

THE ROLE OF CORRELATION IN COMMUNICATION OVER FADING
CHANNELS

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Wenyi Zhang, B.S., M.S.

J. Nicholas Laneman, Director

Graduate Program in Electrical Engineering

Notre Dame, Indiana

July 2006

© Copyright by

Wenyi Zhang

2006

All Rights Reserved

THE ROLE OF CORRELATION IN COMMUNICATION OVER FADING CHANNELS

Abstract

by

Wenyi Zhang

In wireless communication systems with no channel state information (CSI) available at the transmitter or receiver, channel correlation plays a pivotal role in analyzing system behavior. This thesis explores the theme of channel correlation in several scenarios.

The first part of the thesis investigates a Rayleigh fading channel without correlation in time. Using a suitable logarithmic transformation, the channel model can be converted into an equivalent channel model with additive noise only. Such a perspective lends geometric intuition and often substantially reduces mathematical machinery to evaluate performance in terms of achievable rates and capacity. On one hand, a series of known results are revisited and reinterpreted; and on the other hand, several new results are obtained using this induced additive-noise perspective.

The second part of the thesis focuses on channel correlation in time. Using a block decision-feedback architecture, the original channel is decomposed into a series of parallel sub-channels, each being conditionally coherent linear Gaussian. Both discrete-time and continuous-time channels are examined, and the asymptotics of mutual information at low signal-to-noise ratio (SNR) are obtained. Compared to known capacity upper bounds under peak constraints, these asymptotics lead to

negligible loss in the low-SNR regime for slowly time-varying fading channels.

The third part of the thesis studies channel correlation in space, *i.e.*, across multiple transmit and receive antennas. In the low-SNR regime, it is shown using asymptotic analysis that spatially correlated antennas lead to both multiplicative rate gain as well as peak power reduction, at no cost of additional transmit power. Then a low-complexity communication scheme employing on-off signaling with hard-decision demodulation is evaluated, which demonstrates that most of the benefits promised by asymptotic analysis are realizable.

Finally, the last part of the thesis summarizes the main contributions and points out directions for future work.

To my parents and my wife

CONTENTS

FIGURES	v
ACKNOWLEDGMENTS	viii
CHAPTER 1: INTRODUCTION	1
1.1 Background Overview	1
1.2 Overview of the Thesis	8
CHAPTER 2: AN INDUCED ADDITIVE-NOISE CHANNEL MODELING APPROACH TO FADING CHANNELS WITHOUT CORRELATION . .	14
2.1 Introduction	14
2.2 Log-Scale Transform: from Fading Channel to Additive-Noise Channel	18
2.3 Channel Mutual Information	21
2.3.1 Conditions for Good Input Distributions	22
2.3.2 Log-Scale Continuous Uniform Inputs	24
2.3.3 Log-Scale Discrete Uniform Inputs	26
2.4 Multiple-Input Multiple-Output (MIMO) Channels	29
2.5 A Low-Complexity Scheme	33
2.5.1 Fundamental Limits of OOK	37
2.5.2 Measuring Gaussianness of the Normalized Induced Additive Noise \tilde{W}	39
2.5.3 Simulation Results	40
2.5.4 Robustness Considerations	42
2.6 Conclusion	44
CHAPTER 3: EXPLOITATION OF CHANNEL TEMPORAL CORRELA- TION	49
3.1 Introduction	49
3.2 Channel Model	51
3.3 An Achievable Rate and its Recursive Training Scheme Interpretation	53
3.3.1 An Achievable Rate Based on Average Channel Mutual Infor- mation	53
3.3.2 A Recursive Training Interpretation of the Achievable Rate R_{rt}	56
3.4 Channel Mutual Information at Low SNR	61

3.4.1	Comparison with a Capacity Upper Bound	63
3.4.2	Comparison with the High-SNR Channel Behavior	63
3.4.3	Case Study: Discrete-Time Gauss-Markov Fading Processes	67
3.5	Filling the Gap to Capacity by Widening the Input Bandwidth	70
3.5.1	An Intuitive Explanation of Proposition 3.5.1	75
3.5.2	Case Study: The Continuous-Time Gauss-Markov Fading Model	77
3.5.3	Case Study: Clarke's Fading Model	77
3.6	Extension to Rician Fading Channels	79
3.7	Concluding Remarks	83
CHAPTER 4: EXPLOITATION OF CHANNEL SPATIAL CORRELATION		
IN THE LOW-SNR REGIME 85		
4.1	Introduction	85
4.1.1	Connections with Related Research	86
4.2	Channel Model	87
4.3	Asymptotic Analysis in the Low-SNR Regime: First and Second	
	Derivatives	89
4.3.1	Interpretation of the Asymptotics	91
4.4	Non-Asymptotic Study of OOK with Symbol-by-Symbol Hard Decisions	95
4.4.1	Scalar Channel	98
4.4.2	Multiple Transmit Antennas and Transmit Correlation	100
4.4.3	Multiple Receive Antennas and Receive Correlation	101
4.5	Conclusion	106
CHAPTER 5: CONCLUSION 108		
.1	A Capacity Upper Bound Revisited	111
.2	Proof of Proposition 2.3.7	112
.3	Proof of Lemma 2.4.1	113
.4	Capacity per Unit Energy Under Peak Constraints for Rician Fading	
	Channels	115
.4.1	Proof of Proposition .4.1	116
.4.2	Interpretation of Proposition .4.1 and a Related Capacity Up-	
	per Bound	117
.4.3	Extension to Continuous-Time Channels	119
.5	Proof of Proposition 4.3.1	119
.6	Proof of Proposition 4.3.2	120
.7	PDF and CDF of Complex Gaussian Quadratic Forms with a Positive	
	Semi-Definite Kernel Matrix	121
BIBLIOGRAPHY 122		

FIGURES

2.1	The PDF of the induced additive noise W for the scalar channel case	20
2.2	Comparison of different channel mutual information and capacity bounds vs. SNR	26
2.3	Mutual information vs. SNR for L -ary LDU ($L = 2, 3, 4, 5$), LCU, and optimized OOK inputs	28
2.4	Cutoff rates R_0 vs. SNR for L -ary LDU ($L = 2, 3, 4, 5, \infty$) and optimized OOK inputs	29
2.5	Lower bounds of $I(U;T)$ (based on the entropy power inequality) under LCU inputs vs. SNR, for different N	34
2.6	Illustration of the memoryless signal conditioning front end at the receiver.	38
2.7	The channel mutual information for symmetric OOK and different receive diversity orders.	39
2.8	The PDF of the normalized induced additive noise \tilde{W} for different receive diversity orders when the symmetric OOK operates at 6 dB.	40
2.9	The Gaussianness metric $D(\tilde{W})$ of the normalized induced additive noise \tilde{W} for different receive diversity orders.	41
2.10	The bit error rate (BER) versus the channel SNR for different receive diversity orders. Vertical dashed-dot lines indicate the information-theoretic limits of the required channel SNR.	42
2.11	The mean and variance of the induced additive noise for the “on” input as functions of the LOS component d in Rician fading channels. The channel SNR is $\rho = 3$ dB. The case of $d = 0$ corresponds to Rayleigh fading channels.	45
2.12	The mean and variance of the induced additive noise for the “on” input as functions of the fading figure m in Nakagami fading channels. The channel SNR is $\rho = 3$ dB. The case of $m = 1$ corresponds to Rayleigh fading channels.	46

2.13	The mean and variance of the induced additive noise for the “on” input as functions of the channel correlation factor μ in Rayleigh fading channels with correlated receive branches. The channel SNR is $\rho = 3$ dB. The case of $\mu = 0$ corresponds to independent receive branches.	47
3.1	Illustration of the interleaving structure in the recursive training scheme. Input symbols are encoded/decoded column-wise, and transmitted/received row-wise.	58
3.2	Spectral density function of a deterministic fading process that leads to poor low-SNR performance. The narrow notches on the spectrum make the process deterministic, while the remaining almost unit spectrum makes it behave as if nearly memoryless in the low-SNR regime for large n	65
3.3	Spectral density function of an almost memoryless fading process that leads to good low-SNR performance. The almost unit spectrum makes the process nearly memoryless, while the narrow impulse-like spectrum peak significantly contributes to the integral $(1/2\pi) \cdot \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega})d\Omega$, leading to good low-SNR performance for large n	66
3.4	Illustration of the three operating regimes for the discrete-time Gauss-Markov fading model with $\epsilon = 10^{-4}$	69
3.5	Normalized rate R_{rt}/ρ vs. SNR for recursive training with QPSK on the discrete-time Gauss-Markov fading channel. For comparison, the dashed-dot curve is the channel capacity normalized by SNR with an average SNR constraint ρ and perfect receive CSI, achieved by circular complex Gaussian inputs.	71
3.6	Rate R_{rt} vs. SNR for recursive training with different PSK constellations on the discrete-time Gauss-Markov fading channel with $\epsilon = 10^{-4}$. For comparison, the dashed-dot curve is the channel capacity with an average SNR constraint ρ and perfect receive CSI, achieved by circular complex Gaussian inputs.	72
3.7	The asymptotic achievable rate $\lim_{T \rightarrow 0}(R_{rt}/T)$ vs. the envelope P , for recursive training with complex proper PSK on the continuous-time Gauss-Markov fading channel with innovation rate $\epsilon_c = 0.9$. The dashed-dot curves indicate the limiting behaviors for small and large P , <i>i.e.</i> , (3.5.20) and (3.5.15).	78
3.8	The asymptotic achievable rate $\lim_{T \rightarrow 0}(R_{rt}/T)$ vs. the envelope P , for recursive training with complex proper PSK on Clarke’s fading channel with maximum Doppler frequency $\omega_m = 100$. The dashed-dot curves indicate the limiting behaviors for small and large P , <i>i.e.</i> , (3.5.23) and (3.5.15).	80

4.1	Relative loss \mathcal{L} vs. normalized peak SNR $M\rho_p \cdot \lambda_{\max}(\Phi_T)$ for different receive correlation models: the constant correlation model and the Jakes model. The number of receive antennas is $N = 10$	92
4.2	$I(\rho)/\rho$ vs. ρ for the scalar channel case. The dashed-dot horizontal lines indicate the corresponding $\dot{I}(0)$ limit as $\rho \rightarrow 0$	94
4.3	The binary discrete memoryless channel (DMC) resulting from symbol-by-symbol hard decisions of on-off keying (OOK) of the original fading channel.	97
4.4	The maximum normalized mutual information and the corresponding optimal peak SNR for the scalar channel case. The curves for OOK with hard decisions is in solid, and those for OOK with soft decisions is in dashed-dot, as a comparison.	99
4.5	The optimal decision threshold τ vs. ρ for the scalar channel case. The dashed-dot curve is the threshold for the MAP decision rule. . .	100
4.6	The maximum normalized mutual information and the corresponding optimal peak SNR for M spatially fully correlated transmit antennas and a single receive antenna. The curve for $M = 1$ also corresponds to that for $M > 1$ uncorrelated transmit antennas.	102
4.7	The maximum normalized mutual information and the corresponding optimal peak SNR for a single transmit antenna and N receive antennas. The solid curves correspond to the spatially fully correlated case, and the dashed-dot curves correspond to the spatially uncorrelated case.	104
4.8	The maximum normalized mutual information and the corresponding optimal peak SNR for a single transmit antenna and $N = 2$ receive antennas. The correlation coefficient is $v = 0$ (uncorrelated), 0.5, 0.9, and 1 (fully correlated).	105

ACKNOWLEDGMENTS

First of all, I would like to thank my advisor Nick Laneman. Both my Ph.D. work (the major part of which is condensed into this thesis) and my personal development have benefited greatly from his guidance and insight. He has set up a figure for me to follow in my future career.

I would like to thank my thesis committee members, Professors Dan Costello, Tom Fuja, and Martin Haenggi. Their suggestions have helped to make this a better thesis.

Lots of thanks go to the ex and current JNL group fellows: Deqiang Chen, Shiva Kotagiri, Brian Dunn, Dawei Shen, and Michael Dickens; also to many other friends in the EE department, especially to Qiang Ling for his frequent help on research.

The financial support of the National Science Foundation, the State of Indiana, and the University of Notre Dame (both the Department of Electrical Engineering and the Center for Applied Mathematics), are also gratefully acknowledged.

Last but (definitely!) not least, I thank my family: my parents, grandparents, and my wife, Yi Shen, for their love and support. I could never have accomplished my Ph.D. study without them. There is simply no word to express my gratitude.

CHAPTER 1

INTRODUCTION

1.1 Background Overview

Recent years have witnessed an explosive increase of various communication applications operating over wireless media. Consequently, it becomes crucial to investigate the fundamental performance limits on reliable communication over wireless channels, in which the phenomenon of fading [4] poses critical challenges as well as promising opportunities to communication engineers. Fading is one of the major forms of volatility in a wireless environment, taking place in all dimensions, *i.e.*, time, frequency, and space. Fading channels as mathematical models have been systematically studied since the earliest era of communication theory and information theory; see [79], [63, Part 3], [62], [32] and references therein. As indisputable evidence of the important role of fading channels, an astonishing number of 549 references are listed in the well-known review paper [4] in 1998. Taking into account the burgeoning efforts devoted to space-time systems and wireless networks since the late 1990's, this number of references would be doubled if the survey were updated today.

On one hand, due to fading, transmitted signals can occasionally suffer from deep signal-to-noise ratio (SNR) attenuation, severe phase distortion, and inter-symbol interference, all of which notoriously degrade the quality of communication. On the other hand, by intelligently exploiting the inherent diversity of fading processes, the

chance of successful reception as well as the achievable rate of transmission can both be dramatically improved [71] [83] [34].

In order to combat the detrimental effects of fading and furthermore to exploit the inherent channel diversity, a critical issue is understanding the effect of channel state information (CSI) upon system behavior. Throughout the thesis, the term “channel state information” refers in particular to the realization of fading processes, which are often modeled as random processes. In system analysis, a pivotal distinction is whether or not CSI is accurately known at the receiver, and first-order models often replace accurate CSI with perfect CSI for analytical convenience. From a high-level perspective, when the receiver has perfect CSI, the study of fading channels boils down to investigating certain time-varying linear Gaussian channels, which are relatively well understood. In contrast, when CSI is not perfectly known at the receiver, the resulting channels are in general non-Gaussian, and neither channel capacity nor capacity-achieving input distributions are fully known. In the thesis, channels in which CSI is perfectly known at the receiver are referred to as *coherent*, otherwise they are referred to as *non-coherent*.

From a communication engineer’s point of view, all fading channels are inevitably non-coherent. Coherent channels, in which CSI is perfectly known, are approximations of channels with accurate channel identification and estimation. Conceptually, in adopting a coherent channel model, we essentially decouple communication into two stages, one of pure channel identification and estimation, the other of pure information transmission over the estimated channels, ignoring any residual channel estimation errors. When the inherent channel volatilities are mild, accurate estimation of CSI is usually feasible under acceptable cost of additional power and bandwidth, hence the coherent channel model is reasonably justified. However, for transceivers experiencing high mobility in a rich scattering propagation medium,

the channel volatilities usually become significant. As a consequence, the estimation error resulting from conventional algorithms may no longer be negligible. This channel uncertainty poses a series of fundamental difficulties in understanding fading channels, as commented in [48]:

“...As channels vary in time and frequency, the necessary finite amount of energy available to a system for measurement fundamentally limits the accuracy of channel estimation and, hence, what use we can make of such measurements. Moreover, in systems where very wide bands are used, we may in effect consider that, owing to decorrelation in frequency, we are dealing with not just one, but many channels. The applicability of capacity-achieving schemes, particularly in the asymptotic regime, is heavily affected by channel uncertainty. Channel measurement uncertainty is also often pointed to as the root cause of the difficulty of implementing, in practice, many schemes that are appealing from an information-theoretic perspective...”

For non-coherent fading channels, there exist several largely separate lines of work, due to the various settings of channel models. We briefly outline them in the sequel.

Detection-Theoretic Approach

This line of work dates back to the earliest study of fading channels [59] [30]. In the detection-theoretic approach, the major concern is to decide, with minimum error probability, which of two or more distinct waveforms was transmitted. In general it is assumed that CSI is unknown, except its statistics. Particularly, the auto-correlation function of the temporal fading process is exploited, and the resulting detection procedure turns out to have an estimator-correlator structure. The detection-theoretic approach is elegant, requiring a minimal amount of assumptions,

and leading to particularly simple implementations. Unfortunately, its connection to information-theoretic performance metrics, such as channel capacity, has yet to be completely established.

Orthogonal Modulation Schemes

This line of work is focused on orthogonal modulation, such as pulse-position modulation (PPM) and frequency-shift keying (FSK) [32]. In practice such schemes are usually attractive because the transmitted waveforms remain orthogonal in certain fading conditions, and the optimal receiver often comprises simple energy detectors. Furthermore, an important fact is that, as the channel bandwidth grows to infinity, orthogonal schemes with a vanishing duty cycle and an unbounded peak power turn out to be asymptotically capacity-achieving, and the limiting channel capacity is exactly the same as the capacity of a linear Gaussian channel of the same signal-to-noise ratio (SNR) in the wideband limit. In practice, however, the finiteness of available bandwidth tends to severely degrade the asymptotic wideband capacity of wideband fading channels.

The plaguing discrepancy between the asymptotic and actual channel behaviors is explained in terms of error exponent in [42], and in terms of channel capacity in [75]. In [75], Taylor expansions of channel mutual information at vanishing SNR are rigorously carried out. The infinite-bandwidth asymptotic channel capacity is captured in the first-order Taylor coefficient, which is shown to be identical regardless of the availability of CSI. For non-coherent fading channels, the first-order Taylor coefficient is maximized only if a so-called “flash” signaling is employed, which, roughly speaking, corresponds to the aforementioned orthogonal schemes with a vanishing duty cycle and an unbounded peak power. Refining the analysis in comparing the higher-order Taylor expansion terms for different channels, it is observed that, unless

CSI is perfectly known at the receiver, the second-order Taylor coefficient always approaches minus infinity in the wideband limit. Such diverging behavior of wideband, or low-SNR, channel capacity expansions reveals a fundamental discrepancy between coherent and non-coherent fading channels. However, from a practical perspective, it is of great importance to study the finer distinctions among non-coherent fading channels with different configurations, which are not fully characterized in the asymptotic analysis outlined above.

Memoryless Fading Channels

This line of work particularly treats an extreme case, *i.e.*, the so-called discrete-time memoryless Rayleigh fading channel model, in which the fading coefficient changes independently for adjacent channel uses. For this channel model, neither capacity nor the capacity-achieving input distributions are fully known. It is shown in [1] that the channel capacity is achieved using a discrete input distribution with a finite number of mass points, always including a mass point at zero. Unfortunately the capacity-achieving input distribution changes for different SNR, and can in general only be obtained through numerical nonlinear optimization. An important observation made rigorously in [1] is that, for sufficiently low SNR, the capacity-achieving input distribution reduces to on-off keying (OOK) with optimized peak power. This observation, to some extent, suggests the utilization of orthogonal modulation schemes for wideband channels, as discussed in the previous subsection.

Block Fading Channels

In block fading channel models, a generic assumption is that the fading process is piece-wise constant. That is, its realization remains constant for a block of T channel uses, and changes to independent realizations for the adjacent blocks. For the case of $T = 1$, the model reduces to the discrete-time memoryless fading chan-

nel model discussed above. However, the main interest has been for the case of $T > 1$. In a pioneering work [43], the basic structure of capacity-achieving input distributions is characterized, and the asymptotic channel behavior for the case of a single antenna is analyzed. A fundamental contribution of [43] is its recognition of unitary matrices as a good choice of channel input, especially at high SNR. This has led to a series of subsequent work on constellation design using unitary matrices [25] [26] [24]. In [82] it is further shown that for high SNR, unitary constellations asymptotically achieve the full multiplexing gain in capacity. The elegance of [82] is that it establishes a geometric perspective for non-coherent block fading channels, utilizing the sophisticated mathematical tool of the Grassmann manifold. In practical applications, the major concern about unitary constellations is the exponentially increasing complexity of the demodulator implementation, due to the lack of efficient representation of the constellations. In theoretic analysis, the major concern is the validity of the block fading model itself, especially as SNR grows to infinity, which motivates the following line of work focusing on high-SNR channel analysis.

Asymptotic Analysis at High SNR

In the block fading channel model, by letting the fading coefficient remain constant for a block of channel uses, we ignore the channel fluctuation within a block. Care should be taken in extending this approximate model to the regime of extremely high SNR, in which the impact of intra-block channel fluctuations is significantly amplified such that they can no longer be ignored. In [70], which investigates the discrete-time memoryless Rayleigh fading channel, it is first observed that the channel capacity grows double-logarithmically at high SNR. This double-logarithmic growth rate is then refined in [36] [35] for more general models such as Ricean fading, non-memoryless fading processes, and multiple-antenna channels.

It is shown that for very general channel models, the channel capacity behaves as $C = \log \log \rho + \chi + o(1)$ as the channel SNR $\rho \rightarrow \infty$, where χ is called the fading number, which varies depending upon the particular channel model. The double-logarithmic behavior is an alerting signal that one needs to be extremely careful in analyzing the asymptotic performance of non-coherent fading channels, which are highly sensitive to the choice of channel models. It should be noted that the double-logarithmic behavior is usually pessimistic; for example, as pointed out in [13] for typical slowly time-varying fading channels [32], the double-logarithmic behavior is effective only for impractically high SNR, *e.g.*, hundreds of dBs.

Pilot-Assisted Modulation and Coding

For slowly time-varying fading channels, explicit channel training using pilot symbols has long been a common practice; see [72] and references therein. Compared with “blind” approaches that adopt non-coherent coding and modulation techniques, pilot-assisted schemes are pseudo-coherent, and often simpler to implement. In a typical pilot-assisted training scheme, a sequence of pilot symbols known to the receiver is inserted sparingly among channel uses. Based upon the received pilot symbols, the receiver estimates the fading process, which is then treated as the true fading process to facilitate coherent reception. Since channel estimation cannot be error-free, in general the channel input affects the resulting effective noise, making it non-Gaussian and input-dependent. This imposes considerable difficulties in performance analysis of pilot-assisted schemes. Much effort has been devoted to this issue [38], [47], [3], [23], [49], [13]. In general, these existing works are focused on the case of circular complex Gaussian channel inputs, which are known to be capacity-achieving for coherent channels. Since the effective noise for Gaussian channel inputs is non-Gaussian and input-dependent, the channel mutual information is difficult

to evaluate explicitly. As a result, lower and upper bounds are usually derived, which may shed light on the actual capacity behavior. Perhaps the most important known rule of thumb so far is that, when the channel SNR is much less than the reciprocal of the variance of the channel estimation error, we may view the channel using pilot-assisted schemes as essentially coherent.

1.2 Overview of the Thesis

In this thesis, our aim is to further the existing work summarized in the previous section, and to gain more insight on the behavior of non-coherent fading channels. To this end, we focus on channel correlation as the main thread, which plays a pivotal role in characterizing distinct channel behaviors.

In non-coherent fading channels, abundant channel correlation is a prerequisite for realizing the potential performance gains promised by coherent channel analysis, since it provides the possibility for transceivers to keep track of the underlying fading processes, either explicitly or implicitly. As a thought experiment, consider transmitting information using pilot-assisted training over a block fading channel (*e.g.*, [23]). Inevitably, certain amounts of channel resources, such as bandwidth and power, need to be used by the pilot symbols, which facilitate channel estimation but carry no information. On one hand, if too little resource is devoted to channel training, the fading process is poorly tracked; hence, the effective noise level of the trained channel is high. On the other hand, if too much resource is devoted to channel training, the remaining channel resource cannot support the transmission of information. It is obvious that this tradeoff can be alleviated, for example, when the fading block length becomes large, *i.e.*, when the channel correlation becomes significant, or when the available resource is abundant, *e.g.*, when the channel SNR is sufficiently high.

Given the scope of the problem, even under the umbrella of channel correlation, a ubiquitous framework is by no means easy to establish. First of all, channel correlation may exist in all possible dimensions, like time, frequency, and space. Their underlying physical mechanisms are totally different, and difficult to model using simple and unified mathematical models. To describe the temporal and frequency-domain channel correlation, we can subscribe to the Jakes' model [28], the first-order Gauss-Markov model [47], or the block fading model [43], *etc.* Modeling issues associated with spectral and spatial channel correlation are even more complicated; see [32] [33] [77] and references therein.

Another difficulty preventing a ubiquitous treatment is that non-coherent fading channels exhibit qualitatively distinct behavior in different SNR regimes. The overall channel behavior therefore is jointly affected by both channel correlation and SNR. Intuitively, for high SNR, it tends to be more tolerable to use more power for channel training. This is because at high SNR the capacity increase from increasing SNR tends to be marginal, *i.e.*, logarithmic for coherent channels, whereas the channel estimation error decreases essentially in inverse proportion to the increased training power. On the other hand, for low SNR, the channel tends to be power-limited instead of bandwidth-limited. In that regime we are not able to afford precise channel training which consumes a large portion of scarce available power, and more sophisticated transmission schemes turn out to be necessary. Recently, the interaction of channel correlation and SNR is analyzed in the low SNR limit [84], by way of asymptotic scaling laws. To the best of our knowledge, so far no established result fully accounts for the non-asymptotic SNR regimes, which are of greater practical interest.

Due to the difficulties mentioned above, in this thesis we restrict ourselves to several special cases rather than the most general framework. Specifically, we inves-

tigate the following three distinct scenarios: discrete-time memoryless Rayleigh fading channels, discrete-time/continuous-time Rayleigh/Rician fading channels with correlation in time, and discrete-time memoryless Rayleigh fading channels with correlation in space. For each scenario, we adopt a different perspective and a correspondingly different approach, as will be revealed in the following chapters. In remainder of this introduction, we give some brief motivation for the three scenarios.

Discrete-Time Memoryless Rayleigh Fading Channels

This work may be the best motivated by the cepstral techniques in signal processing; refer to [52] and references therein. Cepstral techniques, and homomorphic analysis more generally, essentially convert the operation of convolution into the operation of addition. That is, for the input-output relationship

$$y(t) = h(t) \star x(t),$$

or

$$Y(j\omega) = H(j\omega)X(j\omega),$$

if we ignore the phase information and take logarithms on both sides, we obtain

$$\log |Y(j\omega)| = \log |H(j\omega)| + \log |X(j\omega)|,$$

an addition operation.

Now consider a multiplicative channel

$$X = S \cdot H,$$

where S is the channel input and H is a random multiplicative disturbance (fading).

If we take logarithms on both sides, we obtain

$$\log |X| = \log |S| + \log |H|,$$

which is analogous to the cepstral analysis above. Compared to fading channels, here in the multiplicative channel model we have not introduced the additive noise Z . In Chapter 2, we will see how to apply the same idea of transformation to convert a discrete-time memoryless Rayleigh fading channel into an induced additive-noise channel, and demonstrate a series of resulting applications.

Temporally Correlated Fading Channels

To motivate our approach to exploitation of channel temporal correlation, let us examine at a high level the key deficiency of conventional pilot-assisted training schemes. Again, such schemes divide the precious channel resource into two parts: pilots and information symbols. The pilots, used only for channel estimation, have nothing to do with information transmission. On the other hand, information symbols are typically not utilized for channel estimation, except in some joint channel estimation-coding schemes with high complexity. One way to improve upon the conventional solution, therefore, is to exploit the channel state information embedded in the information symbols.

As a simple example, consider a block of 3 channel uses within which the fading coefficient H is a fixed random variable, *i.e.*, block fading with $T = 3$, and the channel input-output relationship is

$$X_i = H \cdot S_i + Z_i, \quad i = 1, 2, 3.$$

In a conventional pilot-assisted training scheme, the first channel use is allocated to a pilot symbol, say, $S_1 = 1$. The first channel output, $X_1 = H + Z_1$, is then used by the receiver to estimate H . Consequently, the following two channel uses are allocated for information transmission, and the channel inputs S_2 and S_3 are detected coherently based upon the estimate \hat{H}_1 of H .

Can we improve upon the scheme described above? Our solution is a decision

feedback receiver that detects S_2 and S_3 in a successive manner. Based upon the channel estimate \hat{H}_1 , we coherently detect S_2 . Then we can treat the detected S_2 as a fresh pilot to update the estimate of H . Here assuming that S_2 is correctly detected, then we have two pilots, $S_1 = 1$ and S_2 . Therefore the corresponding estimate \hat{H}_2 is more precise compared to \hat{H}_1 , and consequently the detection of S_3 can be performed with higher reliability. In Chapter 3, we will apply this essential idea to investigate general temporally correlated fading channels, and in particular, demonstrate that near-optimal performance can be guaranteed in the low-SNR regime.

Spatially Correlated Fading Channels

In multiple-antenna communication systems, channel spatial correlation is usually viewed as detrimental. This observation stems from the fact that, when the channel SNR is high, channel spatial correlation decreases the degrees of freedom in spatial dimensions, thus reduces the multiplexing gain. For low-SNR communication, however, the situation is dramatically reversed. Spatially correlation among multiple antennas effectively boosts the powers in the channel as well as at the receiver, without spending additional transmit power or excess channel bandwidth. These effects turn out to be a key to improving communication efficiency in the low-SNR regime. To motivate our detailed analysis in Chapter 4, let us consider the following two simple examples.

In the first example, consider a memoryless fading channel with two transmit antennas and a single receive antenna,

$$X = H_1 \cdot S_1 + H_2 \cdot S_2 + Z,$$

where H_1 and H_2 are possibly correlated circular complex Gaussian random variables with zero mean and unit variance, and Z is also $\mathcal{CN}(0, 1)$. As SNR vanishes, OOK becomes asymptotically capacity-achieving [74]. Let us employ the following OOK

input: $(S_1, S_2) = (0, 0)$ or $(s/\sqrt{2}, s/\sqrt{2})$. If H_1 and H_2 are independent, the channel becomes

$$X = \sqrt{2} \cdot H \cdot S + Z,$$

where $H \sim \mathcal{CN}(0, 1)$ and $S = 0$ or s . Compared to a single-antenna channel, the effective channel SNR is doubled. On the other hand, if H_1 and H_2 are identical, *i.e.*, fully correlated, then the channel becomes

$$X = 2 \cdot H \cdot S + Z.$$

Here the effective channel SNR is increased by a factor of four, which even doubles that for the case of spatially uncorrelated antennas.

In the second example, consider a memoryless fading channel with a single transmit antenna and two receive antennas,

$$X_i = H_i \cdot S + Z_i, \quad i = 1, 2,$$

where H_1 and H_2 are possibly correlated circular complex Gaussian random variables with zero mean and unit variance, and Z_1 and Z_2 are $\mathcal{CN}(0, 1)$ and independent. Again let us employ OOK: $S = 0$ or s . If H_1 and H_2 are independent, the sufficient statistic of the channel output is the total output energy $|X_1|^2 + |X_2|^2$, so that compared to a single-antenna channel the effective channel SNR is doubled. On the other hand, if H_1 and H_2 are identical, *i.e.*, fully correlated, then the channel collapses to

$$(X_1 + X_2)/\sqrt{2} = \sqrt{2} \cdot H \cdot S + (Z_1 + Z_2)/\sqrt{2}.$$

So that the effective channel SNR is also doubled, or alternatively, we can save half of the average (as well as the peak) power in achieving the same performance.

CHAPTER 2

AN INDUCED ADDITIVE-NOISE CHANNEL MODELING APPROACH TO FADING CHANNELS WITHOUT CORRELATION

2.1 Introduction

In contrast to the thorough understanding of the additive white Gaussian noise (AWGN) channel, much less is understood about the non-coherent discrete-time memoryless Rayleigh fading channel. The simplest case is that for which the channel is scalar, *i.e.*,

$$X = S \cdot H + Z, \tag{2.1.1}$$

where $S \in \mathbb{C}$ is the channel input, and $X \in \mathbb{C}$ is the channel output. The fading coefficient $H \in \mathbb{C}$ and the additive noise $Z \in \mathbb{C}$ are both zero-mean circular complex Gaussian random variables, and independent for different channel uses. Hence in (2.1.1) we suppress all of the time indexes. Throughout the chapter we consider the case in which the realization of H , the fading coefficient, is available to neither the transmitter nor the receiver. For sake of simplicity and without loss of generality, we let H and Z both be of unit variance, and S be average-power-limited so that

$$\mathcal{E} [|S|^2] = P.$$

Consequently, the average signal-to-noise ratio (SNR) of the channel becomes

$$\rho = \frac{\mathcal{E} [|H|^2] \cdot \mathcal{E} [|S|^2]}{\mathcal{E} [|Z|^2]} = P.$$

Neither the capacity nor the capacity-achieving input distribution of the channel (2.1.1) is fully known. The authors of [1] prove that the capacity is achieved by a discrete input distribution with a finite number of mass points, including a mass point at $S = 0$. However, they do not provide any specific signaling design, partly because the proof technique employed is non-constructive in nature. Furthermore, the capacity-achieving input distribution changes for different SNR; the number of mass points, their locations, and their probabilities, all depend on the operating SNR, and can only be determined through numerical nonlinear optimization. In [70] the authors obtain upper and lower bounds on the channel capacity, constructing an explicit discrete input distribution to give the lower bound. They show that, at high SNR, the channel capacity grows double-logarithmically, that is, $C = \mathcal{O}(\log \log \rho)$ as $\rho \rightarrow \infty$. This double-logarithmic growth rate is refined in [36] [35] for more general models such as Ricean fading, non-memoryless fading processes, and multiple-antenna channels. It is shown that for very general channel models, the channel capacity behaves as $C = \log \log \rho + \chi + o(1)$ as $\rho \rightarrow \infty$, where χ is called the *fading number* of the channel. Another relevant work is [8], in which the authors derive necessary and sufficient conditions for continuous input distributions that lead to unbounded mutual information as SNR goes to infinity.

In addition to the high-SNR analyses mentioned above, for sufficiently low SNR, it is known that the channel capacity is achieved by on-off keying (OOK), for which the non-zero input has amplitude that becomes unbounded as SNR approaches zero [1] [75]. Numerical results [1] indicate that, for SNR less than roughly -3.5 dB, optimized OOK is capacity-achieving, and for SNR less than 10 dB, there is at most a 4% gap to capacity from the mutual information achieved by optimized OOK.

The main idea of this chapter stems from the following observation: As SNR approaches infinity, most of the channel randomness results from the fading coefficient

H rather than the additive noise Z . Intuitively we may approximate the limiting channel behavior at high SNR as $\log |X| = \log |S| + \log |H|$, *i.e.*, a channel with additive noise only. This limiting approximation is employed as a proof technique in [8]. In this chapter, we further observe that such an additive-noise channel interpretation holds for *all* SNR.¹ Specifically, by taking the logarithm of the channel output's magnitude $|X|$, we can decompose the transformed channel output into the sum of two terms. One term is merely a deterministic function of the channel input S , and the other term is independent of the channel input and can therefore be viewed as an additive noise. This equivalent additive-noise channel model is conceptually more intuitive to communication engineers, and yields a natural geometric interpretation for many earlier results. The effects of both the multiplicative fading and the additive noise in the original channel (2.1.1) are now described by a single induced additive noise. The inefficiency of communication over the original fading channel (2.1.1) can be intuitively understood as the log-scale transform rescales the original SNR ρ to roughly $\log \rho$. Consequently, the double-logarithmic capacity behavior immediately follows.

Remark: For memoryless Rayleigh fading channels, the additive-noise channel interpretation is not an entirely new idea. In addition to information-theoretic analyses, practical coding design has been investigated in [14] and [54] for the special case that the input S is symmetric OOK and all codewords have constant weight. In particular, [14] proposes another induced additive-noise channel model to simplify the receiver design, but this model only applies to the special case of constant weight OOK inputs.

We briefly summarize the main results of the chapter as follows.

¹The authors of [1] apparently observed the same additive-noise channel interpretation. However, they only used it implicitly in a proof [1, Appendix I, Lemma 2], and did not publish their other developments in this direction.

1) In Section 2.2 we derive the induced additive-noise channel model from the original fading channel model (2.1.1). The two channel models are equivalent, thus the problem is converted to communication over an additive-noise channel with non-Gaussian noise.

2) Based upon the induced additive-noise channel perspective, in Section 2.3 we establish new necessary and sufficient conditions for channel input distributions to achieve unbounded mutual information as SNR $\rho \rightarrow \infty$, and to asymptotically achieve the channel capacity. These conditions are qualitatively similar to those established in [8]. However, the induced channel model involves much less mathematical machinery, and appears to be more convenient to apply.

3) In Section 2.3 we further show that a simple input distribution called the log-scale (continuous) uniform distribution is asymptotically capacity-achieving, *i.e.*, it achieves both the double-logarithmic growth with SNR and the fading number. To provide practical signaling design, we re-interpret a class of discrete input distributions, which, in the induced channel perspective, turns out to be nothing but pulse amplitude modulation (PAM) with uniformly spaced pulses. For moderate SNR we demonstrate that these discrete input constellations with appropriate size achieve mutual information higher than that achieved by the log-scale (continuous) uniform input.

4) In Section 2.4 we further extend the induced channel modeling approach to memoryless multiple-antenna channels, possibly with transmit, but without receive correlation. It is shown that essentially all the developments for the scalar channel case can be paralleled for the multiple antenna case.

5) In Section 2.5 we utilize the induced additive-noise channel model to demonstrate a low-complexity scheme for communication over non-coherent memoryless Rayleigh fading channels. The logarithmic transformation can be easily imple-

mented as a memoryless signal conditioning front end at the receiver, and its outputs can then be approximately treated as that of an AWGN channel. When the receiver has multiple diversity branches, only a few parameters of the memoryless signal conditioning front end need be adjusted accordingly. Simulation results show that generic turbo codes designed for AWGN channels perform reasonably well for the scheme, and the scheme also exhibits a certain degree of robustness with respect to fading distributions other than Rayleigh.

In this chapter we frequently use Euler's constant $\gamma = 0.5772\dots$

2.2 Log-Scale Transform: from Fading Channel to Additive-Noise Channel

For the scalar channel (2.1.1), because the fading coefficient H totally distorts the phase information contained in the channel input S , we can restrict S to be real and non-negative. The channel output X has a sufficient statistic $|X|^2 = XX^\dagger$, *i.e.*, its energy, or equivalently, $|X|$, *i.e.*, its magnitude. Conditioned on $S = s$, we have $X \sim \mathcal{CN}(0, s^2 + 1)$, and its energy $R = |X|^2$ is exponentially distributed with probability density function (PDF)

$$f_{R|S}(r|s) = \begin{cases} \frac{1}{s^2+1} \exp\left(-\frac{r}{s^2+1}\right), & \text{if } r \geq 0 \\ 0, & \text{otherwise.} \end{cases} \quad (2.2.1)$$

Hence the channel $S \rightarrow R$ can be viewed in multiplicative form as

$$R = (S^2 + 1)V, \quad (2.2.2)$$

where V is independent of S and has PDF $f_V(v) = \exp(-v)$ for $v \geq 0$.

After taking the logarithm of $|X| = \sqrt{R} = \sqrt{S^2 + 1} \cdot \sqrt{V}$, we obtain an induced additive-noise channel equivalent to the original fading channel (2.1.1) as

$$T = U + W, \quad (2.2.3)$$

where:

1) The channel input is

$$U = \frac{1}{2} \log(S^2 + 1). \quad (2.2.4)$$

From the average power constraint for S , U should satisfy

$$\mathcal{E} [\exp(2U)] = \rho + 1, \quad (2.2.5)$$

and have support set $U \in [0, \infty)$.

2) The channel output is

$$T = \log |X|. \quad (2.2.6)$$

3) The additive noise $W = (\log V)/2$ is independent of U , and has PDF

$$f_W(w) = 2 \exp [2w - \exp(2w)], \quad \text{for } w \in (-\infty, \infty). \quad (2.2.7)$$

Remark: The induced channel (2.2.3) is equivalent to the normalized channel (2.1.1), in which the fading coefficient H and the noise Z are both $\mathcal{CN}(0, 1)$ random variables. For general parameterizations, *i.e.*, $H \sim \mathcal{CN}(0, \sigma_H^2)$ and $Z \sim \mathcal{CN}(0, \sigma_Z^2)$, the induced channel (2.2.3) still holds, except that the induced channel output T becomes

$$T = \log |X|^2 - \frac{1}{2} \log \sigma_Z^2;$$

and the induced channel input U becomes

$$U = \frac{1}{2} \log\left(\frac{\sigma_H^2}{\sigma_Z^2} S^2 + 1\right),$$

with constraint

$$\mathcal{E} [\exp(2U)] = \frac{\sigma_H^2}{\sigma_Z^2} \cdot P + 1,$$

and support set $U \in [0, \infty)$. The definition of W does not change under different parameterizations.

Figure 2.1 illustrates the PDF of the additive noise W . We observe that the noise density is not symmetric, and it decreases much slower for $w < 0$ than for $w > 0$. These qualitative observations are more precisely quantified by the following lemma.

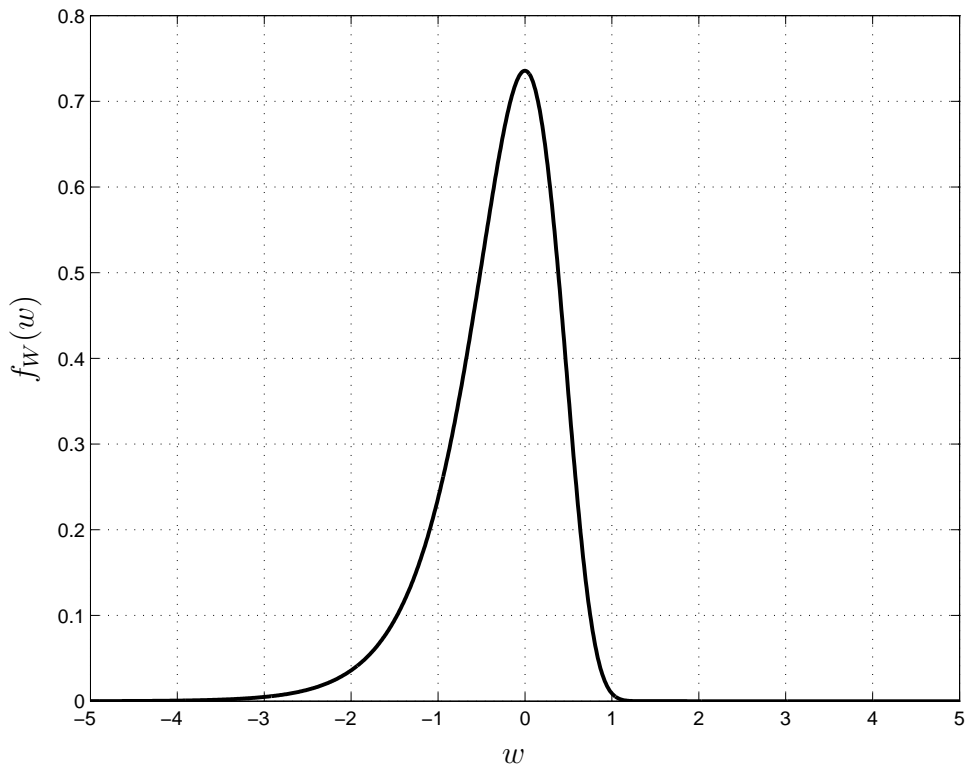


Figure 2.1. The PDF of the induced additive noise W for the scalar channel case

Lemma 2.2.1 (*Properties of the induced additive noise W , the scalar channel case*)
The additive noise W in (2.2.7) has the following properties:

1) The cumulative distribution function (CDF) of W is

$$F_W(w) = 1 - \exp(-\exp(2w)). \quad (2.2.8)$$

2) The characteristic function of W is

$$\varphi_W(j\theta) = \Gamma(1 + j\theta/2), \quad (2.2.9)$$

where $\Gamma(z) = \int_0^\infty x^{z-1} \exp(-x) dx$ is the gamma function [18].

3) The mean of W is

$$\mathcal{E}[W] = -\frac{\gamma}{2}, \quad (2.2.10)$$

and the maximum of $f_W(w)$ is attained at $w = 0$.

4) The variance of W is

$$\sigma_W^2 = \frac{\pi^2}{24}. \quad (2.2.11)$$

5) The differential entropy of W is

$$\begin{aligned} h(W) &= -\int f_W(w) \log f_W(w) dw \\ &= 1 - \log 2 + \gamma. \end{aligned} \quad (2.2.12)$$

Proof: By specializing to $N = 1$ in Appendix .3.

2.3 Channel Mutual Information

The induced additive-noise channel (2.2.3) is convenient for studying the channel behavior. In this section we investigate the channel mutual information, focusing on the high SNR regime.

An asymptotic characterization of the channel capacity is given by [36]

$$\lim_{\rho \rightarrow \infty} (C - \log \log \rho) = \chi, \quad (2.3.1)$$

where χ is called the *fading number*, and $\chi = -1 - \gamma$ for the scalar memoryless Rayleigh case.

From (2.2.3) we can write the channel mutual information as

$$\begin{aligned} I(S; X) &= I(U; T) \\ &= h(T) - h(T|U) = h(T) - h(W). \end{aligned} \quad (2.3.2)$$

From Lemma 2.2.1, $h(W) = 1 - \log 2 + \gamma$, hence finding channel capacity corresponds to maximizing the differential entropy of T . However, maximization of $h(T)$ is non-trivial because one of the constraints is that $T = U + W$, with the support set of U being $[0, \infty)$. In Appendix .1 we relax this constraint to revisit a capacity upper bound previously obtained in [70].

2.3.1 Conditions for Good Input Distributions

We have the following upper/lower bounds for any given distribution of the input U :

$$I(U;T) \leq \frac{1}{2} \log \sigma_U^2 + \log \sqrt{2\pi e \left(1 + \frac{\sigma_W^2}{\sigma_U^2}\right)} - h(W) \quad (2.3.3)$$

$$I(U;T) \geq h(U) + \log \sqrt{1 + e^{-2[h(U)-h(W)]}} - h(W). \quad (2.3.4)$$

The upper bound (2.3.3) is a straightforward application of the property that Gaussian distribution maximizes the entropy over all distributions with the same variance [10, Theorem 9.6.5], and the lower bound (2.3.4) is obtained using the entropy power inequality [10, Theorem 16.7.1]. According to the constraint (2.2.5), the channel input U implicitly depends on the SNR ρ . From (2.3.4) we observe that if $h(U) \rightarrow \infty$ as $\rho \rightarrow \infty$ then $I(U;T) \rightarrow \infty$; on the other hand, from (2.3.3) we observe that if $I(U;T) \rightarrow \infty$ as $\rho \rightarrow \infty$ then $\sigma_U^2 \rightarrow \infty$. We thus obtain necessary and sufficient conditions for unboundedness of channel mutual information as SNR scales.

Proposition 2.3.1 *For the channel (2.2.3), the channel input U in (2.2.4) achieves $I(U;T) \rightarrow \infty$ as $\rho \rightarrow \infty$ if*

$$h(U) \rightarrow \infty \quad \text{as } \rho \rightarrow \infty; \quad (2.3.5)$$

on the other hand, if U achieves $I(U;T) \rightarrow \infty$ as $\rho \rightarrow \infty$, then

$$\sigma_U^2 \rightarrow \infty \quad \text{as } \rho \rightarrow \infty. \quad (2.3.6)$$

Neither of the above sufficient and necessary conditions can be simultaneously necessary and sufficient. Two simple counterexamples using discrete channel inputs are given as follows. On one hand, since the channel capacity is always achieved by discrete inputs whose differential entropies are $-\infty$, $h(U) \rightarrow \infty$ is not necessary. On the other hand, a symmetric OOK input leads to unbounded variance as $\rho \rightarrow \infty$, but the resulting channel mutual information is always upper bounded by $\log 2$ nats.

It is useful to compare the conditions in Proposition 2.3.1 to those obtained in [8]. A necessary and sufficient condition is that $h(T) \rightarrow \infty$ as $\rho \rightarrow \infty$ [8, Theorem 4.3 (2)]. This condition is obvious from (2.3.2). On the other hand, a sufficient condition is that $h(\log |S|) \rightarrow \infty$ as $\rho \rightarrow \infty$ [8, Theorem 4.3 (3)]. This condition is obviously similar to the sufficient condition in Proposition 2.3.1, except for slightly different transforms of the channel input. Finally, there is no necessary condition in [8] that parallels the necessary condition in Proposition 2.3.1.

Proposition 2.3.1 immediately explains why circular complex Gaussian inputs perform poorly for the channel (2.1.1). In fact we have the following corollary.

Corollary 2.3.2 *If $S \sim \mathcal{CN}(0, \rho)$ in (2.1.1), then its induced input U in (2.2.4) has*

$$h(U) \rightarrow h(W) = 1 - \log 2 + \gamma \quad (2.3.7)$$

$$\sigma_U^2 \rightarrow \sigma_W^2 = \frac{\pi^2}{24}, \quad (2.3.8)$$

as $\rho \rightarrow \infty$.

Proof: Noting that for $S \sim \mathcal{CN}(0, \rho)$, $f_U(u) = \exp(1/\rho) \cdot f_W(u - \log \rho/2)$, the result then follows from straightforward manipulations using Lemma 2.2.1. **Q.E.D.**

Comparing the mutual information bounds (2.3.3, 2.3.4) and the capacity asymptote (2.3.1), we further obtain necessary and sufficient conditions for input distributions that asymptotically achieve capacity, *i.e.*, achieve both the double-logarithmic growth with SNR and the fading number.

Proposition 2.3.3 *In order to asymptotically achieve the channel capacity for the channel (2.2.3), *i.e.*,*

$$\lim_{\rho \rightarrow \infty} (C - I(U; T)) = 0,$$

a sufficient condition on the input U in (2.2.4) is

$$\lim_{\rho \rightarrow \infty} (h(U) - \log \log \rho) \geq -\log 2, \quad (2.3.9)$$

and a necessary condition is

$$\lim_{\rho \rightarrow \infty} \left(\sigma_U^2 - \frac{(\log \rho)^2}{8\pi e} \right) \geq -\sigma_W^2. \quad (2.3.10)$$

2.3.2 Log-Scale Continuous Uniform Inputs

Now we introduce the *log-scale continuous uniform* input distribution as follows.

Definition 2.3.4 *A log-scale continuous uniform (LCU) input U for the induced additive-noise channel (2.2.3) has PDF*

$$f_U(u) = \begin{cases} \frac{1}{A} & \text{if } 0 \leq u \leq A \\ 0 & \text{otherwise,} \end{cases} \quad (2.3.11)$$

where $A > 0$ is determined by

$$\frac{\exp(2A) - 1}{2A} - 1 = \rho. \quad (2.3.12)$$

For LCU inputs, the channel output T has PDF

$$\begin{aligned} f_T(t) &= \int f_U(u)f_W(t-u)du \\ &= \frac{(F_W(t) - F_W(t-A))}{A}. \end{aligned}$$

We can then numerically evaluate the corresponding $I(U;T)$. Furthermore, LCU distribution is asymptotically capacity-achieving.

Corollary 2.3.5 *For the induced additive-noise channel (2.2.3), LCU inputs achieve*

$$\lim_{\rho \rightarrow \infty} (C - I(U;T)) = 0. \quad (2.3.13)$$

Proof: The parameter A in the LCU distribution is related to the SNR through (2.3.12), so that

$$\rho = \frac{\exp(2A) - 1}{2A} - 1 < \frac{\exp(2A)}{2A} < \exp(2A),$$

where the last inequality holds if SNR is sufficiently large. Applying Proposition 2.3.3, we have

$$\begin{aligned} \lim_{\rho \rightarrow \infty} (h(U) - \log \log \rho) &= \lim_{\rho \rightarrow \infty} (\log A - \log \log \rho) \\ &\geq \lim_{\rho \rightarrow \infty} \left(\log \left(\frac{1}{2} \log \rho \right) - \log \log \rho \right) \\ &= -\log 2. \end{aligned}$$

Hence, LCU inputs are sufficient for asymptotically achieving channel capacity.

Q.E.D.

A related class of asymptotically capacity-achieving inputs is employed in [36]. Specifically, in [36] the input distribution is uniform for $\log |S|$, over an interval $[\log s_{\min}, (1/2) \log \rho]$, where s_{\min} escapes to infinity, while $\log s_{\min} / \log \rho$ vanishes, as SNR increases. In fact, we can show that LCU distribution is still asymptotically capacity-achieving even if its lower limit is chosen the same as $\log s_{\min}$. For finite SNR, however, there is a rate loss in increasing the lower limit from zero.

For LCU inputs in the high SNR limit, the lower bound (2.3.4) attains the channel capacity, and the upper bound (2.3.3) cannot be met, with a gap of $\log \sqrt{\pi e/6} \approx 0.1765$ nats. For finite SNR, both these bounds are useful in characterizing the achievable rate. In Figure 2.2, we plot the lower/upper bounds and actual mutual information for LCU inputs. For comparison, we also plot the numerically computed channel capacity at low SNR [1], the capacity upper bound from [70], the capacity upper bound from [36] based upon duality methods, and the capacity asymptote based upon the fading number (2.3.1) [36].

Several observations can be made based upon Figure 2.2. First, the actual mutual information for LCU inputs is close to its lower bound for low and high SNR, and its upper bound is a tight estimate for SNR between 10 and 30 dB. Second, at high SNR the gap between the capacity asymptote based upon the fading number (2.3.1) and the mutual information lower bound (2.3.4) is approximately 0.2 nats, even at $\rho \approx 80$ dB. This indicates that the approximation of capacity based on the fading number may under-estimate the actual capacity for a wide range of SNR. As a further example, these two curves remain separated by about 0.05 nats at $\rho \approx 400$ dB. Finally, at low SNR, there exists a significant rate loss from LCU inputs compared to optimized OOK, which is capacity-achieving in that regime.

This loss results because LCU inputs only yield $A = \rho + o(\rho)$ in the low SNR limit, which leads to a vanishing fourthegy [68] for the original fading channel (2.1.1).

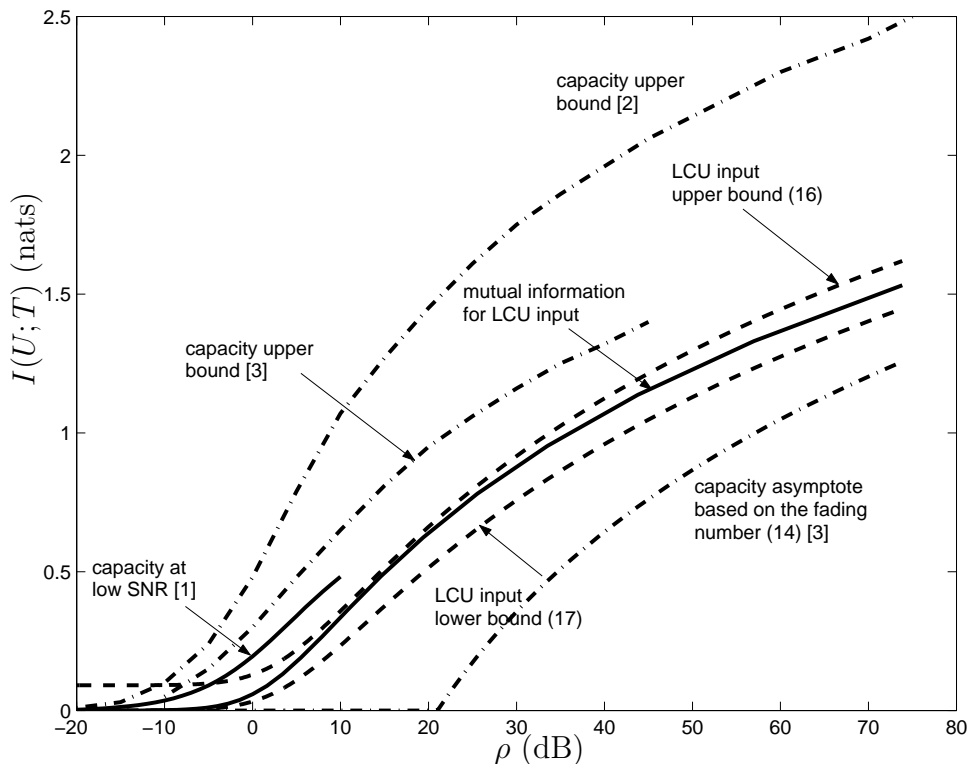


Figure 2.2. Comparison of different channel mutual information and capacity bounds vs. SNR

2.3.3 Log-Scale Discrete Uniform Inputs

In this section, we consider discrete input distributions. First, practical systems usually employ discrete signal sets as modulation constellations. Second, as shown in [1], the channel capacity is achieved by a discrete input distribution. Motivated by the LCU distribution, and the idea of maximizing minimum distance in the induced additive-noise channel, we introduce the following *log-scale discrete uniform* input distribution.

Definition 2.3.6 An L -ary log-scale discrete uniform (LDU) input U for the induced additive-noise channel (2.2.3) has probability mass function (PMF)

$$p_l := \text{Prob}(U = l\Delta) = \frac{1}{L}, \quad \text{for } l = 0, \dots, L-1, \quad (2.3.14)$$

where the spacing Δ is determined by SNR through the equation

$$\frac{\exp(2L\Delta) - 1}{\exp(2\Delta) - 1} = L(\rho + 1). \quad (2.3.15)$$

We note that LDU inputs are precisely those proposed in [5, Theorem 3], where it is shown that these inputs maximize the Kullback-Leibler (KL) distance between the conditional output PDFs. From the induced additive-noise channel perspective, an LDU input is simply log-scale PAM with uniformly spaced signals. For $L = 2$, an LDU input reduces to symmetric OOK, and for $L \rightarrow \infty$ it approximates an LCU input in the sense of convergence in distribution.

Figure 2.3 displays the results of numerically computing $I(U; T)$ for LDU inputs with different L . We observe that, at low SNR, binary inputs ($L = 2$) outperform inputs with larger L . As SNR increases beyond roughly 12.5 dB, inputs with $L = 3, 4, 5, \dots$ successively dominate. Although an LCU input is asymptotically capacity-achieving, for moderate SNR it is outperformed by LDU inputs. At low SNR, we also plot the channel capacity, achieved by optimized OOK [1]. We observe that the rate gain of optimized OOK over symmetric OOK can be significant at low SNR. Finally we note that the capacity lower bound obtained in [70] numerically approximates the upper envelope of all the curves in Figure 2.3.

For LDU inputs, we can also examine the cutoff rates for different constellation sizes, as given by the following proposition.

Proposition 2.3.7 For the induced additive-noise channel (2.2.3) with L -ary LDU inputs, the cutoff rate is

$$R_0 = \log L - \log \left[1 + \frac{4}{L} \sum_{l=0}^{L-1} \sum_{m=0}^{l-1} \frac{e^{-(l+m)\Delta}}{e^{-2l\Delta} + e^{-2m\Delta}} \right]. \quad (2.3.16)$$

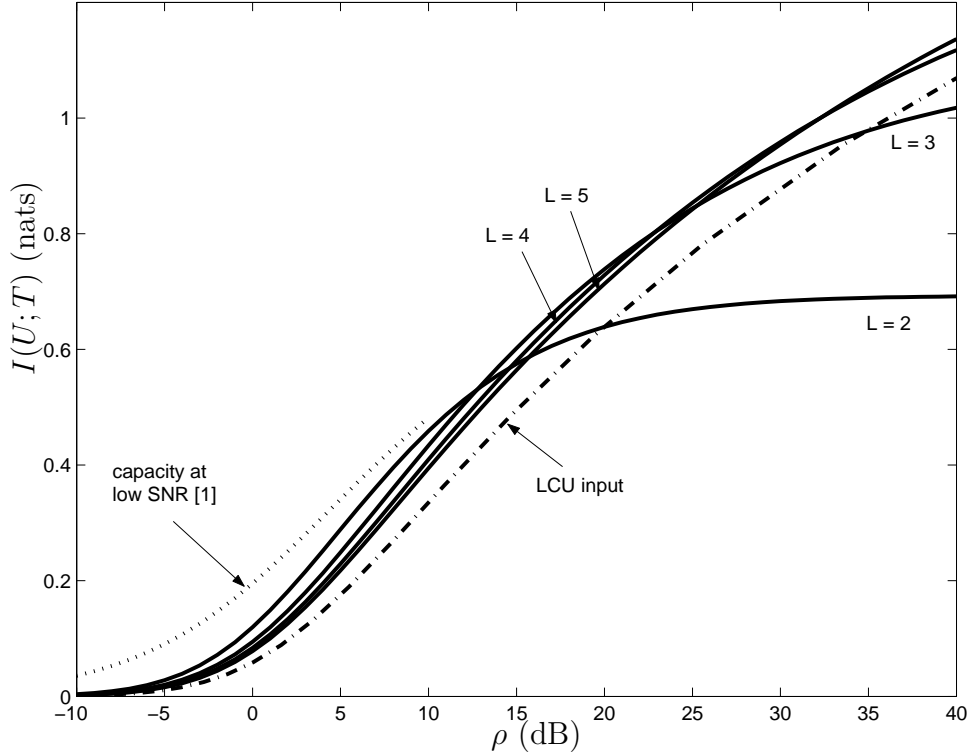


Figure 2.3. Mutual information vs. SNR for L -ary LDU ($L = 2, 3, 4, 5$), LCU, and optimized OOK inputs

Proof: In Appendix .2.

Figure 2.4 shows plots of R_0 vs. SNR for different L . We observe behavior similar to the plot of $I(U; T)$ vs. SNR in Figure 2.3. Here in terms of cutoff rates, the threshold SNR beyond which inputs with $L > 2$ outperform binary inputs is roughly 18.5 dB. For moderate SNR the curve for the $L \rightarrow \infty$ limit is again dominated by the curves corresponding to small L . At low SNR, we also plot the cutoff rate for optimized OOK. An interesting observation is that, in terms of cutoff rates, the gain of using asymmetric optimized OOK turns out to be rather limited, never exceeding 0.03 nats per channel use.

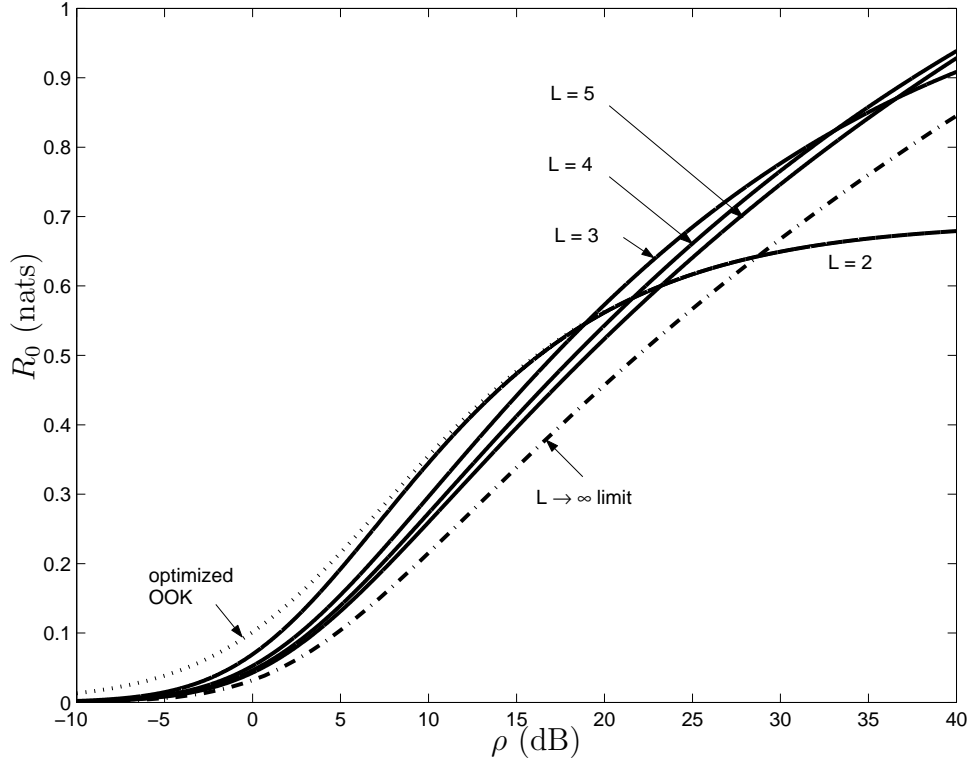


Figure 2.4. Cutoff rates R_0 vs. SNR for L -ary LDU ($L = 2, 3, 4, 5, \infty$) and optimized OOK inputs

2.4 Multiple-Input Multiple-Output (MIMO) Channels

In this section, we demonstrate how the induced additive-noise channel model can be extended to certain MIMO channels. Specifically, we assume that the channel is memoryless and the receive antennas are spatially uncorrelated. For an M -transmit N -receive antenna system, we adopt the channel model

$$\mathbf{X} = \mathbf{S} \cdot \mathbf{H} + \mathbf{Z}, \quad (2.4.1)$$

where:

- 1) $\mathbf{S} \in \mathbb{C}^{1 \times M}$ is the channel input, with an average power constraint

$$\mathcal{E}[\mathbf{S}\mathbf{S}^\dagger] = P.$$

- 2) $\mathbf{X} \in \mathbb{C}^{1 \times N}$ is the channel output.
- 3) $\mathbf{Z} \in \mathbb{C}^{1 \times N}$ is the additive noise vector, which is zero-mean circular complex Gaussian with independent unit-variance elements, *i.e.*, $\mathbf{Z} \sim \mathcal{CN}(0, \mathbf{I}_{N \times N})$.
- 4) $\mathbf{H} \in \mathbb{C}^{M \times N}$ is the fading matrix, which we consider to be of the form

$$\mathbf{H} = \mathbf{\Phi}_T^{1/2} \cdot \mathbf{H}_0,$$

where $\mathbf{H}_0 \in \mathbb{C}^{M \times N}$ is a random matrix in which the elements are zero-mean unit-variance circular complex Gaussian random variables that are mutually independent. $\mathbf{\Phi}_T$ is an $M \times M$ positive semi-definite matrix that characterizes the spatial correlation among the transmit antennas. We normalize the channel model so that the diagonal elements of $\mathbf{\Phi}_T$ are all unity.

For a given channel input vector $\mathbf{S} = \mathbf{s}$, the conditional channel output vector \mathbf{X} is circular complex Gaussian

$$\mathbf{X} | \mathbf{S} = \mathbf{s} \sim \mathcal{CN}(0, (1 + \mathbf{s}\mathbf{\Phi}_T\mathbf{s}^\dagger) \cdot \mathbf{I}_{N \times N}),$$

that is,

$$f_{\mathbf{X}|\mathbf{S}}(\mathbf{x}|\mathbf{s}) = \frac{1}{\pi^N \cdot (1 + \mathbf{s}\mathbf{\Phi}_T\mathbf{s}^\dagger)^N} \cdot \exp \left[\frac{\mathbf{x}\mathbf{x}^\dagger}{(1 + \mathbf{s}\mathbf{\Phi}_T\mathbf{s}^\dagger)} \right].$$

This implies that a scalar sufficient statistic for the channel is $R = \mathbf{X}\mathbf{X}^\dagger = \sum_{n=1}^N X_n X_n^\dagger$. Conditioned upon a given input $\mathbf{S} = \mathbf{s}$, R is the sum of N i.i.d. exponentially distributed random variables, *i.e.*, central chi-square distribution with $2N$ degrees of freedom [60],

$$f_{R|\mathbf{S}}(r|\mathbf{s}) = \begin{cases} \frac{r^{N-1}}{(1 + \mathbf{s}\mathbf{\Phi}_T\mathbf{s}^\dagger)^N (N-1)!} \cdot \exp \left[-\frac{r}{1 + \mathbf{s}\mathbf{\Phi}_T\mathbf{s}^\dagger} \right], & \text{if } r \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

Parallel to the scalar channel case, let us take the logarithm of \sqrt{R} . Then we obtain an equivalent sufficient statistic $T = \frac{1}{2} \log R$, for which the conditional PDF

is

$$f_{T|\mathbf{s}}(t|\mathbf{s}) = \frac{2}{(N-1)!} \exp \left\{ 2Nt - \log(\mathbf{s}\Phi_T\mathbf{s}^\dagger + 1) N - \exp[2t - \log(\mathbf{s}\Phi_T\mathbf{s}^\dagger + 1)] \right\}.$$

As before, this logarithmic transformation converts the fading channel (2.4.1) into an induced additive-noise channel

$$T = U + W, \quad (2.4.2)$$

where:

1) The channel input is

$$U = \frac{1}{2} \log(\mathbf{S}\Phi_T\mathbf{S}^\dagger + 1). \quad (2.4.3)$$

From the average power constraint of the fading channel (2.4.1), U should satisfy

$$\mathcal{E}[\exp(2U)] = \mathcal{E}[\mathbf{S}\Phi_T\mathbf{S}^\dagger] + 1 \quad (2.4.4)$$

For any given input random variable U , amplifying it by a constant factor greater than one does not decrease the channel mutual information, hence the channel capacity of (2.4.2) is achieved when the right hand side of (2.4.4) is maximized. This occurs when \mathbf{S}^\dagger is the eigenvector of Φ_T corresponding to the maximum eigenvalue $\lambda_{\max}(\Phi_T)$, which can be accomplished by beamforming at the transmitter. Thus in the sequel we take the input constraint as

$$\mathcal{E}[\exp(2U)] = \lambda_{\max}(\Phi_T)P + 1 = \rho + 1, \quad (2.4.5)$$

for the support set $U \in [0, \infty)$.

2) The channel output is

$$T = \frac{1}{2} \log \mathbf{X}\mathbf{X}^\dagger. \quad (2.4.6)$$

3) The additive noise W is independent of the channel input U , and has PDF

$$f_W(w) = \frac{2}{(N-1)!} \exp[2Nw - \exp(2w)] \quad (2.4.7)$$

for $w \in (-\infty, \infty)$.

Some basic properties of W are summarized as follows.

Lemma 2.4.1 (*Properties of W , the MIMO case*) *The additive noise W in (2.4.7) has the following properties:*

1) *The CDF of W is*

$$F_W(w) = 1 - \exp(-\exp(2w)) \sum_{n=0}^{N-1} \frac{\exp(2nw)}{n!}. \quad (2.4.8)$$

2) *The characteristic function of W is*

$$\varphi_W(j\theta) = \frac{1}{(N-1)!} \Gamma(N + j\theta/2). \quad (2.4.9)$$

3) *The mean of W is*

$$\mathcal{E}[W] = \frac{\psi(N)}{2}, \quad (2.4.10)$$

where $\psi(\cdot)$ is Euler's psi function [18]. Here for integer-valued N ,

$$\psi(N) = -\gamma + \sum_{n=1}^{N-1} \frac{1}{n} \sim \log N \quad \text{as } N \rightarrow \infty.$$

On the other hand, the maximum of $f_W(w)$ is attained at $w = \frac{\log N}{2}$.

4) *The variance of W is*

$$\sigma_W^2 = \frac{\zeta(2, N)}{4}, \quad (2.4.11)$$

where $\zeta(\cdot, \cdot)$ is Riemann's Zeta function [18]. Here for integer-valued N ,

$$\zeta(2, N) = \sum_{n=0}^{\infty} \frac{1}{(n+N)^2}.$$

5) *The differential entropy of W is*

$$h(W) = N(1 - \psi(N)) - \log 2 + \sum_{n=1}^{N-1} \log n. \quad (2.4.12)$$

Proof: In Appendix .3.

For the induced additive-noise channel (2.4.2), essentially all of the developments in Section 2.3 can be paralleled. The necessary and sufficient conditions for achieving

unbounded channel mutual information (Proposition 2.3.1) and for asymptotically achieving channel capacity (Proposition 2.3.3) directly apply without change. We can also study performance of LCU and LDU inputs. By noting that the fading number is $\chi = N(\psi(N) - 1) - \sum_{n=1}^{N-1} \log n$ [36], we find that LCU distribution is again asymptotically capacity-achieving.

Proposition 2.4.2 *For the induced additive-noise channel (2.4.2), an LCU input distribution achieves*

$$\lim_{\rho \rightarrow \infty} (C - I(U; T)) = 0. \quad (2.4.13)$$

In Figure 2.5 we plot the lower bound of $I(U; T)$ (based on the entropy power inequality) for LCU inputs vs. SNR, for different N . Increasing the number of receive antennas leads to quite noticeable rate gain, since the double-logarithmic capacity growth in SNR is rather limited for moderate SNR. By contrast, in coherent channels where the channel capacity grows logarithmically with SNR, the relative rate gain of merely increasing the number of receive antennas is not that significant.

2.5 A Low-Complexity Scheme

The non-coherent discrete-time memoryless Rayleigh fading channel can often be used to model scenarios with narrowband fast fading due to rich scattering and without line-of-sight (LOS) path. A few possible scenarios are briefly summarized as follows.

1) Communication systems that forego phase synchronization and instead employ sample-level interleaving plus non-coherent reception techniques. The purpose of interleaving is to virtually convert the possibly correlated fading process into a memoryless one. Communication systems forego phase synchronization and coherent reception mainly for two reasons: to reduce receiver complexity and additional

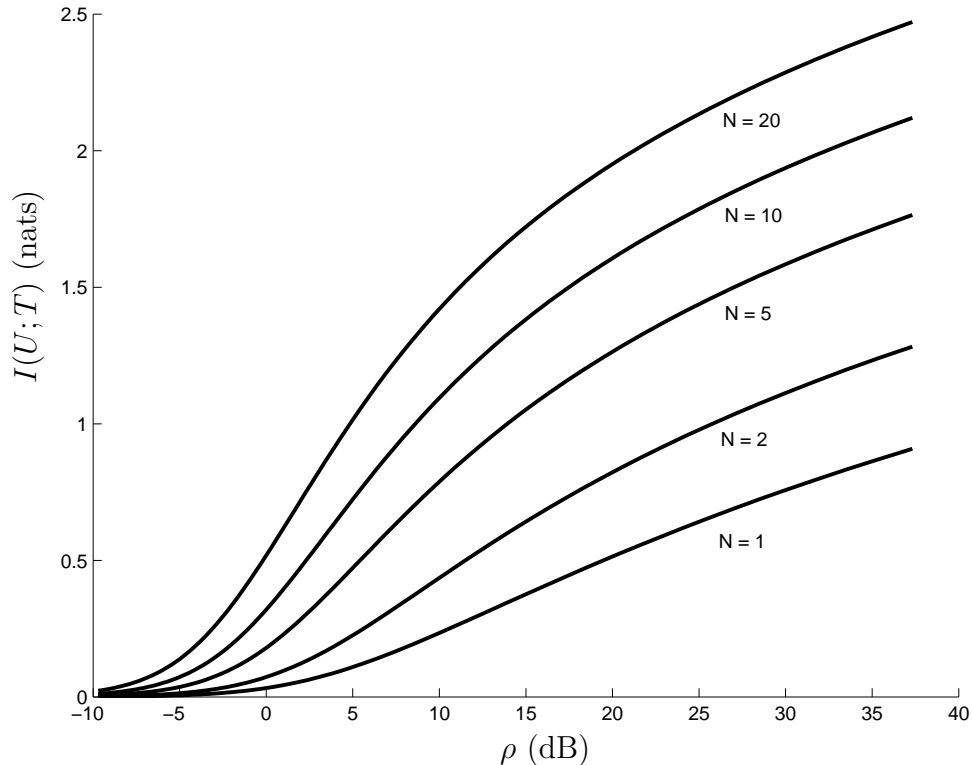


Figure 2.5. Lower bounds of $I(U; T)$ (based on the entropy power inequality) under LCU inputs vs. SNR, for different N

circuitry, for example, in wireless sensor networks based upon the Mote Systems [11] and other platforms in which sensor nodes have limited computing and hardware resources; or because the channel varies relatively fast so that common channel tracking algorithms may not perform well, for example, in certain mobile or aeronautical applications.

2) Fast frequency-hopping (FH) narrow-band systems in which the carrier frequency hops pseudo-randomly over a wide range of bandwidths for every discrete-time channel use, in order to guarantee communication security and resistance to jamming. Examples include certain military and other confidential applications, *e.g.*, Joint Tactical Radio System (JTRS) [29] and Situational Awareness System

for Small Unit Operations (SUOSAS) [69].

3) Time Division Multiple Access (TDMA) and other time-sharing systems in which the channel uses are separated widely enough that they experience essentially memoryless fades. Examples include systems with very sparse and bursty data arrivals, and systems consisting of a large number of communicating parties each of which has certain delay constraints, *e.g.*, Global System for Mobile (GSM) and General Packet Radio Service (GPRS) [61].

From the communication-theoretic viewpoint, it is difficult to develop efficient coded modulation schemes over the channel. For channels with low SNR, a common practice is orthogonal modulation which can be implemented by frequency-shift keying (FSK) or pulse-position modulation (PPM). As SNR approaches zero, orthogonal modulation with an infinite number of dimensions asymptotically achieves the channel capacity; see, *e.g.*, [32, 42]. For coded transmission, earlier work primarily investigates constant-weight codes with OOK based upon certain algebraic constructions; refer to [60, Ch. 14-6-4] and references therein. Another more recent relevant work is [53], where the authors design a coded modulation scheme that extracts soft information from the orthogonal modulation for turbo decoding. Compared to earlier results, the scheme provides significantly improved performance, especially for low information rates.

Let us consider a memoryless Rayleigh fading channel with a single transmit antenna and N independent receive branches. The channel model is

$$X_n = H_n \cdot S + Z_n, \quad \text{for } n = 1, \dots, N. \quad (2.5.1)$$

We assume that $H_n \sim \mathcal{CN}(0, \sigma_H^2)$ and $Z_n \sim \mathcal{CN}(0, \sigma_Z^2)$. Hence from the preceding development, we have an induced additive-noise channel model equivalent to the

original fading channel (2.5.1) as

$$T = U + W, \quad (2.5.2)$$

where

$$T := \frac{1}{2} \log \left(\sum_{n=1}^N |X_n|^2 \right) - \frac{1}{2} \log \sigma_Z^2 \quad (2.5.3)$$

$$U := \frac{1}{2} \log \left(\frac{\sigma_H^2}{\sigma_Z^2} |S|^2 + 1 \right), \quad (2.5.4)$$

and the PDF of W is given by (2.4.7).

In the following we focus on binary symmetric OOK for information transmission. Under the average power constraint, for channel model (2.5.1) the input alphabet is

$$S = \begin{cases} 0, & \text{w.p. } 1/2 \\ \sqrt{2P}, & \text{w.p. } 1/2 \end{cases}; \quad (2.5.5)$$

correspondingly for channel model (2.5.2) the input alphabet becomes

$$U = \begin{cases} 0, & \text{w.p. } 1/2 \\ \frac{1}{2} \log \left(\frac{2\sigma_H^2}{\sigma_Z^2} P + 1 \right), & \text{w.p. } 1/2 \end{cases}. \quad (2.5.6)$$

Therefore we further transform and scale (2.5.2) to obtain a normalized additive-noise channel

$$\tilde{T} = \tilde{U} + \tilde{W} \quad (2.5.7)$$

with antipodal inputs. To this end, we redefine the induced channel output as

$$\tilde{T} := \frac{4}{\log \left(\frac{2\sigma_H^2}{\sigma_Z^2} P + 1 \right)} \cdot \left[\frac{1}{2} \log \left(\sum_{n=1}^N |X_n|^2 \right) - \frac{1}{2} \log \sigma_Z^2 - \frac{\psi(N)}{2} \right] - 1. \quad (2.5.8)$$

The input alphabet then becomes

$$\tilde{U} = \begin{cases} -1, & \text{if } S = 0 \\ 1, & \text{if } S = \sqrt{2P} \end{cases}. \quad (2.5.9)$$

The normalized induced additive noise \tilde{W} is obtained from W as

$$\tilde{W} = \frac{4}{\log\left(\frac{2\sigma_H^2}{\sigma_Z^2}P + 1\right)} \cdot \left(W - \frac{\psi(N)}{2}\right), \quad (2.5.10)$$

which has mean zero and variance

$$\sigma_{\tilde{W}}^2 = \frac{4 \cdot \zeta(2, N)}{\left[\log\left(\frac{2\sigma_H^2}{\sigma_Z^2}P + 1\right)\right]^2}. \quad (2.5.11)$$

From (2.5.8) we observe that the normalized channel model (2.5.7) can be straightforwardly obtained by a memoryless signal conditioning front end at the receiver, as illustrated in Figure 2.6. Since the normalized output \tilde{T} is the addition of $\tilde{U} \in \{-1, 1\}$ and a (non-Gaussian) noise, we can treat the noise as Gaussian and employ an AWGN decoder to decode the message. Compared to maximum-likelihood (ML) decoding, the non-Gaussian channel and the AWGN decoder being mismatched will inevitably result to a certain performance loss. However, we stress that AWGN decoder is by far the most thoroughly studied and the most widely available. Furthermore, as will be shown in the following numerical simulation results, generic turbo codes designed for AWGN channels perform reasonably well for the induced additive-noise channel (2.5.7).

A thorough study of the behavior of mismatched decoding is beyond our scope here and many of the related problems still remain open; refer to [37] and references therein. In the following we therefore mainly rely on numerical simulation to obtain the performance of the channel (2.5.7) under AWGN decoding.

2.5.1 Fundamental Limits of OOK

Figure 2.7 plots the channel mutual information

$$I(\tilde{U}; \tilde{T}) = I(S; X_1, \dots, X_N),$$

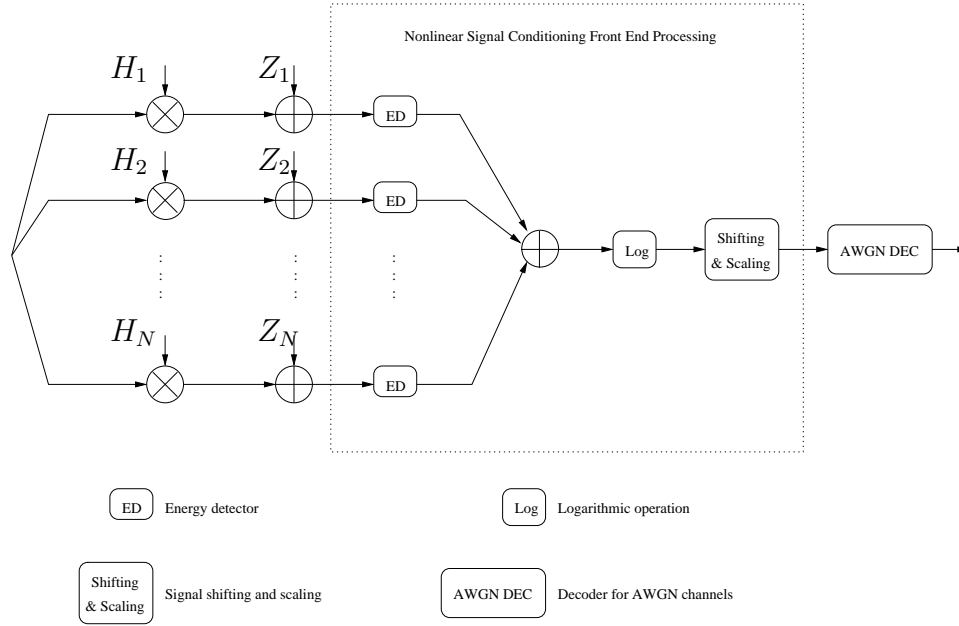


Figure 2.6. Illustration of the memoryless signal conditioning front end at the receiver.

which is the maximum achievable rate of the channel (2.5.1) under OOK, versus the channel SNR

$$\rho := \frac{\sigma_H^2}{\sigma_Z^2} \cdot P$$

for different N , the receive diversity order. We observe that, for the SNR range plotted, increasing N results to a significant performance gain. The source of the performance gain is partially reflected in Figure 2.8, which plots the PDF of the normalized induced additive noise \tilde{W} for different N when the symmetric OOK operates at 6 dB. It is clearly indicated in Figure 2.8 that as N increases, the essential support of the noise density shrinks and its shape becomes more symmetric which resembles a Gaussian density function.

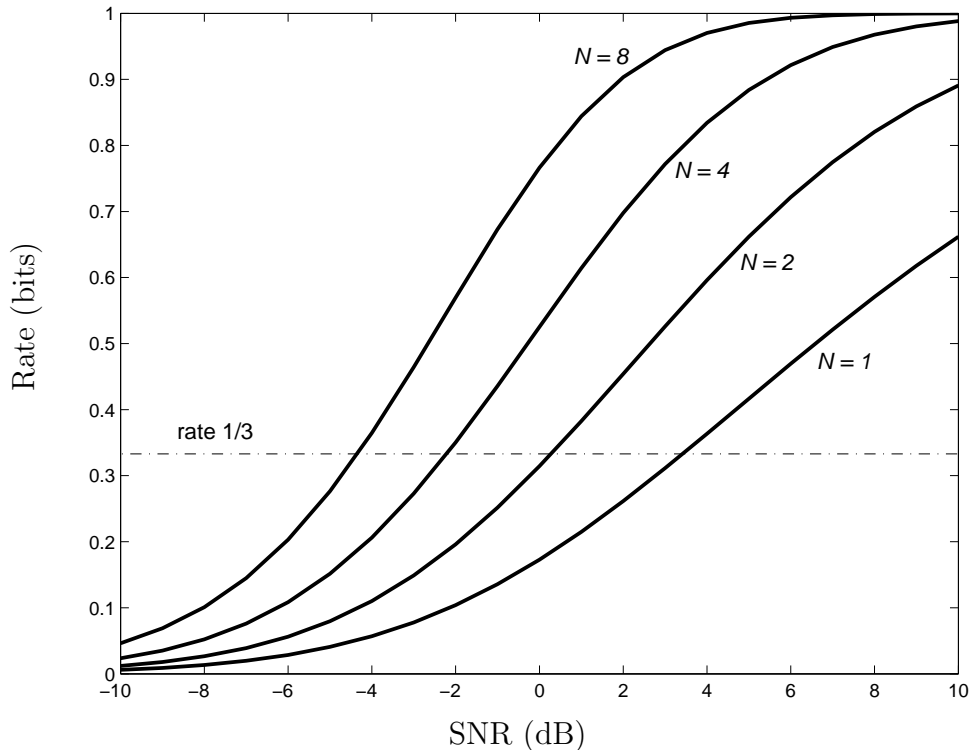


Figure 2.7. The channel mutual information for symmetric OOK and different receive diversity orders.

2.5.2 Measuring Gaussianness of the Normalized Induced Additive Noise \tilde{W}

Since we employ an AWGN decoder for the non-Gaussian channel (2.5.7), it will be intuitively pleasing to have a quantitative metric for measuring the Gaussianness of the normalized induced additive noise \tilde{W} . Here we adopt a metric $D(X)$ proposed in [56], which is defined for a random variable X as the following KL distance

$$D(X) := \mathcal{D}(P_X \parallel \Phi_X), \quad (2.5.12)$$

where P_X is the distribution of X , and Φ_X is the distribution of a Gaussian random variable with the same mean and variance as X . Obviously, $D(X) \geq 0$ with equality if and only if X is Gaussian. For the normalized induced additive noise \tilde{W} , we find

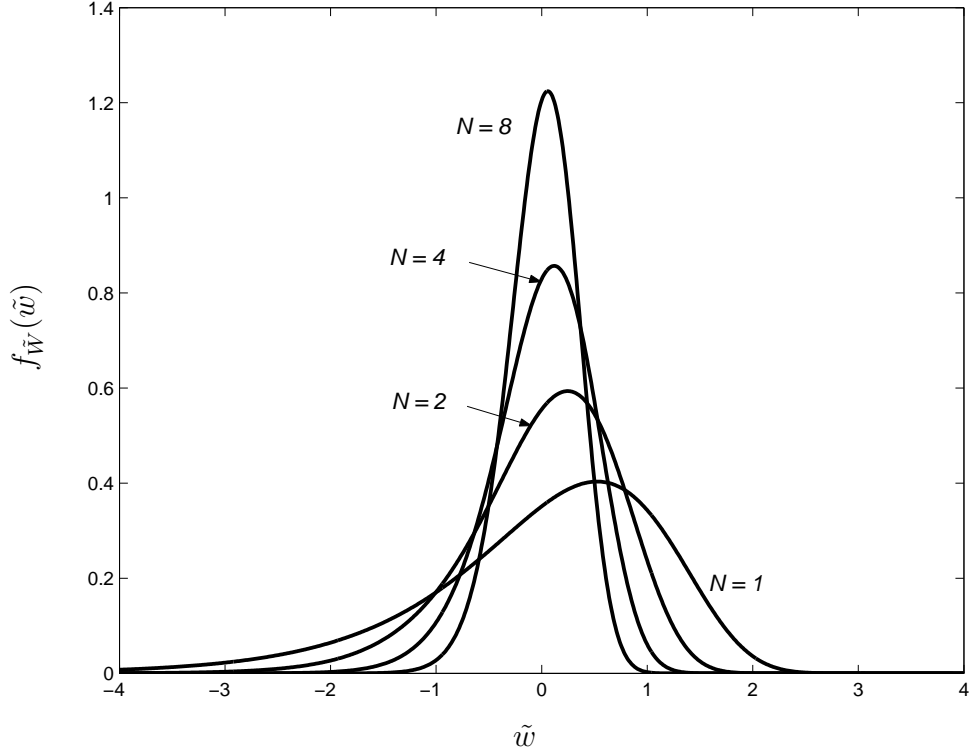


Figure 2.8. The PDF of the normalized induced additive noise \tilde{W} for different receive diversity orders when the symmetric OOK operates at 6 dB.

that

$$D(\tilde{W}) = \frac{1}{2} + N \cdot \psi(N) + \log \frac{\sqrt{2\pi\zeta(2, N)}}{(N-1)!} - \mathbf{E}_{\tilde{W}} \left[\exp \left(\frac{\log(2\rho + 1)}{2} \tilde{W} + \psi(N) \right) \right] \quad (2.5.13)$$

Figure 2.9 plots the Gaussianness metric $D(\tilde{W})$ versus the channel SNR ρ for different N . It is clearly indicated that unless SNR is extremely low, $D(\tilde{W})$ remains essentially constant and is typically small especially for $N > 1$. Hence it may be a justified option to adopt AWGN decoding for the non-Gaussian channel (2.5.7), as we will further illustrate through numerical simulation in the following.

2.5.3 Simulation Results

In simulation we choose a generic rate 1/3 parallel concatenated turbo codec designed for AWGN channels. The two component codes are generated by identical

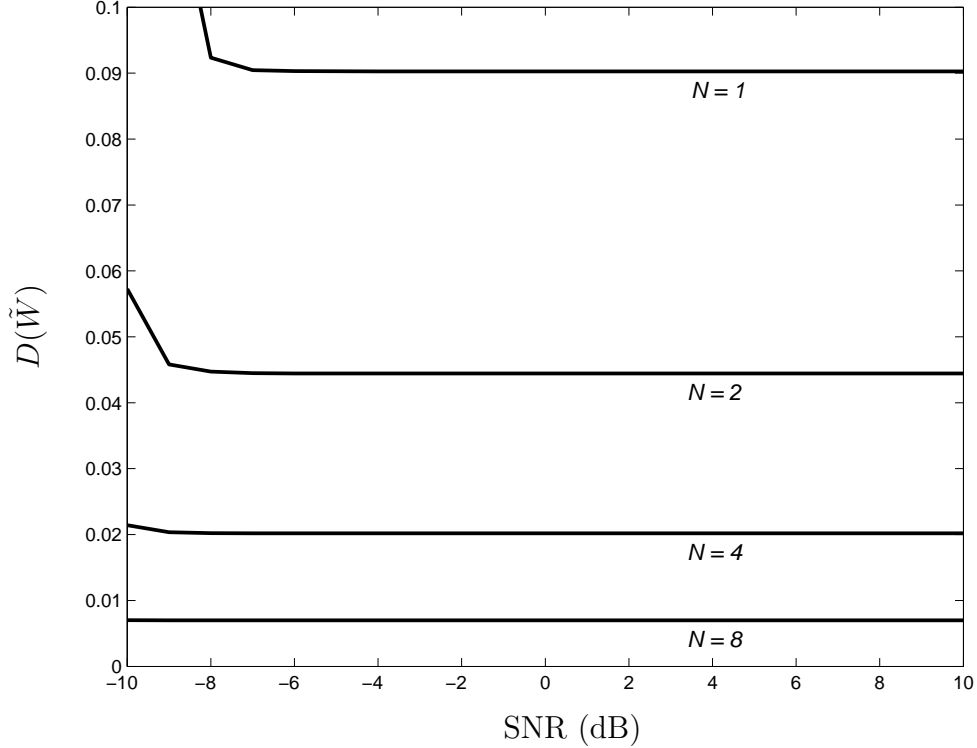


Figure 2.9. The Gaussianness metric $D(\tilde{W})$ of the normalized induced additive noise \tilde{W} for different receive diversity orders.

recursive systematic encoders with generator matrix

$$\begin{bmatrix} 1 & \frac{1 + D^2 + D^3 + D^4}{1 + D + D^4} \end{bmatrix}.$$

For simplicity, we employ a uniformly random interleaver with depth $L = 1024$. Thus the codeword block length is $3L = 3096$ with 1024 information symbols and 2048 parity check symbols. Figure 2.10 plots the bit error rate (BER) of the turbo codec versus the channel SNR ρ for different N . For comparison the information-theoretic limits of the required channel SNR are also indicated, which can be read from Figure 2.7. We observe that to achieve BER of 10^{-5} , the required SNR margin beyond the information-theoretic limits is typically within 1–3 dB, and it decreases as N increases, a phenomenon not fully surprising because as N increases the nor-

malized induced additive noise \tilde{W} tends to be more Gaussian as shown in Section 2.5.2. Furthermore, for $N = 8$ the margin is within 1 dB. Considering the relatively short coding block length, we conclude that the AWGN decoder performs reasonably well for the non-Gaussian channel (2.5.7).

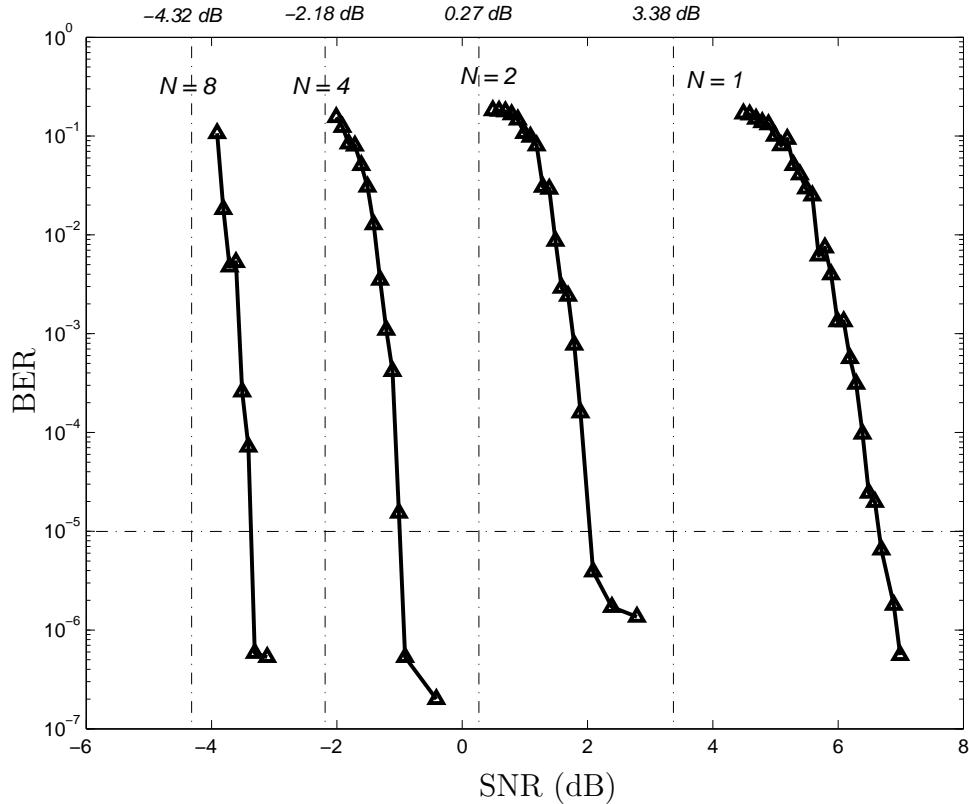


Figure 2.10. The bit error rate (BER) versus the channel SNR for different receive diversity orders. Vertical dashed-dot lines indicate the information-theoretic limits of the required channel SNR.

2.5.4 Robustness Considerations

In this subsection we provide a first-cut discussion of the robustness of the scheme. As the preceding analyses show, the induced additive-noise channel model (2.5.2) heavily depends upon the assumption that the fades among N receive branches are i.i.d. Rayleigh. When the assumption is violated the scheme illustrated in Fig-

ure 2.6 can still be employed, but the mathematical derivations no longer hold since the distribution of the “induced additive noise” becomes input-dependent. Hence it is naturally expected that the scheme only applies when the fading statistics are relatively close to i.i.d. Rayleigh distribution. In the following we use numerical plots to demonstrate that the scheme may actually exhibit a certain degree of robustness, which is a desirable property for practical purposes.

1) *Rician fading*: When (2.5.1) is a Rician fading channel with LOS component $d \in [0, 1]$, each fading coefficient H_n is

$$H_n = d + \sqrt{1 - d^2} \cdot \tilde{H}_n,$$

where $\{\tilde{H}_n\}_{n=1,\dots,N}$ are i.i.d. $\mathcal{CN}(0, \sigma_H^2)$ random variables. We employ symmetric OOK inputs. It is clear that when the channel input is “off” (*i.e.*, zero) the corresponding induced additive noise is the same as the Rayleigh fading case. When the channel input is “on” with average SNR of 3 dB, Figure 2.11 plots the mean and variance of the induced additive noise versus the line-of-sight component d , for different N . It is surprising to notice that the mean and variance remain essentially constant for a wide range of d near zero, implying that the scheme may be robust even if the channel has a weak LOS path.

2) *Nakagami fading*: For Nakagami fading [60] the PDF of the energy of each fading coefficient, $R_n := |H_n|^2$, is

$$f_R(r) = \frac{2}{\Gamma(m)} \left(\frac{m}{\sigma_H^2} \right)^m r^{2m-1} e^{-mr^2/\sigma_H^2},$$

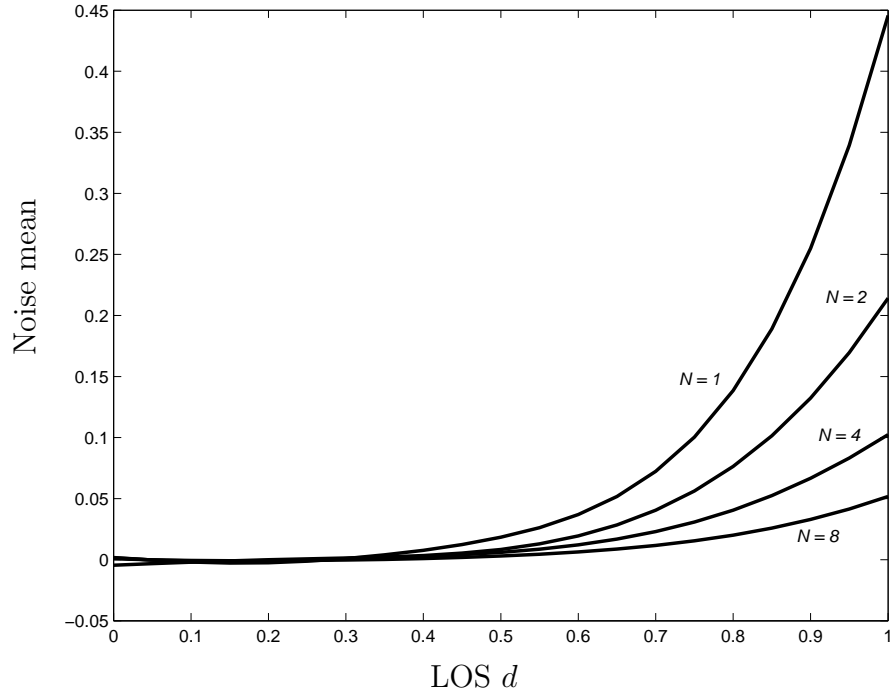
where m is called the fading figure and when $m = 1$ the PDF reduces to Rayleigh. For OOK inputs with SNR of 3 dB, Figure 2.12 plots the mean and variance of the induced additive noise for the “on” input versus the fading figure m , for different N . Here we notice that, especially for small N , the changes near $m = 1$ are less negligible, implying that the scheme may be less robust against Nakagami fading.

3) *Correlated receive branches*: Sometimes there exists certain degree of spatial correlation among the N receive branches. Here we assume that each H_n is marginally $\mathcal{CN}(0, 1)$ and for any pair of $n \neq n'$ $\text{cov}[H_n^\dagger H_{n'}] = \mu \in [0, 1]$. When $\mu = 0$ the N receive branches are independent. Figure 2.13 plots the mean and variance of the induced additive noise for the “on” input versus the correlation factor μ , for different N . Again it can be noticed that the scheme exhibits a certain degree of robustness against weak channel spatial correlation.

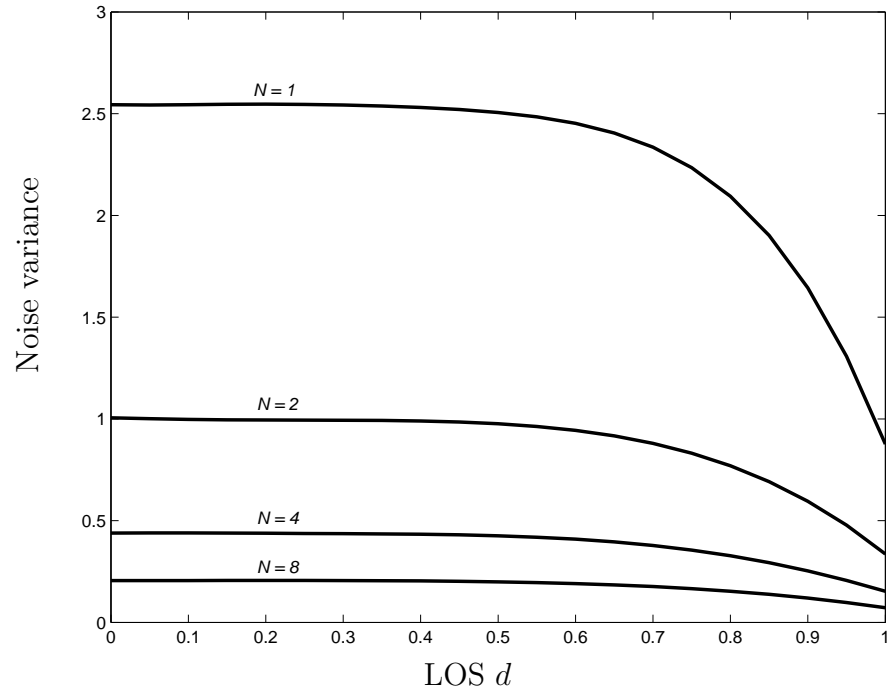
2.6 Conclusion

In this chapter we address non-coherent discrete-time memoryless Rayleigh fading channels. By taking the logarithm of the channel output’s magnitude, we transform the fading channel into an equivalent induced channel with additive noise that is independent of the channel input. This additive-noise channel model holds for all SNR. Using this perspective, we revisit several known results and establish several new results. In particular, we examine the moderate and high SNR channel behavior. We obtain simple conditions for channel input distributions to achieve unbounded mutual information as SNR scales, and to asymptotically achieve the channel capacity in the high SNR limit. We show that a simple input distribution called the log-scale (continuous) uniform distribution is asymptotically capacity-achieving. We further re-interpret a class of log-scale discrete uniform input distributions, and demonstrate that for moderate SNR they perform better than the log-scale (continuous) uniform input. We extend the induced additive-noise channel approach to memoryless multiple-antenna channels possibly with transmit, but without receive, spatial correlation. Finally we demonstrate that the induced additive-noise channel approach can be applied to design practical low-complexity near-optimal systems.

A generalization of the log-scale transform approach, if it exists, doubtlessly

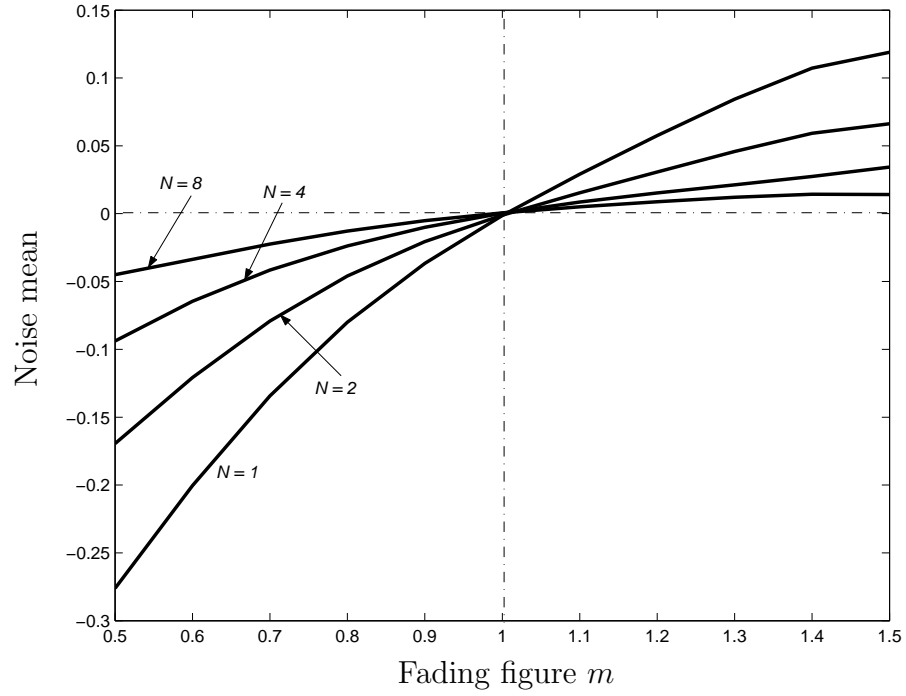


(a) Noise mean vs. d

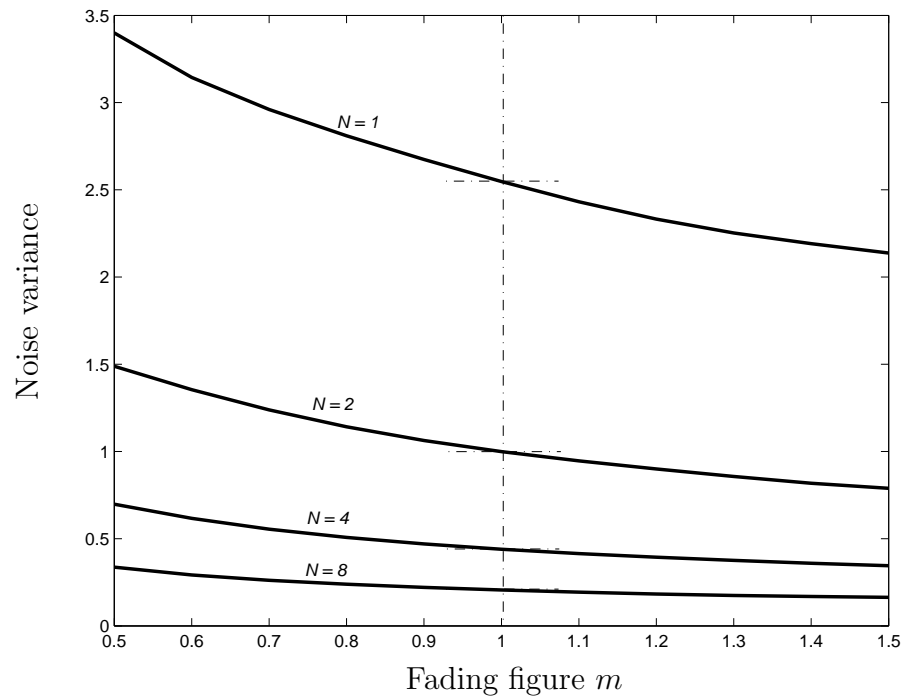


(b) Noise variance vs. d

Figure 2.11. The mean and variance of the induced additive noise for the “on” input as functions of the LOS component d in Rician fading channels. The channel SNR is $\rho = 3$ dB. The case of $d = 0$ corresponds to Rayleigh fading channels.

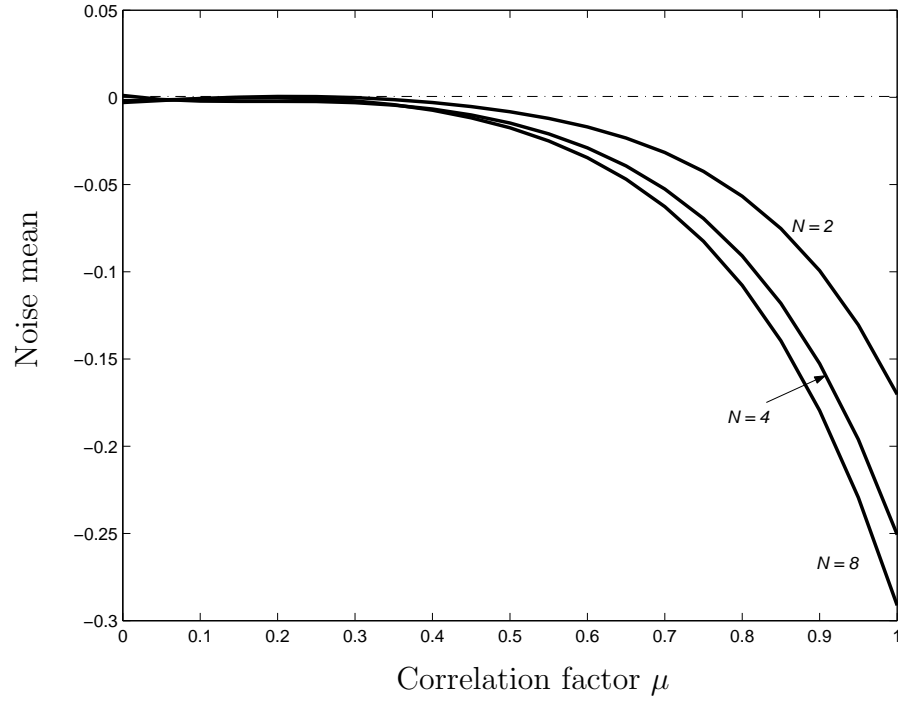


(a) Noise mean vs. m

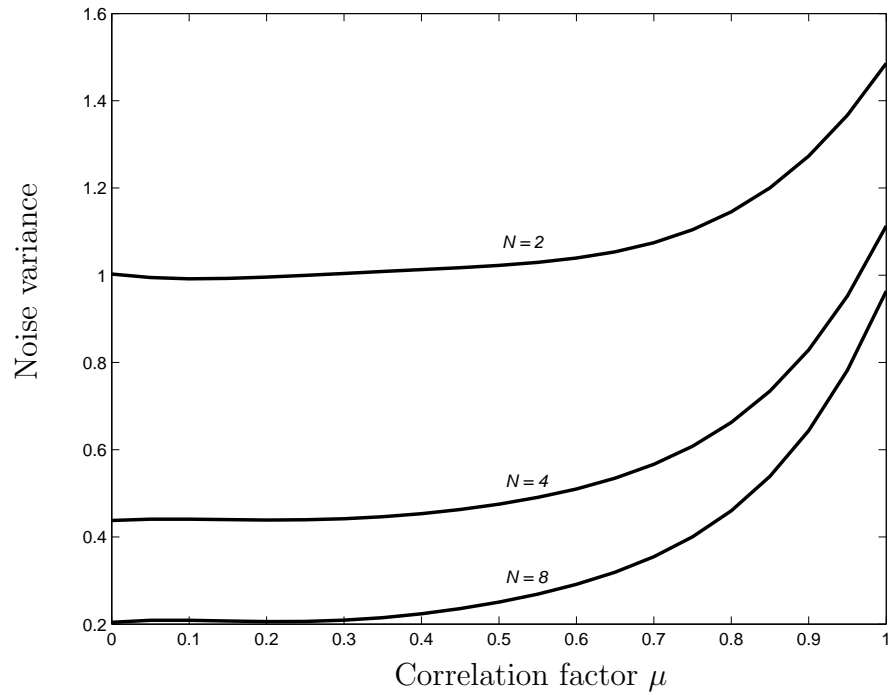


(b) Noise variance vs. m

Figure 2.12. The mean and variance of the induced additive noise for the “on” input as functions of the fading figure m in Nakagami fading channels. The channel SNR is $\rho = 3$ dB. The case of $m = 1$ corresponds to Rayleigh fading channels.



(a) Noise mean vs. μ



(b) Noise variance vs. μ

Figure 2.13. The mean and variance of the induced additive noise for the “on” input as functions of the channel correlation factor μ in Rayleigh fading channels with correlated receive branches. The channel SNR is $\rho = 3$ dB. The case of $\mu = 0$ corresponds to independent receive branches.

would enhance our understanding of channel behavior and communication system design for general non-coherent fading channels. However, we note that currently this approach seems to be applicable only to memoryless Rayleigh fading channels; extensions to more general scenarios are not immediate. There are primarily two obstacles. First, for other fading distributions, the “non-Rayleighness” precludes the convenient shift-invariance property of the log-scale transform. For example, in Ricean channels or phase non-coherent Gaussian channels, the induced additive noise after the log-scale transform is input-dependent. More fundamentally, for channels in which the output sufficient statistic is multivariate rather than scalar, we have not yet found an immediate analog of the log-scale transform.

CHAPTER 3

EXPLOITATION OF CHANNEL TEMPORAL CORRELATION

3.1 Introduction

For fairly general Rayleigh fading channels without CSI at low SNR, the capacity-achieving input gradually tends to bursts of “on” intervals sporadically inserted into the “off” background, even under vanishing peak power constraints [64]. This highly unbalanced input usually imposes implementation challenges. For example, it is difficult to maintain carrier frequency and symbol timing during the long “off” periods. Furthermore, the unbalanced input is incompatible with linear codes, unless appropriate symbol mapping (*e.g.*, orthogonal modulation with appropriately chosen constellation size) is employed to match the input distribution.

This chapter investigates the achievable information rate of phase-shift keying (PSK) on temporally correlated fading channels. PSK is appealing because it has constant envelope and is amenable to linear codes without additional symbol mappings. Focusing on low signal-to-noise ratio (SNR) asymptotics, we develop a capacity lower bound and interpret it as the achievable rate of a recursive training scheme¹ that converts the original fading channel without CSI into a series of parallel sub-channels, each with estimated CSI but additional noise that remains circular complex white Gaussian.

¹A similar framework was also partially pursued in [65, 40].

The central results in this chapter are as follows. First, for a discrete-time Rayleigh fading channel whose unit-variance fading process $\{H_d[k] : -\infty < k < \infty\}$ has a spectral density function $S_{H_d}(e^{j\Omega})$ for $-\pi \leq \Omega \leq \pi$, the achievable rate is

$$\frac{1}{2} \cdot \left[\frac{1}{2\pi} \cdot \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega - 1 \right] \cdot \rho^2 + o(\rho^2) \quad \text{nats per symbol}, \quad (3.1.1)$$

as the average received channel SNR $\rho \rightarrow 0$. This achievable rate is at most $(1/2) \cdot \rho^2 + o(\rho^2)$ away from the channel capacity under peak SNR constraint ρ .² Second, for a continuous-time Rayleigh fading channel whose unit-variance fading process $\{H_c(t) : -\infty < t < \infty\}$ has a spectral density function $S_{H_c}(j\omega)$ for $-\infty < \omega < \infty$, the achievable rate as the input symbol duration $T \rightarrow 0$ is

$$\left[1 - \frac{1}{2\pi P} \cdot \int_{-\infty}^{\infty} \log(1 + P \cdot S_{H_c}(j\omega)) d\omega \right] \cdot P \quad \text{nats per unit time}, \quad (3.1.2)$$

where $P > 0$ is the envelope power. This achievable rate coincides with the channel capacity under peak envelope P .

We apply the above results to specific case studies of Gauss-Markov fading models (both discrete-time and continuous-time) as well as a continuous-time Clarke's fading model. For discrete-time Gauss-Markov fading processes with innovation rate $\epsilon \ll 1$, the quadratic behavior of the achievable rate becomes dominant only for $\rho \ll \epsilon$. Our results, combined with previous results for the high-SNR asymptotics [13], suggest that coherent communication can essentially be realized for $\epsilon \lesssim \rho \lesssim 1/\epsilon$. For Clarke's model, we find that the achievable rate scales sub-linearly, but super-quadratically, as $O(\log(1/P) \cdot P^2)$ nats per unit time as $P \rightarrow 0$.

Finally, we extend the above results to Rician fading channels in which the direct LOS component has strength $d \in [0, \infty)$ relative to the Rayleigh fading process

²Recently, a more refined capacity upper bound [39] suggests that the achievable rate (3.1.1) is in fact optimal up to the second-order term if $\left[\frac{1}{2\pi} \cdot \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega - 1 \right] \geq 1$, a condition met by many slowly time-varying fading channels. A more comprehensive analysis taking into account both peak and average power constraints appears in [66].

component. For a discrete-time channel, the achievable rate is

$$\frac{d}{d+1} \cdot \rho + \frac{1}{2} \left[\frac{1}{2\pi(d+1)^2} \cdot \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega - 1 \right] \cdot \rho^2 + o(\rho^2) \quad \text{nats per symb} \quad (3.1.3)$$

which is again at most $(1/2) \cdot \rho^2 + o(\rho^2)$ away from the channel capacity under peak SNR constraint ρ . For a continuous-time channel, as the input symbol duration $T \rightarrow 0$, the achievable rate is

$$\left[1 - \frac{1}{2\pi P} \cdot \int_{-\infty}^{\infty} \log \left(1 + \frac{P}{d+1} \cdot S_{H_c}(j\omega) \right) d\omega \right] \cdot P \quad \text{nats per unit time,} \quad (3.1.4)$$

which again coincides with the channel capacity under peak envelope P .

The remainder of the chapter is organized as follows. Section 3.2 describes the Rayleigh fading channel model. Section 3.3 develops an achievable rate and its recursive training scheme interpretation. Section 3.4 deals with the discrete-time Rayleigh fading channel model, and Section 3.5 deals with the continuous-time Rayleigh fading channel model. Section 3.6 extends the results to Rician fading channels. Finally Section 3.7 provides some concluding remarks.

3.2 Channel Model

We consider a scalar time-selective, frequency-nonselective Rayleigh fading channel, written in baseband-equivalent continuous-time form as

$$X(t) = H_c(t) \cdot S(t) + Z(t), \quad \text{for } -\infty < t < \infty, \quad (3.2.1)$$

where $S(t) \in \mathbb{C}$ and $X(t) \in \mathbb{C}$ denote the channel input and the corresponding output at time instant t , respectively. The additive noise $\{Z(t) : -\infty < t < \infty\}$ is modeled as a zero-mean circular complex Gaussian white noise process with $\mathcal{E}\{Z(s)Z^\dagger(t)\} = \delta(s-t)$. The fading process $\{H_c(t) : -\infty < t < \infty\}$ is modeled as a wide-sense stationary and ergodic zero-mean circular complex Gaussian process with unit variance $\mathcal{E}\{H_c(t)H_c^\dagger(t)\} = 1$ and with spectral density function $S_{H_c}(j\omega)$

for $-\infty < \omega < \infty$. Additionally, we impose a technical condition that $\{H_c(t) : -\infty < t < \infty\}$ is mean-square continuous, so that its autocorrelation function $K_{H_c}(\tau) = \mathcal{E}\{H_c(t + \tau)H_c^\dagger(t)\}$ is continuous for $\tau \in (-\infty, \infty)$.

Throughout the chapter, we restrict our attention to i.i.d. PSK over the continuous-time channel (3.2.1). For technical convenience, we let the channel input $S(t)$ have constant envelope $P > 0$ and piecewise constant phase, *i.e.*,

$$S(t) := \sqrt{P} \cdot e^{j\Theta[k]}, \quad \text{if } kT \leq t < (k+1)T, \quad (3.2.2)$$

for $-\infty < k < \infty$. The symbol duration $T > 0$ is determined by the reciprocal of the channel input bandwidth.³

Applying the above channel input to the continuous-time channel (3.2.1), and processing the channel output through a unit-energy matched filter,⁴ we obtain a discrete-time channel⁵

$$X[k] = \sqrt{\rho} \cdot H_d[k] \cdot S[k] + Z[k], \quad \text{for } -\infty < k < \infty. \quad (3.2.3)$$

The channel equations (3.2.1) and (3.2.3) are related through

$$S[k] := e^{j\Theta[k]} \quad (3.2.4)$$

$$X[k] := \frac{1}{\sqrt{T}} \int_{kT}^{(k+1)T} X(t) dt \quad (3.2.5)$$

$$Z[k] := \frac{1}{\sqrt{T}} \int_{kT}^{(k+1)T} Z(t) dt \quad (3.2.6)$$

$$H_d[k] := \frac{1}{\sqrt{\int_0^T \int_0^T K_{H_c}(s-t) ds dt}} \int_{kT}^{(k+1)T} H_c(t) dt. \quad (3.2.7)$$

³For multipath fading channels, T should be substantially greater than the multipath delay spread [4], otherwise the frequency non-selective channel model (3.2.1) may not be valid. Throughout the chapter we assume that this requirement is met.

⁴A matched filter suffers no information loss for white Gaussian channels [60]. For the fading channel (3.2.1), it is no longer optimal in general [30]. However, in this chapter we still focus on the matched filter, which is commonly used in many practical systems.

⁵Here we note a slight abuse of notation in this chapter, that a symbol (*e.g.*, S) can be either continuous-time or discrete-time. The two cases are distinguished by $S(t)$ for continuous-time and $S[k]$ for discrete-time.

For the discrete-time channel (3.2.3) we can verify that

- The additive noise $\{Z[k] : -\infty < k < \infty\}$ is circular complex Gaussian with zero mean and unit variance, *i.e.*, $Z[k] \sim \mathcal{CN}(0, 1)$, and is i.i.d. for different k .
- The fading process $\{H_d[k] : -\infty < k < \infty\}$ is wide-sense stationary and ergodic zero-mean circular complex Gaussian, with $H_d[k]$ being marginally $\mathcal{CN}(0, 1)$. We further notice that $\{H_d[k] : -\infty < k < \infty\}$ is obtained through sampling the output of the matched filter, hence its spectral density function is

$$S_{H_d}(e^{j\Omega}) = \frac{T}{\int_0^T \int_0^T K_{H_c}(s-t) ds dt} \cdot \sum_{k=-\infty}^{\infty} S_{H_c} \left(j \frac{\Omega - 2k\pi}{T} \right) \cdot \text{sinc}^2(\Omega - 2k\pi) \quad (3.2.8)$$

for $-\pi \leq \Omega \leq \pi$.

- The average channel SNR experienced at the receiver is given by

$$\rho := \frac{P}{T} \cdot \left(\int_0^T \int_0^T K_{H_c}(s-t) ds dt \right) > 0. \quad (3.2.9)$$

The i.i.d. channel input $\{S[k] : -\infty < k < \infty\}$ is always on the unit circle. As will be explained later, we will only restrict it to be zero-mean, *i.e.*, $\mathcal{E}\{S[k]\} = 0$. Such inputs encompass all the PSK inputs that are complex proper [51], *i.e.*, $\mathcal{E}\{S^2[k]\} = [\mathcal{E}\{S[k]\}]^2$, like quadrature phase-shift keying (QPSK), as well as binary phase-shift keying (BPSK), which is not complex proper.

Throughout the chapter, we assume that the realization of the fading process $\{H_c(t) : -\infty < t < \infty\}$ is not directly available to the transmitter or the receiver, but its statistical characterization in terms of $S_{H_c}(j\omega)$ is precisely known at both the transmitter and the receiver.

3.3 An Achievable Rate and its Recursive Training Scheme Interpretation

3.3.1 An Achievable Rate Based on Average Channel Mutual Information

For an arbitrary coding blocklength L with channel inputs $\{S[l] : l = 0, 1, \dots, L-1\}$, and corresponding channel outputs $\{X[l] : l = 0, 1, \dots, L-1\}$, the average

mutual information of the discrete-time channel (3.2.3) is

$$R_L := \frac{1}{L} I(\{S[l]\}_{l=0}^{L-1}; \{X[l]\}_{l=0}^{L-1}). \quad (3.3.1)$$

As the coding blocklength L scales, standard channel coding theorems [78] demonstrate that the rate $R_\infty := \lim_{L \rightarrow \infty} R_L$ is achievable. However, in general R_∞ is rather difficult to compute since it involves a limiting procedure $L \rightarrow \infty$. In this section we develop and interpret a simple lower bound for R_∞ , and will focus on it in the remainder of the chapter.

Iteratively applying the chain rule for mutual information [10], we have

$$\begin{aligned} R_L &= \frac{1}{L} \sum_{l=0}^{L-1} I(\{S[i]\}_{i=0}^{L-1}; X[l] | \{X[i]\}_{i=0}^{l-1}) \\ &= \frac{1}{L} \sum_{l=0}^{L-1} [I(\{S[i]\}_{i=0}^{l-1}; X[l] | \{X[i]\}_{i=0}^{l-1}) + I(S[l]; X[l] | \{X[i]\}_{i=0}^{l-1}, \{S[i]\}_{i=0}^{l-1}) \\ &\quad + I(\{S[i]\}_{i=l+1}^{L-1}; X[l] | \{X[i]\}_{i=0}^{l-1}, \{S[i]\}_{i=0}^l)] \\ &\geq \frac{1}{L} \sum_{l=0}^{L-1} I(S[l]; X[l] | \{X[i]\}_{i=0}^{l-1}, \{S[i]\}_{i=0}^{l-1}). \end{aligned} \quad (3.3.2)$$

By induction, let us evaluate the l th term in the final summation of (3.3.2), *i.e.*, the mutual information $I(S[l]; X[l] | \{X[i]\}_{i=0}^{l-1}, \{S[i]\}_{i=0}^{l-1})$. We can interpret this term as that at time l , in addition to all the past channel outputs, we have also successfully reconstructed all the past channel inputs at the receiver as side information. Since PSK channel inputs are always on the unit circle, the receiver can compensate for their phases in the channel outputs, and the resulting observations become

$$\underbrace{e^{-j\Theta[i]} \cdot X[i]}_{X'[i]} = \sqrt{\rho} \cdot H_d[i] + \underbrace{e^{-j\Theta[i]} \cdot Z[i]}_{Z'[i]}, \quad \text{for } i = 0, 1, \dots, l-1. \quad (3.3.3)$$

Since zero-mean circular complex Gaussian distributions are invariant under rotation, the rotated noise $Z'[\cdot]$ is still zero-mean unit-variance circular complex Gaussian. Furthermore, since the original noise $Z[\cdot]$ is i.i.d. and independent of $H_d[\cdot]$, it

can be easily verified that the rotated noise $Z'[\cdot]$ is also i.i.d. and independent of $H_d[\cdot]$. Then we can utilize standard linear prediction theory (*e.g.*, [31]) to obtain the one-step minimum mean-square error (MMSE) prediction of $H_d[l]$ defined as

$$\hat{H}_d[l] := \mathcal{E} \left\{ H_d[l] \middle| \{X'[i] : i = 0, 1, \dots, l-1\} \right\}. \quad (3.3.4)$$

The prediction $\hat{H}_d[l]$ and the prediction error $\tilde{H}_d[l] = H_d[l] - \hat{H}_d[l]$ are jointly circular complex Gaussian distributed as $\mathcal{CN}(0, 1 - \sigma^2[l])$ and $\mathcal{CN}(0, \sigma^2[l])$, respectively, and are uncorrelated and further independent. Here $\sigma^2[l]$ denotes the mean-square prediction error. Hence the channel equation (3.2.3) at time l can be rewritten as

$$\begin{aligned} X[l] &= \sqrt{\rho} \cdot H_d[l] \cdot S[l] + Z[l] \\ &= \sqrt{\rho} \cdot \hat{H}_d[l] \cdot S[l] + \underbrace{\sqrt{\rho} \cdot \tilde{H}_d[l] \cdot S[l] + Z[l]}_{\bar{Z}[l]}, \end{aligned} \quad (3.3.5)$$

where the effective noise $\bar{Z}[l]$ is i.i.d. circular complex Gaussian, and is independent of both the channel input $S[l]$ and the predicted fading coefficient $\hat{H}_d[l]$. To show these properties, we notice that $\hat{H}_d[l]$, $\tilde{H}_d[l]$, and $Z[l]$ are independent circular complex Gaussian, and $S[l]$ is merely a rotation on the unit circle. Thus, the channel (3.3.5) becomes a coherent Gaussian fading channel with receive CSI $\hat{H}_d[l]$ and effective SNR

$$\rho[l] = \frac{1 - \sigma^2[l]}{\sigma^2[l] \cdot \rho + 1} \cdot \rho. \quad (3.3.6)$$

In the chapter we mainly focus on the ultimate performance limit without delay constraints, which is achieved as $L \rightarrow \infty$. Under mild technical conditions, the one-step MMSE prediction error sequence $\{\sigma^2[l] : l = 0, 1, \dots\}$ converges to the limit [12, Chap. XII, Sec. 4]

$$\sigma_\infty^2 := \lim_{l \rightarrow \infty} \sigma^2[l] = \frac{1}{\rho} \cdot \left\{ \exp \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(1 + \rho \cdot S_{H_d}(e^{j\Omega})) d\Omega \right\} - 1 \right\}. \quad (3.3.7)$$

Consequently the effective SNR sequence $\{\rho[l] : l = 0, 1, \dots\}$ converges to

$$\rho_\infty := \lim_{l \rightarrow \infty} \rho[l] = \frac{1 - \sigma_\infty^2}{\sigma_\infty^2 \cdot \rho + 1} \cdot \rho. \quad (3.3.8)$$

Summarizing the above arguments and (3.3.2), as $L \rightarrow \infty$ we have that by Cesaro means

$$\lim_{L \rightarrow \infty} R_L \geq R_{\text{rt}} := \lim_{l \rightarrow \infty} I \left(S[l]; X[l] \middle| \hat{H}_d[l] \right) = I \left(S; X \middle| \hat{H}_d \right), \quad (3.3.9)$$

where S , X , and \hat{H}_d are the input, output, and predicted fading coefficient of the steady-state channel

$$X = \sqrt{\rho} \hat{H}_d S + \underbrace{\left(\sqrt{\rho} \tilde{H}_d S + Z \right)}_{\tilde{Z}}, \quad (3.3.10)$$

where $\hat{H}_d \sim \mathcal{CN}(0, 1 - \sigma_\infty^2)$ and $\tilde{H}_d \sim \mathcal{CN}(0, \sigma_\infty^2)$ are independent, and \tilde{Z} is i.i.d. circular complex Gaussian and independent of S and \hat{H}_d . The channel SNR of (3.3.10) is the steady-state effective SNR ρ_∞ as given by (3.3.8).

In Section 3.4 we evaluate the achievable rate R_{rt} of the steady-state channel (3.3.10) which operates at the effective SNR ρ_∞ , as the actual channel SNR $\rho \rightarrow 0$. In fact, for our purposes it is sufficient to examine the first-order expansion of the mutual information in (3.3.9) [45, 38, 75]. Specifically, because we restrict the PSK inputs to be zero-mean, the asymptotically achievable rate becomes

$$R_{\text{rt}} = \rho_\infty + o(\rho_\infty), \quad \text{as } \rho \rightarrow 0. \quad (3.3.11)$$

3.3.2 A Recursive Training Interpretation of the Achievable Rate R_{rt}

Before analyzing the low-SNR behavior of the achievable rate R_{rt} in (3.3.11), in this subsection we present a recursive training interpretation for the achievability of R_{rt} in (3.3.9). From the channel coding theorem for general channels [78], the information rate R_{rt} , and in fact $R_\infty = \lim_{L \rightarrow \infty} R_L$, can be reliably achieved by

coding over one single sufficiently long block of channel uses and by a corresponding decoder taking into full account the channel memory. Here we show that, for channels with finite memory structure, the information rate R_{rt} can also be achieved by a simple recursive training scheme.

The recursive training scheme has been used in a series of earlier works, *e.g.*, [13, 16], for channels with memory. Its basic idea is as follows. By interleaving the transmitted symbols of the discrete-time channel (3.2.3) as illustrated in Figure 3.1, we effectively convert the original non-coherent channel into a series of parallel sub-channels (PSC). The recursive training scheme performs channel estimation and demodulation/decoding in an alternating manner. To initialize transmission, known pilot symbols are transmitted on PSC 0, the first parallel sub-channel. Based upon the k th received pilot symbol $X[k \cdot L]$ in PSC 0, the receiver predicts $H_{\text{d}}[k \cdot L + 1]$, the fading coefficients of PSC 1, for each $k = 0, 1, \dots, K - 1$. The receiver then proceeds to demodulate and decode the transmitted information symbols in PSC 1 coherently using the predicted $\{\hat{H}_{\text{d}}[k \cdot L + 1]\}_{k=0}^{K-1}$. If the rate of PSC 1 does not exceed the corresponding channel mutual information, then the channel coding theorem ensures that, as the coding blocklength $K \rightarrow \infty$, there exist codes that have arbitrarily small decoding error probability. Hence the receiver can, at least in principle, form an essentially error-free reconstruction of the transmitted symbols in PSC 1, which then effectively become “fresh” pilot symbols to facilitate the prediction of $\{H_{\text{d}}[k \cdot L + 2]\}_{k=0}^{K-1}$ and subsequent coherent demodulation/decoding of PSC 2. Extending this procedure, decoded information symbols from PSCs $0, \dots, l - 1$ are used as pilot symbols to predict $\{H_{\text{d}}[k \cdot L + l]\}_{k=0}^{K-1}$ and coherently demodulate/decode PSC l .

From the preceding description, each of the L PSCs obtains estimated receive CSI but additional noise that asymptotically remains i.i.d. circular complex Gaussian as $L \rightarrow \infty$. These L PSCs suffer correlated fading, and this correlation is exactly what

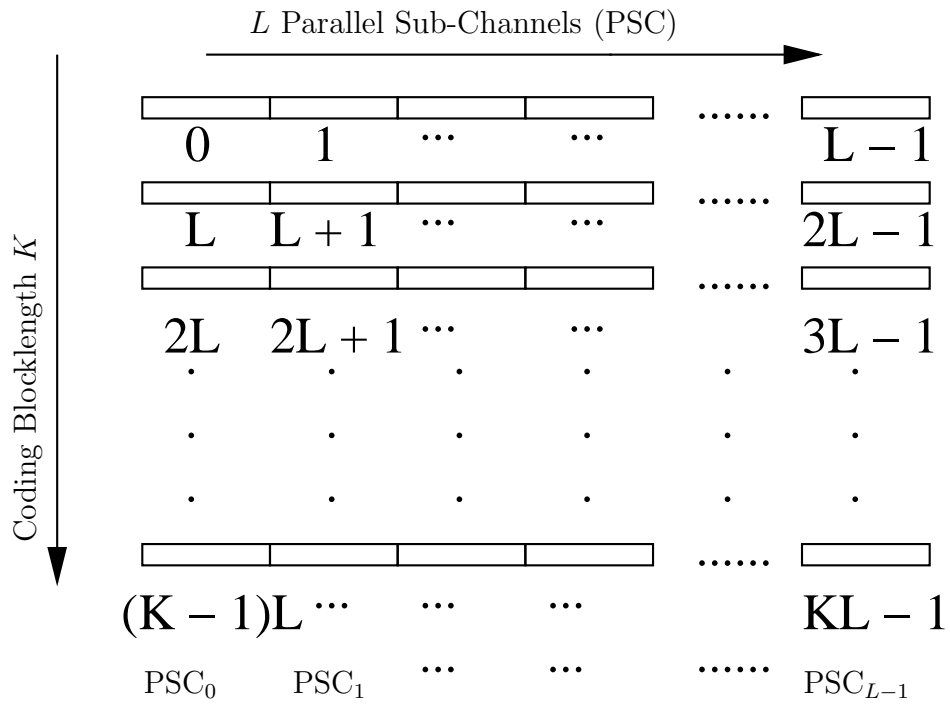


Figure 3.1. Illustration of the interleaving structure in the recursive training scheme. Input symbols are encoded/decoded column-wise, and transmitted/received row-wise.

we seek to exploit using recursive training. Meanwhile for general fading processes that possess infinite memory, some residual correlation remains within each PSC among its K symbols. Due to the ergodicity of the channel (3.2.3), this correlation asymptotically vanishes as the interleaving depth $L \rightarrow \infty$, and each PSC can then be viewed as essentially memoryless.

Unfortunately, to establish a channel coding theorem for the achievability of R_{rt} , the above intuitive reasoning appears to be technically incomplete. This appears to result from a tension between decorrelation and error propagation: Although increasing the interleaving depth L reduces the residual channel correlation within each PSC, it also imposes more stringent requirements on the average decoding error probability for each PSC to prevent catastrophic error propagation. It remains an open problem to characterize under which general conditions the recursive training scheme achieves the information rate R_{rt} . In this section, we focus on the special case that the discrete-time fading process $\{H_{\text{d}}[k] : -\infty < k < \infty\}$ is m -dependent, *i.e.*, there exists a finite integer $m \geq 1$ such that $H_{\text{d}}[k]$ and $H_{\text{d}}[k']$ are independent for every k and $k' \geq k + m$. For the stationary Gaussian process $\{H_{\text{d}}[k] : -\infty < k < \infty\}$, m -dependence is equivalent to the condition that the spectral density function $S_{H_{\text{d}}}(e^{j\Omega})$ is of the form as the square of a polynomial in $e^{j\Omega}$; see, *e.g.*, [6].

With the assumption that the fading process is m -dependent, in Figure 3.1 we simply need to append m extra channel uses at the end of each row as a guard interval. For clarity of exposition, we still keep the time indexes unchanged, and the reader should bear in mind that between any two adjacent rows there is a guard interval with m channel uses that eliminate the correlation between the rows. Later as we let the interleaving depth $L \rightarrow \infty$, the rate loss due to guard intervals becomes negligible.

We note that, in each step of the channel estimation procedure, the receiver does not need to perform a vector one-step MMSE prediction that predicts $\{H_d[k \cdot L + l]\}_{k=0}^{K-1}$ jointly based upon all the previous channel outputs and decoded inputs of the $l - 1$ PSCs, $\{X[k \cdot L + i], S[k \cdot L + i]\}_{k=0, i=0}^{K-1, l-1}$. The memoryless property of each PSC reduces the vector prediction to K separate scalar one-step MMSE predictions. That is, for each $k = 0, 1, \dots, K - 1$, the receiver predicts $H_d[k \cdot L + l]$ solely based upon $\{X[k \cdot L + i], S[k \cdot L + i]\}_{i=0}^{l-1}$.

In order to establish the achievability of rate R_{rt} , we fix a sufficiently small decoding error rate $\delta > 0$ and a sufficiently small overall rate loss factor $\lambda > 0$. That is, we want to have the system achieve the rate $(1 - \lambda) \cdot R_{\text{rt}}$ with an average decoding error probability no greater than δ . From the definition of R_{rt} in (3.3.9), the channel mutual information $I\left(S[l]; X[l] \mid \hat{H}_d[l]\right)$ for PSC l converges to R_{rt} by Cesaro means as $l \rightarrow \infty$. Therefore we can choose a sufficiently large interleaving depth L and a sequence of sufficiently small $\lambda_l > 0$ for $l = 0, \dots, L - 1$ such that the average mutual information of the L PSCs is

$$\frac{1}{L + m} \sum_{l=0}^{L-1} (1 - \lambda_l) \cdot I\left(S[l]; X[l] \mid \hat{H}_d[l]\right) \geq (1 - \lambda) \cdot R_{\text{rt}}, \quad (3.3.12)$$

where the normalizing factor $1/(L + m)$ takes into account the length- m guard intervals. For each PSC l , conditioning upon that no decoding error has occurred in the previous $(l - 1)$ PSCs, we can choose a sufficiently large K such that there exists a codebook with information rate $(1 - \lambda_l) \cdot I\left(S[l]; X[l] \mid \hat{H}_d[l]\right)$ and with average decoding error probability no greater than δ/L . Therefore the overall average

decoding error probability P_e can be upper bounded by

$$\begin{aligned}
P_e &= \sum_{l=1}^{L-1} \text{Prob} \{ \text{no error in PSCs } 0 \text{ through } (l-1), \text{ error in PSC } l \} \\
&\leq \sum_{l=1}^{L-1} \text{Prob} \{ \text{error in PSC } l \mid \text{no error in PSCs } 0 \text{ through } (l-1) \} \\
&\leq (L-1) \cdot \frac{\delta}{L} < \delta.
\end{aligned} \tag{3.3.13}$$

That is, the information rate R_{rt} is achievable using the recursive training scheme.

Remarks:

- A major drawback of the recursive training scheme is that its interleaved structure leads to large delay. The coding blocklength K should be large enough so that the decoding error probability is small enough to prevent catastrophic error propagation. Furthermore, the number of PSCs L should be large enough such that the prediction of the fading process essentially converges to its steady-state limit, and such that the rate loss due to guard intervals is negligible. Only after receiving all the $K \cdot L$ symbols in the interleaved block can the receiver perform the alternating estimation-demodulation/decoding procedure.
- A possible situation is that in wideband channels with frequency-decorrelated fading processes, we can employ multi-carrier techniques, *e.g.*, orthogonal frequency-division multiplexing (OFDM), to decompose the original wide bandwidth into a large number of sub-bands, suffering independent frequency-nonsselective fading processes. If we treat each sub-band as one row in Figure 3.1, then the preceding analysis also applies since the sub-bands are independent. Because the K coding symbols for each PSC occur simultaneously in time, the receiver need not wait until receiving all the $K \cdot L$ symbols to perform the alternating estimation-demodulation/decoding procedure.

3.4 Channel Mutual Information at Low SNR

As shown in (3.3.11), the asymptotic channel mutual information depends on the limiting effective SNR (3.3.8), which further relates to the limiting one-step MMSE prediction error (3.3.7). The following proposition evaluates the asymptotic behavior of the channel mutual information (3.3.11).

Proposition 3.4.1 *For the discrete-time channel (3.2.3), as $\rho \rightarrow 0$, its steady-state induced channel (3.3.10) resulting from PSK with recursive training achieves*

the rate

$$R_{\text{rt}} = \frac{1}{2} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega - 1 \right] \cdot \rho^2 + o(\rho^2), \quad (3.4.1)$$

if the integral $(1/2\pi) \cdot \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega$ exists.

Proof: We will prove that

$$\sigma_{\infty}^2 = 1 - \frac{1}{2} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega - 1 \right] \cdot \rho + o(\rho), \quad (3.4.2)$$

which together with (3.3.8) leads to

$$\rho_{\infty} = \frac{1}{2} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega - 1 \right] \cdot \rho^2 + o(\rho^2). \quad (3.4.3)$$

Then (3.4.1) immediately follows from (3.3.11).

For simplicity let us denote by $g(\rho)$ the integral $(1/2\pi) \cdot \int_{-\pi}^{\pi} \log(1 + \rho \cdot S_{H_d}(e^{j\Omega})) d\Omega$,

hence

$$\begin{aligned} \lim_{\rho \rightarrow 0} g(\rho) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \log 1 d\Omega = 0 \\ \lim_{\rho \rightarrow 0} \frac{dg(\rho)}{d\rho} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}(e^{j\Omega}) d\Omega = 1 \\ \lim_{\rho \rightarrow 0} \frac{dg^2(\rho)}{d^2\rho} &= -\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega. \end{aligned}$$

To prove (3.4.2), we apply l'Hospital's rule in (3.3.7) to evaluate

$$\lim_{\rho \rightarrow 0} \sigma_{\infty}^2 = \lim_{\rho \rightarrow 0} \frac{e^{g(\rho)} - 1}{\rho} = 1; \quad (3.4.4)$$

$$\begin{aligned} \lim_{\rho \rightarrow 0} \frac{d(\sigma_{\infty}^2)}{d\rho} &= \lim_{\rho \rightarrow 0} \left[\frac{e^{g(\rho)}}{\rho} \cdot \frac{dg(\rho)}{d\rho} - \frac{e^{g(\rho)} - 1}{\rho^2} \right] \\ &= -\frac{1}{2} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega - 1 \right]. \end{aligned} \quad (3.4.5)$$

Substituting the above quantities into the first-order Taylor series expansion of σ_{∞}^2 , we obtain (3.4.2). **Q.E.D.**

Proposition 3.4.1 states that for PSK at low SNR, the achievable channel mutual information vanishes quadratically with SNR. This is consistent with [50, 22]. Furthermore, it is of particular interest to compare the asymptotic expansion (3.4.1) with several previous results.

3.4.1 Comparison with a Capacity Upper Bound

For the discrete-time channel (3.2.3), PSK with constant SNR ρ is a particular peak-limited channel input. The capacity per unit energy of the channel (3.2.3) under a peak SNR constraint ρ is [64]

$$\dot{C} = 1 - \frac{1}{2\pi\rho} \cdot \int_{-\pi}^{\pi} \log(1 + \rho \cdot S_{H_d}(e^{j\Omega})) d\Omega, \quad (3.4.6)$$

achieved by OOK in which each “on” or “off” symbol corresponds to an infinite number of channel uses, and the probability of choosing “on” symbols vanishes. Such “bursty” channel inputs are in sharp contrast to PSK. From (3.4.6), an upper bound to the channel capacity can be derived as [64]

$$C \leq U(\rho) := \frac{1}{2} \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega \cdot \rho^2. \quad (3.4.7)$$

Comparing (3.4.1) and (3.4.7), we notice that the penalty for using PSK instead of the bursty \dot{C} -achieving channel input is at most $(1/2) \cdot \rho^2 + o(\rho^2)$. For fast time-varying fading processes, this penalty can be relatively significant. For instance, if the fading process is memoryless, *i.e.*, $S_{H_d}(e^{j\Omega}) = 1$ for $-\pi \leq \Omega \leq \pi$, then $(1/2\pi) \cdot \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega - 1 = 0$, implying that no information can be transmitted using PSK. Fortunately, for slowly time-varying fading processes, the integral $(1/2\pi) \cdot \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega$ is typically much greater than 1, as we will see.

3.4.2 Comparison with the High-SNR Channel Behavior

From (3.4.1) and (3.4.7), it can be said that $(1/2\pi) \cdot \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega$ is a fundamental quantity associated with a fading process at low SNR. This is in contrast to the high-SNR regime, where a fundamental quantity is [35]

$$\sigma_{\text{pred}}^2 := \exp \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log S_{H_d}(e^{j\Omega}) d\Omega \right\}. \quad (3.4.8)$$

The quantity σ_{pred}^2 is the one-step MMSE prediction error of $H_d[0]$ given its entire noiseless past $\{H_d[-1], H_d[-2], \dots\}$. If $\sigma_{\text{pred}}^2 > 0$ the process is said to be regular. If

$\sigma_{\text{pred}}^2 = 0$ it is said to be deterministic, that is, the entire future $\{H_d[0], H_d[1], \dots\}$ can be exactly reconstructed (in the mean-square sense) by linearly combining the entire past $\{H_d[-1], H_d[-2], \dots\}$. It has been established in [35, 36] that, for regular fading processes,

$$C = \log \log \rho - 1 - \gamma + \log \frac{1}{\sigma_{\text{pred}}^2} + o(1) \quad \text{as } \rho \rightarrow \infty, \quad (3.4.9)$$

where $\gamma = 0.5772\dots$ is Euler's constant, and for deterministic fading processes,

$$\frac{C}{\log \rho} \rightarrow \frac{1}{2\pi} \cdot \mu(\{\Omega : S_{H_d}(e^{j\Omega}) = 0\}) \quad \text{as } \rho \rightarrow \infty, \quad (3.4.10)$$

where $\mu(\cdot)$ denotes the Lebesgue measure on the interval $[-\pi, \pi]$.

It is then an interesting issue to investigate the connection between $(1/2\pi) \cdot \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega$ and σ_{pred}^2 . However, as the following two examples reveal, there is no explicit relationship between these two quantities.

Example 1: Even a deterministic fading process can lead to poor low-SNR performance

Consider the following class of spectral density functions $S_{H_d}(e^{j\Omega})$ as illustrated in Figure 3.2:

$$S_{H_d}(e^{j\Omega}) = \begin{cases} \frac{n}{n-1}, & \text{if } |\Omega| \leq \pi - \frac{\pi}{n} \\ 0, & \text{if } \pi - \frac{\pi}{n} < |\Omega| \leq \pi \end{cases}, \quad n = 2, 3, \dots \quad (3.4.11)$$

Since $S_{H_d}(e^{j\Omega}) = 0$ for certain intervals with non-zero measure, the corresponding fading process is deterministic with $\sigma_{\text{pred}}^2 = 0$ [12]. However, this class of $S_{H_d}(e^{j\Omega})$ leads to

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega = \frac{n}{n-1} \rightarrow 1 \quad \text{as } n \rightarrow \infty, \quad (3.4.12)$$

resulting in vanishing values of the quadratic coefficient in (3.4.1).

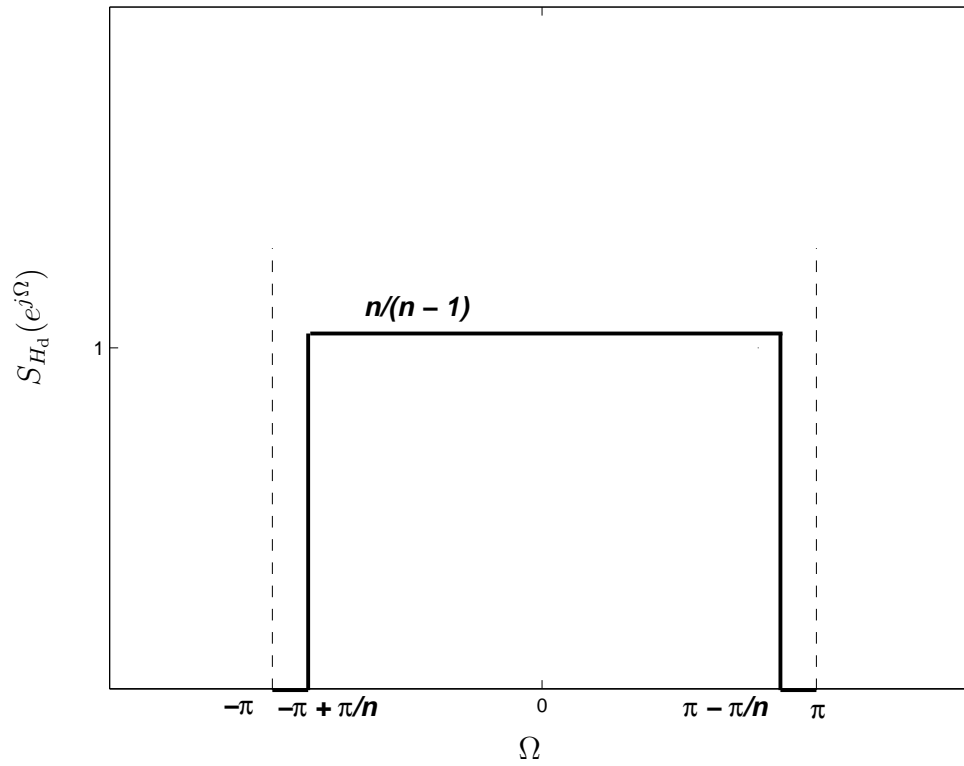


Figure 3.2. Spectral density function of a deterministic fading process that leads to poor low-SNR performance. The narrow notches on the spectrum make the process deterministic, while the remaining almost unit spectrum makes it behave as if nearly memoryless in the low-SNR regime for large n .

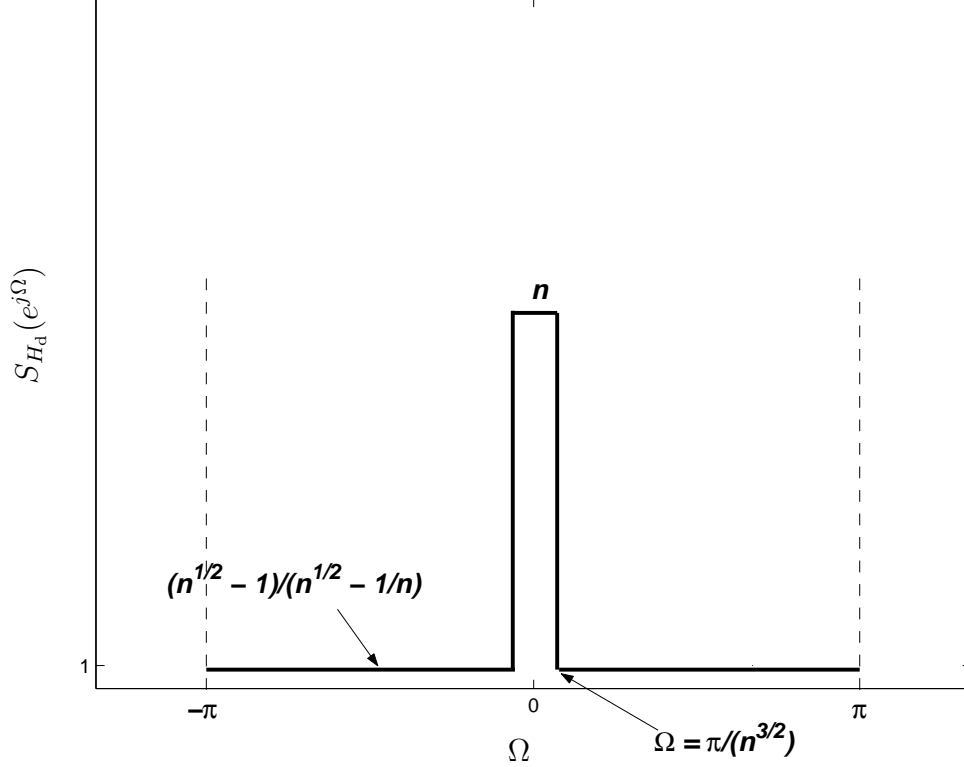


Figure 3.3. Spectral density function of an almost memoryless fading process that leads to good low-SNR performance. The almost unit spectrum makes the process nearly memoryless, while the narrow impulse-like spectrum peak significantly contributes to the integral $(1/2\pi) \cdot \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega$, leading to good low-SNR performance for large n .

Example 2: Even an almost memoryless fading process can lead to good low-SNR performance

Consider the following class of spectral density functions $S_{H_d}(e^{j\Omega})$ as illustrated in Figure 3.3:

$$S_{H_d}(e^{j\Omega}) = \begin{cases} n, & \text{if } |\Omega| \leq \frac{\pi}{n\sqrt{n}} \\ \frac{\sqrt{n}-1}{\sqrt{n}-1/n}, & \text{if } \frac{\pi}{n\sqrt{n}} < |\Omega| \leq \pi \end{cases}, \quad n = 2, 3, \dots \quad (3.4.13)$$

For large n the fading process becomes almost memoryless since

$$\sigma_{\text{pred}}^2 = \exp \left\{ \frac{\log n}{n\sqrt{n}} + \log \frac{\sqrt{n}-1}{\sqrt{n}-1/n} \cdot \left(1 - \frac{1}{n\sqrt{n}}\right) \right\} \rightarrow 1 \quad \text{as } n \rightarrow \infty. \quad (3.4.14)$$

However, this class of $S_{H_d}(e^{j\Omega})$ also leads to

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega = \sqrt{n} + \left(1 - \frac{1}{n\sqrt{n}}\right) \cdot \left(\frac{\sqrt{n}-1}{\sqrt{n}-1/n}\right)^2 \rightarrow \infty \quad (3.4.15)$$

as $n \rightarrow \infty$.

3.4.3 Case Study: Discrete-Time Gauss-Markov Fading Processes

In this subsection, we apply Proposition 3.4.1 to analyze a specific class of discrete-time fading processes, namely, the discrete-time Gauss-Markov fading processes. The fading process in the channel model can be described by a first-order auto-regressive (AR) evolution equation of the form

$$H_d[k+1] = \sqrt{1-\epsilon} \cdot H_d[k] + \sqrt{\epsilon} \cdot V[k+1], \quad (3.4.16)$$

where the innovation sequence $\{V[k] : -\infty < k < \infty\}$ consists of i.i.d. $\mathcal{CN}(0, 1)$ random variables, and $V[k+1]$ is independent of $\{H_d[i] : -\infty < i \leq k\}$. The innovation rate ϵ satisfies $0 < \epsilon \leq 1$.

The spectral density function $S_{H_d}(e^{j\Omega})$ for such a process is

$$S_{H_d}(e^{j\Omega}) = \frac{\epsilon}{(2-\epsilon) - 2\sqrt{1-\epsilon} \cdot \cos \Omega}, \quad -\pi \leq \Omega \leq \pi. \quad (3.4.17)$$

Hence

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega = \frac{\epsilon^2}{2\pi} \int_{-\pi}^{\pi} \frac{1}{((2-\epsilon) - 2\sqrt{1-\epsilon} \cdot \cos \Omega)^2} d\Omega = \frac{2}{\epsilon} - 1.$$

Applying Proposition 3.4.1, we find that for the discrete-time Gauss-Markov fading model, PSK achieves the rate

$$R_{\text{rt}} = \left(\frac{1}{\epsilon} - 1\right) \cdot \rho^2 + o(\rho^2) \quad \text{as } \rho \rightarrow 0. \quad (3.4.18)$$

For practical systems in which the fading processes are underspread [4], the innovation rate ϵ typically ranges from 1.8×10^{-2} to 3×10^{-7} [13]. So the $(1/2) \cdot \rho^2 + o(\rho^2)$

rate penalty of PSK with respect to optimal, peak-limited signaling is essentially negligible at low SNR.

Due to the simplicity of the discrete-time Gauss-Markov fading model, we are able to carry out a non-asymptotic analysis to gain more insights. Applying (3.4.17) to (3.3.7), the steady-state limiting channel prediction error is

$$\sigma_\infty^2 = \frac{(\rho - 1) \cdot \epsilon + \sqrt{(\rho - 1)^2 \cdot \epsilon^2 + 4\rho\epsilon}}{2\rho}. \quad (3.4.19)$$

Further applying (3.4.19) to (3.3.8), we can identify the following three qualitatively distinct operating regimes of the induced channel (3.3.5) for small $\epsilon \ll 1$:

- The quadratic regime: For $\rho \ll \epsilon$, $\sigma_\infty^2 \approx 1 - \rho/\epsilon$, and $\rho_\infty \approx \rho^2/\epsilon$;
- The linear regime: For $\epsilon \ll \rho \ll 1/\epsilon$, $\sigma_\infty^2 \approx \sqrt{\epsilon/\rho}$, and $\rho_\infty \approx \rho$;
- The saturation regime: For $1/\epsilon \ll \rho$, $\sigma_\infty^2 \approx \epsilon$, and $\rho_\infty \approx 1/\epsilon$.

Figure 3.4 illustrates these three regimes for $\epsilon = 10^{-4}$. The different slopes of ρ_∞ on the log-log plot are clearly visible for the three regimes. The linear regime covers roughly 80 dB, from -40 dB to $+40$ dB, in this particular example.

An interesting observation is that the two SNR thresholds dividing the three regimes are determined by a single parameter ϵ , which happens to be the one-step MMSE prediction error σ_{pred}^2 for the discrete-time Gauss-Markov fading process. The $1/\epsilon$ threshold dividing the linear and the saturation regimes coincides with that obtained in [13], where it is obtained for circular complex Gaussian inputs with nearest-neighbor decoding. In this chapter we investigate PSK, which results in a penalty in the achievable rate at high SNR. More specifically, it can be shown that the achievable rate for $\rho \gg 0$ behaves like $(1/2) \cdot \log \min\{\rho, 1/\epsilon\} + O(1)$ [81].

A further observation relevant to low-SNR system design is that the ϵ threshold dividing the quadratic and the linear regimes clearly indicates when the low-SNR

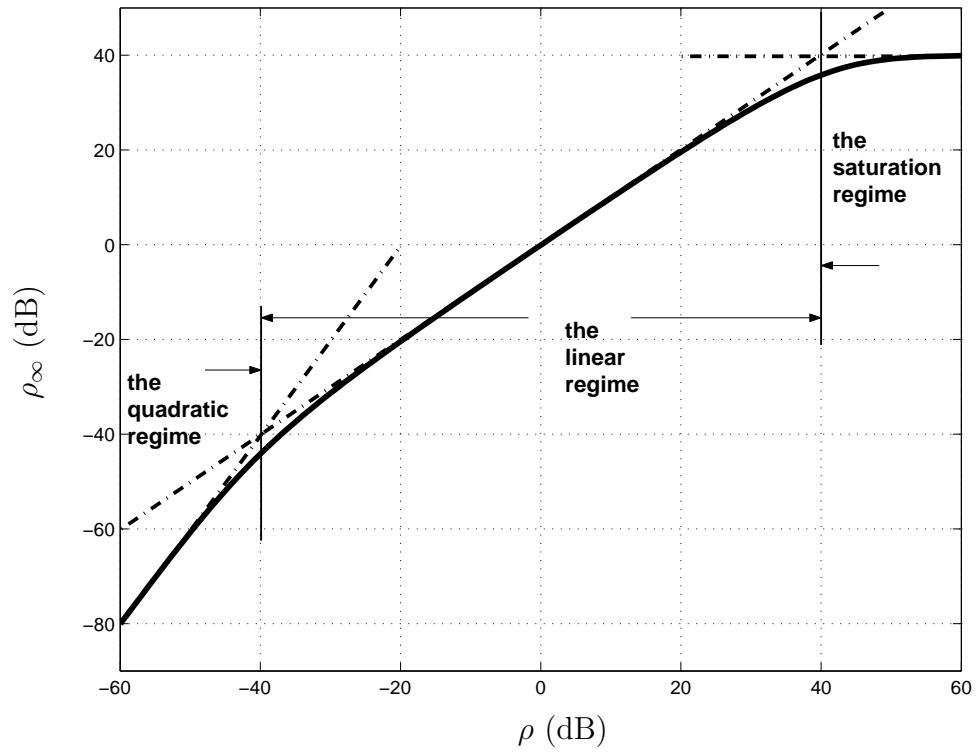


Figure 3.4. Illustration of the three operating regimes for the discrete-time Gauss-Markov fading model with $\epsilon = 10^{-4}$.

asymptotic channel behavior becomes dominant. Since the innovation rate ϵ for underspread fading processes is typically small, we essentially have a low-SNR channel with perfect receive CSI above $\rho = \epsilon$. This suggests that there may be an “optimal” SNR at which the low-SNR capacity limit with an average SNR constraint is the most closely approached. Figure 3.5 plots the normalized achievable rate R_{rt}/ρ vs. SNR, in which the achievable rate R_{rt} is numerically evaluated for the induced channel (3.3.5) using QPSK. Although all the curves vanish rapidly below the threshold $\rho = \epsilon$, for certain $\rho > \epsilon$, the normalized achievable rate can be reasonably close to 1. For example, taking $\epsilon = 10^{-4}$, the “optimal” SNR is $\rho \approx -15$ dB, and the corresponding normalized achievable rate is above 0.9, *i.e.*, more than 90% of the low-SNR capacity limit is achieved. Figure 3.6 further plots the achievable rate R_{rt} vs. SNR for PSK inputs with different constellation size. The innovation rate is fixed as $\epsilon = 10^{-4}$. It can be seen that around $\rho = 0$ dB, letting the constellation size be four, *i.e.*, QPSK, appears sufficient. For higher SNR, larger constellations perform better, and PSK loses degrees of freedom compared to the coherent capacity curve.

3.5 Filling the Gap to Capacity by Widening the Input Bandwidth

In Section 3.4 we have investigated the achievable information rate of the discrete-time channel (3.2.3), which is obtained from the continuous-time channel (3.2.1) as described in Section 3.2. In that context, the symbol duration T is a fixed system parameter. In this section we will show that, if we are allowed to reduce T , *i.e.*, widen the input bandwidth, then the recursive training scheme using PSK achieves an information rate that is asymptotically consistent with the channel capacity under peak envelope P . More specifically, we have the following proposition.

Proposition 3.5.1 *For the continuous-time channel (3.2.1) with envelope $P > 0$, as the symbol duration $T \rightarrow 0$, its steady-state induced channel (3.3.10) resulting*

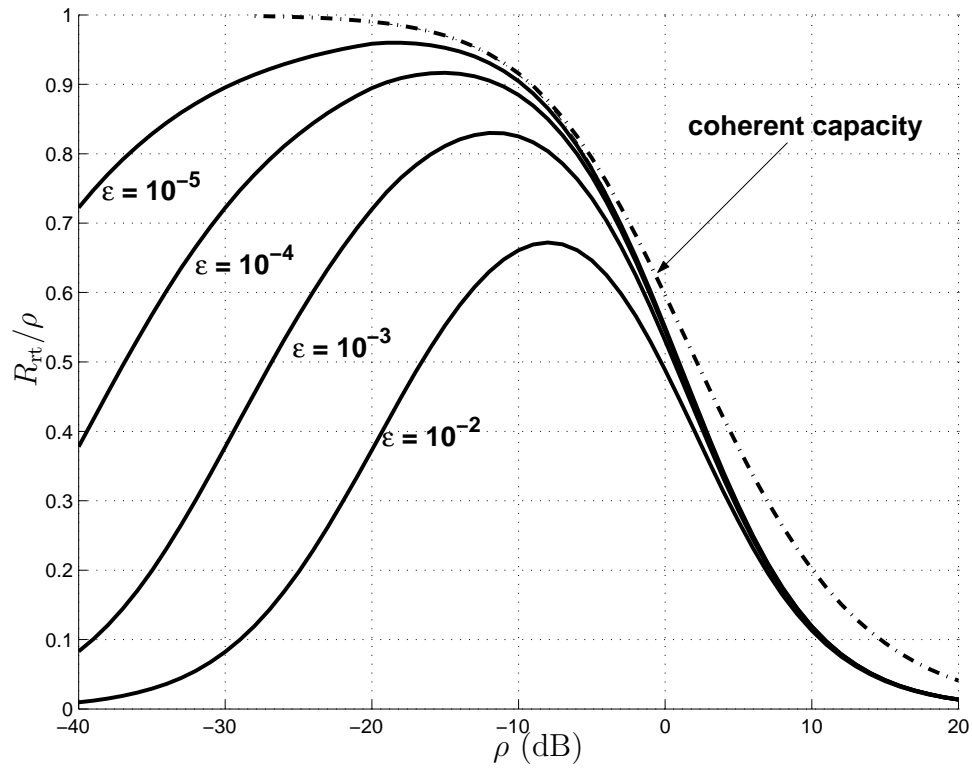


Figure 3.5. Normalized rate R_{rt}/ρ vs. SNR for recursive training with QPSK on the discrete-time Gauss-Markov fading channel. For comparison, the dashed-dot curve is the channel capacity normalized by SNR with an average SNR constraint ρ and perfect receive CSI, achieved by circular complex Gaussian inputs.

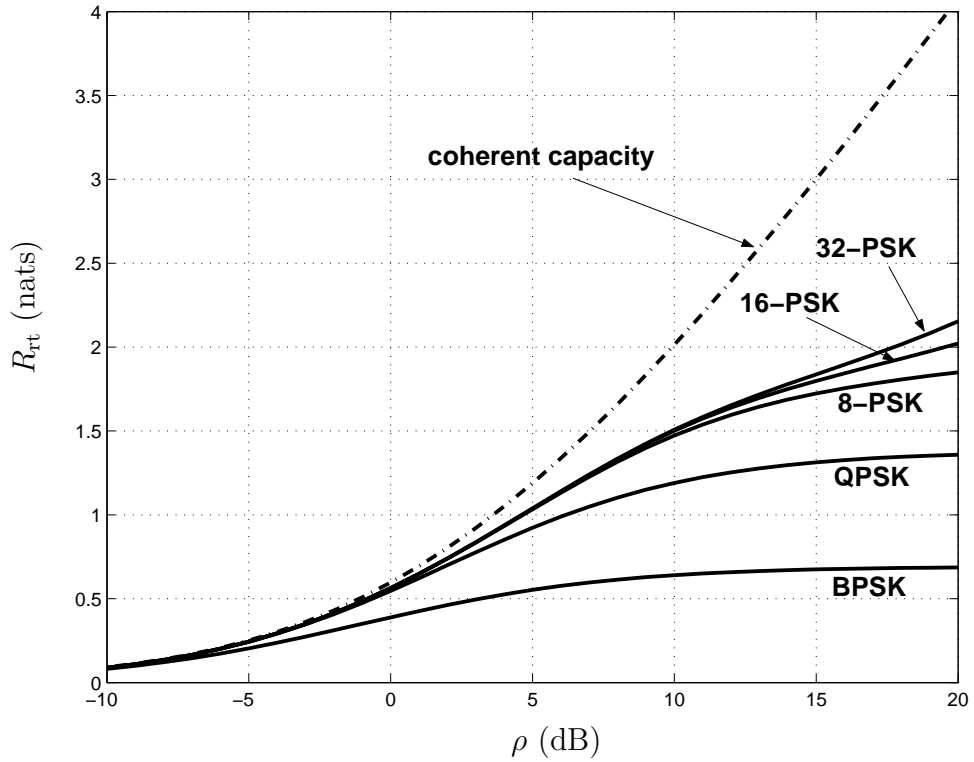


Figure 3.6. Rate R_{rt} vs. SNR for recursive training with different PSK constellations on the discrete-time Gauss-Markov fading channel with $\epsilon = 10^{-4}$. For comparison, the dashed-dot curve is the channel capacity with an average SNR constraint ρ and perfect receive CSI, achieved by circular complex Gaussian inputs.

from PSK with recursive training achieves the rate

$$\lim_{T \rightarrow 0} \frac{R_{\text{rt}}}{T} = \left[1 - \frac{1}{P} \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} \log(1 + P \cdot S_{H_c}(j\omega)) d\omega \right] \cdot P. \quad (3.5.1)$$

Proof: In Section 3.2 we have noted that the spectral density function $S_{H_d}(e^{j\Omega})$ of the discrete-time fading process is related to $S_{H_c}(j\omega)$ through

$$S_{H_d}(e^{j\Omega}) = \frac{T}{\int_0^T \int_0^T K_{H_c}(s-t) ds dt} \cdot \sum_{k=-\infty}^{\infty} S_{H_c} \left(j \frac{\Omega - 2k\pi}{T} \right) \cdot \text{sinc}^2(\Omega - 2k\pi) \quad (3.5.2)$$

and that the SNR of the discrete-time channel (3.2.3) is given by

$$\rho = \left(\int_0^T \int_0^T K_{H_c}(s-t) ds dt \right) \cdot \frac{P}{T}. \quad (3.5.3)$$

For the proof, the following two identities are useful:

$$\lim_{T \rightarrow 0} \frac{1}{T^2} \cdot \left(\int_0^T \int_0^T K_{H_c}(s-t) ds dt \right) = 1 \quad (3.5.4)$$

$$\lim_{T \rightarrow 0} \frac{1}{T} \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(1 + \rho \cdot S_{H_d}(e^{j\Omega})) d\Omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \log(1 + P \cdot S_{H_c}(j\omega)) d\omega \quad (3.5.5)$$

The second identity (3.5.5) has been established in [64, Claims 8.1 and 8.2]. To prove (3.5.4), note that for the mean-square continuous fading process $\{H_c(t) : -\infty < t < \infty\}$, its autocorrelation function $K_{H_c}(\tau) = \mathcal{E}\{H_c(t+\tau)H_c^\dagger(t)\}$ is continuous for all $-\infty < \tau < \infty$. Hence for any $T > 0$, there exists $T^* \in [0, T]$ such that

$$\int_0^T \int_0^T K_{H_c}(s-t) ds dt = K_{H_c}(T^*) \cdot T^2 \quad (3.5.6)$$

$$\rightarrow K_{H_c}(0) \cdot T^2 = T^2 \quad \text{as } T \rightarrow 0. \quad (3.5.7)$$

Now substituting (3.5.4) and (3.5.5) into (3.3.7), we have

$$\begin{aligned} \sigma_\infty^2 &= \frac{1}{\rho} \cdot \left\{ \exp \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(1 + \rho \cdot S_{H_d}(e^{j\Omega})) d\Omega \right\} - 1 \right\} \\ &\stackrel{(a)}{=} \frac{1}{\rho} \cdot \left\{ \exp \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \log(1 + P \cdot S_{H_c}(j\omega)) d\omega \cdot T + o(T) \right\} - 1 \right\} \\ &\stackrel{(b)}{=} \frac{1}{\rho} \cdot \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} \log(1 + P \cdot S_{H_c}(j\omega)) d\omega \cdot T + o(T) \right] \\ &\stackrel{(c)}{=} \frac{1}{P} \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} \log(1 + P \cdot S_{H_c}(j\omega)) d\omega + o(1) \quad \text{as } T \rightarrow 0, \end{aligned} \quad (3.5.8)$$

where (a) results from (3.5.5), (b) results from $e^x = 1 + x + o(x)$ as $x \rightarrow 0$, and (c) results from

$$\begin{aligned}\rho &= \left(\int_0^T \int_0^T K_{H_c}(s-t) ds dt \right) \cdot \frac{P}{T} \\ &= \frac{T^2 + o(T^2)}{T} \cdot P = (T + o(T)) \cdot P.\end{aligned}\quad (3.5.9)$$

That is,

$$\lim_{T \rightarrow 0} \sigma_\infty^2 = \frac{1}{P} \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} \log(1 + P \cdot S_{H_c}(j\omega)) d\omega. \quad (3.5.10)$$

Then substituting (3.5.4) and (3.5.10) into (3.3.8), we have

$$\begin{aligned}\lim_{T \rightarrow 0} \frac{\rho_\infty}{T} &= \lim_{T \rightarrow 0} \frac{(1 - \sigma_\infty^2) \cdot \rho}{(\sigma_\infty^2 \cdot \rho + 1) \cdot T} \\ &= \left[1 - \frac{1}{P} \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} \log(1 + P \cdot S_{H_c}(j\omega)) d\omega \right] \cdot P.\end{aligned}\quad (3.5.11)$$

Finally Proposition 3.5.1 immediately follows from substituting (3.5.11) into (3.3.11).

Q.E.D.

Again we compare the asymptotic achievable rate (3.5.1) to a capacity upper bound based upon the capacity per unit energy. For the continuous-time channel (3.2.1), the capacity per unit energy under a peak envelope constraint $P > 0$ is [64]

$$\dot{C} = 1 - \frac{1}{2\pi P} \cdot \int_{-\infty}^{\infty} \log(1 + P \cdot S_{H_c}(j\omega)) d\omega, \quad (3.5.12)$$

and the related capacity upper bound (measured per unit time) is [64]

$$C \leq U(P) := \left[1 - \frac{1}{2\pi P} \cdot \int_{-\infty}^{\infty} \log(1 + P \cdot S_{H_c}(j\omega)) d\omega \right] \cdot P. \quad (3.5.13)$$

Comparing (3.5.1) and (3.5.13), it is surprising to notice that these two quantities coincide. Recalling that in Section 3.4 we have noticed a $(1/2) \cdot \rho^2 + o(\rho^2)$ rate penalty in discrete-time channels, we conclude that widening the input bandwidth eliminates

this penalty and essentially results in an asymptotically capacity-achieving scheme in the wideband regime.⁶

The channel capacity of continuous-time peak-limited wideband fading channels (3.5.1) was originally obtained in [76]. However, in [76] the capacity is achieved by FSK, which is bursty in frequency. In our Proposition 3.5.1, we show that the capacity is also achievable if we employ recursive training and PSK, which is bursty in neither time nor frequency.

After some manipulations of (3.5.1), we further have that

- As $P \rightarrow 0$,

$$\frac{\lim_{T \rightarrow 0}(R_{\text{rt}}/T)}{P^2} \rightarrow \frac{1}{2} \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{H_c}^2(j\omega) d\omega, \quad (3.5.14)$$

if the above integral exists.

- As $P \rightarrow \infty$,

$$\frac{\lim_{T \rightarrow 0}(R_{\text{rt}}/T)}{P} \rightarrow 1. \quad (3.5.15)$$

In the sequel we will see that (3.5.14) and (3.5.15) are useful for asymptotic analysis.

3.5.1 An Intuitive Explanation of Proposition 3.5.1

In our proof of Proposition 3.5.1, we have utilized identities (3.5.4) and (3.5.5) to conveniently relate the continuous-time channel (3.2.1) to the discrete-time channel (3.2.3). However, these identities also have concealed much of the intuition contained in the derivation. To further illustrate the underlying mechanism in Proposition 3.5.1, here we give an alternative argument. Although the following reasoning is not mathematically rigorous, it does provide an intuitive way to understand the channel behavior as the symbol duration $T \rightarrow 0$.

⁶Again, the same caveat as in footnote 3 applies. So from a practical viewpoint, the result of Proposition 3.5.1 may be applicable when the symbol duration is much smaller than the channel coherence time, but much greater than the multipath delay spread.

In Section 3.2, we have described the conversion from the continuous-time channel (3.2.1) to the discrete-time channel (3.2.3). Strictly speaking, the discrete-time fading coefficient $H_d[k]$ is the k th sample of the matched-filtered, continuous-time fading process. The matched-filtering effect can be viewed as averaging $H_c(t)$ within a symbol interval of length T . Since we have assumed that the continuous-time fading process $\{H_c(t) : -\infty < t < \infty\}$ is mean-square continuous in t , roughly speaking, as $T \rightarrow 0$, the discrete-time fading coefficient $H_d[k] \approx H_c(kT)$, and the SNR per symbol $\rho \approx P \cdot T$. Furthermore, compared to sufficiently small T , the fading process $\{H_c(t) : -\infty < t < \infty\}$ can be viewed as essentially band-limited. So the discrete-time fading process $\{H_d[k] : -\infty < k < \infty\}$ is approximately the sampled continuous-time fading process $\{H_c(t) : -\infty < t < \infty\}$ with sampling rate well beyond its Nyquist rate, and we may write $S_{H_d}(e^{j\Omega}) \approx (1/T) \cdot S_{H_c}(j\Omega/T)$ for $-\pi \leq \Omega \leq \pi$.

Now let us apply the above approximations to (3.3.7) to evaluate σ_∞^2 for small T :

$$\begin{aligned}
\sigma_\infty^2 &= \frac{1}{\rho} \cdot \left\{ \exp \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(1 + \rho \cdot S_{H_d}(e^{j\Omega})) d\Omega \right\} - 1 \right\} \\
&\stackrel{(a)}{\approx} \frac{1}{P \cdot T} \cdot \left\{ \exp \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left(1 + P \cdot S_{H_c}(j\frac{\Omega}{T}) \right) d\Omega \right\} - 1 \right\} \\
&\stackrel{(b)}{=} \frac{1}{P \cdot T} \cdot \left\{ \exp \left\{ \frac{1}{2\pi} \int_{-\pi/T}^{\pi/T} \log(1 + P \cdot S_{H_c}(j\omega)) d\omega \cdot T \right\} - 1 \right\} \\
&\stackrel{(c)}{\approx} \frac{1}{P \cdot T} \cdot \left\{ \exp \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \log(1 + P \cdot S_{H_c}(j\omega)) d\omega \cdot T \right\} - 1 \right\} \\
&\stackrel{(d)}{\approx} \frac{1}{P \cdot T} \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} \log(1 + P \cdot S_{H_c}(j\omega)) d\omega \cdot T \\
&= \frac{1}{P} \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} \log(1 + P \cdot S_{H_c}(j\omega)) d\omega, \tag{3.5.16}
\end{aligned}$$

where (a) results from $\rho \approx P \cdot T$ and $S_{H_d}(e^{j\Omega}) \approx (1/T) \cdot S_{H_c}(j\Omega/T)$, (b) results from a change of variables $\omega = \Omega/T$, (c) results from $\pi/T \rightarrow \infty$ as $T \rightarrow 0$, and (d) results from $e^x = 1 + x + o(x)$ as $x \rightarrow 0$. We notice that the above derivation leads to the

same σ_∞^2 as (3.5.10) in our proof.

3.5.2 Case Study: The Continuous-Time Gauss-Markov Fading Model

In this subsection, we apply Proposition 3.5.1 to analyze the continuous-time Gauss-Markov fading processes. Such a process has autocorrelation function

$$K_{H_c}(\tau) = (1 - \epsilon_c)^{|\tau|/2}, \quad (3.5.17)$$

where the parameter $0 < \epsilon_c \leq 1$ characterizes the channel variation, analogously to ϵ for the discrete-time case in Section 3.4. The spectral density function of the process is

$$S_{H_c}(j\omega) = \frac{|\log(1 - \epsilon_c)|}{\omega^2 + (\log(1 - \epsilon_c))^2 / 4}. \quad (3.5.18)$$

Applying Proposition 3.5.1, we find that the recursive training scheme using PSK with a wide bandwidth asymptotically achieves an information rate

$$\lim_{T \rightarrow 0} \frac{R_{\text{rt}}}{T} = P - \frac{|\log(1 - \epsilon_c)|}{2} \cdot \left(\sqrt{1 + \frac{4P}{|\log(1 - \epsilon_c)|}} - 1 \right) \quad (3.5.19)$$

$$= \frac{1}{|\log(1 - \epsilon_c)|} \cdot P^2 + o(P^2) \quad \text{as } P \rightarrow 0. \quad (3.5.20)$$

Figure 3.7 illustrates the achievable rate (3.5.19) vs. P for $\epsilon_c = 0.9$.

3.5.3 Case Study: Clarke's Fading Model

In this subsection, we apply Proposition 3.5.1 to analyze Clarke's fading processes. Such a process is usually characterized by its spectral density function [28]

$$S_{H_c}(j\omega) = \begin{cases} \frac{2}{\omega_m} \cdot \frac{1}{\sqrt{1 - (\omega/\omega_m)^2}}, & \text{if } |\omega| \leq \omega_m \\ 0, & \text{otherwise,} \end{cases} \quad (3.5.21)$$

where ω_m is the maximum Doppler frequency.

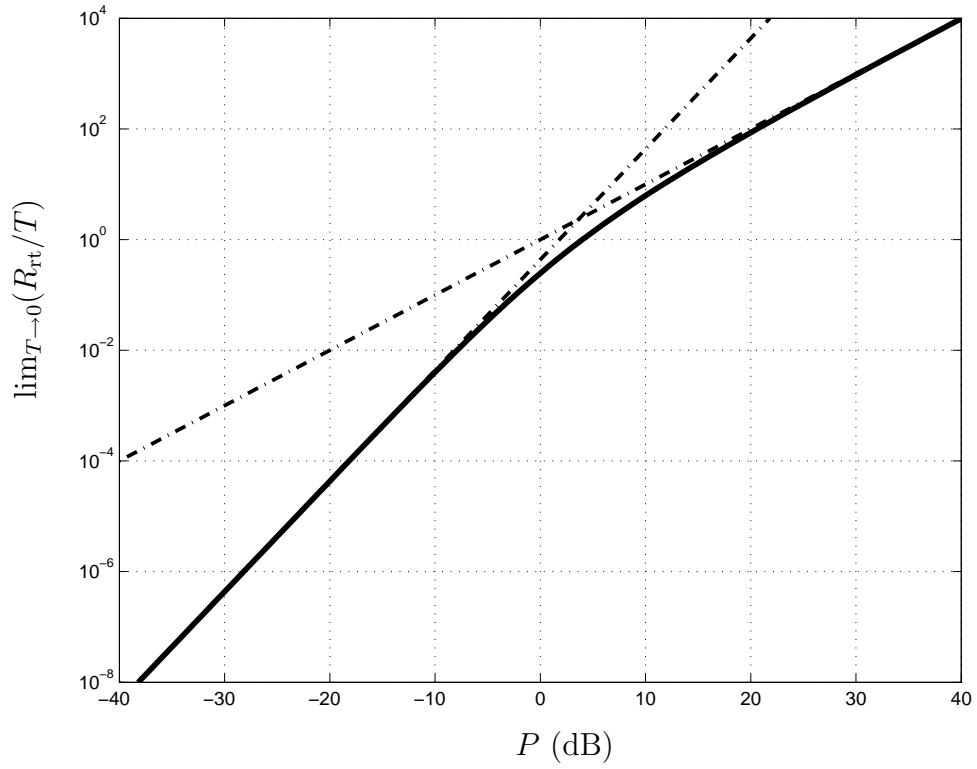


Figure 3.7. The asymptotic achievable rate $\lim_{T \rightarrow 0}(R_{rt}/T)$ vs. the envelope P , for recursive training with complex proper PSK on the continuous-time Gauss-Markov fading channel with innovation rate $\epsilon_c = 0.9$. The dashed-dot curves indicate the limiting behaviors for small and large P , *i.e.*, (3.5.20) and (3.5.15).

Applying Proposition 3.5.1, we find that

$$\lim_{T \rightarrow 0} \frac{R_{\text{rt}}}{T} = \begin{cases} \frac{\omega_m}{\pi} \cdot \left\{ \log \frac{\omega_m}{P} - \sqrt{1 - (2P/\omega_m)^2} \cdot \log \frac{\omega_m \cdot [1 + \sqrt{1 - (2P/\omega_m)^2}]}{2P} \right\}, & \text{if } P \leq \omega_m/2 \\ \frac{\omega_m}{\pi} \cdot \left\{ \log \frac{\omega_m}{P} + \sqrt{(2P/\omega_m)^2 - 1} \cdot \arctan \sqrt{(2P/\omega_m)^2 - 1} \right\}, & \text{if } P > \omega_m/2. \end{cases} \quad (3.5.22)$$

For large P , the asymptotic behavior of (3.5.22) is consistent with (3.5.15). For small P , however, the integral in (3.5.14) diverges, hence the asymptotic behavior of (3.5.22) scales super-quadratically with P . After some manipulations of (3.5.22), we find that

$$\lim_{T \rightarrow 0} \frac{R_{\text{rt}}}{T} = \frac{2}{\pi\omega_m} \cdot \log \frac{1}{P} \cdot P^2 + O(P^2) \quad \text{as } P \rightarrow 0. \quad (3.5.23)$$

Figure 3.8 illustrates the achievable rate (3.5.22) vs. P for $\omega_m = 100$. We notice that, for small P , the asymptotic expansion (3.5.23) is accurate.

3.6 Extension to Rician Fading Channels

In this section we extend the above developments to Rician fading channels. Similar to the Rayleigh fading case, we start with a continuous-time Rician fading channel model,

$$X(t) = \sqrt{\frac{d_c}{d_c + 1}} S(t) + \sqrt{\frac{1}{d_c + 1}} H_c(t) S(t) + Z(t), \quad (3.6.1)$$

where the assumptions for $S(\cdot)$, $X(\cdot)$, $H_c(\cdot)$ and $Z(t)$ are all the same as in Section 3.2. Compared to the Rayleigh fading channel model (3.2.1), the key difference in the Rician case is the presence of a parameter $d_c \in [0, \infty)$ corresponding to the direct LOS component. We assume that d_c is deterministic and precisely known. If $d_c = 0$ the channel reduces to being Rayleigh, and as $d_c \rightarrow \infty$ the channel tends to be Gaussian.

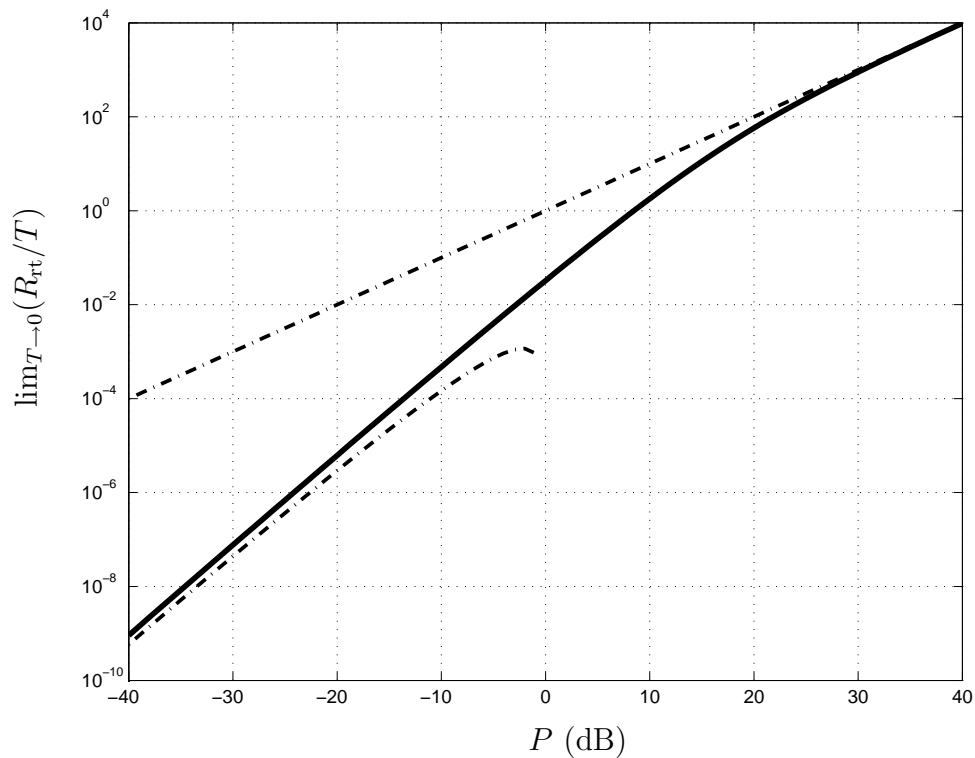


Figure 3.8. The asymptotic achievable rate $\lim_{T \rightarrow 0}(R_{rt}/T)$ vs. the envelope P , for recursive training with complex proper PSK on Clarke's fading channel with maximum Doppler frequency $\omega_m = 100$. The dashed-dot curves indicate the limiting behaviors for small and large P , *i.e.*, (3.5.23) and (3.5.15).

Again, processing the channel output through a matched filter, we obtain a discrete-time model

$$X[k] = \sqrt{\frac{d_d \rho}{d_d + 1}} S[k] + \sqrt{\frac{\rho}{d_d + 1}} H_d[k] S[k] + Z[k], \quad (3.6.2)$$

where the definitions for $S[\cdot]$, $X[\cdot]$, $H_d[\cdot]$ and $Z[\cdot]$ are all the same as in Section 3.2.

The average channel SNR experienced at the receiver is given by

$$\rho = \frac{P}{T} \cdot \frac{d_c T^2 + \int_0^T \int_0^T K_{H_c}(s-t) ds dt}{d_c + 1}, \quad (3.6.3)$$

correspondingly the direct LOS component for the discrete-time channel model is

$$d_d = \frac{T^2}{\int_0^T \int_0^T K_{H_c}(s-t) ds dt} \cdot d_c, \quad (3.6.4)$$

which converges to d_c as $T \rightarrow 0$.

In Appendix we derive the capacity per unit energy under peak constraints for the Rician fading channels (3.6.1) and (3.6.2), and a corresponding upper bound to the capacity as a function of SNR. These results complement the corresponding results for Rayleigh fading channels [64], and are useful for our subsequent developments in this section.

For the discrete-time channel (3.6.2), since the Rician fading process is also circular complex Gaussian, the recursive training procedure with PSK inputs for the Rayleigh fading case still applies. Following the same developments as in Section 3.3, we find that as the interleaving depth $L \rightarrow \infty$, the one-step MMSE prediction error converges to

$$\sigma_\infty^2 = \frac{d_d + 1}{\rho} \left\{ \exp \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left(1 + \frac{\rho}{d_d + 1} S_{H_d}(e^{j\Omega}) \right) d\Omega \right] - 1 \right\}. \quad (3.6.5)$$

Consequently, the steady-state channel can be written as

$$X = \left(\sqrt{\frac{d_d \rho}{d_d + 1}} + \sqrt{\frac{\rho}{d_d + 1}} \hat{H}_d \right) S + \underbrace{\left(\sqrt{\frac{\rho}{d_d + 1}} \tilde{H}_d S + Z \right)}_{\tilde{Z}}, \quad (3.6.6)$$

where the estimated fading coefficient $\hat{H}_d \sim \mathcal{CN}(0, 1 - \sigma_\infty^2)$ and the estimation error $\tilde{H}_d = H_d - \hat{H}_d \sim \mathcal{CN}(0, \sigma_\infty^2)$ are independent, and the effective noise \bar{Z} is still i.i.d. circular complex Gaussian and independent of S and \hat{H}_d .

Applying the first-order asymptotic expansion [75], we find that as $\rho \rightarrow 0$ the steady-state channel (3.6.6) achieves

$$R_{\text{rt}} := I\left(S; X \middle| \hat{H}_d\right) = \frac{d_d + 1 - \sigma_\infty^2}{d_d + 1 + \sigma_\infty^2} \rho + O(\rho^2). \quad (3.6.7)$$

Meanwhile, as $\rho \rightarrow 0$ the steady-state one-step MMSE prediction error behaves like (see the proof of Proposition 3.4.1)

$$\sigma_\infty^2 = 1 - \frac{1}{2(d_d + 1)} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega - 1 \right] \cdot \rho + o(\rho), \quad (3.6.8)$$

assuming that the integral $(1/2\pi) \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega$ exists.

After manipulating terms, we obtain the following Proposition.

Proposition 3.6.1 *For the discrete-time Rician fading channel (3.6.2), as $\rho \rightarrow 0$, its induced channel resulting from PSK with recursive training achieves the rate*

$$R_{\text{rt}} = \frac{d_d}{d_d + 1} \rho + \frac{1}{2} \left[\frac{1}{(d_d + 1)^2} \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega - 1 \right] \rho^2 + o(\rho^2), \quad (3.6.9)$$

if the integral $(1/2\pi) \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega$ exists.⁷

Now let us apply the capacity upper bound (4.10) derived in Appendix 4,

$$C \leq U(\rho) := \frac{d_d}{d_d + 1} \rho + \frac{1}{2(d_d + 1)^2} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega \right] \rho^2. \quad (3.6.10)$$

Comparing (3.6.9) and (3.6.10), we again notice that the penalty of using PSK instead of bursty \dot{C} -achieving channel input is at most $(1/2) \cdot \rho^2 + o(\rho^2)$, exactly the same as in the Rayleigh fading case.

Additional insight for Rician fading channels follows from inspecting the quadratic term in (3.6.9). As d_d increases from zero to infinity, the coefficient of the linear term,

⁷For simplicity of exposition, in the sequel we assume that the integral $(1/2\pi) \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega$ always exists. If it is not the case, *e.g.*, for Clarke's fading model, then additional care as in Section 3.5.3 should be taken.

$d_d/(d_d + 1)$, increases from zero to one, meanwhile the coefficient of the quadratic term decreases from a certain positive value to $-1/2$. There exists a threshold

$$d_d^* = \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega} - 1. \quad (3.6.11)$$

For $d_d \in [0, d_d^*)$, the achievable rate R_{rt} is a locally convex function of ρ at $\rho = 0$; and for $d_d \in (d_d^*, \infty)$, R_{rt} becomes locally concave. This type of behavior resembles that observed in [21] for memoryless Rician fading channels under different peak or fourth-order constraints.

Finally, in parallel to the Rayleigh fading case, we have by the same developments as in Section 3.5 that, in the wideband limit, recursive training using PSK inputs is capacity-achieving.

Proposition 3.6.2 *For the continuous-time Rician fading channel (3.6.1) with peak envelope power $P > 0$, as the symbol duration $T \rightarrow 0$, recursive training using PSK inputs achieves the rate*

$$\lim_{T \rightarrow 0} \frac{R_{rt}}{T} = \left[1 - \frac{1}{P} \frac{1}{2\pi} \int_{-\infty}^{\infty} \log \left(1 + \frac{P}{d_c + 1} S_{H_c}(j\omega) \right) d\omega \right] \cdot P, \quad (3.6.12)$$

which is the channel capacity in the wideband limit.

3.7 Concluding Remarks

For fading channels that exhibit temporal correlation, a key to enhancing communication performance is efficiently exploiting the implicit CSI embedded in the fading processes. From the preceding developments in this chapter, we see that a recursive training scheme, which performs channel estimation and demodulation/decoding in an alternating manner, accomplishes this job reasonably well, especially when the channel fading varies slowly. The main idea of recursive training is to repeatedly use decisions of previous information symbols as pilots, and to ensure the reliability of these decisions by coding over sufficiently long blocks. As such, recursive training can be viewed as a form of a block decision-feedback receiver.

Throughout this chapter, we restrict the channel inputs to zero-mean PSK, which is not optimal in general for fading channels without CSI. There are two main motivations for this choice. First, compared to other channel inputs such as circular complex Gaussian, PSK leads to a significant simplification of the analytical developments. As we saw, recursive training using PSK converts the original fading channel without CSI into a series of parallel sub-channels, each with estimated receive CSI but additional noise that remains circular complex white Gaussian. In this chapter we mainly investigate the steady-state limiting channel behavior of channel prediction; however, it may worth mentioning that, using the induced channel model presented in Section 3.3, exact evaluation of the transient channel behavior is straightforward, with the aid of numerical methods.

Second, PSK inputs perform reasonably well in the moderate to low SNR regime. This is due to the fact that, for fading channels with receive CSI, as SNR vanishes, the linear leading term in the channel capacity can be asymptotically achieved by rather general zero-mean inputs, not just circular complex Gaussian. The main contribution of our work is that it clearly separates the effect of an input peak-power constraint and the effect of replacing optimal peak-limited inputs with PSK, which is non-bursty in both time and frequency. Previous results [64, 75, 50, 22] indicate that in the low-SNR regime, requiring inputs be peak-limited drags the achievable rate down from linear to quadratic, and it is asymptotically optimal to use vanishing duty cycle, *i.e.*, bursty signaling, in transmission. Our results indicate that, for slowly time-varying fading processes, the rate loss from PSK inputs compared to bursty signaling is essentially negligible. Furthermore, as revealed by the non-asymptotic analysis for discrete-time Gauss-Markov fading processes, there appear to be non-vanishing SNRs at which near-coherent performance is attainable with recursive training and PSK.

CHAPTER 4

EXPLOITATION OF CHANNEL SPATIAL CORRELATION IN THE LOW-SNR REGIME

4.1 Introduction

In the previous chapter we have seen that channel temporal correlation plays a crucial role to compensate for the absence of CSI. In this chapter, we continue to demonstrate that spatial correlation among multiple transmit and receive antennas also eases the detrimental effects from the lack of CSI.

The result of channel spatial correlation in this chapter is fundamentally different from that in coherent channels at high SNR [71]. In that case, channels are bandwidth-limited. Multiple antennas increase the spatial degrees of freedom by exploiting the richness of the propagation environment. Channel spatial correlation is generally detrimental as it reduces the spatial degrees of freedom [9]. By contrast, in the non-coherent low-SNR case, channels are power-limited. Correlated antennas become beneficial because of two separate effects. First, they increase the effective channel SNR, through combining of the outputs of the multiple receive antennas, and through “beamforming” across the spatially correlated transmit antennas. Second, they reduce the requirement on peak input power, by distributing a peaky input among spatially correlated transmit antennas, and by virtually amplifying the received SNR at the spatially correlated receive antennas.

In addition to a study of the asymptotically achievable information rates of

general inputs, we also analyze the behavior of OOK with symbol-by-symbol hard decisions, and demonstrate that the corresponding channel mutual information essentially exhibits the benefits of spatial correlation, even in the non-asymptotic regime. This part of the work is mainly motivated by practice. We are often supposed to use wide bandwidths to support correspondingly high data rates, for which the vanishing spectral efficiency asymptotics may not be met. For example, in the Federal Communications Commission (FCC) approved ultra wide-bandwidth (UWB) systems (*e.g.*, [73]), wireless transceivers operate in the 3.1 – 10.6 GHz band to communicate at data rates typically higher than 100 Mbps. Thus, even if the whole 7.5 GHz bandwidth is occupied, the spectral efficiency is still greater than 0.01 b/s/Hz, which is small but non-vanishing. In fact, existing evidence shows that for non-coherent fading channels, asymptotic analyses become effective only at extremely low SNR. For example, as illustrated by numerical results [1, Figs. 2 and 7], in a scalar memoryless channel, for SNR from -20 dB to -100 dB, the peak power of the capacity-achieving OOK grows from 6 dB to 10 dB (which should grow unbounded as SNR vanishes), and the ratio of capacity to SNR increases from 0.45 to 0.7 (which should approach 1 as SNR vanishes). Such extremely slow growths clearly justify the necessity of a non-asymptotic study.

4.1.1 Connections with Related Research

Channel correlation leads to dramatically different channel behavior, depending on the regimes encountered. For coherent channels at high SNR, [9] characterizes the effects of channel spatial correlation upon the multiplexing gain. For coherent channels at low SNR, [41] characterizes the effects of channel spatial correlation upon both the first derivative $\dot{I}(0)$ and the second derivative $\ddot{I}(0)$ of the mutual information. As long as the fading process is ergodic, temporal correlation does

not affect the ergodic capacity of coherent channels, since its effect is eliminated by averaging over time. For non-coherent channels, however, temporal correlation is important. At high SNR, temporal correlation directly affects the achievable multiplexing gain of block fading channels without spatial correlation [82]. The authors of [27] address some general issues of spatial correlation in non-coherent block-fading channels, indicating that at least for block length one, *i.e.*, memoryless fading, spatial correlation is beneficial to channel capacity.

For temporally correlated scalar channels, the capacity per unit energy under a peak input power constraint is obtained in [64]. The optimal input gradually tends to bursts of “on” intervals sporadically inserted into the “off” background, with the duration of “on” intervals growing unbounded and with the ratio between “on” and “off” intervals vanishing. In our work we consider channel spatial correlation, and for certain cases we establish the capacity per unit energy, which is shown to be achieved by a simple spatial power allocation strategy.

The idea of exploiting channel spatial correlation in the non-coherent low-SNR scenario is also partially explored in [80], where the authors consider a block fading model and investigate the low-power reliability function. Channel spatial correlation leads to significant benefits on the low-power reliability function, for which the maximum performance gain is achieved when the channel is fully correlated in the spatial dimension.

4.2 Channel Model

We consider a discrete-time frequency-nonselective Rayleigh fading channel with multiple transmit and receive antennas. Focusing on channel spatial correlation, for simplicity we assume that there is no channel temporal correlation among adjacent channel uses. Hence the channel is memoryless, and in the sequel we suppress the

time indexes; the development to follow can also be directly paralleled for a block fading model. Equipped with M transmit and N receive antennas, the channel input-output relationship is

$$X^T = S^T \cdot H_d + Z^T, \quad (4.2.1)$$

where $S \in \mathbb{C}^{M \times 1}$ is the channel input, and $X \in \mathbb{C}^{N \times 1}$ is the channel output. The additive noise $Z \in \mathbb{C}^{N \times 1}$ is a circular complex Gaussian vector with zero-mean and identity covariance matrix, hence it is spatially white. $H_d \in \mathbb{C}^{M \times N}$ is the fading matrix, with possibly non-white covariance structure, *i.e.*, it is in general spatially correlated. To be specific, we utilize the following Kronecker spatial correlation model:

$$H_d = \Phi_T^{1/2} \cdot H_{d0} \cdot \Phi_R^{1/2}, \quad (4.2.2)$$

where the real symmetric matrices $\Phi_T \in \mathbb{R}^{M \times M}$ and $\Phi_R \in \mathbb{R}^{N \times N}$ are both positive semi-definite, and describe the spatial correlation among the transmit antennas and the receive antennas, respectively. We normalize Φ_T and Φ_R so that their diagonal elements are all unity. H_{d0} is an $M \times N$ random matrix with i.i.d. $\mathcal{CN}(0, 1)$ elements. This separable spatial correlation model is identical to that used in [9, 41, 27], and has been validated for certain scenarios by channel measurements [33].

Remark: There exist other spatial correlation models that take into account coupling between transmit and receive sides; see [77] and references therein. These more sophisticated models may characterize realistic channels more accurately; however, they are less tractable for the analytical approach in this chapter. We note that numerical computation can be conducted for these more general models following our analyses, at least in principle.

Throughout the chapter, unless otherwise stated, we assume that neither the

transmitter nor the receiver has access to the realization of H_d , whereas the covariance matrices Φ_T and Φ_R are perfectly identified at both sides.

Under the Kronecker spatial correlation model (4.2.2) in the channel equation (4.2.1), the channel output conditioned upon a given input is circular complex Gaussian with conditional density

$$p_{X|S}(\mathbf{x}|\mathbf{s}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N \times N} + (\mathbf{s}^\dagger \Phi_T \mathbf{s}) \cdot \Phi_R). \quad (4.2.3)$$

4.3 Asymptotic Analysis in the Low-SNR Regime: First and Second Derivatives

In this section we focus on the channel behavior at asymptotically low SNR. Specifically, we evaluate the first and the second derivatives of the channel mutual information at zero SNR. It is unnecessary to specify any average SNR constraint since it is already set at zero. Instead we impose a peak SNR constraint, measured by SNR ρ_p per transmit antenna, *i.e.*,

$$S_m^\dagger S_m \leq \rho_p, \quad \text{for } m = 1, \dots, M,$$

where S_m denotes the input at the m -th transmit antenna.

Consider OOK with an “off” signal $S = \mathbf{0}$ and a non-zero “on” signal $S = \mathbf{s}$, then from [74, (2)], the first derivative of the resulting channel mutual information at zero SNR is

$$\dot{I}(0) = \frac{\mathcal{D}(p_{X|S}(\mathbf{x}|\mathbf{s}) \| p_{X|S}(\mathbf{x}|\mathbf{0}))}{\mathbf{s}^\dagger \mathbf{s}},$$

where $\mathcal{D}(\cdot \| \cdot)$ is the relative entropy, or KL divergence. Furthermore, \dot{C} , the channel capacity per unit energy, is simply the supremum of $\dot{I}(0)$ taken over all allowable “on” signals \mathbf{s} .

Specializing the general results above to the channel model in this chapter, we

find that its capacity per unit energy is

$$\dot{C} = \max_{\substack{\mathbf{s}: s_m^\dagger s_m \leq \rho_p \\ m=1, \dots, M}} \left[N \cdot \frac{\mathbf{s}^\dagger \Phi_T \mathbf{s}}{\mathbf{s}^\dagger \mathbf{s}} - \sum_{n=1}^N \frac{\log(1 + \lambda_n(\Phi_R) \cdot (\mathbf{s}^\dagger \Phi_T \mathbf{s}))}{\mathbf{s}^\dagger \mathbf{s}} \right], \quad (4.3.1)$$

where $\lambda_n(\Phi_R)$ is the n th eigenvalue of Φ_R .

The maximization problem of finding the capacity per unit energy is finite-dimensional, and can be solved with the aid of optimization software packages. In general, however, little can be said about its analytical solution, because this optimization problem lacks certain convenient properties (*e.g.*, convexity). As the peak SNR $\rho_p \rightarrow \infty$, the second term in (4.3.1) vanishes compared to the first dominating term, and we have $\dot{C} \rightarrow N \cdot \lambda_{\max}(\Phi_T)$. As ρ_p decreases, the effect of the second term cannot be ignored, and the maximization problem typically yields no closed-form solution.

An important exception to the above general case is as follows, for which we can evaluate \dot{C} and characterize its corresponding “on” signal \mathbf{s} in a simple way. We call an “on” signal \mathbf{s} “flat” if its M elements are all with the same energy, *i.e.*, $s_m^\dagger s_m = s_{m'}^\dagger s_{m'}$ for $m, m' = 1, \dots, M$. We have the following proposition.

Proposition 4.3.1 *If there exists a flat signal \mathbf{s} with $s_m^\dagger s_m = \rho_p$ for $m = 1, \dots, M$, and such that*

$$\frac{\mathbf{s}^\dagger \Phi_T \mathbf{s}}{\mathbf{s}^\dagger \mathbf{s}} = \lambda_{\max}(\Phi_T), \quad (4.3.2)$$

then OOK with the “on” signal $S = \mathbf{s}$ achieves the capacity per unit energy, and (4.3.1) reduces to

$$\dot{C} = N \cdot \lambda_{\max}(\Phi_T) - \sum_{n=1}^N \frac{\log(1 + M\rho_p \cdot \lambda_{\max}(\Phi_T) \cdot \lambda_n(\Phi_R))}{M\rho_p}. \quad (4.3.3)$$

The proof of Proposition 4.3.1 is in Appendix .5.

Next we consider $\ddot{I}(0)$, the second derivative of channel mutual information at zero SNR, for OOK. For this purpose, we need to specify that the “on” signal $S = \mathbf{s}$

is used with probability $0 < \delta < 1$, hence the average SNR is $\rho = (\mathbf{s}^\dagger \mathbf{s}) \cdot \delta$. By its definition, $\ddot{I}(0)$ can be evaluated as

$$\ddot{I}(0) := 2 \cdot \lim_{\rho \rightarrow 0} \frac{I(\rho) - \dot{I}(0) \cdot \rho}{\rho^2}.$$

In Appendix .6 we prove the following proposition.

Proposition 4.3.2 *For a given OOK with the “on” signal $S = \mathbf{s}$,*

$$\ddot{I}(0) = \begin{cases} - \left[\prod_{n=1}^N (1 - (\mathbf{s}^\dagger \Phi_{\mathbf{T}} \mathbf{s})^2 \cdot \lambda_n^2(\Phi_{\mathbf{R}}))^{-1} - 1 \right] / (\mathbf{s}^\dagger \mathbf{s})^2, & \text{if } (\mathbf{s}^\dagger \Phi_{\mathbf{T}} \mathbf{s}) \cdot \lambda_{\max}(\Phi_{\mathbf{R}}) < 1 \\ -\infty, & \text{otherwise} \end{cases}. \quad (4.3.4)$$

4.3.1 Interpretation of the Asymptotics

To make the following discussion concrete, let us focus on the case that there exists a flat “on” input \mathbf{s} that achieves \dot{C} as expressed in (4.3.3). We then introduce the relative loss factor

$$\mathcal{L} = \frac{1}{MN\rho_p \cdot \lambda_{\max}(\Phi_{\mathbf{T}})} \cdot \sum_{n=1}^N \log(1 + M\rho_p \cdot \lambda_{\max}(\Phi_{\mathbf{T}}) \cdot \lambda_n(\Phi_{\mathbf{R}})),$$

and rewrite (4.3.3) as $\dot{C} = N \cdot \lambda_{\max}(\Phi_{\mathbf{T}}) \cdot (1 - \mathcal{L})$.

The benefits of channel spatial correlation stem from the fact that, when the channel is correlated, the sum numerator in \mathcal{L} may grow much slower than its denominator $MN\rho_p \cdot \lambda_{\max}(\Phi_{\mathbf{T}})$. We can bound \mathcal{L} by

$$\frac{\log(1 + M^2 N \rho_p)}{M^2 N \rho_p} \leq \mathcal{L} \leq \frac{\log(1 + M \rho_p)}{M \rho_p}.$$

The upper bound is achieved when there is no spatial correlation, *i.e.*, $\Phi_{\mathbf{T}} = \mathbf{I}_{M \times M}$ and $\Phi_{\mathbf{R}} = \mathbf{I}_{N \times N}$. The lower bound is achieved when there is full spatial correlation, *i.e.*, $\Phi_{\mathbf{T}} = \mathbf{E}_{M \times M}$ and $\Phi_{\mathbf{R}} = \mathbf{E}_{N \times N}$, where $\mathbf{E}_{M \times M}$ denotes a $M \times M$ matrix in which all the elements are one. Compared to the case of no spatial correlation, we notice that channel spatial correlation can virtually amplify the peak SNR by at most a factor of MN .

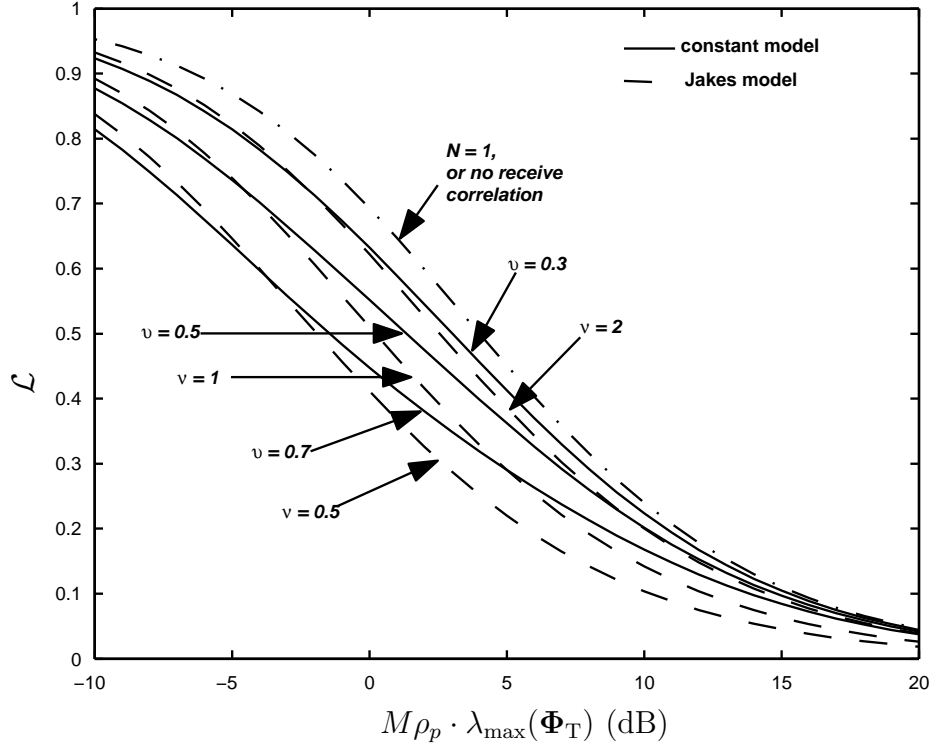


Figure 4.1. Relative loss \mathcal{L} vs. normalized peak SNR $M\rho_p \cdot \lambda_{\max}(\Phi_T)$ for different receive correlation models: the constant correlation model and the Jakes model. The number of receive antennas is $N = 10$.

To illustrate the benefits of spatial correlation on \mathcal{L} , we plot \mathcal{L} vs. $M\rho_p \cdot \lambda_{\max}(\Phi_T)$ in Figure 4.1, for two typical correlation models of Φ_R . In the constant correlation model, $\Phi_R(i, j) = \nu$ if $i \neq j$; and in the Jakes model, $\Phi_R(i, j) = J_0(\nu \cdot (i - j))$, where $\nu > 0$ is determined by the radio wavelength and the receive antenna spacing [67], and $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind. Compared to the case of no receive correlation, the existence of correlation typically leads to a reduction of 3 – 10 dB on the peak SNR for achieving the same \mathcal{L} . It appears that the constant correlation model leads to more evident reduction at lower peak SNR, whereas for higher peak SNR the Jakes model performs more effectively.

Now let us turn to Proposition 4.3.2. For OOK channel inputs, (4.3.4) reveals an

interesting phase transition phenomenon. For low peak SNR, $\ddot{I}(0)$ is finite, hence the channel mutual information $I(\rho)$ may well be approximated by $\dot{I}(0) \cdot \rho + (\ddot{I}(0)/2) \cdot \rho^2$ for reasonably low SNR. Once the peak SNR exceeds the threshold as given in (4.3.4), $\ddot{I}(0)$ becomes divergent,¹ implying that extremely low SNR is required for $I(\rho)/\rho$ to approach the first derivative $\dot{I}(0)$.

To illustrate the phase transition phenomenon, we plot $I(\rho)/\rho$ vs. ρ in Figure 4.2, for the scalar channel case ($M = N = 1$). Solid curves indicate the normalized channel mutual information using OOK with different fixed peak SNR ρ_p , and dashed-dot horizontal lines correspond to $\dot{I}(0)$ for the given ρ_p . Hence as $\rho \rightarrow 0$ each solid curve will eventually approach its corresponding $\dot{I}(0)$ limit. For the scalar channel, $\dot{I}(0)$ and $\ddot{I}(0)$ reduce to

$$\dot{I}(0) = 1 - \frac{\log(1 + \rho_p)}{\rho_p},$$

$$\text{and } \ddot{I}(0) = \begin{cases} -\frac{1}{1-\rho_p^2}, & \text{if } \rho_p < 1 \\ -\infty, & \text{otherwise} \end{cases},$$

respectively. In Figure 4.2 we notice that, for $\rho_p < 1$ (*e.g.*, $-10, -3$ dB), the gap between $I(\rho)/\rho$ and $\dot{I}(0)$ essentially vanishes below $\rho = -20$ dB; at the threshold $\rho_p = 1$ (*i.e.*, 0 dB), the gap slightly increases whereas it still becomes small below $\rho = -30$ dB; for $\rho_p > 1$ (*e.g.*, 3, 10 dB), the gap becomes quite noticeable, implying that the effect of divergent $\ddot{I}(0)$ is significant for finite ρ .

Comparing $\dot{I}(0)$ and $\ddot{I}(0)$, we can clearly observe the key role of signal peakiness. Higher peak SNR increases $\dot{I}(0)$, but it also leads to divergent $\ddot{I}(0)$. When channels exhibit spatial correlation, these effects are even more significant, since channel spatial correlation virtually amplifies the effective peak SNR. As a result, the asymptotic analyses in this section usually are accurate only for rather low peak SNR (such that $|\ddot{I}(0)| < \infty$) or for extremely low average SNR (such that

¹In other words, the wideband slope as defined in [75] vanishes.

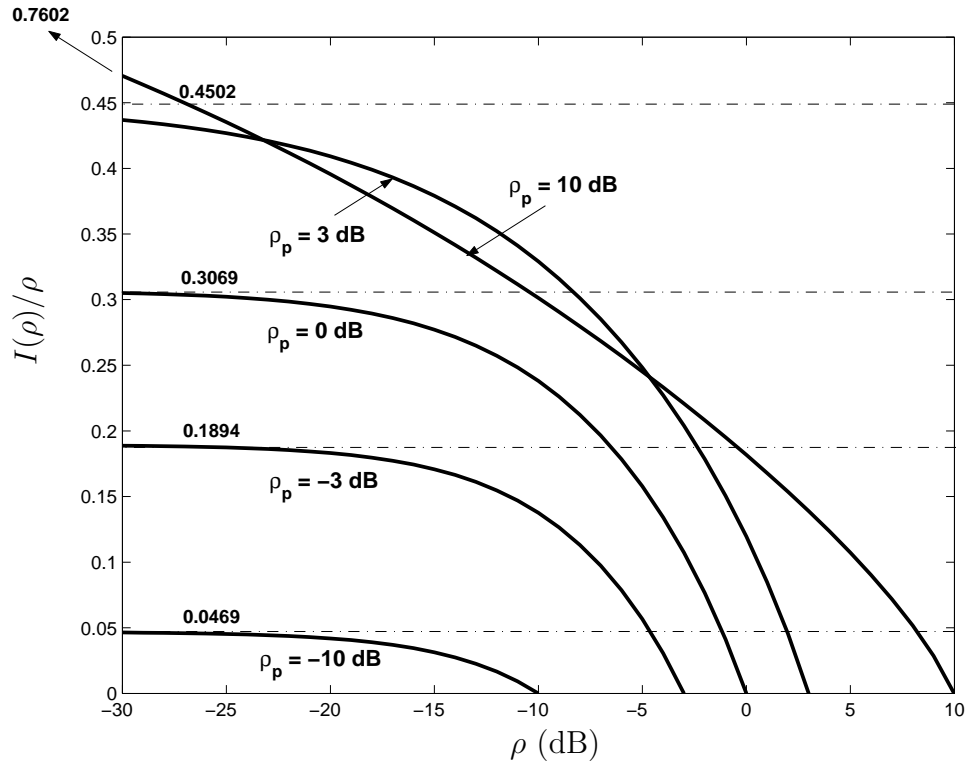


Figure 4.2. $I(\rho)/\rho$ vs. ρ for the scalar channel case. The dashed-dot horizontal lines indicate the corresponding $\dot{I}(0)$ limit as $\rho \rightarrow 0$.

$I(\rho)/\rho \approx \dot{I}(0)$); hence, they appear to be insufficient for characterizing the channel behavior for many practical applications.

4.4 Non-Asymptotic Study of OOK with Symbol-by-Symbol Hard Decisions

In this section, we turn to a simple scheme of OOK with symbol-by-symbol hard decisions, taking advantage of its simplicity, in order to gain more insight on the channel behavior in the low, but non-vanishing, SNR regime.

OOK with symbol-by-symbol hard decisions is inherently sub-optimal compared to techniques that exploit maximum-likelihood (ML) soft decoding. From a practical perspective, this scheme incurs a minimal amount of processing complexity, thus it is a choice especially suitable for receivers with limited resources, which is often the case in the non-coherent scenario. As will be shown, even this particularly simple scheme is capable of realizing many of the benefits suggested by the asymptotic analyses in Section 4.3.

OOK inputs have been introduced in Section 4.3. The channel input is

$$\mathbf{S} = \begin{cases} \mathbf{s}, & \text{w.p. } \delta \\ \mathbf{0}, & \text{w.p. } 1 - \delta, \end{cases},$$

where $0 < \delta < 1$. For a given average SNR ρ , we then have

$$(\mathbf{s}^\dagger \mathbf{s}) \cdot \delta = \rho. \tag{4.4.1}$$

In this section we do not adopt a strict peak SNR constraint, instead we will compare the peak SNR for channels with different correlation structures.

Each symbol-by-symbol hard decision is a binary detection problem, for which the maximum a posteriori (MAP) decision rule minimizes the average probability of decision error. Our performance metric here, however, is the mutual information of the resulting binary discrete memoryless channel (DMC), which is not necessarily

maximized by the MAP decision rule. Thus, we consider the general likelihood ratio test, *i.e.*, the Neyman-Pearson decision rule [58]:

$$\hat{\mathbf{S}} = \begin{cases} \mathbf{s}, & \text{if } \frac{p_{\mathbf{x}|\mathbf{S}}(\mathbf{x}|\mathbf{S}=\mathbf{s})}{p_{\mathbf{x}|\mathbf{S}}(\mathbf{x}|\mathbf{S}=\mathbf{0})} > \tau' \\ \mathbf{0}, & \text{if } \frac{p_{\mathbf{x}|\mathbf{S}}(\mathbf{x}|\mathbf{S}=\mathbf{s})}{p_{\mathbf{x}|\mathbf{S}}(\mathbf{x}|\mathbf{S}=\mathbf{0})} < \tau' \\ \text{arbitrary,} & \text{if } \frac{p_{\mathbf{x}|\mathbf{S}}(\mathbf{x}|\mathbf{S}=\mathbf{s})}{p_{\mathbf{x}|\mathbf{S}}(\mathbf{x}|\mathbf{S}=\mathbf{0})} = \tau' \end{cases},$$

where $\tau' \geq 0$.² From (4.2.3) this detector further reduces to

$$\hat{\mathbf{S}} = \begin{cases} \mathbf{s}, & \text{if } \mathbf{x}^\dagger \left[\mathbf{I}_{N \times N} - (\mathbf{I}_{N \times N} + (\mathbf{s}^\dagger \Phi_{\text{TS}}) \Phi_{\text{R}})^{-1} \right] \mathbf{x} > \tau \\ \mathbf{0}, & \text{if } \mathbf{x}^\dagger \left[\mathbf{I}_{N \times N} - (\mathbf{I}_{N \times N} + (\mathbf{s}^\dagger \Phi_{\text{TS}}) \Phi_{\text{R}})^{-1} \right] \mathbf{x} < \tau \\ \text{arbitrary,} & \text{if } \mathbf{x}^\dagger \left[\mathbf{I}_{N \times N} - (\mathbf{I}_{N \times N} + (\mathbf{s}^\dagger \Phi_{\text{TS}}) \Phi_{\text{R}})^{-1} \right] \mathbf{x} = \tau \end{cases},$$

where the threshold $\tau \geq 0$ is a system parameter subject to optimization in the sequel.

After some manipulations, it can be shown that the probability that a zero input is erroneously decided as non-zero is

$$\alpha = \Pr \left[\hat{\mathbf{S}} = \mathbf{s} | \mathbf{S} = \mathbf{0} \right] = \int_{\tau}^{\infty} p_{R_0}(r) dr, \quad (4.4.2)$$

where the non-negative random variable R_0 corresponds to the complex Gaussian quadratic form

$$R_0 = \mathbf{G}^\dagger \left[\mathbf{I}_{N \times N} - (\mathbf{I}_{N \times N} + (\mathbf{s}^\dagger \Phi_{\text{TS}}) \Phi_{\text{R}})^{-1} \right] \mathbf{G},$$

and $\mathbf{G} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N \times N})$. Similarly, the probability that a non-zero input is erroneously decided as zero is

$$\beta = \Pr \left[\hat{\mathbf{S}} = \mathbf{0} | \mathbf{S} = \mathbf{s} \right] = \int_0^{\tau} p_{R_1}(r) dr, \quad (4.4.3)$$

²For this problem randomization is not necessary. Hence for the sake of simplicity, we let the decision be “arbitrary” when the likelihood ratio equals τ' .

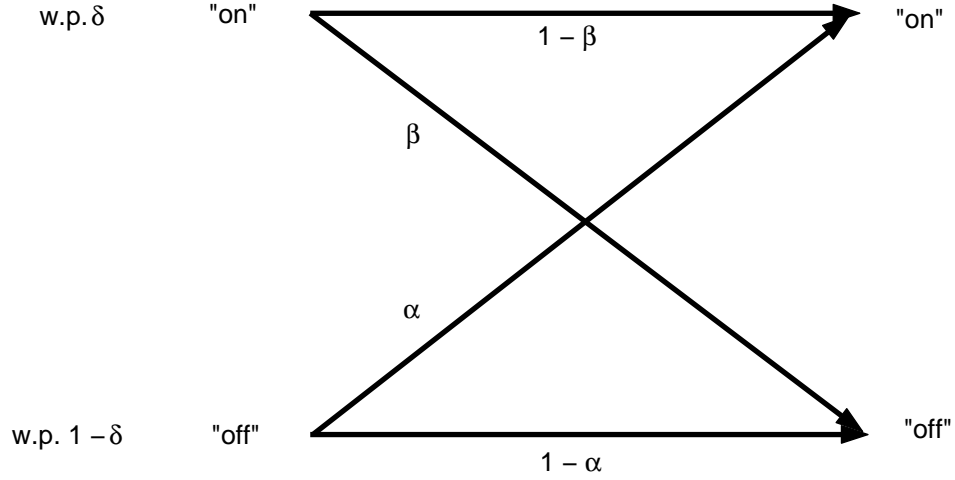


Figure 4.3. The binary discrete memoryless channel (DMC) resulting from symbol-by-symbol hard decisions of on-off keying (OOK) of the original fading channel.

where

$$R_{\mathbf{t}_1} = \mathbf{G}^\dagger (\mathbf{s}^\dagger \Phi_{\mathbf{T}\mathbf{S}}) \Phi_{\mathbf{R}} \mathbf{G},$$

and again, $\mathbf{G} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N \times N})$. To evaluate α and β , in Appendix .7 we present analytic expressions for the PDF and CDF of general complex Gaussian quadratic forms with a positive semi-definite kernel matrix.

From the developments above, symbol-by-symbol hard decisions convert the original channel into a binary DMC, with crossover probabilities α and β , as illustrated in Figure 4.3. The input distribution of this binary DMC is $(\delta, 1 - \delta)$, and the resulting channel mutual information can be obtained as

$$\begin{aligned} I(\rho, \delta, \tau) &= H_b((1 - \delta) \cdot \alpha + \delta \cdot (1 - \beta)) \\ &\quad - (1 - \delta) \cdot H_b(\alpha) - \delta \cdot H_b(\beta), \end{aligned} \quad (4.4.4)$$

where $H_b(p) = -p \log p - (1 - p) \log(1 - p)$ is the binary entropy function.

In the low-SNR regime, we are primarily interested in the maximum normalized mutual information, *i.e.*, the channel mutual information per unit average SNR,

optimized over the “on” probability $0 < \delta < 1$ and the decision threshold $\tau \geq 0$. This can be written as

$$I_{\text{norm}}(\rho) = \max_{\delta, \tau} \frac{I(\rho, \delta, \tau)}{\rho}. \quad (4.4.5)$$

4.4.1 Scalar Channel

For the scalar channel case with $M = N = 1$ and $\Phi_{\text{T}} = \Phi_{\text{R}} = 1$, R_0 and R_1 reduce to

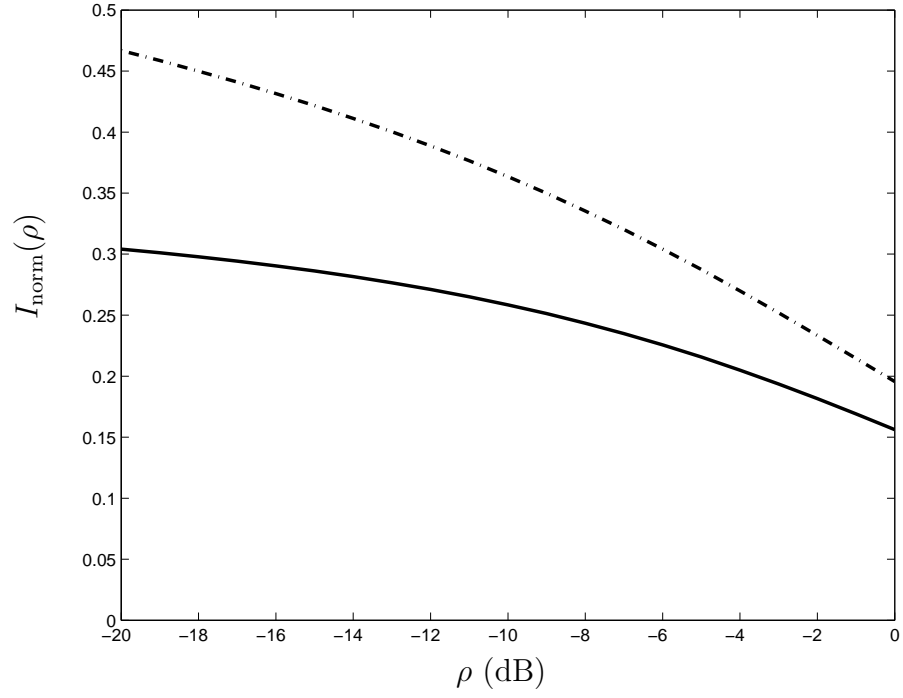
$$\begin{aligned} R_0 &= \frac{s^\dagger s}{s^\dagger s + 1} G^\dagger G \\ R_1 &= (s^\dagger s) G^\dagger G, \end{aligned}$$

where $G \sim \mathcal{CN}(0, 1)$. That is, R_0 and R_1 are exponentially distributed, and consequently α and β can be evaluated as

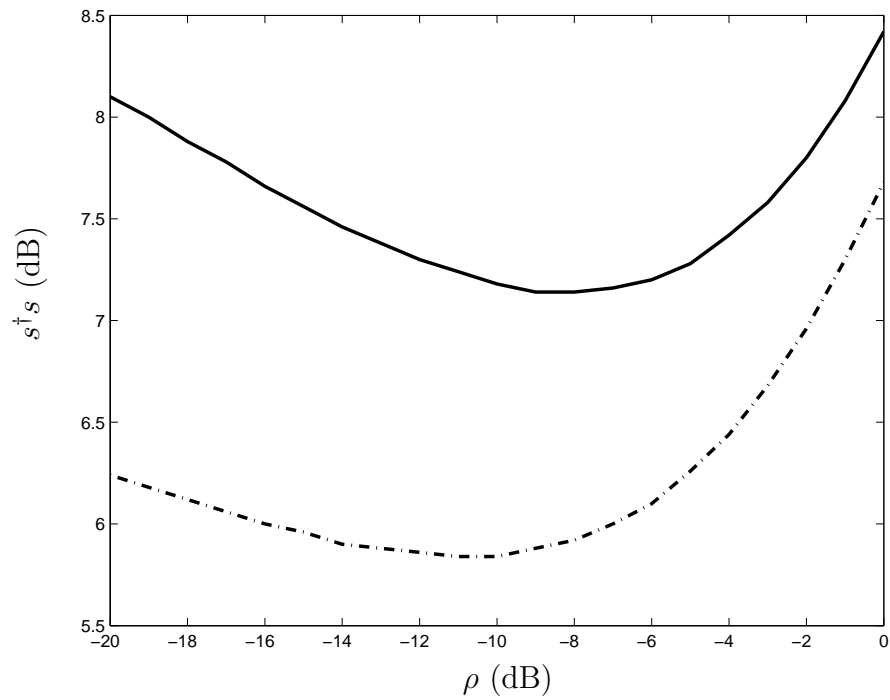
$$\begin{aligned} \alpha &= \exp\left(-\frac{s^\dagger s + 1}{s^\dagger s} \tau\right) \\ \beta &= 1 - \exp\left(-\frac{1}{s^\dagger s} \tau\right). \end{aligned}$$

The maximum normalized mutual information $I_{\text{norm}}(\rho)$ is then evaluated through a numerical search over δ and τ . Figure 4.4 displays $I_{\text{norm}}(\rho)$ and the corresponding optimal peak SNR $s^\dagger s$, for OOK inputs with hard decision (in solid curves) as well as soft decision (in dashed-dot curves). For the SNR range plotted, the soft-decision curve corresponds to the channel capacity [1]. Compared to the soft-decision curves, we see that hard-decision OOK results in a rate loss and a penalty in the peak SNR. The rate loss is due to the loss of information in the hard decision demodulation; and the penalty in the peak SNR can be intuitively understood as hard-decision OOK requiring more discrimination between the “on” and “off” inputs.

We also study the behavior of the decision threshold τ . Figure 4.5 displays the optimal τ vs. the average SNR ρ . As a comparison it also displays the threshold



(a) $I_{\text{norm}}(\rho)$ vs. ρ



(b) Optimal peak SNR $s^\dagger s$ vs. ρ

Figure 4.4. The maximum normalized mutual information and the corresponding optimal peak SNR for the scalar channel case. The curves for OOK with hard decisions is in solid, and those for OOK with soft decisions is in dashed-dot, as a comparison.

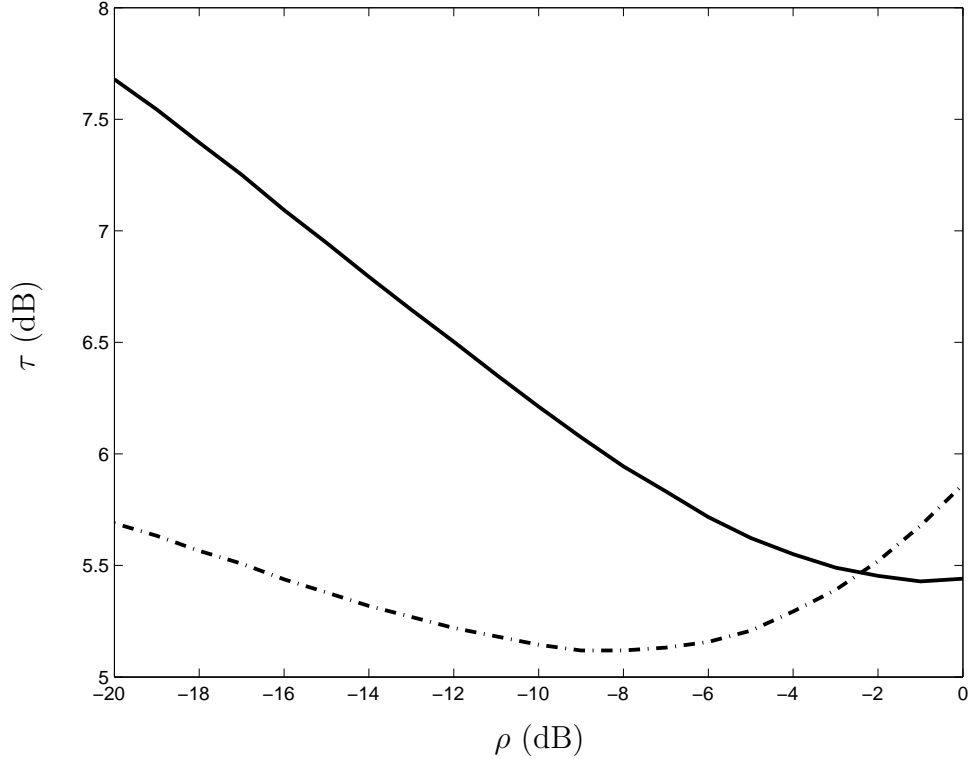


Figure 4.5. The optimal decision threshold τ vs. ρ for the scalar channel case. The dashed-dot curve is the threshold for the MAP decision rule.

corresponding to the MAP decision.³ We observe that the optimal threshold τ maximizing the normalized mutual information is indeed different from the MAP decision threshold. As SNR vanishes the optimal τ grows much faster than the MAP decision threshold.

4.4.2 Multiple Transmit Antennas and Transmit Correlation

Correlated transmit antennas lead to a relatively simple scaling effect in $I_{\text{norm}}(\rho)$, which corresponds to the asymptotic analyses in Section 4.3. From the definitions of α , β , and $I(\rho, \delta, \tau)$, it is clear that an “on” signal \mathbf{s} for the multiple-transmit channel is equivalent to $\sqrt{\mathbf{s}^\dagger \Phi_T \mathbf{s}}$ for a single-transmit channel, in the sense that both lead to

³In the MAP decision we take the a priori probability δ as the one that maximizes $I_{\text{norm}}(\rho)$.

the same $I(\rho, \delta, \tau)$. Consequently, we find that the resulting $I_{\text{norm}}(\rho)$ is $\frac{s^\dagger \Phi_{\text{TS}}}{s^\dagger s}$ times that of the single-transmit channel evaluated at the average SNR $\frac{s^\dagger \Phi_{\text{TS}}}{s^\dagger s} \rho$. That is, both the multiplicative rate gain as well as the peak SNR reduction are achieved, compared to the single-transmit channel case. Figure 4.6 displays $I_{\text{norm}}(\rho)$ and the corresponding optimal peak SNR $s^\dagger s$ for the fully correlated transmit antennas with different M . Both the multiplicative rate gain and the peak SNR reduction are clearly exhibited.

4.4.3 Multiple Receive Antennas and Receive Correlation

Depending upon the receive correlation matrix Φ_{R} , multiple receive antennas lead to rather diverse behavior of $I_{\text{norm}}(\rho)$. Without loss of generality (see 4.4.2) we let the number of transmit antennas $M = 1$. When the N receive antennas are uncorrelated, *i.e.*, $\Phi_{\text{R}} = \mathbf{I}_{N \times N}$, R_0 and R_1 are both χ^2 random variables with $2N$ degrees of freedom, *i.e.*,

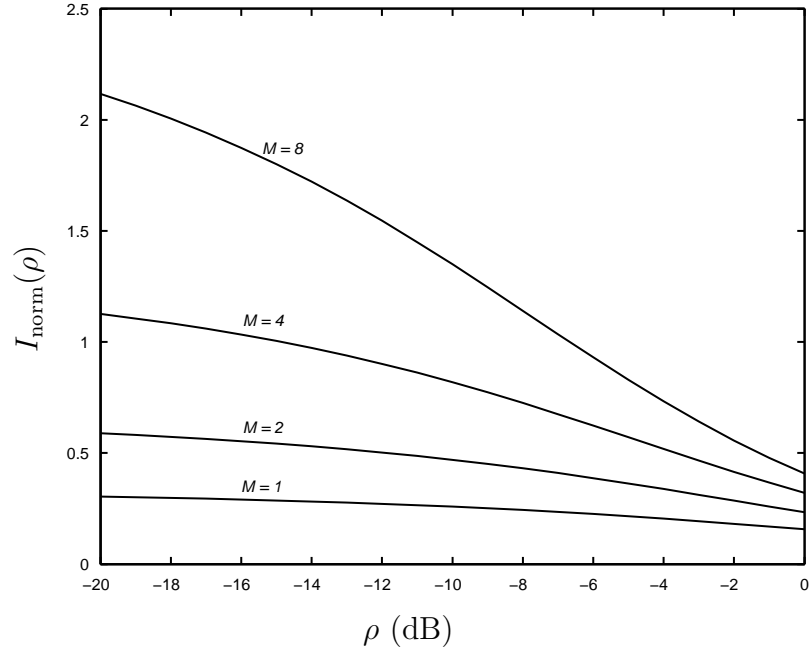
$$p_{R_i}(r) = \frac{1}{(\sigma_i^2)^N (N-1)!} r^{N-1} e^{-\frac{r}{\sigma_i^2}}, \quad i = 0, 1, \text{ for } r \geq 0,$$

where $\sigma_0^2 = \frac{s^\dagger s}{s^\dagger s + 1}$ and $\sigma_1^2 = s^\dagger s$, respectively. Consequently α and β can be evaluated as

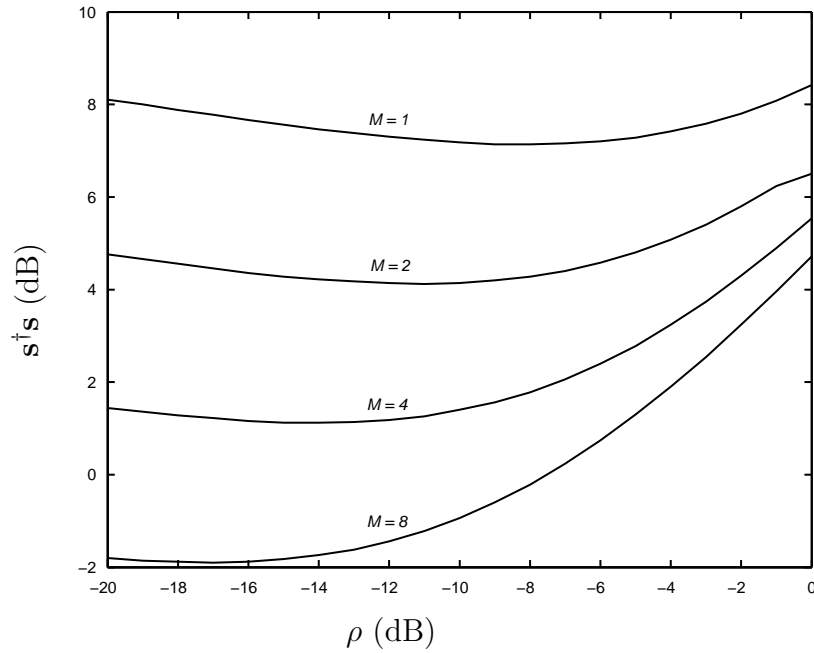
$$\begin{aligned} \alpha &= \exp\left(-\frac{s^\dagger s + 1}{s^\dagger s} \tau\right) \sum_{k=0}^{N-1} \frac{1}{k!} \left(\frac{s^\dagger s + 1}{s^\dagger s} \tau\right)^k \\ \beta &= 1 - \exp\left(-\frac{1}{s^\dagger s} \tau\right) \sum_{k=0}^{N-1} \frac{1}{k!} \left(\frac{\tau}{s^\dagger s}\right)^k. \end{aligned}$$

On the other hand, when the N receive antennas are fully correlated, *i.e.*, $\Phi_{\text{R}} = \mathbf{E}_{N \times N}$, R_0 and R_1 reduce to exponential random variables, and consequently α and β can be evaluated as

$$\begin{aligned} \alpha &= \exp\left(-\frac{Ns^\dagger s + 1}{Ns^\dagger s} \tau\right) \\ \beta &= 1 - \exp\left(-\frac{1}{Ns^\dagger s} \tau\right). \end{aligned}$$



(a) $I_{\text{norm}}(\rho)$ vs. ρ



(b) Optimal peak SNR $s^\dagger s$ vs. ρ

Figure 4.6. The maximum normalized mutual information and the corresponding optimal peak SNR for M spatially fully correlated transmit antennas and a single receive antenna. The curve for $M = 1$ also corresponds to that for $M > 1$ uncorrelated transmit antennas.

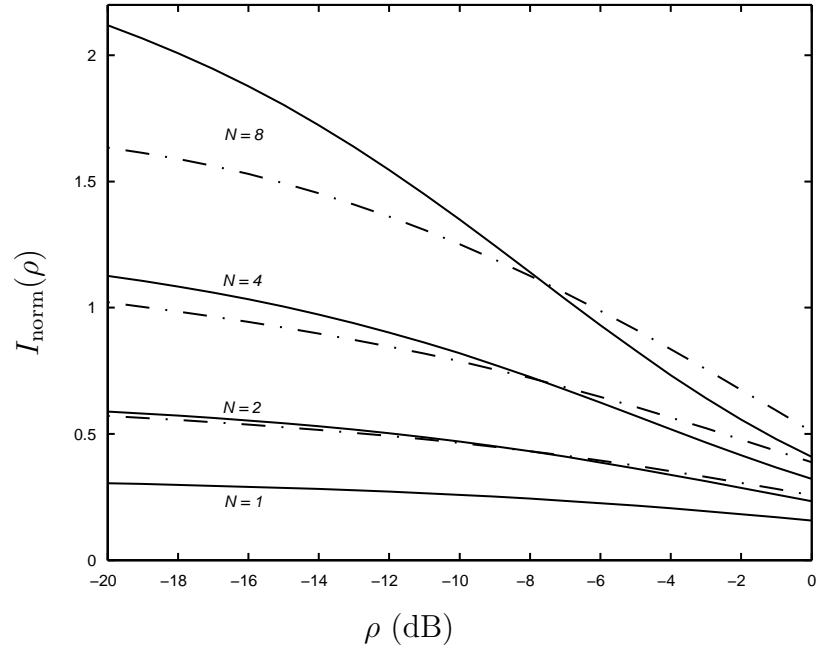
Compared to the scalar channel case, it is clear that fully correlated receive antennas effectively amplify the peak SNR (as well as the average SNR) by a factor of N . Figure 4.7 compares $I_{\text{norm}}(\rho)$ and the corresponding optimal peak SNR s^\dagger s for the uncorrelated and the fully correlated receive antennas. For both cases, we observe that increasing the number of receive antennas dramatically increases $I_{\text{norm}}(\rho)$ as well as decreases the required peak SNR. Compared to the uncorrelated case, fully correlated receive antennas appear to be more beneficial. This is consistent with the asymptotic analyses in Section 4.3.

When multiple receive antennas exhibit general correlation, $I_{\text{norm}}(\rho)$ and the corresponding optimal peak SNR can be similarly evaluated, with the aid of Appendix .7. Figure 4.8 displays the curves for $N = 2$. Due to non-asymptotic effects, for the SNR range plotted, $I_{\text{norm}}(\rho)$ for moderate channel correlation (*e.g.*, $v = 0.5$ and 0.9) is slightly inferior to that for the uncorrelated case (*i.e.*, $v = 0$). Whereas it can also be seen that the curve for $v = 0.9$ has a growth trend sharper than that for $v = 0$, and in fact it will eventually outperform for certain lower SNR not plotted in Figure 4.8(a). On the other hand, channel correlation does lead to noticeable reduction in the required peak SNR, as shown in Figure 4.8(b).

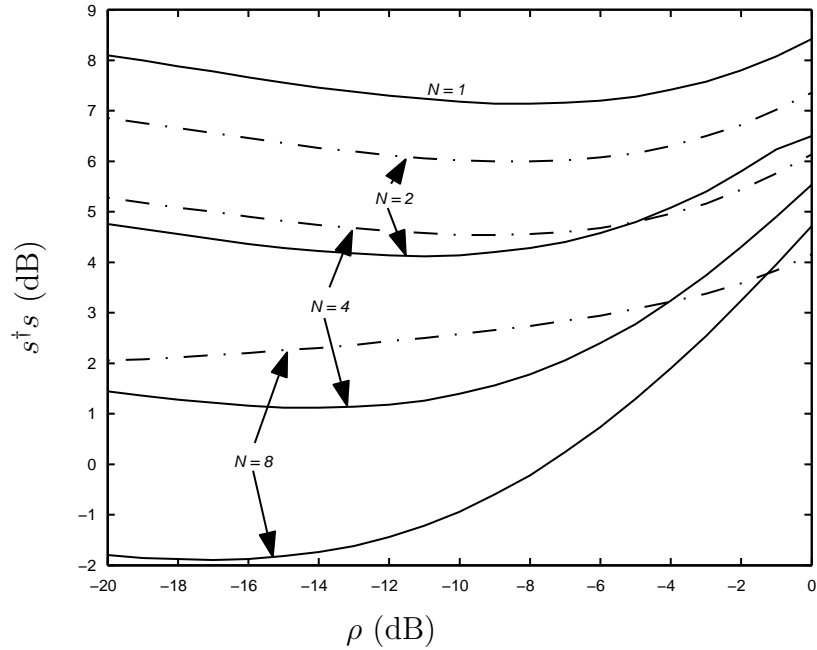
Remarks:

(a) In the preceding analysis we treat the transmit correlation and the receive correlation separately. It can be easily shown that they can be simultaneously utilized by employing a multiple-input-multiple-output (MIMO) architecture.

(b) In the preceding analysis we compare the performance for a fixed number of antennas. For practical applications a more reasonable comparison may be based upon a fixed space constraint [57]. Given a certain physical space to locate antennas, a system architect can choose to have fewer antennas with weaker spatial correlation, or choose to have more antennas with stronger spatial correlation. The results in

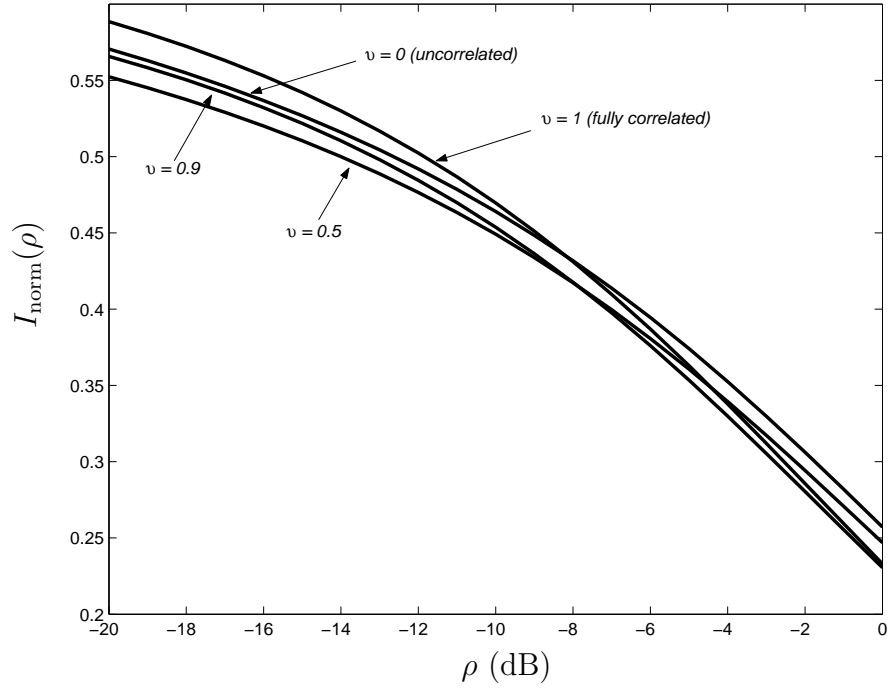


(a) $I_{\text{norm}}(\rho)$ vs. ρ

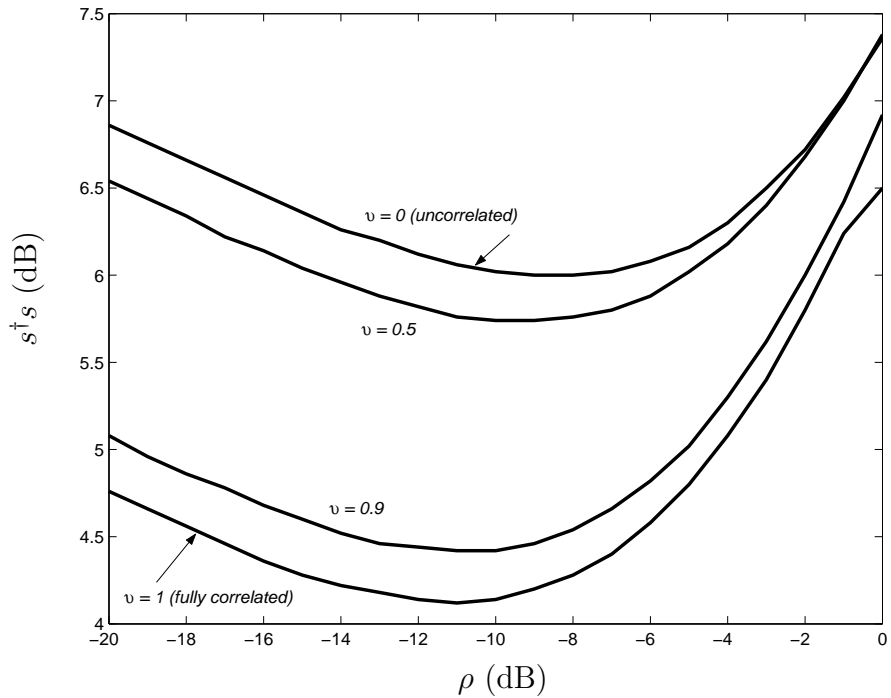


(b) Optimal peak SNR $s^\dagger s$ vs. ρ

Figure 4.7. The maximum normalized mutual information and the corresponding optimal peak SNR for a single transmit antenna and N receive antennas. The solid curves correspond to the spatially fully correlated case, and the dashed-dot curves correspond to the spatially uncorrelated case.



(a) $I_{\text{norm}}(\rho)$ vs. ρ



(b) Optimal peak SNR $s^\dagger s$ vs. ρ

Figure 4.8. The maximum normalized mutual information and the corresponding optimal peak SNR for a single transmit antenna and $N = 2$ receive antennas. The correlation coefficient is $\nu = 0$ (uncorrelated), 0.5, 0.9, and 1 (fully correlated).

this chapter suggest that for non-coherent fading channels at low SNR, the later choice may be more favorable.⁴

(c) Due to the inherent information loss in symbol-by-symbol hard decisions, even as the average SNR $\rho \rightarrow 0$, the maximum normalized mutual information $I_{\text{norm}}(\rho)$ in (4.4.5) cannot achieve $\dot{I}(0)$ and \dot{C} derived in Section 4.3. The quantity, $\lim_{\rho \rightarrow 0} I_{\text{norm}}(\rho)$, can rarely be determined analytically, and usually requires careful numerical study. Further elaboration on this issue is beyond our scope.

(d) It can be numerically shown that δ and τ maximizing $I_{\text{norm}}(\rho)$ typically lead to $\alpha \approx 0$ in the resulting DMC in Figure 4.3. However a direct approximation using the “Z”-channel model (*e.g.*, [46]) is generally not accurate. This is because the input distribution is highly skewed, *i.e.*, the “on” probability $\delta \ll 1$. Thus, the joint probabilities $\Pr[\mathbf{S} = \mathbf{0}, \hat{\mathbf{S}} = \mathbf{s}] = (1 - \delta)\alpha$ and $\Pr[\mathbf{S} = \mathbf{s}, \hat{\mathbf{S}} = \mathbf{0}] = \delta\beta$ are typically comparable and neither can be negligible. Such skewed inputs also pose challenges to practical codes design, which is a potential research topic for future investigation.

4.5 Conclusion

Challenges arise when applications call for efficient communication in the low-SNR regime with fast time-varying fading processes. It is often the case that channel state information becomes a luxury, and communication is required to take place in a non-coherent manner. For such scenarios, we investigate in this chapter the benefits of spatially correlated multiple antennas. The main theme is that spatial correlation along with multiple antennas effectively boosts the powers in the channel as well as at the receiver, without spending additional transmit power or excess channel bandwidth. At high SNR this effect is only marginal, and is usually ignored due to

⁴This observation relies on the channel model we adopt in the chapter. For very densely packed antenna array, complicated near-field electromagnetic effects may become significant, and the channel model in Section 4.2 will eventually collapse. Thus it remains an important problem to characterize the ultimate benefits of packing antennas in such regimes.

its more significant side effect — the simultaneous loss in degrees of freedom. At low SNR, however, it turns out to be a key to improving communication efficiency.

In addition to the rate gains and peak power reductions revealed in this chapter, other considerations make spatial correlation appealing in practice. From a system architect’s perspective, real-world wireless channels always exhibit correlation. For many applications the spatial dimensions are rather limited, and in that scenario channel correlation should be viewed as a potential benefit, rather than a hindrance. From a communication engineer’s perspective, in the proposed scheme employing OOK with symbol-by-symbol hard decisions, the transmitter does not require sophisticated space-time codes. Instead, a common “on” or “off” signal is multiplexed onto all the transmit antennas, perhaps with appropriate “beamforming” to maximize the effective SNR if the transmit antennas exhibit general spatial correlation. Similarly, the receiver simply performs non-coherent energy combining followed by threshold-based hard decisions. These characteristics suggest that schemes exploiting spatially correlated antennas are particularly suitable to non-coherent low-SNR communication systems requiring low-cost and low-complexity solutions.

CHAPTER 5

CONCLUSION

In this thesis we investigate the role of correlation in wireless communication systems with no channel state information. We examine three distinct scenarios, obtaining a series of results and identifying a set of open problems for future study, as summarized below.

For discrete-time memoryless Rayleigh fading channels, we establish an induced additive-noise channel model so that the problem reduces to communication over a linear channel with non-Gaussian additive noise. We obtain simple necessary and sufficient conditions for good and asymptotically optimal performance at high SNR, and construct continuous and discrete input signaling schemes to achieve the performance. Our approach is easily extended to MIMO channels. We also demonstrate that the approach leads to low-complexity and near-optimal practical transceivers.

An open problem, as already pointed in Section 2.6, is extending the induced additive-noise channel modeling approach to more general channels, such as MIMO block fading channels. On one hand, the approach we take in Chapter 2 appears like a rather *ad hoc* technique yielding elegant analytical results only for memoryless Rayleigh fading. On the other hand, the strong similarity with cepstrum techniques and homomorphic analysis of our approach definitely motivates us to extend it in some more general way.

For temporally correlated fading channels, we obtain a convenient lower bound of the achievable information rate for PSK inputs, and give a recursive training scheme interpretation for that lower bound. We then focus on the low-SNR performance, pointing out that using PSK is usually near-optimal compared to other peak-limited inputs. Furthermore, we show that by widening the input bandwidth, using PSK actually achieves the wideband channel capacity under a peak power constraint. Parallel results are also established for Rician fading channels.

An open problem is fully characterizing the behavior of the recursive training scheme. In Section 3.3.2 we establish the achievability in a coding theoretic sense for channels with finite memory. For general fading processes with infinite memory, technical difficulties prevent us from asserting that the recursive training scheme achieves the desired information rate, and additional work is needed.

For spatially correlated fading channels, we demonstrate the potential benefits of spatial correlation by focusing on the Kronecker correlation model. We first carry out asymptotic analyses, evaluating the first and second derivatives of the channel mutual information for OOK at vanishing SNR. Then we examine the behavior of OOK with symbol-by-symbol hard decisions. Both approaches clearly indicate that for low-SNR communication systems, employing MIMO architecture and increasing the spatial correlation among antennas lead to sizeable performance gains compared to single-antenna systems, in both rate gains and peak power reductions.

Two open problems are identified in Chapter 4. One is to construct codes for realizing the benefits revealed by the information-theoretic approach. The challenge is that the channel is a highly asymmetric binary input channel, with both skewed transition probabilities as well as skewed input probabilities, thus is rather incompatible with conventional linear codes. The other open problem is to refine the channel modeling. In Chapter 4 we employ a Kronecker correlation model to

characterize the channel spatial correlation. It is important to examine more general correlation models, in order to validate the applicability of the benefits for practical systems. Furthermore, as the placement of antennas becomes more dense, the discrete model for channel spatial correlation will collapse eventually, therefore it is of interest to investigate the asymptotic behavior of spatial correlation as the antennas are placed in a continuous way and their size vanishes, for which the channel model is more sophisticated and unknown yet.

APPENDIX

.1 A Capacity Upper Bound Revisited

In this section, we obtain a capacity upper bound from [70] using the additive-noise perspective.

In order to apply standard maximum entropy techniques [10], a natural way is to replace the constraint $T = U + W$ by taking moments at both sides. This type of relaxation will yield upper bounds to the channel capacity. In this section let us consider taking their expectations, as

$$\mathcal{E}[T] = \mathcal{E}[U] + \mathcal{E}[W] = \lambda \geq -\frac{\gamma}{2}.$$

Here the last inequality is from $U \in [0, \infty)$. On the other hand, the channel output should have a constrained average power as given by

$$\mathcal{E}[\exp(2T)] = \mathcal{E}[|X|^2] = \rho + 1.$$

Thus following maximum entropy techniques, the resulting entropy-maximizing distribution of T should have PDF

$$f_T^*(t) = 2 \left(\frac{\mu}{\rho + 1} \right)^\mu \frac{1}{\Gamma(\mu)} \exp \left[2\mu t - \frac{\mu}{\rho + 1} \exp(2t) \right],$$

where $\mu > 0$ is determined by the equation

$$\psi(\mu) - \log \mu = 2\lambda - \log(\rho + 1),$$

and $\psi(\mu) = \frac{d}{d\mu} \log \Gamma(\mu)$ is the psi function [18]. Thus, we can treat $f_T^*(t)$ as the actual output PDF to obtain an upper bound to the channel capacity as

$$\begin{aligned} C = \sup_{f_T(t)} I(U; T) &< h(f_T^*(t)) - h(W) \\ &= \log \Gamma(\mu) - \mu\psi(\mu) + \mu - \gamma - 1, \end{aligned} \quad (.1.1)$$

where

$$\log \mu - \psi(\mu) = \log(\rho + 1) - 2\lambda. \quad (.1.2)$$

To tighten the upper bound, we first notice that the right hand side of (.1.1) is increasing with $\mu > 0$, hence the tightest upper bound is obtained when μ is minimized. We then notice that the left hand side of (.1.2) is decreasing with $\mu > 0$, so that the minimum allowed μ is attained when $\lambda = -\gamma/2$. This capacity upper bound is exactly the same as that obtained in [70] using variational methods. Here we revisit it from the induced additive-noise channel model. We also note that by introducing more moment constraints, it is possible to obtain tighter capacity upper bounds. However, their derivation and evaluation will become much less tractable.

As a side note, a tighter capacity upper bound is obtained in [36], based upon a dual expression of channel capacity. Such result can also be reproduced by optimizing the output PDF of the induced additive-noise channel, but this approach does not lead to simplification compared to the original one.

.2 Proof of Proposition 2.3.7

By the definition of cutoff rate (*e.g.*, [60]) we have

$$R_0 = -\log \int_{-\infty}^{\infty} \left[\sum_{l=0}^{L-1} \frac{1}{L} \sqrt{f_W(t - l\Delta)} \right]^2 dt.$$

The inner integral can be evaluated as

$$\begin{aligned}
& \int_{-\infty}^{\infty} \left[\sum_{l=0}^{L-1} \frac{1}{L} \sqrt{f_W(t-l\Delta)} \right]^2 dt \\
&= \int_{-\infty}^{\infty} \frac{1}{L^2} \left\{ \sum_{l=0}^{L-1} f_W(t-l\Delta) \right. \\
&\quad \left. + 2 \sum_{l=0}^{L-1} \sum_{m=0}^{l-1} \sqrt{f_W(t-l\Delta) f_W(t-m\Delta)} \right\} dt \\
&= \frac{1}{L^2} \left\{ L + 2 \sum_{l=0}^{L-1} \sum_{m=0}^{l-1} \int_{-\infty}^{\infty} 2 \exp \left[2t - (l+m)\Delta \right. \right. \\
&\quad \left. \left. - \frac{1}{2} \exp(2t) (\exp(-2l\Delta) + \exp(-2m\Delta)) \right] dt \right\} \\
&= \frac{1}{L} \left\{ 1 + \frac{8}{L} \sum_{l=0}^{L-1} \sum_{m=0}^{l-1} \int_0^{\infty} x \exp[-(l+m)\Delta \right. \\
&\quad \left. - x(\exp(-2l\Delta) + \exp(-2m\Delta))] \frac{1}{2x} dx \right\} \\
&= \frac{1}{L} \left\{ 1 + \frac{4}{L} \sum_{l=0}^{L-1} \sum_{m=0}^{l-1} \frac{\exp(-(l+m)\Delta)}{\exp(-2l\Delta) + \exp(-2m\Delta)} \right\}.
\end{aligned}$$

The cutoff rate R_0 then follows immediately. **Q.E.D.**

.3 Proof of Lemma 2.4.1

1. CDF:

$$\begin{aligned}
& F_W(w) \\
&= \int_{-\infty}^w \frac{2}{(N-1)!} e^{2Nw_1} e^{-e^{2w_1}} dw_1 \\
&= \frac{2}{(N-1)!} \int_0^{e^{2w}} w_2^N e^{-w_2} \frac{1}{2w_2} dw_2 \\
&= \frac{1}{(N-1)!} \left[(N-1)! - e^{-e^{2w}} \sum_{n=0}^{N-1} \frac{(N-1)!}{n!} e^{2nw} \right] \\
&= 1 - e^{-e^{2w}} \sum_{n=0}^{N-1} \frac{e^{2nw}}{n!},
\end{aligned}$$

where we use $\int_0^u x^n e^{-\mu x} dx = \frac{n!}{\mu^{n+1}} - e^{-u\mu} \sum_{k=0}^n \frac{n!}{k!} \frac{u^k}{\mu^{n-k+1}}$, for $u > 0$, $\Re\mu > 0$ [18, 3.351].

2. Characteristic function:

$$\begin{aligned}
& \varphi_W(j\theta) \\
&= \int_{-\infty}^{\infty} \exp(j\theta w) \frac{2}{(N-1)!} e^{2Nw} e^{-e^{2w}} dw \\
&= \frac{1}{(N-1)!} \int_0^{\infty} w_1^{N-1+j\theta/2} \exp(-w_1) dw_1 \\
&= \frac{1}{(N-1)!} \Gamma(N + j\frac{\theta}{2}),
\end{aligned}$$

where we use $\int_0^{\infty} x^{\nu-1} \exp(-\mu x) dx = \frac{1}{\mu^\nu} \Gamma(\nu)$ for $\Re\mu > 0$, $\Re\nu > 0$ [18, 3.381].

3. Mean:

$$\begin{aligned}
\mathcal{E}[W] &= (-j) \left. \frac{d\varphi_W(j\theta)}{d\theta} \right|_{\theta=0} \\
&= \frac{1}{2(N-1)!} \int_0^{\infty} w^{N-1} \exp(-w) \log w dw \\
&= \frac{1}{2(N-1)!} \Gamma(N) \psi(N) = \frac{\psi(N)}{2},
\end{aligned}$$

where we use $\int_0^{\infty} x^{\nu-1} e^{-\mu x} \log x dx = \frac{1}{\mu^\nu} \Gamma(\nu) [\psi(\nu) - \log \mu]$ for $\Re\mu > 0$, $\Re\nu > 0$ [18, 4.352].

The maximization of $f_W(w)$ at $w = \log N/2$ is easily verified by taking derivative.

4. Variance:

$$\begin{aligned}
& \sigma_W^2 \\
&= \mathcal{E}[W^2] - \mathcal{E}[W]^2 \\
&= - \left. \frac{d^2\varphi_W(j\theta)}{d\theta^2} \right|_{\theta=0} - \frac{\psi(N)^2}{4} \\
&= \frac{1}{4(N-1)^2} \Gamma(N) [\psi(N)^2 + \zeta(2, N)] - \frac{\psi(N)^2}{4} \\
&= \frac{\zeta(2, N)}{4},
\end{aligned}$$

where we use $\int_0^{\infty} x^{\nu-1} e^{-\mu x} (\log x)^2 dx = \frac{\Gamma(\nu)}{\mu^\nu} \{[\psi(\nu) - \log \mu]^2 + \zeta(2, \nu)\}$, for $\Re\mu > 0$, $\Re\nu > 0$ [18, 4.358].

5. Differential entropy:

$$\begin{aligned}
& h(W) \\
&= - \int_{-\infty}^{\infty} \frac{2}{(N-1)!} \exp(2Nw - \exp(2w)) \\
&\quad \left[\log \frac{2}{(N-1)!} + 2Nw - \exp(2w) \right] dw \\
&= \sum_{n=1}^{N-1} \log n - \log 2 - N \cdot \psi(N) \\
&\quad + \frac{1}{(N-1)!} \int_0^{\infty} w^N \exp(-w) dw \\
&= \sum_{n=1}^{N-1} \log n - \log 2 + N[1 - \psi(N)].
\end{aligned}$$

Q.E.D.

.4 Capacity per Unit Energy Under Peak Constraints for Rician Fading Channels

We evaluate the capacity per unit energy of Rician fading channels with a peak constraint only. To this end, let us slightly modify the parametrization of channel model (3.6.2) to

$$X[k] = \sqrt{\frac{d_d}{d_d+1}} S[k] + \sqrt{\frac{1}{d_d+1}} H_d[k] S[k] + Z[k], \quad (.4.1)$$

where all the symbols represent the same quantities as in (3.6.2), except that the channel input $S[\cdot]$ is now subject to a peak constraint $S[\cdot]S^\dagger[\cdot] \leq \rho_p$.

The capacity per unit energy for channel (.4.1) is given by the following Proposition.

Proposition .4.1 *The capacity per unit energy of the discrete-time Rician fading channel (.4.1), under the peak constraint*

$$S[k]S^\dagger[k] \leq \rho_p, \quad \text{for } k = 0, \pm 1, \dots$$

is

$$\dot{C} = 1 - \frac{1}{\rho_p} \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left(1 + \frac{\rho_p}{d_d+1} S_{H_d}(e^{j\Omega}) \right) d\Omega. \quad (.4.2)$$

.4.1 Proof of Proposition .4.1

The proof is a direct application of the general formula in [64], and its main steps are indeed the same as that for the Rayleigh fading case in [64].

In [64, Proposition 2.2, (15)] the capacity per unit energy for general fading channels under a peak constraint is established. For channel (.4.1), it is specialized to

$$\dot{C} = \lim_{n \rightarrow \infty} \sup_{\substack{\mathbf{s}_0: s_0[i] s_0^\dagger[i] \leq \rho_P \\ i=1, \dots, n}} \frac{\mathcal{D}(p_{\mathbf{X}|\mathbf{S}}(\mathbf{x}|\mathbf{s}_0) \| p_{\mathbf{X}|\mathbf{S}}(\mathbf{x}|\mathbf{0}))}{\|\mathbf{s}_0\|^2}, \quad (.4.3)$$

where $\mathbf{S} = [S_1, \dots, S_n]^T$ and $\mathbf{X} = [X_1, \dots, X_n]^T$ are the vectorized inputs and outputs over n channel uses, respectively.

For $\mathbf{S} = \mathbf{0}$, the conditional output distribution is circular complex Gaussian, with zero mean vector and identity variance matrix, *i.e.*, $\mathbf{X} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{n \times n})$. For $\mathbf{S} = \mathbf{s}_0$, the conditional output distribution is also circular complex Gaussian, with mean vector

$$\mathbf{m}_0 = \sqrt{\frac{d_d}{d_d + 1}} \mathbf{s}_0,$$

and covariance matrix

$$\Sigma_0 = \mathbf{I}_{n \times n} + \frac{1}{d_d + 1} \mathbf{\Gamma},$$

where

$$\mathbf{\Gamma} = \text{diag}\{s_0[1], \dots, s_0[n]\} \cdot \Sigma_{H_d} \cdot \text{diag}\{s_0^\dagger[1], \dots, s_0^\dagger[n]\},$$

and the (i, j) -th element of Σ_{H_d} is $K_{H_d}(|i - j|)$, where $K_{H_d}(\cdot)$ is the autocorrelation function of the fading process $\{H_d[k] : -\infty < k < \infty\}$.

Thus, the KL distance in (.4.3) becomes [75, (59)]

$$\begin{aligned}
& \mathcal{D} (p_{\mathbf{X}|\mathbf{S}}(\mathbf{x}|\mathbf{s}_0) \| p_{\mathbf{X}|\mathbf{S}}(\mathbf{x}|\mathbf{0})) \\
&= \frac{d_d}{d_d + 1} \mathbf{s}_0^\dagger \mathbf{s}_0 + \text{trace} \left(\frac{1}{d_d + 1} \mathbf{\Gamma} \right) - \log \det \left(\mathbf{I}_{n \times n} + \frac{1}{d_d + 1} \mathbf{\Gamma} \right) \\
&= \frac{d_d}{d_d + 1} \|\mathbf{s}_0\|^2 + \frac{1}{d_d + 1} \|\mathbf{s}_0\|^2 - \log \det \left(\mathbf{I}_{n \times n} + \frac{1}{d_d + 1} \mathbf{\Gamma} \right) \\
&= \|\mathbf{s}_0\|^2 - \log \det \left(\mathbf{I}_{n \times n} + \frac{1}{d_d + 1} \mathbf{\Sigma}_{H_d} \text{diag} \{ |s_0[1]|^2, \dots, |s_0[n]|^2 \} \right). \quad (.4.4)
\end{aligned}$$

That is,

$$\begin{aligned}
& \frac{\mathcal{D} (p_{\mathbf{X}|\mathbf{S}}(\mathbf{x}|\mathbf{s}_0) \| p_{\mathbf{X}|\mathbf{S}}(\mathbf{x}|\mathbf{0}))}{\|\mathbf{s}_0\|^2} \\
&= 1 - \frac{1}{\|\mathbf{s}_0\|^2} \log \det \left(\mathbf{I}_{n \times n} + \frac{1}{d_d + 1} \mathbf{\Sigma}_{H_d} \text{diag} \{ |s_0[1]|^2, \dots, |s_0[n]|^2 \} \right), \quad (.4.5)
\end{aligned}$$

where the second term for any fixed n has been shown to be minimized by letting $|s_0[i]|^2 = \rho_p$, $i = 1, \dots, n$ [64]. Hence

$$\sup_{\substack{\mathbf{s}_0: s_0[i] s_0^\dagger[i] \leq \rho_p \\ i=1, \dots, n}} \frac{\mathcal{D} (p_{\mathbf{X}|\mathbf{S}}(\mathbf{x}|\mathbf{s}_0) \| p_{\mathbf{X}|\mathbf{S}}(\mathbf{x}|\mathbf{0}))}{\|\mathbf{s}_0\|^2} = 1 - \frac{1}{n \rho_p} \log \det \left(\mathbf{I}_{n \times n} + \frac{\rho_p}{d_d + 1} \mathbf{\Sigma}_{H_d} \right) \quad (.4.6)$$

Finally the expression (.4.2) follows from applying the asymptotic theory of Toeplitz matrices (*e.g.*, [19]) to (.4.6) as $n \rightarrow \infty$.

.4.2 Interpretation of Proposition .4.1 and a Related Capacity Upper Bound

For Rayleigh fading channels ($d = 0$), (.4.2) reduces to

$$\dot{C} = 1 - \frac{1}{\rho_p} \frac{1}{2\pi} \int_{-\pi}^{\pi} \log (1 + \rho_p S_{H_d}(e^{j\Omega})) d\Omega, \quad (.4.7)$$

which has been obtained in [64].

Additional insight follows from considering the small-peak asymptotics of (.4.2).

A Taylor series expansion shows that

$$\dot{C} = \frac{d_d}{d_d + 1} + \frac{\rho_p}{2(d_d + 1)^2} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega \right] + o(\rho_p), \quad \text{as } \rho_p \rightarrow 0. \quad (.4.8)$$

In fact, for all $\rho_p \geq 0$ we have

$$\dot{C} \leq \frac{d_d}{d_d + 1} + \frac{\rho_p}{2(d_d + 1)^2} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega \right], \quad (.4.9)$$

which becomes tight for small ρ_p as shown in (.4.8).

Now let us consider that the channel (.4.1) is further subject to an average SNR constraint

$$\frac{1}{n} \mathcal{E} \left[\sum_{i=1}^n S[i] S^\dagger[i] \right] \leq \rho$$

for any given coding block length $n = 1, \dots$. For any fixed peak SNR ρ_p , the channel capacity C is a concave non-decreasing function of ρ [64] [74], hence an upper bound to C can be obtained based upon the capacity per unit energy \dot{C} as

$$C \leq \dot{C} \rho \leq \frac{d_d}{d_d + 1} \rho + \frac{\rho_p \rho}{2(d_d + 1)^2} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega \right]. \quad (.4.10)$$

An interesting observation appears when both ρ and ρ_p vanish, with their ratio $r = \rho_p/\rho \geq 1$ fixed and finite. In this case (.4.10) reduces to

$$C \leq \frac{d_d}{d_d + 1} \rho + o(\rho). \quad (.4.11)$$

On the other hand, even if the channel (.4.1) is memoryless, for $r \in [1, \infty)$, \dot{C} is still $d_d \rho / (d_d + 1)$ [21]. Thus we conclude that

Corollary .4.2 *For the discrete-time Rician fading channel (.4.1) under peak and average SNR constraints*

$$\begin{aligned} \frac{1}{n} \mathcal{E} \left[\sum_{i=1}^n S[i] S^\dagger[i] \right] &\leq \rho \\ S[i] S^\dagger[i] &\leq \rho_p = r\rho, \quad \text{for } i = 1, \dots, \end{aligned}$$

where $r \geq 1$ is fixed and finite, its capacity always behaves like

$$C = \frac{d_d}{d_d + 1} \rho + o(\rho), \quad \text{as } \rho \rightarrow 0, \quad (.4.12)$$

regardless of the channel temporal correlation.

That is, temporal correlation does not affect the first-order behavior of capacity, which is solely determined by the direct LOS component.

.4.3 Extension to Continuous-Time Channels

We can further apply the techniques in [64] to address continuous-time Rician fading channels. The result is as follows.

Proposition .4.3 *The capacity per unit energy of the continuous-time Rician fading channel (3.6.1) under the peak envelope constraint*

$$S(t)S^\dagger(t) \leq P, \quad \text{for } -\infty < t < \infty,$$

is

$$\dot{C} = 1 - \frac{1}{P} \frac{1}{2\pi} \int_{-\infty}^{\infty} \log \left(1 + \frac{P}{d_c + 1} S_{H_c}(j\omega) \right) d\omega. \quad (.4.13)$$

.5 Proof of Proposition 4.3.1

For simplicity, we may denote $a = \mathbf{s}^\dagger \mathbf{s}$ and $d = (\mathbf{s}^\dagger \Phi_{\mathbf{T}} \mathbf{s}) / (\mathbf{s}^\dagger \mathbf{s})$. So a is solely determined by the amplitude of \mathbf{s} , and d is solely determined by the direction of \mathbf{s} . Then the function to be maximized in (4.3.1) reads as

$$G(a, d) = N \cdot d - \sum_{n=1}^N \frac{\log(1 + \lambda_n(\Phi_{\mathbf{R}}) \cdot d \cdot a)}{a}.$$

By taking derivatives, we find that

- For fixed d , $G(a, d)$ is monotonically increasing with a , *i.e.*, it is maximized when each element of \mathbf{s} meets the peak SNR ρ_p .
- For fixed a , $G(a, d)$ is monotonically increasing with d , *i.e.*, it is maximized when $(\mathbf{s}^\dagger \Phi_{\mathbf{T}} \mathbf{s}) / (\mathbf{s}^\dagger \mathbf{s}) = \lambda_{\max}(\Phi_{\mathbf{T}})$.

So a flat signal \mathbf{s} with per-antenna energy ρ_p and satisfying (4.3.2) is a mutual maximizer of $G(a, d)$, and (4.3.3) immediately follows after simplification of (4.3.1). **Q.E.D.**

.6 Proof of Proposition 4.3.2

From the definition of $\ddot{I}(0)$, we have

$$\begin{aligned}
\ddot{I}(0) &:= 2 \cdot \lim_{\rho \rightarrow 0} \frac{I(\rho) - \dot{I}(0) \cdot \rho}{\rho^2} \\
&= 2 \cdot \lim_{\delta \rightarrow 0} -\mathcal{D}((1 - \delta) \cdot p_{X|S}(\mathbf{x}|S = 0) + \\
&\quad \delta \cdot p_{X|S}(\mathbf{x}|S = \mathbf{s}) \| p_{X|S}(\mathbf{x}|S = 0)) / (\mathbf{s}^\dagger \mathbf{s})^2 \cdot \delta^2 \\
&= -\frac{1}{(\mathbf{s}^\dagger \mathbf{s})^2} \cdot \Delta(p_{X|S}(\mathbf{x}|S = 0) \| p_{X|S}(\mathbf{x}|S = \mathbf{s})),
\end{aligned}$$

where the last equality is from [75, (162)], and $\Delta(p_0(x) \| p_1(x))$ is the Pearson's χ -divergence defined by

$$\Delta(p_0(x) \| p_1(x)) := \int_{\mathcal{X}} p_0(x) \left(\frac{p_1(x)}{p_0(x)} - 1 \right)^2 dx.$$

Then we have

$$\begin{aligned}
&\Delta(p_{X|S}(\mathbf{x}|S = 0) \| p_{X|S}(\mathbf{x}|S = \mathbf{s})) \\
&= \int_{\mathbb{C}^{N \times 1}} \frac{p_{X|S}^2(\mathbf{x}|S = \mathbf{s})}{p_{X|S}(\mathbf{x}|S = 0)} d\mathbf{x} - 1 \\
&= \frac{1}{\pi^N \det^2(\mathbf{I}_{N \times N} + (\mathbf{s}^\dagger \Phi_{\mathbf{T}\mathbf{S}}) \cdot \Phi_{\mathbf{R}})} \\
&\quad \cdot \int_{\mathbb{C}^{N \times 1}} \exp\{-\mathbf{x}^\dagger [(2 \cdot (\mathbf{I}_{N \times N} + (\mathbf{s}^\dagger \Phi_{\mathbf{T}\mathbf{S}}) \cdot \Phi_{\mathbf{R}})^{-1} \\
&\quad - \mathbf{I}_{N \times N}] \mathbf{x}\} d\mathbf{x} - 1,
\end{aligned}$$

where the integral is finite only if $(2 \cdot (\mathbf{I}_{N \times N} + (\mathbf{s}^\dagger \Phi_{\mathbf{T}\mathbf{S}}) \cdot \Phi_{\mathbf{R}})^{-1} - \mathbf{I}_{N \times N})$ is positive definite, *i.e.*, $((\mathbf{s}^\dagger \Phi_{\mathbf{T}\mathbf{S}}) \cdot \Phi_{\mathbf{R}} - \mathbf{I}_{N \times N})$ is negative definite, which holds only if $(\mathbf{s}^\dagger \Phi_{\mathbf{T}\mathbf{S}}) \cdot \lambda_n(\Phi_{\mathbf{R}}) < 1$ for all $n = 1, \dots, N$. When the integral is finite, we have

$$\begin{aligned}
&\Delta(p_{X|S}(\mathbf{x}|S = 0) \| p_{X|S}(\mathbf{x}|S = \mathbf{s})) \\
&= \frac{1}{\det(\mathbf{I}_{N \times N} - (\mathbf{s}^\dagger \Phi_{\mathbf{T}\mathbf{S}})^2 \cdot \Phi_{\mathbf{R}}^2)} - 1,
\end{aligned}$$

and (4.3.4) follows. **Q.E.D.**

.7 PDF and CDF of Complex Gaussian Quadratic Forms with a Positive Semi-Definite Kernel Matrix

Consider such a random variable as $R = \mathbf{G}^\dagger \Lambda \mathbf{G}$, where $\mathbf{G} \sim \mathcal{CN}(0, \mathbf{I}_{N \times N})$. Tentatively denote the N eigenvalues of Φ_R by $\{\lambda_1, \lambda_2, \dots, \lambda_N\} \geq 0$. Since the distribution of \mathbf{G} is invariant under unitary transformations, without loss of generality, we may rewrite $R = \mathbf{G}^\dagger \text{diag}(\lambda_1, \dots, \lambda_N) \mathbf{G}$.

Now let us order these N eigenvalues. Suppose that they can be grouped into K distinct non-zero eigenvalues $\{\lambda_1, \dots, \lambda_K\}$, each with multiplicity $\{N_1, \dots, N_K\}$. Thus we may further rewrite R as

$$R = \sum_{k=1}^K \left(\lambda_k \sum_{i=1}^{N_k} G_i^\dagger G_i \right).$$

For each k , $\lambda_k \sum_{i=1}^{N_k} G_i^\dagger G_i$ is a χ^2 random variable with $2N_k$ degrees of freedom [60], hence the characteristic function of R is

$$\begin{aligned} \varphi_R(j\theta) &= \prod_{k=1}^K \frac{1}{(1 - j\lambda_k \theta)^{N_k}} \\ &= \sum_{k=1}^K \sum_{n_k=1}^{N_k} \frac{\alpha_{n_k}}{(1 - j\lambda_k \theta)^{n_k}}, \end{aligned}$$

where the coefficients $\{\alpha_{n_k}\}_{n_k=1}^{N_k}$ are uniquely determined by decomposing the product. The PDF of R is then obtained by inverse Fourier transform, as

$$p_R(r) = \sum_{k=1}^K \sum_{n_k=1}^{N_k} \frac{\alpha_{n_k}}{\lambda_k^{n_k} \cdot (n_k - 1)!} r^{n_k-1} e^{-r/\lambda_k}, \quad (.7.1)$$

for $r \geq 0$; and the CDF of R is

$$P_R(r) = \sum_{k=1}^K \sum_{n_k=1}^{N_k} \alpha_{n_k} \left[1 - e^{-r/\lambda_k} \sum_{i=0}^{n_k-1} \frac{1}{i!} \left(\frac{r}{\lambda_k} \right)^i \right], \quad (.7.2)$$

for $r \geq 0$.

BIBLIOGRAPHY

- [1] Ibrahim C. Abou-Faycal, Mitchell D. Trott, and Shlomo Shamai (Shitz). The Capacity of Discrete-Time Memoryless Rayleigh-Fading Channels. *IEEE Trans. Inform. Theory*, 47(4):1290–1301, May 2001.
- [2] R. B. Ash and M. F. Gardner, *Topics in Stochastic Processes*, Academic Press, New York, 1975.
- [3] Jens Baltersee, Gunnar Fock, and Heinrich Meyr. Achievable Rate of MIMO Channels with Data-Aided Channel Estimation and Perfect Interleaving. *IEEE J. Select. Areas Commun.*, 19(12):2358–2368, December 2001.
- [4] Ezio Biglieri, John Proakis, and Shlomo Shamai (Shitz). Fading Channels: Information-Theoretic and Communications Aspects. *IEEE Trans. Inform. Theory*, 44(6):2619–2692, October 1998.
- [5] M. J. Borran, A. Sabharwal, and B. Aazhang. On Design Criteria and Construction of Noncoherent Space-Time Constellations. *IEEE Trans. Inform. Theory*, 49(10):2332–2351, October 2003.
- [6] R. C. Bradley, “Basic Properties of Strong Mixing Conditions: A Survey and Some Open Questions,” *Probability Surveys*, vol. 2, pp. 107–144, 2005.
- [7] Giuseppe Caire, Giorgio Taricco, and Ezio Biglieri. Bit-Interleaved Coded Modulation. *IEEE Trans. Inform. Theory*, 44(3):927–946, May 1998.
- [8] R.-R. Chen, Bruce Hajek, Ralf Koetter, and U. Madhow. On Fixed Input Distributions for Noncoherent Communication over High SNR Rayleigh Fading Channels. *IEEE Trans. Inform. Theory*, 50(12):3390–3396, December 2004.
- [9] C-N. Chuah, D. N. C. Tse, R. A. Valenzuela, and J. M. Kahn. Capacity Scaling in MIMO Wireless Systems Under Correlated Fading. *IEEE Trans. Inform. Theory*, 48(3):637–650, March 2002.
- [10] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, 1991.
- [11] Crossbow Technology, <http://www.xbow.com>
- [12] J. L. Doob. *Stochastic Processes*. John Wiley & Sons, Inc., New York, 1953.
- [13] R. Etkin and D. N. C. Tse, “Degrees of Freedom in Some Underspread MIMO Fading Channels,” *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1576–1608, April 2006.

- [14] N. T. Gaarder. Signal Design for Fast-Fading Gaussian Channels. *IEEE Trans. Inform. Theory*, 17(3):247–256, May 1971.
- [15] M. Godavarti, T. L. Marzetta, and S. Shamai. Capacity of a Mobile Multiple-Antenna Wireless Link with Isotropically Random Rician Fading. *IEEE Trans. Inform. Theory*, 49(12):3330–3334, December 2003.
- [16] Andrea J. Goldsmith and Pravin P. Varaiya. Capacity, Mutual Information, and Coding for Finite-State Markov Channels. *IEEE Trans. Inform. Theory*, 42(3):868–886, May 1996.
- [17] A. J. Goldsmith and P. P. Varaiya, “Capacity of Fading Channels with Channel Side Information,” *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.
- [18] I. S. Gradshteyn and I. M. Ryzhik. *Tables of Integrals, Series, and Products*. Academic Press, fifth edition, 1994.
- [19] Robert M. Gray. Toeplitz and Circular Matrices: A Review. <http://ee.stanford.edu/~gray/toeplitz.pdf>.
- [20] M. C. Gursoy, H. V. Poor, and S. Verdú. The Noncoherent Rician Fading Channel - Part I: Structure of the Capacity Achieving Input; Part II: Spectral Efficiency in the Low Power Regime. *IEEE Trans. Wireless Commun.*, Sept. 2005.
- [21] M. C. Gursoy, H. V. Poor, and S. Verdú, “Noncoherent Rician Fading Channel - Part II: Spectral Efficiency in the Low-Power Regime,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2207–2221, Sept. 2005.
- [22] Bruce Hajek and Vijay G. Subramanian. Capacity and Reliability Function for Small Peak Signal Constraints. *IEEE Trans. Inform. Theory*, 48(4):828–839, April 2002.
- [23] Babak Hassibi and Bertrand M. Hochwald. How Much Training is Needed in Multiple-Antenna Wireless Links? *IEEE Trans. Inform. Theory*, 49(4):951–963, April 2003.
- [24] Bertrand M. Hochwald and Wim Sweldens. Differential Unitary Space-Time Modulation. *IEEE Trans. Commun.*, 48(12):2041–2052, December 2000.
- [25] Bertrand M. Hochwald and Thomas L. Marzetta. Unitary Space-Time Modulation for Multiple-Antenna Communications in Rayleigh Flat Fading. *IEEE Trans. Inform. Theory*, 46(2):543–564, March 2000.
- [26] Bertrand M. Hochwald, Thomas L. Marzetta, Thomas J. Richardson, Wim Sweldens, and Rüdiger Urbanke. System Design of Unitary Space-Time Constellations. *IEEE Trans. Inform. Theory*, 46(6):1962–1973, September 2000.
- [27] S. A. Jafar and A. Goldsmith. Multiple-Antenna Capacity in Correlated Rayleigh Fading with Channel Covariance Information. *IEEE Trans. Wireless Commun.*, 2003. submitted.

- [28] William C. Jakes and Donald C. Cox, editors. *Microwave Mobile Communications*. IEEE Press, New York, 1994.
- [29] JTRS, <http://jtrs.army.mil>
- [30] Thomas Kailath. Correlation Detection of Signals Perturbed by a Random Channel. *IEEE Trans. Inform. Theory*, 6(3):361–366, June 1960.
- [31] Thomas Kailath. *Lectures on Wiener and Kalman Filtering*. Springer-Verlag, New York, 1981.
- [32] Robert Kennedy. *Fading Dispersive Communication Channels*. John Wiley & Sons, 1969.
- [33] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen. A Stochastic MIMO Radio Channel Model with Experimental Validation. *IEEE J. Select. Areas Commun.*, 20(6):1211–1226, August 2002.
- [34] J. Nicholas Laneman, David N. C. Tse, and Gregory W. Wornell. Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior. *IEEE Trans. Inform. Theory*, 50(12):3062–3080, December 2004. Available at <http://www.nd.edu/~jnl/pubs/it2002.pdf>.
- [35] Amos Lapidoth. On the Asymptotic Capacity of Stationary Gaussian Fading Channels. *IEEE Trans. Inform. Theory*, 51(2):437–446, February 2005.
- [36] Amos Lapidoth and Stefan M. Moser. Capacity Bounds via Duality with Applications to Multiple-Antenna Systems on Flat Fading Channels. *IEEE Trans. Inform. Theory*, 49(10):2426–2567, October 2003.
- [37] Amos Lapidoth and Prakash Narayan. Reliable Communications Under Channel Uncertainty. *IEEE Trans. Inform. Theory*, 44(6):2148–2177, October 1998.
- [38] Amos Lapidoth and S. Shamai (Shitz). Fading Channels: How Perfect Need “Perfect Side Information” Be? *IEEE Trans. Inform. Theory*, 48(5):1118–1134, May 2002.
- [39] A. Lapidoth and L. Wang, “On the Low SNR Capacity of Peak-Limited Non-Coherent Fading Channels with Memory,” [arXiv:cs.IT/0604031](https://arxiv.org/abs/cs.IT/0604031)
- [40] T. Li and O. Collins, “A Successive Decoding Strategy for Channels with Memory,” in *Proc. IEEE Int. Symp. Inform. Theory*, 2005.
- [41] A. Lozano, A. M. Tulino, and S. Verdú. Multiple-Antenna Capacity in the Low-Power Regime. *IEEE Trans. Inform. Theory*, 49(10):2527–2544, October 2003.
- [42] D. S. Lun, M. Médard, and I. C. Abou-Faycal. On the Performance of Peak Capacity-Achieving Signaling on Multipath Fading Channels. *IEEE Trans. Commun.*, 52(6):931–938, June 2004.

- [43] Thomas L. Marzetta and Bertrand M. Hochwald. Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading. *IEEE Trans. Inform. Theory*, 45(1):139–157, January 1999.
- [44] Thomas L. Marzetta and Shlomo Shamai (Shitz). Multiuser Capacity in Block Fading with No Channel State Information. *IEEE Trans. Inform. Theory*, 48(4):938–942, April 2002.
- [45] J. L. Massey, “All Signal Sets Centered About the Origin Are Optimal at Low Energy-to-Noise Ratios on the AWGN Channel,” in *Proc. IEEE Int. Symp. Inform. Theory*, pp. 80–81, 1976.
- [46] Robert McEliece. Are Turbo-like Codes Effective on Nonstandard Channels. In *ISIT Plenary Lecture*, 2001.
- [47] Muriel Medard. The Effect Upon Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel. *IEEE Trans. Inform. Theory*, 46(3):933–946, May 2000.
- [48] Muriel Medard. Channel Uncertainty in Communications. *IEEE Information Theory Society Newsletter*, 53(2):1–12, June 2003.
- [49] Muriel Medard, Ibrahim C. Abou-Faycal, and Upamanyu Madhow. Adaptive Coding for PSAM without Feedback. *IEEE Trans. Commun.*, 2002. Submitted for publication.
- [50] Muriel Medard and R. G. Gallager. Bandwidth Scaling for Fading Multipath Channels. *IEEE Trans. Inform. Theory*, 48(4):840–852, April 2002.
- [51] F. Neeser and J. L. Massey. Proper Complex Random Processes with Applications to Information Theory. *IEEE Trans. Inform. Theory*, 39(4):1293–1302, July 1993.
- [52] A. V. Oppenheim and R. W. Schafér, From Frequency to Quefrequency: A History of the Cepstrum. *IEEE Signal Processing Magazine*, 95–106, September 2004.
- [53] M. Peleg and Shlomo Shamai (Shitz). Reliable Communication over the Discrete-Time Memoryless Rayleigh Fading Channel with Turbo Coding/Decoding. *European Trans. on Telecomm.*, 11(5):475–485, September-October 2000.
- [54] J. F. Pieper, J. G. Proakis, R. R. Reed, and J. K. Wolf. Design of Efficient Coding and Modulation for a Rayleigh Fading Channel. *IEEE Trans. Inform. Theory*, 24(4):457–468, July 1978.
- [55] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*, Holden-Day, San Francisco, 1964.
- [56] M. S. Pinsker, V. V. Prelov and S. Verdú, “Sensitivity of Channel Capacity,” *IEEE Trans. Inform. Theory*, Nov. 1995.

- [57] A. S. Y. Poon, R. W. Brodersen, and D. N. C. Tse, “Degrees of Freedom in Multiple-Antenna Channels: A Signal Space Approach,” *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 523–536, Feb. 2005.
- [58] H. Vincent Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, 1994.
- [59] Robert Price. Optimum Detection of Random Signals in Noise, with Applications to Scatter-Multipath Communication, I. *IEEE Trans. Inform. Theory*, 2(4):125–135, December 1956.
- [60] John G. Proakis. *Digital Communications*. McGraw-Hill, Inc., New York, third edition, 1995.
- [61] T. S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall, 2002.
- [62] John S. Richters. *Communication over Fading Dispersive Channels*. RLE Technical Report 464, MIT, 1967.
- [63] Mischa Schwartz, William Ralph Bennett, and Seymour Stein. *Communication Systems and Techniques*. McGraw-Hill, New York, 1966.
- [64] V. Sethuraman and B. Hajek. Capacity per Unit Energy of Fading Channels with a Peak Constraint. *IEEE Trans. Inform. Theory*, 51(9):3102–3120, September 2005.
- [65] V. Sethuraman, B. Hajek, and K. Narayanan, “Capacity Bounds for Noncoherent Fading Channels with a Peak Constraint,” in *Proc. IEEE Int. Symp. Inform. Theory*, 2005.
- [66] V. Sethuraman and B. Hajek, “Low SNR Capacity of Fading Channels with Peak and Average Power Constraints,” [arXiv:cs.IT/0604049](https://arxiv.org/abs/cs.IT/0604049)
- [67] Hyundong Shin and Jae Hong Lee. Capacity of Multiple-Antenna Fading Channels: Spatial Fading Correlation, Double Scattering, and Keyhole. *IEEE Trans. Inform. Theory*, 49(10):2636–2647, October 2003.
- [68] Vijay Subramanian and Bruce Hajek. Broad-Band Fading Channels: Signal Burstiness and Capacity. *IEEE Trans. Inform. Theory*, 48(4):809–827, April 2002.
- [69] SUOSAS, <http://www.darpa.mil/ato/programs/SUOSAS/index.htm>
- [70] Giorgio Taricco and M. Elia. Capacity of Fading Channels with No Side Information. *Electron. Lett.*, 33(16):1368–1370, July 1997.
- [71] I. Emre Telatar. Capacity of Multi-Antenna Gaussian Channels. *European Trans. on Telecomm.*, 10(6):585–596, November-December 1999.
- [72] Lang Tong, Brian M. Sadler, and Min Dong. Pilot-Assisted Wireless Transmissions. *IEEE Mag. Signal Processing*, pages 12–25, November 2004.

- [73] Ultra-Wideband (UWB) Technology, Technology & Research at Intel, <http://www.intel.com/technology/comms/uwb/>
- [74] Sergio Verdú. Channel Capacity per Unit Cost. *IEEE Trans. Inform. Theory*, 36(5):1019–1030, September 1990.
- [75] Sergio Verdú. Spectral Efficiency in the Wideband Regime. *IEEE Trans. Inform. Theory*, 48(6):1319–1343, June 2002.
- [76] A. J. Viterbi, “Performance of an M-ary Orthogonal Communication System Using Stationary Stochastic Signals,” *IEEE Trans. Inform. Theory*, vol. 13, no. 3, pp. 414–422, July 1967.
- [77] Werner Weichselberger, Markus Herdin, Hüseyin Özcelik, and Ernst Bonek. A Stochastic MIMO Channel Model with Joint Correlation of Both Link Ends. *IEEE Trans. Wireless Commun.*, August 2003. Submitted for publication.
- [78] J. Wolfowitz, *Coding Theorems of Information Theory*, Springer-Verlag, Berlin, 2nd ed., 1964.
- [79] J. M. Wozencraft and I. M. Jacobs. *Principles of Communication Engineering*. Wiley, New York, 1965.
- [80] X. Wu and R. Srikant. MIMO Channels in the Low SNR Regime: Communication Rate, Error Exponent and Signal Peakiness. *IEEE Trans. Inform. Theory*, 2004. submitted.
- [81] Aaron D. Wyner. Bounds on Communication with Polyphase Coding. *Bell Sys. Tech. J.*, 45:523–559, April 1966.
- [82] Lihong Zheng and David N. C. Tse. Communication on the Grassmann Manifold: A Geometric Approach to the Noncoherent Multiple-Antenna Channel. *IEEE Trans. Inform. Theory*, 48(2):359–383, February 2002.
- [83] Lihong Zheng and David N. C. Tse. Diversity and Multiplexing: A Fundamental Tradeoff in Multiple-Antenna Channels. *IEEE Trans. Inform. Theory*, 49(5):1073–1096, May 2003.
- [84] Lihong Zheng, David N. C. Tse, and Muriel Medard. Channel Coherence in the Low SNR Regime. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Chicago, IL, June 2004.