

# Nonlinear Interference in Fiber-Optic Communications

by

Ronen Dar

THESIS SUBMITTED TO THE SENATE OF TEL-AVIV UNIVERSITY  
in partial fulfillment of the requirements for the degree of  
“DOCTOR OF PHILOSOPHY”

April, 2015



# Nonlinear Interference in Fiber-Optic Communications

by

Ronen Dar

THESIS SUBMITTED TO THE SENATE OF TEL-AVIV UNIVERSITY  
in partial fulfillment of the requirements for the degree of  
“DOCTOR OF PHILOSOPHY”

Under the Supervision of Prof. Meir Feder and Prof. Mark Shtaif

April, 2015



This dissertation is submitted as a collection of the peer-reviewed journal papers that I published during my PhD studies.

The work was carried out under the supervision of  
**Prof. Meir Feder and Prof. Mark Shtaiif**



*To my advisors, **Meir** & **Mark**.* I feel immensely fortunate to have had the opportunity to work with you. Your leadership, charisma and pursuit for excellence have inspired me countless times. Words can not express the admiration and gratitude I feel toward you. Thank you.

*To my “third advisor”, **Antonio**.* Your wisdom, knowledge, and enthusiasm for research have tremendously enriched me. Thank you for all the interesting discussions, long email correspondences and fruitful notes exchanging. Working with you was an amazing experience.

*To my parents, **Susan** & **Yigal**.* A thousands thanks for your endless support, guidance, care and love. Huge credit for this accomplishment goes to you.

*To my beautiful wife, **Yifat**.* I feel blessed to have you by my side. Thank you for your advice and devoted support. This achievement could not have been accomplished without your help.

Finally and most importantly, this thesis is dedicated to my biggest love - **Ella**.

# Abstract

The nonlinearity of optical fibers is arguably the most significant factor limiting the capacity of modern wavelength division multiplexed (WDM) communications systems. The Kerr nonlinear effect of the fiber introduces complex distortions into the transmitted optical waveforms, generating nonlinear interference noise (NLIN) between the various WDM channels; this diminishes the receiver's capability of decoding the transmitted data and limits the reach of the system and the achievable data-rates. The modeling of NLIN and the understanding of its impact on system performance are therefore key components in the efficient design of fiber-optic communication systems and in the evaluation of their ultimate reach and capacity.

In this dissertation we formulate a discrete-time model for the NLIN in an arbitrarily link. We use the model to study the accurate extraction of the NLIN variance, its dependence on modulation format, the role of nonlinear phase-noise and polarization rotation, and the existence of temporal correlations. By taking advantage of these results, we develop new lower bounds on the channel capacity and further propose new practical schemes for mitigating the NLIN by exploiting its temporal correlations. In addition, we present a pulse-collision theory that provides qualitative and quantitative insight into the build-up of NLIN, offering a simple and intuitive explanation to many of the reported and previously unexplained phenomena.

In this thesis we also address the topic of space-division multiplexing (SDM) in fiber-optic communications. We examine the case in which the number of optical paths that are addressed by the transmitter and receiver is allowed to be smaller than the total number of paths supported by the link. We calculate the ergodic capacity, outage probability, and diversity-multiplexing tradeoff as a function of the number of addressed paths.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	The under-addressed MIMO channel . . . . .	2
1.2	Nonlinear interference noise in fiber-optic communications . . . . .	3
<b>2</b>	<b>The Under Addressed Optical MIMO Channel: Capacity and Outage</b>	<b>7</b>
2.1	Abstract . . . . .	7
2.2	Introduction and results . . . . .	7
2.3	Conclusions . . . . .	11
<b>3</b>	<b>The Jacobi MIMO Channel</b>	<b>15</b>
3.1	abstract . . . . .	15
3.2	Introduction . . . . .	15
3.3	System Model and Channel Statistics . . . . .	17
3.3.1	Case I - $m_t + m_r \leq m$ . . . . .	18
3.3.2	Case II - $m_t + m_r > m$ . . . . .	19
3.4	The Ergodic Case . . . . .	20
3.4.1	Case I - $m_t + m_r \leq m$ . . . . .	20
3.4.2	Case II - $m_t + m_r > m$ . . . . .	21
3.5	The Non-Ergodic Case . . . . .	22
3.5.1	Case I - $m_t + m_r \leq m$ . . . . .	23
3.5.2	Case II - $m_t + m_r > m$ . . . . .	24
3.6	Achieving The No-Outage Promise . . . . .	25
3.7	Diversity Multiplexing Tradeoff . . . . .	28
3.7.1	Case I - $m_t + m_r \leq m$ . . . . .	31
3.7.2	Case II - $m_t + m_r > m$ . . . . .	33
3.8	Relation To The Rayleigh Model . . . . .	35
3.9	Discussion . . . . .	37
3.10	Appendix: Proof of Theorem 1 . . . . .	38
3.11	Appendix: Proof of Theorem 4 . . . . .	40
<b>4</b>	<b>Properties of Nonlinear Noise in Long, Dispersion-Uncompensated Fiber Links</b>	<b>47</b>
4.1	Abstract . . . . .	47
4.2	Introduction . . . . .	47
4.3	Time-domain analysis . . . . .	49

4.4	Frequency domain analysis . . . . .	53
4.5	Numerical validation . . . . .	56
4.5.1	Modulation format dependence . . . . .	56
4.5.2	The variance of phase-noise and assessment of the residual NLIN . . . .	59
4.5.3	The difference with respect to the NLIN power predicted by the GN model . . . . .	60
4.6	Discussion . . . . .	62
<b>5</b>	<b>New Bounds on the Capacity of the Nonlinear Fiber-Optic Channel</b>	<b>67</b>
5.1	Abstract . . . . .	67
5.2	Introduction and results . . . . .	67
5.3	Conclusions . . . . .	73
<b>6</b>	<b>Accumulation of Nonlinear Interference Noise in Fiber-Optic Systems</b>	<b>77</b>
6.1	Abstract . . . . .	77
6.2	Introduction . . . . .	77
6.3	Theoretical background . . . . .	79
6.4	Results . . . . .	79
6.5	Discussion . . . . .	83
6.6	Conclusions . . . . .	85
6.7	Appendix: Computation of $\chi_1$ and $\chi_2$ . . . . .	86
<b>7</b>	<b>Inter-Channel Nonlinear Interference Noise in WDM Systems: Modeling and Mitigation</b>	<b>93</b>
7.1	Abstract . . . . .	93
7.2	Introduction . . . . .	93
7.3	NLIN modeling . . . . .	96
7.4	Modulation-format dependence and the importance of nonlinear phase-noise	99
7.5	Time-varying ISI channel . . . . .	103
7.5.1	Channel model . . . . .	103
7.5.2	NLIN mitigation - simulation results . . . . .	105
7.6	Conclusions . . . . .	107
7.7	Appendix: Time and frequency-domain representations of $\chi_1$ and $\chi_2$ . . . . .	108
<b>8</b>	<b>Pulse Collision Picture of Inter-Channel Nonlinear Interference in Fiber- Optic Communications</b>	<b>113</b>
8.1	Abstract . . . . .	113
8.2	Introduction . . . . .	114
8.3	Time-domain theory . . . . .	116
8.4	Two-pulse collisions . . . . .	118
8.5	Three-pulse collisions . . . . .	119
8.6	Four-pulse collisions . . . . .	119
8.7	A few numerical examples . . . . .	120
8.8	The effect of polarization-multiplexing . . . . .	123
8.9	System implications of collision classification . . . . .	126

8.10	Relation to the Gaussian noise model and the role of chromatic dispersion . . .	132
8.11	Conclusions . . . . .	133
8.12	Appendix: The dependence on $\Omega$ in complete versus incomplete collisions . .	133
8.12.1	Complete collisions . . . . .	135
8.12.2	Incomplete collisions . . . . .	136
8.13	Appendix: Relative rotation between the modulated polarization axes of the interacting WDM channels . . . . .	136
<b>9</b>	<b>Shaping in the Nonlinear Fiber-Optic Channel</b>	<b>141</b>
9.1	Channel model . . . . .	141
9.2	Shaping gain in the fiber-optic channel . . . . .	143
9.2.1	$2N$ -dimensional cube . . . . .	144
9.2.2	$2N$ -dimensional ball . . . . .	145
9.2.3	Shaping gain . . . . .	146
9.3	Simulation results . . . . .	148
<b>10</b>	<b>Summary and Conclusions</b>	<b>151</b>



# List of Figures

1.1	Achievable data-rates versus signal power . . . . .	2
2.1	Ergodic capacity versus $\rho^2$ . . . . .	9
2.2	Outage probability versus normalized rate . . . . .	11
3.1	Ergodic capacity versus $\rho$ . . . . .	21
3.2	$\rho$ versus the normalized number of received modes . . . . .	24
3.3	Outage probability versus normalized rate . . . . .	25
3.4	The average error probability of a repetition scheme versus $\rho$ . . . . .	30
3.5	Optimal DMT curves for $m_t = m_r = 4$ . . . . .	32
3.6	A comparison between the Alamouti scheme and the repetition scheme . . .	34
3.7	Comparing the Rayleigh and Jacobi models . . . . .	36
4.1	Constellation diagrams for inter-channel NLIN . . . . .	57
4.2	The electric field intensities of a single channel operating with Nyquist sinc-shaped pulses at a baud-rate of 100 GHz . . . . .	58
4.3	The autocorrelation of nonlinear phase-noise . . . . .	59
4.4	The total NLIN variance and phase-noise variance . . . . .	60
4.5	NLIN power versus the average power per-channel . . . . .	61
5.1	The variance of the effective noise versus block-size . . . . .	70
5.2	Capacity lower bound versus linear SNR and the maximum achievable transmission distance as a function of spectral efficiency . . . . .	71
5.3	Noise-to-signal ratio versus average input power . . . . .	72
6.1	NLIN power versus the average power per-channel . . . . .	80
6.2	Constellation diagrams for inter-channel NLIN . . . . .	81
6.3	Accumulation of the NLIN power with the number of spans . . . . .	82
6.4	The ratio between the FON and SON coefficients . . . . .	84
6.5	The effect of pre-dispersion . . . . .	85
7.1	An illustration of SNR versus power . . . . .	94
7.2	The ratio between the intra-channel and inter-channel contributions to the NLIN power . . . . .	95
7.3	Constellation diagrams for inter-channel NLIN in polarization-multiplexed transmission . . . . .	99
7.4	NLIN power versus number of spans . . . . .	100

7.5	The relative weight of complete collisions in a polarization-multiplexed system	101
7.6	Theoretical autocorrelation and variance of $\mathbf{H}_{l,1}^{(n)}$	104
7.7	Autocorrelation and variance of $\mathbf{H}_{0,1}^{(n)}$ as extracted from simulations	106
7.8	Effective SNR and BER as a function of input power after NLIN mitigation	107
8.1	Generic four-pulse interference	116
8.2	Evolution of the NLIN coefficients along the fiber axis	120
8.3	The magnitude of the nonlinear perturbation $\gamma X_{h,k,m} $ as a function of frequency separation	121
8.4	The magnitude of the nonlinear perturbation $\gamma X_{h,k,m} $ as a function of $m$	122
8.5	The relative contribution of the various pulse-collisions versus number of spans	127
8.6	The ratio between unitary noise and circular noise and the ratio between phase-noise and polarization-rotation versus number of spans	129
8.7	The contributions of phase-noise, polarization-rotation, and circular noise to the NLIN power, as a function of the modulation format factor	131
9.1	Auto-correlation of the AMIN term	143
9.2	Linear and nonlinear shaping gains	147
9.3	NLIN variance versus average input power, spectral efficiency versus average input power, and shaping gain versus number dimensions	148

# List of Acronyms

ACF	Auto-Correlation Function
AMIN	Amplitude Modulation Induced Noise
ASE	Amplified Spontaneous Emission
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BPSK	Binary Phase-Shift Keying
COE	Circular Orthogonal Ensemble
CSI	Channel State Information
CUE	Circular Unitary Ensemble
DSP	Digital Signal Processing
DMT	Diversity Multiplexing Tradeoff
EDFA	Erbium-Doped Fiber Amplifier
FON	Fourth Order Noise
FWM	Four-Wave-Mixing
GN	Gaussian Noise
i.i.d.	independent identically distributed
ISI	Inter Symbol Interference
JOE	Jacobi Orthogonal Ensemble
JUE	Jacobi Unitary Ensemble
LMS	Least Mean Squares
MIMO	Multiple-Input Multiple-Output
ML	Maximum Likelihood
NLIN	Non-Linear Interference Noise
NSR	Noise-to-Signal Ratio
pdf	probability density function
PSK	Phase-Shift Keying
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase-Shift Keying
RLS	Recursive Least Squares
SDM	Space Division Multiplexing

SISO	Single-Input Single-Output
SMF	Single-Mode Fiber
SNR	Signal-to-Noise Ratio
SON	Second Order Noise
SPM	Self-Phase-Modulation
WDM	Wavelength Division Multiplexing
XPM	Cross-Phase-Modulation



# Chapter 1

## Introduction

The Kerr nonlinear effect in optical fibers has long been recognized as the most important factor limiting the capacity of fiber-optic communication systems [1, 2]. It produces complicated distortions of the optical waveform, essentially diminishing the receiver's capability of decoding the transmitted information. A convenient way of demonstrating the fundamental limitation imposed by fiber nonlinearity is by examining the dependence of the achievable data-rates on the optical power that is launched into the fiber. Fig. 1.1 depicts a characteristic curve that describes this dependence in modern optical communication systems [1]. At low power levels, the nonlinearity is negligible and the achievable data-rate improves with the optical power (as expected from the increase in the SNR). When the power level is increased further, fiber nonlinearities become significant and the achievable data-rate saturates or reduces to zero. The maximal data-rate is therefore obtained at a power level in which the nonlinear distortions are significant, emphasizing the need of commercial systems to work in the nonlinear regime.

Communicating through a nonlinear channel raises many practical and theoretical problems. Paradigms and basic communication methods which are known from linear channels have to be re-examined. This includes basic questions such as how to detect the signal at the output of the link (linear filtering?), how to modulate the signal (is Gaussian modulation the optimal input distribution?), what is the impact of the nonlinear distortions on system performance, and more. In addition, information theoretical problems such as estimating the channel capacity become much more challenging and cumbersome when considering a nonlinear channel [3, 4]. The study of these questions necessitates an accurate modeling of the nonlinear phenomena [5], a task in which a number of groups have recently engaged [6–8]. This dissertation considers the modeling problem while setting a target of understanding the impact of nonlinearity on the achievable data-rates.

Another topic which is addressed in this dissertation is the implementation of space-division multiplexing (SDM) in fiber-optic communications [10, 11]. While the demand for world-wide data traffic grows at a constant rate of approximately 60% per year, the capacity growth rate of commercial fiber-optic communication systems has experienced a dramatic slow-down in recent years [12]. This slow-down is explained by latest studies on the channel capacity, showing that it is being rapidly approached by the data-rates of commercial systems [13]. One of the most promising approaches for dealing with the imminent “capacity crunch” is the implementation of SDM in multi-mode or multi-core optical fibers. In this context our

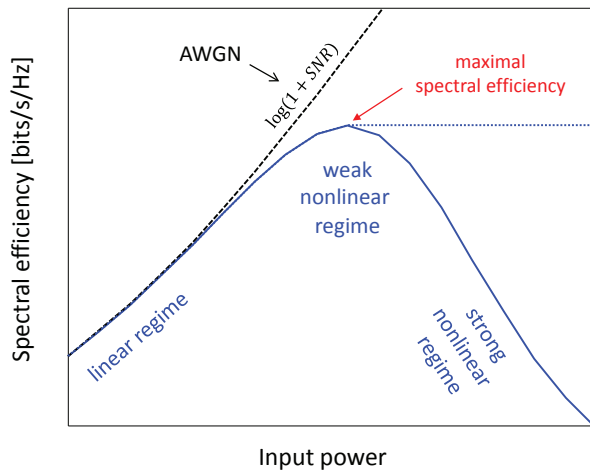


Figure 1.1: Achievable data-rates versus signal power. The solid (blue) line shows a characteristic shape that describes most of the systems. Notice that with simple information theoretical arguments it can be shown that the capacity is always a non-decreasing function of the signal power [9] (as is indicated by the dotted (blue) line). Dashed (black) curve represents the capacity of a linear additive white Gaussian noise (AWGN) channel, where SNR is the signal-to-noise ratio.

focus is on the under-addressed multiple-input and multiple output (MIMO) channel model which was proposed by Winzer and Foschini in [14].

The dissertation is organized as follows. Chapters 2-8 contain the peer-reviewed journal papers that were accepted for publication in the process of my PhD research. They are presented according to the chronological order of their publication, representing the chronological progress of my research and the order in which the results were obtained. It should be noted that I had a major role in the development of all results presented in these chapters. The papers that appear in Chapters 2-3 present our results on the under-addressed SDM channel. The papers that appear in Chapters 4-8 summarize our results on nonlinear interference noise in fiber-optic communications. Chapter 9 discusses some results that were obtained on shaping gains in the nonlinear regime of the fiber-optic channel. These results were presented at the IEEE International Symposium on Information Theory (ISIT 2014) [15], but they were not yet submitted to a publication in a journal. A more detailed introduction for Chapters 2-3 and for Chapters 4-9 is provided in Secs. 1.1 and 1.2, respectively. Chapter 10 is devoted to summary and conclusions. We note that Chapters 2-8 contain their own list of references as each one of them is a copy of a published paper. The reference list for Chapters 1, 9 and 10 appears at the end of the dissertation.

## 1.1 The under-addressed MIMO channel

In Chapters 2-3 we consider the problem of SDM in fiber-optic communications [14]. With this technology, information is transmitted simultaneously over multiple spatial modes of the optical fiber, while relying on MIMO signal processing algorithms to resolve issues related to mode coupling in the process of propagation. With today's technology, the number of

fiber modes that can be effectively supported for the transmission of information is limited almost exclusively by the complexity of the MIMO algorithms and by the speed of the signal-processing hardware. As these technologies continuously improve with time, one may consider deploying multi-mode fibers admitting a larger number of spatial modes than can be processed today, with the intention of harvesting the full capacity of the fiber in the future. Such a solution, initially proposed and considered by Winzer and Foschini in [14], does not come without a price. Part of the transmitted signal energy couples into fiber modes that are not detected at the receiver, thereby resulting in the reduction of the achievable capacity.

The prime goal of the paper in Chapter 2 is to analyze the performance of the under-addressed channel with respect to the number of modes that are addressed by the transmitter and receiver. In the absence of sufficient experimental characterization, we adopt the assumptions of [14] and model the coupling between the modes as unitary. In this case the transfer matrix is modeled as a block of size  $m_r \times m_t$  within a uniformly drawn  $m \times m$  unitary matrix, where  $m_t$  and  $m_r$  represent the number of modes that are addressed by the transmitter and receiver, respectively, and  $m$  represents the total number of available modes. By establishing the relation between the channel matrix and the Jacobi ensemble of random matrices [16–18], we find the distribution of the eigenvalues of the channel and derive explicit expressions to the channel capacity in the ergodic regime and to the outage probability in the non-ergodic regime. An interesting outcome of the derivation is that zero outage capacity can be obtained even when not all of the modes are detected by the receiver.

The paper which appears in Chapter 3 expands the results of Chapter 2 and further analyzes the new channel model from a more information theoretical point of view. We review the calculations of the ergodic capacity and outage probability in a more detailed manner and supplement them with the calculation of the diversity-multiplexing tradeoff of the channel. We further provide a simple communication scheme that uses a channel-state feedback to obtain the zero-outage promise. In addition, we discuss the possibility of using the Jacobi channel as a model for fading MIMO channels where the size of the unitary matrix  $m$  defines the “*fading measure*” of the channel, providing a way to model the statistical structure of the power loss. For example, when  $m$  is equal to  $m_r$ , the transfer matrix is simply composed of orthonormal columns: its elements (i.e., the path gains) are highly dependent and there is no randomness in the received power. As  $m$  becomes greater, the orthogonality of the columns and rows of the transfer matrix fades, the dependency between the path gains becomes weaker and the power loss in the unaddressed receive outputs increases. Indeed, when  $m$  is very large with respect to  $m_t$  and  $m_r$ , with proper normalization that compensates for the average power loss, the Jacobi fading model approaches to the Rayleigh model.

## 1.2 Nonlinear interference noise in fiber-optic communications

In a wavelength division multiplexed (WDM) environment, where several channels are multiplexed into the fiber, nonlinear propagation phenomena can be classified as either intra-channel [19], or inter-channel [20] effects. Intra-channel distortions are those affecting an individual WDM channel. Inter-channel distortions are those in which a given WDM chan-

nel is affected by the electric fields of the neighboring WDM channels that simultaneously propagate in the fiber. In principle, intra-channel distortions can be mitigated by means of digital back propagation at the receiver [21], or pre-distortion at the transmitter [22]. However, similar elimination of inter-channel distortions is considered impractical. The difficulty is not only in the need of coherently detecting and jointly processing multiple WDM channels, but also because of the unpredictable add-drop operations taking place in modern optical networks, and the fact that the polarization states of different WDM channels evolve randomly and differently in the presence of polarization mode dispersion [23]. Inter-channel interference is therefore customarily treated as noise [2], and hence we refer to it in what follows as *nonlinear interference noise*, or NLIN [24–30]. In this dissertation we focus on inter-channel nonlinear effects as they come into play in modern, dispersion uncompensated fiber links with coherent optical detection. We additionally assume that the effect of non-linearity on the ASE noise, and its nonlinear interaction with the signal are negligible. This assumption is safely satisfied in most systems of practical relevance as well as in all system configurations considered throughout the dissertation [2].

Two approaches to the analytical characterization of NLIN have so far been reported in the literature. The first approach, which relies on analysis in the spectral domain, originated from the group of P. Poggiolini at the Politecnico di Torino [6, 31–33] and its derivation has been recently generalized by Johannisson and Karlsson [34] and by Bononi and Serena [7]. The model generated by this approach is commonly referred to as the Gaussian noise (GN) model and its implications have already been addressed in a number of studies [35–37]. The second approach has been proposed by Mecozzi and Essiambre [38], and it is based on a time-domain analysis<sup>1</sup>. The results of the latter approach [38] are different from those of the former [6, 7, 34]. Most conspicuously, in the results of [6, 7, 34], the NLIN is treated as additive Gaussian noise and its power-spectrum is totally independent of modulation format, whereas the theory of Mecozzi et al. predicts a strong dependence of the NLIN variance on the modulation format, as well as the fact that in the presence of non-negligible intensity modulation a large fraction of NLIN can be characterized as phase-noise.

In the paper presented in Chapter 4 we review the essential parts of the time-domain theory of [38], as well as those of the frequency domain GN approach. We show that the assumption of statistical independence between non overlapping frequency components of the propagating electric field is responsible for the fact that the NLIN in Refs. [6, 7, 34] appears to be independent of modulation format. We supplement the NLIN variance obtained in the frequency domain analysis [31] with an extra term that follows from fourth-order frequency correlations. This work has been followed upon by the authors of [6], and a new enhanced GN model that relies on the findings of our analysis has been published subsequently in [41]. We believe that the results of Chapter 4 and the followup work of [41] settle the discrepancy between the frequency domain approach [6, 7, 34] and the time-domain theory of [38]<sup>2</sup>.

In the paper presented in Chapter 5, we revisit the problem of estimating the nonlinear channel capacity. A common feature of capacity estimates published so far [1, 2] is that they treat the intra-channel nonlinear interference as a cancelable noise while treating the

---

<sup>1</sup>Additional time-domain models for inter-channel NLIN were developed in Refs. [8, 39] and for intra-channel NLIN in Ref. [40].

<sup>2</sup>An additional frequency-domain analysis has been recently published [42], showing similar dependence of NLIN spectrum on the high order moments of the transmitted signal.

inter-channel nonlinear interference as additive, white noise which is independent of the data transmitted on the channel of interest. In reality, in the presence of chromatic dispersion, different WDM channels propagate at different velocities so that every symbol in the channel of interest interacts with multiple symbols of every interfering channel. Consequently, adjacent symbols in the channel of interest are disturbed by essentially the same collection of interfering pulses and therefore they are affected by nonlinearity in a highly correlated manner. In addition, as demonstrated in Chapter 4, one of the most pronounced manifestations of nonlinearity is in the form of phase-noise. We show that by taking advantage of the fact that a large fraction of the nonlinear interference is in the form of phase-noise, and by accounting for the long temporal correlations of this noise, the capacity of the fiber-optic channel is notably higher than what is currently assumed. This advantage is translated into the doubling of the link distance for a fixed transmission rate. The impact of these results on practical systems is further examined in Chapter 7.

We follow-up the work of Chapter 4 and further explore the statistical properties of NLIN in Chapter 6. We examine the significance of nonlinear phase-noise and the importance of the modulation format dependence of NLIN variance with respect to system parameters such as the link-length, span-length, amplification scheme, and more. We show that the significance of phase-noise and the modulation format dependence are largest in the case of a single amplified span, or in a system of arbitrary length that uses distributed amplification; their distinctness reduces somewhat in multi-span systems with lumped amplification and with a span-length much larger than the fibers effective length. In order to facilitate future research of this problem, we provide a Matlab program that implements a computationally efficient algorithm for computing the overall NLIN variance. This Matlab program was further developed into an online web application that is available at <http://nlinwizard.eng.tau.ac.il/>.

In the first part of the invited paper presented in Chapter 7 we extend the work reported in Chapters 4 and 6 and generalize the time-domain model to the case of polarization-multiplexed transmission. We characterize the nature of NLIN in this case and explore the dependence of the relevant NLIN features on the parameters of the system. In the second part of the paper we show that the effect of NLIN is equivalent to a linear time-varying inter-symbol-interference (ISI) (where the zeroth-order ISI coefficient manifests itself as phase-noise, as was discussed in Chapter 4-6). We propose a nonlinear interference mitigation scheme that exploit the temporal correlations of the ISI terms to track, estimate and cancel their effect by using linear fast-adaptive equalization. We evaluate the potential benefit of such schemes and discuss the implications on system design.

In the paper of Chapter 8 we model the build-up of NLIN by considering the pulse-collision dynamics in the time domain. We show that the fundamental interactions can be classified as two-pulse, three-pulse, or four-pulse collisions and they can be either complete, or incomplete. Each type of collision is shown to have its unique signature and the overall nature of NLIN is determined by the relative importance of the various classes of pulse collisions in a given WDM system. The most important contributions to NLIN follow from two-pulse and four-pulse collisions. While the contribution of two-pulse collisions is in the form of phase-noise and polarization-state-rotation with strong dependence on modulation format, four-pulse collisions generate complex circular noise whose variance is independent of modulation format. In addition, two-pulse collisions are strongest when the collision is complete, whereas

four-pulse collisions are strongest when the collision is incomplete. We show that two-pulse collisions dominate the formation of NLIN in short links with lumped amplification, or in links with distributed amplification extending over arbitrary length. In long links using lumped amplification the relative significance of four-pulse collisions increases, emphasizing the circularity of the NLIN while reducing its dependence on modulation format.

Finally, in Chapter 9 we examine methods for suppressing NLIN by shaping the input constellation. We show that by using a ball shaped input the nonlinear penalty can be significantly reduced. The potential shaping gain in the nonlinear fiber-optic channel is therefore not only in reducing the average transmission power, as in linear AWGN channels [43,44], but also in suppressing the affect of nonlinearity. We show that the maximum shaping gain in certain fiber-optic scenarios can be more than 1.53dB, which is the ultimate shaping gain in linear AWGN channels. Furthermore, while the “linear” shaping gain monotonically increases with the number of dimensions of the shaping region, the maximum shaping gain from suppressing the NLIN is obtained with a finite-dimensional ball shaped input. The results of this chapter also relate to the problem of finding the optimum input distribution which maximizes the channel mutual information, showing that Gaussian distribution may be sub-optimal in the nonlinear optical channel.

We note that our analysis throughout this dissertation focuses on inter-channel nonlinearities resulted from XPM only while ignoring FWM contributions. As is evident from simulations [24, 41], the contribution of FWM to NLIN is of negligible importance in all scenarios examined throughout this dissertation as well as in most of current commercial systems. We note however that FWM contributions may become important in certain systems employing multi-subcarrier transmission [45] or operating over very low dispersion fibers [24]. The significance of FWM in these cases can be further investigated by using the theoretical predictions of the NLIN Wizard.



# Chapter 2

## The Under Addressed Optical MIMO Channel: Capacity and Outage

Ronen Dar, Meir Feder and Mark Shtaif, “*The under-addressed optical multiple-input, multiple-output channel: capacity and outage*,” Optics letters, vol. 37, pp. 3150–3152 (August 2012)

### 2.1 Abstract

We study an optical space-division multiplexed system where the number of modes that are addressed by the transmitter and receiver is allowed to be smaller than the total number of optical modes supported by the fiber. This situation will be of relevance if, for instance, fibers supporting more modes than can be processed with current MIMO technology are deployed with the purpose of future-proof installation. We calculate the ergodic capacity and the outage probability of the link and study their dependence on the number of addressed modes at the transmitter and receiver.

### 2.2 Introduction and results

One of the most intensely explored approaches for dealing with the imminent capacity crunch of optical communications systems [1] is the implementation of space division multiplexing (SDM) in multi-mode optical fibers [2–6]. With this approach, information is transmitted simultaneously over multiple spatial modes of the optical fiber, while relying on multiple-input and multiple-output (MIMO) signal processing algorithms to resolve issues related to mode coupling in the process of propagation. With today’s technology, the number of fiber modes that can be effectively supported for the transmission of information is limited almost exclusively by the complexity of the MIMO algorithms and by the speed of signal-processing hardware. As these technologies continuously improve with time, one may consider deploying multi-mode fibers admitting a larger number of spatial modes than can be processed today, with the intention of harvesting the full capacity of the fiber in the future. Such a solution, initially proposed and considered by Winzer and Foschini in [3], does not come without a price. Part of the transmitted signal energy couples into fiber modes that are not detected

at the receiver, thereby resulting in the reduction of achievable capacity. In what follows, we refer to the multi-mode fiber-optic channel in which not all supported modes are coupled to transmitters or receivers, as the under-addressed MIMO channel. Analysis of its performance is the prime goal of this work.

We consider a system using a total of  $m$  scalar modes (counting both spatial modes and polarizations) and where the number of modes addressed by the transmitter and receiver are  $m_t$  and  $m_r$ , respectively. We explore two distinct regimes of operation, referred to as the ergodic and the non-ergodic regimes [7]. In the ergodic regime, a single frame of the error-correcting code samples the entire channel statistics, whereas in the non-ergodic regime, the channel within each code-frame is assumed to be constant. The ergodic regime can be relevant if schemes actively randomizing fiber mode-coupling on a time-scale much shorter than the error-correcting code-frame are introduced. Alternatively, in the fiber-optical scenario, if coding is implemented in the spectral domain, the ergodic assumption applies when the channel decorrelates quickly in frequency. A situation that characterizes systems with high modal dispersion [5]. In the ergodic regime, performance is evaluated in terms of the *ergodic capacity*, which is the channel capacity averaged with respect to all channel realizations. In the non-ergodic regime, performance is characterized in terms of the system *outage probability*. We derive these quantities analytically and present their dependence on  $m$ ,  $m_t$  and  $m_r$ . A particularly interesting outcome of our study is that when  $m_t + m_r > m$ , a throughput equivalent to  $m_t + m_r - m$  decoupled single-mode channels can be guaranteed. In the fast-changing channel regime this implies that the ergodic capacity is never smaller than  $(m_t + m_r - m)$  single-input single-output (SISO) channels, whereas in the non-ergodic regime a throughput equivalent to  $(m_t + m_r - m)$  SISO channels can be achieved with zero outage probability.

In the absence of sufficient experimental characterization, we adopt the description of the multi-mode fiber as a unitary system with strong mode-coupling, as was used in most previous studies [3–6] of the multi-mode transmission problem. By doing so, we ignore the effects of mode-dependent losses and justify the description of the overall  $m \times m$  transfer matrix  $\mathbf{H}$  as a random instantiation drawn uniformly from the ensemble of all  $m \times m$  unitary matrices (Haar distributed). In addition, as was done in [3], we assume that the average power generated by each of the  $m_t$  transmitters is constant, regardless of the value of  $m_t$ . Under these conditions the channel can be described as

$$\underline{\mathbf{y}} = \rho \mathbf{H}_{11} \underline{\mathbf{x}} + \underline{\mathbf{z}}, \quad (2.1)$$

where the vector  $\underline{\mathbf{x}}$  containing  $m_t$  complex components, represents the transmitted signal, the vector  $\underline{\mathbf{y}}$  containing  $m_r$  complex components, represents the received signal, and  $\underline{\mathbf{z}}$  accounts for the presence of additive Gaussian noise. The  $m_r$  components of  $\underline{\mathbf{z}}$  are statistically independent, circularly symmetric complex zero-mean Gaussian variables of unit energy  $\mathbb{E}(|z_j|^2) = 1$ , and the components of  $\underline{\mathbf{x}}$  are constrained such that the average energy of each component is equal to 1, i.e.,  $\mathbb{E}(|x_j|^2) = 1$  for all  $j$ . The term  $\rho$  is proportional to the optical power per excited mode so that  $\rho^2$  is equal to the signal-to-noise ratio (SNR) in the single mode ( $m = 1$ ) case. The matrix  $\mathbf{H}_{11}$  is a block of size  $m_r \times m_t$  within the  $m \times m$  random unitary matrix  $\mathbf{H}$

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix}.$$



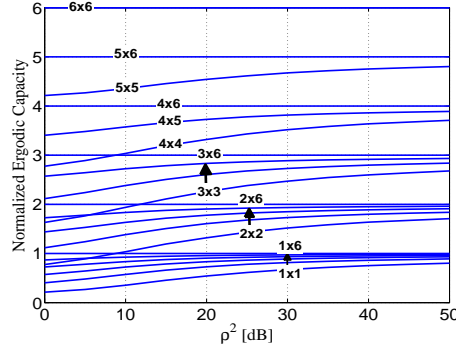


Figure 2.1: The ergodic capacity normalized by  $C(1, 1, 1; \rho) = \log(1 + \rho^2)$  vs.  $\rho^2$  for various combinations of  $m_t \times m_r$  with  $m = 6$ .

As a first stage in our analysis we establish the relation between the transfer matrix  $\mathbf{H}_{11}$  and the Jacobi ensemble of random matrices [8, 9]. The Jacobi ensemble with (integer) parameters  $k_1, k_2, n$  (satisfying  $k_1, k_2 \geq n$ ) is denoted  $\mathcal{J}(k_1, k_2, n)$  and can be constructed as  $G_1^\dagger G_1 (G_1^\dagger G_1 + G_2^\dagger G_2)^{-1}$ , where  $G_1$  and  $G_2$  are statistically independent  $k_1 \times n$  and  $k_2 \times n$  random Gaussian matrices, respectively. By the term Gaussian matrix we are referring to a matrix whose entries are zero-mean complex Gaussian variables with unit-variance. The joint probability density function (PDF) of the eigenvalues of this ensemble is given in [8].

It is known [10] that for  $m_t, m_r$  satisfying the condition  $m_t + m_r \leq m$ , the squared non-zero singular values of  $\mathbf{H}_{11}$  have the same distribution as the eigenvalues of the Jacobi ensemble  $\mathcal{J}(m_{\max}, m - m_{\max}, m_{\min})$ , where  $m_{\min} = \min\{m_t, m_r\}$  and  $m_{\max} = \max\{m_t, m_r\}$ . For  $m_t + m_r > m$  it can be shown [11] that  $m_t + m_r - m$  singular values of  $\mathbf{H}_{11}$  are equal to unity with probability 1, whereas the remaining  $m - m_{\max}$  non-zero singular values of  $\mathbf{H}_{11}$  are equal to the non-zero singular values of  $\mathbf{H}_{22}$ , and hence follow the distribution of the Jacobi ensemble  $\mathcal{J}(m - m_{\min}, m_{\min}, m - m_{\max})$ . The latter property can be seen by noting that the unitarity of  $\mathbf{H}$  implies that  $\mathbf{H}_{11}^\dagger \mathbf{H}_{11} + \mathbf{H}_{21}^\dagger \mathbf{H}_{21} = \mathbf{I}_{m_t}$  and  $\mathbf{H}_{21} \mathbf{H}_{21}^\dagger + \mathbf{H}_{22} \mathbf{H}_{22}^\dagger = \mathbf{I}_{m-m_r}$ .

Since the noise is additive circularly symmetric Gaussian, the capacity for a given channel realization is known [7] and given by  $\log[\det(\mathbf{I}_{m_t} + \rho^2 \mathbf{H}_{11}^\dagger \mathbf{H}_{11})]$ . The ergodic capacity is obtained by averaging this expression over all channel realizations  $\mathbf{H}_{11}$ . It can be expressed in the form

$$C(m_t, m_r, m; \rho) = \mathbb{E}\left[\sum_{i=1}^{m_{\min}} \log(1 + \rho^2 \lambda_i)\right]. \quad (2.2)$$

where the expectation is over  $\lambda_1, \dots, \lambda_{m_{\min}}$ , the squared nonzero singular values of  $\mathbf{H}_{11}$ . In the case  $m_t + m_r \leq m$ , using the the joint PDF of  $\lambda_j$  one finds that the ergodic capacity satisfies [11]

$$C(m_t, m_r, m; \rho) = \int_0^1 \log(1 + \lambda \rho^2) \times \\ \times \lambda^\alpha (1 - \lambda)^\beta \sum_{k=0}^{m_{\min}-1} b_{k, \alpha, \beta}^{-1} [P_k^{(\alpha, \beta)}(1 - 2\lambda)]^2 d\lambda, \quad (2.3)$$

where  $P_k^{(\alpha, \beta)}(x)$  are the Jacobi polynomials (see [12, 8.96]),  $\alpha = |m_r - m_t|$ ,  $\beta = m - m_t - m_r$ ,

and the coefficients  $b_{k,\alpha,\beta}$  are given by

$$b_{k,\alpha,\beta} = \frac{1}{2k + \alpha + \beta + 1} \binom{2k + \alpha + \beta}{k} \binom{2k + \alpha + \beta}{k + \alpha}^{-1}.$$

Note that the above result, as well as all results in what follows are symmetric with respect to  $m_t$  and  $m_r$ , which is a consequence of our choice to maintain the power per transmitter constant in all cases.

To obtain the ergodic capacity in the case where  $m_t + m_r > m$ , we use the distribution of the singular values in that case, turning Eq. (2.2) into

$$C(m_t, m_r, m; \rho) = (m_t + m_r - m)C(1, 1, 1; \rho) + C(m - m_r, m - m_t, m; \rho), \quad (2.4)$$

where  $C(1, 1, 1; \rho) = \log(1 + \rho^2)$  and where  $C(m - m_r, m - m_t, m; \rho)$  is given by Eq. (2.3). Note that the second term on the right-hand-side of (2.4) reduces to 0 when  $m_t$ , or  $m_r$  is equal to  $m$ . The form of Eq. (2.4) follows from the fact that  $m_t + m_r - m$  of the singular values are equal to unity, while the remaining singular values correspond to the Jacobi ensemble, as discussed earlier. In Fig. 2.1 we show an example where the ergodic capacity of the channel is plotted as a function of  $\rho^2$  for the case of  $m = 6$ . The ergodic capacity in the figure was normalized by  $\log(1 + \rho^2)$  (which is also the capacity in the single-mode case). As is evident from the figure, a capacity equivalent to  $m_t + m_r - m$  SISO channels is guaranteed in all cases.

We now turn to the analysis of the non-ergodic case, in which the channel realization is assumed to be constant within each given code-frame. Traditional optical communications systems, which are implemented over single-mode optical fibers ( $m = 2$ ), usually operate in this regime. As discussed earlier, the figure of merit in this regime is the outage probability  $P_{\text{out}}$ , defined as the probability that the capacity induced by the channel realization is lower than the rate  $R$  at which the link is chosen to operate. Note that we assume that the channel instantiation is unknown to the transmitter, thus it can not adapt the transmission rate. The outage probability can be formally expressed as

$$P_{\text{out}}(m_t, m_r, m; R) = Pr[\log \det(\mathbf{I}_{m_t} + \rho^2 \mathbf{H}_{11}^\dagger \mathbf{H}_{11}) < R]. \quad (2.5)$$

For  $m_t + m_r \leq m$  Eq. (2.5) can be readily evaluated by using once again the distribution of the eigenvalues of the Jacobi ensemble [8]. Defining a parameter  $r$  such that the system transmission rate is given by  $R = r \log(1 + \rho^2)$ , we obtain

$$P_{\text{out}}(m_t, m_r, m; R) = K_{m_t, m_r, m}^{-1} \int_{\mathcal{B}} \prod_{i=1}^{m_{\min}} \lambda_i^{|m_r - m_t|} \times \\ \times (1 - \lambda_i)^{m - m_r - m_t} \prod_{i < j} (\lambda_i - \lambda_j)^2 d\lambda, \quad (2.6)$$

where  $K_{m_t, m_r, m}$  is a normalization factor and where  $\mathcal{B}$ , the range of integration includes all ordered singular values of  $\mathbf{H}_{11}$ ,  $0 \leq \lambda_1 \leq \dots \leq \lambda_{m_{\min}} \leq 1$ , for which  $\prod_{i=1}^{m_{\min}} (1 + \rho^2 \lambda_i) < (1 + \rho^2)^r$ .

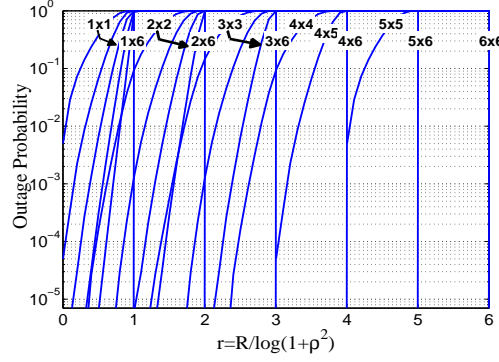


Figure 2.2: Outage probability versus the normalized rate  $r$ . The number of supported modes is  $m = 6$  and  $\rho^2 = 20\text{dB}$ .

For the case of  $m_t + m_r > m$ , since  $m_t + m_r - m$  eigenvalues of  $\mathbf{H}_{11}^\dagger \mathbf{H}_{11}$  are unity and the other  $m - m_{\max}$  non-zero eigenvalues are equal to the non-zeros eigenvalues of  $\mathbf{H}_{22}^\dagger \mathbf{H}_{22}$ , it can be shown [11] that

$$\begin{aligned} P_{out}(m_t, m_r, m; r \log(1 + \rho^2)) &= \\ &= P_{out}(m - m_r, m - m_t, m; \tilde{r} \log(1 + \rho^2)) , \end{aligned} \quad (2.7)$$

where  $\tilde{r}$  is the larger between  $r - (m_t + m_r - m)$  and 0, and where the right-hand-side drops to 0, when  $m_r$ , or  $m_t$  equals  $m$ . Equation (2.7) implies that in the case  $m_t + m_r > m$  the outage probability is identical to that of a channel with  $m - m_r$  modes addressed by the transmitter and  $m - m_t$  modes addressed by the receiver, which is designed to support a transmission rate equivalent to  $\tilde{r}$  single-mode channels. Note that when  $r < (m_t + m_r - m)$ ,  $\tilde{r} = 0$ ,  $P_{out} = 0$  implying that for such rates zero outage probability is achievable. A practical scheme for achieving zero outage under these conditions is proposed and discussed in [11]. In Fig. 2.2 we show an exemplary calculation of the outage probability as a function of the normalized rate  $r$ . These curves, obtained from our analysis were plotted in the same form as the numerical results reported in [3], except that here we assumed a fiber supporting  $m = 6$  scalar modes. Note how the outage probability abruptly drops to 0 whenever  $r$  becomes smaller than  $m_t + m_r - m$ .

## 2.3 Conclusions

We have studied the under-addressed optical MIMO channel, where the number of fiber modes addressed by the transmitter and receiver is allowed to be smaller than the overall number of modes existing in the fiber. This scenario was motivated by the idea of installing fibers that admit more modes that can be MIMO processed with currently available technology, in order to achieve future-proof operation [3]. While the price of not detecting all modes is notable, certain trade-offs exist and can be used to one's advantage. Most importantly, we have shown that a performance equivalent of  $m_t + m_r - m$  uncoupled single-mode channels can be achieved in all cases.

## Acknowledgement

The authors wish to thank Amir Dembo for his useful comments. Mark Shtaif acknowledges financial support by TeraSanta consortium and by Alcatel-Lucent within the framework of Green Touch. The research was partially supported by the Israel Science Foundation, grant No. 634/09.

# Bibliography

- [1] A. R. Chraplyvy, “The coming capacity crunch,” European Conference on Optical Communication 2009 (ECOC09), plenary talk (2009).
- [2] S. Randel, R. Ryf, A. Sierra, P.J. Winzer, A.H. Gnauck, C.A. Bolle, R-J. Essiambre, D.W. Peckham, A. McCurdy, and R. Lingle, “ $6 \times 56$ -Gb/s mode-division multiplexed transmission over 33-km few-mode fiber enabled by  $6 \times 6$  MIMO equalization,” Opt. Express **19**, 16697–16707 (2011).
- [3] P. J. Winzer and G. J. Foschini, “MIMO capacities and outage probabilities in spatially multiplexed optical transport systems,” Opt. Express **19**, 16680-16696 (2011).
- [4] K.-P. Ho and J. M. Kahn, “Frequency Diversity in Mode-Division Multiplexing Systems,” J. Lightwave Technol. **29**, 3719–3726 (2011).
- [5] C. Antonelli, A. Mecozzi, M. Shtaif, and P. J. Winzer, “Stokes-space analysis of modal dispersion in fibers with multiple mode transmission,” Opt. Express **20**, 11718–11733 (2012).
- [6] K.-P. Ho and J. M Kahn, “Statistics of group delays in multi-mode fibers with strong mode coupling,” J. Lightwave Technol. **29**, 3119–3128 (2011).
- [7] I. E. Telatar, “Capacity of multi-antenna Gaussian channels,” European Transactions on Telecommunications **10**, 585-595 (1999).
- [8] R. J. Muirhead, *Aspects of Multivariate Statistical Theory* (Wiley, 1982).
- [9] A. Edelman and N. Raj Rao, “Random matrix theory,” Acta Numerica **14**, 233-297 (2005).
- [10] A. Edelman and B. D. Sutton, “The beta-Jacobi matrix model, the CS decomposition, and generalized singular value problems,” Foundations of Computational Mathematics **8**, 259-285 (2008).
- [11] R. Dar, M. Feder and M. Shtaif, “The Jacobi MIMO Channel,” available on <http://arxiv.org/abs/1202.0305>.
- [12] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic Press, 1980).



# Chapter 3

## The Jacobi MIMO Channel

Ronen Dar, Meir Feder and Mark Shttaif, “*The Jacobi MIMO channel*,” IEEE Transactions on Information Theory, vol. 59, pp. 2426–2441 (April 2013)

### 3.1 abstract

This paper presents a new fading model for MIMO channels, the Jacobi fading model. It asserts that  $\mathbf{H}$ , the transfer matrix which couples the  $m_t$  inputs into  $m_r$  outputs, is a sub-matrix of an  $m \times m$  random (Haar-distributed) unitary matrix. The (squared) singular values of  $\mathbf{H}$  follow the law of the classical Jacobi ensemble of random matrices; hence the name of the channel. One motivation to define such a channel comes from multimode/multicore optical fiber communication. It turns out that this model can be qualitatively different than the Rayleigh model, leading to interesting practical and theoretical results. This paper first evaluates the ergodic capacity of the channel. Then, it considers the non-ergodic case, where it analyzes the outage probability and the diversity-multiplexing tradeoff. In the case where  $k = m_t + m_r - m > 0$  it is shown that at least  $k$  degrees of freedom are guaranteed not to fade for any channel realization, enabling a zero outage probability or infinite diversity order at the corresponding rates. A simple scheme utilizing (a possibly outdated) channel state feedback is provided, attaining the no-outage guarantee. Finally, noting that as  $m$  increases, the Jacobi model approaches the Rayleigh model, the paper discusses the applicability of the model in other communication scenarios.

### 3.2 Introduction

In Multi-Input Multi-Output (MIMO) channels a vector  $\underline{x}$  of  $m_t$  signals is transmitted, a vector  $\underline{y}$  of  $m_r$  signals is received, and an  $m_r \times m_t$  random matrix  $\mathbf{H}$  represents the coupling of the input into the output so that the received vector is  $\underline{y} = \mathbf{H}\underline{x} + \underline{z}$  where  $\underline{z}$  is a noise vector. In this paper we consider a channel matrix  $\mathbf{H}$  which is a sub-matrix of a Haar-distributed unitary matrix, i.e., drawn uniformly from the ensemble of all  $m \times m$  unitary matrices,  $m \geq m_t, m_r$ .

The three classical and most well-studied random matrix ensembles are the *Gaussian*, *Wishart* and *Jacobi* (also known as MANOVA) ensembles [1–4]. A common model for the

channel matrix  $\mathbf{H}$  in fading wireless communication is a matrix with independent Gaussian elements (also known as the Rayleigh model). In that case,  $\mathbf{H}^\dagger \mathbf{H}$  is a *Wishart matrix*. For the model assumed in this paper,  $\mathbf{H}^\dagger \mathbf{H}$  follows the *Jacobi ensemble*. It turns out that this model is both practically useful and qualitatively different than other fading models such as the Rayleigh [5–7], Rician [8–10] and Nakagami [10–13].

Jacobi ensembles, forming an important part of classical random matrix ensembles, are of considerable interest in connection with multivariate statistics and random matrix theory [1–3]. These ensembles have been successfully applied in several problems. One notable example is related to quantum conductance/transmission in mesoscopic systems [4, 14–18], where scattering is modeled by a random unitary matrix (owing to flux conservation) and the blocks of unitary matrix are then the transmission/reflection matrices which govern the conductance/transmission properties. This paper provides another application in communication theory.

The motivation to introduce the Jacobi channel comes from recent developments in optical fiber communication. The expected capacity crunch in long haul optical fibers [19, 20] led to proposals for “space-division multiplexing” (SDM) [21, 22], that is to have several links at the same fiber, by either multiple single-mode fiber strands within a fiber cable, multiple cores within a multi-core fiber, or multiple modes within a multi-mode waveguide. An SDM system with  $m$  parallel transmission paths per wavelength can potentially multiply the throughput of a certain link by a factor of  $m$ . Since  $m$  can potentially be chosen very large, SDM technology is highly scalable. Now, a significant crosstalk between the optical paths raises the need for MIMO signal processing techniques. Unfortunately, for large size MIMO (large  $m$ ) this is unfeasible currently in the optical rates. Assuming that faster computation will be available in the future and having in mind that replacing optical fibers to support SDM is a long and expensive procedure, a long term design is sought after. To that end and more, it was proposed to design an optical system that can support relatively large number of paths for future use, but at start to address only some of the paths. In this scenario the channel can be modeled as a sub-matrix of a larger unitary matrix, i.e., the Jacobi model is applicable.

This under-addressed channel is discussed in [23] where simulations of the capacities and outage probabilities were presented. In this paper we further analyze the channel in the ergodic and non-ergodic settings, where we provide analytical expression for the capacity, outage probability and the diversity-multiplexing tradeoff. It should be noted that in optical systems the outage probability is an important measure, required to be very low. Evidently, since the entire channel matrix is unitary, when all paths are addressed a zero outage probability can be attained for any transmission rate. An interesting result that comes out of this work is that there are situations, where a partial number of paths are addressed, yet a number of streams are guaranteed to experience zero outage. Thus, choosing the number of addressed paths and the corresponding rate is a very critical design element that highly reflects on the system outage and performance. A preliminary description of our work, in the context of the SDM optical channel is provided in [24].

A possibly practical outcome of this work is a simple communication scheme, with channel state feedback, that achieves the highest rate possible with no outage. The scheme works even when the feedback is “outdated”, and it allows simple decoding with no complicated MIMO signal processing, making it plausible for optical communication. We note that while



our theoretical findings indicate that the no-outage promise can be attained with no feedback, the quest for such simple schemes is open.

As noted above, the motivation for this work comes from optical fiber communication. Yet, the application of the Jacobi model and the insights that follow from it may be relevant in other cases, such as in-line communication and even wireless communication. While a constant can be applied to account for the average power loss in the medium, the randomness structure of the loss can be modeled using the Jacobi model. As  $m$  increases with respect to  $m_t, m_r$ , the randomness of the absolute received power increases. Evidently, in typical wireless communication, a large fraction of the energy is not captured, and so the channel can be modeled as a small sub-matrix of a large unitary matrix. Indeed, it will be shown that as  $m$  becomes larger in comparison to  $m_t, m_r$ , the Jacobi model (up to a normalizing constant) approaches the Rayleigh model.

The paper is organized as follows. We start by defining the system model and presenting the channel statistics in Section 3.3. An interesting transition threshold is revealed: when the number of addressed paths is large enough, so that  $k = m_t + m_r - m > 0$ , the statistics of the problem changes. Using this observation we give analytic expressions for the ergodic capacity in Section 3.4. In Section 3.5 we analyze the outage probabilities in the non-ergodic channel and show that for  $k > 0$  a strictly zero outage probability is obtainable for  $k$  degrees of freedom. Following this finding, we present in Section 3.6 a new communication scheme which exploits a channel state feedback to achieve zero outage probability. Section 3.7 discuss the diversity-multiplexing tradeoff of the channel where we show an absorbing difference in the maximum diversity gain between the Rayleigh fading and Jacobi channels. Section 3.9 discusses the results.

### 3.3 System Model and Channel Statistics

We consider a space-division multiplexing (SDM) system that supports  $m$  spatial propagation paths. In tribute to optical communication, in particular multi-mode optical fibers, the initial motivation for this work, we shall refer to these links as modes. Assuming a unitary coupling among all transmission modes the overall transfer matrix  $\mathbf{H}$  can be described as an  $m \times m$  unitary matrix, where each entry  $\mathbf{h}_{ij}$  represents the complex path gain from transmitted mode  $i$  to received mode  $j$ . We further assume a uniformly distributed unitary coupling, that is,  $\mathbf{H}$  is drawn uniformly from the ensemble of all  $m \times m$  unitary matrices (Haar distributed). Considering a communication system where  $m_t \leq m$  and  $m_r \leq m$  modes are being addressed by the transmitter and receiver, respectively, the effective transfer matrix is a truncated version of  $\mathbf{H}$ . Under these conditions the channel can be described as

$$\underline{\mathbf{y}} = \sqrt{\rho} \mathbf{H}_{11} \underline{\mathbf{x}} + \underline{\mathbf{z}}, \quad (3.1)$$

where the vector  $\underline{\mathbf{x}}$  containing  $m_t$  complex components, represents the transmitted signal, the vector  $\underline{\mathbf{y}}$  containing  $m_r$  complex components, represents the received signal, and  $\underline{\mathbf{z}}$  accounts for the presence of additive Gaussian noise. The  $m_r$  components of  $\underline{\mathbf{z}}$  are statistically independent, circularly symmetric complex zero-mean Gaussian variables of unit energy  $\mathbb{E}(|z_j|^2) = 1$ . The components of  $\underline{\mathbf{x}}$  are constrained such that the average energy of each

component is equal to 1, i.e.,  $\mathbb{E}(|x_j|^2) = 1$  for all  $j$ <sup>1</sup>. The term  $\rho \geq 0$  is proportional to the power per excited mode so that it equals to the signal-to-noise ratio in the single mode case ( $m = 1$ ). The matrix  $\mathbf{H}_{11}$  is a block of size  $m_r \times m_t$  within the  $m \times m$  random unitary matrix  $\mathbf{H}$

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix}. \quad (3.2)$$

As a first stage in our analysis we establish the relation between the transfer matrix  $\mathbf{H}_{11}$  and the Jacobi ensemble of random matrices [1–3]. Limiting our discussion to complex matrices we state the following definitions:

**Definition 1** (Gaussian matrices).  $\mathcal{G}(m, n)$  is  $m \times n$  matrix of i.i.d complex entries distributed as  $\mathcal{CN}(0, 1)$ .

**Definition 2** (Wishart ensemble).  $\mathcal{W}(m, n)$ , where  $m \geq n$ , is  $n \times n$  Hermitian matrix which can be constructed as  $A^\dagger A$ , where  $A$  is  $\mathcal{G}(m, n)$ .

**Definition 3** (Jacobi ensemble).  $\mathcal{J}(m_1, m_2, n)$ , where  $m_1, m_2 \geq n$ , is  $n \times n$  Hermitian matrix which can be constructed as  $A(A+B)^{-1}$ , where  $A$  and  $B$  are  $\mathcal{W}(m_1, n)$  and  $\mathcal{W}(m_2, n)$ , respectively.

The first two definitions relate to wireless communication [7]. We claim here that the third classical ensemble, the Jacobi ensemble, is relevant to this channel model by relating its eigenvalues to the singular values of  $\mathbf{H}_{11}$ . To that end we quote the well-known [1, 4] joint probability density function (PDF) of the ordered eigenvalues  $0 \leq \lambda_1 \leq \dots \leq \lambda_n \leq 1$  of the Jacobi ensemble  $\mathcal{J}(m_1, m_2, n)$

$$f(\lambda_1, \dots, \lambda_n) = K_{m_1, m_2, n}^{-1} \prod_{i=1}^n \lambda_i^{m_1-n} (1 - \lambda_i)^{m_2-n} \prod_{i < j} (\lambda_i - \lambda_j)^2, \quad (3.3)$$

where  $K_{m_1, m_2, n}$  is a normalizing constant. We say that  $n$  variables follow the law of the Jacobi ensemble  $\mathcal{J}(m_1, m_2, n)$  if their joint distribution follows (3.3).

We shall now present the explicit distribution of the channel's singular values by distinguishing between the following two cases:

### 3.3.1 Case I - $m_t + m_r \leq m$

In [25, Theorem 1.5] it was shown that for  $m_t, m_r$  satisfying the conditions  $m_t \leq m_r$  and  $m_t + m_r \leq m$ , the eigenvalues of  $\mathbf{H}_{11}^\dagger \mathbf{H}_{11}$  have the same distribution as the eigenvalues of the Jacobi ensemble  $\mathcal{J}(m_r, m - m_r, m_t)$ . For  $m_t, m_r$  satisfying  $m_t > m_r$  and  $m_t + m_r \leq m$ , since  $\mathbf{H}^\dagger$  share the same distribution with  $\mathbf{H}$ , the eigenvalues of  $\mathbf{H}_{11} \mathbf{H}_{11}^\dagger$  follow the law of the Jacobi ensemble  $\mathcal{J}(m_t, m - m_t, m_r)$ . Combining these two results, we can say that the squared non-zero singular values of  $\mathbf{H}_{11}$  have the same distribution as the eigenvalues of the Jacobi ensemble  $\mathcal{J}(m_{\max}, m - m_{\max}, m_{\min})$ , where here and throughout this paper we denote  $m_{\max} = \max\{m_t, m_r\}$  and  $m_{\min} = \min\{m_t, m_r\}$ .

---

<sup>1</sup>The constant per-mode power constraint, as opposed to the constant total power constraint often used in wireless communication, is motivated by the optical fiber nonlinearity limitation. Nevertheless, the total power constraint will be considered as well when needed.

### 3.3.2 Case II - $m_t + m_r > m$

When the sum of transmit and receive modes,  $m_t + m_r$ , is larger than the total available modes,  $m$ , the statistics of the singular values change. Clearly, when  $\mathbf{H}_{11}$  is the complete unitary matrix ( $m_t = m_r = m$ ), all singular values are one. Thus, having in mind that the columns of  $\mathbf{H}$  are orthonormal, one can think of  $m_t + m_r > m$  as a transition threshold in which the size of  $\mathbf{H}_{11}$  is large enough with respect to the size of the complete unitary matrix to change the singularity statistics. The following Lemma provides the joint distribution of the singular values of  $\mathbf{H}_{11}$ , showing that for any realization of  $\mathbf{H}_{11}$  there are  $m_t + m_r - m$  singular values which are 1. This Lemma is a Corollary of a result of Paige and Saunders [26], however its proof is given here for completeness.

**Lemma 1.** *Let  $\mathbf{H}$  be an  $m \times m$  unitary matrix, divided into blocks as in (3.2), where  $\mathbf{H}_{11}$  is an  $m_r \times m_t$  block with  $m_t + m_r > m$ . Then  $m_t + m_r - m$  eigenvalues of  $\mathbf{H}_{11}^\dagger \mathbf{H}_{11}$  are 1,  $m_t - m_{\min}$  are 0, and  $m - m_{\max}$  equal to the non-zero eigenvalues of  $\mathbf{H}_{22} \mathbf{H}_{22}^\dagger$ ; if  $\mathbf{H}$  is Haar distributed these  $m - m_{\max}$  eigenvalues follow the law of the Jacobi ensemble  $\mathcal{J}(m - m_{\min}, m_{\min}, m - m_{\max})$ .*

*Proof.* Since  $\mathbf{H}$  is unitary we can write

$$\mathbf{H}_{11}^\dagger \mathbf{H}_{11} + \mathbf{H}_{21}^\dagger \mathbf{H}_{21} = \mathbf{I}_{m_t} \quad (3.4)$$

and

$$\mathbf{H}_{21} \mathbf{H}_{21}^\dagger + \mathbf{H}_{22} \mathbf{H}_{22}^\dagger = \mathbf{I}_{m-m_r} . \quad (3.5)$$

Let  $\{\lambda_i^{(11)}\}_{i=1}^{m_t}$  and  $\{\lambda_i^{(21)}\}_{i=1}^{m_t}$  be the eigenvalues of  $\mathbf{H}_{11}^\dagger \mathbf{H}_{11}$  and  $\mathbf{H}_{21}^\dagger \mathbf{H}_{21}$ , respectively. From (3.4) we can write

$$\lambda_i^{(11)} = 1 - \lambda_i^{(21)} \quad \forall i = 1, \dots, m_t . \quad (3.6)$$

Since  $\mathbf{H}_{21}$  is a block of size  $(m - m_r) \times m_t$  where  $m - m_r < m_t$ ,  $\mathbf{H}_{21}^\dagger \mathbf{H}_{21}$  has (at least)  $m_t + m_r - m$  zero eigenvalues. Following (3.6),  $\mathbf{H}_{11}^\dagger \mathbf{H}_{11}$  has  $m_t + m_r - m$  eigenvalues which are 1. Now, let  $\{\tilde{\lambda}_i^{(21)}\}_{i=1}^{m-m_r}$  and  $\{\tilde{\lambda}_i^{(22)}\}_{i=1}^{m-m_r}$  be the eigenvalues of  $\mathbf{H}_{21} \mathbf{H}_{21}^\dagger$  and  $\mathbf{H}_{22} \mathbf{H}_{22}^\dagger$ , respectively. From (3.5) we can write

$$\tilde{\lambda}_i^{(21)} = 1 - \tilde{\lambda}_i^{(22)} \quad \forall i = 1, \dots, m - m_r . \quad (3.7)$$

Since for any matrix  $A$ ,  $A^\dagger A$  and  $AA^\dagger$  share the same non-zero eigenvalues we can combine (3.6) and (3.7) to conclude that the additional  $m - m_r$  eigenvalues of  $\mathbf{H}_{11}^\dagger \mathbf{H}_{11}$  are equal to the  $m - m_r$  eigenvalues of  $\mathbf{H}_{22} \mathbf{H}_{22}^\dagger$ . Note that  $m_t - m_{\min}$  of them are 0. Since the above arguments hold for any unitary matrix, and since  $\mathbf{H}_{22}$  is a block of size  $(m - m_r) \times (m - m_t)$ , when  $\mathbf{H}$  is Haar distributed the results of subsection 3.3.1 can be applied, which completes the proof.  $\square$

Lemma 1 reveals an interesting algebraic phenomenon:  $k = \max\{m_t + m_r - m, 0\}$  singular values of  $\mathbf{H}_{11}$  are 1 for any realization of  $\mathbf{H}$ . This provides some powerful results in the context of Jacobi fading channels. For example, the channel's power  $\|\mathbf{H}_{11}\|_F^2$ , where  $\|A\|_F$  denotes the Frobenius norm of  $A$ , is guaranteed to be at least  $k$ . Furthermore,  $\mathbf{H}_{11}$  always comprises an unfaded  $k$ -dimensional subspace. In what follows we show that this implies a lower bound on the ergodic capacity, an achievable zero outage probability and an “unbounded” diversity gain for certain rates.

### 3.4 The Ergodic Case

In the ergodic scenario the channel is assumed to be rapidly changing so that the transmitted signal samples the entire channel statistics. We further assume that the channel realization at each symbol time is known only at the receiver end. It is well known [6] that the channel capacity in that case is achieved by taking  $\underline{\mathbf{x}}$  to be a vector of circularly symmetric complex zero-mean Gaussian components; and is given by

$$C(m_t, m_r, m; \rho) = \max_{\substack{Q: Q \succeq 0 \\ Q_{ii} \leq 1 \ \forall \ i=1, \dots, m_t}} \mathbb{E}[\log \det(\mathbf{I}_{m_r} + \rho \mathbf{H}_{11} Q \mathbf{H}_{11}^\dagger)] , \quad (3.8)$$

where the maximization is over all covariance matrices of  $\underline{\mathbf{x}}$ ,  $Q$ , that satisfy the power constraints. Now, the capacity in (3.8) also satisfies

$$C(m_t, m_r, m; \rho) \leq \max_{\substack{Q: Q \succeq 0 \\ \text{trace}(Q) \leq m_t}} \mathbb{E}[\log \det(\mathbf{I}_{m_r} + \rho \mathbf{H}_{11} Q \mathbf{H}_{11}^\dagger)] , \quad (3.9)$$

where it is well known [6, Theorem 1] that if the distribution of  $\mathbf{H}_{11}$  is invariant under unitary permutations,  $Q = \mathbf{I}_{m_t}$  is the optimal choice for (3.9). Since  $\mathbf{H}$  is Haar-distributed, that is invariant under unitary permutations, also  $\mathbf{H}_{11}$  is invariant under unitary permutations. Thus  $Q = \mathbf{I}_{m_t}$  is the optimal choice for (3.8) and by using the following equation

$$\det(\mathbf{I}_{m_r} + \rho \mathbf{H}_{11} \mathbf{H}_{11}^\dagger) = \det(\mathbf{I}_{m_t} + \rho \mathbf{H}_{11}^\dagger \mathbf{H}_{11}),$$

we can conclude that the ergodic capacity is given by

$$C(m_t, m_r, m; \rho) = \mathbb{E}[\log \det(\mathbf{I}_{m_t} + \rho \mathbf{H}_{11}^\dagger \mathbf{H}_{11})] . \quad (3.10)$$

#### 3.4.1 Case I - $m_t + m_r \leq m$

The following theorem gives an analytical expression to the ergodic capacity for cases where  $m_t + m_r \leq m$ . Using the joint distribution of the eigenvalues of the Jacobi ensemble we associate the ergodic capacity with the Jacobi polynomials [27, 8.96].

**Theorem 1.** *The ergodic capacity of the channel defined in (3.1) with  $m_t, m_r$  satisfying  $m_t + m_r \leq m$ , reads*

$$C(m_t, m_r, m; \rho) = \int_0^1 \lambda^\alpha (1 - \lambda)^\beta \log(1 + \lambda \rho) \sum_{k=0}^{m_{\min}-1} b_{k, \alpha, \beta}^{-1} [P_k^{(\alpha, \beta)}(1 - 2\lambda)]^2 d\lambda \quad (3.11)$$

where  $P_k^{(\alpha, \beta)}(x)$  are the Jacobi polynomials

$$P_k^{(\alpha, \beta)}(x) = \frac{(-1)^k}{2^k k!} (1 - x)^{-\alpha} (1 + x)^{-\beta} \frac{d^k}{dx^k} [(1 - x)^{k+\alpha} (1 + x)^{k+\beta}] , \quad (3.12)$$

the coefficients  $b_{k, \alpha, \beta}$  are given by

$$b_{k, \alpha, \beta} = \frac{1}{2k + \alpha + \beta + 1} \binom{2k + \alpha + \beta}{k} \binom{2k + \alpha + \beta}{k + \alpha}^{-1} ,$$

and  $\alpha = |m_r - m_t|$ ,  $\beta = m - m_t - m_r$ .

*Proof.* See Appendix 3.10, where the Gaudin-Mehta method [28] is applied to analyze the ergodic capacity.  $\square$

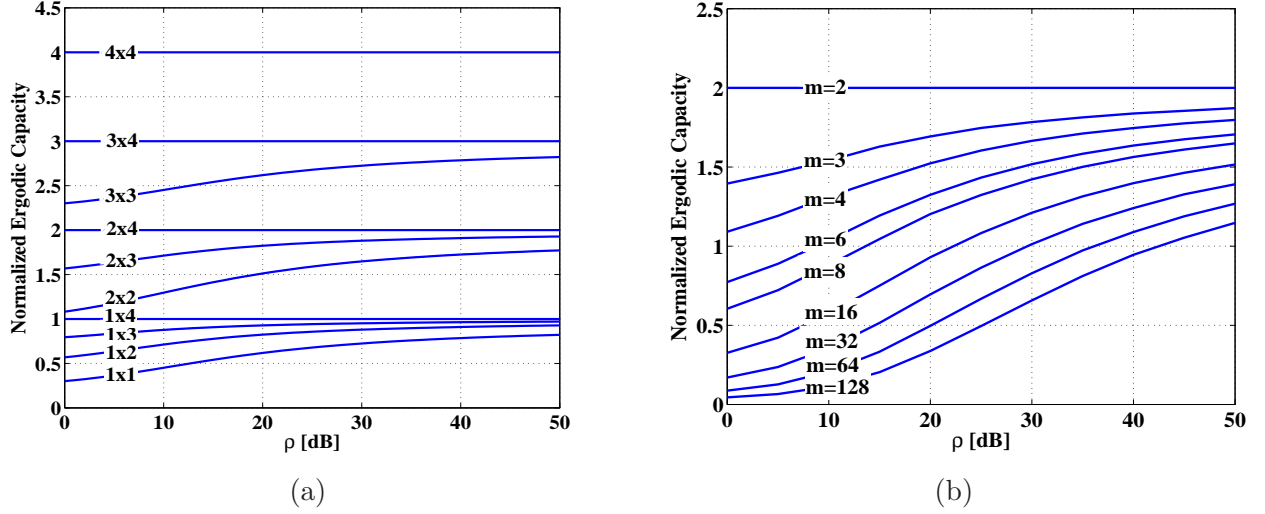


Figure 3.1: The ergodic capacity, normalized by  $C(1, 1, 1; \rho) = \log(1 + \rho)$ , as a function of  $\rho$ . In (a) the number of supported modes is fixed  $m = 4$ , various numbers of transmit  $\times$  receive modes; in (b) the number of addressed modes is fixed  $m_t = m_r = 2$ , various values of supported modes  $m$ .

### 3.4.2 Case II - $m_t + m_r > m$

Applying Lemma 1 to the channel capacity given in (3.10) readily results in the following theorem.

**Theorem 2.** *The ergodic capacity of the channel defined in (3.1) with  $m_t, m_r$  satisfying  $m_t + m_r > m$ , reads*

$$C(m_t, m_r, m; \rho) = (m_t + m_r - m)C(1, 1, 1; \rho) + C(m - m_r, m - m_t, m; \rho), \quad (3.13)$$

where  $C(1, 1, 1; \rho)$  is the SISO channel capacity  $\log(1 + \rho)$ .

*Proof.* According to (3.10) the ergodic capacity satisfies

$$C(m_t, m_r, m; \rho) = \mathbb{E}[\log \det(\mathbf{I}_{m_t} + \rho \mathbf{H}_{11}^\dagger \mathbf{H}_{11})] \quad (3.14)$$

$$= \mathbb{E}\left[\sum_{i=1}^{m_t} \log(1 + \rho \lambda_i)\right], \quad (3.15)$$

where  $\{\lambda_i\}_{i=1}^{m_t}$  are the eigenvalues of  $\mathbf{H}_{11}^\dagger \mathbf{H}_{11}$ . According to Lemma 1,  $m_t + m_r - m$  eigenvalues are 1 and the rest are equal to the  $m - m_r$  eigenvalues of  $\mathbf{H}_{22}^\dagger \mathbf{H}_{22}$ . Applying that into (3.15) results

$$C(m_t, m_r, m; \rho) = (m_t + m_r - m) \log(1 + \rho) + \mathbb{E}[\log \det(\mathbf{I}_{m-m_r} + \rho \mathbf{H}_{22}^\dagger \mathbf{H}_{22})]. \quad (3.16)$$

□

Note that the second term on the right-hand-side of (3.13),  $C(m - m_r, m - m_t, m; \rho)$ , is given by Theorem 1 and reduces to 0 when  $m_t$ , or  $m_r$  is equal to  $m$ . Thus, (3.13) suggests that for systems with  $k = m_t + m_r - m > 0$ , the ergodic capacity is the sum capacities of  $k$  unfaded SISO capacities and a Jacobi MIMO channel with  $m - m_r$  transmit modes and  $m - m_t$  receive modes. Fig. 3.1a depicts the ergodic capacity as a function of  $\rho$  for  $m = 4$  and various combinations of  $m_t$ ,  $m_r$  (note that the ergodic capacity, in our case, is symmetric in  $m_t$ ,  $m_r$ ; thus all combinations are plotted). As is evident from the figure, a capacity equivalent to  $k$  SISO channels is guaranteed in all cases. In Fig. 3.1b the ergodic capacities for  $m_t = m_r = 2$  and various values of supported modes are plotted. Note that as  $m$  increases, the power loss increases and the ergodic capacity becomes smaller. Unlike the common practice of expressing the capacity in terms of the received SNR, here the capacities are presented as a function of  $\rho$ . This normalizes the capacity expression to reflect the capacity loss due to power loss including power leaked into the unobserved modes. In particular, this presentation enables to examine the total effect (capacity loss) of increasing  $m$ . See further discussion in section 3.8.

### 3.5 The Non-Ergodic Case

In the non-ergodic scenario the channel matrix is drawn randomly but rather assumed to be constant within the entire transmission period of each code-frame. The figure of merit in the non-ergodic case is the *outage probability* defined as the probability that the mutual information induced by the channel realization is lower than the rate  $R$  at which the link is chosen to operate. Note that we assume that the channel instantiation is unknown at the transmitter, thus it can not adapt the transmission rate. However, the channel is assumed to be known at the receiver end. By taking an input vector of circularly symmetric complex zero-mean Gaussian variables with covariance matrix  $Q$  the mutual information is maximized and the outage probability can be expressed as

$$P_{out}(m_t, m_r, m; R) = \inf_{Q: Q \succeq 0} Pr[\log \det(\mathbf{I}_{m_r} + \rho \mathbf{H}_{11} Q \mathbf{H}_{11}^\dagger) < R], \quad (3.17)$$

where the minimization is over all covariance matrices  $Q$  satisfying the power constraints. Since the statistics of  $\mathbf{H}_{11}$  are invariant under unitary permutations, the optimal choice of  $Q$ , when applying *constant per-mode power constraint*, is simply the identity matrix. We note that when imposing *total power constraint*, the optimal choice of  $Q$  may depend on  $R$  and  $\rho$  and in general is unknown, even for the Rayleigh channel. Nevertheless, when  $\rho \gg 1$  the identity matrix is approximately the optimal  $Q$  (see section 3.7). Thus, in the following we make the simplified assumption that the transmitted covariance matrix is the commonly used choice  $Q = \mathbf{I}_{m_t}$ .

Now, let the transmission rate be  $R = r \log(1 + \rho)$  (bps/Hz) and let  $\underline{\lambda} = \{\lambda_i\}_{i=1}^{m_{\min}}$  be the ordered non-zeros eigenvalues of  $\mathbf{H}_{11}^\dagger \mathbf{H}_{11}$ ; we can write

$$P_{out}(m_t, m_r, m; r \log(1 + \rho)) = Pr[\log \det(\mathbf{I}_{m_t} + \rho \mathbf{H}_{11}^\dagger \mathbf{H}_{11}) < R] \quad (3.18)$$

$$= Pr\left[\prod_{i=1}^{m_{\min}} (1 + \rho \lambda_i) < (1 + \rho)^r\right], \quad (3.19)$$

and evaluate this expression by applying the statistics of  $\underline{\lambda}$ .

### 3.5.1 Case I - $m_t + m_r \leq m$

Using (3.3) we can apply the joint distribution of  $\underline{\lambda}$  into (3.19) to get

$$P_{out}(m_t, m_r, m; r \log(1 + \rho)) = K_{m_t, m_r, m}^{-1} \int_{\mathcal{B}} \prod_{i=1}^{m_{\min}} \lambda_i^{|m_r - m_t|} (1 - \lambda_i)^{m - m_r - m_t} \prod_{i < j} (\lambda_i - \lambda_j)^2 d\underline{\lambda} , \quad (3.20)$$

where  $K_{m_t, m_r, m}$  is a normalizing factor and  $\mathcal{B}$  describes the outage event

$$\mathcal{B} = \left\{ \underline{\lambda} : \prod_{i=1}^{m_{\min}} (1 + \rho \lambda_i) < (1 + \rho)^r \right\} .$$

This gives an analytical expression to the outage probability. See Fig. 3.3 and the example below.

**Example 1.** Suppose  $m_t = 1$  and  $m_r, m$  satisfy  $m \geq 1 + m_r$ . In that case the outage probability is given by

$$P_{out}(1, m_r, m; R) = K_{1, m_r, m}^{-1} \int_0^{(2^R - 1)/\rho} \lambda^{m_r - 1} (1 - \lambda)^{m - m_r - 1} d\lambda . \quad (3.21)$$

Thus, we can write

$$P_{out}(1, m_r, m; R) = \frac{B((2^R - 1)/\rho; m_r, m - m_r)}{B(1; m_r, m - m_r)} , \quad (3.22)$$

where  $B(x; a, b)$  is the incomplete beta function. Hence, to support an outage probability smaller than  $\epsilon$ ,  $R$  and  $\rho$  have to satisfy

$$\frac{\rho}{2^R - 1} \geq \rho_{norm} = 1/B^{-1}(\epsilon B(1; m_r, m - m_r); m_r, m - m_r) ,$$

where  $B^{-1}(x; a, b)$  is the inverse function of  $B(x; a, b)$ .  $\rho_{norm}$  is the normalized signal-to-noise ratio at the transmitter, is proportional to the received normalized signal-to-noise ratio, and essentially measures the minimal additional power required to support a target rate  $R$  with outage probability smaller than  $\epsilon$  (additional power over the minimal required in SISO unfading channel ( $m = m_r$ )). As  $\rho_{norm}$  is smaller one can afford higher data rate or smaller  $\rho$  (smaller transmission power).

In Fig. 3.2 we plot  $\rho_{norm}$  as a function of  $m_r/m$  for various numbers of available modes  $m$  and desired outage probabilities  $\epsilon$ . For fixed  $m$  and  $m_r/m$ ,  $\rho_{norm}$  increases as  $\epsilon$  decreases (since more power or lower data rate are needed to achieve smaller outage probability). For fixed  $\epsilon$  and  $m$ ,  $\rho_{norm}$  decreases as  $m_r/m$  increases (since more modes are addressed by the receiver, therefore the power loss decreases). This is also true as  $m$  increases while  $\epsilon$  and  $m_r/m$  are fixed (since the diversity at the receiver increases, see Section 3.7). Note that for  $m_r/m = 1$  there is no power loss and we get  $\rho_{norm} = 1$ , that is, the minimal transmission power required to support the rate  $R$ , for any  $\epsilon$ , is  $\rho = 2^R - 1$ .



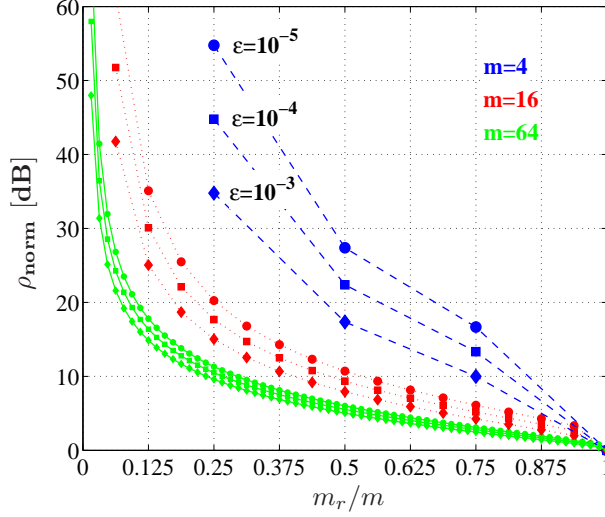


Figure 3.2:  $\rho_{\text{norm}}$  as a function of  $m_r/m$  for  $m = 4, 16, 64$  (blue dashed, red dotted, green solid). Curves are drawn for outage probabilities  $\epsilon = 10^{-5}, 10^{-4}, 10^{-3}$  (circle, square, diamond).

### 3.5.2 Case II - $m_t + m_r > m$

Applying Lemma 1 to (3.19) gives the following.

**Theorem 3.** *The outage probability of the channel defined in (3.1), with  $m_t, m_r$  satisfying  $m_t + m_r > m$ , satisfies*

$$P_{\text{out}}(m_t, m_r, m; r \log(1 + \rho)) = P_{\text{out}}(m - m_r, m - m_t, m; \tilde{r} \log(1 + \rho)) , \quad (3.23)$$

where  $\tilde{r}$  is the larger between  $r - (m_t + m_r - m)$  and 0.

*Proof.* According to (3.19), the outage probability is given by

$$P_{\text{out}}(m_t, m_r, m; r \log(1 + \rho)) = \Pr \left[ \prod_{i=1}^{m_t} (1 + \rho \lambda_i) < (1 + \rho)^r \right] , \quad (3.24)$$

where  $\{\lambda_i\}_{i=1}^{m_t}$  are the eigenvalues of  $\mathbf{H}_{11}^\dagger \mathbf{H}_{11}$ . By applying Lemma 1 we get

$$P_{\text{out}}(m_t, m_r, m; r \log(1 + \rho)) = \Pr \left[ \prod_{i=1}^{m-m_r} (1 + \rho \tilde{\lambda}_i) < (1 + \rho)^{r-(m_t+m_r-m)} \right] , \quad (3.25)$$

where  $\{\tilde{\lambda}_i\}_{i=1}^{m-m_r}$  are the eigenvalues of  $\mathbf{H}_{22} \mathbf{H}_{22}^\dagger$ . When  $\tilde{r} = r - (m_t + m_r - m) < 0$  we get  $P_{\text{out}}(m_t, m_r, m; r \log(1 + \rho)) = 0$ .  $\square$

Note that the right-hand-side drops to 0, when  $m_r$ , or  $m_t$  equals  $m$ . Most importantly, when  $r < m_t + m_r - m$ ,  $\tilde{r} = 0$ , implying that for such rates zero outage probability is achievable. In addition, when  $r \geq m_t + m_r - m > 0$ , Eq. (3.23) implies that the outage



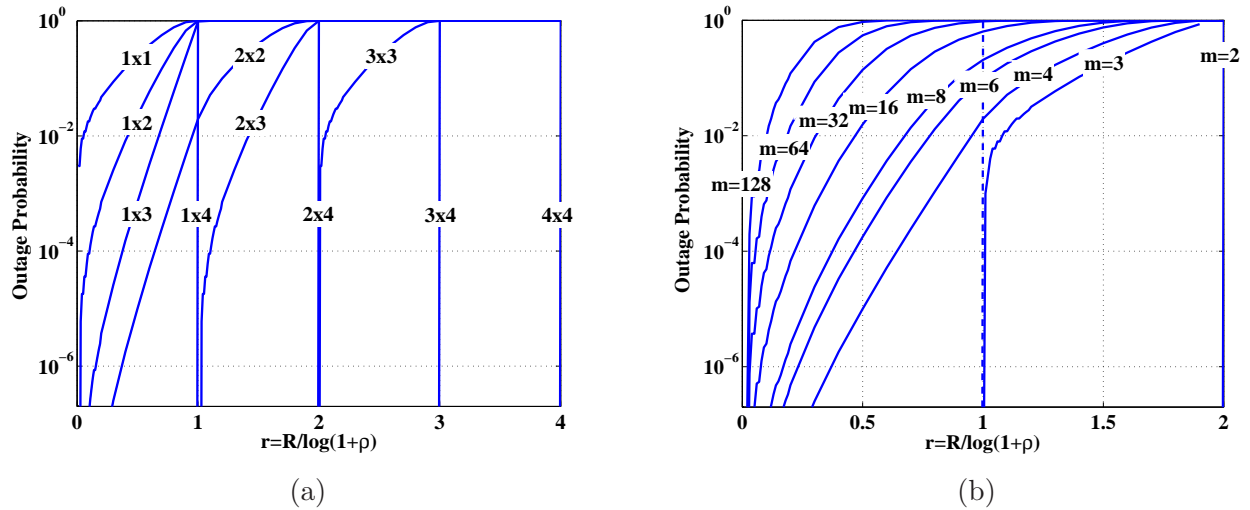


Figure 3.3: Outage probability vs. normalized rate for  $\rho = 20\text{dB}$ . In (a) the number of supported modes is fixed  $m = 4$ , various numbers of transmit $\times$ receive modes; in (b) the number of addressed modes is fixed  $m_t = m_r = 2$ , various values of supported modes  $m$ .

probability is identical to that of a channel with  $m - m_r$  modes addressed by the transmitter and  $m - m_t$  modes addressed by the receiver, which is designed to support a transmission rate equivalent to  $\tilde{r}$  single-mode channels. Thus the right-hand-side of (3.23) applies to Eq. (3.20). In Fig. 3.3a we show an exemplary calculation of the outage probability. These curves, obtained from our analysis were plotted in the same form as the numerical results reported in [23]. Note how the outage probability abruptly drops to 0 whenever  $r$  becomes smaller than  $m_t + m_r - m$ . Also note that the outage probability is symmetric in  $m_t$ ,  $m_r$  since we applied a constant *per-mode* power constraint; thus all combinations of  $m_t$ ,  $m_r$  are plotted in Fig. 3.3a. In Fig. 3.3b outage probability curves are plotted for  $m_t = m_r = 2$  and various values of supported modes,  $m$ . Note that as  $m$  is larger, more power is lost in the unaddressed modes, therefore, as evident from the figure, the outage probability increases.

### 3.6 Achieving The No-Outage Promise

In the previous section we saw that for systems satisfying  $k = m_t + m_r - m > 0$ , a zero outage probability is achievable for any transmission rate below  $R = k \log(1 + \rho)$ . In this section we present a new communication scheme that achieves this promise with a transmission rate arbitrarily close to  $R = k \log(1 + \rho)$ . Using simple manipulations, the scheme exploits a (delayed) channel state information (CSI) feedback to transform the channel into  $k$  independent SISO channels, supporting  $k$  streams (degrees of freedom) with zero outage probability.

Let

$$\mathbf{H}^{(i)} = \begin{bmatrix} \mathbf{H}_{11}^{(i)} & \mathbf{H}_{12}^{(i)} \\ \mathbf{H}_{21}^{(i)} & \mathbf{H}_{22}^{(i)} \end{bmatrix}$$

be the unitary matrix realization at channel use  $i$  and let

$$\underline{\mathbf{y}}^{(i)} = \sqrt{\rho} \mathbf{H}_{11}^{(i)} \underline{\mathbf{x}}^{(i)} + \underline{\mathbf{z}}^{(i)}$$

be the received signal. We assume a perfect knowledge of  $\mathbf{H}_{11}^{(i)}$  at the receiver and a noiseless CSI feedback with a delay of a single channel use. Since  $\mathbf{H}^{(i)}$  is unitary, a matrix  $\mathbf{H}_{21}^{(i)}$  can be computed to complete  $\mathbf{H}_{11}^{(i)}$  into orthonormal columns. Therefore, we shall assume that the receiver noiselessly communicates this computed  $\mathbf{H}_{21}^{(i)}$  to the transmitter. Nevertheless, note that for  $m_t + m_r - m > 1$  and certain matrix instantiations, the computed  $\mathbf{H}_{21}^{(i)}$  is not unique and can be chosen wisely (see Remark 4).

Now, let the transmitter excites the following signal from the addressed modes at each channel use  $i = 1, \dots, n$

$$\underline{\mathbf{x}}^{(i)} = \begin{bmatrix} \mathbf{x}_1^{(i)} \\ \vdots \\ \mathbf{x}_{m_t+m_r-m}^{(i)} \\ \mathbf{H}_{21}^{(i-1)} \underline{\mathbf{x}}^{(i-1)} \end{bmatrix}.$$

That is, the transmitter conveys  $m_t + m_r - m$  new information bearing symbols and  $\mathbf{H}_{21}^{(i-1)} \underline{\mathbf{x}}^{(i-1)}$ , a linear combination of the signal that was transmitted in the previous channel use ( $\underline{\mathbf{x}}^{(0)}$  is a vector of zeros). Note that  $\mathbf{H}$  is unitary, thus the power constraint is left satisfied.

We shall now assume that after the last signal  $\underline{\mathbf{y}}^{(n)}$  is received, the receiver gets as a side information the following noisy measures

$$\underline{\mathbf{y}}_{\text{si}} = \sqrt{\rho} \mathbf{H}_{21}^{(n)} \underline{\mathbf{x}}^{(n)} + \underline{\mathbf{z}}_{\text{si}}, \quad (3.26)$$

where the components of  $\underline{\mathbf{z}}_{\text{si}}$  are i.i.d.  $\mathcal{CN}(0, 1)$ . Thus the receiver can linearly combine  $\underline{\mathbf{y}}^{(n)}$  and  $\underline{\mathbf{y}}_{\text{si}}$  in the following manner

$$\tilde{\underline{\mathbf{y}}}^{(n)} = \begin{bmatrix} \mathbf{H}_{11}^{(n)\dagger} & \mathbf{H}_{21}^{(n)\dagger} \end{bmatrix} \begin{bmatrix} [c] \underline{\mathbf{y}}^{(n)} \\ \underline{\mathbf{y}}_{\text{si}} \end{bmatrix} \quad (3.27)$$

to yield

$$\tilde{\underline{\mathbf{y}}}^{(n)} = \sqrt{\rho} \underline{\mathbf{x}}^{(n)} + \tilde{\underline{\mathbf{z}}} \quad (3.28)$$

where the entries of  $\tilde{\underline{\mathbf{z}}}$  are i.i.d.  $\mathcal{CN}(0, 1)$ . We remind that the first  $m_t + m_r - m$  entries of  $\underline{\mathbf{x}}^{(n)}$  are new information bearing symbols and the last entries are equal to  $\mathbf{H}_{21}^{(n-1)} \underline{\mathbf{x}}^{(n-1)}$ . Thus, the last  $m - m_r$  entries of  $\tilde{\underline{\mathbf{y}}}^{(n)}$ , denoted  $\check{\underline{\mathbf{y}}}$ , satisfy

$$\check{\underline{\mathbf{y}}} = \sqrt{\rho} \mathbf{H}_{21}^{(n-1)} \underline{\mathbf{x}}^{(n-1)} + \check{\underline{\mathbf{z}}},$$

where  $\check{\underline{\mathbf{z}}}$  are the last  $m - m_r$  entries of  $\tilde{\underline{\mathbf{z}}}$ . Again, the receiver can linearly combine  $\underline{\mathbf{y}}^{(n-1)}$  and  $\check{\underline{\mathbf{y}}}$  as

$$\tilde{\underline{\mathbf{y}}}^{(n-1)} = \begin{bmatrix} \mathbf{H}_{11}^{(n-1)\dagger} & \mathbf{H}_{21}^{(n-1)\dagger} \end{bmatrix} \begin{bmatrix} [c] \underline{\mathbf{y}}^{(n-1)} \\ \check{\underline{\mathbf{y}}} \end{bmatrix} \quad (3.29)$$

to yield measures of  $\underline{\mathbf{x}}^{(n-1)}$  as in Eq. (3.28). Repeating this procedure for  $i = n - 1 \rightarrow 1$  results in  $m_t + m_r - m$  independent streams of measures

$$\begin{bmatrix} \tilde{\mathbf{y}}_1^{(1)} \\ \vdots \\ \tilde{\mathbf{y}}_{m_t+m_r-m}^{(1)} \end{bmatrix}, \dots, \begin{bmatrix} \tilde{\mathbf{y}}_1^{(n)} \\ \vdots \\ \tilde{\mathbf{y}}_{m_t+m_r-m}^{(n)} \end{bmatrix}.$$

The scheme above is feasible if the side information after channel use  $n$  is being conveyed by the transmitter through a neglectable number of channel uses (with respect to  $n$ , see Remark 3). In that case the receiver can construct  $m_t + m_r - m$  independent SISO channels, each with signal-to-noise ratio  $\rho$ . Thus the scheme supports a rate arbitrarily close (as  $n$  is larger) to  $(m_t + m_r - m) \log(1 + \rho)$  with zero outage probability. Note that the scheme essentially completes the singular values of the channel to 1. This is feasible since  $m - m_r < m_t$ , thus at each channel use the transmitter can transmit  $\mathbf{H}_{21}^{(i-1)} \underline{\mathbf{x}}^{(i-1)}$ , a signal of  $m - m_r$  entries, and new symbols.

The scheme presented above can be easily expanded to the case where the feedback delay is  $l$  channel uses. In that case the transmitter conveys at each channel use  $m_t + m_r - m$  new information bearing symbols and  $\mathbf{H}_{21}^{(i-l)} \underline{\mathbf{x}}^{(i-l)}$ , a linear combination of the signal that was transmitted  $l$  channel uses before. After channel use  $n$ , the transmitter would have to convey  $l$  noisy measures of the last  $l$  signals, so that the receiver could construct  $m_t + m_r - m$  independent SISO channels. This can be done in a fixed number of channel uses (see Remark 3), thus as  $n$  is larger, the transmission rate of the scheme approaches  $(m_t + m_r - m) \log(1 + \rho)$ .

**Remark 1** (Outdated feedback). *Our scheme exploits a noiseless CSI feedback system to communicate a (possibly) outdated information - the channel realization in previous channel uses. Thus, the feedback is not required to be fast, that is, no limitations on the delay time  $l$ . However, if  $l$  is smaller than the coherence time of the channel, the feedback may carry information about the current channel realization. Thus, the transmitter can exploit the up-to-date feedback to use more efficient schemes. Nevertheless, for systems with a long delay time (e.g., relatively long distance optical fibers), the channel can be regarded as non-ergodic with an outdated feedback. In these cases our scheme efficiently achieves zero outage probability.*

**Remark 2** (Simple decoding). *The scheme linearly process the received signals to construct  $m_t + m_r - m$  independent streams of measures, each with signal-to-noise  $\rho$ . This allows the decoding stage to be simple, where a SISO channel decoder can be used, removing the need for further MIMO signal processing.*

**Remark 3** (Side information measures). *For a feedback with a delay of  $l$  channel uses, the transmitter has to convey  $\mathbf{H}_{21}^{(i)} \underline{\mathbf{x}}^{(i)}$ , for each  $i = n - (l - 1), \dots, n$ , such that the receiver can extracted a vector of noisy measures with signal-to-noise ratio that is not smaller than  $\rho$ . This is feasible with a finite number of channel uses. For example, the repetition scheme can be used to convey these measures (see Section 3.7 Example 2). Suppose each  $\mathbf{H}_{21}^{(i)} \underline{\mathbf{x}}^{(i)}$  is conveyed to the receiver within  $N_{si}$  channel uses (e.g., for the repetition scheme  $N_{si} = m_t(m - m_r)$ ). By taking large enough  $n$  (with respect to  $l \cdot N_{si}$ ) one can approach the rate  $(m_t + m_r - m) \log(1 + \rho)$ .*

**Remark 4** (Uniqueness of  $\mathbf{H}_{21}$ ). *The scheme can be further improved to support even an higher data rate with zero outage probability. For example, the last  $m - m_r$  entries of the transmitted signal at the first channel use can be used to excite information bearing symbols instead of the zeros symbols. Furthermore, as was mentioned above, when  $m_t + m_r - m > 1$ ,  $\mathbf{H}_{21}^{(i)}$  is not unique; there are many  $(m - m_r) \times m_t$  matrices that complete the columns of  $\mathbf{H}_{11}^{(i)}$  into orthonormal vectors. Thus, the receiver can choose  $\mathbf{H}_{21}^{(i)}$  to be the one with the largest number of zeros rows. Now, at time  $i + 1$  the transmitter excites  $m_t + m_r - m$  new symbols and  $\mathbf{H}_{21}^{(i)} \mathbf{x}^{(i)}$ , a retransmission of  $\mathbf{x}^{(i)}$ , the transmitted signal at time  $i$ . With an appropriate choice of  $\mathbf{H}_{21}^{(i)}$ ,  $\mathbf{H}_{21}^{(i)} \mathbf{x}^{(i)}$  contains entries that are zero. Instead, these entries can contain additional new information bearing symbols. An open question is how to further enhance the data rate. One would like to exploit the feedback to approach the empirical capacity for any realization of  $\mathbf{H}_{11}$ . Note that this rate is achievable with an up-to-date feedback. Further approaching this rate with an outdated feedback system (and with zero outage probability) is left for future research.*

### 3.7 Diversity Multiplexing Tradeoff

Using multiple modes/antennas is an important means to improve performance in optical/wireless systems. The performance can be improved by increasing the transmission rate or by reducing the error probability. A coding scheme can achieve both performance gains, however there is a fundamental tradeoff between how much each can get. This tradeoff is known as the diversity-multiplexing tradeoff (DMT). The optimal tradeoff for the Rayleigh fading channel was found in [29]<sup>2</sup>. In this section we seek to find the optimal tradeoff for the Jacobi channel.

To better understand the concepts of diversity and multiplexing gains in the Jacobi channel we start with the following example.

**Example 2** (Repetition scheme). *Suppose the transmitter excites the following ( $m_t$  entries) signals in each  $m_t$  consecutive channel uses:*

$$\begin{bmatrix} \mathbf{x} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \mathbf{x} \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \mathbf{x} \end{bmatrix}.$$

*Let us make the simplifying assumptions that  $\mathbf{x}$  is an uncoded QPSK symbol and that  $m_t \leq m_r$  (similar results can be obtained also for  $m_t > m_r$  and for higher constellation sizes). We further assume that the channel realization is known at the receiver and is constant within the  $m_t$  channel uses. It can be shown that in that case the average error probability satisfies*

$$P_e(\rho) \doteq \mathbb{E}[\exp(-\frac{\rho}{2} \sum_{i=1}^{m_t} \lambda_i)] , \quad (3.30)$$

---

<sup>2</sup>See further studies on the DMT of infinite constellations and correlated MIMO channels in [30–33].

where the expectation is over  $\{\lambda_i\}_{i=1}^{m_t}$ , the eigenvalues of  $\mathbf{H}_{11}^\dagger \mathbf{H}_{11}$ . Here and throughout the rest of the paper we use  $\doteq$  to denote exponential equality, i.e.,  $f(\rho) \doteq \rho^d$  denotes

$$\lim_{\rho \rightarrow \infty} \frac{\log f(\rho)}{\log \rho} = d. \quad (3.31)$$

Now, for  $m_t + m_r \leq m$ , we can apply the joint distribution of the unordered eigenvalues of a Jacobi matrix  $\mathcal{J}(m_r, m - m_r, m_t)$ , to write

$$P_e(\rho) \doteq \frac{K_{m_t, m_r, m}^{-1}}{m_t!} \int_0^1 \cdots \int_0^1 \prod_{i=1}^{m_t} \lambda_i^{m_r - m_t} (1 - \lambda_i)^{m - m_r - m_t} e^{-\frac{\rho}{2} \lambda_i} \prod_{i < j} (\lambda_j - \lambda_i)^2 \prod_{i=1}^{m_t} d\lambda_i. \quad (3.32)$$

Note that the term

$$\prod_{1 \leq i < j \leq m_t} (\lambda_j - \lambda_i)$$

is the determinant of the Vandermonde matrix

$$\begin{bmatrix} 1 & \cdots & 1 \\ \lambda_1 & \cdots & \lambda_{m_t} \\ \vdots & & \vdots \\ \lambda_1^{m_t-1} & \cdots & \lambda_{m_t}^{m_t-1} \end{bmatrix}.$$

Thus we can write

$$\prod_{1 \leq i < j \leq m_t} (\lambda_j - \lambda_i)^2 = \sum_{\sigma_1, \sigma_2 \in S_{m_t}} (-1)^{\text{sgn}(\sigma_1) + \text{sgn}(\sigma_2)} \prod_{i=1}^{m_t} \lambda_i^{\sigma_1(i) + \sigma_2(i) - 2}, \quad (3.33)$$

where  $S_{m_t}$  is the set of all permutations of  $\{1, \dots, m_t\}$  and  $\text{sgn}(\sigma)$  denotes the signature of the permutation  $\sigma$ . Applying (3.33) into (3.32) results

$$P_e(\rho) \doteq \frac{K_{m_t, m_r, m}^{-1}}{m_t!} \sum_{\sigma_1, \sigma_2 \in S_{m_t}} (-1)^{\text{sgn}(\sigma_1) + \text{sgn}(\sigma_2)} \prod_{i=1}^{m_t} \int_0^1 \lambda_i^{m_r - m_t + \sigma_1(i) + \sigma_2(i) - 2} \times \\ \times (1 - \lambda_i)^{m - (m_r + m_t)} e^{-\frac{\rho}{2} \lambda_i} d\lambda_i. \quad (3.34)$$

It can be further shown that the right-hand-side of above is dominated (for large  $\rho$ ) by the following term

$$\frac{K_{m_t, m_r, m}^{-1}}{m_t!} \sum_{\sigma_1, \sigma_2 \in S_{m_t}} (-1)^{\text{sgn}(\sigma_1) + \text{sgn}(\sigma_2)} \prod_{i=1}^{m_t} (m_r - m_t + \sigma_1(i) + \sigma_2(i) - 2)! \left(\frac{\rho}{2}\right)^{-(m_r - m_t + \sigma_1(i) + \sigma_2(i) - 1)}. \quad (3.35)$$

Thus, for  $m_t + m_r \leq m$ , the average error probability satisfies

$$P_e(\rho) \doteq \rho^{-\sum_{i=1}^{m_t} (m_r - m_t + 2i - 1)} \quad (3.36)$$

$$\doteq \rho^{-m_r m_t}. \quad (3.37)$$

For  $m_t + m_r > m$ , by applying Lemma 1 into (3.30) we get

$$P_e(\rho) \doteq e^{-\frac{\rho(m_t+m_r-m)}{2}} \mathbb{E}[\exp(-\frac{\rho}{2} \sum_{i=1}^{m-m_r} \tilde{\lambda}_i)] ,$$

where  $\{\tilde{\lambda}_i\}_{i=1}^{m-m_r}$  are the eigenvalues of  $\mathbf{H}_{22}\mathbf{H}_{22}^\dagger$ . Thus, we can conclude that the error probability of the repetition scheme satisfies

$$P_e(\rho) \doteq \begin{cases} \rho^{-m_r m_t} & , m_t + m_r \leq m \\ e^{-\frac{\rho(m_t+m_r-m)}{2}} \rho^{-(m-m_t)(m-m_r)} & , m_t + m_r > m . \end{cases} \quad (3.38)$$

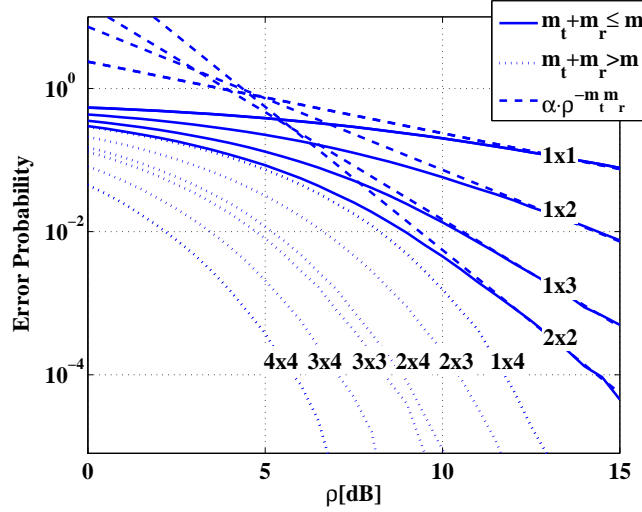


Figure 3.4: The average error probability of the repetition scheme vs.  $\rho$ , for fixed number of supported modes  $m = 4$  and various numbers of transmit×receive modes. The error probability curves are depicted in solid lines for systems satisfying  $m_t + m_r \leq m$  and in dotted lines for systems satisfying  $m_t + m_r > m$ . The dashed lines are given to emphasize the decaying order of the non-exponentially decaying curves.

In Fig. 3.4 we present the average error probability vs.  $\rho$  for  $m = 4$  and various combinations of  $m_t$ ,  $m_r$  (the error probability is symmetric in  $m_t$ ,  $m_r$ , thus all combinations of  $m_t$ ,  $m_r$  are plotted). Note the decaying order of the curves and how they turn exponentially decaying when  $m_t + m_r > m$ .

Eq. (3.38) implies that when using  $m_t$  transmit and  $m_r$  receive modes, where  $m_t + m_r \leq m$ , the exponent of the dominant term in the average error probability is  $-m_r m_t$ . Comparing to a system with a single transmit and a single receive mode, the decaying order of the average error probability is improved by a factor of  $m_r m_t$ . This gain is termed *diversity gain*. When enough modes are being addressed by the transmitter and the receiver to satisfy  $m_t + m_r > m$ , we get an average error probability that exponentially decays with  $\rho$ ; that is, an unbounded diversity gain. Thus, as more modes are being addressed, the diversity gain of the repetition scheme is greater. Since the total transmitted power is spread over all  $m$  available modes,

addressing only some modes at the receiver results in a power loss. As the number of these modes is larger, the probability for a substantial power loss is smaller; hence, smaller error probability. As the signal is transmitted from more modes, the average power in each receive mode is larger since the propagation paths are orthogonal. This is in analogy to the Rayleigh channel where as the signal passes through more (independent) faded paths, the decaying order of the error probability increases. However, it turns out that in the Jacobi channel there is a transition threshold in which enough modes are being addressed to ensure a certain received power. This results in an exponentially decaying error probability for certain rates.

Now, using multiple modes can also improve the data rate of the system. In the example above the rate is fixed,  $R(\rho) = 1/m_t$  (bps/Hz) for any  $\rho$ . Increasing the data rate with  $\rho$  to support a rate of  $R(\rho) = r \log \rho$  (bps/Hz) for some  $0 < r < 1/m_t$ , can be achieved by increasing the constellation size of the transmitted signal. In that case the data rate is improved by a factor of  $r$  comparing to a system with a single transmit and a single receive mode. This gain is termed *multiplexing gain*<sup>3</sup>. By increasing the constellation size, however, the minimum distance between the constellation points decreases, resulting an error probability with a smaller decaying order; that is, a smaller diversity gain. Thus, there is a tradeoff between diversity and multiplexing gains.

We now turn to analyze the DMT in the Jacobi model. To that end, we formalize the concepts of diversity gain and multiplexing gain by quoting some definitions from [29]<sup>4</sup>.

**Definition 4.** *Let a scheme be a family of codes  $\{\mathcal{C}(\rho)\}$  of block length  $l$ , one at each  $\rho$  level. Let  $R(\rho)$  (bps/Hz) be the rate of the code  $\mathcal{C}(\rho)$ . A scheme  $\{\mathcal{C}(\rho)\}$  is said to achieve spatial multiplexing gain  $r$  and diversity gain  $d$  if the data rate satisfies*

$$\lim_{\rho \rightarrow \infty} \frac{R(\rho)}{\log \rho} = r$$

*and the average error probability satisfies*

$$\lim_{\rho \rightarrow \infty} \frac{P_e(\rho)}{\log \rho} = -d .$$

*For each  $r$ , define  $d^*(r)$  to be the supremum of the diversity advantage achieved over all schemes.*

### 3.7.1 Case I - $m_t + m_r \leq m$

The following Theorem provides the optimal DMT of a Jacobi channel with  $m_t$ ,  $m_r$  and  $m$  satisfying  $m_t + m_r \leq m$ . In [29] it was shown that the average error probability in the high SNR regime (large  $\rho$ ) is dominated by the outage probability. Furthermore, the outage probability for a transmission rate  $R = r \log(1 + \rho)$ , where  $r$  is integer, is dominated by the probability that  $r$  singular values of the channel are 1 and the other approach zero. We

---

<sup>3</sup>The multiplexing gain in the given example is 0.

<sup>4</sup>Note that in [29] the definitions in 4 were made with respect to the average signal-to-noise ratio at each receive mode, denoted  $\bar{\rho}$ . However, since  $\bar{\rho} = \rho \mathbb{E}[tr(\mathbf{H}_{11} \mathbf{Q} \mathbf{H}_{11}^\dagger)]/m_r$ , where  $\mathbf{Q}$  is the transmitted covariance matrix, we can write  $\lim_{\bar{\rho} \rightarrow \infty} \log \bar{\rho} = \lim_{\rho \rightarrow \infty} \log \rho$ . Hence the definitions in 4 coincide with those in [29].

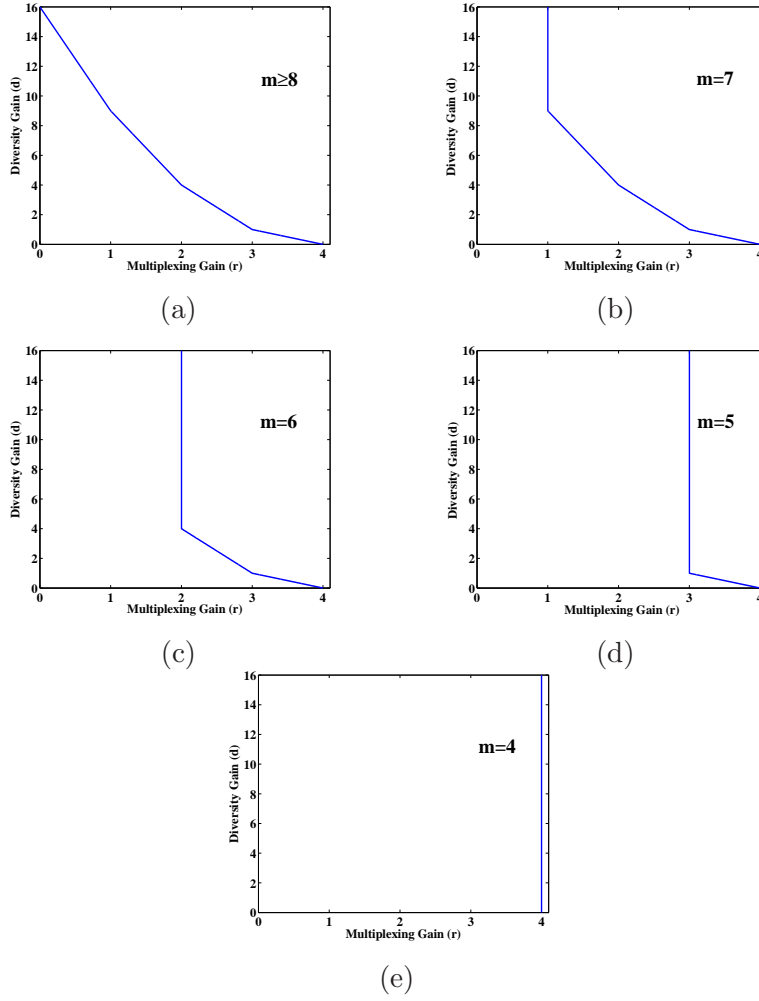


Figure 3.5: Optimal DMT curves for  $m_t = m_r = 4$ ,  $l \geq 7$  and various numbers of supported modes  $m$ .

show that the distribution of the singular values of the Jacobi and Rayleigh channels are approximately identical near 0; essentially proving that the optimal tradeoff is identical in both models.

**Theorem 4.** Suppose  $l \geq m_t + m_r - 1$ . The optimal DMT curve  $d^*(r)$  for the channel defined in (3.1), with  $m_t, m_r$  satisfying  $m_t + m_r \leq m$ , is given by the piecewise linear function that connects the points  $(k, d^*(k))$  for  $k = 0, 1, \dots, m_{\min}$ , where

$$d^*(k) = (m_t - k)(m_r - k) . \quad (3.39)$$

*Proof.* See Appendix 3.11. □

Theorem 4 suggests that for  $m_t + m_r \leq m$ , the optimal DMT curve does not depend on  $m$ . Note that  $m$  relates to the extent in which the elements of  $\mathbf{H}_{11}$  are mutually independent – the dependency is smaller as  $m$  is larger. Hence, at high SNR (large  $\rho$ ) the dependency



between the path gains has no effect on the decaying order of the average error probability. Furthermore, the optimal DMT is identical to the optimal tradeoff in the analogous Rayleigh channel (where the path gains are independent).

### 3.7.2 Case II - $m_t + m_r > m$

According to Theorem 3 a zero outage probability is achievable for rates below  $(m_t + m_r - m) \log(1 + \rho)$ . Hence, for any  $(m_t + m_r - m) > \delta > 0$  there is a scheme  $\{\mathcal{C}(\rho)\}$  with code rates  $(m_t + m_r - m - \delta) \log(1 + \rho)$  that achieves a zero outage probability; therefore, assuming  $l$  is very large, achieves an exponentially decaying error probability. In that case the discussion about diversity is no longer of relevance. Nonetheless, one can think of the gain as infinite. This reveals an interesting difference between the Jacobi and Rayleigh channels - the maximum diversity gain is “unbounded” as opposed to  $m_r m_t$  in the later case.

**Theorem 5.** *The optimal diversity multiplexing tradeoff curve  $d^*(r)$  for the channel defined in (3.1), with  $m_t, m_r$  satisfying  $m_t + m_r > m$ , is given by*

$$d^*(r) = \begin{cases} d_{\text{residual}}^*(r - (m_t + m_r - m)) & , r \geq m_t + m_r - m \\ \infty & , r < m_t + m_r - m \end{cases} \quad (3.40)$$

$d_{\text{residual}}^*(r)$  is the optimal curve for a Jacobi channel with  $m - m_r$  transmit and  $m - m_t$  receive modes.

*Proof.* At high SNR, in terms of minimal outage probability, we can take the covariance matrix of the transmitted signal to be  $Q = \mathbf{I}_{m_t}$ , see Appendix 3.11. Thus Theorem 3 can be applied: for  $r < m_t + m_r - m$  the minimal outage probability is zero hence the error probability turns exponentially decaying with  $\rho$  (assuming  $l$  is very large); for  $r \geq m_t + m_r - m$  the outage probability equals the outage probability for  $\tilde{r} = r - (m_t + m_r - m)$  in a system with  $m - m_r$  transmit and  $m - m_t$  receive modes. Noting that at high SNR the error probability is dominated by the outage probability (see Appendix 3.11) completes the proof.  $\square$

Note that  $d_{\text{residual}}^*(r)$  in Eq. (3.40) is given by Theorem 4 for any block length  $l$  satisfying  $l \geq m_t + m_r - 1$ . Fig. 3.5 depict the optimal DMT curve for  $m_t = m_r = 4$  and various numbers of supported modes  $m$ .

In the following example we try to illuminate the concept of infinite diversity gain.

**Example 3** ( $m_t = m_r = 2$ ). *We consider the  $2 \times 2$  Alamouti scheme [34]. Assuming a code block of length  $l \geq 3$  and rate  $R = r \log \rho$  (bps/Hz), the transmitter excites in each two consecutive channel uses two information bearing symbols in the following manner:*

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}, \quad \begin{bmatrix} -\mathbf{x}_2^\dagger \\ \mathbf{x}_1^\dagger \end{bmatrix}.$$

ML decoding linearly combines the received measures and yields the following equivalent scalar channels:

$$\mathbf{y}_i = \sqrt{\|\mathbf{H}_{11}\|_F^2 \rho} \mathbf{x}_i + \mathbf{z}_i, \quad \forall i = 1, 2 \quad (3.41)$$

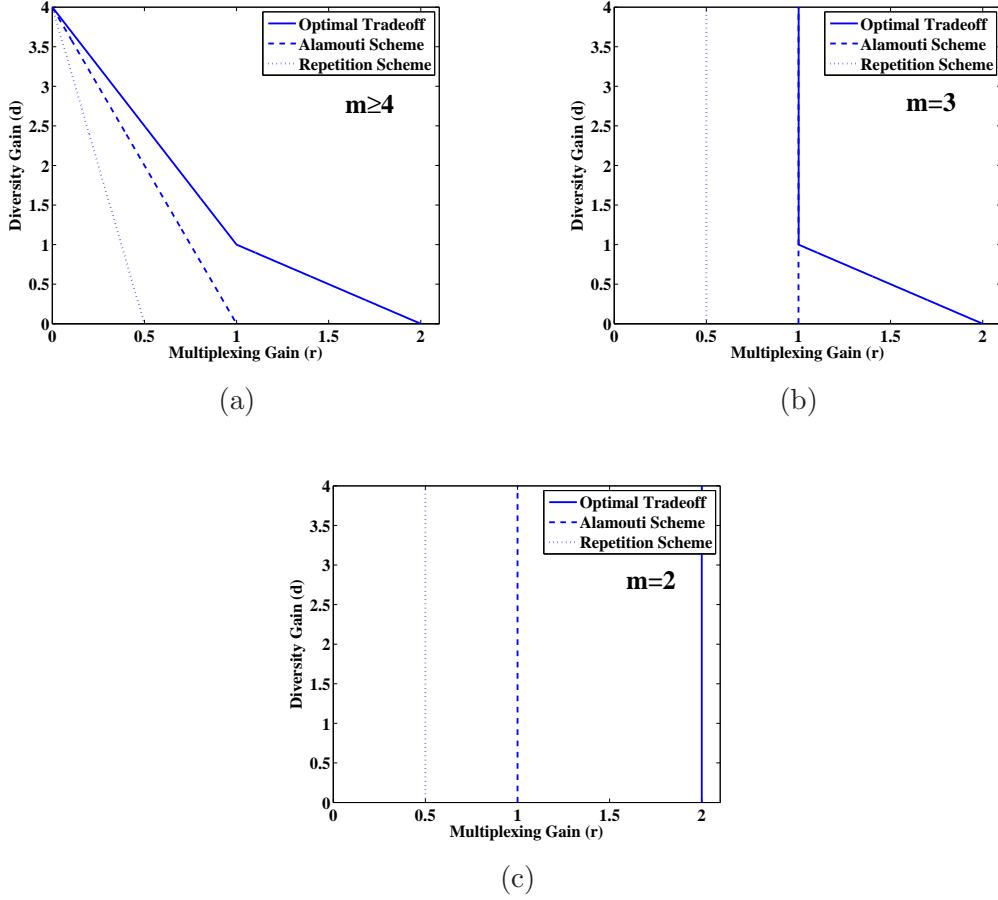


Figure 3.6: Comparison between Alamouti and the repetition scheme:  $l \geq 3$ ,  $m_t = m_r = 2$  and various numbers of supported modes  $m$ .

where each  $\mathbf{z}_i$  is i.i.d.  $\mathcal{CN}(0, 1)$  independent of  $\mathbf{x}_i$  and  $\mathbf{H}_{11}$ . The probability for an outage event is given by

$$P_{out}(2, 2, m; R) = Pr(\log(1 + \|\mathbf{H}_{11}\|_F^2 \rho) < r \log \rho) \quad (3.42)$$

$$\doteq Pr(\|\mathbf{H}_{11}\|_F^2 < \rho^{-(1-r)^+}) . \quad (3.43)$$

Now, in the Rayleigh channel  $\|\mathbf{H}_{11}\|_F^2$  is chi-square distributed with  $2m_t m_r$  degrees of freedom. In that case, as was shown in [29], the  $2 \times 2$  Alamouti scheme can achieve maximum diversity gain of 4. However, in the Jacobi channel:

- for  $m = 2$  we have  $\|\mathbf{H}_{11}\|_F^2 = 2$  ( $\mathbf{H}_{11} = \mathbf{H}$  unitary).
- for  $m = 3$  we have  $\|\mathbf{H}_{11}\|_F^2 \geq 1$  (by Lemma 1).
- for  $m \geq 4$  there is always a non-zero probability for an outage event.

Therefore, for  $m = 2$  and  $m = 3$ , for any  $r \leq 1$ , we get equivalent unfading scalar channels with strictly zero outage probability and one can think of the maximum diversity

gain as infinite. For  $m \geq 4$  it can be shown that the maximum diversity gain is 4 and the DMT curve linearly connects the points  $(1, 0)$  and  $(0, 4)$ .

In Example 2 we saw that for multiplexing gain  $r = 0$  the repetition scheme achieves a diversity gain of  $m_r m_t$  for systems satisfying  $m_t + m_r \leq m$  and an unbounded gain for systems satisfying  $m_t + m_r > m$ . Thus, for  $m \geq 4$  the maximum diversity gain of this scheme is 4 and it can be shown that the DMT curve linearly connects the points  $(1/2, 0)$  and  $(0, 4)$ . For  $m = 2$  and  $m = 3$  we get an unbounded diversity gain for any multiplexing gain below  $r = 1/2$ .

In Fig. 3.6 we compare these DMT curves to the optimal curves. Note that for  $m = 3$  the Alamouti scheme achieves the optimal DMT for  $r = 1$ .

### 3.8 Relation To The Rayleigh Model

The Jacobi fading model is defined by the transfer matrix  $\mathbf{H}_{11}$ , a truncated  $m_r \times m_t$  version of a Haar distributed  $m \times m$  unitary matrix. We shall now examine the case where  $m$  is very large with respect to  $m_t$  and  $m_r$ .

Assuming  $m_t \leq m_r$  and  $m_t + m_r \leq m$ , the statistics of the squared singular values of the Jacobi channel model follow the law of the Jacobi ensemble  $\mathcal{J}(m_r, m - m_r, m_t)$ . This ensemble can be constructed as

$$\mathbf{G}_1^\dagger \mathbf{G}_1 (\mathbf{G}_1^\dagger \mathbf{G}_1 + \mathbf{G}_2^\dagger \mathbf{G}_2)^{-1}, \quad (3.44)$$

where  $\mathbf{G}_1$  and  $\mathbf{G}_2$  are  $m_r \times m_t$  and  $(m - m_r) \times m_t$  independent Gaussian matrices. Thus, the squared singular values of  $\mathbf{H}_{11}$  share the same distribution with the eigenvalues of (3.44). Intuitively, in terms of the singularity statistics, the Jacobi channel can be viewed as an  $m_r \times m_t$  sub-channel of an  $m \times m_t$  normalized Gaussian channel. Furthermore, for  $m \gg m_r$  we have

$$\mathbf{G}_1^\dagger \mathbf{G}_1 (\mathbf{G}_1^\dagger \mathbf{G}_1 + \mathbf{G}_2^\dagger \mathbf{G}_2)^{-1} = \mathbf{G}_1^\dagger \mathbf{G}_1 \left( \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix}^\dagger \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix} \right)^{-1} \quad (3.45)$$

$$\approx \mathbf{G}_1^\dagger \mathbf{G}_1 (m \mathbb{E}[\mathbf{g}\mathbf{g}^\dagger])^{-1} \quad (3.46)$$

$$= \frac{1}{m} \mathbf{G}_1^\dagger \mathbf{G}_1, \quad (3.47)$$

where in (3.45) we applied the law of large numbers ( $\mathbf{g}$  is a vector of  $m_t$  independent components, each distributed  $\mathcal{CN}(0, 1)$ ). In the same manner, for  $m_t > m_r$ ,  $m_t + m_r \leq m$  and  $m \gg m_t$  the squared singular values of the Jacobi channel share the same distribution with the following ensemble of random matrices

$$\mathbf{G}_1 \mathbf{G}_1^\dagger (\mathbf{G}_1 \mathbf{G}_1^\dagger + \mathbf{G}_2 \mathbf{G}_2^\dagger)^{-1} \approx \frac{1}{m} \mathbf{G}_1 \mathbf{G}_1^\dagger. \quad (3.48)$$

This allows us to conclude that up to a normalizing factor the Jacobi model approaches (with  $m$ ) the Rayleigh model.

The discussion above provides an “intuitive” proof, using the law of large numbers, to the fact that the Wishart ensemble approaches the Jacobi ensemble as  $m \rightarrow \infty$ . However, this result can be proved rigorously and is well known in random matrix theory. The proof is based

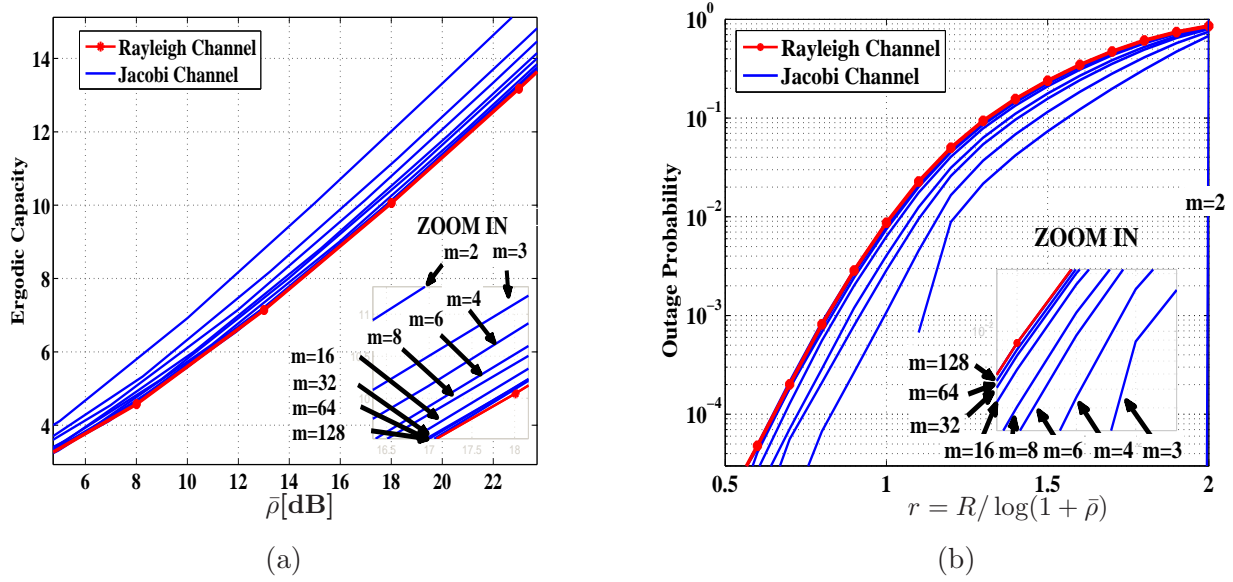


Figure 3.7: Comparing the  $2 \times 2$  Rayleigh and Jacobi models for various numbers of supported modes  $m$ .  $\bar{\rho}$  is the average SNR at each receive antenna. The ergodic capacity is given in (a) and the outage probability for  $\bar{\rho} = 20\text{dB}$  in (b).

on the following. Given a weight function  $w(x)$  and an interval  $[a, b]$ , orthogonal polynomials satisfy the relationship  $\int_a^b p_j(x)p_i(x)w(x)dx = \delta_{ji}$ . The Laguerre polynomials are orthogonal with respect to the weight function  $x^k e^{-x}$  and interval  $[0, \infty)$ , the Jacobi polynomials are orthogonal with respect to the weight function  $(1-x)^{k_1}(1+x)^{k_2}$  and interval  $(-1, 1)$  [27, 35]. Now, the joint eigenvalue densities of the Wishart and Jacobi ensembles are proportional to  $\prod_i w(\lambda_i) \prod_{i < j} |\lambda_i - \lambda_j|^2$ , where  $w(x)$  are the weight functions related to the Laguerre and (shifted) Jacobi polynomials, respectively. By taking the limit  $k_1 \rightarrow \infty$  after scaling  $x$  by  $1/k_1$ , it can be shown that the (shifted) Jacobi polynomials approach the Laguerre polynomials, which further proves above conclusion.

Now, the issue of the normalizing constant in Eq. (3.47) and (3.48),  $1/m$ , should be further explained. With fixed  $m_t, m_r$ , increasing  $m$  has two effects. One effect is power loss into the unaddressed modes. This effect is actually pretty strong, so that for a fixed  $\rho$  the channel matrix, the received SNR, and hence the capacity vanish with  $m$ . The other effect, is that with increasing  $m$  the channel matrix becomes more “random”, e.g., the matrix elements becomes statistically independent, and so the model is closer to the Rayleigh model. To compare the Jacobi model to the Rayleigh mode, we need to compensate for the power loss with increasing  $m$ , and concentrate only on the “randomness” effect. For this, we evaluate the channel characteristics (capacity, outage probability) in terms of  $\bar{\rho}$ , the average SNR at each receive mode, given by

$$\bar{\rho} = \frac{\rho}{m_r} \mathbb{E} \|\mathbf{H}_{11}\|_F^2 = \frac{\rho}{m_r} \mathbb{E} \sum_{i=1}^{m_{\min}} \lambda_i. \quad (3.49)$$

In the Rayleigh channel  $\mathbf{H}_{11}$  is Gaussian, thus  $\bar{\rho} = \rho m_t$ . For the Jacobi channel  $\bar{\rho}$  can be evaluated by applying the marginal PDF  $f_{\lambda_i}(\lambda_i)$  of the channel’s singular values. This PDF

is computed in Appendix 3.10. Nonetheless, for  $m \gg m_t, m_r$  we can apply Equations (3.47) and (3.48) to have  $\bar{\rho} \approx \rho m_t / m$ .

Following that, in Fig. 3.7 we compare the Rayleigh and Jacobi models, for  $m_t = m_r = 2$ . As  $m$  increases, the Jacobi model approaches the Rayleigh model in terms of the ergodic capacity and outage probability, as a function of  $\bar{\rho}$ . For example, with this normalization, the difference between the ergodic capacities of the Rayleigh and Jacobi models is less than 0.1dB already for  $m \geq 32$ .

### 3.9 Discussion

The Jacobi MIMO channel is defined by the transfer matrix  $\mathbf{H}_{11}$ , a truncated  $m_r \times m_t$  portion of an  $m \times m$  Haar distributed unitary matrix. By establishing the relation between the channel's singular values and the Jacobi ensemble of random matrices we derived the ergodic capacity, outage probability and optimal diversity-multiplexing tradeoff. An interesting phenomenon is observed when the parameters of the model satisfy  $m_t + m_r > m$ : for any realization of  $\mathbf{H}_{11}$ ,  $m_t + m_r - m$  singular values are 1. This results in an ergodic capacity which is at least  $m_t + m_r - m$  times the SISO capacity. In the non-ergodic scenario this enables a strictly zero outage probability and an exponentially decaying error probability ("infinite diversity") for any transmission rate below  $(m_t + m_r - m) \log(1 + \rho)$ .

In this work we considered the case where the full  $\mathbf{H}$ -matrix is drawn according to Haar measure on the group of all  $m \times m$  unitary matrices. This  $\mathbf{H}$  belongs to what is known in random matrix theory as Circular Unitary Ensemble (CUE). The eigenvalues of  $\mathbf{H}_{11}^\dagger \mathbf{H}_{11}$  then correspond to Jacobi Unitary Ensemble (JUE). It is known that, depending on certain symmetries of the system to be modeled, one can have other kind of ensembles of unitary matrices also, for example Circular Orthogonal Ensemble (COE). In this case the ensemble comprises of unitary matrices which are symmetric, and then the eigenvalues of  $\mathbf{H}_{11}^\dagger \mathbf{H}_{11}$  correspond to Jacobi Orthogonal Ensemble (JOE). Since the marginal density in this case is known, it can be used, as was done in this work, to calculate the ergodic capacity, outage probability and DMT of the induced fading model. Furthermore, just as the Rayleigh model is obtained in the limit  $m \gg m_t, m_r$  when  $\mathbf{H}$  is taken from CUE, one-sided Gaussian model will follow if  $\mathbf{H}$  is taken from COE. Thus, it will be interesting to find communication media with "reciprocity" or symmetry feature that can be naturally modeled by the Jacobi orthogonal ensemble.

The main motivation to define the Jacobi model comes from optical communication. Nonetheless, the results presented in this paper may provide conceptual insights on fading channels in other communication scenarios. The size of the unitary matrix,  $m$ , can be viewed as the number of orthogonal propagation paths in the medium, whereas  $m_t$  and  $m_r$  are the number of addressed paths at the transmitter and receiver, respectively. The Jacobi fading model can be regarded as providing statistical model for the structure of the power loss in a system, where for fixed  $m_t$  and  $m_r$ , the size of the unitary matrix  $m$  defines the "*fading measure*" of the channel. For example, when  $m$  is equal to  $m_r$ , the transfer matrix  $\mathbf{H}_{11}$  is simply composed of orthonormal columns: its elements (i.e., the path gains) are highly dependent and there is no randomness in the received power. As  $m$  becomes greater, the orthogonality of the columns and rows of  $\mathbf{H}_{11}$  fades, the dependency between the path gains

becomes weaker and the power loss in the unaddressed receive outputs increases. Indeed, when  $m$  is very large with respect to  $m_t$  and  $m_r$ , with proper normalization that compensates for the average power loss, the Jacobi fading model approaches to the Rayleigh model. The Jacobi model, thus, introduces new concepts in fading channels, providing a degree of freedom to scale the model from a unitary channel up to the Rayleigh channel.

## Acknowledgement

We wish to thank Amir Dembo and Yair Yona for interesting discussions on Lemma 1.

### 3.10 Appendix: Proof of Theorem 1

The Gaudin-Mehta method [28] is used here for proving Theorem 1.

According to (3.10), the ergodic capacity satisfies

$$C(m_t, m_r, m; \rho) = \mathbb{E}[\log \det(\mathbf{I}_{m_t} + \rho \mathbf{H}_{11}^\dagger \mathbf{H}_{11})] \quad (3.50)$$

$$= \mathbb{E}\left[\sum_{i=1}^{m_t} \log(1 + \rho \lambda_i)\right] \quad (3.51)$$

where we denote by  $\{\lambda_i\}_{i=1}^{m_t}$  the eigenvalues of  $\mathbf{H}_{11}^\dagger \mathbf{H}_{11}$ . To simplify notations let us assume  $m_t \leq m_r$  (one can simply replace  $m_t$  with  $m_r$  to obtain the proof for  $m_t > m_r$ ). Thus, we can write

$$C(m_t, m_r, m; \rho) = m_t \mathbb{E}[\log(1 + \rho \lambda_1)]. \quad (3.52)$$

Now, the joint distribution of the ordered eigenvalues  $f_\lambda(\lambda_1, \dots, \lambda_{m_t})$  is given by (3.3). The joint distribution of the *unordered* eigenvalues equals

$$\frac{1}{m_t!} f_\lambda(\lambda_1, \dots, \lambda_{m_t}),$$

thus we can compute the density of  $\lambda_1$  by integrating out  $\{\lambda_i\}_{i=2}^{m_t}$ , that is

$$f_{\lambda_1}(\lambda_1) = \int_0^1 \dots \int_0^1 \frac{1}{m_t!} f_\lambda(\lambda_1, \dots, \lambda_{m_t}) \prod_{i=2}^{m_t} d\lambda_i. \quad (3.53)$$

By taking

$$\lambda_i = \frac{1}{2}(1 - \tilde{\lambda}_i) \quad (3.54)$$

we can write

$$f_{\tilde{\lambda}_1}(\tilde{\lambda}_1) = \int_{-1}^1 \dots \int_{-1}^1 f_{\tilde{\lambda}}(\tilde{\lambda}_1, \dots, \tilde{\lambda}_{m_t}) \prod_{i=2}^{m_t} d\tilde{\lambda}_i, \quad (3.55)$$

where

$$f_{\tilde{\lambda}}(\tilde{\lambda}_1, \dots, \tilde{\lambda}_{m_t}) = \tilde{K}_{m_t, m_r, m}^{-1} \prod_{i=1}^{m_t} (1 - \tilde{\lambda}_i)^\alpha (1 + \tilde{\lambda}_i)^\beta \prod_{i < j} (\tilde{\lambda}_i - \tilde{\lambda}_j)^2, \quad (3.56)$$

and  $\alpha = m_r - m_t$ ,  $\beta = m - m_r - m_t$ . Now, the term

$$\prod_{1 \leq i < j \leq m_t} (\tilde{\lambda}_i - \tilde{\lambda}_j)$$

is the determinant of the Vandermonde matrix

$$\begin{bmatrix} 1 & \dots & 1 \\ \tilde{\lambda}_1 & \dots & \tilde{\lambda}_{m_t} \\ \vdots & & \vdots \\ \tilde{\lambda}_1^{m_t-1} & \dots & \tilde{\lambda}_{m_t}^{m_t-1} \end{bmatrix}. \quad (3.57)$$

With row operations we can transform (3.57) into the following matrix

$$\begin{bmatrix} P_0^{(\alpha, \beta)}(\tilde{\lambda}_1) & \dots & P_0^{(\alpha, \beta)}(\tilde{\lambda}_{m_t}) \\ \vdots & & \vdots \\ P_{m_t-1}^{(\alpha, \beta)}(\tilde{\lambda}_1) & \dots & P_{m_t-1}^{(\alpha, \beta)}(\tilde{\lambda}_{m_t}) \end{bmatrix}. \quad (3.58)$$

where  $P_n^{(\alpha, \beta)}(x)$  are the Jacobi polynomials [27, 8.96]. These polynomials form a complete orthogonal system in the interval  $[-1, 1]$  with respect to the weighting function  $w(x) = (1-x)^\alpha(1+x)^\beta$ , that is

$$\int_{-1}^1 w(x) P_n^{(\alpha, \beta)}(x) P_k^{(\alpha, \beta)}(x) dx = a_{k, \alpha, \beta} \delta_{kn}, \quad (3.59)$$

where the coefficients  $a_{k, \alpha, \beta}$  are given by

$$a_{k, \alpha, \beta} = \frac{2^{\alpha+\beta+1}}{2k + \alpha + \beta + 1} \binom{2k + \alpha + \beta}{k} \binom{2k + \alpha + \beta}{k + \alpha}^{-1}. \quad (3.60)$$

Thus we can write

$$\prod_{1 \leq i < j \leq m_t} (\tilde{\lambda}_i - \tilde{\lambda}_j) = C_{m_t, m_r, m} \sum_{\sigma \in S_{m_t}} (-1)^{\text{sgn}(\sigma)} \prod_{i=1}^{m_t} P_{\sigma(i)-1}^{(\alpha, \beta)}(\tilde{\lambda}_i), \quad (3.61)$$

where  $S_{m_t}$  is the set of all permutations of  $\{1, \dots, m_t\}$ ,  $\text{sgn}(\sigma)$  denotes the signature of the permutation  $\sigma$  and  $C_{m_t, m_r, m}$  is a constant picked up from the row operations on the Vandermonde matrix (3.57). By applying (3.61) into (3.56) we get

$$f_{\tilde{\lambda}}(\tilde{\lambda}_1, \dots, \tilde{\lambda}_{m_t}) = \tilde{C}_{m_t, m_r, m}^{-1} \sum_{\sigma_1, \sigma_2 \in S_{m_t}} (-1)^{\text{sgn}(\sigma_1) + \text{sgn}(\sigma_2)} \prod_{i=1}^{m_t} (1 - \tilde{\lambda}_i)^\alpha (1 + \tilde{\lambda}_i)^\beta P_{\sigma_1(i)-1}^{(\alpha, \beta)}(\tilde{\lambda}_i) P_{\sigma_2(i)-1}^{(\alpha, \beta)}(\tilde{\lambda}_i). \quad (3.62)$$

Further integrating over  $\{\tilde{\lambda}_i\}_{i=2}^{m_t}$  results

$$f_{\tilde{\lambda}_1}(\tilde{\lambda}_1) = \tilde{C}_{m_t, m_r, m}^{-1} \sum_{\sigma_1, \sigma_2 \in S_{m_t}} (-1)^{\text{sgn}(\sigma_1) + \text{sgn}(\sigma_2)} (1 - \tilde{\lambda}_1)^\alpha (1 + \tilde{\lambda}_1)^\beta \times \\ \times P_{\sigma_1(1)-1}^{(\alpha, \beta)}(\tilde{\lambda}_1) P_{\sigma_2(1)-1}^{(\alpha, \beta)}(\tilde{\lambda}_1) \prod_{i=2}^{m_t} a_{(\sigma_1(i)-1), \alpha, \beta} \delta_{\sigma_1(i) \sigma_2(i)} \quad (3.63)$$

$$= \tilde{C}_{m_t, m_r, m}^{-1} (m_t - 1)! \sum_{k=0}^{m_t-1} (1 - \tilde{\lambda}_1)^\alpha (1 + \tilde{\lambda}_1)^\beta [P_k^{(\alpha, \beta)}(\tilde{\lambda}_1)]^2 \prod_{i \neq k} a_{i, \alpha, \beta} \quad (3.64)$$

$$= \frac{1}{m_t} \sum_{k=0}^{m_t-1} a_{k, \alpha, \beta}^{-1} [P_k^{(\alpha, \beta)}(\tilde{\lambda}_1)]^2 (1 - \tilde{\lambda}_1)^\alpha (1 + \tilde{\lambda}_1)^\beta, \quad (3.65)$$

where the first equality follows from (3.59) and thus implies that  $\sigma_1(i) = \sigma_2(i)$  for all  $i$ . This results in the second equality while the third follows from (3.59) and the fact that  $f_{\tilde{\lambda}_1}(\tilde{\lambda}_1)$  must integrate to unity. Turning back to  $\lambda_1$  we get:

$$f_{\lambda_1}(\lambda_1) = \frac{1}{m_t} \sum_{k=0}^{m_t-1} b_{k, \alpha, \beta}^{-1} (P_k^{(\alpha, \beta)}(1 - 2\lambda_1))^2 \lambda_1^\alpha (1 - \lambda_1)^\beta, \quad (3.66)$$

where

$$b_{k, \alpha, \beta} = \frac{1}{2k + \alpha + \beta + 1} \binom{2k + \alpha + \beta}{k} \binom{2k + \alpha + \beta}{k + \alpha}^{-1}. \quad (3.67)$$

### 3.11 Appendix: Proof of Theorem 4

The outage probability for a transmission rate  $R$  is

$$P_{out}(m_t, m_r, m; R) = \inf_{Q: Q \succeq 0} Pr[\log \det(\mathbf{I}_{m_r} + \rho \mathbf{H}_{11} Q \mathbf{H}_{11}^\dagger) < R], \quad (3.68)$$

where the minimization is over all covariance matrices  $Q$  of the transmitted signal that satisfy the power constraints. As was already mentioned, since the statistics of  $\mathbf{H}_{11}$  is invariant under unitary permutations, the optimal choice of  $Q$ , when applying constant per-mode power constraint, is simply the identity matrix. When imposing power constraint on the total power over all modes, we can take  $Q = \mathbf{I}_{m_t}$  if  $\rho \gg 1$  since

$$P_{out}(m_t, m_r, m; R) \doteq Pr[\log \det(\mathbf{I}_{m_r} + \rho \mathbf{H}_{11} \mathbf{H}_{11}^\dagger) < R], \quad (3.69)$$

where we use  $\doteq$  to denote *exponential equality*, i.e.,  $f(\rho) \doteq \rho^d$  denotes

$$\lim_{\rho \rightarrow \infty} \frac{\log f(\rho)}{\log \rho} = d. \quad (3.70)$$

Eq. (3.69) can be proved by picking  $Q = \mathbf{I}_{m_t}$  to derive an upper bound on the outage probability and  $Q = m_t \mathbf{I}_{m_t}$  to derive a lower bound. It can be easily shown that these bounds are exponentially tight (see [29]), hence, in the scale of interest, we can take  $Q = \mathbf{I}_{m_t}$ .



Now, let the transmission rate be  $R = r \log(1 + \rho)$  and without loss of generality, let us assume that  $m_t \leq m_r$  (the outage probability is symmetric in  $m_t$  and  $m_r$ ). Since

$$\log \det(\mathbf{I}_{m_r} + \rho \mathbf{H}_{11} \mathbf{H}_{11}^\dagger) = \log \det(\mathbf{I}_{m_t} + \rho \mathbf{H}_{11}^\dagger \mathbf{H}_{11})$$

we can apply the joint distribution of the ordered eigenvalues of  $\mathbf{H}_{11}^\dagger \mathbf{H}_{11}$  to write

$$P_{out}(m_t, m_r, m; r \log(1 + \rho)) \doteq K_{m_t, m_r, m}^{-1} \int_{\mathcal{B}} \prod_{i=1}^{m_t} \lambda_i^{m_r - m_t} (1 - \lambda_i)^{m - m_r - m_t} \prod_{i < j} (\lambda_i - \lambda_j)^2 d\lambda, \quad (3.71)$$

where  $K_{m_t, m_r, m}$  is a normalizing factor and

$$\mathcal{B} = \{\underline{\lambda} : 0 \leq \lambda_1 \leq \dots \leq \lambda_{m_t} \leq 1, \prod_{i=1}^{m_t} (1 + \rho \lambda_i) < (1 + \rho)^r\}$$

is the set that describes the outage event. Letting

$$\lambda_i = \rho^{-\alpha_i} \quad (3.72)$$

for  $i = 1, \dots, m_t$  allows us to write

$$P_{out}(m_t, m_r, m; r \log(1 + \rho)) \doteq (\log \rho)^{m_t} K_{m_t, m_r, m}^{-1} \int_{\mathcal{B}} \prod_{i=1}^{m_t} \rho^{-\alpha_i(m_r - m_t + 1)} \quad (3.73)$$

$$(1 - \rho^{-\alpha_i})^{m - m_r - m_t} \prod_{i < j} (\rho^{-\alpha_i} - \rho^{-\alpha_j})^2 d\alpha. \quad (3.74)$$

Since

$$1 + \rho^{1 - \alpha_i} \doteq \rho^{(1 - \alpha_i)^+},$$

where  $(x)^+ = \max\{0, x\}$ , we can describe the set of outage events by

$$\mathcal{B} = \{\underline{\alpha} : \alpha_1 \geq \dots \geq \alpha_{m_t} \geq 0, \sum_{i=1}^{m_t} (1 - \alpha_i)^+ < r\}.$$

Now, the term  $(\log \rho)^{m_t} K_{m_t, m_r, m}^{-1}$  satisfies

$$\lim_{\rho \rightarrow \infty} \frac{\log((\log \rho)^{m_t} K_{m_t, m_r, m}^{-1})}{\log \rho} = 0, \quad (3.75)$$

thus we can write

$$P_{out}(m_t, m_r, m; r \log(1 + \rho)) \doteq \int_{\mathcal{B}} \prod_{i=1}^{m_t} \rho^{-\alpha_i(m_r - m_t + 1)} \times (1 - \rho^{-\alpha_i})^{m - m_r - m_t} \prod_{i < j} (\rho^{-\alpha_i} - \rho^{-\alpha_j})^2 d\alpha \quad (3.76)$$

$$\leq \int_{\mathcal{B}} \prod_{i=1}^{m_t} \rho^{-\alpha_i(m_r - m_t + 1)} \prod_{i < j} (\rho^{-\alpha_i} - \rho^{-\alpha_j})^2 d\alpha. \quad (3.77)$$

In [29, Theorem 4] it was shown that the right hand side of above satisfies

$$\int_{\mathcal{B}} \prod_{i=1}^{m_t} \rho^{-\alpha_i(m_r-m_t+1)} \prod_{i < j} (\rho^{-\alpha_i} - \rho^{-\alpha_j})^2 d\alpha \doteq \rho^{-f(\underline{\alpha}^*)} , \quad (3.78)$$

where

$$f(\underline{\alpha}) = \sum_{i=1}^{m_t} (2i - 1 + m_r - m_t) \alpha_i \quad (3.79)$$

and

$$\underline{\alpha}^* = \arg \inf_{\underline{\alpha} \in \mathcal{B}} f(\underline{\alpha}) . \quad (3.80)$$

By defining  $S_\delta = \{\underline{\alpha} : \alpha_i > \delta \ \forall i = 1, \dots, m_t\}$  for any  $\delta > 0$ , we can write

$$\begin{aligned} P_{out}(m_t, m_r, m; r \log(1 + \rho)) &\geq \int_{\mathcal{B} \cap S_\delta} \prod_{i=1}^{m_t} \rho^{-\alpha_i(m_r-m_t+1)} \times \\ &\quad \times (1 - \rho^{-\alpha_i})^{m-m_r-m_t} \prod_{i < j} (\rho^{-\alpha_i} - \rho^{-\alpha_j})^2 d\alpha \end{aligned} \quad (3.81)$$

$$\begin{aligned} &\geq (1 - \rho^{-\delta})^{m_t(m-m_r-m_t)} \int_{\mathcal{B} \cap S_\delta} \prod_{i=1}^{m_t} \rho^{-\alpha_i(m_r-m_t+1)} \times \\ &\quad \times \prod_{i < j} (\rho^{-\alpha_i} - \rho^{-\alpha_j})^2 d\alpha \end{aligned} \quad (3.82)$$

$$\doteq \rho^{-f(\underline{\alpha}_\delta^*)} , \quad (3.83)$$

where

$$\underline{\alpha}_\delta^* = \arg \inf_{\underline{\alpha} \in \mathcal{B} \cap S_\delta} f(\underline{\alpha}) . \quad (3.84)$$

Using the continuity of  $f$ ,  $\underline{\alpha}_\delta^*$  approaches  $\underline{\alpha}^*$  as  $\delta$  goes to zero and we can conclude that

$$P_{out}(m_t, m_r, m; r \log(1 + \rho)) \doteq \rho^{-f(\underline{\alpha}^*)} . \quad (3.85)$$

This result was obtained in [29] for the Rayleigh model. From here one can continue as was presented in [29], showing that the error probability is dominated by the outage probability at high SNR (large  $\rho$ ) for  $l \geq m_t + m_r - 1$  ([29, Lemma 5 and Theorem 2], these proofs rely on (3.85) without making any assumptions on the channel statistics, therefore are true also for the Jacobi model).

# Bibliography

- [1] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.
- [2] M. L. Mehta, *Random Matrices*. 3rd ed. New York: Academic Press, 1991.
- [3] A. Edelman and N. R. Rao, “Random matrix theory,” *Acta Numerica*, vol. 14, pp. 233–297, 2005.
- [4] P. J. Forrester, “Quantum conductance problems and the Jacobi ensemble,” *Journal of Physics A: Mathematical and General*, vol. 39, pp. 6861–6870, 2006.
- [5] G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,” *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.
- [6] I. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *European Transactions on Telecommunications*, vol. 10, pp. 585–595, 1999.
- [7] A. M. Tulino and S. Verdú, “Random matrix theory and wireless communications,” *Foundations and Trends in Communications and Information Theory*, vol. 1, pp. 1–182, June 2004.
- [8] S. Jayaweera and H. Poor, “On the capacity of multiple-antenna systems in Rician fading,” *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 1102 – 1111, May 2005.
- [9] M. Kang and M. Alouini, “Capacity of MIMO Rician channels,” *IEEE Transactions on Wireless Communications*, vol. 5, no. 1, pp. 112 – 122, January 2006.
- [10] M. K. Simon and M. S. Alouini, *Digital Communications Over Fading Channels*. New York: Wiley, 2000.
- [11] M. Nakagami, “The  $m$ -distribution: A general formula of intensity distribution of rapid fading,” *Statistical Methods in Radio Wave Propagation*. New York: Pergamon, 1960, pp. 3–36.
- [12] G. Fraidenraich, O. Leveque, and J. M. Cioffi, “On the MIMO channel capacity for the Nakagami- $m$  channel,” *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3752 – 3757, August 2008.

- [13] S. Kumar and A. Pandey, “Random matrix model for Nakagami-Hoyt fading,” *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2360–2372, May 2010.
- [14] C. W. J. Beenakker, “Random matrix theory of quantum transport,” *Review of Modern Physics*, vol. 69, pp. 731–808, July 1997.
- [15] H. Schanze, H.-J. Stöckmann, M. Martínez-Mares, and C. H. Lewenkopf, “Universal transport properties of open microwave cavities with and without time-reversal symmetry,” *Physical Review E*, vol. 71, p. 016223, January 2005.
- [16] S. H. Simon and A. L. Moustakas, “Crossover from conserving to lossy transport in circular random-matrix ensembles,” *Physical Review Letters*, vol. 96, p. 136805, April 2006.
- [17] S. Kumar and A. Pandey, “Jacobi crossover ensembles of random matrices and statistics of transmission eigenvalues,” *Journal of Physics A: Mathematical and Theoretical*, vol. 43, no. 8, p. 085001, 2010.
- [18] —, “Crossover ensembles of random matrices and skew-orthogonal polynomials,” *Annals of Physics*, vol. 326, no. 8, pp. 1877–1915, 2011.
- [19] R. W. Tkach, “Scaling optical communications for the next decade and beyond,” *Bell Labs Technical Journal*, vol. 14, no. 4, pp. 3–10, 2010.
- [20] A. R. Chraplyvy, “The coming capacity crunch,” *European Conference on Optical Communication (ECOC)*, plenary talk, 2009.
- [21] P. Winzer, “Energy-efficient optical transport capacity scaling through spatial multiplexing,” *IEEE Photonics Technology Letters*, vol. 23, no. 13, pp. 851–853, July 2011.
- [22] T. Morioka, “New generation optical infrastructure technologies: EXAT initiative towards 2020 and beyond,” in *OptoElectronics and Communications Conference (OECC)*, 2009.
- [23] 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. 16 680–96, 2011.
- [24] R. Dar, M. Feder, and M. Shtaif, “The underaddressed optical multiple-input, multiple-output channel: Capacity and outage,” *Optics Letters*, vol. 37, no. 15, pp. 3150–3152, 2012.
- [25] A. Edelman and B. D. Sutton, “The beta-Jacobi matrix model, the CS decomposition, and generalized singular value problems,” *Foundations of Computational Mathematics*, vol. 8, no. 1, pp. 259–285, 2008.
- [26] C. C. Paige and M. A. Saunders, “Towards a generalized singular value decomposition,” *SIAM Journal on Numerical Analysis*, vol. 18, no. 3, pp. 398–405, 1981.

- [27] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. New York: Academic Press, 1980, vol. 48.
- [28] M. Mehta and M. Gaudin, “On the density of eigenvalues of a random matrix,” *Nuclear Physics*, vol. 18, no. 0, pp. 420 – 427, 1960.
- [29] L. Zheng and D. N. C. Tse, “Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels,” *IEEE Transactions on Information Theory*, vol. 49, pp. 1073–1096, 2002.
- [30] R. Narasimhan, “Finite-SNR diversity-multiplexing tradeoff for correlated Rayleigh and Rician MIMO channels,” *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3965–3979, September 2006.
- [31] Z. Rezki, D. Haccoun, F. Gagnon, and W. Ajib, “Impact of spatial correlation on the finite-SNR diversity-multiplexing tradeoff,” *IEEE Transactions on Wireless Communications*, vol. 7, no. 4, pp. 1184–1189, April 2008.
- [32] S. Loyka and G. Levin, “Finite-SNR diversity-multiplexing tradeoff via asymptotic analysis of large MIMO systems,” *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 4781–4792, October 2010.
- [33] Y. Yona and M. Feder, “Fundamental limits of infinite constellations in MIMO fading channels,” *Submitted to IEEE Transaction on Information Theory, Available on arxiv.org*.
- [34] S. Alamouti, “A simple transmit diversity technique for wireless communications,” *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451 –1458, October 1998.
- [35] G. Szegő, *Orthogonal Polynomials*, 4th ed. Providence, RI: AMS Colloquium Publications, 1975.



# Chapter 4

## Properties of Nonlinear Noise in Long, Dispersion-Uncompensated Fiber Links

Ronen Dar, Meir Feder, Antonio Mecozzi and Mark Shtaif, “*Properties of nonlinear noise in long, dispersion-uncompensated fiber links*,” Optics express, vol. 21, pp. 25685–25699 (October 2013)

### 4.1 Abstract

We study the properties of nonlinear interference noise (NLIN) in fiber-optic communications systems with large accumulated dispersion. Our focus is on settling the discrepancy between the results of the Gaussian noise (GN) model (according to which NLIN is additive Gaussian) and a recently published time-domain analysis, which attributes drastically different properties to the NLIN. Upon reviewing the two approaches we identify several unjustified assumptions that are key in the derivation of the GN model, and that are responsible for the discrepancy. We derive the true NLIN power and verify that the NLIN is not additive Gaussian, but rather it depends strongly on the data transmitted in the channel of interest. In addition we validate the time-domain model numerically and demonstrate the strong dependence of the NLIN on the interfering channels’ modulation format.

### 4.2 Introduction

The modeling of nonlinear propagation in optical fibers is a key component in the efficient design of fiber-optic communications. Although computer simulations have long reached a state of maturity allowing very accurate prediction of system performance, their use is prohibitively complex in many cases of relevance, where approximate analytical models become invaluable. In a wavelength division multiplexed (WDM) environment, nonlinear propagation phenomena can be classified as either intra-channel [1], or inter-channel [2] effects. Intra-channel effects manifest themselves as nonlinear inter-symbol interference, which can in principle be eliminated by means of post-processing (such as back-propagation [3]), or

pre-distortion [4]. Inter-channel effects consist of cross-phase-modulation (XPM) and four-wave-mixing (FWM) between WDM channels, and in a complex network environment, where joint processing is prohibitively complex, distortions due to inter-channel effects are random and it is customary to treat them as noise. The chief goal of analytical models of fiber propagation is to accurately characterize this noise in terms of its statistical properties.

While early attempts of characterizing the properties of nonlinear interference noise (NLIN) in the context of fiber-communications date back to the previous millennium [5], two recent analytical approaches are of particular relevance to this paper. The first approach, which relies on analysis in the spectral domain, originated from the group of P. Poggiolini at the Politecnico di Torino [6–11] and its derivation has been recently generalized by Johannisson and Karlsson [12] and by Bononi and Serena [13]. The model generated by this approach is commonly referred to as the Gaussian noise (GN) model and its implications have already started to be addressed in a number of studies [14–16]. The second approach has been reported by Mecozzi and Essiambre [17], and it is based on a time-domain analysis. The results of the latter approach [17] are distinctly different from those of the former [6–13]. Most conspicuously, in the results of [6–13], the NLIN is treated as additive Gaussian noise and its power-spectrum is totally independent of modulation format. Conversely, the theory of Mecozzi et al. predicts a strong dependence of the NLIN variance on the modulation format, consistently with recent experimental observations [18]. It also predicts that in the presence of non-negligible intensity modulation a large fraction of NLIN can be characterized as phase noise. This property has a very important practical consequence. If NLIN indeed has a large phase-noise component, as argued in [17], then it can be canceled out easily by making use of its long temporal correlation [19], and the effective NLIN becomes much weaker than suggested by its overall variance. The consequences of this reality in terms of the predicted channel capacity have been recently studied in [19, 20].

In this paper we review the essential parts of the time-domain theory of [17], as well as those of the frequency domain GN approach. We argue that the difference between the two models results from three subtle, but very important shortcomings of the frequency domain analysis. The first is the implicit assumption that NLIN can be treated as additive noise, while ignoring its statistical dependence on the data in the channel of interest. While it is true that within the framework of a perturbation analysis NLIN can always be expressed as an additive noise term, its dependence on the channel of interest is critical. In the case of phase noise, for example, the signal of interest  $s(t)$  changes into  $s(t) \exp(i\Delta\theta)$ , and the noise  $s(t) \exp(i\Delta\theta) - s(t) \simeq i s(t) \Delta\theta$  may be uncorrelated with  $s(t)$ , but it is certainly not statistically independent of it. The second shortcoming of the frequency domain approach is the assumption that in the limit of large chromatic dispersion the electric field of the signal and the NLIN that accompanies it can be treated as a Gaussian processes whose distribution is uniquely characterized in terms of its power density spectrum. The third shortcoming that we find in the GN analysis, is the claim that non overlapping frequency components of the propagating electric field are statistically independent of each other. We show here that these components are statistically dependent in general and it is the assumption of independence that is responsible for the fact that the NLIN in [6–13] appears to be independent of modulation format. We supplement the NLIN variance obtained in the frequency domain analysis of [6] with an extra term that follows from fourth-order frequency correlations and which, as we believe, settles the discrepancy with respect to the time-domain



theory of [17].

The study contained in this paper was performed only for the case of single carrier transmission, where XPM constitutes the predominant contribution to NLIN, a fact which is confirmed by our simulations. For this reason the analytical parts of this paper focus exclusively on XPM. Moreover, in order to isolate only the NLIN caused by inter-channel nonlinear interference, we back propagate the channel of interest so as to eliminate the distortions that are induced by SPM and chromatic dispersion.

The paper is organized as follows. In Section 4.3 we review the main analytical steps of [17], occasionally recasting them in a form that emphasizes the aspects of most relevance to this paper, and supplement them by the calculation of the autocorrelation function of the nonlinear phase-noise [19]. We then review the spectral approach in Section 4.4 and explain the consequences of the assumptions on Gaussianity and statistical independence that were made in [6, 7, 12, 13]. In Sec. 4.5 we describe a numerical study that validates the analytical prediction of Secs. 4.3 and 4.4. Section 4.6 is devoted to a summary and discussion.

### 4.3 Time-domain analysis

We consider a channel of interest, whose central frequency is arbitrarily set to zero, and a single interfering channel whose central frequency is set to  $\Omega$ . Since XPM only involves two-channel interactions, the NLIN contributions of multiple WDM channels add up independently, and there is no need to conduct the initial analysis with more than a single pair. We also ignore nonlinear interactions that involve amplified spontaneous emission noise, which are negligible within the framework of a perturbation analysis such as we are conducting here. While a second-order analysis such as in [21] is possible in principle, we find the first-order approach sufficiently accurate in the context of the study conducted here. As a starting point we express the zeroth order (i.e. linear) solution for the electric field as

$$u^{(0)}(z, t) = \sum_k a_k g^{(0)}(z, t - kT) + \sum_k b_k e^{-i\Omega t + i\frac{\beta''\Omega^2}{2}z} g^{(0)}(z, t - kT - \beta''\Omega z), \quad (4.1)$$

where the superscript  $^{(0)}$  throughout the equation signifies “zeroth order”. The first sum on the right-hand-side of (4.1) represents the channel of interest, and the second sum represents the interfering channel. The symbols  $a_k$  and  $b_k$  represent the data that is carried by the  $k$ -th symbol of the two channels, respectively,  $z$  and  $t$  are the space and time coordinates,  $\beta''$  is the dispersion coefficient and  $T$  is the symbol duration. For simplicity of notation, and without loss of generality we will assume throughout this section that  $\beta''$  is negative and  $\Omega$  positive. The fundamental pulse representing an individual symbol is  $g^{(0)}(z, t) = \mathbf{U}(z)g(0, t)$ , where  $g(0, t)$  is the input waveform and  $\mathbf{U}(z) = \exp(i\frac{1}{2}\beta''z\partial_t^2)$  (with  $\partial_t$  denoting the time derivative operator) [22] is the propagation operator in the presence of chromatic dispersion. We assume that the waveform  $g(0, t)$  is normalized to unit energy, whereas the actual energy of the transmitted symbols is accounted for by the coefficients  $a_k$  and  $b_k$ . In addition it is assumed that the input waveform  $g(0, t)$  is orthogonal with respect to time shifts by an integer number of symbol durations, namely  $\int_{-\infty}^{\infty} g^*(0, t - kT)g(0, t - k'T)dt = \delta_{k,k'}$ . Owing to the unitarity of  $\mathbf{U}(z)$  this property of orthogonality is also preserved in the linearly propagated waveform  $g^{(0)}(z, t)$ .

The first order correction for the field,  $u^{(1)}(z, t)$ , is obtained by solving the nonlinear Schrödinger equation in which the nonlinear term is evaluated from the zeroth order approximation

$$\frac{\partial u^{(1)}(z, t)}{\partial z} = -\frac{i}{2}\beta''\partial_t^2 u^{(1)}(z, t) + i\gamma f(z)|u^{(0)}(z, t)|^2 u^{(0)}(z, t), \quad (4.2)$$

where  $\gamma$  is the nonlinearity coefficient and the function  $f(z)$  accounts for the loss/gain profile along the optical link [17]. It is equal to 1 in the case of perfectly uniform distributed amplification, whereas in the case of lumped amplifiers  $f(z) = \exp(-\alpha z')$ , where  $\alpha$  is the loss coefficient and  $z'$  is the difference between the point  $z$  and the position of the last amplifier that precedes it. It is assumed that only terms that contribute to the channel of interest (i.e. in the vicinity of zero frequency) are retained in the nonlinear term in (4.2). The solution to Eq. (4.2) at  $z = L$  is straightforward and it is given by

$$u^{(1)}(L, t) = i\gamma \int_0^L dz \mathbf{U}(L - z) f(z) |u^{(0)}(z, t)|^2 u^{(0)}(z, t). \quad (4.3)$$

We now focus, without loss of generality on the detection of the zeroth data symbol  $a_0$ , which is obtained by passing the received field,  $u(L, t) \simeq u^{(0)}(L, t) + u^{(1)}(L, t)$ , through a matched filter whose impulse response is proportional to  $g^{(0)}(L, T)$ . The contribution of  $u^{(0)}(L, t)$  to the output of the matched filter is  $a_0$  itself, whereas the contribution of  $u^{(1)}(L, t)$  is the estimation error  $\Delta a_0$  resulting from NLIN. It is given by

$$\Delta a_0 = \int_{-\infty}^{\infty} u^{(1)}(L, t) g^{(0)*}(L, t) dt = i\gamma \int_0^L dz f(z) \int_{-\infty}^{\infty} dt g^{(0)*}(z, t) |u^{(0)}(z, t)|^2 u^{(0)}(z, t), \quad (4.4)$$

where we have used the identity  $\mathbf{U}(L - z) g^{(0)*}(L, t) = g^{(0)*}(z, t)$ , which follows from the definition of the linear propagation operator. Substitution of the zeroth order field expression from Eq. (4.1) in Eq. (4.4) produces the result

$$\Delta a_0 = i\gamma \sum_{h,k,m} \left( a_h a_k^* a_m S_{h,k,m} + 2 a_h b_k^* b_m X_{h,k,m} \right). \quad (4.5)$$

where

$$S_{h,k,m} = \int_0^L dz f(z) \int dt g^{(0)*}(z, t) g^{(0)}(z, t - hT) g^{(0)*}(z, t - kT) g^{(0)}(z, t - mT), \quad (4.6)$$

is responsible for intra-channel interference effects, whereas

$$\begin{aligned} X_{h,k,m} = & \int_0^L dz f(z) \int dt g^{(0)*}(z, t) g^{(0)}(z, t - hT) \\ & \times g^{(0)*}(z, t - kT - \beta''\Omega z) g^{(0)}(z, t - mT - \beta''\Omega z), \end{aligned} \quad (4.7)$$

accounts for (inter-channel) XPM induced interference. Intra-channel interference involves only symbols transmitted in the channel of interest and they need not be considered as noise. It can be reduced either by performing joint decoding of a large block of symbols, or eliminated by means of back-propagation or pre-distortion. We will hence ignore the terms

proportional to  $S_{h,k,m}$  in what follows and focus on the NLIN due to XPM. Notice that given the injected pulse waveform  $g(0, t)$ , the symbol duration  $T$ , the channel spacing  $\Omega$  and the parameters of the fiber, the value of  $X_{h,k,m}$  can be found numerically. It can be seen to reduce monotonically with the walk-off between channels, where the relevant parameter is the ratio between the group velocity difference  $\beta''\Omega$  and the symbol duration  $T$ .

A very important feature in  $X_{h,k,m}$  is that it is proportional to the overlap between four temporally shifted waveforms. It is therefore reasonable to expect based on Eq. (4.7) that the largest elements of  $X_{h,k,m}$  are those for which  $h = 0$  and  $k = m$ . That is because in this situation only two temporally shifted waveforms need to overlap. We write the contribution of these terms to  $\Delta a_0$  as

$$\Delta a_{0_p} = ia_0 \left( 2\gamma \sum_m |b_m|^2 X_{0,m,m} \right) = ia_0 \theta, \quad (4.8)$$

where we define  $\theta = 2\gamma \sum_m |b_m|^2 X_{0,m,m}$ . Notice that since  $X_{0,m,m}$  is a real quantity according to Eq. (4.7),  $\theta$  is a real quantity as well and it represents a nonlinear phase rotation. This was the inspiration for using the sub-index  $p$  (as in “phase”) in the symbol  $\Delta a_{0_p}$  [23]. The first and second moments of  $\theta$  are given by

$$\langle \theta \rangle = 2\gamma \langle |b_0|^2 \rangle \sum_m X_{0,m,m}, \quad \text{and} \quad \langle \theta^2 \rangle = 4\gamma^2 \sum_{m,m'} \langle |b_m|^2 |b_{m'}|^2 \rangle X_{0,m,m} X_{0,m',m'}$$

and the variance of the phase rotation is

$$\Delta \theta^2 = \langle \theta^2 \rangle - \langle \theta \rangle^2 = 4\gamma^2 (\langle |b_0|^4 \rangle - \langle |b_0|^2 \rangle^2) \sum_m X_{0,m,m}^2. \quad (4.9)$$

where we have used the independence between different data symbols

$$\langle |b_m|^2 |b_{m'}|^2 \rangle = \langle |b_m|^2 \rangle \langle |b_{m'}|^2 \rangle (1 - \delta_{m,m'}) + \langle |b_m|^4 \rangle \delta_{m,m'}, \quad (4.10)$$

as well as their stationarity  $\langle |b_m|^n \rangle = \langle |b_0|^n \rangle$ . Equation (4.9) constitutes an extremely important result that the phase noise grows with the variance of the square amplitude of the information symbols and that it *vanishes* in the case of pure phase-modulation where  $|b_0|$  is a constant (and hence  $\langle |b_0|^4 \rangle - \langle |b_0|^2 \rangle^2 = 0$ ). This is a rather counter-intuitive result in view of the fact that upon propagation through a dispersive fiber, the intensity of the electric field appears to fluctuate randomly, independent of the way in which it is modulated ([6, 12] and see discussion related to Fig. 2 in Sec. 4.5 of this paper).

Apart from the pure phase-noise that follows from XPM between WDM channels there are additional noise contributions involving a single pulse from the channel of interest with a pair of pulses from the interfering channel. We refer to the NLIN due to these contributions as residual NLIN, so as to distinguish it from the phase NLIN that was described earlier. In general, since residual NLIN occurs in the process of temporal overlap between three or four distinct waveforms (see Eq. (4.7)) its magnitude in the presence of amplitude modulation (as in 16QAM or larger QAM constellations) is expected to be notably smaller than that of phase noise, as we demonstrate numerically in section 4.5.

A further simplification of the expression for the variance of phase-noise follows in the limit of large accumulated chromatic dispersion, which accurately characterizes the situation in most modern fiber-communications links that do not include inline dispersion compensation. In this situation the propagating waveform  $g^{(0)}(z, t)$  quickly becomes proportional to its own Fourier transform [26], namely

$$g^{(0)}(z, t) \simeq \sqrt{\frac{i}{2\pi\beta''z}} \exp\left(-\frac{it^2}{2\beta''z}\right) \tilde{g}\left(0, \frac{t}{\beta''z}\right). \quad (4.11)$$

where  $\tilde{g}(0, \omega) = \int_{-\infty}^{\infty} g(0, t) \exp(i\omega t) dt$ . Equation (4.11) simply reflects the fact that dispersion causes different frequency components of the incident signal to propagate at different velocities, so that the frequency spectrum of the injected signal is mapped into time. In this limit the coefficients  $X_{0,m,m}$  are given by

$$X_{0,m,m} = \int_{z_0}^L dz f(z) \int \frac{d\nu}{4\pi^2\beta''z} |\tilde{g}(0, \nu)|^2 \left| \tilde{g}\left(0, \nu - \Omega - \frac{mT}{\beta''z}\right) \right|^2. \quad (4.12)$$

where we defined  $\nu = t/\beta''z$ . In Eq. (4.12) we neglected the nonlinear distortion generated in the vicinity of the fiber input and defined  $z_0 \sim T^2/|\beta''| \ll L$  as the distance after which the large dispersion approximation Eq. (4.11) becomes valid. Using Eq. (4.12) we derive an approximate analytic expression for  $\Delta\theta^2$  in the case of perfectly distributed amplification. The approximation relies on the notion that the largest overlap between the two waveforms in the integrand of (4.12) occurs at a position  $z = z_m = -mT/\beta''\Omega$ . We replace the integral from  $z_0$  to  $L$  with an integral from  $-\infty$  to  $+\infty$  and approximate  $f(z)$  with  $f(z_m)$ , which is set to 1 when  $z_m \in [z_0, L]$  and to 0, otherwise. Physically this is equivalent to stating that all collision whose center is inside the region  $[z_0, L]$  are counted as complete collisions in spite of the fact that in reality some of them (those that are centered close to the edges of the fiber) are partial. Multiplying the integrand by  $z_m/z$  (which is close to unity when there is strong overlap between pulses), and changing the order of integration, we obtain

$$\begin{aligned} X_{0,m,m} &= \int \frac{d\nu}{2\pi} |\tilde{g}(0, \nu)|^2 \int_{-\infty}^{\infty} dz \frac{z_m f(z_m)}{2\pi\beta''z^2} \left| \tilde{g}\left(0, \nu - \Omega - \frac{mT}{\beta''z}\right) \right|^2 \\ &\simeq \begin{cases} \frac{1}{\beta''\Omega} & 0 \leq m \leq \frac{|\beta''\Omega|L}{T} \\ 0 & \text{otherwise} \end{cases}. \end{aligned} \quad (4.13)$$

Substitution into Eq. (4.9) yields the result

$$\Delta\theta^2 = (\langle |b_0|^4 \rangle - \langle |b_0|^2 \rangle^2) \frac{4\gamma^2 L}{|\beta''\Omega|T}. \quad (4.14)$$

The simplified expression for  $X_{0,m,m}$ , Eq. (4.13), also allows calculation of the temporal autocorrelation function of the phase noise  $R_\theta(l) = \langle \theta_n \theta_{n+l} \rangle - \langle \theta \rangle^2$ , where we use the notation  $\theta_n$  to denote the nonlinear phase rotation induced upon the  $n$ -th symbol in the channel of interest. Using Eq. (4.8) we have [19, 20]

$$R_\theta(l) = 4\gamma^2 \sum_m \sum_n \langle |b_m|^2 |b_{n+l}|^2 \rangle X_{0,m,m} X_{0,n,n}^* - \langle \theta \rangle^2 = \Delta\theta^2 \left[ 1 - \frac{|l|T}{|\beta''\Omega|L} \right]^+, \quad (4.15)$$

where  $[a]^+ = \max\{a, 0\}$ . In the case of multiple WDM channels, Eq. (4.15) generalizes to

$$R_\theta(l) = \sum_s \Delta\theta^2(\Omega_s) \left[ 1 - \frac{|l|T}{|\beta''\Omega_s|L} \right]^+, \quad (4.16)$$

where  $\Omega_s$  is the frequency separation between the  $s$ -th WDM channel and the channel of interest and the summation is over all the interfering channels. Notice that in the limit of large accumulated dispersion,  $|\beta''\Omega_s|L/T \gg 1$ , the phase noise is characterized by a very long temporal correlation. This property allows a cancelation of nonlinear phase-noise with available equalization technology [24, 25] and contributes to the achievement of higher information capacity [19, 20]. It also allows the extraction of phase noise from simulations, as we explain in Sec. 4.5.

## 4.4 Frequency domain analysis

In this section we review the approach adopted in [6–13] of analyzing in the frequency domain the interaction leading to NLIN and relate it to the analysis in Sec. 4.3.

Following [6], we assume that the transmitted symbols  $a_n$  and  $b_n$  are periodic with period  $M$ , so that  $a_{n+M} = a_n$ ,  $b_{n+M} = b_n$  and the propagating field  $u^{(0)}(z, t)$ , which is defined in Eq. (4.1) is periodic in time with a period  $MT$ . As pointed out in [6], for large enough  $M$ , the assumption of periodicity is immaterial from the physical standpoint, but facilitates calculations by allowing the representation of the signal by means of discrete frequency tones,

$$u^{(0)}(z, t) = \frac{1}{\sqrt{MT}} \left[ \sum_n \nu_n(z) e^{-i\frac{2\pi}{MT}nt} + e^{-i\Omega t} \sum_n \xi_n(z) e^{-i\frac{2\pi}{MT}nt} \right]. \quad (4.17)$$

The coefficients  $\nu_n$  represent the spectrum of the channel of interest at frequency  $\omega = 2\pi\frac{n}{MT}$ , whereas  $\xi_n$  represent the spectrum of the interfering channel at  $\omega = \Omega + 2\pi\frac{n}{MT}$ . Both  $\nu_n$  and  $\xi_n$  are zero mean random variables whose statistics depends on the transmitted symbols in a way on which we elaborate in what follows. The complex amplitude of the NLIN that the interfering channel imposes on the channel of interest is the sum of all the nonlinear interactions between triplets of individual frequency tones,

$$\Delta u(t) = 2 \sum_{lmn, m \neq n} \rho_{lmn} \nu_l \xi_m \xi_n^*, \quad (4.18)$$

where consistently with [6], the terms  $m = n$  that only contribute to a time independent phase-shift, where excluded from the summation. The factor of 2 in front of the sum in Eq. (4.18) is characteristic of XPM when the nonlinearly interacting channels are co-polarized. The coefficients  $\rho_{lmn}$  are given by [7]

$$\begin{aligned} \rho_{lmn} = & \frac{\gamma}{(MT)^{3/2}} e^{-i\frac{2\pi}{MT}(l+m-n)t} \\ & \times \frac{1 - e^{i\left(\frac{2\pi}{MT}\right)^2 \beta'' N L_s (m-n)(l-qM-n)}}{1 - e^{i\left(\frac{2\pi}{MT}\right)^2 \beta'' L_s (m-n)(l-qM-n)}} \frac{1 - e^{-\alpha L_s} e^{i\left(\frac{2\pi}{MT}\right)^2 \beta'' L_s (m-n)(l-qM-n)}}{\alpha - i\left(\frac{2\pi}{MT}\right)^2 \beta'' (m-n)(l-qM-n)}, \end{aligned} \quad (4.19)$$

where the WDM channel spacing is assumed to be  $\Omega = q\frac{2\pi}{T}$ ,  $L_s$  is the length of a single amplified span and  $N$  is the overall number of amplified spans in the system. The NLIN power is given by the square average of  $\Delta u(t)$

$$\langle |\Delta u(t)|^2 \rangle = 4 \sum_{lmnl'm'n'} \rho_{lmn} \rho_{l'm'n'}^* \langle \nu_l \nu_{l'}^* \rangle \langle \xi_m \xi_n^* \xi_{m'}^* \xi_{n'} \rangle, \quad (4.20)$$

where  $m \neq n$  and  $m' \neq n'$  and where we have made use of the fact that  $\nu_l$  and  $\xi_m$  are statistically independent for all  $l$  and  $m$  since they correspond to different WDM channels that transmit statistically independent data. Lack of correlation between different frequency tones implies that  $\langle \nu_l \nu_{l'}^* \rangle = \langle |\nu_l|^2 \rangle \delta_{ll'}$ , and the assumption of true statistical independence (which is key in obtaining the results of [6]) implies that  $\langle \xi_m \xi_n^* \xi_{m'}^* \xi_{n'} \rangle = \langle |\xi_m|^2 \rangle \langle |\xi_n|^2 \rangle \delta_{mm'} \delta_{nn'} (1 - \delta_{mn})$  (where the irrelevant cases with  $m = n$ , or  $m' = n'$  were ignored for simplicity). Equation (4.20) then simplifies to

$$\langle |\Delta u|^2 \rangle = 4 \sum_{lmn, m \neq n} |\rho_{lmn}|^2 \langle |\nu_l|^2 \rangle \langle |\xi_m|^2 \rangle \langle |\xi_n|^2 \rangle, \quad (4.21)$$

an expression that *only* depends on the mean power spectrum of the interacting channels and is totally independent of modulation format. As we now show for the case of single carrier modulation, the above assumption of statistical independence is unjustified (even as an approximation) with most of the relevant modulation formats.

We consider a generic interfering channel as in (4.1)  $x(t) = \sum_k b_k g(t - kT)$ , which is periodic as in [6] with  $b_{k+M} = b_k$ . The Fourier coefficients of  $x(t)$  are

$$\xi_n = \frac{1}{\sqrt{MT}} \int_0^{MT} x(t) e^{i\frac{2\pi}{MT}nt} dt = \tilde{g}(\omega_n) \frac{1}{\sqrt{MT}} \sum_{k=0}^{M-1} b_k e^{i\frac{2\pi}{M}kn}, \quad (4.22)$$

where  $\tilde{g}(\omega) = \int g(t) \exp(i\omega t) dt$  is the Fourier transform of  $g(t)$ ,  $\omega_n = n\frac{2\pi}{MT}$ , and the final expression on the right-hand-side follows from a straight-forward, albeit slightly cumbersome algebraic manipulation. The correlation relations between the various  $\xi_n$  are obtained by averaging the product  $\xi_n \xi_{n'}^*$  with respect to the transmitted data. In order to simplify the algebra we will assume Nyquist, sinc-shaped pulses  $g(t) = \text{sinc}(\pi t/T)$  in which case [6, 7]

$$\langle \xi_n \xi_{n'}^* \rangle = \frac{|\tilde{g}(\omega_n)|^2}{T} \langle |b_0|^2 \rangle \delta_{nn'}. \quad (4.23)$$

The restriction to Nyquist pulses ensured wide-sense stationarity for  $x(t)$  and allowed avoiding the appearance of correlations between frequency tones  $\xi_n$  and  $\xi_{n'}$  that are separated by an integer multiple of  $M$  [6]. Assuming circularly symmetric complex modulation, the central limit theorem can be applied to Eq. (4.22), implying (as argued in [6, 7]) that in the limit of large  $M$ , the coefficients  $\xi_n$  are Gaussian distributed random variables. Yet, unlike the claim made in [6, 7], the fact that the coefficients  $\xi_n$  are Gaussian and uncorrelated does not imply their statistical independence. That is because the coefficients  $\xi_n$  are Gaussian individually, but not jointly and hence their lack of correlation does not imply anything regarding the statistical dependence between them. In order to see the lack of joint Gaussianity note that if all  $\xi_n$  were jointly Gaussian then  $x(t)$  (which can be expressed as their

linear combination) would have to be Gaussian as well. Therefore, unless the data-carrying symbols  $b_n$  are themselves Gaussian distributed, the Fourier coefficients  $\xi_n$  cannot obey a jointly Gaussian distribution. We now write the fourth order correlation, which is obtained from Eq. (4.22) (again, after some algebra and for the case of Nyquist pulses)

$$\begin{aligned} \langle \xi_m \xi_n^* \xi_{m'}^* \xi_{n'} \rangle &= \frac{\langle |b_0|^2 \rangle^2}{T^2} |\tilde{g}(\omega_m)|^2 |\tilde{g}(\omega_{n'})|^2 (\delta_{m-n} \delta_{m'-n'} + \delta_{m-m'} \delta_{n-n'}) \\ &\quad + \frac{\langle |b_0|^4 \rangle - 2\langle |b_0|^2 \rangle^2}{MT^2} \mathcal{P}_{mnm'n'} \end{aligned} \quad (4.24)$$

where

$$\begin{aligned} \mathcal{P}_{mnm'n'} &= \tilde{g}(\omega_m) \tilde{g}^*(\omega_n) \tilde{g}^*(\omega_{m'}) \tilde{g}(\omega_{n'}) \\ &\quad \times (\delta_{n-n'+m'-m-M} + \delta_{n-n'+m'-m} + \delta_{n-n'+m'-m+M}). \end{aligned} \quad (4.25)$$

The first term on the right-hand-side of (4.24) is what would follow if the coefficients  $\xi_n$  were indeed statistically independent, as assumed in [6], whereas the second term reflects the deviation from this assumption. Upon substitution into Eq. (4.20) we find that the noise variance can be written as

$$\langle |\Delta u|^2 \rangle = \langle |a_0|^2 \rangle \langle |b_0|^2 \rangle^2 \chi_1 + \langle |a_0|^2 \rangle (\langle |b_0|^4 \rangle - 2\langle |b_0|^2 \rangle^2) \chi_2 \quad (4.26)$$

where

$$\chi_1 = \frac{4}{T^3} \sum_{lmn, m \neq n} |\tilde{g}(\omega_l)|^2 |\tilde{g}(\omega_m)|^2 |\tilde{g}(\omega_n)|^2 |\rho_{lmn}|^2, \quad (4.27)$$

$$\chi_2 = \frac{4}{MT^3} \sum_{lmnm'n'} |\tilde{g}(\omega_l)|^2 \mathcal{P}_{mnm'n'} \rho_{lmn} \rho_{lm'n'}^* \quad (4.28)$$

where terms with  $m = n$  or  $m' = n'$  are excluded from the summation. The first term on the right-hand-side of (4.26) is due to second-order correlations between the frequency tones and we will refer to it as the second-order noise (SON). This term coincides with the result of [6, 12] (and can be obtained by substituting Eq. (4.23) in Eq. (4.21)). The second term is absent in the calculations of [6, 12] and since it results from fourth order correlations between the frequency tones we will refer to it as fourth-order noise (FON). Consistently, we will refer to  $\chi_1$  and  $\chi_2$  as the SON and FON coefficients, respectively. Due to the delta functions in the definition of  $\mathcal{P}_{mnm'n'}$  in Eq. (4.25), the number of free indices in the summation in Eq. (4.28) is four (e.g.  $l, m, n, m'$ , in which case  $n'$  is determined by the other four and given by  $n' = n + n' - m - M$ ,  $n' = n + n' - m$ , or  $n' = n + n' - m + M$ ). Since every free index runs over  $O(M)$  values, the factor of  $1/M$  in the expression for the FON coefficient  $\chi_2$  is countered by  $M$  more summations than in Eq. (4.27) and hence  $\chi_1$  and  $\chi_2$  are of similar order of magnitude. Moreover, as we demonstrate numerically in Sec. 4.5, in the limit of distributed amplification the coefficients  $\chi_2$  is almost identical to  $\chi_1$  and they become practically indistinguishable when the frequency separation between the interfering channels grows (see Fig. 4.5). Interestingly, in the special case of purely Gaussian modulation, when the symbols  $b_k$  are circularly symmetric complex Gaussian variables,  $\langle |b_0|^4 \rangle - 2\langle |b_0|^2 \rangle^2 = 0$  and



the FON vanishes, in which case the NLIN spectrum found in [6] is exact. Consistently, we remind that this is also the only case in which  $x(t)$  is truly Gaussian distributed and the lack of correlation between different frequency tones indeed implies their statistical independence.

The last point that we address in this section is the assumption of Gaussianity in the context of NLIN in the limit of high chromatic dispersion. The argument against this assumption is similar to the argument made in the context of Gaussianity in the frequency domain. That is because in the limit of large dispersion, the signal frequency spectrum is simply mapped to the time domain. Therefore, the field becomes Gaussian point-wise, but it does not form a Gaussian process. It is in fact a general principle that a linear unitary time independent operation, such as chromatic dispersion, cannot transform a non-Gaussian process into a Gaussian one. In the absence of joint Gaussianity between all of the field samples, the power density spectrum does not sufficiently characterize the nature of NLIN.

## 4.5 Numerical validation

In order to validate the analytical results of the previous section, a set of simulations, all based on the standard split-step Fourier transform method, was performed. In order to demonstrate the principle and to be able to test the phase-noise variance predicted in Eq. (4.14) we perform all simulations for the case of perfectly distributed gain, namely where the loss coefficient  $\alpha$  is set to 0. The simulations are performed for a 500 km system over a standard single mode fiber, whose dispersion coefficient is  $\beta'' = 21 \text{ ps}^2/\text{km}$  and whose nonlinearity coefficient is given by  $\gamma = 1.3 \text{ W}^{-1}\text{km}^{-1}$ . As we are only interested in characterizing the NLIN, we did not include ASE noise in any of the simulations. In all our simulations the symbol-rate was 100 Gb/s, similarly to [27], and the channel spacing was set to 102 GHz. Nyquist pulses of a perfectly square optical spectrum (of 100 GHz width) were assumed. The number of simulated symbols in each run was 8192 and up to 500 runs (each with independent and random data symbols) were performed with each set of system parameters, so as to accumulate sufficient statistics. The data symbols of the various channels were generated independently of each other using Matlab's random number generator whose periodicity is much larger than the collective number of symbols produced in our simulations. Use of very long sequences in every run is critical in such simulations so as to achieve acceptable accuracy in view of the long correlation time of NLIN, as well as to avoid artifacts related to the periodicity of the signals that is imposed by the use of the discrete Fourier transform. In all system simulations that we present, the number of WDM channels was five, with the central channel being the channel of interest. At the receiver the channel of interest was isolated with a matched optical filter and back-propagated so as to eliminate the effects of SPM and chromatic dispersion.

### 4.5.1 Modulation format dependence

In order to demonstrate the dependence of NLIN on the modulation format we plot in Fig. 4.1 the received signal constellations in six different cases. The figures in the left column represent the case in which the *channel of interest* undergoes QPSK modulation, whereas the right column refers to the case in which the modulation of the channel of interest is 16-



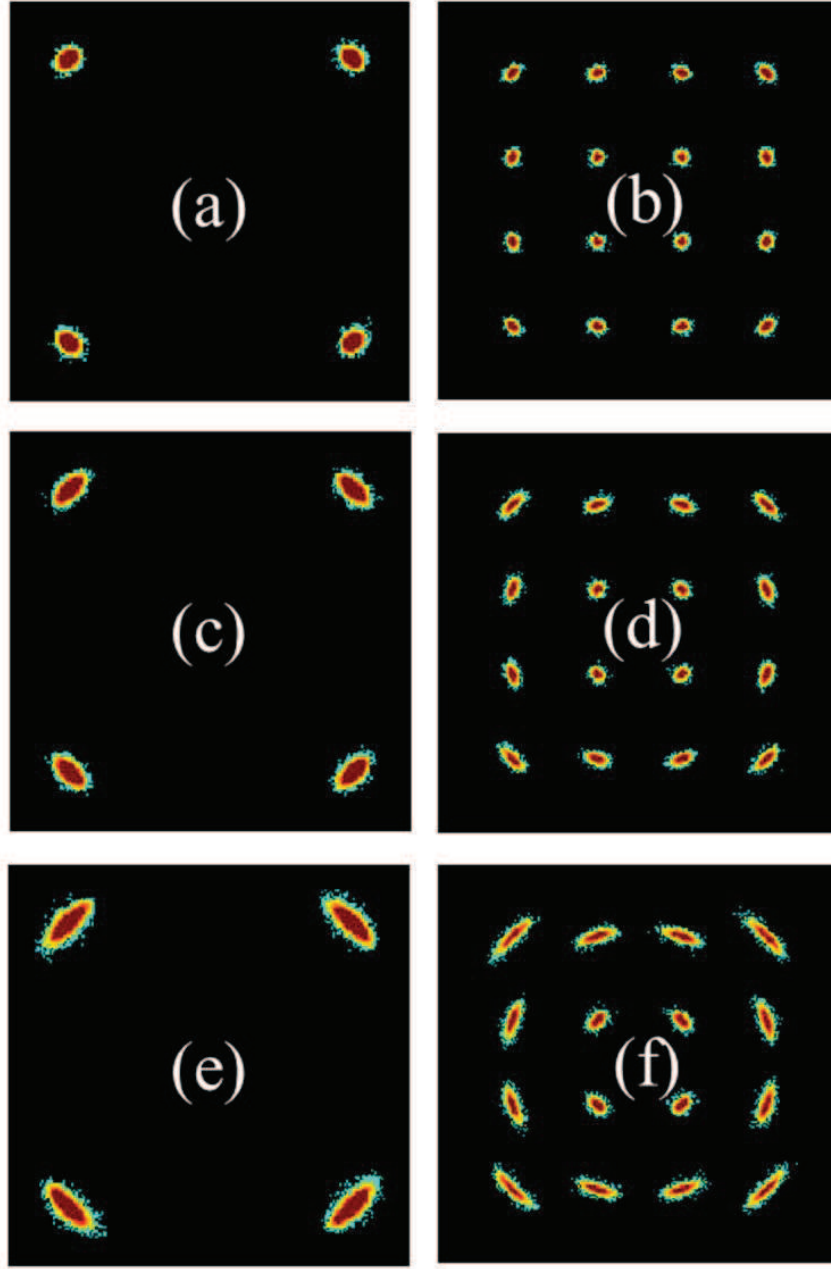


Figure 4.1: Received constellations (after compensating for the average nonlinear phase-rotation) of the channel of interest after 500 km of fiber. The per-channel power was -2dBm. The channel of interest is QPSK-modulated in the left column and 16-QAM modulated in the right column. In the top panel (Figs. (a) and (b)) the modulation of the interfering channels is QPSK. In the middle panel (Figs. (c) and (d)) the modulation of the interfering channels is 16-QAM, and in the bottom panel (Figs. (e) and (f)) the modulation of the interfering channel is Gaussian. The dominance of phase noise is evident in the middle and bottom panels, whereas in the top panel phase-noise is negligible.

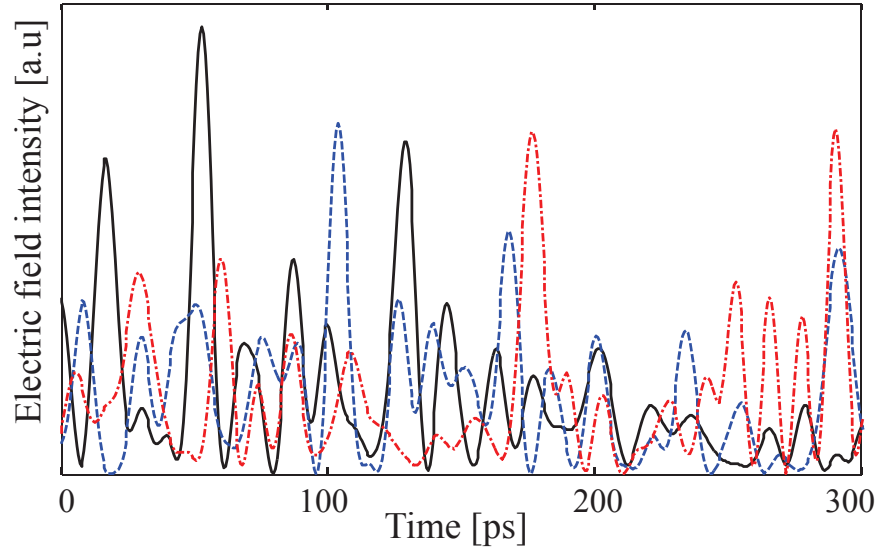


Figure 4.2: The electric field intensities of a single channel operating with Nyquist sinc-shaped pulses at a baud-rate of 100 GHz after being dispersed by 8500 ps/nm/km (equivalent to 500 km in standard fiber). The solid (black), dashed (blue), and dash-dotted (red) curves correspond to QPSK, 16-QAM and Gaussian modulation, respectively. In spite of the apparent similarity between the dispersed waveforms as demonstrated in this figure, the NLIN strongly depends on the modulation format.

QAM. The figures in the top panel correspond to the case in which the *interfering channels* are QPSK modulated, whereas the figures in the middle panel were produced with 16-QAM modulated interferers. The bottom two figures were produced in the case where the symbols of the interfering channels were Gaussian modulated. In the top panel, where the interfering channels undergo pure phase modulation, the NLIN is almost circular, albeit a small amount of phase-noise can still be observed. This small phase-noise is due to coefficients  $X_{0,k,m}$  ( $k \neq m$ ) that were neglected in Sec. 4.3 [23]. In the center and bottom panels, where the intensity of the interfering channels is modulated, the phase-noise nature of NLIN is very evident, and it is largest in the case of Gaussian modulation.

The modulation format dependence that is predicted in [17] and summarized in Sec. 4.3 is of somewhat subtle origin and is fairly counter-intuitive. As was argued correctly in [6,12,13] the electric field of the strongly dispersed signal appears fairly random independently of the modulation format as can be seen in Fig. 4.2. Moreover, as noted earlier, the point-wise distribution of the field is indeed Gaussian. Nonetheless the types of NLIN produced by the various modulations are very different as can be clearly seen in Fig. 4.1.

We note that the phase-noise nature of NLIN was not evident in the simulation results reported in [7]. While the difference between the results cannot be determined unambiguously based on the simulation details provided in [7], it may result from certain differences in the simulated system. Most importantly, the simulations in [7] do not eliminate intra-channel effects through back-propagation, as we do here, but use adaptive equalization, which may leave some of the intra-channel interference uncompensated. Furthermore, it is possible that the phase-noise that we report (which is characterized by a very long temporal correlation)

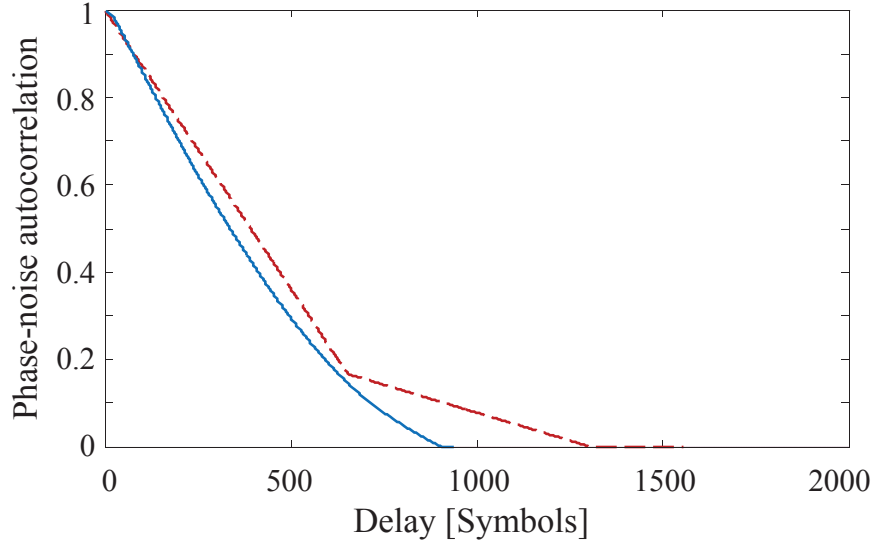


Figure 4.3: The phase-noise autocorrelation function  $R_\theta(l)$  of the phase-noise Eq. (4.16) (dashed-red) and as obtained from the simulations (solid blue) for -6dBm per-channel average power [29].

is inadvertently eliminated in the process of adaptive equalization. Additionally, some of the discrepancy could result from the fact that the system simulated in [7] assumed lumped amplification, as opposed to distributed amplification that we assumed here. It is possible that these differences explain the agreement between the simulations reported in [7] and the analytical results of the GN model.

#### 4.5.2 The variance of phase-noise and assessment of the residual NLIN

In this section we validate the analytical expression for the phase-noise variance in Eq. (4.14), and assess the residual noise. We remind that the residual noise is the part of the NLIN that does not manifest itself as phase-noise and hence remains after phase-noise cancellation. To this end, we define a procedure for extracting the phase noise from the results of the simulations. Denoting by  $r_n$  the  $n$ -th sample of the received signal (in the channel of interest and after back propagation and matched filtering) we have  $r_n = a_n \exp(i\theta_n) + \Delta a_{n,r}$ , where  $\Delta a_{n,r}$  is the residual noise. We extract  $\theta_n$  through a least-squares procedure by performing a sliding average of the quantity  $a_n^* r_n$  over a moving window of  $N = 50$  adjacent symbols. We then normalize the absolute value of the averaged quantity to 1, so as to ensure that we are only extracting phase noise. The residual noise  $\Delta a_{n,r}$  is evaluated by subtracting  $a_n \exp(i\hat{\theta}_n)$  (with  $\hat{\theta}_n$  being the estimated phase) from the received sample  $r_n$ . The width of the sliding window needs to be narrow enough relative to the correlation time of  $\theta_n$ , but broad enough to ensure meaningful statistics. Using this procedure we computed the autocorrelation function of the nonlinear phase  $\theta$ , which is plotted in Fig. 4.3 together with the analytical expression (4.16). The agreement between the analytical and numerical autocorrelation functions is self

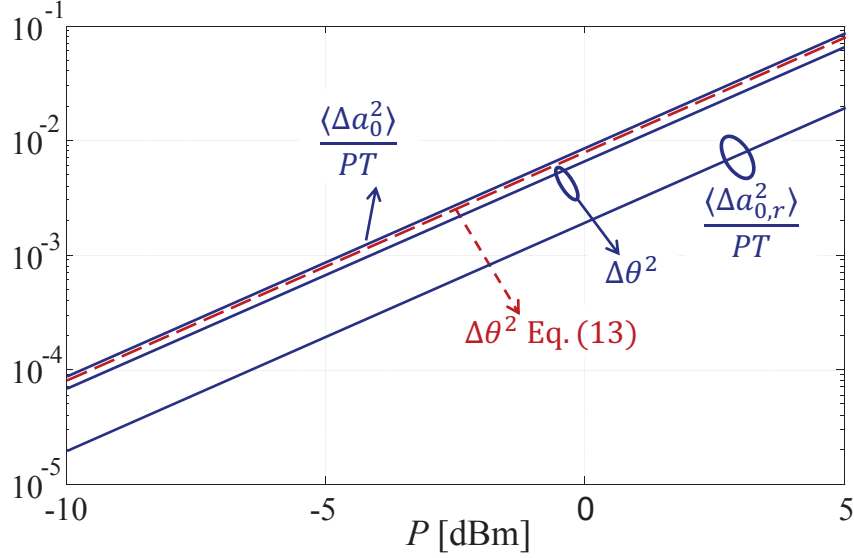


Figure 4.4: The complete NLIN variance  $\langle \Delta a_0^2 \rangle$  normalized to the average symbol energy (top solid curve), the phase noise  $\Delta \theta^2$  as obtained from the simulations (center solid curve) and the variance of the residual noise  $\langle \Delta a_r^2 \rangle$  normalized to the average symbol energy [29]. The dashed curve (red) shows the analytical result for  $\Delta \theta^2$ , Eq. (4.14). It is within 20% of the numerically obtained  $\Delta \theta^2$ .

evident. Notice that over a block of 50 symbols the autocorrelation of  $\theta_n$  drops only by 6% relative to its maximal value, thereby justifying the choice of  $N = 50$  for the moving average window. Further considerations in optimizing the window-size can be found in [19].

Figure 4.4 shows the normalized overall NLIN variance  $\langle \Delta a_0^2 \rangle / PT$ , the phase-noise variance  $\Delta \theta^2$ , and the normalized variance of the residual noise  $\langle \Delta a_{0,r}^2 \rangle / PT$ , where  $P$  is the average power in each of the interfering channels. The analytical expression for the phase-noise variance Eq. (4.14) is also plotted by the dashed red curve. All the curves in Fig. 4.4 were obtained in the case of Gaussian modulation of the data-symbols. The accuracy of the analytical result is self evident, as is the clear dominance (that was predicted in [17]) of the phase-noise component of NLIN.

### 4.5.3 The difference with respect to the NLIN power predicted by the GN model

In order to assess the error in the estimation of the NLIN power by the GN model, we compute the NLIN power, as it is predicted by the GN model and as it is predicted by the theory in Sec. 4.4. In the case where  $P$  is the average power used in each of the channels, these quantities are specified by Eq. (4.26) and given by

$$\langle |\Delta u|^2 \rangle_{\text{GN}} = P^3 \sum_s \chi_1(\Omega_s) \quad (4.29)$$

$$\langle |\Delta u|^2 \rangle_{\text{Full}} = P^3 \sum_s [\chi_1(\Omega_s) - \chi_2(\Omega_s)] + P^3 \left( \frac{\langle |b_0|^4 \rangle}{\langle |b_0|^2 \rangle^2} - 1 \right) \sum_s \chi_2(\Omega_s), \quad (4.30)$$

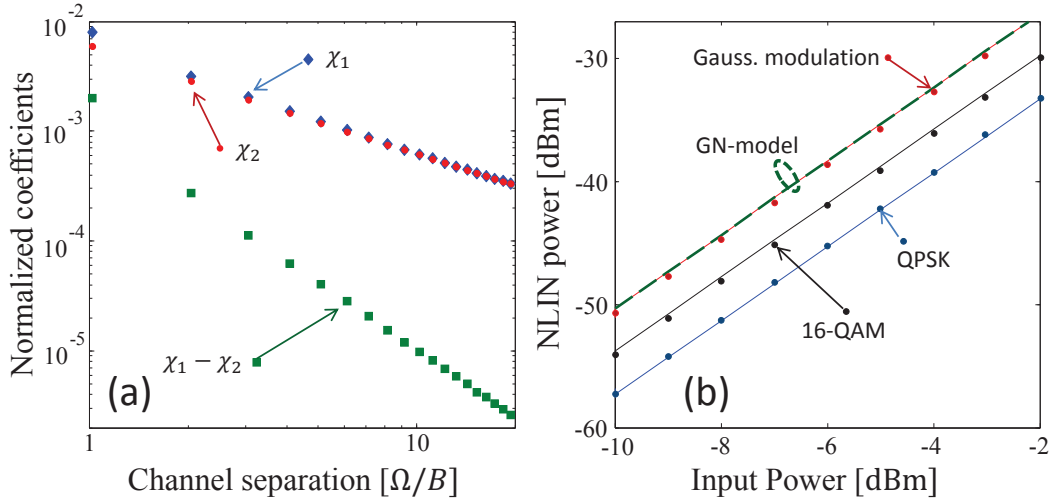


Figure 4.5: (a) The SON coefficient  $\chi_1$  (blue diamonds) and the FON coefficient  $\chi_2$  (red circles) as a function of the spacing between channels. The coefficients in the figure are normalized by  $\gamma^2 L^2 / T^3$  and hence they are unit-less. The green squares show  $\chi_1 - \chi_2$ . (b) The NLIN power versus the average power per-channel for QPSK, 16-QAM and Gaussian modulation. The symbols show the results of a split-step-simulation performed with the same parameters as in Fig. 1. The solid lines represent Eq. (4.30), whereas the dashed green line shows the prediction of the GN model Eq. (4.29). It is correct only with Gaussian modulation, but severely overestimates the actual noise in other formats.

where the summation index  $s$  runs over all neighboring channels (which are spectrally separated by  $\Omega_s$  from the channel of interest). The SON coefficient  $\chi_1$  and the FON coefficient  $\chi_2$  are plotted in Fig. 4.5a as a function of the frequency separation between the interacting channels, where the blue diamonds are used to represent  $\chi_1$  and the red circles represent  $\chi_2$ . The two coefficients are seen to be very similar to each other so that the difference between them, which is illustrated by the green squares, is significantly smaller than the coefficients themselves. The Monte-Carlo integration method [28] was deployed in order to compute the sums in Eqs. (4.27) and (4.28) in the limit of  $M \rightarrow \infty$  with the estimation error being always lower than 3%.

In Fig. 4.5b we show the NLIN power in our simulated 5-channel system in the cases of QPSK, 16-QAM, and Gaussian modulation. The solid curves show the theoretical result  $\langle |\Delta u|^2 \rangle_{\text{Full}}$  of Eq. (4.30) and the circles represent the variance obtained in the full split-step simulation. The dashed green line represents the prediction of the GN model  $\langle |\Delta u|^2 \rangle_{\text{GN}}$ , Eq. (4.29), which is correct only for Gaussian modulation. In the case of QPSK the actual NLIN power is lower by approximately 6.5dB than the prediction of the GN model. Since the NLIN powers in Fig. 4.5b include the contribution of phase-noise, the relation to the error-rate is not straightforward.

## 4.6 Discussion

Having reviewed the essential parts of the time domain model and the frequency domain GN model, we have pointed out that the difference between the models results from three unjustified assumptions in the frequency domain approach. The assumption that NLIN can be described as an additive noise term that is statistically independent on the signal, the assumption that in the large dispersion limit the electric field of the signal and the noise forms a Gaussian process that is uniquely characterized in terms of its spectrum, and the claim of statistical independence between non-overlapping tones in the spectrum of the interfering signal. We have shown that by correctly accounting for fourth-order correlations in the signals' spectrum an extra term — the FON — arises. The FON (which can be positive, or negative depending on the modulation format) needs to be added to the noise power obtained in the GN model (the SON) in order to obtain the correct overall NLIN. The inclusion of the FON recovers the dependence of the NLIN power on modulation format, a property that is absent from the existing GN model and reconciles between the frequency domain and the time domain theories. We stress that the current GN model of [6–13] which does not contain the FON term, cannot be considered a valid approximation, since with standard modulation formats (e.g QPSK, 16-QAM), the magnitude of the FON is comparable to that of the SON, which is the quantity calculated in [6, 7]. The numerical validation of the theoretical results has been performed in the case of a five-channel WDM system with idealized distributed amplification. In this case the FON term was almost identical to the SON term, implying that the error in the NLIN power predicted by the GN model is very significant.

While the study presented in this paper focused on the single polarization case, the effect of polarization multiplexing can be anticipated by considering the relevant factors. The SON part of the NLIN variance changes in the presence of polarization multiplexing by a factor of 16/27 [7], whereas it can be shown that the FON part changes by 40/81. The small difference between these factors has practically no effect on the conclusions made in this paper regarding the importance of accounting for FON. The numerical study of polarization multiplexed transmission, as well as the effects of lumped amplifications and the many other practical system parameters, is beyond the scope of this work and will be addressed in the future.

Finally, we note that when treating the NLIN as an additive, signal-independent noise process, its bandwidth appears to be comparable to that of the signal itself. Thus, one cannot take advantage of the fact that phase noise that dominates the variance of NLIN in many cases of interest is very narrow-band as we have shown here (see Eq. (4.16) and Fig. 4.3). The importance of this property of NLIN is immense as it allows cancelation of the phase-noise part of NLIN by means of available equalization technology [24, 25], such that the residual NLIN (whose variance is much smaller than that of the NLIN as a whole) determines system performance. The system consequences of this reality have been addressed in [20].

## Acknowledgement

Financial support from the Israel Science Foundation (grant 737/12) is gratefully acknowledged. Ronen Dar would like to acknowledge the support of the Adams Fellowship Program

of the Israel Academy of Sciences and Humanities, and the Yitzhak and Chaya Weinstein Research Institute for Signal Processing. The authors acknowledge useful comments by P.J. Winzer and A. Bononi.

# Bibliography

- [1] A. Mecozzi, C. B. Clausen, and M. Shtaif, “Analysis of intrachannel nonlinear effects in highly dispersed optical pulse transmission,” *IEEE Photon. Technol. Lett.* **12**, 392–394 (2000).
- [2] F. Forghieri, R.W. Tkack, A. R. Chraplyvy, “Fiber nonlinearities and their impact on transmission systems,” Ch. 8 in *Optical Fiber Telecommunications IIIA*, P. Kaminow and T. L. Koch eds. Academic Press, 1997
- [3] E. Ip, J.M. Kahn “Compensation of dispersion and nonlinear impairments using digital backpropagation,” *J. Lightwave Technol.* **26**, 3416–3425 (2008).
- [4] R. I. Killey, P. M. Watts, V. Mikhailov, M. Glick, P. Bayvel “Electronic dispersion compensation by signal predistortion using digital processing and a dual-drive Mach-Zehnder modulator,” *IEEE Photon. Technol. Lett.*, **17**, 714–716 (2005).
- [5] A. Splett, C. Kurtzke, K. Petermann, “Ultimate transmission capacity of amplified fiber communication systems taking into account fiber nonlinearities,” in 19th European Conference on Optical Communication (ECOC), ECOC Technical Digest, (1993), Paper MoC2.4.
- [6] P. Poggiolini, A. Carena, V. Curri, G. Bosco, and F. Forghieri, “Analytical modeling of nonlinear propagation in uncompensated optical transmission links,” *IEEE Photon. Technol. Lett.*, **23**, 742–744, (2011).
- [7] A. Carena, V. Curri, G. Bosco, P. Poggiolini, and F. Forghieri, “Modeling of the Impact of Nonlinear Propagation Effects in Uncompensated Optical Coherent Transmission Links,” *J. Lightwave Technol.* **30**, 1524–1539 (2012).
- [8] P. Poggiolini, G. Bosco, A. Carena, V. Curri, and F. Forghieri, “A Detailed Analytical Derivation of the GN Model of Non-Linear Interference in Coherent Optical Transmission Systems,” July 2, 2012 [Online]. Available: arXiv:1209.0394v12 [physics.optics].
- [9] P. Poggiolini, “The GN Model of Non-Linear Propagation in Uncompensated Coherent Optical Systems,” *J. of Lightwave Technol.*, **30**, 3857–3879, (2012).
- [10] A. Carena, G. Bosco, V. Curri, P. Poggiolini, M. Tapia Taiba, and F. Forghieri, “Statistical characterization of PM-QPSK signals after propagation in uncompensated fiber links,” in *Proc. ECOC*, 2010, Paper P4.07.



- [11] E. Torrenco, R. Cigliutti, G. Bosco, A. Carena, V. Curri, P. Poggiolini, A. Nespola, D. Zeolla, and F. Forghieri, "Experimental validation of an analytical model for nonlinear propagation in uncompensated optical links," in Proc. ECOC, 2011, Paper We.7.B.2.
- [12] P. Johannisson and M. Karlsson, "Perturbation analysis of nonlinear propagation in a strongly dispersive optical communication system," *IEEE J. of Lightwave Technol.*, **31**, 1273–1282 (2013).
- [13] A. Bononi, P. Serena "An alternative derivation of Johannissons regular perturbation model," July 19, 2012 [Online]. Available: arXiv:1207.4729v1 [physics.optics].
- [14] G. Bosco, R. Cigliutti, A. Nespola, A. Carena, V. Curri, F. Forghieri, Y. Yamamoto, T. Sasaki, Y. Jiang, P. Poggiolini, "Experimental investigation of nonlinear interference accumulation in uncompensated links," *IEEE Photon. Technol. Lett.* **24**, 1230 – 1232, (2012)
- [15] A. Bononi, N. Rossi, P. Serena, "Nonlinear Threshold Decrease with Distance in 112 Gb/s PDM-QPSK Coherent Systems," Proceeding of the European Conf. on Opt. Comm. (ECOC), Paper We.2.C.4, Amsterdam (2012)
- [16] S.J. Savory, "Approximations for the Nonlinear Self-Channel Interference of Channels With Rectangular Spectra," *IEEE Photon. Technol. Lett.* **25**, 961 – 964, (2013)
- [17] A. Mecozzi and R. Essiambre, "Nonlinear Shannon Limit in Pseudolinear Coherent Systems," *J. Lightwave Technol.* **30**, 2011–2024 (2012).
- [18] Xiang Liu, A. R. Chraplyvy, P. J. Winzer, R. W. Tkach, S. Chandrasekhar, "Phase-conjugated twin waves for communication beyond the Kerr nonlinearity limit," *Nature Photonics* **7**, 560–568 (2013).
- [19] Ronen Dar, Mark Shtaif, and Meir Feder, "New Bounds on the Capacity of Fiber-Optics Communications," arXiv:1305.1762v1 [physics.optics].
- [20] Ronen Dar, Mark Shtaif, and Meir Feder, "Improved Bounds on the Nonlinear Fiber-Channel Capacity," in Proc. ECOC, 2013, paper P.4.16.
- [21] S. Kumar and D. Yang, "Second-order theory for self-phase modulation and cross-phase modulation in optical fibers," *J. of Lightwave Technol.*, **23**, 2073 – 2080, (2005)
- [22] We have assumed a fixed position independent dispersion parameter  $\beta''$ . The generalization to position dependent dispersion is straightforward, but we avoid it as it considerably complicates the notation.
- [23] Although terms involving  $X_{0,k,m}$  with  $m \neq k$  also contribute to phase noise, they are much smaller than the terms proportional to  $X_{0,m,m}$  (particularly in the limit of large dispersion) and we omit them from the analysis presented in this paper.
- [24] X. Zhou, "Hardware Efficient Carrier Recovery Algorithms for Single-Carrier QAM systems," in Advanced Photonics Congress, OSA Technical Digest (online) (Optical Society of America, 2012), paper SpTu3A.1.

- [25] A. Tolmachev, I. Tselniker, M. Meltsin, I. Sigron, M. Nazarathy, “Efficient multiplier-free fpga demonstration of polar-domain multi-symbol-delay-detector (MSDD) for high performance phase recovery of 16-QAM,” Proceedings of Optical Fiber Comm. Conf. OFC 2012, Paper OMC8
- [26] A. Papoulis, “Pulse compression, fiber communications, and diffraction: a unified approach,” J. Opt. Soc. Am. A **11**, 3–13 (1994).
- [27] R.-J. Essiambre, G. Kramer, P.J. Winzer, G.J. Foschini, and B. Goebel, “Capacity Limits of Optical Fiber Networks,” J. Lightwave Technol. **28**, 662–701, (2010).
- [28] S. Asmussen, P.W. Glynn, *Stochastic simulation, algorithms and analysis*, Springer, 2007
- [29] The numerical curves in Figs. 4.3 and 4.4 reproduce the plots reported in [20], but with larger statistics.

# Chapter 5

## New Bounds on the Capacity of the Nonlinear Fiber-Optic Channel

Ronen Dar, Mark Shtaif and Meir Feder, “*New bounds on the capacity of the nonlinear fiber-optic channel*,” Optics letters, vol. 39, pp. 398–401 (January 2014)

### 5.1 Abstract

We revisit the problem of estimating the nonlinear channel capacity of fiber-optic systems. By taking advantage of the fact that a large fraction of the nonlinear interference between different wavelength-division multiplexed channels manifests itself as phase noise, and by accounting for the long temporal correlations of this noise, we show that the capacity is notably higher than what is currently assumed. This advantage is translated into the doubling of the link distance for a fixed transmission rate.

### 5.2 Introduction and results

Estimation of the fiber-optic channel capacity has come to be one of the most challenging and important problems in the field of optical communications [1–7]. Recently, its importance has grown even higher as the latest capacity estimates are being rapidly approached by the rates of commercial communications systems [8]. The difficulty in estimating the capacity of the fiber-optic channel is mostly due to the effect of fiber nonlinearity which generates complicated distortions of the transmitted optical waveforms. Perhaps the most comprehensive and familiar attempt of estimating the fiber-channel capacity to date is due to Essiambre et al. [3], where it was argued that, under plausible assumptions on network architecture, nonlinear interference between different wavelength-division multiplexed (WDM) channels must be treated as noise, which was then identified as the predominant nonlinear factor in limiting the capacity of the fiber-optic channel. This point of view has been adopted by most subsequent studies [4–6], and we also adopt it in the study presented herein.

A common feature of capacity estimates published so far is that they treat the nonlinear noise as additive, white and independent of the data transmitted on the channel of interest. In reality, in the presence of chromatic dispersion, different WDM channels propagate at

different velocities so that every symbol in the channel of interest interacts with multiple symbols of every interfering channel. Consequently, adjacent symbols in the channel of interest are disturbed by essentially the same collection of interfering pulses and therefore they are affected by nonlinearity in a highly correlated manner. In addition, as has been recently demonstrated in [6, 9], one of the most pronounced manifestations of nonlinearity is in the form of phase noise due to cross-phase modulation (XPM). The dominance of the phase noise nature of the nonlinear interference is particularly pronounced in systems with distributed gain, which is the regime where the capacity of the fiber-optic channel has been evaluated [3], and which is also assumed in the present work.

We demonstrate in what follows that by taking advantage of the long temporal correlations that allow the cancelation of nonlinear phase noise, it is possible to communicate at a higher rate than predicted in [3], or equivalently (almost) double the distance achievable at a given rate of communications. We stress that practical methods for canceling the nonlinear phase noise are not discussed in this paper as we are only interested in the capacity implications. We also note that nonlinear phase noise is canceled inadvertently in coherent optical systems where an appropriately fast phase tracking algorithm is deployed.

We start by expressing the received signal samples after coherent detection and matched filtering as

$$y_j = x_j \exp(i\theta_j) + n_j^{\text{NL}} + n_j, \quad (5.1)$$

where  $j$  is the time index and the term  $n_j^{\text{NL}}$  accounts for all nonlinear noise contributions that do not manifest themselves as phase noise. As in [3–6] we assume that the samples  $n_j^{\text{NL}}$  are zero-mean statistically independent complex Gaussian variables with variance  $\sigma_{\text{NL}}^2$ . A similar assumption holds for the amplified spontaneous emission (ASE) samples  $n_j$ , whose variance is denoted by  $\sigma_{\text{ASE}}^2$ . All three noise contributions  $\theta_j$ ,  $n_j^{\text{NL}}$ , and  $n_j$  are assumed to be statistically independent of each other. All of the above assumptions, regarding the whiteness of  $n_j^{\text{NL}}$  and  $n_j$ , the statistical independence of all noise contributions and the Gaussianity of  $n_j^{\text{NL}}$  constitute a **worst case** in terms of the resultant capacity [10, Ch. 10] and hence they are in accord with our goal of deriving a capacity lower bound. Finally, consistently with what is suggested by the analysis in [6], we will also assume that  $\theta_j$  is a Gaussian distributed variable and its variance will be denoted by  $\sigma_\theta^2$ , whose expression can be found in [6, 9].

For arriving at an analytical lower bound for the capacity, we assume that the nonlinear phase noise  $\theta_j$  is blockwise constant. In other words, it is assumed that the noise  $\theta_j$  does not change at all within a block of  $N$  symbols and then in the subsequent block it changes in a statistically independent manner. The assumption that  $\theta_j$  does not change within a block is consistent with the very long temporal correlations of the phase noise, as was demonstrated in [9, 11, 12]. The assumption of statistical independence of  $\theta_j$  in adjacent blocks is again a worst-case scenario which is in accord with our interest in a lower bound.

The capacity of the block-wise independent phase noise channel (5.1) is given by

$$C = \frac{1}{N} \sup_{\underline{p}_X} I(\underline{X}; \underline{Y}), \quad (5.2)$$

where  $\underline{X}$  and  $\underline{Y}$  are column random vectors representing a block of  $N$  channel inputs and outputs, respectively, in which the phase noise is constant. The term  $I(\underline{X}; \underline{Y}) = h(\underline{Y}) -$

$h(\underline{Y}|\underline{X})$  is the mutual information between the channel's input and output, where  $h(\cdot)$  is the differential entropy. The supremum is over all input distributions satisfying the power constraint  $\mathbb{E}[\|\underline{X}\|^2] = NP$ . In order to obtain a lower bound on the capacity we assign  $\underline{X}$  a circular Gaussian distribution with statistically independent elements. Note that with this input distribution, and taking into account the fact that  $n_j^{\text{NL}}$  and  $n_j$  are uncorrelated with the input signal [6,9], our assumption that these quantities are white circular Gaussian constitutes a worst case scenario from the standpoint of the resultant capacity [10, Ch. 10, Ex. 1]. In this case  $\underline{Y}$  is also a circularly symmetric complex Gaussian vector with differential entropy  $h(\underline{Y}) = N \log_2(\pi e(P + \sigma_{\text{eff}}^2))$ , where  $\sigma_{\text{eff}}^2$  is the variance of the effective additive noise, i.e.,  $\sigma_{\text{eff}}^2 = \sigma_{\text{ASE}}^2 + \sigma_{\text{NL}}^2$ . The conditional distribution of  $\underline{Y}$  given  $\underline{X}$  is obviously not Gaussian (see Eq. (5.1)), but since the Gaussian distribution maximizes the differential entropy of a vector of zero-mean random variables with a given covariance matrix [10, Ch. 9, Theorem 9.6.5], the differential entropy  $h(\underline{Y}|\underline{X})$  satisfies

$$h(\underline{Y}|\underline{X}) = \mathbb{E}_{\underline{x}}(h(\underline{Y}|\underline{X} = \underline{x})) \quad (5.3)$$

$$\leq \frac{1}{2} \mathbb{E}_{\underline{x}}(\log_2 \det(2\pi e Q_{\hat{\underline{Y}}|\underline{X}=\underline{x}})) , \quad (5.4)$$

where  $\hat{\underline{Y}} = \begin{bmatrix} \Re(\underline{Y}) \\ \Im(\underline{Y}) \end{bmatrix}$  and  $Q_{\hat{\underline{Y}}|\underline{X}=\underline{x}}$  is the covariance matrix of  $\hat{\underline{Y}}$  given  $\underline{X} = \underline{x}$ . By applying some algebraic manipulations the determinant of  $Q_{\hat{\underline{Y}}|\underline{X}=\underline{x}}$  can be shown to satisfy

$$\det(Q_{\hat{\underline{Y}}|\underline{X}=\underline{x}}) = \left(\frac{\sigma_{\text{eff}}^2}{2}\right)^{2N} (1 + 2\frac{\|\underline{x}\|^2}{\sigma_{\text{eff}}^2} \sigma_c^2) (1 + 2\frac{\|\underline{x}\|^2}{\sigma_{\text{eff}}^2} \sigma_s^2) , \quad (5.5)$$

where the terms  $\sigma_c^2 = 0.5(1 - e^{-\sigma_\theta^2})^2$  and  $\sigma_s^2 = 0.5(1 - e^{-2\sigma_\theta^2})$  are the variances of  $\cos(\theta)$  and  $\sin(\theta)$ , respectively, and their calculation relies on the Gaussianity of  $\theta_j$ . Note that throughout this paper the Gaussianity assumption of the phase noise is needed only here, for calculating  $\sigma_c^2$  and  $\sigma_s^2$ . Finally, by plugging (5.5) into (5.4), the following capacity lower bound is obtained

$$\begin{aligned} C &\geq \log_2 \left( 1 + \frac{P}{\sigma_{\text{eff}}^2} \right) \\ &\quad - \frac{1}{2N} \mathbb{E}_v \left\{ \log_2 \left( 1 + v \sigma_c^2 \frac{P}{\sigma_{\text{eff}}^2} \right) \right\} \\ &\quad - \frac{1}{2N} \mathbb{E}_v \left\{ \log_2 \left( 1 + v \sigma_s^2 \frac{P}{\sigma_{\text{eff}}^2} \right) \right\} , \end{aligned} \quad (5.6)$$

where the symbol  $\mathbb{E}_v$  stands for ensemble averaging with respect to a standard Chi-square distributed variable  $v$  with  $2N$  degrees of freedom. Notice that the first line on the right-hand-side of Eq. (5.6) follows from treating the nonlinear noise as white circular Gaussian – similarly to the analysis in [3]. Yet, the difference is that in our case  $\sigma_{\text{NL}}^2$  corresponds only to the part of the nonlinear noise that does not manifest itself as phase noise and hence it is smaller than the nonlinear noise that is accounted for in [3]. The effect of phase noise on the capacity is captured in our case by the bottom two lines on the right-hand-side of (5.6). This capacity loss, which may be viewed as a rate reduction needed for tracking the phase noise, vanishes when the phase exhibits very long term correlations (i.e., when  $N \rightarrow \infty$ ).

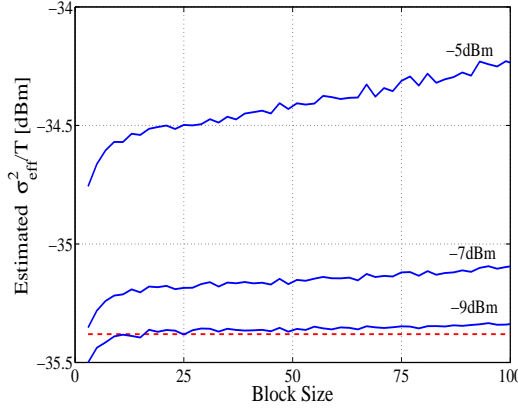


Figure 5.1: The numerically estimated  $\sigma_{\text{eff}}^2$  (normalized by  $T$ ) vs. block-size in a 500km link for input average power levels of -9dBm, -7dBm and -5dBm. Red dashed line shows  $\sigma_{\text{ASE}}^2/T$ . Due to insufficient statistics for small values of  $N$ , the estimated  $\sigma_{\text{eff}}^2$  grows rapidly with block-size. Then, when the accumulated statistics is sufficient, the growth is much slower and it is due to the fact that phase fluctuations inflate the estimated  $\sigma_{\text{eff}}^2$ .

We have performed a set of numerical simulations in order to extract the variances  $\sigma_{\theta}^2$  and  $\sigma_{\text{eff}}^2$  and to obtain lower bounds on the capacity of the nonlinear fiber channel. The simulations were performed using the parameters of a standard single mode fiber; dispersion  $D = 17$  ps/nm/km, attenuation of 0.2 dB/km, nonlinear coefficient  $\gamma = 1.27$  W<sup>-1</sup>km<sup>-1</sup> and signal wavelength  $\lambda_0 = 1.55$   $\mu$ m. Perfectly distributed and quantum limited (i.e. fully inverted) amplification with spontaneous emission factor  $n_{\text{sp}} = 1$  was assumed. Sinc-shaped pulses with a perfectly square 100 GHz wide spectrum were used for transmission and the spacing between adjacent WDM channels was 102 GHz (i.e leaving a 2 GHz guard band). The number of simulated WDM channels was 5, with the central channel being the channel of interest. All of the above assumptions are identical to those made by Essiambre et al. in the computation of the capacity lower bound reported in [3]. The number of simulated symbols in each run was 8192 for the 500km system and 16384 for the 1000km and 2000km systems. Up to 500 runs (each with independent and random data symbols) were performed with each set of system parameters, so as to accumulate sufficient statistics. We assumed a circularly symmetric complex Gaussian distribution of points in the transmitted constellation. This constellation was used to derive our capacity lower bound (5.6). At the receiver, the central channel was filtered out with a perfectly square filter (which is also the matched filter with sinc pulses) and back-propagated ideally (using the same step-size criteria as in the forward propagation). Then, the signal was optimally sampled and analyzed. As in [3], all simulations have been performed with the scalar nonlinear Schrödinger equation and correspond to a single polarization. Consideration of the polarization multiplexed case, where cross-polarization phenomena are directly accounted for in the simulations and in the analysis, is left for future study.

In order to extract the angle  $\theta$  we exploit the fact that the nonlinear phase noise changes very slowly on the scale of the symbol duration, and apply the least-squares estimation procedure. Namely, we evaluated  $\exp(i\theta)$  by averaging the variable  $x_j^* y_j$  over  $N = 50$

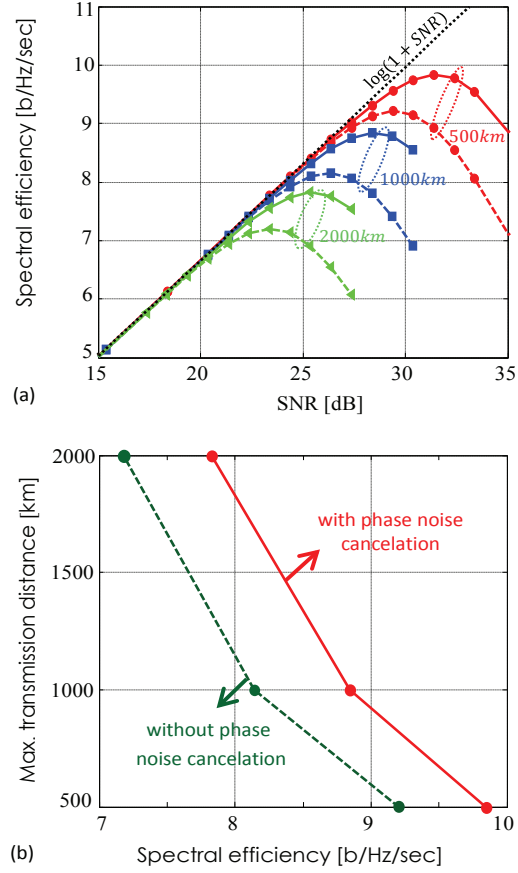


Figure 5.2: (a) Capacity lower bound vs. linear SNR for 500km (red dots), 1000km (blue squares), and 2000km (green triangles). Dashed curves result from treating the entire non-linear noise as noise. Solid curves represent the new bounds derived here. Dotted curve represents the Shannon limit  $\log_2(1 + \text{SNR})$ . (b) The maximum achievable transmission distance as a function of spectral efficiency with (solid) and without (dashed) phase noise cancellation.

adjacent symbols and then normalizing the absolute value of the averaged quantity to 1, so as to ensure that we are only extracting phase noise. The estimate of  $\exp(i\theta)$  will be denoted by  $\exp(i\hat{\theta})$  in what follows. We then subtracted  $x_j \exp(i\hat{\theta})$  from  $y_j$  to obtain  $n_j^{\text{NL}} + n_j$  and to evaluate  $\sigma_{\text{eff}}^2$ . Note that the choice of  $N$  affects the estimated noise variance in two ways. On the one hand, the estimation of  $\sigma_{\text{eff}}^2$  improves as  $N$  increases (the mean square error of the estimation is proportional to  $N^{-1}$ ). On the other hand, the assumption of constant phase noise becomes less accurate as  $N$  increases. As a result the variations of the phase noise inflate the estimate of  $\sigma_{\text{eff}}^2$  and reduce the tightness of our capacity lower bound. Fig. 5.1 shows the dependence of the estimated value of  $\sigma_{\text{eff}}^2$  on the assumed block-size  $N$  for several values of average signal power per-channel. The various curves share an important and very distinct feature. In all cases, the estimated value of  $\sigma_{\text{eff}}^2$  grows with  $N$  at small  $N$  values and then it abruptly transitions to a much slower rate of growth. The fast growth in



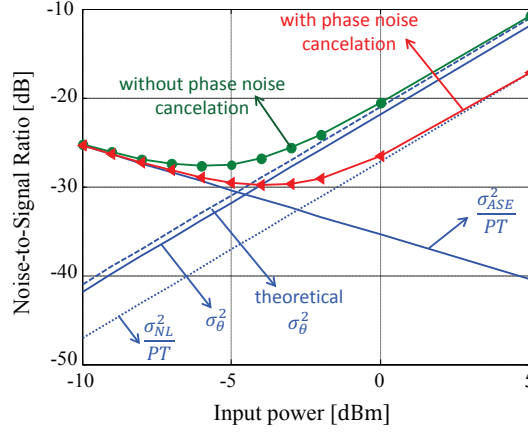


Figure 5.3: Noise-to-signal ratio vs. average input power in a 500km link. Decreasing solid line shows  $\sigma_{\text{ASE}}^2/(PT)$ , increasing solid line shows  $\sigma_{\theta}^2$ . Dashed line is the theoretical expression for  $\sigma_{\theta}^2$  found in [6,9]. Dotted line is  $\sigma_{\text{NL}}^2/(PT)$ . Triangles and dots show  $\sigma_{\text{eff}}^2/(PT)$  with and without phase noise cancellation, respectively.

the first stage is due to the lack of sufficient statistics at small  $N$  values, whereas the slow growth in the second stage is due to the slow variations of the phase noise whose significance increases with increasing block-size. Our choice of  $N = 50$  is always higher than the value of  $N$  that corresponds to the transition between the two growth rates, thereby guaranteeing that sufficient statistics is used in all cases (albeit at the expense of a slightly overestimated  $\sigma_{\text{eff}}^2$ ). Finally, the variance  $\sigma_{\theta}^2$  was evaluated by extracting  $\hat{\theta}_j$  from a sliding window average (of width  $N = 50$ ) performed over all simulated symbols.

Fig. 5.2a shows the capacity lower bound curves as a function of the linear SNR (which is the ratio between the average signal power,  $P$ , and the power of the ASE noise within the channel bandwidth,  $\sigma_{\text{ASE}}^2$ ). The dashed curves correspond to the case in which we do not separate the phase noise and treat the entire nonlinear distortions as noise. These curves are consistent with the study reported in [3], except that our usage of Gaussian modulation leads to slightly higher capacity values. The solid curves represent our new lower bound, achieving a peak capacity that is higher by approximately 0.7 bit/sec/Hz in all cases. In Fig. 5.2b we invert the peak capacity results of Fig. 5.2a so as to plot the maximum achievable system length as a function of the spectral efficiency. As is evident from the figure, the achievable system length is approximately doubled by exploiting our scheme. Finally in Fig. 5.3 we show the various noise contributions as a function of the average power per channel in the case of a 500km link. The monotonically decreasing blue curve shows the noise to signal ratio (NSR)  $\sigma_{\text{ASE}}^2/(PT)$  due to the ASE noise by itself. The monotonically increasing solid blue curve shows  $\sigma_{\theta}^2$ , which in the limit of small variations in  $\theta_j$ , represents the NSR due to phase noise. The dashed blue line is the theoretical expression for  $\sigma_{\theta}^2$  as given in [6,9], which is seen to be in very good agreement with our numerical result. The blue dotted line shows the NSR due to the residual nonlinear noise  $\sigma_{\text{NL}}^2/(PT)$  after separating the phase noise. As is evident in the figure, separation of nonlinear phase noise reduces the nonlinear noise by approximately 6dB. Triangles and dots show  $\sigma_{\text{eff}}^2/(PT)$  with and without phase noise cancellation, respectively. Evidently, the minimum effective NSR of the



system is improved by approximately 2dB. We note that to facilitate the distinction between the noise contributions the simulation that produced Fig. 5.3 was performed without ASE propagation. In Figs. 5.2a and 5.2b ASE noise was propagated, although similar results were observed when the ASE was added at the end. We also note that these figures were obtained using more statistics than those presented in [11].

## 5.3 Conclusions

We have derived a new lower bound for the capacity of the nonlinear fiber channel by taking into account the fact that phase noise is one of the the most significant consequences of nonlinear interference, and by taking advantage of the fact that this noise is characterized by strong temporal correlations. We showed that the peak capacity per polarization can be increased by approximately 0.7bit/s/Hz or equivalently the length of a system can be (almost) doubled for a given transmission rate.

## Acknowledgement

The authors would like to acknowledge financial support from the Israel Science Foundation (grant 737/12). Ronen Dar would like to acknowledge the support of the Adams Fellowship Program of the Israel Academy of Sciences and Humanities, and the Yitzhak and Chaya Weinstein Research Institute for Signal Processing. Fruitful discussions with R. J. Essiambre are also gratefully acknowledged.

# Bibliography

- [1] P. P. Mitra and J. B. Stark, “Nonlinear limits to the information capacity of optical fibre communications,” *Nature* **411**, 1027–1030 (2001).
- [2] K. S. Turitsyn, S. A. Derevyanko, I. V. Yurkevich, and S. K. Turitsyn, “Information capacity of optical fiber channels with zero average dispersion,” *Physical review letters* **91**, 203901–203904 (2003).
- [3] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, “Capacity limits of optical fiber networks,” *J. Lightwave Technol.* **28**, 662–701 (2010).
- [4] A. D. Ellis, J. Zhao, and D. Cotter, “Approaching the non-linear Shannon limit,” *J. Lightwave Technol.* **28**, 423–433 (2010).
- [5] G. Bosco, P. Poggiolini, A. Carena, V. Curri, and F. Forghieri, “Analytical results on channel capacity in uncompensated optical links with coherent detection,” *Opt. Express* **19**, B440–B451 (2011).
- [6] A. Mecozzi and R.-J. Essiambre, “Nonlinear Shannon limit in pseudolinear coherent systems,” *J. Lightwave Technol.* **30**, 2011–2024 (2012).
- [7] E. Agrell and M. Karlsson, “WDM channel capacity and its dependence on multichannel adaptation models,” *Optical Fiber Communication Conference 2013 (OFC13)*.
- [8] P. J. Winzer and G. J. Foschini, “MIMO capacities and outage probabilities in spatially multiplexed optical transport systems,” *Opt. Express* **19**, 16680–96 (2011).
- [9] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, “Properties of nonlinear noise in long, dispersion-uncompensated fiber links,” *Opt. Express* **21**, 25685–25699 (2013).
- [10] T. M. Cover and J. A. Thomas, “*Elements of information theory*,” (Wiley, 1991).
- [11] R. Dar, M. Shtaif, and M. Feder, “Improved bounds on the nonlinear fiber-channel capacity,” *European Conference on Optical Communication 2013 (ECOC13)*, paper P.4.16 (2013).
- [12] With system parameters similar to those used for evaluating the capacity in [3] and assuming for example a 500km link, the phase has been shown to be nearly constant on the scale of a few tens of symbol durations [9].

- [13] A. Lapidoth and S. M. Moser, “Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels,” *IEEE Trans. on IT* **49**, 25685–25699 (2003).



# Chapter 6

## Accumulation of Nonlinear Interference Noise in Fiber-Optic Systems

Ronen Dar, Meir Feder, Antonio Mecozzi and Mark Shtaif, “*Accumulation of nonlinear interference noise in fiber-optic systems*,” Optics express, vol. 22, pp. 14199–14211 (June 2014)

### 6.1 Abstract

Through a series of extensive system simulations we show that all of the previously not understood discrepancies between the Gaussian noise (GN) model and simulations can be attributed to the omission of an important, recently reported, fourth-order noise (FON) term, that accounts for the statistical dependencies within the spectrum of the interfering channel. We examine the importance of the FON term as well as the dependence of NLIN on modulation format with respect to link-length and number of spans. A computationally efficient method for evaluating the FON contribution, as well as the overall NLIN power is provided.

### 6.2 Introduction

Inter-channel nonlinear interference is arguably the most important factor in limiting the performance of fiber-optic communications [1]. Since joint processing of the entire WDM spectrum of channels is prohibitively complex, nonlinear interference between channels is customarily treated as noise. The statistical characterization of this noise — to which we refer in what follows as nonlinear interference noise (NLIN) — is the goal of most recent theoretical studies of nonlinear transmission [2–9]. Understanding the features of NLIN is critical for the efficient design of fiber-optic systems and for the accurate prediction of their performance.

Most of the available work on NLIN in fiber-optic systems was published in the context of the Gaussian noise (GN) model [2–5], which describes NLIN as an additive Gaussian

noise process whose variance and spectrum it evaluates. The validation of the GN model and the characterization of its accuracy have been the subject of numerous studies (e.g. see [3, 4]). It was found that while the model's accuracy is satisfactory in some scenarios, it is highly inadequate in others. Some of the GN model's most conspicuous shortcomings are its independence of modulation format [6], its independence of pre-dispersion [4], and its large inaccuracy in predicting the growth of the NLIN variance with the number of spans in an amplified multi-span link [3, 4]. Although phenomenological fixes for the latter problem have been proposed (most notably through the practice of accumulating the NLIN contributions of various spans incoherently [2–4]), the remedy that they offered remained limited, and the fundamental reason for the observed behavior has never been understood.

We argue, similarly to [6], that the reason for the inaccuracy of the GN approach is in ignoring the statistical dependence between different frequency components of the interfering channel. Accounting for this dependence produces an important correction term to which we refer (for reasons explained in the following section) as the *fourth-order noise* or FON. By simulating a number of fiber-systems in the relevant range of parameters, we demonstrate that the FON term resolves all of the reported inaccuracies of the GN model, including the dependence on modulation format, signal pre-dispersion, and the accumulation of NLIN with the number of spans.

We stress that as demonstrated in [6] the NLIN is not an additive Gaussian process, and hence its variance (and even its entire spectrum) does not characterize its properties in a satisfactory manner. For example, as pointed out in [8], part of the NLIN manifests itself as phase-noise, whose effect in terms of transmission performance is very different from that of additive noise [10]. We show in what follows that the phase-noise character of NLIN, as well as the dependence of NLIN on modulation format is largest in the case of a single amplified span, or in a system of arbitrary length that uses distributed amplification (as in [6]). The distinctness of these properties reduces somewhat in multi-span systems with lumped amplification and with a span-length much larger than the fiber's effective length.

In order to facilitate future research of this problem, we provide a computer program that implements a computationally efficient algorithm for computing the SON and the FON coefficients that are needed for reproducing the theoretical curves that we present in this paper. For the reader's convenience, the program also includes the option of computing the entire NLIN variance, including intra-channel interference terms, as well as inter-channel interference terms that are not directly addressed in the main text of this paper, and which have been recently posted in [7]. The contribution of these terms reduces very rapidly with channel spacing and while one may generate situations in which inclusion of these terms becomes relevant, they were negligible in the system studied in [6] (which assumed distributed amplification, and a guard-band as small as 2%), and they are also negligible in systems with more realistic parameters, as considered here. We note that while the full scale simulations of the systems of interest are computationally intense and time consuming, the extraction of NLIN power on the basis of the FON and SON coefficients that we provide is practically instantaneous.

## 6.3 Theoretical background

In a recent paper [6] we have demonstrated that by removing the assumption of statistical independence between frequency components within the interfering channel (which has been used in the derivation of the GN model) the variance of NLIN is given by

$$\sigma_{\text{NLIN}}^2 = \underbrace{P^3\chi_1}_{\text{SON}\backslash\text{GN}} + \underbrace{P^3\chi_2 \left( \frac{\langle |b|^4 \rangle}{\langle |b|^2 \rangle^2} - 2 \right)}_{\text{FON}}, \quad (6.1)$$

where  $P$  is the average power,  $b$  denotes the data symbol in the interfering channel (e.g. for QPSK modulation  $b$  is a random variable that receives each of the four values  $\pm \frac{1}{\sqrt{2}} \pm \frac{j}{\sqrt{2}}$  with probability of 1/4), and the angled brackets denote statistical averaging. The terms  $\chi_1$  and  $\chi_2$  are given by Eqs. (26–27) in [6] multiplied by  $T^3$ , where  $T$  is the symbol duration. These coefficients are functions of the transmitted pulse waveform and of the fiber parameters. The first term on the right-hand-side of (6.1) is identical to the result of the GN model, and since it follows only from second-order statistics, we refer to it as the second-order noise (SON) term. The second term depends on fourth-order statistics and is hence referred to as the fourth-order noise (FON) term. The presence of  $\langle |b|^4 \rangle$  in the FON term implies modulation format dependence. For example, the NLIN variance is  $P^3(\chi_1 - \chi_2)$  with QPSK modulation,  $P^3(\chi_1 - 0.68\chi_2)$  with 16-QAM, and  $P^3\chi_1$  when Gaussian modulation is used. Note that only with Gaussian modulation the NLIN variance is independent of  $\chi_2$  and hence the GN-model's prediction is exact. In the section that follows we demonstrate the accuracy of Eq. (6.1) with respect to a range of fiber-optic systems that we simulate, and discuss the role and relative importance of the FON term in the various scenarios.

We note that in order to compare with the theory of [2–5], the SON and FON coefficients were written in Eqs. (26) and (27) of [6] without including the band-limiting effect of the receiver matched filter. To include its effect, products that fall outside of the received channel bandwidth should be excluded from the summation. The computer program which we provide in the appendix to this paper for the extraction of  $\chi_1$  and  $\chi_2$  accounts for the presence of a matched filter.

## 6.4 Results

The results are obtained from a series of simulations considering a five-channel WDM system implemented over standard single-mode fiber (dispersion of 17 ps/nm/km, nonlinear coefficient  $\gamma = 1.3 \text{ [Wkm]}^{-1}$ , and attenuation of 0.2dB per km). We assume Nyquist pulses with a perfectly square spectrum, a symbol-rate of 32 GSymbols/s and a channel spacing of 50 GHz. The number of simulated symbols in each run was 4096 and the total number of runs that were performed with each set of system parameters (each with independent and random data symbols) ranged between 100 and 500 so as to accumulate sufficient statistics. As we are only interested in characterizing the NLIN, we did not include amplified spontaneous emission (ASE) noise in any of the simulations. At the receiver, the channel of interest was isolated with a matched optical filter and ideally back-propagated so as to eliminate the effects of self-phase-modulation and chromatic dispersion. All simulations were performed

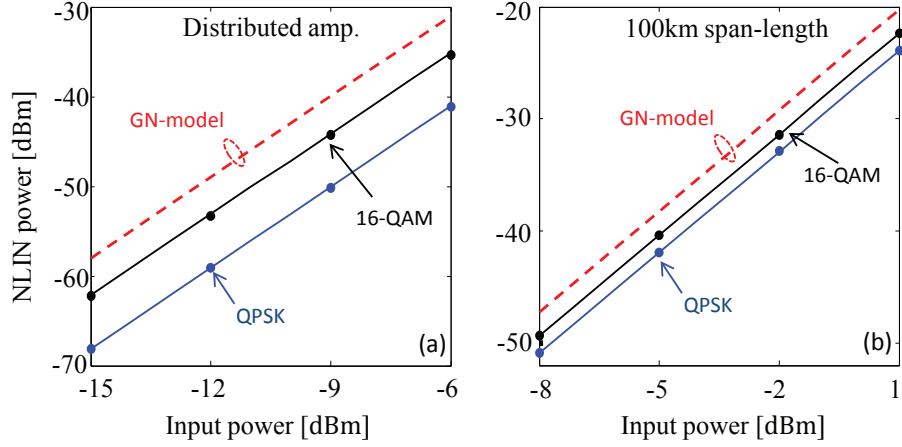


Figure 6.1: The NLIN power versus the average power per-channel in a  $5 \times 100\text{km}$  system for QPSK and 16-QAM modulation. The solid lines show the theoretical results given by Eq. (6.1) and the dots represent simulations. The dashed red line corresponds to the SON contribution  $P^3\chi_1$ , which is identical to the result of the GN model. (a) Distributed amplification. (b) Lumped amplification.

with a single polarization, whereas the scaling of the theoretical results to the dual polarization case has been discussed in [6]. For both forward and backward propagation, the scalar nonlinear Schrödinger equation has been solved using the standard split-step-Fourier method with a step size that was set to limit the maximum nonlinear phase variation to 0.02 degrees (and bounded from above by 1000 m). The sampling rate was 16 samples per-symbol. To extract the NLIN, we first removed the average phase-rotation induced by the neighboring WDM channels and then evaluated the offset between the received constellation points and the ideal constellation points that would characterize detection in the absence of nonlinearity.

In Fig. 6.1 we show the NLIN power as a function of the average input power for a system consisting of  $5 \times 100\text{ km}$  spans in the cases of QPSK and 16-QAM modulation. Figure 6.1(a) corresponds to the case of purely distributed amplification whereas Fig. 6.1(b) represents the same system in the case of lumped amplification. The solid curves represent the analytical results obtained from Eq. (6.1) while the dots represent the results of the simulations. The dashed red curve shows the prediction of the GN model, i.e.  $P^3\chi_1$ . The dependence on modulation format is evident in both figures, as is the GN model's offset. However, while the error of the GN model in the case of QPSK is 10dB for distributed amplification, it reduces to 3.7dB when lumped amplification is used. Note that the difference between the modulation formats, as well as the error of the GN-model result are both independent of the input power. The excellent agreement between the theory (Eq. (6.1)) and simulation is self evident.

More insight as to the significance of the span-length can be extracted from Fig. 6.2 which shows the received constellations in a 500 km system for QPSK (top panels) and 16-QAM (bottom panels) transmission. The first column of panels from the left [Figs. 6.2(a) and 6.2(b)] correspond to the case of distributed amplification, the second column of panels [Figs. 6.2(c) and 6.2(d)] correspond to 25 km spans, the third column of panels [Figs. 6.2(e)



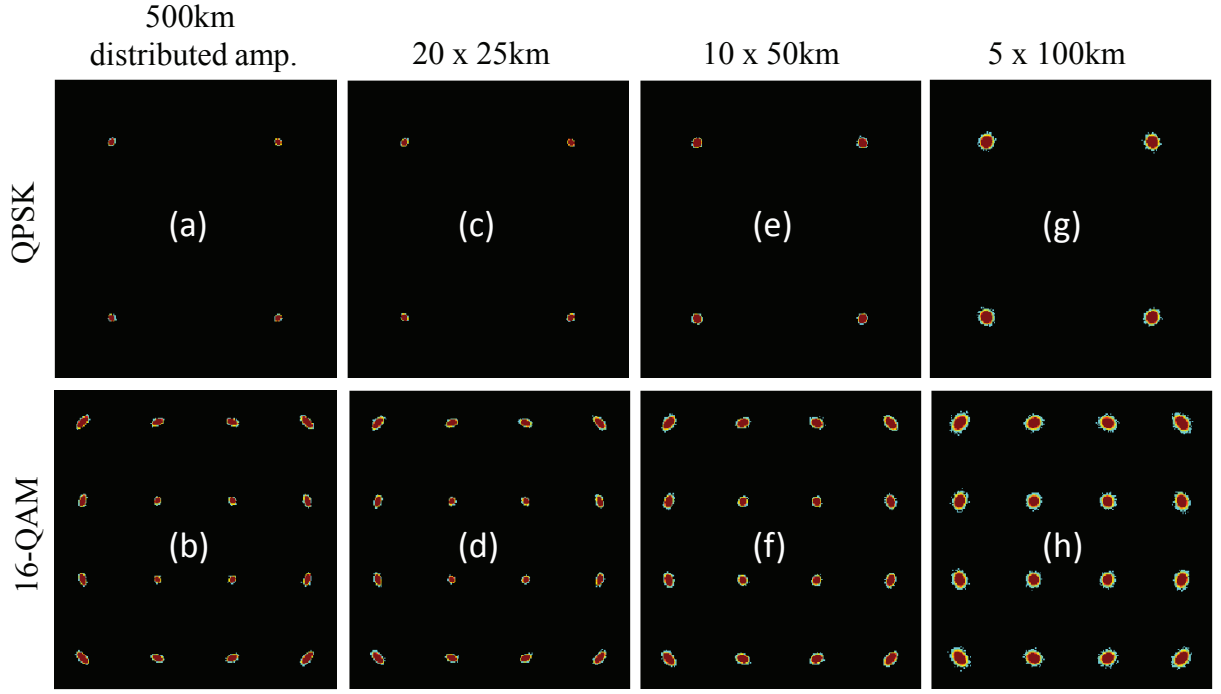


Figure 6.2: Constellation diagrams for QPSK (top panels) and 16-QAM (bottom panels) in the cases of distributed amplification (a and b), 25 km spans (c and d), 50 km (e and f) and 100 km (g and h). In all cases a path-averaged power of -10dBm was used. The phase noise nature of NLIN is evident in the case of 16-QAM modulation, but its relative significance reduces when the span-length is large. The 100 km span case is closer to the circular noise distributions observed in [3].

and 6.2(f)] correspond to 50 km spans and the case of 100 km spans is shown in the rightmost panels [Figs. 6.2(g) and 6.2(h)]. Here and in the figures that follow, the launched powers were selected such that the path-averaged power per-channel was -10dBm in all cases (input power of -7.7dBm, -5.9dBm and -3.3dBm, for 25 km, 50 km, and 100 km spans, respectively). Use of a constant path-averaged power is customary when comparing systems with different span lengths [11], and the value of -10dBm was found to be roughly optimal from the standpoint of capacity maximization in the distributed amplification case [8]. As can be seen in Fig. 6.1, the comparison between modulation formats is practically independent of the launched optical power. Consistently with the predictions in [6], the phase noise is negligible in the case of QPSK transmission, and the constellation points are nearly circular for all span-lengths. With 16-QAM modulation, phase-noise is dominant in the case of distributed amplification and 25 km spans, whereas at longer span-lengths the phase-noise character of NLIN becomes less evident. We note that even in the case of 100 km spans, phase noise is clearly visible implying that NLIN cannot be accurately described as additive, although the error from doing so is notably smaller.

Figure 6.3 shows the accumulation of NLIN along systems of different span-lengths. Fig. 6.3(a) shows the case of distributed amplification whereas Figs. 6.3(b)-6.3(d) show

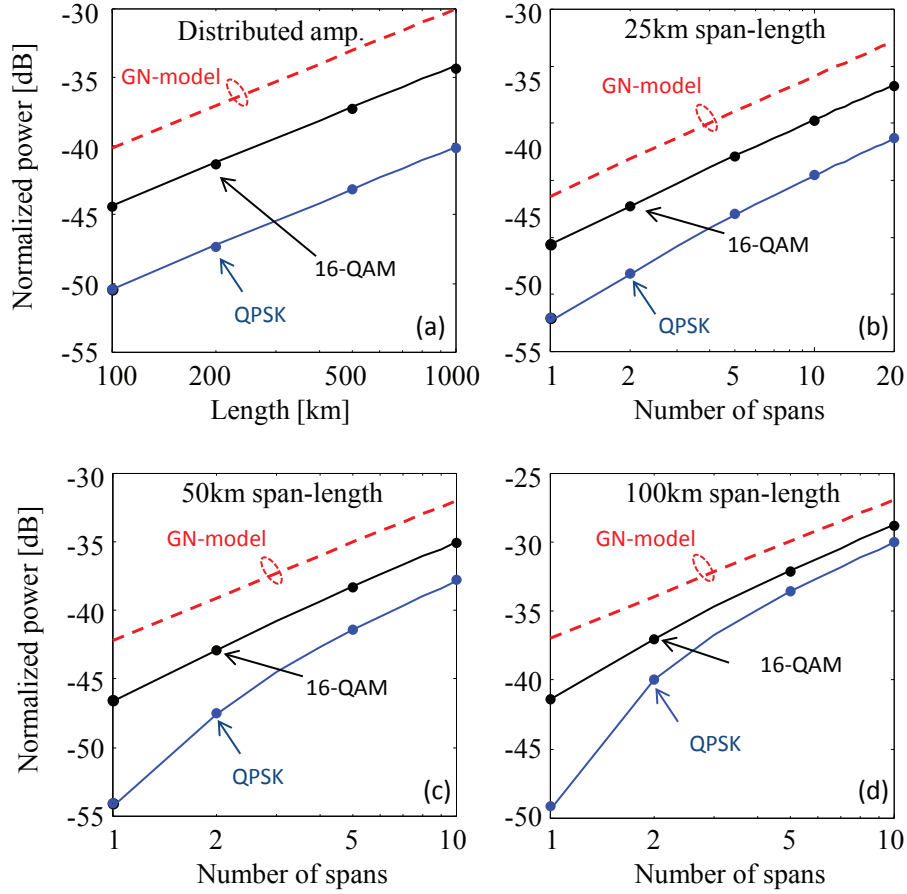


Figure 6.3: Accumulation of the NLIN power (normalized to the received power) with the number of spans. Figure a corresponds to the case of distributed amplification whereas figures b,c and d correspond to the cases of 25 km, 50 km, and 100 km span-lengths, respectively. The solid lines show the theoretical results given by Eq. (6.1) and the dots represent simulations. The red dashed curve represents the SON, or equivalently the GN model result.

the cases of 25 km, 50 km, and 100 km span-lengths, respectively. The NLIN power in these figures is normalized to the received optical power in each case so that the vertical axes can be interpreted as noise to signal ratio. Notice that in the cases of single-span transmission the dependence on modulation format and the inaccuracy of the GN model are largest. As the number of spans increases the NLIN increases more rapidly for the QPSK format and the difference between the modulation formats reduces. After 500 km of propagation the difference between the NLIN power for QPSK and 16-QAM modulation formats is approximately 6dB, 4.8dB, 2.8dB and 1.5dB for distributed amplification, 25 km, 50 km, and 100 km spans. The error of the GN model in the case of QPSK is approximately 10dB, 8.6dB, 5.8dB, and 3.7dB for the cases of distributed amplification, 25 km, 50 km and 100 km spans.

## 6.5 Discussion

The explanation to the observed behavior can be attributed to the dynamics of nonlinear collisions in WDM systems, whose details will be discussed in a separate publication. In particular, NLIN that is induced by cross-phase-modulation (XPM) is strongest when the nonlinearly interacting pulses experience incomplete collisions [12]. In a single-span system, or in a system with distributed gain, such collisions occur only twice; once in the beginning of the system and again at its end. In a multi-span system with lumped amplification incomplete collisions occur at every point of power discontinuity, namely at the beginning of every amplified span. Incomplete collisions taking place at different locations, produce NLIN contributions of independent phase and therefore, when the NLIN is dominated by incomplete collisions, it appears more isotropic in phase-space and its distribution becomes closer to Gaussian. The relative significance of incomplete collisions is determined mainly by two factors; the number of incomplete collisions (which grows with the number of spans), and the magnitude of power discontinuity (increases with span length). For a fixed length link, when the number of spans is so large that attenuation within the span is negligible, the power discontinuity at the amplifier sites vanishes and the system becomes equivalent to a distributed gain system, where only two incomplete collisions occur (at the beginning and at the end of the entire link). As the span length increases, the power discontinuity at the amplifier locations grows and the overall significance of incomplete collisions increases. This explains the fact that in Fig. 6.2 the 16-QAM constellation spots appear more and more circular as the span length increases from 0 (distributed amplification) to 100 km. When the span-length increases further, to the extent that it becomes much longer than the fiber's effective length ( $1/\alpha \sim 20$  km in most fibers), the growth in the power discontinuity becomes negligible, but the number of incomplete collisions continues to decrease with the number of spans, until eventually, in a single-span link (where only one incomplete collision occurs at the link's beginning) the non-Gaussianity of NLIN reappears and the deviation from the GN model is very significant. This point can be seen in Figs. 6.3(b)–6.3(d), where the NLIN variance is shown as a function of the number of spans and the span-length is kept constant. The error in the GN model is always largest in a single-span link, and reduces considerably with the number of spans in the case of lumped amplification.

Figure 6.4 summarizes these ideas by showing the ratio  $\chi_2/\chi_1$  as a function of the number of spans in a fixed-length system [Fig. 6.4(a)], and as a function of system length [Fig. 6.4(b)]. When incomplete collisions dominate so that the NLIN approaches a circular Gaussian distribution and the significance of phase-noise reduces,  $\chi_2/\chi_1 \ll 1$  and the NLIN variance is dominated by the SON contribution (the GN model). In Fig. 6.4(a) the deviation from the GN model is seen to be largest ( $\chi_2/\chi_1 \sim 1$ ) in the single-span case, and when the number of spans is so large that the scheme approaches the conditions of distributed amplification. When the span-length is fixed, as in Fig. 6.4(b), the ratio  $\chi_2/\chi_1$  reduces with the length of the link, with the highest rate of reduction occurring when the span-length is long so that the power discontinuity is largest.

Another interesting aspect of the nonlinear dynamics is revealed in the context of the effect of signal pre-dispersion. One of the most central claims made in [2–4, 7] is that signal Gaussianity, which is crucial for the validity of the GN model, follows from the accumulated effect of chromatic dispersion, and hence the large inaccuracy of the GN model in the first

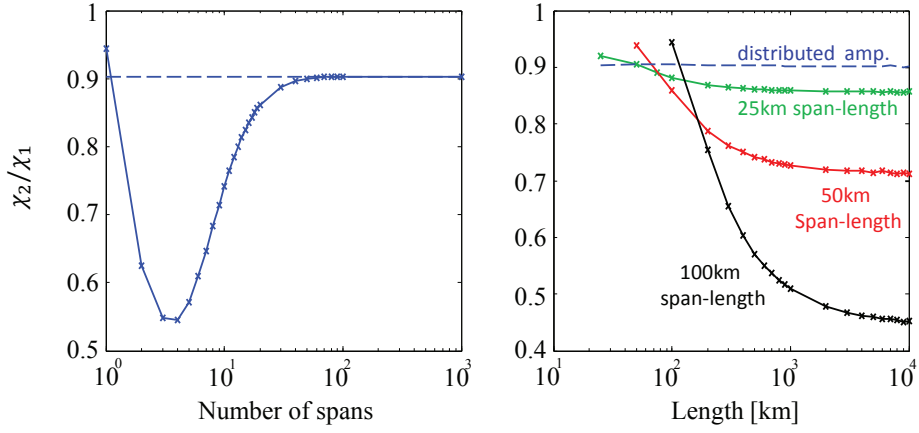


Figure 6.4: The importance of incomplete collisions is reflected in the ratio between the FOM and SON coefficients  $\chi_2/\chi_1$ . When incomplete collisions dominate  $\chi_2/\chi_1 \ll 1$ , and when their contribution is small (as occurs in single span, or distributed gain systems)  $\chi_2/\chi_1 \sim 1$ . In (a) The total link-length is held fixed at 500 km. In (b) the span length is kept constant. In both Figs. (a) and (b), the dashed curve corresponds to distributed amplification.

few spans of a WDM system was attributed to the fact that the signal is not sufficiently dispersed. Indeed, in [4] it has been demonstrated that in the presence of very aggressive pre-dispersion, the NLIN variance is accurately described by the GN model even in the very first few spans (where without pre-dispersion the inaccuracy of the GN model is largest). In our understanding the role of dispersion in this context has been misconstrued. While it is true that significant pre-dispersion reduces the GN model's inaccuracy in the first span, as shown in [4], it is not the absence of sufficient dispersion that explains the GN model's inaccuracy. Here we present an alternative view at the role of pre-dispersion. We plot in Fig. 6.5 the NLIN variance as a function of system length, once in the case of distributed amplification and once in the case of lumped amplification with 100km spans. In both cases the signals were pre-dispersed by 8500 ps/nm — equivalent to a 500 km long link. Notice that indeed pre-dispersion improves the accuracy of the SON term representing the GN model in the first few spans. However, when the link becomes longer and the accumulated dispersion exceeds the amount of pre-dispersion assigned to the signal, the deviation from the GN-model increases and eventually, the simulated NLIN variance approaches the same value that it has without pre-dispersion. This behavior is seen to be in clear contrast to the interpretation of [2–4, 7]. If the pre-dispersed signals are Gaussian enough at the end of the first span so as to satisfy the assumptions of the GN model, how come they are less Gaussian further along the system given that the accumulated dispersion increases monotonically? Our own interpretation to this behavior relies once again on the time domain picture of pulse collisions. When the temporal spreading of the launched pulses is larger than the walk-off between channels, all collisions become incomplete, and for the same reasons that we explained earlier the GN model becomes more accurate. Nonetheless, when the system length increases to the extent that the inter-channel walk-off becomes large enough to accommodate full collisions, the deviation from the GN result reappears once again.

When examining the situation in the frequency domain picture, pre-dispersion implies

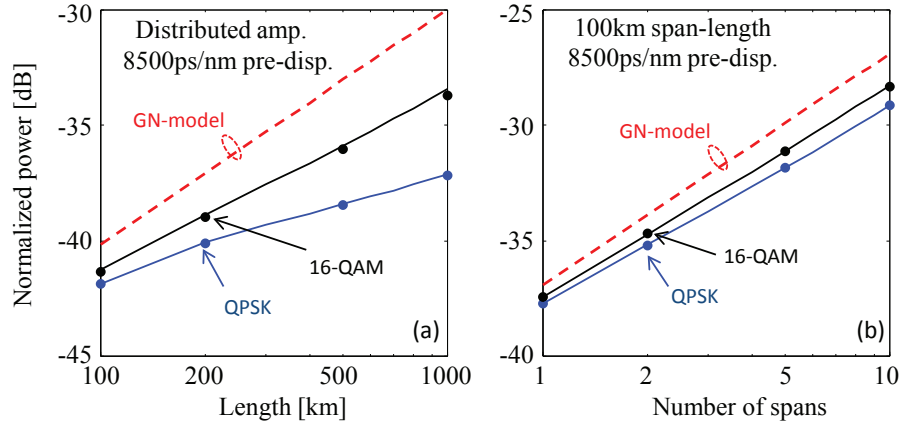


Figure 6.5: The effect of pre-dispersion. Accumulation of the NLIN power (normalized to the received power) with the number of spans. Figures a and b correspond to distributed amplification and span-length of 100 km, respectively, where pre-dispersion of 8500 ps/nm was applied to the injected pulses. The solid lines show the theoretical results given by Eq. (6.1) and the dots represent simulations. The red dashed curve represents the SON, or equivalently the GN model result.

rapid phase variations in the interfering channel's spectrum (i.e. variations in the phase of  $\tilde{g}(\omega)$  in the notation of [6]). While the SON coefficient  $\chi_1$  is not affected by the spectral phase, the FON coefficient  $\chi_2$  reduces considerably in this situation, since the fourth-order correlation terms (Eq. 24 in [6]) lose coherence. We note however, that since with all relevant modulation formats, the quantity  $\frac{\langle |b|^4 \rangle}{\langle |b|^2 \rangle^2} - 2$  is negative, the reduction of  $\chi_2$  through pre-dispersion always leads to an increase in the NLIN variance and is therefore undesirable.

## 6.6 Conclusions

We have shown that the previously unexplained dependence of the NLIN variance on pre-dispersion, modulation format and on the number of amplified spans, is accounted for by the FON term, which follows from the correct treatment of the signal's statistics [6]. Excellent agreement between theory and simulations has been demonstrated in all of our simulations, suggesting that in the range of parameters that we have tested, the inclusion of additional correction terms, which were presented in [7] is not necessary. The relative magnitude of the FON term is largest in single span systems, or in systems using distributed amplification, and it reduces notably in the case of lumped amplification with a large number of spans. Similarly, the relative significance of phase noise (which is included both in the SON and the FON terms) is largest in single span systems, or in systems with distributed amplification, although it remains significant in all the cases that we have tested.

## 6.7 Appendix: Computation of $\chi_1$ and $\chi_2$

The extraction of the analytical curves in Figs. 6.1 and 6.3 –6.5 relies on the computation of the SON and FON coefficients  $\chi_1$  and  $\chi_2$ , which requires summation over three and five indices, respectively. Multi-dimensional summations are extremely inefficient in brute-force computation, and hence we have adopted the Monte-Carlo integration method [13] for evaluating these quantities. We provide a code (written in Matlab) that computes  $\chi_1$ ,  $\chi_2$  (using Eqs. (26) and (27) of [6] including receiver matched filtering that removes products that fall outside of the received channel bandwidth), and allows the evaluation of the implied NLIN variance according to Eq. (6.1). The program also evaluates the relative error in the computed NLIN variance, where in all of the numerical curves presented in this paper the number of integration points  $N$  was large enough to ensure that the relative error was well under 1%.

For the reader's convenience, in addition to providing the tools for reproducing the curves presented in this paper, we include blocks that compute the variance of nonlinear intra-channel noise, as well as additional inter-channel terms that contribute to NLIN when the guard-band between WDM channels is much smaller than the channel bandwidth, and which were first reported in [7]. It can be easily verified that the contribution of these terms is negligible with the typical system parameters assumed in this paper, although they may play a role in the case of future densely spaced super-channels. The option of describing a polarization multiplexed link is also included. The program assumes perfect Nyquist pulses and homogeneous fiber spans, but it can be readily extended to an arbitrary pulse-shape and to systems with different span length and fiber dispersions.

The runtime for the computation of the SON and FON coefficients  $\chi_1$  and  $\chi_2$  on a standard PC with an i5 processor is of the order of 0.5 seconds, whereas the computation of all (intra and inter-channel) NLIN terms is performed in less than 2 seconds. We note that polarization multiplexing does not affect the run-time of the code (although it more than doubles the runtime of a full split-step simulation).

```
function main()
%% System parameters
clear;%clc;tic;
PolMux = 0; % 0 when single polarization, 1 with polarization mulltiplexing
gamma = 1.3; % Nonlinearity coefficient in [1/W/km]
beta2 = 21; % Dispersion coefficient [ps^2/km]
alpha = 0.2; % Fiber loss coefficient [dB/km]
Nspan = 24; % Number of spans
L = 100; % Span length [km]
PD = 0; % Pre-dispersion [ps^2]
PdBm = -2; % Average input power [dBm]
BaudRate = 32; % Baud-rate [GHz]
ChSpacing = 33.6; % Channel spacing [GHz]
kur = 1;%1.32; % Second order modulation factor <|a|^4>/<|a|^2>^2
kur3 = 1;%1.96; % Third order modulation factor <|a|^6>/<|a|^2>^3
N = 1000000; % Number of integration points in algorithm [14]. Should
% be set such that the relative error is desirably small.
```

```

%% Monte-Carlo integration
alpha_norm = alpha/10*log(10);
T=1000/BaudRate;
P0 = 10^((PdBm-30)/10);
beta2_norm = beta2/T^2;
PD_norm = PD/T^2;
ChSpacing_norm = ChSpacing./BaudRate;

%% calculate inter-channel nonlinear noise variance according to Eq. (1)
[NLIN_var chi1 chi2 Err] = calc_interChannel(gamma,beta2_norm,alpha_norm,...
      Nspan,L,PD_norm,P0,kur,kur3,N,PolMux,ChSpacing_norm);

%% display
disp('%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%');
if(PolMux == 1)
    disp('%%%Polarization Multiplexed case is considered%%');
end
disp('%%Results correspond to a single interferer%%')
disp(['(1) chi_1 = ' num2str(chi1) ', chi_2 = ' num2str(chi2)]);
disp(['(2) NLIN variance according to Eq. (1) is ' num2str(NLIN_var)...
      ' Watts (' num2str(10*log10(NLIN_var*1000)) 'dBm).'...
      ' Relative computation error is ', num2str(Err*100), '%']);

%% calculate the contribution of the additional inter-channel terms of [7]
NLIN_var_addTerm = calc_interChannel_addTerms(gamma,beta2_norm,alpha_norm,...
      Nspan,L,PD_norm,P0,kur,kur3,N,PolMux,ChSpacing_norm);

%% calculate intra-channel nonlinear noise
%% NLIN_var_intra(1) is the intra-channel nonlinear noise variance
%% NLIN_var_intra(2) is due to the nonlinear broadening of the adjacent interferer
NLIN_var_intra = calc_intraChannel(gamma,beta2_norm,alpha_norm,...
      Nspan,L,PD_norm,P0,kur,kur3,N,PolMux,ChSpacing_norm);
NLIN_var_addTerm = NLIN_var_addTerm + NLIN_var_intra(2);

%%Continue display
disp(['(3) Contribution of additional inter-channel interference '...
      'terms of [7] is ' num2str(NLIN_var_addTerm) ' Watts ('...
      num2str(10*log10(NLIN_var_addTerm*1000)) 'dBm)']);
disp(['(4) Total NLIN variance (2)+(3) is ' num2str(NLIN_var+NLIN_var_addTerm)...
      ' Watts (' num2str(10*log10((NLIN_var+NLIN_var_addTerm)*1000)) 'dBm)']);
disp(['(5) Intra-Channel nonlinear noise variance is ' num2str(NLIN_var_intra(1))...
      ' Watts (' num2str(10*log10(NLIN_var_intra(1)*1000)) 'dBm)']);
toc;
disp('%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%');
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [NLIN_var chi1 chi2 Err] = ...
    calc_interChannel(gamma,beta2,alpha,Nspan,L,PD,P0,kur,kur3,N,PolMux,q)

R = 2*pi*(rand(4, N)-0.5*ones(4, N));
Volume = (2*pi)^4;

```



```

%% calculate chi1
w0 = R(1,:)-R(2,:)+R(3,:);
arg1 = (R(2,:)-R(3,:)).*(R(2,:)+2*pi*q-R(1,:));
argPD1 = arg1;
ss1 = exp(1i*argPD1*PD).*(exp(1i*beta2*arg1*L-alpha*L)-1)...
    ./((1i*beta2*arg1-alpha).*(w0<pi).*(w0>pi));
s1 = abs(ss1.*(1-exp(1i*Nspan*arg1*beta2*L)))...
    ./((1-exp(1i*arg1*beta2*L))).^2/Volume;
avgF1 = sum(s1)/N;
chi1 = avgF1*Volume*(4*gamma^2*P0^3);

%% calculate chi2
w3p = -R(2,:)+R(4,:)+R(3,:)+2*pi*q;
arg2 = (R(2,:)-R(3,:)).*(R(4,:)-R(1,:)+2*pi*q);
argPD2 = arg2;
ss2 = exp(-1i*argPD2*PD).*(exp(-1i*beta2*arg2*L-alpha*L)-1)...
    ./((-1i*beta2*arg2-alpha).*(w3p>pi+2*pi*q).*(w3p<pi+2*pi*q));
s2 = (1-exp(1i*Nspan*arg1*beta2*L))./(1-exp(1i*arg1*beta2*L)).*ss1...
   .*(1-exp(-1i*Nspan*arg2*beta2*L))...
    ./((1-exp(-1i*arg2*beta2*L)).*ss2/Volume;
avgF2 = real(sum(s2))/N;
chi2 = avgF2*Volume*(4*gamma^2*P0^3);

%% calculate NLIN
NLIN_var = chi1+(kur-2)*chi2;
if(PolMux == 1)
    NLIN_var = (9/8)^2*16/81*(NLIN_var+2*chi1/4+(kur-2)*chi2/4);
end

%% calculate the root mean square relative error
if(PolMux == 0)
    Err = (sum((s1-avgF1+(kur-2)*(real(s2)-avgF2)).^2)...
        /(N-1)).^4/(avgF1+(kur-2)*avgF2)/sqrt(N);
else
    Err = (sum(((9/8)^2*16/81*(6/4*(s1-avgF1)+5/4*(kur-2)*...
        (real(s2)-avgF2))).^2)/(N-1)).^4/((9/8)^2*16/81*...
        (6/4*avgF1+5/4*(kur-2)*avgF2))/sqrt(N);
end

end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [NLIN_var] = calc_interChannel_addTerms(gamma,beta2,alpha,...
    Nspan,L,PD,P0,kur,kur3,N,PolMux,q)

R = 2*pi*(rand(4, N)-0.5*ones(4, N));

%% calculate X21
w0 = R(1,:)-R(2,:)+R(3,:)+2*pi*q;
arg1 = (R(2,:)-R(3,:)-2*pi*q).*(R(2,:)-R(1,:));
argPD1 = arg1;
ss1 = exp(1i*argPD1*PD).*(exp(1i*beta2*arg1*L-alpha*L)-1)...
    ./((1i*beta2*arg1-alpha).*(w0<pi).*(w0>pi));
s1 = abs(ss1.*(1-exp(1i*Nspan*arg1*beta2*L))./(1-exp(1i*arg1*beta2*L))).^2;

```



```

X21 = sum(s1)*(gamma^2*P0^3)/N;

%% calculate X22
w1 = R(1,:)-R(2,:)+R(4,:);
arg2 = (w1-R(3,)-2*pi*q).*(R(2,:)-R(1,:));
argPD2 = arg2;
ss2 = exp(-1i*argPD2*PD).*(exp(-1i*beta2*arg2*L-alpha*L)-1)...
    ./(-1i*beta2*arg2-alpha).*(w1<pi).*(w1>pi);
s2 = (1-exp(1i*Nspan*arg1*beta2*L))./(1-exp(1i*arg1*beta2*L)).*ss1...
    .*(1-exp(-1i*Nspan*arg2*beta2*L))./(1-exp(-1i*arg2*beta2*L)).*ss2;
X22 = real(sum(s2))*(gamma^2*P0^3)/N;

%% calculate X23
w2 = R(1,:)+R(2,:)-R(3,)-2*pi*q;
arg1 = (R(3,)+2*pi*q-R(2,)).*(R(3,)+2*pi*q-R(1,));
argPD1 = arg1;
ss3 = exp(1i*argPD1*PD).*(exp(1i*beta2*arg1*L-alpha*L)-1)...
    ./(1i*beta2*arg1-alpha).*(w2<pi).*(w2>pi);
s3 = abs(ss3.*(1-exp(1i*Nspan*arg1*beta2*L))./(1-exp(1i*arg1*beta2*L))).^2;
X23 = sum(s3)*(gamma^2*P0^3)/N;

%% calculate X24
w3 = R(1,:)-R(4,:)+R(2,:);
arg2 = (R(3,)+2*pi*q-R(4,)).*(R(3,)+2*pi*q-w3);
argPD2 = arg2;
ss4 = exp(-1i*argPD2*PD).*(exp(-1i*beta2*arg2*L-alpha*L)-1)...
    ./(-1i*beta2*arg2-alpha).*(w3<pi).*(w3>pi);
s4 = (1-exp(1i*Nspan*arg1*beta2*L))./(1-exp(1i*arg1*beta2*L)).*ss3...
    .*(1-exp(-1i*Nspan*arg2*beta2*L))./(1-exp(-1i*arg2*beta2*L)).*ss4;
X24 = real(sum(s4))*(gamma^2*P0^3)/N;

%% calculate NLIN
NLIN_var = 4*X21+4*(kur-2)*X22+2*X23+(kur-2)*X24;
if(PolMux == 1)
    NLIN_var = (9/8)^2*16/81*(NLIN_var+2*X21+(kur-2)*X22+X23+0*(kur-2)*X24);
end

end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [NLIN_var] = ...
    calc_intraChannel(gamma,beta2,alpha,Nspan,L,PD,P0,kur,kur3,N,PolMux,q)

if(exist('q')==0) q = 0; end;
R = 2*pi*(rand(5, N)-0.5*ones(5, N));

%% calculate X1
w0 = R(1,:)-R(2,:)+R(3,:);
argInB = (w0<pi).*(w0>pi);
argOutB = (w0<pi+2*pi*q).*(w0>pi+2*pi*q);
arg1 = (R(2,:)-R(3,)).*(R(2,:)-R(1,));
argPD1 = arg1;
ss1 = exp(1i*argPD1*PD).*(exp(1i*beta2*arg1*L-alpha*L)-1)...
    ./(1i*beta2*arg1-alpha);

```

```

s1 = abs(ss1.*(1-exp(1i*Nspan*arg1*beta2*L))./(1-exp(1i*arg1*beta2*L))).^2;
X1 = [sum(s1.*argInB) sum(s1.*argOutB)]*(gamma^2*P0^3)./N;

%% calculate X0
s0 = ss1.*(1-exp(1i*Nspan*arg1*beta2*L))./(1-exp(1i*arg1*beta2*L));
X0 = [abs(sum(s0.*argInB)/N).^2 abs(sum(s0.*argOutB)/N).^2]*(gamma^2*P0^3);

%% calculate X2
w1 = -R(2,:)+R(4,:)+R(3,:);
arg2 = (R(2,:)-R(3,:)).*(R(4,:)-R(1,:));
argPD2 = arg2;
ss2 = exp(-1i*argPD2*PD).*(exp(-1i*beta2*arg2*L-alpha*L)-1)...
    ./(-1i*beta2*arg2-alpha).*(w1<pi).*(w1>pi);
s2 = (1-exp(1i*Nspan*arg1*beta2*L))./(1-exp(1i*arg1*beta2*L)).*ss1...
    .*(1-exp(-1i*Nspan*arg2*beta2*L))./(1-exp(-1i*arg2*beta2*L)).*ss2;
X2 = [real(sum(s2.*argInB)) real(sum(s2.*argOutB))]*(gamma^2*P0^3)./N;

%% calculate X21
w2 = R(4,:)-R(1,:)-R(3,:);
arg2 = (R(2,:)-R(4,:)).*(R(2,:)-w2);
argPD2 = arg2;
ss2 = exp(-1i*argPD2*PD).*(exp(-1i*beta2*arg2*L-alpha*L)-1)...
    ./(-1i*beta2*arg2-alpha).*(w2<pi).*(w2>pi);
s21 = (1-exp(1i*Nspan*arg1*beta2*L))./(1-exp(1i*arg1*beta2*L)).*ss1...
    .*(1-exp(-1i*Nspan*arg2*beta2*L))./(1-exp(-1i*arg2*beta2*L)).*ss2;
X21 = [real(sum(s21.*argInB)) real(sum(s21.*argOutB))]*(gamma^2*P0^3)./N;

%% calculate X3
w3 = R(1,:)-R(2,:)+R(4,:)+R(3,:)-R(5,:);
arg3 = (R(4,:)-R(5,:)).*(R(4,:)-w3);
argPD3 = arg3;
ss3 = exp(-1i*argPD3*PD).*(exp(-1i*beta2*arg3*L-alpha*L)-1)...
    ./(-1i*beta2*arg3-alpha).*(w3<pi).*(w3>pi);
s3 = (1-exp(1i*Nspan*arg1*beta2*L))./(1-exp(1i*arg1*beta2*L)).*ss1...
    .*(1-exp(-1i*Nspan*arg3*beta2*L))./(1-exp(-1i*arg3*beta2*L)).*ss3;
X3 = [real(sum(s3.*argInB)) real(sum(s3.*argOutB))]*(gamma^2*P0^3)./N;

%% calculate NLIN
NLIN_var = 2*X1+(kur-2)*(4*X2+X21)+(kur3-9*kur+12)*X3-(kur-2)^2*X0;
if(PolMux == 1)
    NLIN_var = (9/8)^2*16/81*(NLIN_var+X1+(kur-2)*X2);
end

end

```

## Acknowledgment

The authors would like to acknowledge financial support from the Israel Science Foundation (grant 737/12). Ronen Dar would like to acknowledge the support of the Adams Fellowship of the Israel Academy of Sciences and Humanities, the Yitzhak and Chaya Weinstein Research Institute, and the Feder Family Award.

# Bibliography

- [1] R.-J. Essiambre, G. Kramer, P.J. Winzer, G.J. Foschini, B. Goebel, “Capacity limits of optical fiber networks,” *J. Lightwave Technol.* **28**, 662–701 (2010).
- [2] P. Poggiolini, A. Carena, V. Curri, G. Bosco, F. Forghieri, “Analytical modeling of nonlinear propagation in uncompensated optical transmission links,” *IEEE Photon. Technol. Lett.* **23**, 742–744 (2011).
- [3] A. Carena, V. Curri, G. Bosco, P. Poggiolini, F. Forghieri, “Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links,” *J. Lightwave Technol.* **30**, 1524–1539 (2012).
- [4] A. Carena, G. Bosco, V. Curri, P. Poggiolini, F. Forghieri, “Impact of the transmitted signal initial dispersion transient on the accuracy of the GN-model of non-linear propagation,” 39th European Conference and Exhibition on Optical Communication (ECOC 2013), Paper Th.1.D.4.
- [5] P. Johannisson and M. Karlsson, “Perturbation analysis of nonlinear propagation in a strongly dispersive optical communication system,” *IEEE J. Lightwave Technol.* **31**, 1273–1282 (2013).
- [6] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, “Properties of nonlinear noise in long, dispersion-uncompensated fiber links,” *Optics Express*, **21**, pp. 25685–25699 (2013).
- [7] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, F. Forghieri, “On the accuracy of the GN-model and on analytical correction terms to improve it,” arXiv preprint 1401.6946v1 (2014).
- [8] A. Mecozzi and R.-J. Essiambre, “Nonlinear Shannon limit in pseudolinear coherent systems,” *J. Lightwave Technol.* **30**, 2011–2024 (2012).
- [9] M. Secondini, E. Forestieri, G. Prati, “Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps,” *J. Lightwave Technol.* **31**, 3839–3852 (2013).
- [10] R. Dar, M. Shtaif, M. Feder, “New bounds on the capacity of the nonlinear fiber-optic channel,” *Optics Lett.* **39**, 398–401 (2014).
- [11] J. P. Gordon and L. F. Mollenauer, “Effects of fiber nonlinearities and amplifier spacing on ultra-long distance transmission,” *J. Lightwave Technol.* **9**, 170–173 (1991)

- [12] M. Shtaif, “Analytical description of cross-phase modulation in dispersive optical fibers.” *Optics Lett.* **23**, 1191–1193 (1998).
- [13] R. E. Caflisch, “Monte Carlo and quasi-Monte Carlo methods,” *Acta Numerica* **7**, Cambridge University Press, 1–49 (1998).

# Chapter 7

## Inter-Channel Nonlinear Interference Noise in WDM Systems: Modeling and Mitigation

Ronen Dar, Meir Feder, Antonio Mecozzi and Mark Shtaif, “*Inter-channel nonlinear interference noise in WDM systems: modeling and mitigation*,” invited paper, IEEE Journal of Lightwave Technology, vol. 33, pp. 1044–1053 (March 2015)

### 7.1 Abstract

We review the modeling of inter-channel nonlinear interference noise (NLIN) in fiber-optic communication systems, focusing on the accurate extraction of the NLIN variance, the dependence on modulation format, the role of nonlinear phase-noise and the existence of temporal correlations. We show ways in which temporal correlations can be exploited for reducing the impact of NLIN, and discuss the prospects of this procedure in future systems.

### 7.2 Introduction

Signal distortions imposed by the nonlinearity of optical fibers are arguably the most significant factor limiting the capacity of modern wavelength division multiplexed (WDM) communications systems [1, 2]. Nonlinear distortions are customarily categorized as belonging to the intra-channel [3] or the inter-channel [4] types. Intra-channel distortions are those affecting an individual WDM channel independently of the other WDM channels propagating in the fiber. Inter-channel distortions are those in which a given WDM channel is affected by the electric fields of the neighboring WDM channels that propagate in the fiber at the same time. In principle, while intra-channel distortions can be eliminated by means of digital back propagation at the receiver [5], or pre-distortion at the transmitter [6], similar elimination of inter-channel distortions is considered impractical. The difficulty is not only in the need of coherently detecting and jointly processing multiple WDM channels, but also because of the unpredictable add-drop operations taking place in modern optical networks, and the fact that the polarizations states of different WDM channels evolve randomly and

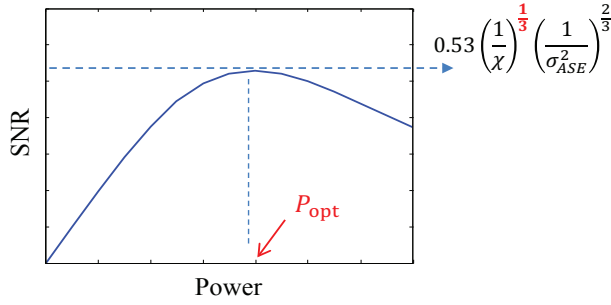


Figure 7.1: An illustration of the SNR vs. power as given in Eq.7.1. The proportionality of the peak SNR to  $\chi^{-1/3}$  implies that for every decibel of reduction in the NLIN power, the peak SNR reduces only by one third of a decibel.

differently in the presence of polarization mode dispersion [7]. For this reason, interference due to inter-channel nonlinear distortions must be considered as noise.

A convenient way of illustrating the importance of inter-channel interference and its mitigation is by considering the signal-to-noise ratio (SNR) of a system impaired by nonlinearity and by amplified spontaneous emission (ASE) noise. Although the SNR is not a sufficient parameter for characterizing a nonlinear system (i.e. the statistical properties of the ASE and of the nonlinear distortions can be distinctly different from each other [8]), it provides valuable insight as to the significance of nonlinear distortions and the prospects of their mitigation. We start from the case of no nonlinearity mitigation, where both the intra-channel and the inter-channel distortions are treated as noise. Since the variances of both types of distortions are proportional to  $P^3$ , with  $P$  denoting the average power per-channel (under the simplifying assumption that all WDM channels operate at the same average power), the SNR can be written as [9, 10]<sup>1</sup>

$$\text{SNR} = \frac{P}{\sigma_{\text{ASE}}^2 + \chi P^3}, \quad (7.1)$$

where  $\chi$  is a proportionality coefficient. Equation (7.1) produces the famous SNR curve illustrated in Fig. 7.1, which peaks at<sup>2</sup>  $0.53\chi^{-1/3}(\sigma_{\text{ASE}}^2)^{-2/3}$ . The effect of nonlinearity mitigation is to reduce the variance of the nonlinear distortions, or equivalently to reduce the magnitude of the coefficient  $\chi$ . But the proportionality of the above expression to  $\chi^{-1/3}$  implies that for every decibel of improvement in the nonlinear variance, the system's peak SNR improves only by one third of a decibel – a very unfavorable relation in the context of nonlinearity mitigation. Using this notion one can assess the effect of intra-channel nonlinear mitigation on the system SNR. Fig. 7.2(a) shows the ratio between the intra-channel and inter-channel contributions to the NLIN power as a function of the number of spans in the case of a standard single-mode fiber, assuming that the entire C-band (35 nm) is populated with 16-QAM modulated channels running at a baud-rate of 32 G-Symbols per second. The

<sup>1</sup>The dependence on  $P^3$  is rigorous within the framework of first-order perturbation analysis. In addition, the expression for the SNR also assumes that the effect of nonlinearity on the ASE noise, and its nonlinear interaction with the signal are negligible. Both these assumption are safely satisfied in most systems of practical relevance.

<sup>2</sup>The exact value of the coefficient is  $4^{1/3}/3 \simeq 0.53$ .

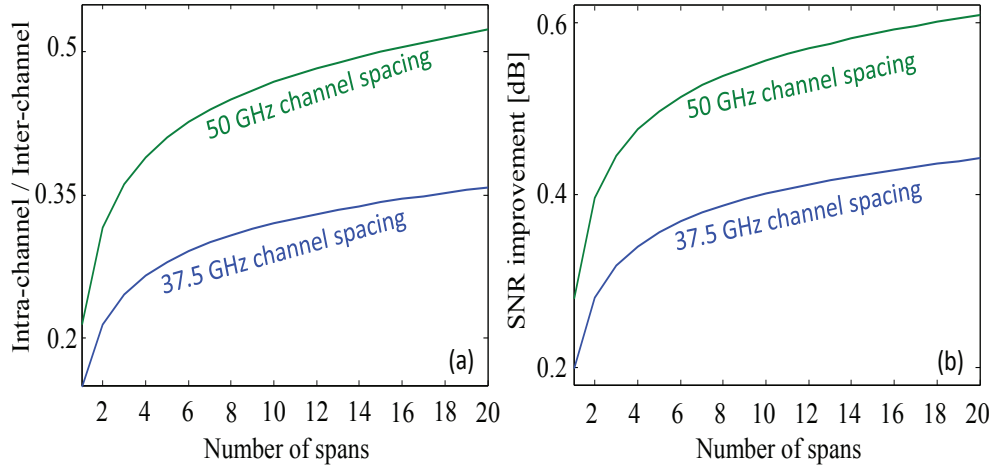


Figure 7.2: (a) The ratio between the intra-channel and inter-channel contributions to the NLIN power. (b) The improvement in the peak SNR resulting from ideal single-channel back-propagation. The number of WDM channels in the cases of 37.5 GHz, and 50 GHz spacing were 115 and 87 (43 and 57 channels on each side of the channel of interest), respectively. Each span consisted of 100 km of standard single-mode fiber (dispersion of 17 ps/nm/km, nonlinear coefficient  $1.3 \text{ [Wkm]}^{-1}$ ). The modulation format was polarization multiplexed 16-QAM, using perfect Nyquist pulses at 32 Giga-baud.

results were obtained using the NLIN wizard [11] and are plotted for two typical values of channel frequency separation; 37.5 GHz and 50 GHz, with the total number of WDM channels being 115, and 87, respectively (57, and 43 neighboring channels on each side of the channel of interest). Figure 7.2(b) shows the improvement in the effective peak SNR that would follow from ideal single-channel back-propagation. As is evident in the figure, the benefit is only of the order of  $\sim 0.4$  dB (see further discussion in [12]), and therefore, in order to improve system performance beyond this level, it is imperative to address the mitigation of inter-channel nonlinear effects.

In what follows we extend the work reported in [13]. We concentrate on the characterization and mitigation of noise generated by inter-channel distortions and refer to it as nonlinear interference noise, or NLIN [8, 14–17]. We argue that typical dispersion-uncompensated fiber-communications links installed over standard single-mode fiber can be divided into two categories. The first consists of metro systems extending over less than 1000 km, or systems of any length using short spans, or relying mostly on distributed (Raman) amplification. The second category consists of long-haul systems using lumped (EDFA) amplification with long spans, and extending over much more than 1000 km. We show that the NLIN in the first category of systems exhibits clear phase-noise-like features and a strong dependence on modulation format. In the second category of systems these features are suppressed considerably. In addition, while in the first category of systems, the accurate assessment of the NLIN variance can only be achieved using the complete model reported in [8], or its extension provided in [18], the standard Gaussian noise (GN) model [9, 19–22] can be used to approximate the NLIN variance in systems of the second category. In [14] we provide a Matlab script that computes the inter-channel (as well as the intra-channel) NLIN variance

within seconds. The implementation of this script with a number of important extensions is also available as an online application at [11].

Another topic that we address in this paper is the correlation properties of inter-channel NLIN [8, 15–17], which offer a potential path for inter-channel nonlinearity mitigation. The presence of chromatic dispersion causes different WDM channels to propagate at different group velocities, so that every symbol in the channel of interest interacts with multiple symbols of every interfering channel. Consequently, adjacent symbols in the channel of interest are disturbed by essentially the same collection of interfering pulses, and are therefore affected by nonlinearity in a highly correlated manner<sup>3</sup>. This property has been exploited to some extent in [15], where it was shown that when taking this correlation into account with respect to the phase-noise component of NLIN, the lower bound for the information capacity increases by nearly 1 bit. The implications of temporal correlations on the channel capacity have also been recently considered in [25]. We take this idea further and argue that a channel impaired by inter-channel NLIN is equivalent to a linear channel with time-varying inter-symbol interference (ISI) [16, 17, 26]. Using this concept, we evaluate the extent to which mitigation of inter-channel nonlinearities may improve system performance.

The paper is organized as follows. In Sec. 7.3 we briefly review the essentials of the time domain model, initially introduced in [8, 27], and extend the main equations to the polarization multiplexed case. In Sec. 7.4 we quantify the significance of modulation format dependence and the phase-noise-like nature of NLIN in polarization multiplexed systems. In Sec. 7.5 we introduce the time-varying ISI model and evaluate the potential of inter-channel NLIN mitigation. Finally, Sec. 7.6 is devoted to conclusions.

### 7.3 NLIN modeling

Throughout this work we neglect the effect of four-wave-mixing, whose significance in the kind of systems used in modern communications (particularly when implemented over standard single-mode fiber (SMF)) is known to be secondary, and focus our attention on interference due to the cross-phase-modulation (XPM) effect. Since this effect only involves two-channel interactions, the NLIN contributions of multiple WDM channels add up independently, and hence we conduct the initial analysis with only two channels, one of which is the channel of interest (whose central frequency is arbitrarily set to 0) and the other is the interfering channel, and its central frequency is denoted by  $\Omega$ . The electric field vector that is injected into the fiber is thereby given by<sup>4</sup>

$$\underline{u}(0, t) = \sum_n \underline{a}_n g(0, t - nT) + e^{-i\Omega t} \sum_n \underline{b}_n g(0, t - nT) , \quad (7.2)$$

---

<sup>3</sup>The existence of temporal correlations can also be seen from the power spectral density of NLIN, which has been considered in [23, 24].

<sup>4</sup>For simplicity, Eq. (7.2) assumes that the modulated polarization axes of the interacting channels are parallel to each other. While this assumption significantly simplifies the analysis it does not affect the results presented in this paper. In particular, the NLIN power and its dependence on modulation format are independent of the relative rotation between the channels. The magnitude of phase-noise may change with the the relative orientation of the channels, but the major conclusions regarding the significance of phase-noise and its dependence on modulation format remain valid.



where the symbols  $\underline{a}_n$  and  $\underline{b}_n$  represent the transmitted data in the channel of interest and interfering channel, respectively, in the  $n$ -th time-slot, and  $T$  is the symbol duration. The underlined quantities in Eq. (7.2) and throughout this paper are column vectors, whose two elements correspond to the two polarization components of the field. The injected fundamental symbol waveform is denoted by  $g(0, t)$  and it is assumed to be normalized in energy and orthogonal with respect to time shifts by an integer number of symbol durations, i.e.  $\int g^*(0, t - nT)g(0, t - n'T)dt = \delta_{n,n'}$ . The dispersed waveform of the individual pulse when reaching point  $z$  along the fiber is  $g(z, t) = \exp(-\frac{i}{2}\beta''z\frac{\partial^2}{\partial t^2})g(0, t)$ , where  $\beta''$  is the fiber dispersion coefficient<sup>5</sup>.

We now focus, without loss of generality, on the detection of the zeroth-index data symbols  $\underline{a}_0$ . We assume coherent detection preceded by a matched filter that isolates the channel of interest and removes the effect of chromatic dispersion. We note that throughout this work we assume a matched filter that is optimized to the unperturbed system, namely one without the nonlinear interference. The received signal samples in the two polarizations can then be expressed as  $\underline{r}_0 = \underline{a}_0 + \Delta\underline{a}_0$ , where  $\Delta\underline{a}_0$  represents the NLIN contribution. It can be shown that within the first order perturbation analysis [27], the results of [8, 27] can be readily extended to the two-polarization case, yielding

$$\begin{aligned}\Delta\underline{a}_0 &= i\frac{8}{9}\gamma \sum_{l,k,m} S_{l,k,m} \underline{a}_k^\dagger \underline{a}_m \underline{a}_l \\ &\quad + i\frac{8}{9}\gamma \sum_{l,k,m} X_{l,k,m} (\underline{b}_k^\dagger \underline{b}_m \mathbf{I} + \underline{b}_m \underline{b}_k^\dagger) \underline{a}_l,\end{aligned}\tag{7.3}$$

where  $\gamma$  is the fiber's nonlinearity coefficient,  $\mathbf{I}$  is a  $2 \times 2$  identity matrix, and the coefficients  $S_{l,k,m}$  and  $X_{l,k,m}$  are given by

$$\begin{aligned}S_{l,k,m} &= \int_0^L dz f(z) \int_{-\infty}^{\infty} dt g^*(z, t) g(z, t - lT) \\ &\quad \times g^*(z, t - kT) g(z, t - mT)\end{aligned}\tag{7.4}$$

and

$$\begin{aligned}X_{l,k,m} &= \int_0^L dz f(z) \int_{-\infty}^{\infty} dt g^*(z, t) g(z, t - lT) \\ &\quad \times g^*(z, t - kT - \beta''\Omega z) g(z, t - mT - \beta''\Omega z).\end{aligned}\tag{7.5}$$

Here  $L$  is the length of the link and the function  $f(z)$  accounts for its loss/gain profile. For example, it is equal to 1 in the case of perfectly distributed amplification and  $f(z) = \exp(-\alpha \text{mod}(z, L_s))$  in the case of lumped amplification (where  $\alpha$  is the loss coefficient,  $L_s$  is the span-length, and where  $\text{mod}(z, L_s)$  is the modulo operation). The first and second summations on the right-hand side of Eq. (7.3) represent intra-channel and inter-channel interference, respectively. As mentioned earlier, we shall focus in what follows on the NLIN due to inter-channel interactions and ignore the terms proportional to  $S_{l,k,m}$ .

---

<sup>5</sup>Note the change in notation with respect to [8], where the waveform which we now denote by  $g(z, t)$  was denoted by  $g^{(0)}(z, t)$ .

Equation (7.3) illuminates two relevant features of NLIN, whose significance will be quantified in Sec. 7.4. These are the existence of a phase-noise component within the NLIN and the dependence of the NLIN on the modulation format of the interfering channels. The phase-noise contribution arrives from the diagonal elements of the  $2 \times 2$  matrices  $X_{0,k,m}(\underline{b}_k^\dagger \underline{b}_m \mathbf{I} + \underline{b}_m \underline{b}_k^\dagger)$  corresponding to  $l = 0$ . These contributions can be divided into two subgroups. The first subgroup (which constitutes the largest contribution to phase-noise [8, 28]) consists of terms where  $k = m$ , whereas the second subgroup consists of the terms where  $k$  and  $m$  are different. Since the coefficients  $X_{0,m,m}$  can be readily seen to be real-valued, the diagonal contributions from the first group are at a complex angle of  $\pi/2$  from the detected symbol  $\underline{a}_0$ , and hence they constitute phase-noise. A similar conclusion can be drawn for the terms in which  $k \neq m$ , when noting that  $X_{0,k,m} = X_{0,m,k}^*$  and after summing the contributions in which  $k$  and  $m$  are permuted with each other.

The dependence of the NLIN on modulation format follows from the terms in Eq. (7.3) where  $k = m$ . This dependence is particularly significant to the magnitude of the phase-noise part of NLIN, which (as mentioned earlier) is dominated by this kind of terms. Indeed, when the interfering channel is phase-modulated (e.g. QPSK), the phase-noise experienced by the channel of interest has been demonstrated [8, 14] to be much smaller than it is when the amplitude of the interfering channel is modulated as well. The dependence of the overall inter-channel NLIN variance on modulation format has been explored in [8], where it has been shown that when the symbols transmitted over the two channels are independent, identically and isotropically distributed, this variance can be expressed as

$$\sigma_{\text{NLIN}}^2 = \underbrace{P^3 \chi_1}_{\text{SON/GN}} + \underbrace{P^3 \chi_2 \left( \frac{\langle |b|^4 \rangle}{\langle |b|^2 \rangle^2} - 2 \right)}_{\text{FON}}, \quad (7.6)$$

where  $P$  is the average input power (assuming it is equal in both WDM channels),  $b$  represents a single-polarization constellation point of the interfering channel, and the angled brackets denote statistical averaging over the data constellation points<sup>6</sup>. The terms  $\chi_1$  and  $\chi_2$  in a single-polarization transmission are given by Eqs. (26) and (27) in [8] (multiplied by  $T^3$ ), whereas in the case of dual polarizations they need to be multiplied by  $8/27$  and  $20/81$ , respectively. In the appendix of this paper we provide a time-domain representation for these terms and illustrate the relation with the frequency domain analysis adopted in [8].

The first term on the right-hand-side of (7.6) is identical to the value that the GN model [9, 19, 20] assigns to the variance of the inter-channel NLIN, and since it follows only from second-order statistics, we refer to it as the second-order noise (SON) term. The second term depends on fourth-order statistics and is hence referred to as the fourth-order noise (FON) term. The presence of the fourth moment  $\langle |b|^4 \rangle$  in the FON term is responsible for the dependence on modulation format. The value of  $(\langle |b|^4 \rangle / \langle |b|^2 \rangle^2 - 2)$  is  $-1$ ,  $-1.32$ ,  $-1.38$  and  $0$  for the cases of QPSK, 16-QAM, 64-QAM, and Gaussian modulation, respectively. Note that only with Gaussian modulation the second term in Eq. (7.6) vanishes and the result of the GN model becomes exact.

---

<sup>6</sup>As is customary in the fiber-optics community,  $P$  represents the total average signal power (over both polarizations) and  $\sigma_{\text{NLIN}}^2$  is the sum of the NLIN variances in the two polarizations. It is assumed that the symbols transmitted over the two polarizations are statistically independent of each other and that the modulation formats are identical.

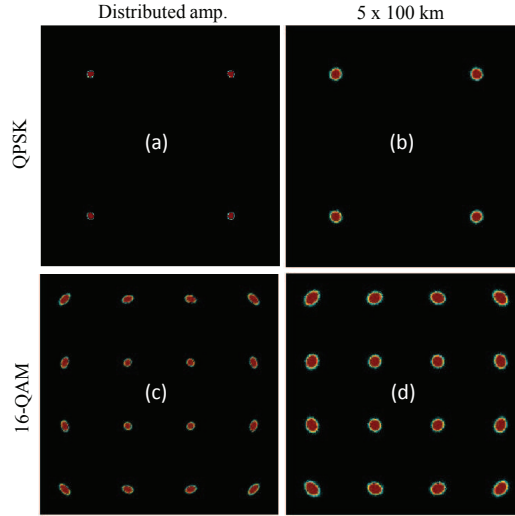


Figure 7.3: Received constellations in a  $5 \times 100$  km link with distributed amplification (left column) and with lumped amplification (right column), for polarization multiplexed transmission of 5 WDM channels using QPSK (top row) and 16-QAM (bottom row) modulation formats (constellations are shown after optimal single-channel back-propagation).

It has been recently pointed out in [18] that in the limit where the WDM channel spacing approaches the bandwidth of the individual WDM channel, Eq. (7.6) acquires a small inaccuracy that can be eliminated by adding additional correction terms. This additional contribution to NLIN is caused by the nonlinear spectral broadening of the nearest neighboring channels, which in the limit of very tight channel spacing, results in the penetration of these channels' spectra into the bandwidth of the channel of interest. The significance of these correction terms diminishes quickly with the channel separation, justifying their neglect when the channel spacing is greater than  $\sim 1.2B$ , where  $B$  is the individual channel bandwidth.

An effective algorithm allowing a practically instantaneous computation of  $\chi_1$  and  $\chi_2$  is provided in the appendix of [14], and a web-based application that conveniently computes the variance of NLIN (including intra-channel NLIN contributions and the correcting terms of [18]) is available at [11].

## 7.4 Modulation-format dependence and the importance of nonlinear phase-noise

In Figs. 7.3–7.5 we consider a system operating over 100km standard SMF spans, with five 32 Giga-baud channels positioned on a 50 GHz ITU grid. The average input power was -0.3dBm and -7dBm for lumped and distributed amplification, respectively, so as to have constant path-averaged power per-channel of -7dBm in both cases (use of a constant path-averaged power is customary when comparing systems with different span lengths [29]). At the receiver, the channel of interest was isolated with a matched optical filter and ideally back-propagated so as to eliminate the effects of self-phase-modulation (neglecting the non-

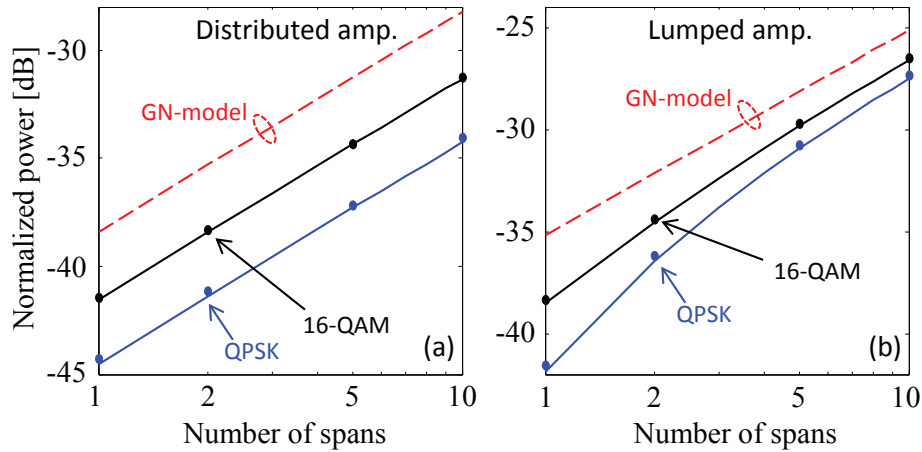


Figure 7.4: The NLIN power (normalized to the received power) vs. the number of spans for  $5 \times 32$  Giga-baud channels using polarization multiplexed QPSK and 16-QAM modulation. Span-length is 100 km and the channel spacing is 50 GHz. Solid curves show the theoretical predictions of Eq. (7.6) whereas the dots represent simulations. Dashed (red) curves show the prediction of the GN model  $P^3\chi_1$  (without intra-channel effects).

linear interaction between the ASE noise and the signal) and chromatic dispersion. These figures are similar to some of the figures presented in [14], except that here they were computed for the polarization multiplexed case. The detailed simulation parameters can be found in section 7.5.2.

The phase-noise character of NLIN is examined in Fig. 7.3, where we show the received constellations corresponding to two  $5 \times 100$ km systems, one with distributed amplification (left panels), and one with lumped amplification (right panels). The two top panels correspond to the case where all channels are QPSK modulated, whereas in the bottom panels the channels are modulated in the 16-QAM format. Consistently with our discussion in the previous section, the phase-noise is much smaller in the case of QPSK modulation than it is in the case of 16-QAM. In addition, the phase-noise in the case of distributed amplification is notably more dominant than it is in the lumped amplification case, although it is clearly visible in both cases.

In Fig. 7.4 the normalized NLIN power for QPSK and 16-QAM modulation formats is plotted as a function of the number of spans. Solid curves show the theoretical predictions according to Eq. (7.6), dots represent the results obtained in a full simulation, whereas dashed curves show the GN model's predictions. The excellent agreement between the analytical predictions and the simulation results is self-evident. With lumped amplification, the difference between the various modulation formats is largest in the case of single-span transmission while it gradually reduces with the number of spans, whereas with distributed amplification the difference is practically independent of the link length.

The observed differences in the nature of NLIN can be attributed to the dynamics of collisions between pulses propagating in different WDM channels [14], as detailed in [28]. When pulses pass smoothly over each other, the collision is considered complete, whereas incomplete pulse collisions occur at points of power discontinuity [30]. In single-span links, or

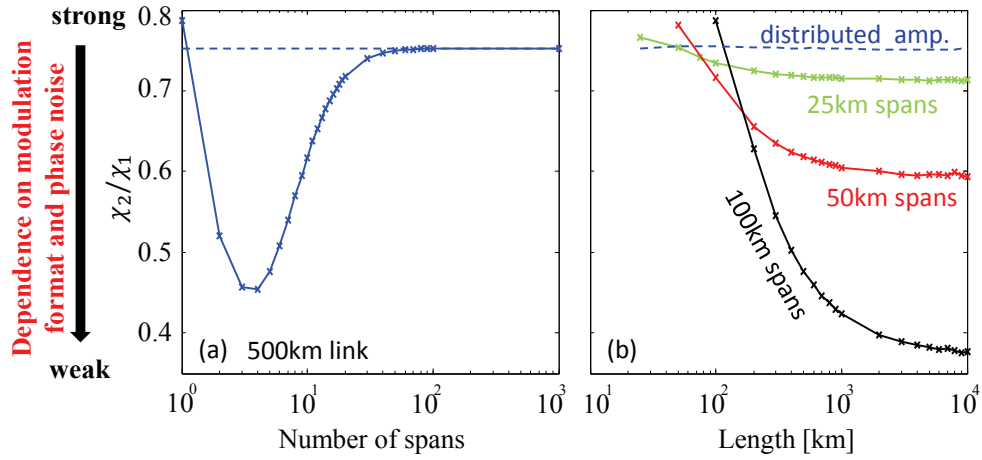


Figure 7.5: The relative weight of complete collisions (represented by the ratio  $\chi_2/\chi_1$ ) in a polarization multiplexed system. When this ratio is large complete collisions dominate and the effects of phase-noise, modulation dependence and inaccuracy of GN model are maximized. (a) for a fixed link length of 500km (dashed curve show the asymptotic result of distributed amplification) (b) fixed span-length (the horizontal axes represents the total length of the link).

in links deploying distributed amplification, almost all pulse collisions are complete, whereas incomplete collisions occur only at the link beginning and end. Conversely, in multi-span links using lumped amplification, incomplete collisions occur at every amplification site. It can be shown that while the dominant effect of complete collisions is that of phase-noise, incomplete collisions mostly produce incoherent circularly distributed disturbances that are not modulation format dependent. The relative significance of complete collisions is represented by the ratio between the FON and SON coefficients  $\chi_2/\chi_1$  [28]. When most collisions are complete, the SON and FON coefficients become very similar to each other<sup>7</sup> implying strong dependence on modulation format, large phase-noise, and the inadequacy of the standard GN model represented by the SON term in Eq. (7.6).

Figure 7.5 illustrates the ratio  $\chi_2/\chi_1$ , showing the relative importance of complete collisions as a function of the number of spans. In Fig. 7.5(a) the overall system length is fixed to 500 km, whereas in Fig. 7.5(b) the span-length is fixed to the specified values. Looking at Fig. 7.5(a), it is quite evident that complete collisions dominate in the case of very few spans, or when the number of spans is large to the extent that the link gradually approaches the conditions of distributed amplification, namely when the span-lengths shorten to the extent that the power discontinuity at the amplifier sites diminishes. In between, when the span-length is of the order of 100 km, incomplete collisions are dominant — a regime in which the phase-noise, the modulation format dependence and the inaccuracy of the GN model are minimized. Figure 7.5(b) shows that for a fixed span-length, the ratio  $\chi_2/\chi_1$  decreases with the length of the link, as can be expected from the fact that as the link becomes longer the number of incomplete collisions increases accordingly.

<sup>7</sup>In the dual polarization case the ratio  $\chi_2/\chi_1$  approaches 5/6 when complete collisions are dominant. In the single-polarization case  $\chi_2 \sim \chi_1$  in this limit [14].

<b>Metro systems</b> (less than 1000 km) <b>OR Raman amplified</b> (of any length)	<b>Long-Haul systems</b> (more than 1000 km, no Raman)
<ul style="list-style-type: none"> <li>• Strong dependence on modulation format</li> <li>• Distinct phase-noise-like nature</li> <li>• The complete model [8,18] is needed for the accurate estimation of NLIN variance</li> </ul>	<ul style="list-style-type: none"> <li>• Weak dependence on modulation format</li> <li>• Mostly complex circular noise</li> <li>• The simple GN model provides sufficiently accurate predictions for NLIN variance</li> </ul>

Table 7.1: Summary of the main properties of inter-channel NLIN in two major categories of dispersion-uncompensated systems

The difference between the various span-lengths considered in Fig. 7.5(b) is once again consistent with the observations regarding the magnitude of the power discontinuity at the amplifier sites. With distributed amplification the ratio  $\chi_2/\chi_1$  is independent of the length of the link since no incomplete collisions occur during propagation. In the case of lumped amplification, the number of incomplete collisions that occur at the amplification sites grows with the system's length, and the significance of each incomplete collision grows with the magnitude of the power discontinuity, or equivalently with the span-length. That is why the largest drop in the ratio  $\chi_2/\chi_1$  with the system length is observed in the case of the 100 km system, for which the size of the power discontinuity at the amplifications sites is the largest.

Interestingly, as can be explained on the basis of the time-domain picture [28] and the results presented above, fiber-systems can be divided into two categories; those that are dominated by complete collisions, and those that are dominated by incomplete collisions. In typical dispersion-uncompensated links deployed over standard single-mode fiber with no pre-dispersion, the first category consists of Metropolitan-type systems whose reach is limited to less than 1000 km, or systems of any length relying on distributed (Raman) amplification. The second category consists of long optical systems extending over distances well exceeding 1000 km, and using lumped (EDFA) amplification with long (of the order of 100 km) spans. Undersea cable-systems, which use lumped amplification with relatively short (of the order of 50 km) spans, fall between the two categories. The first category of systems is characterized by a relatively strong modulation format dependence, and a visible phase-noise component. In addition, the standard GN model's prediction of the NLIN variance is highly inaccurate in such systems, and the use of the more complete theory of [8, 11, 14, 18] is necessary for assessing the inter-channel NLIN variance. In the second category of systems, the inter-channel NLIN is mostly complex circular, it is weakly dependent on modulation format and its variance is better approximated by the standard GN model. These guidelines are summarized in Table 7.1.

The comparison between the results displayed in Figs. 7.3 – 7.5 and the corresponding

figures in [14], which were obtained in the single polarization case, deserves some attention. Although the qualitative results are similar, the dependence on modulation format and the significance of phase-noise reduce somewhat in the presence of polarization multiplexing. This difference can be explained in terms of the pulse collisions picture, as detailed in [28]. In the most part, complete collisions, whose product in the single polarization case was almost exclusively modulation-format dependent nonlinear phase-noise, may occur not only between copolarized pulses, but also between orthogonally polarized pulses. The latter type of complete collisions has a modulation independent circular noise contribution and therefore the significance of phase-noise and modulation format dependence reduces slightly relative to the single-polarization case. Nonetheless, as Figs. 7.3–7.5 indicate, accounting for the presence of these effects is still highly important.

## 7.5 Time-varying ISI channel

In this section we reformulate the channel model given by Eq. (7.3), presenting the nonlinear interference as a time-varying inter-symbol interference (ISI). A similar model was independently proposed in [26], relying on a logarithmic perturbation analysis (which converges to the same results when the nonlinear perturbations are sufficiently small). We further show that the ISI coefficients are slowly changing, suggesting that they may be tracked, estimated and canceled by the receiver, while using standard ISI mitigation techniques. Finally, we present simulation results and discuss the potential gains of practical NLIN mitigation in systems with lumped amplification, perfectly distributed amplification, and Raman amplification using a practical pumping scheme [31].

### 7.5.1 Channel model

We start by rewriting the inter-channel contribution of Eq. (7.3) in the general case (for an arbitrary symbol  $n$ ) as

$$\Delta \underline{a}_n = \sum_l \mathbf{H}_l^{(n)} \underline{a}_{n+l}, \quad (7.7)$$

where the ISI coefficients are the  $2 \times 2$  matrices  $\mathbf{H}_l^{(n)}$  given by

$$\mathbf{H}_l^{(n)} = i\gamma \frac{8}{9} \sum_{k,m} \left( \underline{b}_{n+k}^\dagger \underline{b}_{n+m} \mathbf{I} + \underline{b}_{n+m} \underline{b}_{n+k}^\dagger \right) X_{l,k,m}. \quad (7.8)$$

Evidently,  $\mathbf{H}_l^{(n)}$  depends on the data symbols that were transmitted over the interfering channel and it is therefore unknown to the receiver. Furthermore, the set of interfering symbols that participate in the summation shown in Eq. (7.8) changes with  $n$  and therefore the ISI matrices are time-varying. This time dependence is indicated by the superscript  $(n)$  in  $\mathbf{H}_l^{(n)}$ .

In the limit of large accumulated dispersion, the dependence of  $\mathbf{H}_l^{(n)}$  on the time  $n$  becomes weak, implying slow variation with time. This property can be attributed to the



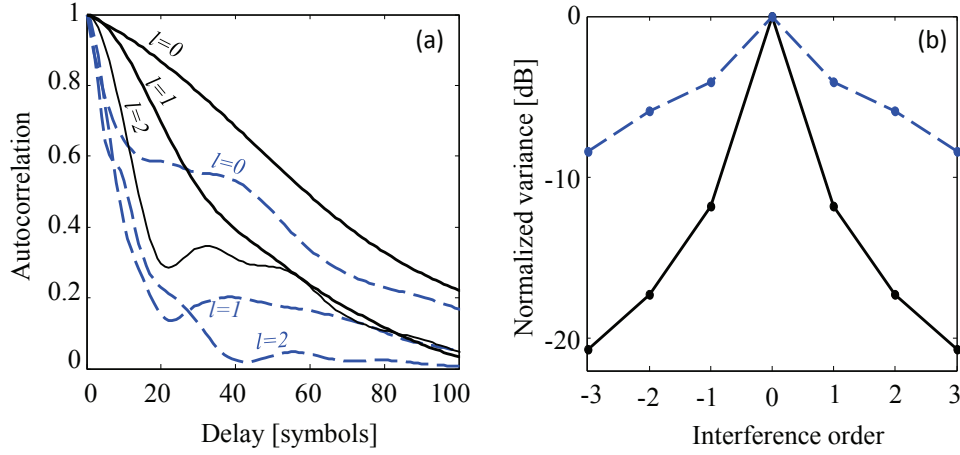


Figure 7.6: (a) The temporal autocorrelation function  $R_l(k)$  of  $\mathbf{H}_{l,1}^{(n)}$  (b) the variance of  $\mathbf{H}_{l,1}^{(n)}$  normalized by the variance of  $\mathbf{H}_{0,1}^{(n)}$ . Computed based on the numerical evaluation of the coefficients  $X_{l,k,m}$  for 16-QAM transmission in a  $5 \times 100$ km link. Similar behavior was observed for all of the elements in the matrix  $\mathbf{H}_l^{(n)}$ . Dashed (blue) and solid (black) curves represent the cases of lumped and distributed amplification.

fact that in the presence of chromatic dispersion, different WDM channels propagate at different velocities so that every symbol in the channel of interest interacts with multiple symbols of the interfering channel. Consequently, adjacent symbols in the channel of interest are disturbed by essentially the same collection of interfering pulses and are affected by nonlinearity in a highly correlated manner. Fig. 7.6(a) shows the autocorrelation function (ACF) of the top diagonal element  $R_l(k) = \langle \mathbf{H}_{l,1}^{(n+k)} \mathbf{H}_{l,1}^{(n)*} \rangle$  for  $l = 0$ ,  $l = 1$ , and  $l = 2$  in the cases of distributed (solid black lines) and lumped (dashed blue lines) amplification. Similar behavior was observed for all elements in the interference matrices. These curves were produced based on the computation of the coefficients  $X_{l,k,m}$  given in Eq. (7.5), for a standard  $5 \times 100$ km system with five 32 Giga-baud channels positioned on a 50 GHz ITU grid (further details on system configuration appear in Sec. 7.5.2). As can be seen in the figure, the elements of the zeroth interference matrix  $\mathbf{H}_0^{(n)}$  exhibits the longest temporal correlations whereas the correlations of the higher-order ISI elements are notably shorter. In addition, it can be seen that the correlations are slightly longer in the case of distributed amplification (especially for  $l > 0$ ) than they are with lumped amplifiers. These results suggest that only few of the lowest orders of interference exhibit sufficiently long temporal correlations which may allow their estimation and cancellation.

Fig. 7.6(b) shows the variance of the top diagonal element in each interference matrix, normalized by the variance of the same diagonal element in the zeroth-order interference matrix  $\langle |\mathbf{H}_{l,1}^{(n)}|^2 \rangle / \langle |\mathbf{H}_{0,1}^{(n)}|^2 \rangle$  (similar behavior was observed for all other elements in the interference matrices). Solid (black) and dashed (blue) curves depict the cases of distributed and lumped amplification, respectively. While the zeroth-order interference term is the largest in both cases, the rate at which the significance drops with the interference order is much larger in the distributed amplification case. For example the variance of the third-order interference



is smaller than the zeroth-order interference by  $\sim 20\text{dB}$  in the distributed amplification case whereas it is smaller by only  $\sim 10\text{dB}$  in the case of lumped amplification. This implies that the benefit of canceling only the zeroth order would be much greater with distributed amplification than it would be when lumped amplification is used. The fact that the dominance of the zeroth interference term is larger in the distributed amplification case is consistent with the observations made in the previous section regarding the larger significance of phase-noise in this situation. This is because, as we discussed in Sec. 7.3, the diagonal terms  $\mathbf{H}_{01,1}^{(n)}$  and  $\mathbf{H}_{02,2}^{(n)}$  can be identified as contributing to phase-noise.

In what follows we present simulation results of practical NLIN mitigation, employing standard ISI estimation, tracking and cancelation techniques, focusing exclusively on the cancelation of the zeroth order matrix  $\mathbf{H}_0^{(n)}$  as it has the strongest impact and it exhibits the longest temporal correlations. Attempts to compensate for higher ISI orders produced negligible benefits in the lumped amplification case. Since the potential gain in systems with distributed amplification is much larger than it is in systems with lumped amplification, a practical compromise on which we test the performance of our mitigation, is a standard Raman amplified system, with parameters taken from [31]. As we show in what follows, the performance improvement extracted from NLIN mitigation in this case is quite notable. We further note that while the cancelation of the diagonal elements in the zeroth-order interference matrix  $\mathbf{H}_0^{(n)}$  amounts to canceling the phase-noise part of the NLIN, cancelation of the off-diagonal elements amounts to canceling circular NLIN originating from the lowest order cross-polarization interference.

### 7.5.2 NLIN mitigation - simulation results

We have performed a series of simulations assuming a five-channel polarization multiplexed WDM system implemented over standard single-mode fiber (dispersion of  $17\text{ ps/nm/km}$ , nonlinear coefficient  $\gamma = 1.3\text{ [Wkm]}^{-1}$ , and attenuation of  $0.2\text{dB per km}$ ). The transmission consisted of Nyquist pulses at  $32\text{ Giga-baud}$  and a standard channel spacing of  $50\text{ GHz}$ . The modulation format used by all WDM channels was 16-QAM where the number of simulated symbols in each run and for each polarization was 4096 when simulating a  $500\text{ km}$  system, and 8192 when simulating  $6000\text{ km}$ . More than 100 runs (each with independent and random data symbols) were performed with each set of system parameters, so as to accumulate sufficient statistics. In the Raman amplified cases, a combination of co-propagating and counter-propagating Raman pumps was assumed, providing gains of  $5\text{ dB}$  and  $15\text{ dB}$ , respectively (as was done in [31]). At the receiver, the channel of interest was isolated with a matched optical filter and (perfectly, using 16 samples per symbol) back-propagated so as to eliminate the effects of self-phase-modulation and chromatic dispersion. The noise figure of the lumped amplifiers was  $4\text{dB}$  and the Raman noise was generated with  $n_{sp} = 1.26$  (corresponding to a local NF of  $4\text{dB}$ ). In order to concentrate on the nonlinear noise mitigation aspects, we assumed ideal carrier phase and frequency estimation. To cancel the effect of the interference matrix  $\mathbf{H}_0^{(n)}$  we trained and tracked a canceling matrix, using a standard decision-aided approach based on the recursive least squares (RLS) algorithm [32]. The RLS algorithm converged faster than the more familiar least mean-squares (LMS) algorithm, and with a small filter order, the added complexity is expected to be tolerable in future DSP used in

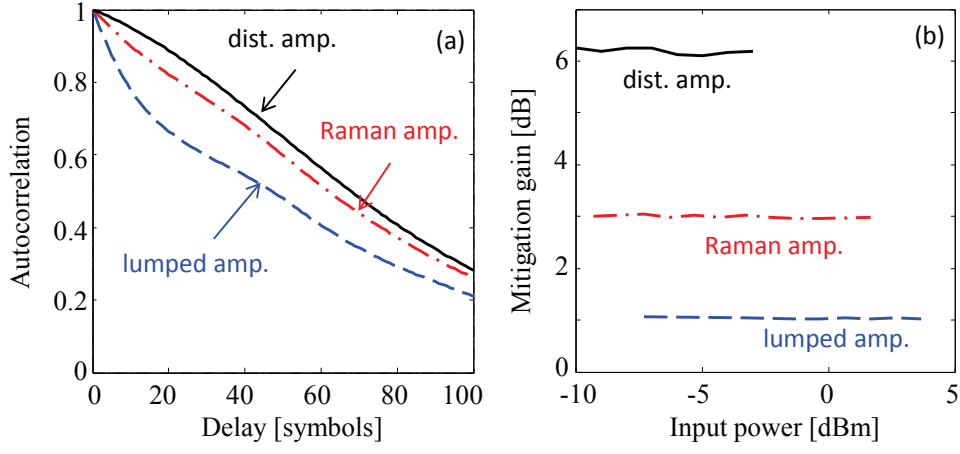


Figure 7.7: (a) The temporal autocorrelation function  $R_0(k)$  of  $\mathbf{H}_{0,1}^{(n)}$ , as extracted from simulations (similar behavior was observed for all elements in  $\mathbf{H}_0^{(n)}$ ). (b) The ratio between the NLIN powers before and after mitigation. Dashed (blue), dashed-dotted (red), and solid (black) curves represent the cases of lumped, Raman, and perfectly distributed amplification. Here, unlike in Fig. 7.8, the extraction of the ISI coefficient matrices was performed without ASE noise.

fiber optics. The algorithm uses a “forgetting factor”  $\lambda$  which weights the past samples exponentially, where a larger factor allows faster adaptation while a smaller factor results in higher stability and convergence. In the simulations we used  $\lambda = 0.98, 0.96, 0.9$ , specified for the cases of lumped, Raman and distributed amplification, respectively.

Due to the very long computation times that are implied by the need to monitor error-rates, we first test our insight regarding time correlations and amplification strategies on a relatively short link, and then simulate a long Raman amplified link so as to characterize the effect on BER. In Fig. 7.7(a) we plot the autocorrelation function of one of the diagonal elements of  $\mathbf{H}_0^{(n)}$  in the three amplification strategies after  $5 \times 100\text{km}$  of propagation (the extracted ACF of all elements of  $\mathbf{H}_0^{(n)}$  was practically indistinguishable). As can be seen in the figure, the correlations extend over a few tens of symbols, which is the key factor behind the implementation of the proposed mitigation approach. In addition, careful examination of the curves for the cases of lumped and distributed amplification shows a good alignment with the theoretical predictions provided in Fig. 7.6(a). In Fig. 7.7(b) we show the NLIN mitigation gain, defined as the ratio between the NLIN variances before and after mitigation. As is evident from the figure, the gain from canceling the zeroth interference term is 6dB, 3dB and 1dB in the cases of distributed, Raman and lumped amplification, respectively. In Fig. 7.8(a) we show the effective SNR curves with (solid) and without (dotted) NLIN mitigation. The peak SNR after mitigation is larger than the peak SNR prior to mitigation by 0.3 dB with lumped amplification, 0.9 dB with Raman amplification and 1.6 dB with distributed amplification. The dashed curves in the figure represent the case in which mitigation addresses only the diagonal terms of  $\mathbf{H}_0^{(n)}$ , showing that the advantage of accounting for NLIN induced interference between polarizations is notable.

In Fig. 7.8(b) we address the consequences of mitigation in terms of the system BER.

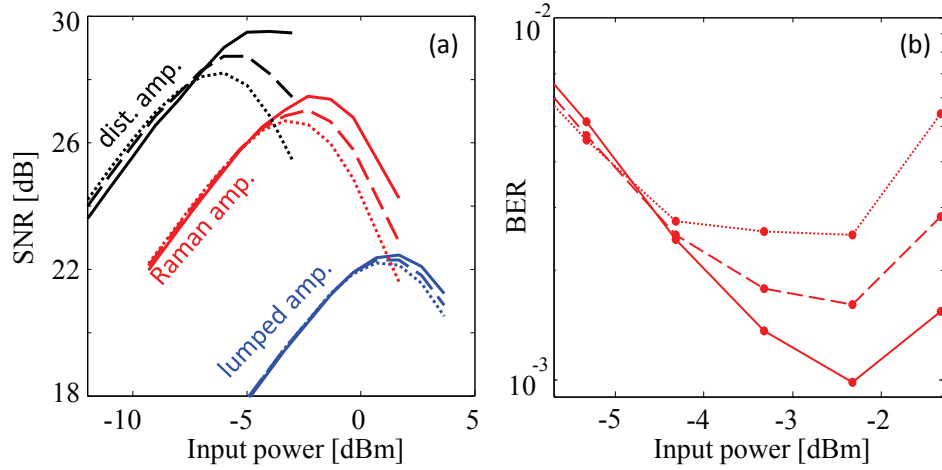


Figure 7.8: (a) Effective SNR in a  $5 \times 100\text{km}$  link. Black, red and blue curves correspond to distributed, Raman, and lumped amplification, respectively. Without inter-channel NLIN mitigation (dotted), after mitigation of the zeroth NLIN term  $\mathbf{H}_0^{(n)}$  (solid), and after cancellation of only the diagonal elements in  $\mathbf{H}_0^{(n)}$  (dashed). Evidently, the gains in the peak effective SNR are 0.3 dB (lumped), 0.9 dB (Raman) and 1.6 dB (distributed), where approximately half is attributed to the mitigation of the diagonal elements of  $\mathbf{H}_0^{(n)}$ . In the linear regime, imperfect estimation of the equalizer coefficients (due to ASE) is seen to slightly spoil performance. (b) BER as a function of input power in a 6000 km Raman amplified system. Dotted – without mitigating inter-channel NLIN, solid – mitigation of  $\mathbf{H}_0^{(n)}$ , dashed – mitigation of only the diagonal elements in  $\mathbf{H}_0^{(n)}$ . At very low powers mitigation spoils performance as a result of poor estimation caused by ASE.

Since truly distributed amplification is of limited practical interest, we chose to concentrate this highly computationally intensive set of simulations on the case of the practical Raman-amplified scenario, where the gain of mitigation is higher than it is in the lumped amplification case. The simulations in this case were conducted for  $60 \times 100$  km spans, so as to reach relevant BER values. The dotted curve represents the case without mitigation and the solid curve represents the results obtained after canceling the effect of the zeroth interference term  $\mathbf{H}_0^{(n)}$ , showing an improvement of the minimum BER from  $2.6 \times 10^{-3}$  to less than  $1 \times 10^{-3}$ , equivalent to approximately 1dB gain in the Q factor. Equivalently, to achieve the same BER without NLIN mitigation the OSNR would have to be increased by 2.3dB. Similarly to Fig. 7.8(a), the dashed curve shows the result obtained when only the effect of the diagonal terms in  $\mathbf{H}_0^{(n)}$  was mitigated, showing that approximately half of the mitigation gain is attributed to cross polarization interference.

## 7.6 Conclusions

We reviewed the time-domain model of NLIN in WDM communications systems, while explicitly extending the fundamental equations to the polarization multiplexed case. We discussed the dependence of the NLIN on modulation format and the significance of its

phase-noise component, as well as the accuracy and range of validity of the various models for predicting the variance of NLIN. We showed that both the dependence on modulation format and the phase-noise nature of inter-channel NLIN are particularly important in metro-length systems, or in systems of any reach that use short spans, or Raman amplification. In the case of long-haul systems (well over 1000 km) operating with long spans and using lumped (EDFA) amplification the significance of these phenomena is considerably smaller. The observed phenomena were related to the pulse collisions picture, whose details are provided separately [28]. Finally, we introduced the slowly varying ISI model for the nonlinear channel and examined the improvement that can be achieved in system performance when mitigating inter-channel NLIN by means of adaptive linear equalization (without joint processing of multiple WDM channels). We note that the proposed mitigation scheme has the advantage of using the same type of hardware currently used for equalizing polarization effects, yet the equalization algorithm and the required speed of convergence are substantially different.

## Acknowledgement

The authors would like to acknowledge financial support from the Israel Science Foundation (grant 737/12). Ronen Dar would like to acknowledge the support of the Adams Fellowship of the Israel Academy of Sciences and Humanities, the Yitzhak and Chaya Weinstein Research Institute, the Advanced Communication Center and the Electro-Optic Fund of Tel-Aviv University, and the Terasanta consortium.

## 7.7 Appendix: Time and frequency-domain representations of $\chi_1$ and $\chi_2$

The process of evaluating the variance of the NLIN involves the calculation of the second moment  $\langle \Delta \underline{a}^\dagger \Delta \underline{a} \rangle$  from the (inter-channel) summation appearing in Eq. (7.3). The idea is that the FON coefficient  $\chi_2$  can be found by identifying all terms in the second sum in Eq. (7.3) that contribute to the existence of a fourth moment  $\langle |b|^4 \rangle$ . Then, the term  $\chi_1$  is found such that the terms proportional to the second moment  $\langle |b|^2 \rangle^2$  are accounted for, in accordance with the form of Eq. (7.6). A cumbersome, yet straightforward calculation, produces the results

$$\chi_1 = \frac{32}{27} \gamma^2 \sum_{l,k,m} |X_{l,k,m}|^2 \quad (7.9)$$

$$\chi_2 = \frac{80}{81} \gamma^2 \sum_{l,m} |X_{l,m,m}|^2, \quad (7.10)$$

where the nonlinear coefficients  $X_{l,k,m}$  are given in Eq. (7.5).

The frequency-domain representations of  $\chi_1$  and  $\chi_2$  can now be easily derived from the time-domain representations given in Eqs. (7.9–7.10). Using the fact that  $g(L, t)$  is equal to  $\int \tilde{g}(\omega) \exp(-i\omega t + i\beta''\omega^2 L/2) d\omega / (2\pi)$ , where  $\tilde{g}(\omega)$  is the inverse Fourier transform of the

fundamental pulse  $g(t)$ , the expression for  $X_{l,k,m}$  can be rewritten as follows

$$X_{l,k,m} = \int \rho(\omega_1, \omega_2, \omega_3) e^{i(\omega_1 l - \omega_2 k + \omega_3 m)T} \frac{d^3\omega}{(2\pi)^3}, \quad (7.11)$$

where  $d^n\omega$  stands for  $d\omega_1 \cdots d\omega_n$  and where

$$\begin{aligned} \rho(\omega_1, \omega_2, \omega_3) &= \tilde{g}^*(\omega_1 - \omega_2 + \omega_3) \tilde{g}(\omega_1) \tilde{g}^*(\omega_2 - \Omega) \tilde{g}(\omega_3 - \Omega) \\ &\times \int_0^L f(z) e^{i\beta''(\omega_2 - \omega_3)(\omega_2 - \omega_1)z} dz. \end{aligned} \quad (7.12)$$

Note that within the time-domain picture, the various contributions to NLIN are proportional to  $X_{l,k,m}$ , whereas the frequency-domain representation consists of the contributions of multiple FWM products that are proportional to  $\rho(\omega_1, \omega_2, \omega_3)$  (each contribution is produced by FWM between the frequency tone at  $\omega_1$  in the channel of interest and the frequency tones at  $\omega_2$  and  $\omega_3$  in the interfering channel). Equation (7.11) reveals an interesting relation between the time-domain and frequency-domain analyses where  $X_{l,k,m}$  and  $\rho(\omega_1, \omega_2, \omega_3)$  are related through the 3-dimensional Fourier transform.

Substitution of Eq. (7.11) into Eq. (7.9) results in

$$\begin{aligned} \chi_1 &= \frac{32}{27} \gamma^2 \sum_{l,k,m} \int e^{i(\omega_1 - \omega'_1)lT} e^{-i(\omega_2 - \omega'_2)kT} e^{i(\omega_3 - \omega'_3)mT} \\ &\times \rho(\omega_1, \omega_2, \omega_3) \rho^*(\omega'_1, \omega'_2, \omega'_3) \frac{d^3\omega}{(2\pi)^3} \frac{d^3\omega'}{(2\pi)^3}, \end{aligned} \quad (7.13)$$

where for simplicity we assumed perfect Nyquist pulses so that  $\tilde{g}(\omega)\tilde{g}(\omega \pm 2\pi/T) = 0$ . In this case, by using the identity

$$\sum_{k=-\infty}^{\infty} e^{ikT\omega} = \frac{2\pi}{T} \sum_{n=-\infty}^{\infty} \delta\left(\omega - \frac{2\pi n}{T}\right) \quad (7.14)$$

we find that

$$\chi_1 = \frac{32}{27} \frac{\gamma^2}{T^3} \int |\rho(\omega_1, \omega_2, \omega_3)|^2 \frac{d^3\omega}{(2\pi)^3}. \quad (7.15)$$

In the same manner it can be shown that  $\chi_2$  is equal to

$$\chi_2 = \frac{80}{81} \frac{\gamma^2}{T^2} \int \rho(\omega_1, \omega_2, \omega_3) \rho^*(\omega_1, \omega_4, \omega_4 + \omega_3 - \omega_2) \frac{d^4\omega}{(2\pi)^4}. \quad (7.16)$$

The SON and FON coefficients  $\chi_1$  and  $\chi_2$  can be therefore calculated by computing a three and four dimension integrals, respectively. Such integration is readily performed using the Monte Carlo integration method [33], whose implementation is described in the appendix of [14].

# Bibliography

- [1] R.-J. Essiambre, G. Kramer, P.J. Winzer, G.J. Foschini, B. Goebel, “Capacity limits of optical fiber networks,” *J. Lightwave Technol.*, vol. 28, pp. 662–701 (2010).
- [2] A. D. Ellis, J. Zhao, D. Cotter, “Approaching the non-linear Shannon limit,” *J. Lightwave Technol.*, vol 28, pp. 423–433 (2010).
- [3] A. Mecozzi, C. B. Clausen, and M. Shtaif, “Analysis of intrachannel nonlinear effects in highly dispersed optical pulse transmission,” *IEEE Photon. Technol. Lett.*, vol 12, pp. 392–394 (2000).
- [4] F. Forghieri, R.W. Tkack, A. R. Chraplyvy, “Fiber nonlinearities and their impact on transmission systems,” Ch. 8 in *Optical Fiber Telecommunications IIIA*, P. Kaminow and T. L. Koch eds. (Academic Press, 1997).
- [5] E. Ip, J. M. Kahn, “Compensation of dispersion and nonlinear impairments using digital backpropagation,” *J. Lightwave Technol.*, vol. 26, pp. 3416–3425 (2008).
- [6] R. I. Killey, P. M. Watts, V. Mikhailov, M. Glick, P. Bayvel, “Electronic dispersion compensation by signal predistortion using digital processing and a dual-drive Mach-Zehnder modulator,” *IEEE Photon. Technol. Lett.*, vol 17, pp. 714–716 (2005).
- [7] E. Ip, “Nonlinear compensation using backpropagation for polarization-multiplexed transmission,” *J. Lightwave Technol.*, vol 28, pp. 939–951 (2010).
- [8] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, “Properties of nonlinear noise in long, dispersion-uncompensated fiber links,” *Optics Express*, vol. 21, pp. 25685–25699 (2013).
- [9] A. Carena, V. Curri, G. Bosco, P. Poggiolini, F. Forghieri, “Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links,” *J. Lightwave Technol.*, vol. 30, pp. 1524–1539 (2012).
- [10] A. Bononi, N. Rossi, P. Serena, “On the nonlinear threshold versus distance in long-haul highly-dispersive coherent systems,” *Optics Express*, vol. 20, pp. B204–B216 (2012).
- [11] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, Nonlinear interference noise wizard. Available at: <http://nlinwizard.eng.tau.ac.il>
- [12] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, F. Forghieri, “The GN-model of fiber non-linear propagation and its applications,” *J. Lightwave Technol.*, vol. 4, pp. 100–127 (2014).

- [13] M. Shtaif, R. Dar, A. Mecozzi, M. Feder, “Nonlinear interference noise in WDM systems and approaches for its cancelation,” in European Conference on Optical Communications (ECOC), Paper We.1.3.1, Cannes (2014).
- [14] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, “Accumulation of nonlinear interference noise in fiber-optic systems,” *Optics Express*, vol. 22, pp. 14199–14211 (2014).
- [15] R. Dar, M. Shtaif, M. Feder, “New bounds on the capacity of the nonlinear fiber-optic channel,” *Optics Lett.*, vol. 39, pp. 398–401 (2014).
- [16] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, “Time varying ISI model for nonlinear interference noise,” in Optical Fiber Communication (OFC) Conference, pp. W2A.62, San Francisco (2014).
- [17] R. Dar, O. Geller, M. Feder, A. Mecozzi, M. Shtaif, “Mitigation of inter-channel nonlinear interference in WDM systems,” in European Conference on Optical Communications (ECOC), P.5.6, Cannes (2014).
- [18] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, F. Forghieri, “EGN model of non-linear fiber propagation,” *Optics Express*, vol. 22, pp. 16335–16362 (2014).
- [19] P. Poggiolini, A. Carena, V. Curri, G. Bosco, F. Forghieri, “Analytical modeling of nonlinear propagation in uncompensated optical transmission links,” *IEEE Photon. Technol. Lett.*, vol. 23, pp. 742–744 (2011).
- [20] P. Poggiolini, “The GN model of non-linear propagation in uncompensated coherent optical systems,” *J. Lightwave Technol.*, vol. 30, pp. 3857–3879 (2012).
- [21] P. Johannisson and M. Karlsson, “Perturbation analysis of nonlinear propagation in a strongly dispersive optical communication system,” *J. Lightwave Technol.*, vol. 31, pp. 1273–1282 (2013).
- [22] P. Serena, A. Bononi, “An alternative approach to the Gaussian noise model and its system implications,” *J. Lightwave Technol.*, vol. 31, pp. 3489–3499 (2013).
- [23] A. Bononi, P. Serena, N. Rossi, “Nonlinear signalnoise interactions in dispersion-managed links with various modulation formats,” *Optical Fiber Technology*, vol. 16, pp. 73–85 (2010).
- [24] A. Bononi, M. Bertolini, P. Serena, G. Bellotti, “Cross-phase modulation induced by OOK channels on higher-rate DQPSK and coherent QPSK channels,” *J. Lightwave Technol.*, vol. 27, pp. 3974–3983 (2009).
- [25] E. Agrell, A. Alvarado, G. Durisi, M. Karlsson, “Capacity of a nonlinear optical channel with finite memory,” *J. Lightwave Technol.*, vol. 32, pp. 2862–2876 (2014).
- [26] M. Secondini and E. Forestieri, “On XPM mitigation in WDM fiber-optic systems,” *IEEE Photon. Technol. Lett.*, vol. 26, pp. 2252–2255, 2014.



- [27] A. Mecozzi and R.-J. Essiambre, “Nonlinear Shannon limit in pseudolinear coherent systems,” *J. Lightwave Technol.*, vol. 30, pp. 2011–2024 (2012).
- [28] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, “Pulse collision picture of nonlinear interference noise in fiber-optic communications,” November 5, 2014 [Online]. Available: arXiv:1411.1204 [physics.optics], submitted to *J. Lightwave Technol.* (2014).
- [29] J. P. Gordon and L. F. Mollenauer, “Effects of fiber nonlinearities and amplifier spacing on ultra-long distance transmission,” *J. Lightwave Technol.*, vol. 9, pp. 170–173 (1991)
- [30] M. Shtaif, “Analytical description of cross-phase modulation in dispersive optical fibers.” *Optics Lett.*, vol. 23, pp. 1191–1193 (1998).
- [31] C. Xie, G. Raybon, P. J. Winzer, “Transmission of mixed 224-Gb/s and 112-Gb/s PDM-QPSK at 50-GHz channel spacing over 1200-km dispersion-managed LEAF spans and three ROADMs,” *J. Lightwave Technol.*, vol. 30, pp. 547–552 (2012).
- [32] S. Haykin, *Adaptive filter theory*, Pearson Education (2005).
- [33] R. E. Caflisch, “Monte Carlo and quasi-Monte Carlo methods,” *Acta Numerica* vol. 7, Cambridge University Press, pp. 1–49 (1998).



# Chapter 8

## Pulse Collision Picture of Inter-Channel Nonlinear Interference in Fiber-Optic Communications

Ronen Dar, Meir Feder, Antonio Mecozzi and Mark Shtaif, “*Pulse collision picture of inter-channel nonlinear interference in fiber-optic communications*,” accepted for publication in IEEE Journal of Lightwave Technology (April 2015)

### 8.1 Abstract

We model the build-up of inter-channel nonlinear interference noise (NLIN) in wavelength division multiplexed systems by considering the pulse collision dynamics in the time domain. The fundamental interactions can be classified as two-pulse, three-pulse, or four-pulse collisions and they can be either complete, or incomplete. Each type of collision is shown to have its unique signature and the overall nature of NLIN is determined by the relative importance of the various classes of pulse collisions in a given WDM system. The pulse-collision picture provides qualitative and quantitative insight into the character of NLIN, offering a simple and intuitive explanation to all of the reported and previously unexplained phenomena. In particular, we show that the most important contributions to NLIN follow from two-pulse and four-pulse collisions. While the contribution of two-pulse collisions is in the form of phase-noise and polarization-state-rotation with strong dependence on modulation format, four-pulse collisions generate complex circular noise whose variance is independent of modulation format. In addition, two-pulse collisions are strongest when the collision is complete, whereas four-pulse collisions are strongest when the collision is incomplete. We show that two-pulse collisions dominate the formation of NLIN in short links with lumped amplification, or in links with distributed amplification extending over arbitrary length. In long links using lumped amplification the relative significance of four-pulse collisions increases, emphasizing the circularity of the NLIN while reducing its dependence on modulation format.

## 8.2 Introduction

The nonlinearity of optical fibers has been long recognized as one of the most important factors limiting the growth of data-rates transmitted in wavelength division multiplexed (WDM) systems [1]. The nonlinear distortions are usually classified as either intra-channel [2], or inter-channel [3]. While intra-channel distortions are generated by each of the WDM channels individually, inter-channel distortions are caused by interference between different WDM channels. In this work we focus on the physical mechanism responsible for inter-channel nonlinear effects as it comes into play in modern, dispersion uncompensated fiber links with coherent detection. Since joint processing of multiple WDM channels is currently considered to be prohibitively complex in commercial systems, inter-channel interference is customarily treated as noise [1], and hence we refer to it in what follows as *nonlinear interference noise*, or NLIN [4–9].

We follow up on the theory published in [4,6,10] and model in the time domain<sup>1</sup> the build-up of NLIN in WDM systems which are using single-carrier transmission. In this approach inter-channel NLIN is attributed to multiple pulse collisions whose dynamics we examine in detail (a similar analysis of intra-channel pulse-interactions can be found in [2,14]). As we demonstrate in what follows, the pulse-collision picture provides deep quantitative and qualitative insight into the build-up of NLIN, and explains the various properties of NLIN that have previously only been observed empirically in simulations. For example, issues like the importance of the phase-noise component of the NLIN versus the legitimacy of describing NLIN as complex circular noise, the dependence of the NLIN power on modulation format, the effect of pre-dispersion, the dependence on system-length and type of amplification, etc. [4–6,15–18], are all clarified once the essence of pulse collision dynamics is understood.

While the practice of analyzing nonlinear interference by means of pulse collisions has been known for a while (initially in the context of soliton [19,20] and dispersion-managed-soliton [21,22] transmission, and subsequently in the more general context of arbitrary pulse interactions [23]), the pulse collision dynamics encountered in modern coherent fiber communications systems, has a number of fundamentally different characteristics. Most prominently, in modern systems, which avoid the use of inline dispersion compensation, simultaneous multi-pulse collisions play a unique and very significant role. In order to illustrate this, consider the case where the channel of interest is perturbed by a nonlinear interaction with a single interfering WDM channel. In the old generation of dispersion managed systems [21–23], interference was generated by collisions between pairs of pulses, one from the channel of interest and one from the interfering channel. In the new generation of dispersion uncompensated systems, pulse spreading generates significant temporal overlap within each of the channels, allowing for example, situations in which two different pulses in the interfering channel interact nonlinearly with a third pulse, belonging to the channel of interest, so as to generate interference that affects a fourth pulse, also belonging to the channel of interest.

In what follows we demonstrate that substantial understanding of a broad range of observed nonlinear interference phenomena can be achieved by means of proper classification of the type of collisions taking place in the fiber. Most relevantly, there is a qualitative and

---

<sup>1</sup>Other time-domain models for inter-channel effects were considered in [11–13]

a quantitative difference between the interference caused by two-pulse, three-pulse and four-pulse interactions. Additionally, the effect of complete collisions (where interacting pulses that belong to different WDM channels pass by each other completely within the fiber span) may differ considerably from the effect of incomplete collisions. In the case of two-pulse collisions, nonlinear interference manifests itself as phase-noise, and polarization-rotation noise, and it is strongly dependent on modulation format. Among the three-pulse collisions some distinctly produce phase and polarization noise, whose variance is modulation format independent, whereas others produce complex circular noise, whose variance is modulation format dependent. Finally, all four-pulse collisions produce complex circular noise whose power is independent of modulation format. While the NLIN produced by two-pulse collisions grows monotonically during the collision process and is maximized when the collision is complete, the behavior of three and four-pulse collisions is distinctly different. The NLIN generated by these collisions is built up constructively in the first half of the collision process and then most of it cancels through destructive interference when the collision is completed. Hence the effect of three and four-pulse collision ends up being much more significant when the collision is incomplete.

The relative importance of complete versus incomplete collisions and therefore the relative importance of two, three and four-pulse collisions, in a given system is determined by the link parameters, predominantly by the length and number of amplified spans, and by the type of amplification that is used. For example, with lumped amplification complete collisions are most significant in short, few-span systems, whereas in the case of distributed amplification (which is the limiting case of Raman amplification [24]) the contribution of complete collisions is dominant regardless of the link length. In links dominated by complete collisions, two-pulse collisions are more significant and the NLIN has a distinct phase and polarization noise nature, and it is characterized by a strong dependence on modulation format. As the significance of incomplete collisions increases, four-pulse collisions become more significant so that the NLIN progressively evolves into complex circular noise, and its dependence on modulation format reduces.

Our analysis in what follows focuses only on nonlinear interference between two WDM channels, one of which is referred to as the channel of interest, and the other is referred to as the interfering channel. In the traditional jargon of WDM systems [25], this would qualify as studying only the effect of cross-phase-modulation (XPM), and omitting inter-channel four wave mixing (FWM) processes in which three or four different WDM channels are involved. As is evident from simulations [4, 16], the contribution of FWM to NLIN in most relevant fiber types is of negligible importance<sup>2</sup>. Under these circumstances, the NLIN that is generated by multiple WDM channels is simply the sum of the NLIN contributions produced by the individual interferers. Furthermore, in this work we concentrate on a regime in which the nonlinear interaction between the signal and the amplified spontaneous emission (ASE) noise is negligible.

The paper is organized as follows. In Sec. 8.3 we briefly review the essentials of the time domain model, initially introduced in [4, 6, 10] and establish the background for the

---

<sup>2</sup>As pointed out in [16], in certain types of non-zero-dispersion-shifted fibers, one can observe a contribution to NLIN that originates from FWM processes. However, even in these situations, the FWM contribution is a relatively small correction to the contribution of XPM, as can be further verified in <http://nlinwizard.eng.tau.ac.il>.

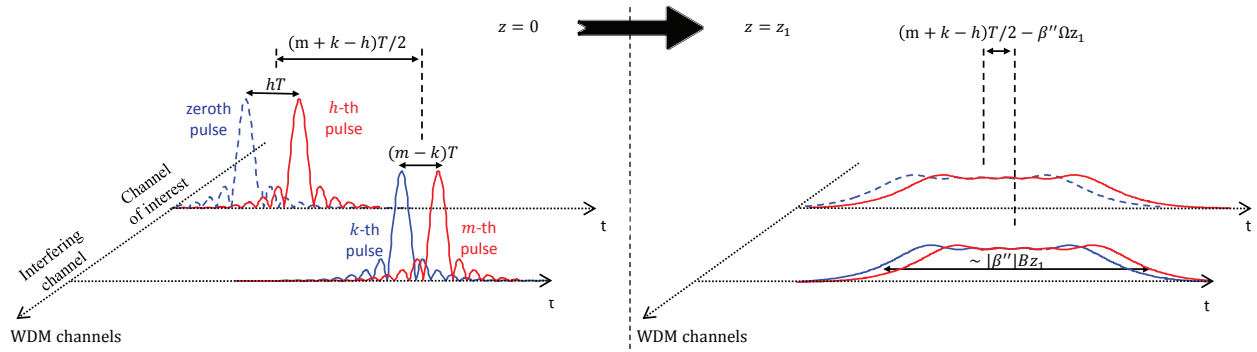


Figure 8.1: The generic four-pulse interference. The  $h$ -th pulse in the channel of interest interacts with the  $k$ -th and  $m$ -th pulses of the interfering channel so as to create NLIN that affects the zeroth symbol in the channel of interest. Since we assume detection that involves filtering matched to the waveform of the fundamental pulse centered at the zeroth symbol, as discussed in Sec. 8.3, one can refer to this as a four-pulse interaction. The zeroth pulse is plotted only to keep track of the matched-filter waveform, and it is plotted by a dashed line in order to stress that it is not a pulse that is physically participating in the nonlinear interaction during propagation. The left panel shows the situation in the beginning of the link, before the pulses overlap, whereas the right panel illustrates the situation once an overlap between the pulses is formed. When two of the indices overlap ( $h = 0$ , or  $k = m$ ) one observes a three-pulse interaction and when both index pairs overlap ( $h = 0$  and  $k = m$ ) one obtains a two-pulse interaction.

subsequent analysis. In Secs. 8.4, 8.5, and 8.6 we classify the nonlinear interference processes as two, three and four-pulse collisions, respectively. In each case we describe their distinctive features depending on whether the collision is complete, or incomplete. Several numerical examples illustrating the main ideas are presented in Sec. 8.7. For the simplicity of illustration Secs. 8.3–8.6 consider scalar (single-polarization) transmission. The extension to the polarization multiplexed case is presented in Sec. 8.8. In Sec. 8.9 we discuss the implications of the pulse-collision theory to fiber-communications systems, whereas in Sec. 8.10 we explain the role of chromatic dispersion and relate the time domain theory to the Gaussian noise (GN) model [26–30]. Section 8.11 is devoted to conclusions.

### 8.3 Time-domain theory

For the simplicity of illustration we start by considering the single-polarization case, postponing the generalization to polarization multiplexed systems to Sec. 8.8. In the case of two WDM channels, we may express the electric field at the fiber input as

$$\sum_n a_n g(0, t - nT) + \exp(-i\Omega t) \sum_n b_n g(0, t - nT), \quad (8.1)$$

where  $T$  is the symbol duration and  $a_n$  and  $b_n$  are the complex valued data symbols transmitted over the channel of interest and over the interfering channel, respectively. The central

frequency of the channel of interest is arbitrarily set to 0, whereas the central frequency of the interfering channel is denoted by  $\Omega$ . The injected fundamental symbol waveforms in both channels are identical and denoted by  $g(0, t)$ , whereas  $g(z, t) = \exp(-\frac{i}{2}\beta''z\frac{\partial^2}{\partial t^2})g(0, t)$  is the dispersed waveform of the individual pulse when reaching point  $z$  along the fiber with  $\beta''$  being the fiber dispersion coefficient<sup>3</sup>. The fundamental waveforms are further assumed to be orthogonal with respect to time shifts by an integer number of symbol durations and their energy is assumed to be normalized to 1

$$\int g^*(z, t - nT)g(z, t - n'T)dt = \delta_{n,n'}. \quad (8.2)$$

After coherent detection, the channel of interest is match-filtered with the filter's impulse response being proportional to  $g^*(L, t)$ , where  $L$  is the link length. The extracted  $n$ -th data symbol can be then expressed as  $a_n + \Delta a_n$ , with  $\Delta a_n$  accounting for the presence of NLIN. In [4] we have shown that  $\Delta a_n$  due to inter-channel NLIN is given by

$$\Delta a_0 = 2i\gamma \sum_{h,k,m} a_h b_k^* b_m X_{h,k,m}. \quad (8.3)$$

where  $\gamma$  is the nonlinear coefficient and where we have arbitrarily (and without loss of generality) set the index of the received symbol to 0. The coefficients  $X_{h,k,m}$  have been introduced in Ref. [4], where they have been shown to be

$$X_{h,k,m} = \int_0^L dz f(z) \int_{-\infty}^{\infty} dt g^*(z, t)g(z, t - hT) \\ \times g^*(z, t - kT - \beta''\Omega z)g(z, t - mT - \beta''\Omega z). \quad (8.4)$$

The function  $f(z)$  accounts for the loss/gain profile along the link [4, 10]. For example,  $f(z) = 1$  in the case of perfectly distributed amplification and  $f(z) = \exp(-\alpha \text{mod}(z, L_s))$  in the case of lumped amplification where  $\alpha$  is the loss coefficient and  $L_s$  is the span length.

The coefficient  $X_{h,k,m}$  is associated with the NLIN that is observed in the measurement of the zeroth data symbol in the channel of interest following the nonlinear interaction between the  $h$ -th pulse in the channel of interest and the  $k$ -th and  $m$ -th pulses in the interfering channel, as illustrated in Fig. 8.1. The triple product  $\xi(z, t) = g(z, t - hT)g^*(z, t - kT - \beta''\Omega z)g(z, t - mT - \beta''\Omega z)$ , appearing in the integrand of Eq. (8.4), describes a classic nonlinear Kerr interaction taking place at time  $t$  and position  $z$  along the link (where  $\beta''\Omega$  is the difference between the reciprocal group velocities). This product then propagates towards the receiver, while accumulating the dispersion present in the remaining fiber-length, which is equal to  $L - z$ . Since the impulse response of the matched filter is proportional to  $g^*(L, t) = \exp(\frac{i}{2}\beta''(L - z)\frac{\partial^2}{\partial t^2})g^*(z, t)$ , this accumulated dispersion is removed. The sampled perturbation at the receiver, as the inner integral in Eq. (8.4) suggests, is therefore  $\int g^*(z, t)\xi(z, t)dt$ , which can be interpreted as if the nonlinear product  $\xi(z, t)$  was matched filtered by a 'fourth' pulse  $g^*(z, t)$  exactly at the position along the link at which the Kerr interaction occurred. The integral over  $z$  in Eq. (8.4) gathers the nonlinear products produced along the entire link.

---

<sup>3</sup>As we are relying on a first order perturbation analysis – the waveform  $g(z, t)$  includes only the effect of dispersion and not the effect of nonlinearity. In [4] we used  $g^{(0)}(z, t)$  for the same quantity.

We now proceed to classifying the various contributions to the NLIN in terms of the type of collisions that produce them. In the most general case, these contributions are produced by nonlinear interactions between four pulses – the “dummy” zeroth pulse from the channel of interest (accounting for the fact that the signal is matched filtered at the receiver), the  $h$ -th pulse from the channel of interest, and the  $k$ -th and  $m$ -th pulses from the interfering channel. When some of the indices coincide, as we shall see in the following sections, a two-pulse, or three-pulse collision is formed.

## 8.4 Two-pulse collisions

A two-pulse collision occurs when one of the interacting pulses is the pulse of interest, which we arbitrarily select as the zeroth-index pulse, whereas the other pulse belongs to the interfering channel, so that  $h = 0$  and  $k = m$ . The strength of this interaction is governed by the coefficient

$$X_{0,m,m} = \int_0^L dz f(z) \int_{-\infty}^{\infty} dt |g(z, t)|^2 |g(z, t - mT - \beta''\Omega z)|^2, \quad (8.5)$$

and, according to Eq. (8.3), it produces a perturbation equal to  $ia_0(2\gamma X_{0,m,m}|b_m|^2)$ . As  $X_{0,m,m}$  is a real-valued quantity the perturbation is at a complex angle of  $\pi/2$  from the transmitted symbol  $a_0$ . Namely, the perturbation generated by two-pulse interactions has the character of phase-noise.

An interesting perspective into this reality can be obtained by considering the ‘old class’ of two-pulse collisions initially studied in the context of solitons [19–22]. There, it has become common knowledge that when a pulse in the channel of interest undergoes a complete collision with an interfering pulse belonging to an adjacent WDM channel, the only notable consequence observed on the pulse of interest is a *time-independent* phase-shift. However, it is also well known that when the collision between the two pulses is not complete [22, 23], the phase of the interacting pulses is distorted via XPM leading to the formation of a *time-dependent* phase distortion (nonlinear chirp), which subsequently produces an amplitude (or timing-jitter) perturbation after being affected by the dispersion of the remaining fiber. Note that in the scheme that we consider here, no amplitude perturbation is formed in two-pulse collisions, regardless of whether the collisions are complete, or incomplete. This difference with respect to the old generation of systems can be attributed to coherent detection and the use of a matched filter, which is equivalent to measuring the perturbation at the point along the fiber where the collision takes place, as explained earlier. Namely, the effect of dispersion, which translates the time-dependent phase distortion into amplitude variations is reversed by the matched filter.

Since in addition to being real-valued, the integrand in Eq. (8.5) is also non-negative, the nonlinear phase-shift is proportional to the “completeness” of the collision, namely the phase-shift is zero when the waveforms do not overlap at all during propagation and it is maximized when the collision is complete. As we demonstrate in Appendix 8.12, two-pulse collision coefficients  $X_{0,m,m}$  scale as  $\Omega^{-1}$ . In addition, the fact that two-pulse contributions to NLIN are proportional to  $a_0|b_m|^2$  implies that the variance of NLIN that is associated with these contributions is proportional to  $\langle |a_0|^2 \rangle (\langle |b_m|^4 \rangle - \langle |b_m|^2 \rangle^2)$ , with the angled brackets



denoting statistical averaging. While  $\langle |a_0|^2 \rangle$  is simply proportional to the average signal power  $\langle |b_m|^4 \rangle - \langle |b_m|^2 \rangle^2$ , is clearly dependent on modulation format. For example, when the interfering channel is modulated only in phase (e.g. QPSK) the NLIN variance caused by two-pulse collisions is 0.

## 8.5 Three-pulse collisions

Three-pulse interactions are represented by coefficients of the form  $X_{0,k,m}$  (where  $k \neq m$ ) and  $X_{h,m,m}$  (where  $h \neq 0$ ). The first form represents an interaction between the pulse of interest (the zeroth pulse from the channel of interest) and the  $k$ -th and  $m$ -th pulses from the interfering channel. The perturbation produced by this interaction is equal to  $ia_0(2\gamma X_{0,k,m}b_k^*b_m + 2\gamma X_{0,m,k}b_m^*b_k)$  and since it can be easily verified that  $X_{0,k,m} = X_{0,m,k}^*$ , the perturbation is equal to  $ia_0\Re\{4\gamma X_{0,k,m}b_k^*b_m\}$  which is at a complex angle of  $\pi/2$  from the transmitted symbol  $a_0$  and therefore it too has the character of phase-noise. Yet, unlike the case of two-pulse collisions, since  $k \neq m$  the variance of this phase-noise contribution depends only on the average signal power and not on the modulation format. The second form of three-pulse interactions  $X_{h,m,m}$  involves an interaction between two pulses in the channel of interest (the zeroth and  $h$ -th pulses) and a single pulse from the interfering channel (the  $m$ -th pulse). The perturbation produced by this interaction is equal to  $\Delta a_0 = ia_h(2\gamma X_{h,m,m}|b_m|^2)$ . This contribution has no fixed phase relation with  $a_0$  and hence it can be viewed as complex circular noise. Yet, similarly to the case of two-pulse collisions, the dependence on  $|b_m|^2$  implies dependence of the NLIN variance due to this contribution on the data modulation format.

We show in Sec. 8.7 and in Appendix 8.12 that the perturbation to the symbol of interest is formed within the initial part of a three-pulse collision, but then cancels through destructive interference once the collision is complete, so that the residual magnitude of the perturbation scales as  $\Omega^{-2}$ . In the regime where the relative number of complete collisions is significant, the overall contribution of three-pulse collisions practically vanishes relative to that of two-pulse collisions (in spite of the fact that the number of three-pulse collisions is larger, see Sec. 8.9). The scaling of incomplete three-pulse collisions with frequency is not monotonous, and depends on the point at which the collision is discontinued.

## 8.6 Four-pulse collisions

The case of four-pulse collisions involves two pulses from the channel of interest (the zeroth and  $h$ -th pulses, with  $h \neq 0$ ) and two pulses from the interfering channel (the  $k$ -th and  $m$ -th pulses with  $k \neq m$ ). Each of these interactions generates a perturbation that is equal to  $2i\gamma a_h(X_{h,k,m}b_k^*b_m + X_{h,m,k}b_m^*b_k)$  and since there is no fixed phase relation between the various symbols, these contribution can be viewed as producing complex circular noise. Since  $h \neq 0$  and  $k \neq m$ , and the data transmitted on different symbols is assumed to be statistically independent, the average value of NLIN due to four-pulse contributions is 0, and the variance is proportional to  $\langle |a_0|^2 \rangle \langle |b_0|^2 \rangle^2$ , which is determined by the average power per symbol independently of the modulation format.

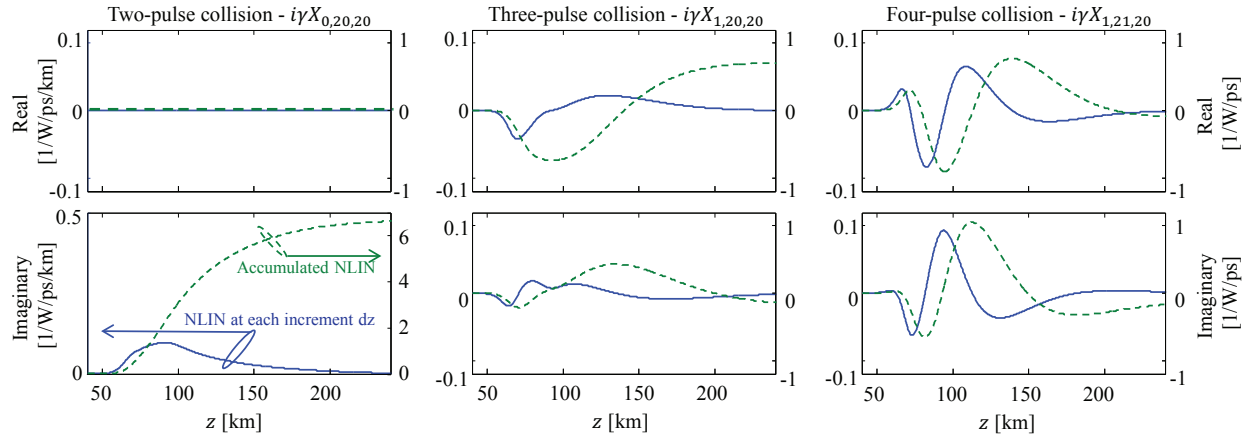


Figure 8.2: The evolution of the NLIN coefficients along the fiber axis  $z$ , assuming ideally distributed amplification and channel separation of 50GHz. The solid (blue) curves show the contribution from a single fiber increment  $dz$  at position  $z$  along the fiber to the NLIN component  $i\gamma X_{h,k,m}$  (this is proportional to the inner integral in Eq. (8.4)), whereas the dashed (green) curves show the accumulated NLIN component  $i\gamma X_{h,k,m}$  itself, under the assumption that the link is ended at point  $z$ . The two-pulse component  $i\gamma X_{0,20,20}$  is imaginary and hence produces pure phase-noise. It grows monotonically with distance, reaching a maximum after the collision has been completed. The three and four-pulse components  $i\gamma X_{1,20,20}$  and  $i\gamma X_{1,21,20}$  are complex and their value is largest during the collision itself and reduces significantly when the collision is complete.

We show in Sec. 8.7 and in Appendix 8.12 that as in the case of three-pulse collisions, the interference formed in the first half of a four-pulse collision process cancels through destructive interference in the second half. The magnitude of the interference produced by a four-pulse collision is therefore determined by the point in which the collision is discontinued. In Appendix 8.12 we show that the maximal interference produced by an incomplete collision scales as  $\Omega^{-1}$  whereas the magnitude of the residual interference left after a complete collision scales as  $\Omega^{-3}$ . This implies that the overall contribution of complete four-pulse collisions is negligible in spite of their much larger number (see Sec. 8.9).

## 8.7 A few numerical examples

In order to better illustrate the distinctive features of the various types of collisions we first examine the dependence of three particular NLIN coefficients on space and frequency separation. These coefficients are  $X_{0,20,20}$ ,  $X_{1,20,20}$  and  $X_{1,21,20}$ , representing a two-pulse, three-pulse and four-pulse collision, respectively. Notice that  $X_{1,20,20}$  can be viewed as representing both types of three-pulse collisions (i.e., those with  $h = 0$  and those with  $h \neq 0$ ), as it can be easily shown that  $X_{h,m,m} = X_{0,m-h,m}^*$ . Since the concept of complete collisions is best isolated when losses during the collision are negligible, we plot the examples in this section in the ideally distributed amplification case. The issue of lumped versus distributed amplification is further discussed in Sec. 8.9. We assume square-root raised-cosine pulses



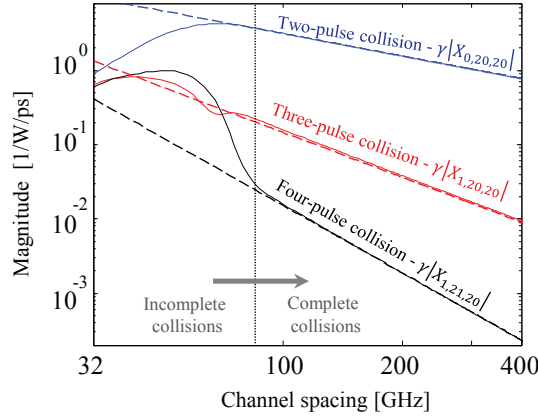


Figure 8.3: The magnitude of the nonlinear perturbation  $\gamma|X_{0,20,20}|$ ,  $\gamma|X_{1,20,20}|$ , and  $\gamma|X_{1,21,20}|$  as a function of frequency separation for a 100 km link with ideally distributed amplification. It can be clearly observed that once the collision becomes complete (according to Eq. (8.6), this happens when the channel separation is larger than 85GHz), the two-pulse NLIN coefficient drops as  $\Omega^{-1}$ , the three-pulse coefficient drops as  $\Omega^{-2}$  and the four-pulse coefficient drops as  $\Omega^{-3}$  (dashed lines show the scaling with  $\Omega^{-j}$  for  $j = 1, 2$  and  $3$ ). Note that for longer links, the transition from incomplete to complete collisions would take place at a correspondingly lower channel spacing.

of 32GHz baud-rate with a roll-off factor of 0.2, transmitted over a standard single-mode fiber with dispersion coefficient  $\beta'' = 21\text{ps}^2/\text{km}$ , nonlinear coefficient  $\gamma = 1.3 \text{ W}^{-1}\text{km}^{-1}$  and attenuation of 0.2 dB/km.

In Fig. 8.2 we assume channel separation of 50GHz and show the evolution of the various NLIN coefficients along the fiber, where the real and imaginary parts of the perturbation  $i\gamma X_{h,k,m}$  are plotted in the top and bottom panels, respectively. In each figure the solid (blue) curve shows the contribution to the NLIN produced in a single fiber increment of length  $dz$  (the inner (time) integral of Eq. (8.4)). The dashed (green) curves show the accumulated contribution to the NLIN, obtained by integrating the corresponding solid (blue) curves from the link input up to point  $z$  along the fiber. In the case of two-pulse collisions, as we have argued earlier, the incremental contribution is imaginary (i.e. it is in quadrature with the symbol of interest), and hence it only contributes to the formation of phase-noise. In addition, the imaginary part of the incremental contributions is strictly non-negative and hence, the corresponding phase-perturbation increases monotonically with position along the link. Therefore, in the case of an incomplete collision (where the integration ends while the collision is still taking place) the phase-noise is smaller than it is when the collision is complete. With three and four-pulse collisions the situation is distinctly different. First, the incremental contributions are complex, and secondly, the incremental contributions change signs during the collision, so that the integrated NLIN coefficient grows in the beginning of the collision, but then reduces towards its end. The fact that with three and four-pulse collisions, the end-value of the accumulated contribution (green-dashed curves) is lower than the corresponding peak value (around the center of the collision) is clearly evident in the center and right panels of Fig. 8.2, suggesting that the interference induced by incomplete

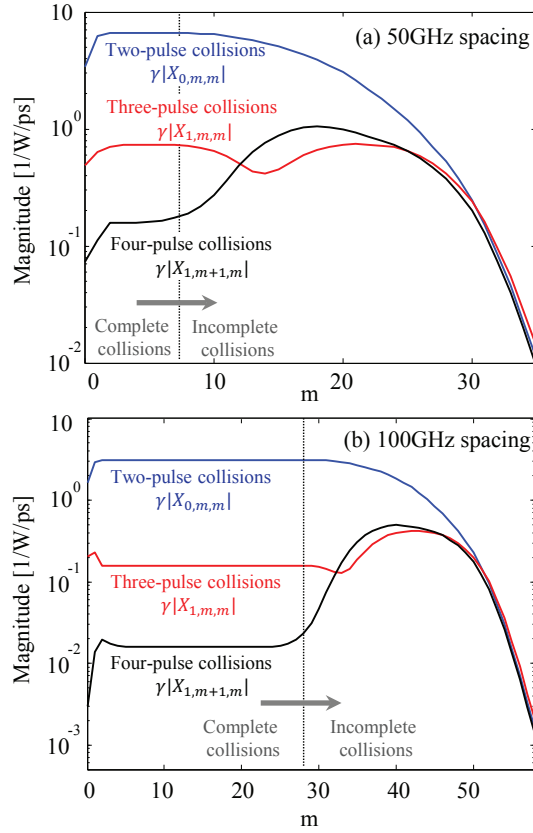


Figure 8.4: The magnitude of the nonlinear perturbation  $\gamma|X_{0,m,m}|$ ,  $\gamma|X_{1,m,m}|$ , and  $\gamma|X_{1,m+1,m}|$  as a function of  $m$  for a 100 km link with ideally distributed amplification. Top and bottom panels assume 50GHz and 100GHz channel separations, accordingly. In longer links, the transition from incomplete to complete collisions would take place at a correspondingly higher values of  $m$ .

collisions is likely to be much stronger than it is in the case of complete collisions. This feature is particularly strong in the case of four-pulse collisions and for both three-pulse and four-pulse collisions it becomes increasingly more distinct when the frequency separation between the interacting WDM channels becomes larger.

The dependence on frequency separation is illustrated in Fig. 8.3 where we show the dependence of the coefficients  $X_{h,k,m}$  on  $\Omega$  for a 100 km link. As can be deduced from Appendix 8.12, when

$$\Omega > B + \frac{(m+k-h)T}{2|\beta''|L}, \quad (8.6)$$

where  $B$  is the spectral width<sup>4</sup> of the individual pulses, the group velocity differences between the channels are large enough to guarantee that the collision is completed before the fiber's end. In the case considered in Fig. 8.3, this happens when the channel spacing is larger than 85GHz. Evidently, in this regime the perturbation due to two, three, and four-pulse

<sup>4</sup>We stress that in Eq. (8.6) and throughout this paper  $\Omega$  and  $B$  are indicated in angular frequency units.

collisions scales as  $\Omega^{-1}$ ,  $\Omega^{-2}$ , and  $\Omega^{-3}$ , respectively. When the frequency separation is below 85GHz, the collision is no longer complete, as it is discontinued abruptly at the end of the fiber. In this regime the magnitude of the perturbation depends on the point in which the collision is disrupted and hence, consistently with the discussion of Fig. 8.2, the dependence on  $\Omega$  is only monotonous in the two-pulse collision case.

We further illustrate the properties of the various types of collisions by examining their dependence on the index of the interfering pulses while fixing the channel separation and the link length. In Fig. 8.4 we plot the magnitude of  $i\gamma X_{0,m,m}$ ,  $i\gamma X_{1,m,m}$  and  $i\gamma X_{1,m+1,m}$  versus  $m$  in a 100km link. Figure 8.4a shows the case of 50GHz channel spacing whereas Fig. 8.4b shows the case of 100GHz channel spacing. As can be deduced from Eq. (8.6), the collisions are complete when  $m < 7$  for channel separation of 50GHz and  $m < 28$  for channel separation of 100GHz. In this regime  $\gamma|X_{0,m,m}|$  receives its maximum value, while  $\gamma|X_{1,m,m}|$  and  $\gamma|X_{1,m+1,m}|$  are approximately one and two orders of magnitude below that value, respectively. When  $m$  increases, the collisions become incomplete and the perturbation in the case of two-pulse collisions reduces monotonically while it oscillates in the cases of three-pulse and four-pulse collisions. Eventually, when  $m$  grows further the pulses do not collide at all and all of the coefficients vanish<sup>5</sup>.

## 8.8 The effect of polarization-multiplexing

Throughout Secs. 8.4-8.6 we have assumed for simplicity the case of a signal transmitted in a single polarization. While the consideration of polarization multiplexed transmission is considerably more cumbersome, the principles remain very similar. Equation (8.3) is generalized to

$$\Delta \underline{a}_0 = i\gamma \sum_{h,k,m} X_{h,k,m} \left( \underline{b}_k^\dagger \underline{b}_m \mathbf{I} + \underline{b}_m \underline{b}_k^\dagger \right) \underline{a}_h, \quad (8.7)$$

where the underline denotes a two-element column vector (i.e. its elements represent the two polarization component of the optical field) and  $\mathbf{I}$  is the  $2 \times 2$  identity matrix. For simplicity Eq. (8.7) corresponds to the case in which the modulated polarization axes in the interfering channel are parallel to those in the channel of interest. The more general case, in which the modulated polarization axes in the two WDM channels are rotated relative to each other is treated in Appendix 8.13. The NLIN in Eq. (8.7) is identical to that given in (8.3), except that each pulse is now modulated by a vector and carries two complex degrees of freedom. The coefficient  $X_{h,k,m}$  given in Eq. (8.4) remains unchanged, and hence its scaling with the frequency separation  $\Omega$  is also unaltered.

As in the case of single-polarization, interference terms that are proportional to coefficients of the form  $X_{0,k,m}$  (corresponding to either two-pulse collisions (when  $k = m$ ) or three-pulse collisions (when  $k \neq m$ )) do not change the norm of the detected vector of symbols  $\underline{a}_0$ , but they do change its phase and state of polarization. Thus, the phase-noise content of

---

<sup>5</sup>We note that the reasoning presented in this paper does not strictly apply to the cases of very low  $m$  values (such as  $m = 0$ , or  $m = 1$  in Fig. 8.4), where more than two-pulse collisions cannot be formed for lack of sufficient dispersion.

	Two-pulse collisions $h = 0, k = m$	Three-pulse collisions		Four-pulse collisions $h \neq 0, k \neq m$
		$h = 0, k \neq m$	$h \neq 0, k = m$	
• Nature of noise	Phase-noise and polarization-rotation	Phase-noise and polarization-rotation	Circular noise	Circular noise
• Dependence on modulation format	Yes	No	Yes	No
• Strongest when the collision is	Complete	Incomplete	Incomplete	Incomplete

Table 8.1: Summary of XPM contributions

NLIN in the single-polarization case is generalized into phase-noise *and* polarization-rotation. We first show this for the case of two-pulse collisions, whose contribution to the NLIN is given by

$$\begin{aligned}
\Delta \underline{a}_0^{(2\text{PC})} &= i\gamma \sum_m X_{0,m,m} (\|\underline{b}_m\|^2 \mathbf{I} + \underline{b}_m \underline{b}_m^\dagger) \underline{a}_0 \\
&= \frac{i}{2} \gamma \sum_m X_{0,m,m} \left( 3\|\underline{b}_m\|^2 \mathbf{I} + \vec{B}_m \cdot \vec{\sigma} \right) \underline{a}_0,
\end{aligned} \tag{8.8}$$

where the superscript ‘2PC’ stands for two-pulse-collisions, and where  $\|\underline{b}_m\|^2 = \underline{b}_m^\dagger \underline{b}_m$  is the square modulus of the  $m$ -th vector of symbols. The Second line in Eq. (8.8) uses the Stokes-space representation [31]. The term  $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$  is a vector whose three components are the Pauli matrices. The term  $\vec{B}_m$  is the vector that represents  $\underline{b}_m$  in the Stokes space. Its three components are real-valued and given by  $B_{m,j} = \underline{b}_m^\dagger \sigma_j \underline{b}_m$ , with  $j = 1, 2$ , or  $3$ . The notation  $\vec{B}_m \cdot \vec{\sigma}$  is a short-hand for  $B_{m,1}\sigma_1 + B_{m,2}\sigma_2 + B_{m,3}\sigma_3$ . All of the relevant properties of the Stokes representation are conveniently summarized in [31]. Within the first order perturbation analysis, the perturbed vector can be written as<sup>6</sup>

$$\underline{a}_0 + \Delta \underline{a}_0^{(2\text{PC})} = e^{i\varphi_{2\text{PC}}} e^{\frac{i}{2} \vec{S}_{2\text{PC}} \cdot \vec{\sigma}} \underline{a}_0. \tag{8.9}$$

where

$$\varphi_{2\text{PC}} = \frac{3}{2} \gamma \sum_m X_{0,m,m} \|\underline{b}_m\|^2 \tag{8.10}$$

$$\vec{S}_{2\text{PC}} = \gamma \sum_m X_{0,m,m} \vec{B}_m. \tag{8.11}$$

While the first exponent in Eq. (8.9) represents a phase rotation of both polarizations by an angle  $\varphi_{2\text{PC}}$ , the second exponent represents a polarization-rotation. In the Stokes jargon, the vector representing  $\underline{a}_0$  on the Poincaré sphere, rotates about the vector  $\vec{S}_{2\text{PC}}$  at an angle equal to its length  $|\vec{S}_{2\text{PC}}|$  [31].

As can be readily deduced from Eq. (8.10), the phase-noise  $\varphi_{2\text{PC}}$  strongly depends on the modulation format, and in particular, its variance vanishes in the case of phase-modulated

<sup>6</sup>Since the entire theory relies on the first-order perturbation analysis, Eq. (8.9) is not an approximation of Eq. (8.8), but rather it is just as accurate as Eq. (8.8) is.

transmission (such as QPSK) as  $\|\underline{b}_m\|^2$  is constant. The dependence of the polarization rotating effect on modulation format can be seen by observing how the orientation of the vector  $\vec{B}_m$  is scattered on the Poincaré sphere. In particular, the isotropy of the orientations of  $\vec{B}_m$  on the Poincaré sphere increases with increasing QAM-order, whereas with BPSK, or QPSK, all the vectors  $\vec{B}_m$  reside in a single plane. Since the polarization-rotation is determined by  $\sum_m X_{0,m,m} \vec{B}_m$ , the variance of the rotations angle  $|\vec{S}_{2PC}|$  also depends on modulation format (as shown in the section that follows), but it is always non-zero, even in the case of pure phase-modulation.

The effect of three-pulse collisions of the type  $h = 0$  and  $k \neq m$  generalizes identically to the above, except that  $\varphi_{2PC}$  and  $\vec{S}_{2PC}$  are replaced by

$$\varphi_{3PC} = \frac{3}{2}\gamma \sum_{\substack{k,m \\ k \neq m}} X_{0,k,m} \underline{b}_k^\dagger \underline{b}_m \quad (8.12)$$

$$\vec{S}_{3PC} \cdot \vec{\sigma} = 2\gamma \sum_{\substack{k,m \\ k \neq m}} X_{0,k,m} \left( \underline{b}_m \underline{b}_k^\dagger - \frac{1}{2} \underline{b}_k^\dagger \underline{b}_m \mathbf{I} \right), \quad (8.13)$$

where the ‘real-valuedness’ of  $\varphi_{3PC}$  and  $\vec{S}_{3PC}$  follows from the fact that  $X_{0,k,m} = X_{0,m,k}^*$ . The variance of the phase-shift and of the polarization-rotation angle caused by three-pulse collisions ( $k \neq m$ ) are independent of modulation format since  $\underline{b}_k$  and  $\underline{b}_m$  are statistically independent. The combined perturbation that is caused by both of the above processes satisfies  $\underline{a}_0 + \Delta \underline{a}_0^{(2PC)} + \Delta \underline{a}_0^{(3PC)} = \exp(i\varphi) \exp(\frac{i}{2} \vec{S} \cdot \vec{\sigma}) \underline{a}_0$ , with  $\varphi = \varphi_{2PC} + \varphi_{3PC}$  and  $\vec{S} = \vec{S}_{2PC} + \vec{S}_{3PC}$ . The unitarity of these processes implies that  $\|\underline{a}_0 + \Delta \underline{a}_0^{(2PC)} + \Delta \underline{a}_0^{(3PC)}\| = \|\underline{a}_0\|$ .

Table 8.1 summarizes the classification of pulse collisions in the polarization multiplexed case, while pointing out the main properties of each collision type. As seen in the table, the effects of four-pulse collisions and of three-pulse collisions of the type  $h \neq 0$ ,  $k = m$  remains unchanged relative to the single-polarization case as they produce complex circular noise on each of the two polarizations.

It is important to note that the effect of nonlinear polarization-rotation is only unitary when both polarization components are considered jointly as a vector. In systems that process the individual polarizations separately, the effect of polarization-rotation does not appear unitary. Rather, from the standpoint of an individual polarization component, the diagonal terms of the polarization-rotation matrix  $\exp(\frac{i}{2} \vec{S} \cdot \vec{\sigma})$  appear as additional phase-noise (which has opposite signs for each of the individual polarization components). Whereas, the non-diagonal terms cause mixing between the two polarization components, which under the assumption that the components are statistically independent, manifests itself as complex circular noise. Consequently, in polarization multiplexed systems where the two polarization channels are processed separately from each other, two-pulse and three-pulse collisions with  $h = 0$ , generate not only pure phase-noise (as in the single-polarization case), but some complex circular noise as well. This implies that the relative significance of phase-noise in such systems is slightly smaller than in the single polarization case.

As explained in Secs. 8.4 and 8.5, and as is summarized in Table 8.1, the dependence of the NLIN power on modulation format is caused only by the interference terms proportional

to  $X_{h,m,m}$  (whether  $h = 0$ , or not)<sup>7</sup>. In the polarization multiplexed case, the modulation format dependence of these terms may be slightly weaker than in the case of single polarization. In order to demonstrate this, we examine the NLIN contribution to the  $y$  component of  $\underline{a}_0$  that is proportional to  $X_{h,m,m}$ . This contribution is readily extracted from Eq. (8.7), and it is given by

$$i\gamma X_{h,m,m} \left( 2a_h^{(y)} |b_m^{(y)}|^2 + a_h^{(y)} |b_m^{(x)}|^2 + a_h^{(x)} b_m^{(x)*} b_m^{(y)} \right). \quad (8.14)$$

The first term in Eq. (8.14) falls into the category of single polarization interactions and its contribution to the variance of the NLIN is modulation-format dependent, exactly as discussed in Secs. 8.4-8.5 for the single-polarization case. The same is true for the second term in Eq. (8.14), which represents a classical XPM interaction between the  $y$ -polarized symbols from the channel of interest and the  $x$ -polarized symbols from the interfering channel. But the third term in Eq. (8.14) is different in the sense that it involves interference between four *different* data symbols. Assuming that the data transmitted on different polarizations is statistically independent, the variance of the third term is independent of modulation format, and therefore it does not contribute to the modulation format dependence of the overall NLIN variance. It should be noted however, that the modification to the significance of both phase-noise (when the polarization channels are processed separately) and modulation-format dependence in polarization multiplexed systems is relatively small, as can be seen in [6] and in the section that follows.

Finally, as we demonstrate in Appendix 8.13, the conclusions of this section do not change when the modulated polarization axes in the two channels are not identical. In particular, the phase-noise  $\varphi$ , the angle of polarization-rotation  $|\vec{S}|$ , and the NLIN variance in each of the two polarizations, are all invariant to whether the modulated polarization axes in the two WDM channels are identical, or not. A feature that changes somewhat when the modulated polarization axes are different in the two channels, is the relative significance of phase-noise in systems where the two polarization components are processed separately. For further details, the reader is referred to Appendix 8.13.

## 8.9 System implications of collision classification

The nature of NLIN in systems changes depending on whether it is dominated by complete or incomplete collisions. A collision can be safely characterized as complete when the evolution of the individual pulses during the collisions is negligible. This condition can be translated into two requirements. The first is that the dispersive broadening of the pulses during the collision is negligible relative to the pulse width prior to the collisions. This requirement is automatically satisfied when the frequency spacing between the interacting WDM channels is sufficiently larger than the bandwidth of the channels themselves. To see that, we define the collision length,  $L_c$ , as the length of the section of fiber in which the collision takes place

---

<sup>7</sup>Notice that the variance of the interference contribution that is proportional to  $X_{h,m,m}$  depends on the statistical average of  $\|\underline{b}_m\|^4 = |b_m^{(y)}|^4 + 2|b_m^{(y)}|^2|b_m^{(x)}|^2 + |b_m^{(x)}|^4$ , where  $b_m^{(y)}$  and  $b_m^{(x)}$  are the  $y$  and  $x$  components of  $\underline{b}_m$ , accounting for a dependence on the *fourth-order moment* of the transmitted symbols and for a dependence on the *second-order correlation* between the two polarization components.

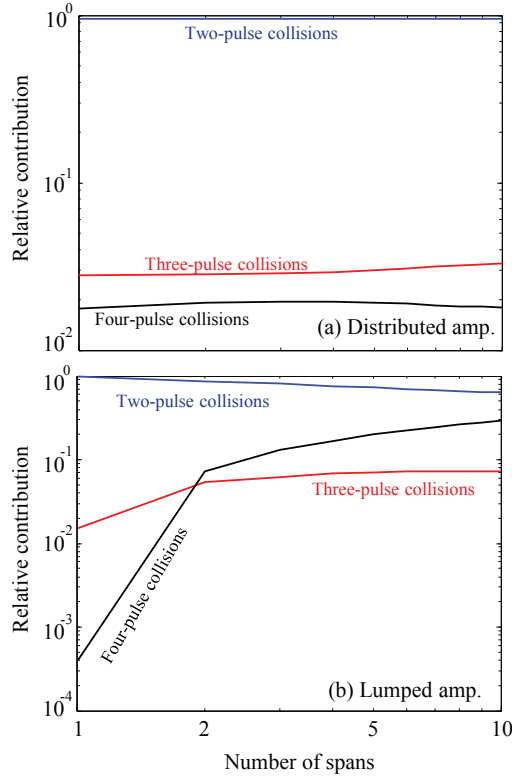


Figure 8.5: The contribution of two-pulse collisions  $\sum_m |X_{0,m,m}|^2$ , three-pulse collisions  $\left(\sum_{k \neq m} |X_{0,k,m}|^2 + \sum_{h \neq 0} |X_{h,m,m}|^2\right)$ , and four-pulse collisions  $\sum_{h \neq 0, k \neq m} |X_{h,k,m}|^2$ , normalized to the overall NLIN contributions  $\sum_{h,k,m} |X_{h,k,m}|^2$ , assuming polarization-multiplexed transmission of 87 WDM channels, at 32GHz baud-rate, positioned on a 50GHz ITU grid (43 channels on each side of the channel of interest). Span length is 100km where top-panel (a) corresponds to ideally distributed amplification and bottom panel (b) corresponds to lumped amplification.

$L_c \simeq \frac{\Delta t}{\beta'' \Omega}$ , where  $\Delta t$  is the temporal duration of the pulses immediately prior to the collision. Denoting by  $B$  the spectral width of the individual pulses, the temporal broadening that is caused by dispersion during the process of a collision can be approximated by  $\beta'' B L_c = \Delta t \frac{B}{\Omega}$ , and it is certain to become much smaller than  $\Delta t$  when  $\Omega/B$  is large<sup>8</sup>.

The second requirement is that the effects of attenuation or gain are negligible during a collision. In systems using lumped amplification this condition is not satisfied when the collision overlaps with an amplifier site, or when the collision length  $L_c$  is so long that attenuation during the collision cannot be neglected. In modern uncompensated fiber links, where  $\Delta t$  increases due to the accumulated dispersion,  $L_c$  is bound to reach values over which fiber loss becomes significant and the collision can no longer be considered complete. For this reason, truly complete collisions may exist in distributed amplification systems (as can be approximately achieved with Raman technology). In systems using lumped amplification,

<sup>8</sup>As can be seen in Sec. 8.7, consistency with the pulse-collisions theory is observed even in the case of a 32 G-baud system with a 50 GHz channel separation, where  $\Omega/B$  is only of the order of 1.5.



complete collisions occur mostly in the first few spans, where the pulses are still narrow enough to ensure that the collision length is short so that the attenuation during the collision is negligible and that only a small fraction of the collisions overlap with an amplification site.

As can be concluded from the combinations of indices, the number of four-pulse collisions is the largest, whereas the smallest number of collisions are those involving only two pulses. Nevertheless, in a regime where complete collisions dominate, the nature of NLIN is dictated primarily by two-pulse collision processes, whose scaling with frequency separation is the most favorable (proportional to  $\Omega^{-1}$ ). In this regime the NLIN is expected to have a strong phase and polarization-rotation noise character, and its variance should be strongly dependent on the modulation format. In the opposite case, where incomplete collisions dominate, the larger number of four-pulse collisions emphasizes their significance and makes the overall NLIN more similar to complex circular noise. In order to illustrate these principles we consider in what follows the cases of systems with distributed amplification (governed by complete collisions) and with lumped amplification (where short links are governed by complete collisions and long links are governed by incomplete collisions). We assume a polarization multiplexed transmission of square-root raised-cosine pulses with a roll-off factor of 0.2, propagating over 100 km spans of standard single-mode fiber ( $\beta'' = 21\text{ps}^2/\text{km}$ ,  $\gamma = 1.3\text{W}^{-1}\text{km}^{-1}$ , and attenuation of 0.2 dB/km). We further assume that the entire C-band (35 nm) is populated with 32GBaud WDM channels positioned on a 50GHz ITU grid (amounting to a total of 87 WDM channels), where the channel of interest is at the center. The choice of using 32GBaud transmission was made for consistency with current commercial systems. The principles governing NLIN build-up as described in this paper will not change at higher symbol-rate. In fact, a shorter collision length will make the contribution of complete collisions even closer to the theoretical prediction. Additionally, in order to facilitate the comparison between the various cases, the examples that we show in this section are plotted for the case of 10-span links regardless of the modulation format, or type of amplification. However, the principles of NLIN build-up do not change when the link extends over longer distances.

We first examine the various pulse-collision contributions to the quantity  $\sum_{h,k,m} |X_{h,k,m}|^2$ , which in the case of Gaussian modulation is proportional to the overall NLIN variance (see appendix in [6]). Other modulation formats are considered subsequently. The relative contributions of two-pulse collisions ( $h = 0, k = m$ ), three-pulse collisions<sup>9</sup> ( $h = 0, k \neq m$ , or  $h \neq 0, k = m$ ), and four-pulse collisions ( $h \neq 0, k \neq m$ ) to this quantity are plotted in Fig. 8.5 for the cases of distributed and lumped amplification. The dominance of two-pulse collisions is clearly evident in the distributed amplification case, whereas in the case of lumped amplification two-pulse collisions are dominant primarily in few span systems (where a large fraction of the collisions are complete). In longer lumped amplification systems the role of four-pulse collisions is emphasized, as the number of incomplete collisions grows with the number of spans.

In order to distinguish between complex circular noise on the one hand, and phase and polarization noise on the other, we express the received vector of symbols at the  $n$ -th time-

---

<sup>9</sup>The two types of three-pulse collisions (those with  $h = 0$  and those with  $h \neq 0$ ) contribute similarly to the overall NLIN variance in the case of Gaussian modulation since  $X_{h,m,m} = X_{0,m-h,m}^*$ .



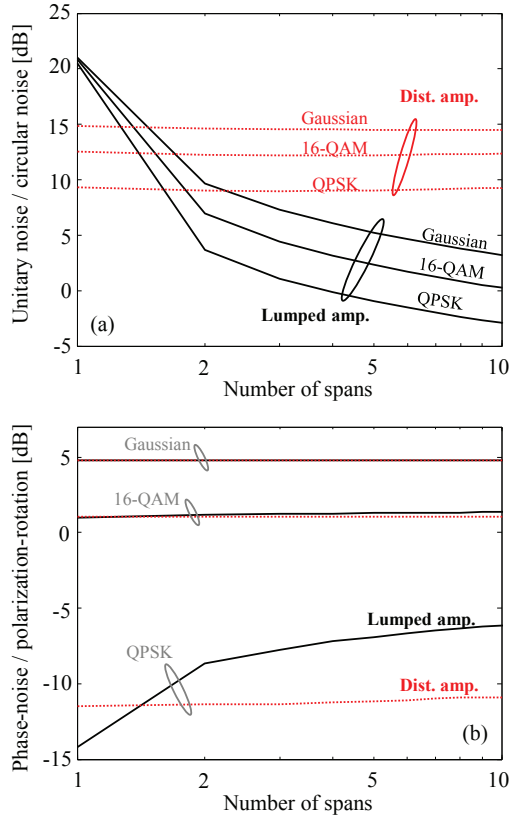


Figure 8.6: (a) The ratio between the contribution of unitary noise (phase-noise and polarization-rotation) and circular noise to the overall NLIN power, i.e.  $PT(\sigma_\phi^2 + \sigma_S^2)/\sigma_\nu^2$  in the notation of Eq. (8.16). (b) The ratio between the contributions of phase-noise and polarization-rotation to the overall NLIN power, i.e.  $\sigma_\phi^2/\sigma_S^2$ . Dotted (red) curves correspond to ideally distributed amplification and solid (black) curves correspond to lumped amplification, assuming 43 WDM channels on each side of the channel of interest.

slot as

$$\underline{a}_n + \Delta \underline{a}_n = e^{i\phi_n} e^{\frac{i}{2} \vec{S}_n \cdot \vec{\sigma}} \underline{a}_n + \underline{\nu}_n \quad (8.15)$$

where  $\phi_n$  represents the phase-noise component of the NLIN,  $\vec{S}_n$  is the polarization-rotation vector, as explained in the previous section, and  $\underline{\nu}_n$  represents circular nonlinear noise. In the relevant limit of signal to noise ratio being much larger than unity  $\Delta \underline{a}_n \simeq i(\phi_n \mathbf{I} + \frac{1}{2} \vec{S}_n \cdot \vec{\sigma}) \underline{a}_n + \underline{\nu}_n$  and assuming that the data vectors  $\underline{a}_j$  and  $\underline{b}_j$  are statistically independent, identically distributed and isotropically symmetric in the phase-space, the NLIN variance can be expressed as

$$\begin{aligned} \sigma_{\text{NLIN}}^2 &= \langle \|\Delta \underline{a}_n\|^2 \rangle - \|\langle \Delta \underline{a}_n \rangle\|^2 \\ &= PT(\sigma_\phi^2 + \sigma_S^2) + \sigma_\nu^2, \end{aligned} \quad (8.16)$$

where  $\sigma_\phi^2 = \langle \phi_n^2 \rangle - \langle \phi_n \rangle^2$ ,  $\sigma_S^2 = \frac{1}{4} \langle |\vec{S}_n|^2 \rangle$ ,  $\sigma_\nu^2 = \langle \|\underline{\nu}_n\|^2 \rangle$ ,  $P$  is the average signal power and  $T$  is the symbol duration (so that  $\langle \|\underline{a}_n\|^2 \rangle = PT$ ). The first two terms on the right-hand-side of

the bottom row of Eq. (8.16) are the phase-noise and polarization-rotation contributions to the total NLIN variance, whereas the third term represents the contribution of the complex circular noise.

In Fig. 8.6a we plot the ratio between the contributions of the unitary noise (i.e., phase-noise and polarization-rotation) and of the circular noise to the overall NLIN variance (namely,  $PT(\sigma_\phi^2 + \sigma_S^2)/\sigma_\nu^2$ ). Notice that this ratio is power independent as all the contributions to the NLIN variance are proportional to  $P^3$ . In the distributed amplification case, where complete collisions govern the generation of NLIN so that the effect of two-pulse collisions is dominant, phase-noise and polarization-rotation constitute the largest NLIN contribution, as expected. In the case of lumped amplification, where complete collisions dominate only the beginning of the link, phase-noise and polarization-rotation are largest after the first amplified span. As the number of spans increases, collisions become increasingly incomplete and therefore circular noise caused by four-pulse collisions grows gradually.

Figure 8.6b illustrates the relative significance of the two unitary noise contributions, where the ratio  $\sigma_\phi^2/\sigma_S^2$  is plotted versus the number of spans. Clearly, the role of phase-noise increases with the extent to which the intensity of the symbols is modulated, growing from QPSK to 16-QAM and then to Gaussian modulation. This point is further explored in Fig. 8.7. In addition, in the cases of Gaussian modulation and 16-QAM there is almost no difference between the cases of lumped and distributed amplification, whereas a notable difference exists in the case of QPSK. The reason is that in QPSK, only three-pulse collisions (of the type  $h = 0$ ) generate phase-noise and their relative significance increases with the number of spans in the case of lumped amplification, whereas it is independent of the link-length in the case of distributed amplification (see Fig. 8.5).

We further address the dependence of the unitary noise and the circular noise on modulation format. In order to do that we introduce the concept of the *Fourth-order modulation factor*

$$M = \frac{\langle |b|^4 \rangle}{\langle |b|^2 \rangle^2} \quad (8.17)$$

where the random variable  $b$  is a symbol in one polarization component of the interfering channel. The fourth-order modulation factor is equal to unity in the case of pure phase-modulation and grows as the power variations of the modulated signal increase. For example in the case of QAM modulation,  $M = 1$  for pure phase modulation (like QPSK) and increases asymptotically towards 1.4 with the QAM order (specific values are  $M = 1.32$  for 16-QAM, and  $M = 1.38$  for 64-QAM). In the case of complex Gaussian modulation  $M$  is equal to 2. Excluding cases of extremely tight channel-spacing, which require the correction term discussed in [16], and assuming that the two polarization components of the transmitted data vectors are statistically independent, all aspects of the modulation-format dependence of the NLIN power have been shown to be captured by the fourth-order modulation factor [4, 5].

In Fig. 8.7 we plot the contributions of phase-noise  $PT\sigma_\phi^2$ , polarization-rotation  $PT\sigma_S^2$ , and circular noise  $\sigma_\nu^2$  to the NLIN power as a function of  $M$  in a  $10 \times 100$ km link with distributed (Fig. 8.7a) and lumped (Fig. 8.7b) amplification. Since the NLIN power scales with  $P^3$ , the curves are displayed for the case in which  $P = 1$  mW, and therefore to obtain the results for other input powers, the vertical axes simply need to be shifted upward by three times the average input power expressed in dBm units. Evidently, the contribution

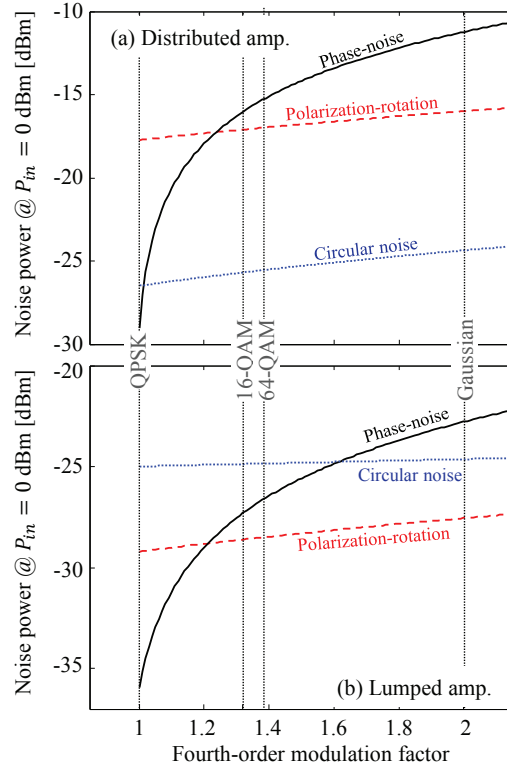


Figure 8.7: The contributions of phase-noise (solid black curves), polarization-rotation (dashed red curves), and circular noise (dotted blue curves) to the NLIN power, as a function of the modulation format factor  $M$ , for a  $10 \times 100$ km link, assuming polarization-multiplexed transmission with 43 WDM channels on each side of the channel of interest, baud rate of 32GHz and 50GHz channel spacing. Ideally distributed amplification is considered in figure (a) and lumped amplification is considered in figure (b). Values of  $M$  corresponding to known modulation formats are marked on the figure.

of phase-noise strongly depends on modulation format in both cases of distributed and lumped amplification, whereas the dependence of circular noise and polarization-rotation on modulation format is notably weaker. The dependence of the phase-noise on modulation format can be attributed to the fact that it is dominated by two-pulse collisions which, as we have explained earlier, are strongly dependent on modulation format. The circular noise, on the other hand, is strongly affected by four-pulse collisions (particularly in the multi-span lumped amplification scenario), and therefore the dependence of its variance on modulation format is weak. As to polarization-rotations noise, which is also dominated by two-pulse collisions, its weak dependence on modulation format can be understood in view of Eq. (8.11), which shows that the rotation angle (given by  $|\vec{S}_{2PC}|$ ) is determined by a summation that includes the random vectors  $\vec{B}_m$ , resulting in a substantial rotation angle variance even in the case of QPSK modulation.

## 8.10 Relation to the Gaussian noise model and the role of chromatic dispersion

The GN model has been derived under the assumption that the electric fields of the interfering channels are statistically independent Gaussian random processes [26–30]. Hence its predictions regarding the NLIN power are rigorously accurate (within the obvious limits of the first-order perturbation analysis) only when the interfering channels undergo Gaussian modulation. With other modulation formats, the GN model is known to have certain inaccuracies, which are particularly conspicuous in the case of distributed amplification systems [4], or few-span systems (less than 1000 km) using lumped amplification [5, 6, 27]. The improvement in the GN model’s accuracy with the number of spans is often attributed to chromatic dispersion [15, 18, 27], with the main argument being that with large chromatic dispersion the electric fields of the propagating channels approach Gaussian statistics and hence the assumption of Gaussianity, on which the GN model relies, is better satisfied. The difficulty with this argument is that it does not explain the fact that in systems using distributed amplification the inaccuracy of the GN model with non-Gaussian modulation does not seem to reduce with the length of the system [4], regardless of the amount of chromatic dispersion that the signals accumulate. The role of chromatic dispersion and its relation to the GN model’s accuracy can be explained in terms of the pulse collision picture introduced in this paper, as we elaborate in what follows.

The GN model is accurate in the regime where modulation-format independent three and four-pulse collisions dominate the formation of NLIN. This does not occur in systems with distributed amplification, where the completeness of collisions implies that the effects of three and four-pulse collisions practically vanishes and the NLIN is dominated by two-pulse collision processes, independently of the link’s length, as can be seen in Fig. 8.5(a). In the regime of lumped optical amplification incomplete collisions become significant to the formation of NLIN only after the first few spans (see Fig. 8.5(b)), which is where the GN model’s accuracy begins to improve.

Another interesting aspect of the role of chromatic dispersion has been introduced in [15] and followed in [5]. It has been shown that in the presence of large pre-dispersion the predictions of the GN model become accurate already in the first span, where in the absence of pre-dispersion the GN model is known to be highly inaccurate. It has been suggested [15] that the better accuracy of the GN model in this case results from the fact that pre-dispersion transforms the distribution of the electric field of the interfering channels into Gaussian, independently of their modulation format. While tempting in its simplicity, this argument is inconsistent with an observation that was reported in [5], where it was shown that the inaccuracy of the predictions of the GN model in pre-dispersed systems is smaller than without pre-dispersion only in short links whose accumulated dispersion is much smaller than the pre-dispersion that was applied. For example, when the interfering channels are pre-dispersed by an amount equivalent to 500 km of transmission fiber, the GN model’s prediction improved in the first few hundred kilometers, but in longer links, of the order of 1000 km, it deteriorated to the same level of inaccuracy characterizing systems without pre-dispersion, as illustrated in Fig. 5 of [5]. Clearly, if the GN model’s accuracy improved because of better Gaussianity that followed from pre-dispersion, there is no reason for it

to deteriorate in longer systems, where even more dispersion is accumulated. The theory presented in this paper suggests a different explanation, that was briefly alluded to in [5]. Large pre-dispersion creates overlaps between all pulses at the link input so that the collisions taking place in the first few spans of the systems are incomplete. The incompleteness of the collisions emphasizes the significance of modulation format independent three and four-pulse collisions whose effect tends to agree with the predictions of the GN model. As the system becomes longer, the impact of pre-dispersion reduces and the inaccuracy of the GN model resumes its regular value.

## 8.11 Conclusions

We have shown that the various contributions to inter-channel NLIN can be classified as pulse collisions involving two, three, or four colliding pulses. Furthermore, we have argued that pulse collisions may either be complete (if they occur on a length-scale where loss/gain and dispersive broadening are negligible), or incomplete. When the majority of collisions are complete (as occurs with distributed gain systems [4] or in few-span systems using lumped amplification [5]) the NLIN is dominated by two-pulse collisions which give it the character of phase-noise and polarization-rotation noise with strong dependence on modulation format. When the majority of collisions are incomplete (high span-count systems with lumped amplification), the role of four-pulse interactions becomes more prominent, emphasizing the circular NLIN component, as assumed in [26–30]. In this case the dependence of NLIN on the modulation format is much weaker.

## Acknowledgement

The authors would like to acknowledge financial support from the Israel Science Foundation (grant 737/12). Ronen Dar would like to acknowledge the support of the Adams Fellowship of the Israel Academy of Sciences and Humanities, the Yitzhak and Chaya Weinstein Research Institute and the Advanced Communication Center and the Electro-Optic Fund of Tel-Aviv University.

## 8.12 Appendix: The dependence on $\Omega$ in complete versus incomplete collisions

We consider a generic setup of the kind plotted in Fig. 8.1. Two pulses, with indices 0 and  $h$ , belonging to the channel of interest collide with two pulses of the interfering channel, whose indices are  $k$  and  $m$ . Recall that the pulse with the index 0 is not a real pulse that propagates along the fiber, but rather (as we described earlier) it is included in the analysis so as to account for the effect of matched filtering.

We assume the regime of large accumulated dispersion, which is relevant in the case of

dispersion uncompensated links and take advantage of the far-field approximation [4, 10],

$$g(z, t) \simeq \sqrt{\frac{i}{2\pi|\beta''|z}} \exp\left(-\frac{it^2}{2\beta''z}\right) \tilde{g}\left(\frac{t}{\beta''z}\right), \quad (8.18)$$

where  $\tilde{g}(\omega) = \int g(t) \exp(i\omega t) dt$  is the Fourier transform of the fundamental waveform  $g(t)$ . By substituting Eq. (8.18) into Eq. (8.4) and using the change of variables  $\omega = t/\beta''z$ , we get

$$X_{h,k,m} \simeq \frac{e^{i\Omega hT}}{2\pi|\beta''|} \int_{z_0}^{z_1} \frac{f(z)}{z} \tilde{\psi}\left(\Omega + \frac{\bar{m}T}{\beta''z}\right) e^{i\frac{\bar{m}T}{\beta''z}hT} dz, \quad (8.19)$$

where  $\bar{m}$  stands for  $(m+k-h)/2$  and

$$\begin{aligned} \tilde{\psi}(u) = & \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{g}^*\left(\omega + u + \frac{hT}{2\beta''z}\right) \tilde{g}\left(\omega + u - \frac{hT}{2\beta''z}\right) \\ & \times \tilde{g}^*\left(\omega + \frac{(m-k)T}{2\beta''z}\right) \tilde{g}\left(\omega - \frac{(m-k)T}{2\beta''z}\right) \\ & \times e^{i\omega(h-k+m)T} d\omega. \end{aligned} \quad (8.20)$$

Note that the function  $\tilde{\psi}(u)$  reduces to zero when  $z$  satisfies  $\frac{|h|T}{2|\beta''|z} > \frac{B}{2}$  or  $\frac{|m-k|T}{2|\beta''|z} > \frac{B}{2}$ , where  $B$  is the spectral width of the fundamental pulse  $g(t)$ . In the limit of large dispersion, we can assume that  $\frac{|m-k|T}{2|\beta''|z} \ll \frac{B}{2}$  so that the product  $\tilde{g}^*\left(\omega + \frac{(m-k)T}{2\beta''z}\right) \tilde{g}\left(\omega - \frac{(m-k)T}{2\beta''z}\right)$  is approximately equal to  $|\tilde{g}(\omega)|^2$ . In the same manner, by assuming  $\frac{|h|T}{2|\beta''|z} \ll \frac{B}{2}$  we get

$$\tilde{\psi}(u) \simeq \frac{1}{2\pi} \int_{-\infty}^{\infty} |\tilde{g}(\omega + u)|^2 |\tilde{g}(\omega)|^2 e^{i\omega(h-k+m)T} d\omega. \quad (8.21)$$

Since  $\tilde{\psi}(u) \simeq 0$  when  $|u| \leq B$  we can conclude that the lower integration limit  $z_0$  in Eq. (8.19) is the largest of the three quantities  $\frac{|h|T}{|\beta''|B}$ ,  $\frac{|k-m|T}{|\beta''|B}$ , and  $\frac{\bar{m}T}{|\beta''|(\Omega+B)}$ , whereas the upper limit  $z_1$  is the smaller between the link length  $L$  and  $\frac{\bar{m}T}{|\beta''|(\Omega-B)}$ . These integration limits  $z_0$  and  $z_1$  can be intuitively understood when considering the nature of the nonlinear interaction, as illustrated in Fig. 8.1. In the general case, two pulses from the channel of interest interact with two pulses of the interfering channel, and the nonlinear interaction takes place while all four pulses overlap with each other temporally. The dispersive broadening of the individual pulses after propagating to point  $z$  along the fiber is approximately  $|\beta''|Bz$ . Within the far field approximation, significant overlap between the zeroth and the  $h$ -th pulses in the channel of interest is achieved when  $|\beta''|Bz \gg |h|T$  and similarly the  $k$ -th and  $m$ -th pulses in the interfering channel overlap when  $|\beta''|Bz \gg |m-k|T$ . In this case, the zeroth and  $h$ -th pulses form a single block of width  $\sim |\beta''|Bz$  centered at  $hT/2$ , whereas the  $k$ -th and  $m$ -th pulses form a block of the same width, centered at  $(m+k)T/2 - \beta\Omega z$ . The overlap between the two blocks starts when  $(m+k)T/2 - \beta\Omega z - \beta''Bz/2 > hT/2 + \beta''Bz/2$  and ends when  $(m+k)T/2 - \beta\Omega z + \beta''Bz/2 < hT/2 - \beta''Bz/2$ . The combination of these conditions translates into the integration limits indicated above, implying that when  $h$ ,  $k$  and  $m$  satisfy  $\bar{m} > \max\{|h|, |k-m|\}(1+\Omega/B)$ , a collision is formed if the condition  $\bar{m} < (\Omega+B)|\beta''|L/T$

is being satisfied. The collision is completed before the end of the link if  $h$ ,  $k$  and  $m$  further satisfy  $\bar{m} < (\Omega - B) |\beta''| L/T$ .

Finally, defining  $u = \Omega + \frac{\bar{m}T}{\beta''z}$ , Eq. (8.19) can be rewritten as

$$X_{h,k,m} = \frac{1}{2\pi |\beta''|} \int_{u_0}^{u_1} \frac{1}{\Omega - u} f\left(\frac{\bar{m}T}{|\beta''|(\Omega - u)}\right) e^{-iuhT} \tilde{\psi}(u) du, \quad (8.22)$$

where  $u_0 = \Omega + \frac{\bar{m}T}{\beta''z_1}$  and  $u_1 = \Omega + \frac{\bar{m}T}{\beta''z_0}$ . This form will be useful in what follows.

### 8.12.1 Complete collisions

In the case of a complete collision, the integrand in Eq. (8.22) vanishes at the integration boundaries, which can therefore be replaced by  $\pm\infty$ . In addition, since in the regime of complete collisions, attenuation and gain during the collision must be assumed negligible, we may replace  $f\left(\frac{\bar{m}T}{|\beta''|(\Omega - u)}\right)$  with  $f\left(\frac{\bar{m}T}{|\beta''|\Omega}\right)$  and move it outside of the integral. Having done so, we obtain

$$X_{h,k,m} = \frac{f\left(\frac{\bar{m}T}{|\beta''|\Omega}\right)}{2\pi |\beta''| \Omega} \sum_{n=0}^{\infty} \int_{-\infty}^{\infty} \frac{u^n}{\Omega^n} e^{-iuhT} \tilde{\psi}(u) du. \quad (8.23)$$

Defining  $\psi(v) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{\psi}(u) \exp(-iuv) du$ , as the inverse Fourier transform of  $\tilde{\psi}(u)$ , Eq. (8.23) can be rewritten as

$$X_{h,k,m} = \frac{1}{|\beta''| \Omega} f\left(\frac{\bar{m}T}{|\beta''|\Omega}\right) \sum_{n=0}^{\infty} \left(\frac{i}{\Omega}\right)^n \psi^{(n)}(hT), \quad (8.24)$$

where  $\psi^{(n)}(hT)$  is the  $n$ -th derivative of  $\psi(v)$ , evaluated at  $v = hT$ . From Eq. (8.21) it is evident that

$$\psi(v) = R^*(v) R(v - (h - k + m)T), \quad (8.25)$$

where  $R(v) = \int g^*(t) g(v - t) dt$  is the temporal autocorrelation function of the fundamental pulse  $g(t)$ . The orthogonality condition of Eq. (8.2) implies that  $R(v)$  has a maximum when  $v = 0$  ( $R(0) = 1$ ) and that it vanishes when  $v/T$  is any integer other than 0.

It is evident from Eq. (8.25) that the only contribution that scales with  $\Omega^{-1}$  is the one proportional to  $\psi^{(0)}(hT) = \psi(hT)$ . Substitution of  $v = hT$  in Eq. (8.25) reveals that  $\psi(hT) = R^*(hT) R((k - m)T)$  is non-zero only when  $h = 0$  and  $k = m$ . This is exactly the case of two-pulse collisions discussed earlier. The next contribution to NLIN, which scales as  $\Omega^{-2}$ , follows from  $\psi^{(1)}(hT) = R'(hT) R((k - m)T) + R(hT) R'((k - m)T)$  and it is nonzero only in the case of three-pulse collisions with  $h = 0$  and  $k \neq m$  (one pulse in the channel of interest and two pulses in the interfering channel), or with  $h \neq 0$  and  $k = m$  (two pulses in the channel of interest and a single pulse in the interfering channel). Only the term with  $n = 2$  in the summation in Eq. (8.24), which scales as  $\Omega^{-3}$  is nonzero in the case of complete four-pulse collisions (with  $h \neq 0$  and  $k \neq m$ ). Hence we may summarize that complete two-pulse collisions scale as  $\Omega^{-1}$ , complete three-pulse collisions scale as  $\Omega^{-2}$  and



complete four-pulse collisions scale as  $\Omega^{-3}$ . As can be deduced from the integration limits of Eq. (8.22), the number of complete two, three and four-pulse collisions is proportional to  $\Omega$ ,  $\Omega^2$  and  $\Omega^3$ , respectively, and therefore their overall contribution to NLIN (which is proportional to  $|X_{h,k,m}|^2$ ) is dominated by two-pulse collisions (see Sec. 8.7).

### 8.12.2 Incomplete collisions

The analysis of incomplete collisions is somewhat more complicated and difficult to extend beyond what is given by expressions (8.19), or (8.22). Some insight can be gained from considering the case of distributed amplification where  $f(z) = 1$  and can be taken out of the integral. In this case Eq. (8.23) still holds, except that the actual integration limits  $u_0$  to  $u_1$  need to be applied.

Using the formulation of the previous subsection  $\psi(v)$  is no longer given by Eq. (8.25), but rather it is the convolution of the expression on the right-hand-side of Eq. (8.25) with a sinc function corresponding to the inverse Fourier transform of a unit frequency window extending between  $u_0$  and  $u_1$ . In this situation  $\psi(hT)$  may differ from zero for all combinations of  $h, k$ , and  $m$ , implying that the contributions of all types of incomplete pulse collisions scale as  $\Omega^{-1}$ .

## 8.13 Appendix: Relative rotation between the modulated polarization axes of the interacting WDM channels

In the analysis of polarization multiplexing in Sec. 8.8, we assumed for simplicity that the modulated polarization axes in the interfering channel are parallel to those in the channel of interest. This assumption clearly does not represent the situation in reality where the modulated polarization axes in each of the two channels undergo different paths between the individual transmitters and the point at which they are wavelength multiplexed into a single transmission fiber. In addition, polarization mode dispersion in practical links may also cause relative polarization rotations between the various WDM channels. We examine these situations in this appendix.

Rigorously, the rotation of relative polarizations can be taken into account by replacing the vectors  $\underline{b}_j$  with  $\mathbf{U}\underline{b}_j$ , where  $\mathbf{U}$  is an arbitrary unitary matrix representing a relative rotation between the channel of interest and the interfering channel, changing the expression for the NLIN into

$$\Delta \underline{a}_0 = i\gamma \sum_{h,k,m} X_{h,k,m} \mathbf{U} \left( \underline{b}_k^\dagger \underline{b}_m \mathbf{I} + \underline{b}_m \underline{b}_k^\dagger \right) \mathbf{U}^\dagger \underline{a}_h. \quad (8.26)$$

According to Eq. (8.26), and by assuming that the components of  $\underline{a}_j$  and  $\underline{b}_j$  are isotropic in their phase-space, the NLIN variance is given by

$$\begin{aligned} \langle \|\Delta \underline{a}_0\|^2 \rangle - \|\langle \Delta \underline{a}_0 \rangle\|^2 \\ = \gamma^2 \sum_{\substack{h,k,m \\ h',k',m'}} X_{h,k,m}^* X_{h',k',m'} \langle \underline{a}_h^\dagger \bar{\Lambda}_{k,m}^\dagger \bar{\Lambda}_{k',m'} \underline{a}_{h'} \rangle, \end{aligned} \quad (8.27)$$



where  $\mathbf{\Lambda}_{k,m} = \mathbf{U} \left( \underline{b}_k^\dagger \underline{b}_m \mathbf{I} + \underline{b}_m \underline{b}_k^\dagger - \langle \underline{b}_k^\dagger \underline{b}_m \mathbf{I} + \underline{b}_m \underline{b}_k^\dagger \rangle \right) \mathbf{U}^\dagger$ , and the subtraction of the average  $\langle \underline{b}_k^\dagger \underline{b}_m \mathbf{I} + \underline{b}_m \underline{b}_k^\dagger \rangle$  inside the parenthesis eliminates the average (deterministic) phase shift of the entire constellation<sup>10</sup>. By further assuming that the polarization components of  $\underline{a}_j$  and  $\underline{b}_j$  are statistically independent and identically distributed, it can be readily shown that  $\langle \underline{a}_h^\dagger \bar{\mathbf{\Lambda}}_{k,m}^\dagger \bar{\mathbf{\Lambda}}_{k',m'} \underline{a}_{h'} \rangle$  is zero whenever  $h' \neq h$ ,  $k' \neq k$ , or  $m' \neq m$ , meaning that the NLIN variance is

$$\begin{aligned}
& \langle \|\Delta \underline{a}_0\|^2 \rangle - \|\langle \Delta \underline{a}_0 \rangle\|^2 \\
&= \gamma^2 \sum_{h,k,m} \langle \underline{a}_h^\dagger \bar{\mathbf{\Lambda}}_{k,m}^\dagger \bar{\mathbf{\Lambda}}_{k,m} \underline{a}_h \rangle |X_{h,k,m}|^2 \\
&= \gamma^2 \sum_{h,k,m} \text{Tr} \left( \langle \bar{\mathbf{\Lambda}}_{k,m} \underline{a}_h \underline{a}_h^\dagger \bar{\mathbf{\Lambda}}_{k,m}^\dagger \rangle \right) |X_{h,k,m}|^2 \\
&= \gamma^2 \langle |a|^2 \rangle \sum_{h,k,m} \langle \text{Tr} \left( \bar{\mathbf{\Lambda}}_{k,m} \bar{\mathbf{\Lambda}}_{k,m}^\dagger \right) \rangle |X_{h,k,m}|^2, \tag{8.28}
\end{aligned}$$

where  $\text{Tr}(\mathbf{A})$  denotes the *trace* of the matrix  $\mathbf{A}$ . Finally, since the trace of a matrix is invariant to unitary rotations, i.e.  $\text{Tr}(\mathbf{U}\mathbf{A}\mathbf{U}^\dagger) = \text{Tr}(\mathbf{A})$ , Eq. (8.28), and therefore also the NLIN variance, are independent of the relative rotation between the channels. Notice that symmetry between the two polarization components suggests that the NLIN variance in each polarization component is also invariant to  $\mathbf{U}$ .

Further insight into the effect of relative polarization rotation can be extracted from Eq. (8.15), which describes the effect of nonlinearity as a phase-noise  $\phi_n$ , polarization-rotation noise  $\vec{S}_n$ , and circular noise  $\nu_n$ . The variance of the circular noise  $\nu_n$  is independent of  $\mathbf{U}$ , as discussed in the previous paragraph. In addition, using Eqs. (8.10-8.13) it can be easily verified that both the phase shift  $\phi_n$  and the rotation angle  $|\vec{S}_n|$  are also independent of  $\mathbf{U}$ . Yet, in cases where the two polarization channels are processed separately from each other, the inclusion of  $\mathbf{U}$  may affect the phase-noise that is observed in each of the individual polarization components. This can be seen most conveniently from the contribution of two-pulse collisions to the phase-noise in the  $y$  component of the vector  $\underline{a}_0$ , which are given by

$$\gamma \sum_m X_{0,m,m} \left( 2|\tilde{b}_m^{(y)}|^2 + |\tilde{b}_m^{(x)}|^2 \right), \tag{8.29}$$

where  $\tilde{b}_m^{(x)}$  and  $\tilde{b}_m^{(y)}$  are the  $x$  and  $y$  component of the vector  $\mathbf{U}\underline{b}$ , respectively. Assume for simplicity the example of phase-modulated transmission (e.g. QPSK) where there is no rotation of the relative modulation axes ( $\mathbf{U} = \mathbf{I}$ ). In this case  $|\tilde{b}_m^{(x)}|^2 = |\tilde{b}_m^{(y)}|^2 = \text{Const}$  and the variance of the phase-noise due to the contribution of two-pulse collisions is 0. In the case where  $\mathbf{U} \neq \mathbf{I}$ ,  $|\tilde{b}_m^{(x)}|^2$  and  $|\tilde{b}_m^{(y)}|^2$  become dependent on the transmitted data so that two-pulse collisions contribute to a phase-noise with nonzero variance.

---

<sup>10</sup>The subtracted sum  $i\gamma \sum_{h,k,m} X_{h,k,m} \mathbf{U} \langle \underline{b}_k^\dagger \underline{b}_m \mathbf{I} + \underline{b}_m \underline{b}_k^\dagger \rangle \mathbf{U}^\dagger \underline{a}_h$  represents a deterministic phase-shift when  $k = m$  and  $h = 0$ . The summation over the terms  $k \neq m$  is clearly 0 because of the statistical independence between symbols and their isotropy. The summation over the terms  $h \neq 0$  can also be shown to be 0 by using the property that  $\sum_m X_{h,m,m} = 0$  for all  $h \neq 0$ .

# Bibliography

- [1] R.-J. Essiambre, G. Kramer, P.J. Winzer, G.J. Foschini, B. Goebel, “Capacity limits of optical fiber networks,” *J. Lightwave Technol.*, vol. 28, pp. 662–701 (2010).
- [2] A. Mecozzi, C. B. Clausen, and M. Shtaif, “Analysis of intrachannel nonlinear effects in highly dispersed optical pulse transmission,” *IEEE Photon. Technol. Lett.*, vol 12, pp. 392-394 (2000).
- [3] F. Forghieri, R.W. Tkack, A. R. Chraplyvy, “Fiber nonlinearities and their impact on transmission systems,” Ch. 8 in *Optical Fiber Telecommunications IIIA*, P. Kaminow and T. L. Koch eds. (Academic Press, 1997).
- [4] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, “Properties of nonlinear noise in long, dispersion-uncompensated fiber links,” *Optics Express*, vol. 21, pp. 25685–25699 (2013).
- [5] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, “Accumulation of nonlinear interference noise in fiber-optic systems,” *Optics Express*, vol. 22, pp. 14199–14211 (2014).
- [6] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, “Inter-channel nonlinear interference noise in WDM systems: modeling and mitigation,” accepted for publication in *J. Lightwave Technol.* (October 2014).
- [7] R. Dar, M. Shtaif, M. Feder, “New bounds on the capacity of the nonlinear fiber-optic channel,” *Optics Lett.*, vol. 39, pp. 398–401 (2014).
- [8] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, “Time varying ISI model for nonlinear interference noise,” in *Optical Fiber Communication (OFC) Conference*, pp. W2A.62, San Francisco (2014).
- [9] R. Dar, O. Geller, M. Feder, A. Mecozzi, M. Shtaif, “Mitigation of inter-channel nonlinear interference in WDM systems,” in *European Conference on Optical Communications (ECOC)*, P.5.6, Cannes (2014).
- [10] A. Mecozzi and R.-J. Essiambre, “Nonlinear Shannon limit in pseudolinear coherent systems,” *J. Lightwave Technol.*, vol. 30, pp. 2011–2024 (2012).
- [11] S. Kumar, D. Yang, “Second-order theory for self-phase modulation and cross-phase modulation in optical fibers,” *J. Lightwave Technol.*, vol. 23, pp. 2073–2080 (2005).

- [12] S. N. Shahi, S. Kumar, X. Liang, “Analytical modeling of cross-phase modulation in coherent fiber-optic system,” *Optics express*, vol. 22, pp. 1426–1439 (2014).
- [13] X. Liang, S. Kumar, “Analytical modeling of XPM in dispersion-managed coherent fiber-optic systems,” *Optics Express*, vol. 22, pp. 10579–10592 (2014).
- [14] A. Mecozzi and F. Matera, “Polarization scattering by intra-channel collisions,” *Optics Express*, vol. 20, pp. 1213–1218 (2012).
- [15] A. Carena, G. Bosco, V. Curri, P. Poggiolini, F. Forghieri, “Impact of the transmitted signal initial dispersion transient on the accuracy of the GN-model of non-linear propagation,” in *European Conference on Optical Communication (ECOC)*, Paper Th.1.D.4, London (2013).
- [16] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, F. Forghieri, “EGN model of non-linear fiber propagation,” *Optics Express*, vol. 22, pp. 16335–16362 (2014).
- [17] N. Rossi, P. Ramantanis, J.-C. Antona, “Nonlinear interference noise statistics in unmanaged coherent networks with channels propagating over different lightpaths,” in *European Conference on Optical Communications (ECOC)*, Paper Mo.4.3.4, Cannes (2014).
- [18] P. Serena, A. Bononi, “On the accuracy of the Gaussian nonlinear model for dispersion-unmanaged coherent links,” in *European Conference on Optical Communications (ECOC)*, Paper Th.1.D.3, London (2013).
- [19] L. F. Mollenauer, S. G. Evangelides, J.P. Gordon, “Wavelength division multiplexing with solitons in ultra-long distance transmission using lumped amplifiers,” *J. Lightwave Technol.*, vol. 9, pp. 362–367 (1991)
- [20] S. G. Evangelides Jr, J. P. Gordon, “Energy transfers and frequency shifts from three soliton collisions in a multiplexed transmission line with periodic amplification,” *J. Lightwave Technol.*, vol. 14, pp. 1639–1643 (1996).
- [21] J. F. L. Devaney, W. Forysiak, A. M. Niculae, N. J. Doran, “Soliton collisions in dispersion-managed wavelength-division-multiplexed systems,” *Optics Lett.*, vol. 22, pp. 1695–1697 (1997)
- [22] P. V. Mamyshev and L. F. Mollenauer, “Soliton collisions in wavelength-division-multiplexed dispersion-managed systems,” *Optics Lett.*, vol. 24, pp. 448–450 (1999)
- [23] M. Shtaif, “Analytical description of cross-phase modulation in dispersive optical fibers.” *Optics Lett.*, vol. 23, pp. 1191–1193 (1998).
- [24] J. Bromage, “Raman amplification for fiber communications systems,” *J. Lightwave Technol.*, vol. 22, p. 79 (2004).
- [25] F. Forghieri, R. W. Tkach, A. R. Chraplyvy, “Fiber nonlinearities and their impact on transmission system,” in *Optical Fiber Telecommunications IIIA*, I.P. Kaminow and T. L. Koch eds., Academic (1997)

- [26] P. Poggiolini, A. Carena, V. Curri, G. Bosco, F. Forghieri, “Analytical modeling of nonlinear propagation in uncompensated optical transmission links,” *IEEE Photon. Technol. Lett.*, vol. 23, pp. 742–744 (2011).
- [27] A. Carena, V. Curri, G. Bosco, P. Poggiolini, F. Forghieri, “Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links,” *J. Lightwave Technol.*, vol. 30, pp. 1524–1539 (2012).
- [28] P. Poggiolini, “The GN model of non-linear propagation in uncompensated coherent optical systems,” *J. Lightwave Technol.*, vol. 30, pp. 3857–3879 (2012).
- [29] P. Johannisson and M. Karlsson, “Perturbation analysis of nonlinear propagation in a strongly dispersive optical communication system,” *IEEE J. Lightwave Technol.*, vol. 31, pp. 1273–1282 (2013).
- [30] P. Serena, A. Bononi, “An alternative approach to the Gaussian noise model and its system implications,” *J. Lightwave Technol.*, vol 31, pp. 3489–3499 (2013).
- [31] J. P. Gordon and H. Kogelnik, “PMD fundamentals: Polarization mode dispersion in optical fibers,” *Proc. Natl. Acad. Sci. USA*, vol. 97, pp. 4541–4550 (2000).

# Chapter 9

## Shaping in the Nonlinear Fiber-Optic Channel

### 9.1 Channel model

We consider a WDM transmission where several channels are launched into a nonlinear optical fiber. The received signal at the  $n$ -th time-slot in the channel of interest is given by

$$y_n = a_n + \Delta a_n + z_n , \quad (9.1)$$

where  $a_n$  is the data symbol transmitted over the channel of interest at the  $n$ -th symbol time, and where  $\Delta a_n$  is the interference induced by the nonlinearity of the fiber. The amplified spontaneous emission (ASE) noise in the system is represented by  $z_n$  and is modeled as a linear additive white complex Gaussian noise with zero mean and variance  $\sigma_{\text{ASE}}^2$ .

In Eq. (9.1) we assume that the channel of interest has been back-propagated so as to eliminate the intra-channel nonlinearities, implying that only inter-channel interference is included in  $\Delta a_n$ <sup>1</sup>. The interference resulting from a single interfering channel was shown to have the form (see Chapter 4)

$$\Delta a_n = \sum_{l,k,m} a_{n-l} b_{n-k}^* b_{n-m} X_{l,k,m} . \quad (9.2)$$

The symbol  $b_n$  is the  $n$ -th data symbol of the interfering channel and  $X_{l,k,m}$  is a coefficient whose explicit form is provided in Chapter 4 and whose value is defined by the waveforms of the individual pulses, the channel separation, the dispersion coefficient, the link distance and the link power loss/gain profile. For simplicity of notation, Eq. (9.2) was written for the case of a single interfering channel, whereas in the presence of multiple interferers  $\Delta a_n$  is approximately equal to the sum of the individual contributions<sup>2</sup>.

---

<sup>1</sup>The analysis presented in this chapter can be derived in a similar way also for intra-channel nonlinearities. In fact, these NLIN contributions may also be suppressed by spherical constellations. The investigation of the achievable shaping gains resulting from both intra- and inter-channel NLIN suppression is left for future study.

<sup>2</sup>Four-wave mixing effects, i.e. nonlinear interactions that do not involve the channel of interest, and nonlinear interactions of ASE noise with itself and with the signal, are considered to be much smaller than those discussed here.

The first important feature of NLIN is the dependance on the average power of the transmitted signals. In particular, Eq. (9.2) suggests that the variance of  $\Delta a_n$  is proportional to  $\langle |a|^2 \rangle \langle |b|^2 \rangle^2$ , where  $\langle |a|^2 \rangle$  and  $\langle |b|^2 \rangle$  denote the average symbol energies in the channel of interest and in the interfering channel, respectively. Assuming that the average transmission power of both channels is  $P$ , the power of NLIN is proportional to  $P^3$ . Therefore, the NLIN is small at low average signal powers, whereas at high power levels it becomes the predominant effect. The tradeoff is between increasing  $P$  to obtain higher SNR with respect to the linear Gaussian noise, but also keeping it low enough so as to avoid the effects of NLIN. The optimization of the transmission power will be further discussed in the following section.

A second interesting feature of NLIN is the dependance on the modulation format. This dependance is the result of the terms in Eq. (9.2) with  $k = m$ , i.e., the sum

$$\sum_l a_{n-l} \sum_m |b_{n-m}|^2 X_{l,m,m}. \quad (9.3)$$

Assuming that both channels independently transmit independent and identically distributed (i.i.d.) symbols, the variance of the noise induced by these terms is proportional to

$$\langle |a|^2 \rangle (\langle |b|^4 \rangle - \langle |b|^2 \rangle^2). \quad (9.4)$$

For phase-shift keying (PSK) input constellations, for example, the variance is strictly zero while it is proportional to  $\langle |b|^2 \rangle^2$  for Gaussian modulation format. Because of the dependance on the amplitude modulation of the interfering data symbols, we refer to the noise induced by these terms as *amplitude modulation induced noise* (AMIN). In Chapter 6 it was shown that the dependance on modulation format is largest in short distance links and in systems of arbitrary length that use a perfectly distributed amplification scheme. In these cases NLIN is predominated by the AMIN and transmission schemes that suppress its affect are highly alluring.

As was mentioned above, PSK modulation format completely removes the AMIN terms and significantly reduces the NLIN in certain systems. In doing so the amplitude of the data symbols does not carry information and half of the degrees of freedom are lost. More efficient input constellations can be applied when taking into account the fact that the coefficient  $X_{l,m,m}$  does not change significantly with  $m$ . Figure 9.1 shows the auto-correlation of the AMIN terms, i.e.,  $R_l(n) = \frac{\sum_m X_{l,m,m} X_{l,m+n,m+n}^*}{\sum_m |X_{l,m,m}|^2}$ , for a single amplified span of 100km. The bandwidth of the WDM channels is 100GHz, where the channel separation between the two channels is 204GHz (for the complete set of system parameters, see section 9.3). Note that the correlation reduces to 0.5 after 40 symbols for  $l = 0$  and after 16 symbols for  $l = 1, 2, 3$ , indicating the fact that for a given  $l$ , the coefficients  $X_{l,m,m}$  are slowly changing with  $m$ . In particular, one can assume that  $X_{l,m,m} \approx X_{l,m+1,m+1}$  for each  $m$ . Now, suppose that for even  $m$ 's the symbols  $b_m$  and  $b_{m+1}$  are chosen from the envelope of a 4-dimensional ball, i.e., satisfying  $|b_m|^2 + |b_{m+1}|^2 = 2P$ . In this case the AMIN terms are proportional to

$$\sum_{\substack{m \\ m \text{ is even}}} [X_{l,m,m} |b_m|^2 + X_{l,m+1,m+1} |b_{m+1}|^2] \approx 2P \sum_{\substack{m \\ m \text{ is even}}} X_{l,m,m} \quad (9.5)$$

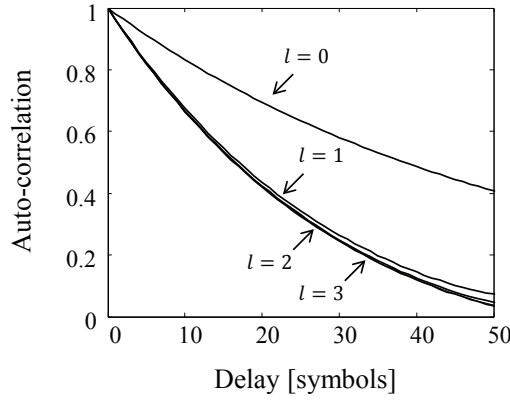


Figure 9.1: Auto-correlation of the AMIN for a single span of 100km.

which is constant (and is even equal to zero for  $l \neq 0$  [28]). The AMIN terms are therefore “shaped-out” by a multidimensional sphere constellation at the expense of losing only one quarter of the degrees of freedom. Furthermore, if the input symbols are chosen from a ball shaped input constellation, they are relatively close to the ball envelope with high probability (a probability that increases with the number of dimensions of the ball). In this case, with high probability the AMIN terms are constant and the induced noise variance reduces. Therefore, one may suppress the NLIN effects by shaping the input constellation. This idea is the subject of this work and it is formalized and discussed in the following section.

## 9.2 Shaping gain in the fiber-optic channel

We start by computing a lower bound on the mutual information between the input and output of the channel. The vectors  $\underline{a}$  and  $\underline{b}$  represent in what follows the sequence of transmitted data symbols from the channel of interest and from the interfering channel, respectively. The vector  $\underline{y} = \underline{a} + \Delta\underline{a} + \underline{z}$  contains the symbols received by the channel of interest, the vector  $\underline{z}$  contains the ASE noise samples and  $\Delta\underline{a}$  contains the NLIN samples in the channel of interest (similarly,  $\Delta\underline{b}$  contains the NLIN samples received in the interfering channel). We shall also use the common assumptions that the channel of interest and the interferer are non-cooperative and are transmitting at the same average power level.

The differential entropy of the received samples can be lower bounded by the entropy of the input samples since  $h(\underline{y}) \geq h(\underline{y}|\underline{z}, \underline{b} + \Delta\underline{b}) = h(\underline{a})$ . This follows from the fact that conditioning reduces entropy and the fact that given  $\underline{b} + \Delta\underline{b}$ , joint back propagation can be applied with the samples  $\underline{a} + \Delta\underline{a}$ , so as to get the transmitted symbols at the beginning of the link,  $\underline{a}$  and  $\underline{b}$ , without loosing information. The differential entropy of the output given the input can be upper bounded by  $h(\underline{y}|\underline{a}) \leq n_0 \log[\pi e(\sigma_{\text{NL}}^2 + \sigma_{\text{ASE}}^2)]$ , where  $n_0$  is the length of the vectors  $\underline{a}$  and  $\underline{y}$ , and where  $\sigma_{\text{NL}}^2$  and  $\sigma_{\text{ASE}}^2$  are the variances of the entries of  $\Delta\underline{a}$  and  $\underline{z}$ , respectively<sup>3</sup>. This inequality follows from the fact that conditioning reduces entropy and

<sup>3</sup>When the transmitted sequences  $\underline{a}$  and  $\underline{b}$  are not i.i.d. the NLIN variance may be time dependant. In this case,  $\sigma_{\text{NL}}^2$  can be viewed as the average of the variances which, according to Jensen’s inequality, is aligned



the fact that Gaussian distribution maximizes the differential entropy of a random variable with a given covariance matrix.

The resultant lower bound on the mutual information of the channel is therefore

$$\frac{1}{n_0} I(\underline{a}; \underline{y}) \geq \frac{1}{n_0} h(\underline{a}) - \log [\pi e (\sigma_{\text{NL}}^2 + \sigma_{\text{ASE}}^2)] \quad (9.6)$$

We note that this bound may be loose to some extent as we have applied some relaxing assumptions along the way. In particular, the bound on  $h(\underline{y}|\underline{a})$  assumes that the NLIN samples,  $\Delta \underline{a}$ , are memoryless, statistically independent of  $\underline{a}$  and follow a Gaussian distribution. As discussed in the previous sections, in some scenarios the memoryless assumption is not too restrictive. In the case of short links, for example, the dynamics of NLIN are fast changing and the memory property cannot be exploited to enhance the information rates. Secondly, the assumption of statistical independence between NLIN and the input samples is a strictly restrictive assumption, as the dependance can be clearly observed in Eq. (9.2). The third assumption, on the Gaussianity of the NLIN, is not too harsh since each sample of NLIN involves many independent nonlinear interactions and therefore it is indeed approximately point-wise Gaussian. In addition, the lower bound on  $h(\underline{y})$  assumes that the entropy of the received samples is dominated by the entropy of the input samples. This assumption however is well justified in the regimes of interest, namely at the high SNR regime where the average input power is much larger than the ASE noise as well as the NLIN (as can be seen from Fig. 9.3b).

Upon reviewing the assumptions above, we stress that the resulted bound on the mutual information may be relatively loose. It allows us, however, to derive some very interesting insights on the ability to shape the input constellation not only to reduce the average transmission power but also to “shape-out” the nonlinear interference. The lower bound presented in Eq. (9.6) further allows us to derive an analytical expression for the achievable shaping gain in the fiber-optic channel, an expression that accounts also for the dependance of NLIN on the input constellation. In what follows we compute the bound in Eq. (9.6) for the cases of uniform input distribution within a  $2N$ -dimensional cube and ball, and derive an analytical expression for the shaping gain in the fiber-optic channel.

### 9.2.1 $2N$ -dimensional cube

Let the data symbols of the channel of interest and the interfering channel be uniformly distributed over a  $2N$ -dimensional cube. In order to compute the lower bound of Eq. (9.6) we need to derive the entropy of the input distribution and the induced NLIN variance. The latter is given by

$$\sigma_{\text{NL}}^2 = P^3 \bar{X}_1 + P^3 \left( \frac{\langle |b_0|^4 \rangle}{\langle |b_0|^2 \rangle^2} - 2 \right) \bar{X}_2 \quad (9.7)$$

where we used the fact that the symbols  $a_l$  and  $b_m$  are i.i.d. and statistically independent of each other, and where we denote the quantities  $\sum_{l,k,m} |X_{l,k,m}|^2$  and  $\sum_{l,m} |X_{l,m,m}|^2$  by  $\bar{X}_1$  and  $\bar{X}_2$ , respectively. The quantity  $\mathcal{K} = \langle |b_0|^4 \rangle / \langle |b_0|^2 \rangle^2 - 2$  accounts for the dependance

---

with our goal of upper bounding the entropy  $h(\underline{y}|\underline{a})$ .



on the fourth order moment (that implies a dependance on the amplitude modulation of the interfering symbols), and in our case it is equal to -0.6. For 4-QAM and 16-QAM constellations,  $\mathcal{K}$  is equal to -1 and -0.68, respectively, whereas it approaches to -0.6 as the size of the QAM constellation increases.

The entropy of a random vector that is uniformly distributed within some multi-dimensional region is equal to the logarithm of the region volume [46]. The entropy of the input in our case is therefore  $\log(6P)$  and the resulting lower bound on the mutual information follows

$$\frac{1}{n_0} I(\underline{a}; \underline{y}) \geq \log \left( \frac{6}{\pi e \sigma_{\text{ASE}}^2 + P^3 (\bar{X}_1 - 0.6\bar{X}_2)} \frac{P}{P} \right) . \quad (9.8)$$

The optimal average transmission power is the one that optimally accounts both for the linear and nonlinear noises and maximizes the mutual information at the output of the channel. The average power that maximizes the expression on the right hand side of Eq. (9.8) (and actually maximizes the effective SNR at the receiver) is equal to

$$P_{\text{opt}}^{\square} = \left( \frac{1}{2} \frac{\sigma_{\text{ASE}}^2}{\bar{X}_1 - 0.6\bar{X}_2} \right)^{1/3} , \quad (9.9)$$

and the maximum mutual information is hereby bounded by

$$I_{\text{max}}^{\square} \triangleq \sup_P \frac{1}{n_0} I(\underline{a}; \underline{y}) \geq \log \left( \frac{4}{\pi e} \frac{P_{\text{opt}}^{\square}}{\sigma_{\text{ASE}}^2} \right) . \quad (9.10)$$

### 9.2.2 $2N$ -dimensional ball

In the case of uniform input distribution within a  $2N$ -dimensional ball, the input symbols are statistically dependant and using straight forward calculations the NLIN variance can be shown to satisfy

$$\begin{aligned} \sigma_{\text{NL}}^2 &= P^3 \frac{\langle |b_0|^2 |b_1|^2 \rangle}{\langle |b_0|^2 \rangle^2} \bar{X}_1 \\ &+ P^3 \left( \frac{\langle |b_0|^4 \rangle - \langle |b_0|^2 |b_1|^2 \rangle}{\langle |b_0|^2 \rangle^2} - 1 \right) \bar{X}_2 \\ &+ P^3 \left( \frac{\langle |b_0|^2 |b_1|^2 \rangle}{\langle |b_0|^2 \rangle^2} - 1 \right) \bar{X}_{\text{COV}} \end{aligned} \quad (9.11)$$

where  $b_0$  and  $b_1$  are two correlated input symbols and the term  $\bar{X}_{\text{COV}}$  is given by

$$\bar{X}_{\text{COV}} = 2 \sum_{l,s} \sum_{m=sN}^{(s+1)N-2} \sum_{k=m+1}^{(s+1)N-1} \Re\{X_{l,m,m} X_{l,k,k}^*\} . \quad (9.12)$$

Now, suppose that each  $M$  consecutive coefficients  $X_{l,m,m}, \dots, X_{l,m+M-1,m+M-1}$  are approximately equal. In this case, if  $N \leq M$  then

$$\bar{X}_{\text{COV}} \approx (N-1) \sum_{l,m} |X_{l,m,m}|^2 = (N-1) \bar{X}_2 . \quad (9.13)$$

If  $N > M$ , then  $\bar{X}_{\text{COV}}$  is approximately equal to  $(M-1)\bar{X}_2$ <sup>4</sup>. By using straightforward calculations we find that  $\langle |b_0|^4 \rangle = 2P^2(N+1)/(N+2)$  and  $\langle |b_0|^2 |b_1|^2 \rangle = P^2(N+1)/(N+2)$ , and the NLIN variance follows

$$\sigma_{\text{NL}}^2 \approx P^3 \frac{N+1}{N+2} \bar{X}_1 - P^3 \frac{\bar{M}}{N+2} \bar{X}_2 \quad (9.14)$$

where  $\bar{M} = \min\{N, M\}$ .

Since the volume of the ball is  $V_{2N}(P) = (\pi(N+1)P)^N/N!$  and the entropy of the input is  $\log(V_{2N}(P))$  we get that the optimal average power is

$$P_{\text{opt}}^\odot = \left( \frac{1}{2} \frac{\sigma_{\text{ASE}}^2}{\frac{N+1}{N+2} \bar{X}_1 - \frac{\bar{M}}{N+2} \bar{X}_2} \right)^{1/3} \quad (9.15)$$

and the lower bound on the maximum mutual information is

$$I_{\text{max}}^\odot \triangleq \sup_P \frac{1}{n_0} I(\underline{a}; \underline{y}) \geq \log \left( \frac{2(N+1)}{3e(N!)^{1/N}} \frac{P_{\text{opt}}^\odot}{\sigma_{\text{ASE}}^2} \right). \quad (9.16)$$

### 9.2.3 Shaping gain

In fiber-optic systems one can arbitrarily choose the average transmission power, and so we would like to compare the *maximum* information rate of the various input distributions. The difference between the bounds on the maximum mutual information of a uniform distribution within a  $2N$ -dimensional ball and within a cube is given by

$$I_{\text{max}}^\odot - I_{\text{max}}^\square = \log \left( \frac{\pi(N+1)}{6(N!)^{1/N}} \frac{P_{\text{opt}}^\odot}{P_{\text{opt}}^\square} \right). \quad (9.17)$$

Notice that the first quotient in the expression inside the logarithm is exactly the shaping gain of a  $2N$ -dimensional ball in the linear AWGN channel. The second quotient accounts for the fact that the optimal average power depends on the NLIN variance which is affected by the shaping of the input constellation. Hence, on the one hand, shaping the input constellation may reduce the average power of the transmitted symbols, and on the other hand, it may reduce the NLIN variance and therefore increase the optimal average power.

We therefore can write the shaping gain in the optical fiber channel as

$$\gamma_s^{\text{OFC}}(\odot) = \gamma_s^{\text{AWGN}}(\odot) \gamma_s^{\text{NL}}(\odot) \quad (9.18)$$

where  $\gamma_s^{\text{AWGN}}(\odot)$  is the shaping gain of a  $2N$ -dimensional ball in the AWGN channel and where  $\gamma_s^{\text{NL}}(\odot)$  is the gain from “shaping-out” the NLIN. By plugging the optimal average power levels from Eqs. (9.9) and (9.15) we get that the latter is given by

$$\gamma_s^{\text{NL}}(\odot) = \left( \frac{\bar{X}_1 - 0.6\bar{X}_2}{\frac{N+1}{N+2} \bar{X}_1 - \frac{\bar{M}}{N+2} \bar{X}_2} \right)^{1/3} \quad (9.19)$$

---

<sup>4</sup>A more precise expression should also account for the cross terms  $X_{l,m,m} X_{l,k,k}^*$  with  $|k-m| > M$ , however, the significance of these terms is relatively small when  $M$  is chosen to be large enough.

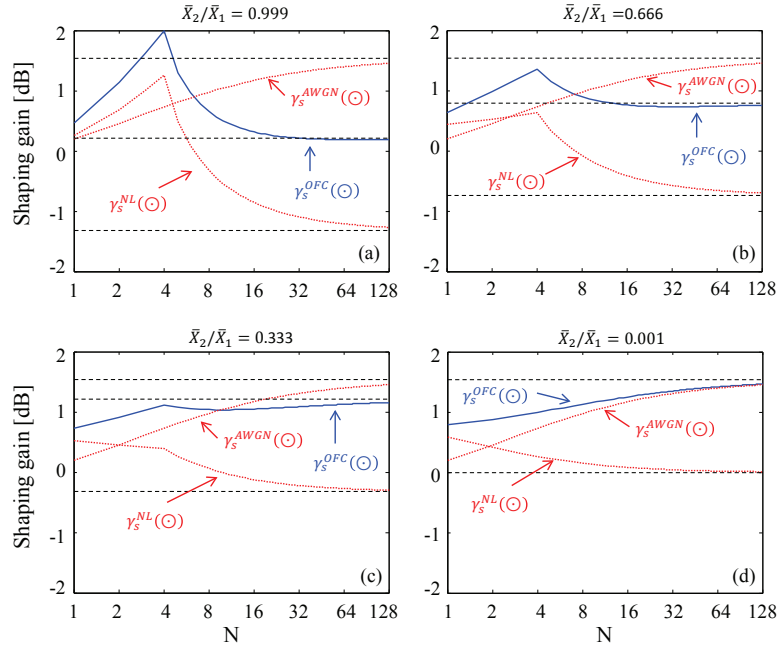


Figure 9.2: Shaping gain of a  $2N$ -dimensional ball shaped input for  $M = 4$  and  $\frac{\bar{X}_2}{\bar{X}_1} = 1000, 2, 0.5, 0.001$  (figs. (a), (b), (c) and (d), respectively). Blue curves represent the resultant shaping gain, which is equal to the multiplication of the linear AWGN shaping gain and the nonlinear shaping gain (both are represented by the dotted red curves). Dashed black curves depict the asymptotic value of the curves.

where  $\bar{M} = \min\{N, M\}$ .

The nonlinear shaping gain presented in Eq. (9.19) depends on the ratio between  $\bar{X}_1$  and  $\bar{X}_2$ . When this ratio is very small, the nonlinear shaping gain is approximately equal to 1. In this case the total shaping gain is dominated by the linear shaping gain and large values of  $N$  are desirable. When  $\bar{X}_2/\bar{X}_1$  is close to 1 the maximum nonlinear shaping gain is obtained with  $N = M$  and is approximately equal to  $0.4M$ . In this case  $N \rightarrow \infty$  may not be the optimal choice that maximizes the total shaping gain. The dependance of  $\gamma_s^{\text{NL}}(\odot)$  on the ratio  $\bar{X}_2/\bar{X}_1$  can be explained as follows. First recall that  $\bar{X}_1$  is equal to  $\sum_{l,k,m} |X_{l,k,m}|^2$  and that  $\bar{X}_2$  is equal to  $\sum_{l,m} |X_{l,m,m}|^2$ . Since the variance of the AMIN contributions is proportional to  $\bar{X}_2$ , the ratio  $\bar{X}_2/\bar{X}_1$  actually defines the significance of these terms with respect to all other nonlinear noise contributions. Since most of the gain in “shaping-out” the nonlinear noise is in the AMIN terms, it actually defines the importance of the nonlinear shaping gain – as the ratio  $\bar{X}_2/\bar{X}_1$  increases, the significance of the AMIN terms becomes larger and so as the gain from shaping them out.

Figure 9.2 shows the shaping gain for the case of  $M = 4$  and various ratios of  $\bar{X}_2/\bar{X}_1$ . Solid blue curves represent  $\gamma_s^{\text{OFC}}(\odot)$  whereas dotted red curves represent  $\gamma_s^{\text{NL}}(\odot)$  and  $\gamma_s^{\text{AWGN}}(\odot)$ . Notice that the nonlinear shaping gain is largest for small values of  $N$  and the maximum gain increases with the ratio  $\bar{X}_2/\bar{X}_1$ . In addition, in the cases depict in Figs. 9.2a and 9.2b, the maximum shaping gain is higher than the ultimate linear shaping gain in AWGN channels (1.53dB) and it is obtained with  $N = 4$ .

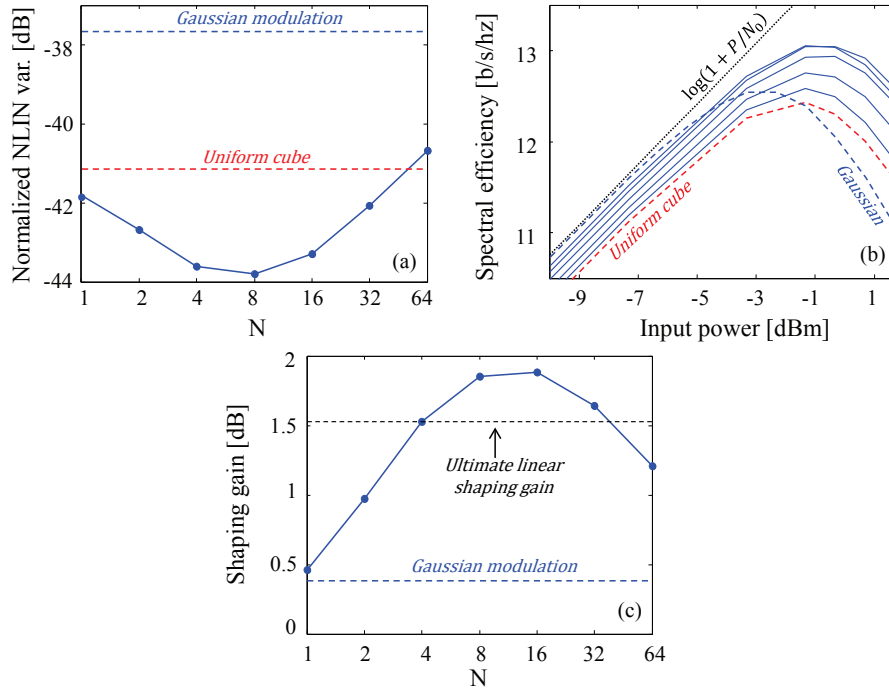


Figure 9.3: Dashed red curve represents the case of a uniform input distribution within a cube, dashed blue curve represents the case of a Gaussian input distribution, and solid blue curves represent the case of a uniform input distribution within a  $2N$ -dimensional ball. (a) NLIN variance (normalized by the input power) for an average input power of  $-0.3\text{dBm}$ . Dots indicate the simulated cases. (b) Spectral efficiency versus average input power, where the solid blue curves present the results for a ball shaped input with  $N = 1, 2, 4, 8, 16$ . Dotted black curve shows the linear AWGN channel capacity  $\log(1 + P/\sigma_{\text{ASE}}^2)$ . (c) Shaping gain versus  $N$ . Dots indicate the simulated cases.

### 9.3 Simulation results

We have performed a set of numerical simulations in order to extract the NLIN variance  $\sigma_N^2$  and to evaluate the shaping gain in a single span link of  $100\text{km}$ . The simulations were performed using the parameters of a standard single mode fiber; dispersion  $D = 17\text{ ps/nm/km}$ , attenuation of  $0.2\text{ dB/km}$ , nonlinear coefficient  $\gamma = 1.27\text{ W}^{-1}\text{km}^{-1}$  and signal wavelength  $\lambda_0 = 1.55\text{ }\mu\text{m}$ . Sinc-shaped pulses with a perfectly square  $100\text{ GHz}$  wide spectrum were used for transmission and the spacing between adjacent WDM channels was  $102\text{ GHz}$ . The number of simulated WDM channels was five, with the central channel being the channel of interest. The number of simulated symbols in each run was 2048 and up to 500 runs were performed so as to accumulate sufficient statistics. At the receiver, the central channel was isolated with a matched filter and back-propagated so as to eliminate the effects of intra-channel nonlinearity and chromatic dispersion. As in [2], all simulations have been performed with the scalar nonlinear Schrödinger equation and correspond to a single polarization.

Figure 9.3a depicts the extracted variance of NLIN (normalized by the input power) when the input power is  $0\text{dBm}$  (similar curves were obtained for other levels of input power).

Dashed red curve represent the case of a uniform input distribution within a cube, solid blue curve represents the case of a uniform input distribution within a  $2N$ -dimensional ball (dots indicate the simulated cases), and dashed blue curve represents the case of a Gaussian input distribution (equivalent to the case of  $N \rightarrow \infty$ ). Evidently, a ball shaped input with  $N = 8$  attains the smallest NLIN variance, a value which is smaller by 2.7dB and 6.1dB than the NLIN variance for the case of uniform input distribution within a cube and a Gaussian input distribution, respectively.

In Fig. 9.3b we use the extracted NLIN variances in order to compute the lower bound on the mutual information given by (9.6), where we draw only the curves of  $N = 1, 2, 4, 8, 16$ . For low levels of average input power the nonlinear effects are small and the fiber-optic channel is approximately a linear AWGN channel. In this regime the mutual information in the case of a Gaussian input distribution is identical to the linear capacity and the mutual information in the case of a uniform distribution within a cube approaches to the linear capacity up to the ultimate shaping gain of 1.53dB. For higher levels of average input power, the nonlinear effects become stronger and limits the achievable information rates. The maximum spectral efficiency is obtained with a ball shaped input of  $N = 16$ . Note that the maximum spectral efficiency is notably higher than the maximum spectral efficiency obtained with a Gaussian input.

Figure 9.3c shows the resultant shaping gain, i.e.,  $2^{I_d}$  where  $I_d$  is the difference between the maximum mutual information obtained by the various input distributions and the maximum mutual information obtained by the uniform distribution within a cube. The maximum shaping gain is obtained with  $N = 16$  – while the case of Gaussian input distribution obtains only a 0.38dB shaping gain, a ball shaped input with  $N = 16$  obtains a gain of 1.88dB which is clearly better than the 1.53dB ultimate shaping gain in a linear AWGN channel.



# Chapter 10

## Summary and Conclusions

In the context of SDM transmission in fiber-optic communications, we have studied the under-addressed optical MIMO channel, where the number of fiber modes addressed by the transmitter and receiver is allowed to be smaller than the overall number of modes existing in the fiber. This scenario was motivated by the idea of installing fibers that admit more modes than can be processed with currently available technology, mainly in order to achieve future-proof operation [14]. By establishing the relation between the channels singular values and the Jacobi ensemble of random matrices we derived the ergodic capacity, outage probability and optimal diversity-multiplexing tradeoff. An interesting phenomenon is observed when the parameters of the model satisfy  $m_t + m_r > m$  (where  $m$  is the total number of modes that are available in the fiber, and where  $m_t$  and  $m_r$  are the number of modes that are addressed by the transmitter and receiver, respectively): for any realization of the transfer matrix,  $m_t + m_r - m$  singular values are 1. This results in an ergodic capacity which is at least  $m_t + m_r - m$  times the SISO capacity. In the non-ergodic scenario this enables a strictly zero outage probability and an exponentially decaying error probability (“infinite diversity”) for any transmission rate below  $(m_t + m_r - m)\log(1 + \text{SNR})$ . The results of Chapters 2 and 3 therefore quantify the information theoretical limits of optical SDM systems with respect to  $m$ ,  $m_t$  and  $m_r$  and they may be used to understand the tradeoff between system performance and the complexity and cost that come along with the utilization of more modes. For example, one important outcome that may be relevant also for SDM systems that address most of the available modes, is the fact that a performance equivalent of  $m_t + m_r - m$  uncoupled single-mode channels is always achievable. It should be noted however that Chapters 2 and 3 study the information theoretical limitations imposed by not addressing all modes while ignoring the limitations imposed by the nonlinearity of the fiber. The impact of nonlinearities on the results of Chapters 2 and 3 is left for future research.

The main focus of my PhD research was the modeling problem of nonlinear propagation in fiber-optic communications. We have reviewed the essential parts of two previous modeling approaches, the time domain model and the frequency domain GN model, and pointed out the differences between the models. We have shown that by correctly accounting for fourth-order correlations in the NLIN spectrum an extra term arises. The inclusion of this term recovers the dependence of the NLIN power on modulation format, a property that is absent from the existing GN model and reconciles between the frequency domain and the time

domain theories. It should be noted however that further work is needed to understand the impact of nonlinearities on measures of system performance beyond the effective SNR (such as BER and achievable information rates), accounting for the unique properties of NLIN. In addition, our analysis throughout this dissertation considers systems that are optimized to work over linear channels, assuming, for example, single-carrier transmission and linear matched filtering at the receiver. The investigation of nonlinearities in other system configurations (such as systems employing Nonlinear Fourier transform [47–49]) is left for future research.

We have derived a new lower bound on the capacity of the fiber-optic channel, which exploits the fact that part of the nonlinear distortions are in the form of slowly varying phase-noise. We showed that the peak capacity per polarization can be increased by approximately 0.7bit/s/Hz or equivalently the length of a system can be (almost) doubled for a given transmission rate. The idea of mitigating the nonlinear distortions by taking advantage of temporal correlations was further generalized by describing the NLIN as time-varying ISI channel. We have evaluated the temporal correlations of the various ISI coefficients and examined the improvement that can be achieved in system performance when mitigating inter-channel NLIN by means of adaptive linear equalization (without joint processing).

The dependence of NLIN on the structure of the system was further examined. We have shown that the importance of the modulation format dependence of the NLIN variance and the relative significance of the phase-noise are largest in few-span systems, or in systems employing Raman amplification<sup>1</sup>, and they reduce notably in the case of lumped amplification with a large number of spans. A pulse-collision theory that explains the dependence of NLIN on the parameters of the system has been developed. In particular, we have shown that the various contributions to NLIN can be classified as pulse-collisions involving two, three, or four colliding pulses and they can either be complete, or incomplete. When the majority of collisions are complete (as occurs with distributed gain systems [24] or in few-span systems using lumped amplification [25]) the NLIN is dominated by two-pulse collisions which give it a character of phase and polarization noise with strong dependence on modulation format. When the majority of collisions are incomplete (high span-count systems with lumped amplification), four-pulse interactions provide the dominant contribution to NLIN, emphasizing the circularity of the noise and reducing its dependence on modulation format.

---

<sup>1</sup>The dynamics of NLIN depend on the power profile along the link and particularly on the magnitude of the power discontinuities between spans. As the variations in the power profile are less significant the modulation format dependence and phase-noise nature of NLIN become more notable. Practical Raman amplification may significantly reduce the power variations, emphasizing these aforementioned properties of NLIN.



# Bibliography

- [1] P. P. Mitra and J. B. Stark, “Nonlinear limits to the information capacity of optical fibre communications,” *Nature*, vol. 411, pp. 1027–1030 (2001).
- [2] R.-J. Essiambre, G. Kramer, P.J. Winzer, G.J. Foschini, B. Goebel, “Capacity limits of optical fiber networks,” *Journal of Lightwave Technology*, vol. 28, pp. 662–701 (2010).
- [3] K. S. Turitsyn, S. A. Derevyanko, I. V. Yurkevich, and S. K. Turitsyn, “Information capacity of optical fiber channels with zero average dispersion,” *Physical review letters*, vol. 91, pp. 203901–203904 (2003).
- [4] M. Secondini, E. Forestieri, G. Prati, “Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps,” *Journal of Lightwave Technology*, vol. 31, pp. 3839–3852 (2013).
- [5] E. Agrell and M. Karlsson, “WDM channel capacity and its dependence on multichannel adaptation models,” in *Optical Fiber Communication (OFC) conference*, pp. OTu3B-4, Anaheim, CA (2013).
- [6] A. Carena, V. Curri, G. Bosco, P. Poggiolini, and F. Forghieri, “Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links,” *Journal of Lightwave Technology*, vol. 30, pp. 1524–1539 (2012).
- [7] P. Serena, A. Bononi, “An alternative approach to the Gaussian noise model and its system implications,” *Journal of Lightwave Technology*, vol 31, pp. 3489–3499 (2013).
- [8] S. N. Shahi, S. Kumar, X. Liang, “Analytical modeling of cross-phase modulation in coherent fiber-optic system,” *Optics express*, vol. 22, pp. 1426–1439 (2014).
- [9] E. Agrell, “The channel capacity increases with power,” *arXiv preprint:1108.0391* (2011).
- [10] S. Berdagué and P. Facq, “Mode division multiplexing in optical fibers,” *Applied Optics*, vol. 21 pp. 1950–1955 (1982).
- [11] A. Tarighat, R. C. Hsu, A. Shah, A. H. Sayed, and B. Jalali, “Fundamentals and challenges of optical multiple-input multiple-output multimode fiber links,” *IEEE Communications Magazine*, vol. 45 (2007).

- [12] R. W. Tkach, “Scaling optical communications for the next decade and beyond,” Bell Labs Technical Journal, vol. 14, pp. 3–9 (2010).
- [13] A. Chralyvy, “The coming capacity crunch,” in European Conference on Optical Communication (ECOC), plenary paper, Vienna (2009).
- [14] P. J. Winzer and G. J. Foschini, “MIMO capacities and outage probabilities in spatially multiplexed optical transport systems,” Optics Express, vol. 19, pp. 16680–16696 (2011).
- [15] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, “On shaping gain in the nonlinear fiber-optic channel,” in IEEE International Symposium on Information Theory (ISIT), pp. 2794–2798, Honolulu (2014).
- [16] R. J. Muirhead, “Aspects of Multivariate Statistical Theory.” Wiley (1982).
- [17] M. L Mehta , “Random Matrices.” Academic Press (1991).
- [18] A. Edelman and N. R. Rao, “Random matrix theory,” Acta Numerica, vol. 14, pp. 233–297 (2005).
- [19] A. Mecozzi, C. B. Clausen, and M. Shtaif, “Analysis of intrachannel nonlinear effects in highly dispersed optical pulse transmission,” IEEE Photonics Technology Letters, vol. 12, pp. 392–394 (2000).
- [20] F. Forghieri, R.W. Tkack, A. R. Chraplyvy, “Fiber nonlinearities and their impact on transmission systems,” Chapter 8 in Optical Fiber Telecommunications IIIA, P. Kaminow and T. L. Koch eds., Academic Press (1997).
- [21] E. Ip, J.M. Kahn “Compensation of dispersion and nonlinear impairments using digital backpropagation,” Journal of Lightwave Technology, vol. 26, pp. 3416–3425 (2008).
- [22] R. I. Killey, P. M. Watts, V. Mikhailov, M. Glick, P. Bayvel “Electronic dispersion compensation by signal predistortion using digital processing and a dual-drive Mach-Zehnder modulator,” IEEE Photonics Technology Letters, vol. 17, pp. 714–716 (2005).
- [23] E. Ip, “Nonlinear compensation using backpropagation for polarization-multiplexed transmission,” Journal of Lightwave Technology, vol. 28, pp. 939–951 (2010).
- [24] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, “Properties of nonlinear noise in long, dispersion-uncompensated fiber links,” Optics Express, vol. 21, pp. 25685–25699 (2013).
- [25] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, “Accumulation of nonlinear interference noise in fiber-optic systems,” Optics Express, vol. 22, pp. 14199–14211 (2014).
- [26] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, “Inter-channel nonlinear interference noise in WDM systems: modeling and mitigation,” Journal of Lightwave Technology, vol. 33, pp. 1044–1053 (2015).
- [27] R. Dar, M. Shtaif, M. Feder, “New bounds on the capacity of the nonlinear fiber-optic channel,” Optics Letters, vol. 39, pp. 398–401 (2014).

- [28] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, “Time varying ISI model for nonlinear interference noise,” in Optical Fiber Communication (OFC) conference, pp. W2A.62, San Francisco, CA (2014).
- [29] R. Dar, O. Geller, M. Feder, A. Mecozzi, M. Shtaif, “Mitigation of inter-channel nonlinear interference in WDM systems,” in European Conference on Optical Communications (ECOC), P.5.6, Cannes (2014).
- [30] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, “Pulse collision picture of inter-channel nonlinear interference noise in fiber-optic communications,” accepted for publication in Journal of Lightwave Technology (April 2015).
- [31] P. Poggiolini, A. Carena, V. Curri, G. Bosco, and F. Forghieri, “Analytical modeling of nonlinear propagation in uncompensated optical transmission links,” IEEE Photonics Technology Letters, vol. 23, pp. 742–744, (2011).
- [32] P. Poggiolini, “The GN model of non-linear propagation in uncompensated coherent optical systems,” Journal of Lightwave Technology, vol. 30, pp. 3857–3879 (2012).
- [33] P. Poggiolini, G. Bosco, A. Carena, V. Curri, and F. Forghieri, “A detailed analytical derivation of the GN model of non-linear interference in coherent optical transmission systems,” arXiv preprint:1209.0394v12 (2012).
- [34] P. Johannisson and M. Karlsson, “Perturbation analysis of nonlinear propagation in a strongly dispersive optical communication system,” Journal of Lightwave Technology, vol. 31, pp. 1273–1282 (2013).
- [35] G. Bosco, R. Cigliutti, A. Nespola, A. Carena, V. Curri, F. Forghieri, Y. Yamamoto, T. Sasaki, Y. Jiang, P. Poggiolini, “Experimental investigation of nonlinear interference accumulation in uncompensated links,” IEEE Photonics Technology Letters, vol.24, pp. 1230–1232 (2012)
- [36] A. Bononi, N. Rossi, P. Serena, “Nonlinear threshold decrease with distance in 112 Gb/s PDM-QPSK coherent systems,” in European Conference on Optical Communications (ECOC), pp. We.2.C.4, Amsterdam (2012).
- [37] S.J. Savory, “Approximations for the nonlinear self-channel interference of channels with rectangular spectra,” IEEE Photon. Technol. Lett. **25**, 961 – 964 (2013)
- [38] A. Mecozzi and R.-J. Essiambre, “Nonlinear Shannon limit in pseudolinear coherent systems,” Journal of Lightwave Technology, vol. 30, pp. 2011–2024 (2012).
- [39] X. Liang, S. Kumar, “Analytical modeling of XPM in dispersion-managed coherent fiber-optic systems,” Optics Express, vol. 22, pp. 10579–10592 (2014).
- [40] L. Beygi, E. Agrell, P. Johannisson, M. Karlsson, H. Wymeersch, “A discrete-time model for uncompensated single-channel fiber-optical links,” IEEE Transactions on Communications, vol. 60, pp. 3440–3450 (2012).

- [41] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, F. Forghieri, “EGN model of non-linear fiber propagation,” *Optics Express*, vol. 22, pp. 16335–16362 (2014).
- [42] M. I. Yousefi, F. R. Kschischang, and G. Kramer, “Kolmogorov-Zakharov model for optical fiber communication,” *arXiv preprint:1411.6550* (2014).
- [43] G. D. Forney JR. and G. Ungerboeck, “Modulation and coding for linear Gaussian channels,” *IEEE Transactions on Information Theory*, vol. 44, pp. 2384–2415 (1998).
- [44] G. D. Forney JR. and L.-F. Wei., “Multidimensional constellations – part I: Introduction, figures of merit, and generalized cross constellations,” *IEEE Journal on Selected Areas in Communications*, vol. 7.6, pp. 877–892 (1989).
- [45] P. Poggiolini, Y. Jiang, A. Carena, G. Bosco, and F. Forghieri, “Analytical results on system maximum reach increase through symbol rate optimization,” in *Optical Fiber Communication Conference (OFC)*, pp. Th3D-6 (2015).
- [46] T. M. Cover and J. A. Thomas, *Elements of information theory*. Jhon Wiley and Sons (2012).
- [47] E. Meron, M. Feder, and M. Shtaiif, “On the achievable communication rates of generalized soliton transmission systems,” *arXiv preprint:1207.0297* (2011).
- [48] M. Yousefi and F. R. Kschischang, “Information transmission using the nonlinear Fourier transform, Part I: Mathematical tools,” *IEEE Transaction on Information Theory*, vol. 60, pp. 4312–4328 (2014).
- [49] S. T. Le, J. E. Prilepsky, and S. K. Turitsyn, “Nonlinear inverse synthesis for high spectral efficiency transmission in optical fibers,” *Optics Express*, vol. 22, pp. 26720–26741 (2014).