mathbf{X}_1^k$, the random vectors $\mathbf{Y}_k$ and $\mathbf{X}_1^{k-1}$ are independent; the second step follows from the data processing inequality and by noting that, conditional on $\mathbf{X}_1^k$,

$$\mathbf{Y}_1^{k-1} \multimap \left\{\mathbb{H}_\ell+\frac{\sigma}{\mathsf{A}}\mathbb{W}_\ell\right\}_{\ell=1}^{k-1} \multimap \mathbf{Y}_k$$

forms a Markov chain; and the last step follows because the expected value cannot be greater than the supremum, because additional information cannot decrease mutual information, and because the channel is stationary.

We continue by expressing the fading $\mathbb{H}_0$ as

$$\mathbb{H}_0 = \overline{\mathbb{H}}_0 + \widetilde{\mathbb{H}}_0, \tag{B.5}$$

where $\overline{\mathbb{H}}_0$ is the best estimate of the fading $\mathbb{H}_0$ given the noisy observation $\mathbb{H}_1+\sigma/\mathsf{A}\,\mathbb{W}_1$, $\mathbb{H}_2+\sigma/\mathsf{A}\,\mathbb{W}_2,\ldots$, i.e.,

$$\overline{\mathbb{H}}_0 = \mathsf{E}\left[\mathbb{H}_0\,\middle|\,\left\{\mathbb{H}_\ell+\frac{\sigma}{\mathsf{A}}\mathbb{W}_\ell\right\}_{\ell=-\infty}^{-1}\right], \tag{B.6}$$

and where $\widetilde{\mathbb{H}}_0$ is the prediction error. We thus have

$$\begin{aligned}
&I\left(\mathbf{Y}_0; \left\{\mathbb{H}_\ell + \frac{\sigma}{\mathsf{A}}\mathbb{W}_\ell\right\}_{\ell=-\infty}^{-1} \,\middle|\, \mathbf{X}_0 = \mathbf{x}_0\right) \\
&\quad = I\left(\left(\overline{\mathbb{H}}_0 + \widetilde{\mathbb{H}}_0\right)\mathbf{X}_0 + \mathbf{Z}_0; \left\{\mathbb{H}_\ell + \frac{\sigma}{\mathsf{A}}\mathbb{W}_\ell\right\}_{\ell=-\infty}^{-1} \,\middle|\, \mathbf{X}_0 = \mathbf{x}_0\right) \\
&\quad = h\left(\left(\overline{\mathbb{H}}_0 + \widetilde{\mathbb{H}}_0\right)\mathbf{X}_0 + \mathbf{Z}_0 \,\middle|\, \mathbf{X}_0 = \mathbf{x}_0\right) \\
&\qquad - h\left(\left(\overline{\mathbb{H}}_0 + \widetilde{\mathbb{H}}_0\right)\mathbf{X}_0 + \mathbf{Z}_0 \,\middle|\, \left\{\mathbb{H}_\ell + \frac{\sigma}{\mathsf{A}}\mathbb{W}_\ell\right\}_{\ell=-\infty}^{-1}, \mathbf{X}_0 = \mathbf{x}_0\right). \qquad \text{(B.7)}
\end{aligned}$$

To compute the first entropy on the RHS of (B.7), we note that, conditional on $\mathbf{X}_0 = \mathbf{x}_0$, the random vector $\left(\overline{\mathbb{H}}_0 + \widetilde{\mathbb{H}}\right)\mathbf{x}_0 + Z_0$ has a Gaussian law of mean

$$\mathsf{E}[\mathbb{H}_0\mathbf{x}_0 + \mathbf{Z}_0] = \mathsf{D}\mathbf{x}_0 \qquad \text{(B.8)}$$

and of covariance matrix

$$\mathsf{K}_{\mathbf{YY}} = \mathsf{E}\left[\mathbb{H}_0\mathbf{x}_0\mathbf{x}_0^\dagger\mathbb{H}_0^\dagger\right] + \sigma^2\mathsf{I}_{n_\mathrm{R}}. \qquad \text{(B.9)}$$

We further note that $\mathsf{K}_{\mathbf{YY}}$ is a diagonal matrix with diagonal entries

$$\begin{aligned}
&\mathsf{E}\left[\left|\sum_{t=1}^{n_\mathrm{T}}\big(H_0(r,t) - d(r,t)\big)\, x_0(t) + Z_0(r)\right|^2\right] \\
&\quad = \mathsf{E}\left[\left|\sum_{t=1}^{n_\mathrm{T}}\big(H_0(r,t) - d(r,t)\big)\, x_0(t)\right|^2\right] + \sigma^2 \\
&\quad = \sum_{t=1}^{n_\mathrm{T}} |x_0(t)|^2 + \sigma^2 \\
&\quad = \|\mathbf{x}_0\|^2 + \sigma^2, \quad r = 1, \ldots, n_\mathrm{R}, \qquad \text{(B.10)}
\end{aligned}$$

where the first step follows because $\mathbb{H}_0$ and $\mathbf{Z}_0$ are independent; and the second step follows because $\{\mathbb{H} - \mathsf{D},\, k \in \mathbb{Z}\}$ is spatially independent, and because the processes $\{H(r,t) - d(r,t),\, k \in \mathbb{Z}\}$ are of unit variance.

Using the expression for the differential entropy of a multivariate Gaussian random variable yields

$$h\Big(\big(\overline{\mathbb{H}}_0 + \widetilde{\mathbb{H}}_0\big)\mathbf{X}_0 + \mathbf{Z}_0 \;\Big|\; \mathbf{X}_0 = \mathbf{x}_0 \Big)$$
$$= n_{\mathrm{R}} \log(\pi e) + \sum_{r=1}^{n_{\mathrm{R}}} \log\left(\|\mathbf{x}_0\|^2 + \sigma^2 \right). \tag{B.11}$$

To compute the second entropy on the RHS of (B.7), we note that, conditional on

$$\left(\mathbf{X}_0, \left\{ \mathbb{H}_\ell + \frac{\sigma}{\mathsf{A}} \mathbb{W}_\ell \right\}_{\ell=-\infty}^{-1} \right) = \left(\mathbf{x}_0, \left\{ \mathsf{H}_\ell + \frac{\sigma}{\mathsf{A}} \mathsf{W}_\ell \right\}_{\ell=-\infty}^{-1} \right),$$

the random vector $\big(\overline{\mathbb{H}}_0 + \widetilde{\mathbb{H}}_0\big)\mathbf{X}_0 + \mathbf{Z}_0$ has a Gaussian law of mean

$$\left(\mathsf{D} + \overline{\mathsf{H}}_0 \right) \mathbf{x}_0 \tag{B.12}$$

and of covariance matrix

$$\mathsf{K}_{\mathbf{YY}|\overline{\mathsf{H}}_0} = \mathsf{E}\left[\widetilde{\mathbb{H}}_0 \mathbf{x}_0 \mathbf{x}_0^\dagger \widetilde{\mathbb{H}}_0^\dagger \right] + \sigma^2 \mathsf{I}_{n_{\mathrm{R}}}. \tag{B.13}$$

We further note that $\mathsf{K}_{\mathbf{YY}|\overline{\mathsf{H}}_0}$ is a diagonal matrix with diagonal entries

$$\mathsf{E}\left[\left| \sum_{t=1}^{n_{\mathrm{T}}} \widetilde{H}_0(r,t)\, x_0(t) + Z_0(r) \right|^2 \right]$$
$$= \mathsf{E}\left[\left| \sum_{t=1}^{n_{\mathrm{T}}} \widetilde{H}_0(r,t)\, x_0(t) \right|^2 \right] + \sigma^2$$
$$= \sum_{t=1}^{n_{\mathrm{T}}} |x_0(t)|^2\, \epsilon_{r,t}^2(1/\mathrm{SNR}) + \sigma^2, \quad r = 1, \ldots, n_{\mathrm{R}}, \tag{B.14}$$

where the last step follows because if $\{\mathbb{H}_k - \mathsf{D},\, k \in \mathbb{Z}\}$ is spatially independent, then also the entries in $\widetilde{\mathbb{H}}_0$ are independent [22, Lemma 3.1].

Using, as above, the expression for the differential entropy of a multivariate Gaussian random variable yields

$$h\left(\big(\overline{\mathbb{H}}_0 + \widetilde{\mathbb{H}}_0\big)\mathbf{X}_0 + \mathbf{Z}_0 \;\middle|\; \left\{ \mathbb{H}_\ell + \frac{\sigma}{\mathsf{A}} \mathbb{W}_\ell \right\}_{\ell=-\infty}^{-1}, \mathbf{X}_0 = \mathbf{x}_0 \right)$$
$$= n_{\mathrm{R}} \log(\pi e) + \sum_{r=1}^{n_{\mathrm{R}}} \log\left(\sum_{t=1}^{n_{\mathrm{T}}} |x_0(t)|^2\, \epsilon_{r,t}^2(1/\mathrm{SNR}) + \sigma^2 \right), \tag{B.15}$$

where we additionally use that the covariance matrix $\mathsf{K}_{\mathbf{YY}|\overline{\mathsf{H}}_0}$ does not depend on the realization of $\left\{\mathbb{H}_\ell + \sigma/\mathsf{A}\,\mathbb{W}_\ell\right\}_{\ell=-\infty}^{-1}$.

Applying (B.11) and (B.15) to (B.7) yields

$$\begin{aligned}
I\biggl(\mathbf{Y}_0; \Bigl\{\mathbb{H}_\ell + \frac{\sigma}{\mathsf{A}}\mathbb{W}_\ell\Bigr\}_{\ell=-\infty}^{-1} \biggm| \mathbf{X}_0 = \mathbf{x}_0\biggr)
&= \sum_{r=1}^{n_\mathrm{R}} \log \frac{\|\mathbf{x}_0\|^2 + \sigma^2}{\sum_{t=1}^{n_\mathrm{T}} |x_0(t)|^2\, \epsilon_{r,t}^2(1/\mathrm{SNR}) + \sigma^2} \\
&= \sum_{r=1}^{n_\mathrm{R}} \log \frac{1 + \sigma^2/\|\mathbf{x}_0\|^2}{\sum_{t=1}^{n_\mathrm{T}} |x_0(t)|^2/\|\mathbf{x}_0\|^2\, \epsilon_{r,t}^2(1/\mathrm{SNR}) + \sigma^2/\|\mathbf{x}_0\|^2} \\
&\le \sum_{r=1}^{n_\mathrm{R}} \log \frac{1 + 1/\mathrm{SNR}}{\sum_{t=1}^{n_\mathrm{T}} |x_0(t)|^2/\|\mathbf{x}_0\|^2\, \epsilon_{r,t}^2(1/\mathrm{SNR}) + 1/\mathrm{SNR}}, \qquad \text{(B.16)}
\end{aligned}$$

where the last step follows because the function

$$x \mapsto \log \frac{1+x}{\alpha + x}, \quad 0 < \alpha \le 1$$

is monotonically decreasing in x, and because, by the power constraint, $\|\mathbf{x}_0\| \le \mathsf{A}$. (Note that $\sum_{t=1}^{n_\mathrm{T}} |x_0(t)\|^2/\|\mathbf{x}_0\|^2\, \epsilon_{r,t}^2(1/\mathrm{SNR})$ cannot be greater than 1, because the prediction error cannot be larger than the variance of the random variable one wishes to predict—i.e., $\epsilon_{r,t}^2(1/\mathrm{SNR}) \le 1$—and because $\sum_{t=1}^{n_\mathrm{T}} |x_0(t)|^2/\|\mathbf{x}_0\|^2 = 1$.)

Combining (B.2), (B.3), (B.4), and (B.16), we obtain

$$\begin{aligned}
I\bigl(\mathbf{Y}_k; \mathbf{X}_1^n \bigm| \mathbf{Y}_1^{k-1}\bigr)
&\le C_\mathrm{PP}^{(\mathrm{IID})}(\mathrm{SNR}) + \sup_{\|\mathbf{x}_0\| \le \mathsf{A}} I\biggl(\mathbf{Y}_0; \Bigl\{\mathbb{H}_\ell + \frac{\sigma}{\mathsf{A}}\mathbb{W}_\ell\Bigr\}_{\ell=-\infty}^{-1} \biggm| \mathbf{X}_0 = \mathbf{x}_0\biggr) \\
&\le C_\mathrm{PP}^{(\mathrm{IID})}(\mathrm{SNR}) \\
&\quad + \sup_{\|\mathbf{x}_0\| \le \mathsf{A}} \sum_{r=1}^{n_\mathrm{R}} \log \frac{1 + 1/\mathrm{SNR}}{\sum_{t=1}^{n_\mathrm{T}} |x_0(t)|^2/\|\mathbf{x}_0\|^2\, \epsilon_{r,t}^2(1/\mathrm{SNR}) + 1/\mathrm{SNR}} \\
&= C_\mathrm{PP}^{(\mathrm{IID})}(\mathrm{SNR}) \\
&\quad + \max_{\|\hat{\mathbf{x}}\| = 1} \sum_{r=1}^{n_\mathrm{R}} \log \frac{1 + 1/\mathrm{SNR}}{\sum_{t=1}^{n_\mathrm{T}} |\hat{x}(t)|^2\, \epsilon_{r,t}^2(1/\mathrm{SNR}) + 1/\mathrm{SNR}}, \qquad \text{(B.17)}
\end{aligned}$$

which together with (B.1) and (4.4) yields the final upper bound

$$C_{\mathrm{PP}}(\mathrm{SNR}) \leq C_{\mathrm{PP}}^{(\mathrm{IID})}(\mathrm{SNR}) + \max_{\|\hat{\mathbf{x}}\|=1} \sum_{r=1}^{n_{\mathrm{R}}} \log \frac{1+1/\mathrm{SNR}}{\sum_{t=1}^{n_{\mathrm{T}}} |\hat{x}(t)|^2 \, \epsilon_{r,t}^2(1/\mathrm{SNR}) + 1/\mathrm{SNR}}. \tag{B.18}$$

B.2 Proof of Theorem 4.4

The proof of Theorem 4.4 follows along the same lines as the proof of Theorem 4.2 in Section B.1, but with spatially correlated fading.

As in Section B.1, we begin with the chain rule for mutual information

$$I\big(\mathbf{X}_1^n; Y_1^n\big) = \sum_{k=1}^{n} I\big(Y_k; \mathbf{X}_1^n \bigm| Y_1^{k-1}\big) \tag{B.19}$$

and upper bound each mutual information in the sum by

$$I\big(Y_k; \mathbf{X}_1^n \bigm| Y_1^{k-1}\big) \leq C_{\mathrm{PP}}^{(\mathrm{IID})}(\mathrm{SNR}) + \sup_{\|\mathbf{x}\|\leq \mathsf{A}} I\left(Y_0; \left\{\mathbf{H}_\ell + \frac{\sigma}{\mathsf{A}}\mathbf{W}_\ell\right\}_{\ell=-\infty}^{-1} \,\middle|\, \mathbf{X}_0 = \mathbf{x}_0\right), \tag{B.20}$$

see (B.2)–(B.4). (The random process $\{\mathbf{W}_k,\, k \in \mathbb{Z}\}$ is defined in Section B.1.) We continue by expressing the fading as

$$\mathbf{H}_0 = \overline{\mathbf{H}}_0 + \widetilde{\mathbf{H}}_0$$

(where $\overline{\mathbf{H}}_0$ and $\widetilde{\mathbf{H}}_0$ are given in (B.5)) and writing the second mutual information on the RHS of (B.20) as the difference of two differential entropies

$$\begin{aligned} I&\left(\big(\overline{\mathbf{H}}_0 + \widetilde{\mathbf{H}}_0\big)^{\mathsf{T}}\mathbf{X}_0 + Z_0; \left\{\mathbf{H}_\ell + \frac{\sigma}{\mathsf{A}}\mathbf{W}_\ell\right\}_{\ell=-\infty}^{-1} \,\middle|\, \mathbf{X}_0 = \mathbf{x}_0\right) \\ &= h\left(\big(\overline{\mathbf{H}}_0 + \widetilde{\mathbf{H}}_0\big)^{\mathsf{T}}\mathbf{X}_0 + Z_0 \,\middle|\, \mathbf{X}_0 = \mathbf{x}_0\right) \\ &\quad - h\left(\big(\overline{\mathbf{H}}_0 + \widetilde{\mathbf{H}}_0\big)^{\mathsf{T}}\mathbf{X}_0 + Z_0 \,\middle|\, \left\{\mathbf{H}_\ell + \frac{\sigma}{\mathsf{A}}\mathbf{W}_\ell\right\}_{\ell=-\infty}^{-1}, \mathbf{X}_0 = \mathbf{x}_0\right). \end{aligned} \tag{B.21}$$

To compute the first entropy on the RHS of (B.21), we note that, conditional on $\mathbf{X}_0 = \mathbf{x}_0$, the random variable $\big(\overline{\mathbf{H}}_0 + \widetilde{\mathbf{H}}_0\big)^{\mathsf{T}}\mathbf{X}_0 + Z_0$ has a Gaussian law of mean

$$\mathsf{E}\Big[\big(\overline{\mathbf{H}}_0 + \widetilde{\mathbf{H}}_0\big)^{\mathsf{T}}\mathbf{x}_0 + Z_0\Big] = \mathbf{d}^{\mathsf{T}}\mathbf{x}_0 \tag{B.22}$$

and of variance

$$\sigma_Y^2 = \mathbf{x}_0^{\mathsf{T}}\mathsf{K}\mathbf{x}_0^* + \sigma^2, \tag{B.23}$$

where $\mathsf{K} \triangleq \mathsf{E}\big[(\mathbf{H}_0 - \mathbf{d})(\mathbf{H}_0 - \mathbf{d})^\dagger\big]$. We thus have

$$h\Big(\big(\overline{\mathbf{H}}_0+\widetilde{\mathbf{H}}_0\big)^{\mathsf{T}}\mathbf{X}_0+Z_0 \,\Big|\, \mathbf{X}_0 = \mathbf{x}_0\Big) = \log(\pi e)+\log\big(\mathbf{x}_0^{\mathsf{T}}\mathsf{K}\mathbf{x}_0^*+\sigma^2\big). \tag{B.24}$$

To compute the second entropy on the RHS of (B.21), we note that, conditional on

$$\bigg(\Big\{\mathbf{H}_\ell + \frac{\sigma}{\mathsf{A}}\mathbf{W}_\ell\Big\}_{\ell=-\infty}^{-1}, \mathbf{X}_0\bigg) = \bigg(\Big\{\mathbf{h}_\ell + \frac{\sigma}{\mathsf{A}}\mathbf{w}_\ell\Big\}_{\ell=-\infty}^{-1}, \mathbf{x}_0\bigg),$$

the random variable $\big(\overline{\mathbf{H}}_0 + \widetilde{\mathbf{H}}_0\big)^{\mathsf{T}}\mathbf{X}_0 + Z_0$ has a Gaussian law of mean

$$\overline{\mathbf{h}}_0 + \mathbf{d}^{\mathsf{T}}\mathbf{x}_0 \tag{B.25}$$

and of variance

$$\sigma_{Y|\overline{h}}^2 = \mathbf{x}_0^{\mathsf{T}}\Sigma(1/\mathrm{SNR})\mathbf{x}_0^* + \sigma^2, \tag{B.26}$$

where $\Sigma(1/\mathrm{SNR})$ denotes the covariance matrix of the prediction error in predicting $\mathbf{H}_0$ from $\mathbf{H}_{-1} + \sigma/\mathsf{A}\,\mathbf{W}_{-1}$, $\mathbf{H}_{-2} + \sigma/\mathsf{A}\,\mathbf{W}_{-2}, \ldots$ We thus obtain

$$\begin{aligned} h\bigg(\big(\overline{\mathbf{H}}_0 + \widetilde{\mathbf{H}}_0\big)^{\mathsf{T}}\mathbf{X}_0 + Z_0 \,\bigg|\, \Big\{\mathbf{H}_\ell + \frac{\sigma}{\mathsf{A}}\mathbf{W}_\ell\Big\}_{\ell=-\infty}^{-1}, \mathbf{X}_0 = \mathbf{x}_0\bigg) \\ = \log(\pi e) + \log\big(\mathbf{x}_0^{\mathsf{T}}\Sigma(1/\mathrm{SNR})\mathbf{x}_0^* + \sigma^2\big), \end{aligned} \tag{B.27}$$

where we additionally use that the variance $\sigma_{Y|\overline{h}}^2$ does not depend on the realization of $\big\{\mathbf{H}_\ell + \sigma/\mathsf{A}\,\mathbf{W}_\ell\big\}_{\ell=-\infty}^{-1}$.

Applying (B.24) and (B.27) to (B.21) yields

$$
\begin{aligned}
&I\Bigl((\overline{\mathbf{H}}_0+\widetilde{\mathbf{H}}_0)^{\mathsf{T}}\mathbf{X}_0+Z_0;\Bigl\{\mathbf{H}_\ell+\frac{\sigma}{\mathsf{A}}\mathbf{W}_\ell\Bigr\}_{\ell=-\infty}^{-1}\,\Big|\,\mathbf{X}_0=\mathbf{x}_0\Bigr)\\
&\qquad=\log\frac{\mathbf{x}_0^{\mathsf{T}}\mathsf{K}\mathbf{x}_0^*+\sigma^2}{\mathbf{x}_0^{\mathsf{T}}\Sigma(1/\mathrm{SNR})\mathbf{x}_0^*+\sigma^2}\\
&\qquad=\log\frac{\hat{\mathbf{x}}_0^{\mathsf{T}}\mathsf{K}\hat{\mathbf{x}}_0^*+\sigma^2/\|\mathbf{x}_0\|^2}{\hat{\mathbf{x}}_0^{\mathsf{T}}\Sigma(1/\mathrm{SNR})\hat{\mathbf{x}}_0^*+\sigma^2/\|\mathbf{x}_0\|^2}\\
&\qquad\le\log\frac{\|\mathsf{K}\|+\sigma^2/\|\mathbf{x}_0\|^2}{\lambda_{\min}(1/\mathrm{SNR})+\sigma^2/\|\mathbf{x}_0\|^2}\\
&\qquad\le\log\frac{\|\mathsf{K}\|+1/\mathrm{SNR}}{\lambda_{\min}(1/\mathrm{SNR})+1/\mathrm{SNR}},
\end{aligned}
\tag{B.28}
$$

where $\hat{\mathbf{x}}_0=\mathbf{x}_0/\|\mathbf{x}_0\|$, and where $\lambda_{\min}(1/\mathrm{SNR})$ denotes the smallest eigenvalue of $\Sigma(1/\mathrm{SNR})$. Here the third step follows because

$$
\begin{aligned}
\hat{\mathbf{x}}^{\mathsf{T}}\mathsf{K}\hat{\mathbf{x}}^* &\le \|\mathsf{K}\|, & \|\hat{\mathbf{x}}\|&=1\\
\hat{\mathbf{x}}^{\mathsf{T}}\Sigma(1/\mathrm{SNR})\hat{\mathbf{x}}^* &\ge \lambda_{\min}(1/\mathrm{SNR}), & \|\hat{\mathbf{x}}\|&=1;
\end{aligned}
$$

and the last step follows because the function

$$
x\mapsto\log\frac{\alpha+x}{\beta+x},\quad \alpha\ge\beta
$$

is monotonically decreasing in x, and because, by the power constraint, $\|\mathbf{x}_0\|\le\mathsf{A}$. (We have $\|\mathsf{K}\|\ge\lambda_{\min}(1/\mathrm{SNR})$ because $\mathsf{K}-\Sigma(1/\mathrm{SNR})$ is positive semidefinite, so $\|\mathsf{K}\|\ge\|\Sigma(1/\mathrm{SNR})\|$ and consequently $\|\mathsf{K}\|\ge\lambda_{\min}(1/\mathrm{SNR})$.)

Combining (B.20) and (B.28), we obtain

$$
\begin{aligned}
&I\bigl(\mathbf{Y}_k;\mathbf{X}_1^n\,\big|\,\mathbf{Y}_1^{k-1}\bigr)\\
&\qquad\le C_{\mathrm{PP}}^{(\mathrm{IID})}(\mathrm{SNR})+\log\frac{\|\mathsf{K}\|+1/\mathrm{SNR}}{\lambda_{\min}(1/\mathrm{SNR})+1/\mathrm{SNR}},
\end{aligned}
\tag{B.29}
$$

which together with (B.19) and (4.4) yields the final upper bound

$$
C_{\mathrm{PP}}(\mathrm{SNR})\le C_{\mathrm{PP}}^{(\mathrm{IID})}(\mathrm{SNR})+\log\frac{\|\mathsf{K}\|+1/\mathrm{SNR}}{\lambda_{\min}(1/\mathrm{SNR})+1/\mathrm{SNR}}.
\tag{B.30}
$$

B.3 Proof of Theorem 4.13

An upper bound on the capacity pre-log of MIMO fading channels follows from the upper bound given in Theorem 4.2, namely,

$$C_{\mathrm{PP}}(\mathrm{SNR}) \leq C_{\mathrm{PP}}^{(\mathrm{IID})}(\mathrm{SNR}) + \max_{\|\hat{\mathbf{x}}\|=1} \sum_{r=1}^{n_{\mathrm{R}}} \log \frac{1+1/\mathrm{SNR}}{\sum_{t=1}^{n_{\mathrm{T}}} |\hat{x}(t)|^2\, \epsilon_{r,t}^2(1/\mathrm{SNR}) + 1/\mathrm{SNR}},$$

by computing its limiting ratio to $\log \mathrm{SNR}$ as SNR tends to infinity.

It follows from (4.10) that

$$\overline{\lim_{\mathrm{SNR}\to\infty}} \left\{ C_{\mathrm{PP}}^{(\mathrm{IID})}(\mathrm{SNR}) - \log\log \mathrm{SNR} \right\} < \infty$$

and hence

$$\overline{\lim_{\mathrm{SNR}\to\infty}} \frac{C_{\mathrm{PP}}^{(\mathrm{IID})}(\mathrm{SNR})}{\log \mathrm{SNR}} = 0. \qquad \text{(B.31)}$$

It thus suffices to show that

$$\overline{\lim_{\mathrm{SNR}\to\infty}} \frac{\max_{\|\hat{\mathbf{x}}\|=1} \sum_{r=1}^{n_{\mathrm{R}}} \log \frac{1+1/\mathrm{SNR}}{\sum_{t=1}^{n_{\mathrm{T}}} |\hat{x}(t)|^2\, \epsilon_{r,t}^2(1/\mathrm{SNR})+1/\mathrm{SNR}}}{\log \mathrm{SNR}} \leq \max_{1\leq t\leq n_{\mathrm{T}}} \sum_{r=1}^{n_{\mathrm{R}}} \mu(\{\lambda : F'_{r,t}(\lambda) = 0\}). \qquad \text{(B.32)}$$

Indeed, we have for any $\|\hat{x}\| = 1$

$$\begin{aligned} \sum_{r=1}^{n_{\mathrm{R}}} \log\left(\sum_{t=1}^{n_{\mathrm{T}}} |\hat{x}(t)|^2\, \epsilon_{r,t}^2(1/\mathrm{SNR}) + 1/\mathrm{SNR} \right) &= \sum_{r=1}^{n_{\mathrm{R}}} \log\left(\sum_{t=1}^{n_{\mathrm{T}}} |\hat{x}(t)|^2 \big(\epsilon_{r,t}^2(1/\mathrm{SNR}) + 1/\mathrm{SNR}\big) \right) \\ &\geq \sum_{r=1}^{n_{\mathrm{R}}} \sum_{t=1}^{n_{\mathrm{T}}} |\hat{x}(t)|^2 \log\Big(\epsilon_{r,t}^2(1/\mathrm{SNR}) + 1/\mathrm{SNR}\Big) \\ &= \sum_{r=1}^{n_{\mathrm{R}}} \sum_{t=1}^{n_{\mathrm{T}}} |\hat{x}(t)|^2 \int_{-1/2}^{1/2} \log\big(F'_{r,t}(\lambda) + 1/\mathrm{SNR}\big)\, \mathrm{d}\lambda, \end{aligned} \qquad \text{(B.33)}$$

where the second step follows from Jensen's inequality; and where the last step follows by applying the expression for the noisy prediction error (4.22).

With this, we obtain

$$
\begin{aligned}
&\varlimsup_{\mathrm{SNR}\to\infty} \frac{\max_{\|\hat{\mathbf{x}}\|=1} \sum_{r=1}^{n_\mathrm{R}} \log \frac{1+1/\mathrm{SNR}}{\sum_{t=1}^{n_\mathrm{T}} |\hat{x}(t)|^2\, \epsilon_{r,t}^2(1/\mathrm{SNR})+1/\mathrm{SNR}}}{\log \mathrm{SNR}} \\
&= \varlimsup_{\mathrm{SNR}\to\infty} \max_{\|\hat{\mathbf{x}}\|=1} \frac{-\sum_{r=1}^{n_\mathrm{R}} \log\Big(\sum_{t=1}^{n_\mathrm{T}} |\hat{x}(t)|^2\, \epsilon_{r,t}^2(1/\mathrm{SNR}) + 1/\mathrm{SNR}\Big)}{\log \mathrm{SNR}} \\
&\le \varlimsup_{\mathrm{SNR}\to\infty} \max_{\|\hat{\mathbf{x}}\|=1} \frac{-\sum_{r=1}^{n_\mathrm{R}} \sum_{t=1}^{n_\mathrm{T}} |\hat{x}(t)|^2 \int_{-1/2}^{1/2} \log\big(F'_{r,t}(\lambda) + 1/\mathrm{SNR}\big)\, \mathrm{d}\lambda}{\log \mathrm{SNR}} \\
&= \varlimsup_{\delta\downarrow 0} \max_{\|\hat{\mathbf{x}}\|=1} \sum_{r=1}^{n_\mathrm{R}} \sum_{t=1}^{n_\mathrm{T}} |\hat{x}(t)|^2 \frac{\int_{-1/2}^{1/2} \log\big(F'_{r,t}(\lambda) + \delta\big)\, \mathrm{d}\lambda}{\log \delta} \\
&= \varlimsup_{\delta\downarrow 0} \max_{\|\hat{\mathbf{x}}\|=1} \sum_{t=1}^{n_\mathrm{T}} |\hat{x}(t)|^2 \sum_{r=1}^{n_\mathrm{R}} \frac{\int_{-1/2}^{1/2} \log\big(F'_{r,t}(\lambda) + \delta\big)\, \mathrm{d}\lambda}{\log \delta} \\
&= \varlimsup_{\delta\downarrow 0} \max_{1\le t\le n_\mathrm{T}} \sum_{r=1}^{n_\mathrm{R}} \frac{\int_{-1/2}^{1/2} \log\big(F'_{r,t}(\lambda) + \delta\big)\, \mathrm{d}\lambda}{\log \delta} \\
&= \max_{1\le t\le n_\mathrm{T}} \sum_{r=1}^{n_\mathrm{R}} \varlimsup_{\delta\downarrow 0} \frac{\int_{-1/2}^{1/2} \log\big(F'_{r,t}(\lambda) + \delta\big)\, \mathrm{d}\lambda}{\log \delta} \\
&= \max_{1\le t\le n_\mathrm{T}} \sum_{r=1}^{n_\mathrm{R}} \mu(\{\lambda : F'_{r,t}(\lambda) = 0\}), \qquad \text{(B.34)}
\end{aligned}
$$

where the first step follows because

$$\lim_{\mathrm{SNR}\to\infty} \frac{\log(1+1/\mathrm{SNR})}{\log \mathrm{SNR}} = 0;$$

the second step follows from (B.33); the third step follows by substituting $\delta = 1/\mathrm{SNR}$; the fourth step follows by interchanging the order of summation; the fifth step follows because the unit vector that maximizes the sum is 1 for the largest summand and 0 for the other summands; the sixth step follows because maximization and summation are both taken over finite index sets, in which case it is valid to compute the limit first; and the last step follows by [26, Eq. (63)].

This concludes the proof.

B.4 Proof of Theorem 4.16

To derive an upper bound on Λ_{PP}, we begin with the upper bound (4.27), namely,

$$C_{\mathrm{PP}}(\mathrm{SNR}) \leq C_{\mathrm{PP}}^{(\mathrm{IID})}(\mathrm{SNR}) + \log \frac{1 + 1/\mathrm{SNR}}{\epsilon^2(1/\mathrm{SNR}) + 1/\mathrm{SNR}}. \tag{B.35}$$

It follows from (4.10) that

$$\varlimsup_{\mathrm{SNR}\to\infty} \frac{C_{\mathrm{PP}}^{(\mathrm{IID})}(\mathrm{SNR})}{\log\log \mathrm{SNR}} \leq 1. \tag{B.36}$$

By using the expression (4.22) for $\epsilon^2(1/\mathrm{SNR})$, we obtain for the second term on the RHS of (B.35)

$$\begin{aligned} &\varlimsup_{\mathrm{SNR}\to\infty} \frac{\log \frac{1+1/\mathrm{SNR}}{\epsilon^2(1/\mathrm{SNR})+1/\mathrm{SNR}}}{\log\log \mathrm{SNR}} \\ &\quad = \varlimsup_{\mathrm{SNR}\to\infty} \frac{-\int_{-1/2}^{1/2} \log\bigl(F'(\lambda) + 1/\mathrm{SNR}\bigr)\,\mathrm{d}\lambda}{\log\log\mathrm{SNR}} \\ &\quad = \varlimsup_{\delta\downarrow 0} \frac{-\int_{-1/2}^{1/2} \log\bigl(F'(\lambda) + \delta\bigr)\,\mathrm{d}\lambda}{\log\log\frac{1}{\delta}}. \end{aligned} \tag{B.37}$$

Combining (B.36) and (B.37) thus yields

$$\Lambda_{\mathrm{PP}} \leq 1 + \varlimsup_{\delta\downarrow 0} \frac{-\int_{-1/2}^{1/2} \log\bigl(F'(\lambda) + \delta\bigr)\,\mathrm{d}\lambda}{\log\log\frac{1}{\delta}}. \tag{B.38}$$

To derive a lower bound on Λ_{PP}, we evaluate the lower bound (4.30) in Proposition 4.1 for

$$\alpha^2 = \mathrm{SNR}^{-(1-\beta)} \quad \text{for some } 0 < \beta < 1,$$

which yields

$$\begin{aligned}
&C_{\mathrm{PP}}(\mathrm{SNR})\\
&\geq \log\frac{1}{\epsilon^2(\xi)+\xi} - \frac{1+\beta}{2}\log\mathrm{SNR}\\
&\quad - \exp\left(\frac{e}{(1-\beta)\,\log(\mathrm{SNR})\,\mathrm{SNR}^{1/2(1+\beta)}}\right)\\
&\qquad\qquad \times \mathrm{Ei}\left(-\frac{e}{(1-\beta)\,\log(\mathrm{SNR})\,\mathrm{SNR}^{1/2(1+\beta)}}\right)\\
&= \log\frac{1}{\epsilon^2(\xi)+\xi} + \log(1-\beta) + \log\log\mathrm{SNR} - 1\\
&\quad + \log\frac{e}{(1-\beta)\,\log(\mathrm{SNR})\,\mathrm{SNR}^{1/2(1+\beta)}}\\
&\quad - \exp\left(\frac{e}{(1-\beta)\,\log(\mathrm{SNR})\,\mathrm{SNR}^{1/2(1+\beta)}}\right)\\
&\qquad\qquad \times \mathrm{Ei}\left(-\frac{e}{(1-\beta)\,\log(\mathrm{SNR})\,\mathrm{SNR}^{1/2(1+\beta)}}\right)\\
&\geq \log\frac{1}{\epsilon^2(\xi)+\xi} + \log(1-\beta) + \log\log\mathrm{SNR} - 1 - \gamma,
\end{aligned} \tag{B.39}$$

where $\xi = \mathrm{SNR}^{-\beta}$. Here the second step follows by adding

$$\log\frac{e}{(1-\beta)\,\log(\mathrm{SNR})\,\mathrm{SNR}^{1/2(1+\beta)}} - 1 + \log(1-\beta) + \log\log\mathrm{SNR} + \frac{1+\beta}{2}\log\mathrm{SNR} = 0;$$

and the last step follows because the function

$$g(x) = \log(x) - \exp(x)\,\mathrm{Ei}(-x)\,, \quad x \geq 0$$

is monotonically increasing in x with $g(0) = -\gamma$ [28, Eqs. (210)–(213)].

We continue by showing that

$$\varlimsup_{\mathrm{SNR}\to\infty} \frac{-\log\bigl(\epsilon^2(\xi)+\xi\bigr)}{\log\log\mathrm{SNR}} = \varlimsup_{\delta\downarrow 0} \frac{-\int_{-1/2}^{1/2}\log\bigl(F'(\lambda)+\delta\bigr)\,\mathrm{d}\lambda}{\log\log\frac{1}{\delta}}. \tag{B.40}$$

Indeed, from (4.22) we obtain

$$\begin{aligned}
&\overline{\lim_{\text{SNR}\to\infty}} \frac{-\log\bigl(\epsilon^2(\xi)+\xi\bigr)}{\log\log\text{SNR}}\\
&= \overline{\lim_{\text{SNR}\to\infty}} \frac{-\int_{-1/2}^{1/2}\log\bigl(F'(\lambda)+\text{SNR}^{-\beta}\bigr)\,\mathrm{d}\lambda}{\log\log\text{SNR}}\\
&= \overline{\lim_{\text{SNR}\to\infty}} \left\{\frac{-\int_{-1/2}^{1/2}\log\bigl(F'(\lambda)+\text{SNR}^{-\beta}\bigr)\,\mathrm{d}\lambda}{\log\log\text{SNR}^{\beta}}\,\frac{\log\log\text{SNR}^{\beta}}{\log\log\text{SNR}}\right\}\\
&= \overline{\lim_{\delta\downarrow 0}} \frac{-\int_{-1/2}^{1/2}\log\bigl(F'(\lambda)+\delta\bigr)\,\mathrm{d}\lambda}{\log\log\frac{1}{\delta}},
\end{aligned}$$

where the last step follows by substituting $\delta = \text{SNR}^{-\beta}$ and because

$$\lim_{\text{SNR}\to\infty} \frac{\log\log\text{SNR}^{\beta}}{\log\log\text{SNR}} = 1.$$

Dividing the RHS of (B.39) by $\log\log\text{SNR}$ and computing the limit as SNR tends to infinity, yields

$$\begin{aligned}
\Lambda_{\text{PP}} &\geq \overline{\lim_{\text{SNR}\to\infty}} \frac{-\log\bigl(\epsilon^2(\xi)+\xi\bigr)}{\log\log\text{SNR}} + 1\\
&= 1 + \overline{\lim_{\delta\downarrow 0}} \frac{-\int_{-1/2}^{1/2}\log\bigl(F'(\lambda)+\delta\bigr)\,\mathrm{d}\lambda}{\log\log\frac{1}{\delta}},
\end{aligned} \tag{B.41}$$

where the last step follows from (B.40). Together with (B.38), this proves Theorem 4.16.

Appendix C

Appendix to Chapter 6

To prove (6.43), we lower bound

$$h\left(\sum_{\ell=0}^{k-1} H_k^{(\ell)} X_{k-\ell} + Z_k \,\middle|\, X_1^n = x_1^n, \mathbf{H}_1^{k-1} = \mathbf{h}_1^{k-1}\right) \tag{C.1}$$

for a given $\mathbf{h}_1^{k-1}$ and average then the result over $\mathbf{H}_1^{k-1}$. Let $\mathcal{H}_k$ denote the set

$$\mathcal{H}_k \triangleq \left\{H_k^{(\ell)}, 0 \leq \ell < k : \alpha_\ell = 0\right\}. \tag{C.2}$$

We have

$$\begin{aligned}
&h\left(\sum_{\ell=0}^{k-1} H_k^{(\ell)} X_{k-\ell} + Z_k \,\middle|\, X_1^n = x_1^n, \mathbf{H}_1^{k-1} = \mathbf{h}_1^{k-1}\right)\\
&\geq h\left(\sum_{\ell=0}^{k-1} H_k^{(\ell)} X_{k-\ell} + Z_k \,\middle|\, X_1^n = x_1^n, \mathbf{H}_1^{k-1} = \mathbf{h}_1^{k-1}, \mathcal{H}_k\right)\\
&= h\left(\sum_{\ell\in\mathcal{S}_k} H_k^{(\ell)} X_{k-\ell} + Z_k \,\middle|\, X_1^n = x_1^n, \mathbf{H}_1^{k-1} = \mathbf{h}_1^{k-1}, \mathcal{H}_k\right)\\
&\geq \log\left(\sum_{\ell\in\mathcal{S}_k} e^{h\left(H_k^{(\ell)} X_{k-\ell} \,\middle|\, X_1^n = x_1^n, \left\{H_{k'}^{(\ell)}\right\}_{k'=1}^{k-1} = \left\{h_{k'}^{(\ell)}\right\}_{k'=1}^{k-1}\right)} + e^{h(Z_k)}\right),
\end{aligned} \tag{C.3}$$

where $\mathcal{S}_k$ is defined in (6.42). Here the first step follows because conditioning cannot increase differential entropy; the second step follows because differential entropy is invariant under deterministic translation [5, Thm. 9.6.3] and because the terms where we have $x_{k-\ell} = 0$ do not contribute to the sum; and the last step follows by the entropy power inequality [5, Thm. 16.6.3] and because the processes

$$\left\{H_k^{(0)},\, k \in \mathbb{Z}\right\}, \left\{H_k^{(1)},\, k \in \mathbb{Z}\right\}, \ldots$$

are independent. (Note that, for a given $\mathbf{H}_1^{k-1} = \mathbf{h}_1^{k-1}$, the conditional entropies on the RHS of (C.3) are possibly infinite. However, by (6.6) this event is of zero probability and is therefore immaterial to (C.3) when averaged over $\mathbf{H}_1^{k-1}$.)

Since the processes of the path gains are independent and jointly independent of X_1^n, we can compute the expectation of (C.3) over $\mathbf{H}_1^{k-1}$ by averaging (C.3) first over $\big(H_1^{(0)}, \ldots, H_{k-1}^{(0)}\big)$, then averaging the result over $\big(H_1^{(1)}, \ldots, H_{k-1}^{(1)}\big)$, and so on. To lower bound the individual expectations, we note that the function

$$x \mapsto \log\big(e^x + \zeta\big) \tag{C.4}$$

is convex for all $\zeta \geq 0$. Thus, by defining for each $\ell' = 0, \ldots, k-1$

$$\zeta_{\ell'} \triangleq \sum_{\substack{\ell \in \mathcal{S}_k,\\ \ell < \ell'}} e^{h\left(H_k^{(\ell)} X_{k-\ell} \,\middle|\, X_1^n = x_1^n, \left\{H_{k'}^{(\ell)}\right\}_{k'=1}^{k-1}\right)} + \sum_{\substack{\ell \in \mathcal{S}_k,\\ \ell > \ell'}} e^{h\left(H_k^{(\ell)} X_{k-\ell} \,\middle|\, X_1^n = x_1^n, \left\{H_{k'}^{(\ell)}\right\}_{k'=1}^{k-1} = \left\{h_{k'}^{(\ell)}\right\}_{k'=1}^{k-1}\right)} + e^{h(Z_k)}, \tag{C.5}$$

it follows from Jensen's inequality

$$\begin{aligned} &\mathsf{E}_{\left\{H_{k'}^{(\ell')}\right\}_{k'=1}^{k-1}}\left[\log\left(e^{h\left(H_k^{(\ell')} X_{k-\ell'} \,\middle|\, X_1^n = x_1^n, \left\{H_{k'}^{(\ell')}\right\}_{k'=1}^{k-1} = \left\{h_{k'}^{(\ell)}\right\}_{k'=1}^{k-1}\right)} + \zeta_{\ell'}\right)\right] \\ &\quad \geq \log\left(e^{h\left(H_k^{(\ell')} X_{k-\ell'} \,\middle|\, X_1^n = x_1^n, \left\{H_{k'}^{(\ell')}\right\}_{k'=1}^{k-1}\right)} + \zeta_{\ell'}\right), \quad \ell' \in \mathcal{S}_k, \end{aligned} \tag{C.6}$$

where $\mathsf{E}_{\left\{H_{k'}^{(\ell')}\right\}_{k'=1}^{k-1}}$ denotes expectation with respect to $\left\{H_{k'}^{(\ell')}\right\}_{k'=1}^{k-1}$.

Averaging (C.3) over $\mathbf{H}_1^{k-1}$, and employing (C.6) to compute this average, yields thus

$$\begin{aligned} &h\left(\sum_{\ell=0}^{k-1} H_k^{(\ell)} X_{k-\ell} + Z_k \,\middle|\, X_1^n = x_1^n, \mathbf{H}_1^{k-1}\right) \\ &\quad \geq \log\left(\sum_{\ell \in \mathcal{S}_k} e^{h\left(H_k^{(\ell)} X_{k-\ell} \,\middle|\, X_1^n = x_1^n, \left\{H_{k'}^{(\ell)}\right\}_{k'=1}^{k-1}\right)} + e^{h(Z_k)}\right). \end{aligned} \tag{C.7}$$

This proves the lower bound (6.43).

List of Symbols

See also the remarks about notation in Section 1.3.

Elementary

i	imaginary unit, $\mathsf{i} = \sqrt{-1}$
$\triangleq$	definition
$\mathrm{Ei}(\cdot)$	exponential integral function
$\phi(\cdot)$	Euler's psi-function
γ	Euler's constant, $\gamma \approx 0.577$
$\Gamma(\cdot)$	Gamma function
$\lfloor\cdot\rfloor$	floor function
$\lceil\cdot\rceil$	ceiling function
$\mathrm{I}\{\text{statement}\}$	indicator function, is 1 if the statement is true and 0 otherwise
$\underline{A_m^n}$	sequence $A_m, A_{m+1}, \ldots, A_n$
$\overline{\lim}$	limit superior
$\underline{\lim}$	limit inferior
$\log(\cdot)$	natural logarithm function

Sets

$\mathbb{R}$	set of real numbers
$\mathbb{C}$	set of complex numbers
$\mathbb{Z}$	set of integers
$\mathbb{N}$	set of positive integers
$\mathbb{N}_0$	set of nonnegative integers

Vectors and Matrices

$\|\cdot\|$	Euclidean norm
$\|\cdot\|_{\mathrm{F}}$	Frobenius norm
$\det(\mathsf{A})$	determinant of A
$\mathrm{tr}\,(\mathsf{A})$	trace of A
A^*	complex conjugate of A
A^{T}	transpose of A
$\mathsf{A}^{\dagger}$	Hermitian transpose of A
I_n	$n \times n$ identity matrix

Probability and Information Theory

$\mathcal{U}\,(\mathcal{X})$	uniform probability distribution over the set $\mathcal{X}$
$\mathcal{N}(\boldsymbol{\mu}, \mathsf{K})$	Gaussian distribution of mean $\boldsymbol{\mu}$ and covariance matrix K
$\mathsf{E}[X]$	expectation of the random variable X
$\mathsf{E}_X[\cdot]$	expectation with respect to X
$\{X_k,\, k \in \mathbb{Z}\}$	stochastic process
I(X;Y)	mutual information between the random variables X and Y
$D(P\|Q)$	relative entropy between the probability distributions P and Q
$h(X)$	differential entropy of the random variable X

Bibliography

[1] B. E. S. Akgul, L. N. Chakrapani, P. Korkmaz, and K. V. Palem, "Probabilistic CMOS technology: A survey and future directions," in *Proceedings IFIP International Conference on VLSI*, October 2006.

[2] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: information-theoretic and communications aspects," *IEEE Transactions on Information Theory*, vol. 44, pp. 2619–2692, October 1998.

[3] J. R. Brown, *Ergodic Theory and Topological Dynamics.* Academic Press, 1976.

[4] R.-R. Chen, B. Hajek, R. Koetter, and U. Madhow, "On fixed input distributions for noncoherent communication over high-SNR Rayleigh-fading channels," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3390–3396, December 2004.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. John Wiley & Sons, 1991.

[6] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Academic Press, 1981.

[7] X. Deng and A. M. Haimovich, "Information rates of time varying Rayleigh fading channels," in *Proceedings IEEE International Conference on Communications (ICC)*, Paris, France, June 20–24, 2004.

[8] J. Doob, *Stochastic Processes.* John Wiley & Sons, 1990.

[9] A. Edelman, "Eigenvalues and condition numbers of random matrices," Ph.D. dissertation, Massachusetts Institute of Technology, May 1989.

[10] C. C. Enz and Y. Cheng, "MOS transistor modeling for RF IC design," *IEEE Journal of Solid State Circuits*, vol. 35, no. 2, pp. 186–201, February 2000.

[11] T. H. E. Ericson, "A Gaussian channel with slow fading," *IEEE Transactions on Information Theory*, vol. 16, no. 3, pp. 353–355, May 1970.

[12] R. Etkin and D. Tse, "Degrees of freedom in underspread MIMO fading channels," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Yokohama, Japan, June 29 – July 4, 2003.

[13] ——, "Degrees of freedom in some underspread MIMO fading channels," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1576–1608, April 2006.

[14] R. G. Gallager, *Information Theory and Reliable Communication.* John Wiley & Sons, 1968.

[15] K. E. Goodson, "Thermal conduction in electronic microstructures," in *CRC Handbook of Thermal Engineering*, 1st ed., ser. Mechanical Engineering Handbook Series, F. Keith, Ed., December 1999.

[16] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6th ed., A. Jeffrey, Ed. Academic Press, San Diego, 2000.

[17] A. Grant, "Rayleigh fading multi-antenna channels," *EURASIP Journal on Applied Signal Processing*, no. 3, pp. 316–329, March 2002.

[18] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications.* University of California Press, 1958.

[19] S. Ihara, *Information Theory for Continuous Systems.* World Scientific Publishing Co., 1993.

[20] Y.-H. Kim, "A coding theorem for a class of stationary channels with feedback," *IEEE Transactions on Information Theory*, vol. 54, no. 4, pp. 1488–1499, April 2008.

[21] L. B. Kish, "End of Moore's law: thermal (noise) death of integration in micro and nano electronics," *Physics Letters A*, no. 3–4, pp. 144–149, December 2002.

[22] T. Koch, "On the asymptotic capacity of multiple-input single-output fading channels with memory," Master's thesis, Signal and Information Processing Laboratory, ETH Zurich, Switzerland, April 2004, supervised by Prof. Dr. Amos Lapidoth.

[23] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Transactions on Information Theory*, vol. 42, pp. 1520–1529, September 1996.

[24] ——, "On phase noise channels at high SNR," in *Proceedings Information Theory Workshop (ITW)*, Bangalore, India, October 20–25, 2002.

[25] ——, "On the high SNR capacity of stationary Gaussian fading channels," in *Proceedings Forty-First Allerton Conference on Communication, Control and Computing*, Allerton House, Monticello, Illinois, October 1–3, 2003.

[26] ——, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 437–446, February 2005.

[27] ——, "On the high-SNR capacity of noncoherent networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3025–3036, September 2005.

[28] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2426–2467, October 2003.

[29] ——, "The fading number of single-input multiple-output fading channels with memory," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 437–453, February 2006.

[30] A. Lapidoth and S. Shamai (Shitz), "Fading channels: how perfect need 'perfect side-information' be?" *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[31] J. H. Lienhard IV and J. H. Lienhard V, *A Heat Transfer Textbook*, 3rd ed. Cambridge Massachusets, USA: Phlogisten Press, 2008.

[32] S. E. Lyshevski, *Nano and Molecular Electronics Handbook*, 1st ed. CRC Press, 2007.

[33] G. Maruyama, "The harmonic analysis of stationary stochastic processes," *Memoirs of the faculty of science, Series A*, vol. 4, no. 1, pp. 45–106, 1949.

[34] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139–157, January 1999.

[35] K. Petersen, *Ergodic Theory*, ser. Cambridge Studies in Advanced Mathematics 2. Cambridge University Press, 1983.

[36] E. C. Posner, "Random coding strategies for minimum entropy," *IEEE Transactions on Information Theory*, vol. 21, pp. 388–391, July 1975.

[37] B. Razavi, "CMOS technology characterization for analog and RF design," *IEEE Journal of Solid State Circuits*, vol. 34, no. 3, pp. 268–276, March 1999.

[38] W. Rudin, *Real and Complex Analysis*, 3rd ed. McGraw-Hill, 1987.

[39] V. Sethuraman and B. Hajek, "Capacity per unit energy of fading channels with a peak constraint," *IEEE Transactions on Information Theory*, vol. 51, pp. 3102–3120, September 2005.

[40] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and October 1948.

[41] B. Simon, *Orthogonal Polynomials on the Unit Circle. Part 1: Classical Theory.* American Mathematical Society, 2005.

[42] M. K. Simon, *Probability Distributions Involving Gaussian Random Variables: A Handbook for Engineers and Scientists.* Kluwer Academic Publishers, 2002.

[43] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, November–December 1999.

[44] F. Topsøe, "An information theoretical identity and a problem involving capacity," *Studia Scientarum Mathematicarum Hungarica*, vol. 2, pp. 291–292, 1967.

[45] Y. Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd ed. USA: Oxford University Press, 2003.

[46] R. Venkatesan, A. Kaloyeros, M. Beylansky, S. J. Souri, K. Banerjee, K. C. Saraswat, A. Rahman, R. Reif, and J. D. Meindl, "Interconnect limits on gigascale integration (GSI) in the 21st century," *Proceedings of the IEEE*, vol. 89, no. 3, pp. 305–324, March 2001.

[47] S. Verdú, "On channel capacity per unit cost," *IEEE Transactions on Information Theory*, vol. 36, pp. 1019–1030, September 1990.

[48] ——, "Spectral efficiency in the wideband regime," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[49] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.

[50] N. Wiener and P. Masani, "The prediction theory of multivariate stochastic processes I," *Acta Mathematica*, vol. 98, pp. 111–150, November 1957.

[51] W. Xu, "Firm lower bounds on the capacity of non-coherent fading channels," Semester Project, Signal and Information Processing Laboratory, ETH Zurich, Switzerland, 2005.

[52] L. Zheng and D. N. C. Tse, "Communicating on the Grassmann manifold: a geometric approach to the non coherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, February 2002.

[53] M. M. Ziegler and M. R. Stan, "The CMOS/nano interface from a circuits perspective," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, Bangkok, Thailand, May 25–28, 2003.

About the Author

Tobias Koch was born in Zurich, Switzerland on December 22, 1978. From 1992 to 1998 he attended the Kantonsschule in Wiedikon, Zurich, where he graduated with a Matura Type B (with Latin). An elective course on digital technology, which was offered at that time, sparked his interest in electrical engineering and disposed him to study this subject at the ETH in Zurich. During his studies, his areas of interest changed on a half-yearly bases and included high-frequency technology, optical communication, and signal processing, until he came to rest on the subject of communications and information theory, on which he also wrote his master's thesis. He graduated in April 2004 and received his M.Sc. in Electrical Engineering from ETH Zurich.

Having been partial to theoretical aspects of communications and signal processing, he decided to join the Signal and Information Processing Lab (ISI) at ETH, where he started his Ph.D. under the supervision of Prof. Amos Lapidoth in October 2004. Before that, he went to Bell Labs in New Jersey, USA for an internship. From his mentor there, Prof. Gerhard Kramer, he not only learned a great deal about information theory, but even more about doing research and writing. It is fair to say that he left for New Jersey as a raw rookie and returned as a beginner. During his time at the ISI he had the invaluable opportunity to visit Prof. Ezio Biglieri and Prof. Angel Lozano at the Universitat Pompeu Fabra in Barcelona.

Even more than information theory, he enjoys music (he plays piano and guitar), Salsa, traveling, and spending time with his friends.