Characterization and Coding Techniques for Long-Haul Optical Telecommunication Systems

by

Miloš Ivković

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF MATHEMATICS

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

In the Graduate College

The University of Arizona

2 0 0 7

THE UNIVERSITY OF ARIZONA

GRADUATE COLLEGE

As members of the dissertation committee, we certify that we have read the dissertation prepared by Miloš Ivković entitled, *Characterization and Coding Techniques for Long-Haul Optical Telecommunication Systems*, and recommend that it be accepted as fulfilling the dissertation requirement for the degree of Doctor of Philosophy.

_____ Date: 11/19/2007

Ildar Gabitov

_____ Date: 11/19/2007

Ivan Djordjević

_____ Date: 11/19/2007

Robert Indik

_____ Date: 11/19/2007

Mikhail Stepanov

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College. I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_____ Date: 11/19/2007

Dissertation Director: Bane Vasić

Statement by Author

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Milos Ivkovic

Table of Contents

List of Figures 7

List of Tables 9

Abstract 10

1. Introduction 12

2. Signal propagation 16

2.1. Dispersion Managed Systems 17

2.2. Analysis of intersymbol interference 20

3. Channel characterization 28

3.1. Introduction 28

3.2. Probability density functions by using the Instanton approach . 31

3.3. Edgeworth expansion 35

3.3.1. Orthogonal polynomials 36

3.4. Numerical results 39

Table of Contents—*Continued*

3.5. Final remarks on channel characterization 45

4. Channel capacity 49

4.1. Introduction 49

4.2. Information rate and capacity for optical transmission systems 50

4.2.1. BCJR algorithm 52

4.3. Numerical results 54

4.4. Final remarks on the proposed method for estimating achievable information rate 56

5. Coding schemes 58

5.1. Hard Decision Coding Scheme Description 60

5.1.1. Gallager B decoding algorithm for LDPC codes . . 61

5.2. Soft decoding scheme 62

5.3. Numerical results 64

5.3.1. Hard decision scheme 64

5.3.2. Soft decision scheme 66

5.4. Final remarks on decoding schemes 70

Table of Contents—*Continued*

6. Conclusion 72

References 74

List of Figures

Figure 2.1. Pulse propagation through dispersion managed system 19

Figure 2.2. Energy versus number of bit patters, for bit patterns with "0" at the center timeslot. 25

Figure 2.3. Energy versus number of bit patters, for bit patterns with "1" at the center timeslot. 26

Figure 2.4. Hundred bit patterns with "0" at the center timeslot and highest energy. 27

Figure 3.1. Scheme of the optical transmission system considered 40

Figure 3.2. Probability density functions after 100 spans 42

Figure 3.3. Probability density functions after 100 spans. Normalization is chosen so that the peak of the PDF for the bit pattern "0001000" corresponds to 1. 43

Figure 3.4. Comparison of various PDF approximations for energy of bit pattern "0110110" after 300 spans 45

Figure 3.5. Comparison of various PDF approximations for energy of bit pattern "0001000" after 300 spans 46

List of Figures—*Continued*

Figure 3.6. Comparison of various PDF approximations for center bit energy within bit pattern "0001000" after 300 spans. Normalization is chosen so that the peak of the PDF corresponds to 1. 47

Figure 3.7. Comparison of various PDF approximations for center bit energy within bit pattern "0110110" after 300 spans. Normalization is chosen so that 1 on the x-axis corresponds to 1 as in Fig. 3.6. 48

Figure 4.1. Achievable information rates for different memory lengths . . . 55

Figure 4.2. Achievable information rates for different noise strength 56

Figure 5.1. Block diagram of the proposed "hard decision" coding scheme . 60

Figure 5.2. Block diagram of the proposed "soft decision" coding scheme . 63

Figure 5.3. BER performance of proposed hard-decision error correcting scheme with different LDPC codes 66

Figure 5.4. BER performance of Viterbi equalizer supplemented with hard-decision Gallager B LDPC decoder, for different memory assumptions . 67

Figure 5.5. Comparison of different coding schemes 68

Figure 5.6. Comparison of different coding schemes 69

Figure 5.7. Comparison of coding schemes with different memory 70

List of Tables

Table 2.1. Bit configurations with highest energy 26

Table 3.1. Parameters of the fibers used 41

Abstract

This dissertation is a study of error in long haul optical fiber systems and how to coupe with them. First we characterize error events occurring during transmission, then we determine lower bounds on information capacity (achievable information rates) and at the end we propose coding schemes for these systems.

Existing approaches for obtaining probability density functions (PDFs) for pulse energy in long-haul optical fiber transmission systems rely on numerical simulations or analytical approximations. Numerical simulations make far tails of the PDFs difficult to obtain, while existing analytic approximations are often inaccurate, as they neglect nonlinear interaction between pulses and noise.

Our approach combines the instanton method from statistical mechanics to model far tails of the PDFs, with numerical simulations to refine the middle part of the PDFs. We combine the two methods by using an orthogonal polynomial expansion constructed specifically for this problem. We demonstrate the approach on an example of a specific submarine transmission system.

Once the channel is characterized estimating achievable information rates is done by a modification of a method originally proposed by Arnold and Pfitser. We give numerical results for the same optical transmission system (submarine system at

transmission rate 40Gb/s). The achievable information rate varies with noise and length of the bit patterns considered (among other parameters). We report achievable numerical rates for systems with different noise levels, propagation distances and length of the bit patterns considered.

We also propose two iterative decoding schemes suitable for high-speed long-haul optical transmission. One scheme is a modification of a method, originally proposed in the context of magnetic media, which incorporates the BCJR algorithm (to overcome intersymbol interference) and Low-Density Parity-Check (LDPC) codes for additional error resilience. This is a "soft decision scheme" -meaning that the decoding algorithm operates with probabilities (instead of binary values). The second scheme is "hard decision" -it operates with binary values. This scheme is based on the maximum likelihood sequence detection-Viterbi algorithm and a hard decision "Gallager B" decoding algorithm for LDPC codes.

1. Introduction

Optical fiber is a low-loss, high capacity, cost-efficient transmission media. Due to its huge natural bandwidth and low loss, an optical fiber is capable of keeping up with growing bandwidth demand. Because of this property, optical fiber is of crucial importance for current and next generation communication systems.

Information can be encoded in many different ways. One way, which is considered in this dissertation, is *on-off keying* with return to zero (RZ) pulses, where binary information is transmitted by encoding ones as pulses and zeros as empty time slots.

Although conceptually simple, transmission of information in this way is very complex. Numerous physical effects can lead to pulse degradation and possible loss of information. Chromatic dispersion, pulse attenuation, noise from amplifiers, (non-linear) interaction between pulses, polarization mode dispersion and other effects make construction of the optical fiber transmission systems a challenging task.

Within this dissertation we first deal with characterization of optical fiber transmission system(s) as a communication channel (Chapters II and III). Then we estimate the fundamental limits of information capacity of such systems, by calculating achievable information rates (lower bounds on capacity) for systems with different

parameters (Chapter IV). In Chapter V, we turn our attention to coding schemes for such systems, proposing two schemes -one, operating with probabilities ("soft decision" scheme), with excellent correcting capability, but high complexity, and another, operating with bit values ("hard decision" scheme), with considerably reduced complexity, but somewhat worse performance. Conclusions are in Chapter VI. A brief overview of the dissertation follows until the end of this introduction.

Chapter II is introductory. Here we discuss the propagation of light through optical fiber and define basic terms that will be used in the dissertation, such as dispersion managed system and length of non-linearity. We also analyze intersymbol interference, clarifying why characterization of a long haul optical fiber transmission system as a communication channel requires the probability density functions that are derived in Chapter III.

In Chapter III we characterize long haul optical fiber transmission systems as a communication channel(s) by developing a new approach to model pulse energy. Existing approaches for obtaining probability density functions (PDFs) rely on numerical simulations or analytical approximations. Numerical simulations make far tails of the PDFs difficult to obtain, while analytical approximations are often inaccurate, as they neglect nonlinear interaction between pulses and noise.

Our approach combines the instanton method from statistical mechanics to model the far tails of the PDFs, with numerical simulations to refine the middle part of the

PDFs. We combine the two methods by using an orthogonal polynomial expansion constructed specifically for this problem.

Motivation to use this method comes from the fact that bit errors in modern optical transmission systems are very rare. The statistics of rare events are dominated by contributions from very special randomness realizations (optimal fluctuations). The concept of optimal fluctuations was first developed in the context of condensed matter physics; recently it has been successfully applied to evaluation of the efficiency (bit error rate) of forward-correction codes [43]. We demonstrate the approach on an example of a specific submarine transmission system. This work has been presented at the *IEEE ICC 2007* conference and published in *IEEE Photonic Technology Letters* [1].

As rates at which information can be transmitted increase, a fundamental question about physical limitations of optical fiber arises. The nonlinear nature of the propagation of light in optical fiber systems plays the crucial role in limiting the capacity, and makes these limits difficult to calculate.

In Chapter IV we discuss a method for estimation of achievable information rates of high-speed optical transmission systems. Once the channel is characterized (as in Chapter III for example) estimating achievable information rates is done by a modification of a method originally proposed by Arnold and Pfitser[35, 36]. We give numerical results for the same optical transmission system (submarine system

at transmission rate 40Gb/s). The achievable information rate varies with noise and length of the bit patterns considered (among other parameters). We report achievable numerical rates for systems with different noise levels, propagation distances and length of the bit patterns considered. Work in this chapter is based on two papers published in *IEEE Journal of Lightwave Technology* [2], [3].

In Chapter V we propose two iterative decoding schemes suitable for high-speed long-haul optical transmission. One scheme is a modification of a method, originally proposed in the context of magnetic media, which incorporates the BCJR algorithm to overcome (intrachannel) intersymbol interference and Low-Density Parity-Check (LDPC) codes (for additional error resilience). This is a "soft decision scheme" - meaning, roughly speaking, that the decoding algorithm operates with probabilities (instead of binary values). The second scheme is "hard decision" -it is based on the maximum likelihood sequence detection-Viterbi algorithm and a hard decision "Gallager B" decoding algorithm for LDPC codes. Work from this chapter is presented in two conference publications [4], [5].

2. Signal propagation

The emphasis within this dissertation is on original results. When it comes to physics of pulse propagation through fiber we rely on well known references such as [10].

The transmission of a signal through the fiber is modeled by the nonlinear Schrödinger equation (NLSE) [10]:

$$\frac{\partial A}{\partial z} = -\frac{\alpha}{2}A - \frac{i}{2}\beta_2(z)\frac{\partial^2 A}{\partial t^2} + \frac{\beta_3(z)}{6}\frac{\partial^3 A}{\partial t^3} + i\gamma|A|^2 A \tag{2.1}$$

where z is the propagation distance along the fiber, retarded time $t = t_{real} - z/v_g$ gives a frame of reference moving at the group velocity v_g, $A(z,t)$ is the complex field amplitude of the pulse, α is the attenuation coefficient of the fiber, $\beta_2(z)$ is the group velocity dispersion (GVD) coefficient, $\beta_3(z)$ is the second-order GVD, γ is the non-linearity coefficient.

As mentioned in the introduction, information can be encoded in various ways. Within this dissertation we study *on-off keying,* where binary information is transmitted by encoding ones as pulses and zeros as empty time slots. How do errors in these systems occur? It is obvious that attenuation of the pulses might lead to misdetection of an original "1" as "0", but misdetection of "0" as "1" is also possible

due to four-wave mixing (described by the non-linear term of NLSE Eq.(2.1)), which redistributes energy over the pulse sequence causing pattern dependant amplitude jitter. This energy redistribution causes generation of pulses in empty slots (false bits) and a resulting increase of the bit-error rate. This mechanism has a random nature due to the randomness in the structure of the bit pattern. So, estimation of the error probability requires understanding of bit-pattern deterioration.

2.1. Dispersion Managed Systems

Chromatic dispersion -fact that different frequencies travel through fiber at different speeds, is arguably the biggest problem in optical fiber transmission. The dispersion is represented in NLSE Eq.(2.1), via the dispersion coefficient $\beta_2(z)$ and the second derivative in time $\frac{\partial^2 A}{\partial t^2}$. Higher order dispersion is relevant to the systems operating at high speeds (roughly speaking at 40 Gb/s and higher). This is due to the fact the pulses shorter in time have larger bandwidth. In the section on numerical simulation of this dissertation we will consider the term with third order derivative ($\frac{\beta_3(z)}{6}\frac{\partial^3 A}{\partial t^3}$ in Eq.(2.1)).

To deal with, or better said to *manage* dispersion, a scheme called *dispersion management* is used. The dispersion map consists of alternating section of fiber with positive (anomalous) and negative (normal) dispersion. The dispersion map is defined as a piecewise constant function of the propagation distance z, so that $\beta_2(z) = D_+ > 0$

for the first section of the map and $\beta_2(z) = D_- < 0$ for the second section. Such systems are called *dispersion managed system.*

One *span* consist of one section of fiber with positive dispersion, one section of fiber with negative dispersion, and corresponding amplifiers (see Fig. 3.1). The system is designed so that the residual dispersion $\int_0^z \beta_2(z)dz$ at the end of each span is small, or zero in the case of total dispersion compensation. After being launched into the fiber, a pulse experiences dispersive broadening while it propagates through the first section of the map, then recompresses back to its original shape when it traverses the second section of the map, see Fig. 2.1.

The dispersion management technique demonstrate an excellent performance in optical fiber communications (see for example [21]). However, there are problems. As nearby pulses broaden, they begin to overlap with their neighbors, and four-wave mixing and other nonlinear processes occur. The most malign effect of this nonlinear interaction is the generation of *ghost pulses*, which are pulses that form in timeslots which initially contain no pulse. After enough dispersion map periods, a ghost pulse in a "0" slot may contain enough energy to register as a "1" at the receiver, producing an error. The onset of ghost pulse energy accumulation was shown to grow quadratically with propagation distance [9].

An important parameter of dispersion managed systems is a measure of pulses overlapping called *strength of the map.* Pulses with initial full width at half max-

imum (FWHM) τ_0 are broadened to the maximal FWHM τ_{dm} and then return to their original FWHM at the end of the map period (in the case of total dispersion compensation). This maximum pulse width is used to define the *strength* $S = \tau_{dm}/\tau_0$ of the map. In the case of high speed data transmission (narrow pulses) the ratio of the maximal pulse width to its minimal value over the map cycle is large (strong dispersion map). The amount of broadening is large except for small regions at the beginning and at the end of the dispersion map. For a given transmission line, the strength of the dispersion map is larger for shorter pulses which results in stronger overlapping of optical pulses during the "breathing" cycle. It is important to mention that pulses at the launching point usually occupy only a fraction of the designated timeslot. This fraction is called *duty cycle* and has to be taken into account in when calculating the number of timeslots affected by a specific pulse.

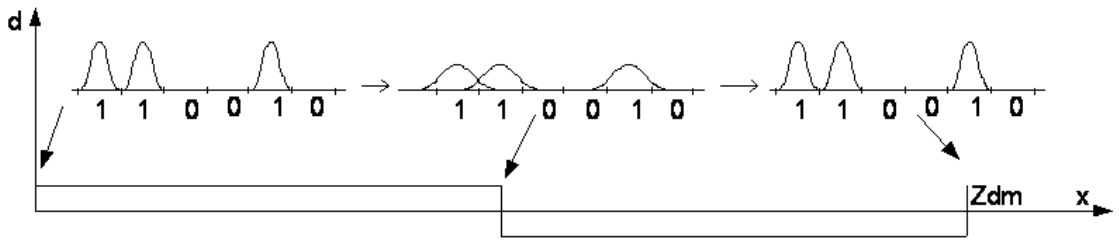

FIGURE 2.1. Pulse propagation through dispersion managed system

2.2. Analysis of intersymbol interference

As the demand for transmitting information increases (primary due to developments in internet technology), there is a strong push to increase speed of transmission over existing long haul optical fiber transmission systems. At this time, there are systems operating at 2.5 Gb/s and 10 Gb/s, the majority of the systems in the next generation will most likely operate at 40Gb/s, although there is research activity at speeds up to 160Gb/s.

As speed increases, bit timeslots are packed together more densely. For example, at this dissertation we run computer simulations for systems operating at 40Gb/s, for which one timeslot has length of 25 picoseconds. This requires pulses which are very short in time, consequently these pulses have (relatively) large bandwidth. This, in turn, means that dispersion will be large, i.e., pulses will overlap during longer periods of transmission, leading to stronger four-wave mixing and other nonlinear processes. In other words, intersymbol interference (ISI) will be pronounced. The analysis that follows will shred some light on this process, but before we start the analysis we define *length of non-linearity,* an important parameter that will be mentioned in the following chapters. Characteristic length of non-linearity is defined as:

$$z_{nl} = \frac{1}{\gamma P_0}$$

where γ is the non-linearity coefficient (from NLSE (Eq.2.1)) and P_0 is the mean

launched optical pulse power. z_{nl} has physical dimension of (propagation) length, as name suggests. Roughly speaking z_{nl} is the distance at which non-linear effects become important. Reducing the launched power might seem like a good idea to increase the length of non-linearity, however, spontaneous amplifier noise imposes lower bounds on launched power -simply said we need pulses powerful enough to distinguish them from noise.

To start analyzing (quite complicated, practical) optical fiber transmission systems we first consider a somewhat simplified dispersion managed system, where losses are completely compensated by amplifiers; higher order dispersion and noise are neglected.

In this case, the system is accurately modeled by the dimensionless nonlinear Schrödinger equation with the following boundary condition

$$i\psi_z + \theta'(z)\psi_{tt} + 2|\psi|^2\psi = 0 \tag{2.2}$$

$$\psi_0(t) = \psi(t,0) = \pi^{-1/4}\sum_{k=1}^{N}\zeta_k \exp[-(t-kT)^2/2] \tag{2.3}$$

$$\psi_k(t,0) = \pi^{-1/4}\exp[-(t-kT)^2/2]. \tag{2.4}$$

Here ψ is the dimensionless envelope of the electric field, t is the retarded time variable moving at the group velocity, $\theta'(z) = d(z)$ is the periodic renormalized dispersion. The dispersion map is defined as a piecewise constant function of the propagation distance z, so that $\theta'(z) = D_+ > 0$ for the first section of the map and $\theta'(z) = D_- < 0$ for the second section. The residual dispersion over one span is zero,

that is $\int_0^{z_{dm}} \theta'(z)dz$ Boundary conditions correspond to Gaussian shaped pulses. Here $\zeta_k = 1$ with probability $1/2$ and $\zeta_k = 0$ with probability $1/2$, i.e., we assume a binary source with uniform distribution. This sequence consists of N slots on either side of an initially empty center slot.

At leading order, evolution of optical pulses is governed by the linear part of the equation Eq.(2.2) and evolution of each individual pulse $\Psi_k(t,x)$ is given by

$$\Psi_k = \zeta_k \frac{\exp\left(-(t-kT)^2/(2\chi)\right)}{\pi^{1/4}\chi} \tag{2.5}$$

$$\chi = \sqrt{1+2i\theta(z)} \tag{2.6}$$

The equation for the next order $\psi(t,x) = \Psi(t,x) + \phi(t,x)$ is:

$$i\phi_z + \theta'(z)\phi_{tt} + 2|\Psi|^2\Psi = 0 \tag{2.7}$$

The solution of the Eq. (2.7) in the originally empty slot, at the end of m-th dispersion map period $\phi_m(t) \equiv \phi(t, z_m)$, where $\theta(z_m) = 0$ is represented (see [19, 18]) as:

$$\begin{aligned} \phi_m = (t,z) &= \frac{2iz}{\pi^{3/4} z_{dm}} \sum_{k,l,n=-N}^{N} \zeta_k\zeta_l\zeta_n \int_0^{z_{dm}} (12u^2 - 4iu + 1)^{-1/2} \\ &\times \exp\left[-\frac{3-2iu}{2-12iu}(t-t_{kln})^2 - T^2\frac{(n-k)(l-n)+(k-l)^2/(1+2iu)}{3-2iu}\right] dy, \end{aligned} \tag{2.8}$$

with

$$t_{kln} = T\frac{k+l-n+i(k+l+n)/2u}{1+3i/2u}.$$

Here it is understood that u is a function of the integration variable y. The expression for the energy in the center timeslot is given by

$$\begin{aligned} E &= \int_{-T/2}^{T/2} |\psi(t,z)+\phi(t,z)|^2\, dt \\ &\simeq \int_{-T/2}^{T/2} |\phi(t,z)|^2\, dt, \end{aligned} \tag{2.9}$$

where the approximation can be done for the bit patterns with "0" at the center timeslot since the tails of the Gaussian pulses in other timeslots contribute a negligible amount of energy to the center timeslot at the end of the map.

If $|\theta| \gg 1$, the value of the exponent is largest when $k+l=n$. This is referred to as the *resonant condition* and participating bits are called *resonant triples*. As we already mentioned, the energy of the ghost pulse (in the case where initially there was no pulse at the 0-th slot) depends quadratically on the number of dispersion map periods [9].

We studied the energy of these bit patterns numerically, by using Fourier split-step scheme. The bit rate was 40 Gb/s, duty cycle 33%, peak power of launched pulses was 1mW. The length of one dispersion map period (span) was $z_{dm} = 50$km, consisting of two fiber pieces, one with positive and one with negative dispersion $d_+ = -d_- = 20$ps/(nm km). This resulted in map strength of $S \approx 36$ and from that,

considering the duty cycle, we get that strength of the mao is ≈ 6 bit timeslots on each side. We considered all the bit configurations of length 15 , i.e, there were 7 timeslots on each side of the center timeslot. Nonlinear refractive index was $2.2\times10^{-20}\mathrm{m}^2/\mathrm{W}$. Pulses were propagated for L=750km (=50km*15spans).

In Fig. 2.2 we present levels of energy accumulated during propagation at the center timeslot for bit patterns with (initial) zero at the center slot. The x-axis represents energy normalized so that 1 corresponds to the average energy at the center slot for the bit patterns with "1" at the center slot, the y-axis is number of bit patterns with given energy. A similar plot, for bit levels with (initial) one at the center slot is in Fig. 2.3. Note that the distribution of energy at Fig. 2.2 is non-Gaussian and has exponential extended tail.

In the Table 3.1 we present five of bit patters with "0" at the center timeslot -those that have the highest energy. As it can be expected, the sequence with zero surrounded by ones has the highest energy. The next two bit patterns are also expected, they differ from the one with highest energy only at the most distant timeslots. The fourth and fifth bit pattern have empty timeslots five positions away from the center slot. This is probably because contributions from certain bit triples have different phase, so there can be destructive interference between different contributions. A similar effect was noticed in [9].

At the end of this section, notice that Eq. (2.2) has time symmetry, so it is

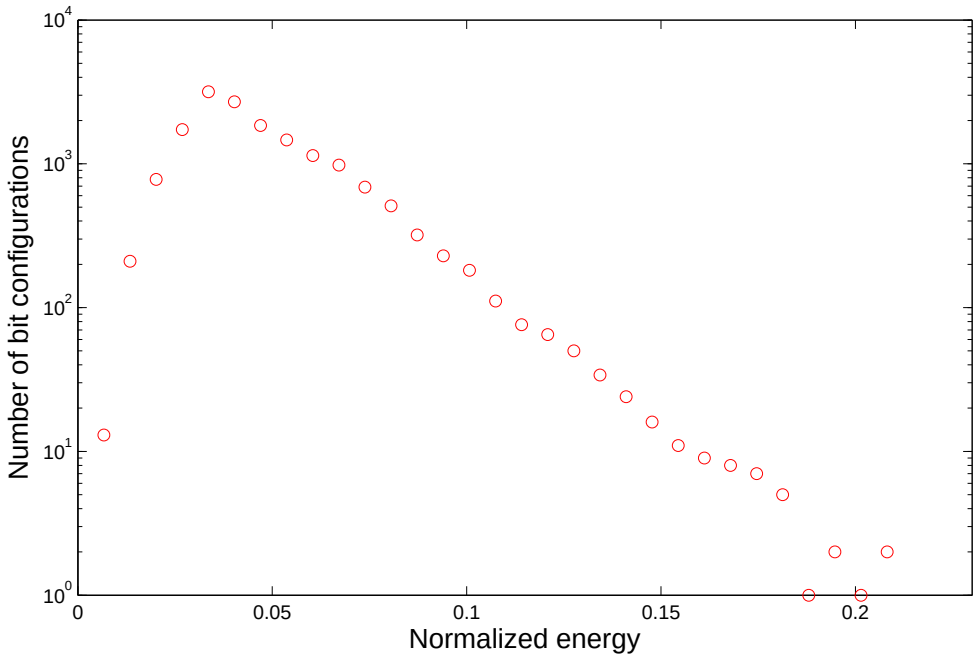

FIGURE 2.2. Energy versus number of bit patters, for bit patterns with "0" at the center timeslot.

expected that bit patterns symmetric with respect to center time slot have the same energy. This can be seen at Fig. 2.4., where hundred bit patterns with "0" at the center timeslot and highest energy are represented. Dark areas correspond to pulses and white correspond to empty timeslots. The certain discrepancy from perfect symmetry in Fig. 2.4. comes from the fact the many bit patterns (presented on the plot as well as some that are not) carry some energy (see Fig. 2.2).

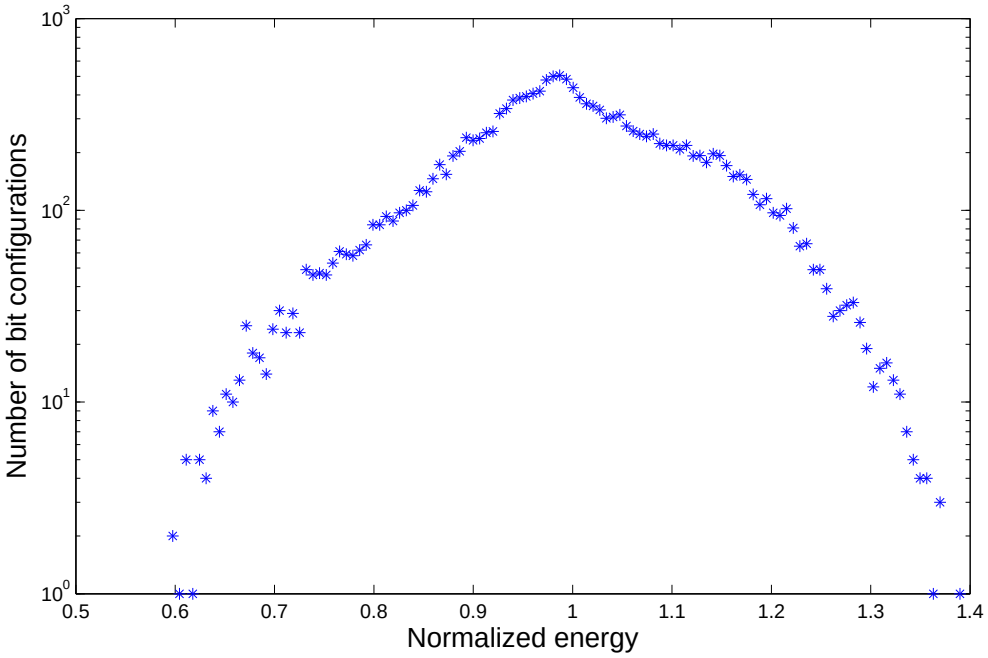

FIGURE 2.3. Energy versus number of bit patters, for bit patterns with "1" at the center timeslot.

TABLE 2.1. Bit configurations with highest energy

1111111	0	1111111
1111111	0	1111110
0111111	0	1111111
1111111	0	1111011
1101111	0	1111111

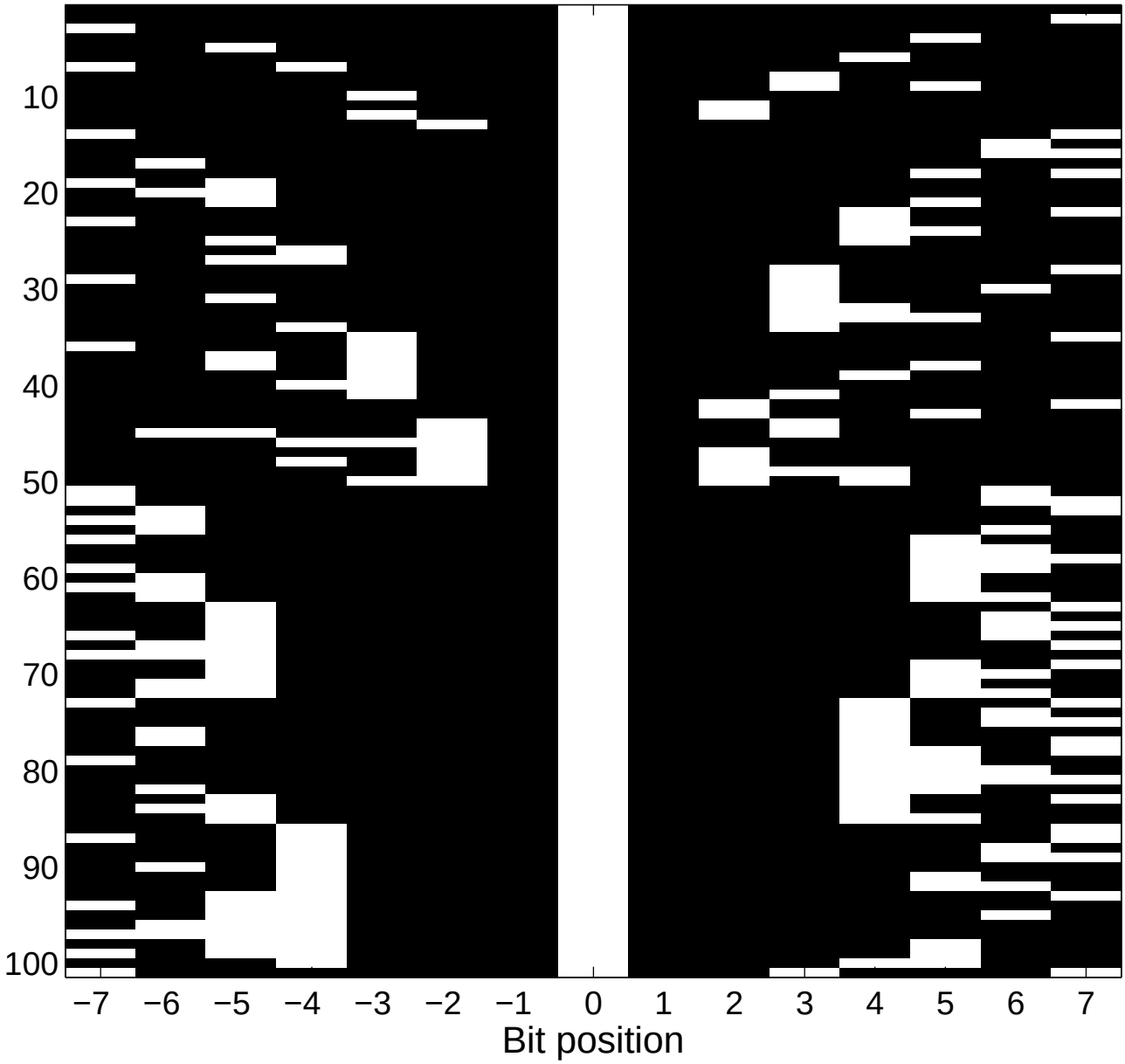

FIGURE 2.4. Hundred bit patterns with "0" at the center timeslot and highest energy.

3. Channel characterization

3.1. Introduction

Growing demand for information transmission leads to constant increases in optical fiber transmission systems bit rate. On the physical level this results in shortening of the bit carrier width τ_0. In order to keep bit error rate in compliance with systems requirements each pulse must carry energy $\mathcal{E}$ above a certain threshold $\mathcal{E} > \mathcal{E}_\prime$. The energy of the optical pulse can be evaluated as $\mathcal{E} \sim \tau_\prime \times \mathcal{I}_\prime$, here I_0 is characteristic value of the pulse intensity. Therefore, pulse intensity is growing with bit rate ($BR \sim 1/\tau_0$) as $I_0 \sim \mathcal{E} \times \mathcal{BR}$. Taking into account Kerr nonlinearity of the refractive index of a silica fiber $n = n_0 + \alpha I$ one can conclude that nonlinear response of optical fiber is growing with bit rate. On another hand this nonlinearity results in a intrachannel bit interaction (inter-symbol interaction due to four-wave mixing mechanism as mentioned in a previous chapter), which must be taken into account. The evaluation of the channel capacity taking into account nonlinear coupling of the temporal noise and intersymbol interaction (Gordon-Mollenauer effect [20]) presents a challenging mathematical problem.

Presence of ISI, leads to a necessity to study probability density functions (PDFs)

of signal samples corresponding to longer bit configurations and not just individual bits. Also, although noise is uncorrelated initially (at the amplifiers), due to nature of propagation and filtering, noise samples at adjacent timelots on the receiver end are correlated.

All of this means that a precise characterization of these systems would require a multidimensional PDFs $P(E_{-k}, E_{-k+1}, \ldots, E_0, \ldots, E_{k-1}, E_k | x_{-k}, x_{-k+1}, \ldots, x_0, \ldots, x_{k-1}, x_k)$ where E_i stands for observed energy at timelot i and x_i stands for bit transmitted at timeslot i.

However, even if theoretically developed, these PDFs would require optical signal processing or very complicated electronics to operate in real time. In particular it will require optical buffering during the signal processing. Unfortunately the current development of technology has not delivered optical buffers reliably operating at high bit-rates.

Taking into account these constrains in this work we present a method dealing with simplified probability distribution functions of form $P(E_0 | x_{-k}, x_{-k+1}, \ldots, x_0, \ldots, x_{k-1}, x_k)$.

Existing approaches for modeling these PDFs rely either on extensive numerical simulations [2], [8] or on simplified and often inaccurate analytical approximations (for a discussion see [28] and references therein). Numerical approaches approximate the middle part of distributions well, but far tails of PDFs are very hard to obtain

numerically. All the existing analytical approaches neglect non-linear interaction between pulses and noise (which is implicitly incorporated in numerical modeling). This leads to PDFs approximations that are applicable only under very severe restrictions in terms of system speed, distance and types of fiber used (see [26], [11]).

Several recent papers [56], [12], [11], [17], [24] used Karhunen-Loève series expansion (KL) to determine PDFs. In cases where the covariance function is known this method works well [11]. In the case of a general optical fiber communication system, covariance between received pulses is unknown and the KL expansion approach has several drawbacks. Firstly, a covariance matrix needs to be approximated numerically (in which case Karhunen-Loève expansion is in fact the Principle Component Analysis (PCA)) and numerical calculation of eigen-quantities is often unstable. Secondly, a separate set of eigen-quantities needs to be numerically calculated for each bit configuration. For a good discussion see [24], further details can be found in [14], [56] and [11]).

In this work we use a method of optimal fluctuations, or *instantons,* to model far tails, and numerical simulations to refine the middle part of PDFs. We combine these two approaches by using Edgeworth expansion with orthogonal polynomials specially constructed for this problem.

3.2. Probability density functions by using the Instanton approach

Short correlated ASE noise at the position z in the moment t, $\xi(t,z)$, can be considered to be Gaussian with zero mean and uncorrelated (the time in which noise from EFDA is correlated is short even compared with the duration of the bit slot [15]):

$$\langle \xi(t_1,z), \xi(t_2,z) \rangle = N\delta(t_1 - t_2)$$

where N is noise intensity and correlation is denoted by <>. We assume that noise statistics do not depend on transmitted pulses.

We consider that propagation of noise through fiber is governed by:

$$i\vartheta_z(t,z) + u'(z)\vartheta_{tt}(t,z) = i\xi(t,z) \tag{3.1}$$

Nonlinear interaction of the noise is neglected at this point, it will be taking into account in the section. The dispersion map function is $u'(z)$ and $\vartheta(t,0) = 0$.

The solution of Eq. (3.1) is

$$\vartheta(t,z) = \int_0^z \int_R G(t-s,y)\xi(s,y)dsdy$$

where the Green's function of the linear operator $i\partial_z + u'(z)\partial_t^2$ is given by

$$G(t,z) = \frac{\exp\left(\frac{it^2}{4[u(z)-u(0)]}\right)}{\sqrt{4\pi i[u(z)-u(0)]}}.$$

Taking into account that

$$\lim_{u(z_1)\to u(z_2)} \frac{\exp\left(\frac{i(t_1-t_2)^2}{4[u(z_1)-u(z_2)]}\right)}{\sqrt{4\pi[u(z_1)-u(z_2)]}} = \delta(t_1 - t_2),$$

we can get:

$$\lim_{u(z_1)\to u(z_2)} \langle \vartheta(t_1, z_1), \vartheta(t_2, z_2) \rangle = Nz\delta(t_1 - t_2) \tag{3.2}$$

The probability density function of the energy at the center slot is:

$$p(E|s) = \left\langle \delta \left(E - \int_{-T/2}^{T/2} |A_s(t, z) + \vartheta(t, z)|^2 dt \right) \right\rangle_{\vartheta} \tag{3.3}$$

where T is the size of a time slot and s is a bit configuration surrounding the center bit that gives $A_s(t, z)$ via non-linear interaction of the pulses in the absence of noise, here δ stands for the delta function. Averaging $<>_{\vartheta}$ is done over all noise configurations.

Taking into account Eq.(3.2), Eq.(3.3) can be rewritten as:

$$\begin{aligned} p(E|s) = \int d\lambda \int D\vartheta \exp \Bigg[-\frac{1}{2Nz} \int_{-T/2}^{T/2} |\vartheta|^2 dt) \times \\ \times \exp(-i\lambda(E - \int_{-T/2}^{T/2} |A_s(t, z) + \vartheta(t, z)|^2 dt)) \Bigg] \end{aligned} \tag{3.4}$$

where D denotes path integral.

This integral can be estimated by evaluating it around its saddle point(s) (optimal fluctuations). Denoting

$$\Phi = \left(-\frac{1}{2Nz} \int_{-T/2}^{T/2} |\vartheta|^2 dt - i\lambda(E - \int_{-T/2}^{T/2} |A_s + \vartheta|^2 dt) \right)$$

we have:

$$\begin{aligned} \frac{\delta\Phi}{\delta\vartheta} &= 0 \\ \frac{\delta\Phi}{\delta\vartheta^*} &= 0 \\ \frac{\partial\Phi}{\partial\lambda} &= 0 \end{aligned}$$

where δ denotes functional derivative.

This system of equations gives:

$$\begin{aligned} -\frac{1}{2Nz}\vartheta^* + i\lambda(A_s^* + \vartheta^*) &= 0 \\ -\frac{1}{2Nz}\vartheta + i\lambda(A_s + \vartheta) &= 0 \\ E - \int_{-T/2}^{T/2} |A_s + \vartheta|^2 dt &= 0 \end{aligned} \tag{3.5}$$

for all $t \in [-T/2, T/2]$.

This system has only one solution for ϑ given by

$$\vartheta = \frac{2i\lambda Nz}{1 - 2i\lambda Nz} A_s. \tag{3.6}$$

This is the noise configuration that has the greatest influence on energy of the received pulse. It is not surprising that it has the same shape as the pulse A_s.

From the system (3.5) we get:

$$|\vartheta + A_s|^2 = |\frac{2i\lambda Nz}{1 - 2i\lambda Nz} A_s + A_s|^2 = \frac{|A_s|^2}{(1 - 2i\lambda Nz)^2} \tag{3.7}$$

for all $t \in [-T/2, T/2]$. By integrating the last expression and by using the last equation from (3.5) we get that:

$$E = \int_{-T/2}^{T/2} \frac{|A_s|^2}{(1 - 2i\lambda Nz)^2} dt \tag{3.8}$$

This allows us to express the "optimal" λ in terms of A_s and E

$$2i\lambda Nz = 1 - \sqrt{\frac{\int_{-T/2}^{T/2} |A_s|^2}{E}} \tag{3.9}$$

Also

$$\int_{-T/2}^{T/2} |\vartheta|^2 dt = (2i\lambda Nz)^2 E \tag{3.10}$$

Finally, the expression for $P(E|s)$ (4.2) can be estimated by using the last equality from (3.5) and equations (3.9) and (3.10) as:

$$p(E|s) \approx C \exp\left[-\frac{1}{2Nz}\left(\sqrt{E} - \sqrt{\int_{-T/2}^{T/2} |A_s(t,z)|^2}\right)^2\right] \tag{3.11}$$

where C is normalization constant

$$C^{-1} = \sqrt{2\pi N z E_s}\left[\mathrm{erf}\left(\sqrt{\frac{E_s}{2Nz}}\right) + 1\right] + 2\sqrt{Nz}\exp\left(-\frac{E_s}{2Nz}\right) \tag{3.12}$$

.

Here $E_s = \int_{-T/2}^{T/2} |A_s(t,z)|^2 dt$ and erf is the "error function" $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \mathrm{e}^{-t^2} dt$.

It is important to stress that expression Eq. 3.11) is a first order approximation consistent with estimates of probability density functions that exists in literature [26, 11, 56, 12, 17, 25]. For example, considering that the time-bandwidth product in modern systems is between 3 and 5 [11], the instanton approximation is consistent with the approximation in [25].

Noise with any statistics can be easily incorporated as long as it is uncorrelated. The expression $\exp(-\frac{1}{2Nz}\int |\vartheta|^2 dt)$ in Eq.(4.2) is a consequence of using

Gaussian noise. Noise with different statistics will have an expression other than $\exp(-\frac{1}{2Nz}\int|\vartheta|^2dt)$, reflecting different density. If noise is correlated, expressions for the partial derivatives of Φ become integral equations that need to be solved numerically.

3.3. Edgeworth expansion

In this section we use Edgeworth expansion [27] to refine PDFs derived in the previous section.

An unknown distribution $w(x)$ can be represented as

$$w(x) = u(x)\left[\sum_{i=1}^{\infty} C_i P_i(x)\right] \tag{3.13}$$

where $u(x)$ is a starting approximate distribution and $P_i(x), n \in N$ is a family of polynomials orthogonal with respect to the weight $u(x)$. Let

$$P_i(x) = \sum_{k=0}^{i-1} a_k x^k$$

.

By multiplying Eq. (3.13), by $P_i(x)$ and integrating over the domain of orthogonality (in our case $x \in (0, +\infty)$ since x represents energy) we get

$$\sum_{k=0}^{i-1} a_k \eta_k = C_i \tag{3.14}$$

where $\eta_k, k \in N$ are moments of the distribution $w(x)$. We can obtain a finite number j of these moments numerically or experimentally, therefore deriving an approximate PDF $\tilde{w}_j(x)$ by truncating the infinite sum Eq. (3.13) to j terms. These approximate PDFs $\tilde{w}_j(x), j \in N$ are guaranteed to converge to $w(x)$ uniformly [27].

3.3.1. Orthogonal polynomials

In order to use PDFs given in Eq. 3.11 as starting distributions we need polynomials $P_n(x; m; p), n \in N$ orthogonal with respect to the weight

$$u(x; m; p) = \mathrm{e}^{-m(\sqrt{x}-p)^2}, \quad x \in (0, +\infty) \qquad (m, p > 0). \tag{3.15}$$

These polynomials can be seen as generalization of Laguerre polynomials and to the best of our knowledge have not been studied before. Here we shall explain how to construct them. Moment of the distribution $u(x; m; p)$ can be written as

$$\mu_n(m; p) = 2 \int_{-p}^{+\infty} (t + p)^{2n+1} \mathrm{e}^{-t^2} \, dt.$$

We denote

$$\sigma_n(p) = 2 \int_{-p}^{+\infty} (t + p)^{n} \mathrm{e}^{-t^2} \, dt,$$

and by partial integration rule, it can be shown that

$$\sigma_{n+2}(p) = p \; \sigma_{n+1}(p) + \frac{n+1}{2} \sigma_n(p) \quad (n = 0, 1, \ldots),$$

$$\sigma_0(p) = \sqrt{\pi} \; \big(1 + \mathrm{erf}(p)\big), \quad \sigma_1(p) = \mathrm{e}^{-p^2} + p \; \sigma_0(p).$$

Expressing

$$\sigma_n(p) = r\ \gamma_n(p) + s\ \delta_n(p),$$

where

$$r = \mathrm{e}^{-p^2}, \qquad s = \sigma_0(p),$$

and using the previous relation, we find that the polynomial sequences $\{\gamma_n(p)\}$ $(\deg(\gamma_n) = n-1)$ and $\{\delta_n(p)\}$ $(\deg(\delta_n) = n)$ satisfy the same recurrence relation with only difference in the initial values:

$$\gamma_{n+2}(p) = p\ \gamma_{n+1}(p) + \frac{n+1}{2}\gamma_n(p), \qquad \gamma_0 = 0, \quad \gamma_1 = 1;$$

$$\delta_{n+2}(p) = p\ \delta_{n+1}(p) + \frac{n+1}{2}\delta_n(p), \qquad \delta_0 = 1, \quad \delta_1 = p.$$

By mathematical induction and the fact $\mu_n := \sigma_{2n+1}$, we prove:

The moments $\{\mu_n\}$ can be expressed by

$$\mu_n(p) = \frac{1}{2^n}\left\{R_n(p^2)\ \mathrm{e}^{-p^2} + S_n(p^2)\ p\ \sqrt{\pi}\big(1 + \mathrm{Erf}(p)\big)\right\},$$

where

$$\frac{R_n(p^2)}{2^n} = \gamma_{2n+1}(p), \qquad \frac{pS_n(p^2)}{2^n} = \delta_{2n+1}(p) \qquad (n \in \mathbb{N}_0).$$

Example 1.1. The first members of the sequences $R_n(p)$ and $S_n(p)$ are:

$$R_0(p) = 1, \qquad S_0(p) = 1,$$

$$R_1(p) = 2(p+1), \qquad S_1(p) = 2p+3,$$

$$R_2(p) = 2(2p^2+9p+4), \qquad S_2(p) = 4p^2+20p+15.$$

Knowing all the moments $\mu_n(m;p)$ we can calculate the polynomials by:

$$P_n(x;m;p) = \frac{1}{\sigma_1^n(p)} \begin{vmatrix} \sigma_1(p) & \frac{\sigma_3(p)}{m^2} & \cdots & \frac{\sigma_{2n-1}(p)}{m^{2n-2}} & 1 \\ \frac{\sigma_3(p)}{m^2} & \frac{\sigma_5(p)}{m^4} & & \frac{\sigma_{2n+1}(p)}{m^{2n}} & x \\ \vdots & & \ddots & & \\ \frac{\sigma_{2n+1}(p)}{m^{2n}} & & & \frac{\sigma_{2(2n-1)+1}(p)}{m^{2(2n-1)}} & x^n \end{vmatrix}.$$

Coefficients of these polynomials quickly become very cumbersome if expressed in general terms i.e. as functions of e^{-p^2} and erf(p). However for a given numerical value of $p = \sqrt{\int_{-T/2}^{T/2} |A_s(t,z)|^2}$ there are no computational problems.

Example 2.1. We present the first members of the sequence $Q_n(x;m;p) = 2^{n(2n-1)} P_n(x;m;p)$ developed over r and s. Multiplication by $2^{n(2n-1)}$ is done so that numeric coefficients are integers.

$$Q_0(x;p) = r + ps,$$

$$Q_1(x;p) = r(-2-2p^2+2x) + s(-3p-2p^3+2px),$$

$$Q_2(x;p) =$$

$$r^2(128+624p^2+480p^4+128p^6+(-256-976p^2-288p^4)x+(64+160p^2)x^2)+$$

$$rs(456p+1552p^3+1088p^5+256p^7+(-792p-1952p^3-544p^5)x+(176p+288p^3)x^2)$$

$$+s^2(360p^2+960p^4+608p^6+128p^8+(-480p^2-960p^4-256p^6)x+(96p^2+128p^4)x^2).$$

Or, developed over x, they look like:

$$
\begin{aligned}
&Q_0(x;p) = r + ps,\\
&Q_1(x;p) = 2(r+ps)x - 2(1+p^2)r - (3+2p^2)ps,\\
&Q_2(x;p) = \Big((64+160p^2)r^2 + (176p+288p^3)rs + (96p^2+128p^4)s^2\Big)x^2\\
&+\Big((-256-976p^2-288p^4)r^2 + (-792p-1952p^3-544p^5)rs\\
&+(-480p^2-960p^4-256p^6)s^2\Big)x\\
&+(128+624p^2+480p^4+128p^6)r^2 + (456p+1552p^3+1088p^5+256p^7)rs\\
&+(360p^2+960p^4+608p^6+128p^8)s^2.
\end{aligned}
$$

3.4. Numerical results

In this section we shall illustrate the developed method. We considered the system in Fig. 3.1. As mentioned, this system consists of periodically distributed sections of fiber with positive $\beta_2(z) = D_+$ and negative dispersion $\beta_2 = D_-$ separated by amplifiers (EDFA). One *span* consist of one section of fiber with positive dispersion, one section of fiber with negative dispersion, and corresponding amplifiers (see Fig. 3.1).

The transmission of a signal through the fiber is modeled by the nonlinear Schrödinger equation (NLSE)

$$\frac{\partial A}{\partial z} = -\frac{\alpha}{2}A - \frac{i}{2}\beta_2(z)\frac{\partial^2 A}{\partial t^2} + \frac{\beta_3(z)}{6}\frac{\partial^3 A}{\partial t^3} + i\gamma|A|^2A \tag{3.16}$$

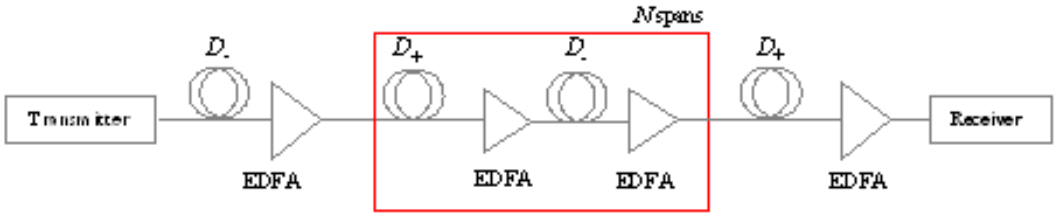

FIGURE 3.1. Scheme of the optical transmission system considered

where z is the propagation distance along the fiber, relative time $t = t_{real} - z/v_g$ gives a frame of reference moving at the group velocity v_g, $A(z,t)$ is the complex field amplitude of the pulse, α is the attenuation coefficient of the fiber, $u'(z)$ is the group velocity dispersion (GVD) coefficient, $\beta_3(z)$ is the second-order GVD, γ is the nonlinearity coefficient giving rise to Kerr effect nonlinearities: self-phase modulation (SPM), intrachannel cross-phase modulation (IXPM) and intrachannel four-wave mixing (IFWM). In short, this calculation takes into account modulation, extinction ratio, realistic models of transmitter, optical filter and electrical filter, crosstalk effects, Kerr nonlinearities, ASE noise, and dispersion effects (GVD and second order GVD). In the system simulator, propagation of pulses through the system, i.e. solving NLSE, was done numerically by the split-step Fourier method [10].

The parameters of positive dispersion D_+ and negative dispersion D_- fibers are given in Table 3.1. Pre-compensation of -330 ps/nm and corresponding post-compensation were also applied. The RZ modulation format has duty cycle of 33%,

and the launched power was set to 0dBm. EDFAs with noise figure of 8dB were deployed after every fiber section, the bandwidth of optical filter was set to $3R_b$ and the bandwidth of electrical filter to $0.65R_b$, with R_b being the bit rate (40 Gb/s).

TABLE 3.1. Parameters of the fibers used

Parameters	D_+ fiber	D_- fiber
Dispersion [ps/(nm km)]	20	-40
Dispersion Slope [ps/(nm^2 km)]	0.06	-0.12
Effective Cross-sectional Area [μm^2]	110	50
Nonlinear refractive index [m^2/W]	2.2×10^{-20}	2.2×10^{-20}
Attenuation Coefficient [dB/km]	0.19	0.25
Length (in one span) [km]	33.4	16.7

We first ran the system simulator in the training mode (without noise) with a random bit sequence (of length 2^{17}) to obtain values $E_s = \int_{-T/2}^{T/2} |A_s(t,z)|^2$ needed in the probability density functions Eq. (3.11).

The nonlinear distance of this system is roughly 6000km. As expected, for distances below this number instanton approximation itself approximates the true PDF well. To illustrate this, in the Fig. 3.2 and Fig.3.3, we plot both PDFs obtained by instanton method and histograms obtained numerically, for two bit configurations: (i) with zero in the center slot $s = 0110110$ and (ii) with "1" at the center slot $s = 0001000$. Both bit configurations were propagated through 100 spans, that is 5000km. The difference between Fig. 3.2 and Fig. 3.3 is that Fig. 3.2 uses linear scale while Fig. 3.3 is uses log scale and that the energy is normalized so that the peak of the PDF for the bit pattern "0001000" corresponds to 1.

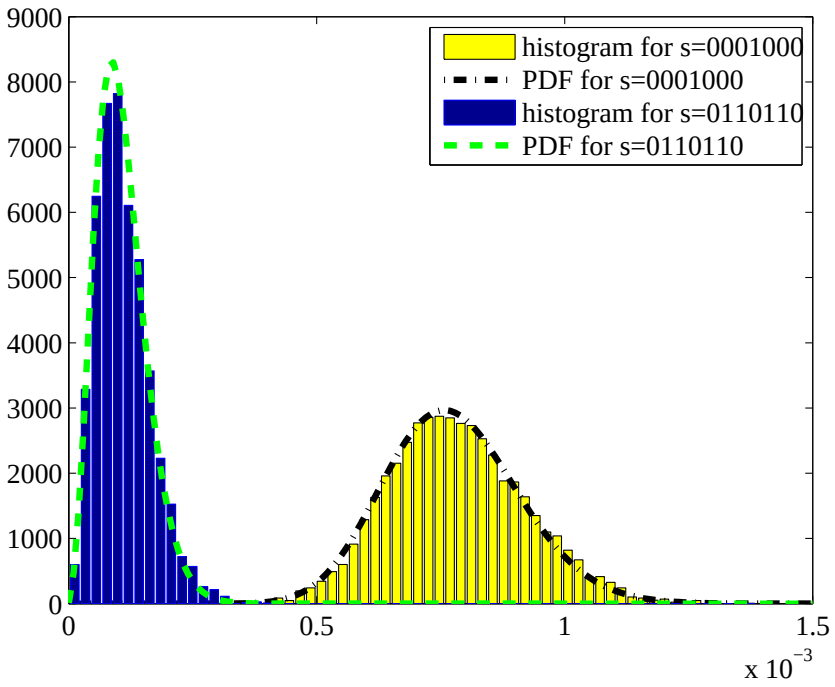

FIGURE 3.2. Probability density functions after 100 spans

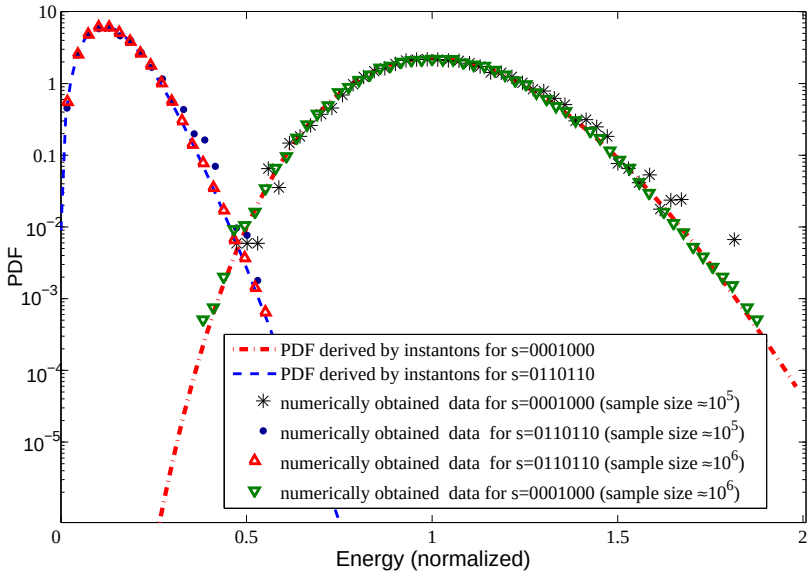

FIGURE 3.3. Probability density functions after 100 spans. Normalization is chosen so that the peak of the PDF for the bit pattern "0001000" corresponds to 1.

However, when the propagation distance is longer, the instanton approximation is not sufficient. As Fig. 3.4 and Fig. 3.5 show, after 300 spans (approximately two and a half nonlinear distances) neglecting the nonlinear interaction between noise and pulses in the instanton approximation makes this approximation overly "optimistic", i. e., too narrow. The difference is even more apparent if log-scale is used as Fig. 3.6 and Fig. 3.7 show.

After a polynomial correction is added, the refined PDFs differ very little from the numerically obtained data. We used 2^{17} numerically obtained samples to calculate the distribution moments needed for the polynomial correction. Further numerical simulations of length 2^{20}, shown with green triangles in Fig. 3.5 and Fig. 3.4, were not used to determine the curves, but show further agreement. Note that only polynomials up to fourth order are needed to refine the PDF for "0110110", and only the first two polynomials are needed for "0001000". This is in sharp difference from the result reported in [28], where, for a similar system, more than a dozen Hermite polynomials are needed to sufficiently improve the starting Gaussian distribution (also presented in figures).

Of course, the question of how many polynomials are needed to approximate an unknown distribution sufficiently well. From the engineering point of view, a satisfactory answer can be to add higher order polynomials until the moment difference between two consecutive refinements falls under certain threshold.

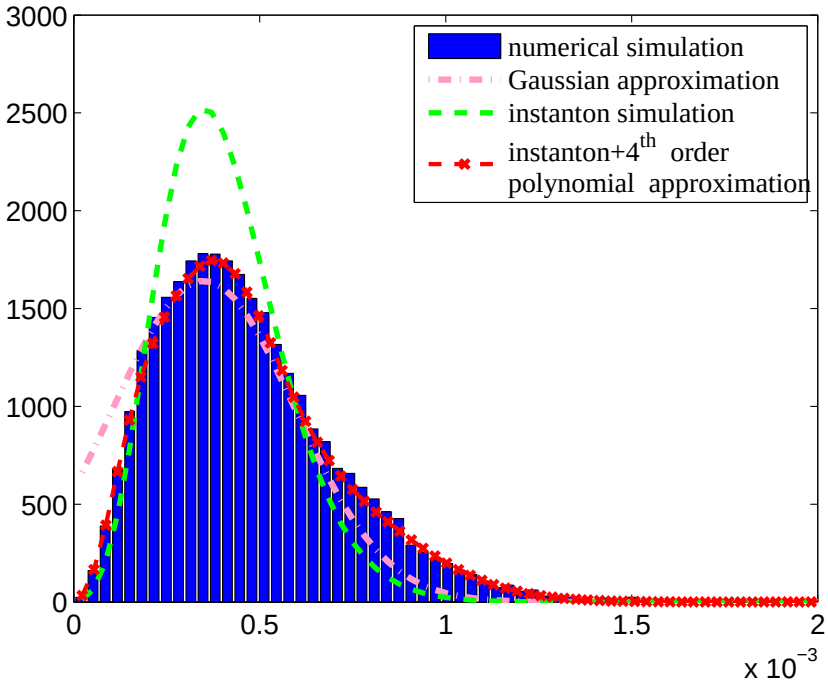

FIGURE 3.4. Comparison of various PDF approximations for energy of bit pattern "0110110" after 300 spans

3.5. Final remarks on channel characterization

We developed an approach for approximating probability density functions that is both practical and accurate. The instanton method gives suitable asymptotic behavior for the tails of distributions. Use of the parameterized family of orthogonal polynomials reduces computer processing cost, making this approach applicable for high speed applications.

The method is also very general (it is not restricted to specific pulse shaping, bit

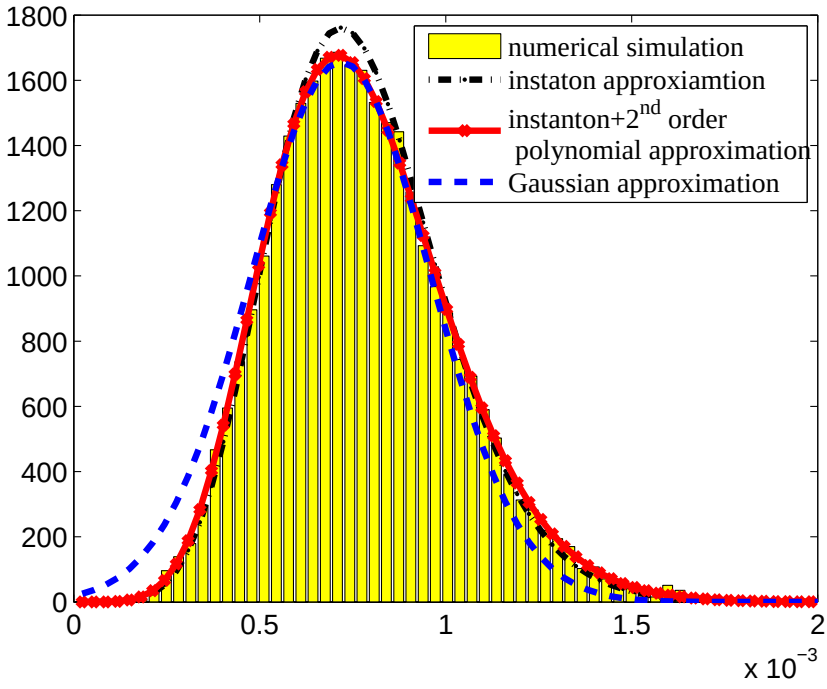

FIGURE 3.5. Comparison of various PDF approximations for energy of bit pattern "0001000" after 300 spans

rate, propagation distance) and therefore applicable to a wide range of systems.

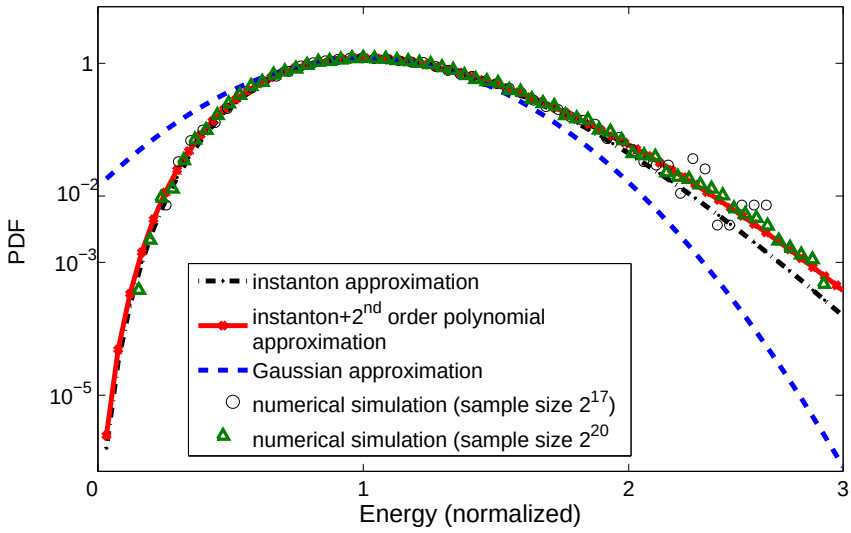

FIGURE 3.6. Comparison of various PDF approximations for center bit energy within bit pattern "0001000" after 300 spans. Normalization is chosen so that the peak of the PDF corresponds to 1.

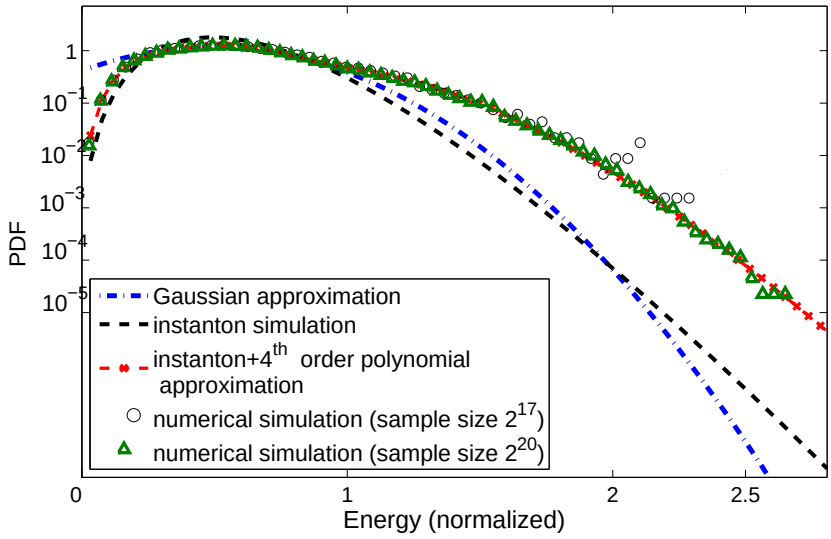

FIGURE 3.7. Comparison of various PDF approximations for center bit energy within bit pattern "0110110" after 300 spans. Normalization is chosen so that 1 on the x-axis corresponds to 1 as in Fig. 3.6.

4. Channel capacity

4.1. Introduction

The problem of determining capacity of optical transmission systems has been addressed by numerous researchers. For example, Mitra and Stark [29] approximated a nonlinear noisy channel by a linear one with an effective nonlinear noise. Such an approach, as indicated in [30], is valid only for specific transmission systems. The results obtained by Tang in [33] are based on solving the NLSE by Volterra series expansion up to the first order. The channel capacity is determined using the Pinsker's formula, which is valid only if noise is Gaussian. The properties of optical transmission in an analog noisy nonlinear channel with weak dispersion management, zero average dispersion and unlimited power was considered by Turitsyn *et al.* in [30].

We investigate single channel systems with return-to-zero (RZ) pulses (with the same phase) and erbium-doped fiber amplifiers (EDFA). Data is encoded by presence (or absence) of a pulse in a bit slot (on-off keying).

Following the approach from our recent publication [2], we use a method based on the Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm [34] originally proposed in the context of magnetic channels (see [35] and [36]). This method is centered around

probability density functions (PDF) of the energy at a bit slot given bit configuration surrounding that slot. In [2] those functions were numerically approximated, but the calculations were computationally intensive and the results were very sensitive to the precision of the calculated PDFs.

In the numerical results section we apply the proposed method on a set of optical communication systems similar to the system consider in Chapter II, the only difference being varying noise levels.

We estimate achievable information rate of a high-speed (40Gbs/s) long-haul optical transmission with an independent uniformly distributed (i.u.d.) source when the combined effects of amplified spontaneous emission (ASE) noise, Kerr nonlinearity, stimulated Raman scattering (SRS) and chromatic dispersion (group velocity dispersion (GVD) and second-order GVD) are considered.

4.2. Information rate and capacity for optical transmission systems

An optical transmission system (channel) is modeled as an intersymbol interference channel (ISI), where m previous and m next bits influence the observed bit. The fact that observed bit depends on next bits as well as on previous bits is one of the differences between optical channels and magnetic channels ([35] and [37]), where only previous bits influence the observed bit's value. Because of this, the allowed transitions within the complete trellis still depend on the right most bit (since we are

using forward recursion), but the branch metrics depends mostly on the bit at the center slot.

A communication channel is described by the conditional density function of the channel output vector $\mathbf{y} = (y_1, \ldots y_n)$ with $y_i \in Y$, given the channel input (source) vector $\mathbf{x} = (x_1, \ldots x_n)$ with $x_i \in X$. In our case inputs are binary, that is $X = \{0, 1\}$. The channel is completely defined by X, Y and the conditional probability function $P(\mathbf{Y}|\mathbf{X})$. We shall use $\mathbf{X}_i^j$ to denote $(X_i, \ldots X_j)$.

Information rate is defined as

$$I(\mathbf{Y}; \mathbf{X}) = H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X}) \tag{4.1}$$

where $H(U)$ is entropy defined as $H(U) = -\mathrm{E}(\log_2 P(U))$ for a random variable U. Channel capacity is defined as:

$$C = \max I(\mathbf{Y}; \mathbf{X})$$

where the maximization is performed over all possible input distributions.

One way of achieving performance close to channel capacity, in a channel with memory (such as an optical channel), is to use non-linear codes (see for example [40]). Here we address a problem often encountered in practice: calculating information rate in the case of independent and uniformly distributed (i.u.d) channel input source. Therefore, we calculate a lower bound on the channel capacity.

The Shannon-McMillan-Breiman theorem [41] gives that:

$$\lim_{n\to\infty} \frac{1}{n} \log_2 P(\mathbf{Y}_1^n) = \mathrm{E}(\log_2 P(\mathbf{Y})) \quad \text{almost surely.}$$

So the problem of estimating the information rate given by Eq.(4.1) can be reduced to generating a long sequence $\mathbf{y_1^n}$ and calculating $\log_2 P(\mathbf{y_1^n})$.

Expression (4.1) reduces to:

$$I(Y;X) = \lim_{n\to\infty} \frac{1}{n} \left[\sum_{t=1}^{n} \log_2 P(y^t|\mathbf{y}_1^{t-1}, \mathbf{x}_1^n) - \sum_{t=1}^{n} \log_2 P(y_t|\mathbf{y}_1^{t-1}) \right] \tag{4.2}$$

4.2.1. BCJR algorithm

To calculate an estimate of $P(y_t|\mathbf{y}_1^{t-1})$ we use a variant of the BCJR algorithm [34].

Bit configurations (states) can be ordered corresponding to integer value of states seen as binary numbers. We shall refer to states by their index, i.e, state s_t shall be referred to as t.

The forward recursion of the BCJR algorithm is a computation of $P(y_t|y_1^{t-1})$ for $t = 1 \ldots, n$ given by:

$$P(y_t|y_1^{t-1}) = \sum_{i,j} \alpha_{t-1}(i) P_{ij} P(y_t|j)$$
$$\alpha_t(s) = \frac{\sum_i \alpha_{t-1}(i) P_{is} P(y_t|s)}{\sum_{i,j} \alpha_{t-1}(i) P_{ij} P(y_t|j)},$$

where the posterior source/channel state probability mass function $\alpha_t(s)$ is defined as:

$$\alpha_t(s) = P(S_t = s|Y_1^t = y_1^t).$$

The factor P_{ij} is a probability of transition from i to j. In the case of a binary i.u.d. source it is $\frac{1}{2}$ for two possible transitions and zero for others. Probability functions $P(y_t|j)$ come from the physical properties of the channel -we can use the PDFs derived in the previous chapter.

Before we move further, note that the second term from Eq.(4.2), $P(y^t|y_1^{t-1}, x_1^n)$, is exactly equal to $P(y_t|x_{t-m}^{t+m})$. Once we find probability density functions (as in Chapter II) we can determine an estimate of achievable information rate as follows:

- Optical transmission system is simulated in the absence of noise to determine the values $\int_{-T/2}^{T/2} |A_s(t,z)|^2$.
- The system is simulated with noise to obtain sequence $\mathbf{y}_1^n$. Value for n should be $\sim 10^6$ [37].

- An estimate for the achievable information rate is obtained from Eq.(4.2) and Eq.(3.11).

4.3. Numerical results

To illustrate the method, we considered the system as in Chapter II.

We first ran the system simulator without noise and with a random bit sequence (of length 2^{17}) to obtain the values $E_s = \int_{-T/2}^{T/2} |A_s(t,z)|^2$ needed in the probability density functions Eq. (3.11).

We then ran the system simulator to obtain the sequences $\mathbf{y}_1^n$ for different numbers of spans. BCJR was applied to estimate achievable information rate via Eq. (4.2). We applied BCJR for different values of m -number of slots (memory) on each side of center slot. This is of interest because designing electronics that would perform BCJR on large number of states (2^{2m+1}) is not a trivial task, especially for systems operating at 40Gb/s. Results are presented in Fig. 4.1. It can be seen that the achievable information rate is close to one bit per channel use for reasonable distances (total length of one span is 50km). However, for high numbers of spans, achievable information rate drops significantly with shortening memory.

At the end we compare sensitivity of the results to the amplifier noise. Current EDFA introduce noise at levels of 4-8dB, so we considered these levels. Results are

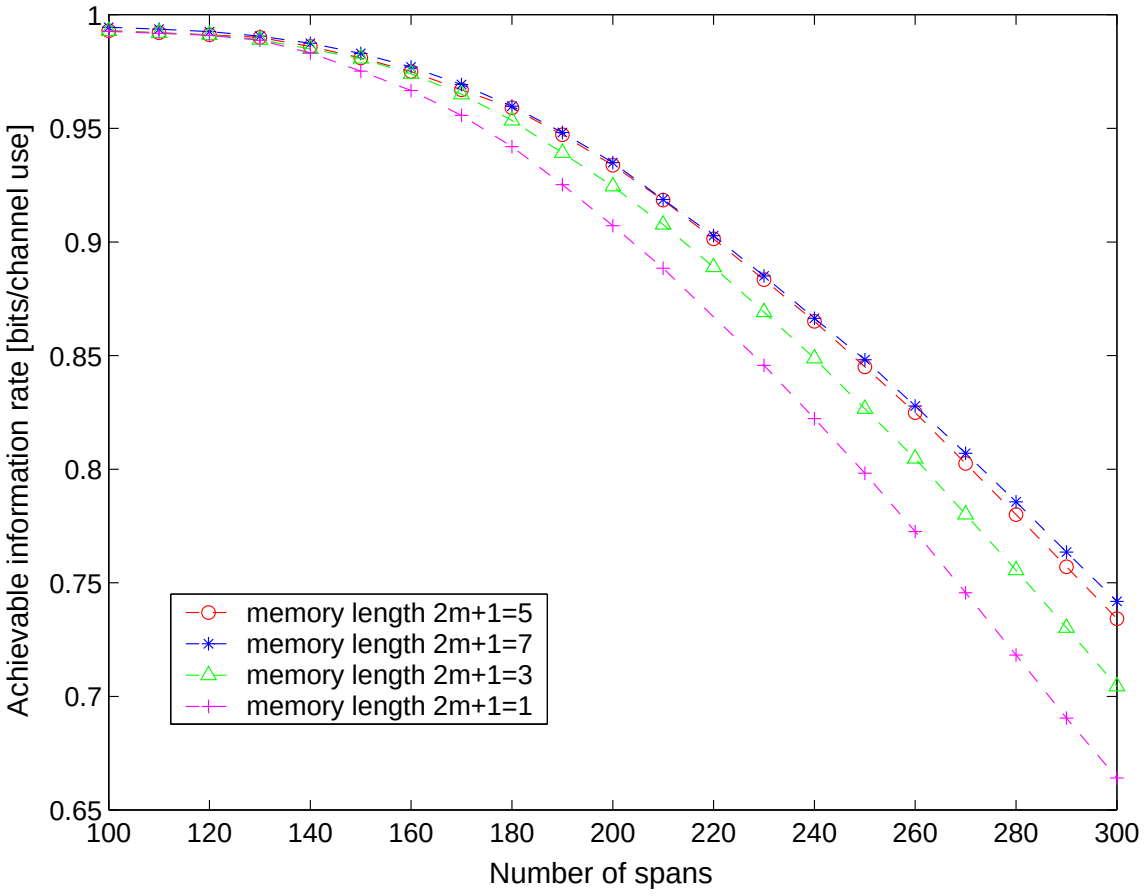

FIGURE 4.1. Achievable information rates for different memory lengths

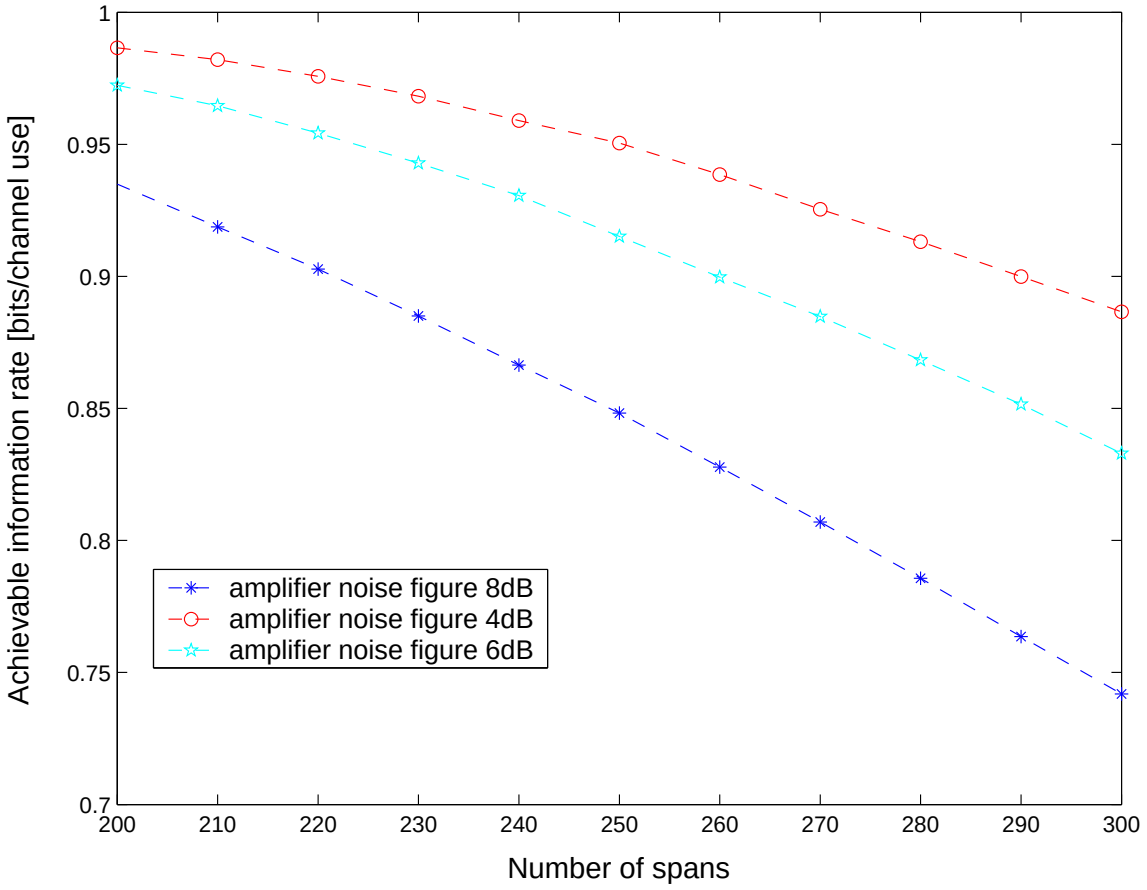

FIGURE 4.2. Achievable information rates for different noise strength

presented in the Fig. 4.2. Memory length in these simulation was $2m + 1 = 7$.

4.4. Final remarks on the proposed method for estimating achievable information rate

We propose an approach for estimating achievable information rate that is flexible and applicable to a wide range of transmission systems. The approach is based on

a combination of a Monte-Carlo method for estimating entropy and an analytical method for estimating channel properties.

As an illustration, the method was applied to a submarine optical transmission system operating at a transmission rate of 40 Gb/s. Numerical results indicate that, with an appropriate receiver and for reasonable distances, this kind of a system can achieve information rates very close to one bit per channel use.

5. Coding schemes

To deal with chromatic dispersion and polarization mode dispersion (PMD) the use of maximum-likelihood sequence detection (MLSD) based on the Viterbi algorithm has been recently proposed ([12], [44], [56]). On the other hand, the discovery by MacKay and Neal that long Low-Density Parity-Check (LDPC) codes with sum-product decoding algorithm achieve near-optimum performance on an additive white Gaussian noise (AWGN) channel [45] led to a lot of research on these codes. However, the output of the Viterbi algorithm is the most likely sequence (hard output). This makes a Viterbi detector unsuitable for coupling with any soft iterative decoding scheme, such as Low-Density Parity-Check codes (in specific with sum-product decoding) or turbo codes.

In this chapter we suggest two similar coding schemes: the first scheme is a "hard decision" scheme. It uses LDPC codes with hard decision "Gallager B" decoding algorithm [46, 55] coupled with the Viterbi algorithm. The data stream is first encoded by an LDPC code and then transmitted through the system. At the receiver, the Viterbi algorithm is applied first and then data is decoded with Gallager B algorithm.

Unlike soft decoding schemes, all the calculations performed in Gallager B are done over binary field. This makes this decoder suitable for high speed applications

like optical transmission systems.

The second coding scheme is a "soft decision," i.e., it is operating on probabilities. This scheme is a modification of a method, originally proposed in the context of magnetic channels (see [57] and [58]), based on the Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm [34], coupled with LDPC codes. BCJR is used as a method for suppressing nonlinear ISI, because its output is probabilities (likelihoods) of individual symbols being "0" or "1". These likelihoods are then used as inputs for an LDPC decoder that further improves performance (bit error rate). LDPC codes are chosen because of their excellent performance [45]; in particular, these codes perform better than turbo codes as the length of codewords increases [57].

In the numerical results section we apply the proposed schemes to a realistic optical communication system. This system is very similar to the systems used in previous sections -it belongs to the class of dispersion-managed systems, where pieces of optical fiber with positive and negative dispersion are periodically incorporated. The system operates at 40Gb/s, but this time the length of one span is 120km. We present numerical results for various LDPC codes and it is shown that the proposed schemes provides significant coding gain with respect to the uncoded system and MLSD (Viterbi equalizer only). As expected the soft-decision decoding scheme performs (much) better that the hard decision one.

5.1. Hard Decision Coding Scheme Description

In the proposed scheme data is first encoded by an LDPC code and then transmitted over an optical transmission system. See Fig. 5.1.

As usual, the optical transmission system is modeled as an ISI channel, where m previous and m following bits influence the observed bit. At the receiver the Viterbi equalization is applied first. The biggest problem arising when the Viterbi algorithm is applied to this channel is the determination of branch metrics [56], i.e. probability density functions (PDFs) for received energy at the observed bit. This is where the method from Chapter III can be applied. After that, if necessary the LDPC decoder is used.

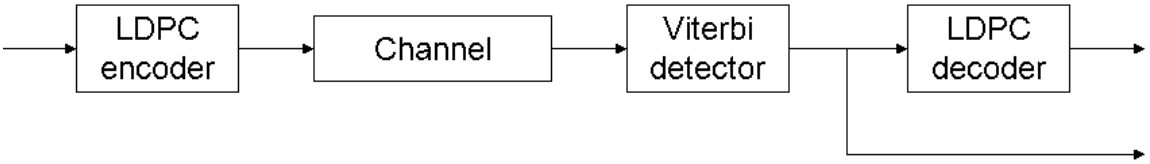

FIGURE 5.1. Block diagram of the proposed "hard decision" coding scheme

There are many variations of iterative message passing decoding algorithm for LDPC codes. The one that we use, Gallager B, was introduced by Gallager [46], and it was named later [55]. For completeness we cite this decoding algorithm here.

5.1.1. Gallager B decoding algorithm for LDPC codes

We consider a regular binary LDPC code of length n. Let the received word consist of bits $(x_1, x_2, \ldots, x_n)$ referred to also as variables. We have parity check matrix H of dimension $(n-k) \times n$, $k < n$, which has K 1's in each row and J 1's in each column. As such, each variable bit x_i, $1 \leq i \leq n$, is involved in J parity-check equations:

$$[x_1, x_2, \ldots, x_n]H^T = [c_1, c_2, \ldots, c_{n-k}]$$

where all operations are in binary field and c_t corresponds to the value of t-th parity-check sum for $1 \leq t \leq n-k$. The vector $[c_1, c_2, \ldots, c_{n-k}]$ is called *the syndrome*. Parity check c_t is said to be *satisfied* if $c_t = 0$ and *unsatisfied* if $c_t = 1$.

To visualize message passing decoding algorithms for LDPC codes, the parity-check matrix is usually represented as a bipartite graph with two types of vertices [48]. The first subset of vertices is comprised of code bits (variables), and the second subset of vertices is comprised of parity-check equations. An edge between a bit and an equation exists if the bit is involved in the check. Let $E(v)$ and $E(c)$ denote the set of edges incident on node v and c, respectively, an edge will be denote by e. Also, let $r(v)$ denote the received value of node v.

Gallager B decoding algorithm for LDPC codes consists of the following steps [55].

1. Each variable node v sends $r(v)$ along each edge in $E(v)$.

2. Each check node c sends $m(e)$ along each edge in $E(c)$ where

$$m(e) = (\sum_{e' \in E(c) \setminus \{e\}} m(v')) \mod 2$$

3. Each variable node v sends $m(e)$ along each edge in $E(v)$ where

$$m(e) = \begin{cases} 1, |\{e' \in E(v) : m(e') = 1\}| > \lceil J/2 \rceil \\ 0, |\{e' \in E(v) : m(e') = 0\}| > \lceil J/2 \rceil \\ r(v), \quad \text{otherwise} \end{cases}$$

The final decision is made on the basis of estimate majority. Steps 2 and 3 are iterated until a valid codeword is obtained or some fixed number of iterations are reached, in which case a decoding failure is declared.

5.2. Soft decoding scheme

The proposed coding scheme is as follows. Data is encoded by a LDPC code and then transmitted through the channel (optical transmission system), see Fig. 5.2. At the receiver end a BCJR detector begins by taking received channel samples (output of the electrical filter) as inputs and produces output probabilities. If a valid codeword is reached decoding halts, if not, the sum-product LDPC decoder [52] starts by taking the produced probabilities as inputs. We shall distinguish two variants of the scheme:

(a) The LDPC decoder iterates until a valid codeword is obtained or a maximal number of iterations is reached (in which case decoding failure is reported).

(b) The LDPC decoder iterates until a valid codeword is obtained or a certain number of iterations is reached and then the soft outputs of the LDPC decoder are passed back to the BCJR as inputs [57]. So soft information is iterated between BCJR and LDPC decoders a certain number of times in a way that is analogous to turbo decoding.

In the next section we shall the examine numerical performance of both options. As expected, decoding scheme (b) performs better, but at the expense of greater complexity and decoding delay (this is because the BCJR decoder is serial, see [58]).

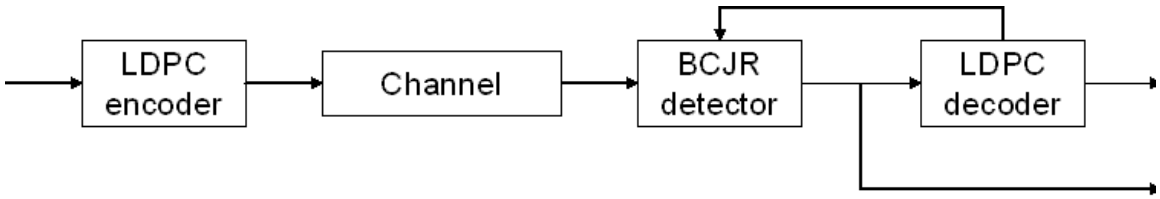

FIGURE 5.2. Block diagram of the proposed "soft decision" coding scheme

The BCJR algorithm, LDPC codes and sum-product decoding algorithm are well known [52] and will not be repeated here. Note, however, that an optical transmission system is modeled as an ISI channel, where m previous and m next bits influence the observed bit. The fact that the observed bit depends on following bits as well as on previous bits is one of the differences between optical channels and magnetic channels [57], [37], where only previous bits influence observed bit's value. This difference requires adjustments in the BCJR algorithm.

5.3. Numerical results

To illustrate the methods, we again considered a dispersion managed system. This is a single channel system with return-to-zero (RZ) pulses, operating at 40Gb/s, with duty cycle of 33%, and the launched power of 0dBm. The erbium-doped fiber amplifiers (EDFAs) with noise figure of 6dB are deployed after every fiber section to compensate for the fiber loss. One span consist of one section of fiber with positive dispersion, one section of fiber with negative dispersion, and corresponding amplifiers. The other parameters of the fibers are (also) the same like in previous chapters.

5.3.1. Hard decision scheme

In the Fig. 5.3, we present performance improvement due to proposed hard-decision error correcting scheme. We used several different codes.

The code denoted by "LC" is one of, so called, lattice LDPC codes of length 8547, rate 0.81, column weight 4 and girth 8 [49].

The code denoted by "BC" is one of block circulant LDPC codes of length 4320, rate 0.75, column weight 3 and girth 8 [50].

The code denoted by "MK" is one of MacKay's codes LDPC codes of length more than 8000, rate 0.5, column weight 3 and girth 6 [51].

Code denoted by "PG" is is a projective geometry code of length 4161, rate 0.825

column weight 65 and girth 6 [52].

The number of slots (memory, m) on either side of the center slot was set to 3 in these simulations. The maximal number of iterations for all the decoders was 200.

Although the code rates used are relatively high all the codes except for MK outperform the system with Viterbi equalizer significantly. It is important to notice that up to 30 spans all errors were corrected for BC, LC and PG codes.

The theory of iterative decoding for LDPC codes is not developed in full (so extensive simulations are needed), but we know that the relatively weak performance of MK code under the hard decision algorithm is due to low column weight and low girth [53]. The good performance of the PG code is probably due to its high column weight (65). It is important to notice that such a high column weight makes the decoder for this code very complex and not practical. Note that LC and BC codes have very similar performance although they have different lengths and rates. This is a demonstration of the high flexibility that LDPC codes offer.

We also ran the Viterbi algorithm for different values of m. This is of interest because designing electronics that would perform Viterbi on high number of states (2^{2m+1}) is not a trivial task, especially for systems operating at high speed. Results are presented in the Fig. 5.4. The curve with cycles on this plot was obtained with m=0, so in that case Viterbi equalizer was omitted, and performance improvement is due to the LDPC code solely.

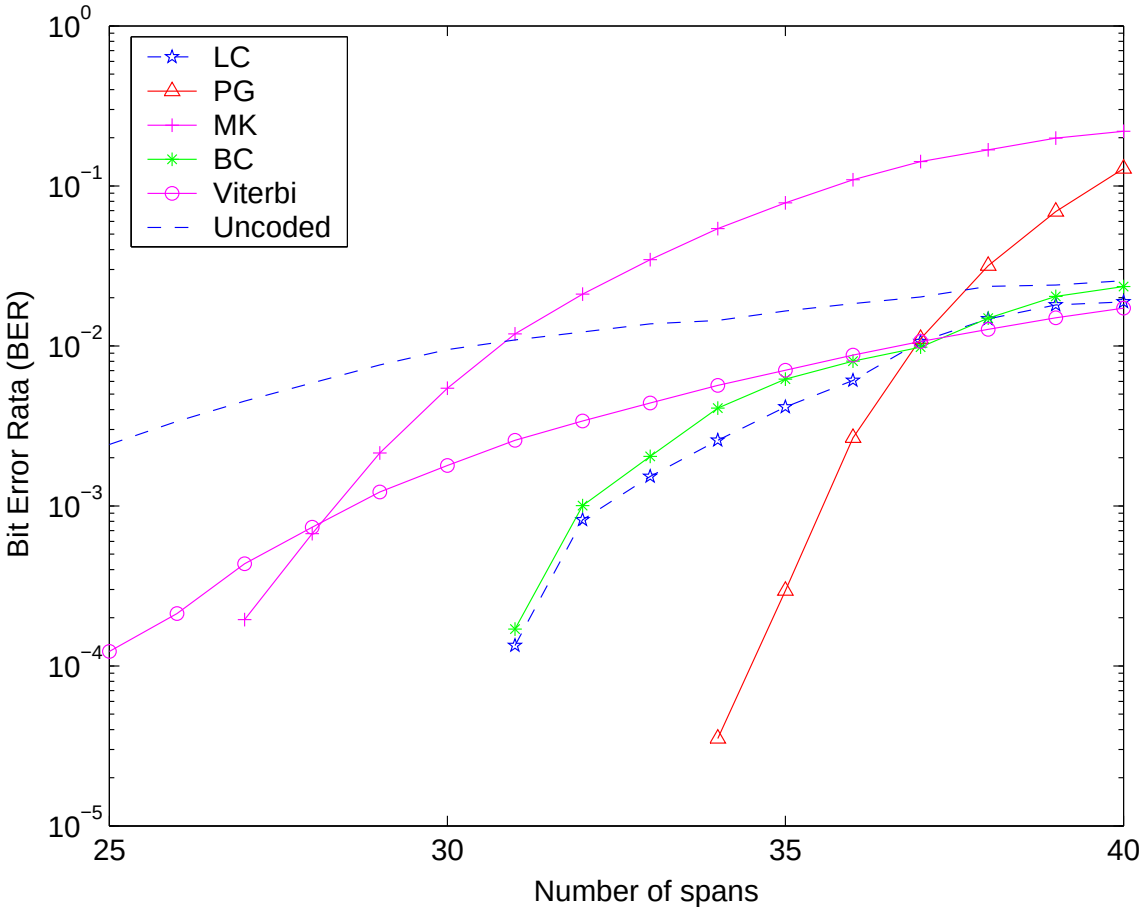

FIGURE 5.3. BER performance of proposed hard-decision error correcting scheme with different LDPC codes

5.3.2. Soft decision scheme

We ran the system simulator with different (soft) coding schemes and different numbers of spans. In Fig. 5.5, the Viterbi detector had a sliding window of 32 bits, and the BCJR algorithm operated on bit configurations with $m = 3$ slots on each side of center slot. We used one of the lattice LDPC codes of length 8547, rate 0.81, column

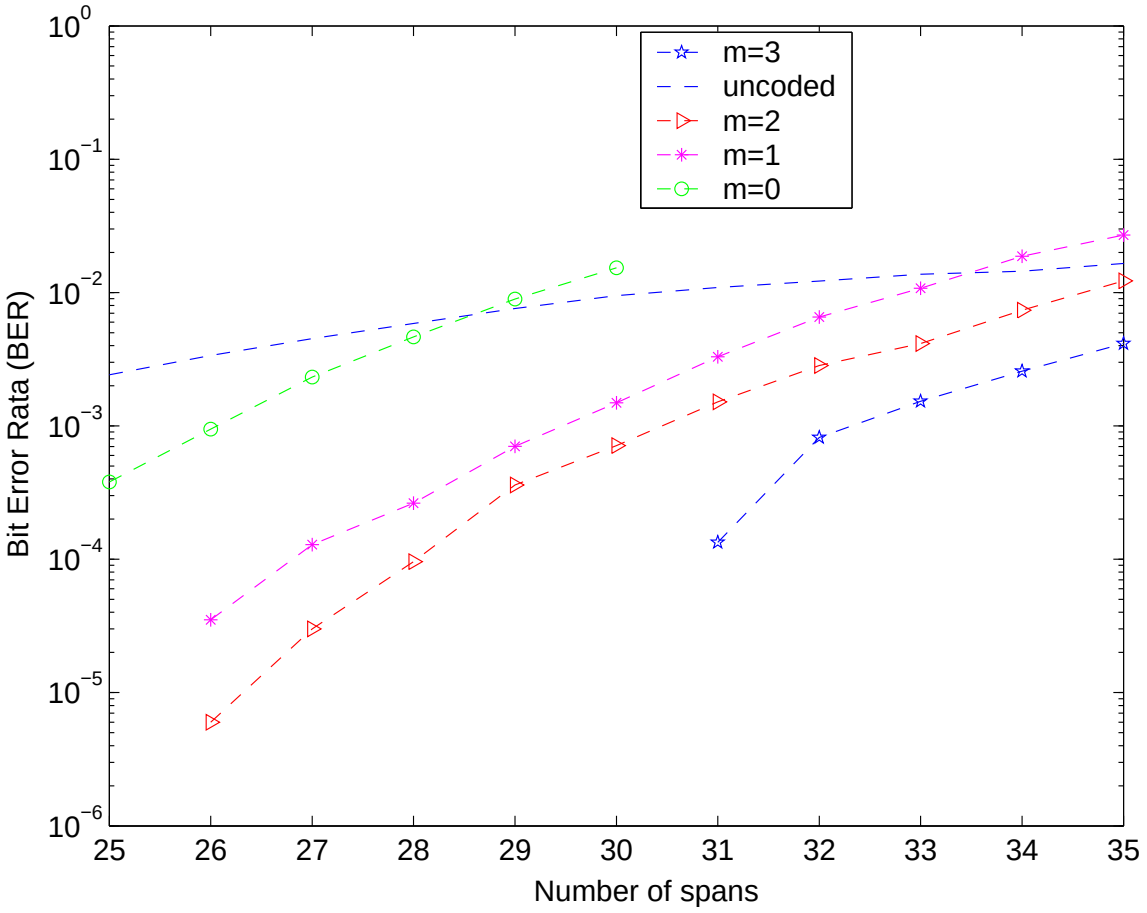

FIGURE 5.4. BER performance of Viterbi equalizer supplemented with hard-decision Gallager B LDPC decoder, for different memory assumptions

weight 5 and girth 8 [49]. It can be seen that Viterbi and BCJR detectors perform very similarly, but that all the coding schemes with the BCJR and the LDPC code perform much better than the Viterbi algorithm.

In Fig. 5.6 we compare different versions of the soft decision scheme. The curve with asterisks was obtained by decoding scheme (a) with the maximal number of iterations of the sum-product algorithm set to 25. The curve with circles was also

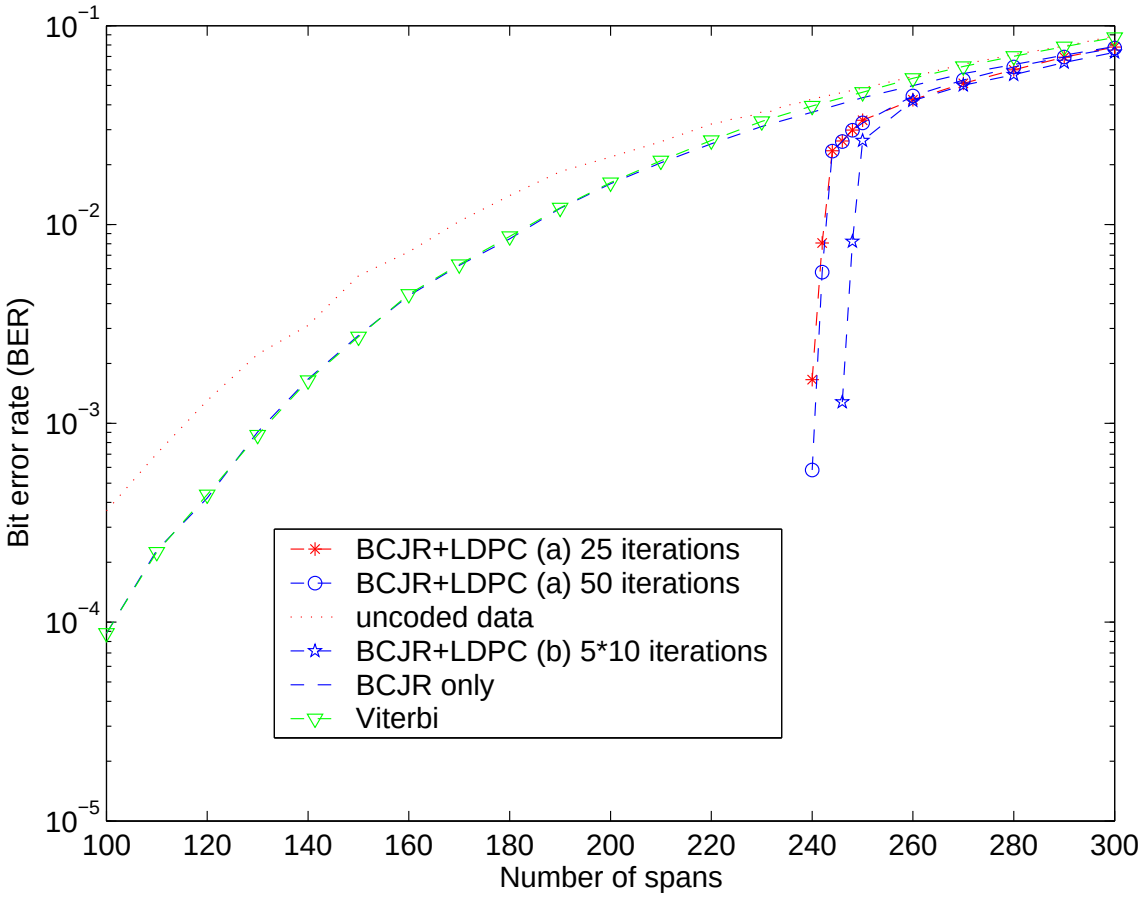

FIGURE 5.5. Comparison of different coding schemes

obtained by decoding scheme (a), but this time with the maximal number of iterations of the sum-product algorithm set to 50. It can be seen that the former performs slightly better. However, both have zero bit error rate if the number of spans is less than 240. The curve with five pointed stars was obtained by decoding scheme (b) with the maximal number of iterations of the sum-product algorithm set to 10, and 5 outer iterations, that is, the BCJR algorithm was run five times and the LDPC decoder ran at most 5×10 iterations. As expected this decoding scheme performs

the best, but at the expense of higher complexity [58].

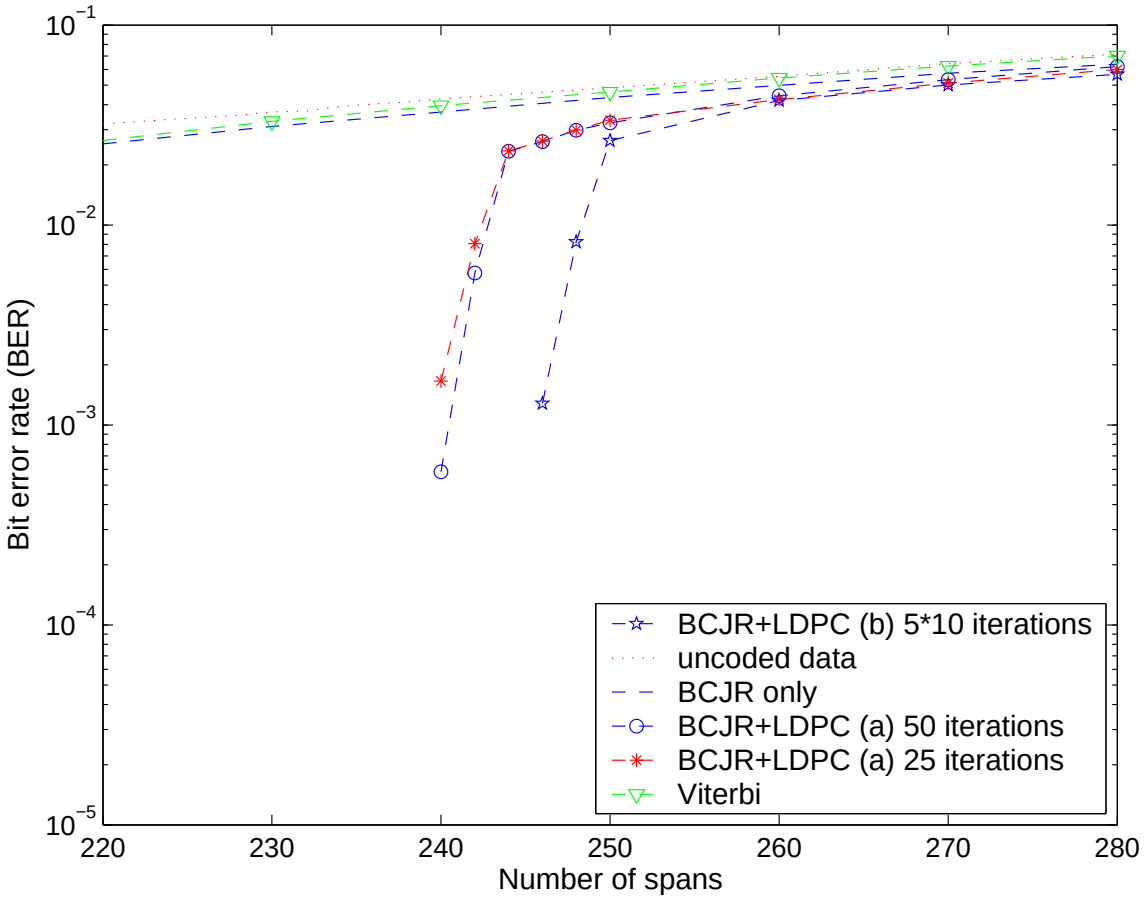

FIGURE 5.6. Comparison of different coding schemes

We also ran the BCJR algorithm for different values of m -the number of slots (memory) on each side of center slot. As already mentioned, this is of interest because designing electronics that would perform BCJR on a high number of states (2^{2m+1}) is not a trivial task, especially for systems operating at 40Gb/s. The results are presented in Fig. 5.7. Probably the most interesting curve on this plot is the one with circles. It was obtained with $m = 0$, so in that case the BCJR did not run at

all, and the performance is due LDPC code solely.

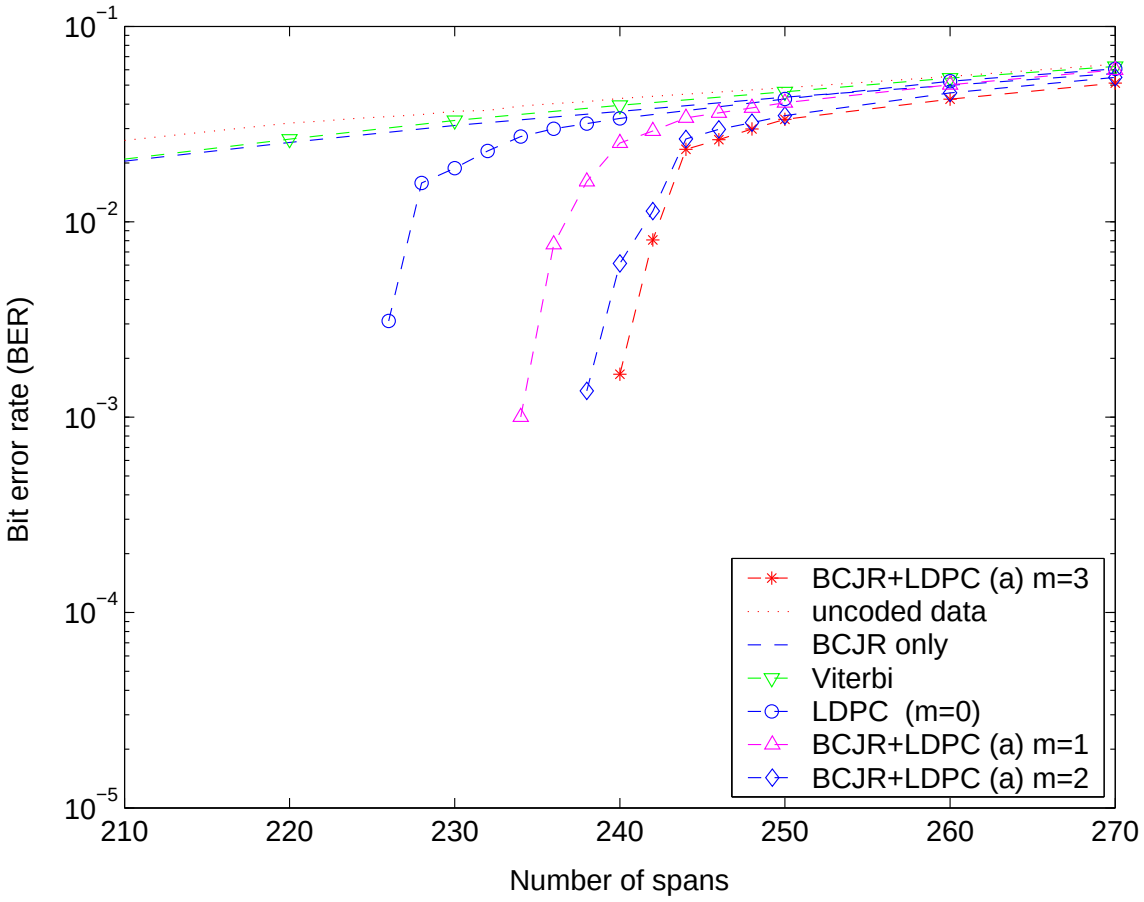

FIGURE 5.7. Comparison of coding schemes with different memory

5.4. Final remarks on decoding schemes

LDPC codes offer great flexibility in terms of length, rate and decoding algorithms which makes the proposed schemes even more attractive and applicable in a wide range of systems.

The first Viterbi equalizers operating at 10.7 Gb/s are already commercially available [54]. This fact, together with relatively low complexity of the Gallager B decoder makes the proposed hard decision scheme a viable option for high speed optical transmission.

The numerical results confirm that the soft decision scheme performs much better than the hard decision one. However, at this moment, to the best of our knowledge, there are no hardware implementations of BCJR equalization and sum-product algorithm operating at 40Gb/s. This makes the soft decision scheme not practical right now. It can be expected that in the (near) future hardware operating at such speeds will be available. Once that happens the soft decoding schemes will become practical.

6. Conclusion

It is important to stress that this dissertation presents, more or less, a complete story. By this it is meant that we first developed a technique to characterize the systems, then found achievable information rates. Numerical results indicate that, with an appropriate receiver and for reasonable distances, this kind of a system can achieve information rates very close to one bit per channel use. This means that sophisticated coding schemes, using codes with rates close to one are desirable. In the Chapter V we propose coding techniques with very good performance.

Second, we would like to mention that the proposed techniques have both quite developed theoretical bases and practical importance.

For example, the theoretical instanton method gives a suitable asymptotic behavior for the tails of PDFs. On the other hand, system outputs enter into PDFs via coefficients of a parameterized family of orthogonal polynomials. The parameterized family of orthogonal polynomials also reduces computer processing cost, making this approach applicable for high speed applications.

Finally, the proposed techniques are very general. For example, Edgeworth expansion can be used independently of modulation and/or starting distribution. Noise

with any statistics can be easily incorporated in the instanton approach as long as it is uncorrelated. Nowadays, there is a wide range of LDPC codes with different length, rate. There are several decoding algorithms that can be used. All of this makes the proposed schemes even more attractive.

Looking forward, we can see only increasing need for information transmission. We hope that techniques presented in this dissertation, or similar one, will play role in future transmission systems.

References

[1] M. Ivkovic, I. B. Djordjevic, P. Rajkovic, B. Vasic, "Pulse Energy Probability Density Functions for Long-Haul Optical Fiber Transmission Systems by Using Instantons and Edgeworth Expansion," *IEEE Photonic Technology Letters,* vol. 19, no. 20, pp. 1604 - 1606, Oct.15, 2007.

[2] I. Djordjevic, M. Ivkovic, B. Vasic, I. Gabitov, "Achievable information rates for high-speed long-haul optical transmission", *Journal of Lightwave Technology*, vol. 23 no.11, pp. 3755-3763, Nov. 2005.

[3] M. Ivkovic, I. B. Djordjevic, B. Vasic,"Calculation of Achievable Information Rates of Long-Haul Optical Transmission Systems using Instanton Approach", *Journal of Lightwave Technology*, vol. 25, pp. 1163-1168, May 2007.

[4] M. Ivkovic, I. B. Djordjevic, B. Vasic, "A Soft Decision Coding Scheme for Long-Haul Optical Transmission Systems Based on Instanton Approach," in *Proc. Globecom 2006*, Paper no. CTHp1-1, 27 November - 1 December 2006, San Francisco, CA, USA.

[5] M. Ivkovic, I. B. Djordjevic, B. Vasic, "Hard decision error correcting scheme based on LDPC codes for long-haul optical transmission," in *Proc. Optical Trans-*

mission Systems and Equipment for Networking V-SPIE Optics East Conference, 1-4 October 2006, Boston, Massachusetts, USA.

[6] M. Ivkovic, I. B. Djordjevic, P. Rajkovic, B. Vasic, "Modelling errors in long-haul optical fiber transmission systems by using instantons and Edgeworth expansion," in *Proc. ICC 2007,* 24th-27th June 2007, Glasgow, Scotland, UK.

[7] M. Ivkovic, I. B. Djordjevic, B. Vasic, "Calculation of Achievable Information Rates of Long-Haul Optical Transmission Systems using Instanton Approach," in *Proc. ISIT 2006,* paper no. IT06-1983, July 2006.

[8] O. E. Agazzi, M. R. Hueda, H. S. Carrer, D. E. Crivelli, "Maximum-likelihood sequence estimation in dispersive optical channels" *Journal of Lightwave Technology*, vol. 23, issue 2, pp. 749 - 763, Feb. 2005.

[9] M. J. Ablowitz and T. Hirooka, "Resonant nonlinear intrachannel interactions in strongly dispersion-managed transmission systems", *Optics Express*, vol. 25, no. 24 pp. 1750-1752, Dec 2000.

[10] G. P. Agrawal, *Nonlinear Fiber Optics*. San Diego, CA: Academic, 2001.

[11] N. Alić *et al.* "Signal statistics and maximum likelihood sequence estimation in intensity modulated fiber optic links containing a single optical preamplifier", *Optics Express*, vol. 13, pp. 4568-4579, Jun. 2005.

[12] F. Buchali, G. Thielecke, H. Blow, "Viterbi equalizer for mitigation of distortions from chromatic dispersion and PMD at 10 Gb/s," in *Proc. Opt. Fiber Comm. Conf.*, vol. 1, Los Angeles, CA, Feb. 2004.

[13] V. Chernyak, *et al.* "PMD-induced fluctuations of bit-error rate in optical fiber systems", *Journal of Lightwave Technology* vol. 22, no.4, pp.1155-68, 2004.

[14] E.R. Dougherty, *Random Processes for Image and Signal Processing.* SPIE Opt. Eng. Press, IEEE Press, New York 1998.

[15] *EDFA Noise Gain Profile and Noise Gain Peak Measurements,* Agilent Technologies, http://cp.literature.agilent.com/litweb/pdf/5963-7148E.pdf

[16] T. Foggi *et al.* "Maximum likelihood sequence detection with closed-form metrics in OOK optical systems impaired by GVD and PMD", *Journal of Lightwave Technology* vol. 24. no. 8., pp 3073-3087, Aug. 2006.

[17] E. Forestieri, "Evaluating the error probability in lightwave systems with chromatic dispersion, arbitrary pulse shape and pre- and postdetection filtering," *Journal of Lightwave Technology*, vol. 18, no. 11, pp. 1493-1503, Nov. 2000.

[18] A. Mecozzi, C. B. Clausen and M. Shtaif, S.G. Park and A. H. Gnauck "Cancellation of timing and amplitude jitter in symmetric links using highly dispersed pulses," *IEEE Photonic Technology Letters*, vol. 13, pp. 445-447, 2001.

[19] E. Podivilov, I. Gabitov, unpublished notes.

[20] L. Mollenauer, J. Gordon, *Solitons in Optical Fibers: Fundamentals and Applications*. Academic Press, New York 2006.

[21] S. -G. Park, A.H. Gnauck, J.M. Wiesenfeld and L.D. Garrett, 40-Gb/s transmission over multiple 120-km spans of conventional single-mode fiber using highly dispersed pulses", *IEEE Photonics Technology Letters*, vol. 12, no. 8, pp. 1085-1087 (2000).

[22] T. Freckmann, J. Spiedel, "Viterbi equalizer with analytically calculated branch meterics for optical ASK and DBPSK," *IEEE Photonic Technology Letters*, vol. 18, no. 1, pp. 277-279, Jan. 2006.

[23] C.L. Ho, "Calculating the performance of optical communication systems with modal noise by saddlepoint method," *Journal of Lightwave Technology*, no. 9, pp. 1820-1825, Sep. 1995.

[24] R. Holzlöhner, V. S. Grigoryan, C. R. Menyuk, and W. L. Kath, "Accurate Calculation of Eye Diagrams and Bit Error Rates in Optical Transmission Systems Using Linearization," *Journal of Lightwave Technology*, no.20, pp. 389-400, Mar. 2002.

[25] M. R. Hueda, D. E. Crivelli, H. S. Carrer, "Performance of MLSE-based receivers in lightwave systems with nonlinear dispersion and amplified spontaneous emis-

sion noise," in *Proc. IEEE GLOBECOM* 2004, vol. 1, pp. 299-303, Nov./Dec. 29-3, 2004.

[26] P. A. Humblet, M. Azizoglu, "On the bit error rate in lightwave systems with optical amplifiers," *Journal of Lightwave Technology* vol. 9, 1576-82, Nov. 1991.

[27] J. Kolassa, *Series Approximation Methods in Statistics,* 2nd ed. Springer-Verlag, New York 1997.

[28] M. Nazarathy, "Accurate evaluation of bit-error rates of optical communication systems using the Gram-Chariler series", *IEEE Transanctions on Communications*, vol. 54, no.2, pp. 295- 301, Feb. 2006.

[29] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fiber communications", *Nature*, vol. 411, pp. 1027-1030, June 2001.

[30] K. S. Turitsyn, S. A. Derevyanko, I. V. Yurkevich, and S. K. Turitsyn, "Information capacity of optical fiber channels with zero average dispersion," *Physics Revue Letters*, vol. 91, no. 20, p. 203 901, Nov. 2003.

[31] J. Tang, "The Shannon channel capacity of dispersion-free nonlinear optical fiber transmission", *Journal of Lightwave Technology*, vol. 19, pp. 1104-1109, Aug. 2001.

[32] J. Tang, "The multispan effects of Kerr nonlinearity and amplifier noises on Shannon channel capacity for a dispersion-free nonlinear optical fiber", *Journal of Lightwave Technology*, vol. 19, pp. 1110-1115, Aug. 2001

[33] J. Tang, "The channel capacity of a multispan DWDM system employing dispersive nonlinear optical fibers and an ideal coherent optical receiver", *Journal of Lightwave Technology*, vol. 20, pp. 1095-1101, July 2002.

[34] L. R. Bahl, J. Cocke, F. Jelinek, J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate" *IEEE Transanction on Information Theory*, pp. 284-287, March 1974.

[35] D. Arnold, H. A. Loeliger, "On the information rate of binary-input channels with memory" in *Proc. 2001 Int. Conf. Communications* , Helsinki, Finland, June 11-14, 2001, pp. 2692-2695.

[36] H. D. Pfitser, J. B. Soriaga, P. H. Siegel, "On the achievable information rates of finite state ISI channels" in *Proc. Globecom* 2001, San Antonio, TX, Nov. 25-29, 2001, pp. 2992-2996.

[37] S. Yang, A. Kavcic, "Capacity of Partial Response Channels", in Handbook on Coding and Signal Processing for Recording Systems, CRC Press, 2004.

[38] D. Arnold, A. Kavcic, H. A. Loeliger, P. O. Vontobel, W. Zeng, "Simulation-based computation of information rates: upper and lower bounds" in *Proc. IEEE Intern. Symp. Inform. Theory* (ISIT 2003), 2003, p. 119.

[39] V. Sharma, S. K. Singh, "Entropy and channel capacity in the regenerative setup with applications to Markov channels" in Proc. IEEE Int. Symp. Information Theory, Washington, DC, June 24-29, 2001, p. 283.

[40] N. Varnica, X. Ma, A. Kavcic, "Capacity-Approaching Codes for Partial Response Channels" in Handbook on Coding and Signal Processing for Recording Systems, CRC Press, 2004.

[41] T. M. Cover, J. A. Thomas, *Elemenents of Information Theory*. New York: John John Wiley and Sons, 1991.

[42] I. M. Lifshitz, "Energy spectrum structure and quantom states of disordered condensed systems", Usp. Fiz. Nauk, vol. 83 no.4, pp.617-663, 1964.

[43] V. Chernyak, M. Chertkov, M. Stepanov, B. Vasic, "Error correction on a tree: an instanton approach", Phy. Rev. Letters, vol. 22, pp. 228701/1-4. Nov. 2005.

[44] O. E. Agazzi, D. E. Crivelli and H. S. Carrer, "Maximum-likelihood sequence estimation in dispersive optical channels," *Journal of Lightwave Technology*, vol. 23, pp. 749-763, Feb. 2005.

[45] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *Electron. Lett.*, vol. 32, pp. 16451646, Aug. 1996.

[46] R.G. Gallager, Low-Density Parity-Check Codes, MIT Press,1963.

[47] R. Urbanke, *Iterative Coding Systems*, unpublished notes. http://lthcwww.epfl.ch/publications/index.php

[48] R. M. Tanner, "A Recursive Approach to Low Complexity Codes," *IEEE Trans. Inform. Theory,* Vol. IT-27, no. 5, pp. 533-547, Sept. 1981.

[49] B. Vasic, I. B. Djordjevic and R. Kostuk, "Low-density parity-check codes and iterative decoding for long haul optical communication systems," *J. Lightwave Technol.,* vol. 21, no.2, pp. 438-446, Feb. 2003.

[50] O. Milenkovic *et al.* "Block-Circulant Low-Density Parity-Check codes for optical communication systems," *IEEE Journal of selected topics in quantum electronics*, vol. 10, no. 2, March 2004.

[51] http://www.inference.phy.cam.ac.uk/mackay/codes/

[52] S. Lin and D.J. Costello, *Error Control Coding*, 2. ed. Upper Saddle River, NJ: Prentice Hall, 2004.

[53] S. Sankaranarayanan *et al.* "Failures of the Gallager B decoder: analysis and applications", Information Theory and Aplications workshop, UCSD, San Diego, 2005.

[54] A. Faerbert *et al.*, "Performance of a 10.7 Gb/s Receiver with Digital Equaliser using Maximum Likelihood Sequence Estimation," ECOC'O4, Paper no.Th.4.1.5.

[55] R. Urbanke, *Iterative Coding Systems*, unpublished notes. http://lthcwww.epfl.ch/publications/index.php

[56] T. Freckmann and J. Spiedel, "Viterbi equalizer with analytically calculated branch meterics for optical ASK and DBPSK," *Photonic Technology Letters*, vol. 18, no. 1, pp. 277-279, Jan. 2006.

[57] J. L. Fan, A. Friedmann, E. Kurtas, and S. W. McLaughlin, "Low density parity check codes for magnetic recording," in Proc. 37th Annual Allerton Conf. Communication, Control, and Computing, 1999, pp. 13141323.

[58] B. M. Kurkoski, P. H. Siegel, J.K. Wolf, "Joint message passing decoding of LDPC codes and partial-response channles," *IEEE Transanctions on Information Theory*, vol.48, pp. 1410-1422, June 2002.