Ph.D. Thesis
Doctor of Philosophy

HELLENIC REPUBLIC
National and Kapodistrian
University of Athens

Applications of Random Matrix Theory in Wireless and Fiber Optical MIMO Channels

Apostolos Karadimitrakis
(B.Sc, M.Eng)

Athens 2018

Department of Physics
Section of Electronics, Computer, Telecommunications and Control
National & Kapodistrian University of Athens

Panepistimioupolis
Building PHY IV-V
Athens, Greece
Phone +30 210 7276755

http://www.phys.uoa.gr

Abstract

Aim of this work is to present metrics like Outage Capacity and Gallager Bound for the characterization of the performance of the link of wireless and fiber optical channels. Due to the increased complexity and computational burden, these metrics can only be approximated through methods stemming from Random Matrix Theory, Large Deviations and Replica Theory. In this work we provide analytic equations for the approximation of these metrics and we show that they are very accurate for real systems.

Summary

The thesis in divided into 7 chapters: *Introduction* (Chapter 1), *Theory - Mathematical and Telecommunications Background* (Chapter 2), *The Optical Channel* (Chapter 3), *Crosstalk Existence* (Chapter 4), *Outage Capacity* (Chapter 5), *Gallager Bound* (Chapter 6) and *Optical Fiber MIMO Channel Model- Hamiltonian Approach* (Chapter 7).

In Chapter 1 it is given a brief description of the telecommunication trends and how they will change in the upcoming years. Moreover, it is a made a small introduction into wireless and fiber optical communications.

In Chapter 2 it is given the theory for the required mathematical and telecommunications background which is necessary in this work. In particular some basic notions from random matrix theory (RMT) and statistical physics are demonstrated along with some examples. Moreover, it is presented the large deviation (LD) approach and the saddle point analysis in addition to basic concepts of information theory in SISO and MIMO telecommunications.

In Chapter 3 the model of the optical MIMO channel is demonstrated and we discuss the two different approaches for its analysis: the scattering matrix approach and the Hamiltonian approach.

Further, in Chapter 4 it is given evidence of the presence of crosstalk in modern optical fiber communications and thus, rendering them ideal for the application of the MIMO technology from the wireless domain.

In Chapter 5 and 6 some metrics are presented for the evaluation of the performance of fiber optical MIMO networks.

Finally, in Chapter 7 it is demonstrated a channel model for the fiber optical MIMO channel.

Preface

This PhD thesis was prepared at the Department of Physics, Section of Electronics, Computer, Telecommunications and Control, at the National & Kapodistrian University of Athens in fulfillment of the requirements for acquiring a Ph.D degree in Communication Networks.

Athens, November 9, 2018

Apostolos Karadimitrakis
(B.Sc, M.Eng)

Accepted by the jury:

Aris Moustakas	**Dimitrios Frantzeskakis**	**Ektoras Nistazakis**
Associate Professor,	Professor,	Associate Professor,
Department of Physics,	Department of Physics,	Department of Physics,
University of Athens	University of Athens	University of Athens

Andreas Polydoros	**Nicholaos Kalouptsidis**	**Takis Mathiopoulos**
Professor,	Professor,	Professor,
Department of Physics,	Department of Informatics & Telecommunications,	Department of Informatics & Telecommunications,
University of Athens	University of Athens	University of Athens

George S. Tombras
Professor,
Department of Physics

University of Athens

Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisor As. Prof. Aris L. Moustakas for the support, patience and immense knowledge. Many times that I had found myself swamped or lost, he was there to provide me with guidance and aid, and reach solutions to problems otherwise faraway. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to thank Prof. Merouane Debbah who provided me the opportunity to continue my research in the world famous CentraleSupélec in collaboration with world leader companies such as Alcatel-Lucent and Huawei. His precious support had- and still has- a great impact on me.

I thank my fellow researchers both in Greece and in France for the discussions and support all these years. Their friendship is invaluable to me.

Last but not least, I would like to thank my family. This life time endeavor would have been impossible without the immense support from my parents and my brother. I know for sure that as they were beside me all these years, they will be there also for the years to come.

Notations

Linear algebra

$\mathbf{X}$	matrix
$\mathbf{I}_N$	identity matrix of size $N \times N$
$\mathrm{diag}(x_1, x_2, \ldots, x_N)$	diagonal matrix with entries $x_1, x_2, \ldots, x_N$
$\mathbf{X}^T$	transpose of $\mathbf{X}$
$\mathbf{X}^\dagger$	complex conjugate transpose of $\mathbf{X}$
$\mathbf{X}^*$	complex conjugate of $\mathbf{X}$
$\mathrm{tr}\mathbf{X}$	trace of $\mathbf{X}$
$\det \mathbf{X}$	determinant of $\mathbf{X}$
$\mathbf{X} \otimes \mathbf{Y}$	Kronecker or Tensor product of $\mathbf{X}$ and $\mathbf{Y}$
$\|\|\mathbf{X}\|\|_p$	ℓ-p norm of $\mathbf{X}$
$\mathbf{x}$	column vector
x_i	ith entry of vector $\mathbf{x}$
$\leq, \geq, >, <$	component-wise inequalities, e.g. $\mathbf{x} \geq \mathbf{y}$ implies that $x_i \geq y_i \ \forall i$
$\preceq, \succeq, \succ, \prec$	matrix inequalities, e.g., $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is nonnegative definite
$D\mathbf{X} \equiv \prod_{i=1}^{m_{rows}} \prod_{j=1}^{m_{cols}} dX_{ij}$	integrals over real matrix elements. We integrate over the elements of an $m_{rows} \times m_{cols}$ matrix $\mathbf{X}$

Analysis

$\mathbb{C}, \mathbb{R}, \mathbb{N}$	the complex, real and natural numbers
$\|x\|$	absolute value
$\Re\{z\}$	real part of z
$\Im\{z\}$	imaginary part of z
i	$i = \sqrt{-1}$ with $\Im\{i\} = 1$
$f'(x)$	first derivative of $f(x)$
$\log(x)$	natural logarithm
$\mathbb{C}^+$	$z \in \mathbb{C} : \Im\{z\} > 0$

Probability theory

F_X	distribution function of X, i.e., $F_X(x) = P(X \leq x)$
$\mathbb{E}[X]$	expectation of X, i.e, $\mathbb{E}[X] = \int_\Omega X(\omega) dP(\omega)$
$\xrightarrow{\text{a.s.}}$	almost sure convergence
$\sim$	distributed as, e.g., $X \sim N(0,1)$
$N(m, Q)$	Gaussian distribution with mean m and covariance Q
$Q(x)$	Q-function, i.e., $Q(x) = \int_0^\infty \frac{1}{2\pi} e^{-\frac{t^2}{2}} dt$

Contents

CHAPTER 1

Introduction

1.1 Communications Carving our World

Undoubtedly the era that we are living in can be identified as the era of information. Never before in the history of mankind, vast amount of information was delivered across the earth in the blink of an eye and the whole knowledge of humanity was in the disposal of every man and woman. The outstanding fact is that all these are happening individually and simultaneously. In other words, it is not necessary for someone to be in a specific place i.e. home, library etc to access the vast ocean of information, but on the contrary, the information is available to everyone and at any time, even to hand held devices such as mobile phones.

That breakthrough has altered the human societies forever. First of all, we (everyone separately and together collectively) get in contact with anyone, no mater where they are, forming a global civilization. Secondly, the search of knowledge is unthinkably easier now and the scientific and technological advance more rapid than ever. Personally, I am having difficulties to imagine myself writing the same thesis and performing the same scientific work without the use of internet. Finally, the easiness of information exchange is opening new frontiers to the human evolution. Sociologically, scientifically, economically, Earth will not be the same.

Modern communications are carving everyday life. It is almost impossible nowadays for someone not to own a mobile phone or not to use a broadband internet connection: We are all connected to each other, constantly exchanging information either for work or for recreation purposes. Analysts predict that soon the number of people and devices trying to "go online" will exceed the 15 billions in number 1.1 and at the same time, the demand for greater connection speeds with minimal latency will be undiminished, rendering the today technology obsolete.

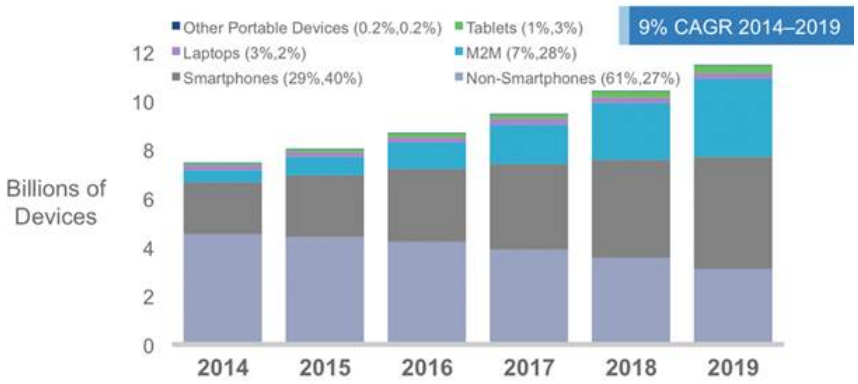


Figure 1.1: Global Mobile Devices and Connections Growth [1].

Although, xDSL , 3G and 4G services are totally adequate to accommodate today the increased demand for connectivity, this is impossible to be valid in the near future, thus making the establishment of the new 5G wireless and fiber-to-the-premise wired networks an one-way road.

1.2 The Challenge of Backbone Networks.

One major common feature in our communications either through wireless infrastructure (2G, 3G, LTE) or through the wired one (copper, coaxial) is that our traffic passes, almost surely, through a network of optical fibers. Thus, the optical networks, even if they do not reach our premises directly (FTTx), form the backbone of modern information exchange and feel the urgent need of expanding their potential capacity. So far, we keep optimizing the end-user part of the network and we believe that the already laid optic fiber network is totally adequate to administer the increased traffic. But as the number of users soars high and the information exchange rate follows this trend, the networks become more and more congested. A perfect example for this, is the traffic jam occurring in highways [10]: as the number of car users increases, the traffic jams become an everyday phenomenon even for the highways. Studies show [66], that by the year 2020 a capacity crunch in the optical fiber networks is eminent as can be seen in Fig. 1.2 and we are forced to find a drastic solution in this urgent problem.

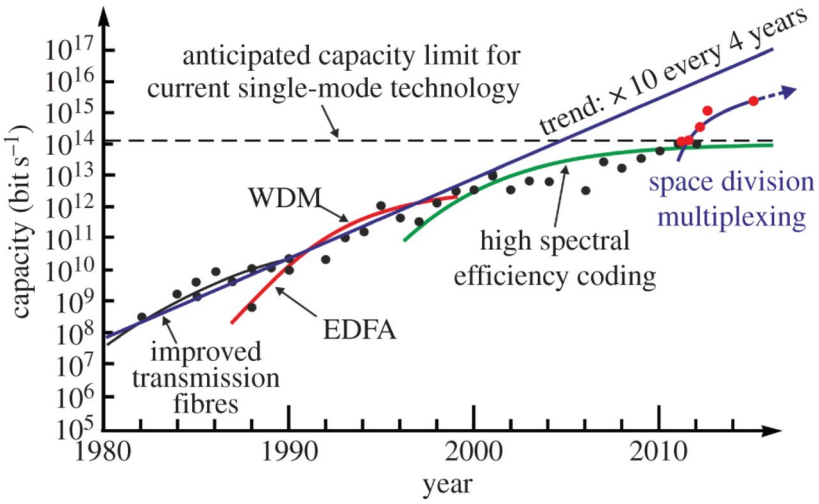


Figure 1.2: A capacity crunch is unavoidable in the upcoming future [66].

1.3 Modern Communications

Wireless Communications

The birth of wireless communications dates back to early 1800's with the pioneering work of Oersted, Faraday, Gauss, Maxwell and Hertz. The explosive growth of radio communication systems, however, happened the next century, when Fleming demonstrated the first vacuum tube amplifier which led to the invention of amplitude modulated (AM) signal and so the first radio station appeared i.e., Radio KDKA in Pittsburg.

Nowadays almost no one thinks of himself without using some sort of wireless communication. This involves the transmission of information over a distance without the help of wires, cables or any other forms of electrical conductors. This definition is so broad that includes communications from just a few meters of distance e.g., television's remote control, to thousands of kilometers e.g., ionospheric and satellite communications. Of course cellular telephony is the type of wireless communication with the most usage in our times.

System of Wireless Communications

The wireless communication system is designed so as to send information from a source to one or more recipients. In general, such a (digital) system can be seen in Fig. 1.3.

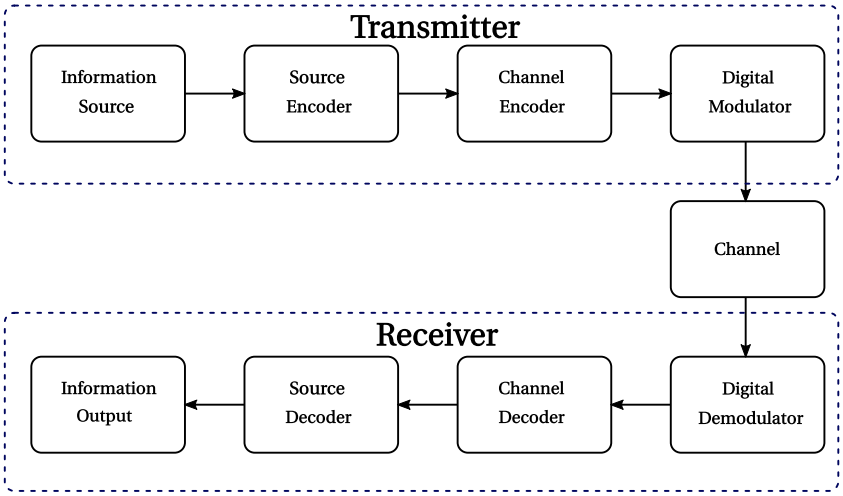


Figure 1.3: Basic elements of a digital communication system.

We can divide the system into three major parts:

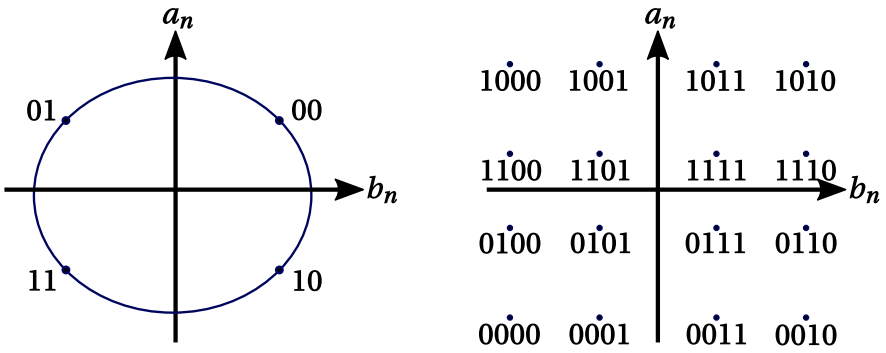
- The transmitter. Produces the information/message, encodes and modulates it.
- The channel. Is the physical medium that is used for the propagation of the signal.
- The receiver. Collects the messages and after successful demodulation and decoding, acquires the information.

Transmitter

In contemporary digital communication systems, the messages produced by the source are converted into a binary digits (bits) sequence. Ideally the output of the source consists of the minimum number of bits possible, thus the redundancy is minimized. This procedure is called source encoding or data compression.

The data sequence then, enters the channel encoder which inserts deliberately some sort of redundancy in order to increase the reliability of the received data and to enhance the fidelity of the corresponding signal. For example, a trite method of encoding a binary sequence is just repeating every bit m-times. A more complicated encoder, receives k bits of information and represents them in another sequence of n bits ($n > k$) which is called codeword. The amount of redundancy inserted with this method is measured through the ratio n/k and the inverse of this, is called code rate.

After that, the encoded message enters the modulator which represents the binary sequence into a signal waveform e.g., Phase Shift Keying (PSK) or Quadrature Amplitude Modulation (QAM), see Fig. 1.4



(a) 4-PSK (b) 16-QAM

Figure 1.4: Signal constellation. The corresponding waveform can be expressed as $s(t) = a(t)cos(2\pi f_c t) - b(t)sin(2\pi f_c t)$.

We can assume that the transmission of the bit sequence is happening at a uniform rate R $bits/sec$. If the output of the modulator is updated every $b-$bits, the corresponding available transmission time is b/R sec.

Receiver

Before we describe the channel we can take a look at the side of the receiver, which in practice executes the reverse operation of the transmitter: The demodulator processes the corrupted from the channel, waveform and estimates the corresponding bit. Using the inherited redundancy, the decoder is able to fill in, where the demodulator was unable to decide. A way of measuring (metric) the error-free operation of the demodulator and decoder is the rate of bit-error occurrence (Bit Error Rate - BER) at the output of the decoder.

Channel

As we described before, the channel provides the link between the transmitter and the receiver. Usually, the channel is governed by additive noise which corrupts any signal that propagates through it. This type of noise can be due to the electronics used in the different components of the system (thermal noise) as well as it can be due to interference from other users of the channel. Alternative sources of signal degradation are e.g., the signal attenuation and the distortion due to multipaths.

Many times, the noise effect can be overcome by increasing the power of the transmitted signal. However, limitations due to electronics and other reasons such as proximity to more channel users constrict this method. Finally, the available channel bandwidth imposes one more major limitation to our communications. That way, the maximum amount of information we can transmit through a given channel each time is finite, depends on the signal noise and bandwidth of the channel and is called capacity of the channel.

MIMO-Diversity and Spatial Multiplexing

The wireless communication channel is very hostile. The transmitted signal over a wireless communication link suffers from fading (severe fluctuations in signal level), dispersion in time and frequency, path loss and other phenomena of degradation. Of course, as was discussed briefly in the previous section, the limited available bandwidth posses one more obstacle in our wireless communications.

Multiple antenna systems are the current trend in our effort to design communication systems with high spectral efficiency and high quality of link availability. Mutliple Input-Multiple Output (MIMO) systems, whch can be seen in Fig. 1.5, improve the spectral efficiency and offer high quality links when compared to traditional Single Input-Single Output (SISO) systems [77, 78].

Apart from antenna configurations (antenna multiplicity either on the transmitter T_x-MISO- or on the receiver R_x-SIMO- or on both sides-MIMO) there are two ways of performing MIMO (in general) communications. Existence of multiple antennas in a system, means existence of a plurality of propagation paths, hence aiming at improving the reliability of the system we may choose to send the same data across the different propagation (spatial) paths. This is called spatial diversity. On the other hand, aiming at improving the data rate of the system, we may choose to transmit different portions of the data through different propagation paths. This is called spatial multiplexing.

In spatial multiplexing, each subchannel carries independent information, thus increasing the data rate of the system. This can be compared to Orthogonal Frequency Division Multiplexing (OFDM) technique, where, different frequency subchannels carry different parts of the modulated data. But in spatial multiplexing, if the scattering by the environment is rich enough, several independent subchannels are created in the same allocated bandwidth. Thus the multiplexing gain comes at no additional cost on bandwidth or power.

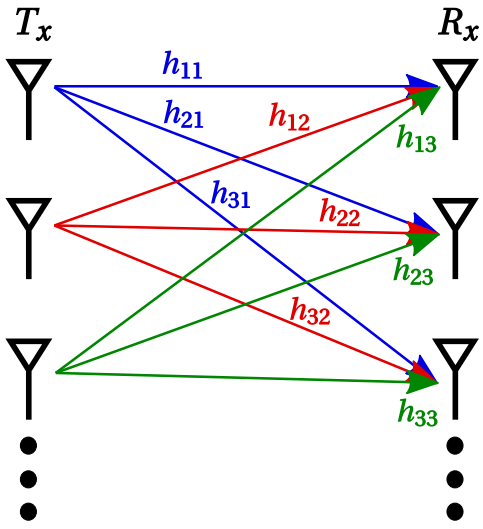


Figure 1.5: Schematic of the MIMO channel model.

Optical Communications

The Fiber

Optical fibers is not something new in communications. In fact, it was first proposed by Kao and Hockam [42] in 1966 and for that study they received the Nobel prize in physics in 2009. The idea is both simple and revolutionizing: light beam carrying the information, is inserted into a silica fiber and after many total reflections, the light comes out with minimal power loss as can be seen in the schematic Fig.1.6.

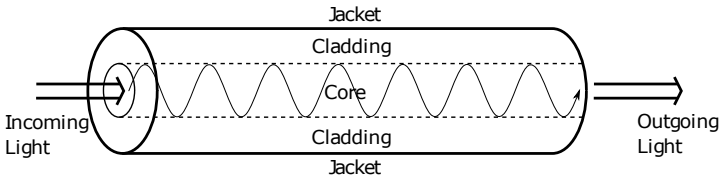


Figure 1.6: Schematic of a simple fiber.

In order to achieve total reflection, the fiber consists of three parts: the core with reflective index $n_1 \backsim 1.62$, where the waveguiding takes place, the cladding with reflective index $n_2 \backsim 1.52$ and the jacket which protects and insulates the rest system. According to the core diameter, there are different types of fiber. For core diameter $8 \backsim 12\ \mu m$ the fiber is called *Single Mode Fiber* (SMF) and as the name suggests, it can support only one propagating mode inside the core. When the core diameter gets bigger, then more than one propagating modes can exist inside the core (see Fig. 4.1) and the fiber is called *Multi Mode Fiber*. A sub-category of the MMF is the *Few Mode Fiber* (FMF) where the core can support only a few propagating modes.

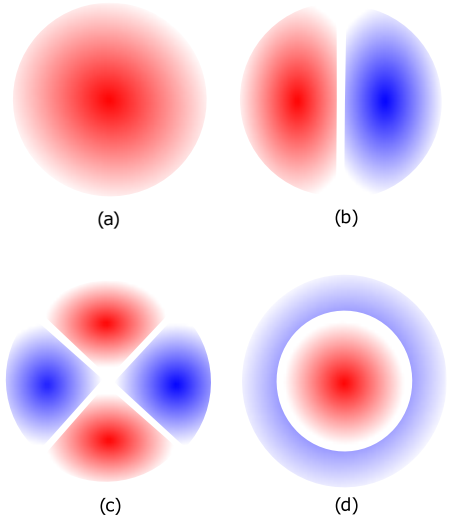


Figure 1.7: Mode profiles of the lower -order fiber modes (a) LP_{01} (b) LP_{11} (c) LP_{21} (d) LP_{02}

Because the different propagating modes have slightly different propagating velocities, great fiber lengths result in significant difference in the time of arrival of the various waveforms. Therefore SMFs are more suitable for long distance communications [30]. On the other hand, MMFs have greater splicing and coupling tolerance and are easier to manufacture, making them ideal for short range communications. A new type of optical fiber is the *Multi Core Fiber* where inside the cladding there is not only one, but multiple cores, each one acting as a single and different waveguide. But as the number of the cores increases, so does the energy leakage between them. Thus, the crosstalking phenomenon between the various cores arises.

In order to have total reflections resulting into successful light propagation inside the fiber, the incoming light beam from the source must come from a maximum angle in the core, called *Acceptance Angle* (θ_α). If n_0 is the reflective index of the air, then we call *Numerical Aperture* (NA) the quantity

$$NA = n_0 \sin\theta_\alpha = \sqrt{n_1^2 - n_2^2} \tag{1.1}$$

Light Sources

In optical communications the light source (see Fig. 1.8) can be either a LASER or a LED source depending on the type of the fiber we are going to use. In SMFs the NA is very small, thus precision in the alignment source-fiber is needed and the usage of LASER source is necessary.

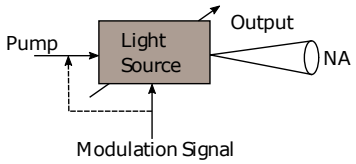


Figure 1.8: Ideal light source

Optical Modulators

The most common way of modulating an optical signal after its production from a LASER source, is by using a Mach-Zehnder modulator: The input light is split up into two interferometer arms. When we apply voltage across one of the arms, a phase shift is induced for the signal passing through that arm. If the two arms are recombined, the phase difference between the two waves is converted to an amplitude modulation. That way we can produce simple ASK (*Amplitude Shift Keying*) symbols or more complicate QAM (*Quadrature Amplitude Modulated*) ones.

Optical Amplifiers

In modern optical communications, the light amplifiers that are in use, are the *Erbium Doped Fiber Amplifiers* (EDFAs) which are, in fact, some lengths of Erbium doped

fiber. A secondary light source is used as a pump to raise the Erbium ions from their ground state to a metastable state and so, the incoming light signal absorbs energy (amplification) as the ions return to their ground state, see Fig 1.9.

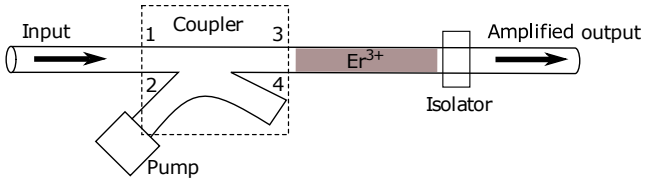


Figure 1.9: EDFA schematic

The main advantages of the EDFAs are the all-optical and low noise operation.

Photo Detection and Demodulation

The basic method of photo detection is just by using a photodetecting diode. That way the optical signal is converted into electrical one which is then fed to an equalizing system for the demodulation process. In optical fiber communications we do not use pilot signals to acquire channel information. Instead we use blind equalization techniques.

Spatial Multiplexing (SDM) in Optical Domain

The ongoing exponential growth in both wire and wireless data traffic is forcing the currently deployed infrastructure to its limits. To counter this trend, scientists have been working towards exhausting all available degrees of freedom of fiber-optical transmission, including the bandwidth (through WDM modulation), available power (subject to power constraints imposed by non-linearities), and polarization diversity [86]. One possibility to increase throughput is spatial modulation, which would allow multiple transmission streams within the same fiber or fiber bundle. This can be achieved by designing MMFs and/or MCFs. An important issue that arises is that typically there is cross-talk between fiber modes, as can be graphically seen in Fig 1.10, which increases with segment length [75] and can be attributed to imperfections, as well as to the twist and the bending of the fiber [21, 33], and slight variations in the local temperature [44].

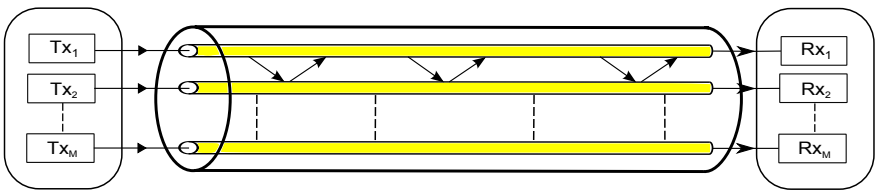


Figure 1.10: Optical SDM using M parallel transmission paths in MCF. There is crosstalking between adjacent cores.

There have been two trends of work in this direction. In the first, effort has been made to minimize cross talk between cores to extremely low levels [92], thus not having to deal with self-interference. While appealing from a signal processing point of view, the downside is that this methodology does not scale, in the sense that coupling becomes unavoidable with increasing number of cores in a fiber. Another more pragmatic approach is to design fibers without bothering about the appearance of cross-talk. Indeed, bringing cores close to each other can lead to power being spread at the receiver side evenly in the outlook of the channel [57]. Recently, it was proposed [38,76,86] to use sophisticated transceiver techniques developed in the context of wireless communications between multiple transmitting and receiving antennas (MIMO), which can mitigate self-interference, thus providing significant throughput increases. Of course, optical fiber multi-core systems have several differences compared to multi-antenna wireless systems, which need to be addressed later.

1.4 Summary of Contributions and Questions for Future Work

The main idea of this work stems from the need to provide cost-effective and efficient ways of evaluating the performance of modern communication links both fiber optical and wireless. Therefore, we had to work on different channel models which describe the physics governing the various connections, by using the assumption of large systems and hence utilizing the approximations provided by the random matrix theory. It was shown, however that the analytic expressions which were obtained, are applicable as well to small-sized systems, thus making our proposed methods especially useful to modern communication systems.

More precisely, for the optical MIMO channel we proved, first of all, the existence of crosstalking (Chapter 4) just like in an ordinary wireless system with multiple transmitters and receivers, and further we proposed a method of improving the demodulation procedure, based on compressed sensing techniques. Therefore, having established the basic notion that optic fiber links can follow the trend of the wireless ones, namely incorporate Spatial Division Multiplexing (SDM), we calculated the outage capacity of such a system (Chapter 5) under the assumptions of increased crosstalking and zero loss inside the fiber. Further, we investigated the Gallager bound for the coded fiber optic MIMO link (Chapter 6.2) with fixed power constraints as well as average power constraint among all transmitters, and the Sphere-Packing bound. Since, the calculation of the information capacity of a communication channel, can impose huge computational burden, we are obliged to find good enough approximations, which provide us with lower and upper bounds for the corresponding channel capacity. Once again, the proposed approximation method follows well the theory. Finally, as concern as the optic fiber channel we investigated a more realistic channel model based on the idea of a chaotic cavity (Chapter 7). Specifically, we modeled the optic fiber MIMO channel as a chaotic cavity where energy is injected and taken out from leads in the

form of particles, and the crosstalking happens in a random way with energy exchange between the particles. That way we can address phenomena occurring inside the fiber such as increased loss, non-linearities and some level of crosstalking.

Unlike in the optical domain, the use of random matrix theory methods in the wireless one is a well established practice by numerous of researchers. Hence, in the context of this work, we calculated an approximated Gallager bound for the wireless channel (Chapter 6.1) in a closed form by using random matrix theory for fixed power constraints as well as average power constraint among all transmitters, and the Sphere-Packing bound. That way, we provided a metric of the performance of such a system which although requires minimal computational burden, however it follows well the expected bounds.

Surely, the more we dive into the exciting world of telecommunications with the help of random matrix theory, the more questions rise. For example, a more realistic practical implementation of an optical MIMO system along with a compressed sensing equalization technique is a question for future work, in order to grasp better the mechanics of the crosstalking and thus the mechanics also of the optical MIMO. Furthermore, the methodology for the approximation of the Gallager bound can be applied to include the uplink MU-MIMO [23] and the Amplify-and-Forward channels [13]. It should be noted that more general Gaussian channels, which do not have a known joint eigenvalue distribution can be analyzed in similar ways using the replica method [59]. Finally, the newly introduced MIMO fiber optic channel model, although it gives us the means to analyze the statistics of throughput in the corresponding channel in the presence of arbitrary level of crosstalk and mode dependent loss, it is however, also amenable to extensions, such as dispersion and nonlinear effects.

1.5 Publications

Journal Papers

- A. Karadimitrakis, A. L. Moustakas and R. Couillet, "Gallager Bound for MIMO Channels: Large-N Asymptotics," in *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 1323-1330, Feb. 2018.

- A. Karadimitrakis, A. L. Moustakas, and P. Vivo, "Outage Capacity for Optical MIMO Channel," in *IEEE Transactions on Information Theory*, vol.60, no. 7, pp. 4370-4382, 2014.

Conference Papers

- A. Karadimitrakis, R. Couillet, A. Moustakas and L. Sanguinetti, "The Gallager Bound in Fiber Optical MIMO", in *21th International ITG Workshop on Smart Antennas (WSA)* 2017, Berlin, Germany, 2017, pp. 1-8.

- A. Karadimitrakis, A. L. Moustakas, H. Hafermann and A. Mueller, "Optical Fiber MIMO Channel Model and its Analysis", in *IEEE International Symposium on Information Theory (ISIT)* 2016, Barcelona, 2016, pp. 2164-2168.
- A. Karadimitrakis, and A. L. Moustakas, "Large Deviation Approach to the Outage Optical MIMO Capacity," in *11th IEEE International Symposium on Information Theory (ISIT)*, 2014, pp. 966-970.
- A. Karadimitrakis, M. Debbah, and A. L. Moustakas, "Optical MIMO: Results and Analysis " in *2013 IEEE International Symposium on Wireless Communications Systems (ISWCS)* 2013, pp. 839-843.

CHAPTER 2

Telecommunications and Mathematical Background

2.1 Introduction in Modern Telecommunications

The "Classic" SISO Channel

Modern telecommunications owe a lot, if not their own existence to Claude Shannon, who in 1949 estimated the information capacity of a telecommunications channel [71]. The equation that describes the discrete memoryless channel (DMC), as seen in Fig.2.1, is

$$\mathbf{y} = \mathbf{h}\mathbf{x} + \mathbf{n}, \tag{2.1}$$

where $\mathbf{y}$,$\mathbf{h}$,$\mathbf{x}$ and $\mathbf{n}$ are the output, the complex impulse response, input and noise of the channel respectively, and the mutual information $I(X;Y)$ that denotes the amount of information that one random variable contains about the other random variable is

$$I(X;Y) = H(X) - H(Y|X). \tag{2.2}$$

As a "loan" from statistical physics, the average amount of information per symbol is called entropy and is measured in bits/symbol:

$$H(Z) = \sum_{i=1}^{N} p_i \log_2 \left[\frac{1}{p_i}\right] = -\sum_{z \in Z} p(z) \log_2 p(z), \tag{2.3}$$

which is the measure of uncertainty of a random variable $\mathbf{Z} \in \{z_1, z_2, \ldots z_N\}$ with probabilities $p(z) = \{p_1, p_2, \ldots p_N\}$. In other words, it characterizes the amount of information required on an average to describe the random variable. Hence, information

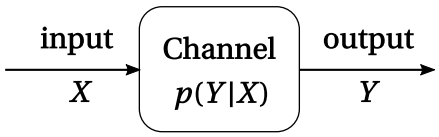


Figure 2.1: The discrete memoryless channel (DMC). The noise corrupts the input symbols independently.

channel capacity C is the maximum mutual information taken over all possible input

distributions $p(x)$ for given channel realization:

$$C = \max_{p(x)} I(X;Y|H). \tag{2.4}$$

SISO capacity evaluation

Let's assume the SISO flat fading channel (2.1). The noise is considered to be white,zero mean, Gaussian, independent of the input, with variance σ_n^2. Moreover, the receiver has perfect knowledge of the channel and the average power $P = \sigma_x^2$ at the transmitter is upper limited to P_t. That way the channel capacity reads

$$C = \max_{p(x), P \leq P_t} I(X;Y|H). \tag{2.5}$$

Since the input and output values are discrete in time, we have to use differential entropy H_d (in the literature the differential entropy is often depicted with h; here in order to avoid confuse with the channel coefficients, we chose to use the subscript d)

$$I(X;Y|H) = H_d(Y|H) - H_d(Y|X,H) = H_d(Y|H) - H_d(hX+N|X,H). \tag{2.6}$$

Since the receiver has perfect knowledge of the channel: $H_d(hX|X,H) = 0$ and the noise N is independent of the input X: $H_d(N|X,H) = H_d(N|H)$. Thus, for a given channel realization, we get

$$I(X;Y|H) = H_d(Y|H) - H_d(N|H). \tag{2.7}$$

If, on the other hand, the noise is zero mean, e.g. $N \sim \mathcal{N}(0, \sigma_n^2)$, the PDF of the noise is given by

$$f_N(n) = \frac{1}{\pi\sigma_n^2} e^{-\frac{n^2}{\sigma_n^2}}, \tag{2.8}$$

and the differential entropy of the noise for a given channel matrix, is given by

$$\begin{aligned} H_d(N|H) &= -\int f_N(n) \log_2 \left[f_N(n)\right] \mathrm{d}n \\ &= -\int f_N(n) \log_2 \left[\frac{1}{\pi\sigma_n^2} e^{-\frac{n^2}{\sigma_n^2}}\right] \mathrm{d}n \\ &= -\int f_N(n) \left[-\log_2(\pi\sigma_n^2) - \frac{n^2}{\sigma_n^2}\log_2(e)\right] \mathrm{d}n \\ &= \log_2(\pi\sigma_n^2) \int f_N(n)\mathrm{d}n + \frac{\log_2(e)}{\sigma_n^2} \int n^2 f_N(n)\mathrm{d}n \\ &= \log_2(\pi e \sigma_n^2). \end{aligned} \tag{2.9}$$

In the above calculations we used that $\int f_N(n)\mathrm{d}n = 1$ and $\int n^2 f_N(n)\mathrm{d}n = \sigma_n^2$. Therefore, the differential entropy is independent of the mean of the noise and for the aforementioned channel, the mutual information $I(X;Y|H)$ is maximized when the differential entropy $H_d(Y|H)$ is also maximized. It is proven, that the differential entropy attains the maximum value for Gaussian random variables, thus $H_d(Y|H) = \log_2(\pi e\sigma_y^2)$, where the received average power is

$$\sigma_y^2 = \mathbb{E}[Y^2] = \mathbb{E}[(hX+N)(hX+N)^*] = \sigma_x^2|h|^2 + \sigma_n^2, \tag{2.10}$$

and finally,the information channel capacity is

$$\begin{aligned} C &= H_d(Y|H) - H_d(N|H) \\ &= \log_2(\pi e\sigma_y^2) - \log_2(\pi e\sigma_n^2) \\ &= \log_2\left[\pi e(\sigma_x^2|h|^2 + \sigma_n^2)\right] - \log_2(\pi e\sigma_n^2) \\ &= \log_2\left(1 + \frac{\sigma_x^2}{\sigma_n^2}|h|^2\right) \\ &= \log_2\left(1 + \frac{P_t}{\sigma_n^2}|h|^2\right) \ \textit{bits per channel use.} \end{aligned} \tag{2.11}$$

In the above equation the base of the logarithm is 2 and the capacity is measured in bits per channel use, but usually for convenience we use as base the neperian number, thus we denote the logarithm simply as log and the capacity is measured in *nats per channel use*. In the last equation, the channel term h is considered to be a random variable, thus the capacity equation is also a random variable. Hence, we can distinguish the block-fading channel model:

Block Fading Channel Model

An important channel parameter is the time-scale of the variation of the channel. In other words, how fast the channel statistics change. The block-fading channel model assumes that the channel coefficients remain constant for a block of T consecutive transmitted symbols and change to an independent realization in the next block [7, 9, 89]. We can consider the parameter T as the channel's coherence time, or if we want to incorporate the idea of time slots, T is the number of time-frequency slots over which the channel remains constant. Let $\mathbf{y}$,$\mathbf{x}$, $\mathbf{z} \in \mathbb{R}^{TN}$ are the output, input and noise (i.i.d. $\sim \mathcal{Z}(0,1)$ sequences respectively. By arranging these components as $T \times N$ arrays we can write the channel equation as

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{Z}, \tag{2.12}$$

where $\mathbf{A} = \mathrm{diag}(\sqrt{\alpha_0}, \ldots, \sqrt{\alpha_{T-1}})$ is an $T \times T$ matrix whose diagonal elements are the fading amplitudes over the current slot. Input symbols on the same row of $\mathbf{X}$ experience the same fading coefficient.

Ergodic Capacity

As ergodic capacity we define the statistical average of the mutual information, where the expectation is taken over all channel realizations

$$C_{erg} = \mathbb{E}_H \left[\log \left(1 + \frac{P_t}{\sigma_n^2} |h|^2 \right) \right]. \tag{2.13}$$

Outage Capacity

As outage capacity we define the information rate (R) below which the instantaneous mutual information falls below a prescribed value of probability expressed as percentage:

$$P_{out} = \Pr \left[\log \left(1 + \frac{P_t}{\sigma_n^2} |h|^2 \right) < C_{out,q\%} \right] = q\%. \tag{2.14}$$

In other words, we can assume that the channel allows $\log(1+\rho|h|^2)$ *nats per channel use* through when the fading gain is h. For high SNR we can approximate

$$P_{out}(R) \approx \frac{2^R - 1}{\rho} \tag{2.15}$$

The MIMO Channel

Consider a MIMO channel with N transmitting and K receiving antennas. For simplicity the channel is considered to be frequency flat and the input-output equation reads

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{2.16}$$

where $\mathbf{y}$ is the $K \times 1$ received vector, $\mathbf{x}$ is the $N \times 1$ transmit vector, $\mathbf{n}$ is $K \times 1$ white, zero mean, circularly symmetric complex Gaussian noise vector with variance σ_n^2, $\mathbf{H}$ is the $K \times N$ channel matrix. In order to maintain the total average energy constrain, for the covariance matrix of $\mathbf{x}$, $\mathbf{R_{xx}} = \mathbb{E}\{\mathbf{x}\mathbf{x}^H\}$, must apply that $\mathrm{tr}\mathbf{R_{xx}} = P_t$. As seen for the SISO channel, the information capacity for a given channel realization is defined as

$$C = \max_{f(\mathbf{x})} I(\mathbf{x}; \mathbf{y}|\mathbf{H}), \tag{2.17}$$

where $f(\mathbf{x})$ is the probability distribution of $\mathbf{x}$. By defining $H_d(\mathbf{y})$ the differential entropy of $\mathbf{y}$ and $H_d(\mathbf{y}|\mathbf{x})$ the conditional differential entropy of $\mathbf{y}$ given $\mathbf{x}$, then

$$I(\mathbf{x}; \mathbf{y}|\mathbf{H}) = H_d(\mathbf{y}|\mathbf{H}) - H_d(\mathbf{y}|\mathbf{x}, \mathbf{H}) = H_d(\mathbf{y}|\mathbf{H}) - H_d(\mathbf{H}\mathbf{x} + \mathbf{n}|\mathbf{x}, \mathbf{H}). \tag{2.18}$$

We can assume that the channel is perfectly known to the receiver, i.e., $H_d(\mathbf{H}\mathbf{x}|\mathbf{x}, \mathbf{H}) = 0$ and since the noise is independent of the input, i.e., $H_d(\mathbf{n}|\mathbf{x}, \mathbf{H}) = H_d(\mathbf{n}|\mathbf{H})$, the equation for the mutual information becomes

$$I(\mathbf{x}; \mathbf{y}|\mathbf{H}) = H_d(\mathbf{y}|\mathbf{H}) - H_d(\mathbf{n}|\mathbf{H}). \tag{2.19}$$

Following the same procedure as for the SISO case we can write

$$H_d(\mathbf{n}|\mathbf{H}) = \log\det(\pi e\mathbf{R}_{\mathbf{nn}}) \tag{2.20}$$

$$H_d(\mathbf{y}|\mathbf{H}) = \log\det[\pi e(\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^{\mathrm{H}} + \mathbf{R}_{\mathbf{nn}})] \tag{2.21}$$

where $\mathbf{R}_{\mathbf{nn}}$ is the covariance matrix of the noise vector and the covariance matrix of $\mathbf{y}$ is

$$\mathbf{R}_{\mathbf{yy}} = \mathbb{E}\{(\mathbf{H}\mathbf{x}+\mathbf{n})(\mathbf{H}\mathbf{x}+\mathbf{n})^{\mathrm{H}}\} = \mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^{\mathrm{H}} + \mathbf{R}_{\mathbf{nn}}. \tag{2.22}$$

Therefore, we can write for the capacity

$$C = \log\det(\mathbf{I}_K + \frac{1}{\mathbf{R}_{\mathbf{nn}}}\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^{\mathrm{H}}). \tag{2.23}$$

For the case, where the noise is uncorrelated between the antenna branches, $\mathbf{R}_{\mathbf{nn}} = \frac{1}{\sigma_n^2}\mathbf{I}_K$, thus the capacity for MIMO flat fading channel is

$$C = \log\det(\mathbf{I}_K + \frac{1}{\sigma_n^2}\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^{\mathrm{H}}). \tag{2.24}$$

If the transmitter has no CSI, the optimal strategy is to distribute evenly the available transmit power at the corresponding antennas. Hence, $\mathbf{R}_{\mathbf{xx}} = \frac{P_t}{N}\mathbf{I}_N$. Analogously to the SISO case we can define also the ergodic and the outage capacity.

2.2 The Gallager Bound

For fading channels, the standard metric to characterize the performance of the link is the outage capacity [62], which corresponds to the throughput for a fixed outage probability. However, the outage capacity corresponds to infinitely long codewords. To deal with the realistic case of finite length codewords, Gallager [26] proposed a simple yet effective bound to the probability of error, as a function of rate and codeword length T. In its original version, as well as in more recent variations [69] this bound focused on single antenna links. There has been a number of extensions of the Gallager bound. For example, in [72] the Gallager's random coding error exponent was derived for MIMO Rayleigh block-fading channels, however, the expressions, while valid for all antenna sizes, are quite cumbersome to compute and analyze for any reasonably sized antenna array. In [88,90] expressions for Gallager's exponent were derived for space-time-block-coding (STBC) MIMO channels for non-Rayleigh fading models. However, STBC reception effectively corresponds to a single antenna link with increased diversity.

More recently, optimal bounds of the error probability for large but finite codewords have been established for single-link communications [32, 64]. These results are of a central–limit–theoretic nature, in that they are valid for large blocklengths with the rate converging to the ergodic rate at a fixed error probability. Similar results were obtained for MIMO systems in [37], where the number of antennas also goes to infinity at a fixed ratio with T. In contrast to the Gallager bound this approach does not capture the tails of the error probability, i.e. when the rate deviation per antenna from the ergodic rate is finite.

Gallager Bound Calculation for SISO Channel

Let us remind ourselves the discrete memoryless channel (DMC) depicted in Fig.2.1. The input of the channel is a random variable X who selects its value from a discrete limited set $\mathcal{X}$ whose cardinality is the number of the point in the used constellation. In an ideal channel the output is equal to the input but since in real life channels the input is corrupted by noise, the output can be different from the input with a given probability. All these probabilities $P(Y = y_i|X = x_i)$ are called transition probabilities.

Thus, for the Maximum Likelihood (ML) decoding for fixed codes, the error probability is upper bounded by

$$P_{e|m} \leq \sum_{\mathbf{y}} p_N(\mathbf{y}|\mathbf{x}^m) \left(\sum_{m' \neq m} \left(\frac{p_N(\mathbf{y}|\mathbf{x}^{m'})}{p_N(\mathbf{y}|\mathbf{x}^m)} \right)^{\lambda} \right)^{\rho} \quad \lambda, \rho \geq 0, \tag{2.25}$$

where $P_{e|m}$ is the block error probability conditioned on the transmitted length-N code word $\mathbf{x}^m$ $(m = 1, 2, \ldots, M)$, $\mathbf{y}$ is the observation vector (N components) and $p_N(\mathbf{y}|\mathbf{x})$ is the channel's transition probability measure for a block of length N.

For the calculation of the homonym bound, Gallager assumed codewords selected independently by a distribution $\mu_N(\mathbf{x})$. By applying the Jensen inequality $\mathbb{E}[x^\rho] \leq (\mathbb{E}[x])^\rho$ for $(0 < \rho < 1)$ and setting $\lambda = \frac{1}{1+\rho}$, the Gallager random coding bound results in

$$P_e \leq (M-1)^\rho \sum_{\mathbf{y}} \left(\sum_{\mathbf{x}} \mu_N(\mathbf{x}) p_N(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} \right)^{1+\rho}, \quad 0 \leq \rho \leq 1, \tag{2.26}$$

where P_e is the average ML decoding error probability and M is the number of codewords. For a momoryless channel it is

$$p_N(\mathbf{y}|\mathbf{x}) = \prod_{l=1}^{N} p(y_l|x_l), \tag{2.27}$$

and the corresponding input distribution is

$$\mu_N(\mathbf{x}) = \prod_{l=1}^{N} \mu(x_l). \tag{2.28}$$

By setting the code rate

$$R = \frac{\log M}{N}, \tag{2.29}$$

and

$$E(R, \mu) = \max_{0 \leq \rho \leq 1} (E_0(\rho, \mu) - \rho R), \tag{2.30}$$

with

$$E_0(\rho,\mu) = -\log\left(\sum_y\left(\sum_x \mu(x)p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right), \tag{2.31}$$

the Gallager random coding bound reads

$$P_e \leq e^{-NE(R,\mu)}. \tag{2.32}$$

Sphere-Packing Bound [70]

A fixed composition code over a k-alphabet has the property that the number of occurrences of the k symbols within a codeword is the same for each codeword. Let $\mathcal{C}$ be a fixed composition code of M codewords and block length N and its transmission takes place over a DMC. Assume that the DMC is specified by the set of transition probabilities $P(j|k)$ where $k \in \{1,\ldots,K\}$ and $j \in \{1,\ldots,J\}$ designate the channel input and output alphabets respectively. Assume that the code $\mathcal{C}$ forms a set of M codewords of length N and consider an arbitrary list decoder where the size of the list is limited to L. Then the maximal error probability satisfies

$$P_{e,max} \geq e^{-N\left(E_{sp}\left(R-\frac{\log 4}{N}-\epsilon\right)+\sqrt{\frac{8}{N}}\log\left(\frac{e}{\sqrt{P_{min}}}\right)+\frac{\log 4}{N}\right)}, \tag{2.33}$$

where $R = \frac{\log\left(\frac{M}{L}\right)}{N}$ is the code rate, P_{min} designates the smallest non-zero transition probability, the parameter ϵ is an arbitrarily small positive number, and the function E_{sp} is given by

$$E_{sp}(R) = \sup_{\rho>0}(E_0(\rho) - \rho R) \tag{2.34}$$

$$E_0(\rho) = \max_{\boldsymbol{\mu}} E_0(\rho,\boldsymbol{\mu}) \tag{2.35}$$

$$E_0(\rho,\boldsymbol{\mu}) = -\log\left[\sum_{j=1}^{J}\left(\sum_{k=1}^{K}\mu_k P(j|k)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right]. \tag{2.36}$$

The maximum in the RHS of (2.35) is taken over all probability vectors $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_K)$, i.e. over all $\boldsymbol{\mu}$ with K non-negative components summing to 1.

Gallager Exponent with Fixed Power Constraints

Let X denote the input space to a DMC and let $f(x)$ be a real-valued function on the input letters. We impose the constraint that $\mathbb{E}\{f(x)\} \leq \mathcal{E}$, where $\mathcal{E}$ is a fixed value. For example, by choosing $f(x) = x^2$ then we are describing an energy constraint. From a coding point of view, we can write the constraint as a demand that each codeword $\mathbf{x} = (x_1,\ldots,x_N)$ satisfies

$$\sum_{n=1}^{N} f(x_n) \leq N\mathcal{E}. \tag{2.37}$$

That way, we will construct an ensemble of codes in which each codeword satisfies the equation above. Denoting $P(j|k)$ the transition probabilities for a DMC with input alphabet $0, \ldots, K-1$ and output alphabet $0, \ldots, J-1$, and $\mu(k)$ the input distribution, then

$$\sum_k \mu(k) f(k) \leq \mathcal{E}. \tag{2.38}$$

Further, let $\mu_N(\mathbf{x})$ be the probability distribution of N channel inputs given by

$$\mu_N(\mathbf{x}) = c^{-1} \phi(\mathbf{x}) \prod_{n=1}^{N} \mu(x_n), \tag{2.39}$$

where

$$\phi(\mathbf{x}) = \begin{cases} 1; & \text{for } N\mathcal{E} - \delta < \sum_n f(x_n) \leq N\mathcal{E} \\ 0; & \text{otherwise} \end{cases} \tag{2.40}$$

$$c = \sum_{\mathbf{x}} \phi(\mathbf{x}) \prod_{n=1}^{N} \mu(X_n) \tag{2.41}$$

and δ is an arbitrary positive number. Finally, consider an ensemble of codes with M codewords of block length N in which the codewords are independently chosen with joint pdf $\mu_N(\mathbf{x})$, P_N the joint pdf of the transition probabilities, then the average error probability for each message, $1 \leq m \leq M$ over the ensemble of codes, is upper bounded for all ρ, $0 \leq \rho \leq 1$ by

$$P_{e|m} \leq (M-1)^{\rho} \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} \mu_N(\mathbf{x}) P_N(\mathbf{y}|\mathbf{x})^{1/(1+\rho)} \right]^{1+\rho}. \tag{2.42}$$

By choosing an appropriate upper bound for $\phi(\mathbf{x})$ we can write

$$P_{e|m} \leq \left(\frac{e^{r\delta}}{c} \right)^{1+\rho} e^{-N(E_0(\rho, \boldsymbol{\mu}, r) - \rho R)} \tag{2.43}$$

$$E_0(\rho, \boldsymbol{\mu}, r) = -\log \sum_j \left[\sum_k \mu(k) e^{r(f(k)-\mathcal{E})} P(j|k)^{1/(1+\rho)} \right]^{1+\rho} \tag{2.44}$$

where $r, \delta \geq 0$ are arbitrary parameters and $M = \lceil e^{NR} \rceil$.

2.3 Random Matrix Theory (RMT)

The origins of RMT can be found back in 1928 in the work of Wishart [87] and James (1954-1964) [40] but it was Wigner in '50s who used RMT in Nuclear physics and

showed how helpful can be this mathematical tool. In particular, Wigner proposed that the fluctuations in positions of compound nuclei resonances can be described in terms of statistical properties of eigenvalues of very large symmetric matrices with i.i.d. entries.

In RMT we calculate the statistics of the eigenvalue of matrices whose entries are drawn randomly from various probability distributions. There are three classical probability distributions (random matrix ensembles): The Gaussian Orthogonal Ensemble (GOE), the Gaussian Symplectic Ensemble (GSE) and the Gaussian Unitary Ensemble (GUE). In summary, the properties of each ensemble are

- GOE: The Hamiltonian matrix $\mathbf{H}$ of the system is a Hermitian symmetric random matrix with its elements $H_{jk},\ j \geq k$ statistically independent. The joint probability function of the eigenvalues is invariant under all real orthogonal transformations of $\mathbf{H}$, thus modeling Hamiltonians with time-reversal symmetry.

- GUE: The Hamiltonian matrix $\mathbf{H}$ of the system is a Hermitian random matrix, with its diagonal elements H_{jj} and the real and imaginary parts of its off-diagonal elements $H_{jk},\ j > k$, are statistically independent. Moreover, the The joint probability function of the eigenvalues is invariant under all unitary transformations of $\mathbf{H}$, thus modeling Hamiltonians without time-revearsal symmetry.

- GSE: The Hamiltonian matrix $\mathbf{H}$ of the system is a Hermitian self-dual random matrix, with its diagonal elements H_{jj} and the four quaternionic components of its off-diagonal elements $H_{jk},\ j > k$, are statistically independent and the joint probability function of the eigenvalues is invariant under all symplectic transformations of $\mathbf{H}$, thus modeling Hamiltonians with time-reversal symmetry but no rotational symmetry.

Marčenko-Pastur Law

The Marčenko-Pastur law describes the asymptotic behavior of singular values of large rectangular matrices, see Fig. 2.2.

Theorem 2.1. *If* $\mathbf{X}$ *is a* $M \times N$ *random matrix which entries are i.i.d. random variables with zero mean and variance* $\sigma^2 < \infty$*, then the empirical distribution of eigenvalues of* $\mathbf{X}\mathbf{X}^\dagger$ *converges almost surely as* $M, N \to \infty$ *with* $\frac{M}{N} \to \beta$*:*

$$f(x) = \left(1 - \frac{1}{\beta}\right)^+ \delta(x) + \frac{\sqrt{(x-a_0)^+(b_0-x)^+}}{2\pi\beta x}, \tag{2.45}$$

where $(z)^+ = \max(0, z)$ *and*

$$a_0 = (1 - \sqrt{\beta})^2 \qquad b_0 = (1 + \sqrt{\beta})^2. \tag{2.46}$$

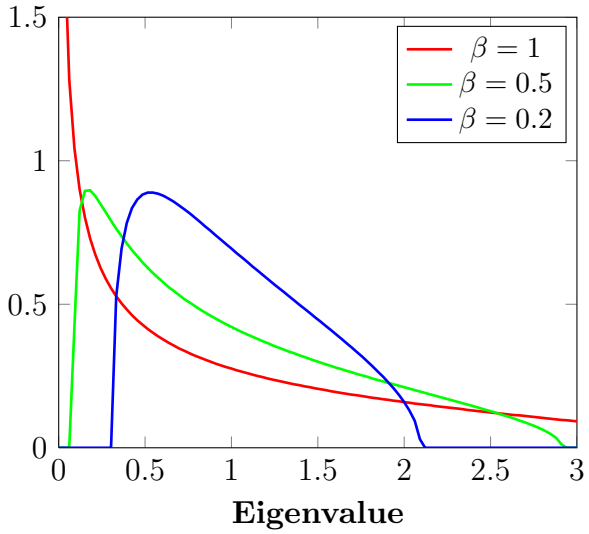


Figure 2.2: Marčenko-Pastur density function

Wishart Matrices

Assume a Gaussian real $M \times N$ matrix $\mathbf{X}$ with i.i.d. columns, each with zero mean and covariance $\mathbf{\Sigma}$, and $N \geq M$. The real matrix $\mathbf{X}\mathbf{X}^T$ is called Wishart and its distribution is indicated as $\mathcal{W}_M(N, \mathbf{\Sigma})$. Denoting $\Gamma_m(a) = \pi^{m(m-1)/4} \prod_{i=1}^{m} \Gamma(a-(i-1)/2)$, the joint pdf of the ordered eigenvalues $\lambda_i \geq 0, \quad i \in \{1, \ldots, M\}$ of the real Wishart matrix with $\mathcal{W}_M(N, \mathbf{I})$ is

$$f(x_1, \ldots, x_M) = K \prod_{i=1}^{M} e^{-\frac{x_i}{2}} x_i^{\frac{N-M-1}{2}} \prod_{i<j}^{M} (x_i - x_j), \tag{2.47}$$

where K is a normalizing constant given by

$$K = \frac{\pi^{\frac{M^2}{2}}}{2^{\frac{MN}{2}} \Gamma_M(\frac{N}{2}) \Gamma_M(\frac{M}{2})} \tag{2.48}$$

The marginal pdf of the unordered eigenvalues z of the Wishart matrix $\frac{1}{N}\mathbf{X}\mathbf{X}^\dagger$ is

$$g_{M,N}(zN) = \frac{1}{\beta} \sum_{k=0}^{M-1} \frac{k!}{(k+N-M)!} [L_k^{N-M}(zN)]^2 (zN)^{N-M} e^{-zN}, \tag{2.49}$$

where $L_m^n(x)$ is the Laguerre polynomials. In Fig. 2.3 we see that as we increase the dimensions of the system, the marginal pdf converges to the Marčenko-Pastur limit.

Haar Measure

In the case of no time reversal symmetry, the unitary matrices form the group $U(N)$. We define the Haar measure as the unique uniform measure on $U(N)$. In this work

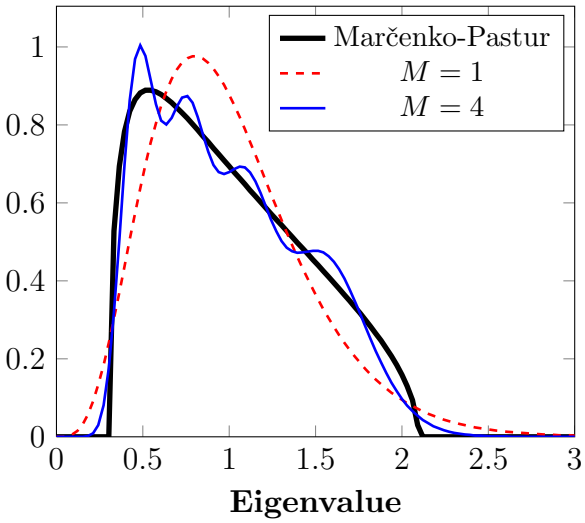


Figure 2.3: Marginal pdf of thε unordered eigenvalues of an $M \times N$, $N \geq M$, Wishart matrix for $\beta = 0.2$. The convergence to the Marčenko-Pastur limit increases as we increase the dimensions of the matrix.

we are interested in the left-translation-invariant property of a Haar measure which states that for a countably additive, nontrivial measure μ on the Borel subsets of a locally compact topological group G it is

$$\mu(gS) = \mu(S), \tag{2.50}$$

for every $g \in G$ and all Borel sets $S \subset G$.

Random Matrices and Telecommunications

A typical telecommunications channel, both wireless and optical, is described by the equation

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{2.51}$$

where $\mathbf{x}$ is the K-dimensional input vector, $\mathbf{y}$ is the N-dimensional output vector, $\mathbf{n}$ is the additive Gaussian noise, which for memoryless channel, the components are independent complex Gaussian random variables with the same variance $\frac{\sigma^2}{2}$, i.e. circularly distributed, and $\mathbf{H}$ is the $N \times K$ complex random matrix describing the channel.

The empirical cumulative distribution function of the eigenvalues of an $N \times N$ Hermitian matrix $\mathbf{A}$ is defined as

$$f(x) = \frac{1}{N} \sum_{i=1}^{N} \Theta(\lambda_i(\mathbf{A}) \leq x), \tag{2.52}$$

where $\lambda_1(\mathbf{A}), \ldots, \lambda_N(\mathbf{A})$ are the eigenvalues of $\mathbf{A}$ and $\Theta(x)$ is the indicator (step) function.

In (2.51), if the channel is known by the receiver and the input $\mathbf{x}$ is Gaussian with independent and identically distributed entries, the normalized mutual information conditioned on $\mathbf{H}$ is, as we saw in the previous section,

$$\begin{aligned} I(\mathbf{x};\mathbf{y}|\mathbf{H}) &= \log\det\left(\mathbf{I}+\rho\mathbf{H}\mathbf{H}^\dagger\right) \\ &= \sum_{i=1}^{N}\log\left[1+\rho\lambda_i(\mathbf{H}\mathbf{H}^\dagger)\right] \\ &= N\int_0^\infty \log\left[1+\rho f(x)\right]\mathrm{d}x, \end{aligned} \tag{2.53}$$

where ρ is the transmitted signal-to-noise ratio and $f(x)$ the pdf of the eigenvalues.

Multi-Antenna Channel Example [81]

Let us consider a singe-user channel with n_T transmitting antennas and n_R receiving. Thus, for frequency-flat fading assumption, the entries of $\mathbf{H}$ represent the fading coefficients between each n_{Ti} and each n_{Ri}, typically normalized, zero-mean Gaussian such that

$$\mathbb{E}\left[\mathrm{tr}\left\{\mathbf{H}\mathbf{H}^\dagger\right\}\right] = n_R. \tag{2.54}$$

If all antennas are co-polarized the entries of $\mathbf{H}$ are identically distributed and the resulting variance is $\frac{1}{n_T}$. Hence, the signals transmitted by different antennas can be correlated. Normalized by its energy per dimension, the input covariance is

$$\boldsymbol{R}_{\mathbf{xx}} = \mathbb{E}[\mathbf{x}\mathbf{x}^\dagger], \tag{2.55}$$

and normalized, so that $\mathbb{E}[\mathrm{tr}\{\boldsymbol{R}_{\mathbf{xx}}\}] = n_T$. Furthermore, we can decompose the input covariance in its eigenvectors: $\boldsymbol{R}_{\mathbf{xx}} = \mathbf{V}\mathbf{P}\mathbf{V}^\dagger$, where each eigenvalue represents the normalized power allocated to the corresponding signaling eigenvector.

In order to achieve capacity, $\boldsymbol{R}_{\mathbf{xx}}$ must be properly determined depending on the channel state information (CSI), which is available to the transmitter. So, in any case of full, partial or zero CSI knowledge respectively, the capacity per receive antenna is given by the maximum over $\boldsymbol{R}_{\mathbf{xx}}$ of the Shannon transform of the averaged empirical distribution of $\mathbf{H}\boldsymbol{R}_{\mathbf{xx}}\mathbf{H}^\dagger$ ($\mathcal{V}_{\mathbf{H}\boldsymbol{R}_{\mathbf{xx}}\mathbf{H}^\dagger}$):

$$C(\rho) = \max_{\boldsymbol{R}_{\mathbf{xx}}:\mathrm{tr}\boldsymbol{R}_{\mathbf{xx}}=n_T} \mathcal{V}_{\mathbf{H}\boldsymbol{R}_{\mathbf{xx}}\mathbf{H}^\dagger}(\rho). \tag{2.56}$$

2.4 The Replica Method

The replica method is a very useful and effective mathematical tool, although yet, there is no rigorous mathematical proof. The method originates itself from statistical

physics, where the partition function $\mathcal{Z}$ depends also on some random parameters, hence $\mathcal{Z}$ (or $\log \mathcal{Z}$) becomes a random variable too. Moreover, in the thermodynamic limit of $N \to \infty$, $\frac{1}{N} \log \mathcal{Z} \xrightarrow{a.s.} \frac{1}{N} \langle \log \mathcal{Z} \rangle$ (self-averaging property). Therefore, if we are examining a system with many particles, the free energy of any particle, for a typical realization of the random parameters that apply on the system, becomes the computation of $\langle \log \mathcal{Z} \rangle$. But, in general that computation is very hard. Exactly at this point, replica method shows its usefulness. Because the partition function Z is a sum, we do not have many things to say for $\langle \log \mathcal{Z} \rangle$. But, thanks to the simple relation

$$\langle \log \mathcal{Z} \rangle = \lim_{n \to 0} \frac{\langle \mathcal{Z}^n \rangle - 1}{n} = \lim_{n \to 0} \frac{\log \langle \mathcal{Z}^n \rangle}{n} \tag{2.57}$$

we can interchange the expected value of log-sum to the expected value of integer moments. To do this, we have to use the trick of doing the aforementioned computation of a general, positive integer n and acquire the expression as function of integer n. But since our system consists of many particles, we can assume that the expression is also valid for n being a real variable. Finally, in order to compute the free-energy of any particle, we take the limit $n \to 0$.

In modern telecommunications the goal for a successful demodulation at the receiver is to minimize the energy function of a given received signal. This is an optimization problem which can be solved by using Lagrangian form: the free energy is the object to be minimized and the system temperature is the Lagrange multiplier.

In most cases the function in the exponent is multivariate, therefore the replica method dictates the computation of the extremum of a multivariate function for an arbitrary number of arguments and unless we exploit some symmetries of the optimization problem, this is an impossible task. This replica symmetry means that one can safely assume that the optimization is feasible if all variables take on the same value, thus reducing the multivariate problem to a single variate one. One way to circumvent the problem of the existence of a replica symmetry is to assume there is one and prove that is a correct, thus sufficient solution. In many cases, we reach true solutions which show physical impracticality, e.g. negative entropy. This is called replica symmetry breaking and there has been an extensive research in literature on this [22, 56, 60]. For the rest of this work, however, replica symmetry breaking is a too advanced issue.

Example - SISO Channel Capacity

In order to better understand the usage and the usefulness of the replica method in the telecommunications, we are going to provide next with an example. Let us remind ourselves the SISO noisy channel (2.1), where the mutual information is given by

$$I(Y;X) = H(Y|H) - H(Y|X,H) \tag{2.58}$$

where X, Y and H are the random variables for the input, output and channel respectively. From the definition of the entropy we can write

$$H(Y|X,H) = -\int \left[\int p(y|x,h) \log p(y|x,h) \mathrm{d}y \right] p(x) \mathrm{d}x \mathrm{d}h \tag{2.59}$$

$$H(Y) = -\int p(y)\log p(y)\mathrm{d}y \tag{2.60}$$

Using the replica trick we can write for the expectation of the entropy

$$\begin{aligned}\mathbb{E}\{H(Y)\} &= -\lim_{n\to 0}\frac{\partial}{\partial n}\log\mathbb{E}\left\{\int p(y)[p(y)]^n\mathrm{d}y\right\}\\ &= -\lim_{n\to 0}\frac{\partial}{\partial n}\log\mathcal{Z}_n,\end{aligned} \tag{2.61}$$

and for the conditional entropy

$$\mathbb{E}\left\{H(Y|X,H)\right\} = -\lim_{n\to 0}\frac{\partial}{\partial n}\log\mathbb{E}_y\left\{\prod_{a=1}^{n+1}\iint p(y|x_a,h)p(x_a)p(h)\mathrm{d}y\mathrm{d}x_a\right\}. \tag{2.62}$$

Further in our calculations we can use the Gaussian input and the white, zero mean, Gaussian noise (z) assumption:

$$p(y|x_a,h) = \frac{1}{\sqrt{\pi\sigma_z^2}}e^{-\frac{(y-x_a)^2}{\sigma_z^2}} = \frac{1}{\sqrt{\pi\sigma_z^2}}e^{-\frac{z^2}{\sigma_z^2}} \tag{2.63}$$

$$p(y) = \frac{1}{\sqrt{\pi\sigma_y^2}}e^{-\frac{y^2}{\sigma_y^2}}. \tag{2.64}$$

But we can reduce the computations if we remind ourselves (2.7). Thus,

$$\begin{aligned}\mathbb{E}\{H(Y)\} &= -\lim_{n\to 0}\frac{\partial}{\partial n}\log\mathbb{E}\left\{\int\left(e^{-\frac{y^2}{\sigma_y^2}}\right)^{n+1}\mathrm{d}y\right\}\\ &= -\lim_{n\to 0}\frac{\partial}{\partial n}\log\frac{1}{\sqrt{n+1}}\left(\frac{1}{\pi\sigma_y^2}\right)^n\\ &= \frac{1}{2}+\log\left(\frac{\pi}{\sigma_y^2}\right)\end{aligned} \tag{2.65}$$

and because of (2.10)

$$\mathbb{E}\{H(Y)\} = \frac{1}{2}+\log\left(\frac{\pi}{\sigma_x^2|h|^2+\sigma_z^2}\right); \tag{2.66}$$

same way

$$\mathbb{E}\{H(Z)\} = \frac{1}{2}+\log\left(\frac{\pi}{\sigma_z^2}\right) \tag{2.67}$$

and finally we end up at the well-known equation for the capacity

$$C = \mathbb{E}\{H(Y)\} - \mathbb{E}\{H(Z)\} = \log\left(1+\frac{\sigma_x^2}{\sigma_z^2}|h|^2\right). \tag{2.68}$$

2.5 The Saddle Point Approximation- Method of Steepest Descend

Let us consider an integral in the complex plane of the form

$$I(N) = \int_a^b g(z) e^{Nf(z)} \mathrm{d}z, \tag{2.69}$$

where f is an analytic complex function.

We expect that the integral to be dominated by the highest stationary points of f. To better understand this, let us write $f = u + iv$ and thus we expect the integral to be dominated by points where u is maximized and also we require that v is stationary so that the oscillating contributions do not cancel. Hence, $f'(z) = 0$.

Actually, the only extrema possible for f are *saddle points*. To see this, recall that for $f = u + iv$ the Cauchy-Riemann conditions imply that both u, v satisfy Laplace's equation i.e.,

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0. \tag{2.70}$$

In other words, at a stationary point, if e.g., $\frac{\partial^2 u}{\partial x^2} > 0$ then $\frac{\partial^2 u}{\partial y^2} < 0$. We call these stationary points as *saddle points*, see Fig. 2.4.

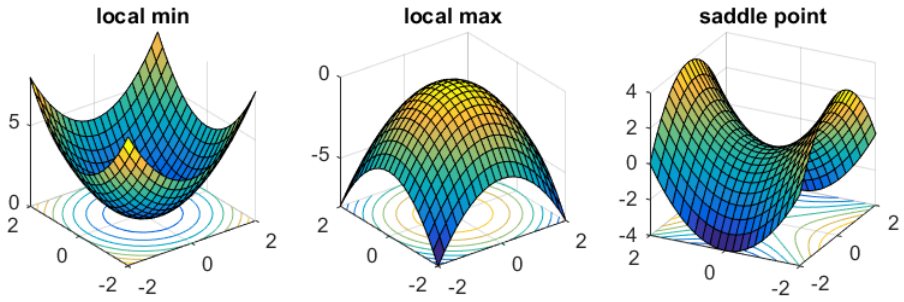


Figure 2.4: Graphical explanation of a saddle point.

Assume z_0 the highest saddle point and using Taylor expansion near z_0 we can write

$$f(z) \simeq f(z_0) + \frac{1}{2} f''(z_0)(z - z_0)^2, \tag{2.71}$$

where we took into account that $f'(z_0) = 0$. Thus, $I(N)$ becomes

$$I(N) \simeq g(z_0) e^{Nf(z_0)} \int e^{\frac{1}{2} N |f''(z_0)| e^{i\theta} r^2 e^{2i\phi}} e^{i\phi} \mathrm{d}r. \tag{2.72}$$

Further, we may choose $\phi = \frac{\pi - \theta}{2}$ and we can re-write

$$I(N) \simeq g(z_0) e^{Nf(z_0)} e^{i\phi} \int e^{-\frac{1}{2} N |f''(z_0)| r^2} \mathrm{d}r. \tag{2.73}$$

Extending the limits of the above integral to infinity and performing the gaussian integral we get

$$I(N) \simeq g(z_0)e^{Nf(z_0)}e^{i\phi}\left(\frac{2\pi}{N|f''(z_0)|}\right)^{1/2}. \tag{2.74}$$

The equation above is called *Saddle Point Approximation*. The path we chose to make the integration gaussian i.e., $\phi = \frac{\pi-\theta}{2}$ corresponds to the path that descends most steeply from the saddle point, hence the name *Steepest Descend*.

2.6 Constrained Optimization and Lagrange Multipliers

Nowadays, there is an increasing interest in optimization problems, mainly driven by new found applications such as machine learning and artificial intelligence. An optimization problem is related to minimizing or maximizing a function, e.g. error function and likelihood respectively with respect to some variable $\mathbf{x}$. If there are constraints in the possible values of $\mathbf{x}$, then the Lagrange Multipliers method can restrict the search of solutions in the feasible set of values of $\mathbf{x}$. Hence,

$$\begin{aligned} \mathbf{x}^* &= \underset{\mathbf{x}}{\operatorname{argmin}}\, f(\mathbf{x}) \\ \text{subject to } h_i(\mathbf{x}) &= 0, \forall i = 1, \ldots, m. \end{aligned} \tag{2.75}$$

The Lagrange Multipliers method dictates the insertion of both cost function and constraints in a single minimization problem and multiply each constraint by a factor of λ_i. Hence, the optimization problem becomes

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}}\, \mathcal{L}(\mathbf{x}, \lambda) = \underset{\mathbf{x}}{\operatorname{argmin}}\, f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i h_i(\mathbf{x}), \tag{2.76}$$

where $\mathcal{L}$ is the Lagrangian and λ the Lagrange multiplier. The search for the extrema in the last equation leads us to the following optimality condition

$$\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \lambda) = \nabla_{\mathbf{x}} f(\mathbf{x}) + \sum_{i} \lambda_i \nabla_{\mathbf{x}} h_i(\mathbf{x}) = 0. \tag{2.77}$$

Finally, since λ_i are "dummy" variables and irrelevant to the initial problem, we can set

$$\frac{\partial \mathcal{L}(\mathbf{x}, \lambda)}{\partial \lambda_i} = 0, \tag{2.78}$$

which makes the problem determined and thus solvable. A simple explanation of the method of Lagrange Multipliers can be seen in Fig. 2.5. The problem is to minimize $f(x, y)$ subject to $g(x, y) = c$. The necessary conditions is $\nabla f = \lambda \nabla g$. Treat ∇f as the direction of the movement of f and ∇g as the normal vector of the curve $g(x, y) = c$. One can observe the optimal value is attained when ∇f is parallel to ∇g.

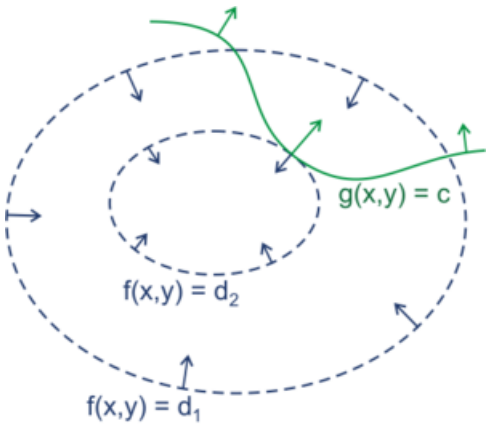


Figure 2.5: How Lagrange Multipliers Method works.

2.7 The Large Deviation (LD) Method

The LD method comes in hand when we have to deal with the probabilities of rare events that are exponentially small as a function of some parameter [79]. In other words, the LD method gives a description of the asymptotic behavior of remote tails of sequences of probability distributions. For example, assume a sum of real random variables

$$S_N = \frac{1}{N}\sum_{i=1}^{N} X_i. \tag{2.79}$$

This is also called sample mean and we want to compute the pdf of S_N in the case where the random variables are i.i.d.

$$p(X_1, \ldots, X_N) = \prod_{i=1}^{N} p(X_i). \tag{2.80}$$

In case of Gaussian pdf then

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad , \quad x \in \mathbb{R} \tag{2.81}$$

where $\mu = \mathrm{E}[X]$ is the mean of X and $\sigma^2 = \mathrm{E}[(X-\mu)^2]$ is the variance. In order to compute the pdf of S_N, we sum the pdf of all the values (realizations $x_1, \ldots, x_N \in \mathbb{R}^N$) of $X_1, \ldots, X_N$ such that $S_N = s$, using the Dirac's delta funtion $\delta(x)$.

$$p_{S_N} = \int_{\mathbb{R}} dx_1 \ldots \int_{\mathbb{R}} dx_N \delta(\sum_{i=1}^{N} x_i - Ns) p(x_1, \ldots, x_N). \tag{2.82}$$

From (2.82) we can calculate the explicit expression for p_{S_N}, i.e. by using the method of generating functions :

$$p_{S_N} \approx e^{-NI(s)}, \tag{2.83}$$

where $I(s) = \frac{(s-\mu)^2}{2\sigma^2}$, $s \in \mathbb{R}$. p_{S_N} is an exponentially decaying function of N.

So, we observe that $I(s)$ controls the exponential decay rate of p_{S_N} to zero. It is $I(s) \geq 0$ and only $I(s) = 0$ when $s = \mu$. Therefore, as $N \to 0$, the normalized S_N concentrates more and more around $s = \mu$, e.g. S_N converges in probability.

The LD Principle

The rigorous mathematical definition of the LD principle is too technical for this present work, as it involves concepts of topology and measure theory [18]. Therefore, for simplicity we will say that a random variable S_n or its pdf $p(S_n)$ satisfies the LD principle if the following limit exists [79]

$$\lim_{N\to\infty} \left(-\frac{1}{N} \log p_{S_N} \right) = I(s) \tag{2.84}$$

where $I(s)$ is called rate function. Because p_{S_N} is dominated for a large N by a decaying exponential, the exact pdf of S_N can be written as

$$p_{S_N} = e^{-NI(s)+o(N)} \tag{2.85}$$

where $o(N)$ is for any correction term that is sub-linear in N. Taking the LD limit of the last equation,

$$\lim_{N\to\infty} \left(-\frac{1}{N} \log p_{S_N} \right) = I(s) - \lim_{N\to\infty} \left(-\frac{o(N)}{N} \right) = I(s). \tag{2.86}$$

In other words, the LD limit is the one that we need to retain the dominant exponential term in p_{S_N}, while we can discard all the other terms.

Varadhan's Theorem

In the LD method the Varadhan's theorem is of big importance. Let us assume the functional expectation of the form

$$W_N[f] = \mathbb{E}\{e^{Nf(S_N)}\} = \int_{\mathbb{R}} p_{S_N}(s) e^{Nf(s)} ds, \tag{2.87}$$

where f is some function of the random variable S_N. Assume also that S_N satisfies the LD principle with rate function $I(s)$:

$$W_N[f] \approx \int_{\mathbb{R}} e^{N[f(s)-I(s)]} ds. \tag{2.88}$$

We have omitted all the sub-exponential correction terms in N. The last integral can be identified as the Laplace integral and it is dominated for large N by its largest integrand when it is unique. Thus, we have

$$W_N[f] \approx e^{N \sup_s [f(s)-I(s)]}. \tag{2.89}$$

If you look closely to the last equation, you can identify the saddle point approximation that we saw in a previous section. So, for a functional of the form

$$\lambda[f] = \lim_{N\to\infty} \frac{1}{N} \log W_N(f), \tag{2.90}$$

we can get

$$\lambda[f] = \sup_{s\in\mathbb{R}} \{f(s) - I(s)\}. \tag{2.91}$$

(2.91) is the well known Varadhan's Theorem [82], [79]. This theorem is valid not just for i.i.d. random variables but also for random vectors and even for random functions.

Small and Large Deviations

We can get a lot of information about the pdf of a random variable, just by looking at the LD principle of it. For example we know that the $p(S_N)$ concentrates on certain point corresponding to the zeros of the rate function $I(s)$. These points correspond to the most probable values of S_N as $N \to \infty$ and can be shown that they are related to the law of large numbers (see Sec. 3.5.7 of [79]).

Many times we need to know not only that S_N converges in probability to some values but also the probability these values are not away but close to the typical ones. Let us consider such a typical value $s^\star$ and we can expand $I(s)$ according to Taylor series around that point:

$$I(s) = I(s^\star) + I'(s^\star)(s - s^\star) + \frac{I''(s^\star)}{2}(s - s^\star)^2 + \dots \tag{2.92}$$

But, since $s^\star$ corresponds to a zero of $I(s)$, the first two terms in (2.92) are also zero. In other words, the small deviations of S_N around the typical values are Gaussian distributed :

$$p_{S_N}(s) \approx e^{-NI''(s^\star)(s-s^\star)^2/2}. \tag{2.93}$$

Therefore, we can identify inside the LD theory, the presence of the Central Limit theorem and we get information not only about the small deviations of S_N, but also about the large deviations far away from its typical values.

LD Example

In order to understand better the LD theory, let us consider a simple example. Suppose we have a biased coin, such that for each toss the probability optaining "heads" is "p". If we assign the value 1 to the outcome "heads" (each outcome is denoted by $X_i,\ i = 1, 2, \ldots$) and 0 to the outcome "tails", then the outcome after N trials is

$$M_N = \frac{1}{N} \sum_{i=1}^{N} X_i. \tag{2.94}$$

As $N \to \infty$, it is expected that $M_N \to p$. However, for large enough N, we want to calculate the probability that M_N differs from p by at least x, where x is a pre-assigned fraction, less than unity. By using the LD method, we can safely calculate how the tail of the probability distribution of X_i behaves, as long as X_i are bounded and i.i.d. random variables, i.e.,

$$\begin{aligned} P(M_N > x) &\approx e^{-NI(x)} \text{ for } x > p \\ P(M_N < x) &\approx e^{-NI(x)} \text{ for } x < p. \end{aligned} \tag{2.95}$$

Further, we can use Varadhan's theorem that ensures that the sequence M_N satisfies a large deviation principle, i.e., $P(M_N \approx x) \sim e^{-NI(x)}$ and further using the Chernoff's bound for a random variable X and for every $t > 0$

$$\Pr(X \geq a) = \Pr(e^{tX} \geq e^{ta}) \leq \frac{\mathbb{E}\{e^{tX}\}}{e^{ta}} \tag{2.96}$$

and hence for the Bernoulli distribution under consideration

$$\mathbb{E}\{e^{tX_i}\} = 1 + p(e^t - 1) \leq e^{p(e^t - 1)}; \tag{2.97}$$

and minimizing w.r.t t, we have for the biased coin toss example

$$I(x) = x \log \frac{x}{p} + (1 - x) \log \frac{1 - x}{1 - p}. \tag{2.98}$$

CHAPTER 3
The Optical Channel

3.1 The Scattering Matrix Approach

We consider a single-segment N-channel lossless optical fiber system, with $N_t \leq N$ transmitting channels excited and $N_r \leq N$ receiving channels coherently excited in the input (left) and output (right) side of the fiber, as seen in Figure 3.1. The propagation through the fiber may be analyzed through its $2N \times 2N$ scattering matrix given by [4, 86]

$$\mathbf{S} = \begin{bmatrix} \mathbf{R}_\ell & \mathbf{T}_\ell \\ \mathbf{T}_r & \mathbf{R}_r \end{bmatrix}. \tag{3.1}$$

This matrix "connects" the N left (ℓ) with the N right (r) modes of the fiber. The kth column (for $k = 1, \ldots, N$) of $\mathbf{R}_\ell$ correspond to the reflection coefficients at the left of the N modes of the fiber when a unit amplitude signal is inserted from the kth left input of the fiber. The same input signal results to transmission through the fiber, with transmission coefficients at the right hand of the fiber given by the kth column of $\mathbf{T}_r$. In an analogous fashion the kth columns of $\mathbf{R}_r$ and $\mathbf{T}_\ell$ correspond to the right-reflection and left-transmission coefficients when a unit amplitude signal is inserted from the kth right input of the fiber. The input signal is represented by an $2N$ dimensional vector, in which the first N entries correspond to the amplitudes of the left-incoming signal and the remaining entries to the amplitudes of the right-incoming signal.

We now assume that the signal propagates through the above N channels. In this case, for any input $\mathbf{v}_{in}$ the total input power into the fiber is equal to the total output power, i.e.

$$\mathbf{v}_{in}^\dagger \mathbf{v}_{in} = \mathbf{v}_{out}^\dagger \mathbf{v}_{out} = \mathbf{v}_{in}^\dagger \mathbf{S}^\dagger \mathbf{S} \mathbf{v}_{in} \tag{3.2}$$

since $\mathbf{v}_{out} = \mathbf{S}\mathbf{v}_{in}$. As a result, the matrix $\mathbf{S}$ has to be unitary, i.e. $\mathbf{S}^\dagger \mathbf{S} = \mathbf{I}_{2N}$.

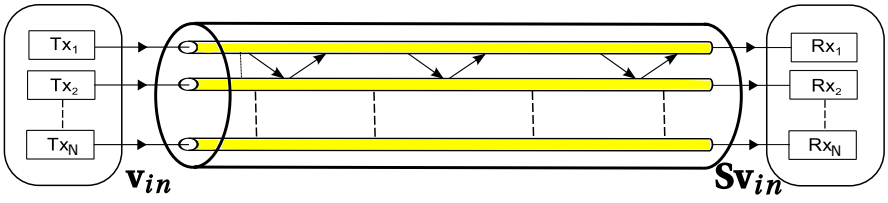


Figure 3.1: Illustration of crosstalk between spatial channels in optical fiber with input power $\mathbf{v}_{in}$ and output $\mathbf{v}_{out} = \mathbf{S}\mathbf{v}_{in}$.

A second important property of the scattering matrix relates to its time-reversal symmetry. It is well known that electromagnetic propagation in the absence of external magnetic fields is symmetric under time reversal. In this context, time reversal corresponds to a change in the direction of propagation and time. For example, under time-reversal the amplitude of a propagating plane-wave $\psi(\mathbf{x},\mathbf{t}) = \exp[i(\mathbf{kx}-\omega t)]$ changes both time $t \to -t$ and propagation direction $\mathbf{k} \to -\mathbf{k}$. Hence, time-reversal amounts to phase conjugation [39]. Therefore, if propagation through the optical fiber is to be time-reversal invariant, feeding the system with the time-reversed version of the output should produce the original version of the input. This implies that $\mathbf{R}_\ell = \mathbf{R}_\ell^t$, $\mathbf{R}_r = \mathbf{R}_r^{\mathrm{T}}$ and $\mathbf{T}_\ell = \mathbf{T}_r^{\mathrm{T}}$.

As a result, we are left with three different matrices, namely $\mathbf{R}_\ell$, $\mathbf{R}_r$ and $\mathbf{T}_\ell = \mathbf{T}_r^t \equiv \mathbf{T}$.These matrices are not independent, since they share the same singular values, since $\mathbf{R}_\ell^\dagger\mathbf{R}_\ell + \mathbf{T}^\dagger\mathbf{T} = \mathbf{R}_r^\dagger\mathbf{R}_r + \mathbf{T}^\dagger\mathbf{T} = \mathbf{I}_N$.It is convenient to define the matrix $\mathbf{\Delta}$ as the diagonal matrix with the reflection eigenvalues, i.e. $\mathbf{\Delta} = \mathbf{R}_\ell^\dagger\mathbf{R}_\ell = \mathbf{R}_r^\dagger\mathbf{R}_r$. It has been shown elsewhere [4, 53] that $\mathbf{S}$ can be expressed in terms of $\mathbf{\Delta}$ by means of a so-called polar decomposition as follows

$$\mathbf{S} = \begin{bmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{V} \end{bmatrix} \begin{bmatrix} -\mathbf{\Delta}^{1/2} & (\mathbf{I}_N - \mathbf{\Delta})^{1/2} \\ (\mathbf{I}_N - \mathbf{\Delta})^{1/2} & \mathbf{\Delta}^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{W}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}^{\mathrm{T}} \end{bmatrix} \tag{3.3}$$

As a result, the information of the scattering matrix $\mathbf{S}$ is encoded in the matrix $\mathbf{\Delta}$ and the unitary matrices $\mathbf{W}$ and $\mathbf{V}$.

We now discuss two important properties of the scattering matrix as seen from experimental data in the literature, which will help describe it better. We start with the strength of backscattering, i.e. reflection in optical fibers. This process is typically due to localized imperfections in the fiber and is sometimes called Rayleigh scattering. The strength of the reflected light is typically proportional to the product of the density of such imperfections and the length of the fiber [31], i.e. proportional to the average number of such imperfections over the fiber length travelled. Due to the high quality of fiber production techniques this imperfection density is extremely small. Hence, in [28] single core fibers have reflection coefficients approximately equal to -120dB/mm, which amounts to -30dB per 1000km. Similarly, in [54] a 25km single mode fiber has Rayleigh backscattered power roughly -30dB. These very low reflected powers appear in single mode fibers, however, we conjecture that they should be quite low for multi-core fibers described below. As a result of this low backscattering amplitudes we may assume that the reflection in the fiber may discarded, and hence $\mathbf{\Delta} \approx \mathbf{0}$. In this limit the transmission matrix $\mathbf{T}$ can be expressed as $\mathbf{T} = \mathbf{W}\mathbf{V}^{\mathrm{T}}$, which is thus random Haar distributed.

A second important property of the scattering matrix in a multicore/multimode fiber is the considerable mixing between core transmissions. For example, in a 60km three coupled core fiber analyzed in [57], the crosstalk is so strong that light injected into one core is equally distributed across all cores in the output. Considerable crosstalk has been seen in other cases, e.g. in [61] where crosstalk of -25dB/km was observed.

Even if this effect is smaller that in [57] above due to the different design of the cores (it results to -8dB coupling for 60km), it highlights the relevance and ubiquitous nature of crosstalk in multicore fibers, when they have their cores placed close to each other. It should be pointed out that the difference in magnitude of backscattering and crosstalk can be attributed to different mechanisms being responsible for the two effects. In the backscattering case, as discussed above, the effect is due to localized scattering [28], while in the latter the mechanism is scattering among the core modes due their proximity, or due to bending [21].

In summary, we consider fibers with negligible backscattering and strong mixing between core modes. We assume this mixing to be random over different frequency subbands, due to strong delay spread. For example, in [67] 10nsec delay spreads were measured over 700km transmission over a 6 mode fiber using 50GHz sub-band widths. Hence the transmission matrix $\mathbf{T}$ will be modelled as a Haar random matrix of dimension $N \times N$. Without loss of generality we assume $N_t \leq N$ transmitting channels and $N_r \leq N$ receiving channels, and therefore we only consider a submatrix of the full transmission matrix $\mathbf{T}$, which we denote by $\mathbf{U}$, since not all transmitting or receiving channels may be available to a given link. For simplicity we assume that this is the upper left corner of $\mathbf{T}$. We should emphasize that the remaining $N - \max(N_t, N_r)$ "untapped" channels in $\mathbf{T}$ can be used to model *loss* to the environment in the fiber propagation [73]. This is done in an analogous way to wireless communications, where the energy transmitted from an antenna operating at wavelength λ is spread away over a large number of modes (e.g. plane waves) and only a tiny fraction thereof $O(\lambda^2/R^2)$ is received at a receiving antenna at a distance R away. Indeed, in the limit of large $N \gg N_t, N_r$ the channel will converge to a Gaussian distributed channel, [73] similar to the case of open space wireless propagation, where the signal loss is significant. As a result, the corresponding MIMO channel for this system reads

$$\mathbf{y} = \mathbf{U}\mathbf{x} + \mathbf{z}, \tag{3.4}$$

with coherent detection and channel state information only at the receiver [24, 78]. $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are the $N_t \times 1$ input, the $N_r \times 1$ output signal vectors and the $N_r \times 1$ unit variance noise vector, respectively, all assumed for simplicity to be complex Gaussian. This assumption is also based on the optical MIMO modulation scheme, which uses MZM (*Mach-Zehnder Modulator*) [47] to modulate a continuous wave (CW) laser to generate the digital signal, which is then, transmitted through the fiber. This modulation is achieved by equally splitting the incoming optical signal and enforcing a time delay (phase shift) in one path, before recombining it. Therefore the signal output from an MZM is complex, which for simplicity we model as a complex Gaussian. We also assume no mode-dependent loss. As a result, the mutual information can be expressed as

$$I_N(\mathbf{U}) = \frac{1}{N_t} \log\det(I + \rho \mathbf{U}^\dagger \mathbf{U}) \tag{3.5}$$

$$= \frac{1}{N_t} \sum_{k=1}^{N_t} \log(1+\rho\lambda_k)$$
$$= \int_0^1 p(x)dx \log(1+\rho x).$$

In the last line $p(x)$ is the spectral density of $\mathbf{U}^\dagger\mathbf{U}$ defined as

$$p(x) = \frac{1}{N_t} \sum_k \delta(x-\lambda_k), \tag{3.6}$$

and "log" is the natural logarithm, ρ is the average total signal-to-noise ratio, λ_k are the eigenvalues of the matrix $\mathbf{U}^\dagger\mathbf{U}$ and we assume for concreteness $N_t \leq N_r$. It is useful to define $\beta = N_r/N_t > 1$, $N_0 = N - N_t - N_r$ and $n_0 = N_0/N_t \geq 0$. If $N_0 < 0$, [15] showed that we may recover the above form by replacing $N_t \to N - N_r$, $N_r \to N - N_t$ and $N_0 \to -N_0$ and $I_N \to I_N + n_0 \log(1+\rho)$. It should be emphasized that the above mutual information is used as a performance metric of the channel.

3.2 The Hamiltonian Approach

The optical fiber may be viewed as a cavity where optical power may enter and exit from both ends. The output power $\mathbf{v}_{\text{out}}$ is related to the input power v_{in} through $\mathbf{v}_{out} = \mathbf{S}\mathbf{v}_{in}$ (see Fig. 3.1) with the $2N \times 2N$ scattering matrix $\mathbf{S}$ given by (3.1)

The fiber exhibits random distributed crosstalk between modes or cores. We assume this mixing to be random over different frequency bands, due to strong delay spread. The situation is analogous to that of a chaotic cavity, which randomly mixes the cavity states. The analytic expression of the $2N \times 2N$ scattering matrix for a chaotic cavity reads [4]:

$$\mathbf{S} \;=\; \mathbf{I} - 2\pi i \mathbf{W}^\dagger (\boldsymbol{\mathcal{H}} + i\pi \mathbf{W}\mathbf{W}^\dagger)^{-1}\mathbf{W}. \tag{3.7}$$

Here $\boldsymbol{\mathcal{H}}$ is the $2N \times 2N$ channel Hamiltonian and $\mathbf{W}$ is a $2N \times 2N$ matrix containing the coupling constants of the fiber to the outside world. The dimension is $2N \times 2N$ as there are N incoming states from the left and N incoming states from the right, while inside the fiber there are N states propagating from left to right and N states propagating from right to left. So in case of perfect (lossless) leads, $\mathbf{W} \propto \mathbf{I}_{2N}$.

The channel Hamiltonian is

$$\boldsymbol{\mathcal{H}} = \begin{bmatrix} \mathbf{0}_N & \mathbf{H}_{r\to\ell} \\ \mathbf{H}_{\ell\to r} & \mathbf{0}_N \end{bmatrix}. \tag{3.8}$$

The offdiagonal sub-matrices vanish due to absence of reflection. $\mathbf{H}_{(\ell,r\to r,\ell)N}$ are $N \times N$ Hermitian. Because $\mathbf{H}_{r\to\ell} = \mathbf{H}^\dagger_{\ell\to r}$ we can write for simplicity $\mathbf{H} \equiv \mathbf{H}_{\ell\to r}$. (3.7) then becomes

$$\mathbf{S} = \mathbf{I}_{2N} - 2\alpha\pi i \mathbf{W}^\dagger \left(\boldsymbol{\mathcal{H}} + i\alpha\pi \mathbf{W}\mathbf{W}^\dagger\right)^{-1} \mathbf{W}. \tag{3.9}$$

But, as $\mathbf{W}\mathbf{W}^\dagger = \alpha\mathbf{I}_{2N} \ll 1$, for simplification it is $\boldsymbol{\mathcal{H}} + i\pi\mathbf{W}\mathbf{W}^\dagger \approx \boldsymbol{\mathcal{H}}$ and finally we have

$$\mathbf{S} = \mathbf{I}_{2N} - 2\alpha\pi i\mathbf{W}^\dagger\boldsymbol{\mathcal{H}}^{-1}\mathbf{W}. \tag{3.10}$$

To model the MDL we add the $2N \times 2N$ loss matrix Γ [25]

$$\Gamma = \begin{bmatrix} \mathbf{0}_N & \boldsymbol{\Gamma}_{r\to\ell} \\ \boldsymbol{\Gamma}_{\ell\to r} & \mathbf{0}_N \end{bmatrix}. \tag{3.11}$$

Just as for $\boldsymbol{\mathcal{H}}$ we have $\boldsymbol{\Gamma} \equiv \boldsymbol{\Gamma}_{\ell\to r}$. For simplicity we can assume that $\boldsymbol{\Gamma}$ is a diagonal matrix.

$$\mathbf{S} = \mathbf{I} - 2\alpha\pi i\mathbf{W}^\dagger\left(\boldsymbol{\mathcal{H}} + i\Gamma\right)^{-1}\mathbf{W} \tag{3.12}$$

or

$$\mathbf{S} \;=\; \mathbf{I} - 2\alpha\pi i\mathbf{W}^\dagger\left(\boldsymbol{\mathcal{H}} + i\Gamma\right)\left(\boldsymbol{\mathcal{H}}^2 + \Gamma^2\right)^{-1}\mathbf{W}. \tag{3.13}$$

As we saw in the previous section, we are interested in the complex $N_t \times N_r$ submatrix $\mathbf{U}$ (channel matrix) of the $2N \times 2N$ matrix $\mathbf{S}$. To extract $\mathbf{U}$ we use two diagonal $2N \times 2N$ matrices $\mathbf{A}_{\text{diag}}$ and $\mathbf{B}_{\text{diag}}$, which can be seen in Appendix A with $\mathbf{A}_{\text{diag}}\mathbf{B}_{\text{diag}} = 0$.

$$\mathbf{U} = \mathbf{A}_{\text{diag}}\mathbf{S}\mathbf{B}_{\text{diag}} \tag{3.14}$$

and

$$\mathbf{U}^\dagger = \mathbf{B}_{\text{diag}}^\dagger\mathbf{S}^\dagger\mathbf{A}_{\text{diag}}^\dagger, \tag{3.15}$$

which yields

$$\mathbf{U}^\dagger\mathbf{U} = \mathbf{B}_{\text{diag}}^\dagger\mathbf{S}^\dagger\mathbf{A}_{\text{diag}}^\dagger\mathbf{A}_{\text{diag}}\mathbf{S}\mathbf{B}_{\text{diag}}, \tag{3.16}$$

so that we can write,

$$\begin{aligned} \mathbf{U}^\dagger\mathbf{U} &= 2\alpha i\pi\left(\mathbf{H} - i\boldsymbol{\Gamma}\right)\left(\mathbf{H}^2 + \boldsymbol{\Gamma}^2\right)^{-1}\left(-2\alpha i\pi\left(\mathbf{H} + i\boldsymbol{\Gamma}\right)\left(\mathbf{H}^2 + \boldsymbol{\Gamma}^2\right)^{-1}\right) \\ &= 4\alpha^2\pi^2\left(\mathbf{H}^2 + \boldsymbol{\Gamma}^2\right)^{-1}. \end{aligned} \tag{3.17}$$

To incorporate the idea of a fading channel subject to crosstalk in our analysis, we assume that the channel $\mathbf{H}$ consists of a random part $\mathbf{G}$, plus a deterministic part $\mathbf{H}_0$. Thus the final equation becomes

$$\mathbf{U}^\dagger\mathbf{U} \;=\; 4\alpha^2\pi^2\left(\left(\mathbf{H}_0 + \gamma\mathbf{G}\right)^2 + \boldsymbol{\Gamma}^2\right)^{-1}. \tag{3.18}$$

Here γ is a parameter controlling the randomness. $\mathbf{H}_0$ is a diagonal matrix and corresponds to the line-of-sight component inside the fiber while the Gaussian distributed matrix $\mathbf{G}$ describes the crosstalk.

CHAPTER 4
Crosstalking Existence

4.1 Existence of Crosstalking in Modern Optical MIMO

In order to feed the ever hungry demand for uninterrupted and broadband communications, optical backhaul networks are being deployed but even these are not sufficient, reaching the limits of the "classical" - *Single Mode Fiber* (SMF) communications, which forces us to investigate new ways of better exploiting the corresponding bandwidth. For that, *Multi Core Fibers* (MCF) or *Few Mode Fibers* (FMF) have been developed, which along with the *Multi Mode Fibers* (MMF), allow the light to propagate using spatial division, thus increasing considerably the available channel capacity. An MMF, contrary to SMF, has a larger core diameter, which exceeds the wavelength of the propagating light. That characteristic allows the excitation of more than one propagation modes (see Fig. 4.1), but at the same time makes the communication system suffer from *modal dispersion* [27]. Moreover, as the number of transmitters increases, the crosstalk phenomenon occurs [35, 36]: In order to fit the many different propagating ways inside the constricted space of the fiber, the distance between these propagating modes has been reduced to a minimum, thus inevitably, they experience crosstalking with adjacent ones. In this chapter we will show that the crosstalking phenomenon is present even for a 4×4 optical MIMO system of a very short span.

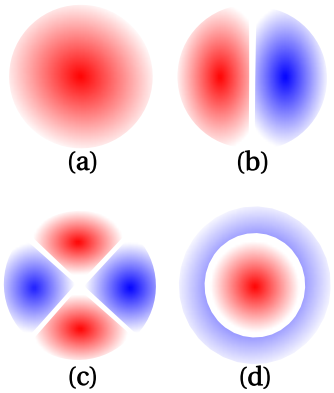


Figure 4.1: Mode profiles of the lower-order fiber modes (a) LP_{01} (b) LP_{11} (c) LP_{21} (d) LP_{02}

In order to show this, we conducted an experiment along with Alcatel-Lucent. More precisely, Our system consists of a very short piece of MMF fiber between MUX

and DeMUX. 5 modes LP01, LP11 a+b and LP21 a+b being injected in the system (LP11 and LP21 have 2 polarizations, "a" and "b"), while the receiver reads LP21 a+b and processes them in parallel in a MIMO 4×4 adaptive (blind) equalizer driven by *Constant Modulus Algorithm*-CMA [41]). Each polarization is equally stimulated. At the receiver there are 4 equalizers, each one with 4 filters (see Fig. 4.3). Each filter consists of 25 taps because we need to recover not only the state of the polarization (or mode) which would be memoryless, but also to recover all the other impairments coming from the signal (*Intersymbol Interference*- ISI) or from the channel (chromatic dispersion and polarization/spatial-mode dispersion). The sequence used is $2^{15} - 1$ PRBS -*Pseudo Random Binary Sequence* of PDM-QPSK at a rate of 32 GBs. Each "Equalizer" variable is the state of the equalizer at the end of the processing of an acquisition. The scope is triggered every 5 seconds and an acquisition lasts 40ms At the end of the processing of an acquisition we register the output of each of the 25 taps ($T/2$ spaced) of the 16 equalizer filters. In Fig. 4.2 we can see a representation of our system, where $s(n)$ is the PRBS sequence/signal, $c(t)$ is the channel, $w(t)$ is the AWGN, FSE is the *Fractionally Spaced Equalizer* with "equalizer" variables f_k and $y_M(n)$ is the signal after the equalizing processing. The multi-channel model of Fig. 4.4 subdivides

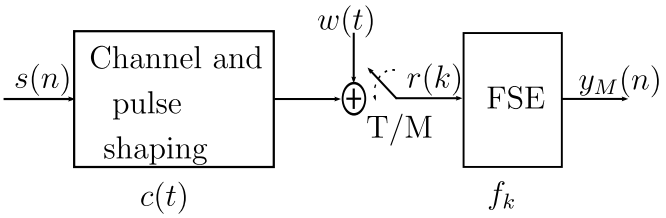


Figure 4.2: Schematic representation of our system

the fractionally spaced channel coefficients $c_k = c(k\frac{T}{2})$ and the discrete-time random process $w_k = w(k\frac{T}{2})$ into even and odd counterparts, [41], so that $c_n^{even} = c_{2n}$ and $c_n^{odd} = c_{2n+1}$ for $n = 0, 1, \ldots$. In an analogous way, the coefficients of the *Fractionally Spaced Equalizer* (FSE), $\mathbf{f}$, are partitioned as $f_n^{even} = f_{2n}$ and $f_n^{odd} = f_{2n+1}$. That way, we can create

$$\mathbf{C}_e = \begin{bmatrix} c_0^{even} & & & \\ c_1^{even} & c_0^{even} & & \\ \vdots & c_1^{even} & & \\ c_{M-1}^{even} & \vdots & \ddots & c_0^{even} \\ & c_{M-1}^{even} & & c_1^{even} \\ & & & \vdots \\ & & & c_{M-1}^{even} \end{bmatrix}, \quad \mathbf{C}_o = \begin{bmatrix} c_0^{odd} & & & \\ c_1^{odd} & c_0^{odd} & & \\ \vdots & c_1^{odd} & & \\ c_{M-1}^{odd} & \vdots & \ddots & c_0^{odd} \\ & c_{M-1}^{odd} & & c_1^{odd} \\ & & & \vdots \\ & & & c_{M-1}^{odd} \end{bmatrix}$$

which are matrices of size $P \times N$ where $P = M+N+1$, M is the length of a counterpart of the channel coefficients and N, analogously, is the length of a counterpart of FSE

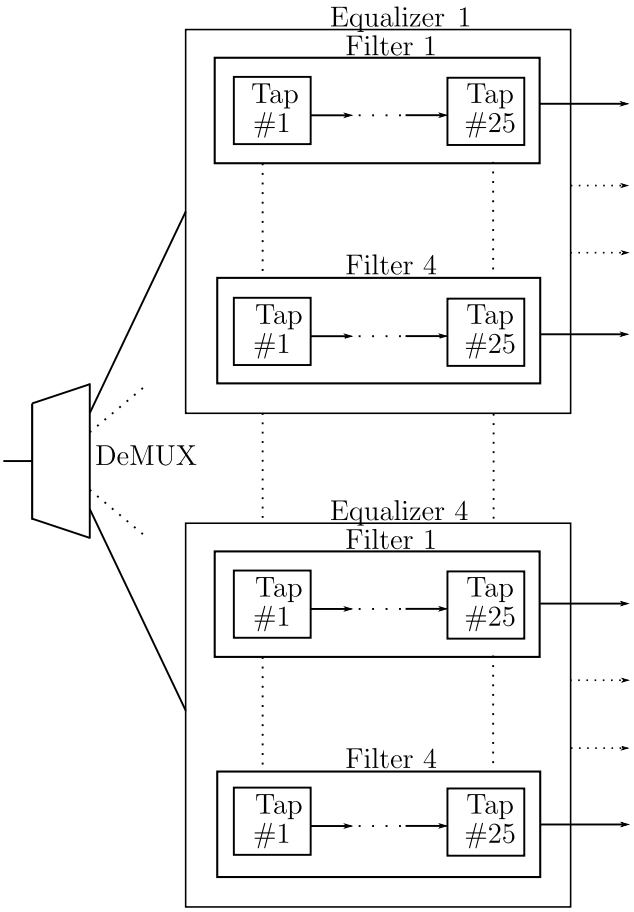


Figure 4.3: Block schema of the receiver

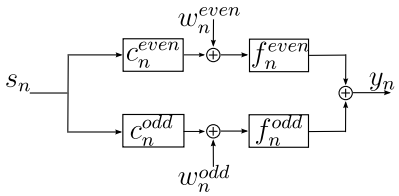


Figure 4.4: Multi-channel paradigm, subdivided into odd and even counterparts.

coefficients. The impulse response, $\mathbf{h}$, of the linear system relating s_n to y_n can be constructed so that

$$\mathbf{h} = \mathbf{Cf}, \tag{4.1}$$

where $\mathbf{C} = [\mathbf{C_o}\ \mathbf{C_e}]$ and $\mathbf{f}^T = [\mathbf{f_e}\ \mathbf{f_o}]$. In case of perfect equalization and *Perfect Source Recovery* (PSR) it is $\mathbf{h}_\delta = [0, 0, \ldots, 1, \ldots, 0]$, where the non-zero coefficient is in the δth position ($0 \leq \delta \leq P-1$). For PSR under arbitrary δ, $\mathbf{C}$ must be of full row rank [48]. To simplify our notation we can construct the matrix $\mathbf{C}_{FS}$ of size $(M+N) \times 2N$ from

a vector of fractionally-spaced channel coefficients $\mathbf{c}_{FS} = [c_0, c_1, \ldots, c_{2M-1}]^T$ so as

$$\mathbf{C}_{FS} = \begin{bmatrix} c_0 & & & \\ c_1 & c_0 & & 0 \\ \vdots & c_1 & & \\ c_{2M-1} & \vdots & \ddots & c_0 \\ & c_{2M-1} & & c_1 \\ 0 & & & \vdots \\ & & & c_{2M-1} \end{bmatrix}, \tag{4.2}$$

and the full rank requirement dictates that

$$2N \geq M + (N-1) \Rightarrow N \geq M - 1. \tag{4.3}$$

Finally, the nth equalizer output can be written as

$$y_n = \mathbf{s}^T(n)\mathbf{C}\mathbf{f} + \mathbf{w}^T(n)\mathbf{f}. \tag{4.4}$$

The CM algorithm for equalization uses a cost function J_{CM} which penalizes deviations [41]. For PSK symbols $s_n \in \{e^{j2\pi m}/2^M\}$, $m \in \{0, 1, \ldots 2^M - 1\}$

$$\begin{aligned} J_{CM|PSK} = &\sum_{i=0}^{P-1} |h_i|^4 + 2\sum_{i=0}^{P-1}\sum_{m=0,m\neq i}^{P-1} |h_i|^2|h_m|^2 + |E\{s^2\}|^2 \sum_{i=0}^{P-1}\sum_{j=0,j\neq i}^{P-1} h_i^2(h_j^*)^2 \\ &+ \kappa_w\sigma_w^4 \sum_{i=0}^{2N-1} |f_i|^4 + 2\sigma_w^4 \sum_{i=0}^{2N-1}\sum_{m=0,m\neq i}^{2N-1} |f_i|^2|f_m|^2 \\ &+ 4\sigma_w^2||\mathbf{h}||_2^2||\mathbf{f}||_2^2 - 2(||\mathbf{h}||_2^2 + \sigma_w^2||\mathbf{f}||_2^2) + 1, \end{aligned} \tag{4.5}$$

where κ_w is the normalized kurtosis of the noise, σ_w is the variance of the noise and $||\cdot||_2$ is the $\ell-2$ norm J_{CM} is optimized through Steepest Descend. and the generalized equalizer tap update for complex signals is

$$\mathbf{f}(k+1) = \mathbf{f}(k) + \mu.\underbrace{y(k)(\gamma^2 - |y(k)|^2)}_{\nabla_{\mathbf{f}} J^{CMA} = e^{CMA}(k)}\mathbf{r}^*(k) \tag{4.6}$$

where $e^{CMA}(k)$ is the error signal. Equation (4.6) is written in terms of the (fractionally-sampled) regressor vector at time n:

$$\mathbf{r}_n = [r_n^{odd}, \ldots, r_{n-(N-1)}^{odd}, r_n^{even}, \ldots, r_{n-(N-1)}^{even}]^T \tag{4.7}$$

the equalizer parameter vector $\mathbf{f}_n$ at time index n, the equalizer output y_n, a step-size μ, and the squared source-modulus γ (also referred to as the dispersion constant).

Methodology

As was stated in the beginning, we have acquired and saved the equalizer coefficients $\mathbf{f}_{2N}$. Our aim is to compute the channel coefficients $\mathbf{c}_{2M}$ only by using the already known $\mathbf{f}_{2N}$. In the case of PSR, we know that $\mathbf{h}_\delta = \mathbf{Cf}$, therefore for our real data it is a classic problem of finding $\mathbf{C}$ while minimizing the squared ℓ_2 norm :

$$||\mathbf{Cf} - \mathbf{h}_\delta||_2^2 = |\mathbf{Cf}|^2 - 2\mathbf{h}_\delta\mathbf{Cf} + 1, \tag{4.8}$$

which is a quadratic equation of the form $\mathbf{c}^T\mathbf{Ac} - 2\mathbf{c}^T\mathbf{v}$, where $\mathbf{A}$ is the quadratic part and $\mathbf{v}$ is the linear part. To minimize (4.8), we take the derivative with respect to $\mathbf{c}$ and set it to zero, which solves for

$$\mathbf{c} = \mathbf{A}^{-1}\mathbf{v}. \tag{4.9}$$

We can take advantage of the toeplitz characteristics of $\mathbf{C}$ and write (4.8) analytically,

$$\sum_{m=0}^{2M-1}\sum_{m'=0}^{2M-1} c_m c_{m'} \sum_{k_{max(m,m')}}^{2N-1+min(m,m')} f_{k-m}f_{k-m'} - 2\sum_{l=max(0,\delta-2M)}^{min(2N-1,\delta)} c_{\delta-l}f_l + 1 = 0. \tag{4.10}$$

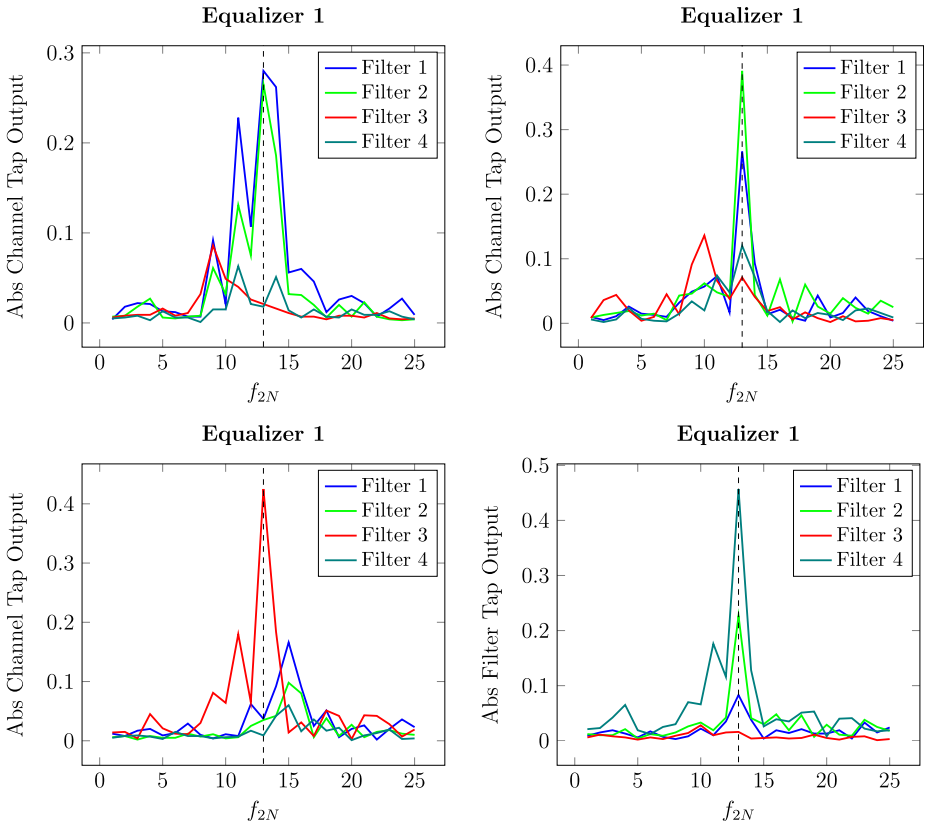


Figure 4.5: Average value of the 500 different instantiations of the 4 filters for the 4 equalizers. The dashed line points the 13th tap.

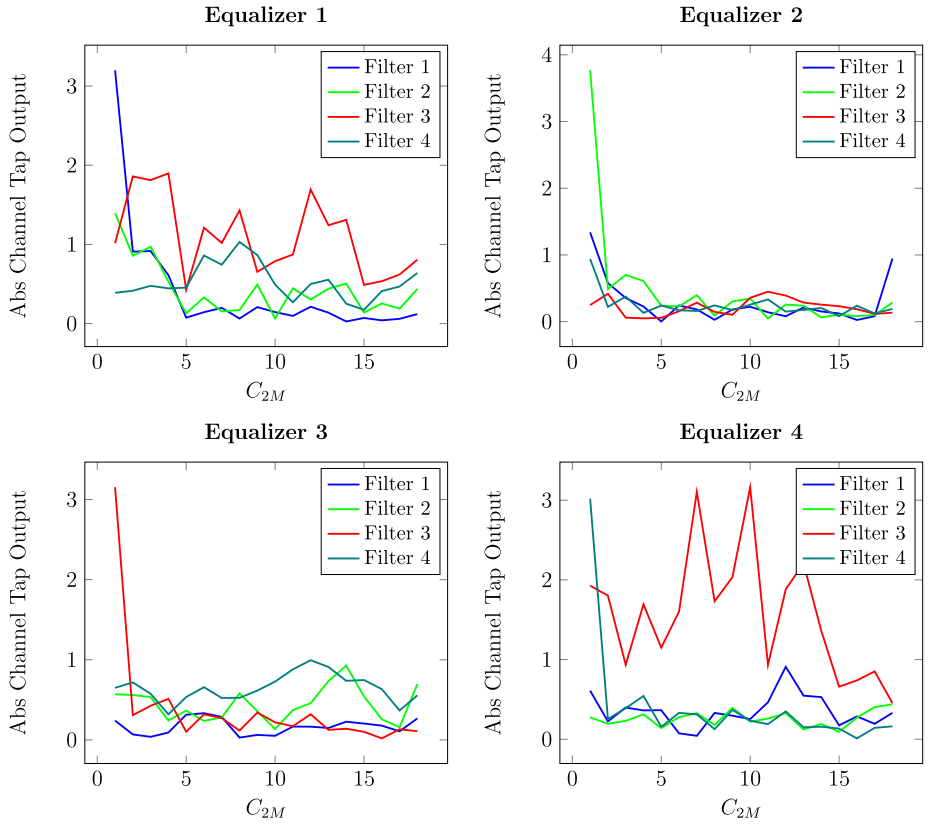


Figure 4.6: Channel coefficients for the 4 equalizers.

4.2 Real Data Analysis and Results

As a first step in our data analysis, we plot the average of the output of the filter taps; see Fig. 4.5. We can distinguish that for every equalizer, there is a dominant filter, which has a maximum at almost the same tap, the 13th. It has to be mentioned though that although the averaged output has always maximum at the 13th tap, this is not the general rule for every instantiation. That characteristic, gives us the hint of $\delta = \#\max(f_{2N})$. In fact, we check our hypothesis: $\min(||\mathbf{Cf} - \mathbf{h}_\delta||)$ and it holds true, the proof of which is omitted. That way, we compute the $\mathbf{c}_{2M}$, for every equalizer, see Fig. 4.6. The number of spatial sub-channels M is set manually to $M = 9$. The reason for that is twofold: In the second part of (4.10) the argument of c is $\delta - l$ but the length of $\mathbf{c}$ is $2M$. Therefore $2M \geq \delta - l \Rightarrow M_{min} = \frac{\delta - l}{2}$. Moreover, we chose $M = M_{min} = 9$ for computational simplicity.

In Fig. 4.7 we compute the CDF of the eigenvalues for different channel coefficients. By setting also the off-diagonal elements of the channel coefficients to zero, we plot 2 different group of curves, each one distinguished by the red and blue color. That way, the red color indicates there is no crosstalking in our system. Therefore, it is easily visible that our system suffers from cross-talking.

Finally, in Fig. 4.8 we plot the MIMO channel capacity c_f for the different frequencies, with and without crosstalk.

$$c_f = \frac{1}{18}\sum_{f=1}^{18} \log\det(\mathbf{I}_4 + \rho\mathbf{H}^{\dagger}(f)\mathbf{H}(f)) \tag{4.11}$$

It is obvious that, by exploiting the MIMO gains, the channel capacity can be enhanced.

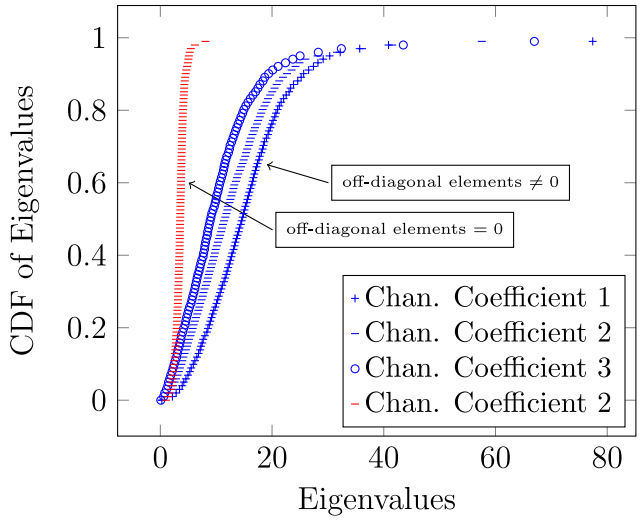


Figure 4.7: CDF of eigenvalues for some channel coefficients.

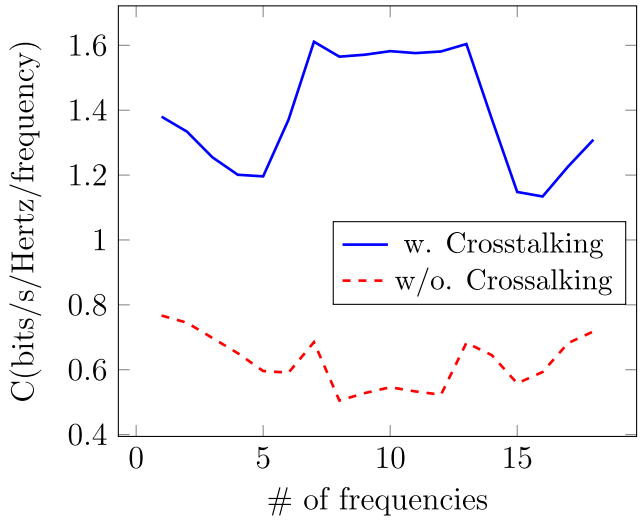


Figure 4.8: Average capacity c_f for the different frequencies f, over the 500 instantiations.

4.3 Effect of Noise on Channel Estimation

In our experiment we registered only the equalizer coefficients, which left us uninformed about the noise, its characteristics and how it affects our proposed method of extracting the channel state information. To test the robustness of our algorithm, we construct a single pseudo-channel with coefficients c_{2M}, $|\mathbf{C}_{2M}^2|^2 = 1$. The channel vector $\mathbf{c}$ is sparse, with 7 non-zero elements and 11 zero ones. Channel matrix $\mathbf{C}$ with dimensions $(M+N)\times 2N$ is constructed according to (4.2) with toeplitz characteristics, where $2N$ is the size of the equalizer vector $\mathbf{f}$. In order to ease our computations we set $M = N$, thus $\mathbf{C}$ is easily invertible. Arbitrarily we set $\delta = 7$. So, the estimated $\mathbf{f}_{est}$, without the presence of AWGN is

$$\mathbf{f}_{est}^T = \mathbf{C}^{-1}\mathbf{h}, \tag{4.12}$$

and reversely using the methodology from the previous section to estimate the channel coefficients $\mathbf{c}_{est}$, we end up in $\mathbf{A}_{est}$ and $\mathbf{V}_{est}$.

Sparse Channel Estimation

In the case of noise presence, by considering our channel as sparse, we can estimate a new de-noised $\tilde{\mathbf{c}}_{est}$ by using a *Compressed Sensing* (CS) method based on *Dantzig Selector* (DS) [11]. Let us remember (4.9). By adding AWGN (z), we get

$$\mathbf{c}_{noise} = \mathbf{A}^{-1}\mathbf{v} + \mathbf{z}. \tag{4.13}$$

$\mathbf{v}$ is the solution to the following optimization

$$\min_{\tilde{\mathbf{v}}} ||\tilde{\mathbf{v}}||_1 \text{ s.t. } ||\mathbf{A}^H(\mathbf{c} - \mathbf{A}^{-1}\tilde{\mathbf{v}})||_\infty \leq \epsilon, \tag{4.14}$$

where $\tilde{\mathbf{v}}$ is the estimate of $\mathbf{v}$ and ϵ is a constant, depending on the noise and the channel. In our case we set $\epsilon = 5.10^{-3}$ As can be seen in Fig. 4.9, AWGN, can insert significant distortion in channel estimation. But on the other hand by using a CS algorithm like DS, the distortion can be minimized. The recreation of the channel is less effective for higher P_{noise} but yet the CS method shows significant effectiveness.

4.4 Conclusions

In this chapter it was shown that in-fiber crosstalking is an existing impairment for systems which incorporate SDM techniques. Therefore, fiber optics is another field where MIMO techniques can find great usage. Moreover, due to the sparse characteristics of the channel, CS methods can further enhance the equalization process . A more realistic, practical implementation of an optical MIMO system along with a CS-equalization technique is a question for future work.

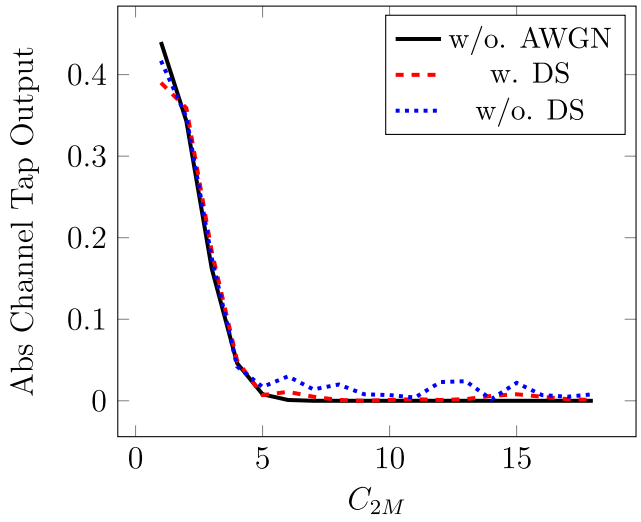


Figure 4.9: Sparse channel coefficient estimation for $P_{noise} = -20dB$ per sample with and without the use of DS.

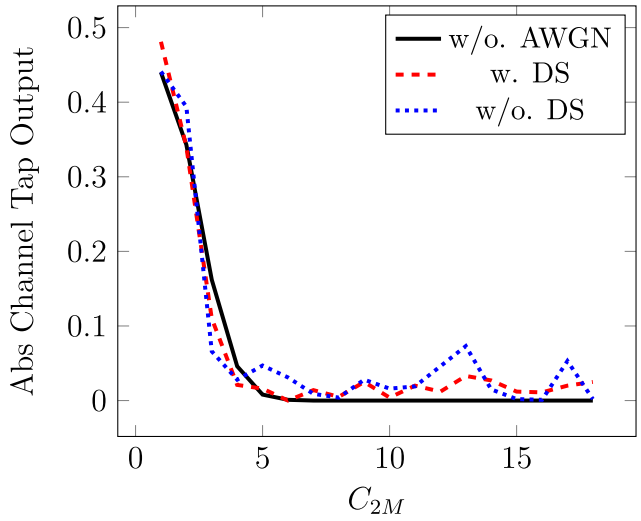


Figure 4.10: Sparse channel coefficient estimation for $P_{noise} = -10dB$ per sample with and without the use of DS.

CHAPTER 5

Outage Capacity

5.1 Outage Capacity for the Optical MIMO Channel

An important issue that arises in optical MIMO, as was already seen in the previous chapter, is that typically there is cross-talk between fiber modes, which increases with segment length [75] and can be attributed to imperfections, as well as to the twist and the bending of the fiber [21, 33], and slight variations in the local temperature [44]. There have been two trends of work in this direction. In the first, effort has been made to minimize cross talk between cores to extremely low levels [92], thus not having to deal with self-interference. While appealing from a signal processing point of view, the downside is that this methodology does not scale, in the sense that coupling becomes unavoidable with increasing number of cores in a fiber. Another more pragmatic approach is to design fibers without bothering about the appearance of cross-talk. Indeed, bringing cores close to each other can lead to power being spread at the receiver side evenly in the outlook of the channel [57].

Recently, it was proposed [38, 76, 86] to use sophisticated transceiver techniques developed in the context of wireless communications between multiple transmitting and receiving antennas (MIMO), which can mitigate self-interference, thus providing significant throughput increases. Of course, optical fiber multi-core systems have several differences compared to multi-antenna wireless systems, which need to be addressed. One important difference is the one-dimensional, near-lossless propagation through the optical fiber. As a result, the incoming and outgoing propagating modes of the fiber are related through a so-called scattering matrix, which we examined in a previous chapter and is unitary in the limit of lossless propagation. In contrast, since wireless propagation incurs significant radiation loss to the environment, the corresponding channel coefficients may be taken to be i.i.d and Gaussian [23, 73]. Second, due to the existence of non-linearities at high powers, one cannot expect arbitrarily large powers to be practical. Hence, the intuitive analysis of the tradeoff between diversity and multiplexing [91], which is popular in the wireless context is not applicable here. Third, in contrast to the wireless setting, where due to physical motion the channel fades significantly over time, the variability of the channel is mostly over different frequencies and fiber segments. Hence, given that a given packet is likely to travel over different segments and frequencies, which cannot be known apriori to the transmitter, it is important to define an outage criterion over the realization of the channel matrix in this context. Finally, the practical metric for the performance is not

the ergodic mutual information, but, rather, the outage capacity at very low outage (e.g. 10^{-4}) [86], due to the fact that feedback from the receiver to the transmitter to request a retransmission in the case of packet loss, is almost always impossible.

It is therefore important to develop a propagation channel model for the fiber-optical MIMO channel, which addresses these issues. Several attempts in this direction have been made [38, 68], however the unitary aspect of propagation has not been dealt with explicitly until [86] and then [15] introduced the unitary aspects of the transmission channel. In particular, [15] introduced in a somewhat ad-hoc fashion the so-called Jacobi MIMO channel, in which the matrix corresponding to the channel is a rectangular submatrix from a Haar distributed random matrix from $\mathbf{U}(N)$.

In this chapter, we derive a channel model for an fiber-optical MIMO channel when the coupling between transmitting channels is strong and backscattering is weak. These two assumptions allow us to provide the general random matrix that characterizes the propagation in such a multimode fiber in the presence of time-reversal symmetry. The resulting model is similar to the one introduced by [15], but it also parameterizes loss in the fiber. We show how for increased loss, the channel interpolates between unitary and Gaussian. This channel allows us to analyze the outage capacity of the optical MIMO channel. As mentioned above, this is the relevant information transmission metric for fiber-optical coupled multi-core channels. We analyze the characterize the problem in a dual way. First, we obtain closed form expressions for the outage probability for small numbers of channels. We also obtain analytical expressions, which are valid technically in the limit of large channel numbers, but also work well over smaller channel numbers. The method is particularly suited to obtain the outage mutual information for very low outages with finite SNR. Essentially, it amounts to calculating the rate function of the logarithm of the average moment generating function of the mutual information. The methodology we use is based on the so-called Coulomb gas approach which was developed in the physics literature in the context of random matrix theory [19] in the 60's. It is quite intuitive because it interprets the eigenvalues as point charges on a line repelling each other logarithmically. The Coulomb gas method has seen recently a renewed interest in its use to obtain large deviations results for random matrix problems [16, 51, 83, 84] and also in communications [13, 14, 45, 49]. We will follow the basic steps discussed in more details in [45]. As a by-product of this analysis we obtain the ergodic mutual information and its variance for this channel.

5.2 System Model

We consider a single-segment N-channel lossless optical fiber system, with $N_t \leq N$ transmitting channels excited and $N_r \leq N$ receiving channels coherently excited in the input (left) and output (right) side of the fiber. The propagation through the fiber may be analyzed through its $2N \times 2N$ scattering matrix which was analyzed in Chapter 3.1.

Statement of Problem

We may now define the main problem we address, namely the calculation of

$$P_{out}(r) = Prob(I_N < r) \tag{5.1}$$
$$= \mathbb{E}_{\mathbf{U}}\left[\Theta(r - I_N(\mathbf{U}))\right] \tag{5.2}$$

where $\Theta(x)$ is the indicator (step) function. We will also analyze the density of r i.e.

$$P(r) = P'_{out}(r) = \mathbb{E}_{\mathbf{U}}[\delta(r - I_N(\mathbf{U}))] \tag{5.3}$$

The aim of this chapter is to calculate the tails of the distribution of the rate r. The first step is to express the joint distribution of eigenvalues of $\mathbf{U}^\dagger\mathbf{U}$ as derived initially in [73] and more recently in this context [15]

$$P_\lambda(\lambda_1 \dots \lambda_{N_t}) = \mathcal{Z}_{N_t}^{-1} \prod_{n<m} |\lambda_n - \lambda_m|^2 \prod_k \lambda_k^{|N_t - N_r|}(1-\lambda_k)^{N_0} \tag{5.4}$$

In the above, $\mathcal{Z}_{N_t}$ is a normalization constant defined in (C.1).

In the above equation, we can see that when N_0 becomes large, the last term can be approximated roughly as $(1-\lambda)^{N_0} \approx e^{-N_0\lambda}$. This corresponds to the eigenvalue distribution of a Wishart matrix [85], which is typically used to model wireless MIMO channel propagation because the latter has significant power loss in the atmosphere. Hence, it can be seen that the parameter N_0 can effectively model power loss through the fiber and provide a continuous cross-over between lossless and lossy fibers [73].

In contrast, as was shown in [15], for $N_0 < 0$, there is no outage below rates $r = |n| \log(1+\rho)$, because all transmission can be performed over channels with unit constant channel.

In the next section, we will show how the above expression can be used to provide a closed form solution for the outage probability, in terms of finite sums of simple functions. However, it will become clear that for increasing channel numbers, the formula becomes quite cumbersome, without providing much intuition. Hence, in Section 5.4 a different approach will be adopted, namely the large-N analysis of the outage probability using the Coulomb gas formalism.

5.3 Exact Solution

In this section, we will obtain a closed form expression for the outage probability $P_{out}(r)$. We start by introducing an integral representation for the Θ function

$$1 - \Theta(x) = \Theta(-x) = -\int_{-\infty}^{+\infty} \frac{dp}{2\pi \mathrm{i}} \frac{e^{\mathrm{i}px}}{p + \mathrm{i}\epsilon} \tag{5.5}$$

where ϵ is an infinitesimal positive number indicating that the k-integral goes over the pole at zero. As a result, the outage probability can be expressed as follows:

$$1 - P_{out}(r) = \int d\lambda P_\lambda(\lambda) \int_{-\infty}^{\infty} \frac{dp}{2\pi} \frac{e^{\mathrm{i}pN_t r}}{\epsilon - \mathrm{i}p} \prod_{n=1}^{N_t} (1 + \rho\lambda_n)^{-\mathrm{i}p} \tag{5.6}$$

where the integral notation $\int d\lambda$ signifies multiple integration over all λ_k for $k = 1, \cdots, N_t$. In Appendix B we show how the above multiple integral can be evaluated. The final result can be expressed as follows:

$$1 - P_{out}(r) = \sum_{\mathbf{k},\mathbf{n}} c_{\mathbf{k},\mathbf{n}} \sum_{\boldsymbol{\sigma}} (-1)^{|\boldsymbol{\sigma}|} \sum_{\ell=\ell(r)}^{N_t} (-1)^{\ell+N_t} d_\ell(\mathbf{s}_{\boldsymbol{\sigma}}) F(N_t r - \ell \log(1+\rho), \mathbf{s}_\sigma)$$

where the sum of $\mathbf{k}$ is over $[0, |N_t - N_r|]^{N_t}$, the sum of $\mathbf{n}$ is over $[0, N_0]^{N_t}$ and the sum over $\boldsymbol{\sigma}$ is over all permutations of $(1, \ldots, N_t)$ with signature $|\boldsymbol{\sigma}|$. The N_t-dimensional integer vector $\mathbf{s}_{\boldsymbol{\sigma}}$ has components $s_j = j + \sigma_j - 1 + k_j + N_0 - n_j$ and $\ell(r)$ is the smallest integer for which $N_t r < \ell \log(1+\rho)$, while

$$c_{\mathbf{k},\mathbf{n}} = \frac{N_t! \prod_{j=1}^{N_t} c_{k_j,n_j}}{\mathcal{Z}_N \rho^{N_t^2 + (|N_t - N_r| + N_0) N_t}} \tag{5.7}$$

$$d_\ell(\mathbf{s}) = \mathbf{e}_\ell\left((1+\rho)^{s_1}, \cdots, (1+\rho)^{s_{N_t}}\right) \tag{5.8}$$

$$F(z, \mathbf{s}) = \prod_{j=1}^{N_t} s_j^{-1} + \sum_{j=1}^{N_t} \frac{e^{s_j z}}{s_j \prod_{k \neq j} (s_k - s_j)} \tag{5.9}$$

In the above $c_{k,n}$ are given in (B.6), $\mathcal{Z}_{N_t}$ is given in (C.1), while $\mathbf{e}_\ell(x_1, x_2, \cdots, x_{N_t})$ is the elementary symmetric polynomial of degree ℓ [50]. The prescription of how to deal with $F(z, \mathbf{s})$ in the case where two or more integers s_i are equal is discussed in Appendix B. We also note that the density of r, $P(r)$, can be obtained directly from the above by differentiation with respect to r.

Although analytic and in closed form, the above result is handy and provides intuition for the answer at best for small values of N_t, N_r, N_0. When this is not the case, one needs an alternate path, which can be achieved using the asymptotic approach in N, which will be discussed next.

5.4 Coulomb Gas Methodology

In this section we will follow a complementary approach to the above and will derive the outage probability in the limit of large channel numbers. The first step is to rewrite the joint distribution of eigenvalues of $\mathbf{U}^\dagger \mathbf{U}$ provided in (5.4) in the following form

$$P_\lambda(\lambda_1 \ldots \lambda_{N_t}) = \mathcal{Z}_{N_t}^{-1} e^{-N_t^2 E(\lambda)} \tag{5.10}$$

where

$$E(\boldsymbol{\lambda}) = -\frac{N_0}{N_t^2} \sum_{k=1}^{N_t} \log(1 - \lambda_k) - \frac{N_r - N_t}{N_t^2} \sum_{k=1}^{N_t} \log(\lambda_k) - \frac{1}{N_t^2} \sum_{k \neq k'} \log|\lambda_k - \lambda_{k'}| \tag{5.11}$$

$E(\boldsymbol{\lambda})$ represents the normalized potential energy of N_t unit charges bound on the unit interval $x \in (0, 1)$, while repelling from each other and from the boundaries

logarithmically. It is reasonable to expect that when N is large, the charges will coalesce to a smooth density $p(x)$. This hypothesis, which is originally due to Dyson [19], and is called the *Coulomb (or Dyson) gas approach*, has been used extensively in statistical physics [23, 51, 55] and more recently in communications [45]. This hypothesis was set in a more mathematical footing by [3] who proved that the large deviations of the law of the spectral density $p(x)$ can be described by a rate function corresponding to the continuum limit of $E(\boldsymbol{\lambda})$. [3] showed this for the case of the Wigner Gaussian matrices, while [34] generalized it to Wishart matrices. Their proof can be directly applied to this model by restricting the support of eigenvalues from $\lambda \in (0, \infty)$ to the unit interval $\lambda \in (0, 1)$. We will apply this formalism to obtain the tails of $P_{out}(r)$. The first result is summarized in the following theorem, which is proved in Appendix C.1. Let us first denote by $\mathcal{X}$ the space of probability measures on $(0, 1)$, endowed with weak topology.

Theorem 5.1 (Large Deviations of Eigenvalue Density)**.** *Let*

$$\begin{aligned}\mathcal{E}[p] = &- n_0 \int p(x) \log(1-x) dx - (\beta - 1) \int p(x) \log(x) dx \\ &- \iint p(x) p(y) \log|x - y| dy dx \end{aligned} \tag{5.12}$$

where $p(x) \in \mathcal{X}$. Then

1. *$\mathcal{E}[p]$ is convex on $\mathcal{X}$.*
2. *$\mathcal{E}[p]$ obtains its minimum value denoted by $\mathcal{E}_0$ at a unique probability density $p_0(x)$ on $(0, 1)$.*
3. *$\lim_{N_t \to \infty} \frac{1}{N_t^2} \log P(I_N \leq r) = \mathcal{E}_0 - \inf_{p \in \mathcal{X}_r} \mathcal{E}[p]$ where*

$$\mathcal{X}_r = \left\{ p \in \mathcal{X} \text{ and } \int_0^1 p(x) \log(1 + \rho x)\, dx \leq r \right\} \tag{5.13}$$

For concreteness and future use, we define as r_{erg} the rate for the probability density $p_0(x)$, namely

$$r_{erg} = \int_0^1 p_0(x) \log(1 + \rho x)\, dx \tag{5.14}$$

In this chapter we mostly interested in the outage probability defined in (5.1) and therefore the above result is of interest. However, an analogous result can be obtained for the $1 - P_{out}(r)$:

Corollary 5.1. *If $\mathcal{X}_r$ includes the density $p_0(x)$ (or is arbitrarily close to it), then from the above we conclude that $\inf_{p \in \mathcal{X}_r} \mathcal{E}[p] = \mathcal{E}_0$ and hence for $r > r_{erg}$ we have $\log P(I_N \leq r)/N_t^2 \to 0$. Hence, in this case, we do not strictly speaking have a large*

deviation event. Nevertheless, in this case it can be shown that the complement of $\mathcal{X}_r$*, namely*

$$\mathcal{X}_r^c = \left\{ p \in \mathcal{X} \ \textit{and} \ \int_0^1 p(x) \log(1+\rho x)\, dx > \ r \right\} \tag{5.15}$$

is a large deviation event if $r > r_{erg}$*, i.e.*

$$\lim_{N_t \to \infty} \frac{1}{N_t^2} \log P(I_N > r) = \mathcal{E}_0 - \inf_{p \in \mathcal{X}_r^c} \mathcal{E}[p] \tag{5.16}$$

Due to the convexity of $\mathcal{E}[p]$ and $\mathcal{X}_r$, it is sufficient to find a local minimum of the functional, subject to the constraints, which then is ensured to be a global minimum. One handy way to do so is to introduce a Lagrangian and include the constraints of normalization and positivity of $p(x)$ using Lagrange multipliers. We thus have

$$\mathcal{L}_0[p, \nu, c] = \mathcal{E}[p] - c\left(\int_0^1 p(x)\, dx - 1 \right) - \int_0^1 \nu(x) p(x)\, dx \tag{5.17}$$

$$\mathcal{L}[p, \nu, c, k] = \mathcal{L}_0[p, \nu, c] - k\left(\int_0^1 p(x) \log(1+\rho x)\, dx - r \right) \tag{5.18}$$

from which we obtain $\mathcal{E}_0$ and $\mathcal{E}(r)$ by maximizing over the dual parameters ν (non-negativity constraint), c (normalization constraint) and k (mutual information constraint):

$$\mathcal{E}_0 = \sup_{\nu \geq 0;\, c} \inf_p \mathcal{L}_0[p, \nu, c] \tag{5.19}$$

$$\mathcal{E}(r) = \sup_{\nu \geq 0;\, c,k} \inf_p \mathcal{L}[p, \nu, c, k] \tag{5.20}$$

Note that the functional form of (5.18) amounts to replacing the mutual information inequality constraints in (5.13) and (5.15) with an equality constraint. Hence, to obtain the large deviation results above, we need to optimize of $\mathcal{E}(r)$ as follows

$$\lim_{N_t \to \infty} \frac{1}{N_t^2} \log P(I_N < r) = \mathcal{E}_0 - \inf_{r' < r} \mathcal{E}(r') \tag{5.21}$$

$$\lim_{N_t \to \infty} \frac{1}{N_t^2} \log P(I_N > r) = \mathcal{E}_0 - \inf_{r' > r} \mathcal{E}(r') \tag{5.22}$$

It is useful to summarize here some properties of $\mathcal{E}(r)$:

Lemma 5.1 (Properties of $\mathcal{E}(r)$)**.**

1. *For $r > r_{erg}$, $\mathcal{E}(r)$ is an increasing function of r, i.e. $\mathcal{E}'(r) > 0$.*
2. *For $r < r_{erg}$, $\mathcal{E}(r)$ is an increasing function of r, i.e. $\mathcal{E}'(r) < 0$.*
3. *$\mathcal{E}(r)$ is a convex function of r.*

The proof is deferred for Appendix C.2

The convexity of $\mathcal{L}_0$, $\mathcal{L}$ over p ensures uniqueness of the minimizing p. Therefore, if we find a local minimum for the corresponding Lagrangian for k, c and ν that satisfy the constraints, this will be a unique one. It is also worth pointing out that the only difference between $\mathcal{E}_0$ and $\mathcal{E}(r)$ above is that the former can be seen as the maximum over $\mathcal{L}[p, \nu, c, k]$ keeping $k = 0$; this relation will come in handy later, because it allows us to work with $\mathcal{L}$ and at the very last step set $k = 0$ to obtain $\mathcal{E}_0$. This result is in agreement with (C.2) derived in Appendix C.1 using other methods. To find a local minimum of $\mathcal{L}$, it suffices to calculate its functional derivative with respect to p and which is then set to zero. Note that the functional derivative of $\mathcal{L}$ at $p \in \mathcal{X}_r$ is the distribution $\delta\mathcal{L}[p, \nu, c, k]$ whose action on test functions which leave $\mathcal{E}[p]$ finite is given by:

$$\langle \delta\mathcal{L}[p], \phi \rangle = \frac{d}{dt}\bigg|_{t=0} \mathcal{L}[p + t\phi]. \tag{5.23}$$

Note that maximizing the result with respect to k and c simply corresponds to enforcing the normalization and mutual information constraints that appear in (5.17) and (5.18):

$$\int_0^1 p(x)\, dx = 1 \tag{5.24}$$

$$\int_0^1 p(x) \log(1 + \rho x)\, dx = r \tag{5.25}$$

The maximization over $\nu(x)$ ensures the non-negativity of $p(x)$. It can be shown [8] that either $\nu(x)$ or $p(x)$ are non-zero, therefore making $\nu(x)p(x) = 0$. For simplicity we will not analyze this constraint, instead enforcing it explicitly. Setting the functional derivative of $\mathcal{L}[p]$ to zero results to

$$2 \int p(x') \log|x - x'| dx' = -n_0 \log(1 - x) - c - k \log(1 + \rho x) - (\beta - 1) \log(x) \tag{5.26}$$

for all x in the support of $p(x)$. Once again, due to the convexity of $\mathcal{E}[p]$ and $\mathcal{X}_r$, it is sufficient to find a local minimum of the functional, subject to the constraints, which then is ensured to be a global minimum. Hence, we look for a solution for the above equation with $p(x)$ having support on the (connected) interval $(a, b) \subseteq (0, 1)$, with a, b to be determined. Taking the derivative with respect to x in the above we obtain

the following integral equation, which has the physically intuitive meaning of force balancing at the charges in x:

$$2\mathcal{P}\int_a^b \frac{p(x')}{x-x'}dx' = \frac{n_0}{1-x} - \frac{\beta-1}{x} - \frac{k\rho}{1+\rho x} \tag{5.27}$$

where $\mathcal{P}$ represents the Cauchy principal value of the integral. Once $p(x)$ has been determined, we can obtain $\mathcal{E}(r)$ by direct integration. To evaluate the double integral in (5.12) we can integrate using (5.26). Then the value of c can be determined by calculating (5.26) at $x' = a$ [45,83]. Following Tricomi's theorem [45,80] this integral equation may be solved to yield the following general expression

$$p(x) = \frac{\frac{n_0\sqrt{(1-a)(1-b)}}{1-x} - \frac{k\sqrt{(1+a\rho)(1+b\rho)}}{1+\rho x} - \frac{(\beta-1)\sqrt{ab}}{x} + C}{2\pi\sqrt{(x-a)(b-x)}} \tag{5.28}$$

where C is a constant. This is a valid solution if the right hand side expression of (5.27) is $L^{1+\epsilon}$ integrable (for some $\epsilon > 0$) over the support (a, b). Clearly, this is not the case if $a = 0$ or $b = 1$, whenever $\beta > 1$ or $n > 0$, respectively. Therefore, in those cases the values of a and b need to be found self-consistently, by demanding that $p(x)$ is continuous at that value, i.e. that $p(a > 0) = 0$ or $p(b < 1) = 0$. As a result, we find four types of solutions, depending on whether $a = 0$ and/or $b = 1$. Before summarizing the solution results for these four cases, we obtain the solution for the case $k = 0$, which corresponds to most probable value of $r = r_{erg}$. In this case, the eigenvalue distribution that minimizes $\mathcal{L}_0$ is simply

$$p_0(x) = \frac{\sqrt{(x-a_0)(b_0-x)}}{2\pi x(1-x)} \tag{5.29}$$

where

$$a_0, b_0 = \frac{\left(\sqrt{1+n_0} \pm \sqrt{\beta(n_0+\beta)}\right)^2}{n_0+1+\beta} \tag{5.30}$$

which has been obtained using other methods in [17, 73]. From the above $p_0(x)$, $\mathcal{E}_0$ can be evaluated. The result thus obtained matches the result obtained using a more direct method in Appendix C.1.

In the next sections we will obtain the solution for $\Delta\mathcal{E}(r)$ for all allowed values of parameters n, β, r. It should be stressed that given the convexity of $\mathcal{E}[p]$ with respect to p, it is sufficient to find an acceptable solution of the constrained extremization procedure discussed above. Below we will analyze the four possible types of solutions, corresponding to $a = 0$ or $a > 0$ and $b = 1$ or $b < 1$. We will see that for any parameter value of n_0, β, r, there is a single solution to the Tricomi equation above (5.27), which is consistent with all constraints, as well as positivity and continuity on $(0, 1)$. We will see that while continuity will exclude some types of solutions, e.g. $a = 0$ when $\beta > 1$ and $b = 1$ when $n_0 > 0$, we will find two or three types of solutions applicable for a given set

of n_0 and β. Of course, only one is valid for any given value of r. We will see that there is a critical value of r, at which one type of solution becomes invalid, while another becomes applicable. This phase transition is characterized with the attachment of the support of $p(x)$ to a boundary of $(0,1)$ and has been in the literature with a third order phase transition and the Tracy-Widom law [16,45,51,83]. In Table 5.1 we summarize the validity of each solution type, denoted by S_{01}, S_{a1}, S_{0b} and S_{ab}, where the first index describes the infimum of the support (0 if $a = 0$ and a if $a > 0$) and the second corresponds to its supremum (1 if $b = 1$ and b if $b < 1$).

	S_{0b}	S_{ab}	S_{01}	S_{a1}
	$a = 0$ $b < 1$	$a > 0$ $b < 1$	$a = 0$ $b = 1$	$a > 0$ $b = 1$
$n_0 = 0;\ \beta = 1$	$r < r_{c1}$	–	$r_{c1} < r < r_{c2}$	$r > r_{c2}$
$n_0 > 0;\ \beta = 1$	$r < r_{c3}$	$r > r_{c3}$	–	–
$n_0 = 0;\ \beta > 1$	–	$r < r_{c4}$	–	$r > r_{c4}$
$n_0 > 0;\ \beta > 1$	–	all r	–	–

Table 5.1: Summary of validity of four types of solutions depending on the values of n_0, β and r.

Solution S_{01}: $a = 0$, $b = 1$

We start with the most trivial type of solution, namely when the support boundaries $a = 0$ and $b = 1$ are enforced. This solution can be valid only when $n_0 = 0$ and $\beta = 1$, since otherwise the right-hand-side of (5.27) and hence also $p(x)$ [80] will not be $L^{1+\epsilon}$-integrable. The resulting optimal normalized spectral density is

$$p(x) = \frac{(1+\rho x)(k+2) - k\sqrt{1+\rho}}{2\pi(1+\rho x)\sqrt{x(1-x)}} \tag{5.31}$$

The resulting relation between r and k obtained by enforcing the rate constraint is (5.25)

$$r = r(k) \equiv \log \frac{(1+\sqrt{1+\rho})^2}{4} + k \log \frac{(1+\sqrt{1+\rho})^2}{4\sqrt{1+\rho}} \tag{5.32}$$

and the corresponding value of the exponent $\Delta\mathcal{E} = \mathcal{E}(r) - \mathcal{E}_0$ becomes quadratic

$$\Delta\mathcal{E} = \frac{\left(r - 2\log\frac{1+\sqrt{1+\rho}}{2}\right)^2}{2\log\frac{(1+\sqrt{1+\rho})^2}{4\sqrt{1+\rho}}} \tag{5.33}$$

The validity of the above result breaks down when the positivity constraint of $p(x)$ is violated. This happens when $k < k_{c1}$ or $k > k_{c2}$, where

$$k_{c1} = -\frac{2\sqrt{1+\rho}}{\sqrt{1+\rho}-1}, \qquad k_{c2} = \frac{2}{\sqrt{1+\rho}-1} \tag{5.34}$$

with corresponding values of the rate obtained through $r < r_{c1} = r(k_{c1})$ and $r > r_{c2} = r(k_{c2})$, respectively. If this is true, we need seek for a solution allowing $b < 1$, or $a > 0$, respectively. This will be analyzed in the next two subsections.

Solution S_{0b}: $a = 0$, $b < 1$

This solution can only be valid for $\beta = 1$. In this case the resulting optimal eigenvalue density is given by

$$p(x) = \frac{1}{2\pi}\sqrt{\frac{b-x}{x}}\left(\frac{n_0}{\sqrt{1-b}}\frac{1}{1-x} - \frac{k}{\sqrt{1+\rho b}}\frac{\rho}{1+\rho x}\right) \tag{5.35}$$

The normalization condition (5.24) gives

$$\frac{n}{\sqrt{1-b}} + k\sqrt{\frac{1}{1+\rho b}} = 2 + n_0 + k \tag{5.36}$$

which is shown in Appendix C.3 to have a unique solution, while the rate equality (5.25) condition gives

$$\begin{aligned} r = r(k) \equiv &\log(\rho b) + \frac{nb}{2\sqrt{1-b}}\left(G(\frac{1}{\rho b}, 0) - G(\frac{1}{\rho b}, -\frac{1}{b})\right) \\ &- \frac{k\rho b}{2\sqrt{1+\rho b}}\left(G(\frac{1}{\rho b}, 0) - G(\frac{1}{\rho b}, \frac{1}{\rho b})\right) \end{aligned} \tag{5.37}$$

and finally

$$\begin{aligned} \mathcal{E}(r) = &\frac{k}{2}\Big[r - \log\left(1+b\rho\right)\Big] - \frac{n_0\log(1-b)}{2} - \frac{(n_0+2)\log b}{2} \\ &- \frac{n_0^2 b}{4\sqrt{1-b}}\left(I_3(\frac{1}{b}-1) + G(\frac{1}{b}-1, \frac{1}{b}-1)\right) \\ &+ \frac{n_0 k\rho b}{4\sqrt{1+\rho b}}\left(I_3(\frac{1}{b}-1) + G(\frac{1}{b}-1, -1-\frac{1}{\rho b})\right) \\ &- \frac{n_0 b}{2\sqrt{1-b}}\left(I_3(0) + G(0, -1-\frac{1}{b})\right) + \frac{k\rho b}{2\sqrt{1+\rho b}}\left(I_3(0) + G(0, -1-\frac{1}{\rho b})\right) \end{aligned} \tag{5.38}$$

In the above we have introduced the functions $G(x, y)$ and $I_3(x)$. Their definition and closed-form expressions can be seen in Appendix E, where we have used results from

[12]. When $n_0 = 0$, (5.36) breaks down (and hence $p(x)$ is not properly normalized) if $k > k_{c1}$, assuming of course $b \leq 1$. Hence, in this case this solution is invalid in agreement with the discussion in the previous subsection.

In contrast when $n > 0$, the above solution breaks down when $p(x) < 0$ for small x. This happens when, in addition to (5.36) $n_0(1+\rho) < (2+n+k_{c3})\rho\sqrt{1-b_{c3}}$, which corresponds to $r > r_{c3} = r(k_{c3})$. In this case, we need to allow $a > 0$, which will be analyzed in a later subsection.

Solution S_{a1}: $a > 0$, $b = 1$

In the spirit of previous subsections, this solution can only be valid when $n_0 = 0$. In this case the resulting optimal eigenvalue density is given by

$$p(x) = \frac{\sqrt{x-a}}{2\pi\sqrt{1-x}}\left[k\sqrt{\frac{1+\rho}{1+\rho a}}\frac{\rho}{1+\rho x} + \frac{\beta-1}{x}\sqrt{\frac{1}{a}}\right] \tag{5.39}$$

Using the normalization equation

$$\beta + 1 + k = \frac{\beta-1}{\sqrt{a}} + \frac{k\sqrt{1+\rho}}{\sqrt{1+\rho a}} \tag{5.40}$$

and the rate constraint (5.25)

$$\begin{aligned} r =& r(k) \equiv \log(\rho(1-a)) \\ &+ \frac{k(1-a)\rho}{2\sqrt{(1+\rho)(1+\rho a)}}\left(I_3(\frac{a+\rho^{-1}}{1-a}) + G(\frac{a+\rho^{-1}}{1-a}, \frac{a+\rho^{-1}}{1-a})\right) \\ &+ \frac{(\beta-1)(1-a)}{2\sqrt{a}}\left(I_3(\frac{a+\rho^{-1}}{1-a}) + G(\frac{a+\rho^{-1}}{1-a}, \frac{a}{1-a})\right) \end{aligned} \tag{5.41}$$

we can finally calculate $\mathcal{E}(r)$

$$\begin{aligned} \mathcal{E}(r) =& \frac{k}{2}\left(r - \log\left(1+a\rho\right)\right) - \frac{\beta-1}{2}\log a - \frac{\beta+1}{2}\log(1-a) \\ &- \frac{(\beta-1)^2(1-a)}{4\sqrt{a}}\left(I_3(\frac{a}{1-a}) + G(\frac{a}{1-a}, \frac{a}{1-a})\right) \\ &- \frac{(\beta-1)k\rho(1-a)}{4\sqrt{(1+\rho a)(1+\rho)}}\left(I_3(\frac{a}{1-a}) + G(\frac{a}{1-a}, \frac{a+\rho^{-1}}{1-a})\right) \\ &- \frac{(\beta-1)(1-a)}{2\sqrt{a}}\left(I_3(0) + G(0, \frac{a}{1-a})\right) \\ &- \frac{k\rho(1-a)}{4\sqrt{(1+\rho a)(1+\rho)}}\left(I_3(0) + G(0, \frac{a+\rho^{-1}}{1-a})\right) \end{aligned} \tag{5.42}$$

When $\beta = 1$, for $k < k_{c2}$, where k_{c2} is defined in (5.34), (5.40) gives $a < 0$, which is obviously not allowed, hence invalidating this solution. This is in agreement with subsection 5.4.

In contrast when $\beta > 1$, the above solution breaks down when $p(x) < 0$ for $x \approx 1$. This happens when, in addition to (5.40) we have $(\beta - 1) + (1 + \beta + k_{c4})\rho\sqrt{a_{c4}} = 0$, which corresponds to $r = r_{c4} = r(k_{c4})$. In this case, we need to also allow $b < 1$, which will be analyzed below.

Solution S_{ab}: $a > 0$, $b < 1$

The final, more general case includes generic a and b. In this case the resulting optimal eigenvalue density is given by

$$p(x) = \frac{\sqrt{(x-a)(b-x)}}{2\pi(1+\rho x)}\left(\frac{n_0(\rho+1)}{(1-x)\sqrt{(1-a)(1-b)}} + \frac{\beta-1}{x\sqrt{ab}}\right) \tag{5.43}$$

with the additional constraint

$$\frac{n_0}{\sqrt{(1-a)(1-b)}} = \frac{\beta-1}{\sqrt{ab}} + \frac{k\rho}{\sqrt{(1+\rho a)(1+\rho b)}} \tag{5.44}$$

obtained by demanding $p(a) = p(b) = 0$. The parameters a, b, k can be evaluated uniquely from the above equation in addition to the normalization constraint (5.24)

$$n_0 + \beta + 1 + k = \frac{\beta-1}{\sqrt{ab}} + \frac{k(1+\rho)}{\sqrt{(1+\rho a)(1+\rho b)}} \tag{5.45}$$

and the rate constraint (5.25)

$$\begin{aligned} r = r(k) \equiv \log\rho\Delta + \frac{n}{2\sqrt{\bar{a}_c\bar{b}_c}}\Big[G\left(\bar{a}_z, \bar{a}_z\right) - G\left(\bar{a}_z, -\bar{a}_c\right)\Big] \\ + \frac{(\beta-1)}{2\sqrt{\bar{a}\bar{b}}}\Big[G\left(\bar{a}_z, \bar{a}\right) - G\left(\bar{a}_z, \bar{a}_z\right)\Big] \end{aligned} \tag{5.46}$$

where $\Delta = b - a$. For notational simplicity we also define $\bar{a} = a/\Delta$, $a_c = 1 - a$, $\bar{a}_c = (1-a_c)/\Delta$, $\bar{a}_z = (a+\rho^{-1})/\Delta$ and $\bar{b} = b/\Delta$, $\bar{b}_c = b_c/\Delta = (1-b)/\Delta$, $\bar{b}_z = (b+\rho^{-1})/\Delta$. The G function can be seen in Appendix E. We may now integrate over $p(x)$ and obtain an expression for $\mathcal{E}(r)$ as follows

$$\begin{aligned} \mathcal{E}(r) =& \frac{k}{2}\left(r - \log\left(1+b\rho\right)\right) - \frac{\log\Delta}{2}\left(n_0+\beta+1\right) - \frac{n}{2}\log b_c \\ &- \frac{n_0^2}{4\sqrt{\bar{a}_c\bar{b}_c}}\left(G(\bar{b}_c, \bar{b}_c) - G(\bar{b}_c, -\bar{b}_z)\right) - \frac{(\beta-1)}{2}\log b \\ &+ \frac{n_0(\beta-1)}{4\sqrt{\bar{a}\bar{b}}}\left(G(\bar{b}_c, -\bar{b}) - G(\bar{b}_c, -\bar{b}_z)\right) \\ &+ \frac{n_0(\beta-1)}{4\sqrt{\bar{a}_c\bar{b}_c}}\left(G(\bar{a}, -\bar{a}_c) - G(\bar{a}, \bar{a}_z)\right) - \frac{(\beta-1)^2}{4\sqrt{\bar{a}\bar{b}}}\left(G(\bar{a}, \bar{a}) - G(\bar{a}, \bar{a}_z)\right) \end{aligned}$$

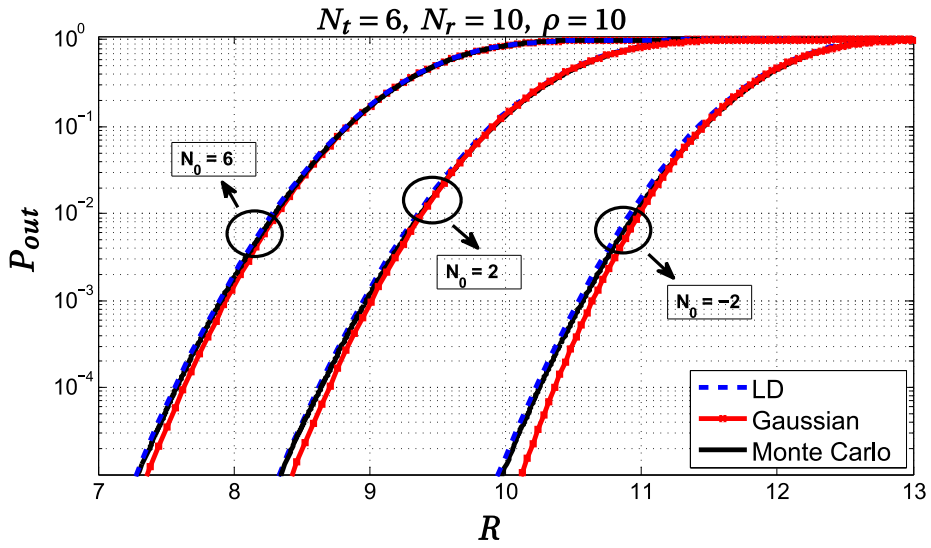


Figure 5.2: Outage probability curves as a function of transmitted rate for moderate signal-to-noise ratio $\rho = 10$ and three different values of N_0. For each set of parameters we plot three curves, namely the LD (blue dashed) curves (corresponding to the current large deviations methodology), the numerically generated curves (in black) and the ones corresponding to the Gaussian approximation (red dot-dashed). First, we observe that with decreasing values of N_0 the whole curve shifts to the right signifying increased throughput. Notice that this is so even when $N_0 = -2 < 0$. In this case, as mentioned also in [15], a number of channels ($|N_0|$) becomes deterministic without fading. Second, we observe that while the LD curves follow the numerical curves very closely, the Gaussian curves tend to deviate at low outages. Even though the curves look to be close, the outage difference between LD and Gaussian come out to about a factor of 2.

$$-\frac{n_0}{2\sqrt{\bar{a}_c\bar{b}_c}}\left(G(0,\bar{b}_c)-G(0,-\bar{b}_z)\right)+\frac{\beta-1}{2\sqrt{\bar{a}\bar{b}}}\left(G(0,-\bar{b})-G(0,-\bar{b}_z)\right) \tag{5.47}$$

To make contact with the solutions of the previous sections, we observe that the conditions (5.44) and (5.45) cannot be simultaneously be satisfied if $\beta = 1$, $n > 0$, and $k < k_{c3}$ (corresponding to $r < r_{c3}$) unless $a < 0$. In this parameter region S_{0b} applies. Also, for $\beta > 1$, $n = 0$, and $k > k_{c4}$ (and correspondingly $r > r_{c4}$), the above equations result to $b > 1$, thereby invalidating the solution and necessitating the solution S_{a1}. In conclusion, we see that the above four solutions are mutually exclusive and cover all possible parameter values, thereby providing the unique solution to the exponent $\mathcal{E}(r)$ of the outage probability.

Probability Distributions $P(r)$ and $P_{out}(r)$

In the previous sections we obtained the asymptotic behavior of the outage probability in the large N_t limit. We found that the outage probability is approximately $P(I_N \leq r) \sim \exp[-N_t^2(\mathcal{E}(r) - \mathcal{E}_0)]$ when $r < r_{erg}$ and we can similarly find for

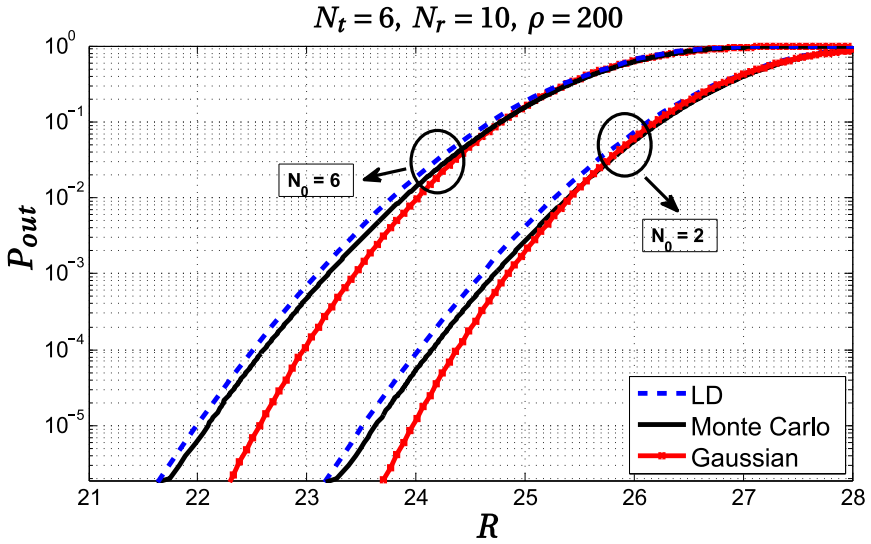


Figure 5.3: Outage probability curves as a function of transmitted rate for large signal-to-noise ratio $\rho = 100$ and two different values of N_0. As in the previous figure, three curves are plotted for each parameter set, namely the LD (blue dashed) curves, the numerically generated curves (in black) and the Gaussian (red dot-dashed) curves. The LD curves continue to follow the numerical curves very closely, however the Gaussian curves deviate much more from the numerical curves and the deviation can reach a factor of 20. Hence, for large SNR the Gaussian approximation does not provide such good results as the LD approach.

$r > r_{erg}$ that $P(I_N > r) \sim \exp[-N_t^2(\mathcal{E}(r) - \mathcal{E}_0)]$. By differentiation we obtain to leading exponential order that the probability density follows the same law, i.e. $P(r) \sim \exp[-N_t^2(\mathcal{E}(r) - \mathcal{E}_0)]$. To obtain the normalization constant for the density, we observe that the distribution close to its peak will be asymptotically Gaussian. This can be checked by calculating $\mathcal{E}(r)$ in the small k limit and showing that it is quadratic in r. Hence the normalization of the distribution will be given for large N_t by the variance of the distribution close to the peak. Therefore, we obtain

$$P(r) \approx N_t \frac{e^{-N_t^2(\mathcal{E}(r)-\mathcal{E}_0)}}{\sqrt{2\pi v_{erg}}} \tag{5.48}$$

where v_{erg} is the variance at the peak of the distribution. Correspondingly r_{erg} is the solution of $r(k = 0)$ in (5.32), (5.37), (5.41) or (5.46) (depending on the values of n, β) and corresponds to the ergodic rate. To obtain the value for v_{erg} we observe that $\mathcal{E}'(r) = k(r)$, which is negative for $r < r_{erg}$ and positive for $r > r_{erg}$. Similarly, we can obtain the local variance by differentiating once again $\mathcal{E}''(r) = dk(r)/dr$. Setting $k = 0$, it follows that

$$v_{erg} = \int_{a_0}^{b_0} dx \frac{dp(x,k)}{dk} \log(1+\rho x) = \log \frac{(\sqrt{1+\rho b_0} + \sqrt{1+\rho a_0})^2}{4\sqrt{1+\rho b_0}\sqrt{1+\rho a_0}} \tag{5.49}$$

where a_0, b_0 are given in (5.30).

To obtain an expression for the outage probability that is continuous at $k = 0$, we may integrate $P(r)$ above from 0 to r and noticing that due to the exponential dependence on N_t, only the region close to r will be important. Thus for $r < r_{erg}$ the outage probability is

$$P_{out}(r) \approx \frac{e^{-N^2[\mathcal{E}(r)-\mathcal{E}_0-\frac{\mathcal{E}'(r)^2}{2\mathcal{E}''(r)}]}Q\left(\frac{N|\mathcal{E}'(r)|}{\sqrt{\mathcal{E}''(r)}}\right)}{\sqrt{\mathcal{E}''(r)\upsilon_{erg}}} \tag{5.50}$$

and for $r > r_{erg}$ it is

$$P_{out}(r) \approx 1 - \frac{e^{-N^2[\mathcal{E}(r)-\mathcal{E}_0-\frac{\mathcal{E}'(r)^2}{2\mathcal{E}''(r)}]}Q\left(\frac{N|\mathcal{E}'(r)|}{\sqrt{\mathcal{E}''(r)}}\right)}{\sqrt{\mathcal{E}''(r)\upsilon_{erg}}} \tag{5.51}$$

and $\mathcal{E}'(r) = k(r)$ and $\mathcal{E}''(r) = k'(r)$ are the first and second derivative of $\mathcal{E}(r)$ with respect to r and $Q(x) = \int_x^\infty dt e^{-t^2/2}/\sqrt{2\pi}$. This approximation, while is essentially the same as (5.21) when N_t is large irrespective of r, but it is convenient, because it gives the crossover for fixed N_t and $r \approx r_{erg}$.

5.5 Numerical Simulations

The above section has provided us with the mathematical formulae to express the outage probability as a function of rate r for various values of β, n and ρ. We will now briefly describe the performance of the optical MIMO system and the accuracy of the analytic approach. Apart from the outage probability, which is expressed in (5.50) we will also show the behavior of the so-called Gaussian approximation. The corresponds to assuming the outage probability is given by an error function with variance υ_{erg} evaluated in (5.49) and mean the value of r_{erg} defined in (5.14), i.e.

$$1 - P_{out,g}(R) = Q\left(\frac{R - N_t r_{erg}}{\upsilon_{erg}}\right) \tag{5.52}$$

In Fig. 5.2 we plot the outage probability for moderate signal-to-noise ratios $\rho = 10$ for different values of N_0. We see that the outage curves are shifted to larger values of the rate as N_0 is reduced, in agreement with out intuition: the fewer channels we leave untapped, the more throughtput we can get from the transmission. This holds also for negative N_0, in which case, as [15] predicted, a number of channels ($|N_0|$) will become non-fading. We also see that the LD outage curves are in good agreement with numerics, while the Gaussian approximation also tends to follow along, with some small deviations.

The deviations of the Gaussian approximation become more pronounced for larger values of ρ, as in Fig. 5.3. For lower outages we see a significant deviation which reaches

in the figure to about a factor of 20. This behavior has been observed in the context of wireless MIMO tails [45]. The reason is that for larger ρ the importance of eigenvalues of the matrix that are of the order of $1/\rho$ become significant. However, in the Gaussian approximation these eigenvalues are "frozen" out of the problem. Nevertheless, the accuracy of the LD results is good.

We also tested the system for the case when $N_t = N_r$, i.e. $\beta = 1$, in Fig. 5.4. Here, and for moderate values of ρ, we see that the Gaussian approximation ceases to be valid deviating significantly from the numerical curves. Instead the LD approximation, although it also deviates, is relatively close to the numerical curves, suggesting that it is robustly correct overall, even for small channel numbers ($N_0 = N_t = N_r = 2$).

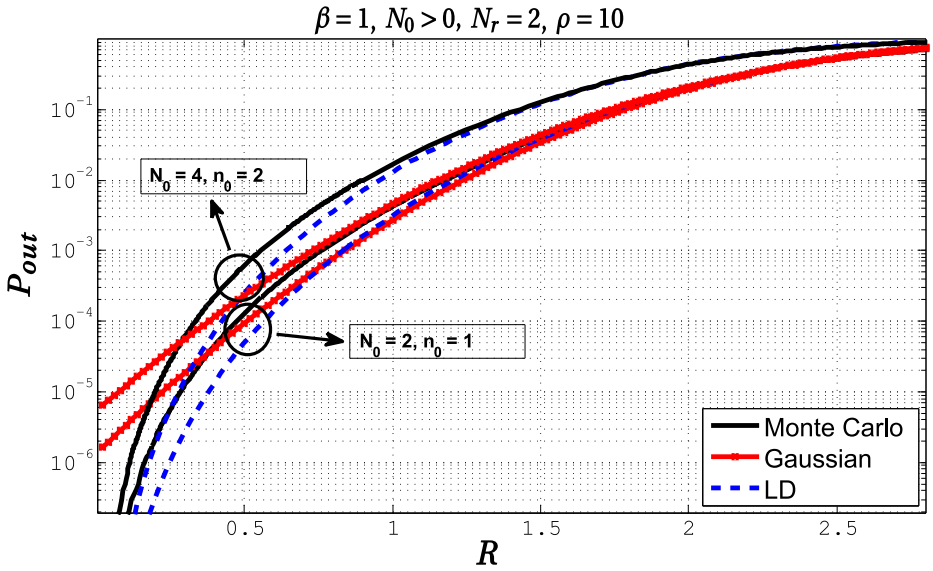


Figure 5.4: Outage probability curves as a function of transmitted rate for moderate signal-to-noise ratio $\rho = 10$ and two different values of N_0. In contrast to Fig. 5.2, here the parameter $\beta = 1$, i.e. $N_t = N_r$. As in the previous figure, three curves are plotted for each parameter set, namely the LD (blue dashed) curves, the numerically generated curves (in black) and the Gaussian (red dot-dashed) curves. In this case, the LD, while following close the numerical values, there is a larger deviation compared to the case $\beta > 1$. Furthermore, we see that the Gaussian deviate even more and finally for small rates become completely unreliable. This behavior is mainly due to the very small values of rates at low outages, making the Gaussian curve "feel" that there are no negative values of r.

5.6 Conclusions

The purpose of this chapter was to analyze the outage capacity for a particular model of the optical MIMO channel applicable to a multimode-multicore optical fiber system. The assumptions underlying the model assume strong forward scattering of light

between the modes/cores, while the backscattering is weak. At the same time, we can model loss inside the fiber by varying a particular parameter of the model, namely N_0. We have provided two complementary approaches to provide analytic solutions for the outage capacity. In the first, we derived closed-form expressions for the outage probability. Despite its exactness, this approach becomes cumbersome to use beyond the size of a few channels. Therefore, we also implemented a large deviation approach to calculate the outage capacity for the optical MIMO channel in the limit of large channel numbers. Our method is especially applicable for the tails of the distribution, which is relevant for low outage requirements due to the absence of feedback and finite SNR. Our analytical results agree very well with numerical experiments. On the other hand the Gaussian approximation fails to follow the respective numerical and the deviation becomes greater as our system increases in size and complexity (β and n). Additionally the method provides the distribution of eigenvalues constrained on the transmission rate and SNR. Although the channel assumptions taken here are somewhat idealized, this result gives an analytic metric to compare with other more complicated channel models. Clearly, more work is necessary, both from the channel sounding side, but also from the channel modeling side, so that the model will become more realistic.

CHAPTER 6

Gallager Bound

6.1 The Gallager Bound in the Wireless Domain

In this chapter we apply random matrix theory to evaluate the error probability exponent of the Gallager bound when the blocklength T, the number of N transmitting, and K receiving antennas, and the rate R all become large but at fixed ratios $\alpha = T/N$, $\beta = K/N$, $r = R/N$. Our large deviation result is valid for all normalized rates $0 < r < r_{erg}$. When we evaluate the error exponent for small $|r_{erg} - r| \ll 1$, our results match with the upper bound obtained by [37]. While the asymptotic limit of large antenna numbers is somewhat idealized, it is known from other works, e.g. [45] that even for moderate antenna numbers the asymptotic results become quite accurate. In addition, we explore the impact of fading in the channel by allowing the channel to take Q independent realizations within a codeword of length T.

Problem Formulation and Results

Channel Model and Capacity

Let us consider a MIMO link with N transmit and K receive antennas and analyze the transmission of T symbols. We assume a block fading channel, which remains constant over $\tau_Q = \left[\frac{T}{Q}\right]$ symbols and changes independently after each such coherence time [6]. Hence τ_Q is a parameter indicated by the bandwidth of the system and the fading statistics of the channel. Therefore, the memoryless channel reads

$$\mathbf{Y}_q = \mathbf{H}_q \mathbf{X}_q + \sigma \mathbf{W}_q \tag{6.1}$$

for $q = 1 \dots Q$, where $\mathbf{Y}_q \in \mathbb{C}^{K \times \tau_Q}$ is the received signal matrix during the qth block, $\mathbf{H}_q \in \mathbb{C}^{K \times N}$ is the channel matrix, whose entries are independent and identically distributed (i.i.d.) $\mathcal{CN}(0, \frac{1}{N})$, $\mathbf{X}_q \in \mathbb{C}^{N \times \tau_Q}$ is the transmitted signal matrix and $\sigma \mathbf{W}_q \in \mathbb{C}^{K \times \tau_Q}$ is the noise matrix with entries i.i.d. following $\mathcal{CN}(0, \sigma^2)$. For notational convenience we will denote $\mathbf{Y} = [\mathbf{Y}_1, \dots, \mathbf{Y}_Q]$, $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_Q]$, etc. The transmitter has only statistical knowledge of the channel, while the receiver knows it perfectly e.g., using a pilot signal. The mutual information per channel use over the qth block for Gaussian input with i.i.d. entries following $\mathcal{C}N(0, 1)$ is given by

$$C_q(\sigma^2, \mathbf{H}_q) = \log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{H}_q\mathbf{H}_q^\dagger\right). \tag{6.2}$$

The joint distribution of eigenvalues of $\mathbf{H}_q\mathbf{H}_q^\dagger$ is

$$P_{\boldsymbol{\lambda}}(\lambda_1 \dots \lambda_N) = \frac{1}{\mathcal{Z}_N} \prod_{N<i<j\leq K} |\lambda_i - \lambda_j|^2 \prod_i w(\lambda_i) = \frac{1}{\mathcal{Z}_N} e^{-N^2 E(\boldsymbol{\lambda})}, \tag{6.3}$$

where $\mathcal{Z}_N$ is the normalization constant, $w(\lambda)$ is a weight function, which depends on the statistics of $\mathbf{H}_q$ and the exponent $E(\boldsymbol{\lambda})$ is an energy functional of the eigenvalues $\{\lambda_i\}$ that will become useful later. Fot the case of complex Gaussian channels, the form of the weight function is $w(x) = x^{K-N}e^{-Nx}$. There are a number of other random matrix models for which the joint distribution of eigenvalues takes the same form with different realizations of $w(x)$. The value of the mutual information per antenna $C_q(\sigma^2, \mathbf{H}_q)/N$ converges weakly to a deterministic value in the large N limit, given by the ergodic average of the mutual information [59] (Eq. 105-106),

$$r_{erg}(\beta, \sigma^2) = \log u + \beta \log \left[1 + \frac{1}{u\,\sigma^2}\right] - (1 - u^{-1}), \tag{6.4}$$

with

$$u = \frac{1}{2\,\sigma^2}\left(\sigma^2 + \beta - 1 + \sqrt{(\sigma^2 + (\beta-1))^2 + 4\,\sigma^2}\right), \tag{6.5}$$

where $\beta = \frac{K}{N} > 1$. The empirical eigenvalue density of $\mathbf{H}_q\mathbf{H}_q^\dagger$ converges weakly to the well-known Marčenko-Pastur distribution [81] (Equation 1.12)

$$p_0(x) = \begin{cases} \frac{\sqrt{(b_0-x)(x-a_0)}}{2\pi x}, & \text{for } x \in [a_0, b_0] \\ 0, & \text{otherwise,} \end{cases} \tag{6.6}$$

where $a_0, b_0 = (\sqrt{\beta} \pm 1)^2$ are the endpoints of its support.

In the infinite codelength limit, the effect of the channel fading is captured through the optimal outage error probability [62] over the channel matrix $\mathbf{H}_q$, given by $p_{out} = \mathbb{P}(C/N < r)$ (in the case of $Q = 1$, $C \equiv C_1$). The exponent of the outage probability was analyzed in [45] when the number of antennas becomes large. There it was shown that when $K, N \to \infty$ with $\beta = K/N$ fixed, the outage probability behaves as

$$\lim_{N\to\infty} \frac{1}{N^2} \log \mathbb{P}\left(\frac{C}{N} < r\right) = -E_{out}(r), \tag{6.7}$$

where $E_{out}(r)$ close to $r = r_{erg}$ behaves as

$$E_{out}(r) = \frac{(r - r_{erg})^2}{2v_\infty} + o\left((r - r_{erg})^2\right), \tag{6.8}$$

where

$$v_\infty = -\log\left(1 - \frac{(1-u)^2}{\beta u^2}\right). \tag{6.9}$$

The above quantity is the dispersion of the mutual information distribution in the infinite codelength limit and will be called hereafter infinite codelength dispersion, in accordance with the names used for similar quantities in [2, 37, 64].

Gallager Exponent for Power-Constrained Input Alphabets

The error probability of transmission at a code rate of $R = Nr$ for a given instantiation of $\{\mathbf{H}_q\}$, $\mathbb{P}(\mathcal{E}|\{\mathbf{H}_q\})$ of a discrete memoryless channel without feedback and maximum likelihood (ML) decoding is bounded by (see Eq. (7.3.20) in [26])

$$\mathbb{P}(\mathcal{E}|\{\mathbf{H}_j\}) \leq e^{TNr} \int d\mathbf{Y} \left[\int d\mathbf{X}\ \mu_{con}(\mathbf{X}) \left[\mu(\mathbf{Y}|\mathbf{X}, \{\mathbf{H}_q\})\right]^{\frac{1}{1+\rho}} \right]^{1+\rho}, \tag{6.10}$$

where $\rho \in [0,1]$, $\mu(\mathbf{Y}|\mathbf{X}, \{\mathbf{H}_q\})$ is the distribution of the noise $\sigma\mathbf{W}$, while $\mu_{con}(\mathbf{X})$ is the distribution of $\mathbf{X}$ constrained to inputs such that only codewords with

$$\mathrm{Tr}\left[\mathbf{X}^\dagger \mathbf{X}\right] \leq NT \tag{6.11}$$

are used. This constraint can be enforced as an inequality by following (Eq. (7.3.17)) in [26] to observe that

$$\mu_{con}(\mathbf{X}) \leq \bar{c}\ \mu(\mathbf{X}) e^{s\left(\mathrm{Tr}[\mathbf{X}^\dagger \mathbf{X}] - NT\right)}, \tag{6.12}$$

for any $s > 0$, where $\mu(\mathbf{X})$ is the unconstrained input distribution assumed henceforth to be Gaussian and $\bar{c}$ a normalization constant. Integrating over $\mathbf{X}$, $\mathbf{Y}$ we obtain

$$\begin{aligned} \log \mathbb{P}(\mathcal{E}|\{\mathbf{H}_q\}) \leq -\frac{T}{Q} \sum_{q=1}^{Q} \Bigg[\rho \log\det\left(1 + \frac{1}{(1+\rho)(1-s)\,\sigma^2} \mathbf{H}_q \mathbf{H}_q^\dagger\right) \\ - \rho\, rN + (1+\rho)N\left(s + \log(1-s)\right) \Bigg], \end{aligned} \tag{6.13}$$

after omitting the normalization term $(\rho+1)\log\bar{c}$, which can be shown to be subleading in N [26]. After averaging $\mathbb{P}(\mathcal{E}|\{\mathbf{H}_q\})$ over $\{\mathbf{H}_q\}$ and optimizing over the values of ρ, s, we find that $\mathbb{P}(\mathcal{E})$, the average error rate after jointly decoding the total message sent over Q blocks is bounded by

$$\mathbb{P}(\mathcal{E}) = \mathbb{E}_{\{\mathbf{H}_q\}}\left[\mathbb{P}\left(\mathcal{E}|\{\mathbf{H}_q\}\right)\right] \leq \mathbb{E}_{\{\mathbf{H}_q\}}\left[e^{-N^2 E(r|\{\mathbf{H}_q\})}\right], \tag{6.14}$$

where

$$\begin{aligned} E(r|\{\mathbf{H}_q\}) = \max_{\substack{\rho\in[0,1] \\ s\in[0,1)}} \Bigg\{ \frac{\alpha}{Q} \sum_{q=1}^{Q} \Bigg[\frac{\rho}{N} \log\det\left(1 + \frac{1}{(1+\rho)(1-s)\,\sigma^2} \mathbf{H}_q \mathbf{H}_q^\dagger\right) \\ - \rho\, r + (1+\rho)\left(s + \log(1-s)\right) \Bigg] \Bigg\}. \end{aligned} \tag{6.15}$$

In the above, $\alpha = T/N$ and $r = R/N$ is the per-antenna rate and σ^{-2} is the SNR. We then define the Gallager exponent as

$$E_N(r) = -\frac{1}{N^2} \log \mathbb{E}_{\{\mathbf{H}_q\}}\left[e^{-N^2 E(r|\{\mathbf{H}_q\})}\right]. \tag{6.16}$$

It should be stressed that while in single link transmission schemes the exponent of the probability of error scales with the blocklength T, in MIMO systems it should be proportional to NT, which is the number of symbols transmitted. To be able to compare with the infinite codelength error exponent defined in the previous section, we have chosen to re-scale the error exponent in the same way (i.e. with N^2), adding a factor of α in (6.44). We then take the limit $N, K, T \to \infty$, while at the same time keeping the ratios $\beta = K/N$ and $\alpha = T/N$ fixed. The analytic evaluation of the error exponent $E_N(r)$ in this limit is the main result of this section and is summarized by the following theorem.

Theorem 6.1. *The limit of the error exponent $E(r) = \lim_{N\to\infty} E_N(r)$ exists and can be expressed as*

$$\begin{aligned} E(r) =& Q \max_{\substack{\rho\in[0,1]\\ s\in[0,1)}} \Bigg[-\int_a^b \int_a^b \log|x-y| p^*(x)p(y)dxdy + \int_a^b (x-(\beta-1))\log(x)p^*(x)dx \\ &+ \frac{\alpha}{Q}\left(\rho \int_a^b \log\left(1+\frac{x}{z_{\rho s}}\right) p^*(x)dx - \rho r + (1+\rho)\left(s+\log(1-s)\right)\right) \\ &- \frac{1}{2}\left(3\beta - \beta^2\log\beta + (\beta-1)^2\log(\beta-1)\right)\Bigg], \end{aligned} \tag{6.17}$$

where

$$p^*(x) = \frac{\sqrt{(x-a)(b-x)}}{2\pi x(x+z_{\rho s})}\left[x + z_{\rho s} + \frac{\alpha\,\rho\, z_{\rho s}}{Q\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}}\right], \tag{6.18}$$

and $z_{\rho s} = (1+\rho)(1-s)\,\sigma^2$. The values of the parameters a, b and s, as functions of ρ, are the unique solutions of the following equations:

$$\frac{\beta-1}{\sqrt{ab}} - \frac{\rho\,\alpha}{Q\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}} = 1, \tag{6.19}$$

$$a + b + 2\frac{\rho\,\alpha}{Q} - 2(\beta+1) = \frac{2\,\rho\,\alpha z_{\rho s}}{Q\sqrt{(a+z_{\rho s})(b+z_{\rho s})}}, \tag{6.20}$$

$$s = \frac{\rho}{4(1+\rho)}\left(\sqrt{z_{\rho s}+b} - \sqrt{z_{\rho s}+a}\right)^2. \tag{6.21}$$

Having determined these parameters as functions of ρ, ρ is determined from r as follows. Defining the function $\bar{r}(\rho)$ as

$$\bar{r}(\rho) = \log(1-s) + \int_a^b p^*(x)\log\left(1+\frac{x}{z_{\rho s}}\right)dx \tag{6.22}$$

$$= \log \frac{\Delta(1-s)}{z_{\rho s}} + \frac{\Delta}{2}\left(1 + \frac{\rho\,\alpha}{\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}}\right) G\left(\frac{z_{\rho s}+a}{\Delta}, \frac{a}{\Delta}\right)$$
$$- \frac{\Delta\,\rho\,\alpha}{2\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}} G\left(\frac{z_{\rho s}+a}{\Delta}, \frac{z_{\rho s}+a}{\Delta}\right), \tag{6.23}$$

where the function $G(x,y)$ can be seen in Appendix E, and setting $r_1 = \bar{r}(1)$ we have

$$\rho(r) = \begin{cases} 1 & r \le r_1 \\ \bar{r}^{-1}(r) & r > r_1, \end{cases} \tag{6.24}$$

where $\bar{r}^{-1}$ indicates the inverse function of $\bar{r}$.

The proof of Theorem 6.1 can be found in Appendix D.1. [1]

Remark 6.1. $p^*(x)$ defined in (6.18) and appearing in (6.17) and (6.22) can be interpreted as a density of eigenvalues and exhibits a square root singularity at the limits of its support, just as the Marčenko – Pastur density [52]. From physical point of view, $p^*(x)$ corresponds to the equilibrium charge density in the Coulomb gas picture, when the energy function is given by $E(r)$. From a practical point of view, it corresponds to the empirical distribution of observed eigenvalues $\{\lambda_i\}$ of the realized channel matrices, which balance the occurrence probability of such channel matrices with the corresponding coding error probability, when operating at a given normalized rate r, α and β.

Remark 6.2. Setting $s = 0$ in (6.12) corresponds to an unconstrained Gaussian input distribution. Hence, the corresponding solution of (6.17) will be the Gallager exponent for unconstrained Gaussian inputs, which is expected to be smaller.

Remark 6.3. From the equations of the above theorem we immediately see that the Q-dependence of $E(r)$ has the following form: $E(r,\alpha,Q) = QE(r,\frac{\alpha}{Q},1)$, where we explicitly included the dependence of $E(r)$ on α and Q. This allows us to make all calculations for $Q = 1$ and in the end to re-scale $E(r)$ and α accordingly.

Corollary 1. *For $\beta > 1$ the above expression for the error exponent can be calculated in closed form to read*

$$E(r) = Q\left[\frac{\Delta^2}{32} - \frac{\alpha\,\rho\,r}{Q} + \frac{a}{2} - \log\Delta - \frac{\beta-1}{2}\log(a\Delta) + \frac{\alpha(1+\rho)}{Q}(s + \log(1-s))\right.$$
$$+ \frac{\alpha\,\rho}{2Q}\left(\log(1+a/z_{\rho s}) + z_{\rho s}\frac{\left(\sqrt{z_{\rho s}+b} - \sqrt{z_{\rho s}+a}\right)^2}{4\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}}\right)$$
$$\left. + \frac{\Delta\alpha\,\rho}{2Q\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}}\left[G\left(0, \frac{z_{\rho s}+a}{\Delta}\right) + \frac{\beta-1}{2}G\left(\frac{a}{\Delta}, \frac{z_{\rho s}+a}{\Delta}\right)\right]\right.$$

[1] a and b are the endpoints of the support of $p^*(x)$ and should not be confused with $\alpha = T/N$ and $\beta = K/N$.

$$
\begin{aligned}
&-\frac{\Delta}{2}\left(1+\frac{\alpha\,\rho}{Q\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}}\right)\left[G\left(0,\frac{a}{\Delta}\right)+\frac{\beta-1}{2}G\left(\frac{a}{\Delta},\frac{a}{\Delta}\right)\right]\\
&-\frac{1}{2}\left(3\beta-\beta^2\log\beta+(\beta-1)^2\log(\beta-1)\right)\\
&+\frac{\alpha\,\rho}{2Q}\left[\log\left(\frac{\Delta}{z_{\rho s}}\right)-\frac{\Delta\alpha\,\rho}{2Q\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}}G\left(\frac{z_{\rho s}+a}{\Delta},\frac{z_{\rho s}+a}{\Delta}\right)\right.\\
&\left.+\left(\frac{\Delta}{2}+\frac{\alpha\,\rho\,\Delta}{2Q\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}}\right)G\left(\frac{z_{\rho s}+a}{\Delta},\frac{a}{\Delta}\right)\right]
\end{aligned}
\tag{6.25}
$$

where $\Delta = b - a$.

Corollary 2. *In the special case* $\beta = 1$ *the lower limit of the support of* $p(x)$ *becomes zero, i.e.* $a = 0$. *In this case* (6.19) *(which results from the continuity condition* $p(a) = 0$*) does not hold. However, we can obtain* $E(r)$ *by setting* $a = 0$, $\beta = 1$ *in equations* (6.20), (6.22), (6.25). *Then* $E(r)$ *reads*

$$
\begin{aligned}
E(r) =&Q\left[\frac{\alpha\,\rho}{Q}\left(\frac{b}{8}+\log\frac{1+\sqrt{1+\frac{b}{z_{\rho s}}}}{2}\right)-\log\frac{b}{4}+\frac{\alpha}{32Q}(b-4)(4z_{\rho s}+3b+12)\right.\\
&+\frac{\alpha\,\rho}{2Q}\left[\frac{b}{2}\left[G(\frac{z_{\rho s}}{b},0)+\frac{1}{2}\log\left(\frac{b}{z_{\rho s}}\right)\right]\right.\\
&+\frac{\alpha\,\rho\,b}{2Q\sqrt{z_{\rho s}(z_{\rho s}+b)}}\left[G(\frac{z_{\rho s}}{b},0)-G(\frac{z_{\rho s}}{b},\frac{z_{\rho s}}{b})\right.\\
&\left.\left.\left.-\left(\frac{z_{\rho s}}{b}-\frac{\sqrt{z_{\rho s}(z_{\rho s}+b)}}{b}\right)\log\left(\frac{b}{z_{\rho s}}\right)\right]\right]-\frac{\alpha\,\rho}{Q}r+\frac{\alpha(1+\rho)}{Q}\left(s+\log(1-s)\right)\left.\right].
\end{aligned}
\tag{6.26}
$$

Analysis

Dependence of $E(r)$ on $\alpha = T/N$

In Fig. 6.1, we plot the Gallager error exponent for various values of α. We see that increasing α brings the error curve closer to the error exponent $E_{out}(r)$ of the infinite codelength outage probability introduced in [45]. This convergence can be seen directly in (6.18)-(6.22). As $\alpha \to \infty$, $\rho \to 0$, so that $\alpha\,\rho = O(1)$ and the solution converges to that of [45].

It is important to point out here that the assumption that the receiver knows the channel matrix necessitates the existence of some training overhead, which becomes significant when the number of channel uses T becomes comparable to the number of transmit antennas N. We do not take into account this issue here, assuming instead that the training takes place through some parallel channel. However, an effective way

to incorporate training is to replace α by $\alpha - 1$, since it takes roughly N channel uses to train the N transmit antenna channels.

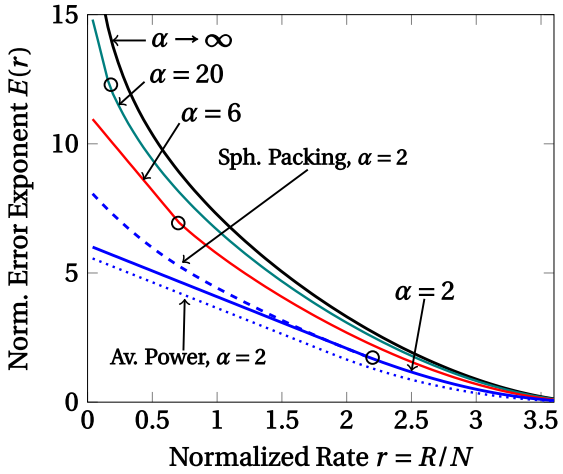


Figure 6.1: The Gallager error exponent $E(r)$. As α is increased, the curves for $E(r)$ approach the outage probability exponent $E_{out}(r)$ [45] (black). The small circles indicate the points where $r = r_1$. For $\alpha = 2$ we also depict Gallager exponent for the average power constraint ($s = 0$) and the Sphere Packing Bound error exponent (dotted). Parameter values used are: $\beta = 3$, SNR $= \sigma^{-2} = 20$, $Q = 1$.

$r \leq r_1$ and Comparison with Sphere Packing Bound

The circles in Fig. 6.1 correspond to the values $r = r_1 = \bar{r}(\rho = 1)$, below which the Gallager error exponent becomes linear in r. This behavior is due to the fact that the value of the error exponent in (6.17) is the result of the maximization with respect to the parameter ρ over the unit interval $\rho \in [0, 1]$. For $r < r_1$ the maximum lies outaside this interval and hence ρ remains fixed to unity. Hence the error exponent in (6.17) becomes linear in r. Extending the ρ-maximization interval to $\mathbb{R}^+$ provides the so-called sphere-packing error exponent [26]. In Fig. 6.1 we include the sphere-packing exponent in the case of $\alpha = 2$ (dash-dot) for comparison. As expected, for rates above the value of $r = r_1$ indicated by a circle, the error exponent coincides with the Gallager random coding exponent, while for $r < r_1$ (corresponding to solutions with $\rho > 1$) the sphere-packing exponent is higher.

Region $r \approx r_{erg}$ and Comparison with [37]

The region close to $r = r_{erg}$ is interesting because the error exponent $E(r)$ vanishes and hence the error probability is maximal. It is easy to see that $\frac{dE(r)}{dr} = -\alpha\,\rho(r)$, where $\rho(r)$ is the solution of the equation $r = \bar{r}(\rho)$ in (6.22) for $r > r_1$. From (6.22), we see that when $\rho \to 0$, then $r \to r_{erg}$. This implies that $E(r_{erg}) = 0$ is a global minimum, since, taking advantage of the convexity of the supremum operation with respect to ρ

and s, it can be shown that $E(r)$ is a convex function of r [43]. Therefore, close to $r = r_{erg}$, we can write

$$\rho(r) = (r - r_{erg})\,\rho'(r_{erg}) \tag{6.27}$$

where $\rho'(r) = \frac{d\rho}{dr}$. Let us define v_α through

$$\rho'(r_{erg}) = -\frac{1}{\alpha v_\alpha}. \tag{6.28}$$

The left-hand-side of the above equation is easy to evaluate since $\frac{d\rho(r)}{dr}\frac{d\bar{r}(\rho)}{d\rho} = 1$. Hence, by differentiating $\bar{r}(\rho)$ and expressing its value at $\rho = 0$, we obtain

$$E(r) = \frac{(r - r_{erg})^2}{2v_\alpha} + o\left((r - r_{erg})^2\right). \tag{6.29}$$

In the above, v_α can be expressed as

$$v_\alpha = v_\infty + \frac{\delta v}{\alpha}, \tag{6.30}$$

where v_∞ is the infinite codelength dispersion given in (6.9) and $\delta v > 0$ has the simple form

$$\delta v = 2g_0 - g_0^2, \tag{6.31}$$

where g_0 is given by

$$g_0 = \int_{a_0}^{b_0} \frac{x p_0(x)}{x + \sigma^2} dx = \frac{\left(\sqrt{\sigma^2 + b_0} - \sqrt{\sigma^2 + a_0}\right)^2}{4}, \tag{6.32}$$

where $p_0(x)$ is the Marcenko-Pastur distribution given in (6.6) and a_0, b_0 its endpoints. It is worth pointing out that the last term in (6.31) is the correction due to the peak-power codeword constraint (6.11). We see that the Gallager error exponent $E(r)$, which is valid for all rates $r < r_{erg}$ takes a quadratic form akin to the exponent of a normal distribution for rates close to r_{erg}. This is analogous to the case of infinite codelengths discussed in Section 6.1. (6.88) is valid when $|r_{erg} - r| \ll 1$, in order for the error exponent to be small. However, it is also implicitly assumed that $N|r_{erg} - r| \gg 1$, so that the term $N^2 E(r)$ in the error probability exponent (see (6.16)) is the dominant one. Hence this is exactly the moderate deviations regime discussed for general single link systems in [2]. An important point that can be drawn from the form of (6.32) is that it depends only on the empirical distribution of eigenvalues, which in this case happens to be the Marcenko-Pastur distribution. Therefore, v_α can be calculated for other channel models for which v_∞ and $p_0(x)$ are known.

In [37], the authors obtained bounds on the optimum average probability of error for MIMO systems when the normalized rate of the code $r = R/N$ approaches the ergodic rate r_{erg} such that $N|r - r_{erg}| = O(1)$ in the limit that N, K, T become large

with fixed ratios. In this limit, they show that the error probability is bounded between two Gaussian distributions with variances (or dispersions) given (in their notation) by θ_-, which can be expressed as

$$\theta_- = \alpha v_\infty + \frac{1}{2}\left(\beta + 1 - \frac{\sigma^2(\beta+1) + (\beta-1)^2}{\sqrt{(\sigma^2+a_0)(\sigma^2+b_0)}}\right), \tag{6.33}$$

and $\theta_+ = \alpha v_\alpha$, respectively. Therefore, the Gallager random coding exponent with Gaussian input saturates the upper bound in the dispersion derived by [37].

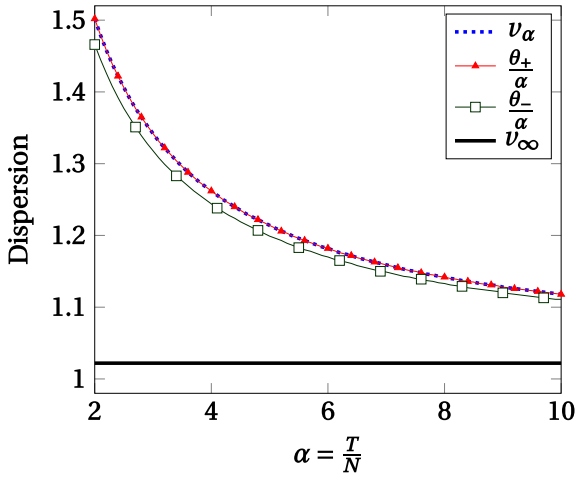


Figure 6.2: The dispersion at the Gaussian limit ($\rho = 0$) using the asymptotic method and the method of induced ergodicity of [37]. The v_α and $\frac{\theta_+}{\alpha}$ curves are identical; SNR $= \sigma^{-2} = 20$, $\beta = 3$, $Q = 1$.

Impact of Fading

The case $Q > 1$ models the realistic situation where the channel varies during the transmission of the codeword. Specifically, the channel matrix $\mathbf{H}$ changes (Q times) during the codelength T. It is assumed here that the receiver knows each channel realization, either using an additional pilot signal or by using part of the codeword as pilot (in which case T will represent the data-transmitting part of the codeword). In Fig. 6.3 we can see the behavior of the error exponent for increasing values of Q. As Q, the number of independent fading blocks within a codeword increases, the error exponent $E(r)$ also increases, signifying lower error probabilities. To understand the behavior for large Q, we prove the following result.

Theorem 6.2 ($Q \to \infty$ Limit of $E(r)$)**.**

$$\lim_{Q\to\infty} E(r) = \alpha \max_{\substack{\rho\in[0,1]\\ s\in[0,1)}} \left[\rho\, r_{erg}(\beta, z_{\rho s}^{-1}) - \rho\, r + (1+\rho)(s + \log(1-s))\right]. \tag{6.34}$$

For fixed ρ, the maximum over s in the above equation is attained at the value

$$s = \frac{\rho}{4(1+\rho)} \left(\sqrt{z_{\rho s} + b_0} - \sqrt{z_{\rho s} + a_0} \right)^2 . \tag{6.35}$$

Defining the function

$$\bar{r}(\rho) = \log(1-s) + r_{erg}\left(\beta, z_{\rho s}^{-1}\right), \tag{6.36}$$

and setting $r_1 = \bar{r}(1)$ we have

$$\rho(r) = \begin{cases} 1 & r \le r_1 \\ \bar{r}^{-1}(r) & r > r_1 \end{cases}, \tag{6.37}$$

where $\bar{r}^{-1}$ indicates the inverse function of $\bar{r}$.

From the above theorem we conclude that for fast-fading, and therefore large values of Q it is the ergodic rate that determines the behavior of the error exponent. When $r \approx r_{erg}$ we can once again expand $E(r)$ in powers of $r - r_{erg}$ to obtain

$$E(r) = \frac{(r - r_{erg})^2}{2\ \delta v} + o\left((r - r_{erg})^2\right), \tag{6.38}$$

where δv is given in (6.31).

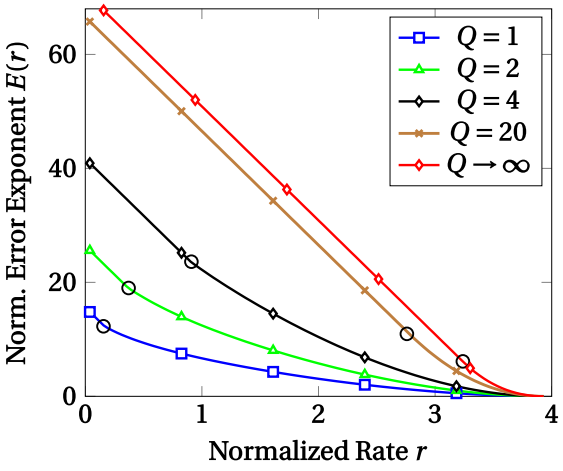


Figure 6.3: The error exponent for the Gallager bound with power constraint at the limit of $N \to \infty$; $\beta = 3$, SNR $= \sigma^{-2} = 20$, $\alpha = 20$.The small circles indicate the points of behavior change ($r = r_1$).

6.2 The Gallager Bound in Fiber Optical MIMO

In the previous section we studied extensively the application of the Gallager error bound to the wireless MIMO case. Hence, it is logical that we use the same bound for the evaluation and the quantification of the performance of MIMO fiber optical systems. Again, since the expression of the Gallager bound is cumbersome to be analyzed we will take the asymptotic limit for large length T of codewords and numbers N and K of transmitting and receiving modes respectively.

In this chapter we evaluate the error probability exponent of the Gallager bound for large T and large numbers N and K and fixed ratios, using random matrix theory. That way, the calculation becomes simpler but the outcome is still valid: we invoke the large deviations theory, just as we did in the previous chapter, and examine the tails of the Gallager bound which correspond to regions with low outage probability which is the region where the fiber optical networks operate.

Problem Formulation

System Model

In this chapter we consider a single-segment N_{tot}-channel lossless optical fiber system, with $N \leq N_{tot}$ transmitting channels excited and $K \leq N_{tot}$ receiving channels coherently excited in the input (left) and output (right) side of the fiber. The propagation through the fiber may be analyzed through its $2N_{tot} \times 2N_{tot}$ scattering matrix given by [15, 86] So, the corresponding MIMO channel for this system reads

$$\mathbf{y} = \mathbf{U}\mathbf{x} + \mathbf{z} \tag{6.39}$$

with coherent detection and channel state information only at the receiver [24, 78]. $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are the $N \times 1$ input, the $K \times 1$ output signal vectors and the $K \times 1$ unit variance noise vector, respectively, all assumed for simplicity to be complex Gaussian. We also assume no differential delays between channels, which effectively leads to frequency flat fading [86]. We also assume no mode-dependent loss. As a result, the mutual information can be expressed as

$$C = \log\det\left(\mathbf{I}_K + \rho\ \mathbf{U}\mathbf{U}^\dagger\right), \tag{6.40}$$

where ρ is the SNR. The total transmission rate is $R_{erg} = Nr_{erg}$ where r_{erg} is the ergodic rate per transmitter. The value of the mutual information per transmitter $C(\rho, \mathbf{U})/N$ converges weakly to a deterministic value in the large N limit: the ergodic average of the mutual information [65]. In addition it is important to note that the empirical eigenvalue density of $\mathbf{U}\mathbf{U}^\dagger$ converges weakly, almost surely so, to the well-known Marčenko-Pastur distribution [81]

$$p_0(x) = \begin{cases} \frac{\sqrt{(b_0-x)(x-a_0)}}{2\pi x}, & \text{for } x \in [a_0, b_0] \\ 0, & \text{otherwise.} \end{cases} \tag{6.41}$$

where $a_0, b_0 = (\sqrt{\beta} \pm 1)^2$ are the endpoints of its support.

Gallager Exponent

In the infinite codeword limit, the effect of the channel fading is captured through the optimal outage error probability [62], which in the large N, K limit has been analyzed in [45]. On the other hand, for finite codewords, one can use the Gallager bound: For *Maximum Likelihood* (ML) decoding for a discrete memoryless, fixed channel without feedback the error probability $\mathbb{P}(\mathcal{E})$, is bounded by

$$\begin{aligned}\mathbb{P}(\mathcal{E}|\mathbf{U}) &\leq e^{-N^2 E(R|\mathbf{U})},\\ E(R|\mathbf{U}) &= \frac{1}{N}\max_{\substack{\rho\in[0,1]\\ s\in[0,1)}}\{E_0[\rho\,|\mathbf{U}] - \rho\, r + (1+\rho)(s+\log(1-s))\}\end{aligned} \tag{6.42}$$

where $E_0(\rho\,|\mathbf{U})$ is Gallager's error exponent defined as

$$\begin{aligned}E_0(\rho\,|\mathbf{U}) &= \log\int d\mathbf{Y}\left[\int d\mathbf{X}\;\mu(\mathbf{X})\,[\mu(\mathbf{Y}|\mathbf{X},\mathbf{U})]^{\frac{1}{1+\rho}}\right]^{1+\rho}\\ &= \log\det\left(1+\frac{\rho}{(1+\rho)(1-s)}\mathbf{U}\mathbf{U}^{\dagger}\right),\end{aligned} \tag{6.43}$$

where the last line follows [6], for independent Gaussian input and for any $s > 0$. So, plugging (6.43) in (6.42) we have

$$\begin{aligned}E(R|\mathbf{U}) = \frac{1}{N}\max_{\substack{\rho\in[0,1]\\ s\in[0,1)}}\Bigg\{&\alpha\,\rho\left[\log\det\left(1+\frac{\rho}{(1+\rho)(1-s)}\mathbf{U}\mathbf{U}^{\dagger}\right) - Nr\right]\\ &+(1+\rho)N(s+\log(1-s))\Bigg\},\end{aligned} \tag{6.44}$$

where $\alpha = \frac{T}{N}$, r is the rate. $\mathbb{P}(\mathcal{E})$ is error rate when we decode the message, therefore

$$\mathbb{P}(\mathcal{E}) = \mathbb{E}_{\mathbf{U}}\left[\mathbb{P}\left(\mathcal{E}|\mathbf{U}\right)\right] \leq \mathbb{E}_{\mathbf{U}}\left(e^{-N^2 E(r|\mathbf{U})}\right) \equiv e^{-N^2 E_N(r)}. \tag{6.45}$$

In this chapter, we calculate the closed form expression of the error exponent $E_N(r)$ as $N, K, T \to \infty$ while $\beta = \frac{K}{N} > 1$ is kept constant. We further define $N_0 = N_{tot} - N - K$ and $n_0 = \frac{N_0}{N}$ the losses inside the fiber. In the case of $N_0 < 0$, [15] showed that we may recover the form of β, N_0 and n_0 by substituting $N \to N_{tot} - K$, $K \to N_{tot} - N$ and $N_{tot} \to -N_{tot}$. Then, the mutual information becomes also $C \to C + n_0 \log(1+\rho)$.

The joint distribution of eigenvalues of $\mathbf{U}\mathbf{U}^{\dagger}$ is

$$P_{\boldsymbol{\lambda}}(\lambda_1 \ldots \lambda_N) = \frac{1}{\mathcal{Z}_N}\prod_{N<i<j\leq K}|\lambda_i - \lambda_j|^2 \prod_i \lambda_i^{K-N}(1-\lambda_i)^{N_0}, \tag{6.46}$$

where $\mathcal{Z}_N$ is the normalizing constant. We can assume that when N_{tot} is large, then the eigenvalues will coalesce to a smooth density $p(x)$, which will be such that the energy $\lim_{N\to\infty} E_N(r) = E(r)$ will be minimum and (6.46) corresponds to the most

probable eigenvalue distribution. In this limit, we can write the minimum energy of the eigenvalues as

$$E_0(r) = \left\{ \sup_c \inf_p \mathcal{L}_0[c,p] \right\} \tag{6.47}$$

$$\begin{aligned}\mathcal{L}_0 = \max_{\rho\in[0,1]} \Bigg\{ &- n_0 \int p(x)\log(1-x)dx - (\beta-1)\int p(x)\log(x)dx \\ &- \int\int p(x)p(y)\log|x-y|dydx \Bigg\} - c\left(\int p(x)dx - 1\right),\end{aligned} \tag{6.48}$$

where we have added a Lagrange multiplier c to ensure that $p(x)$ is properly normalized, while implicitly assuming that $p(x)$ is continuous in $x \in (0,1)$. In order to incorporate the constraint on the rate r, we introduce another Lagrange multiplier so that

$$E(r) = \sup_c \inf_p \mathcal{L}_1[c,\rho,p,s] \tag{6.49}$$

$$\begin{aligned}\mathcal{L}_1[p,c,\rho,s] =& \mathcal{L}_0[p,c] + \alpha\,\rho\left(\int \log\left(1+\frac{\rho x}{1+\rho}\right)p(x)dx - r\right) \\ &+ \alpha(\rho+1)(s+\log(1-s)).\end{aligned} \tag{6.50}$$

It is implied in (6.49) that first we maximize with respect to ρ and s and then search for the infimum in p. But this is not an easy task because the maximization will depend on the specific distribution of $p(x)$. Hence, since it can be shown that $\mathcal{L}_1$ is continuous and convex, it is very useful to point out the Minimax theorem ([20]-Theorem 2) which allows in our case, to exchange places between the max – inf. Therefore, we can re–write

$$E(r) = \max_{\substack{\rho\in[0,1]\\ s\in[0,1)}} \left\{ \sup_c \inf_p \mathcal{L}_1[c,p] \right\} \tag{6.51}$$

and continue with our calculations. It is easy to show that the minimum of the above function is unique since we can reach $E_0(r)$ from $E(r)$ by maximizing over $\mathcal{L}_1$, while keeping $\rho, s = 0$. The minimization of $\mathcal{L}_1$ with respect to p is done by taking the functional derivative and setting it to zero which calculations are simlpe but tedious and will be omitted. It can be shown also that the minimum of $\mathcal{L}_1$ is unique [45]. So, taking the functional derivative with respect to $p(x)$ and setting to zero and differentiating with respect to x, gives us the following integral equation

$$2PV\int \frac{p^*(x)}{x-y}dx = \frac{n_o}{1-x} - \frac{\beta-1}{x} + \frac{\alpha\,\rho}{z_{\rho\,s}+x}, \tag{6.52}$$

where $z_{\rho\,s} = \frac{(1+\rho)(1-s)}{\rho}$ and gives

$$p^*(x) = \frac{1}{2\pi\sqrt{(x-a)(x-b)}}$$

$$\times\left[\frac{n_0\sqrt{(1-a)(1-b)}}{1-x}-(\beta-1)\frac{\sqrt{ab}}{x}+\frac{\alpha\,\rho\,\sqrt{(z+a)(z+b)}}{x+z}+C'\right]. \tag{6.53}$$

We search for solution among continuous, non–negative, normalized functions over $x\in(0,\infty)$. Continuity at $x=b$ results to the constraint that the expression in the square bracket vanishes at $x=b$:

$$p^*(x)=\frac{\sqrt{b-x}}{2\pi\sqrt{x-a}}\left[n_0\sqrt{\frac{1-a}{1-b}}-\frac{(\beta-1)\sqrt{a}}{x\sqrt{b}}+\frac{\alpha\,\rho\,\sqrt{z_{\rho s}+a}}{(x+z_{\rho s})\sqrt{z_{\rho s}+b}}\right]. \tag{6.54}$$

The a is obtained through the continuity condition $p(a)=0$:

$$\frac{n_0}{\sqrt{(1-a)(1-b)}}=\frac{\beta-1}{\sqrt{ab}}-\frac{\alpha\,\rho}{\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}}, \tag{6.55}$$

and back to $p(x)$ we have

$$p^*(x)=\frac{\sqrt{(x-a)(b-x)}}{2\pi(1-x)}\left[\frac{\beta-1}{x\sqrt{ab}}-\frac{\alpha\,\rho}{\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}}\frac{1+z_{\rho s}}{x+z_{\rho s}}\right]. \tag{6.56}$$

The b will be evaluated through the normalization condition of $p^*(x)$,

$$\begin{aligned}&\frac{\beta-1}{2\sqrt{ab}}\left(1-\sqrt{ab}-\sqrt{(1-a)(1-b)}\right)\\&-\frac{\alpha\,\rho}{2\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}}\left[z_{\rho s}+1-\sqrt{(1-a)(1-b)}-\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}\right]=1.\end{aligned} \tag{6.57}$$

The value of s will be calculated from

$$s=\frac{\rho}{1+\rho}\int_a^b p^*(x)\frac{x}{x+z_{\rho s}}, \tag{6.58}$$

which gives

$$\begin{aligned}s=&\frac{\rho}{2(1+\rho)(z_{\rho s}+1)}\left[\frac{\beta-1}{\sqrt{ab}}\left(z_{\rho s}+1-\sqrt{(1-a)(1-b)}-\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}\right)\right.\\&-\frac{\alpha\,\rho}{\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}}\left(z_{\rho s}+1-\sqrt{(1-a)(1-b)}-\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}\right.\\&\left.\left.\frac{z_{\rho s}(z_{\rho s}+1)}{\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}}\left(\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}-z_{\rho s}-\frac{a+b}{2}\right)\right)\right].\end{aligned} \tag{6.59}$$

The value of ρ will be determined by the saddle point equation,

$$r=\int_a^b p^*(x)\left[\log\left(1+\frac{x}{z_{\rho s}}\right)-\frac{\rho}{1+\rho}\frac{x}{x+z_{\rho s}}\right]dx. \tag{6.60}$$

Therefore, we integrate (6.60) to obtain

$$
\begin{aligned}
r = &\log\frac{\Delta\rho}{1+\rho} - \frac{\Delta\alpha\,\rho}{2\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}}G\left(\frac{z_{\rho s}+a}{\Delta},\frac{z_{\rho s}+b}{\Delta}\right)\\
&+\frac{\Delta}{2}\left(1+\frac{\alpha\,\rho}{\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}}\right)G\left(\frac{z_{\rho s}+a}{\Delta},\frac{a}{\Delta}\right)\\
&+\frac{\rho(\beta-1)}{(1+\rho)(z_{\rho s}+1)\sqrt{ab}}\left(1-\sqrt{(1-a)(1-b)}-\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}\right)\\
&-\frac{\alpha\,\rho^2(z_{\rho s}+1)}{(1+\rho)\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}}\left[\frac{z_{\rho s}+1}{2z_{\rho s}^2+1}\left[\frac{b+a}{2}+2z_{\rho s}-\sqrt{\frac{z_{\rho s}+a}{z_{\rho s}+b}}\left(\frac{b-a}{2}+2z_{\rho s}+b\right)\right.\right.\\
&\left.+\frac{\Delta a}{2\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}}\right]\\
&+\frac{1}{2z_{\rho s}^2+1}\left[z_{\rho s}\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}-z_{\rho s}^2-\frac{\Delta}{2}(a(a+b)-3a+1)\right.\\
&\left.\left.-(1-a)^2-(1-a+a\Delta)\sqrt{(1-a)(1-b)}\right]\right],
\end{aligned}
\tag{6.61}
$$

where $\Delta = b-a$ and $G(x,y)$ can be seen in Appendix E

Results

Finally, we integrate over $p(x)$ for $\beta > 1$ and $n_0 > 0$:

$$
\begin{aligned}
E(r) = &-\rho\,\alpha r + \frac{1}{2}\rho\,\alpha\log\left(1+\frac{b}{z_{\rho s}(1+\rho)}\right) - \frac{1}{2}(\beta+1+n_0)\log(b-a) - \frac{\beta-1}{2}\log(b)\\
&-\frac{n_0}{2}\log(1-b) - \frac{n_0^2(b-a)}{4\sqrt{(1-a)(1-b}}\left[G\left(\frac{1-b}{b-a},\frac{1-b}{b-a}\right)-G\left(\frac{1-b}{b-a},\frac{z_{\rho s}+b}{b-a}\right)\right]\\
&+\frac{n_0(\beta-1)(b-a)}{4\sqrt{ab}}\left[G\left(\frac{1-b}{b-a},-\frac{b}{b-a}\right)-G\left(\frac{1-b}{b-a},-\frac{z_{\rho s}+b}{b-a}\right)\right]\\
&+\frac{n_0(\beta-1)(b-a)}{4\sqrt{(1-a)(1-b)}}\left[G\left(\frac{a}{b-a},-\frac{1-a}{b-a}\right)-G\left(\frac{a}{b-a},\frac{z_{\rho s}+a}{b-a}\right)\right]\\
&-\frac{(\beta-1)^2(b-a)}{4\sqrt{ab}}\left[G\left(\frac{a}{b-a},\frac{a}{b-a}\right)-G\left(\frac{a}{b-a},\frac{z_{\rho s}+a}{b-a}\right)\right]\\
&-\frac{n_0(b-a)}{2\sqrt{(1-a)(1-b)}}\left[G\left(0,\frac{1-b}{b-a}\right)-G\left(0,-\frac{z_{\rho s}+b}{b-a}\right)\right]\\
&+\frac{(\beta-1)(b-a)}{2\sqrt{ab}}\left[G\left(0,-\frac{b}{b-a}\right)-G\left(0,-\frac{z_{\rho s}+b}{b-a}\right)\right]\\
&+\frac{\rho\,\alpha}{2}\log\left(\frac{b-a}{z_{\rho s}(1+\rho)}\right)+\alpha(1+\rho)(s+\log(1-s))
\end{aligned}
$$

$$+\frac{n_0(b-a)}{2\sqrt{(1-a)(1-b)}}\left[G\left(\frac{z_{\rho s}+a}{b-a},\frac{z_{\rho s}+a}{b-a}\right)-G\left(\frac{z_{\rho s}+a}{b-a},-\frac{1-a}{b-a}\right)\right]$$
$$+\frac{(\beta-1)(b-a)}{2\sqrt{ab}}\left[G\left(\frac{z_{\rho s}+a}{b-a},\frac{a}{b-a}\right)-G\left(\frac{z_{\rho s}+a}{b-a},\frac{z_{\rho s}+a}{b-a}\right)\right]$$
$$-\frac{1}{2}\Big[(\beta+n_0+1)^2\log(\beta+n_0+1)-(\beta+n_0)^2\log(\beta+n_0)-\beta^2\log(\beta)$$
$$+(\beta-1)^2\log(\beta-1)-(n_0+1)^2\log(n_0+1)+n_0^2\log(n_0)\Big]. \tag{6.62}$$

Special Cases

We distinguish two different cases: while it is $\beta=\frac{K}{N}>1$, the losses inside the fiber can be either $n_0>0$ or $n_0=0$.

Therefore, for $n_0=0$ the density of eigenvalues becomes

$$p^*(x)=\frac{1}{2\pi\sqrt{(x-a)(x-b)}}\left[-\frac{(\beta-1)\sqrt{ab}}{x}+\frac{\alpha\,\rho\,\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}}{x+z_{\rho s}}+C\right]. \tag{6.63}$$

The case $0\le\rho\le1$ **and** $\beta>1\to\alpha>0$

From the constraint $p^*(a)=0$ we have

$$p^*(x)=\frac{\sqrt{x-a}}{2\pi\sqrt{b-x}}\left[\frac{\beta-1}{x}\sqrt{\frac{b}{a}}-\frac{\alpha\,\rho\,\sqrt{z_{\rho s}+b}}{\sqrt{z_{\rho s}+a}}\frac{1}{x+z_{\rho s}}\right]. \tag{6.64}$$

There are two options $b=1$ and $b<1$.

The case $b=1$

In the first case of $b=1$, which is stable close to $\rho=0$, we have the additional constraint that the first term cannot appear since it will lead to non-integrable singularity. Hence,

$$p_1(x)=\frac{\sqrt{x-a}}{2\pi\sqrt{1-x}}\left(\frac{\beta-1}{x\sqrt{a}}-\frac{\alpha\,\rho\,\sqrt{z_{\rho s}+1}}{(z+x_{\rho s})\sqrt{z_{\rho s}+a}}\right), \tag{6.65}$$

and the normalization conditions gives

$$\beta+1-\rho\,\alpha=\frac{\beta-1}{\sqrt{a}}-\rho\,\alpha\frac{\sqrt{z_{\rho s}+1}}{z_{\rho s}+a}. \tag{6.66}$$

To obtain ρ we need to solve the fixed point equation for $r(\rho)$:

$$r_1(\rho)=\int_a^1 p_1(x)\log\left(1+\frac{x}{z_{\rho s}}\right)dx-\frac{\rho}{1+\rho}\int_a^1 p_1(x)\frac{x}{x+z_{\rho s}}dx. \tag{6.67}$$

Sparing with the calculations, we have

$$\begin{aligned}
r_1 =& \log\left(\frac{1-a}{z_{\rho s}}\right) - \frac{\alpha\,\rho(1-a)}{2\sqrt{(z_{\rho s}+1)(z_{\rho s}+a)}}\left[I_3\left(\frac{z_{\rho s}+a}{1-a}\right) + G\left(\frac{z_{\rho s}+a}{1-a}, \frac{z_{\rho s}+a}{1-a}\right)\right] \\
&+\frac{(\beta-1)(1-a)}{2\sqrt{a}}\left[I_3\left(\frac{z_{\rho s}+a}{1-a}\right) + G\left(\frac{z_{\rho s}+a}{1-a}, \frac{a}{1-a}\right)\right] \\
&-\frac{\rho}{1+\rho}\Bigg[1 - \frac{z_{\rho s}(1-a)}{4(1+z_{\rho s})^2}\left(\frac{(\beta-1)z_{\rho s}}{\sqrt{a}} + \beta + 1 - \alpha\,\rho\right) - \frac{(\beta-1)}{4\sqrt{a}}(1-\sqrt{a})^2 \\
&+\frac{(\sqrt{z_{\rho s}+1}-\sqrt{z_{\rho s}+a})^2}{4(z_{\rho s}+1)^2}\left(\frac{(\beta-1)(z_{\rho s}+a)}{\sqrt{a}} + \frac{\alpha\,\rho\,\sqrt{z_{\rho s}+1}}{\sqrt{z_{\rho s}+a}} + \frac{\alpha\,\rho\,z_{\rho s}+1}{z_{\rho s}+a}\right)\Bigg],
\end{aligned} \tag{6.68}$$

where

$$I_3(x) = -G(x,-1). \tag{6.69}$$

From here, following the method we saw in previous section, the calculation of E is straight forward:

$$\begin{aligned}
E(r_1) =& \frac{\rho\,\alpha}{2}\log\left(1+\frac{1}{z_{\rho s}(1+\rho)}\right) - \rho\,\alpha r_1 - \frac{\beta-1}{2}\log(a) - \frac{\beta+1}{2}\log(1-a) \\
&-\frac{(\beta-1)^2(1-a)}{4\sqrt{a}}\left[G\left(\frac{a}{1-a}, \frac{a}{1-a}\right) - G\left(\frac{a}{1-a}, -1\right)\right] \\
&+\frac{\alpha\,\rho(\beta-1)(1-a)}{4\sqrt{(z_{\rho s}(\rho+1)+1)(z_{\rho s}(\rho+1)+a)}} \\
&\times\left[G\left(\frac{a}{1-a}, \frac{z_{\rho s}(\rho+1)+a}{1-a}\right) - G\left(\frac{a}{1-a}, -1\right)\right] \\
&-\frac{(\beta-1)(1-a)}{2\sqrt{a}}\left[G\left(0, \frac{a}{1-a}\right) - G\left(0,-1\right)\right] + \alpha(1+\rho)(s+\log\left(1-s\right)) \\
&+\frac{\rho\,\alpha(1-a)}{2\sqrt{(z_{\rho s}(\rho+1)+1)(z_{\rho s}(\rho+1)+a)}}\left[G\left(0, \frac{z_{\rho s}(\rho+1)+a}{1-a}\right) - G\left(0,-1\right)\right] \\
&+\frac{\rho\,\alpha}{2}\log\left(\frac{1-a}{z_{\rho s}(\rho+1)}\right) - \frac{\rho\,\alpha(1-a)}{2\sqrt{(z_{\rho s}(\rho+1)+1)(z_{\rho s}(\rho+1)+a)}} \\
&\times\left[G\left(\frac{z_{\rho s}(\rho+1)+a}{1-a}, \frac{z_{\rho s}(\rho+1)+a}{1-a}\right) - G\left(\frac{z_{\rho s}(\rho+1)+a}{1-a}, -1\right)\right] \\
&+\frac{(\beta-1)(1-a)}{\sqrt{a}}\left[G\left(\frac{z_{\rho s}(\rho+1)+a}{1-a}, \frac{a}{1-a}\right) - G\left(\frac{z_{\rho s}(\rho+1)+a}{1-a}, -1\right)\right] \\
&-\frac{1}{2}\Big[(\beta+1)^2\log(\beta+1) - 2\beta^2\log(\beta) + (\beta-1)^2\log(\beta-1)\Big].
\end{aligned} \tag{6.70}$$

The above expressions are valid for when $p_1(x) > 0$ for all $a < x < 1$. When $p_1(x) < 0$, it breaks down and the upper limit becomes $b < 1$. The stability condition

for that, is

$$\beta + 1 - \rho\,\alpha = \frac{\beta - 1}{\sqrt{a}} - \frac{\alpha\,\rho}{\sqrt{(z_{\rho\,s})(z_{\rho\,s} + a)}}. \tag{6.71}$$

As α grows, this will not be valid at $\rho = 1$ and hence, there will be a breakdown. The critical value of $\rho = \rho_c$ is the solution of

$$\rho_c\,\alpha = \beta + 1 + \frac{\beta - 1}{\sqrt{a_c}} z_{\rho\,s}, \tag{6.72}$$

where $a_c = \left(\frac{(\beta-1)z_{\rho\,s}}{\rho\,\alpha-\beta-1}\right)^2$. Because $a < 1$ we also have

$$\rho > \frac{(\beta + 1)(z_{\rho\,s} + 1)}{\alpha - z_{\rho\,s}(\beta + 1)}, \tag{6.73}$$

and we finally reach

$$\begin{aligned}\rho_c =& \frac{\alpha(z_{\rho\,s} + 1)(\beta + 1) - z_{\rho\,s}(\beta^2 + 1) - z_{\rho\,s}^2(\beta + 1)^2}{(\alpha - z(\beta + 1)^2 + 4\beta z} \\ &+ \frac{2\sqrt{\beta z}\sqrt{z(\alpha + \beta)(\alpha + 1) + \alpha(\alpha + 1 + \beta)}}{(\alpha - z(\beta + 1)^2 + 4\beta z}.\end{aligned} \tag{6.74}$$

So, we calculate the $E(r)$ for the two different sub–cases

The case $\rho \leq \rho_c < 1$

$$\begin{aligned}E_{\rho\leq\rho_c}(r_1) =& \frac{\rho\,\alpha}{2}\log\left(1 + \frac{1}{z_{\rho\,s}(1+\rho)}\right) - \rho\,\alpha r_1 - \frac{\beta - 1}{2}\log(a) - \frac{\beta + 1}{2}\log(1 - a) \\ &- \frac{(\beta - 1)^2(1 - a)}{4\sqrt{a}}\left[G\left(\frac{a}{1-a}, \frac{a}{1-a}\right) - G\left(\frac{a}{1-a}, -1\right)\right] \\ &+ \frac{\alpha\,\rho(\beta - 1)(1 - a)}{4\sqrt{(z_{\rho\,s}(\rho+1) + 1)(z_{\rho\,s}(\rho+1) + a)}} \\ &\times\left[G\left(\frac{a}{1-a}, \frac{z_{\rho\,s}(\rho+1) + a}{1-a}\right) - G\left(\frac{a}{1-a}, -1\right)\right] + \alpha(1 + \rho)(s + \log(1 - s)) \\ &- \frac{(\beta - 1)(1 - a)}{2\sqrt{a}}\left[G\left(0, \frac{a}{1-a}\right) - G(0, -1)\right] \\ &+ \frac{\rho\,\alpha(1 - a)}{2\sqrt{(z_{\rho\,s}(\rho+1) + 1)(z_{\rho\,s}(\rho+1) + a)}}\left[G\left(0, \frac{z_{\rho\,s}(\rho+1) + a}{1-a}\right) - G(0, -1)\right] \\ &+ \frac{\rho\,\alpha}{2}\log\left(\frac{1 - a}{z_{\rho\,s}(\rho+1)}\right) - \frac{\rho\,\alpha(1 - a)}{2\sqrt{(z_{\rho\,s}(\rho+1) + 1)(z_{\rho\,s}(\rho+1) + a)}} \\ &\times\left[G\left(\frac{z_{\rho\,s}(\rho+1) + a}{1-a}, \frac{z_{\rho\,s}(\rho+1) + a}{1-a}\right) - G\left(\frac{z_{\rho\,s}(\rho+1) + a}{1-a}, -1\right)\right]\end{aligned}$$

$$+\frac{(\beta-1)(1-a)}{\sqrt{a}}\left[G\left(\frac{z_{\rho s}(\rho+1)+a}{1-a},\frac{a}{1-a}\right)-G\left(\frac{z_{\rho s}(\rho+1)+a}{1-a},-1\right)\right]$$
$$-\frac{1}{2}\Big[(\beta+1)^2\log(\beta+1)-2\beta^2\log(\beta)+(\beta-1)^2\log(\beta-1)\Big]. \tag{6.75}$$

The case $\rho \geq \rho_c < 1$

$$\begin{aligned}
E_{\rho\geq\rho_c}(r_1) &= \frac{\rho\,\alpha}{2}\log\left(1+\frac{a'}{b'}\right)-\frac{(\beta-1)}{2}\log(a')-\frac{\beta+1}{2}\log(b'-a') \\
&-\frac{(\beta-1)^2(b'-a')}{4}\left[G\left(\frac{a'}{b'-a'},\frac{a'}{b'-a'}\right)-G\left(\frac{a'}{b'-a'},\frac{z_{\rho s}(1+\rho)+a'}{b'-a'}\right)\right] \\
&-\frac{(\beta-1)(b'-a')}{\sqrt{a'b'}}\left[G\left(0,\frac{a'}{b'-a'}\right)-G\left(0,\frac{z_{\rho s}(1+\rho)+a'}{b'-a'}\right)\right] \\
&+\frac{\rho\,\alpha}{2}\left[\log\left(\frac{b'-a'}{z_{\rho s}(1+k)}\right)+\frac{(\beta-1)(b'-a')}{2}\log(a'b')\right. \\
&\left.\times\left[G\left(\frac{z_{\rho s}(1+\rho)+a'}{b'-a'},\frac{a'}{b'-a'}\right)-G\left(\frac{z_{\rho s}(1+\rho)+a'}{b'-a'},\frac{z_{\rho s}(1+\rho)+a'}{b'-a'}\right)\right]\right] \\
&-\rho\,\alpha r_1-\frac{1}{2}\Big[(\beta+1)^2\log(\beta+1)-2\beta^2\log(\beta)+(\beta-1)^2\log(\beta-1)\Big] \\
&+\alpha(1+\rho)(s+\log(1-s)).
\end{aligned} \tag{6.76}$$

where

$$a' = z_{\rho s}(\rho+1)\left(\frac{\sqrt{\beta(\rho\,\alpha-\beta)}-\sqrt{\rho\,\alpha-1}}{\rho\,\alpha-1-\beta}\right)^2, \tag{6.77}$$

and

$$b' = z_{\rho s}(\rho+1)\left(\frac{\sqrt{\beta(\rho\,\alpha-\beta)}+\sqrt{\rho\,\alpha-1}}{\rho\,\alpha-1-\beta}\right)^2. \tag{6.78}$$

The case $\rho > 1$

Finally, we analyze the case where $\rho > \rho_c$, so that the support of $p(x)$ does not extend to 1:

$$p_2(x) = \frac{\sqrt{(x-a)(x-b)}}{2\pi}\frac{(\beta-1)z_{\rho s}}{x\sqrt{ab}(x+z_{\rho s})}, \tag{6.79}$$

with the constraints

$$\frac{\beta-1}{\sqrt{ab}} = \frac{\rho\,\alpha}{\sqrt{(z_{\rho s}+a)(z_{\rho s}+b)}}, \tag{6.80}$$

and

$$\beta + 1 - \rho\,\alpha = \frac{\beta - 1}{\sqrt{ab}} - \frac{\rho\,\alpha(z_{\rho s} + 1)}{\sqrt{(z_{\rho s} + a)(z_{\rho s} + b)}}. \tag{6.81}$$

So, we have

$$a = z_{\rho s}\left(\frac{\sqrt{\beta(\rho\,\alpha - 1)} - \sqrt{\rho\,\alpha - 1}}{\rho\,\alpha - \beta - 1}\right)^2, \tag{6.82}$$

and

$$b = z_{\rho s}\left(\frac{\sqrt{\beta(\rho\,\alpha - \beta)} + \sqrt{\rho\,\alpha - 1}}{\rho\,\alpha - \beta - 1}\right)^2, \tag{6.83}$$

and finally,

$$\begin{aligned} r_2 =& \log\left(\frac{\Delta}{z_{\rho s}}\right) + \log(1 - s) + \frac{(\beta - 1)\Delta}{2\sqrt{ab}}\left[G\left(\frac{z_{\rho s} + a}{\Delta}, \frac{a}{\Delta}\right) - G\left(\frac{z_{\rho s} + a}{\Delta}, \frac{z_{\rho s} + a}{\Delta}\right)\right] \\ &- \frac{\rho}{1 + \rho}\frac{(\beta - 1)z_{\rho s}}{\sqrt{ab}}\frac{(\sqrt{z_{\rho s} + b} - \sqrt{z_{\rho s} + a})^2}{4\sqrt{(z_{\rho s} + a)(z_{\rho s} + b)}}. \end{aligned} \tag{6.84}$$

The evaluation of $E(r_2)$ is then straight forward:

$$\begin{aligned} E(r_2) =& \frac{\rho\,\alpha}{2}\log\left(1 + \frac{a}{z_{\rho s}(\rho + 1)}\right) - \frac{\beta - 1}{2}\log(a) - \frac{\beta + 1}{2}\log(1 - a) \\ &- \frac{(\beta - 1)^2(1 - a)}{4\sqrt{a}}\left[G\left(\frac{a}{1 - a}, \frac{a}{1 - a}\right) - G\left(\frac{a}{1 - a}, -1\right)\right] \\ &+ \frac{\rho\,\alpha(\beta - 1)(1 - a)}{4\sqrt{(z_{\rho s}(\rho + 1) + 1)(z_{\rho s}(\rho + 1) + a)}} \\ &\times\left[G\left(\frac{a}{1 - a}, \frac{z_{\rho s}(1 + \rho) + a}{1 - a}\right) - G\left(\frac{a}{1 - a}, -1\right)\right] \\ &- \frac{(\beta - 1)(1 - a)}{\sqrt{a}}\left[G\left(0, \frac{a}{1 - a}\right) - G\left(0, -1\right)\right] \\ &+ \frac{\rho\,\alpha(1 - a)}{\sqrt{z_{\rho s}(\rho + 1) + a}}\left[G\left(0, \frac{z_{\rho s}(1 + \rho) + a}{1 - a}\right) - G\left(0, -1\right)\right] \\ &+ \frac{\rho\,\alpha}{2}\left[\log\left(\frac{1 - a}{z_{\rho s}(1 + \rho)}\right) - \frac{\rho\,\alpha(1 - a)}{2\sqrt{(z_{\rho s}(\rho + 1) + 1)(z_{\rho s}(\rho + 1) + a)}}\right. \\ &\times\left[G\left(\frac{z_{\rho s}(\rho + 1) + a}{1 - a}, \frac{z_{\rho s}(1 + \rho) + a}{1 - a}\right) - G\left(\frac{z_{\rho s}(\rho + 1) + a}{1 - a}, -1\right)\right] \\ &\left.+ \frac{(\beta - 1)(1 - a)}{\sqrt{a}}\left[G\left(\frac{z_{\rho s}(1 + \rho) + a}{1 - a}, \frac{a}{1 - a}\right) - G\left(\frac{z_{\rho s}(1 + \rho) + a}{1 - a}, -1\right)\right]\right] \end{aligned}$$

$$-\rho\,\alpha r_2 - \frac{1}{2}\Big[(\beta+1)^2\log(\beta+1) - 2\beta^2\log(\beta) + (\beta-1)^2\log(\beta-1)\Big]$$
$$+\alpha(1+\rho)(s+\log(1-s)). \tag{6.85}$$

The case $\rho \approx 0$

Another issue is the behaviour close to $\rho = 0$. We already have seen that $\frac{dE(r)}{dr} = \rho(r)$. At the ergodic point it is $\rho(r_{erg}) = 0$. Therefore, assuming that the first derivative exists, we can write

$$\rho(r) = (r - r_{erg})\frac{d\,\rho(r)}{dr}\bigg|_{r=r_{erg}} + o((r - r_{erg})), \tag{6.86}$$

and we define

$$\left(\frac{d\,\rho}{dr}\right)^{-1}\bigg|_{\rho=0} = v_{erg}\alpha. \tag{6.87}$$

Thus, by differentiating $r(\rho)$ and expressing their values and the values of their derivatives at $\rho = 0$ and $r = r_{erg}$, we have

$$E(r) = \frac{(r - r_{erg})^2}{2v_{erg}} + o\left((r - r_{erg})^2\right), \tag{6.88}$$

where

$$v_{erg} = v_{opt} + \frac{\delta v}{\alpha}, \tag{6.89}$$

where v_{opt} can be seen in (5.49)

$$v_{opt} = -\log\left[\frac{\left(\sqrt{z_0+b_0}+\sqrt{z_0+a_0}\right)^2}{4\sqrt{(z_0+b_0)(z_0+a_0)}}\right], \tag{6.90}$$

$$\delta v = 2g_0 - g_0^2, \tag{6.91}$$

where following the respective case in the wireless domain

$$g_0 = \int_{a_0}^{b_0} p_0(x)\frac{x}{x+z_s}\mathrm{d}x = a_0 + b_0 + 2z_0 - 2\sqrt{(z+a_0)(z+b_0)}, \tag{6.92}$$

For the evaluation of δv we can remind ourselves

$$\frac{dr}{d\,\rho} = \alpha\frac{\partial r}{\partial(\alpha\,\rho)} + \frac{dr}{dz}\frac{dz_{\rho\,s}}{d\,\rho} \tag{6.93}$$

with

$$r = \int_a^b p(x)\left[\log(1+\frac{x}{z_{\rho\,s}}) - \frac{\rho}{1+\rho}\frac{x}{x+z_{\rho\,s}}\right] + \log(1-s) \tag{6.94}$$

and $\frac{\partial p}{\partial z_{\rho\,s}}|_{\rho\approx 0} = 0$

Strictly speaking, the Gaussian approximation is valid for values of r closer to r_{erg}, specifically for $r - r_{erg} = \mathcal{O}(\frac{1}{N})$, there we can neglect in (6.88) terms of order higher than 2.

Analysis

We evaluated the exponent of the Gallager bound for $\beta > 1$ and the corresponding sub–cases. The analysis was based on the respective analysis of the previous section and therefore some tedious evaluations or proofs are omitted. Values away from the optimal curve, means worst behavior. We can see that as we increase α we come closer to the optimal limit [43] which corresponds to infinite codelength ($\alpha \to \infty$). In Fig. 6.4 we can see the behaviour of the error exponent while there is some loss inside the fiber ($n_0 \neq 0$) and in Fig. 6.5 we can see the corresponding behaviour for zero loss ($n_0 = 0$). In most of the curves we identify a phase transition which is indicated with small, black circles. Since, $\rho \in [0, 1]$, there are two major regions in the analysis of the error exponent $0 \leq \rho \leq 1$ and $\rho > 1$. For these two regions, we have respectively $r_{ph} < r \leq r_{erg}$ and $r < r_{ph}$, where r_{erg} is the ergodic r and $r_{ph} > r_{erg}$.

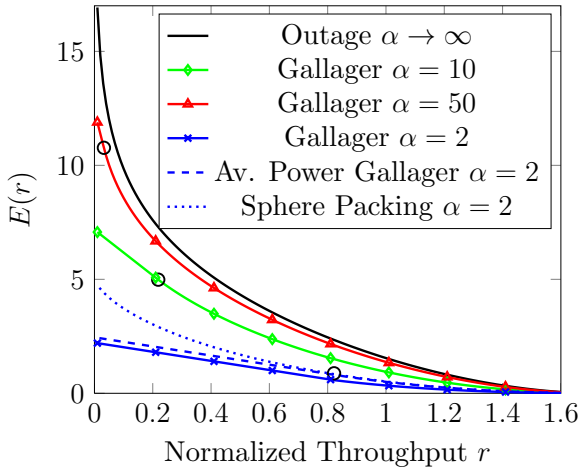


Figure 6.4: As α is increased, the curves for $E(r)$ approach the outage probability exponent [45] (black). The small circles indicate the points where $r = r_1$. For $\alpha = 2$ depicted also the Gallager exponent for the average power constraint ($s = 0$, dashed) and the Sphere Packing Bound error exponent (dotted); $\beta = 3$, $\rho = 10, n_0 = 2$.

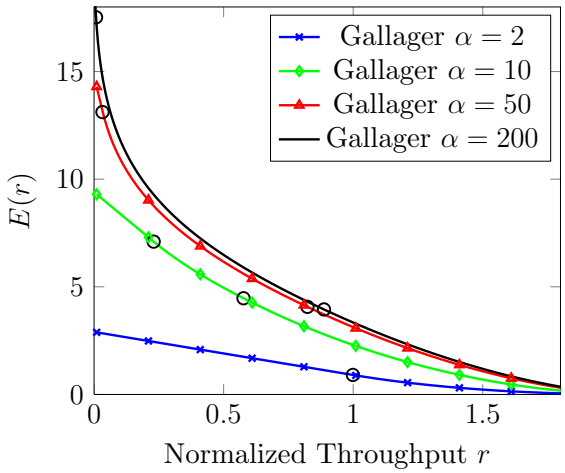


Figure 6.5: $\beta = 3$, $\rho = 10, n_0 = 0$. The small circles indicate the points of phase transition.

6.3 Conclusions

In this chapter we have applied random matrix theory to calculate an analytic expression of the Gallager bound for finite codelength for block fading channels both wireless with Q independent fading blocks within a codeword and the corresponding optical ones. This method is valid for arbitrary normalized rates $r < r_{erg}$, in the large N, K, T limit. As expected, the error exponent increases with Q, resulting to a lower error probability. The limit $Q \to \infty$ is characterized by r_{erg}. Furthermore, when the normalized rate r becomes close to r_{erg} the Gallager exponent becomes asymptotically equal to an upper bound of the fixed optimal error analysis. Other cases for which the joint eigenvalue distribution of the effective channel is known and hence this methodology can be directly applied by using the appropriate weight function $w(x)$ in (6.46), include the uplink MU-MIMO channel [23] and the Amplify-and-Forward channel [13]. It should be noted that more general Gaussian channels, which do not have a known joint eigenvalue distribution can be analyzed in similar ways using the replica method [59].

CHAPTER 7

Optical Fiber MIMO Channel Model- Hamiltonian Approach

7.1 Optical Fiber MIMO Channel Model and its Analysis

Previously, in Chapter 5, we considered the outage capacity of the fiber optical MIMO channel in the limit of full subchannel mixing and in the absence of Mode Dependent Loss (MDL). In this chapter, we move towards a more realistic model of the optical MIMO channel by establishing its analogy with a model from mesoscopic physics: The chaotic cavity [4]. This model is amenable to a random matrix theory analysis and can interpolate from zero to strong mixing between subchannels and includes MDL. To showcase its validity we compute the channel's mutual information via a saddle point analysis.

7.2 Channel Description

The optical fiber may be viewed as a cavity where optical power may enter and exit from both ends. The output power $\mathbf{v}_{\text{out}}$ is related to the input power v_{in} through $\mathbf{v}_{out} = \mathbf{S}\mathbf{v}_{in}$ (see Fig. 3.1). The fiber exhibits random distributed crosstalk between modes or cores. We assume this mixing to be random over different frequency bands, due to strong delay spread. The situation is analogous to that of a chaotic cavity, which randomly mixes the cavity states. The analytic expression of the $2N \times 2N$ scattering matrix for a chaotic cavity reads [4] and we have already discussed in Chapter 3.2

$$\mathbf{S} = \mathbf{I} - 2\pi i \mathbf{W}^{\dagger}(\boldsymbol{\mathcal{H}} + i\pi \mathbf{W}\mathbf{W}^{\dagger})^{-1}\mathbf{W}. \tag{7.1}$$

Here $\boldsymbol{\mathcal{H}}$ is the $2N \times 2N$ channel Hamiltonian and $\mathbf{W}$ is a $2N \times 2N$ matrix containing the coupling constants of the fiber to the outside world.

Statement of Problem

We wish to compute the capacity of the optical MIMO channel. The mutual information is given by the well-known expression

$$\mathcal{I}(\mathbf{y};\mathbf{x}|\mathbf{U}) = \left\langle \log\det\,(\mathbf{I} + \rho_0 \mathbf{U}\mathbf{U}^{\dagger}) \right\rangle, \tag{7.2}$$

where ρ_0 is the signal strength, $\mathbf{U}$ is the complex $N_r \times N_t$ channel matrix where $N = N_t + N_r$ and N_t, N_r are the number transmitted and reflected modes. $\mathbf{x}, \mathbf{y}$ are N_t and N_r dimensional vectors of the transmitted and received signals, respectively. Both are assumed to be zero-mean Gaussian. The maximum of the mutual information over the input distribution yields the capacity of the channel. The capacity is the maximum error-free information transmission rate when the channel matrix $\mathbf{U}$ varies through its whole distribution $p(\mathbf{U})$. We assume $\mathbf{U}$ to be Gaussian distributed and we can remind that

$$\mathbf{U}^\dagger \mathbf{U} = 4\alpha^2\pi^2 \left((\mathbf{H}_0 + \gamma\mathbf{G})^2 + \mathbf{\Gamma}^2\right)^{-1}. \tag{7.3}$$

Here γ is a parameter controlling the randomness, $\mathbf{H}_0$ is a diagonal matrix and corresponds to the line-of-sight component inside the fiber while the Gaussian distributed matrix $\mathbf{G}$ describes the crosstalk and $\mathbf{\Gamma}$ is a diagonal matrix which models the losses inside the fiber.

7.3 Analysis

Replica Theory

We start from the mutual information Eq. (7.2) of an optical MIMO channel. By introducing Eq. (7.3) to Eq. (7.2) we obtain

$$\begin{aligned}
\mathcal{I}(\mathbf{y};\mathbf{x}|\mathbf{H}_0,\mathbf{G}) =& \left\langle \log\det \left(\mathbf{I} + \rho_0 4\alpha^2\pi^2 \left((\mathbf{H}_0 + \gamma\mathbf{G})^2 + \mathbf{\Gamma}^2\right)^{-1}\right)\right\rangle \\
=& \left\langle \log\det\left[(\mathbf{H}_0 + \gamma\mathbf{G})^2 + \mathbf{\Gamma}^2 + \rho\mathbf{I}\right] - \log\det\left[(\mathbf{H}_0 + \gamma\mathbf{G})^2 + \mathbf{\Gamma}^2\right]\right\rangle \\
=& \mathbb{E}\left[\mathcal{I}_1 - \mathcal{I}_2\right],
\end{aligned} \tag{7.4}$$

where

$$\rho = 4\alpha^2\rho_0\pi^2, \tag{7.5}$$
$$\mathbf{F} = \mathbf{\Gamma}^2 + \rho\mathbf{I}, \tag{7.6}$$
$$\mathcal{I}_1 = \log\det\left[(\mathbf{H}_0 + \gamma\mathbf{G})^2 + \mathbf{F},\right] \tag{7.7}$$
$$\mathcal{I}_2 = \log\det\left[(\mathbf{H}_0 + \gamma\mathbf{G})^2 + \mathbf{\Gamma}^2\right]. \tag{7.8}$$

The generating function of (7.2), following [59], is

$$\begin{aligned}
g(\nu) &= \left\langle \left[\det\left(\mathbf{I} + \gamma_0\mathbf{U}^\dagger\mathbf{U}\right)\right]^{-\nu}\right\rangle = \left\langle e^{-\nu\mathcal{I}}\right\rangle \\
&= 1 - \nu\langle\mathcal{I}\rangle + \frac{\nu^2}{2}\langle\mathcal{I}^2\rangle + \ldots
\end{aligned} \tag{7.9}$$

So we have

$$g(\nu_1,\nu_2) = \left\langle e^{-(\nu_1\mathcal{I}_1 + \nu_2\mathcal{I}_2)}\right\rangle. \tag{7.10}$$

We are interested in the mean $\mathcal{I} = \langle \mathcal{I}_1 \rangle - \langle \mathcal{I}_2 \rangle$ and the variance $\mathrm{var}(\mathcal{I}) = \mathrm{var}(\mathcal{I}_1) + \mathrm{var}(\mathcal{I}_2) - 2\mathrm{covar}(\mathcal{I}_1, \mathcal{I}_2)$ of the mutual information. Therefore we have to calculate the mean value of both $\mathcal{I}_1$ and $\mathcal{I}_2$ and also their respective variances and their covariance. In order to set ourselves either to ν_1- or ν_2-space, we set $\nu_2 = 0$ or $\nu_1 = 0$, respectively. The $N \to \infty$ and $\nu \to 0^+$ limits in the evaluation of $g(\nu)$ can be interchanged by first taking the former and then the latter without changing the final answer. Indeed, the two limits of large number of propagating modes and small ν are not different from each other. Higher terms in the ν expansion lead to higher terms in the $\frac{1}{N}$ expansion.

Calculation of $g_{\mathcal{I}_1}(\nu_1)$

Using Identity 1 (Appendix F), we can write $g_{\mathcal{I}_1}(\nu_1) = \left\langle \det\left[(\mathbf{H}_0 + \gamma\mathbf{G})^2 + \mathbf{F}\right]^{-\nu_1} \right\rangle$, as

$$g_{\mathcal{I}_1}(\nu_1) = \int D\mathbf{X} e^{-\frac{1}{2}\mathrm{Tr}\{\mathbf{X}^\dagger\left((\mathbf{H}_0+\gamma\mathbf{G})^2+\mathbf{F}\right)\mathbf{X}\}} = \int D\mathbf{X} e^{-\frac{1}{2}\mathrm{Tr}\{\mathbf{X}^\dagger\mathbf{F}\mathbf{X}\}} \langle e^{-\frac{1}{2}\mathrm{Tr}\{\mathbf{X}^\dagger(\mathbf{H}_0+\gamma\mathbf{G})^2\mathbf{X}\}} \rangle_{\mathbf{G}}. \tag{7.11}$$

Using Identity 2 setting $\mathbf{A}^\dagger = -i\mathbf{X}^\dagger(\mathbf{H}_0 + \gamma\mathbf{G})$ and $\mathbf{B} = i\mathbf{X}(\mathbf{H}_0 + \gamma\mathbf{G})$, we write

$$\langle e^{-\frac{1}{2}\mathrm{Tr}\{\mathbf{X}^\dagger(\mathbf{H}_0+\gamma\mathbf{G})^2\mathbf{X}\}} \rangle = \int D\mathbf{Y} e^{-\frac{1}{2}\left(\mathbf{Y}^T\mathbf{Y}+i\mathbf{Y}^T(\mathbf{H}_0+\gamma\mathbf{G})\mathbf{X}\right)+\left(\mathbf{Y}^T\mathbf{Y}+i\mathbf{X}^T(\mathbf{H}_0+\gamma\mathbf{G})\mathbf{Y}\right)}. \tag{7.12}$$

All-together we obtain

$$g_{\mathcal{I}_1}(\nu_1) = \int D\mathbf{X} \int D\mathbf{Y} e^{-\frac{1}{2}\left(\mathbf{X}^T\mathbf{F}\mathbf{X}+\mathbf{Y}^T\mathbf{Y}\right)} e^{-\frac{i}{2}\left(\mathbf{Y}^T\mathbf{H}_0\mathbf{X}+\mathbf{X}^T\mathbf{H}_0\mathbf{Y}\right)} \langle e^{-\frac{i\gamma}{2}\left(\mathbf{X}^T\mathbf{G}\mathbf{Y}+\mathbf{Y}^T\mathbf{G}\mathbf{X}\right)} \rangle_{\mathbf{G}}. \tag{7.13}$$

The ensemble average over channel realization for an arbitrary function is

$$\langle e^{-\frac{i\gamma}{2}\left(\mathbf{X}^T\mathbf{G}\mathbf{Y}+\mathbf{Y}^T\mathbf{G}\mathbf{X}\right)} \rangle_{\mathbf{G}} = \int D\mathbf{G} e^{-\frac{N}{2}\mathrm{Tr}\{\mathbf{G}^2\}} e^{-\frac{i\gamma}{2}\mathrm{Tr}\{\mathbf{G}\left(\mathbf{X}\mathbf{Y}^T+\mathbf{Y}\mathbf{X}^T\right)\}} \propto e^{-\frac{1}{2}\frac{\gamma^2}{4N}\left(\mathbf{X}\mathbf{Y}^T+\mathbf{Y}\mathbf{X}^T\right)^2}. \tag{7.14}$$

The last exponential can be written as

$$e^{\frac{\gamma}{8N}\mathrm{Tr}\left(2(\mathbf{X}^T\mathbf{X}\mathbf{Y}^T\mathbf{Y})+(\mathbf{X}^T\mathbf{Y})^2+(\mathbf{Y}^T\mathbf{X})^2\right)}. \tag{7.15}$$

In order to evaluate the first term of (7.15) we will use Identity 3 and introduce $\nu_1 \times \nu_1$ matrices $\mathcal{R}, \mathcal{T}$:

$$e^{\frac{\gamma}{8N}\mathrm{Tr}\left(2(\mathbf{X}^T\mathbf{X}\mathbf{Y}^T\mathbf{Y})\right)} = \int D(\mathcal{T},\mathcal{R}) e^{N\mathrm{Tr}(\mathcal{T}\mathcal{R})} e^{-\frac{\gamma}{2}\mathrm{Tr}\left(\mathbf{Y}^T\mathbf{Y}\mathcal{R}+\mathcal{T}\mathbf{X}^T\mathbf{X}\right)}. \tag{7.16}$$

The evaluation of the quadratic parts of the exponential (7.15) is trickier. This time we will introduce $\nu_1 \times \nu_1$ matrices $\mathcal{P}, \mathcal{Q}$. We have

$$e^{\frac{\gamma}{8N}\mathrm{Tr}\left((\mathbf{X}^T\mathbf{Y})^2+(\mathbf{Y}^T\mathbf{X})^2\right)} =$$

$$= \int D\mathcal{P} e^{-N\mathcal{P}^2} e^{-\frac{i\gamma}{2}\mathrm{Tr}\left(\mathcal{P}(\mathbf{X}^T\mathbf{Y}+\mathbf{Y}^T\mathbf{X})\right)} + \int D\mathcal{Q} e^{-N\mathcal{Q}^2} e^{-\frac{i\gamma}{2}\mathrm{Tr}\left(\mathcal{Q}(\mathbf{X}^T\mathbf{Y}-\mathbf{Y}^T\mathbf{X})\right)}. \tag{7.17}$$

Saddle-point analysis

So, bringing everything together, we have

$$g_{\mathcal{I}_1}(\nu_1) = \int D(\mathcal{T},\mathcal{R},\mathcal{P},\mathcal{Q})e^{-\mathcal{S}}, \tag{7.18}$$

where

$$\mathcal{S} = -N\mathrm{Tr}\left(\mathcal{T}\mathcal{R} - \mathcal{P}^2 - \mathcal{Q}^2\right) + \log\det\begin{bmatrix} \mathbf{F}+\gamma\mathcal{T} & i\left(\mathbf{H}_0+\gamma(\mathcal{P}+\mathcal{Q})\right) \\ i\left(\mathbf{H}_0+\gamma(\mathcal{P}-\mathcal{Q})\right) & 1+\gamma\mathcal{R} \end{bmatrix}. \tag{7.19}$$

To consider the vicinity near the saddle-point we rewrite $\mathcal{R},\mathcal{T},\mathcal{P},\mathcal{Q}$ as

$$\begin{aligned} \mathcal{T} &= t\mathbf{I}_{\nu_1} + \delta\mathbf{T} \\ \mathcal{R} &= r\mathbf{I}_{\nu_1} + \delta\mathbf{R} \\ \mathcal{P} &= p\mathbf{I}_{\nu_1} + \delta\mathbf{P} \\ \mathcal{Q} &= q\mathbf{I}_{\nu_1} + \delta\mathbf{Q}, \end{aligned} \tag{7.20}$$

where $\delta\mathbf{T},\delta\mathbf{R},\delta\mathbf{P},\delta\mathbf{Q}$ are $\nu_1\times\nu_1$ matrices which represent deviations around the saddle point.

That way, we can use the Taylor expansion for $\mathcal{S}$ of (7.19) as :

$$\mathcal{S} = \mathcal{S}_0 + \mathcal{S}_1 + \mathcal{S}_2 + \mathcal{S}_3 + \dots \tag{7.21}$$

with

$$\mathcal{S}_0 = -N\mathrm{Tr}(tr - p^2 - q^2) + \log\left[\left(\mathbf{F}+\gamma t\right)\left(\mathbf{I}+\gamma r\right) + \left(\mathbf{H}_0+\gamma(p-q)\right)\left(\mathbf{H}_0+\gamma(p+q)\right)\right]. \tag{7.22}$$

Continuing the evaluation of the next term of the Taylor expansion, since we are looking for saddle point solution, $\mathcal{S}$ must be stationary with respect to variations in $\mathcal{R},\mathcal{T},\mathcal{P},\mathcal{Q}$. Therefore $\mathcal{S}_1 = 0$ and the corresponding saddle-point equations are:

$$r = \frac{1}{N}\mathrm{Tr}\frac{\gamma(1+\gamma r)}{\left(\mathbf{F}+\gamma t\right)\left(\mathbf{I}+\gamma r\right) + \left(\mathbf{H}_0-\gamma p\right)^2} \tag{7.23}$$

$$p = \frac{1}{N}\mathrm{Tr}\frac{\gamma(\mathbf{H}_0-\gamma p)}{\left(\mathbf{F}+\gamma t\right)\left(\mathbf{I}+\gamma r\right) + \left(\mathbf{H}_0-\gamma p\right)^2} \tag{7.24}$$

$$t = \frac{1}{N}\mathrm{Tr}\frac{\gamma(\mathbf{F}+\gamma t)}{\left(\mathbf{F}+\gamma t\right)\left(\mathbf{I}+\gamma r\right) + \left(\mathbf{H}_0-\gamma p\right)^2} \tag{7.25}$$

$$q = 0 \tag{7.26}$$

and the second order term is

$$\mathcal{S}_2 = \frac{1}{2}\mathrm{Tr}\left\{ \begin{bmatrix} \delta\mathbf{T} \\ \delta\mathbf{R} \\ \delta\mathbf{P} \\ \delta\mathbf{W} \end{bmatrix}^T \boldsymbol{\Sigma} \begin{bmatrix} \delta\mathbf{T} \\ \delta\mathbf{R} \\ \delta\mathbf{P} \\ \delta\mathbf{W} \end{bmatrix} \right\}, \tag{7.27}$$

where $\boldsymbol{\Sigma}$ is a 4×4 Hessian matrix the entries of which can be seen in Appendix G.1.

Following [59], the final outcome for the variance is:

$$\left\langle \mathcal{I}^2 \right\rangle - \left\langle \mathcal{I} \right\rangle^2 = -\log\det|\boldsymbol{\Sigma}|\,. \tag{7.28}$$

Calculation of $g_{\mathcal{I}_2}(\nu)$

For the calculation of the $\langle \mathcal{I}_2 \rangle$ and the corresponding variance, we follow the same method as above but instead of $\mathbf{F}$ we only have $\boldsymbol{\Gamma}^2$.

Calculation of $g_{\mathcal{I}_{12}}(\nu_1, \nu_2)$ (Covariance)

Again, using Identity 1

$$g_{\mathcal{I}_1\mathcal{I}_2}(\nu_1,\nu_2) = \int D\mathbf{X}_1 D\mathbf{X}_2 \Bigg\langle \exp\Bigg(-\frac{1}{2}\mathbf{X}_1^\dagger \left[\mathbf{F} + (\mathbf{H}_0 + \gamma\mathbf{G})^2\right]\mathbf{X}_1 - \frac{1}{2}\mathbf{X}_2^\dagger \left[\boldsymbol{\Gamma}^2 + (\mathbf{H}_0 + \gamma\mathbf{G})^2\right]\mathbf{X}_2 \Bigg) \Bigg\rangle. \tag{7.29}$$

Following the previous method we have

$$g_{\mathcal{I}_1\mathcal{I}_2}(\nu_1,\nu_2) = \int D\mathbf{X}_1 D\mathbf{X}_2 \exp\left(-\frac{1}{2}\mathrm{Tr}\left\{ \mathbf{X}_1^\dagger \mathbf{F}\mathbf{X}_1 + \mathbf{X}_2^\dagger \boldsymbol{\Gamma}^2 \mathbf{X}_2 \right\} \right) \times$$
$$\int D\mathbf{Y}_1 D\mathbf{Y}_2 \exp\left(-\frac{1}{2}\mathrm{Tr}\left\{ \mathbf{Y}_1^\dagger\mathbf{Y}_1 + \mathbf{Y}_2^\dagger\mathbf{Y}_2 + i\mathbf{X}_1^\dagger\mathbf{H}_0\mathbf{Y}_1 + i\mathbf{X}_2^\dagger\mathbf{H}_0\mathbf{Y}_2 \right\} \right) \times$$
$$\left\langle \exp\left(-\frac{i\gamma}{2}\left(\mathbf{X}_1^\dagger\mathbf{G}\mathbf{Y}_1 + \mathbf{X}_2^\dagger\mathbf{G}\mathbf{Y}_2 \right) \right) \right\rangle_{\mathbf{G}}. \tag{7.30}$$

Again, using Identity 2 and Identity 3 we introduce the ever-helpful $\nu_1 \times \nu_1$ matrices $\mathcal{R}_1, \mathcal{T}_1, \mathcal{P}_1, \mathcal{Q}_1$, $\nu_2 \times \nu_2$ matrices $\mathcal{R}_2, \mathcal{T}_2, \mathcal{P}_2, \mathcal{Q}_2$ and $\nu_1 \times \nu_2$ matrices $\mathcal{R}_{12}, \mathcal{T}_{12}, \mathcal{P}_{12}, \mathcal{Q}_{12}$ and following the diagonalizing method we end up to matrix $\mathbf{A}$ which can be seen in Appendix G.2.

At the saddle point the cross-terms $\mathcal{R}_{12}, \mathcal{T}_{12}$ etc are equal to zero. So $\mathbf{A}$ becomes a block-diagonal matrix and the evaluation of the determinant (Identity 1) is just the multiplication of these 2 blocks.

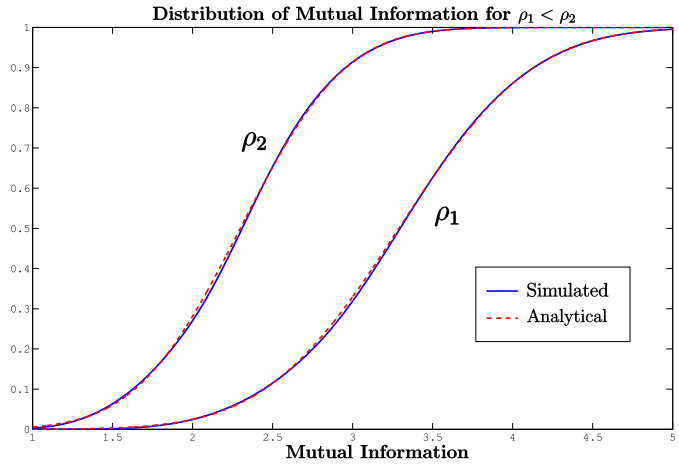


Figure 7.1: Cumulative distribution function (CDF) of mutual information for N = 6 and $\rho_1 < \rho_2$. Runs = 10^6

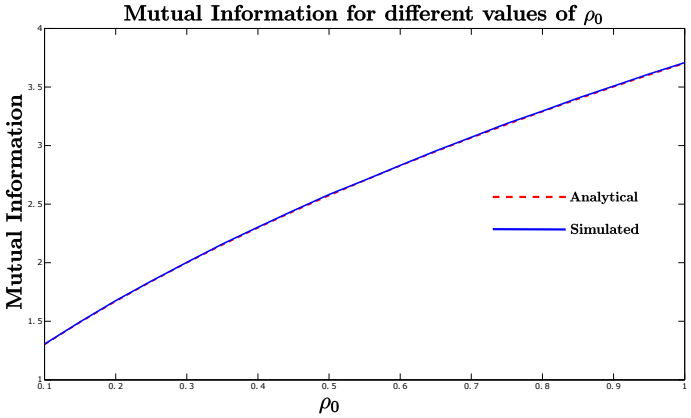


Figure 7.2: Comparison of the calculation of the mutual information. $N = 6$. Runs= 10^6

The rest of the calculations for the computation of the covariance is straightforward and the Hessian matrix for the covariance can be seen in Appendix G.3.

So finally the variance is

$$Var = -\log\det|\mathbf{\Sigma}_{\mathcal{I}_1}| - \log\det|\mathbf{\Sigma}_{\mathcal{I}_2}| + 2\log\det|\mathbf{\Sigma}_{cov}| + 4\log 2\,. \tag{7.31}$$

7.4 Numerical results

In order to check our analytical results we have numerically computed mutual information and its cumulative distribution function (CDF). The results are shown in Figs. 7.1 and 7.2. We compare the Gaussian distribution $\mathcal{N}(\langle\mathcal{I}\rangle, \mathrm{var}(\mathcal{I}))$ evaluated using the analytical calculations in this chapter with results obtained by averaging over a large

number of random matrix realizations for $N = 6$. Remarkably, we find that in both cases, the discrepancy between the different curves is small. This shows that our analytical results are valid even for the practical relevant case of small number of modes or cores N.

7.5 Conclusions

In this chapter we have introduced a novel channel model for the optical MIMO channel by relating it to a chaotic cavity. This new modeling approach captures the fundamental properties of the optical MIMO channel. Using tools from random matrix theory and a saddle point analysis, we have shown that in the limit of large N, the distribution of the mutual information approaches a Gaussian. Numerically we find that this method is valid even for low N. This analytic method gives us the means to analyze the statistics of throughput in the fiber optical MIMO channel in the presence of arbitrary level of crosstalk and MDL. The introduced channel model is also amenable to extensions, such as dispersion and nonlinear effects in deterministic and random part respectively.

APPENDIX A

The diagonal matrices $\mathbf{A}_{\text{diag}}$ and $\mathbf{B}_{\text{diag}}$

Let be $2N \times 2N$ matrices $\mathbf{A}_{\text{diag}}$ and $\mathbf{B}_{\text{diag}}$

$$\mathbf{A}_{\text{diag}} = \begin{bmatrix} 1 & & & & \\ & 1 & & 0 & \\ & & 1 & & \\ & 0 & & \ddots & \\ & & & & 0 \end{bmatrix} \tag{A.1}$$

where the number of "1"s equals the number of the receivers N_r and

$$\mathbf{B}_{\text{diag}} = \begin{bmatrix} 0 & & & & \\ & 0 & & 0 & \\ & & 1 & & \\ & 0 & & \ddots & \\ & & & & 1 \\ & & & & & 0 \end{bmatrix} \tag{A.2}$$

where the number of "1"s equals the number of the transmitters N_t and the number of lower "0"s equals $N - N_t$. In general, to model the power loss inside the fiber we introduce N_0 so as $N = N_r + N_t - N_0$ and after some minor algebra we have $\tilde{N} = \tilde{N}_r + \tilde{N}_t$ [15]. So, without loss of generality we can omit N_0 in our calculations.

APPENDIX B

Details for Derivation of Closed Form Solution

In this appendix we provide details of the derivation of the closed form expression for the outage presented in Section 5.3. We start with (5.6). It is convenient to make the change of variables $1+\rho\lambda_k = y_k$ to get

$$\begin{aligned} 1 - P_{out}(r) = \mathcal{A} \int_{-\infty}^{\infty} \frac{dp}{2\pi} \frac{e^{\mathrm{i}pN_t r}}{\epsilon - \mathrm{i}p} \int_{[1,1+\rho]^{N_t}} d\mathbf{y} \prod_{n<m} |y_n - y_m|^2 \\ \times \prod_k \left[y_k^{-\mathrm{i}p} (y_k - 1)^{|N_t - N_r|} \left((\rho+1) - y_k \right)^{N_0} \right] \\ \mathcal{A} = \frac{1}{Z_N \rho^{N_t^2 + (|N_t - N_r| + N_0)N_t}} \end{aligned}$$

where $d\mathbf{y} = dy_1 \cdots dy_{N_t}$ and $\mathcal{A}$ a normalization constant. Now we invoke the Andréief identity (see also Lemma 8 in [73]), which takes advantage of the fact that the products of the form $\prod_{n,m}(y_n - y_m)$ can be written as a Vandermonde determinant. Defining the function

$$g(x,p) = x^{-\mathrm{i}p}(x-1)^{|N_t - N_r|} \left((\rho+1) - x \right)^{N_0} \tag{B.1}$$

we have

$$\begin{aligned} &\int_{[1,1+\rho]^{N_t}} d\mathbf{y} \prod_{k=1}^{N_t} g(y_k,p) \det\left(y_i^{j-1}\right)^2 \qquad \text{(B.2)} \\ &= \int_{[1,1+\rho]^{N_t}} d\mathbf{y} \prod_{k=1}^{N_t} g(y_k,p) \left(\sum_{\mathbf{a}(N_t)} (-1)^{|\mathbf{a}|} \prod_{i=1}^{N_t} y_i^{a_i - 1} \right)^2 \\ &= \sum_{\mathbf{a}(N_t),\mathbf{b}(N_t)} (-1)^{|\mathbf{a}|+|\mathbf{b}|} \int_{[1,1+\rho]^{N_t}} d\mathbf{y} \prod_{k=1}^{N_t} g(y_k,p) y_k^{a_k + b_k - 2} \\ &= N_t! \det\left(H_{i+j}(p)\right)_{i,j=1,\ldots,N_t} \end{aligned}$$

In the second line we used the Leibnitz expansion of determinants [73], where the sum is over all permutations $\mathbf{a}$ with $(-1)^{|\mathbf{a}|}$ being the sign of the permutation. In the final line we re-summed the integrated quantities to get a determinant of the Hankel matrix $\mathbf{H}$ with elements H_{i+j} which can be expressed as

$$H_\ell(p) = \int_1^{1+\rho} dx\; x^{\ell - 2 - \mathrm{i}p} (x-1)^{|N_t - N_r|} ((1+\rho) - x)^{N_0}$$

$$=\rho^{N_0+|N_t-N_r|+1}B(b,c-b)F(a,b;c;-\rho) \tag{B.3}$$

where $B(b,c-b)$ and $F(a,b;c;z)$ are the Beta and hypergeometric functions respectively [29], for the parameters $a=2+\mathrm{i}p-\ell$, $b=|N_t-N_r|+1$, $c=N_0+|N_t-N_r|+2$. This results to the following expression for the outage probability

$$1-P_{out}(r)=\mathcal{A}'\int_{-\infty}^{\infty}\frac{dp}{2\pi}\,\frac{e^{\mathrm{i}pN_tr}}{\epsilon-\mathrm{i}p}\det\left(H_{i+j}(p)\right)_{i,j=1,\ldots,N_t} \tag{B.4}$$

where $\mathcal{A}'=N_t!\mathcal{A}$.

The integral (B.3) can be evaluated in the most elementary form exploiting the fact that both $|N_t-N_r|$ and N_0 are integers. Using the binomial theorem to expand the second and third powers, we get

$$H_\ell(p)=\sum_{k=0}^{|N_t-N_r|}\sum_{n=0}^{N_0}c_{k,n}\frac{(1+\rho)^{\ell-1-\mathrm{i}p+k+N_0-n}-1}{\ell-1-\mathrm{i}p+k+N_0-n} \tag{B.5}$$

where

$$c_{k,n}=\binom{|N_t-N_r|}{k}\binom{N_0}{n}(-1)^{|N_t-N_r|-k+N_0-n}(1+\rho)^n \tag{B.6}$$

Let us now expand the determinant in (B.4) using (B.5). After rearranging the sums we get

$$1-P_{out}(r)=\sum_{\mathbf{k},\mathbf{n}}C_{\mathbf{k},\mathbf{n}}\sum_{\boldsymbol{\sigma}(N_t)}(-1)^{|\boldsymbol{\sigma}|}J\left(r;\{\mathbf{s}_{\boldsymbol{\sigma}}\}\right) \tag{B.7}$$

where the sum over the integer components of the vector $\mathbf{k}=[k_1,\cdots,k_{N_t}]$ is over the interval $[0,|N_t-N_r|]$, while for the vector $\mathbf{n}=[n_1,\ldots,n_{N_t}]$ its components are summed over the interval $[0,N_0]$. Also, $C_{\mathbf{k},\mathbf{n}}=\mathcal{A}'\prod_i c_{k_i,n_i}$ and

$$J(r;\{\mathbf{s}_{\boldsymbol{\sigma}}\})=\int_{-\infty}^{\infty}\frac{dp}{2\pi}\,\frac{e^{\mathrm{i}pN_tr}}{\epsilon-\mathrm{i}p}\prod_{j=1}^{N_t}\frac{(1+\rho)^{s_j-\mathrm{i}p}-1}{s_j-\mathrm{i}p} \tag{B.8}$$

where the components of the integer vector $\mathbf{s}_{\boldsymbol{\sigma}}$ are $s_j=j+\sigma_j-1+k_j+N_0-n_j$. Expanding the numerator of the above equation, we obtain

$$\begin{aligned}
J(r;\{\mathbf{s}\})&=\sum_{\ell=0}^{N_t}(-1)^{\ell+N_t}d_\ell(\mathbf{s})F(N_tr-\ell\log(1+\rho),\mathbf{s})\\
d_\ell(\mathbf{s})&=\mathbf{e}_\ell\left((1+\rho)^{s_1},\ldots,(1+\rho)^{s_{N_t}}\right) &\text{(B.9)}\\
F(z,\mathbf{s})&=\int_{-\infty}^{\infty}\frac{dp}{2\pi}\,\frac{e^{\mathrm{i}pz}}{(\epsilon-\mathrm{i}p)\prod_{i=1}^{N_t}(s_i-\mathrm{i}p)} &\text{(B.10)}
\end{aligned}$$

where in the second line we have used the elementary symmetric polynomials $\mathbf{e}_\ell(x_1,x_2,\cdots,x_{N_t})$ of degree ℓ.

As a result, in order to evaluate the outage probability in closed form we only need to evaluate the complex integral in $F(z, \mathbf{s})$. Since all poles of the integrand are in the lower half complex p-plane, if $z > 0$ (hence $N_t r > \ell \log(1+\rho)$) then the integral vanishes [63]. Hence only ℓ-terms with $N_t r < \ell \log(1+\rho)$ survive. Having this in mind the integral can be evaluated by summing over the residues of the poles. As a result we obtain

$$F(z, \mathbf{s}) = \prod_{j=1}^{N_t} s_j^{-1} + \sum_{j=1}^{N_t} \frac{e^{s_j z}}{s_j \prod_{k \neq j}(s_k - s_j)} = \prod_{j=1}^{N_t} s_j^{-1} + F_1(z, \mathbf{s}) \tag{B.11}$$

Putting all above formulae together provides the final result expressed in (5.7).

Before concluding this section, it is worth discussing the value of the above equation when two or more integers s_i are equal. To address this issue it will prove useful to express $F_1(z, \mathbf{s})$ as a ratio of determinants [46]. Indeed we get

$$F_1(z, \mathbf{s}) = \frac{\det\left(f_i(s_j, z)\right)}{\prod_{n>m}(s_n - s_m)} \tag{B.12}$$

where the elements of the vector function $\mathbf{f}(x, z)$ is defined as follows

$$f_i(x, z) = \begin{cases} \frac{e^{xz}}{x} & i = 1 \\ x^{i-1} & N_t \geq i > 1 \end{cases} \tag{B.13}$$

When one or more values of s_j are identical, the ratio is ill-defined, because both numerator and denominator vanish. Although we could have dealt with the problem directly at the level of complex integration by considering double poles, it is more instructive to analyze this case as a limit of the s's approaching each other. Following Lemma 1 in [58] we can show that if s_1 has multiplicity m then $F_1(z, \mathbf{s})$ can be expressed as

$$F_1(z, \mathbf{s}) = \frac{\det \mathbf{Z} \prod_{q=1}^{m-1}\{q!\}^{-1}}{\prod_{a>b>m}(s_a - s_b) \prod_{j=m+1}^{N_t}(s_j - s_1)^m} \tag{B.14}$$

where the matrix $\mathbf{Z}$ can be expressed as

$$\mathbf{Z} = \left[\mathbf{f}(s_1, z); \mathbf{f}'(s_1, z); \ldots; \mathbf{f}^{(m-1)}(s_1, z); \mathbf{f}(s_{m+1}, z); \ldots; \mathbf{f}(s_{N_t}, z)\right] \tag{B.15}$$

where the primes represent partial derivative with respect to the first argument. We can similarly obtain expressions for the case when we have several multiplicities in $\mathbf{s}$.

APPENDIX C

Proofs in Chapter 5

C.1 Proof of Theorem 5.1

In this appendix we will provide some details on the proof of the above theorem.

Convexity

The convexity of $\mathcal{E}[p]$ has been shown in [3] over functions in $\mathcal{X}$, as also in [45].

Uniqueness

The uniqueness of the minimum of $\mathcal{E}[p]$ has been shown in [3,34]. The value of $\mathcal{E}_0$ can be obtained form the limit $\mathcal{E}_0 = -\lim_{N_t\to\infty} \log \mathcal{Z}_{N_t}/N_t^2$. However, the normalization factor $\mathcal{Z}_{N_t}$ can be evaluated explicitly using the Selberg integral [23] as follows:

$$\mathcal{Z}_N = \prod_{k=0}^{N-1} \frac{\Gamma(N(\beta-1)+1+k)\Gamma(Nn+1+k)\Gamma(k+2)}{\Gamma(N(\beta+n)+k+1)} \tag{C.1}$$

Using the Stirling approximation for the Γ-functions and approximating the sums with integrals, we get that

$$\begin{aligned}\mathcal{E}_0 =& \frac{(\beta+n+1)^2}{2}\log(\beta+n+1) - \frac{(\beta+n)^2}{2}\log(\beta+n) \\ &- \frac{\beta^2}{2}\log\beta + \frac{(\beta-1)^2}{2}\log(\beta-1) - \frac{(1+n)^2}{2}\log(1+n) + \frac{n^2}{2}\log n\end{aligned} \tag{C.2}$$

Exponential Asymptote of $Prob(I_N < N_t r)$

Let $\mathcal{X}_r$ be the set given by

$$\mathcal{X}_r = \left\{ p \in \mathcal{X} \text{ and } \int_0^1 p(x)\log(1+\rho x)\,dx \le r \right\} \tag{C.3}$$

Given the linearity of the constraint, the above set is convex. Now, in [34] it has been shown that $Prob(I_N \le r)$ obeys the large deviation principle with good rate function $I[p] = \mathcal{E}[p] - \mathcal{E}_0$. Hence,

$$\mathcal{E}_0 - \inf_{p\in\mathcal{X}_r} \mathcal{E}[p] = -\limsup_{N_t\to\infty} \frac{1}{N_t^2}\log P(\mathcal{X}_r) = -\liminf_{N_t\to\infty} \frac{1}{N_t^2}\log P(\mathcal{X}_r) \tag{C.4}$$

The analogous result can be obtained for Corollary 5.1 by noting that the complement of $\mathcal{X}_r$, namely $\mathcal{X}_r^c$ is also convex. Then the above result follows directly for $P(I_N > r)$.

C.2 Proof of Lemma 5.1

We first define

$$\Lambda(k) = \sup_{c,\nu\geq 0}\inf_{p}\left(k\int_0^1 p(x)\log(1+\rho x)\,dx - \mathcal{L}_0[p,\nu,c]\right) \tag{C.5}$$

By the definition of (5.18) we have $\mathcal{E}(r) = \sup_k(kr - \Lambda(k))$.

We start by noting that for $r > r_{erg}$, $\mathcal{E}(r)$ can be obtained by optimizing over only $k > 0$, i.e. $\mathcal{E}(r) = \sup_{k>0}(kr - \Lambda(k))$. Indeed, for any $k \leq 0$

$$kr - \Lambda(k) \leq kr_{erg} - \Lambda(k) \leq \mathcal{E}(r_{erg}) = \mathcal{E}_0 \tag{C.6}$$

Hence, since we have $\inf_r \mathcal{E}(r) = \mathcal{E}_0$ from Theorem 5.1(b), the quantity $kr - \Lambda(k)$ will have its supremum for positive k. Part (1) of this Lemma is then proved by taking the derivative of $\mathcal{E}(r)$ with respect to r, which is

$$\mathcal{E}'(r) = k(r) > 0 \tag{C.7}$$

where $k(r)$ is the k-value of the supremum. We can similarly prove that $k(r) < 0$ for $r < r_{erg}$, hence showing part (b) of Lemma 5.1.

To prove convexity, let $\theta \in [0,1]$. Then we have

$$\begin{aligned}
&\theta\mathcal{E}(r_1) + (1-\theta)\mathcal{E}(r_2) \\
&= \sup_k(\theta k r_1 - \theta\Lambda(k)) + \sup_k((1-\theta)kr_2 - (1-\theta)\Lambda(k)) \\
&\geq \sup_k((\theta r_1 + (1-\theta)r_2)k - \Lambda(k)) = \mathcal{E}(\theta r_1 + (1-\theta)r_2)
\end{aligned} \tag{C.8}$$

C.3 Uniqueness of solution of (5.36)

In this appendix we will show the uniqueness of solution of the normalization equation (5.36)

$$\frac{n}{\sqrt{1-b}} + k\sqrt{\frac{1}{1+b\rho}} = 2 + n + k$$

The left hand side of the above equation can, also, be identified as the in-parenthesis element of the eigenvalues density equation (5.35) for $x = b$. We can set

$$f(b) = \frac{n}{\sqrt{1-b}} + k\frac{1}{\sqrt{1+\rho b}}$$

and taking the first derivative

$$f'(b) = \frac{n}{(1-b)^{3/2}} - \frac{k\rho}{(1+\rho b)^{3/2}}$$

- If $k < 0$ it is $f'(b) > 0$ and so, $f(b)$ is monotonous and (5.36) has unique solution
- If $k > 0$ we also need the second derivative

$$f''(b) = \frac{3}{2}\frac{n}{(1-b)^{5/2}} + \frac{3}{2}\frac{k\rho^2}{(1+\rho b)^{5/2}} > 0$$

The minimum value of $f(b)$ can be found for $b = 0$ equal to $f(b)_{min} = n + k < 2 + n + k$, which is the right hand side of the (5.36), and the maximum value is for $b = 1$, equal to $f(b)_{max} \to \infty$. Finally, because $f'(b) = 0$ has one real root, we can visualize that again (5.36) has a unique solution.

The same procedure can be used to derive the respective solution uniqueness for the other cases.

APPENDIX D
Proofs in Chapter 6.1

D.1 Proof of Theorem 6.1

For $Q = 1$ we study the limit

$$E(r) = -\lim_{N\to\infty} \frac{1}{N^2} \log \mathbb{E}_{\mathbf{H}} \left[e^{-N^2 f[\mu_N]} \right], \tag{D.1}$$

where $\mu_N(x) = \frac{1}{N}\sum_i \delta(x-\lambda_i)$, λ_i are the eigenvalues of $\mathbf{H}\mathbf{H}^\dagger$ and $f[p]$ is defined on $\mathcal{M}(\mathbb{R}^+) \to \mathbb{R}$, where $\mathcal{M}(\mathbb{R}^+)$ is the space of probability measures on $\mathbb{R}^+$ as

$$f[p] = \alpha \max_{\substack{\rho\in[0,1]\\ s\in[0,1)}} \left\{ \rho \int_{\mathbb{R}^+} \log\left(1+\frac{x}{z_{\rho\, s}}\right) p(x)dx - \rho\, r + (1+\rho)(s+\log(1-s)) \right\}, \tag{D.2}$$

where the argument of max is defined as $g[\rho, s, p]$. It is therefore important to show a number of properties of $f[p]$ and $g[\rho, s, p]$. First, when $\rho = s = 0$, the function g vanishes, i.e. $g[0, 0, p] = 0$, so that $f[p] \geq 0$. Second, $f[p]$ is continuous in p, for which Berge's Maximum theorem [5] can be invoked. Third, $f[p]$ is convex in p, which can be shown directly from its definition. Fourth, $g[\rho, s, p]$ is quasi-concave in ρ, s. To show this we start by noting that, excluding the term $\rho\, s$ in (D.2), $g[\rho, s, p]$ is concave in both ρ, s. Hence, since $\rho\, s$ is quasi-concave, so is $g[\rho, s, p]$. Therefore for all $p \in \mathcal{M}(\mathbb{R}^+)$ for which the integral $\left|\int_0^\infty \log(x)p(x)dx\right| < \infty$, $g[\rho, s, p]$ has a global maximum in $\rho \in [0, 1]$, $s \in [0, 1)$.

Varadhan's Lemma

We now wish to invoke Varadhan's Lemma. To do so, we first provide the following definitions:

Definition D.1. *[18]* A rate function $I[p]$ is a lower semicontinuous mapping $I : \mathcal{M}(\mathbb{R}^+) \to [0, \infty]$, for which all level sets are closed. If, in addition, the level sets are compact, then $I[p]$ is called a good rate function.

Definition D.2. *[3, 18]* The probability law μ_N satisfies the large deviation principle in the scale N^2 with rate function I if, for all subsets of $\Gamma \subset \mathcal{M}(\mathbb{R}^+)$

$$-\inf_{p\in\Gamma^o} I[p] \leq \liminf_{N\to\infty} \frac{1}{N^2} \log \mu_N(\Gamma) \leq \limsup_{N\to\infty} \frac{1}{N^2} \log \mu_N(\Gamma) \leq -\inf_{p\in\bar{\Gamma}} I[p], \tag{D.3}$$

where Γ^o and $\bar{\Gamma}$ are the interior and closure of Γ, respectively.

We now note that $f[p]$ is continuous. In addition, since $f[p] \geq 0$ for every p, then for any $\gamma > 0$

$$\limsup_{N\to\infty} \frac{1}{N^2} \log \mathbb{E}\left[e^{-\gamma N^2 f[\mu_N]}\right] < \infty. \tag{D.4}$$

Furthermore, in [34] it was shown that μ_N, the probability law of the $\{\lambda_i\}$ satisfies a large deviation principle with good rate function given by

$$\begin{aligned} I[p] = &\int\int \log|x-y| p(x)p(y)dxdy \\ &+ \int (x-(\beta-1))\log(x)p(x)dx - \frac{1}{2}\left(3\beta - \beta^2\log\beta + (\beta-1)^2\log(\beta-1)\right). \end{aligned} \tag{D.5}$$

As a result, Varadhan's Lemma can be applied to (D.1) to show that the limit exists and is equal to

$$E(r) = \inf_{p\in\mathcal{M}(\mathbb{R}^+)} \left(f[p] + I[p]\right). \tag{D.6}$$

Furthermore, it is possible to show that $I[p]$ is a convex function of p. This follows directly from [3, 45] by observing the quadratic dependence of $I[p]$ in p. Therefore, since $f[p] + I[p]$ is convex in p its infimum has a unique solution. Taking into account the definition of $f[p]$ and its concave-convex properties discussed above, we may apply Sion's theorem [74] to exchange the order in which the max – inf are applied. Therefore, $E(r)$ in (D.6) can be expressed as

$$E(r) = \max_{\substack{\rho\in[0,1]\\ s\in[0,1)}} \inf_{p\in\mathcal{M}(\mathbb{R}^+)} \left(g[\rho,s,p] + I[p]\right). \tag{D.7}$$

Explicit Solution of optimum $p(x)$ and Evaluation of $E(r)$

To solve the above optimization problem (D.7), we introduce the Lagrangian functions

$$\mathcal{L}_0[p,c] = f[p] + I[p] - c\left(\int p(x)dx - 1\right), \tag{D.8}$$

$$\mathcal{L}_1[p,c,\rho,s] = \mathcal{L}_0[p,c] + \alpha(\rho+1)\left(s + \log(1-s)\right) + \alpha\,\rho\left(\int \log\left(1 + \frac{x}{z_{\rho\,s}}\right)p(x)dx - r\right). \tag{D.9}$$

Since $\mathcal{L}_1$ is convex in p and concave in c, ρ and s, the saddle point is unique [8] and we obtain

$$E(r) = \sup_{c,\rho,s}\inf_{p} \mathcal{L}_1[p,c,\rho,s]. \tag{D.10}$$

Taking advantage of the convexity in p, in order to find the infimum of $\mathcal{L}_1$ we will take the functional derivative with respect to p, which is defined as

$$\delta\mathcal{L}_1[p] = \frac{d}{dt}\bigg|_{t=0} \mathcal{L}_1[p^* + t\phi], \tag{D.11}$$

where $(p + t\phi) \in \mathcal{M}(\mathbb{R}^+)$ and ϕ is a test function. This can be re-written as

$$\delta\mathcal{L}_1[p] = \int \phi(x)\Psi[p^*, x]dx, \tag{D.12}$$

where

$$\Psi[p^*, x] = -2\int p(y)\log|x-y|dy + x - (\beta-1)\log(x) - c + \alpha\,\rho\log\left(1 + \frac{x}{z_{\rho\,s}}\right). \tag{D.13}$$

At the minimum, (D.12) must vanish identically for all φ, thus $\Psi[p^*, x] = 0$ and it follows that

$$2\int \log|x-y|p^*(y)dy = x - (\beta-1)\log(x) - c + \alpha\,\rho\log\left(1 + \frac{x}{z_{\rho\,s}}\right). \tag{D.14}$$

Next, we differentiate (D.13) with respect to x to obtain

$$2\mathrm{PV}\int \frac{p^*(x)}{x-y}dx = 1 - \frac{\beta-1}{x} + \frac{\alpha\,\rho}{z_{\rho\,s}+x}, \tag{D.15}$$

where PV denotes the principle value. Following [80], the solution of the last equation is given by

$$p^*(x) = \frac{1}{2\pi\sqrt{(x-a)(b-x)}}\left[-x - (\beta-1)\frac{\sqrt{ab}}{x} + \frac{\alpha\,\rho\sqrt{(z_{\rho\,s}+a)(z_{\rho\,s}+b)}}{(x+z_{\rho\,s})} + C\right], \tag{D.16}$$

where C is an unknown constant and $a,\ b$ are the (unspecified) endpoints of the support of $p(x)$. Since $p^*(x)$, if it exists, is unique, we search for a solution among continuous, non-negative, normalized functions over $x \in (0, \infty)$. Continuity at $x = b$ demands that $p^*(b) = 0$, which fixes the value of C above, while continuity at $x = a$, i.e. $p^*(a) = 0$ results to (6.19). Furthermore, the normalization condition on $p(x)$, $\int_a^b p(x)dx = 1$, results to (6.20). In Appendix D.2 it is shown that for fixed ρ there is a unique solution of (6.19) and (6.20) for $0 < a < b$. For $s \in [0, 1)$, $\mathcal{L}_1$ is concave. Hence the maximum over s results when the first derivative of $\mathcal{L}_1[p^*]$ with respect to s vanishes, hence (6.21).

Once we have determined the value of s as a function of ρ, we now search for the optimal value of ρ. Extending its support to $\rho \in [0, \infty)$, the extremal value of ρ is determined by (6.22). However, since the optimization of ρ is over $[0, 1]$, then there are

two possible types of solution: For $r_{erg} \geq r \geq r_1$, the optimal value of $\rho < 1$, hence the value of ρ is determined by (6.22). In contrast, for $r < r_1 = \bar{r}(1)$, the optimal value of ρ is fixed to the boundary of the region, i.e. $\rho = 1$.

Having determined the values of a, b, s, ρ, we may now integrate the expression in (D.9) to evaluate $E(r)$. The integrals appear in (6.18) in Theorem 1. All single integrals over $p(x)$ can be evaluated in closed-form directly. (D.13) can be used to simplify the double integral over p into a single one, which then can be evaluated directly. The value of c in (D.13) can be obtained by evaluating $\Psi[p, x]$ at $x = a$. Finally, to obtain the value of $E(r)$ for general Q, we make the substitution $\alpha \to \frac{\alpha}{Q}$ and $E(r) \to QE(r)$ as discussed in Remark 6.3.

D.2 Proof of uniqueness of solution of (6.19),(6.20)

To show that (6.19) and (6.20) have a unique solution, we observe that the normalization integral $n(b) = \int_{a(b)}^{b} p(x)dx$ is an increasing function of b since its derivative can be expressed as

$$n'(b) = \frac{1}{4z_{\rho s}} \left(1 + \frac{(\beta - 1)z_{\rho s}}{\sqrt{a(b)b^3}}\right) \frac{z_{\rho s} - a(b)}{z_{\rho s} + b} > 0. \tag{D.17}$$

As $a(b)$ is a decreasing function and bounded below by 0, we have $\lim_{b\to\infty} n(b) = +\infty$ and by continuity there will be a unique b^* such that $n(b^*) = 1$. Thus, both $a^* = a(b^*)$ and $b = b^*$ will be the unique solution to (6.19) and (6.20).

D.3 Proof of Theorem 6.2

Let us re–write (D.6) as:

$$E(r) = Q \underbrace{\left(\frac{1}{Q} f[p^*] + I[p^*]\right)}_{J_Q[p^*]}, \tag{D.18}$$

where p^* is the function p at the infimum of $J_Q[p]$. To examine the behavior of the error exponent $E(r)$ for large Q, we analyze the derivative of J_Q with respect to Q. Since $J_Q[p]$ is stationary at p^*, its variations with respect to p vanish at p^*. Hence, since both $f[p^*]$ and $I[p^*]$ do not depend explicitly on Q we obtain

$$\frac{dJ_Q}{dQ} = \frac{\partial J_Q}{\partial Q} = -\frac{1}{Q^2} f[p^*]. \tag{D.19}$$

Now, as Q grows, it can be seen from (6.19) and (6.20) that $p^*(x)$ converges to the Marčenko-Pastur distribution and a, b converge to a_0 and b_0 respectively. Hence, $f[p^*]$ becomes $f[p_0]$. Therefore if we integrate (D.19) between (Q, ∞) we find that to leading order in Q, $J_Q \approx \frac{f[p_0]}{Q}$. Multiplying $J_Q[p^*]$ with Q as in (D.18), we obtain (6.34).

APPENDIX E

$G(x, y)$ and $I_3(x)$ Function

The function $G(x, y)$ for $x > 0$ and $y > 0$ or $y < -1$ is given by

$$\begin{aligned} G(x,y) &= \frac{1}{\pi}\int_0^1 \sqrt{t(1-t)}\frac{\log(t+x)}{t+y}dt \qquad \text{(E.1)} \\ &= -2\text{sgn}(y)\sqrt{|y(1+y)|}\log\left[\frac{\sqrt{x|1+y|}+\sqrt{|y|(1+x)}}{\sqrt{|1+y|}+\sqrt{|y|}}\right] \\ &\quad + (1+2y)\log\left[\frac{\sqrt{1+x}+\sqrt{x}}{2}\right] - \frac{1}{2}\left(\sqrt{1+x}-\sqrt{x}\right)^2 \end{aligned}$$

This result first appeared in [12] and then later in [45], both for $y > 0$. It can readily be shown that it can be generalized for $y < -1$. $I_3(x)$ appearing in [12, 83] can be shown to be $I_3(x) = -G(x, -1)$.

APPENDIX F

Identities

The proofs for the next identities , can be found in [59].

Identity 1. *Let* $\mathbf{M}$ *be a hermitian, positive, definite square matrix* $m \times m$ *and* $\mathbf{X}$ *a complex* $m \times n$ *matrix, then*

$$(detM)^{-1} = \int D\mathbf{X} e^{-\frac{1}{2}Tr\{\mathbf{X}^\dagger \mathbf{M} \mathbf{X}\}}. \tag{F.1}$$

Identity 2. *Let,* $\mathbf{X}, \mathbf{A}, \mathbf{B}$ *be* $m \times n$ *complex matrices, then*

$$\int D\mathbf{X} e^{-\frac{1}{2}Tr\{\mathbf{X}^\dagger \mathbf{X} + \mathbf{A}^\dagger \mathbf{X} - \mathbf{X}^\dagger \mathbf{B}\}} = e^{-\frac{1}{2}Tr\{\mathbf{A}^\dagger \mathbf{B}\}}. \tag{F.2}$$

Identity 3 (Hubbard-Stratonovich transformation)**.** *Let,* $\mathbf{U}, \mathbf{V}$ *be arbitrary complex* $\nu \times \nu$ *matrices, where* ν *here is assumed to be an arbitrary positive integer. Then,*

$$e^{-Tr\{\mathbf{UV}\}} = \int D\mathbf{T} D\mathbf{R} e^{Tr\{\mathbf{RT} - \mathbf{UT} - \mathbf{RV}\}}. \tag{F.3}$$

APPENDIX G

Matrices in Chapter 7

G.1 The Hessian Matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} -\mathrm{Tr}\left(\frac{\gamma^2}{Z_1^2}(\mathbf{I}+\gamma r)^2\right) & -\mathrm{Tr}\left(\frac{\gamma^2}{Z_1^2}(\mathbf{I}+\gamma r)(\Delta+\gamma t)\right) + \frac{\gamma^2}{Z} - 1\Big) & -\mathrm{Tr}\left(\frac{2\gamma^2}{Z_1^2}(\mathbf{I}+\gamma r)(\mathbf{H}_0-\gamma p)\right) & 0 \\ \mathrm{Tr}\left(\frac{\gamma^2}{Z_1^2}(\mathbf{I}+\gamma r)(\Delta+\gamma t)\right) + \frac{\gamma^2}{Z_1} - 1\Big) & -\mathrm{Tr}\left(\frac{\gamma^2}{Z_1^2}(\Delta+\gamma t)^2\right) & -\mathrm{Tr}\left(\frac{2\gamma^2}{Z_1^2}(\Delta+\gamma t)(\mathbf{H}_0-\gamma p)\right) & 0 \\ -\mathrm{Tr}\left(\frac{2\gamma^2}{Z_1^2}(\mathbf{I}+\gamma R)(\mathbf{H}_0-\gamma p)\right) & -\mathrm{Tr}\left(\frac{2\gamma^2}{Z_1^2}(\Delta+\gamma t)(\mathbf{H}_0-\gamma p)\right) & -\mathrm{Tr}\left(\frac{4\gamma^2}{Z_1^2}(\mathbf{H}_0-\gamma p)^2\right) + \frac{2\gamma^2}{Z_1} - 2\Big) & 0 \\ 0 & 0 & 0 & 2N - \mathrm{Tr}\left(\frac{2\gamma^2}{Z_1}\right) \end{bmatrix}$$

where $Z_1 = (\mathbf{F}+\gamma t)\,(\mathbf{I}+\gamma r) + (\mathbf{H}_0 - \gamma p)^2$.

G.2 The Matrix A

$$\mathbf{A} = \begin{bmatrix} \mathbf{F}+\gamma\mathcal{T}_1 & i\mathbf{H}_0 + i\gamma(\mathcal{P}_1+\mathcal{Q}_1) & 0 & 0 \\ i\mathbf{H}_0 + i\gamma(\mathcal{P}_1-\mathcal{Q}_1) & \mathbf{I}+\gamma\mathcal{R}_1 & 0 & 0 \\ 0 & 0 & \boldsymbol{\Gamma}^2+\gamma\mathcal{T}_2 & i\mathbf{H}_0 + i\gamma(\mathcal{P}_2+\mathcal{Q}_2) \\ 0 & 0 & i\mathbf{H}_0 + i\gamma(\mathcal{P}_2-\mathcal{Q}_2) & \mathbf{I}+\gamma\mathcal{R}_2 \end{bmatrix}$$

G.3 The Hessian Matrix for the Covariance

$$\boldsymbol{\Sigma}_{cov} = \begin{bmatrix} -\mathrm{Tr}\left(\frac{\gamma^2}{Z}(F+\gamma t_1)(\mathbf{I}+\gamma r_2)\right) & \mathrm{Tr}\left(\frac{\gamma^2}{Z}(\mathbf{H}_0-\gamma p_1)(\mathbf{H}_0-\gamma p_2)\right) - N & \cdots \\ \mathrm{Tr}\left(\frac{\gamma^2}{Z}(\mathbf{H}_0-\gamma p_1)(\mathbf{H}_0-\gamma p_2)\right) - N & -\mathrm{Tr}\left(\frac{\gamma^2}{Z}(\mathbf{I}_0+\gamma r_1)(\Gamma^2+\gamma t_2)\right) & \cdots \\ \mathrm{Tr}\left(i\frac{\gamma^2}{Z}(\mathbf{I}+\gamma r_2)(\mathbf{H}_0-\gamma p_1)\right) & \mathrm{Tr}\left(i\frac{\gamma^2}{Z}(\mathbf{I}+\gamma r_1)(\mathbf{H}_0+\gamma p_2)\right) & \cdots \\ \mathrm{Tr}\left(i\frac{\gamma^2}{Z}(F+\gamma t_1)(\mathbf{H}_0-\gamma p_2)\right) & \mathrm{Tr}\left(\frac{\gamma^2}{Z}(\Gamma^2+\gamma t_2)(\mathbf{H}_0-\gamma p_1)\right) & \cdots \end{bmatrix}$$

$$\begin{bmatrix} \cdots & \mathrm{Tr}\left(i\frac{\gamma^2}{Z}(\mathbf{I}+\gamma r_2)(\mathbf{H}_0-\gamma p_1)\right) & \mathrm{Tr}\left(i\frac{\gamma^2}{Z}(F+\gamma t_1)(\mathbf{H}_0-\gamma p_2)\right) \\ \cdots & \mathrm{Tr}\left(i\frac{\gamma^2}{Z}(\mathbf{I}+\gamma r_1)(\mathbf{H}_0-\gamma p_2)\right) & \mathrm{Tr}\left(\frac{\gamma^2}{Z}(\Gamma^2+\gamma t_2)(\mathbf{H}_0-\gamma p_1)\right) \\ \cdots & -\mathrm{Tr}\left(\frac{\gamma^2}{Z}(\mathbf{I}+\gamma r_1)(\mathbf{I}+\gamma r_2)\right) & \mathrm{Tr}\left(\frac{\gamma^2}{Z}(\mathbf{H}_0-\gamma p_1)(\mathbf{H}_0-\gamma p_2)\right) - N \\ \cdots\ \mathrm{Tr}\left(\frac{\gamma^2}{Z}(\mathbf{H}_0-\gamma p_1)(\mathbf{H}_0-\gamma p_2)\right) - N & & \mathrm{Tr}\left(\frac{\gamma^2}{Z}(F+\gamma t_1)(\Gamma^2+\gamma t_2)\right) \end{bmatrix}$$

where

$$Z = \left[(F + \gamma t_1)(\mathbf{I} + \gamma r_1) + (\mathbf{H}_0 - \gamma p_1)^2\right] \times \left[(\Gamma^2 + \gamma t_2)(\mathbf{I} + \gamma r_2) + (\mathbf{H}_0 - \gamma p_2)^2\right]$$

and F, Γ are scalars. Due to simplification reasons we assumed that the loss is not frequency selective.

Bibliography

[1] "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2014–2019 White Paper." [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html

[2] Y. Altuğ and A. B. Wagner, "Moderate Deviations in Channel Coding," *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4417–4426, 2014.

[3] G. B. Arous and A. Guionnet, "Large Deviations for Wigner's Law and Voiculescu's non-Commutative Entropy," *Probability Theory and Related Fields*, vol. 108, no. 4, pp. 517–542, 1997.

[4] C. W. Beenakker, "random-Matrix Theory of Quantum Transport," *Reviews of Modern Physics*, vol. 69, no. 3, p. 731, 1997.

[5] C. Berge, *Topological Spaces: Including a Treatment of Multi-Valued Functions, Vector Spaces, and Convexity.* Courier Corporation, 1963.

[6] R. A. Berry and R. G. Gallager, "Communication Over Fading Channels with Delay Constraints," *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1135–1149, 2002.

[7] E. Biglieri, J. Proakis, and S. Shamai, "Fading Channels: Information-Theoretic and Communications Aspects," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2619–2692, 1998.

[8] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004.

[9] G. Caire, G. Taricco, and E. Biglieri, "Optimum Power Control Over Fading Channels," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1468–1489, 1999.

[10] S. Cairns, C. Hass Klau, and P. Goodwin, *Traffic Impact of Highway Capacity Reductions: Assessment of the Evidence.* Landor Publishing, 1998.

[11] E. Candes and T. Tao, "The Dantzig Selector: Statistical Estimation When p Is Much Larger Than n," *The Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, 12 2007. [Online]. Available: http://dx.doi.org/10.1214/009053606000001523

[12] Y. Chen and S. Manning, "Some Eigenvalue Distribution Functions of the Laguerre Ensemble," *Journal of Physics A: Mathematical and General*, vol. 29, no. 23, p. 7561, 1996.

[13] Y. Chen, N. S. Haq, and M. R. McKay, "Random Matrix Models, Double-Time Painlevé Equations, and Wireless Relaying," *Journal of Mathematical Physics*, vol. 54, no. 6, p. 063506, 2013.

[14] Y. Chen and M. R. McKay, "Coulumb Fluid, Painlevé Transcendents, and the Information Theory of MIMO Systems," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4594–4634, 2012.

[15] R. Dar, M. Feder, and M. Shtaif, "The Jacobi MIMO Channel," *IEEE Transactions on Information Theory*, vol. 59, no. 4, pp. 2426–2441, 2013.

[16] D. S. Dean and S. N. Majumdar, "Extreme Value Statistics of Eigenvalues of Gaussian Random Matrices," *Physical Review E*, vol. 77, no. 4, p. 041108, 2008.

[17] M. Debbah, W. Hachem, P. Loubaton, and M. De Courville, "MMSE Analysis of Certain Large Isometric Random Precoded Systems," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1293–1311, 2003.

[18] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer Science & Business Media, 2009, vol. 38.

[19] F. J. Dyson, "Statistical Theory of the Energy Levels of Complex Systems. I," *Journal of Mathematical Physics*, vol. 3, no. 1, pp. 140–156, 1962.

[20] K. Fan, "Minimax Theorems," *Proceedings of the National Academy of Sciences*, vol. 39, no. 1, pp. 42–47, 1953.

[21] J. M. Fini, B. Zhu, T. F. Taunay, and M. F. Yan, "Statistics of Crosstalk in Bent Multicore Fibers," *Opt. Express*, vol. 18, no. 14, pp. 15 122–15 129, Jul 2010.

[22] K. H. Fischer and J. A. Hertz, *Spin Glasses*. Cambridge University Press, 1993, vol. 1.

[23] P. J. Forrester, *Log-Gases and Random Matrices*. Princeton University Press, 2010.

[24] G. J. Foschini and M. J. Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas," *Wireless personal communications*, vol. 6, no. 3, pp. 311–335, 1998.

[25] Y. V. Fyodorov, D. Savin, and H. Sommers, "Scattering, Reflection and Impedance of Waves in Chaotic and Disordered Systems with Absorption," *Journal of Physics A: Mathematical and General*, vol. 38, no. 49, p. 10731, 2005.

[26] R. G. Gallager, *Information Theory and Reliable Communication*. Springer, 1968, vol. 2.

[27] A. Gholami, D. Molin, and P. Sillard, "Compensation of Chromatic Dispersion by Modal Dispersion in MMF- and VCSEL-Based Gigabit Ethernet Transmissions," *Photonics Technology Letters, IEEE*, vol. 21, no. 10, pp. 645–647, May 2009.

[28] D. K. GIFFORD, "Fiber-Optics Test & Measurement: Rayleigh Backscatter Reflectometry Boosts Fiber Characterization Social Media Tools."

[29] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. Academic press, 2014.

[30] P. E. Green, *Fiber Optic Networks*. Prentice-hall, 1993.

[31] P. Gysel and R. K. Staubli, "Statistical Properties of Rayleigh Backscattering in Single-Mode Fibers," *Lightwave Technology, Journal of*, vol. 8, no. 4, pp. 561–567, 1990.

[32] M. Hayashi, "Information Spectrum Approach to Second-Order Coding Rate in Channel Coding," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 4947–4966, 2009.

[33] T. Hayashi, T. Taru, O. Shimakawa, T. Sasaki, and E. Sasaoka, "Ultra-Low-Crosstalk Multi-Core Fiber Realizing Space-Division Multiplexed Ultra-Long-Haul Transmission," in *CLEO: Science and Innovations*. Optical Society of America, 2012, p. CTh4G.3.

[34] F. Hiai and D. Petz, "Eigenvalue Density of the Wishart Matrix and Large Deviations," *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, vol. 1, no. 04, pp. 633–646, 1998.

[35] K.-P. Ho and J. Kahn, "Delay-Spread Distribution for Multimode Fiber with Strong Mode Coupling," *Photonics Technology Letters, IEEE*, vol. 24, no. 21, pp. 1906–1909, Nov 2012.

[36] K.-P. Ho and J. M. Kahn, "Statistics of Group Delays in Multimode Fiber with Strong Mode Coupling," *J. Lightwave Technol.*, vol. 29, no. 21, pp. 3119–3128, Nov 2011.

[37] J. Hoydis, R. Couillet, and P. Piantanida, "The Second-Order Coding Rate of the MIMO Quasi-Static Rayleigh Fading Channel," *Information Theory, IEEE Transactions on*, vol. 61, no. 12, pp. 6591–6622, 2015.

[38] R. C. J. Hsu, A. Tarighat, A. Shah, A. H. Sayed, and B. Jalali, "Capacity Enhancement in Coherent Optical MIMO (COMIMO) Multimode Fiber Links," *IEEE Comm. Letters*, vol. 10, pp. 195–197, 2006.

[39] J. D. Jackson and J. D. Jackson, *Classical Electrodynamics*. Wiley New York etc., 1962, vol. 3.

[40] A. T. James, "Normal Multivariate Analysis and the Orthogonal Group," *The Annals of Mathematical Statistics*, pp. 40–75, 1954.

[41] J. Johnson, R., P. Schniter, T. Endres, J. Behm, D. Brown, and R. Casas, "Blind Equalization Using the Constant Modulus Criterion: A Review," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1927–1950, Oct 1998.

[42] K. Kao and G. A. Hockham, "Dielectric-Fibre Surface Waveguides for Optical Frequencies," in *Proceedings of the Institution of Electrical Engineers*, vol. 113, no. 7. IET, 1966, pp. 1151–1158.

[43] A. Karadimitrakis, A. L. Moustakas, and P. Vivo, "Outage Capacity for the Optical MIMO Channel," *Information Theory, IEEE Transactions on*, vol. 60, no. 7, pp. 4370–4382, 2014.

[44] T. Kato, Y. Koyano, and M. Nishimura, "Temperature Dependence of Chromatic Dispersion in Various Types of Optical Fiber," *Optics Letters*, vol. 25, 2000.

[45] P. Kazakopoulos, P. Mertikopoulos, A. L. Moustakas, and G. Caire, "Living at the Edge: A Large Deviations Approach to the Outage MIMO Capacity," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 1984–2007, 2011.

[46] M. Kiessling and J. Speidel, "Analytical Performance of MIMO MMSE Receivers in Correlated Rayleigh Fading Environments," in *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, vol. 3. IEEE, 2003, pp. 1738–1742.

[47] N. Kikuchi, E. Yamada, Y. Shibata, and H. Ishii, "High-Speed InP-Based Mach-Zehnder Modulator for Advanced Modulation Formats," in *Compound Semiconductor Integrated Circuit Symposium (CSICS), 2012 IEEE*. IEEE, 2012, pp. 1–4.

[48] Lang Tong, Guanghan Xu, and Thomas Kailath, "Blind Identification and Equalization Based on Second-Order Statistics: A Time Domain Approach," *IEEE Trans. Inform. Theory*, vol. 40, pp. 340–349, 1994.

[49] S. Li, M. R. McKay, and Y. Chen, "On the Distribution of MIMO Mutual Information: An In-Depth Painlevé-Based Characterization," *IEEE Transactions on Information Theory*, vol. 59, no. 9, pp. 5271–5296, 2013.

[50] I. G. Macdonald, *Symmetric Functions and Hall Polynomials*. Oxford University Press, 1998.

[51] S. Majumdar, "Random Matrices, the Ulam Problem, Directed Polymers and Growth Models, and Sequence Matching, 2007," *Complex Systems (Les Houches Lecture Notes)*, pp. 179–216.

[52] V. A. Marchenko and L. A. Pastur, "Distribution of Eigenvalues for Some Sets of Random Matrices," *Matematicheskii Sbornik*, vol. 114, no. 4, pp. 507–536, 1967.

[53] T. Martin and R. Landauer, "Wave-Packet Approach to Noise in Multichannel Mesoscopic Systems," *Physical Review B*, vol. 45, no. 4, p. 1742, 1992.

[54] J. Martínez, I. Garcés, A. López, A. Villafranca, and M. Losada, "Analysis of the Influence of Backscattered Optical Power over Bidirectional PON Links," *Optics Communications*, vol. 283, no. 10, pp. 2243–2250, 2010.

[55] M. L. Mehta, *Random Matrices*. Elsevier, 2004, vol. 142.

[56] M. Mezard, G. Parisi, M. A. Virasoro, and D. J. Thouless, "Spin Glass Theory and Beyond," *Physics Today*, vol. 41, p. 109, 1988.

[57] T. Morioka, Y. Awaji, R. Ryf, P. Winzer, D. Richardson, and F. Poletti, "Enhancing Optical Communications with Brand New Fibers," *Communications Magazine, IEEE*, vol. 50, no. 2, pp. s31 –s42, february 2012.

[58] A. L. Moustakas, S. H. Simon, and T. L. Marzetta, "Capacity of Differential Versus Nondifferential Unitary Space–Time Modulation for MIMO Channels," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3622–3634, 2006.

[59] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO Capacity Through Correlated Channels in the Presence of Correlated Interferers and Noise: A (not so) Large N Analysis," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2545–2561, 2003.

[60] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction*. Clarendon Press, 2001, vol. 111.

[61] M. Ohashi, K. Kawazu, A. Nakamura, and Y. Miyoshi, "Simple Backscattered Power Technique for Measuring Crosstalk of Multi-Core Fibers," in *2012 17th Opto-Electronics and Communications Conference*, 2012.

[62] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information Theoretic Considerations for Cellular Mobile Radio," *IEEE transactions on Vehicular Technology*, vol. 43, no. 2, pp. 359–378, 1994.

[63] G. C. M. K. C. Pearson, "Functions of a Complex Variable," 1966.

[64] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel Coding Rate in the Finite Blocklength Regime," *Information Theory, IEEE Transactions on*, vol. 56, no. 5, pp. 2307–2359, 2010.

[65] P. B. Rapajic and D. Popescu, "Information Capacity of a Random Signature Multiple-Input Multiple-Output Channel," *IEEE Transactions on Communications*, vol. 48, no. 8, pp. 1245–1248, 2000.

[66] D. J. Richardson, "Filling the Light Pipe," *Science*, vol. 330, no. 6002, pp. 327–328, 2010. [Online]. Available: http://www.sciencemag.org/content/330/6002/327.short

[67] R. Ryf, M. A. Mestre, S. Randel, X. Palou, A. H. Gnauck, R. Delbue, P. Pupalaikis, A. Sureka, Y. Sun, X. Jiang *et al.*, "Combined SDM and WDM Transmission Over 700-km Few-Mode Fiber," in *Optical Fiber Communication Conference*. Optical Society of America, 2013, pp. OW1I–2.

[68] A. R. Shah, R. C. J. Hsu, A. Tarighat, S. Member, A. H. Sayed, and B. Jalali, "Coherent Optical MIMO (COMIMO)," *IEEE J. Lightwave Tech*, vol. 23, pp. 2410–2419, 2005.

[69] S. Shamai and I. Sason, "Variations on the Gallager Bounds, Connections, and Applications," *IEEE Transactions on Information Theory*, vol. 48, no. 12, pp. 3029–3051, 2002.

[70] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels," *Information and Control*, vol. 10, no. 1, pp. 65–103, 1967.

[71] C. E. Shannon, "Communication in the Presence of Noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[72] H. Shin and M. Z. Win, "Gallager's Exponent for MIMO Channels: A Reliability-Rate Tradeoff," *Communications, IEEE Transactions on*, vol. 57, no. 4, pp. 972–985, 2009.

[73] S. H. Simon and A. L. Moustakas, "Crossover from Conserving to Lossy Transport in Circular Random-Matrix Ensembles," *Physical review letters*, vol. 96, no. 13, p. 136805, 2006.

[74] M. Sion, "On General Minimax Theorems," *Pacific Journal of Mathematics*, vol. 8, no. 1, pp. 171–176, 1958.

[75] K. Takenaga, Y. Arakawa, S. Tanigawa, N. Guan, S. Matsuo, K. Saitoh, and M. Koshiba, "An Investigation on Crosstalk in Multi-Core Fibers by Introducing Random Fluctuation Along Longitudinal Direction." *IEICE Transactions*, vol. 94-B, no. 2, pp. 409–416, 2011.

[76] A. Tarighat, R. Hsu, A. Shah, A. Sayed, and B. Jalali, "Fundamentals and Challenges of Optical Multiple-Input Multiple-Output Multimode Fiber Links [Topics in Optical Communications]," *Communications Magazine, IEEE*, vol. 45, no. 5, pp. 57 –63, may 2007.

[77] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-Time Block Codes from Orthogonal Designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.

[78] I. E. Telatar *et al.*, "Capacity of Multi-Antenna Gaussian Channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[79] H. Touchette, "The Large Deviation Approach to Statistical Mechanics," *Physics Reports*, vol. 478, no. 1, pp. 1–69, 2009.

[80] F. Tricomi, "Integral Equations, ser," *Pure Appl. Math V. London: Interscience*, 1957.

[81] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Now Publishers Inc, 2004, vol. 1.

[82] S. S. Varadhan, "Asymptotic Probabilities and Differential Equations," *Communications on Pure and Applied Mathematics*, vol. 19, no. 3, pp. 261–286, 1966.

[83] P. Vivo, S. N. Majumdar, and O. Bohigas, "Large Deviations of the Maximum Eigenvalue in Wishart Random Matrices," *Journal of Physics A: Mathematical and Theoretical*, vol. 40, no. 16, p. 4317, 2007.

[84] ——, "Distributions of Conductance and Shot Noise and Associated Phase Transitions," *Physical Review Letters*, vol. 101, no. 21, p. 216809, 2008.

[85] Z. Wang and G. B. Giannakis, "Outage Mutual Information of Space-Time MIMO Channels," *IEEE Transactions on Information Theory*, vol. 50, no. 4, pp. 657–662, 2004.

[86] P. J. Winzer and G. J. Foschini, "MIMO Capacities and Outage Probabilities in Spatially Multiplexed Optical Transport Systems," *Optics express*, vol. 19, no. 17, pp. 16 680–16 696, 2011.

[87] J. Wishart, "The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population," *Biometrika*, pp. 32–52, 1928.

[88] J. Xue, C. Zhong, and T. Ratnarajah, "Performance Analysis of Orthogonal STBC in Generalized-k Fading MIMO Channels," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 3, pp. 1473–1479, 2012.

[89] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Block-Fading Channels at Finite Blocklength," in *Proceedings of the Tenth International Symposium on Wireless Communication Systems (ISWCS 2013)*. VDE, 2013, pp. 1–4.

[90] J. Zhang, M. Matthaiou, G. K. Karagiannidis, H. Wang, and Z. Tan, "Gallager's Exponent Analysis of STBC MIMO Systems Over η-μ and κ-μ Fading Channels," *IEEE Transactions on Communications*, vol. 61, no. 3, pp. 1028–1039, 2013.

[91] L. Zheng and D. N. C. Tse, "Diversity and Multiplexing: A Fundamental Tradeoff in Multiple-Antenna Channels," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073–1096, 2003.

[92] B. Zhu, T. F. Taunay, M. F. Yan, J. M. Fini, and M. Fishteyn, "Seven-Core Multicore Fiber Transmissions for Passive Optical Network," *Opt. Express*, vol. 18, no. 11, pp. 11 117–11 122, May 2010.

List of Figures