# Calculation of Bit Error Rates in Optical Fiber Communications Systems in the Presence of Nonlinear Distortion and Noise

by

Oleg V. Sinkin

Dissertation submitted to the Faculty of the Graduate School
of the University of Maryland in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2006

# ACKNOWLEDGEMENT

The period of my life in the United States was an entirely new world and the experience was very diverse: pleasing and exciting and very difficult at times. I want to thank my advisor, colleagues, friends, and family, who shared my happy and joyful moments with me and who stood by me in hard times. It was most important for me to feel your presence, support and compassion.

First of all, I would like to thank my advisor Curtis Menyuk. Working with him was always inspiring and fruitful as he is a bright scientist and a great promoter and educator. I learned the theory, practice and some politics of optical fiber communications. Many of my current skills and much of my current knowledge are due to him. In science, Curtis not only guided me, but also gave me complete freedom for creativity. I would like to note his exceptional human nature and I am very grateful to him for his understanding of my personal situation, his compassion, and willingness to help.

It was a great pleasure to work with Vladimir Grigoryan, a true scientist and a tremendous person. His enthusiasm and energy can inspire the worst pessimist. Vladimir helped me in tough moments as an excellent mentor and psychologist. In addition, Vladimir and his family have also been just good and hospitable friends.

It is truly great that some Italian universities have a tradition to send students abroad as a part of their PhD program! Thanks to this circumstance, I met Valter Pellegrini, Paola Griggio, and later Marco Secondini. I am grateful to these Italian *ragazzi* as they showed me some colors of life that were unknown to me before. Marco was my officemate for a while and we shared not only scientific ideas but many everyday and fun moments. We became friends that did not need to finish a phrase to understand each other. And I learned some of the essential Italian vocabulary. If I travel to Italy, now I know what to say if I meet a beautiful lady or if somebody steps on my foot.

I am very happy that in the beginning of my stay in Baltimore, I met Ronald Holzlöhner, Jochen Dörring, and Mireia Blanco. It was my great pleasure to work with Ron as well as to share many fun moments. Jochen and Mireia were more than friends to me as we were like a family.

A very interesting experience is being a part of an internationally and culturally diverse group, which has representatives from all continents on the Earth, except for Antarctica. Many thanks to all my colleagues for the privilege of working with you: Ivan Lima, Aurenice Lima, Hai Xu, Yu Sun, Brian Marks, Jonathan Hu, Jiping Wen, Hua Jiao, Vlad Seghete, Venkat Veerasubramanian, Kami Okusaga, Anthony Lenihan, Gaston Tudury, Bill Astar, Heider Ereifej, Edem Ibragimov, and to our group's professors Gary Carter, Vladimir Grigoryan, Li Yan, and John Zweck.

I would like to express my special gratitude to John Zweck for his support throughout the course of my Ph.D. study. He has been a creative and patient advisor, an extraordinary teacher, and a very responsive person. And, of course, many thanks for his enormous effort in organizing and keeping our computer code.

I am very grateful to Gary Carter for helping me with many personal and academic problems, for the opportunity to participate in the Bossnet experimental project, and for being a respectful and cheerful person.

I want to thank Tulay Adalı for her collaboration with us and for just being a talented teacher and a very inspiring person.

I would like to specially thank Ivan Lima and Hai Xu for helping me when I arrived and knew nothing and nobody in the US and for organizing a very warm and hospitable reception.

I am delighted to know Lyn Randers, who is not just a talented organizer but also a very kind and approachable person. Without Lyn the TRC would have been just the Commonwealth of Independent Labs.

I would like to thank Valery Butylkin, professor of Moscow Institute of Physics and Technology, who supported me when I was applying to UMBC and who taught me the subject of nonlinear optics — the background that was essential for my Ph.D. research.

It was my pleasure to collaborate with Mark Ablowitz, Andrew Docherty, and Cory Ahrens from University of Colorado, bright scientists as well as friendly and hospitable people.

I want to thank Silvia Petrova, Matt Zustiak, and Irina Ramos, for sharing skiing adventures and being friends upon whom I can rely.

My stay here would have been very hard if I did not have my Russian friends with me. Very special thanks to Vadim Jelezniakov, who I know from my freshman year at Moscow Institute of Physics and Technology, known as PhysTech. Even when separated by distance, we are always close in spirit. I am lucky to know Filipp Kistenev and Alexey Finogenov, who also represent the PhysTech community. Thank you for your enormous support and for fun time skiing, windsurfing, partying, conquering the salsa dance floor, and… The list of what we did not do together is actually shorter.

I was lucky to have two greatest roommates Andrey Parafiynik and Alexander Sinyukov who made my everyday life more interesting and with whom I shared parties and many “life talks”. Many thanks to Valery Soika and Irina Krasnova for their hospitality, fun time we had, and trips we went to.

I would like to express my enormous gratitude to Maria Kasatkina, without whom I could not imagine anything in my life happening, for her devotion, patience, and great support.

It is impossible to express how grateful I am to my parents, Valery Vasilievich Sinkin and Olga Nikolaevna Sinkina, who not only gave me the opportunity to get the best physics education in Russia and supported me while I was away on another continent, but also were friends to me — companions with whom I could share any concern. And, of course, I am infinitely happy that I have the best brother in the world, Yura — exceptionally dedicated, supportive, and fun — who always finds the right words in any situation. To my family I dedicate this dissertation.

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# 1. INTRODUCTION

A fundamental goal of modeling fiber communications systems is to understand the physics of the system behavior and to develop computational tools to design systems and predict their performance. Transmission of data through a fiber-optic link unavoidably leads to bit errors due to various effects, the dominant of which are noise from optical amplifiers, fiber nonlinearity, polarization effects, and non-ideal transmitters and receivers. There exist numerous studies that provide techniques to characterize all of these effects and to calculate the bit error ratio (BER) due to them [1]–[4]. However, there are still many important unanswered questions and one of them is how to accurately calculate the bit error ratio (BER) in the presence of nonlinear signal distortion.

Why is a careful analysis of nonlinear effects in optical fiber communications systems important? Nearly all modern systems operate in the linear propagation regime, in which the signal evolution is almost linear [5]. However, there always exist small nonlinear interactions and small nonlinear signal distortions accumulated during transmission over long distances can lead to an increase in the error rate. Reducing the optical power decreases the importance of the nonlinear interactions, but it also decreases the signal-to-noise ratio. There exists an optimal power level at which the BER is minimal. Even if the power level is much lower than optimum, the accumulation of nonlinear distortions during transmission over hundreds or thousands of kilometers of fiber can introduce a significant system penalty [6]–[9]. Calculating the BER in such a regime or finding the optimal power level requires an accurate model of the nonlinear interactions.

The main challenge in characterizing the nonlinear penalty is that it is a statistical quantity. In on-off keyed systems, a digital 1 (mark) is represented by the presence of an optical pulse and a digital 0 (space) is represented by its absence. The amount of distortion that an optical pulse suffers depends on the particular pattern of surrounding pulses, which is effectively random as these pulses represent the information bits, and the information sequence of bits is quasi-random. This effect is often referred to as the nonlinear pattern dependence effect [1], [10]. In single-channel transmission, dispersion leads to the spread of optical signals, causing approximately three to seven adjacent pulses to interfere [5], [11]. Therefore, a common approach to account for pattern-dependent nonlinear effects is to use a pseudo-random sequence of bits, which is typically $2^3 - 2^7$ bits, to find the worst-case bit in the sequence. When we consider a multi-channel system, this approach is inappropriate since there are many more pulses interacting with each other due to the dispersive walkoff between the frequency channels. As an illustration, Fig. 1.1 shows three simulated eye diagrams of a noise-free signal in the center channel of a 10 Gb/s wavelength-division multiplexed (WDM) return-to-zero (RZ) system after propagating over 5000 km. We used nine co-polarized channels spaced by 50 GHz and the average power was approximately $-0.7$ dBm per channel. We used three different sets of bit patterns in different WDM channels, while the bit pattern in the center channel remained unchanged. As we move from 1.1a to 1.1c, it is apparent that the eye changes from being almost completely open to completely closed. In this case, finding the worst-case performance becomes not only prohibitively time-consuming since the number of possible interaction patterns grows exponentially, but it is also not useful because the likelihood of the worst-case pattern is negligibly small. Therefore, a probabilistic approach is necessary to treat this problem.

Typically, the dominant nonlinear effect in modern high-speed systems operating

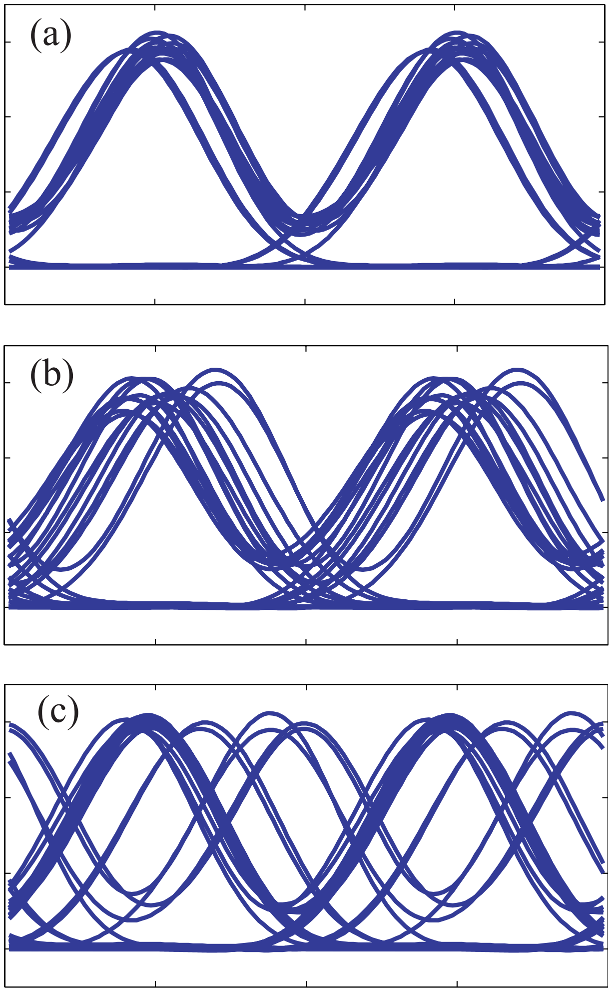

**Fig. 1.1:** Nonlinear pattern dependence effect in a WDM RZ system. The bit pattern in the center channel is fixed. As we change the patterns in the neighboring channels, the eye in the center channel goes from (a) almost completely open to (b) only partially open to (c) almost completely closed.

at 10 GB/s, is cross-phase modulation [7], [8], [12]–[15]. The phase of an optical pulse is changed by the presence of pulses in either the same or neighboring wavelength channels. This phase change leads to intensity distortion by means of fiber dispersion. The manifestation of this effect depends on the light modulation format. In non-return-to-zero (NRZ) transmission, the signal distortion appears in the form of amplitude jitter [14], [16]–[18], while in the RZ systems, the dominant nonlinear effect is typically collision-induced timing-jitter [12], [19]–[23]. This fact requires the development of completely different approaches to account for the nonlinear effects in these two types of systems. We note that at present, the NRZ and RZ modulation formats are still the most commonly used formats in optical fiber communications systems. The NRZ format is the simplest form of intensity modulation and it has been historically the format of choice for many system providers. Recently, it has been discovered that the RZ-modulated signal undergoes less intersymbol interference in the receiver and is more robust to fiber nonlinearities and thus more capable of long-haul data transmission [11], [22], [24]–[27]. Because the NRZ modulation has been used for many years, techniques have been developed to characterize the nonlinear effects in WDM NRZ transmission [14], [16]–[18], [28]–[36], and the BER calculations based on these techniques agree well with experimental results. The basic idea in these approaches [16], [30]–[32] is to utilize a pump-probe method, in which the cross-phase modulation-induced distortion is treated as an additive perturbation. An exception is [36] where the authors treat the distortion as multiplicative. In order to determine the influence of the nonlinearity on the system performance, one further assumes that the XPM-induced distortion may be treated as additive Gaussian noise and a correction to the $Q$-factor is calculated [14], [16], [30], [33]–[35], [37].

The major nonlinear effect in WDM RZ systems, collision-induced timing jitter, has also been well studied in both soliton and linear systems [12], [19]–[21], [38]–[44]. It is well known how to calculate the time shift that results from a collision of a pair of pulses and to calculate the standard deviation of the time shift. However, no accurate BER calculation that takes into account the inter-channel nonlinear bit-pattern effect due to this timing jitter has been reported in the literature.

The purpose of this dissertation research is to develop a method that allows one to accurately account for inter-channel nonlinear crosstalk in calculations of BER in WDM RZ systems.

Our method of computing the BER in the presence of the nonlinear distortion and amplified spontaneous emission noise, is based on calculations of the complete probability density function (pdf) of the nonlinearly-induced amplitude or timing jitter. Using the knowledge of the pdf of the noise-induced amplitude variation, we combine the noise and nonlinear contributions to calculate the resulting BER.

The dissertation is organized as follows:

In the second chapter, we describe a prototypical undersea system to which we apply our BER calculation technique. We discuss some system design tradeoffs and the performance optimization issues.

In the third chapter, we review methods commonly used to characterize the effects of fiber nonlinearity in optical fiber communications systems.

The fourth chapter contains a summary and discussion of the biased Monte Carlo methods that can be used to estimate the pdf of the received current or the time shift of a pulse for the values of the pdf ranging over many orders of magnitude.

In the fifth chapter, we introduce the time shift function, a function that describes the time shift of a pulse in a two-pulse collision, depending on the frequency and initial time separation of the two pulses. We discuss the properties of the time shift function and use it to calculate the pdf of the collision-induced time shift by means of the characteristic function method. Finally, we validate this method of computing the pdf of the time shift with biased Monte Carlo simulations.

In the sixth chapter, we present and validate a method for evaluating the pdf of the received current due to nonlinear effects in transmission. Then, we describe an additive white Gaussian noise model for calculating the pdf of the received current and show how to calculate the BER using the methods that we presented.

The seventh chapter contains the conclusions that summarize the main results of the dissertation.

# 2. TEST SYSTEM

In this chapter, we present a prototypical long-haul undersea system that uses WDM technology and an RZ modulation format. Long-haul submarine transmission systems represent a significant and rapidly growing portion of the world fiber optic network [45] and a large portion of current transoceanic transmission lines operate using the RZ modulation format, including cable systems built by Tyco Telecommunications, Marconi, NEC Submarine Systems, and Fujitsu. These systems typically have shorter values of amplifier spacing and longer transmission distances than terrestrial systems. The propagation length in these systems is limited by the tradeoff between the signal-to-noise ratio and the accumulation of the nonlinear penalties. Hence it is especially important to develop accurate tools to model nonlinear impairments for these types of systems.

The original goal of this dissertation was to perform a comparative study of spectral efficiency of different modulation formats. In order to do this, we had to optimize system parameters for each format and different values of channel spacing. In particular, for a given channel spacing we varied the average map dispersion and the pre- and post-compensation to determine the optimal dispersion profile for each modulation format. During the initial study, we encountered the nonlinear pattern-dependence effect [1], [10]. Exploration of this matter led us to a more general, and, in our view, more important research topic. The current goal of this dissertation is to develop new and accurate tools to evaluate the effect of fiber nonlinearity on the performance of RZ systems. In this chapter, we describe only the RZ system that we used to develop our methodology, and we summarize the optimization of the dispersion profile for this system.

## 2.1 System model

Typical modern fiber communications systems contain a large number of optical components such as lasers, modulators and demodulators, multiplexers and demultiplexers, filters, optical fibers, and amplifiers [1], [2]. In addition, there is an extensive set of electrical equipment at the receiver, such as photodiodes, electrical filters, amplifiers, and decision circuits. When simulating these components, the level of detail of the model must be appropriate for the system under study. For example, a lumped amplifier model is used often for optical amplifiers, in which the optical field is simply multiplied by a factor. We use this model in the simulations in this dissertation. More realistic models can include amplified spontaneous emission noise, gain saturation, the gain profile, polarization hole burning, and transients [46]. The first step of any simulation is to simplify the system and to restrict the model of the optical propagation to the essential effects. The nature of the most important effects strongly depends on the type of the optical system. In particular, the amount of nonlinearity has an important impact on the evolution of the signal and the noise.

Figure 2.1 shows a schematic illustration of a simple model of an optical fiber communications system that is used in this dissertation. The optical signal is generated by a transmitter and inserted into the fiber. It then passes through a transmission line that primarily consists of fiber spans and optical amplifiers. At the end of the transmission line, the signal is optically filtered and enters a receiver, where it is converted to an electrical current by a high-speed photodiode. This current passes

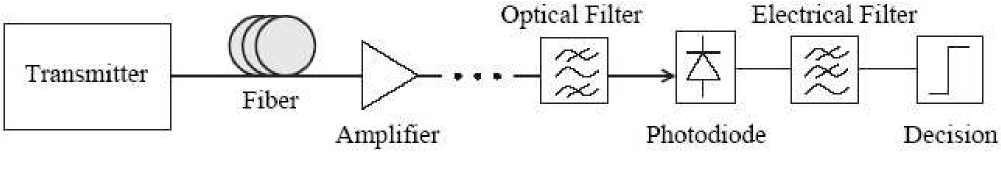

**Fig. 2.1:** Simple communications system. Reproduced from [149].

through a low-pass filter and enters a decision circuit.

In a typical digital optical fiber transmission system, the pulses are either directly created by an optical laser, or the output of a constant wave laser is modulated by an external modulator. We will only discuss on-off keying in this dissertation. In RZ transmission, the marks are represented by isolated optical pulses, while in NRZ signaling, a sequence of marks is represented by a continuous constant light intensity. We assume that the amplitude modulation is produced by a Mach-Zehnder interferometer. The functional form of the RZ or chirped-RZ (CRZ) complex field envelope that is used in this dissertation is

$$U(t) = \sqrt{\frac{1}{2}\left[1 + \cos\left(\pi \sin \frac{\pi t}{T}\right)\right]} \exp\left(iC\pi \cos \frac{2\pi t}{T}\right), \tag{2.1}$$

where $T$ is the bit period and $C$ is the chirp parameter [11], [47]. The parameter $C$ equals zero for the unchirped RZ signal. The pulse stream may then be combined with pulse streams with different central frequencies to make a single WDM signal [1], [2]. In this work, we used nine co-polarized WDM channels, spaced by 50 GHz, each carrying a 10-GB/s unchirped RZ signal. We used different values of peak power in the system optimization step; however, we set the peak power to 5 mW per channel in the rest of the study.

Transmission of light through optical fiber can be described by the nonlinear Schrödinger (NLS) equation [48]

$$\frac{\partial u(z,t)}{\partial z} + i\frac{\beta''(z)}{2}\frac{\partial^2 u(z,t)}{\partial t^2} - i\gamma|u(z,t)|^2 u(z,t) = g(z)u(z,t), \tag{2.2}$$

where $u(z,t)$ is the electric field envelope, $z$ is the physical distance, $t$ is the retarded time with respect to the central frequency of the signal, $\beta''$ is the local dispersion, $\gamma$ is the Kerr coefficient, and $g(z)$ is the fiber loss and gain coefficient. This form of the NLS equation is based on the negative carrier frequency convention in which the

Fourier transform pairs are defined as

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{X}(\omega)\exp(-i\omega t)\,d\omega, \qquad \tilde{X}(\omega) = \int_{-\infty}^{\infty} x(t)\exp(i\omega t)\,dt. \tag{2.3}$$

This convention is common in the physics literature, but electrical engineers and mathematicians typically use the opposite convention. The literature on optical communications is mixed and it is not uncommon to find work in which authors switch from one convention to the other without making any note of it. While this issue is usually not important, it can sometimes lead to errors. The implications of the different carrier conventions are thoroughly discussed in [49], [50]. We note as well that (2.2) does not include polarization effects. In practice, polarization effects are important because it is common to use orthogonal polarization between neighboring channels or even neighboring bits to reduce the effects of nonlinearity [8], [11], [51]. We do not take these effects into account here because they have no impact on the techniques that we present and would complicate the discussion.

The receiver subsystem includes an optical demultiplexer, square-law photodetector, and a low-pass electrical filter as shown in Fig. 2.1. The photodetector is modeled as an ideal square-law detector without noise. We choose the spectral transmission function of the optical demultiplexer to be third-order super-Gaussian, where a super-Gaussian function of $m$-th order is defined as $\exp(-x^{2m})$. The bandwidth of the optical filter was optimized to maximize the eye opening after demultiplexing a WDM signal back-to-back and it is found to be 35 GHz for the channel spacing of 50 GHz and the unchirped RZ signal. The electrical filter is a fifth-order Bessel filter, which is a typical filter in modern optical fiber communications systems, and its 3-dB bandwidth is set to 8 GHz.

Finally, the decision and clock recovery circuit in this work is modeled by simply calculating the central time of a pulse at the receiver in a given channel. Since the time reference frame is moving with the group velocity of the center channel, the pulses in the center channel do not move in this time frame due to dispersion. For pulses in other frequency channels, one can analytically account for the group delay corresponding to the particular channel. We chose this model of clock recovery because it is simple and insensitive to nonlinear distortions, which enables us to obtain a reliable estimate of the system performance. A more realistic model of the clock recovery is based on calculating the phase of one of the strong frequency components of the signal at the receiver [52], [53].

In this chapter, we discuss some aspects of system design and optimization of system parameters. For the purpose of parameter optimization, we take the eye opening at the clock recovery time as a measure of the system performance, which is defined as the difference between the average current of a mark and the average current of a space at the detection time in the receiver.

Calculations of the eye opening are complicated by the effect of pattern dependence during transmission due to nonlinear interactions. In a single channel system, each bit only interacts with its neighbors. All bit patterns of length $n$ are contained in a *de Bruijn* pseudo-random bit sequence (PRBS) of length $N = 2^n$ [54]. If we increase the length of the PRBS, the rails of the eye diagram converge. In other words, there is a certain number of surrounding bits, $n$, that affect the center bit, which is determined by the amount of pulse stretching due to dispersion during transmission. Hence if we consider a PRBS of length $N = 2^n$ or larger, we will include all possible interaction patterns that may occur in any data stream in a single-channel system.

The situation becomes much more complicated when we add WDM channels to the system. In addition to interacting with a limited number of neighboring bits in the same channel, a single bit will also interact with many bits in the neighboring channels. For a fixed bit string in the center channel, the resulting eye diagram, will depend on both the bit strings in the adjacent channels and the relative positions of these bit strings. The consequence of this effect was illustrated in Fig. 1.1. An exact treatment of this problem requires a probabilistic approach, in which all possible bit patterns in the neighboring channels are considered. This approach will be discussed in the following chapters. For the purpose of finding a set of system parameters that optimize the performance, we run a large number of simulations, in which we keep the bit string in the center channel fixed and we randomly vary the bit strings in the side channels. We compute the average eye opening from these simulations. While it is not possible to accurately infer a BER from this procedure, it is possible to reliably infer the system parameters that will produce the best performance.

## 2.2 Dispersion map design

All recently deployed systems employ dispersion management, in which the dispersion varies periodically. Each period consists of a concatenation of several fiber spans with different local fiber dispersions, and the variation of dispersion in one period is referred to as the dispersion map. The dispersion map is characterized by the average map dispersion $D_{\mathrm{map}}$. For a map consisting of two fiber links, $D_{\mathrm{map}}$ given by

$$D_{\mathrm{map}} = \frac{D_1 L_1 + D_2 L_2}{L_1 + L_2}, \tag{2.4}$$

where $D_1$ and $D_2$ are dispersion of the first and the second sections of the map and $L_1$ and $L_2$ are the corresponding fiber lengths. The dispersion parameter $D$ is in the units of ps/nm–km, which is commonly used in optical communications. It is related to the second-order dispersion $\beta''$ in the units of ps/km$^2$ used in (2.2) by the relation

$$D = -\frac{2\pi c}{\lambda^2}\beta''. \tag{2.5}$$

If $D_{\mathrm{map}} \neq 0$ then the residual accumulated dispersion is compensated at the terminals by means of extra links of fiber known as pre- and post-compensation fibers [11], [47], [55]–[59]. It is typically beneficial to operate the system at non-zero average dispersion. Using dispersion pre-compensation results in spreading out the optical pulses initially, which reduces the effects of four-wave mixing and cross-phase modulation [11], [22], [25]. However, excessive spreading leads to intra-channel four-wave mixing [60]. This tradeoff determines the optimal value of $D_{\mathrm{map}}$, which we will discuss further in this section. The amount of both the pre- and post-compensation dispersion must be carefully chosen as an improper choice may result in substantial intra-channel cross-phase modulation [47], [58], [61], [62].

### 2.2.1 Dispersion slope-matched fiber pair

An important milestone in the development of optical fiber communications systems was the introduction of dispersion-sloped matched fibers [63], [64], which allowed an increase in the reach of the systems and a decrease in the cost of the terminal equipment. Dispersion slope compensation has become a key technology for both 10 and 40 Gb/s WDM systems [51], [59], [65]–[67]. In this work, we study a system based on dispersion slope matched fibers [63], [64], often called $D_+$ and $D_-$ fibers, whose characteristics are shown in Table 2.1. The fiber transmission link is based on a dispersion map of length $L$, in which the first two-thirds consist of $D_+$ fiber and the remaining one-third consists of $D_-$ fiber. The exact proportion is determined by the desired map average dispersion, $D_{\text{map}} = (D_+L_+ + D_-L_-)/L$, where $L_+$ and $L_-$ are the lengths of the $D_+$ and $D_-$ fibers respectively.

| Parameter | $D_+$ fiber | $D_-$ fiber | SMF | DCF |
|---|---|---|---|---|
| Dispersion (ps/nm–km) at $\lambda = 1500$ nm | 20.17 | –40.8 | 17.0 | –95.0 |
| Dispersion slope (ps/nm$^2$–km) | 0.062 | –0.124 | 0.075 | –0.2 |
| Attenuation (dB/km) | 0.192 | 0.251 | 0.21 | 0.5 |
| Effective area ($\mu$m$^2$) | 106.7 | 31.1 | 47.0 | 20.0 |
| $n_2$ ($\times 10^{-20}$ W$^{-1}$) | 1.7 | 2.2 | 2.6 | 2.6 |

Tab. 2.1: Characteristics of $D_+/D_-$ and SMF/DCF fiber pairs.

The main advantage of the $D_+/D_-$ pair is that the average dispersion slope is much smaller than when an SMF/DCF pair is used. We demonstrate this feature in Fig. 2.2. The horizontal axis represents the average map dispersion $D_{\text{map}}(\lambda_c)$ of the center channel located at $\lambda_c$ =1550 nm. The vertical axis shows the difference between the average map dispersions for the edge channel $\lambda_e$ and the center channel:

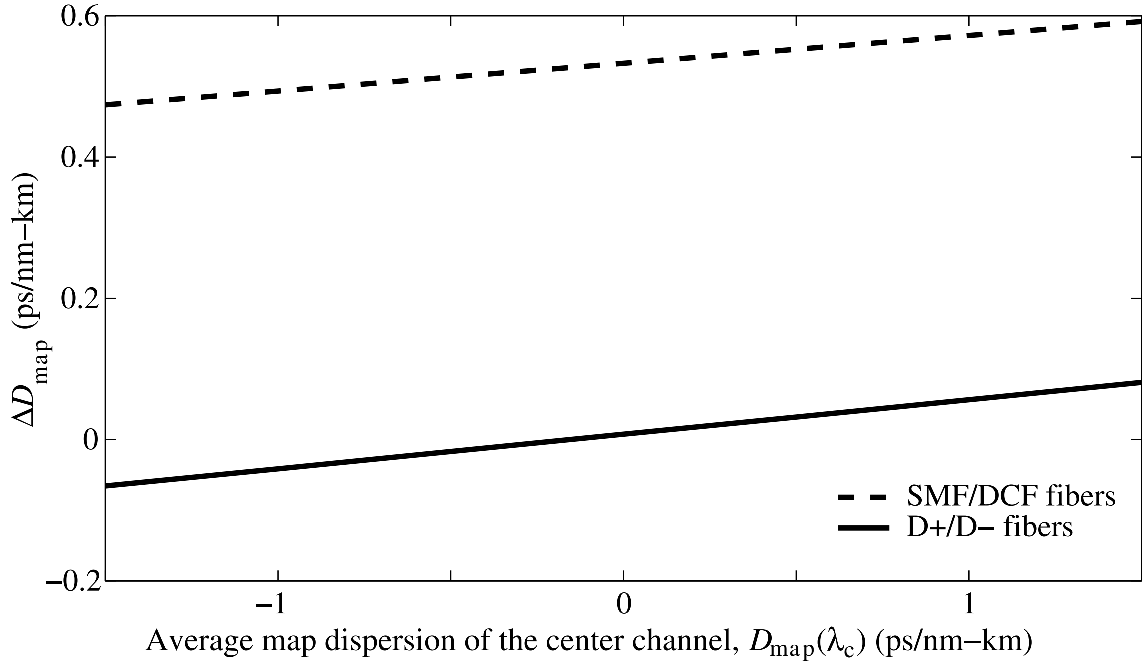

**Fig. 2.2:** Dispersion slope characteristics of two fiber pairs

$\Delta D_{\mathrm{map}} = |D_{\mathrm{map}}(\lambda_e) - D_{\mathrm{map}}(\lambda_c)|$. In this case, $2|\lambda_c - \lambda_e| = 32$ nm, which correspond to a WDM signal with 40 channels spaced by 100 GHz. In the case of $D_+/D_-$ fibers, the variation in the average map dispersion of the center channel is very small and does not exceed a few percent. Consequently, all channels experience almost the same average map dispersion. By contrast, the SMF/DCF pair exhibits a strong wavelength dependence of the average map dispersion. For all values of $D_{\mathrm{map}}(\lambda_c)$, the side channels will have an average map dispersion $D_{\mathrm{map}}(\lambda_e)$ that differs from that of the center channel by 0.4–0.5 ps/nm–km. For this reason it is useless to optimize the map's average dispersion for the SMF/DCF fiber pair since different wavelengths will have completely different average dispersions. By contrast, using $D_+/D_-$ fiber pair one may optimize the average map dispersion for WDM transmission. Other significant advantages of $D_+/D_-$ fibers include reduced nonlinearity due to an increased effective core area and reduced polarization mode dispersion.

### 2.2.2 Optimization of dispersion profile

To optimize the dispersion profile, we considered a noiseless transmission model, based on the NLS equation (2.2), which allows us to separate the effects of nonlinearity and dispersion from the effects of the signal-noise interaction. The measure of performance was the eye opening of the optical signal in the center channel at the end of transmission.

We consider a system with a propagation distance of 5000 km and an amplifier spacing of 50 km. The dispersion map consists of 34 km of $D_+$ fiber and approximately 16 km of $D_-$ fiber followed by an amplifier. This layout is typical for modern undersea systems [11], [51]. Since the effective nonlinearity of the $D_-$ fiber is larger than that of the $D_+$ fiber, the amplifier is placed after the $D_-$ fiber. Thus the signal power at the input of the $D_-$ fiber is low, which reduces the nonlinear impairments in the system.

#### Optimization of dispersion pre- and post-compensation

The nonlinear interactions during transmission result in two major types of signal distortion: amplitude and timing jitter. These effects originate both from intra-channel and inter-channel interactions [10]. With properly selected pre- and post-compensation one can significantly improve the transmission quality [11], [24], [57], [58], [61]. The physical principle behind this improvement is as follows: Phase modulation induced by the nonlinear pulse interactions is converted into amplitude modulation by dispersion; this amplitude modulation then causes waveform distortion. By adjusting the amount of dispersion pre-compensation and total dispersion of the transmission line this amplitude modulation can be partially reversed [37]. It has been shown that the best dispersion-compensation scheme is nearly symmetric [11], [58], [61]. Timing jitter reduction can be explained as follows: If we consider a simplified situation, in which only two pulses interact with each other [58], one can show that the nonlinear interactions induce opposite frequency shifts for the two pulses. Due to the dispersion, these frequency shifts translate into time shifts. The magnitude of the time shifts is proportional to the dispersion and the magnitude of the frequency shift. If the accumulated dispersion is a symmetric function of distance, it changes its sign in the middle of the transmission. Hence the time shifts change their direction, and the resulting timing jitter is thus canceled at the end of the transmission line. However, in realistic systems the optimum accumulated dispersion function is not perfectly symmetric and nonzero residual dispersion is often preferable, due to self-phase modulation and to the non-constant power profile.

In this work, we optimize the compensation scheme by numerically finding the one that results in the maximum eye opening. In order to do so, we use single-channel transmission, so that we are only minimizing the intra-channel nonlinear effects. We fix all the system parameters except for the amount of pre- and post-compensation, and we automatically adjust this amount to achieve the maximum optical eye opening. Figure 2.3 illustrates the improvement that is achieved by optimizing the dispersion

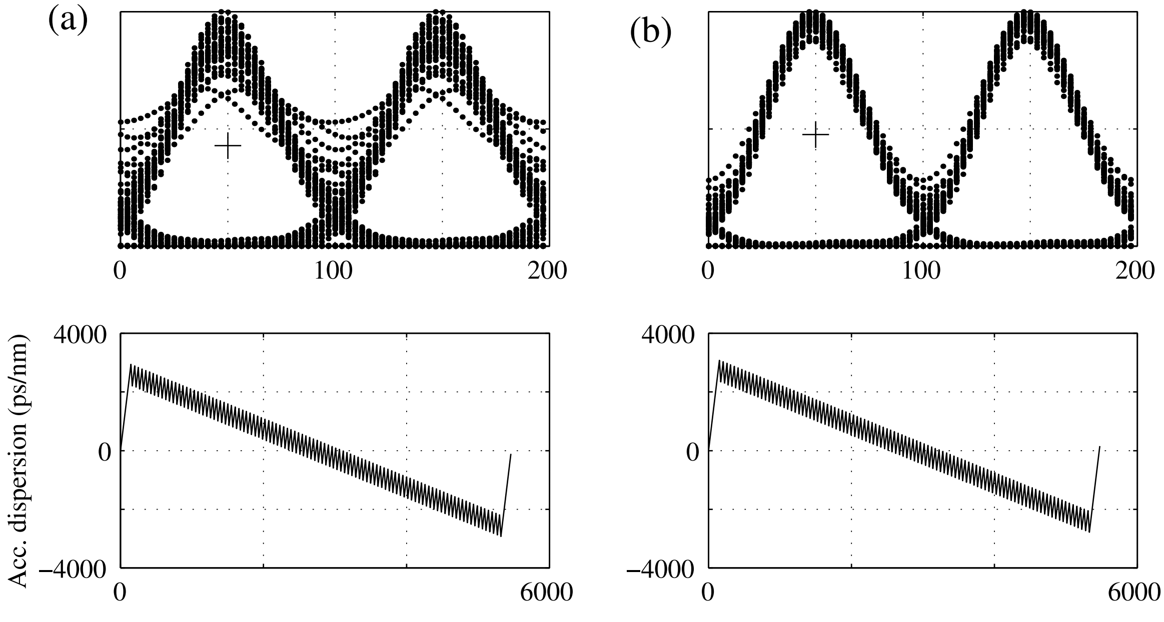

**Fig. 2.3:** RZ eye diagrams with (a) suboptimal and (b) optimal pre- and post-compensation with the corresponding accumulated dispersion functions.

compensation scheme for our test system. We have consistently found in numerical simulations that the accumulated dispersion is a nearly symmetric function of distance and that a positive amount of residual dispersion is required to compensate for self-phase-modulation-induced distortions. We see from Fig. 2.3 that undercompensated dispersion results in an eye closure penalty. These results are consistent with previous studies [11], [24], [57], [58], [61].

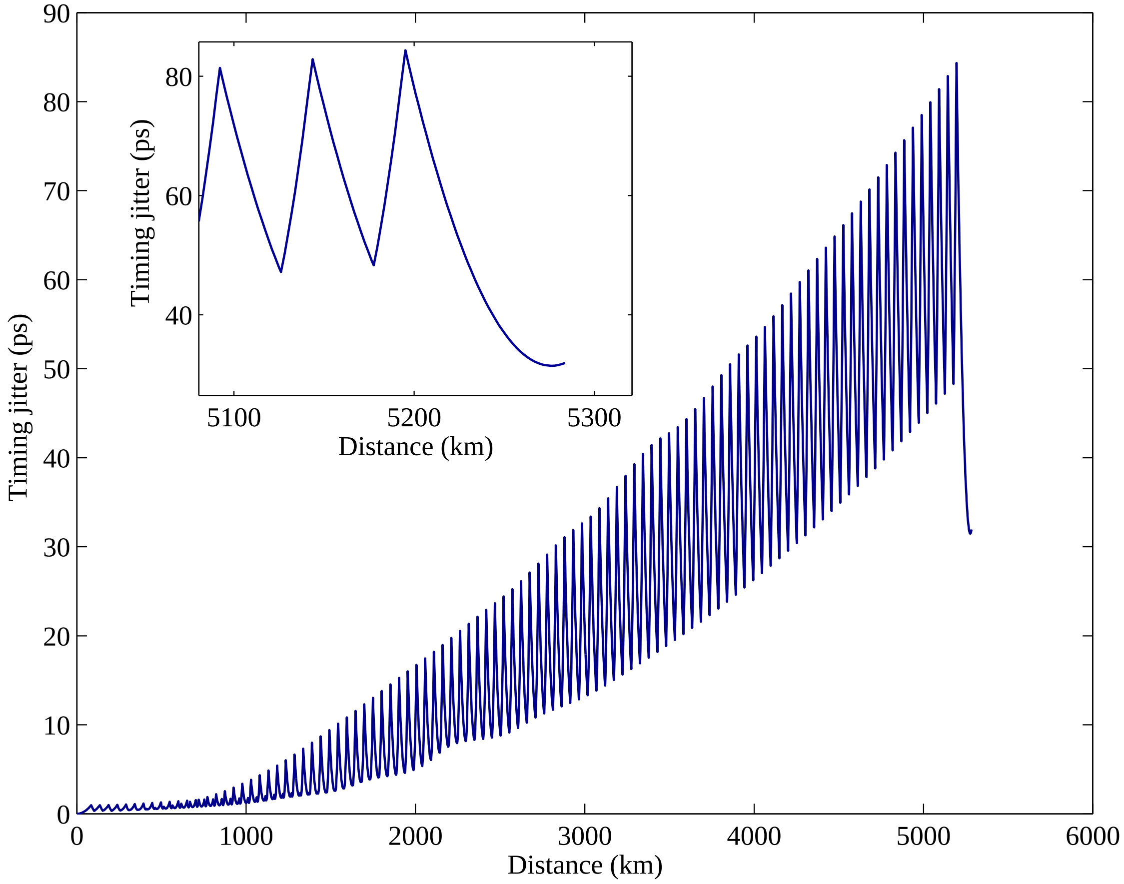

**Fig. 2.4:** Collision-induced timing jitter as a function of distance.

We note that we performed optimization of the pre- and post-compensation only for a single channel system. However, our recent results indicate that when we add WDM channels to this system, the optimum compensation we found for the single-channel case also results in minimizing inter-channel collision-induced timing jitter. Figure 2.4 shows the timing jitter, which is the standard deviation of collision-induced time shift as a function of transmission distance. The minimum of the timing jitter occurs right at the end of the transmission, as shown in the inset, indicating that the previous choice of the pre-and post-compensation is optimum in the WDM case as well. This effect is due to the symmetry of the dispersion function, since time shifts induced in the first and the second half of the transmission tend to cancel out.

### Optimization of average map dispersion

The purpose of this part of the study was to find the optimal average map dispersion for a given value of channel spacing. The measure of performance is the optical eye opening in the center channel, which we compute as a function of the average map dispersion for different power levels and channel spacings. Note that for each simulation, we optimized the dispersion pre- and post-compensation for the center channel.

Figure 2.5 summarizes the results of the average map dispersion study. The different curves in each subplot correspond to different power levels of the signal and different subplots represent the results for the channel spacing values of 25, 50, and 100 GHz. All curves have a minimum around $D_{\text{map}} = 0$, but, for the lowest input power level (diamonds), the minimum is less deep because the nonlinear interactions become weaker with a lower signal power. We also see that the range of the optimal map dispersion is large spanning from –1 to –0.3 ps/nm/km and variation in eye opening in this range are small. Existence of the optimum range of map dispersion values is explained by the tradeoff between the intra- and inter-channel nonlinear interactions. For large values of dispersion, the adjacent channels slide through each other faster, thus reducing inter-channel crosstalk. However, large values of dispersion cause a larger number of bits within one channel to overlap, resulting in an increase in the intra-channel distortion. When the total map dispersion is close to zero, intra-channel effects are weak, and the inter-channel crosstalk dominates the eye closure [10]. Note that some of the curves are not smooth, which is attributed to inter-channel pattern dependence effects.

For all channel spacings, the results are qualitatively the same: If the average map dispersion is close to zero, the system performance is substantially worse for all three channel spacings, but this effect is more noticeable for systems with smaller channel spacing, since the inter-channel interactions increase when the channel spacing is decreased [10]. Moreover, as the channel spacing decreases, the absolute values of optimum map dispersion increase.

The dispersion map parameters and other system data that we used in this work are summarized in Table 2.2.

| | |
|---|---|
| Pre-compensation | 51 km of $D_+$ fiber |
| Post-compensation | 90 km of $D_-$ fiber |
| Length of $D_+$ fiber in one map | 34 km |
| Length of $D_-$ fiber in one map | 17.44 km |
| Number of dispersion map periods | 100 |
| Input peak power | 5 mW |
| Bit rate | 10 Gb/s |
| Pulse width (FWHM) | 35 ps |
| Channel spacing | 50 GHz |
| Number of channels | 9 |
| Optical demultiplexor | 3-rd order super Gaussian, FWHM = 35 GHz |
| Electrical filter | 5-th order Bessel, HWHM = 8 GHz |

Tab. 2.2: System parameters.

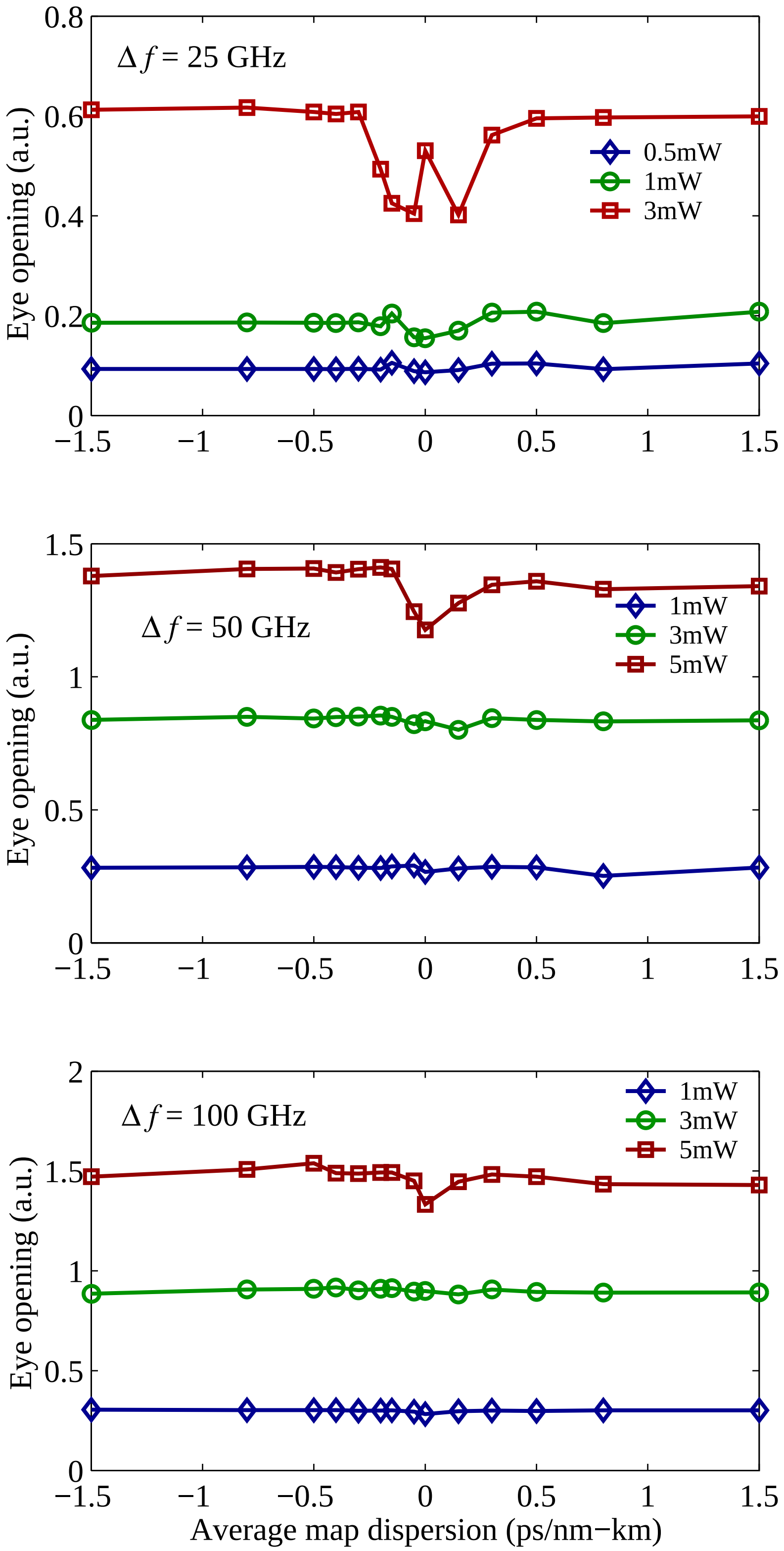

**Fig. 2.5:** Eye opening as a function of average map dispersion.

# 3. MODELING THE EFFECTS OF FIBER NONLINEARITY IN OPTICAL COMMUNICATIONS

In this chapter we discuss commonly-used techniques to model the effect of fiber nonlinearity on the light propagation.

The nonlinearities in silica fibers can be classified into two categories: stimulated scattering and the Kerr effect that manifests itself in a nonlinear index of refraction [48], [68]–[72]. Raman and Brillouin scattering causes a power-dependent gain or loss, and the Kerr effect causes an intensity-dependent phase change, which coupled with the dispersion leads to amplitude distortion of the signal. In Brillouin scattering, an optical wave interacts with a sound wave (acoustic phonons) in the medium and can produce a Stokes wave downshifted from the pump wave in frequency. The typical Stokes shift in Brillouin scattering is on the order of 10 GHz and therefore it does not cause interchannel crosstalk. The Brillouin gain linewidth is on the order of 20 MHz and hence it typically has a negligible effect on modulated signals. Stimulated Raman scattering is due to the interaction of light and the molecular vibrations of the medium (optical phonons). In amorphous materials such as silica, molecular vibrational frequencies spread into bands that overlap and create a continuum [73]. Therefore the Raman gain in silica extends over a frequency range of more than 40 THz, which is large compared to Brillouin gain bandwidth. The peak of the Raman gain occurs near 13 THz. The wideband Raman gain leads to a noticeable crosstalk between WDM channels that are separated in wavelength by as much as 10 nm.

The intensity-dependent refractive index gives rise to three effects: self-phase modulation (SPM), cross-phase modulation (XPM), and four-wave mixing (FWM) [10], [74]. Self-phase modulation refers to the phase modulation of the signal within one frequency channel due to the intensity changes in this channel, while XPM refers to the phase changes induced by the intensity variations in different frequency channels of a WDM system. These effects lead to spectral broadening of the signals, and phase modulation is converted into intensity fluctuations by the fiber dispersion. In FWM, the beating between the optical waves of different frequencies leads to energy exchange between them and to the generation of light at new frequencies. When three waves of frequencies $f_i, f_j$, and $f_k$ interact, they generate a wave at a frequency $f_{ijk} = f_i + f_j - f_k$. Thus, three co-propagating waves generate nine new optical waves [74]. If the frequency channels are evenly spaced, FWM leads to energy exchange between these channels causing crosstalk. Significant FWM occurs only if the relative phase of the mixing waves nearly vanishes. The efficiency of FWM is roughly inversely proportional to the square of fiber dispersion. In modern fibers the efficiency of FWM is greatly reduced by the use of a large local dispersion, so that the impact of FWM on the system is much smaller than the effects of SPM and XPM.

In this work we focus our attention on the effects of SPM and XPM.

## 3.1 Full system model

### 3.1.1 Split-step Fourier method: Theory

By a full system model, we mean a model of light transmission through an optical fiber based on the NLS equation (2.2) and using the exact input pulse shape (2.1). A full system model involves solving (2.2) numerically and then calculating the waveform distortion due to the fiber nonlinearity. One may use Monte Carlo approach in which one repeatedly solves (2.2) while changing the input bit sequences in all the WDM channels at random for each new run. One can then build a histogram of the received current to estimate its pdf as, for example, in [75]. A more common approach is to solve (2.2) once for a randomly chosen set of bit strings to calculate the standard deviation of the sampled current [7], [18]. The length of the bit sequences should be long enough for this standard deviation to converge, as it was shown in [18].

Equations of type (2.2) may be solved numerically using either the finite difference or split-step Fourier methods [76], [77]. The split-step Fourier method is convenient for its simplicity and flexibility in dealing with higher-order dispersion, the Raman effect and filtering and therefore it is the most widely used approach [7], [10], [18], [48], [78]–[81].

The efficiency of the split-step method depends on both the time (or frequency) domain resolution and the distribution of step sizes along the fiber. In simulations of optical fiber transmission systems, the time and frequency resolutions are respectively determined by the bandwidth of the signal and the number of bits that are to be propagated through the system. Consequently, the properties of the signal determine the minimum required number of Fourier modes. Although the number of Fourier modes affects the accuracy of the numerical solution, as was shown in [80], it does not change the qualitative behavior of the spatial step size selection algorithm. Therefore, we only discuss the accuracy and efficiency of different spatial step size selection criteria.

A variety of step size selection criteria, most based on physical intuition, have been proposed for optimizing the split-step method. The figure of merit for each criterion is the computational cost for a given resulting global accuracy. Historically, in numerical methods used to solve (2.2) the step-size distribution was optimized for simulating soliton propagation. However, this optimization is not necessarily appropriate for modeling modern transmission systems, which often feature both high and low dispersion and relatively small nonlinearity, by which we mean that the nonlinear length scale is long compared to typical dispersion length scales.

In the numerical simulations performed in this dissertation, we used a a method called the *local error method*, in which the step size is selected by bounding the relative local error of the step [80]. We will describe this method in this section and compare its performance to four commonly used step-size selection methods that are based on physical intuition. In the first of these four methods, called the *nonlinear phase rotation method*, the step size is chosen so that the phase change due to nonlinearity does not exceed a certain limit [10]. This method was designed with soliton propagation in mind. The second, the *logarithmic step size method*, is designed to efficiently suppress spurious four-wave mixing, by employing a logarithmic distribution of the step sizes [82]. In the third method, the *walk-off method*, the step size is chosen to be inversely proportional to the product of the absolute value of dispersion and the spectral bandwidth of the signal. The idea behind this criterion is to resolve the collisions between pulses in different channels or at least to have a measure for the violation of this criterion. This method was designed for low power, multi-channel systems. In the fourth, *the constant step size method*, the step sizes are kept constant along the whole transmission path.

The local error method is inspired by and closely related to widely-used algorithms for adaptively controlling the step size in ordinary differential equation solvers [83]. In particular, we have adopted the well-known techniques of step-doubling to estimate the local error and linear extrapolation to obtain the higher-order solution. As is typically the case for higher-order schemes, our scheme has the advantage that it is much more computationally efficient than a second-order scheme when the global accuracy is high [84]. On the other hand, it can be less efficient at low accuracy. This behavior is consistent with the results of Fornberg and Driscoll [85], who compared split-step methods of order 2, 4, and 6 with several higher-order linear multistep methods. For a two-soliton collision, Fornberg and Driscoll showed that for the global error range of $10^{-3}$–$10^{-2}$, the second-order split-step scheme is more efficient than the fourth- and sixth-order schemes. However, for global errors smaller than $10^{-4}$, the higher-order schemes become more efficient. We found similar qualitative behavior for the second-order schemes and third-order local error method that we discuss here.

### Origin of the split-step error

To estimate the local and global errors in the split-step Fourier method it is convenient to represent (2.2) in the form

$$\frac{\partial u(z,t)}{\partial z} \;=\; (\hat{D}+\hat{N}[u])u(z,t), \tag{3.1}$$

where $\hat{D} = -i(\beta''/2)\partial^2/\partial t^2 + g$ is the linear operator and $\hat{N}[u] = i\gamma|u|^2$ is the nonlinear operator. In the symmetric split-step scheme, the solution to (3.1) is approximated by

$$u(z+h,t) \;\approx\; \exp\Big(\frac{h}{2}\,\hat{D}\Big)\exp\Big\{h\hat{N}\Big[u\Big(z+\frac{h}{2},\,t\Big)\Big]\Big\}\exp\Big(\frac{h}{2}\,\hat{D}\Big). \tag{3.2}$$

Since operators $\hat{D}$ and $\hat{N}$ do not commute in general, the solution (3.2) is only an approximation to the exact solution. An argument based on the Baker-Campbell-Hausdorff formula shows that the *local error*, which is the error incurred in a single step of the symmetric split-step scheme, has a leading order term which is of third order in the step size $h$, *i.e.,* the error is $O(h^3)$ [86]. Since the total number of steps in a fiber span is inversely proportional to the average step size, the *global error* accumulated over a fiber span is second order in the step size, $O(h^2)$.

Finding an optimal step size distribution depends on the particular optical transmission system. We will review several criteria for choosing the step size in the split-step Fourier method, and we will introduce a new criterion based on a measure of the local error.

### Nonlinear Phase Rotation Method

The nonlinear phase rotation method is a variable step size method that is designed for systems in which nonlinearity plays a major role. For a step of size $h$, the effect of the nonlinear operator $\hat{N}$ is to increment the phase of $u$ by an amount $\phi_{\mathrm{NL}} = \gamma |u|^2 h$. If we impose an upper limit $\phi_{\mathrm{NL}}^{\max}$ on the nonlinear phase increment $\phi_{\mathrm{NL}}$, we obtain the bound on the step size:

$$h \leq \frac{\phi_{\mathrm{NL}}^{\max}}{\gamma |u|^2}. \tag{3.3}$$

This criterion for selecting the step size was originally applied to simulate soliton propagation and is widely used in optical fiber transmission simulators. However, as we will show, this approach is far from optimal for many modern communications systems.

### Spurious Four-Wave Mixing and Logarithmic Step Size Distribution

An improper distribution of the step sizes may lead not only to a general reduction of accuracy, but also to numerical artifacts. Forghieri [87] demonstrates that the power of the four-wave mixing products can be greatly overestimated by a constant step size method since four-wave mixing is a resonance effect. To efficiently suppress this numerical artifact, Bosco, *et al.* [82] used a logarithmic distribution of the step sizes to keep the spurious four-wave mixing components below a certain level. For a fiber span of length $L$ and loss coefficient $g$, the step size of $n$-th step is given by

$$h_n = -\frac{1}{2g} \ln \left[ \frac{1 - n\sigma}{1 - (n-1)\sigma} \right], \tag{3.4}$$

where $\sigma = \big[1 - \exp(-2gL)\big]/K$, and $K$ is the number of steps per fiber span. We will call this implementation of the split-step method the *logarithmic step size method.*

### Walk-off Method

In many optical fiber communications systems chromatic dispersion is the dominant effect, and nonlinearity only plays a secondary role, particularly in multi-channel systems in which the wavelength channels cover a broad spectrum. In this case it can be reasonable to use the *walk-off method*, in which the step size is determined by the largest group velocity difference between channels. The basic idea is to choose the step size to be smaller than a characteristic walk-off length. In a multi-channel system with large local dispersion, pulses in different channels move through each other very rapidly. To resolve the collisions between pulses in different channels, the step size in the walk-off method is chosen so that in a single step two pulses in the two edge channels shift with respect to each other by a time which is a specified fraction of the pulse width. Consequently, the step size is given by

$$h = \frac{C}{\Delta V_\mathrm{g}}, \tag{3.5}$$

where $\Delta V_\mathrm{g}$ is the largest group velocity difference between channels and $C$ is a constant that can vary from system to system. In any system, $\Delta V_\mathrm{g} = \big| D_2 \lambda_2 - D_1 \lambda_1 \big|$, where $D_1$ and $D_2$ are the dispersions corresponding to the smallest and largest wavelengths $\lambda_1$ and $\lambda_2$. Since $\Delta V_\mathrm{g}$ is constant in any particular kind of fiber, the step size is constant in each fiber segment. The walk-off method can be applied to single-channel as well as multi-channel systems by choosing $\lambda_1$ and $\lambda_2$ at the two edges of the signal spectrum.

### Constant Step Size Method

The simplest way to implement the split-step Fourier method is to use a constant step size along the whole transmission path. The global accuracy can be improved only by increasing the total number of steps. Note that the walk-off and constant step size methods are identical in systems with only one type of fiber.

### Local Error Method

In practice, it is desirable to have a general criterion for choosing the step size distribution that is close to optimal for an arbitrary system. Adaptive methods for controlling the step size using a measure of the local error are widely used in ordinary differential equation solvers [83]. We have implemented a scheme based on bounding the error in each step using the technique of step-doubling and local extrapolation. Given the field $u$ at a distance $z$, our aim is to compute the field at $z + 2h$. Suppose that we perform one step of size $2h$ in a symmetric split-step scheme. We will refer to the solution obtained at $z + 2h$ as the *coarse solution,* $u_c$. Since the local error in the symmetric split-step scheme is third order, there is a constant $\kappa$ so that

$$u_c = u_t + \kappa(2h)^3 + O(h^4), \tag{3.6}$$

where the *true solution* $u_t$ is the exact solution at $z + 2h$ obtained from the given solution at $z$. When we write that $u = v + O(h^4)$ for some functions $u$ and $v$, we mean that $|u - v| < Ch^4$, for some constant $C$. Next, we return to $z$ and compute the *fine solution* $u_f$ at the same distance $z + 2h$ using two steps of size $h$. As above, the fine solution is related to the true solution by

$$u_f = u_t + 2\kappa h^3 + O(h^4). \tag{3.7}$$

By taking an appropriate linear combination of the fine and coarse solutions we can obtain an approximate solution at $z + 2h$ for which the leading order error term is of fourth order in the step size $h$ [83]. From (3.6) and (3.7) it follows that this higher-order solution is given by

$$u_4 = \frac{4}{3}u_f - \frac{1}{3}u_c = u_t + O(h^4), \tag{3.8}$$

which we take as the input to the next step of size $2h$.

In the local error method the step size is adaptively chosen so that the local error incurred from $z$ to $z+2h$ is bounded within a specified range. The relative local error $\delta_4$ of the higher-order solution is defined by

$$\delta_4 = \frac{\|u_4 - u_t\|}{\|u_t\|}, \tag{3.9}$$

where the norm $\|u\|$ is defined as $\|u\| = \left(\int |u(t)|^2 dt\right)^{1/2}$. However, since we cannot compute the true solution $u_t$ in practice, we cannot compute the local error using (3.9). Instead, we define the *relative local error* of a step to be the local error in the coarse solution relative to the fine solution:

$$\delta = \frac{\|u_f - u_c\|}{\|u_f\|}. \tag{3.10}$$

Notice that $\delta$ is a measure of the true local error $\delta_4$, since $\delta$ can be obtained from $3\delta_4$ by replacing $u_t$ by $u_f$. The step size is chosen by keeping the relative local error $\delta$ within a specified range $(1/2\delta_G,\, \delta_G)$, where $\delta_G$ is a specified local error target. If $\delta > 2\delta_G$, the solution is discarded and the step size is halved. If $\delta$ is in the range $(\delta_G,\, 2\delta_G)$, the step size is divided by a factor of $2^{1/3}$ for the next step. If $\delta < 1/2\delta_G$, the step size is multiplied by a factor of $2^{1/3}$ for the next step. The reason for choosing this factor is that the local error should change by a factor less than 2, since it is proportional to $h^3$.

Rather than simply computing the fine solution, our method computes both the fine and coarse solutions. Although it requires 50% more Fourier transforms than does the standard symmetric split-step method, the method yields both a higher-order solution, which is globally third-order accurate and a measure of the relative local error, which is used to control the step size. However, it is important to understand that the higher-order solution $u_4$ is not always more accurate than the fine solution $u_f$, especially when the step size is large, since we are bounding the local error $\delta$ of the coarse solution relative to the fine solution, rather than using the true local error $\delta_4$.

Since we do not make any assumptions about the physical properties of the system, such as the amount of nonlinearity or dispersion, we expect the local error method to work well in an arbitrary system. In order to simulate a system with optimal efficiency, one must first ascertain the major sources of the split-step error. Assuming that the system is dominated by one source of error, one can select an appropriate criterion for choosing the step sizes. The local error method allows us to deal with general systems when the major source of error is unknown or may even change during the propagation, or when performing a series of simulations in which the system parameters are varied. The method can be applied to a variety of systems without sacrificing too much computational efficiency.

### 3.1.2 Split-step Fourier method: Numerical examples

In this part we compare the efficiency of the five implementations of the split-step method described above. Since most of the computational time is consumed by evaluating fast Fourier transforms (FFTs), we use the number of FFTs per simulation as a measure of the total computational cost [85]. We used the following scheme to compare the different methods. First, we compute a solution $u_a$ that is accurate to machine precision using the standard symmetric split-step method (with step sizes on the order of 5 cm). Next we compute the numerical solution $u_n$ for each of the different split-step implementations, and calculate the *global relative error* $\varepsilon$ defined by

$$\varepsilon = \frac{\|u_n - u_a\|}{\|u_a\|}, \tag{3.11}$$

where we use the norm defined in Section 3.1.1. We compare the performance of the different methods by plotting the number of FFTs versus the global relative error.

### Higher-Order Solitons

We start with the propagation of second- and fifth-order solitons. These systems are both highly nonlinear. In addition, higher-order solitons are very sensitive to numerical errors, thus requiring an efficient adaptive algorithm. The exact functional form of the $N$-soliton solution can be found in [88]–[93]. We use an anomalous-dispersion fiber with $\beta''= -0.1$ ps$^2$/km. The initial pulse is a hyperbolic secant of the form $u(t) = A\eta\big(|\beta''|/\gamma\big)^{1/2}\mathrm{sech}(\eta t)$, where the nonlinear coefficient is $\gamma = 2.2$ W$^{-1}$km$^{-1}$, the inverse pulse duration is $\eta = 0.44$ ps$^{-1}$, and where $A = 2$ and $A = 5$ for the second-order and fifth-order solitons respectively. The corresponding FWHM pulse duration is 4 ps and the peak powers are 35 mW and 220 mW for the second- and fifth-order solitons respectively. The number of Fourier modes is 1024 and the simulation time window is 50 ps. We show the performance of the different implementations of the split-step method applied to the second-order soliton in Fig. 3.1(a) and to the fifth-order soliton in Fig. 3.1(b).

In Fig. 3.1, we have plotted the number of FFTs versus the global relative error for the different step-size criteria. Although the performance of the local error method is not significantly better in the range of low accuracy values $10^{-2}$–$10^{-3}$, at high accuracy the computational cost of the local error method is one or two orders of magnitude less than for other methods. Notice that the nonlinear phase method performs better

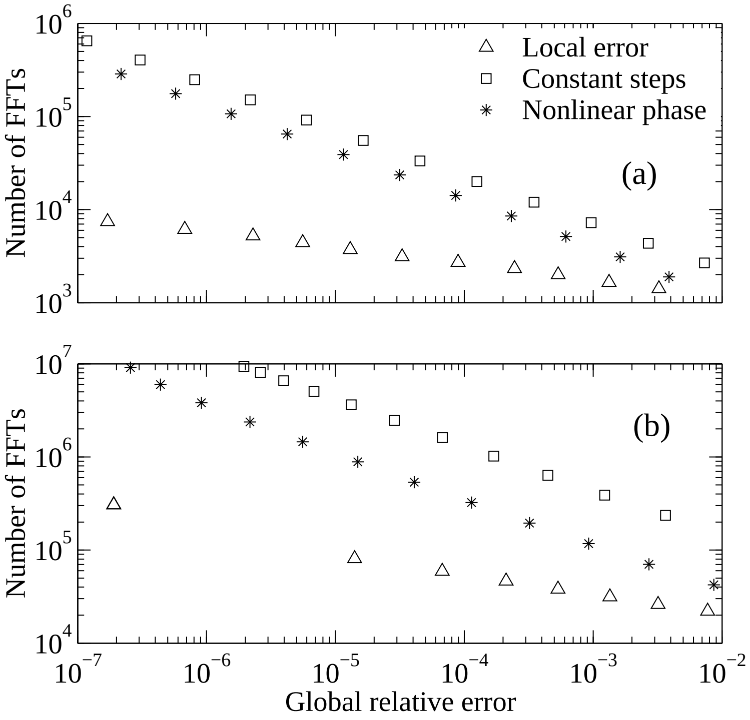

**Fig. 3.1:** Plot of the total number of FFTs versus global relative error $\varepsilon$ for second-order (a) and fifth-order (b) solitons.

than the constant step size method, consistent with the system's large nonlinearity. The slope of the local error method curve is less than those of the other two methods since the constant step size and nonlinear phase methods are globally second-order accurate, while the local error method is globally third-order accurate. The walk-off and constant step size methods are identical since this system includes only one type of fiber. The logarithmic step size method reduces to the constant step size method because the fiber is lossless, and (3.4) leads to a constant step size distribution.

### Soliton Collisions

Soliton collisions can be a good test for numerical methods because the subtle effect of four-wave mixing cancellation after the collision is very sensitive to numerical errors [89]. The fiber type and the initial pulse shape are the same as in the previous section, except that $A = 1$. The pulse duration is 4 ps and the peak power is 8.8 mW.

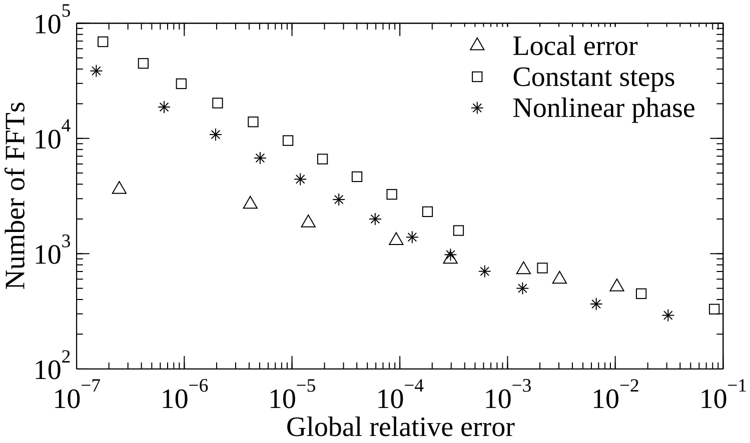

**Fig. 3.2:** Plot of the total number of FFTs versus global relative error $\varepsilon$ for a collision of two first-order solitons.

We launch two soliton pulses separated in time by 100 ps and with a central frequency difference of 800 GHz. The number of Fourier modes is 3072 and the simulation time window is 400 ps. We show the performance of the different methods in Fig. 3.2. The local error, constant step size, and nonlinear phase rotation methods perform equally well at low accuracy, when the global error is in the range $10^{-3}$–$10^{-1}$, while the local error method is much more efficient when the global error is less than $10^{-4}$. Global errors less than $10^{-4}$ are required to estimate the four-wave mixing terms correctly

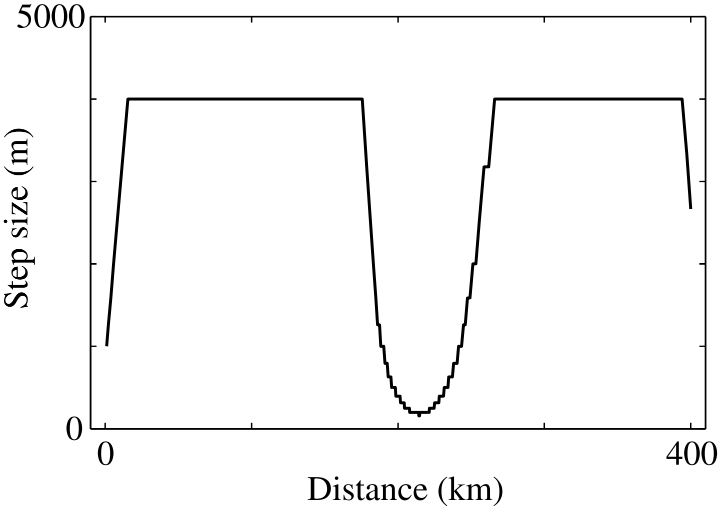

**Fig. 3.3:** Step size $h$ as a function of distance for the local error method applied to a collision of two first-order solitons.

and to have them cancel out after the collision. The nonlinear phase method still works better than the constant step size method because the nonlinear interactions are critical in the propagation. As before, the logarithmic step size and walk-off methods reduce to the constant step size method.

Using the example of a soliton collision, we illustrate the adaptive behavior of the local error algorithm. Fig. 3.3 shows the step size as a function of propagation distance for the soliton collision when the targeted range for the local error is $(0.5\times10^{-5}, 10^{-5})$ and the initial guess for the step size is 1000 m. Since the local error for this initial step is much less than the targeted range of values, at each step the step sizes are increased until the local error is within the targeted range. The pulse collision occurs at a distance of 200 km. At this point, we observe a significant decrease in the step size, which is necessary to accurately resolve the collision. After the collision, the step size is increased to the same value as before the collision. The last step is smaller than the previous step simply because the remaining section of the fiber is shorter than the step size chosen by the algorithm.

### Single-Channel Systems

In this part we study periodically-stationary dispersion-managed soliton (DMS) and chirped-return-to-zero (CRZ) systems that resemble experimental systems [55], [94]. The DMS system is highly nonlinear, meaning that both dispersion and nonlinearity determine the signal evolution, while the CRZ system is quasi-linear and the evolution is mostly determined by dispersion [95]. Thus we are studying the four split-step implementations using two different types of systems. We include fiber attenuation and gain, but we do not consider amplifier noise. We use 64-bit random bit streams that repeat periodically. We stress that our goal is to test the performance of the numerical methods for realistic systems rather than to achieve optimal propagation. Consequently, it is important that we have pulse streams rather than single pulses, that we use dispersion management, and that we include the effects of fiber loss and amplifier gain.

The DMS system is based on a 107 km dispersion map, which consists of four dispersion-shifted fiber spans, each of 25 km, with normal dispersion equal to $-1.10$ ps/nm-km, followed by 7 km of standard single-mode fiber with anomalous dispersion of 16.6 ps/nm-km at 1551 nm [94]. The loss in both fibers is 0.21 dB/km, and the amplifier spacing is 25 km with an additional amplifier after the standard single-mode fiber. We use Gaussian pulses with a FWHM duration of 9 ps, as is appropriate for a 10 Gbit/s bit rate. The peak power is 8 mW. The signal is launched in the middle of a span of anomalous fiber to ensure the periodicity of the pulse shape as it propagates along the fiber. The propagation distance is 1,280 km. The simulation time window is 6400 ps and the number of Fourier modes is 6144. We have not included a dispersion slope in this system since there is only a single channel and previous work indicates that higher-order dispersion plays no role [94].

The CRZ system is based on a 180 km dispersion map consisting of 160 km of dispersion-shifted fiber with dispersion $-2.44$ ps/nm-km, followed by 20 km of standard fiber with dispersion 16.55 ps/nm-km [55]. The dispersion slope is 0.075 $\mathrm{ps}^2$/nm-km and the fiber loss is 0.21 dB/km for both fibers, while the amplifier spacing is 45 km. Symmetric dispersion pre- and post-compensation is performed using fiber spans of length 2.0 km, where the dispersion is 93.5 ps/nm-km, the slope is $-0.2$ $\mathrm{ps}^2$/nm-km and the loss is 0.5 dB/km. The initial pulses are phase-modulated, raised-cosine pulses with 1 mW peak power and a chirp parameter equal to $-0.6$ [95]. The bit rate is 10 Gbit/s and the propagation distance is 1,800 km. The simulation time window is 6400 ps and the number of Fourier modes is 4096.

The performance of the four split-step implementations for the single-channel DMS and CRZ systems is shown in Figs. 3.4(a) and (b) respectively. In both systems, the local error method performs best over the entire range. Due to its higher order of ac-

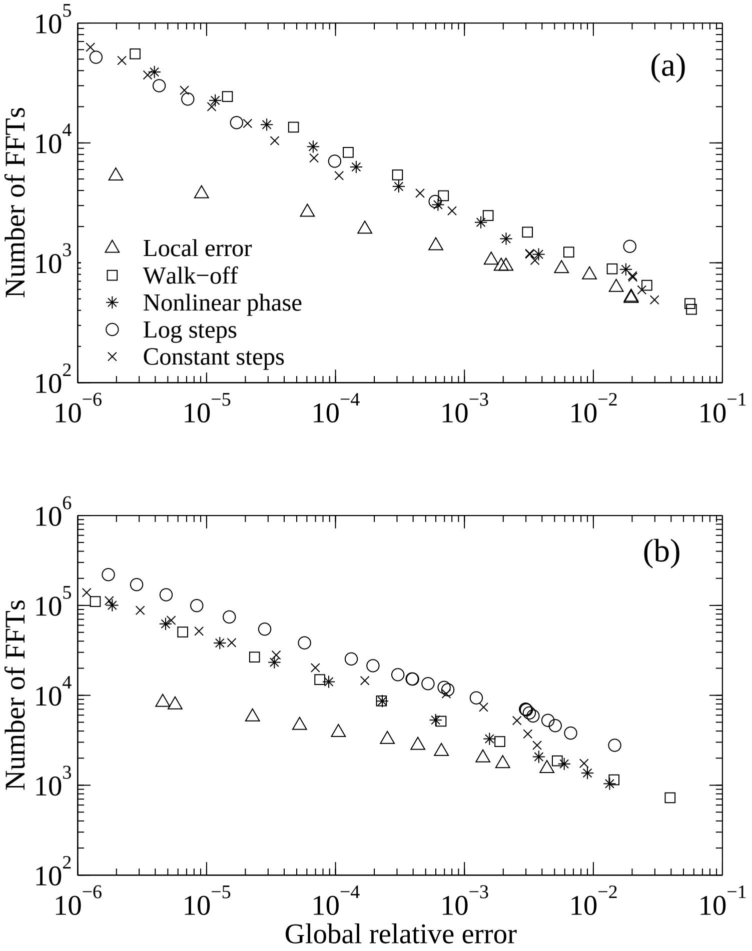

**Fig. 3.4:** Plot of the total number of FFTs versus global relative error $\varepsilon$ for the single-channel (a) DMS and (b) CRZ systems.

curacy, the data points for the local error method lie on a line with a smaller absolute slope than those of the other methods, as expected. However, all methods become comparable in the range of global errors $10^{-3}$–$10^{-1}$, the region of most interest in simulating fiber optic links. We note however, that in the CRZ system the performance of the logarithmic step size method is somewhat poorer than that of the nonlinear phase and walk-off methods.

### Multi-Channel CRZ System

In order to compare the split-step implementations for modeling multi-channel communications systems, we used the same CRZ system as described above. In Fig. 3.5, we show the performance of the split-step selection criteria on a 5-channel CRZ system with a 50 GHz channel spacing. As in the single-channel case, the local error method is much more efficient at high accuracy. However, at low accuracy, with the global error in the range $10^{-3}$–$10^{-1}$, the walk-off method performs best. At low accuracy, the local error method does not perform as well as the walk-off method for the following reasons. First, in the multi-channel CRZ system, the step size within each fiber in the local error method varies approximately within a factor of two, and the average value is comparable to the step size in the walk-off method for a given global error. However, each pair of steps in the local error method is 50% more expensive than in the walk-off method. In addition, when the step size is large and the global accuracy is low, the higher-order solution $u_4$ may not be as accurate as the fine solution $u_f$. Indeed, we have observed that the local error method performs slightly better at low global accuracy if we keep the fine solution $u_f$ instead of the higher-order solution $u_4$ at each step.

Next, we observe that the nonlinear phase rotation method does not perform as well as the walk-off method in the multi-channel CRZ system, although the performance of the two methods is comparable in the single-channel DMS and CRZ systems. There are two major reasons for this behavior. First, in contrast to the single-channel case, the walk-off criterion becomes more physically relevant in a WDM system, in which pulses in different channels collide. Second, the step size in the nonlinear phase rotation method is determined by the peak power in the time domain. In the single-channel CRZ system, the power function contains spikes due to the overlap between neighboring pulses. However, between amplifiers the peak power decreases monotonically with distance due to fiber attenuation. By contrast, the peak power

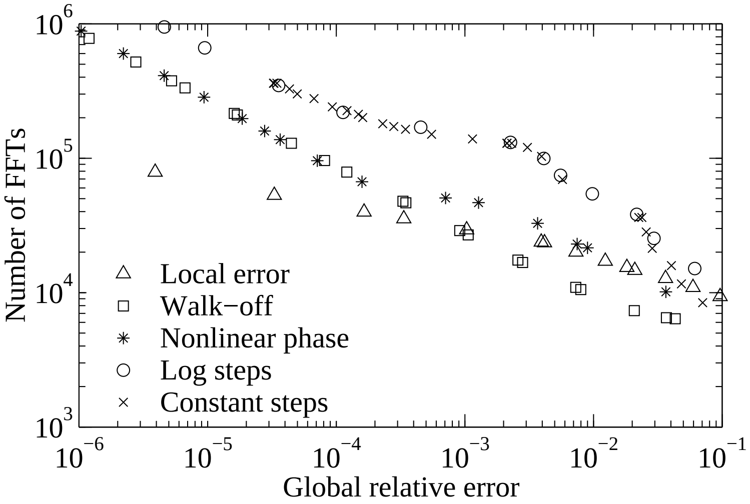

**Fig. 3.5:** Plot of the total number of FFTs versus global relative error $\varepsilon$ for the multichannanel CRZ system.

of the multi-channel system does not decrease monotonically with distance but contains irregular spikes because pulses from different channels rapidly pass through each other. As a consequence, there is a significant proportion of step sizes in the nonlinear phase rotation method that are much smaller than they need to be for a given global accuracy. The logarithmic step size method is not efficient in the CRZ system because the step size choice is only based on limiting spurious four-wave mixing, which is only one of the potential sources of error in a multi-channel simulation. We also found that in the logarithmic step size method, the error grows most rapidly in fibers with high dispersion. We find that the constant step size method is inefficient in the multi-channel CRZ system. The reason it performs so poorly is that for a given step size the global error does not accumulate linearly with distance. Consequently, in some sections of the transmission line the global error grows rapidly, while in others the error accumulates very slowly and computational effort is wasted.

In Figure 3.6, we show the step sizes in the local error method as a function of propagation distance when the targeted range for the local error is $(0.5 \times 10^{-4}, 2 \times 10^{-4})$. The upper two plots show the step sizes for the first two and last two periods of the dispersion map, and the lower two plots show the corresponding portions of

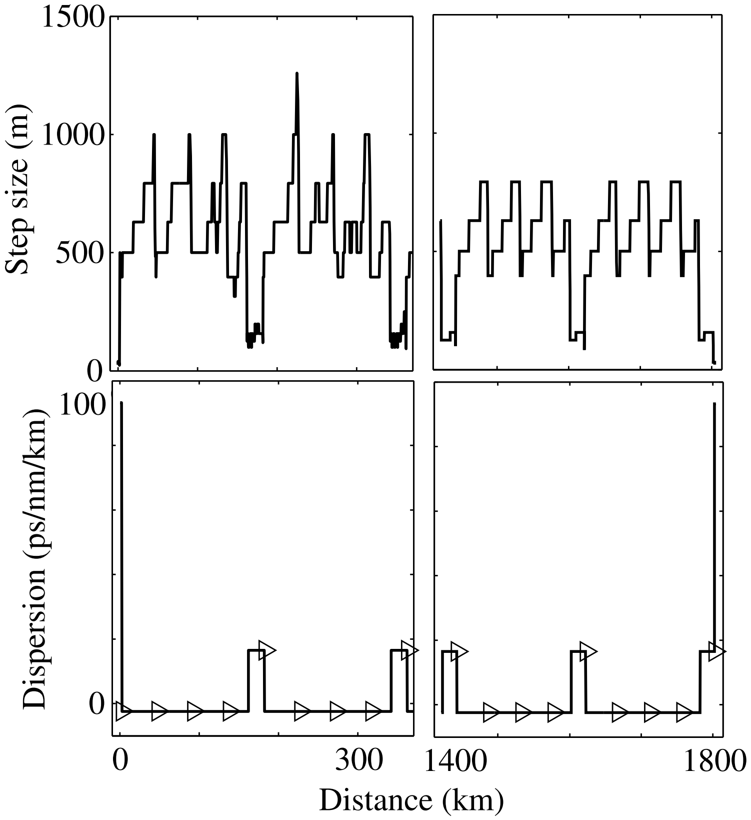

**Fig. 3.6:** Step size $h$ as a function of distance for the local-error method applied to the multichannel CRZ system. The upper two plots show the step sizes for the first two and last two periods of the dispersion map, and the lower two plots show the corresponding portions of the dispersion map. Triangles indicate the positions of amplifiers.

the dispersion map. The amplifiers, marked by triangles, are placed after the precompensation fiber, and then every 45 km. Notice that the step size increases as the signal power and the strength of the nonlinear interactions decrease due to the fiber loss. Also note that step size is smaller in fibers with higher dispersion since the pulses in neighboring channels move faster with respect to each other.

### Variation of Method Parameters

In this part we address two important questions concerning how the method parameter should be chosen to achieve a desired global accuracy. The method parameter is the parameter in a split-step method that we vary to adjust the accuracy of the method. First, for a given global error, how much does the method parameter depend on the particular system? Second, by what factor should the method parameter be decreased to halve the global error?

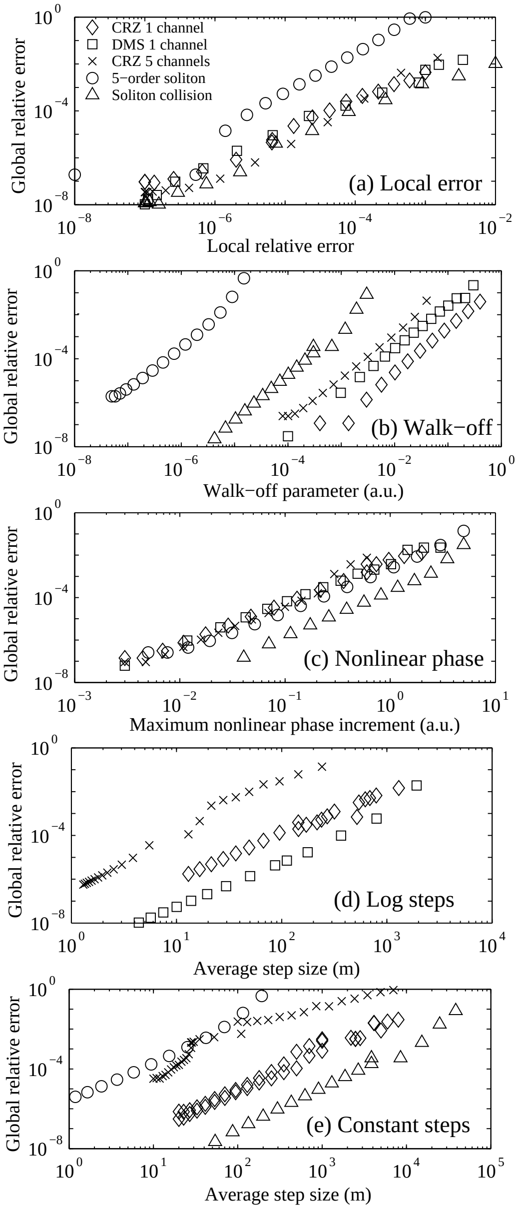

**Fig. 3.7:** Plot of the global error as a function of method parameter for (a) local error, (b) walk-off, (c) nonlinear phase, (d) logarithmic step, and (e) constant step methods.

To answer the first question, in Figs. 3.7(a)–(e), we show the dependence of the global error on the method parameter for the local error, walk-off, nonlinear phase, logarithmic step, and constant step size methods respectively. Although the walk-off method is the most efficient in some cases, it exhibits the worst system dependence. In particular, for the five systems we studied, when the global error is $10^{-3}$ the walk-off parameter varies over five orders of magnitude, whereas the parameter for the other three methods vary only over one to two orders of magnitude. Even omitting the two soliton systems from the comparison, the walk-off method has a greater system dependence than the local error method. Consequently, each new system requires a significantly different walk-off parameter to achieve the same global accuracy.

To answer the second question we examine the slopes of the curves in Fig. 3.7. For the walk-off, nonlinear phase, and logarithmic steps methods, the slopes are approximately 2, as expected, since these three schemes are second order and the step sizes depend linearly on the method parameter. Ideally, the global error should depend linearly on the local error. However, for the local error method the slopes of the curves in Fig. 3.7(a) are approximately 1.3, rather than 1. The reason for this discrepancy is that the true local error (3.9) is unavailable. Instead, we use an estimate of local error given by (3.10). In addition, in our local error algorithm, the local error (3.10) is maintained within a range of values rather than being kept constant.

## 3.2 Common reduced models

Solving the NLS equation numerically enables one to calculate the solution with high accuracy, typically limited by the numerical roundoff error. However, in many cases the computational cost of the numerical solution for WDM systems is unacceptable. A typical approach to efficiently evaluate the nonlinear crosstalk is to separate different nonlinear effects and calculate the signal distortion, the eye opening penalty, or the $Q$-factor degradation [7], [14], [33], [34], [48], [74], [96]. There exists a variety of methods for the evaluation of different nonlinear effects [7], [12], [16], [18]–[20], [28]–[30], [32], [35]–[37], [39], [41], [74], [97], [98], most of which are focused on the effect of XPM, as the dominant nonlinear effect in modern optical fiber communications systems.

### 3.2.1 Non-return-to-zero transmission

Most of the work to date on nonlinear impairments for the NRZ format is based on a pump-probe analysis and linearization of the distortion around the signal. The perturbation of the signal due to nonlinear interactions is assumed to be small compared to the signal and then one can derive linear equations for the perturbation. It is also typical to neglect the effects of self-phase modulation and dispersion on pump signal [16], [28]–[30] or to assume that dispersion is small [31], [37]. We will review the most common methods and we will only focus on the theoretical results of the publications discussed.

#### Additive perturbation approach

Chiang, *et al.* [28] introduced a pump-probe model to estimate the XPM-induced phase shift. The starting point is the equation for the electric field envelope of the probe wave $u_j$ in the presence of a pump wave with the envelope $u_k$,

$$\frac{\partial u_j}{\partial z}+\frac{\alpha}{2}u_j+\frac{1}{v_j}\frac{\partial u_j}{\partial t}=i\gamma(|u_j|^2+2|u_k|^2)u_j, \tag{3.12}$$

where $\alpha$ is the attenuation coefficient, $v_j$ is the group velocity of the $j$-th wave, and $\gamma$ is the nonlinear coefficient. The effect of chromatic dispersion within a single wavelength channel is neglected. The evolution of the power $P_k$ of the pump wave is assumed to be in the form

$$P_k(z,t)=\left|u_k\left(0,t-\frac{z}{v_k}+d_{jk}z\right)\right|^2\mathrm{e}^{-\alpha z}, \tag{3.13}$$

where $d_{jk}=v_j^{-1}-v_k^{-1}\approx D(\lambda_j-\lambda_k)$ is the walkoff parameter, with $D$ being fiber dispersion and $\lambda_{j,k}$ the wavelengths of the probe and pump. Assuming a continuous-wave (cw) probe and a sinusoidally-modulated pump,

$$\begin{aligned} P_j(0,t) &= |u_j(0,t)|^2=P_{j\,0} && \text{(3.14a)}\\ P_k(0,t) &= |u_k(0,t)|^2=P_{k\,0}+P_{km}\cos(\omega t+\theta), && \text{(3.14b)}\end{aligned}$$

Chiang, *et al.* show that the phase of the probe wave is given by

$$\phi_j(z,t)=\gamma(P_{j\,0}+2P_{k\,0})L_{\mathrm{eff}}+2\gamma P_{km}\sqrt{\eta}L_{\mathrm{eff}}\cos\left[\omega(t-\frac{z}{v_j}+\theta+\varphi)\right], \tag{3.15}$$

where $L_{\mathrm{eff}}=(1-\mathrm{e}^{-\alpha L})/\alpha$, $\eta$ is the XPM efficiency given by

$$\eta=\frac{\alpha^2}{\alpha^2+\omega^2 d_{jk}^2}\left[1+\frac{4\sin^2(\omega\,d_{jk}\,L/2)\mathrm{e}^{-\alpha L}}{(1-\mathrm{e}^{-\alpha L})^2}\right], \tag{3.16}$$

and

$$\varphi = \arg(\alpha + i\omega d_{jk}) + \arg\left[\mathrm{e}^{-\alpha L}\cos(\omega d_{jk}L) - 1 - i\mathrm{e}^{-\alpha L}\sin(\omega d_{jk}L)\right]. \tag{3.17}$$

The power of an arbitrarily modulated signal can be represented as a Fourier integral

$$P_k(z,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{P}_k(z,\omega)\exp(-i\omega t)d\omega, \tag{3.18}$$

where $\tilde{P}_k(z,\omega)$ is the spectrum of the pump wave intensity. Then the XPM-induced phase shift is given by

$$\phi_j(z,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}H(\omega)\tilde{P}_k(0,\omega)\cos\left[\omega\left(t - \frac{z}{v_j}\right) + \theta\right]\exp(-i\omega t)d\omega, \tag{3.19}$$

where $H(\omega)$ is the fiber XPM frequency response

$$H(\omega) = 2\gamma\sqrt{\eta(\omega)}\,L_{\mathrm{eff}}\,\exp[i\varphi(\omega)] \tag{3.20}$$

Equations (3.19) and (3.20) suggest that XPM acts on a signal as a phase modulator. The input to this modulator is the pump wave modified by the fiber response function $H(\omega)$. The authors also extend the analysis to the multi-span transmission systems. This analysis does not provide the amplitude distortion that occurs due to the action of dispersion on the cross-phase-modulated signal.

Hui, *et al.* [16] extended the work by Chiang, *et al.* [28] by calculating XPM-induced amplitude distortion and including the effect of dispersion on the probe signal. The theoretical model of XPM-induced crosstalk is also based on a pump-probe approach, in which the probe is cw and the pump is arbitrarily modulated, while the distortion induced by XPM is considered a small perturbation. The starting point is a propagation equation that is similar to (3.12) but it includes dispersion in the probe signal so that

$$\frac{\partial u_j}{\partial z} + \frac{\alpha}{2} u_j + \frac{1}{v_j}\frac{\partial u_j}{\partial t} + i\frac{\beta''}{2}\frac{\partial^2 u_j}{\partial t^2} = i\gamma(|u_j|^2 + 2|u_k|^2)u_j, \tag{3.21}$$

where $\beta''$ is the chromatic dispersion. The pump wave is assumed undistorted and its evolution is described by (3.14b). Assuming that the phase and amplitude perturbations are small and using a variable separation technique similar to the split-step method, one can obtain the following expression for the XPM-induced power variation:

$$\delta\tilde{s}_{jk}(\omega, z) = 4\gamma\, P_j(0)\tilde{P}_k(0,\omega)\,\frac{\sin(\beta''\omega^2 z/2)}{\alpha + i\omega d_{jk}}\,\exp\left(i\omega z/v_j - \alpha z\right). \tag{3.22}$$

This result is then generalized to the case of multi-span systems. For an arbitrary pump and probe waveform, Hui, *et al.* find that the crosstalk is given by

$$C_{jk}(t) = \mathrm{FT}^{-1}\left\{\mathrm{FT}[m_k(t)]\frac{\delta\tilde{s}_{jk}(\omega,z)}{P_j\,\exp(-\alpha z)}\,H_R(\omega)\right\} m_j(t), \tag{3.23}$$

where $H_R(\omega)$ is the receiver frequency response function, $m_j(t)$ and $m_k(t)$ are the normalized probe and pump waveforms at the transmitter, and $\mathrm{FT}(\cdot)$ indicates the Fourier transform, defined by (2.3). The system impact is defined in this work as the maximum eye closure. We note that no justification was given for the undistorted pump assumption. Since Hui, *et al.* did not compare their results to exact simulations it is not possible to assess the importance of this assumption from this work. Cartaxo [32], [99] showed that this assumption leads to inaccurate results in multi-span systems.

A related approach that is often cited in the literature is the *small signal analysis* that was originally developed by Wang and Petermann [100] to study the propagation of a noisy signal from a laser diode through a dispersive fiber. This approach describes the conversion of intensity and phase modulation originating from an arbitrary noise source. The main analytical result of this work is that given a weak intensity modulation $\Delta S_{\text{in}}(\omega)$ normalized to the average input intensity and a weak phase modulation $\phi_{\text{in}}(\omega)$, the intensity and phase at the fiber output are given by

$$\Delta S_{\text{out}}(\omega) = \cos(\omega^2 F)\,\Delta S_{\text{in}}(\omega) + \sin(\omega^2\,F)\,\phi_{\text{in}}(\omega) \tag{3.24a}$$

$$\phi_{\text{out}}(\omega) = -\sin(\omega^2 F)\,\Delta S_{\text{in}}(\omega) + \cos(\omega^2\,F)\,\phi_{\text{in}}(\omega), \tag{3.24b}$$

where $F = -\beta'' z/2$. Equations (3.24a) and (3.24b) relate intensity and phase modulation during transmission through the fiber. Although this approach does not take into account fiber nonlinearity, it can be used to calculate the intensity and phase modulation that is induced by the Kerr effect.

Using the small signal analysis, Cartaxo [32] derives an expression for the intensity of the XPM-induced distortion. The input to the fiber is a cw probe and a modulated pump signal. Using (3.12), one can calculate the phase modulation produced along the fiber, and, assuming that it is small, apply (3.24a) and (3.24b) to calculate the intensity distortion. In this approach, the effect of fiber dispersion on the pump waveform is taken into account only approximately. The evolution of the pump waveform is calculated using the small signal analysis:

$$\tilde{P}_k(z,\,\omega) = \tilde{P}_k(0,\,\omega)\;\cos(qz)\exp\left(-\alpha z + i\omega z/v_k\right), \tag{3.25}$$

where $q = \omega^2 D_k \lambda_k^2/(4\pi c)$ with $D_k$, and $\lambda_k$ being dispersion coefficient and wavelength of the pump channel. Note that this approximation is only valid for weak dispersion and narrow-band signals. It cannot be applied to typical RZ-modulated signals as RZ pulses typically undergo significant dispersive broadening. It can then be shown that the XPM-induced intensity variation is described by [32]

$$\begin{aligned}&\tilde{P}_{\mathrm{XPM}}(\omega) = 2\gamma P_j(0)\tilde{P}_k(0,\omega)\exp\left(-\alpha z + i\omega/v_j z\right)\\ &\cdot\Bigg(\frac{1}{a^2+(b+q)^2}\left\{a\sin(bz)-(b+q)\,\cos(bz)+[a\sin(qz)+(b+q)\cos(qz)]\mathrm{e}^{-\alpha z}\right\}\\ &+\frac{1}{a^2+(b-q)^2}\left\{a\sin(bz)-(b-q)\,\cos(bz)-[a\sin(qz)+(b-q)\cos(qz)]\mathrm{e}^{-\alpha z}\right\}\Bigg),\end{aligned} \tag{3.26}$$

where $a = \alpha + i\omega\, d_{jk}$, $b = \omega^2 D_j \lambda_j^2/(4\pi c)$, with $D_j$ and $\lambda_j$ being dispersion coefficient and wavelength of the probe pump channel. For an arbitrary modulation, the pump intensity can be represented using (3.18). Then the intensity fluctuation is given by

$$P_{\mathrm{XPM}}^{\mathrm{mod}}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H_{\mathrm{XPM}}(\omega)\tilde{P}_k(\omega)\exp(-i\omega t)d\omega, \tag{3.27}$$

where $H_{\mathrm{XPM}}(\omega)$ is determined from (3.26) as $H_{\mathrm{XPM}}(\omega) = \tilde{P}_{\mathrm{XPM}}(\omega)/\tilde{P}_k(\omega)$. The analysis is then generalized for multi-span systems [32]. In the case of negligible pump envelope change and large channel spacing the approaches of Hui [16] and Cartaxo [32] give the same result, as shown by Bellotti [30].

Bellotti, *et al.* [30] used essentially the same theoretical approach to derive the XPM-induced power variation as in [31], [32] employing the small signal analysis [100], except that the pump waveform distortion due to chromatic dispersion was neglected. The work by Jiang, *et al.* [35] also uses the small signal analysis as in [99] neglecting the pump wave distortion but including the Raman effect. Later, Luís and Cartaxo [37] extended the analysis by including the effect of SPM on the probe wave. They found that including the SPM effect can be important when propagating a probe signal with a power of 7 dBm and a pump signal with a power of 4 dBm over sixteen 80 km-long spans of standard SMF and DCF fiber.

The validity of the additive perturbation methods described above is commonly assessed by estimating the accuracy of the normalized power distortion as a function of modulated frequency. The analytical results are in good agreement with both full numerical simulations [16], [28], [30], [32], [37] and experiments [16], [28], [96]. The main limitation is the assumption that the pump is either undistorted or approximated by (3.25). In [16], [30], the authors showed a good agreement between the analytical and numerical approaches by comparing the NRZ waveform distortions. In [37], [96] the authors calculated the standard deviation of the received current analytically and showed a good agreement with the numerical simulations. The approaches described in [16], [28], [30], [32], [37] fail at high power and high dispersion. No accurate BER calculation has been presented in the literature.

### Multiplicative perturbation approach

Ciaramella and Forestieri [36] showed that using a multiplicative perturbation instead of the traditional additive perturbation can be more accurate in evaluating the waveform distortion due to the Kerr effect. Consider a normalized NLS equation in which the field envelope is redefined such that $u(z,t) := u(z,t)\exp(\alpha z/2)$,

$$\frac{\partial u(z,t)}{\partial z} + i\frac{\beta''(z)}{2}\frac{\partial^2 u(z,t)}{\partial t^2} - i\gamma \mathrm{e}^{-\alpha z}|u(z,t)|^2 u(z,t) = 0. \tag{3.28}$$

Given the linear solution $u_0(z,t)$ of (3.28) in the case $\gamma = 0$, the general solution to (3.28) is assumed to be of the form

$$u(z,t) = u_0(z,t)\exp\left[-i\gamma\vartheta(z,t)\right], \tag{3.29}$$

where $\vartheta(z,t)$ is a complex-valued function that also depends on $\gamma$. Assuming that $\gamma$ is small, one may expand $\vartheta$ in a Maclaurin series in $\gamma$ as

$$\vartheta = \sum_{k=0}^{\infty} \vartheta_k \gamma^k. \tag{3.30}$$

The first-order perturbation can be obtained by using the first term $\vartheta_0$ from (3.30). Then the evolution of $\vartheta_0$ is described by a linear equation

$$\frac{\partial \vartheta_0}{\partial z} = i\frac{\beta''}{2}\left(\frac{2}{u_0}\frac{\partial \vartheta_0}{\partial t}\frac{\partial u_0}{\partial t} + \frac{\partial^2 \vartheta_0}{\partial t^2}\right) + \mathrm{e}^{-\alpha z}|u_0|^2. \tag{3.31}$$

Using the split-step Fourier method, one then finds that the solution of (3.31) may be written [36],

$$\vartheta_0(z,t) = \frac{1}{u_0(z,t)}\int_0^z \left[|u_0(x,t)|^2 u_0(x,t)\right] \otimes h(z-x,t)\mathrm{e}^{-\alpha x}dx, \tag{3.32}$$

where $h(z,t) = \mathrm{FT}^{-1}\left[\exp(i\beta''\omega^2 z/2)\right]$ is the linear fiber impulse response, and the symbol $\otimes$ denotes the convolution [36].

Calculating the second term $\vartheta_1$ in (3.30), one can obtain the second order perturbation

$$\begin{aligned}\vartheta_1(z,t) &= \frac{1}{2}i\vartheta_0^2(z,t)\\ &\quad - i\frac{1}{u_0(z,t)}\int_0^z \left\{[2\vartheta_0(x,t) - \vartheta_0^*(x,t)]|u_0(x,t)|^2 u_0(x,t)\right\} \otimes h(z-x,t)\mathrm{e}^{-\alpha x}dx.\end{aligned} \tag{3.33}$$

Even though we consider the first-order term $\gamma\vartheta_1$ in (3.30), the perturbation is called the second order in the original work [36] because of an extra factor of $\gamma$ in the exponent in (3.29).

As a result, (3.29), (3.32), and (3.33) are the approximate solution to the NLS equation (3.28) for an arbitrary input signal, which can be either a single channel or multiple channels in a WDM system. As a measure of accuracy Ciaramella and Forestieri used the normalized square deviation between the accurate and approximate solutions for a 16-bit NRZ waveform and they showed that for a single channel NRZ system, the first-order multiplicative perturbation approach is comparable to the first-order additive approach [101], while the second-order multiplicative approach is up to two orders of magnitude more accurate than the additive theory for the same computational effort. A significant drawback of this approach is that the signal $u_0(z,t)$ in the denominator of (3.32) and (3.33) can become zero leading to numerical problems. A suggested way to overcome these problems is to set a finite extinction ratio or to use the expansion $\exp(x) \approx 1 + x$ in (3.29) when $u_0(z,t)$ becomes small. In this case the multiplicative perturbation reduces to the additive perturbation. However, this approach requires a time- and distance-dependent switch inside the computation and seriously complicates it. This problem could in principle be resolved by using a combined additive and multiplicative perturbation approach developed by Secondini [102], [103], where he introduced an extra variable to avoid the singularity, but this approach has yet to be applied to nonlinear signal propagation.

### Impact on the system performance

In all the work described above, except [36], the effect of XPM is characterized by the variance of the received current. The XPM impact on the system in [30] is estimated by first calculating the probe waveform distortion using a 128-bit pseudo-random sequence in the pump channel and then calculating the standard deviation of the probe intensity at the receiver. The impact of XPM on the system performance can be estimated as a variance $\sigma^2_{\mathrm{XPM}}$ of the received current using the power spectral density $S_P(f)$ of the received current of the probe signal, as has been done in [35], [37], [96]. The power spectral density of a signal is defined as the Fourier transform of its autocorrelation function [104]. Using (3.27), one can obtain [104]

$$S_P(\omega) = 4P_j(0)|H_{\mathrm{XPM}}(\omega)|^2 S_{\mathrm{mod}}(0,\omega). \tag{3.34}$$

The input power spectral density of the intensity-modulated signal with random binary data $S_{\text{mod}}(0,\omega)$ is expressed as [104]

$$S_{\text{mod}}(0,\omega) = \frac{1}{4T}\,|\tilde{P}_p(\omega)|^2 + \frac{1}{4T^2}\sum_{k=-\infty}^{\infty}\frac{1}{2\pi}\left|\tilde{P}_p\left(\frac{k}{T_{\text{bit}}}\right)\right|^2 \delta\left(\omega - \frac{2\pi k}{T_{\text{bit}}}\right), \tag{3.35}$$

where $T_{\text{bit}}$ is the bit period, and $\tilde{P}_p(\omega)$ is the Fourier transform of a single pulse. The current variance is then given by

$$\sigma^2_{\text{XPM}} = \frac{1}{P_j(0)}\int_{-\infty}^{\infty} S_P(\omega)|H_e(\omega)|^2\frac{d\omega}{2\pi}, \tag{3.36}$$

where $H_e(\omega)$ is the electrical filter transfer function.

### 3.2.2 Return-to-zero transmission

Both soliton and linear WDM RZ systems are mainly affected by the fiber nonlinearity in the form of collision-induced timing jitter, which has been a major focus of recent research [12], [19], [20], [39], [41], [74], [98], [105]–[110]. All this work is based on the analysis of two-pulse collisions. The analysis starts with the reduced propagation equation for a pulse $u$ in the presence of another pulse $v$ that is offset from $u$ in both time and central frequency

$$\frac{\partial u(z,t)}{\partial z} + i\frac{\beta''(z)}{2}\frac{\partial^2 u(z,t)}{\partial t^2} - i\gamma\left[|u(z,t)|^2 + 2|v(z,t)|^2\right]u(z,t) = g(z)u(z,t), \tag{3.37}$$

with the same variables and parameters as in (2.2). The propagation equation (3.37) is obtained from (2.2) assuming that the frequency separation between the pulses $u$ and $v$ is much larger than their spectral bandwidths. Also, FWM is neglected in (3.37). Commonly, the central time and frequency of the pulse $u$ is defined as [12],

[20], [105]–[107]

$$\begin{aligned} T_c &= \frac{1}{E}\int_{-\infty}^{\infty} t|u|^2 dt, &(3.38a)\\ \Omega_c &= \frac{1}{E}\int_{-\infty}^{\infty} \mathrm{Im}\left(\frac{\partial u^*}{\partial t}u\right) dt, &(3.38b)\end{aligned}$$

where $E = \int_{-\infty}^{\infty} |u|^2 dt$. Using these definitions, one can obtain the following dynamic equations for the central time and frequency from (3.37)

$$\begin{aligned} \frac{dT_c}{dz} &= \beta''\Omega_c, &(3.39a)\\ \frac{d\Omega_c}{dz} &= -\frac{2\gamma}{E}\int_{-\infty}^{\infty} |u|^2 \frac{\partial |v|^2}{\partial t} dt. &(3.39b)\end{aligned}$$

Equations (3.39a) and (3.39b) are the starting point for most of the published work [12], [20], [105]–[107], and they can be used to calculate collision-induced frequency and time shifts for a given system and pulse shape.

Classical solitons are known to be transparent to complete collisions, so that there is no change in the central frequency, energy, and shape [89], [105]. Mollenauer, *et al.* [106] showed that solitons remain transparent in systems with fiber loss and lumped amplification, as long as the length of the collision is long enough compared to the amplifier spacing or to a period of variation in some other parameter, such as the fiber's chromatic dispersion. Mollenauer, *et al.* [106] also derived the first analytical expression for the frequency shift of a soliton during a two-soliton collision for a periodically-amplified system with periodically varying dispersion. Let us make the transformation

$$u := u/\sqrt{G} \quad \text{and} \quad z := z\,\Delta, \qquad (3.40)$$

where $dG/dz = 2gG$, and $\Delta = D/\langle D\rangle$ is the dispersion parameter normalized to the path-averaged dispersion $\langle D\rangle$. Then the original variations in dispersion and gain and loss are equivalent to variations in the nonlinear coefficient $\gamma(z) := \gamma G(z)/\Delta(z)$.

The solution of the propagation equation is assumed to be in the form $u = \mathrm{sech}(t + \Omega z)\exp(-i\Omega z + i\phi)$ [106], where $2\Omega$ is the frequency separation between the pulses $u$ and $v$. By performing a spatial Fourier transform of $\gamma(z)$ as

$$\gamma(z) = \frac{1}{2\pi}\int \tilde{\gamma}(k)\exp(ikz)dk, \tag{3.41}$$

one can obtain the collision-induced frequency shift $\delta\Omega$ [106],

$$\delta\Omega = \mathrm{Im}\frac{32}{\pi^2}\int_0^\infty \frac{\tilde{\gamma}(k)}{k}\frac{x^4}{\sinh^2(x)}dk, \tag{3.42}$$

where $x = \pi k/(4\Omega)$.

The approach of Mollenauer, *et al.* [106] has been extended to calculate the standard deviation of the collision-induced time shift for a large number of pulse collisions in WDM soliton systems with dispersion decreasing fibers [19], [20], [109].

Devaney, *et al.* [107] have applied Mollenauer's theory [106] to two-step dispersion-managed soliton systems and have shown that the maximum frequency shift decreases with the strength of the dispersion management irrespective of the fiber sequence. Devaney, *et al.* [107] showed that the effectiveness of the dispersion management has two sources: First, the large local dispersion reduces the residual frequency shift that is due to single-soliton collisions, and second, the modified collision dynamics, whereby each complete interaction comprises a number of self-canceling fast crossings. A similar result was obtained by Hirooka and Hasegawa [111] that includes a description of the behavior of chirped solitons and a derivation of more accurate expressions for Gaussian pulses as opposed to using the formulae for hyperbolic-secant pulses.

Collision-induced timing jitter in quasi-linear systems has recently been studied by Grigoryan and Richter [12], Ablowitz, *et al.* [39], and Ahrens, *et al.* [41]. These methods can be applied to both quasi-linear and soliton systems. In [12], the authors used a numerical approach to calculate the time shift directly from (3.39a) by solving the NLS to obtain the pulse shape $u$ as a function of distance. They also calculated the standard deviation of the time shift. This approach can efficiently be applied to arbitrary WDM RZ systems and it takes into account the exact evolution of the pulses. Ablowitz, *et al.* [39] calculated the collision-induced time shift in a quasi-linear system analytically using asymptotic analysis. The pulse shape in [39] is assumed to be Gaussian and the evolution of the pulses is assumed to be linear. Then by integrating (3.39a) and (3.39b) and using Gaussian pulses, one can obtain expressions of form $\int f(z)\exp[-\eta\phi(z)]dz$. The parameter $\eta \gg 1$ is proportional to the square of the channel spacing and $\phi(z) > 0$. Integrals of this type are amenable to asymptotic analysis, specifically, by Laplace's method [112]. For large $\eta$ the exponent decays rapidly away from the minimum points of $\phi(z)$ and the integrals can be evaluated as sums of Gaussian-type integrals. Physically, the minima of $\phi(z)$ are located at the centers of collisions. For a complete collision, the residual collision-induced time shift $\delta t_i$ is approximately given by

$$\delta t_i \sim \frac{\gamma E}{2\Omega\langle D\rangle}, \tag{3.43}$$

where $E$ is the initial pulse energy, $\langle D\rangle$ is the path-averaged dispersion and $2\Omega$ is the frequency separation between the pulses. For an incomplete collision [113],

$$\begin{aligned}\delta t_c \sim \frac{\gamma E}{4\Gamma\Omega\langle D\rangle z_a}\{&\xi(f_a,f_b) + g([\theta/2 - f_b\theta]z_a) - g([\theta/2 - f_a\theta]z_a)\\ &+g([\theta/2 + f_a(1-\theta)]z_a) - g([\theta/2 + f_b(1-\theta)z_a])\ \}\end{aligned} \tag{3.44}$$

where $\Gamma$ is the loss coefficient, $z_a$ is the amplifier spacing, $f_a = 1/2 - z_0/L_c$, $f_b = 1/2 + (L - z_0)/L_c$, $\xi(f_a, f_b) = 0$ if $(f_a - 1/2)(f_b - 1/2) > 0$ and $\xi(f_a, f_b) = 2\Gamma z_a$ if $(f_a - 1/2)(f_b - 1/2) < 0$, and the function $g(z)$ is given by

$$g(z) = \frac{2\Gamma z_a}{1 - \exp(-2\Gamma z_a)}\ \exp(-2\Gamma z). \tag{3.45}$$

The parameter $\theta$ is the fraction of the map with dispersion $D_1$ of a two-step dispersion map. Ahrens, *et al.* [41] extended this approach for the case of quasi-linear systems with dispersion pre- and post-compensation. This asymptotic approach agrees well with full numerical simulations for a 10 Gb/s system with Gaussian pulses and channel spacing larger than 100 GHz. However, approximating a realistic raised-cosine pulse (2.1) with a Gaussian pulse shape and using channel spacings smaller than 100 GHz limits the accuracy of this approach.

In all published work on RZ systems just discussed, the measure of system performance was collision-induced time or frequency shift and their standard deviation, except for [20], in which the authors used the BER, assuming that there is an error if the timing shift exceeds a maximum tolerable time shift at the receiver. No accurate BER calculation for RZ systems that takes into account nonlinear signal distortion has been reported in the literature prior to our work.

# 4. STATISTICAL METHODS

Many of the effects that degrade the performance of communications systems are random in nature. Nonlinear crosstalk, noise generated in the amplifiers, and randomly varying birefringence are just a few. To calculate the BER due to these effects, one must first compute the probability density functions (pdfs) of the variables that describe the performance of the system. In the work presented here, these variables are either the collision-induced time shift or the received current. In this chapter, we describe statistical methods that we have used to compute the required pdfs. The main advantage of a statistical method is that it allows one to use the full system model, for which analytical solutions do not exist, so that one can in principle produce an arbitrarily accurate estimate of the pdfs.

We use the following scheme in our statistical approach: We solve the propagation equation for the system numerically and compute the time shift of a target pulse at the end of the transmission or the value of the received current at the clock recovery time. We run a set of Monte Carlo simulations in which we change the input bit strings in the WDM channels that are neighbors of the target pulse, and we build a histogram of the time shift or the current, from which we estimate the corresponding pdf. We assume that the bit strings have a uniform distribution, *i.e.*, each bit has a probability of 1/2 of being 0 or 1. In standard Monte Carlo simulations, we draw samples (bit strings) from a uniform distribution. However, most of the samples in a standard Monte Carlo simulation are near the mean of the pdf, while tails remain undersampled. In order to estimate the pdf more efficiently, we use biased Monte

Carlo sampling techniques — the standard importance sampling method (IS) and the multicanonical Monte Carlo method (MMC). Both of these techniques have been discussed in the literature [75], [114]–[131], and we only review the basics.

## 4.1 Importance sampling

### 4.1.1 Biasing distribution

Suppose one has a random vector $\mathbf{z}$= $(z_1, z_2, ..., z_n)$ defined on a sample space $\Omega$. In our case, the vector $\mathbf{z}$ is discrete, and it represents the set of bits in all the channels: $z_l = 0$ or 1, where $l = 1, ..., n$. We want to compute the average

$$P = E[f(\mathbf{z})] = \sum_{\Omega} f(\mathbf{z}_k) p(\mathbf{z}_k), \tag{4.1}$$

where $E(\cdot)$ denotes the statistical average with respect to $p(\mathbf{z})$, $f(\mathbf{z})$ is an indicator function that is equal to 1 if some quantity $X(\mathbf{z})$ falls into a desired subspace $S$ of $\Omega$:

$$f(\mathbf{z}) = I_S \left[ X(\mathbf{z}) \right]. \tag{4.2}$$

In our studies, the quantity $X(\mathbf{z})$ is either the collision-induced time shift or the received current. As was mentioned earlier, calculating (4.1) exactly is not possible since $f(\mathbf{z})$ is a complicated nonlinear function of the signal and system parameters. We estimate (4.1) in a set of Monte Carlo simulations,

$$\hat{P} = \frac{1}{M} \sum_{m=1}^{M} f(\mathbf{z}_m), \tag{4.3}$$

where $M$ is the number of Monte Carlo samples. In a standard Monte Carlo simulation, one draws samples $\mathbf{z}_m$ from a distribution $p(\mathbf{z})$. In this case, the estimator $\hat{P}$ is just the number of hits in the subspace of interest $S$ divided by the total number of samples. It follows from (4.3) that $E(\hat{P}) = P$. Typically, we are interested in rare events, for example, large nonlinear distortions that lead to transmission errors, and a standard Monte Carlo simulation would require us to pick a prohibitively large number of samples to observe a small number of events in the region of interest in the sample space that lead to errors. In standard importance sampling (IS), we use a modified probability distribution $p^*(\mathbf{z})$ to draw samples,

$$P = \sum_{\Omega} f(\mathbf{z}_k)\frac{p(\mathbf{z}_k)}{p^*(\mathbf{z}_k)}p^*(\mathbf{z}_k) = E^*\left[f(\mathbf{z}_k)\frac{p(\mathbf{z}_k)}{p^*(\mathbf{z}_k)}\right]. \tag{4.4}$$

The function

$$L(\mathbf{z}) = \frac{p(\mathbf{z})}{p^*(\mathbf{z})} \tag{4.5}$$

is called the *likelihood ratio*, and the distribution $p^*(\mathbf{z})$ is called the *biasing distribution*. Then the estimator of $P$, denoted $\hat{P}^*$, with IS becomes

$$\hat{P}^* = \frac{1}{M}\sum_{m=1}^{M} f(\mathbf{z}_m)L(\mathbf{z}_m), \tag{4.6}$$

where the samples are now drawn from the biasing distribution. Intuitively, we can sample region of interest $S$ in the sample space more efficiently if we choose the biasing distribution such that $p^*(\mathbf{z}) > p(\mathbf{z})$. Then the number of samples falling into $S$ will be larger. Another way to understand the benefit of IS is to look at the variance of the estimates $\hat{P}$ and $\hat{P}^*$ in that region. If the samples $\mathbf{z}_m$ are statistically independent, the variance $\sigma_{\hat{P}}^2$ of $\hat{P}$ is determined by the sample variance $\sigma_f^2$,

$$\sigma_{\hat{P}}^2 = \sigma_f^2/M, \tag{4.7}$$

and similarly

$$\sigma_{\hat{P}^*}^{*2} = \sigma_{f^*}^{*2}/M, \tag{4.8}$$

where $f^* = fL$. In the unbiased case, the sample variance is given by [116], [124]

$$\sigma_f^2 = E[f^2] - E[f]^2 = P - P^2, \tag{4.9}$$

where we used the relation for an indicator function, $f(z)^2 = f(z)$. The biased variance is given by

$$\begin{aligned} \sigma_{f^*}^{*\,2} = E^*[f^{*\,2}] - E^*[f^*]^2 &= \sum_{\Omega} f^2(\mathbf{z}_k)\frac{p^2(\mathbf{z}_k)}{p^{*\,2}(\mathbf{z}_k)}p^*(\mathbf{z}_k) - P^2 \\ &= \sum_{\Omega} f(\mathbf{z}_k)L(\mathbf{z}_k)p(\mathbf{z}_k) - P^2. \end{aligned} \tag{4.10}$$

From Eqs. (4.7) and (4.9), it follows that when we look into rare events such that $P \ll 1$, the relative variance of the estimator $\hat{P}$ is very large:

$$\frac{\sigma_{\hat{P}}}{P} \approx \frac{1}{\sqrt{MP}}. \tag{4.11}$$

With IS, the variance can be significantly reduced if we choose an appropriate biasing distribution such that $L(\mathbf{z})$ is small, as can be seen from Eqs. (4.9) and (4.10).

### 4.1.2 Multiple importance sampling and the balance heuristic

In practice, it is hard to find a single biasing distribution that will efficiently sample all the regions of the sample space that are needed to create a histogram of $X(\mathbf{z})$ that includes a representation of the subspace of interest $S$. It is necessary then to use several biasing distributions, and the technique of using several biasing distributions simultaneously is called *multiple importance sampling*. In this case, one needs an appropriate method to combine the results from all the biasing distributions. A typical way to do it is to weight the $N_b$ different distributions $p_i^*$ as follows,

$$P = \sum_{i=1}^{N_b} P_i = \sum_{i=1}^{N_b} \sum_{\mathbf{z}_k \in \Omega} w_i(\mathbf{z}_k) f(\mathbf{z}_k) L_i(\mathbf{z}_k) p_i^*(\mathbf{z}_k), \tag{4.12}$$

where $L_i(\mathbf{z}) = p(\mathbf{z})/p_i^*(\mathbf{z})$. In general, the weights $w(\mathbf{z}_k)$ are functions of $\mathbf{z}$, which enables one to efficiently combine the results from different biased samples. The estimator for $P$ with multiple IS is then

$$\hat{P} = \sum_{i=1}^{N_b} \hat{P}_i = \sum_{i=1}^{N_b} \frac{1}{M_i} \sum_{m=1}^{M_i} w_i(\mathbf{z}_m^i) f(\mathbf{z}_m^i) L_i(\mathbf{z}_m^i), \tag{4.13}$$

where $\mathbf{z}_m^i$ is the $m$-th sample from the distribution $p_i^*$, and $M_i$ is the total number of samples taken from the $i$-th distribution.

An efficient way to assign weights to the different biasing distributions is the balance heuristic method. The weight associated with a given sample $\mathbf{z}$ is determined by the relative likelihood of realizing this sample with the $i$-th biasing distribution relative to the total probability of realizing the same sample with all other biasing distributions,

$$w_i(\mathbf{z}) = \frac{M_i p_i^*(\mathbf{z})}{\sum_{j=1}^{N_b} M_j p_j^*(\mathbf{z})}. \tag{4.14}$$

The numerator in (4.14), and hence the $i$-th weight, is proportional to the expected number of hits from the $i$-th distribution. Therefore, according to (4.14), each distribution is weighted most heavily in those regions of the sample space where it is largest.

Given an estimator $\hat{P}$ (4.13), one can show that an unbiased estimator of its variance is given by [117]

$$\hat{\sigma}_{\hat{P}}^2 = \sum_{j=1}^{N_b} \frac{1}{M_j(M_j-1)} \sum_{m=1}^{M_i} \left[ w_i(\mathbf{z}_m^i) f(\mathbf{z}_m^i) L_i(\mathbf{z}_m^i) - \hat{P}_j \right]^2. \tag{4.15}$$

One can also obtain $\hat{\sigma}_{\hat{P}}^2$ without storing all the individual samples by using the recursion relation

$$\hat{\sigma}_{\hat{P}}^2 = \sum_{j=1}^{N_b} \frac{S_{j,M_j}}{M_j(M_j-1)}, \tag{4.16}$$

where

$$\hat{S}_{j,m} = \hat{S}_{j,m-1} + \frac{m-1}{m}\left[w_j(\mathbf{z}_m^j)f(\mathbf{z}_m^j)L_j(\mathbf{z}_m^j) - \hat{P}_{j,m-1}\right]^2 \tag{4.17}$$

and

$$\hat{P}_{j,m} = \frac{m-1}{m}\hat{P}_{j,m-1} + \frac{1}{m}w_j(\mathbf{z}_m^j)f(\mathbf{z}_m^j)L_j(\mathbf{z}_m^j). \tag{4.18}$$

The balance heuristic method has been shown to be close to optimal as the number of samples increases [117]. Therefore, in this work, we will use this technique of combining results from different biasing distributions.

## 4.2 Multicanonical Monte Carlo simulations

In this work we also used another statistical method, the multicanonical Monte Carlo (MMC) method, to compute low-probability tails of the pdfs. This method was proposed by Berg and Neuhaus to simulate first-order phase transitions in statistical physics [132] and recently has been applied to different areas of optical communications to model statistical effects of polarization mode dispersion, noise and nonlinearity [75], [118]–[120], [123], [125]–[131].

Why is the MMC algorithm so attractive? Like IS, MMC increases the number of events in the area of interest by biasing the distribution of the input random variables. With IS, we need to guess a set of biases based on some knowledge of the system properties. In our case, we know for example which bit strings contribute towards large values of collision-induced time shifts. Based on that, we can find efficient biasing distributions of the input bit strings. The major advantage of MMC is that it does not require *a priori* knowledge of these biased distributions. Instead, the algorithm iteratively converges to a set of biases that yields a uniform expected number of samples across the histogram.

Our goal is to compute the pdf of a quantity $X(\mathbf{z})$, which in our studies is either the received current or the time shift, as discussed in the previous section. We divide the sample space $\Omega$ into $K$ subspaces $\Omega_k$:

$$\Omega_k = \{\mathbf{z} \in \Omega \mid (k-1)\Delta X \leq X(\mathbf{z}) < k\Delta X\}, \tag{4.19}$$

where $\Delta X$ is the size of the histogram bin, and $k = 1, ..., K$. The function $f_k(\mathbf{z})$ will be the indicator function that is equal to one when $X(\mathbf{z})$ falls into the subspace $\Omega_k$. As in IS, the MMC algorithm uses the biasing distribution

$$P_k = \sum_{\mathbf{z}_s \in \Omega} f_k(\mathbf{z}_s) \frac{p(\mathbf{z}_s)}{p^*(\mathbf{z}_s)} p^*(\mathbf{z}_s). \tag{4.20}$$

The estimate $\hat{P}_k$ for the probability $P_k$ in Monte Carlo simulations is then

$$\hat{P}_k^* = \frac{1}{M} \sum_{m=1}^{M} f_k(\mathbf{z}_m) \frac{p(\mathbf{z}_m)}{p^*(\mathbf{z}_m)}, \tag{4.21}$$

where the samples $\mathbf{z}_m$ are drawn from the biasing distribution $p^*$ and $M$ is the total number of samples. The optimal biasing distribution that minimizes the variance of the estimate $\hat{P}$ is

$$p^*_{\text{opt}} = \frac{f_k(\mathbf{z})p(\mathbf{z})}{P_k}. \tag{4.22}$$

However, (4.22) implies that $P_k$ is unknown. Hence this result is of no practical use. While in standard IS, one attempts to use *a priori* knowledge to find a biasing distribution that is close to $p^*_{\text{opt}}$, MMC iterates over biasing distributions $p^{*\,j}$ that approach the optimal distribution. In the $j$-th iteration, the biasing distribution is given by

$$p^{*\,j} = \frac{p(\mathbf{z})}{c^j P_k^j}, \tag{4.23}$$

where $P_k^j > 0$, $\sum_{k=1}^{K} P_k^j = 1$, and $c^j$ is an unknown normalization constant that is chosen to satisfy the pdf normalization condition for $p^{*\,j}$. Quantities $1/P_k^j$ play the role of the weights in multiple IS. The $P_k^j$ are the probabilities, and they are updated on each iteration in such a way that the expected number of hits in each bin of the histogram is equal. After a large number of iterations, the $P_k^j$ approach $P_k$ [132].

The Metropolis algorithm is employed within each iteration [133]. This algorithm produces samples in a random walk whose limiting distribution is $p^*$ by using the following rule: In the $m$-th step, one goes from $\mathbf{z}_m = \mathbf{z}_a$ to $\mathbf{z}_{m+1} = \mathbf{z}_b = \mathbf{z}_a + \varepsilon^j \Delta\mathbf{z}$, where $\Delta\mathbf{z}$ is a random vector in the sampling space with no preferred direction, and $\varepsilon^j$ is a constant that may change from iteration to iteration depending on the acceptance ratio. The Metropolis algorithm usually performs best when the number of accepted and rejected steps is approximately equal. So, we adjust $\varepsilon^j$ to keep the acceptance ratio near 1/2. Each step is accepted with the probability $\pi_{ab}$,

$$\pi_{ab} = \min\left[\frac{p^{*j}(\mathbf{z}_b)}{p^{*j}(\mathbf{z}_a)},\ 1\right]. \tag{4.24}$$

The limiting distribution of the random walk with the transition probability (4.24) is $p^{j*}$ [133].

Before describing the iteration procedure, it is important to note that in the Metropolis random walk, one does not use the biased distributions $p^{j*}$ explicitly. From (4.23) it follows that the transition probability is given by

$$\pi_{ab} = \min\left[\frac{p(\mathbf{z}_b)P^j_{k_a}}{p(\mathbf{z}_a)P^j_{k_b}},\ 1\right]. \tag{4.25}$$

Thus one controls the random walk using the initial unbiased pdf $p(\mathbf{z})$ and the $P^j_k$. The limiting distribution of the random walk is $p^{j*}$.

In each iteration, one records a histogram $H^j$ of $X$ so that in each bin $k$

$$H^j_k = \sum_{m=1}^{M_j} f_k(\mathbf{z}_m) \tag{4.26}$$

is the number of samples that fall into $\Omega_k$. One sets $P^1_k = 1/M$ for the first iteration. At the end of each iteration one updates $P^j_k$ by first setting $P^{j+1}_1 = 1$. Any positive number can be used, as the quantities $P^{j+1}_k$ are normalized later. One uses then the recursion relations [132]

$$P^{j+1}_{k+1} = \frac{P^{j+1}_k P^j_{k+1}}{P^j_k}\left(\frac{H^j_{k+1}}{H^j_k}\right)^{\hat{g}^j_k}, \tag{4.27}$$

where

$$\hat{g}^j_k = \frac{g^j_k}{\sum_{l=1}^{j} g^l_k},\quad g^l_k = \frac{H^l_k H^l_{k+1}}{H^l_k + H^l_{k+1}}. \tag{4.28}$$

In (4.27), one sets $\hat{g}_k^j = 0$ when $g_k^j = 0$ and $g_k^l = 0$ when $H_k^l + H_{k+1}^l = 0$. Thus one uses statistical information from all previous iterations. Finally, one normalizes $P_k^{j+1}$ so that $\sum_{k=1}^{K} P_k^{j+1} = 1$. The combination of the Metropolis algorithm and the iterative procedure enables the algorithm to converge to the optimal set of biases without *a priori* knowledge of how to bias the distributions.

# 5. DETERMINISTIC METHOD FOR CALCULATION OF THE PDF OF COLLISION-INDUCED TIME SHIFT

In this chapter, we present a method for the calculating the pdf of the collision-induced time shift. We first derive the equations that describe the evolution of frequency and time shifts of a pulse in WDM RZ systems. A similar methodology is typically used in both soliton and quasi-linear systems [12], [19]–[21], [23], [38]–[44], [108], [134]. In this work, we use the semi-analytical approach presented in [12], which allows us to obtain an accurate solution for the collision-induced time shift numerically. One could also use an analytical asymptotic approach presented by Ahrens [41]. While it provides more insight into the nature of the timing jitter and describes well the asymptotic behavior, it is not accurate for small channel spacings and realistic pulse shapes. We then introduce the concept of the time shift function and discuss the shape and scaling of this function.

Next, we use the characteristic function method to derive the pdf of the collision-induced time shift. The time shift pdf can be approximated by a Gaussian function based on the central limit theorem. We show that the Gaussian function diverges significantly in the tails from the true pdf. We validate our results using IS and MMC simulations.

The pdf of the collision-induced time shift is the key component in the evaluation of the BER in the presence of nonlinear pattern-dependent distortion.

## 5.1 Calculation of collision-induced time shift

The starting point is the NLS equation:

$$\frac{\partial u(z,t)}{\partial z} + i\frac{\beta''(z)}{2}\frac{\partial^2 u(z,t)}{\partial t^2} - i\gamma|u(z,t)|^2 u(z,t) = g(z)u(z,t), \tag{5.1}$$

where $u(z,t)$ is the electric field envelope, $z$ is the physical distance, $t$ is the retarded time with respect to the center frequency of the signal, $\beta''$ is the local dispersion, $\gamma$ is the Kerr coefficient, $g(z)$ is the fiber loss and gain coefficient. While including higher-order dispersion poses no difficulty in principal, its effect on the system under study in this work is negligible and we set it to zero for simplicity. The signal we want to propagate through the fiber consists of several WDM channels, well-separated from each other in frequency,

$$u(z,t) = \sum_n u_n(z,t)\exp(-i\Omega_n t), \tag{5.2}$$

where $u_n(z,t)$ is the electric field envelope of the signal in the $n$-th channel and $\Omega_n$ is the frequency offset of the $n$-th channel. By substituting (5.2) into Eq. 5.1 and requiring all terms proportional to $\exp(i\Omega_n t)$ to vanish independently for different $n$, we obtain the following equation for the signal in the $n$-th channel,

$$\begin{aligned}\left[\frac{\partial u_n}{\partial z} + i\frac{\beta''}{2}\left(\frac{\partial^2 u_n}{\partial t^2} - 2i\Omega_n\frac{\partial u_n}{\partial t} - \Omega_n^2 u_n\right) - g u_n\right]\exp(-i\Omega_n t) = \\ \sum_{\substack{j,k,l: \\ \Omega_j-\Omega_k+\Omega_l=\Omega_n}} i\gamma u_j u_k^* u_l \exp\left[-i(\Omega_j - \Omega_k + \Omega_l)t\right]. \end{aligned} \tag{5.3}$$

Modern transmission fibers typically have large values of dispersion that significantly reduce four-wave mixing efficiency, so that the signal distortion due to the four-wave mixing is negligibly small compared to that of the cross-phase modulation. Hence, we only keep terms of the form $|u_k|^2 u_n$, which are responsible for self- and cross-phase modulation, and we obtain

$$\begin{aligned}\frac{\partial u_n}{\partial z} + i\frac{\beta''}{2}\left(\frac{\partial^2 u_n}{\partial t^2} - 2i\Omega_n\frac{\partial u_n}{\partial t} - \Omega_n^2 u_n\right) - gu_n \\ = i\gamma\left(|u_n|^2 + 2\sum_{k\neq n}|u_k|^2\right)u_n,\end{aligned} \tag{5.4}$$

Next, we make the variable transformation

$$t := t - \beta''\Omega_n z, \tag{5.5a}$$

$$u := u\exp\left(\frac{i}{2}\beta''\Omega_n^2 z\right), \tag{5.5b}$$

so that we change the time reference system to one that moves with the speed of light in the $n$-th channel and remove a common constant phase factor

$$\frac{\partial u_n}{\partial z} + i\frac{\beta''}{2}\frac{\partial^2 u_n}{\partial t^2} - gu_n = i\gamma\left(|u_n|^2 + 2\sum_{k\neq n}|u_k|^2\right)u_n. \tag{5.6}$$

Our goal is to compute the time shift of a pulse in the target channel $n$ that is a result of collisions with pulses in other frequency channels. We consider just one pulse in the target channel, called *the target pulse*, and we assume that all other channels have an arbitrary sequence of pulses. Let $\alpha_{kl}$ be the information bit contained in the $l$-th bit slot of the $k$-th channel, so that $\alpha_{kl} = 1$ or 0 depending on whether a mark or a space is transmitted. Then the field envelope in the $k$-th channel may be written as

$$u_k = \sum_l \alpha_{kl}u_{kl}, \tag{5.7}$$

where $u_{kl}$ is the field envelope of the pulse located in the $l$-th bit slot of the $k$-th channel. We are making an assumption that pulses in one channel are well-separated in time during the propagation so that for each channel $k$ we can use the approximation,

$$|u_k|^2 = \left|\sum_l \alpha_{kl} u_{kl}\right|^2 \approx \sum_l \alpha_{kl}|u_{kl}|^2. \tag{5.8}$$

Originally, this assumption was used in the theory of collision-induced timing jitter for solitons and was well justified because in soliton communications systems, pulses in one channel do not overlap. In modern quasi-linear RZ systems, a pulse overlaps with many of its neighbors due to a large dispersive spread. However, the approximation (5.8) is still reasonable because the peak power of a pulse reduces when it disperses thus reducing the nonlinear interactions. This approximation has proven to yield an accurate calculation of collision-induced timing jitter [12] and the pdf of the time shift [23]. Using (5.8), we obtain the following propagation equation for the target pulse, which we will denote as $u_T$,

$$\frac{\partial u_T}{\partial z} + i\frac{\beta''}{2}\frac{\partial^2 u_T}{\partial t^2} - g u_T = i\gamma\left(|u_T|^2 + 2\sum_{k,l}\alpha_{kl}|u_{kl}|^2\right)u_T. \tag{5.9}$$

We define the central time and frequency of the target pulse as

$$\begin{aligned} T_c &= \frac{1}{E}\int_{-\infty}^{\infty} t|u_T|^2 dt, && \text{(5.10a)}\\ \Omega_c &= \frac{1}{E}\int_{-\infty}^{\infty} \mathrm{Im}\left(\frac{\partial u_T^*}{\partial t}u_T\right)dt, && \text{(5.10b)}\end{aligned}$$

where $E = \int_{-\infty}^{\infty}|u_T|^2 dt$. Using these definitions and (5.9), we obtain the following dynamic equations for the central time and frequency,

$$\begin{aligned} \frac{dT_c}{dz} &= \beta''\Omega_c && \text{(5.11a)}\\ \frac{d\Omega_c}{dz} &= -\frac{\gamma}{E}\int_{-\infty}^{\infty} t|u_T|^2\frac{\partial}{\partial t}\left(2\sum_{k,l}\alpha_{kl}|u_{kl}|^2\right)dt. && \text{(5.11b)}\end{aligned}$$

Then the collision-induced time shift $T_{\text{total}}$ after a propagation distance $L$ can be represented as

$$T_{\text{total}} = \sum_{k,l} \alpha_{kl}\tau_{kl}, \tag{5.12}$$

where

$$\tau_{kl} = \int_0^L \Omega_{kl}(z)\beta''(z)dz \tag{5.13a}$$

$$\Omega_{kl}(z) = -\int_0^z \frac{2\gamma}{E_T(x)} \int_{-\infty}^{\infty} |u_T(x,t)|^2 \, \frac{\partial |u_{kl}(x,t)|^2}{\partial t} \, dt \, dx. \tag{5.13b}$$

When the higher-order dispersion is zero, the pulse shape in all the channels is identical so that

$$u_{kl}(z,t) = u_T\left(z, t - \int_0^z 2\pi\Delta f_k \beta''(x)dx + T_{\text{bit}}l\right), \tag{5.14}$$

where $\Delta f_k$ is the frequency spacing between the $k$-th and the target channel, $T_{\text{bit}}$ is the bit period, and $l = 0$ corresponds to the bit slot of the target pulse. From a practical standpoint, this approximation enables us to reduce the total computation time by computing the pulse shape once for the target pulse and then using it for all other channels.

The quantities $\Omega_{kl}$ and $\tau_{kl}$ are the frequency and time shifts of the target pulse due to its collision with a single pulse located in the $l$-th bit slot of the $k$-th channel. As a result of using approximation (5.8), the time shift of a target pulse that occurs due to its interaction with an arbitrary pulse pattern in all the neighboring channels can be represented as a superposition of pairwise collisions. Thus the set of time shifts due to individual pairwise collisions $\tau_{kl}$ can be used to determine the time shift of a target pulse for an arbitrary sequence of pulses through Eq. 5.12.

This assumption vastly reduces the computer time that is required to calculate the distribution of time shifts and the impairments induced in WDM RZ systems [12], [23].

It also leads to important insights into the origin of timing jitter and to the development of ways to reduce it [40], [44], [134], [135]. We call $\tau_{kl}$ *the time shift function*, and in the next section, we will discuss its properties and asymptotic behavior.

## 5.2 Time shift function

In this section, we calculate and analyze the time shift function for collisions between pairs of pulses in different channels in a prototypical WDM RZ system with dispersion management and pre- and post-compensation. Once the time shift function is known, the impairments due to collision-induced timing jitter can be rapidly determined. We characterize the shape of this function and determine how it scales with the initial pulse separation in time and channel separation in wavelength. At the end of the section, we apply it to the calculation of the worst-case time shift. Determining these worst-case time shifts is important because the time shift distribution is approximately Gaussian up to about 1/2 of these times [23].

### 5.2.1 Shape of the time shift function

We write the time shift function (5.13a) as $\tau(\Delta f, l)$, where $l$ is the initial offset of a bit slot in a pump channel that is separated by $\Delta f$ from the probe channel.

In Fig. 5.1, we plot the time shift function using (5.13a) for the two pump channels with $\Delta f = \pm 100$ GHz. We first note that $\tau(\Delta f, l) = -\tau(-\Delta f, -l)$. This symmetry is exact if there is no higher-order dispersion, as has been reported previously for soliton systems [134]. We also observe that as $l$ increases from $-\infty$ in the pump channel with $\Delta f = -100$ GHz, the time shift starts from zero, becomes positive, then changes sign in the neighborhood of $l = -8$, becoming negative, after which it becomes positive again, and finally decays to zero. We have observed this behavior in a variety of systems, and it has been shown to be generic [41]. We now explain its physical origin.

In long-haul systems with dispersion management, two pulses in different WDM channels move rapidly back and forth with respect to each other, resulting in multiple collisions, each of which we refer to as a *micro-collision*. If, as is typically the case, the average map dispersion is non-zero, two pulses gradually pass through each other while experiencing multiple micro-collisions. We refer to this process as a *macro-collision*. Consider now the collision dynamics of the target pulse with pulses in the pump channel with $\Delta f = -100$ GHz. Figure 5.2(a) shows the central times for the pulses with $l = 1$, 11, and $-11$. The evolution of the frequency and time shifts of the target pulses due to collisions with these three pulses is shown in Figs. 5.2(b) and 5.2(c). The pulse with $l = 1$ undergoes a complete macro-collision, while the pulses with $l = -11$ and 11 undergo incomplete macro-collisions that occur at the beginning and the end of the system respectively.

First, we consider a complete macro-collision, as in the case $l = 1$. In its initial phase, the two interacting pulses undergo incomplete micro-collisions. Since the central time of the pump pulse relative to the target pulse is positive, by Eq. 5.13b the frequency shift after each incomplete micro-collision decreases. In the second phase of the macro-collision, the pulses undergo complete micro-collisions with no significant change in the frequency shift. Finally, the pulses undergo incomplete micro-collisions again, but now the frequency shift increases, and is nearly zero when the macro-collision is over. Since the frequency shift is negative during most of the macro-collision and since the dispersion accumulated during the macro-collision is negative, the resulting time shift is negative by Eq. 5.13a. We note that the pre-

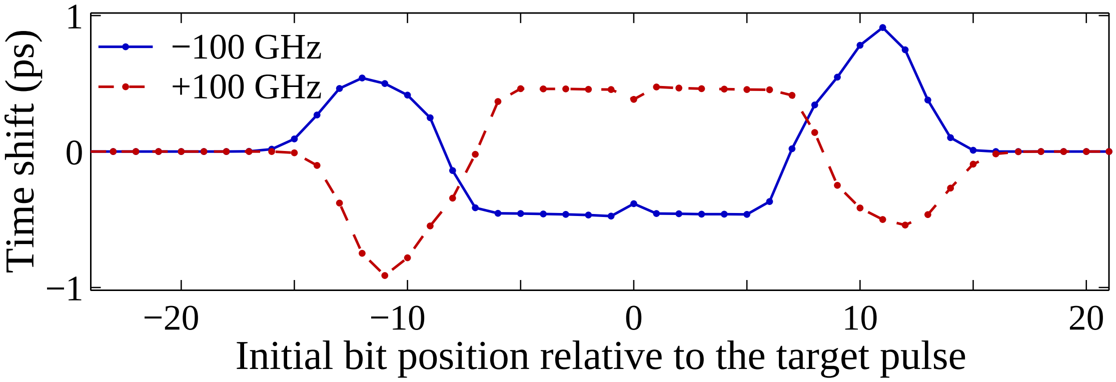

**Fig. 5.1:** Time shift function for two pump channels.

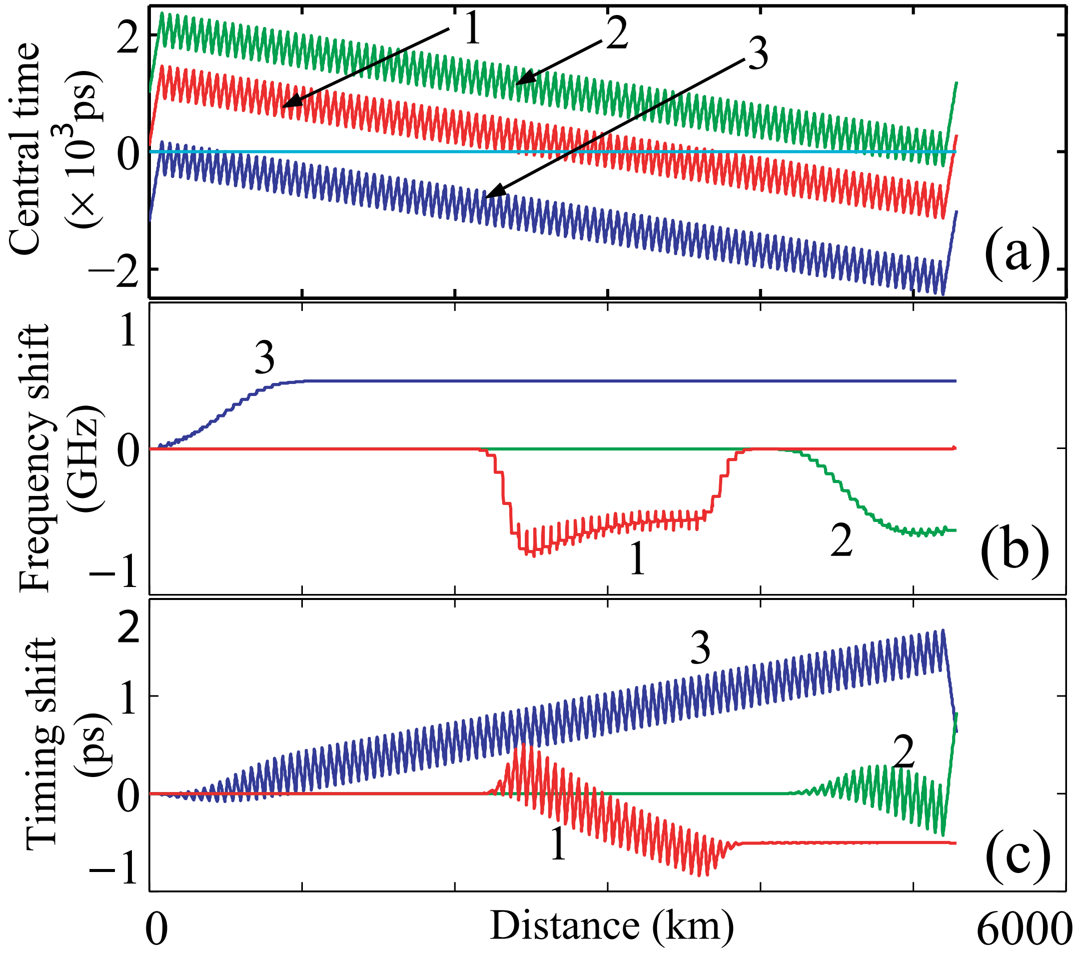

**Fig. 5.2:** Collision dynamics for three different pulses.

and post-compensation fibers have almost no effect on the time shift in a complete macro-collision.

We next consider an incomplete macro-collision at the end of the system, as in the case $l = 11$. The frequency shift is negative because the pulses only interact during the initial phase of a macro-collision. The dispersion accumulated during the macro-collision, including the post-compensation, is positive. Hence, the time shift is positive. Finally, for an incomplete macro-collision at the start of the system, as in the case $l = -11$, the frequency shift is positive because the pulses only interact during the final phase of a complete macro-collision. After pulses have separated, the frequency shift is constant and positive. Since the dispersion accumulated from the

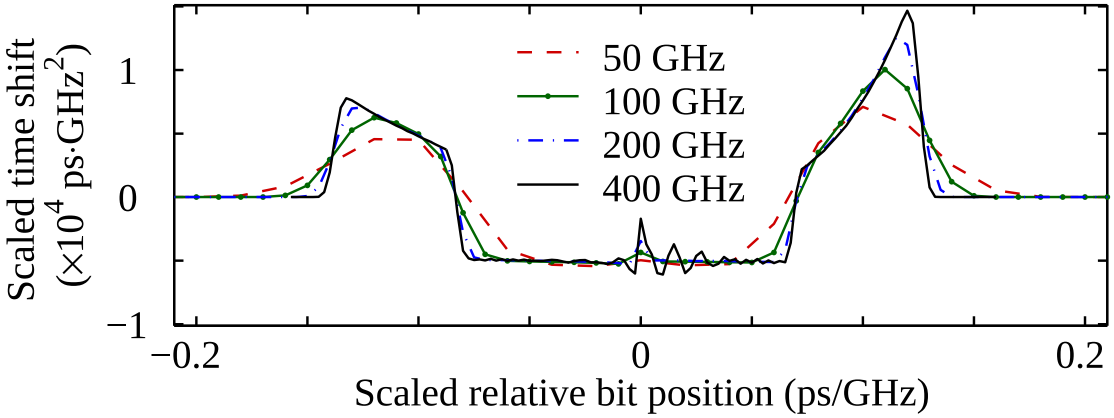

**Fig. 5.3:** The scaled time shift function.

end of the collision to the end of the system is negative, the resulting timing shift is positive.

We found that the details of the dispersion map and transmission distance do not qualitatively affect the shape of $\tau(\Delta f, l)$ as long as the average map dispersion is large enough so that the length of a macro-collision is shorter than the system length, which is usually the case in realistic RZ systems. When the system length is larger then the collision length, we still can observe a similar behavior. However, in this case, the details of the map can significantly affect the shape of $\tau(\Delta f, l)$ [39]. We emphasize that a large average dispersion reduces the impact of nonlinearity, so that the parameter regime considered here is more important in practice.

### 5.2.2 Scaling of the time shift function

We now discuss how the time shift function scales. An asymptotic analysis [39], [113] predicts $\tau(\Delta f, l) = a^2 \tau(a\Delta f, al)$, where $a$ is a scaling parameter. In Fig. 5.3, we show $\tau(\Delta f, l)$ for channel spacings in the range $\Delta f = 50$ GHz – 400 GHz. We have scaled $\tau$ by $(\Delta f)^2$ and $l$ by $(\Delta f)^{-1}$. The scaling relation holds approximately, but there are significant deviations. From a physical standpoint, the scaling with the offset $l$ occurs because the velocity difference between a pulse in a pump channel and the target pulse is proportional to $\Delta f$. In the scaling for $\tau$, the time shift is mainly determined by the accumulation of residual frequency shifts due to incomplete micro-collisions, as seen in Fig. 5.2, while the residual time shift is small. The frequency shift due to each micro-collision is determined by the maximum overlap of pulses (5.13b), which is independent of $\Delta f$, and the interaction length, which is proportional to $(\Delta f)^{-1}$. Hence the frequency shift due to each micro-collision scales as $(\Delta f)^{-1}$. Since the number of incomplete micro-collisions is proportional to $(\Delta f)^{-1}$, the total time shift scales as $(\Delta f)^{-2}$. For this scaling to hold, the number of incomplete micro-collisions in each macro-collision should be large, and the maximum separation of the two pulses in each micro-collision should be large compared to the pulse durations. When $\Delta f$ is large, the first condition is violated, and oscillations are visible when $\Delta f = 400$ GHz. When $\Delta f$ is small, the second condition is violated, blurring the distinction between complete and incomplete macro-collisions, so that the transition between the two occurs more gradually, as seen when $\Delta f = 50$ GHz.

Given $\tau(\Delta f, l)$, we determine the worst-case time shift from Eq. 5.12 by setting $\alpha_{kl} = 1$ when $\tau(\Delta f_k, l) \equiv \tau_{kl} > 0$ and $\alpha_{kl} = 0$ when $\tau(\Delta f_k, l) < 0$ or vice versa. From the relation $\tau(\Delta f, l) = -\tau(-\Delta f, -l)$, we find that the worst case corresponds to an opposite choice of 1's and 0's in channels with $\pm\Delta f$, just as Xu, *et al.* [134], discovered in the case of solitons. From the scaling relation $\tau(\Delta f, l) = a^2\tau(a\Delta f, al)$, we would also conclude with Xu, *et al.* [134] that the maximum time shift increases logarithmically with the number of channels, as seen in Fig. 5.4, where we have fit a logarithmic function of the form $a+b\log N$. This scaling relation breaks down when all micro-collisions are complete, which corresponds to $N > 16$ for the system we studied. However, the residual time shift due to micro-collisions still scales as $(\Delta f)^{-2}$, since the frequency excursion is proportional to $(\Delta f)^{-1}$ and so is the length of the micro-collision. Hence, as seen in Fig. 5.4, the maximum time shift continues to increase logarithmically beyond $N = 16$. However, the scaling of the time shift function only

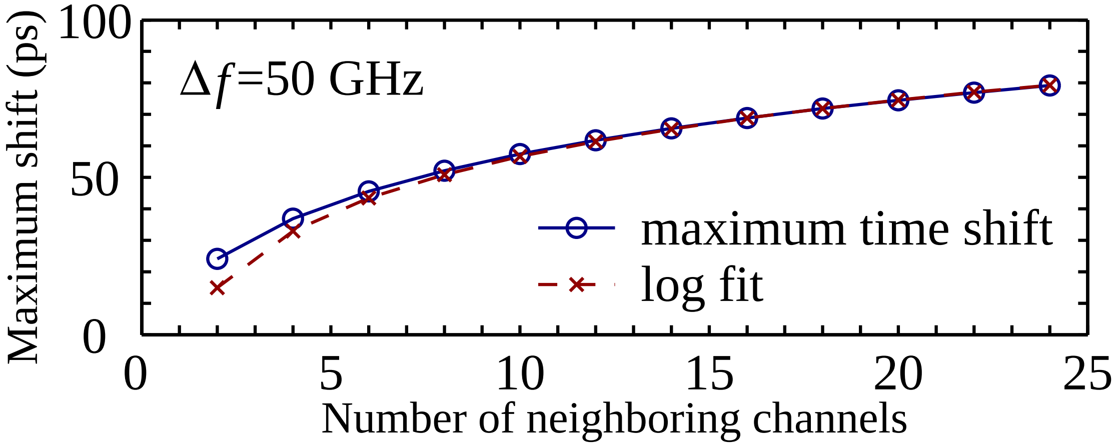

**Fig. 5.4:** Worst-case time shift vs. number of channels $N$.

holds approximately, and in practice it is better and still computationally rapid to use the exact time shift function when calculating the distribution function for the timing jitter and its cutoff.

## 5.3 Probability density function of the time shift

In this section, we calculate the probability density function of the collision-induced time shift using both a Gaussian approximation and the exact theory. Different frequency channels can be synchronized, meaning that the bit slots of all the channels are aligned at the input so that the delay between the signals in different channels is fixed. We will refer to this situation as the *synchronous channels* case. Alternatively, the clock in each channel can drift randomly with time so that the delay between the signals of different channels is random. We call this case the *asynchronous channels* case. The treatment of the synchronous case depends on the exact relative position of the clock in different channels. We will consider the case in which the bit slots in all the channels are aligned at $t = 0$. If they are aligned differently, it can be taken into account by introducing a translation of the time shift function. In the asynchronous case, we treat the drift of the channel clock position as an independent random process.

### 5.3.1 Synchronous channels

We assume that the transmitted data is random so that the $\alpha_{kl}$ are independent, identically-distributed random variables, each having probability 1/2 of being 1 or 0. Thus the total shift of the target pulse $T_{\text{total}}$ is a random variable, which is a linear combination of independent binary random variables. The number of terms in the sum (5.12) is finite due to the finite number of pulses with which the target pulse collides. On the other hand, the number of collisions is large so that based on the central limit theorem, it is reasonable to assume that $T_{\text{total}}$ is Gaussian-distributed. In this case, it is sufficient to know its mean $\mu_T$ and variance $\sigma_T^2$. In the absence of higher-order dispersion, the collision process is antisymmetric with respect to frequency and time inversion, as discussed in Ref. [136] and in Section 5.2, so that $\tau_{kl} = -\tau_{-k\,-l}$ and hence

$$\begin{aligned} \mu_T &= \sum_{k,l} \tau_{kl} = 0, && (5.15a) \\ \sigma_T^2 &= \left\langle T_{\text{total}}^2 \right\rangle = \frac{1}{4} \sum_{k,l} \tau_{kl}^2. && (5.15b) \end{aligned}$$

The pdf of $T_{\text{total}}$ in the Gaussian approximation is then given by

$$p_{T,\text{Gauss}}(t) = \frac{1}{\sqrt{2\pi\sigma_T^2}} \exp\left(\frac{-t^2}{2\pi\sigma_T^2}\right). \tag{5.16}$$

Alternatively, the pdf of $T_{\text{total}}$ can be computed using the characteristic function [23] $w(\xi)$, which is given by

$$w(\xi) = \langle \exp\left(i\xi T_{\text{total}}\right) \rangle = \left\langle \exp\left(i\xi \sum_{k,l} \alpha_{kl}\tau_{kl}\right) \right\rangle,$$

where $\langle\cdot\rangle$ denotes the statistical average so that for a set of discrete random variables $x_1, \ldots, x_n$, with a joint probability mass function $p(x_1, \ldots, x_n)$, and for an arbitrary function $F(x_1, \ldots, x_n)$,

$$\langle\ F(x_1, \ldots, x_n)\rangle = \sum_{k_1,\ldots,k_n} p(x_{k_1}, \ldots, x_{k_n})\, F(x_{k_1}, \ldots, x_{k_n}).$$

Since the random variables $\alpha_{kl}$ are independent and the probability that $\alpha_{kl}$ equals either 1 or 0 is 1/2, we find

$$w\left(\xi\right) = \prod_{k,l} \left\{\frac{1}{2}\left[1 + \exp\left(i\xi\tau_{kl}\right)\right]\right\}.$$

The pdf of the time shift is simply the inverse Fourier transform of the characteristic function,

$$p_{T,\text{Char}}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} w\left(\xi\right) \mathrm{e}^{-i\xi t} d\xi. \tag{5.17}$$

### 5.3.2 Asynchronous channels

In order to account for random drift of the clock position in each channel, we introduce a variable time delay $t_k$ in Eq. (5.14):

$$u_{kl}(z,t) = u_T\left(z, t - \int_0^z 2\pi\Delta f_k \beta''(x)dx + T_{\text{bit}}l + t_k\right).$$

Using the same formalism, we write the characteristic function

$$w(\xi) = \langle \exp\left(i\xi T_{\text{total}}\right)\rangle = \left\langle \exp\left(i\xi \sum_{k,l} \alpha_{kl}\tau_{kl}(t_k)\right)\right\rangle,$$

where $\tau_{kl}(t_k)$ can be viewed as a random variable, which is independent of the $\alpha_{kl}$, assuming independence of the channel clock drift and the data distribution. The average is taken over over the ensemble of both $\alpha_{kl}$ and $t_k$. We then find

$$w\left(\xi\right) = \prod_{k,l}\left\langle \frac{1}{2}\left[1 + \exp\left(i\xi\tau_{kl}(t_k)\right)\right]\right\rangle,$$

where the average is taken over the sample space of $t_k$. To calculate the characteristic function, it is necessary to know the pdf of the time delays $t_k$. For simplicity, we assume that $t_k$ are uniformly distributed on $[0, T_{\text{bit}}]$, so that

$$w\left(\xi\right) = \prod_{k,l} \frac{1}{T_{\text{bit}}}\int_0^{T_{\text{bit}}} \frac{1}{2}\left[1 + \exp\left(i\xi\tau_{kl}(t_k)\right)\right]dt_k. \tag{5.18}$$

The pdf of the time shift is obtained from the characteristic function (5.18) using the inverse Fourier transform (5.17).

## 5.4 Validation

We used the standard IS method to validate the deterministic method for calculating the pdf of collision-induced time shift. We also used the MMC method to verify that the results from IS simulations are correct.

In order to use IS, we have to choose the biasing distributions, as discussed earlier in Chapter 4. In this case, we are interested in the pdf of the collision-induced time shift, and we bias the distributions towards large values of the time shift, using our knowledge of the time shift function (5.13a).

Suppose, we have a set of random variables $x_m$ with the distribution,

$$p(x_m = \sigma_m) = \frac{1+pm}{2}, \;\; p(x_m = -\sigma_m) = \frac{1-pm}{2}. \tag{5.19}$$

We want to bias the distributions of $x_m$ in such a way that the sum $z = \sum_m x_m$ has a mean near a desired value $z_0$. For the sake of simplicity, let us assume that $z$ is approximately Gaussian random variable with the mean $z_0$ and standard deviation $S$:

$$\begin{aligned} p(z) &= \frac{1}{\sqrt{2\pi S}} \exp\left[-\frac{(z-z_0)^2}{2S}\right], \\ S &= \sum_m \sigma_m^2. \end{aligned} \tag{5.20}$$

We then use Bayes rule,

$$p(x_m \mid x) = \frac{p(x_m)p(x-x_m)}{p(x)}, \tag{5.21}$$

so that

$$p(x_m = \sigma_m \mid z) = \frac{g_+}{g_+ + g_-}, \tag{5.22}$$

where

$$g_{\pm} = \frac{1}{\sqrt{2\pi(S-\sigma_m^2)}} \exp\left[-\frac{(z \mp \sigma_m)^2}{2(S-\sigma_m^2)}\right]. \tag{5.23}$$

Then the expected value of $x_m$ is given by

$$E(x_m) = \sigma_m\, p(x_m = \sigma_m \mid z) - \sigma_m p(x_m = -\sigma_m \mid z) = \sigma_m \tanh\left(\frac{\sigma_m z}{S-\sigma_m^2}\right). \tag{5.24}$$

On the other hand, from Eq. (5.19) we have

$$E(x_m) = \sigma_m\, \frac{1+p_m}{2} - \sigma_m\, \frac{1-p_m}{2} = \sigma_m\, p_m. \tag{5.25}$$

From Eqs. (5.24) and (5.25) it follows that

$$p_m = \tanh\left(\frac{\sigma_m z}{S-\sigma_m^2}\right). \tag{5.26}$$

Now we transform the set of variables $x_m$ to the set $\alpha_{kl}\tau_{kl}$ using the mapping

$$\begin{aligned} \beta_{kl} &= 2\alpha_{kl} - 1, \\ x_m &= \frac{1}{2}\beta_{kl}\tau_{kl}. \end{aligned} \tag{5.27}$$

The probability mass function for $\alpha_{kl}$ becomes

$$\begin{aligned} p(\alpha_{kl} = 1) &= \frac{1+p_{kl}}{2}, \\ p(\alpha_{kl} = 0) &= \frac{1-p_{kl}}{2}, \end{aligned} \tag{5.28}$$

where

$$p_{kl} = \tanh\left(\frac{\frac{1}{2}\tau_{kl}T_{\mathrm{goal}}}{\frac{1}{4}\sum_{k,l}\tau_{kl}^2 - \tau_{kl}^2}\right) \tag{5.29}$$

and $T_{\mathrm{goal}}$ is the target value of the mean of the total time shift.

Despite the use of approximation (5.20) that the total time shift is Gaussian-distributed, our simulations show that the choice of biases (5.29) is efficient for both estimating the pdfs of both collision-induced time shift and the received current.

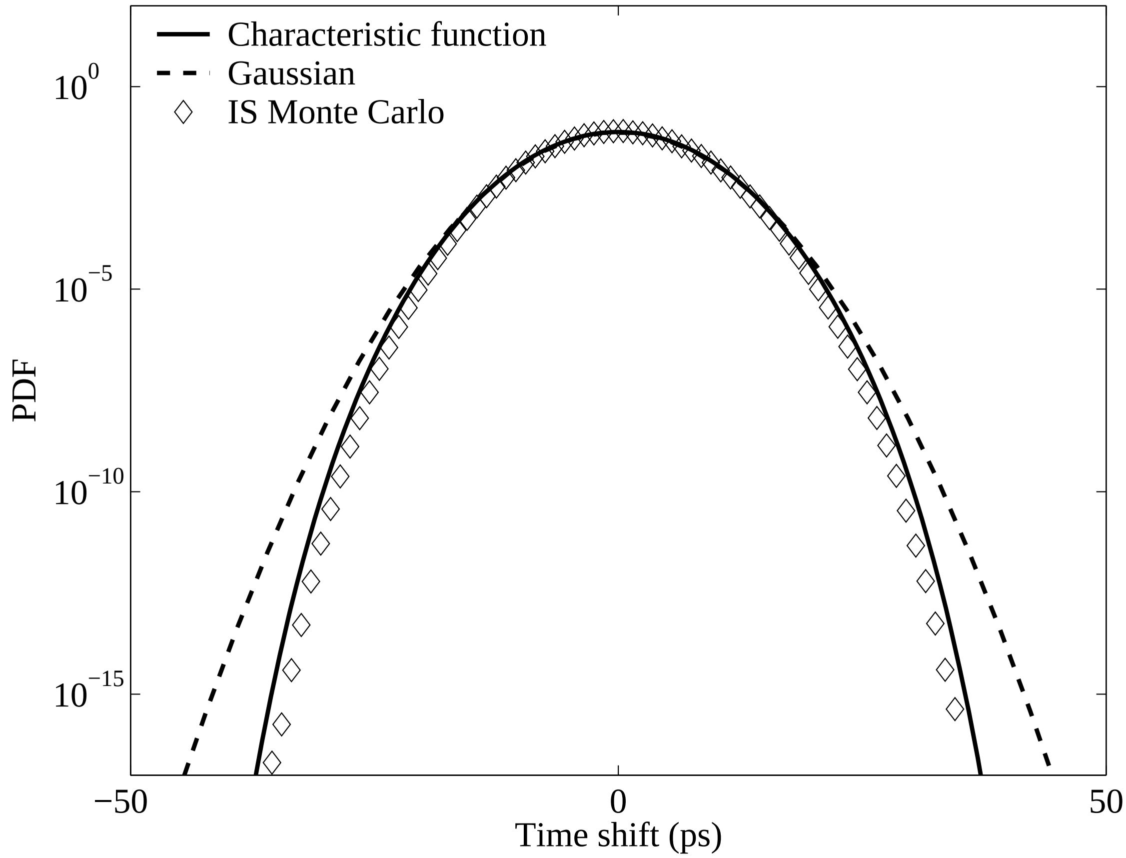

**Fig. 5.5:** Probability density function of the collision-induced time shift.

Figure 5.5 shows the pdf of the time shift obtained using the characteristic function method (5.17), the Gaussian approximation (5.16) and IS simulations. In the IS simulations we used the values of $T_{\text{goal}}$ in (5.29) from –50 ps to 50 ps with a 5 ps interval. We calculated the error of the pdf estimate from IS using (4.15) and it did not exceed 10% for all values of the time shift. We have also verified that the results of IS simulations agree with those of MMC simulations. The results of the reduced deterministic method that employs characteristic function formalism are in excellent agreement with the full statistical model, in which we did not make any simplifications to the propagation equations. This result suggests the validity of the simplifying assumption that the pulse collisions are additive. By contrast, under the same assumptions, the Gaussian approximation to the pdf results in a poor fit of the true pdf. The Gaussian function closely matches the Monte Carlo histogram only near the top of the pdf, which is consistent with the central limit theorem. However, since the number of pulse collisions during the propagation is finite, there exists a worst-case time shift so that time shifts that are larger than this maximum time shift have zero probability. As a result, the Gaussian curve deviates significantly from the true distribution in the tails of the pdf.

We note that even if the number of pulse collisions were infinite or significantly larger in the presence of the hundreds of channels, the results of the Gaussian approximation would still be inaccurate in the tails since one of the conditions of the central limit theorem does not hold. In particular, for a normalized sum

$$Z_n = \frac{1}{s_n} \sum_{k=1}^{n} X_k \tag{5.30}$$

of independent random variables $X_k$ with zero mean and and variance $\sigma_k^2$ and

$$s_n^2 = \sum_{k=1}^{n} \sigma_n^2 \tag{5.31}$$

to converge to the standard normal pdf, it is required that for a given $\varepsilon > 0$ there exists an $n$ such that [137]

$$\sigma_k < \varepsilon s_n, \;\; k = 1, ..., n. \tag{5.32}$$

In our case, the time shifts $\tau_{kl}$ given by (5.13a) are proportional to the square of the wavelength separation, or, equivalently, to the channel index $k$ if the channel count starts from the target channel. Hence the variance for each individual random variable $\alpha_{kl}\tau_{kl}$ is proportional to $k^4$. Since the number of bits in a given channel that collide with the target pulse is proportional to $k$ due to the dispersive walkoff, the total variance for $n$ channels is proportional to $\sum_{k=1}^{n} 1/k^3$, which is limited by a constant for any number of channels $n$. Thus condition (5.32) of the central limit theorem does not hold and the limiting distribution of the time shift is not necessarily Gaussian.

# 6. CALCULATION OF THE PULSE AMPLITUDE DISTORTION AND THE BIT ERROR RATIO

In this chapter, we use the deterministic method for calculating the pdf of the collision-induced time shift to obtain the pdf of the received current, and we validate the result with a full statistical model based on IS simulations. When we have a single pulse in the target channel, the agreement between the reduced deterministic method and the full statistical method is excellent. With multiple pulses in the target channel, it is also important to account for the amplitude jitter due to inter- and intra-channel interactions. We introduce a method for calculating the pdf of the nonlinearly-induced current variation that is not due to the timing jitter. Accounting for both timing and amplitude jitter, we can achieve a very good agreement with the IS simulations. In order to calculate the BER, we combine these deterministic techniques for calculating the nonlinear penalty with a model for noise-induced signal distortion that assumes that the noise is additive, white, and Gaussian. We calculate the pdfs of the current in the marks and spaces and from that determine the BER using an approach described by Forestieri [138].

## 6.1 Probabilistic characterization of the nonlinearly-induced pulse distortion

### 6.1.1 Application of the reduced time shift method

In order to calculate the distortion of the received current signal that is due to timing jitter, we use a simplified method, in which we calculate the pulse shape $i(t)$ at the receiver using a full propagation model and a receiver model that includes an optical filter, an ideal square-law photodetector, and an electrical filter as described in Chapter 2. We then use this pulse shape to determine the value of the sampled current $I$ given a time shift $\Delta T$ by using the expression

$$I(\Delta T) = i(T_0 - \Delta T), \tag{6.1}$$

where $T_0$ is the central time of the pulse, as illustrated in Fig. 6.1. We then obtain the pdf of the current using the cumulative distribution function of the time shift $F_{\Delta T}(t) = \int_{-\infty}^{t} p(\tau)d\tau$, where $p(\tau)$ is the pdf of the time shift,

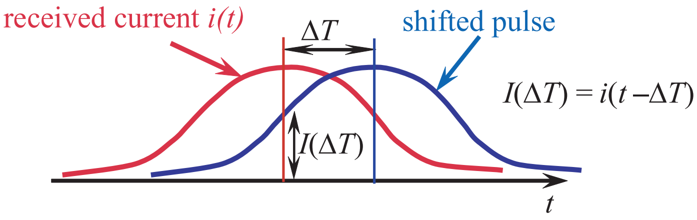

**Fig. 6.1:** Conversion of time shift to the current distortion.

$$F_I(x) = \Pr[i(t) < x] = \Pr(\Delta T < T_1 \cup \Delta T > T_2) = F_{\Delta T}(T_1) + 1 - F_{\Delta T}(T_2), \tag{6.2}$$

where $T_1 < T_2$ are the solutions of the equation

$$i(t - T_0) = x. \tag{6.3}$$

The pdf of the received current $p_I(x)$ is then given by

$$p_I(x) = dF_I(x)/dx. \tag{6.4}$$

We also ran a set of IS simulations, in which we numerically solved the NLS equation with different bit strings. We kept only one pulse in the center channel and randomly varied the bit strings in the neighboring channels. To sample the tails of the distribution efficiently, we biased the distribution of the bit strings towards the large time shifts using (5.29). We have verified that MMC simulations are consistent the IS results, which lends credence to both simulations. In particular, it indicates that we have not left out a region in the sample space that would affect the distribution. Biasing towards large time shifts has proven to be very efficient in calculating the pdf of the received current since the dominant nonlinear effect in our prototypical undersea system is collision-induced timing jitter. An advantage of IS relative to MMC for our problem is that different biasing distributions can be used independently of each other. By contrast, the MMC algorithm is iterative. Also, in contrast to MMC, we use the same set of IS simulations to calculate the pdf of both the time shift and the received current.

The results of the pdf calculation using (6.4) and IS are shown in Fig. 6.2. Statistically, we obtained the pdf of the current using the same IS simulations that we used to obtain the pdf of the time shift as described in Sec. 5.4. We estimated the statistical error of the pdf obtained with IS using (4.15) and it was less than 10% for all values of the current shown in the plot. The agreement between the reduced deterministic approach and the full statistical model is excellent near the low-current tail of the pdf. The two methods disagree near the maximum of the pdf since we did not take into account nonlinearly-induced amplitude jitter in our reduced model. It is necessary to consider the amplitude jitter to obtain the agreement when the

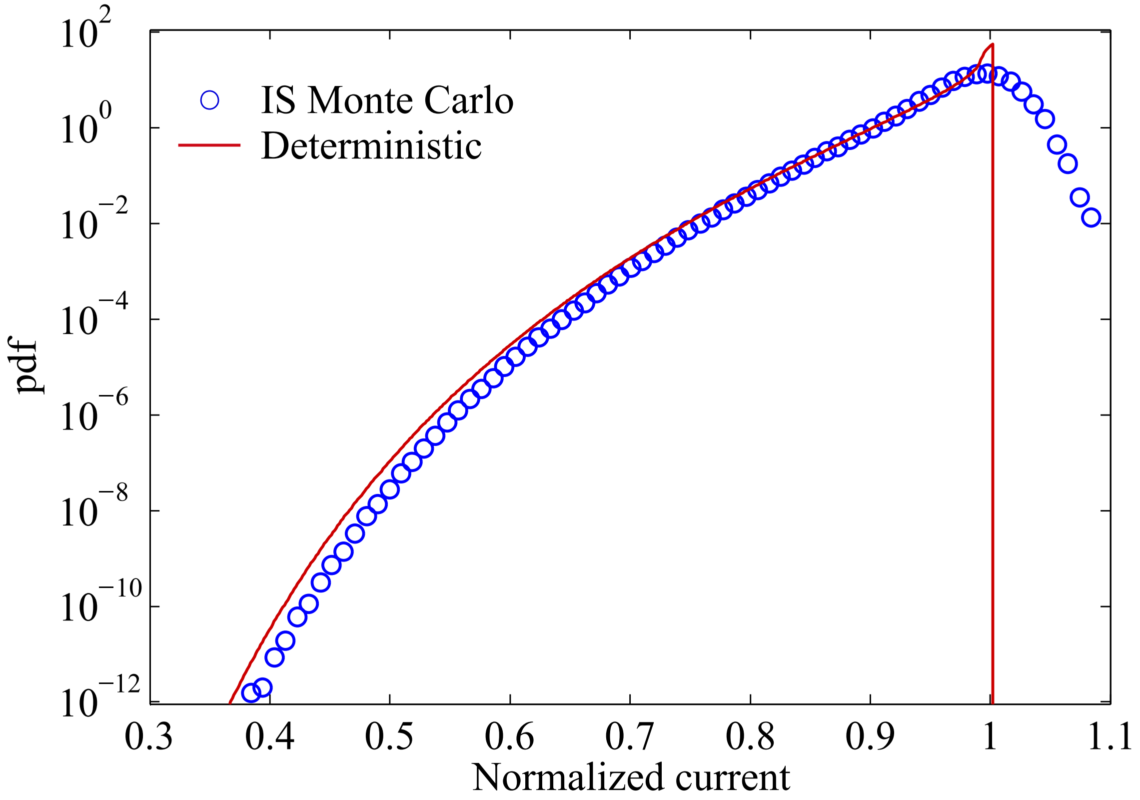

**Fig. 6.2:** Probability density function of the current in the marks due to the nonlinear distortion with a single pulse in the target channel.

normalized current is close to 1.0, since with timing jitter can only decrease the pulse amplitude, leading to a sharp cutoff at 1.0 as seen in Fig. 6.2. In practice, this difference is unimportant since transmission errors are caused by the low-current end of the pdf.

By contrast to the case of a single pulse in the target channel, which we just considered, the nonlinearly-induced amplitude jitter becomes important in the low-current tail of the pdf when we consider multiple pulses in the target channel.

### 6.1.2 Multipulse interactions and the nonlinearly-induced amplitude jitter

Up to this point, we treated pulses of the same frequency channel as if they do not overlap. In reality, in the system under study, each pulse overlaps with a maximum of four of its neighbors due to dispersive spreading. This intra-channel pulse interaction combined with inter-channel nonlinear crosstalk leads to an increase in the amplitude jitter that must be accounted for in the target channel, in addition to the timing jitter, in order to obtain accurate results. Hence we treat the electric field in the target channel as a sum of the electric fields of individual pulses rather than simply adding powers as we did previously (5.8), so that

$$|u_0|^2 = \left|\sum_l \alpha_{0l} u_{0l}\right|^2 \neq \sum_l \alpha_{0l} |u_{0l}|^2. \tag{6.5}$$

In the other channels, we continue to neglect pulse overlap. Since in our system only five pulses in the same channel overlap due to dispersion, we consider a pseudo-random bit sequence (PRBS) of length $2^5 = 32$ bits in the target channel that contains all patterns of five bits. We then treat this PRBS as a superpulse and consider a two-body collision of this superpulse with a single pulse in a neighboring channel.

To determine the effect of the collision of the pulse in the $l$-th bit of the $k$-th channel with the superpulse, we numerically solve the NLS equation, for which the input is the superpulse in the target channel and a single pulse in the $l$-th bit of the $k$-th channel. We then calculate the received current $I_{kl}(t)$ in the target channel after the electrical filter. We repeat this numerical procedure for all $k$ and $l$.

In the next step, we remove the time shift that we accounted for previously using (5.13a) as we are only calculating the amplitude distortion at this point. Since pulses at the receiver are well-separated, and the time shifts $\tau_{kl}$ are small, we can easily remove the time shifts by translating individual pulses by a corresponding time determined by the pre-calculated time shift function $\tau_{kl}$ (5.13a).

Finally, we determine the received current distortion due to the two-body collisions relative to the unperturbed solution. As an unperturbed solution, we calculate the value of the current $I_T(t)$ of a single target channel with the same PRBS of length 32 in it, in the absence of the neighboring channels. We then obtain the current distortion $\delta I_{kl}$ using the expression

$$\delta I_{kl}(t) = I_{kl}(t) - I_T(t). \tag{6.6}$$

We assume that for an arbitrary bit pattern, the total distortion $\delta I(t)$ can be represented as a sum of the individual contributions $\delta I_{kl}$ from pairwise interactions, so that

$$\delta I(t) = \sum \alpha_{kl} \delta I_{kl}(t), \tag{6.7}$$

where $\alpha_{kl} = 1$ or 0.

In order to obtain the pdf of the amplitude deviation $\delta I(t)$, we apply the characteristic function method, as we did for the timing jitter

$$v(\xi, t) = \langle \exp\left[i\xi \delta I(t)\right] \rangle = \left\langle \exp\left( i\xi \sum_{k,l} \alpha_{kl} \delta I_{kl}(t) \right) \right\rangle, \tag{6.8}$$

where $\langle \cdot \rangle$ denotes the statistical average taken over the ensemble of all bit patterns. Using the independence of random variables $\alpha_{kl}$ an assumption that the probability of $\alpha_{kl}$ being equal to either 1 or 0 is 1/2, we obtain

$$v\left(\xi, t\right) = \prod_{k,l} \frac{1}{2} \left\{ 1 + \exp\left[i\xi \delta I_{kl}(t)\right] \right\}. \tag{6.9}$$

Then the pdf of current is simply the inverse Fourier transform of the characteristic function,

$$p_{\delta I}(I, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} v\left(\xi, t\right) \mathrm{e}^{-i\xi I} d\xi. \tag{6.10}$$

In order to obtain the distribution of the current due to nonlinear distortion, we first average the pdf (6.4) of the current at the center of the pulse over all bits. Then we convolve the distribution (6.4) due to collision-induced timing jitter and (6.10) due to the amplitude distortion assuming that the two processes are independent.

The resulting pdf is shown in Fig. 6.3 in comparison with the pdf obtained using IS simulations and the pdf when there is a single pulse in the target channel, as we assumed in Sec. 6.1.1.

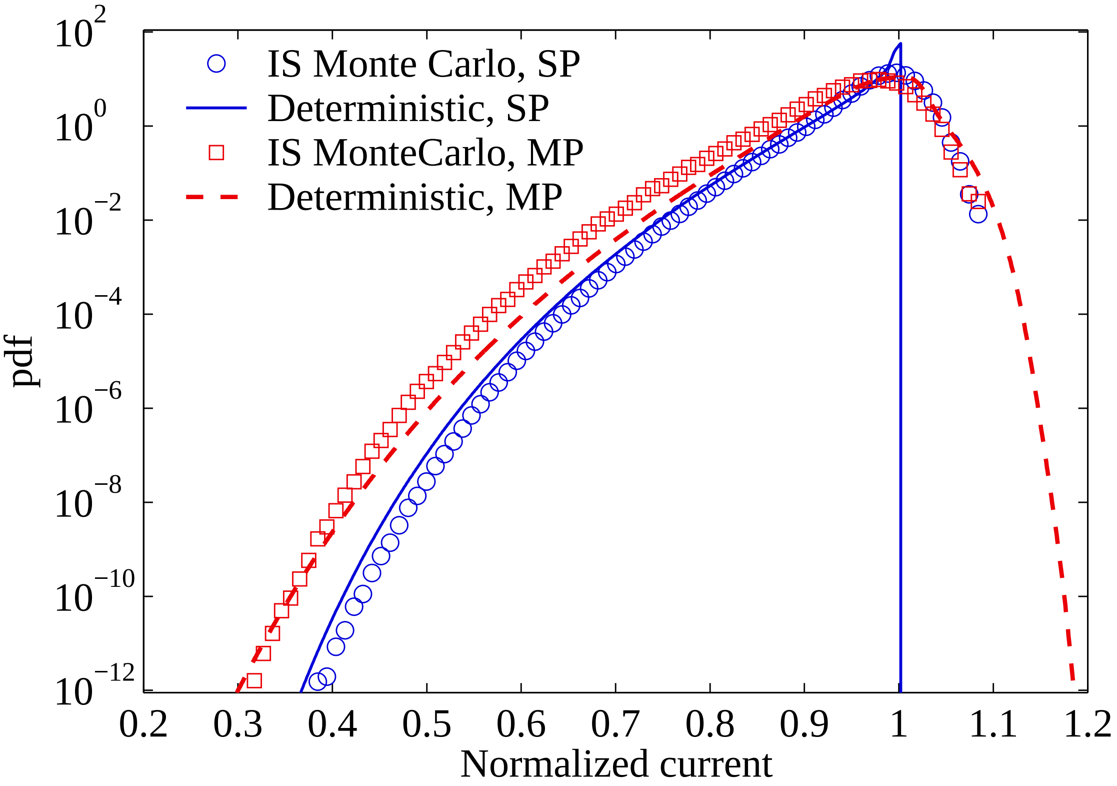

**Fig. 6.3:** Probability density function of the current in the marks at the detection point in the receiver due to the nonlinear distortion with multiple pulses (MP) in the target channel compared to a single pulse in the target channel (SP).

First, we immediately notice the difference that multipulse interactions make in the low-current tail of the pdf. A deterministic model of the pdf of the current at the decision point of the receiver that only includes timing jitter alone agrees well with IS simulations when there is only a single pulse in the target channel. However, the agreement is no longer good when there are multiple pulses in the target channel. It is necessary to include nonlinearly-induced amplitude jitter in the deterministic model, and we then see a good agreement between our deterministic model and IS simulation. The discrepancy in the pdf is less than an order of magnitude over the entire range of interest.

## 6.2 Bit error ratio calculations

In this section, we show how to use the deterministic theory that we developed for the evaluation of the nonlinear penalty to calculate the BER in the presence of amplified spontaneous emission (ASE) noise from optical amplifiers. We will use the assumption that the nonlinear penalty is statistically independent of the ASE noise since it greatly simplifies the analysis. In principle, these two processes are not independent, as one may envision that the noise affects the signals, and that in turn affects the way the signals interact with each other. However, as the noise is a zero-mean process, the effect of noise will result in either reduction or increase of the optical power of a pulse and, consequently, of the nonlinear interference that this pulse generates. Averaging over a large statistical ensemble tends to extinguish the correlation between the noise and the nonlinear penalty.

There is a number of different ways to characterize the performance of a system in the presence of the ASE noise that can be found in the literature [102], [103], [119], [120], [131], [138]–[149]. The most common model is the additive white Gaussian noise (AWGN) model, in which the noise is assumed to be completely independent of the signal and one can simply add up the noise contributions from all the inline amplifiers at the end of transmission prior to the optical filter. Since the optical receiver includes a square-low photodetector, the distribution of the signal after the photodetector is no longer Gaussian, which is well known in radio communications [150]. For optical communications, it was shown by Marcuse [139] and Humblet and Azizoğlu [140] for an idealized receiver model with a flat-top optical filter and an integrate-and-dump electrical filter. Lee and Shim [141] extended their results to the case with arbitrary optical and electrical filters. Forestieri [138] further generalized the results for arbitrary signals and introduced a computationally-efficient method for the BER calculations. Despite the availability of an accurate model, the Gaussian approximation of the pdf of the photodetected signal is still widely used due to its simplicity and straightforward relation to the experimentally measurable $Q$-factor.

In real systems, the noise experiences parametric gain while propagating with the signal. Hence, one cannot simply add the noise to the signal prior to the receiver and the AGWN model is in general not accurate. For the cases, in which the effect of parametric noise amplification is important, there are a several methods available in the literature [102], [103], [119], [120], [131], [143], [144], [146]–[148]. Hui [143], [144] introduced a method based on the continuous wave approximation of the signal, which allows to estimate the power spectral density of the noise due to its interaction with the signal. Holzlöhner, *et al.* [148] introduced a covariance matrix method, which is based on a calculation of the linearized evolution of noise around the signal. After separation of timing and phase jitter, the optical noise is multivariate Gaussian-distributed and hence completely described by its covariance matrix. This method can be used for any intensity-modulated signal. Further, Secondini [102], [103] showed how to calculate the pdf of the noise-corrupted signal in the presence of the parametric noise amplification without phase and timing jitter separation by using a combined additive and multiplicative perturbation method. These methods enable one to accurately calculate the pdf of the received current in the presence of noise. To calculated the BER, we will use the noise pdf in combination with the pdf of the nonlinearly-induced current distribution. Our method for combining the nonlinearly- and noise-induced distortion does not depend on the noise pdf. Therefore, for the sake of simplicity, we will use the AWGN assumption and calculate the noise pdf using Forestieri's technique [138]. We note that this assumption is widely used and often yields accurate results.

### 6.2.1 Additive white Gaussian noise model

We will expand the optical field that consists of the information signal and the noise in a Fourier basis. We then pass the field through the optical filter with a frequency response $H_o(f)$, an ideal square-law photodetector, and an electrical filter with a response $H_e(f)$. Finally, we will evaluate the pdf of the current and BER at the sampling time $t_k$.

We assume that each inline amplifier generates white Gaussian noise and it does not mix nonlinearly with the signal, so that at the end of the transmission the noise $w(t)$ is Gaussian-distributed and white with a power spectral density that we denote as $N_0$. Let $x(t)$ be the complex field envelope of the optical noise-free signal, which is periodic with a period of $NT$, where $N$ is the number of bits in the sequence, and $T$ is the bit period. Note that in computer modeling, we always consider time-limited signals, and due to the nature of the fast Fourier transform being used in the numerical calculations, the signal is periodic with a period of $NT$. Then the signal $x(t)$ can be expanded in a Fourier series as

$$x(t) = \sum_{l=-\infty}^{\infty} x_l \exp(-i2\pi lt/NT), \tag{6.11}$$

where

$$x_l = \frac{1}{NT} \int_{t=0}^{NT} x(t) \exp(i2\pi lt/NT)\, dt. \tag{6.12}$$

For the noise we also use the Fourier expansion in the interval $t_k - T_f < t < t_k$, where $T_f$ is the time constant related to the response time of the filters:

$$T_f = \mu \left( \frac{1}{B_o} + \frac{1}{B_e} \right), \tag{6.13}$$

where $B_o$ and $B_e$ are bandwidths of the optical and electrical filters and $\mu$ is a numerical parameter that is large enough so that the interval $(0, T_f)$ includes the significant details of the impulse response functions of both the optical and electrical filters. In practice, the value of $\mu$ is found by increasing it until the computed value of the BER stabilizes. The noise can then be represented by a Karhunen-Loève expansion as [104]

$$w(t) = \sum_{m=-\infty}^{\infty} w_m \exp[-i2\pi m(t - t_k - T_f)/T_f] \qquad t_k - T_f < t < t_k, \tag{6.14}$$

where the $w_m$ are independent complex Gaussian random variables with zero mean and whose real and imaginary parts have a variance $\sigma^2 = N_0/2T_f$.

We are interested in the value of the received current $y(t_k)$ at a sampling time $t_k$ after passing the signal and noise $x(t) + w(t)$ through the optical filter, the ideal square-law optical photodetector, and the electrical filter. As shown in [138], using expansions (6.11) and (6.14), one can obtain:

$$y(t_k) = d_k + r_k, \tag{6.15}$$

where

$$d_k = \sum_{l=-2L}^{2L} c_l H_e\left(\frac{l}{NT}\right) \exp(-i2\pi l t_k/NT) \tag{6.16}$$

is the signal term. We also have $L = \eta N T B_o$, where $\eta$ is a small number that is chosen to keep all the important frequency components of the signal. The parameter $c_l$ is the autocorrelation of the signal's Fourier coefficients,

$$c_l = \sum_{k=\max(-L,\, l-L)}^{\min(L,\, l+L)} x_l H_o\left(\frac{l}{NT}\right) x_{k-l}^* H_o^*\left(\frac{k-l}{NT}\right). \tag{6.17}$$

The noise term $r_k$ in (6.15) is given by

$$r_k = n_k + \nu_k \tag{6.18}$$

where $n_k$ and $\nu_k$ are given by

$$\begin{aligned} n_k &= \sum_{m=1}^{2M+1} \lambda_m \left| z_m + \frac{b_m}{\lambda_m} \right|^2, &\text{(6.19a)}\\ \nu_k &= \frac{|b_m|^2}{\lambda_m}, &\text{(6.19b)} \end{aligned}$$

with $M = B_o T_0$. The $\lambda_m$ are the eigenvalues of the matrix $\mathsf{A} = \mathsf{H}^{T*}\mathsf{Q}\mathsf{H}$. The matrix $\mathsf{H}$ is defined as

$$\mathsf{H} = \mathrm{diag}\left\{ \left| H_o\left( \frac{m-M-1}{T_f} \right) \right| \right\}, \tag{6.20}$$

and the elements $q_{mn}$ of $\mathsf{Q}$ are

$$q_{mn} = H_e\left( \frac{n-m}{T_f} \right). \tag{6.21}$$

The quantities $z_m$ in (6.15) are independent complex Gaussian random variables with zero mean and real and imaginary parts of variance $\sigma^2 = N_0/2T_f$. Finally, the $b_m$ are the components of the signal vector $\mathbf{b}$ defined as

$$\mathbf{b} = \mathsf{U}^{T*}\mathsf{H}^{T*}\mathbf{v}, \tag{6.22}$$

where $\mathsf{U}$ is a unitary matrix that diagonalizes the matrix $\mathsf{A}$ and the components of the vector $\mathbf{v}$ are given by

$$v_m = \sum_{l=-L}^{L} x_l H_o(l) H_e\left( \frac{l}{NT} - \frac{m-M-1}{T_f} \right) \exp(-i2\pi l t_k/NT). \tag{6.23}$$

Now we evaluate the pdf of the received current $y(t_k)$. The only component of the current $y(t_k)$ that is not deterministic is $\nu_k$ in (6.19b), which is a quadratic combination of independent Gaussian random variables, so that we can use the characteristic function $\Phi_{n_k}(\xi)$ to efficiently calculate its pdf. The characteristic function $\Phi_{n_k}(\xi)$ is defined as

$$\Phi_{n_k}(\xi) = \left\langle \exp\left(i\xi n_k\right)\right\rangle, \tag{6.24}$$

where the average $\langle\cdot\rangle$ is taken over the sample space of the noise. Using the independence of the real and imaginary parts of the random variables $z_m$, one obtains

$$\Phi_{n_k}(\xi) = \prod_{m=1}^{2M+1} \frac{\exp\left(\dfrac{i\xi\alpha_m}{1 - i\xi\beta_m}\right)}{1 - i\xi\beta_m}, \tag{6.25}$$

where

$$\alpha_m = \frac{|b_m|^2}{\lambda_m} \tag{6.26a}$$

$$\beta_m = 2\lambda_m\sigma^2. \tag{6.26b}$$

The pdf $p_{n_k}(\zeta)$ of $n_k$ can be obtained by taking the inverse Fourier transform of $\Phi_{n_k}(\xi)$,

$$p_{n_k}(\zeta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi_{n_k}(\xi)e^{-i\xi\zeta}d\xi. \tag{6.27}$$

In numerical calculations, one can use the fast Fourier transform (FFT) algorithm to evaluate (6.27), which works well in a large dynamic range limited by the round-off error of the FFT. The use of the FFT is sufficient for calculation of the error probabilities as low as $10^{-14}$. If the desired BER range is smaller than $10^{-14}$, one can calculate the pdf very efficiently using the saddle point approximation, described in detail in [138]. In this case, it is more convenient to use the moment generating function that is given by

$$M_{n_k}(s) = \left\langle \exp\left(s n_k\right)\right\rangle \tag{6.28}$$

instead of the characteristic function. Similarly to (6.25), we obtain

$$M_{n_k}(s) = \prod_{m=1}^{2M+1} \frac{\exp\left(\dfrac{s\alpha_m}{1-s\beta_m}\right)}{1-s\beta_m}. \tag{6.29}$$

The pdf $p_{n_k}(\zeta)$ of $n_k$ is then obtained by taking the inverse Laplace transform of $M_{n_k}(s)$,

$$p_{n_k}(\zeta) = \frac{1}{2\pi i}\int_{u_0-i\infty}^{u_0+i\infty} M_{n_k}(s)e^{-s\zeta}ds. \tag{6.30}$$

Since a line integral of an analytical function does not depend on the integration path, we can choose the integration contour $C$ to closely approximate the path of steepest descent passing through the saddle point $u_s$ of the integrand on the real s-axis. Then the main contribution to the integral (6.30) is near the $u_s$. The saddle point $u_s$ is determined from the equation

$$\frac{d}{ds}M_{n_k}(s)e^{-s\zeta} = 0, \tag{6.31}$$

which yields

$$\sum_{m=1}^{2M+1} \frac{\alpha_m + \beta_m(1-\beta_m s)}{(1-\beta_m s)^2} - \zeta = 0. \tag{6.32}$$

The solution for (6.32) can be found numerically using the Newton's method. The path of steepest descent near $u_s$ can be approximated by a parabola,

$$s = u_s + \frac{1}{2}\kappa v^2 + iv, \quad s = u + iv. \tag{6.33}$$

The curvature $\kappa$ of the parabola is given by [138]

$$\kappa = \frac{\Psi'''(u_s)}{3\Psi''(u_s)}, \tag{6.34}$$

where

$$\Psi(s) = \ln[M_{n_k}(s)] - s\zeta, \tag{6.35}$$

and $\Psi''(s)$ and $\Psi'''(s)$ are the second and third derivatives of $\Psi(s)$. The pdf can then be evaluated as [138]

$$p_{n_k}(\zeta) = \frac{1}{\pi}\int_{-\infty}^{\infty} \mathrm{Re}\left\{\exp\left[\Psi\left(u_s + \frac{1}{2}\kappa v^2 + iv\right)\right](1 - i\kappa v)\right\} dv. \tag{6.36}$$

The pdf (6.36) can then be calculated numerically using the trapezoidal rule and by dividing the integration interval into two intervals $(-\infty, 0]$ and $[0, \infty)$. The integration starts from $v = 0$ and stops when the contributions to the sum become negligible.

### 6.2.2 Combining noise and nonlinear effects

As we discussed earlier, we assume that the nonlinearly-induced and noise-induced signal distortions are statistically independent when computing the BER. In addition, we treat the contributions of timing jitter and amplitude jitter to nonlinearly-induced signal distortion as independent. These three impairments are characterized by the pdfs that we obtained previously. The pdf of the collision-induced time shift $p_T(t)$ is given by (5.17), the pdf of the nonlinearly-induced amplitude distortion $p_{\delta I}(I, t)$ is given by (6.10) and the pdf of the noise-induced distortion is obtained from (6.27),

$$p_{\mathrm{noise}}(I) = p_{n_k}(I - d_k - \nu_k). \tag{6.37}$$

In order to compute the pdf $p(I, t)$ of the received current $I$ at the sampling time $t$ due to both nonlinear signal distortion and noise, we convolve the pdf $p_T$ of the time shift, using either (5.17) or the Gaussian pdf given by (5.16) with the noise pdf $p_{\mathrm{noise}}(I, t)$ of the received current that is obtained by propagating a single-channel signal carrying a PRBS of length 32 through the system and with the amplitude jitter distribution (6.10):

$$p(I,\,t)=\int_{\zeta=-\infty}^{\infty}\left[\int_{\tau=-\infty}^{\infty}p_{\mathrm{noise}}(\zeta,t-\tau)\,p_T(\tau)d\tau\right]\,p_{\delta I}(I-\zeta,t)d\zeta. \qquad (6.38)$$

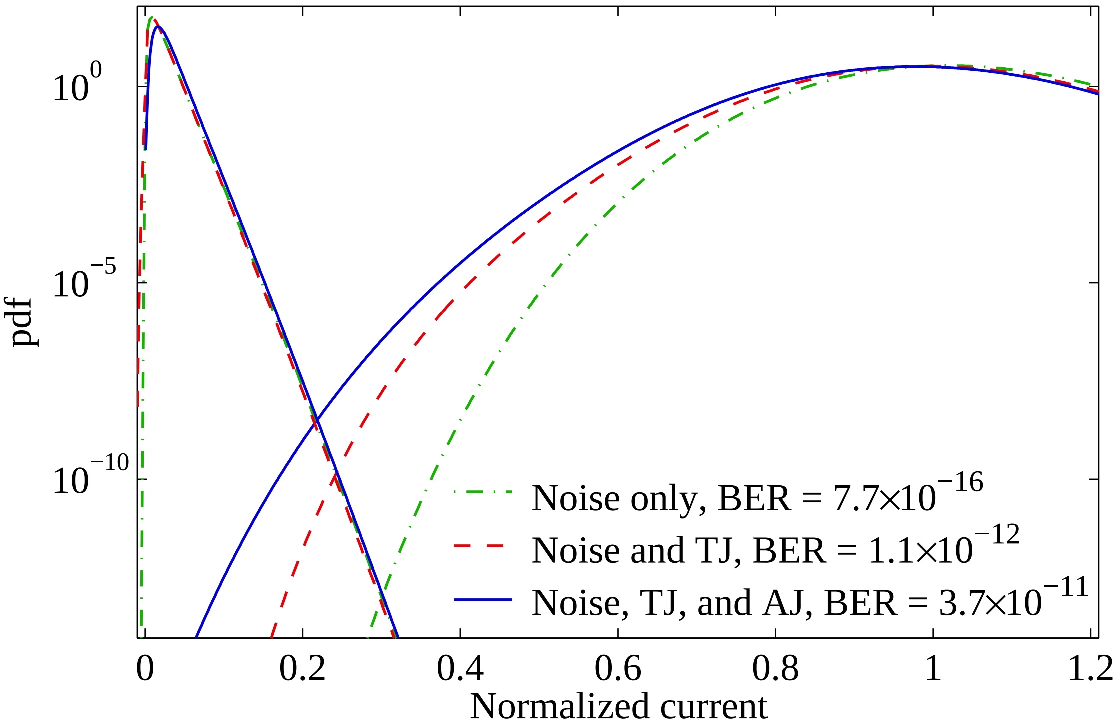

**Fig. 6.4:** Probability density function of the current in the marks and spaces at the decision time due to the nonlinear distortion and noise obtained using (6.38).

We show the results of our calculations in Figure 6.4. The dot-dashed curves show the pdf of the current in the marks and spaces due only to noise. The collision-induced timing jitter increases the BER by over three orders of magnitude. The corresponding pdfs are shown with the dashed curves. Finally, the nonlinearly-induced amplitude jitter degrades the BER by more than one order of magnitude and the total pdfs are displayed with the solid curves. Since the nonlinear interactions can degrade the BER by many orders of magnitude, it is important to model them accurately. Furthermore, even though the collision-induced timing jitter is the dominant nonlinear effect in

WDM RZ systems, it is necessary to consider the inter-channel, nonlinearly-induced amplitude distortion as it leads to a further performance degradation.

Finally, we analyze the accuracy of the Gaussian assumption for the collision-induced time shift pdf. In Fig. 6.5, we plot the BER and $Q$ as a function of channel separation, which was calculated with (6.38) using the time shift pdfs obtained from the Gaussian approximation (5.15b), and the characteristic function method (5.17). For each point in the plot, the total number of pump channels was chosen so that they filled a spectral range of 800 GHz around the center frequency of the target channel. A further increase in the number of channels had a negligible effect on the time shift pdf since the collision-induced time shift decreases quadratically with the channel spacing [41], [136].

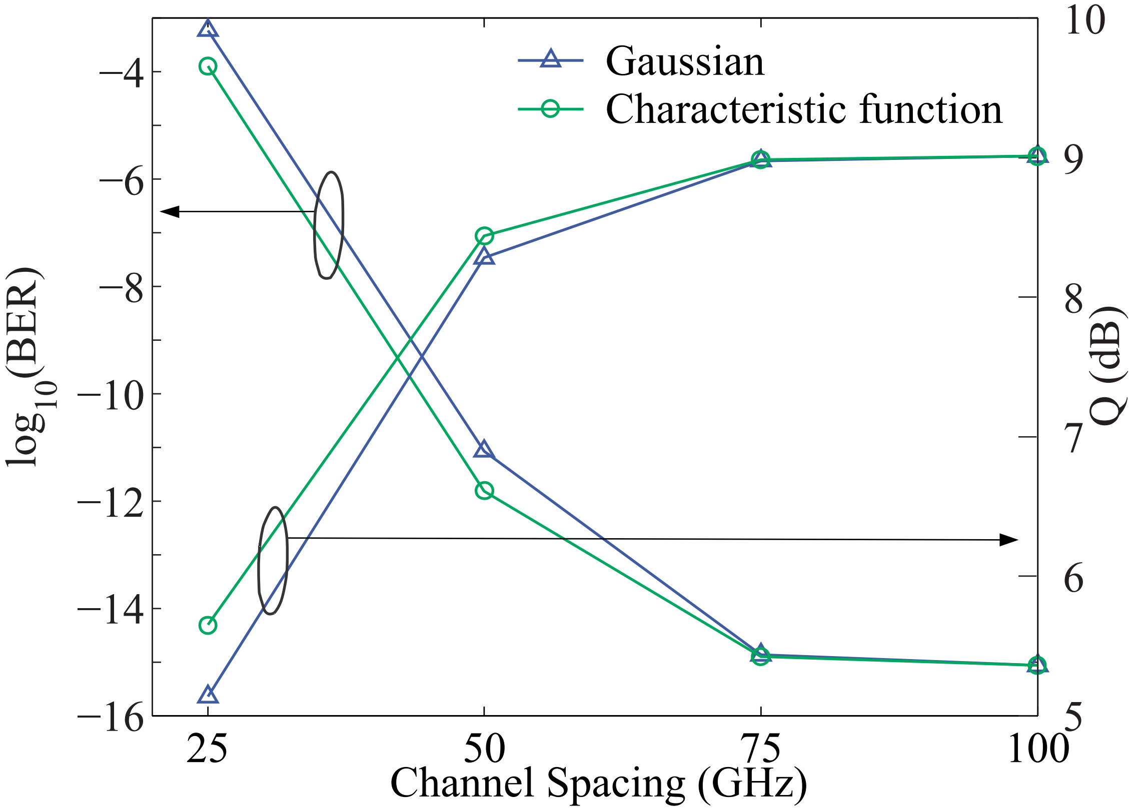

**Fig. 6.5:** Bit error ratio and $Q$-factor as functions of channel spacing.

As seen in Fig. 6.5, the difference in BER computed with the two methods is less than an order of magnitude when amplified spontaneous emission noise is present, despite the difference in the time shift pdfs that are obtained with the Gaussian and characteristic function methods. The reason for that is as follows: The pdf of the electrical current $p(I,\, t)$ in (6.38) is mainly affected by the part of the time shift pdf $p_T$ that has values of 1 down to $10^{-10}$–$10^{-12}$. Even though there is a significant difference between $p_{T,\mathrm{Gauss}}$ and $p_{T,\mathrm{Char}}$ in the tails, they agree near their centers. After the convolution of the noise and nonlinear amplitude jitter contributions to the pdf, the pdf of the current is changed by weighting with the values of the time shift pdf. Since $p_{T,\mathrm{Gauss}}$ and $p_{T,\mathrm{Char}}$ are close near their centers, the difference between the two on average is smaller than it is in the tail. As a result, the BER values obtained with the two methods are much closer than the two time shift pdfs. However, in the case of 25-GHz-spaced channels, the timing jitter leads to a very large error rate. At this point, the relative difference in the BER estimates obtained with the two methods becomes significant. This difference can determine the success or failure of error-correction techniques and cannot be neglected.

As the calculation of the exact pdf of collision-induced time shift using the characteristic function method is not computationally expensive, we believe that it should always be used instead of the Gaussian approximation.

# 7. CONCLUSIONS

The BER is the key figure of merit in an optical fiber communications system and must be accurately calculated when designing systems. For a system researcher, it poses a real challenge as the transmission errors are caused by a large number of phenomena, both deterministic and stochastic in nature. In this work, we have addressed one important aspect of the BER evaluation — the effect of fiber nonlinearity on transmission of optical signals and its contribution to the BER along with the noise from optical amplifiers.

Fiber nonlinearity by itself is a deterministic, unchanging property of silica glass. However, since the information carried by the optical light can be considered random from a statistical standpoint, the optical power in the fiber varies randomly as well. As a consequence, the nonlinear effects lead to non-deterministic signal distortions. Hence, there is a need for a probabilistic description of nonlinear penalties. Mathematically, the behavior of such systems is described by nonlinear partial differential equations with random initial conditions. An exact solution of these equations does not exist in general. A common approach to this problem is to estimate the average distortion of the signal and use the average to calculate the degradation of the $Q$-factor. This approach is, however, not usually sufficient to calculate the BER accurately.

In this dissertation, we have presented a new approach, whose essence is to characterize the nonlinearly-induced penalty with a complete pdf. One can use biased Monte Carlo simulations to estimate the pdf with the required accuracy. In this full statistical approach, we use the full propagation equation that cannot be solved analytically to model the system and we solve it numerically by randomly changing the input conditions on each new simulation run. The Monte Carlo simulations are, however, time consuming and with modern-day computers it is computationally expensive to obtain an estimate of the pdf over many orders of magnitude even with biasing. The main contribution of this dissertation is a new deterministic method for accurate evaluation of the nonlinear effects in BER calculation that is based on finding the pdf of nonlinearly-induced penalty. We have also calculated this pdf using a full statistical method based on biased Monte Carlo simulations. The two methods agree very well over fifteen orders of magnitude.

We focused our effort on WDM systems utilizing the RZ modulation format. While the RZ modulation is the usual choice in undersea systems, no method has appeared in the literature that allows one to accurately calculate the BER due to inter-channel nonlinear crosstalk in such systems.

The method for BER evaluation proposed here is based on calculation of the pdf of collision-induced time shift, which is the major nonlinear effect in WDM RZ systems, and the nonlinearly-induced amplitude jitter. To start with, we have calculated the time shift function for pairwise collisions using efficient analytical and numerical procedures. In this analysis, we assume that all multipulse collisions can be represented as a superposition of two-pulse collisions. The time shift function provides us with an important knowledge of the nonlinear properties of the system. This function can be used to identify the bit patterns that lead to large time shifts. It can be used to design line codes to effectively suppress collision-induced timing jitter [135]. The time shift function can also be used in probabilistic description of the nonlinear penalties due to timing jitter. In particular, we have calculated the pdf of collision-induced time shift using the time shift function and a characteristic function method. This calculation is completely deterministic.

To validate our deterministic theory, we have used biased Monte Carlo methods — the standard importance sampling method and the multicanonical Monte Carlo method. Despite the approximations and assumptions made in our reduced deterministic approach, the pdf of the collision-induced time shift calculated with this approach agrees with the biased Monte Carlo simulations over fifteen orders of magnitude. We also showed that the Gaussian function, often used to approximate the pdf of the collision-induced time shift, deviates significantly in the tails from the true pdf.

In the next step, we evaluated the pdf of the received current due to nonlinear effects, neglecting noise. We converted the pdf of the time shift to pdf of the received current at the detection point using a pulse shape at the receiver. With a single pulse in the target channel, we obtained excellent agreement between the reduced deterministic and full statistical approaches. This agreement demonstrates that the collision-induced timing jitter is the dominant nonlinear effect in this type of system. When considering multiple pulses in the target channel, we must also account for the amplitude jitter, induced by the inter- and intra-channel nonlinear interactions, which does not arise due to timing jitter. To calculate the pdf of the current due to this additional amplitude jitter, we applied an approach based on pairwise interactions, similar to what we did for the timing jitter. In this case, we considered the interaction of a pseudo-random pulse sequence (a superpulse) in the target channel with single pulses in the neighboring channel. Assuming additivity of the pulse distortions at the receiver, one can calculate the pdf of the current using the characteristic function approach. Then we combine the pdfs of the current that are due to collision-induced timing jitter and nonlinearly-induced amplitude jitter, assuming the statistical independence of the two. Despite the approximations, we have achieved agreement to within an order of magnitude between this reduced deterministic approach and a full statistical approach over the entire range of the pdfs that we calculated.

Finally, we showed how to calculate the BER in the presence of both nonlinear signal distortion and amplified spontaneous emission noise. As we showed, the nonlinear effects can significantly degrade the system performance, and in order to take these effects into account, one must use a probabilistic approach.

In this dissertation, we presented and validated the tools that enable one to accurately characterize the inter- and intra-channel nonlinear effects when calculating the BER in WDM RZ systems.

As more advanced modulation formats emerge from the research laboratories, the techniques that have been developed in this work can be used to accurately account for the nonlinear penalties in these new systems.

# BIBLIOGRAPHY

[1] L. Kazovsky, S. Benedetto, and A. Willner, *Optical Fiber Comminication Systems*, Boston: Artech House, 1996.

[2] G. P. Agrawal, *Fiber-Optics Communication Systems*, New York: Wiley, 2nd edition, 1997.

[3] I. P. Kaminov and T. L. Koch (Eds.), *Optical Fiber Telecommunications IIIA*, San Diego: Academic Press, 1997.

[4] I. P. Kaminov and T. Li (Eds.), *Optical Fiber Telecommunications IVB*, San Diego: Academic Press, 2002.

[5] R.-M. Mu and C. R. Menyuk, "Convergence of the chirped return-to-zero and dispersion managed soliton modulation formats in WDM systems," *J. Lightwave Technol.*, vol. 20, pp. 608–617, 2002.

[6] S. Bigo, G. Bellotti, and M. Chbat, "Investigation of cross-phase modulation limitation over various types of fiber infrastructures," *IEEE Photon. Technol. Lett.*, vol. 11, pp. 605–607, 1999.

[7] E. A. Golovchenko, A. N. Pilipetskii, N. S. Bergano, C. R. Davidsen, F. I. Khatri, R. M. Kimball, and V. J. Mazurczyk, "Modeling of transoceanic fiber-optic WDM communications systems," *IEEE J. Sel. Topics Quant. Electron.*, vol. 6, pp. 337–347, 2000.

[8] D. I. Kovsh, L. Liu, B. Bakhshi, A. N. Pilipetskii, E. A. Golovchenko, and N. S. Bergano, “Reducing interchannel crosstalk in long-haul DWDM systems,” *IEEE J. Sel. Topics Quant. Electron.*, vol. 8, pp. 597–602, 2002.

[9] T. Tsuritani, A. Agata, I. Morita, N. Edagawa, and A. Shigeyuki, “Ultra-long-haul 40-Gbit/s-based DWDM transmission using optically prefiltered CS-RZ signals,” *IEEE J. Sel. Topics Quant. Electron.*, vol. 10, pp. 403–411, 2004.

[10] F. Forghieri, R. W. Tkach, and A. R. Chraplyvy, “Fiber nonlinearities and thier impact on transmission systems,” in *Optical Fiber Telecommunications III A*, I. P. Kaminow and T. L. Koch, Eds. Academic Press, London, 1997.

[11] E. A. Golovchenko, A. N. Pilipetskii, and N. S. Bergano, “Transmission properties of chirped return-to-zero pulses and nonlinear intersymbol interference in 10 GBit/s WDM transmission,” in *Proc. Optical Fiber Communication Conference*, FC3, 2000.

[12] V. S. Grigoryan and A. Richter, “Efficient approach for modeling collision-induced timing jitter in WDM return-to-zero dispersion-managed systems,” *J. Lightwave Technol.*, vol. 18, pp. 1148–1154, 2000.

[13] E. Ciaramella, “Nonlinear impairments in extremely dense WDM systems,” *IEEE Photon. Technol. Lett.*, vol. 14, pp. 804–806, 2002.

[14] I. Lyubomirsky, T. Qui, J. Roman, M. Nayfeh, M. Frankel, and M. G. Taylor, “Interplay of fiber nonlinearity and optical filtering in ultradense WDM,” *IEEE Photon. Technol. Lett.*, vol. 15, pp. 147–149, 2003.

[15] L. E. Nelson, R. M. Jopson, A. H. Gnauk, and A. R. Chraplyvy, “Resonances in cross-phase modulation impairment in wavelength-division-multiplexed lightwave transmission,” *IEEE Photon. Technol. Lett.*, vol. 11, pp. 907–909, 1999.

[16] R. Hui, K. R. Demarest, and C. T. Allen, “Cross-phase modulation in multispan WDM optical fiber systems,” *J. Lightwave Technol.*, vol. 17, pp. 1018–1026, 1999.

[17] H. J. Thiele, R. I. Killey, and P. Bayvel, “Influence of transmission distance on XPM-induced intensity distortion in dispersion-managed, amplified fibre links,” *Electron. Lett.*, vol. 35, pp. 408–409, 1999.

[18] E. G. Shapiro, M. P. Fedoruk, and S. K. Turitsyn, “Numerical estimate of BER in optical systems with strong patterning effects,” *Electron. Lett.*, vol. 37, pp. 1179–1181, 2001.

[19] R. B. Jenkins, J. R. Sauer, S. Chakravarty, and M. J. Ablowitz, “Data-dependent timing jitter in wavelength-division-multiplexing soliton systems,” *Opt. Lett.*, vol. 20, pp. 1964–1966, 1995.

[20] M. J. Ablowitz, G. Biondini, S. Chakravarty, and R. L. Horne, “On timing jitter in wavelength-division multiplexed soliton systems,” *Opt. Commun.*, vol. 150, pp. 305–318, 1998.

[21] A. Richter and V. S. Grigoryan, “Efficient approach to estimate collision-induced timing jitter in dispersion-managed WDM RZ systems,” in *Proc. Optical Fiber Communication Conference*, WM33, 1999.

[22] O. V. Sinkin, J. Zweck, and C. R. Menyuk, “Effects of the nonlinearly-induced timing and amplitude jitter on the performance of different modulation formats in WDM optical fiber communications systems,” in *Proc. Optical Fiber Communication Conference*, TuF5, 2003.

[23] O. V. Sinkin, V. S. Grigoryan, R. Holzlöhner, A. Kalra, J. Zweck, and C. R. Menyuk, “Calculation of error probability in WDM RZ systems in presence

of bit-pattern-dependent nonlinearity and of noise," in *Proc. Optical Fiber Communication Conference*, TuN4, 2004.

[24] B. Bakhshi, M. Vaa, E. A. Golovchenko, W. W. Patterson, R. L. Maybach, and N. S. Bergano, "Comparison of CRZ, RZ, and NRZ modulation formats in a $64 \times 12.3$ Gb/s WDM transmission experiment over 9000 km," in *Proc. Optical Fiber Communication Conference*, WF4, 2001.

[25] R.-M. Mu, T. Yu, V. S. Grigoryan, and C. R. Menyuk, "Dynamics of the chirped return-to-zero modulation format," *J. Lightwave Technol.*, vol. 20, pp. 47–57, 2002.

[26] J. X. Cai, M. Nissov, C. R. Davidson, A. N. Pilipetskii, G. Mohs, H. Li, Y. Cai, E. A. Golovchenko, A. J. Lucero, D. Foursa, and N. S. Bergano, "Long-haul 40 Gb/s DWDM transmission with aggregate capacities exceeding 1 Tb/s," *J. Lightwave Technol.*, vol. 20, pp. 2247–2258, 2002.

[27] I. T. Lima, A. O. Lima, Y. Sun, H. Jiao, J. Zweck, and Menyuk C. R., "A receiver model for optical fiber communication systems with arbitrarily polarized noise," *J. Lightwave Technol.*, vol. 23, pp. 1478–1490, 2005.

[28] T.-K. Chiang, N. Kagi, M. E. Marhic, and L. G. Kazovsky, "Cross-phase modulation in fiber links with multiple optical amplifiers and dispersion compensators," *J. Lightwave Technol.*, vol. 14, pp. 249–260, 1996.

[29] M. Shtaif and M. Eiselt, "Analysis of intensity interference caused by cross-phase modulation in dispersive optical fibers," *IEEE Photon. Technol. Lett.*, vol. 10, pp. 979–981, 1998.

[30] G. Bellotti, M. Varani, C. Francia, and A. Bononi, "Intensity distortion induced by cross-phase modulation and chromatic dispersion in optical-fiber transmis-

sions with dispersion compensation," *IEEE Photon. Technol. Lett.*, vol. 10, pp. 1745–1747, 1998.

[31] A. Cartaxo, "Impact of modulation frequency on cross-phase modulation effect in intensity modulation-direct detection WDM systems," *IEEE Photon. Technol. Lett.*, vol. 10, pp. 1268–1270, 1998.

[32] A. Cartaxo, "Cross-phase modulation in intensity modulation direct detection WDM systems with multiple optical amplifiers and dispersion compensators," *J. Lightwave Technol.*, vol. 17, pp. 178–190, 1999.

[33] H. J. Thiele, R. I. Killey, and P. Bayvel, "Simple technique to determine cross-phase modulation induced penalties in WDM transmission," in *Proc. Optical Fiber Communication Conference*, ThM2, 2000.

[34] R. I. Killey, H. J. Thiele, V. Mikhailov, and P. Bayvel, "Prediction of transmission penalties due to cross-phase modulation in WDM systems using a simplified technique," *IEEE Photon. Technol. Lett.*, vol. 12, pp. 804–806, 2000.

[35] Z. Jiang and F. Chongcheng, "A comprehensive study on XPM- and SRS-induced noise in cascaded IM-DD optical fiber transmission systems," *J. Lightwave Technol.*, vol. 21, pp. 953–960, 2003.

[36] E. Ciaramella and E. Forestieri, "Analytical approximation of nonlinear distortions," *IEEE Photon. Technol. Lett.*, vol. 17, pp. 91–93, 2005.

[37] R. S. Luís and A. V. T. Cartaxo, "Analytical characterization of SPM impact on XPM-induced degradation in dispersion-compensated WDM systems," *J. Lightwave Technol.*, vol. 23, pp. 1503–1513, 2005.

[38] M. J. Ablowitz, G. Biondini, A. Biswas, A. Docherty, and T. Hirooka, "Collision-

induced timing shifts in dispersion-managed transmission systems," *Opt. Lett.*, vol. 27, pp. 318–320, 2002.

[39] M. J. Ablowitz, A. Docherty, and T. Hirooka, "Incomplete collisions in strongly dispersion-managed return-to-zero communication systems," *Opt. Lett.*, vol. 28, pp. 1191–1193, 2003.

[40] M. J. Ablowitz, C. Ahrens, G. Biondini, S. Chakravarty, and A. Docherty, "Reduction of collision-induced timing shifts in dispersion-managed quasi-linear systems with periodic-group-delay dispersion compensation," *Opt. Lett.*, vol. 29, pp. 2354–2356, 2004.

[41] C. Ahrens, M. J. Ablowitz, A. Docherty, O. V. Sinkin, J. Zweck, V. S. Grigoryan, and C. R. Menyuk, "Asymptotic analysis of collision-induced timing shifts in return-to-zero quasi-linear systems with pre- and post-dispersion compensation," *Opt. Lett.*, vol. 31, pp. 5–7, 2006.

[42] H. Sugahara, A. Maruta, and Y. Kodama, "Optimal allocation of amplifiers in a dispersion-managed line for a wavelength-division-multiplexed soliton transmission system," *Opt. Lett.*, vol. 24, pp. 145–147, 1999.

[43] H. Sugahara, H. Kato, T. Inoue, A. Maruta, and Y. Kodama, "Optimal dispersion management for a wavelength division multiplexed optical soliton transmission system," *J. Lightwave Technol.*, vol. 17, pp. 1547–1559, 1999.

[44] X. Liu, X. Wei, L. F. Mollenauer, C. J. McKinstrie, and C. Xie, "Collision-induced time shift of a dispersion-managed soliton and its minimization in wavelength-division-multiplexed transmission," *Opt. Lett.*, vol. 28, pp. 1148–1150, 2003.

[45] N. S. Bergano, "Undersea communication systems," in *Optical Fiber Telecom-*

*munications*, I. E. Kaminow and T. L. Koch, Eds., vol. IVB. Academic, San Diego, CA, 2002.

[46] E. Desurvire, *Erbium-doped fiber amplifiers: principles and applications*, New York: Wiley, 1994.

[47] T.-S. Yang and W. L. Kath, "Optimal prechirping for dispersion-managed transmission of return-to-zero pulses," in *Proc. Optical Fiber Communication Conference*, ThQ4, 1999.

[48] G. P. Agrawal, *Nonlinear Fiber Optics*, London: Academic Press, 2nd edition, 1995.

[49] 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.

[50] C .R. Menyuk, B. S. Marks, I. T. Lima, J. Zweck, Y. Sun, G. M. Carter, and D. Wang, "Polarization effects in long-haul undersea systems," in *Undersea Fiber Communication Systems*, J. Chesnoy, Ed. Academic Press, San Diego, CA, 2002.

[51] J. X. Cai, M. Nissov, A. N. Pilipetskii, C. R. Davidson, R. M. Mu, M. A. Mills, L. Xu, D. Foursa, R. Megnes, P. C. Corbett, D. Sutton, and N. S. Bergano, "1.28 Tb/s (32x40 Gb/s) transmission over 4,500 km," in *Proc. European Conference on Optical Communication*, PD.M.1.2, 2001.

[52] P. R. Trischitta and E. L. Varma, *Jitter in digital transmission systems*, Norwood, MA: Artech House, 1989.

[53] I. T. Lima, Jr., *Investigation of the performance degradation due to polarization*

*effects in optical fiber communications systems*, Ph.D. thesis, University of Maryland Baltimore County, Baltimore, USA, 2003.

[54] M. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems*, New York: Plenum, 1992.

[55] N. S Bergano, C. R. Davidson, N. A. Mills, P. Corbert, S. G. Evangelides, B. Pedersen, R. Menges, J. L. Zyskind, J. W. Sulhoff, A. K. Srivastava, C. Wolf, and J. Judkins, "Long-haul WDM transmission using optimal channel modulation," in *Proc. Optical Fiber Communication Conference*, PD16, 1997.

[56] B. Bakhshi, M. Arend, M. Vaa, E. A. Golovchenko, D. J. Duff, H. Li, S. Jiang, W. W. Patterson, R. L. Maybach, and D. Kovsh, "1 Tbit/s (101 × 10 Gbit/s) transmission over transpacific distance using 28 nm C-band EDFAs," in *Proc. Optical Fiber Communication Conference*, PD21, 2001.

[57] A. Sano, Y Miyamoto, S. Kuwahara, and H. Toba, "A 40 Gb/s/ch WDM transmission with SPM/XPM suppression through prechirping and dispersion management," *J. Lightwave Technol.*, vol. 18, pp. 1519–1527, 2000.

[58] R.-M. Mu and C. R. Menyuk, "Symmetric slope compensation in a long-haul WDM system using the CRZ format," *IEEE Photon. Technol. Lett.*, vol. 13, pp. 797–799, 2001.

[59] L. Becouarn, G. Vareille, P. Pecci, and J. F. Marcerou, "3 Tbit/s transmission (301 DPSK channels at 10.709 Gb/s) over 10270 km with a record efficiency of 0.65 (bit/s)/Hz," in *Proc. European Conference on Optical Communication*, Th4.3.2, 2003.

[60] P. V. Mamyshev and N. A. Mamysheva, "Pulse-overlapped dispersion-managed data transmission and intrachannel four-wave mixing," *Opt. Lett.*, vol. 24, pp. 1454–1456, 1999.

[61] A. Mecozzi, C. B. Clausen, M. Shtaif, S. Park, and A. H. Gnauck, "Cancellation of timing and amplitude jitter in symmetric links using highly dispersed pulses," *IEEE Photon. Technol. Lett.*, vol. 13, pp. 445–447, 2001.

[62] A. S. Lenihan, O. V. Sinkin, B. S. Marks, G. E. Tudury, R. J. Runser, A. Goldman, C. R. Menyuk, and G. M. Carter, "Nonlinear timing jitter in an installed fiber network with balanced dispersion compensation," *IEEE Photon. Technol. Lett.*, vol. 17, pp. 1558–1560, 2005.

[63] S. N. Knudsen, M. Pedersen, and L. Gruner-Nielsen, "Optimization of dispersion compensating fibers for cabled long-haul applications," *Electron. Lett.*, vol. 36, pp. 2067–2068, 2000.

[64] S. N. Knudsen, D. W. Peckham, M. Pedersen, D. Philen, T. Veng, L. R. Pritchett, and L. Gruner-Nielsen, "New dispersion-slope matched fiber pair for undersea fiber optic transmission systems," in *Proc. SubOptic'01*, T4.2.2, 2001.

[65] M. Fullenbaum, "Next generation of transoceanic submarine solutions," in *Proc. Pacific Telecommunications Council.* M122, 2006.

[66] J.-X. Cai, D. G. Foursa, C. R. Davidson, Y. Cai, G. Domagala, H. Li, L. Liu, W. W. Patterson, A. N. Pilipetskii, M. Nissov, and N. S. Bergano, "A DWDM demonstration of 3.73 Tb/s over 11,000 km using 373 RZ-DPSK channels at 10 Gb/s," in *Proc. Optical Fiber Communication Conference*, PD22, 2003.

[67] T. Tsuritani, K. Ishida, A. Agata, K. Shimomura, I. Morita, T. Tokura, H. Taga, T. Mizuochi, N. Edagawa, and S. Akiba, "70-Ghz-spaced 40×42.7 Gb/s transpacific transmission over 9400 km using prefiltered CSRZ-DPSK signals, all-Raman repeaters, and symmetrically dispersion-managed fiber spans," *J. Lightwave Technol.*, vol. 22, pp. 215–224, 2004.

[68] S. A. Akhmanov and R. V. Khokhlov, *Problemy Nelineinoi Optiki*, Moscow, Russia: VINITI, 1964, in Russian.

[69] S. Akhmanov, *Problems of Nonlinear Optics*: Gordon and Breach Science, 1972.

[70] Y. R. Shen, *The Principles of Nonlinear Optics*, New York: Wiley, 1984.

[71] N. Bloembergen, *Nonlinear Optics*: World Scientific, 4th edition, 1996.

[72] R. W. Boyd, *Nonlinear Optics*, San Diego, USA: Academic Press, 2nd edition, 2002.

[73] R. Shuker and R. W. Gammon, "Raman-scattering selection-rule breaking and the density of states in amorphous materials," *Phys. Rev. Lett.* , vol. 25, pp. 222–225, 1970.

[74] A. R. Chraplyvy, "Limitations on lightwave communications imposed by fiber nonlinearities," *J. Lightwave Technol.*, vol. 8, pp. 1548–1557, 1990.

[75] O. V. Sinkin, V. S. Grigoryan, R. Holzlöhner, J. Zweck, and C. R. Menyuk, "Probabilistic description of the nonlinear penalties in WDM RZ systems using multicanonical Monte Carlo simulations," in *Proc. IEEE LEOS Annual Meeting*, ThI5, 2003.

[76] N. S. Bakhvalov, N. P. Zhidkov, and G. M. Kobelkov, *Chislennye Metody*, Moscow, Russia: Binom, 2002, in Russian.

[77] V. M. Verzhbitsky, *Foundations of Numerical Methods*, Moscow, Russia: V. Shkola, 2002.

[78] RSoft Design Group, *Optical Communication Design Suite*, http://www.rsoftdesign.com/.

[79] VPIphotonics, *VPItransmissionMaker*, http://www.vpiphotonics.com/.

[80] O. V. Sinkin, R. Holzlöhner, J. Zweck, and C. R. Menyuk, "Optimization of the split-step Fourier method in modeling optical-fiber communication systems," *J. Lightwave Technol.*, vol. 21, pp. 61–68, 2003.

[81] P. V. Mamyshev and J. F. Mollenauer, "Pseudo-phase-matched four-wave mixing in soliton wavelength-division multiplexing transmission," *Opt. Lett.*, vol. 21, pp. 396–398, 1996.

[82] G. Bosco, A. Carena, V. Curri, R. Gaudino, P. Poggiolini, and S. Benedetto, "Suppression of spurious tones induced by the split-step method in fiber systems simulation," *IEEE Photon. Technol. Lett.*, vol. 12, pp. 489–491, 2000.

[83] C. W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations*, New Jersey: Prentice-Hall, 1971.

[84] K. J. Blow and D. Wood, "Theoretical description of transient stimulated Raman scattering in optical fibers," *IEEE J. Quantum Electron.* , vol. 25, pp. 2665–2673, 1989.

[85] B. Fornberg and T. A. Driscoll, "A fast spectral algorithm for nonlinear wave equations with linear dispersion," *J. Comp. Math.*, vol. 155, pp. 456–467, 1999.

[86] J. A. Fleck, J. R. Morris, and M. D. Feit, "Time dependent propagation of high energy laser beams through the atmosphere," *Appl. Phys.*, vol. 10, pp. 129–160, 1976.

[87] F. Forghieri, "Modeling of wavelength multiplexed lightwave systems," in *Proc. Optical Fiber Communication Conference.* TuG1, 1997.

[88] V. E. Zakharov and A. B. Shabat, "Tochnaya teoriya dvumernoy samofokusirovki i odnomernoy avtomodulyatsii voln v nelineinykh sredakh," *Zh. Eksp. Teor. Fiz.*, vol. 61, pp. 118–134, 1971, in Russian.

[89] V. E. Zakharov and A. B. Shabat, "Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media," *Sov. Phys.-JETP*, vol. 34, pp. 62–69, 1972.

[90] A. Hasegawa and F. Tappert, "Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. I. Anomalous dispersion," *Appl. Phys. Lett.*, vol. 23, pp. 142–144, 1973.

[91] S. Novikov, S. V. Manakov, L. P. Pitaevskii, and V. E. Zakharov, *Theory of Solitons : The Inverse Scattering Method*, New York–London: Plenum Press, 1984.

[92] C. Desem and P. L. Chu, "Soliton-soliton interactions," in *Optical Solitions*, J. R. Taylor, Ed., chapter 5, pp. 107–151. Cambridge Univ. Press, Cambridge, UK, 1992.

[93] A. Hasegawa, *Solitons in Optical Communications*, Oxford, UK: Clarendon, 1995.

[94] R.-M. Mu, V. S. Grigoryan, C. R. Menyuk, G. M. Carter, and J. M. Jacob, "Comparison of theory and experiment for dispersion-managed solitons in a recirculating fiber loop," *IEEE J. Sel. Topics Quant. Electron.*, vol. 6, pp. 248–257, 2000.

[95] R. M. Mu, T. Yu, V. S. Grigoryan, and C. R. Menyuk, "Convergence of CRZ and DMS formats in WDM systems using dispersion management," in *Proc. Optical Fiber Communication Conference*, FC1, 2000.

[96] S. Pachnicke and E. Voges, "Fast analytical assessment of the signal quality in transparent optical networks," *J. Lightwave Technol.*, vol. 24, pp. 815–824, 2006.

[97] K.-P. Ho, "Statistical properties of stimulated Raman crosstalk in WDM systems," *J. Lightwave Technol.*, vol. 18, pp. 915–921, 2000.

[98] C. J. McKinstrie, "Frequency shifts caused by collisions between pulses in dispersion-managed systems," *Opt. Commun.*, vol. 205, pp. 123–137, 2002.

[99] A. Cartaxo, "Small-signal analysis for nonlinear and dispersive optical fibres, and its application to design of dispersion supported transmission systems with optical dispersion compensation," *IEE Proc.-Optoelectron.*, vol. 146, pp. 213–222, 1999.

[100] J. Wang and K. Petermann, "Small signal analysis for dispersive optical fiber communication systems," *J. Lightwave Technol.*, vol. 10, pp. 96–100, 1992.

[101] A. Vannucci, P. Serena, and A. Bononi, "The RP method: A new tool for the iterative solution of the nonlinear Schrödinger equation," *J. Lightwave Technol.*, vol. 20, pp. 1102–1112, 2002.

[102] M. Secondini, *Optical communication theory and techniques for high bit-rate systems*, Ph.D. thesis, Scuola Superiore Sant'Anna, Pisa, Italy, 2006.

[103] M. Secondini, E. Forestieri, and C. R. Menyuk, "A novel perturbation method for signal-noise interaction in nonlinear dispersive fibers," in *Proc. Optical Fiber Communication Conference.* OThD3, 2006.

[104] J. G. Proakis, *Digital Communications*, New York: McGraw-Hill, 1995.

[105] L. F. Mollenauer, J. P. Gordon, and M. N. Islam, "Soliton propagation in long

fibers with periodically compensated loss," *IEEE J. Quantum Electron.* , vol. 22, pp. 157–173, 1986.

[106] L. F. Mollenauer, S. G. Evangelides, and 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.

[107] J. F. L. Devaney, W. Forysiak, A. M. Niculae, and N. J. Doran, "Soliton collisions in dispersion-managed wavelength-division-multiplexed systems," *Opt. Lett.*, vol. 22, pp. 1695–1697, 1997.

[108] P. V. Mamyshev and L. Mollenauer, "Soliton collisions in wavelength-division-multiplexed dispersion-managed systems," *Opt. Lett.*, vol. 24, pp. 448–450, 1999.

[109] A. Mecozzi, "Timing jitter in wavelength-division-multiplexed filtered soliton transmission," *J. Opt. Soc. Am. B* , vol. 15, pp. 152–161, 1998.

[110] Y. Cai, T. Adalı, C. R. Menyuk, and J. M. Morris, "Sliding window criterion codes and concatenation scheme for mitigating timing-jitter-induced errors in WDM fiber transmissions," *J. Lightwave Technol.*, vol. 20, pp. 201–212, 2002.

[111] T. Hirooka and A. Hasegawa, "Chirped soliton interaction in strongly dispersion-managed wavelength-division-multiplexing systems," *Opt. Lett.*, vol. 23, pp. 768–770, 1998.

[112] M. J. Ablowitz and A. S. Fokas, *Complex variables: Introduction and applications*, Cambridge, UK: Cambridge University Press, second edition, 2003.

[113] A. Docherty, *Collision induced timing shifts in wavelength-division-multiplexed optical fiber communications systems*, Ph.D. thesis, University of New South Wales, Sydney, Australia, 2004.

[114] B. L. Bratley, B. L. Fox, and L. E. Schrage, *A Guide to Simulation*, New York: Springer-Verlag, 1987.

[115] J.-C. Chen, D. Lu, J. S. Sadowsky, and K. Yao, "On importance sampling in digital communications—Part I: Fundamentals," *IEEE J. Sel. Areas Commun.*, vol. 11, pp. 289–299, 1993.

[116] P. J. Smith, M. Shafi, and H. Gao, "A review of importance sampling techniques in communications sytems," *IEEE J. Sel. Areas Commun.*, vol. 15, pp. 597–613, 1997.

[117] E. Veach, *Robust Monte Carlo methods for light transport simulations*, Ph.D. thesis, Stanford University, Stanford, USA, 1997.

[118] D. Yevick, "Multicanonical communication system modeling-application to PMD statistics," *IEEE Photon. Technol. Lett.*, vol. 14, pp. 1512–1514, 2002.

[119] R. Holzlöhner and C. R. Menyuk, "Accurate bit error rates from multicanonical Monte Carlo simulations," in *Proc. Conference on Lasers and Electro-Optics*, CThJ3, 2003.

[120] R. Holzlöhner and C. R. Menyuk, "The use of multicanonical Monte Carlo simulations to obtain accurate bit error rates in optical communications systems," *Opt. Lett.*, vol. 28, pp. 1894–1896, 2003.

[121] I. T. Lima, A. O. Lima, J. Zweck, and C. R. Menyuk, "Efficient computation of outage probabilities due to polarization effects in a WDM system using a reduced Stokes model and importance sampling," *IEEE Photon. Technol. Lett.*, vol. 15, pp. 45–47, 2003.

[122] A. O. Lima, I. T. Lima, C. R. Menyuk, G. Biondini, B. S. Marks, and W. L. Kath, "Statistical analysis of the performance of PMD compensators using

multiple importance sampling," *IEEE Photon. Technol. Lett.*, vol. 13, pp. 1716–1718, 2002.

[123] R. O. Moore, G. Biondini, and W. L. Kath, "Importance sampling for noise-induced amplitude and timing jitter in soliton transmission systems," *Opt. Lett.*, vol. 28, pp. 105–107, 2003.

[124] G. Biondini, W. L. Kath, and C. R. Menyuk, "Importance sampling for polarization-mode dispersion: techniques and applications," *J. Lightwave Technol.*, vol. 22, pp. 1201–1215, 2004.

[125] A. O. Lima, I. T. Lima, and Menyuk C. R., "Error estimation in multicanonical Monte Carlo simulations with applications to polarization-mode-dispersion emulators," *J. Lightwave Technol.*, vol. 23, pp. 3781–3789, 2005.

[126] A. Bilenca and G. Eisenstein, "Statistical noise properties of an optical pulse propagating in a nonlinear semiconductor optical amplifier," *IEEE J. Quantum Electronics*, vol. 41, pp. 36–44, 2005.

[127] T. Lu and D. Yevick, "Efficient multicanonical algorithms," *IEEE Photon. Technol. Lett.*, vol. 17, pp. 861–863, 2005.

[128] I. Neokosmidis, T. Kamalakis, A. Chipouras, and T. Sphicopoulos, "Estimation of the four-wave mixing noise probability-density function by the multicanonical Monte Carlo method," *Opt. Lett.*, vol. 30, pp. 11–13, 2005.

[129] R. Holzlöhner, A. Mahadevan, C.R. Menyuk, J.M. Morris, and J. Zweck, "Evaluation of the very low BER of FEC codes using dual adaptive importance sampling," *IEEE Commun. Letters*, vol. 9, pp. 163–165, 2005.

[130] Y. Yadin, M. Shtaif, and M. Orenstein, "Bit-error rate of optical DPSK in fiber

systems by multicanonical Monte Carlo simulations," *IEEE Photon. Technol. Lett.*, vol. 17, pp. 1355–1357, 2005.

[131] W. Pellegrini, J. Zweck, C. R. Menyuk, and R. Holzlöhner, "Computation of bit error ratios for a dense WDM system using the noise covariance matrix and multicanonical Monte Carlo methods," *IEEE Photon. Technol. Lett.*, vol. 17, pp. 1644–1646, 2005.

[132] B. A. Berg and F. Neuhaus, "Multicanonical ensemble: a new approach to simulate first-order phase transitions," *Phys. Rev. Lett.* , vol. 68, pp. 9–11, 1992.

[133] N. Metropolis and S. Ulam, "The Monte Carlo method," *J. Am. Stat. Assoc.*, vol. 44, pp. 335–341, 1949.

[134] C. Xu, C. Xie, and L. Mollenauer, "Analysis of soliton collisions in a wavelength-division-multiplexed dispersion-managed soliton transmission system," *Opt. Lett.*, vol. 27, pp. 1303–1305, 2002.

[135] W. Wang, O. V. Sinkin, T. Adali, J. Zweck, and C. R. Menyuk, "Prior-based line-coding for WDM RZ systems," in *Proc. Conference on Lasers and Electro-Optics*, CFN5, 2004.

[136] O. V. Sinkin, V. S. Grigoryan, J. Zweck, C. R. Menyuk, A. Docherty, and M. Ablowitz, "Calculation, characterization, and application of the time shift function in wavelength-division-multiplexed return-to-zero systems," *Opt. Lett.*, vol. 30, pp. 2056–2058, 2005.

[137] A. Papoulis, *Probability, Random Variables, and Stochastic Processes,* New York: McGraw-Hill, 3rd edition, 1991.

[138] E. Forestieri, “Evaluating the error probability in lightwave systems with chromatic dispersion, arbitrary pulse shape and pre- and postdetection filtering,” *J. Lightwave Technol.*, vol. 18, pp. 1493–1503, 2000.

[139] D. Marcuse, “Derivation of analytical expressions for the bit-error probability in lightwave systems with optical amplifiers,” *J. Lightwave Technol.*, vol. 8, pp. 1816–1823, 1990.

[140] P. A. Humblet and M. Azizoǧlu, “On the bit error rate of lightwave systems with optical amplifiers,” *J. Lightwave Technol.*, vol. 9, pp. 1576–1582, 1991.

[141] J. S. Lee and C. S. Shim, “Bit-error analysis of optically preamplified receivers using eigenfunction expansion method in optical frequency domain,” *J. Lightwave Technol.*, vol. 12, pp. 1224–1229, 1994.

[142] A. Carena, V. Curri, R. Gaudino, P. Poggiolini, and S. Benedetto, “New analytical results on fiber parametric gain and its effects on ASE noise,” *IEEE Photon. Technol. Lett.*, vol. 9, pp. 535–537, 1997.

[143] R. Hui, D. Chowdhury, M. Newhouse, M. O’Sullivan, and M. Poettcker, “Nonlinear amplification of noise in fibers with dispersion and its impact in optically amplified systems,” *IEEE Photon. Technol. Lett.*, vol. 9, pp. 392–394, March 1997.

[144] R. Hui, M. O’Sullivan, A. Robinson, and M. Taylor, “Modulation instability and its impact in multispan optical amplified IMDD systems: Theory and experiments,” *J. Lightwave Technol.*, vol. 15, pp. 1071–1081, July 1997.

[145] V. S. Grigoryan, C. R. Menyuk, and R.-M. Mu, “Calculation of timing and amplitude jitter in dispersion-managed optical fiber communications using linearization,” *J. Lightwave Technol.*, vol. 17, pp. 1347–1356, 1999.

[146] G. Bosco, A. Carena, V. Curri, R. Gaudino, P. Poggiolini, and S. Benedetto, "A novel analytical method for the BER evaluation in optical systems affected by parametric gain," *IEEE Photon. Technol. Lett.*, vol. 12, pp. 152–154, 2000.

[147] R. Holzlöhner, V. S. Grigoryan, C. R. Menyuk, and W. L. Kath, "Accurate calculation of eye diagrams and bit error rates in long-haul transmission systems using linearization," *J. Lightwave Technol.*, vol. 20, pp. 389–400, 2002.

[148] R. Holzlöhner, C. R. Menyuk, W. L. Kath, and Grigoryan V. S., "A covariance matrix method to compute bit error rates in a highly nonlinear dispersion-managed soliton system," *IEEE Photon. Technol. Lett.*, vol. 15, pp. 688–690, 2003.

[149] R. Holzlöhner, *A Covariance Matrix Method for the Computation of Bit Errors in Optical Transmission Systems*, Ph.D. thesis, University of Maryland Baltimore County, Baltimore, Maryland, USA, 2003.

[150] M. Kac and A. J. F. Siegert, "On the noise in radio receivers with square law detectors," *J. Appl. Phys.*, vol. 18, pp. 383–397, 1947.