Nonlinear Signal-Noise Interaction in Optical Fiber Transmission

---

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment
of the requirements for the degree

Doctor of Philosophy

by

Deepyaman Maiti

August 2017

APPROVAL SHEET

This Dissertation
is submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Author Signature: _____

This Dissertation has been read and approved by the examining committee:

Advisor: Maite Brandt-Pearce

Committee Member: Gianluca Guadagni

Committee Member: Zongli Lin

Committee Member: Daniel Weller

Committee Member: Stephen Wilson

Committee Member: _____

Accepted for the School of Engineering and Applied Science:

Craig H. Benson, School of Engineering and Applied Science

August 2017

Acknowledgements

Abstract

In a long-haul optical fiber communication system, fiber attenuation, dispersion, and nonlinearity combine with non-deterministic noise from optical amplifiers used for periodic regeneration and cause adverse effects on system performance. In this dissertation, we study and mitigate such undesirable effects.

We present a modified nonlinear decision feedback equalizer designed for use in a legacy optical communication system with periodic dispersion compensation. The effects of noise and nonlinearity on the equalizer coefficients are investigated, and a suboptimal convergence algorithm to reduce such effects is proposed and verified.

Noting the limited ability of existing signal processing tools to combat signal-noise nonlinear interaction effects, we next consider a fundamental scenario to study these effects. We apply Gaussian mixture modeling (GMM) techniques to better understand how noise interacts with the signal in a nonlinear optical fiber span. We validate our technique and learn that at higher levels of nonlinearity, the GMM analysis is more accurate than an additive Gaussian noise or a Volterra series transfer function model.

Finally we present an approach to generalize our analysis. We validate our claims that transmitting a small number of pulses is a good approach to predict the analysis of a practical communication system. We also show that using a 3- or higher-order GMM is necessary to fully understand the nonlinear interaction.

Contents

List of Figures

List of Tables

Table of Acronyms

Acronym	Description
ASE	amplified spontaneous emission (noise)
AWGN	additive white Gaussian noise
BER	bit error rate
DBP	digital backpropagation
DCF	dispersion compensating fiber
DFE	decision feedback equalizer
EDFA	erbium doped fiber amplifier
E-M	expectation maximization (algorithm)
FBF	feedback filter
FFF	feed forward filter
GMM	Gaussian mixture model/ modeling
IS	importance sampling
ISI	inter symbol interference
KDE	kernel density estimation
K-L	Kullback-Leibler divergence
LMS	least mean squares (algorithm)
MAP	maximum a posteriori
MLSE	maximum likelihood sequence estimation
NF	noise figure
NLPN	nonlinear phase noise
NLSE	nonlinear Schrödinger equation
OFDM	orthogonal frequency division multiplexing
OOK	on-off keying modulation
QPSK	quadrature phase shift modulation
RLS	recursive least squares (algorithm)
SSF	split-step Fourier (method)
SSMF	standard single mode fiber
VSTF	Volterra series transfer function
WDM	wavelength division multiplexing

Table of Symbols

Symbol	Description
α	attenuation constant
β_2	group-velocity dispersion parameter
γ	nonlinear parameter
C	channel capacity
ϵ	tolerance
I	mutual information
i_{max}	iterations per symbol for first DFE training stage
i_{avg}	mean iterations per symbol for second DFE training stage
L	span length
μ	mean
N	data block size
n	number of spans
n_1	length of linear parts of DFE filters
n_2	length of nonlinear parts of DFE filters
n_c	number of DFE filter coefficients
$w(t)$	ASE noise
Φ	phase of a symbol (signal)
ψ	amplitude of optical field
P	input power
P_e	bit error rate
$p(t)$	pulse shape of transmitter filter
q	symbol number
ρ	correlation coefficient
Σ	covariance matrix
T	symbol period
t_q	sampling instant
V_q	current state of DFE system
X	transmitted signal
Y	received signal
y_q	equalizer input for q^{th} symbol
r	weights/ proportions

Chapter 1

Introduction

1.1 Research Motivation

We are nearing the performance limits of communication networks, not just in this country, but all over the world [3]. The problem can only get worse, since we need to transmit more information than ever over the same limited amount of resources. The demand for more bandwidth and capacity will only increase with technological advances such as connected cars, smart grids, machine-to-machine communication, and domestic installations such as at-home health monitoring systems. In Fig. 1.1, we note how the number of devices online has increased steadily over the years.

It was realized long ago that the demand for reliably transmitting more and more data will gradually increase over time. However, it was also believed that the design of transmitters, receivers, and other devices would also improve and thus keep pace with the increasing demand, so that the need for excess capacity would be able to be addressed. However, this hasn't been proved to be the case as seen in Fig. 1.2. Effective coding strategies, resourceful use of multiplexing and other
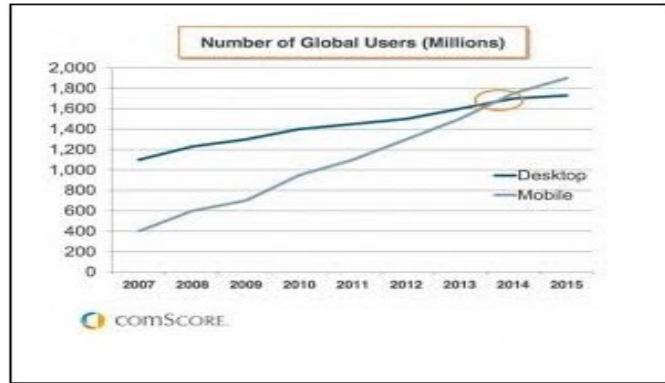
Figure 1.1: Online devices increasing over time [1]
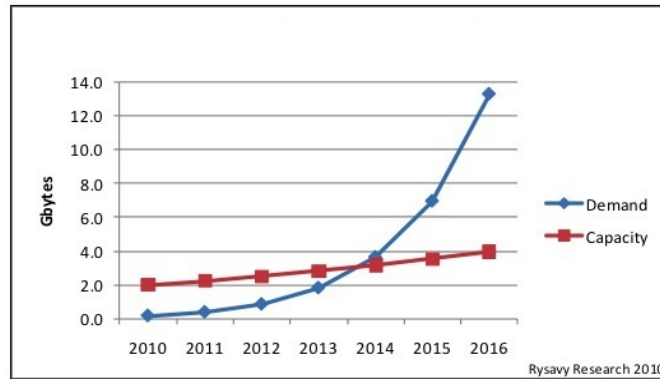
Figure 1.2: Demand vs capacity [2]

techniques can help optimize the bandwidth usage. We turn our attention to the use of optical fiber as an effective communication channel which offer potential for high speed communication over long distances.

However, there are significant challenges which must be overcome. Optical fibers have several kinds of dispersion, including material and waveguide dispersions, the combined effect of which is broadening of pulses during propagation, and when severe would lead to a pulse spreading outside its symbol slot [4]. An optical fiber is a dielectric, and exhibits nonlinearity. In long-haul optical fiber communication systems, optical amplifiers are present at the end of each span (a length of $80 - 100$ kms) to periodically restore the signal strength. The amplifiers

add random amplified spontaneous emission (ASE) noise to the signal.

In the dispersive and nonlinear optical channel, the ASE noise interacts nonlinearly with the deterministic signal, while the signal interacts with itself, as does the noise, to generate crosstalk at the receiver. Intersymbol and interchannel interference, self, cross and interchannel cross phase modulation, four-wave mixing and amplifier noise all affect the performance of the communication system, especially the periodically dispersion compensated legacy systems [5]. Such legacy systems consist of a dispersion-compensating (DCF) fiber after each span to reverse the effects of dispersion, and also an optical amplifier to reverse the effects of fiber attenuation. A large percentage of existing fiber communication networks are such systems. We initially turn our attention to designing tools and techniques which can be applied to these. (Later in this document, we look at systems which do not employ dispersion compensation.)

For long-haul high-speed wavelength division multiplexed (WDM) optical communication systems, coherent detection is often preferred since it supports spectrally efficient modulation formats, and offers potential for superior signal sensitivity and the possibility of electronic compensation algorithms [6]. These spectrally efficient modulation formats require higher signal-to-noise power ratio. However, nonlinear impairment effects are also greater at higher power levels.

We note that our resources are limited, and therefore we should press for the most efficient use of what we have at our disposal. Shannon's theorem [7] states that it is theoretically possible to transmit information at a rate which equals the channel capacity with arbitrarily low error probabilities if suitable coding or signal processing techniques exist. Recently, however, progress in WDM capacity research has markedly slowed down as experiments are closely approaching what has been thought of as the fundamental Shannon limits of nonlinear fiber transmission [8].

In a nonlinear noisy channel, the received signal is a combination of the transmitted signal, noise and interaction terms. In this dissertation, we further our understanding of the nonlinear interactions in our optical fiber communication channel, propose and validate that using improved signal processing techniques can overcome certain limitations, and ultimately increase capacity.

1.2 Technical Background

Faced with the aforementioned challenges, we can design better equalizers to recover the transmitted signal. An equalizer approximates the transmitted information by methods such as removing impairment effects using filters or searching for likely transmitted sequences using probability measures [9]. As outlined above, the optical fiber communication channel has linear impairments such as attenuation and dispersion, and several kinds of nonlinear impairments. The combination of these linear and nonlinear impairments, especially in the presence of non-Gaussian noise, makes our task difficult.

Optical and electrical techniques have been developed to reduce the effects of physical impairments and amplifier noise and extract the transmitted signal with a desired degree of accuracy. Electrical solutions can be integrated into electronic chips and mass-produced at a reasonable cost, and can be applied to existing optical networks. We concentrate on signal processing techniques in this dissertation [10] to learn the nature of, and combat the nonlinear interaction effects.

While both linear and nonlinear equalizers have been proposed and tested, nonlinear equalizers are far more suited to combat the effects of nonlinear impairments in an optical fiber channel. Speaking generally, equalizer designs which

promise high performance also have high computational complexity. This, of course, is not desirable. We set out to correct this issue, and propose and test an intuitive nonlinear equalizer design which allows us good performance at low complexity.

Our equalizer performs as well as the popular digital backpropagation (a high complexity method to jointly compensates linear and nonlinear impairments) benchmark [11]; however, we are still not at the maximum performance limit. We know this since in the presence of nonlinear noise, backpropagation (or our equalizer) cannot exactly reconstruct the transmitted signal, and there is a gap between the no-nonlinearity limit and the actual performance. If we can find and address the reason(s) for this, we can approach the said limit.

This is a situation which can be addressed by obtaining more precise knowledge about the nonlinear interaction between signal and noise. If there were no signal-signal and signal-noise nonlinear interaction, we would have matched the no-nonlinearity performance limit. While we have studied Gaussian noise for decades and the pertinent theory is well-developed, the same cannot be said about non-Gaussian noise. We contend (and later show) that in our optical fiber communication channel, the popular additive white Gaussian noise model assumption is not satisfactory [12].

The nonlinear Schrödinger equation (NLSE) [5] describes the propagation of an optical pulse through a span of fiber. The usual practice to solve the NLSE is using numerical methods such as the split-step Fourier transform. However, the Volterra series transfer function (VSTF) [13] may be used to compute an approximate analytic solution.

There has been considerable work published on the effects of noise in optical fibers. However, some of these works make a low or zero dispersion assumption.

It is the interaction of noise with all the linear and nonlinear impairments which makes this a challenging research problem. Some other works look at important system properties (such as bit error rate or capacity) with or without making assumptions or using existing simplified models [14].

We would like to understand the nature of the nonlinear interaction without any assumptions about the system properties. If we succeed in this, we can use this knowledge to design better receivers or signal processing algorithms which would help us achieving performance closer to the no-nonlinearity limit.

Using a fundamental yet non-trivial situation, we show that we can use clustering techniques to decompose the nonlinear interaction components into several Gaussian components. We validate our technique with suitable tests and show that our algorithm works better than existing techniques when nonlinearity is significant. We then attempt to generalize our study, and approach a practical communication system.

1.3   Dissertation Outline

The rest of the dissertation is organized as follows. In Chapter 2, we present a modified nonlinear decision feedback equalizer designed for use in a legacy optical communication system with periodic dispersion compensation. Chapter 3 explains the theory and application of GMM to the fundamental problem of analyzing signal-noise interaction in a single pulse single channel system and presents compelling evidence that mixture modeling is an useful tool for analyzing the optical fiber channel. Chapter 4 generalizes the fundamental system and studies transmission of multiple pulses. Chapter 5 summarizes the whole dissertation and points out opportunities for some future work.

Chapter 2

Modified Nonlinear Decision Feedback Equalizer

As the transmitted signal pulses travel through several spans of an optical fiber communication channel, they are affected by, and affect each other, due to the fiber physical impairments (both linear and nonlinear). The nonlinear Schrödinger equation (NLSE)

$$\frac{\partial \psi}{\partial z} = -\frac{\alpha}{2}\psi - \frac{i\beta_2}{2}\frac{\partial^2 \psi}{\partial t'^2} + i\gamma|\psi|^2\psi \tag{2.1}$$

describes the propagation of the pulses, where ψ is the varying amplitude of the field, z denotes space, and α, β_2, and γ are respectively the attenuation, second-order dispersion and the nonlinearity in the fiber.

The split-step Fourier method [15] is a numerical method often used to solve the NLSE. This method computes the solution in small steps treating the linear and the nonlinear steps separately. The linear step is made in the frequency domain while the nonlinear step is made in the time domain. Thus it is necessary to

Fourier transform back and forth between the time and frequency domains.

Digital backpropagation [11] jointly compensates linear and nonlinear impairments. The basic principle is to solve an inverse NLSE through the fiber to estimate the input signal. In the absence of stochastic noise, backpropagation can reconstruct the originally transmitted sequence with perfect accuracy. The main disadvantage of backpropagation is its excessive computational complexity. It is also practically difficult to implement with hardware or at high data-rates due to its oversampling rate requirement.

This first task is to see how far we can go in performance improvement by just handling the effects of other pulses. We consider the theoretical best (backpropagation) and practical (equalizer) options. This will indicate a limit after which we need to look at the noise-signal interaction more closely.

Equalizers can be divided into two categories: linear and nonlinear. Linear equalizers may be designed using a linear(ized) or affine model for the communication channel. The simplest form of equalization uses a linear filter which applies the inverse of the channel transfer function on the received symbols to estimate the transmitted data. Equalizers using finite impulse response (FIR) type filters have been studied in works such as [16]. Dispersion compensation techniques [17] also fall under this category. Linear equalizers can only compensate the linear impairments.

On the other hand, a nonlinear equalizer can successfully mitigate nonlinear impairment effects as well and are thus much more suitable for long-haul optical communication system applications. Nonlinear equalizers utilizing methods such as maximum likelihood sequence estimation (MLSE) or maximum a posteriori sequence detection (MAP) have been well studied. MAP equalizers rely on assumed statistical information about the transmitted signal. MLSE equalizers require a

Viterbi decoder and soft decision for decoding [18]. The computational complexity is considerable due to the large number of states and this might be prohibitive at the data rates of interest (hundreds of Gbit/s). Channel memory when coupled with nonlinearity makes this a challenging problem as well.

The complexity of MLSE and MAP-based methods grows exponentially with the equalizer memory size and the number of WDM channels since it needs a look-up table of possible sequences of transmitted symbols. MAP-based equalizers for multi-channel transmission or WDM coherent systems having good performance but high signal processing complexity have been proposed in [19] and [20] respectively.

Frequency domain equalization methods [21] have also been studied, using a look-up table or a constant-modulus based algorithm to adjust tap-weights (in polarization multiplexed transmission systems), or by estimating the channel transfer function (in optical orthogonal frequency division multiplexed (OFDM) transmission) [22].

Some authors have designed nonlinear equalizers based on analytical closed-form approximations from appropriate order Volterra kernels to mitigate nonlinear effects. Using third-order inverse Volterra theory on a mathematical description of an optical system which includes all the impairments effects, a model-centric nonlinear equalizer for single channel systems is found in [23]. The disadvantage of a Volterra-model-based design is its complexity and truncated models are often used [24]. A Wiener-Hammerstein model based equalizer (for OFDM transmission) is presented as a simpler alternative [25].

We propose to apply a decision feedback equalizer (DFE) with nonlinear filters, which we refer to as a nonlinear DFE. In general, for communication systems with additive noise and pre- and post-cursor interference effects (such as ours), DFEs offer a good compromise between complexity and performance [26], and

can be applied to any modulation format. Linear decision feedback approaches (DFEs with linear filters) have been used before in related fields such as wireless optical communication [27] or channel estimation for OFDM [28]. A DFE based structure with a nonlinear Volterra filter feedback for recording systems is found in [29], and with nonlinear Volterra filters in feedforward and feedback paths for reading information stored in optical discs in [30]. The structure and design of a digital DFE is conceptually simple, and it can be implemented using hardware [31]. Compared to a single feedforward equalizer, a nonlinear DFE (with a nonlinear feedforward and a nonlinear feedback filter) can mitigate pre- and post-cursor impairment effects simultaneously without enhancing noise while requiring a lower number of filter coefficients.

Only a few applications of decision feedback in optical fiber communication systems are found in literature. In [32], the authors study the impact of the channel bandwidth on performance and the maximization of spectral efficiency in a multi-channel system using a constant modulus algorithm feed forward equalizer (CMA FF) with a feedback loop. This design compensates for intersymbol interference (ISI) (which would cause one symbol to interfere with subsequent symbols). The equalizer filter coefficients are calculated from the impulse responses of the feed forward and feedback paths. In [33], a DFE architecture integrating error detection code capability and consisting of one or more linear or nonlinear Volterra filters is investigated for a single channel system. This work is conceptually similar to the Volterra filter based equalizers discussed above, with the addition of a decision feedback loop, as done in [30]. The authors conclude that integration of the error detection capability with a Volterra filter in the forward path provide good performance over a range of input power. The transmission distance is 480 km, and the equalizer performance is not compared against that of standard techniques

such as backpropagation.

The novelty of the current work is in identifying a receiver end electrical nonlinear DFE suitable for a long-haul coherent optical fiber system, understanding that system noise and uncompensated nonlinearity affect tuning of the coefficients, and presenting a modified DFE training which uses multiple iteration least mean squares (LMS) algorithm [34] to mitigate it. We report results from long-haul legacy systems having total lengths varying between 2800 and 4000 km and find the equalizer performance (bit error rate) is comparable to that from digital backpropagation while being computationally simpler.

This chapter is organized as follows. In Section 2.1, we decide on the DFE structure from an input-output model of the optical system and describe the equalizer in training and decision modes. In Section 2.2, we study the effect of noise and uncompensated impairments on the filter coefficients. Finding that the coefficients do not achieve steady-state values with the conventional tuning method (in general), we implement and verify a *suboptimal convergence* training approach. In Section 2.3, we evaluate the performance of the DFE, and compare the results with other approaches including backpropagation. In Sections 2.4 and 2.5, we investigate the computational complexity and the performance analysis of the DFE. Section 2.6 summarizes this chapter.

2.1   DFE Structure and Coefficient Tuning

The basic decision feedback equalizer (DFE) structure contains two filters [26], a feed forward filter (FFF), a feedback filter (FBF) and also includes a nonlinear decision device. The FBF feeds back a function of the decisions on the past symbols. The filters may both be linear, or at least one of them can be nonlinear.

If the past decisions were correct, we are in fact feeding back a certain function of the actual past transmitted symbols. Thus the DFE can suppress or cancel the post-cursor interference effects for a number of past symbols.

Our long-haul single-channel optical communication system (Fig. 2.1) is a legacy network. At the transmitter, an M-ary electrical symbol sequence x_q is converted into the corresponding optical signal $s(t)$ after pulse shaping and modulation. We choose a square-root raised cosine pulse shaping filter to minimize intersymbol and other interference effects. The optical signal $s(t)$ is then transmitted through a finite number of spans of standard single mode fibers (SSMF) and is affected by the linear and nonlinear effects (and amplifier noise) to evolve into $r(t)$, the input to the receiver. Our channel model has n spans of SSMF fibers with each span of length L, with dispersion compensating fiber (DCF) sections and optical amplifiers (EDFA- erbium doped fiber amplifiers) at the end of each span (after the DCF). We assume full (100%) and ideal dispersion compensation, and any loss or nonlinearity in the DCF is neglected.

The coherent receiver downconverts the optical signal into electrical signal pulses and it consists of an optical filter to limit the ASE noise, a photo-detection device, and an electrical filter to reduce (electrical) noise [5]. Our receiver-end electrical filter also has a square-root raised cosine pulse shape matched to the transmitter filter. The narrow bandpass nature of the receiver-end electrical filter removes all second-order nonlinearity [35] from the photodetector output. Sampling once per symbol period T at optimal discrete times $t_q = qT$ (ignoring synchronization issues) then generates $y_q = y(t_q)$ which is the input to the equalizer module. Note that, since the pulses are broadened in frequency due to the impairment effects, sampling once per symbol time is suboptimal and does not capture all the nonlinear effects.
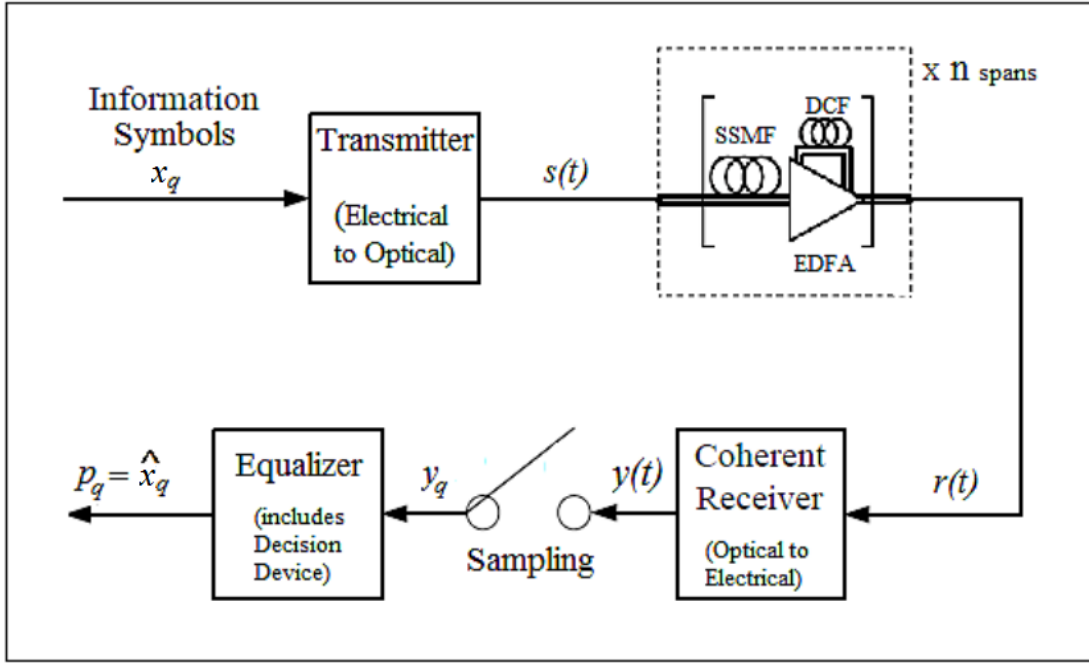
Figure 2.1: Block diagram of the optical fiber communication system

When ASE noise is not considered, using the result in [36] obtained from a discrete-time Volterra model, we determine that y_q is a (mainly third-order) nonlinear function of the modulating data. The general form of this relation would be

$$y_q = \sum_{i=-\infty}^{\infty} \rho_i x_{q+i} + \sum_{i,j,k=-\infty}^{\infty} \rho_{ijk} x_{q+i} x_{q+j}^* x_{q+k} + H.O.T. \qquad (2.2)$$

The coefficients ρ_i's and ρ_{ijk}'s are related to the fiber impairments, and the abbreviation H.O.T indicates fifth and higher-order terms $(O(x^5))$.

For the equalization problem, we need to estimate the transmitted symbols x_q's from the y_q's. Assume that from (2.2), by application of inverse Volterra

theory [23], we can obtain

$$x_q = \sum_{i=-q}^{N-q-1} C_i y_{q+i} + \sum_{i,j,k=-q}^{N-q-1} C_{ijk} y_{q+i} y_{q+j}^* y_{q+k} + H.O.T., \tag{2.3}$$

where N is the total size of transmitted data, still ignoring the effect of ASE noise in the system. C_i and C_{ijk} are respectively the linear and nonlinear coefficients, related to ρ_i and ρ_{ijk} in (2.2).

Analytical methods are sometimes used in comparable scenarios to determine the DFE filter coefficients, but lacking an accurate description of the system noise to include in (2.2), we tune these numerically with training data. From (2.3), we determine that a DFE with third-order filters in both the forward and feedback paths would compensate the impairments, the significant nonlinearity being of third-order. Since the noise affects the calculation of the filter coefficients (tuning), we retain only the largest of the third-order nonlinear terms (from the few adjacent samples), and not the smaller third-order and all higher-order ones, the latter having a greater chance of being inaccurate and adversely affecting the equalization.

Thus, our nonlinear DFE has the structure

$$p_q = decision\big[f_1(\boldsymbol{y_q}) + f_2(\boldsymbol{p_q})\big], \tag{2.4}$$

where the FFF output is

$$f_1(\boldsymbol{y_q}) = \sum_{i=-n_1}^{n_1} c_i y_{q+i} + \sum_{i,j,k=-n_2}^{n_2} c_{ijk} y_{q+i} y_{q+j}^* y_{q+k}, \tag{2.5}$$

and the output of the FBF is

$$f_2(\boldsymbol{p_q}) = \sum_{i=-n_1}^{-1} d_i p_{q+i} + \sum_{i,j,k=-n_2}^{-1} d_{ijk} p_{q+i} p_{q+j}^* p_{q+k}, \tag{2.6}$$

The vector $\boldsymbol{y_q}$ holds values of past, current and future received symbols; that is, $\boldsymbol{y_q} = [y_{q-n_1}, y_{q-n_1+1}, \ldots, y_{q+n_1-1}, y_{q+n_1}]$. Likewise, $\boldsymbol{p_q}$ is the vector of past decisions; $\boldsymbol{p_q} = [p_{q-n_1}, p_{q-n_1+1}, \ldots, p_{q-1}]$. The linear and nonlinear coefficients for the FFF are c_i and c_{ijk}, and those for the FBF are d_i and d_{ijk}, respectively. The length of the linear parts of the two filters is denoted by n_1, and that of the nonlinear parts by n_2.

The FFF and the FBF blocks contain the necessary memory and circuits to form the functions $f_1(\boldsymbol{y_q})$ and $f_2(\boldsymbol{p_q})$ from y_q and p_q respectively, for each q. In the absence of any decision errors, $p_q = x_q$ for each q.

The coefficients of the two filters (FFF and FBF) are tuned during a training period. A known sequence of x_q is transmitted, and the same is fed back to the FBF (instead of the previous decisions from the decision device which is disconnected from the FBF in this mode). The error signal, which is the difference between the training signal and the decision device input, is minimized by adapting the FFF and FBF coefficients, as presented in Fig. 2.2.

Let $\boldsymbol{g_q}$ be a combined coefficient vector composed of the filter coefficients (as yet untuned) $c_i(i = -n_1, \ldots, n_1)$, $c_{ijk}(i, j, k = -n_2, \ldots, n_2)$, $d_i(i = -n_1, \ldots, -1)$ and $d_{ijk}(i, j, k = -n_2, \ldots, -1)$. $\boldsymbol{x_q}$ is the combined vector of the appropriate first and third-order terms from the received symbols and the past decisions and their conjugates such that $\boldsymbol{g_q} \times \boldsymbol{x_q} = f_1(\boldsymbol{y_q}) + f_2(\boldsymbol{p_q})$.

Recursive least squares (RLS) and least mean squares (LMS) are two popular algorithms used in adaptive filtering. RLS-based algorithms have good performance
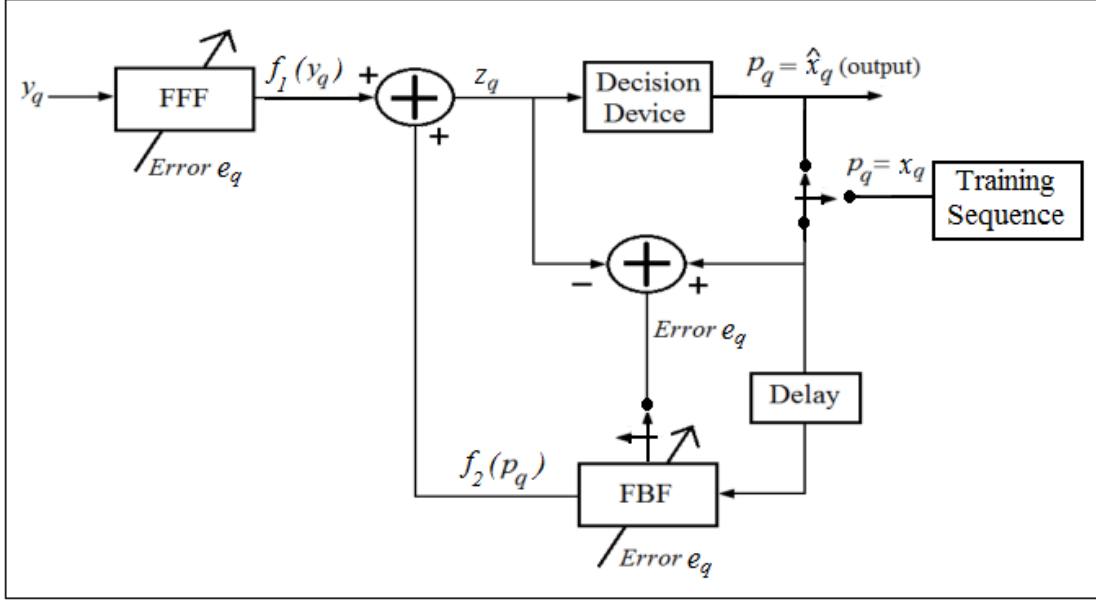
Figure 2.2: Decision feedback based equalizer training and decision modes

when applied in non-stationary channel models [37] but at the costs of increased computational complexity and possible instability [38]. Since the optical fiber channel changes very slowly, we choose the LMS algorithm. It has been successfully used before to adapt the tap weights for nonlinear equalizers [29] [30]. We use multiple iterations for every symbol to increase the convergence rate of the algorithm, as previously done in [34]. The tuning process is stopped after a certain number of iterations i_{max} (Algorithm 1). N_t denotes the size of the training dataset, and e_q^i denotes the (generally complex) error signal after iteration i. The convergence parameter μ must be carefully selected for proper convergence. μ can be a scalar, or if separate parameters for the different first and third-order coefficients are desired, a vector. The coefficients are tuned once for each q.

After the training is completed, it can be expected that the equalizer coefficients are properly tuned. The coefficients can then be obtained from $\boldsymbol{g}_{N_t}^{(i_{max})}$. Let the

---

Algorithm 1 Tuning of FFF and FBF coefficients

> **for** $q = 1$ to N_t **do**
>   **for** $i = 1$ to i_{max} **do**
> $e_q^{(i)} \leftarrow p_q - g_q^{(i)} x_q$
> $g_q^{(i+1)} \leftarrow g_q^{(i)} + 2\mu e_q^{(i)} x_q$
>   **end for**
> $g_{q+1}^{(1)} \leftarrow g_q^{(i)}$
> **end for**

---

decision device input be denoted by z_q. That is,

$$
\begin{aligned}
z_q &= \sum_{i=-n_1}^{n_1} c_i y_{q+i} + \sum_{i,j,k=-n_2}^{n_2} c_{ijk} y_{q+i} y_{q+j}^* y_{q+k} \\
&+ \sum_{i=-n_1}^{-1} d_i p_{q+i} + \sum_{i,j,k=-n_2}^{-1} d_{ijk} p_{q+i} p_{q+j}^* p_{q+k}.
\end{aligned}
\tag{2.7}
$$

A decision is made on z_q according to standard demodulation. We do not update the filter coefficients in the decision mode. This would add to the complexity and overhead. More importantly, if a particular decision is wrong, the resulting adjustments to the filter coefficients would be incorrect and may adversely affect future performance by increasing the probability of subsequent decision errors. The decision device output $p_q = \hat{x}_q$ is the decision on the received symbol y_q, and the FBF forms its input from the decision device output.

2.2 Nature and Convergence of Coefficients

During training, the equalizer must decide on a set of coefficients c_i, c_{ijk}, d_i and d_{ijk} that would–with help from the decision device–correctly map the received symbols y_q's to the transmitted symbols x_q's, in the presence of all the linear and nonlinear impairments, noise and crosstalk.

When the equalizer is trained conventionally as described in Section 2.1, the coefficients never converge to particular fixed values except under special circumstances, such as low fiber nonlinearity coupled with absence of ASE noise. (If we did have a model for the additive noise to use in (2.2), this observation might not have been true.) During the training phase, an optimization problem is solved, for *each* q, so that z_q equals the corresponding training symbol $p_q = x_q$. For any particular symbol pair, the optimization is performed very efficiently indeed, giving rise to very low values of individual $\boldsymbol{e}_q^{(imax)}$. Also, the effects of the residual third-order nonlinearity, the higher-order nonlinearity and the ASE noise unaccounted for in (2.4) all enter into the coefficient tuning process in an unpredictable manner for each q, and are deposited into the coefficients of the other terms.

In Figs. 2.3 – 2.4 we illustrate the behavior of the coefficients c_1 and c_{000} from a (single) simulation experiment run on QPSK (quadrature phase shift keying) modulated data transmitted at 21.4 Gbit/s over 50 SSMF spans at 2 dBm input power. The inline amplifiers have a noise figure (NF) of 5 dB. We found $n_1 = n_2 = 3$ to be a good compromise between performance and computational complexity. Other relevant system parameters are given in Table 2.1. Although noise in the system is relatively low, nonlinearity and nonlinear phase noise (NLPN) are not. Thus the coefficients fail to converge to steady values even as the training size increases. (Increasing the filter lengths does not seem to affect this.)

Table 2.1: Simulation Parameters

Description	Symbol	Value
pseudo-random binary sequence length		2^{15}
transmission wavelength		1550 nm
data-rate	B	21.4 to 50 Gbit/s
bits per symbol	k	
symbol rate	F_s	B/k
oversampling rate (at transmitter)	n_s	16
sampling frequency	F	$n_s F_s$
attenuation constant	α	0.2 dB/km
group-velocity dispersion parameter	β_2	-20 ps^2/km
nonlinear parameter	γ	2/(Wkm)
span length	L	80 km
number of spans	n	35 to 50
transmitter filter 3 dB bandwidth		$1/F_s$
transmitter filter rolloff factor		0.5
amplifier noise figure (practical)		5 dB
amplifier noise figure (ideal)		3 dB
optical filter bandwidth		$2B/F$
receiver filter 3 dB bandwidth		$1/F_s$
receiver filter rolloff factor		0.5
tolerance	ϵ	0.5
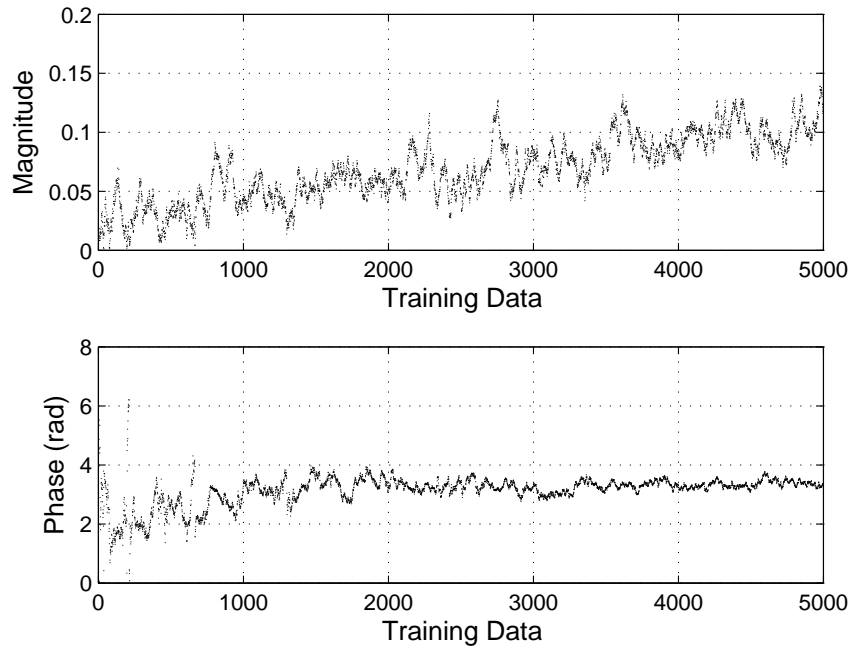
Figure 2.3: Coefficient c_1 vs training data length for conventional training
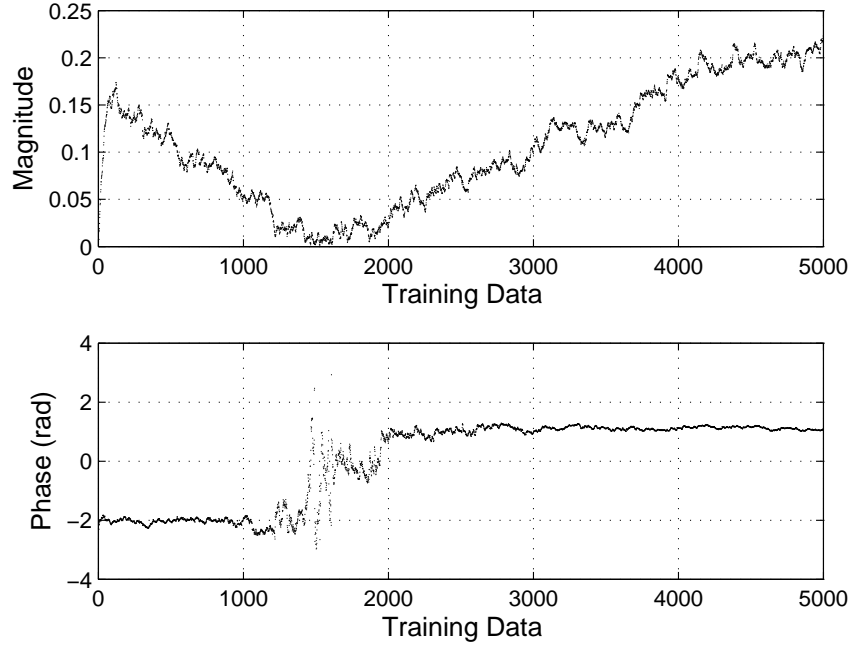
Figure 2.4: Coefficient c_{000} vs training data length for conventional training

To reduce the oscillation in the coefficients, at least to some degree, we modify the way the coefficients are tuned by forcing *suboptimal convergence* of z_q to p_q for each q. This scheme consists of a two-stage training mode. The stages are sequential, the second stage starts after the first ends. The training set is divided into two distinct subsets (which need not be equal in size), one for each stage. In the first stage, we train the equalizer conventionally to get the coefficients in the ballpark. In the second, instead of tuning the coefficients until z_q equals p_q, we set a condition so that the coefficient updating stops as soon as z_q is *sufficiently close* (to be defined) to p_q, and the training moves on to the next $(q+1)th$ symbol.

Recalling that in the training mode as described in Section 2.1, for each q, the coefficients are updated such that after each iteration in the LMS algorithm, z_q moves little by little towards p_q. Also, $\boldsymbol{g}_{q+1}^{(1)} = \boldsymbol{g}_q^{(i_{max})}$. That is, for the first iteration for the $(q+1)th$ training (y_{q+1}, p_{q+1}) pair, the coefficients are the same

as that for the last iteration for the qth pair. Then $\boldsymbol{g}_{q+1}^{(i)}$, $i = 1, 2, \cdots, i_{max}$, may move away from $\boldsymbol{g}_q^{(i_{max})}$ as $z_{q+1} = \boldsymbol{g}_{q+1}^{(i)} \times \boldsymbol{x}_{q+1}$ gets closer to p_{q+1} with every iteration.

We would like $\boldsymbol{g}_{q+1}^{(i_{max})}$ to be close to $\boldsymbol{g}_q^{(i_{max})}$ to identify the filter coefficients to be used in the decision mode. We define a region around p_q so that if z_q is in that region, we consider it close enough (a decision device would map the region to a desired point). For QPSK modulated data; for the *normalized* constellation point $p_q = \pm 1 \pm j$, we consider the optimization sufficiently good if $|p_q - z_q| < \epsilon$; ϵ being a non-negative real number to be tuned empirically, after considering the desired margin against possible noise that could hurt performance. The modified second training stage is illustrated in the pseudocode given in Algorithm 2. The result of this modification is not the same as setting i_{max} to be low: a low i_{max} may not always guarantee the desired amount of convergence, and the additional criterion minimizes the computational effort. Since the non-steadiness of the coefficients cannot always be perfectly eliminated, we use the arithmetic means of the filter coefficients over each q in the second training phase to find the corresponding fixed coefficients for use in decision mode.

Algorithm 2 Modified coefficient training

---

   for $q = 1$ to N_t **do**
      for $i = 1$ to i_{max} **do**
         $e_q^{(i)} \leftarrow p_q - \boldsymbol{g}_q^{(i)} \boldsymbol{x}_q$
         if $|e_q^{(i)}| \geq \epsilon$ **then**
            $\boldsymbol{g}_q^{(i+1)} \leftarrow \boldsymbol{g}_q^{(i)} + 2\mu e_q^{(i)} \boldsymbol{x}_q$
         else
            $\boldsymbol{g}_q^{(i_{max}+1)} \leftarrow \boldsymbol{g}_q^{(i)}$
            $i \leftarrow i_{max}$
         end if
      end for
      $\boldsymbol{g}_{q+1}^{(1)} \leftarrow \boldsymbol{g}_q^{(i)}$
   end for

Besides ensuring the coefficients do not change much with q, the suboptimal convergence modification allows us to run the LMS algorithm, for each q in the second training stage, a smaller number of iterations, since we neither seek not want perfect convergence of z_q to p_q. Simulation runtimes indicate that this cuts down on computations by 30% or more.

In Figs. 2.5 and 2.6 we have the behavior of the same coefficients c_1 and c_{000} from a simulation experiment run at 2 dBm input power when suboptimal convergence is applied. The second training stage starts at $q = 2000$. (As long as the second training set is of a reasonable length, the actual starting point is not important, since all the coefficients are affected at the same time by the training procedure for proper equalization.) These are the counterparts to Figs. 2.3 and 2.4 respectively, and are generated under identical conditions. In Table 2.2, we present the absolute values of the coefficient of variations (ratio of standard deviation to mean) of c_1 and c_{000} for conventional and suboptimal convergence training, and conclude that suboptimal convergence is successful in keeping the coefficients reasonably constant even as q changes.
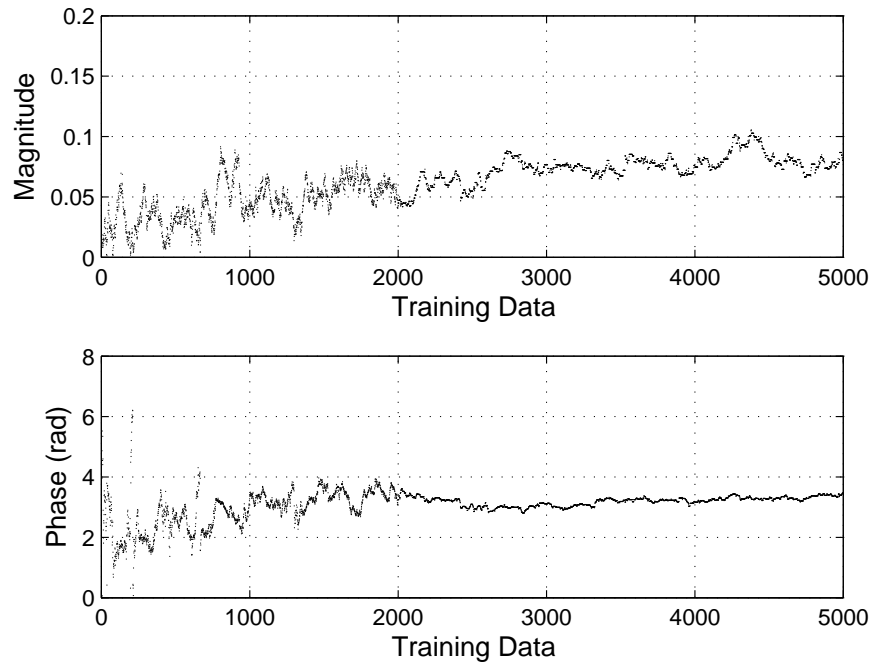
Figure 2.5: Coefficient c_1 vs training data length for suboptimal convergence
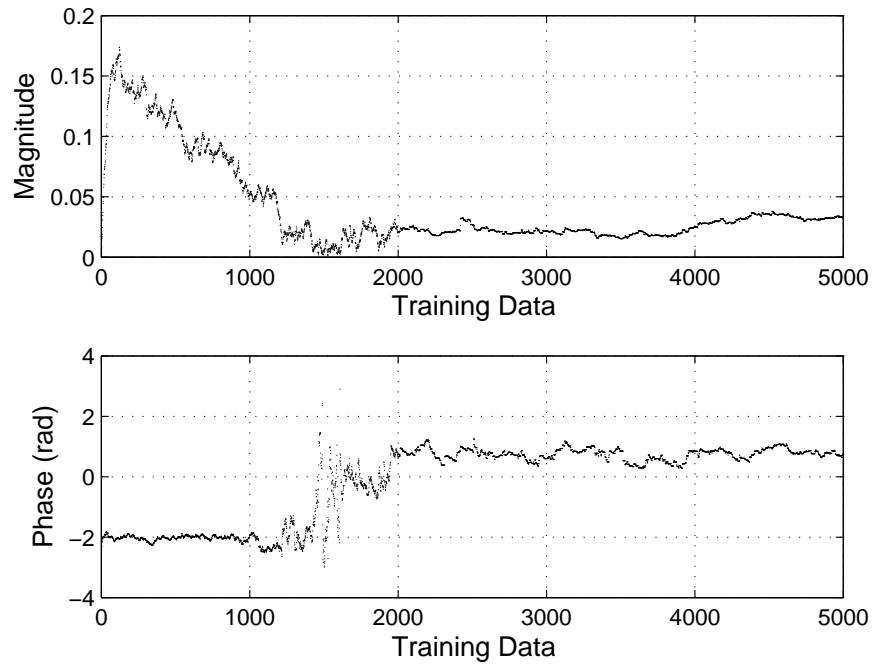
Figure 2.6: Coefficient c_{000} vs training data length for suboptimal convergence

Table 2.2: Coefficient of Variation (CV) of c_1 and c_{000}

	Conventional Training	Suboptimal Convergence
$|CV(c_1)|$	0.5605	0.2301
$|CV(c_{000})|$	1.9560	0.3805

Fig. 2.7 shows the behavior of the coefficient d_{-1} when suboptimal convergence is employed. The results are from a simulation experiment run at 2 dBm input power and the second training stage starts at $q = 2000$.
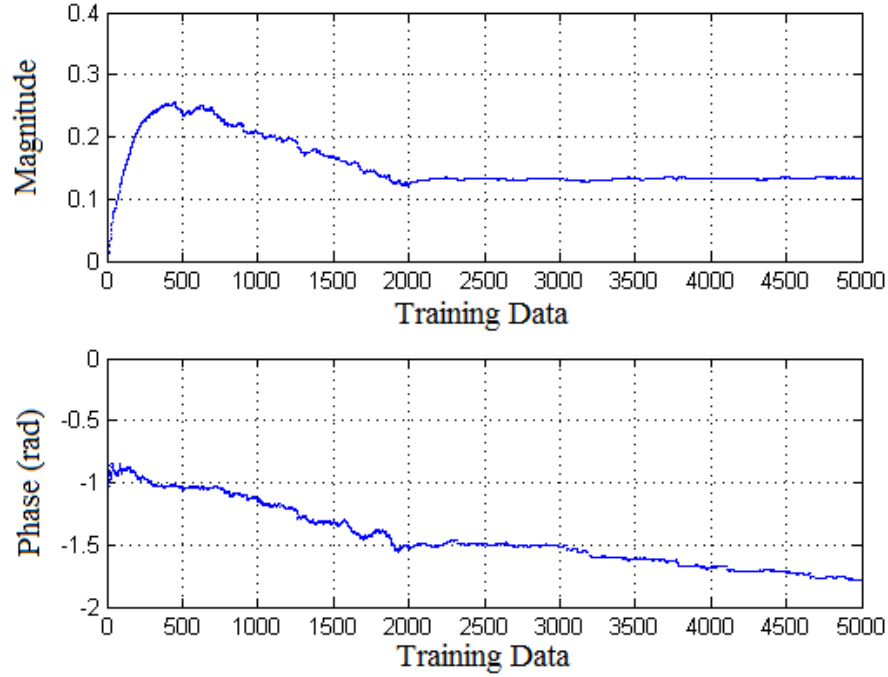
Figure 2.7: Coefficient d_{-1} vs training data length for suboptimal convergence

2.3   Simulation Results

In this section, we apply our DFE to data obtained from simulation experiments (using the split-step Fourier method) conducted under different system conditions and modulation formats (OOK, on-off keying or QPSK, quadrature phase shift

keying) to verify that the performance is satisfactory. We primarily compare the performance of the DFE against that of digital backpropagation, which is considered a benchmark in the literature [11]. The relevant simulation parameters are indicated in Table 2.1.

In Fig. 2.8, we have plotted the bit error rate (BER) curves for the unequalized, linearly equalized (phase rotation filter), conventional nonlinear DFE equalized, modified nonlinear DFE equalized (with $n_1 = n_2 = 3$), backpropagated (with full nonlinearity compensation) and the no-nonlinearity limit cases. (The nonlinear effects become more pronounced with increasing input power.) The BER for the unequalized case is quite high, mostly due to the nonlinear phase rotation [39]. The first-order filter performs well at lower levels of nonlinearity. At the lower end of the input power range, the performances of all four techniques are comparable and close to the no-nonlinearity limit. At higher levels of nonlinearity, backpropagation and the nonlinear DFE outperform the other techniques.

The slight differences in performance between the DFE and the backpropagation method can be attributed to the fact that the modified DFE equalizer makes an adjustment for the uncompensated terms so that the equalized constellations are in a region around the respective symbol points. Backpropagation, on the other hand, works methodically span by span undoing the effects of the linear and nonlinear impairments. Noise in the system causes inaccuracy in the backpropagation solution, regardless of the backpropagation step size.

Next, in Fig. 2.9, we have the BER curves from a simulation experiment run on a higher rate system. This plot is of a similar nature as Fig. 2.8. At this higher data-rate, the nonlinear impairment effects are greater [36], as is the effect from the interaction of the ASE noise with the nonlinearity. The nonlinear DFE and backpropagation again both perform better than the linear filter or the
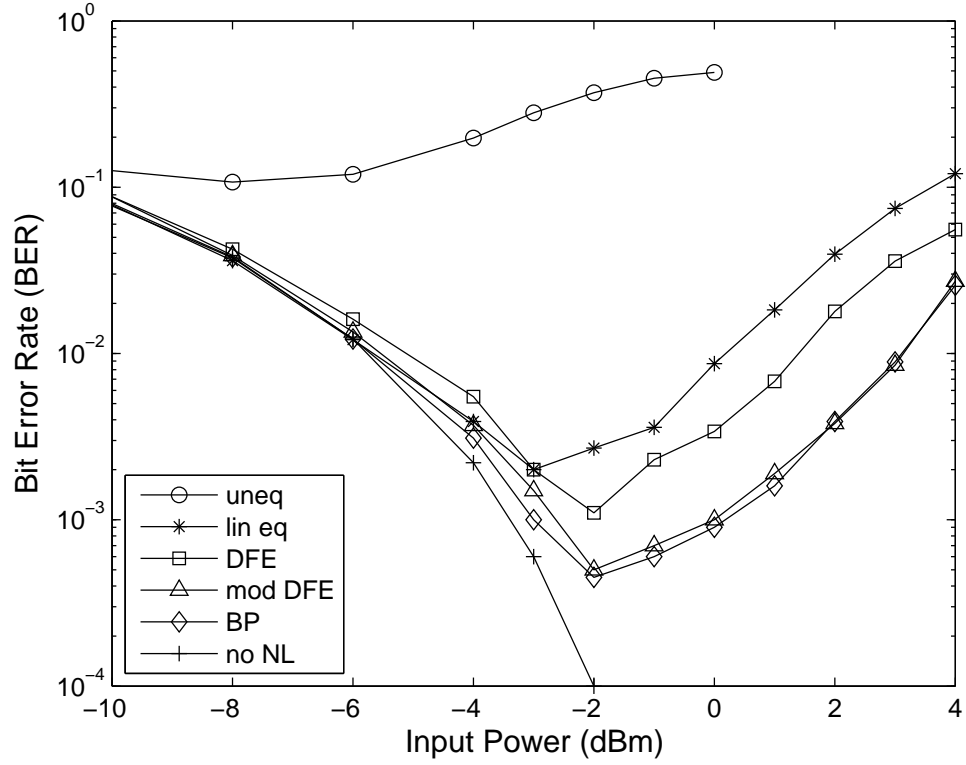
Figure 2.8: BER performance (21.4 Gbit/s QPSK, 50 spans, 5 dB amplifier NF)

conventionally trained DFE.

Additional results on the performance of the nonlinear DFE are presented in Table 2.3. The conditions for the simulation experiments are:

A: 21.4 Gbit/s data-rate, 50 SSMF spans, 3 dB amplifier noise figure,

B: 50 Gbit/s data-rate, 30 SSMF spans, 5 dB amplifier noise figure,

C: 50 Gbit/s data-rate, 40 SSMF spans, 3 dB amplifier noise figure.

Table 2.3: Minimum BER for QPSK Modulated Data ($\times 10^{-3}$)

Conditions	Unequalized	Nonlinear DFE	Backpropagation
A	4.3838	0.133	0.133
B	10.4412	0.343	0.343
C	11.0841	0.457	0.429

From Figs. 2.8- 2.9 and Table 2.3, we conclude that the proposed nonlinear
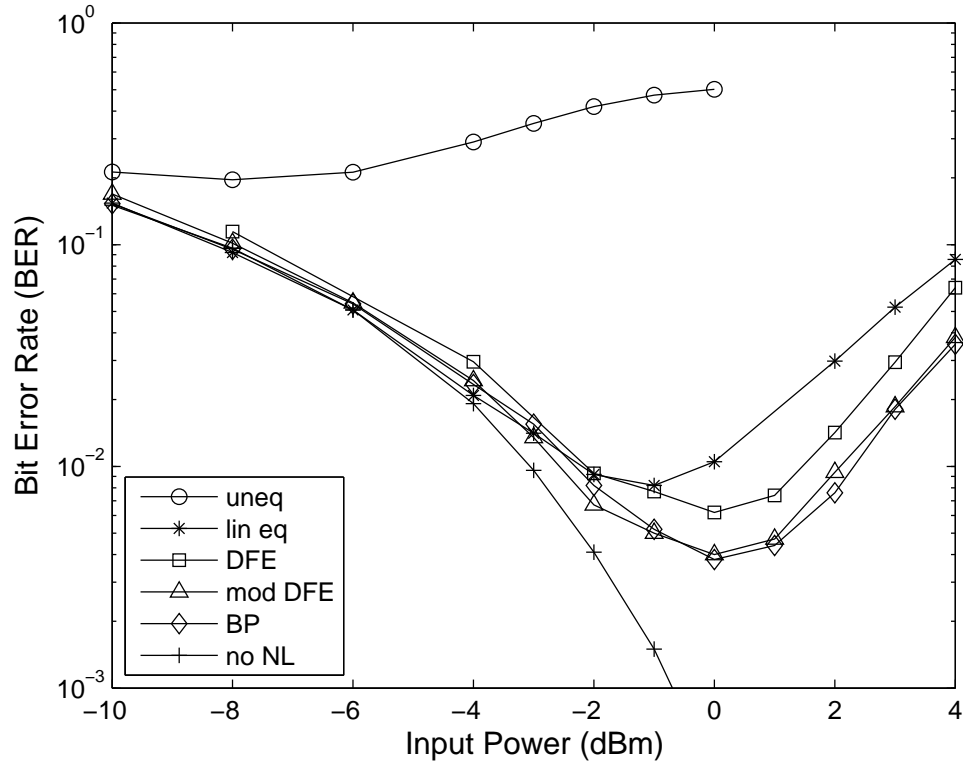
Figure 2.9: BER performance (50 Gbit/s QPSK, 40 spans, 5 dB amplifier NF)

DFE works satisfactorily for QPSK modulated data transmitted at different data-rates over different lengths of fibers, with different noise characteristics, and its performance matches that of backpropagation.

In Fig. 2.10 we have results from a simulation experiment run on OOK modulated data. Comparing with Fig. 2.8, we first note that the unequalized BER is higher for QPSK modulated data. This is due the effect of nonlinear phase noise on PSK modulation schemes. Next, we note that the minimum BER after equalization is lower for QPSK. QPSK is a more power efficient modulation technique, and modulated data has constant amplitude.

Additional results on the performance of the nonlinear DFE are presented in Table 2.4. The conditions for the simulation experiments are:
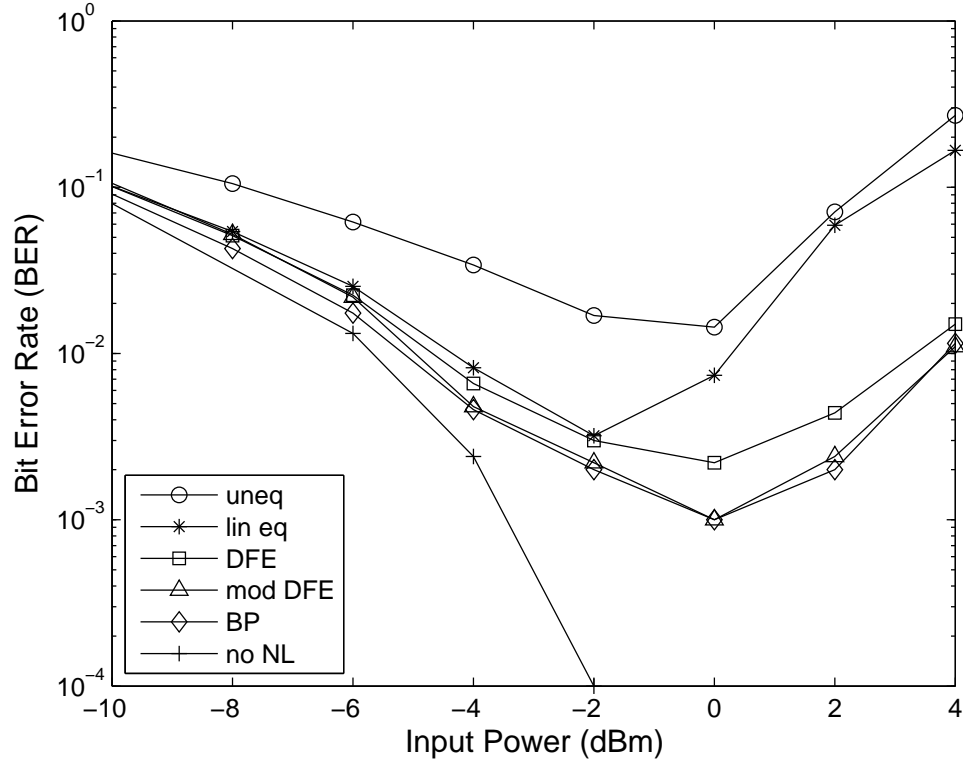
Figure 2.10: BER performance (25 Gbit/s OOK, 40 spans, 5 dB amplifier NF)

D: 25 Gbit/s data-rate, 50 SSMF spans, 3 dB amplifier noise figure,

E: 25 Gbit/s data-rate, 50 SSMF spans, 5 dB amplifier noise figure,

F: 50 Gbit/s data-rate, 40 SSMF spans, 5 dB amplifier noise figure.

Table 2.4: Minimum BER for OOK Modulated Data ($\times 10^{-3}$)

Conditions	Unequalized	Nonlinear DFE	Backpropagation
D	9.658	0.667	0.667
E	28.459	4.345	3.886
F	61.261	5.031	5.029

We can draw similar conclusions as from the corresponding QPSK performance results, that is, lower amount of noise produces a lower BER in the equalized data, as do a lower number of spans. A higher data-rate hurts the minimum equalized BER. From Fig. 2.10 and Table 2.4, we conclude that the DFE is just as effective

Table 2.5: Notations

Description	Symbol
data block size	N
size of data for first training stage	N_1
size of data for second training stage	N_2
total training data size	N_t
number of filter coefficients	n_c
iterations per symbol for first training stage	i_{max}
mean iterations per symbol for second training stage	i_{avg}

in suppressing linear and nonlinear physical impairments for OOK modulated data transmitted under different conditions.

2.4   Complexity Analysis

In this section, we study if the modified DFE algorithm is computationally attractive. We would like to note that the only overhead in our proposed equalization algorithm arises from the need to train the filter coefficients. Once the tuning process is complete, there is no further introduction of overhead since the optical transmission link changes very slowly compared to the data-rate. The notations used are explained in Table 2.5.

For the modified DFE, the number of calculations during first training stage is $(N_1 n_c + N_1 i_{max} n_c + N_1 i_{max})$. Since n_c and i_{max} are much smaller than the size of the training data for either stage, the complexity of this stage is $O(N_1)$. The number of calculations during the second training stage is $(N_2 n_c + N_2 i_{avg} n_c + N_2 i_{avg})$, and the complexity of this stage is $O(N_2)$. Since $N_1 \approx N_2$, the complexities are or the same order. The second training stage involves a comparison to check if $|p_q - z_q| < \epsilon$; however i_{avg} is usually less than half of i_{max} to compensate for this.

For the conventional DFE, the number of calculations during the single training

stage is $(N_t n_c + N_t i_{max} n_c + N_t i_{max})$, and thus the complexity is $O(N_t)$, with $N_t \ll N$ due to the very slowly changing channel. For either DFE configuration, the decision mode (actual operation) is identical. The number of calculations required for a decision on each symbol is of the order of n_c.

For backpropagation, an oversampling rate of 3 or 4 is sufficient [11]. The computational complexity of back- propagation algorithm is $O(N \log N)$ [40]. The complexity of digital backpropagation is also a function of the number of the segments, which can be huge. For reduced complexity backpropagation methods [41], the complexity remains of the same *order*. This shows that the DFE or the modified DFE is computationally less taxing than digital backpropagation for a transmission link with periodic dispersion compensation. For an uncompensated link, the channel has a larger memory necessitating a larger number of equalizer taps and a corresponding increase in the algorithm complexity. The complexity of digital backpropagation remains unaffected.

2.5   Performance Analysis

The error probability evaluation of a DFE can be studied in two subcategories — with and without error propagation. A wrong decision on a particular symbol increases the probability of subsequent errors by worsening the post-cursor impairment effects. Thus the error is propagated. In the absence of error propagation, we assume that as far as post-cursor compensation is concerned, all previous decisions are correct.

Due to residual and higher-order uncompensated nonlinearities, unmodeled ASE noise and NLPN in the system, and possibility of decision errors in the equalizer, z_q in (2.7) is different from x_q in (2.3). Let us call this deviation

$\delta_q = z_q - x_q$. For no error propagation, using (2.5), (2.6) and (2.7),

$$\delta_q = z_q - x_q = f_1(\boldsymbol{y_q}) + f_2(\boldsymbol{p_q}) - x_q = f_1(f_3(x_q)) + f_2(x_q) - x_q, \qquad (2.8)$$

where

$$f_3(x_q) = \sum_{i=-q}^{N-q-1} \rho_i x_{q+i} + \sum_{i,j,k=-q}^{N-q-1} \rho_{ijk} x_{q+i} x_{q+j}^* x_{q+k} \qquad (2.9)$$

from (2.2) when the higher-order terms are ignored. If error propagation is considered,

$$z_q = f_1(\boldsymbol{y_q}) + f_2(\boldsymbol{p_q}) = f_1(f_3(x_q)) + f_2(decision[z_q]), \qquad (2.10)$$

setting up a recursive equation to be solved.

There will be an error in decision for a particular q if δ_q crosses a certain threshold. For normalized QPSK modulated data, the conditions for error in the in-phase and quadrature components are respectively $|Re[\delta_q]| > 1$ and $|Im[\delta_q]| > 1$.

Theoretically, in the absence of noise, we can obtain δ_q analytically to find the error probability. However, if there is no noise in the system, then the FFF and FBF lengths can be chosen large enough such that the error probability is zero at all but the highest input power levels (the effect of the higher-order nonlinearities become more significant with increasing power level). When noise is present, having large filter lengths can be useless or hurt the design, as already discussed.

Related work in existing literature, such as [42], make use of limitations and assumptions about the channel, signals and the noise in the system which are not satisfied by our nonlinear optical communication system that has interacting and co-propagating amplifier noise. The problem is further complicated since we use

training to tune the filter coefficients. Thus we cannot directly use such results.

The condition that a particular decision error does not influence future error probability is enforced by setting the previous p_q's to the corresponding x_q's while a decision is being made for the current y_q. We measure the deviations $\delta_q = z_q - x_q$ and analyze the distributions of the in-phase and the quadrature components of δ_q. A numerical estimate of the DFE error probability (BER) can be obtained by observing the percentage of each component of δ_q having magnitudes greater than unity.

We provide an illustration in Fig. 2.11. The optical amplifiers do *not* introduce any noise in the system, only for this example. The nonlinearity is quite high due to the large number of spans and the high power level, but the effect of the residual third-order and unaccounted higher-order terms are only just beginning to affect the performance. The distribution of indicates that the tails are stretching a little over the -1 and $+1$ limits. The bit error rate is found to be 0.0011.

We illustrate how previous decisions can affect the equalizer performance. We consider a situation where all inline amplifiers have a noise figure of 5 dB. The numerical best case error probability estimates (assuming all previous decisions to be correct) and actual BERs obtained via a simulation experiment are plotted across a range of input power levels in Fig. 2.12. In addition, we plot the BER which would have been obtained if half the previous decisions from the DFE were incorrect.

The equalizer performance is close to the no error propagation BER curve. Even in the presence of a decision error, the DFE can provide a large enough margin to provide correct decisions and recover quickly. We conclude that not considering enough higher-order terms and the contribution from all signals within the range of influence are the more important factors affecting the BER performance.
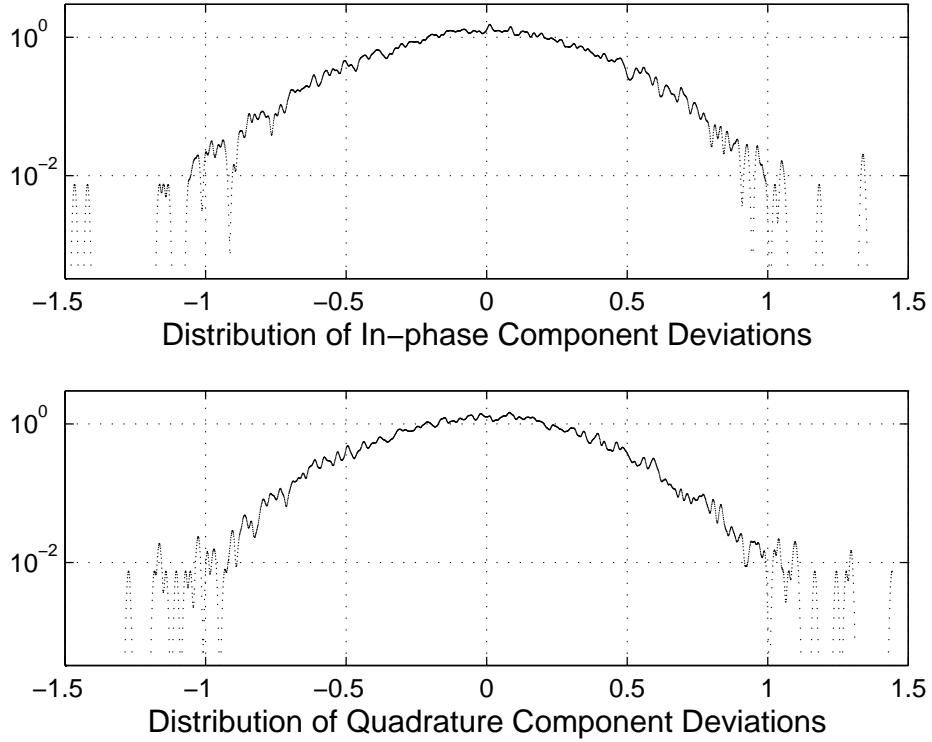
Figure 2.11: Distributions of δ_q (21.4 Gbit/s QPSK, 60 spans, 3 dBm power)

The difference between the measured BER and the no-error-propagation error estimate curves is a little larger at higher power levels since the post-cursor nonlinear impairment enhancement in case of a decision error is magnified with more nonlinearity.

We can model the error propagation mechanism as a Markov chain [43] using stochastic states, transition probabilities (to be determined numerically), and use some simplifications since the number of FBF taps is not unreasonably large. Also, most optical communication systems employ modulation formats for which the number of constellation points (and distinct magnitudes of possible errors) is not very large either.

We consider separately the decisions on the in-phase and quadrature compo-
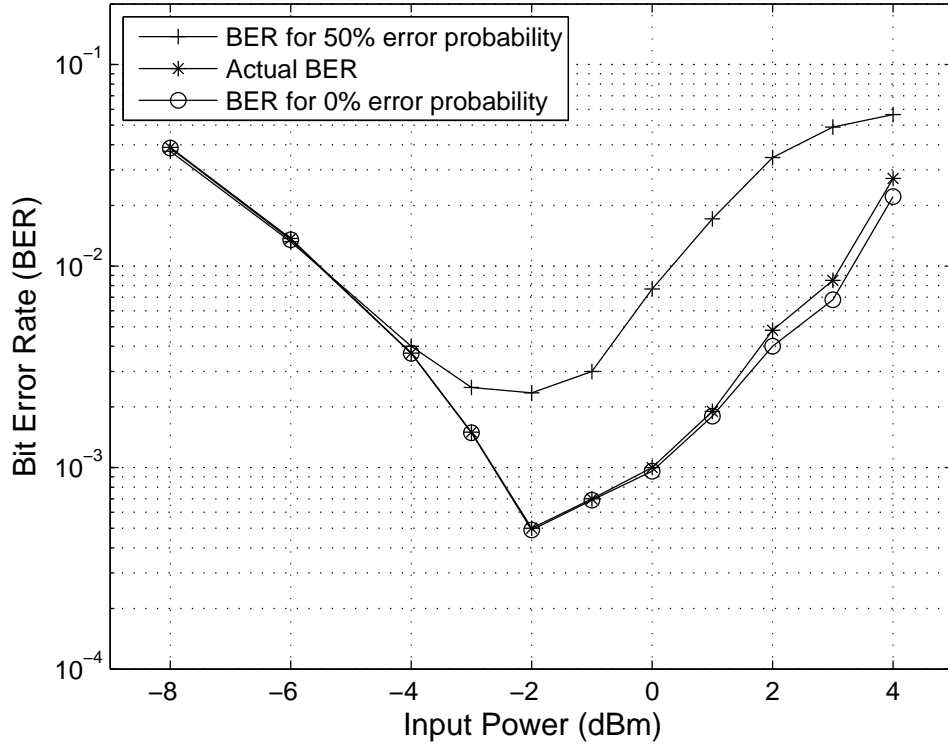
Figure 2.12: Effect of error propagation on DFE performance (21.4 Gbit/s QPSK, 50 spans, 5 dB NF)

nents. For each component, each previous decision which enters into the current decision calculation can either be correct or wrong. Thus the total possible number of combinations of previous decision errors is 2^{2n_2}, including the case where all previous $2n_2$ decisions are correct. A state is assigned to each error combination.

We denote a correct decision with 0, and an erroneous one with 1. As an example, assume that we have $n_2 = 3$, and that the DFE is making a decision on the qth symbol. Also assume the current state of the system is $V_q = [000000]$, where the first two bits indicate the decision on the $(q-1)th$ symbol, the middle two that on the $(q-2)th$ one, and the last two on the $(q-3)th$ one, with the first, third and fifth bits in a state reserved for in-phase symbol bits. The state

$V_q = [000000]$ indicates the decisions on all three past symbols were correct. If the decision on the in-phase bit is incorrect, the next state will be $V_{q+1} = [100000]$, and so on. This allows us to create a state transition table, or equivalently, a state transition diagram, for the error propagation mechanism. With the knowledge of current state, corresponding distributions of the deviation δ_q are generated to yield the transition probabilities.

The error propagation mechanism system can reside in exactly one state at any sampling instant q, and the number of total possible states is finite, as is each of the states. State transition probabilities are determined only by the current state and are independent of all previous states. Also, these probabilities are constant over time, as long as the optical communication system parameters are unchanged. Thus the error propagation mechanism system is a Markov chain process.

The transition matrix of this Markov chain contains the state transition probabilities. Also, by the nature of definition of the states, none of them are absorbing. The limiting distribution is numerically obtained by raising the transition matrix to a sufficiently high power and observing any row.

The above methodology allows us to predict the probability that the error propagation mechanism system is in a certain state, the probabilities being constant in the limiting case. The appropriate state probabilities are then added to compute the required probability of error of the in-phase or the quadrature bit, or the symbol error probability.

Assume that we wish to numerically estimate of the bit error probability when the input power level is 2 dBm in the system in the previous example. With $n_1 = n_2 = 3$, the number of states is $2^{2\times3} = 64$. With the filter coefficients fixed from training, we apply the DFE to a known dataset. The current state V_q contains the previous decisions and we check the probability of a wrong decision for all

Table 2.6: State Transition Table for DFE Error Propagation System

Present State	Next State	Transition Probability
$S_0 = [000000]$	$S_0 = [000000]$	994.23×10^{-3}
$S_0 = [000000]$	$S_1 = [010000]$	2.57×10^{-3}
$S_0 = [000000]$	$S_2 = [100000]$	3.20×10^{-3}
$S_0 = [000000]$	$S_3 = [110000]$	0
$S_1 = [010000]$	$S_4 = [000100]$	933.11×10^{-3}
$S_1 = [010000]$	$S_5 = [010100]$	61.69×10^{-3}
$S_1 = [010000]$	$S_6 = [100100]$	5.20×10^{-3}
$S_1 = [010000]$	$S_7 = [110100]$	0
$\vdots$	$\vdots$	$\vdots$
$S_{62} = [101111]$	$S_{56} = [001011]$	926.25×10^{-3}
$S_{62} = [101111]$	$S_{57} = [011011]$	4.57×10^{-3}
$S_{62} = [101111]$	$S_{58} = [101011]$	69.18×10^{-3}
$S_{62} = [101111]$	$S_{59} = [111011]$	0
$S_{63} = [111111]$	$S_{60} = [001111]$	839.70×10^{-3}
$S_{63} = [111111]$	$S_{61} = [011111]$	76.09×10^{-3}
$S_{63} = [111111]$	$S_{62} = [101111]$	84.21×10^{-3}
$S_{63} = [111111]$	$S_{63} = [111111]$	0

possible V_q's. The resulting percentage of errors in the in-phase and the quadrature bits generate the transition probabilities. Sections of the state transition table are given in Table 2.6.

The transition probabilities in the state transition table form the elements of the transition matrix (with 64 rows and 64 columns). By raising the transition matrix to a sufficiently high power, each row of the limiting distribution of the transition matrix is numerically determined to be [982.31 2.54 3.17 0 0 0 0 0] $\times$ 10^3.

Finally, from the probabilities of the system being in the relevant states, we determine the bit error rate to be 3.09×10^{-3}, which is a good estimate for the simulated BER of 3.83×10^{-3}. We also check the BER corresponding to particular current state V_q's, and this is illustrated in Table 2.7,

Table 2.7: BER as a Function of Current States

V_q	BER
[001000]	3.69×10^{-3}
[000100]	3.43×10^{-3}
[000010]	3.26×10^{-3}
[000001]	3.20×10^{-3}
[001100]	3.49×10^{-3}
[001010]	3.20×10^{-3}
[001001]	3.52×10^{-3}
[000110]	3.57×10^{-3}
[000101]	3.23×10^{-3}
[000011]	3.12×10^{-3}

2.6   Chapter Summary

This chapter discusses the application of a modified nonlinear DFE in a long-haul coherent optical fiber communication system. The channel is lossy, dispersive, nonlinear and contains random amplifier noise along with NLPN. We design the equalizer structure from a mathematical description of the system, and study the effects of noise, NLPN and residual or uncompensated nonlinearity on the filter coefficients when the DFE is trained conventionally, and also using our modified convergence method.

We note the possible benefits of the modified training scheme and apply the designed DFE to a fiber-optic system under various conditions and modulation formats, comparing with results obtained from backpropagation. We conclude that the performance is satisfactory, and that the computational complexity is modest. Finally we study the problem of performance prediction of the DFE when applied to an optical communication system, and use a numerical method for the same both in absence and in presence of error propagation.

In the future, we would like to apply a DFE to a dispersion unmanaged system

where the effects of dispersion, nonlinearity and their interaction are stronger. A rigorous mathematical analysis on the effects of noise and NLPN on filter coefficients is also desirable. This would enable us to design more effective training schemes to tune the coefficients. Finally, when error propagation is taken into account, analytical performance prediction of a DFE when applied to fiber-optic communication systems is a topic which requires further investigation.

The performance limit in the absence of nonlinearity is not approached in a noisy nonlinear channel for high power levels by digital backpropagation or our equalizer, and there is an opportunity of achieving even better performance. To close this gap, we seek to obtain more precise knowledge about the nonlinear interaction between signal and noise.

Chapter 3

Single Pulse Analysis of Nonlinear Fiber Transmission

In the last chapter, we presented a nonlinear equalizer to correctly predict the transmitted signal. We made an approximate adjustment for the nonlinear signal-noise interactions to mitigate the same, and provide good performance. It is important to note that we did not use an accurate model of the interaction, and merely used an approximation. This chapter seeks to understand the nature of such interactions using the simplest non-trivial scenario possible so that we can approach the performance which can be achieved when such nonlinear interactions are absent. We study a single signal pulse plus noise propagating through a single span of fiber.

Optical fibers–due to their inherent advantages–are well-suited for high speed long distance communication. However, the optical fiber channel is dispersive and nonlinear [44] when the communication link stretches for a few thousands of kilometers. In a legacy optical system, periodic restoration of signal strength and dispersion compensation are performed. The random amplified spontaneous

emission (ASE) noise from inline amplifiers interacts nonlinearly with the deterministic signal, while the signal interacts with itself, as does the noise, to generate crosstalk at the receiver [5]. In general, the noise component at the receiver end of an optical communication channel is neither uncorrelated nor Gaussian due to the signal-signal, signal-noise and noise-noise interactions.

Communication in the presence of additive Gaussian noise is well understood and has been extensively studied in the past [45]. If the noise is non-Gaussian, sufficient number of independent noise components from individual sources may combine together through the central limit theorem to approximate Gaussian statistics [46]. However, if the number of additive noise sources is not high enough, or in general, if the Gaussian approximation does not hold for a particular communication system, a different approach may be necessary. We employ Gaussian mixture modeling to study the interaction between amplifier noise, nonlinearity and data in optical fiber communication systems so that signal processing techniques may be used to improve the system performance.

There has been considerable work published on the effects of noise in optical fibers. Some authors have investigated the bit error rate of the transmitted signal or the signal-to-noise ratio at the receiver end of an optical fiber transmission channel under conditions of zero or negligible dispersion [47], [48]. When dispersion is present, [49] derives the received signal statistics in the limit of low dispersion. In [50], the author analyzes the probability density function for a phase modulated communication system. Receiver end statistics such as phase noise distribution, bit error rate, signal detection have been analytically studied utilizing different methods in the literature [51], [52].

For a dispersive channel with ASE noise, researchers have also analyzed the achievable information rate and the multiuser capacity [53], [54]. While works

discussing the actual propagation of the signal and the ASE noise in the channel are comparatively rare, one such piece of literature is [55]. The authors develop a model which describes the ASE noise in systems with distributed Raman gain, and study the interaction between the signal and noise. Simulations and experiments are used to validate the model. It is shown that for larger realistic input signal power levels, the noise statistics deviate significantly from the Gaussian distribution. Another work of a similar nature is [56] where the author studies correlated noise behavior. This work uses computational techniques for the analysis and characterization of nonlinear phase noise and its impact on system performance.

We would like to understand the effects of ASE noise (originating from the amplifiers used to restore signal power) affecting an optical signal as it travels through an optical communication system. As seen above, while there has been considerable research done in this field, critical knowledge about the nature of the signal-noise nonlinear interaction is incomplete. To this end, we investigate the simplest non-trivial scenario possible: the transmission of a single pulse through a single span of standard single mode optical fiber (SSMF), the single signal pulse being affected by a single ASE noise source.

Our communication system consists of a transmitter with a pulse-shaping filter and a noise source which adds additive white Gaussian noise (AWGN) to the single pulse which is transmitted through a single fiber span. We note that the noise source is before the optical fiber span so that the noise can interact with itself and the transmitted signal. We denote the digital data by x, the noise by $w(t)$, and the observed receiver output signal by $y(t)$. We transmit $s(t) = xp(t) + w(t)$, where $p(t)$ is the pulse shape. At the receiver end, the received signal is coherently detected, filtered and sampled. The set-up is illustrated in Fig. 3.1.
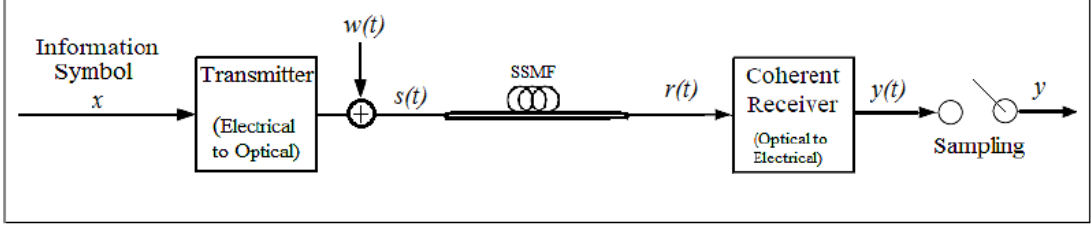
Figure 3.1: Single pulse transmission over a single span

This is the most fundamental situation which can be imagined. There are no interference effects from other adjacent pulses, or the noise affecting them. We have one span of fiber only. We assume no effects from the amplifier other than additive noise, nor do we have dispersion compensation. The propagating signal is affected by itself and the single noise source. In effect, this study can serve as the building block for subsequent research, when other effects are included. This is the motivation behind studying this fundamental scenario.

Let f denote the effect of the fiber (physical impairments). Then, for the set-up in Fig. 3.1, $y(t) = f(x(t) + w(t))$. After filtering and sampling, we assume that the output variable can be written as $y = \widetilde{x} + \widetilde{w}$, and study the statistics of $\widetilde{w}$. The signal term $\widetilde{x}$ is deterministic.

We use different techniques to analyze, describe and understand the various interaction effects. Gaussian mixture modeling (GMM) is a popular tool in data clustering applications [57]. In essence, a GMM is a technique to decompose an unknown distribution into a mixture of two or more Gaussian components, which can then be studied separately. Gaussian mixture models (GMM) have been used to model a variety of non-Gaussian noise environments [58]. Signal processing methods can be developed to work on the actual distribution by combining algorithms suited to each Gaussian component, making understanding, analysis and further processing more convenient [59]. Few existing works discuss application of GMM

in communication theory. In [60], the authors show that the interference in a wireless random-access communication system network could be modeled as a mixture of Gaussian and alpha-stable noise. In [61], independent component analysis using a mixture of Gaussian kernels is applied for separation of quadrature amplitude modulated (QAM) sources. In [62], the authors employ GMM for source localization. GMM is used to approximate a probability density function and this aprroximation is shown to be theoretically arbitrarily accurate. A related work [63] is concerned with finding analytical upper and lower bounds for an MMSE using Gaussian mixture distributions.

In this chapter, we explain the theory and application of GMM to the research problem and present compelling evidence that this is an useful tool for analyzing the optical fiber channel. We reduce the complicated receiver end noise to Gaussian components for which signal processing techniques already exist and obtain channel descriptions (such as mutual information and optimum constellations) which would enable us to find or predict useful characteristics of our fiber-optic communication system. While for lower levels of nonlinearity, the existing methods (such as VSTF) have been proved to be capable, we present evidence that at higher nonlinearity, our proposed analysis technique using GMM is superior.

The rest of the chapter is organized as follows. Section 3.1 provides a brief background on the Expectation Maximization (E-M) algorithm which is our algorithm of choice for obtaining the Gaussian Mixture Model description. Section 3.2 shows how GMM can be utilized to decompose the receiver output into individual Gaussian components. Section 3.3 reviews the use of VSTF for analyzing the mutual information and channel capacity of our communication system. In Section 3.4, we check the accuracy of our decomposition methods using some standard techniques and in Section 3.5, we use the stated techniques to compute

the mutual information of the channel. Section 3.6 summarizes the chapter and discusses possible future research directions.

3.1 Gaussian Mixture Modeling with Expectation Maximization Algorithm

A GMM is a parametric probability density function represented as a weighted sum of Gaussian component densities. Mathematically, this can be expressed as

$$P_{GMM} := P_W(w) = \sum_{m=1}^{M} r_m P(w|\mu_m, \Sigma_m), \qquad (3.1)$$

where W is a data vector we are interested in decomposing into Gaussian components, M is the number of mixture components, r_m is the weight of the mth component, with μ_m and Σ_m being its mean vector and covariance matrix, respectively. Gaussian mixture modeling may be more appropriate than k-means clustering when clusters have different sizes and correlation within them [64].

In our problem, the noise component of the received signal is not entirely Gaussian. However, we can classify it into separate groups, or clusters, with each group being distributed normally. GMMs have an unique capability of representing a large class of sample distributions and are often used to form smooth approximations to arbitrarily shaped densities [65].

We intend to classify the interference into separate groups, or clusters, with each group being distributed normally. The clusters in a GMM are assigned by selecting the component which maximizes the posterior probability of a particular data point (observation) belonging to a cluster. This decomposition technique—which is sometimes considered a soft clustering method—uses an iterative algorithm

which converges to a local optimum. The posterior probabilities for each point indicate that each data point has some probability of belonging to each cluster. Each component in the GMM has its multivariate density expressed as

$$P(w|\mu_m, \Sigma_m) = \frac{1}{2\pi^{\frac{1}{2}}|\Sigma_m|^{\frac{1}{2}}} e^{-\frac{1}{2}(w-\mu_m)\Sigma_m^{-1}(w-\mu_m)^T} \qquad (3.2)$$

The Expectation-Maximization (EM) algorithm [66] is a popular method for obtaining a GMM. Let for a GMM model with M components, Θ denote the complete set of parameters; that is

$$\Theta = \{r_1, r_2, \cdots, r_M, \mu_1, \mu_2, \cdots, \mu_M, \Sigma_1, \Sigma_2, \cdots, \Sigma_M\}, \qquad (3.3)$$

where r_i, μ_i and Σ_i are respectively the weight, mean vector and covariance matrix of the ith GMM component.

For our system, each term is bivariate, because we, in general, transmit complex (phase modulation) data, and in any case, the noise is complex. In our work we assume single polarization. If dual polarizations are used, the model can be extended to four terms.

In the expectation step, denoting the current parameter values as Θ, we compute the membership weights r_m for all data points w_i, $1 \leq i \leq N$ and all mixture components $1 \leq m \leq M$ to which we want to fit a GMM. (The membership weights are calculated using Bayes' rule.) This gives us an NM matrix (N rows and M columns) of membership weights. Each row of the matrix should sum to 1.

In the maximization step, we use the membership weights and the data and calculate new parameter values for the next step. We assign to component m the sum of the membership weights for the mth component. Next we calculate the

updated means $\mu_m^{new}, 1 \leq m \leq M$ as

$$\mu_m^{new} = \frac{1}{N} \sum_{i=1}^{N} r_{i,m} w_i, \tag{3.4}$$

and the updated Σ_m^{new}'s as

$$\Sigma_m^{new} = \frac{1}{N} \sum_{i=1}^{N} r_{i,m} (w_i - \mu_m^{new})(w_i - \mu_m^{new})^T, \tag{3.5}$$

with N being the number of measurements.

After the new parameters have been computed, the maximization step is complete, and we recompute the membership weights in the expectation step, and continue in this manner for a set number of iterations. One pair of expectation and maximization steps constitute one iteration of the algorithm.

3.2   Gaussian Mixture Models Illustration

In this section, we illustrate the GMM method by fitting an M-GMM with $M = 2, 3$ and 4 components to the noise component at the receiver output for the single pulse single span fiber communication system (as illustrated in Fig. 3.1) using the E-M algorithm. Each GMM component has two dimensions, one corresponding to the real part of the noise term, and the other to the imaginary part. The received signal is a combination of the original transmitted signal, the noise from the amplifier, and all the components resulting from the signal-signal and signal-noise interactions. We isolate the noise component (including interaction effects) by removing the original transmitted signal from the received signal. The simulation parameters used are indicated in Table 3.1.

Table 3.1: Simulation Parameters

Description	Symbol	Value
transmission wavelength		1550 nm
oversampling rate (at transmitter)	n_s	16
attenuation constant	α	0.2 dB/km
group-velocity dispersion parameter	β_2	-20 ps^2/km
nonlinear parameter	γ	2/(Wkm)
span length	L	100 km
number of spans	n	1
optical filter bandwidth		$2B/F$
receiver filter 3 dB bandwidth		$1/F_s$
receiver filter rolloff factor		0.5

We plot the data obtained (from split-step Fourier simulation of NLSE equation) for the noise component and superimpose the fitted GMM distributions. In Fig. 3.2, we present the contours of the fitted two or three component Gaussian models for observations obtained at $P = 0$ dBm input power level, and in Fig. 3.3, we present the corresponding contours for $P = 8$ dBm. At the higher power level, the nonlinear effects become more prominent, and this is observed from the scatter plot of the data. The system as a whole becomes more non-Gaussian, and as a result the dominating components have means further away from the origin and the covariance matrices become more non-diagonal, as shown in Figs. 3.4 and 3.5. The larger mean vectors indicate that with increasing nonlinearity, the noise components become more significant, which is to be expected, since the system is nonlinearity-dominated. Also, the real and imaginary parts of the components become more correlated, and an additive independent noise model can no longer be considered sufficient. Nonlinearity and the nonlinear signal-noise interactions increase with larger input power levels.
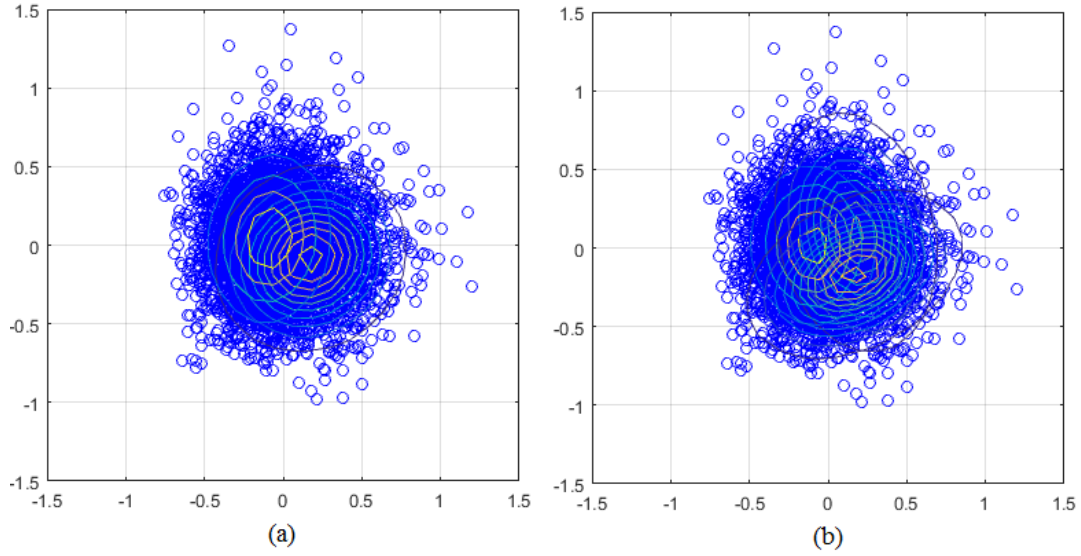
Figure 3.2: (a) 2-GMM and (b) 3-GMM fitted to simulation data at 0 dBm input power
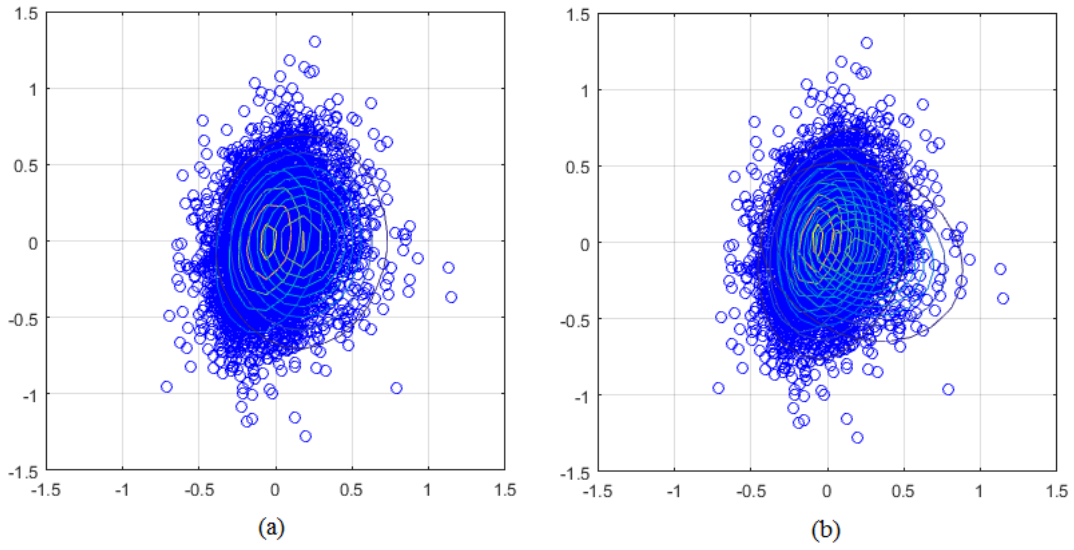

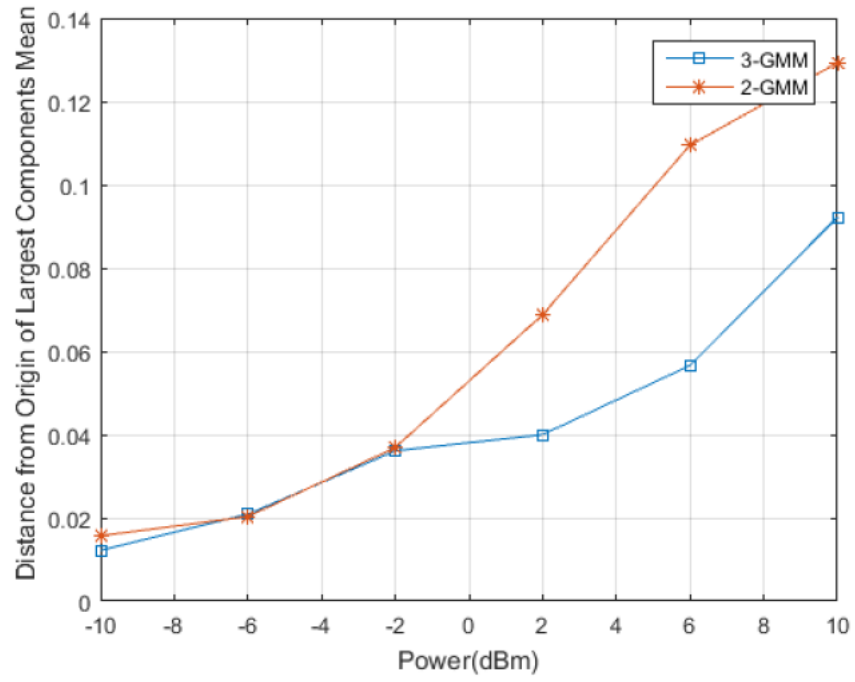Figure 3.3: (a) 2-GMM and (b) 3-GMM fitted to simulation data at 8 dBm input power

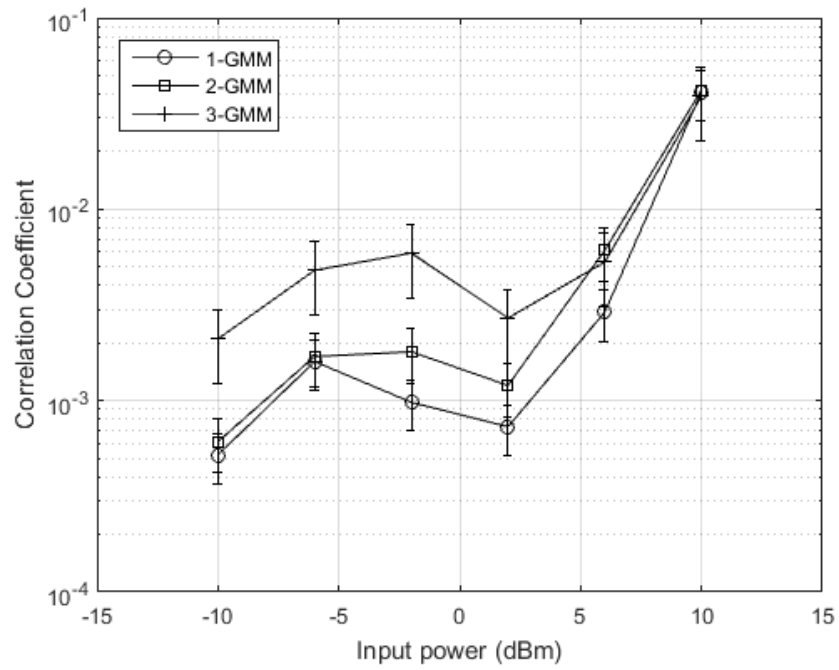Figure 3.4: Mean of dominant GMM component

Figure 3.5: Correlation coefficient of dominant GMM component (95% confidence intervals)

We investigate the weights of the largest component of the different GMM models in Fig. 3.6. We note that the more significant component in the 2-GMM model is always very dominant. However, there is a significant difference between the results obtained for the two ends of the power spectrum.

At a low input power level of $P = 0$ dBm, we note that the weight (proportion) of the less correlated component (the component with a mean closer to zero and a more diagonal covariance matrix) is much higher (0.9994 for a 2-GMM, and 0.8463 for a 3-GMM). At the higher input power level of $P = 8$ dBm when nonlinearity prevails, the corresponding numbers are much lower (0.0562 and 0.0016 respectively), and the correlated Gaussian components dominate. This is why the 2-GMM model fails to perform much better than the 1-GMM model, as we show in Section 3.4. A 3-GMM or a 4-GMM allows for more freedom and a better fit. However, we must also be careful to not overfit to the noise data [67]. We will find in Section 3.4 that a 3-GMM is ideal for our purposes.

Although this chapter is concerned with analyzing a single pulse transmission through a single SSMF span, we present some results which show the consistency of our technique. We know that the nonlinear effects in a fiber scales with the number of spans of the fiber, and also as the 1.5^{th} power of the input optical power [68]. In Table 3.2, we note the consistency of the covariance matrix of the largest 2-GMM component when the span length (L) or the nonlinearity (γ) or the input optical power (P) is increased as described above. This result is the average of five experimental runs.
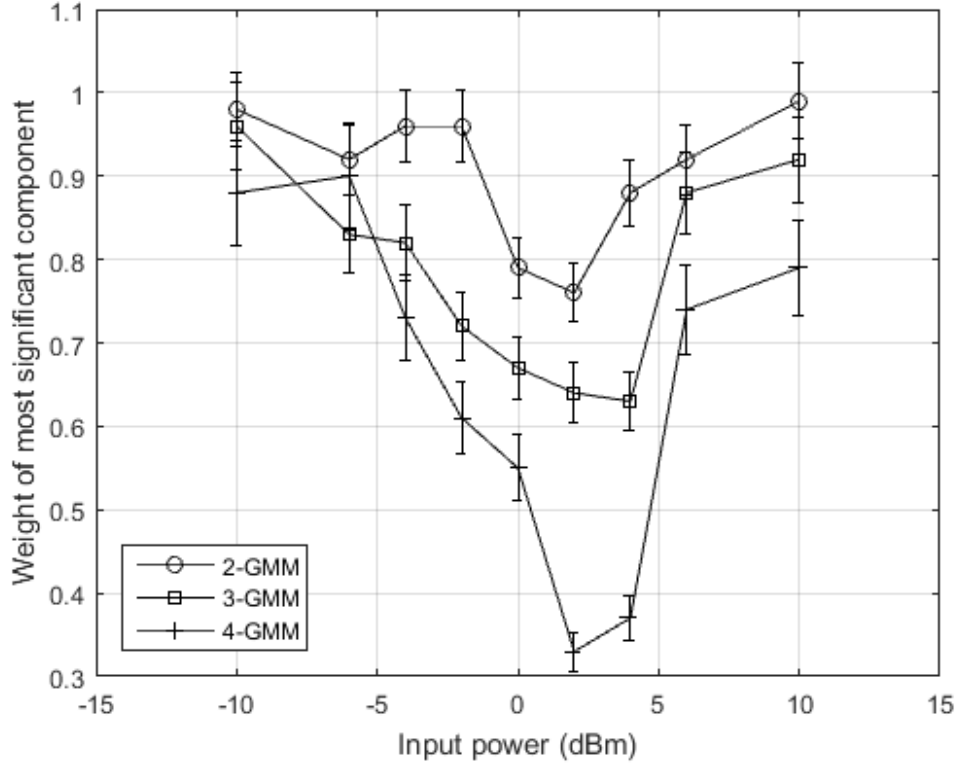
Figure 3.6: Weight of most significant GMM component for different power levels (95% confidence intervals)

3.3 Volterra Series Transfer Function

The Volterra series transfer function in the time or frequency domain is a nonlinear relationship between the input and the output of a system. Let us denote the input and output of the channel by x and y, respectively. Using the third-order Volterra model for the fiber (discarding higher-order kernels), the observed receiver output after sampling for a digital real-valued data x can be modeled as [69]

$$y \propto x + w + \rho_1(x^3 + 2wx^2 + w^*x^2 + 2x \mid w \mid^2 + xw^2 + w \mid w \mid^2), \qquad (3.6)$$

Table 3.2: Covariance Matrix of Most Significant 2-GMM Component

Conditions	Covariance Matrix
2 spans of $L = 80$ km, $\gamma = 2/(\text{Wkm})$, $P = 10$ dBm	$\begin{bmatrix} 0.0330 & 0.0405 \\ 0.0405 & 0.0888 \end{bmatrix}$
$L = 80$ km, $\gamma = 4/(\text{Wkm})$, $P = 10$ dBm	$\begin{bmatrix} 0.0340 & 0.0423 \\ 0.0423 & 0.0914 \end{bmatrix}$
$L = 80$ km, $\gamma = 2/(\text{Wkm})$, $P = 10^{1.5}$ dBm	$\begin{bmatrix} 0.0351 & 0.0444 \\ 0.0444 & 0.0945 \end{bmatrix}$

ρ_1 being a constant coefficient [13]. We can see that y consists of a signal component $x + \rho_1 x^3$ and a noise component, which includes the cross terms resulting from the interaction between complex noise and the transmitted signal.

Instead of numerically solving the nonlinear Schrödinger equation, the Volterra series transfer function (VSTF) computes an approximate analytic solution. Due to high computational complexity associated with integration, we usually truncate the VSTF to third-order. The first-order Volterra kernel represents the linear effect from fiber loss and dispersion, and the third-order kernel represents the nonlinear effect. The VSTF method is equivalent to the perturbation method [70].

At lower power levels, the nonlinearity in the fiber is low. So the higher-order terms in y in (3.6) are small. At a high signal to noise ratio (SNR), the terms containing w^2 are small, and thus an additive Gaussian noise model holds. If we assume a high SNR environment, the last three terms in (3.6) can be ignored. We contend that at higher power levels, using VSTF and GMM is more accurate than the additive Gaussian model in optical fiber communication systems.

Using the results in [71], for our single span single pulse set-up (if the pulse

shape is Gaussian), the received signal can be approximated as

$$
\begin{aligned}
y(t) = \quad & xP^{\frac{1}{2}}exp\left[-\frac{(t-T)^2}{2T^2}+j\Phi\right]+j\gamma xx^*xP^{\frac{3}{2}}exp[j\Phi]exp\left[-\frac{(t-T)^2}{6T^2}\right] \\
& \int_0^L \frac{exp(-\alpha z)}{K_1(z)}exp\left[\frac{-\frac{4(t-T)^2}{3}}{T^2(1+\frac{j3\beta_2 z}{T^2})}\right]dz + \widetilde{w}(t),
\end{aligned}
\tag{3.7}
$$

where

$$
K_1(z) = \sqrt{1+\frac{j2\beta_2 z}{T^2}+\frac{3\beta_2^2 z^2}{T^4}},
\tag{3.8}
$$

Φ is the phase of x and $\widetilde{w}(t)$ is the noise at the output of the fiber, which cannot be modeled using this theory. All other symbols have usual meanings.

3.4   Validation

In probability and information theory, the Kullback–Leibler divergence [72] is a measure of the difference between two probability distributions P and Q. It is not symmetric in P and Q. In applications, P typically represents the actual distribution of data, observations, or a precisely calculated theoretical distribution, while Q typically represents a theory, model, description, or approximation of P. The Kullback–Leibler divergence is defined as

$$
D_{KL}(P||Q) = \sum_{j=1}^{N} P_{\widetilde{W}}(\widetilde{w}_j)\log\frac{P_{\widetilde{W}}(\widetilde{w}_j)}{Q_{\widetilde{W}}(\widetilde{w}_j)}.
\tag{3.9}
$$

In Fig. 3.7, we plot the Kullback–Leibler divergences on 1-GMM (additive Gaussian), 2-GMM, 3-GMM and 4-GMM as compared to SSF simulation for different power levels. No restrictions are set on any of the Gaussian components.

The point of this exercise is to observe how closely our Gaussian component decomposition(s) Q matches the actual noise distribution P. We use the simulation data to obtain P. As indicated above, Q is the sum of the weighted GMM component distributions.

From Fig. 3.7, we find that at lower power levels (when nonlinearity is low), the match between the density estimated using simulated sample points and the Gaussian decomposition is better than at higher power levels. Also, there is no appreciable difference if more components are added. However, when there is more nonlinearity in the system and the noise distribution P_W as a whole is more non-additive Gaussian, the additional components can reduce the Kullback–Leibler divergence and provide a more accurate model of the noise.

We also note that there are only minor differences between the Gaussian model and the 2-GMM decomposition. In addition, 3-GMM and 4-GMM perform similar to each other. As seen above in Fig. 3.6, this is due to the significant domination of one component in the 2-GMM. This leads to the conclusion that for all practical power levels, using a 3-GMM with no restrictions on any components (means and covariance matrices of all components are chosen by the E-M algorithm) would be sufficient for analysis. Using a 4-GMM adds only a little performance gain, but at the cost of much higher computations. In fact, the performance benefits start to saturate after a 3-GMM, while computations keep increasing.

3.5 Mutual Information

In this section we show how the GMM model can be used to approximate the mutual information between the input to the fiber x and the sampled output y. We note the mutual information $I(X;Y)$ between two variables X and Y
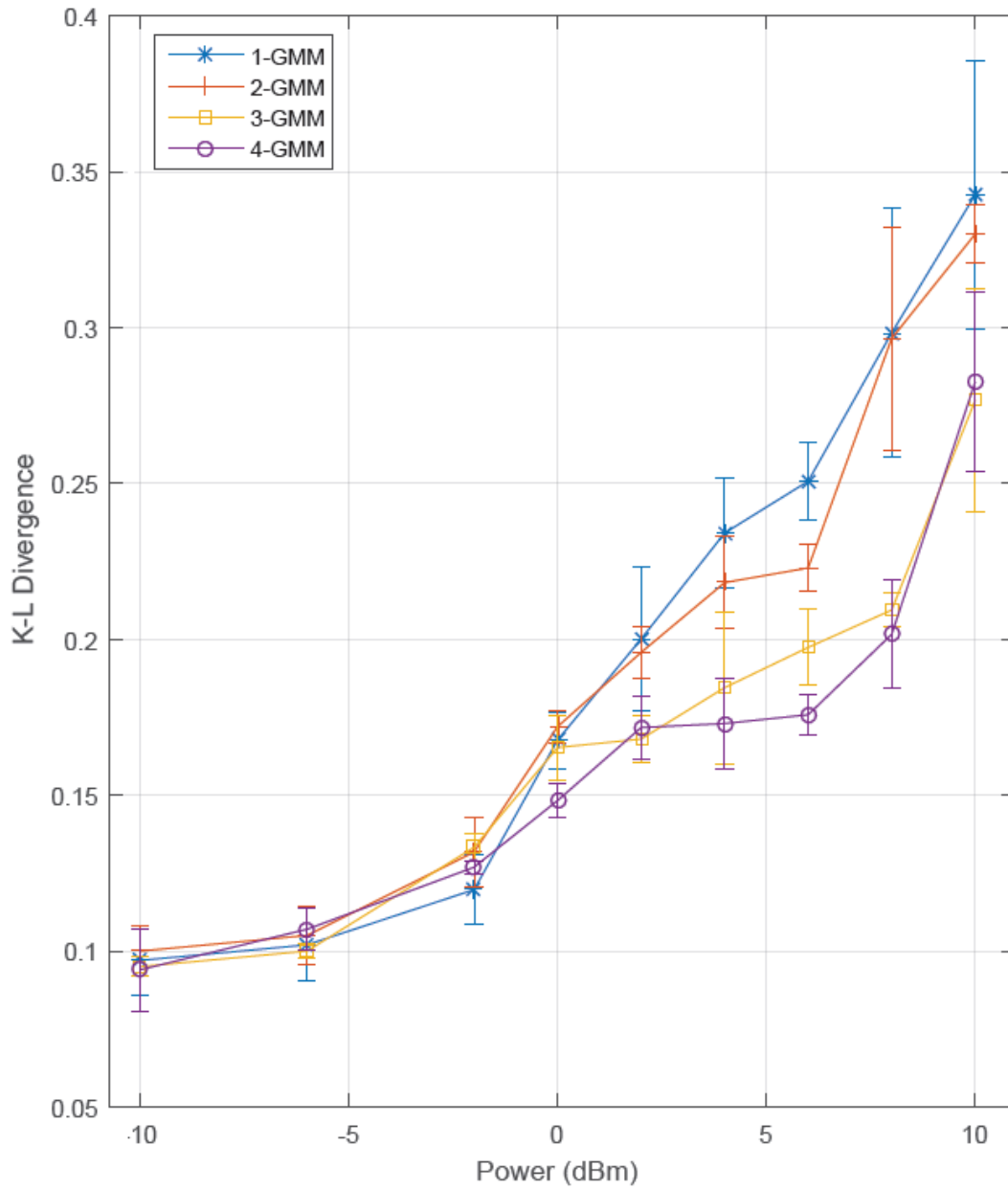
Figure 3.7: K-L divergences for different GMM models vs. input power (95% confidence intervals)

depends only on the probability and conditional probability of the input and output distributions [73].

The capacity of a channel is the maximum rate of communication for which arbitrarily small error probabilities can be achieved. The channel capacity is defined as

$$C = \sup_{P_X(x)} I(X;Y), \tag{3.10}$$

where the supremum is taken over all possible choices of $P_X(x)$, the input distribution of X. The capacity of a channel is the supremum over all *achievable rates.* In general, achievable rates— for conditions placed on $P_X(x)$— are smaller than the channel capacity.

For an ideal channel with additive white Gaussian noise (AWGN), Shannon's law defines the theoretical maximum rate at which error free information can be transmitted over the channel in the presence of noise [7]. The information capacity in bits per transmission can be calculated as

$$C = \frac{1}{2} \log \left(1 + SNR \right), \tag{3.11}$$

where SNR is the signal to noise ratio for the transmission. We note that the above expression does not hold true when the noise is non-Gaussian. As already discussed, while each noise source in our system satisfies AWGN requirements, due to the nonlinear interaction with signal, and itself (or other noise sources), the resultant noise at the receiver end is not Gaussian.

The mutual information between two discrete variables X and Y is defined as

$$I(X;Y) = \sum_{y} \sum_{x} P_{XY}(x,y) \log\left(\frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}\right),$$

(3.12)

where each summation is over the spaces of the two variables, P_{XY} is the joint probability distribution function of the random variables X and Y having marginal distributions P_X and P_Y respectively.

We note that (3.12) can be re-written as an expectation of the log entity. The expectation operation can then be replaced by the sample mean when we use the law of large numbers. We can replace the probability density functions with our GMM estimates to yield the mutual information

$$I(X;Y) \approx \frac{1}{N} \sum_{n=1}^{N} \log\left(\frac{P_{GMM}(x_n, y_n)}{P_X(x)P_{GMM}(y_n)}\right),$$

(3.13)

with N being the number of data samples. P_{GMM} is the weighted sum of the mixture component densities (3.1).

The mutual information between two Gaussian variables is given by

$$I(X;Y) = \frac{1}{2} \log \frac{|\Sigma_{xx}||\Sigma_{yy}|}{|\Sigma|},$$

(3.14)

where Σ is the joint covariance matrix expressed as $\begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$, and $|.|$ denotes the determinant.

For the VSTF method, we use (3.6) and (3.14) to compute the mutual information. The noise at the input $w(t)$ is a zero mean Gaussian random variable,

and

$$\Sigma_{xx} = E((X - E(X))(X - E(X))'), \qquad (3.15)$$

$$\Sigma_{yy} = E((Y - E(Y))(Y - E(Y))'), \qquad (3.16)$$

$$\Sigma_{xy} = E((X - E(X))(Y - E(Y))'), \qquad (3.17)$$

$$\Sigma_{xy} = \Sigma_{yx}, \qquad (3.18)$$

where $E(.)$ denotes the expectation of a random variable.

We present the covariance matrices obtained from the 3-GMM and the VSTF models for comparison. The covariance matrices for largest 3-GMM components at 0 and 10 dBm input power (averaged over 10 trials) are respectively $\begin{bmatrix} 0.0773 & 0.0075 \\ 0.0075 & 0.0730 \end{bmatrix}$ and $\begin{bmatrix} 0.0320 & 0.0388 \\ 0.0388 & 0.0864 \end{bmatrix}$, and the corresponding VSTF matrices are respectively $\begin{bmatrix} 0.0713 & -0.0021 \\ -0.0021 & 0.0860 \end{bmatrix}$ and $\begin{bmatrix} 0.0348 & 0.0413 \\ 0.0413 & 0.0914 \end{bmatrix}$. We conclude that there is a close match between the results obtained using the two tools. We note that the VSTF is not energy preserving, so it is not unexpected that the variance and covariance values are larger than simulation.

Next, we investigate the achievable rate of the network for the Gaussian, VSTF and 3-GMM noise models. We calculate the mutual information and find its upper limit at the different power levels to obtain the achievable rate with a constellation size of 256. We start with random constellations (of the same size 256) and with each constellation point having a random probability (which together add up to unity). We use the closed-form densities provided by our two models: the correlated Gaussian model for the VSTF, and the M-GMM model. Trying to achieve this

using SSF simulation would involve significant amounts of computational effort.

Depending on the noise model, each constellation yields a mutual information value. The upper limit boundary of the mutual information values is traced and this yields the achievable rate. The results for the three different noise models (additive Gaussian, VSTF and 3-GMM) are presented below in Fig. 3.8. We see that at lower power levels, as expected, the independent Gaussian model performs well, and approaches (3.11). However, the independent Gaussian model cannot model the effects from the non-additive terms at higher power levels, and we need to employ more advanced techniques, such as 3-GMM, to effectively understand it. The VSTF model appears to overestimate capacity over the entire range of power levels. At the higher power levels, the GMM model is more accurate.
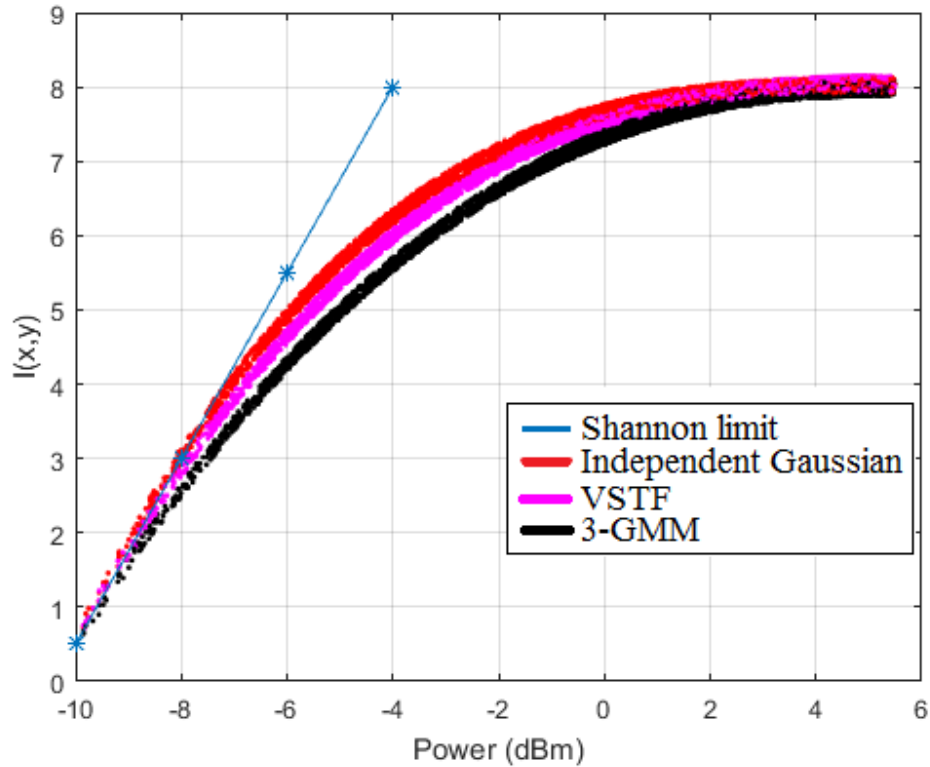

Figure 3.8: Mutual information for varying input powers for different techniques

We plot example input constellations which result in the highest mutual information at 10 dBm input power in Figs. 3.9. We see that the Gaussian and 3-GMM models choose different looking constellations that optimize the mutual information. The 3-GMM model avoids large power constellation points, thereby reducing the effective nonlinear nature at this high power level. The diameter of each circular symbol denotes the probability of the corresponding constellation point.
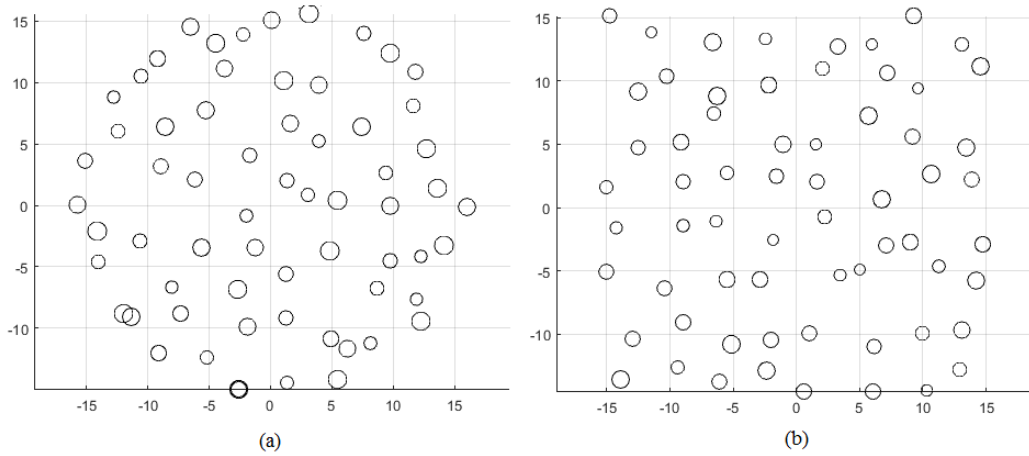

Figure 3.9: (a) 3-GMM and (b) Gaussian constellations (256 constellation points) for highest mutual information at 10 dBm average input power

We describe the power distributions of the optimum constellations for the different models, and note the different factors affecting the distributions. Each constellation point in Fig. 3.9 has a power and a probability associated with it. We use a kernel-based density estimation technique to draw the power density. For each such point, we draw a Gaussian kernel— whose height is determined by the corresponding probability— around the power level. We sum all the kernels (the number of kernels is the same as the number of constellation points) to obtain the power distribution estimation. In Fig. 3.10, we present the power distribution

for the optimum 3-GMM constellation with 256 points at four times the nominal nonlinearity level ($\gamma = 8/(\text{Wkm})$), and input power $P = 10$ mW. We include the average of 20 trial runs, and also superimpose one single sample run result.
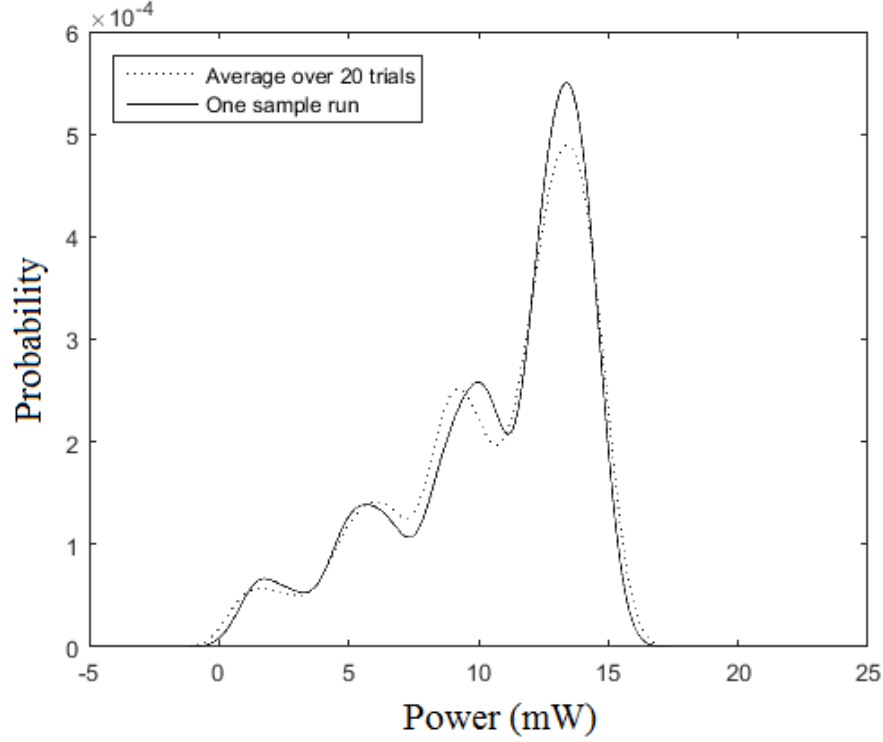

Figure 3.10: Distribution of constellation (256 points) power for 3-GMM at 10 mW average input power ($\gamma = 8/(\text{Wkm})$)

We note that this distribution is dependent on the constellation size for both 3-GMM and Gaussian models and present related results in Fig. 3.11 for four times the nominal nonlinearity level ($\gamma = 8/(\text{Wkm})$), and input power $P = 10$ mW. The peaks become sharper with increasing number of constellation points, and at this combination of high nonlinearity level, high input power and large number of constellation points, we can see four distinct peaks for both 3-GMM and Gaussian models.

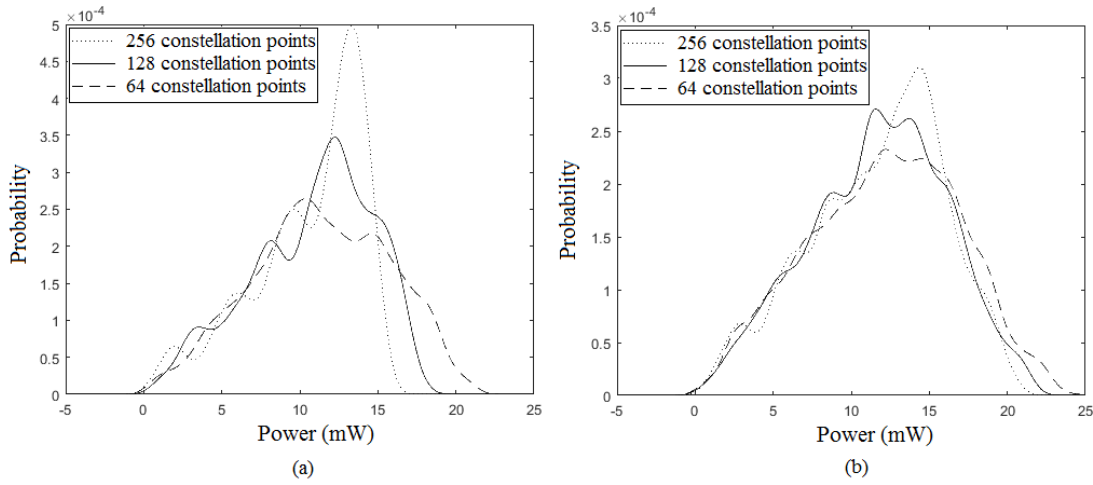Compared to the Gaussian noise model, the 3-GMM noise model does not go

Figure 3.11: (a) 3-GMM and (b) Gaussian constellation power distributions $(64, 128, 256$ constellation points) for highest mutual information at 10 mW average input power $(\gamma = 8/(\text{Wkm}))$

to as high power levels. This is reasonable since at the higher power levels, the noise is not Gaussian at all, as we noted. Therefore, the GMM power distribution has a much higher central tendency compared to the Gaussian model distribution.

Kurtosis is a measure of the tail-heaviness of a distribution, and is calculated at the ratio of its mean to its standard deviation raised to the fourth power. We present kurtosis results in Table 3.3 for both 3-GMM and Gaussian model for the different number of constellation points to illustrate the dependence of the power distribution.

Table 3.3: Kurtosis of 3-GMM and Gaussian Constellations as a Function of the Constellation Size

Constellation Size	3-GMM	Gaussian
64	0.0781	0.0572
128	0.2397	0.1175
256	0.3536	0.1319

Lastly, we investigate how changing the nonlinearity level affects the power distribution. The results are presented in Fig. 3.12 for two different levels
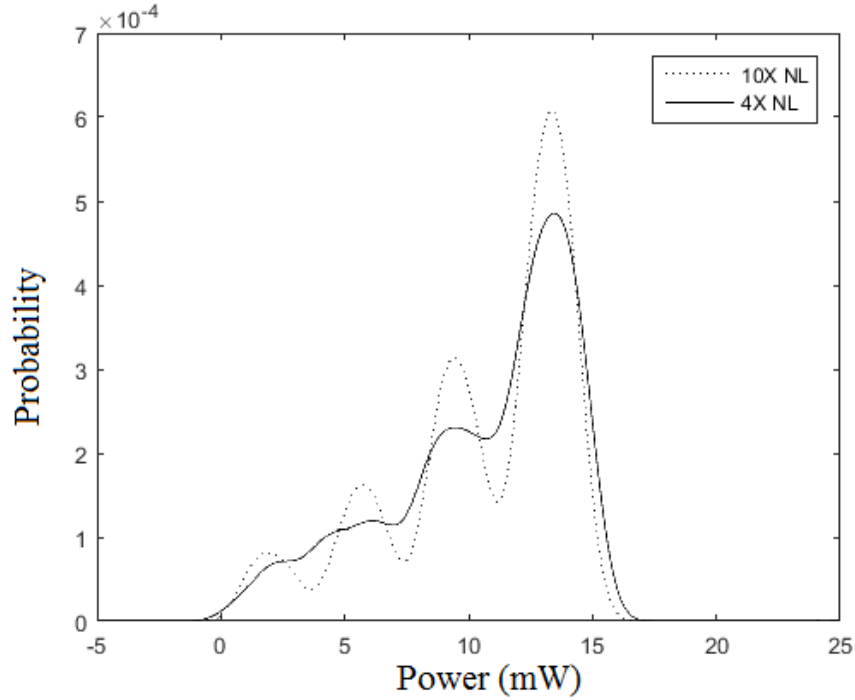
Figure 3.12: Distribution of constellation (256 points) power for 3-GMM at 10 mW average input power ($\gamma = 8/(\text{Wkm})$, $20/(\text{Wkm})$)

of nonlinearity and for 256 constellation points. We find that the higher the nonlinearity, the sharper the peaks are (in general).

3.6   Chapter Summary

We present a novel method to decompose the receiver end noise of an optical fiber communication system into Gaussian components using GMM decomposition. We propose that understanding the noise-nonlinearity-signal interaction on a single optical pulse propagating through a single span of optical fiber is a fundamental step in understanding and combating the effects of ASE. We propose and validate the use of GMM as a tool to understand the various single pulse and noise interaction effects in the simplest optical fiber communication channel. We explain how GMM

and VSTF techniques can be applied to the present problem, and validate them. For lower levels of nonlinearity, the Gaussian model is satisfactory. However, as nonlinearity increases, we need to make use our GMM model. We also find that 3-GMM model avoids large power constellation points, and that the resulting power distribution concentrates more power in the higher end of the power domain. In the next chapter, we extend this work to a multi-pulse transmission system.

Chapter 4

Multi-pulse Transmission with Gaussian Mixture Models

In the last chapter, we use Gaussian mixture modeling to study the nonlinear interaction effects between a signal pulse and a single noise source being transmitted over a single span of SSMF. This chapter furthers that knowledge by generalizing the scenario. We transmit multiple signal pulses, sometimes dissimilar, and find a setup through which we can predict the performance of a practical communication system. Once again, we use Gaussian mixture modeling to decompose the receiver end signal-noise interaction outputs into individual components and study them.

We already investigated the simplest non-trivial scenario possible: the transmission of a single pulse through a single span of standard single mode optical fiber (SSMF), the single signal pulse being affected by a single ASE noise source. This is the most fundamental situation that can be imagined. In this chapter, we add interference effects from other adjacent pulses, and the noise affecting those. We still have one span of fiber only, with no optical dispersion compensation. The propagating signal pulses are affected by themselves, each other, and the noise

source.

Our communication system consists of a transmitter with a pulse-shaping filter and a noise source which adds additive white Gaussian noise (AWGN) to the signals pulses which are transmitted through a single fiber span. We note that the noise source is before the optical fiber span so that the noise can interact with itself and the transmitted signal.

For comparison (including comparison with the single pulse case), we also consider a scenario with digital backpropagation to eliminate all signal-signal interactions. The interactions between the signal pulses were not present when a single pulse transmission was studied. When backpropagation is employed, we have the noise source after the optical fiber span (otherwise the backpropagation would precisely undo the impairment effects leaving us with just the additive noise). In effect, this study generalizes the approach we have already used.

Importance sampling is a technique that allows us to obtain some samples from an interesting or important region of a distribution. In many applications, we want to learn some measure of a distribution $P_X(x)$ where $P_X(x)$ is nearly zero outside a particular region. That is, the probability of the random variable assuming a value in the region of interest is minute. A Monte-Carlo sample from the distribution of X could fail to have any significant number of points inside the region of interest. We solve this problem by sampling from a distribution which overweights the important region [74]. For importance sampling, the sample distribution P_X^* that minimizes the variance of the sample is often chosen. It can be shown that the distribution that minimizes the above variance is [75]

$$P_X^*(x) = \frac{|X|P_X(x)}{\int |x|P_X(x)dx}. \tag{4.1}$$

We use this tool in combination with GMM to calculate the bit error rate.

Kernel density estimation (KDE) is a statistical tool to estimate the probability density function of a random variable. A histogram is not smooth, since it depends on the width of the bins and the end points of the bins. Kernel density estimation solves these problems. Kernel estimators center a kernel function at each data point. We can obtain a smooth density estimate using a smooth kernel function as our building blocks. Kernel estimators smooth out the contribution of each observed data point over a local neighborhood of that data point [76]. We use kernel density estimation after importance sampling to validate our GMM decomposition.

This chapter is organized as follows. In Section 4.1, we study the Gaussian decomposition when two adjacent pulses are transmitted, including the scenario when the twin pulses are not adjacent any more. Section 4.2 describes how the signal pulses affect each other, when the pulses are dissimilar. We also increase the number of pulses (from two) and generalize the analysis. In Section 4.3, we validate our GMM decomposition of nonlinear signal-noise interactions in a multiple pulse transmission. Sections 4.4 and 4.5 respectively presents the theory and the results of the bit error rate calculation in such a transmission. Section 4.6 summarizes the chapter and discusses possible future research directions.

4.1 Transmission of Two Pulses

We study the Gaussian decomposition when two adjacent pulses are transmitted, including the scenario when the twin pulses are not adjacent any more. The goal is to study the effect of one pulse on the other, and also on itself. We concentrate our analysis on QPSK modulation, and transmit two identical pulses $1 + j$. We remove the original signal(s) to isolate the noise and interaction effects. We decompose

the isolated noise, noise-signal and signal-signal interaction terms into two or three Gaussian components using 2- or 3-GMM techniques. We perform 2D 2-or 3-GMM on the first pulse of the complex signal unless noted otherwise. (When digital backpropagation is employed, it removes the signal-signal interaction.)

Digital backpropagation is a technique used to jointly compensate linear and nonlinear impairments. The basic principle is to solve an inverse NLSE through the fiber to estimate the input signal. In the absence of stochastic noise, backpropagation can reconstruct the originally transmitted sequence with perfect accuracy.

We note that digital backpropagation eliminates the effect of other pulses leaving us with only the signal-noise interaction. The set-up using backpropagation is illustrated in Fig. 4.1. In this set-up, we place the noise source after the fiber span. The reasoning behind this is: backpropagation undoes the effects of all physical impairments. If we had placed the noise source before the fiber span, after backpropagation, we would be left with just the additive noise.

Let f denote the effect of the fiber (physical impairments). Then, for the set-up in Fig. 4.1, $y(t) = f^{-1}(f(x(t)) + w(t))$.

After filtering and sampling, we assume that the output variable can be written as $y = \widetilde{x} + \widetilde{w}$, and study the statistics of $\widetilde{w}$. The signal term $\widetilde{x}$ is deterministic.
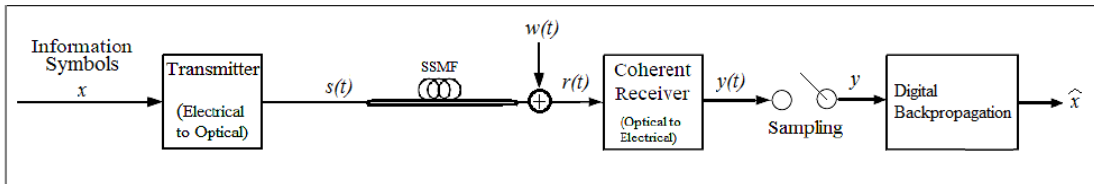

Figure 4.1: Adjacent pulse transmission over a single span with backpropagation

In Figs. 4.2 – 4.3, we plot the weights of the most significant mixture component

for both 2- and 3-GMM over a range of input power including the confidence intervals when we transmit two adjacent pulses. In Chapter 3, we provide similar results for a single pulse transmission scenario. At the lowest power levels, we would expect the weight of the most significant GMM component to be very close to unity from the knowledge that Gaussian noise dominates in this region. However, this is not seen in Fig. 4.2. We contend this is due to the additional effect of the adjacent pulse. To confirm this, we employ the backpropagation method.

When digital backpropagation is not employed, at a lower (more linear) input power level, the largest component has a mean vector closer to the origin. The covariance matrix of the largest component shows that the effect of one pulse on the other is also low, compared to the situation when nonlinearity is significant. When digital backpropagation is used to reduce the signal-signal interaction, this effect is nullified (at all input power levels). As a result, we have a situation where the most significant Gaussian mixture component has a less correlated covariance matrix. This is expected because we remove all interaction effects from the adjacent signal pulse.
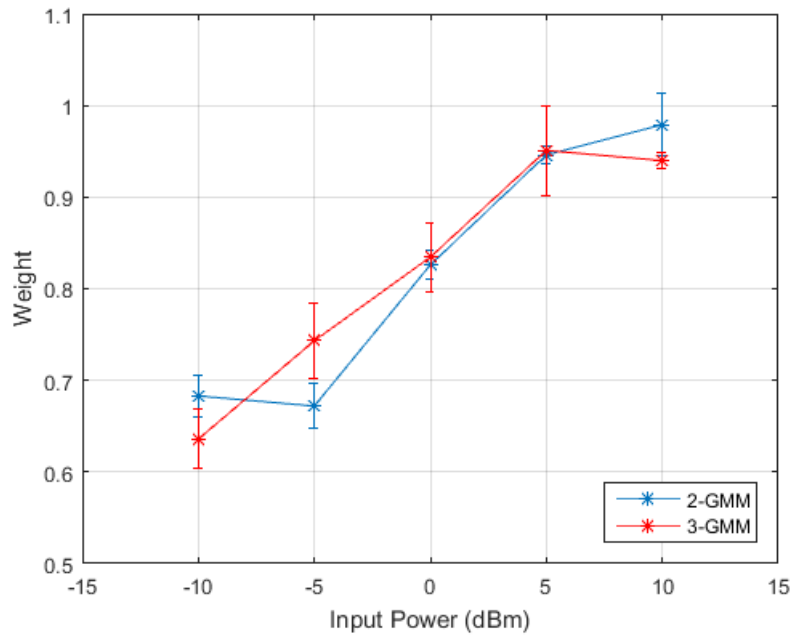
Figure 4.2: Most significant GMM component weights for different power levels without backpropagation (95% confidence intervals)
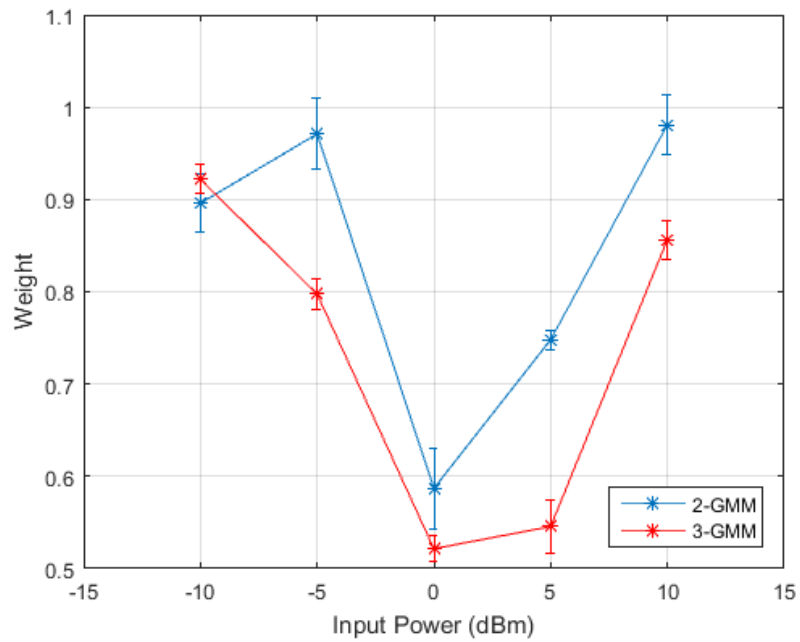

Figure 4.3: Most significant GMM component weights for different power levels with backpropagation (95% confidence intervals)

Next, we study how one pulse affects itself as it is transmitted in the presence of the other (adjacent) pulse. However, first we need to isolate one pulse's effect on itself. This is done by inspecting the corresponding elements in the covariance matrix. Once again, we transmit two identical pulses $1 + j$, and remove the original signal(s) to isolate the noise and interaction effects. This we decompose into two or three Gaussian components using 2- or 3-GMM techniques. We employ the backpropagation method to remove the signal-signal interaction and study each effect separately. The reader should note that now the analysis is on a four-dimensional vector, and the GMM components are 4-dimensional, the real and imaginary components of each pulse.

We present the results in Figs. 4.4 –4.5. We see that as nonlinearity increases, the sum of variances of the real and imaginary components increases. The effects are similar for either pulse, as should be expected, since we are transmitting identical pulses.

When the adjacent pulses are not identical, however, the results are slightly different. In a QPSK modulation scheme, a $-1 + j$ and a $1 - j$ pulse would affect a $1 + j$ pulse similarly (these are mathematically adjacent). The QPSK constellation diagram is presented in Fig. 4.6.
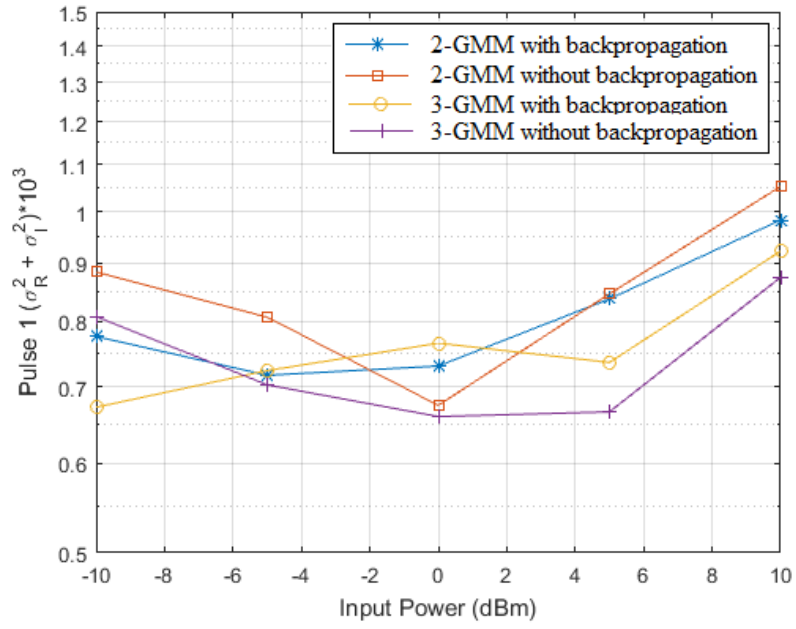
Figure 4.4: Effect of pulse 1 on itself as a function of input power
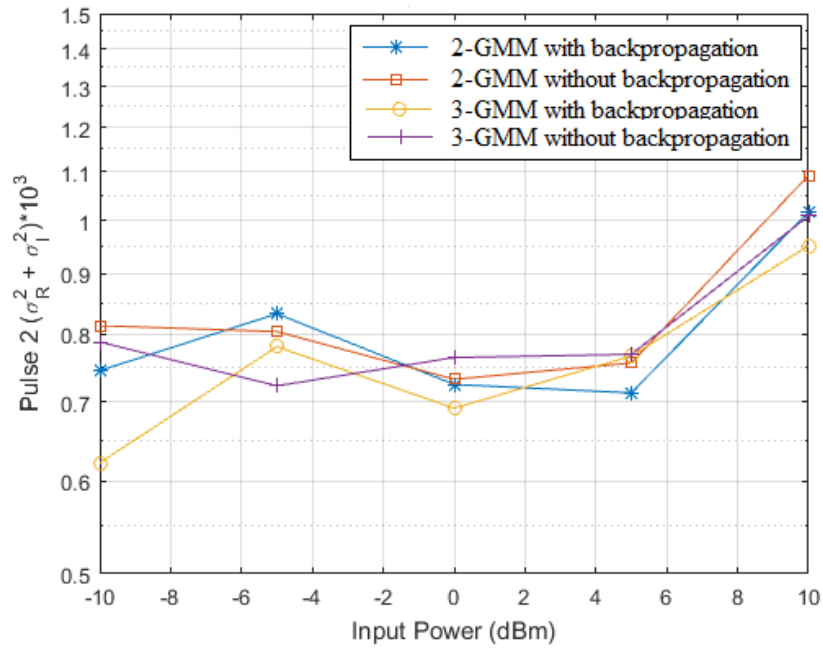

Figure 4.5: Effect of pulse 2 on itself as a function of input power

Figure 4.6: QPSK constellation

We present results from four representative scenarios in Figs. 4.7 – 4.8. These correspond to Figs. 4.4 – 4.5. We conduct four separate experiments to obtain these results, and the two adjacent pulses transmitted are $(-1-j), (-1-j)$ or $(1+j), (-1+j)$ or $(1+j), (-1-j)$ or $(1+j), (1-j)$. Once again, we see that as nonlinearity increases, the sum of variances of the real and imaginary components increases. There is no statistically significant difference in the noise level based on the modulation value of the adjacent pulse.

Figure 4.7: Dissimilar pulses: effect of pulse 1 on itself as a function of input power

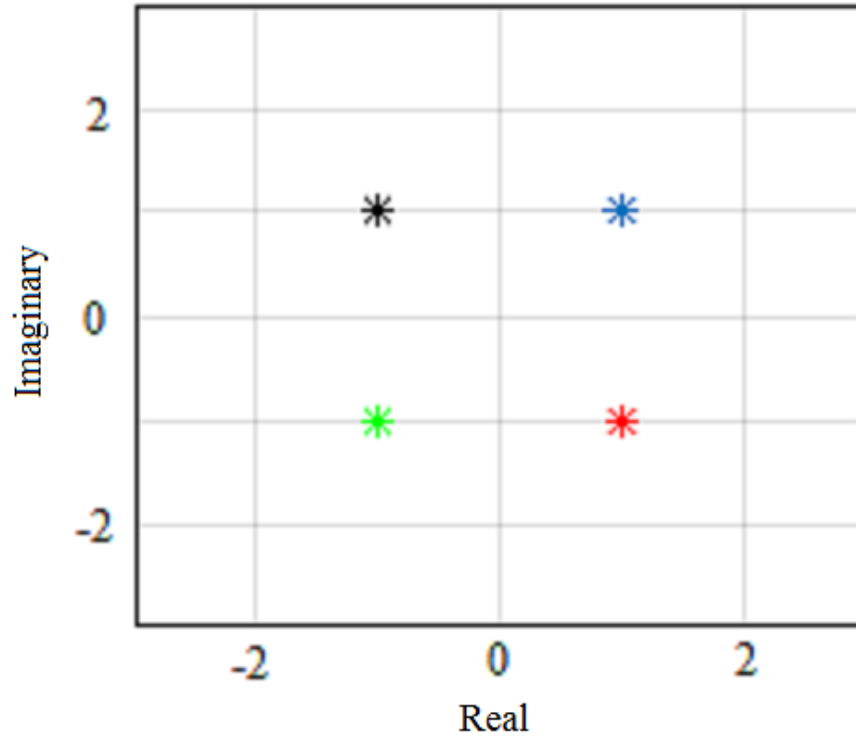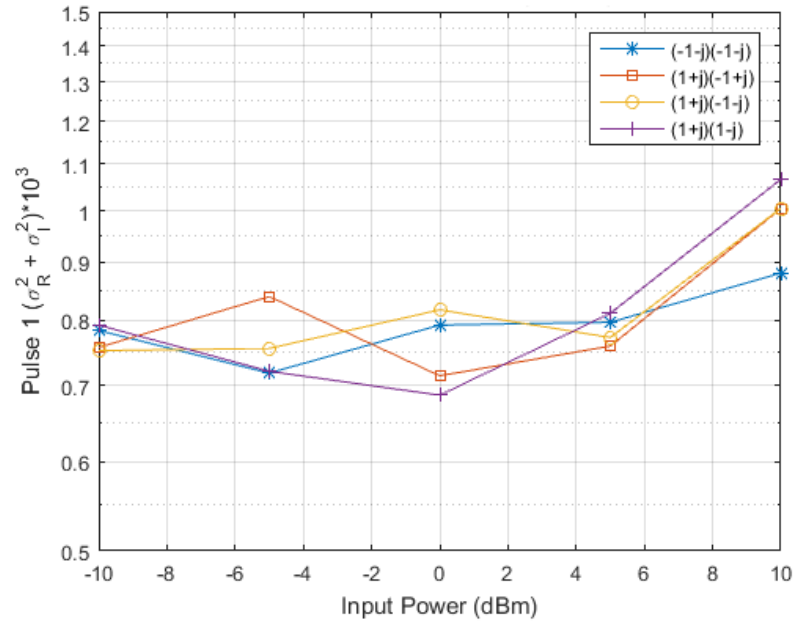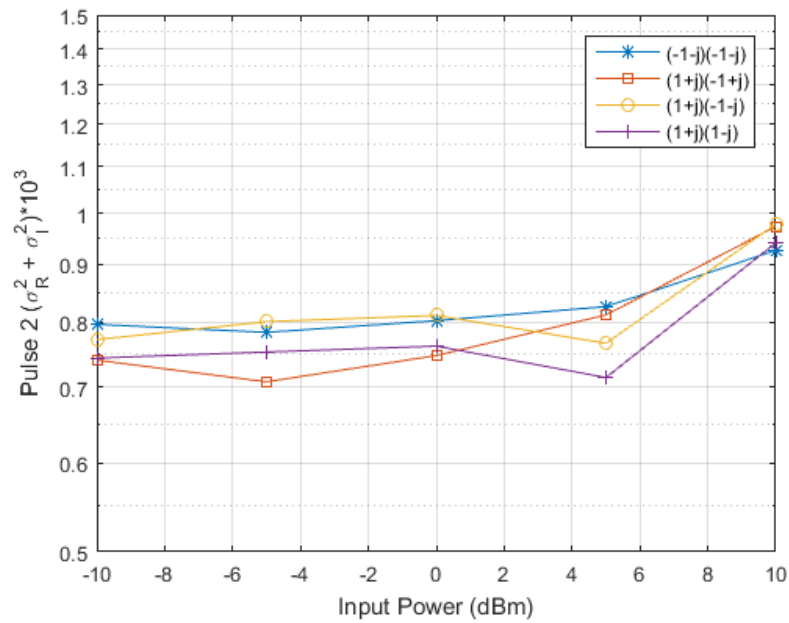Figure 4.8: Dissimilar pulses: effect of pulse 2 on itself as a function of input power
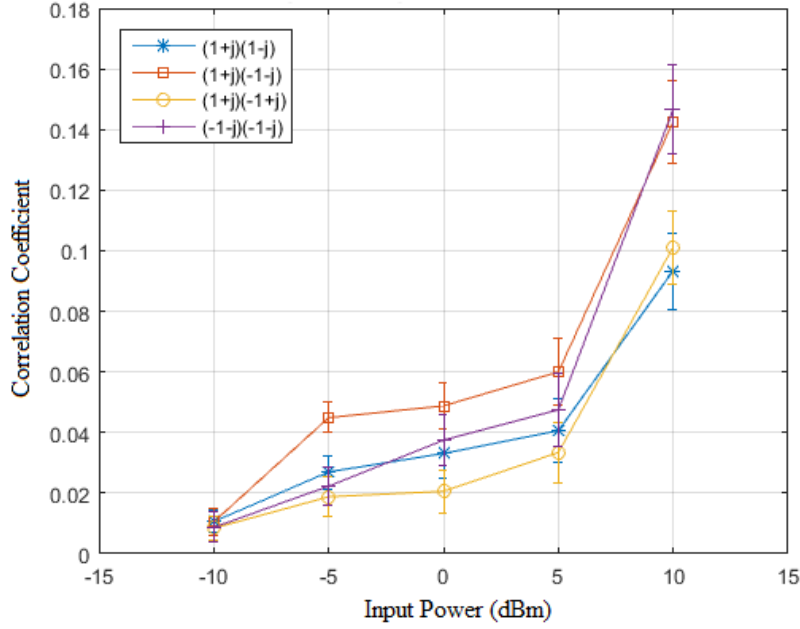
Figure 4.9: Correlation coefficient as a function of input power (dissimilar pulses)

We also analyze how the covariance and the cross-correlation of the most significant GMM component vary as a function of input power. As nonlinearity is increased, both parameters increase as seen in Figs. 4.9 – 4.10. As before, we conduct four separate experiments to obtain these results, and the two adjacent pulses transmitted are $(-1-j),(-1-j)$ or $(1+j),(-1+j)$ or $(1+j),(-1-j)$ or $(1+j),(1-j)$. This is further validation of results presented above. We note that the correlation and the cross-correlation coefficients are much higher (in both cases) when the two pulses are $(1+j),(-1-j)$. This probably happens because the $(-1-j)$ constellation point is furthest away from $(1+j)$, which causes the largest change in the field amplitude. The nonlinear effect is strongest when there is a large change in instantaneous power.

We also experimentally increase the distance between the twin pulses (at 10 dBm input power) and analyze the effect this has on the interaction. In Fig. 4.11,
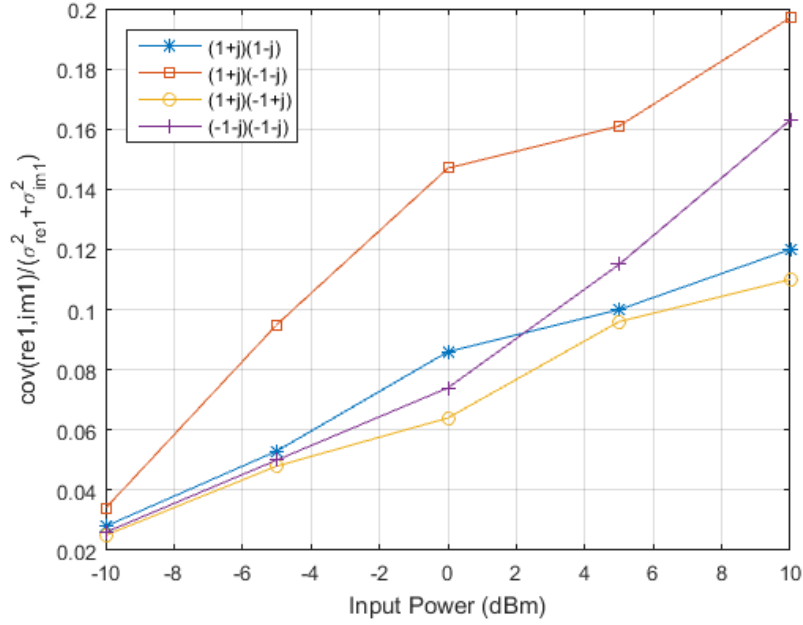
Figure 4.10: Cross-correlation coefficient (normalized to variance) as a function of input power (dissimilar pulses)

we plot the cross-correlation coefficient (normalized to variance) as the distance between the two identical $(1 + j)$ pulses are increased. As expected, the further the pulses are moved from each other, the less effect they have on each other. For comparison, we include the cross-correlation coefficient for a single pulse. As the distance between the pulses is increased, we approach this limiting value.

4.2 Generalization

In this section we generalize our setup by increasing the number of pulses and note the resulting differences. In Fig. 4.12, we plot the cross-correlation coefficient (normalized to variance) as a function of the number of interacting pulses for different input power levels. We perform a 2D 3-GMM of the center pulse are backpropagation to do this analysis. Our aim is to predict what might
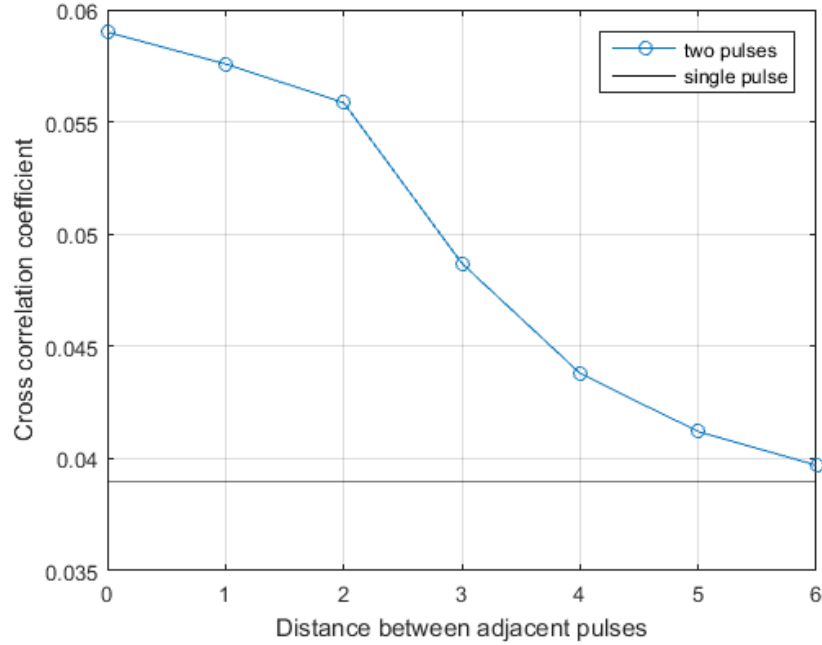
Figure 4.11: Cross-correlation coefficient as a function of distance between pulses (10 dBm input power)

happen when we transmit an infinite number of pulses, as would be done in a real transmission. The cross-correlation is seen to increase both with increasing power, and increasing number of pulses. This happens because we have more nonlinear interactions at a higher input power, and also with a larger number of pulses. However we note that the interaction effects also begin to saturate after about 7 or 9 adjacent pulses depending on the power level.

This fact is corroborated when we plot the correlation coefficient between several interacting pulses (drawn from QPSK constellation) at two different input power levels in Fig. 4.13 at an increased (twofold) level of nonlinearity. The correlation coefficient increases with higher number of pulses due to more interactions, however it also starts to saturate out after a certain number of pulses. At 10 dBm input power, when the nonlinearity is much higher (compared to 0 dBm), there is more
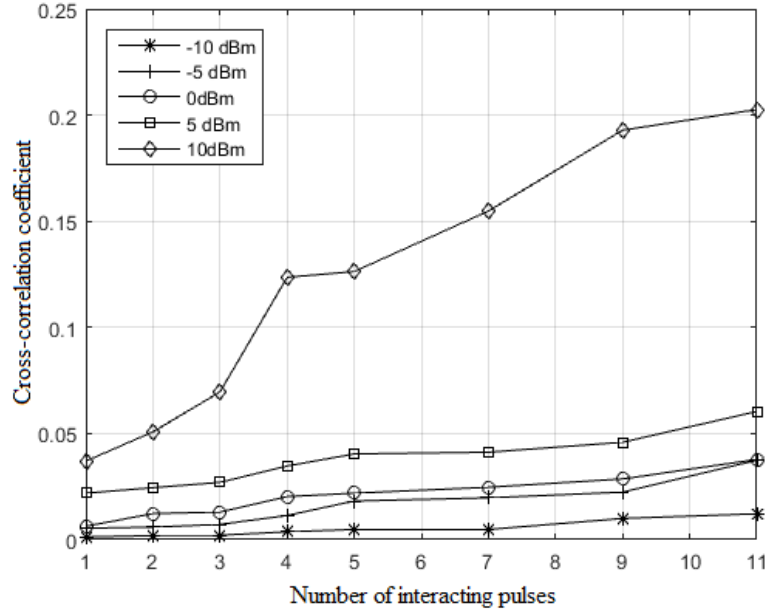
Figure 4.12: Cross-correlation coefficient for different number of pulses at different input power levels

nonlinear interaction, leading to the saturation for a larger number of interacting pulses.

4.3 Validation

We validate our GMM decomposition by calculating the K-L divergences with respect to the distribution of kernel densities. A kernel density estimator smooths out the contribution of each observed data point over a local neighborhood of that data point [77] and can estimate the probability distribution of any continuous random variable. The data points are obtained from importance sampling.

In importance sampling method (IS), the statistics of the noise in the system are biased to increase the probability of bit errors. The samples are generated from a biased density function instead of the original distribution. The sample
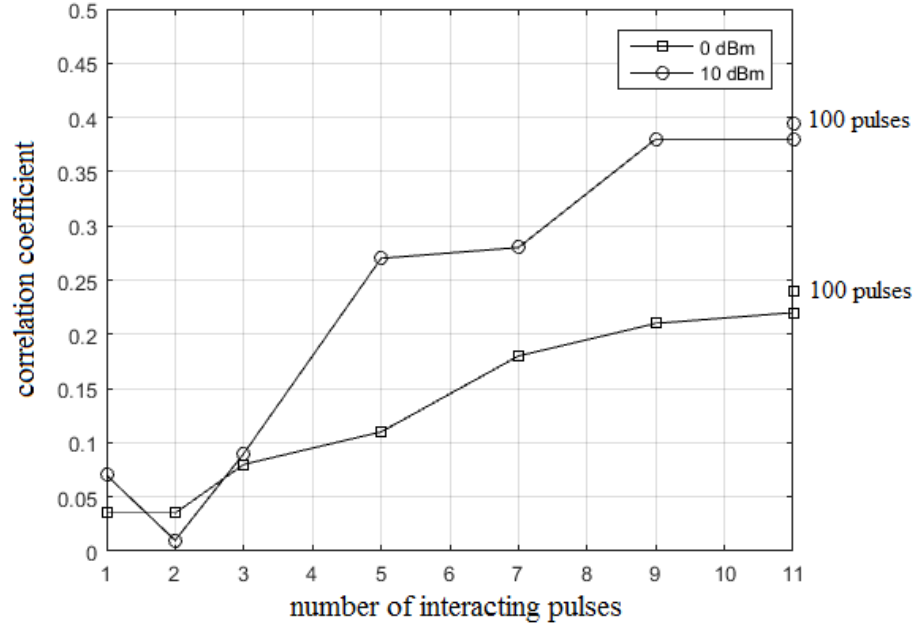
Figure 4.13: Correlation coefficient of most significant component for different number of pulses ($\gamma = 4/(\text{Wkm})$)

distribution is a Gaussian with the mean shifted to the decision threshold.

In Figs. 4.14 – 4.16, we compare between several different Gaussian decompositions, namely the AWGN model, and 1-, 3- and 5-GMM at three different input power levels (-10 dBm, 0 dBm and 10 dBm respectively) for different number of adjacent pulses transmitted (after performing digital backpropagation). We vary the input power to learn the relative effect of increasing nonlinearity. We also include a curve for the no-nonlinearity case (zero nonlinearity) to compare.

When the nonlinearity is low (at -10 dBm), all models perform similarly, and close to the zero nonlinearity scenario. Also, the K-L divergences are low, and vary little as the number of pulses are increased. We conclude this is due to little nonlinear interaction between the pulses.

As we increase the input power level, we can make several observations. First,

the AWGN model performs the worst. However the 1-GMM model is not much better either. From the results of our single pulse analysis with 2-GMM, we know the reason: one of the two Gaussian components significantly dominates the other.

The 5-GMM model performs better than the 3-GMM, but not significantly. We conclude that increasing the number of components in our GMM (beyond the 3-GMM) may promise only diminishing returns, while increasing computational complexity.

Also, the K-L divergences get worse (increase) both with the increase in nonlinearity and an increasing number of pulses. Once again, this is due to the extra nonlinear interactions.
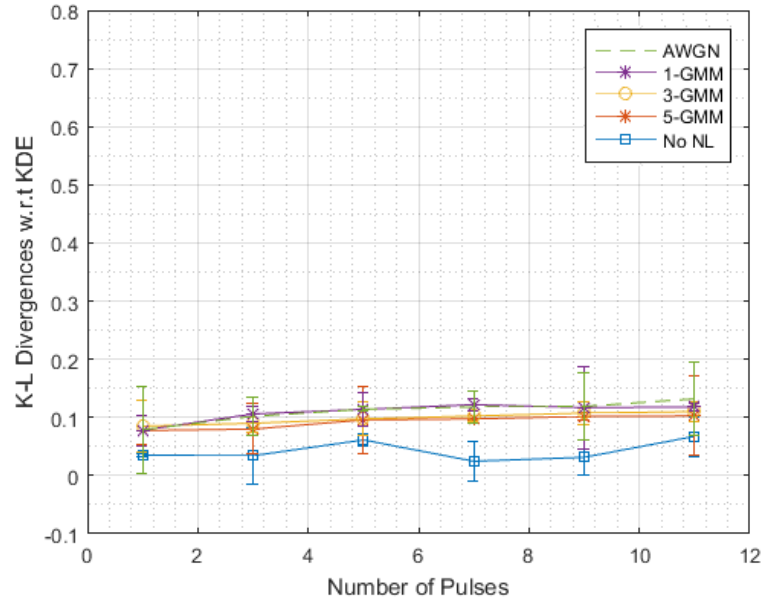

Figure 4.14: K-L divergence for KDE vs GMM at -10 dBm input power (95% confidence intervals)
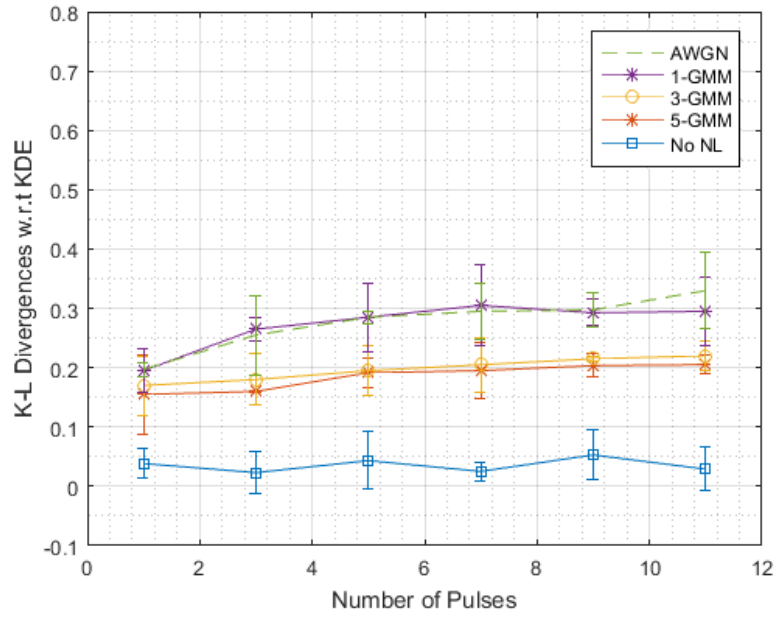
Figure 4.15: K-L divergence for KDE vs GMM at 0 dBm input power (95% confidence intervals)
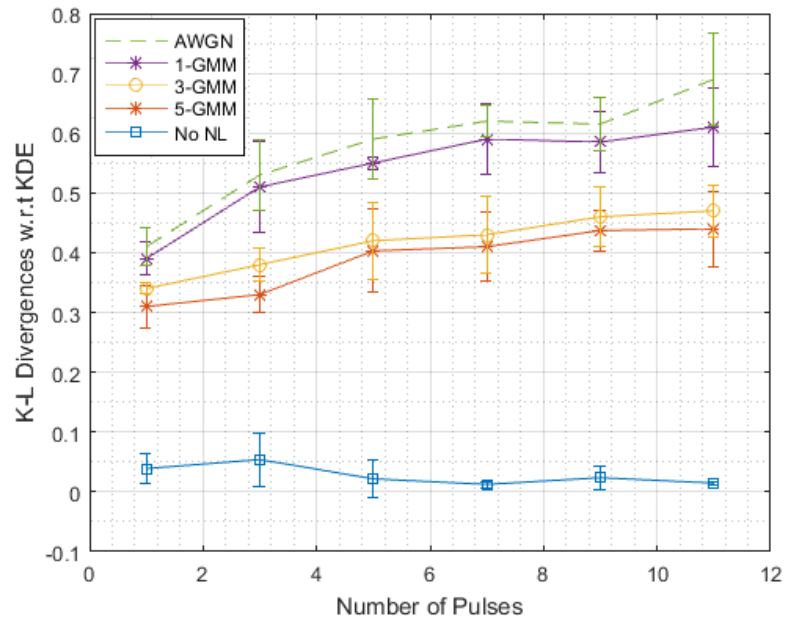
Figure 4.16: K-L divergence for KDE vs GMM at 10 dBm input power (95% confidence intervals)

4.4 Bit Error Rate Calculation

In communication systems, the performance indicator of the quality of the transmission is the bit error rate (BER). The BER represents the probability of receiving an erroneous bit. Thus, BER is an end-to-end performance measurement and can be used to quantify the reliability of the communication system.

A decision device is assumed to make an error whenever the decoded information sequence Y falls inside an error region E. The BER P_e can thus be defined as

$$P_e = P[Y \in E] = \int_E P_Y(y)dy. \tag{4.2}$$

We introduce an error indicator function $1_E(\mathbf{r})$ that equals 1 when $\mathbf{r} \in E$ and equals 0 when $\mathbf{r} \notin E$. Using this definition, (4.2) can be expressed as

$$P_e = \int_{-\infty}^{\infty} 1_{E_X} P_X(x)dx, \tag{4.3}$$

P_X being the distribution of the input random variable.

In general, it is difficult, and sometimes impossible, to calculate the BER analytically and it often must be estimated. The popular Monte-Carlo simulation technique is convenient for estimating BER. We set up a simulation (involving the SSF method) to transmit a large number of bits, count the number of errors, and obtain the BER by dividing the number of incorrectly received bits by the total number of transmitted bits during a given time. Unfortunately, the Monte-Carlo method has a high computational cost for very low BER, simply because a very large number of bits need to be sent to obtain a reasonable number of flipped bits [78].

In the importance sampling method (IS), the statistics of the noise sources in the system are biased in some manner so that bit errors occur with greater probability, thereby reducing computations. Essentially, the importance sampling technique is a modification of the standard Monte-Carlo procedure. This modification leads to a biased estimator. However, since the function which modifies the original distribution is selected by us, we can account and correct for its effect on the overall estimator, and still have an unbiased estimator. The method is a variance reduction technique, since for a given number of trials, the method is expected to produce a BER estimate with lower variance as compared to the Monte-Carlo technique.

For importance sampling simulations, the samples are generated from a biased density function P_X^*, instead of the original distribution P_X. Thus (4.3) can now be expressed as

$$P_e = \int_{-\infty}^{\infty} 1_{E_X} h(x) P_X^*(x) dx, \qquad (4.4)$$

where $h(.)$ is a weight function defined as $h(x) = \frac{P_X(x)}{P_X^*(x)}$.

Using the importance sampling method, we generate a 3-GMM employing the weighted expectation-maximization algorithm [79]. We need a Gaussian model mixture that would match the actual distribution near the tail(s). The sample distribution P_X^* is a Gaussian with twice the variance. We estimate the probability density function of the received signal Y and then compute the BER estimate using Monte-Carlo simulation but with the biased distribution P_X^*. The GMM of course can give us an estimate P_{GMM} of the distribution, as a weighted sum of Gaussian pdfs, as shown in (3.1).

4.5 Results

Finally, we are ready to use the methods discussed in this chapter to predict the bit error rate of a practical system. We calculate the bit error rate for 11 transmitted pulses sent over one span of SSMF for no-nonlinearity case, Monte-Carlo simulation, importance sampling, and importance sampling with 3-GMM.

In Fig. 4.17, we present the bit error rate curves when the nonlinear parameter $\gamma = 2/(\text{Wkm})$. As long as the nonlinearity (input power level) is low, all three approaches yield results close to the no-nonlinearity limit. We do not have enough errors at higher input power levels to use Monte-Carlo simulation (counting the number of errors), and the corresponding curve is incomplete. This is the reason we need to use importance sampling or our 3-GMM model.

We note that there are slight differences when 3-GMM is used. Our guess is that the importance sampling is unbiased only when there is no memory in the system. This condition is not exactly true for our optical communication system. The backpropagation removes the ISI, but the dispersion has already interacted with the nonlinearity creating some memory in the system which cannot be removed by backpropagating.

4.6 Chapter Summary

In this chapter, we analyzed the transmission of multiple pulses (adjacent and sometimes not adjacent) over a single optical fiber span. This is one more step closer to an actual communication system. We learned that our GMM is still a valuable tool in understanding this situation and we validated the same. We saw that adding a second pulse may significantly change our analysis when we do not
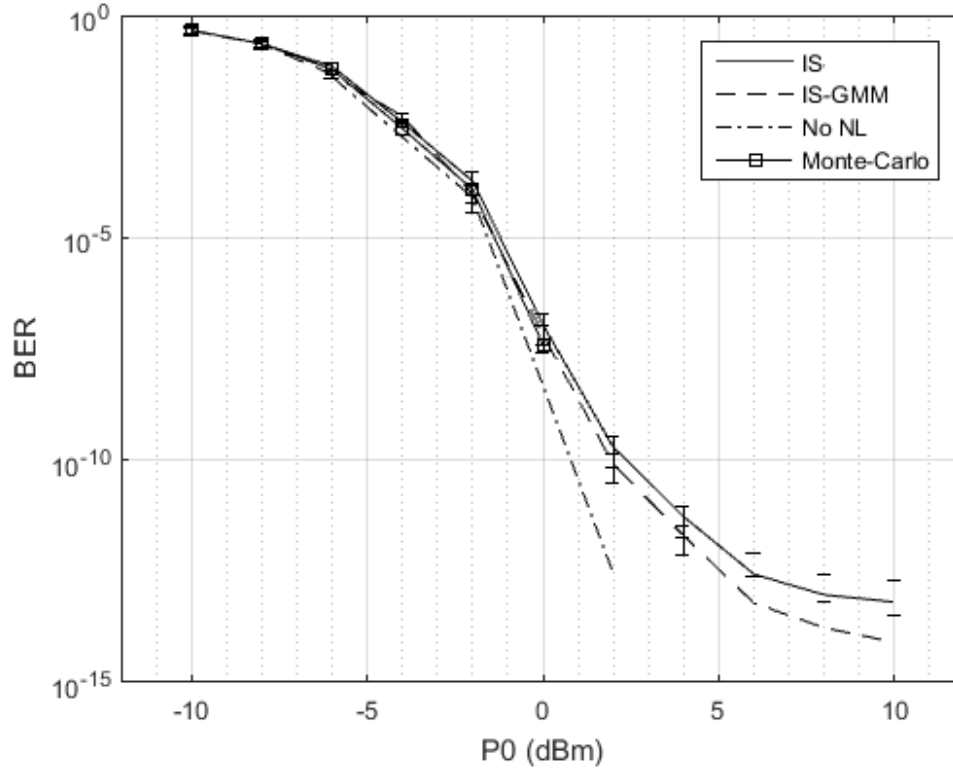
Figure 4.17: BER for different models at regular nonlinearity ($\gamma = 2/(\text{Wkm})$) (95% confidence intervals)

account for the signal-signal interaction, and we used digital backpropagation to isolate the effects of noise.

We also studied transmission of several dissimilar pulses and reached similar conclusions. We showed that sending 7 or 9 pulses is a good approximation to predict the analysis of a practical communication system (when an infinite number of signal pulses would be transmitted).

Finally, we proved that using a 3- or higher-order GMM is necessary to fully understand the nonlinear interaction, and that using an additive Gaussian model may be futile. We used importance sampling to compare our results with kernel density decomposition, and also presented bit error rate curves which show our

GMM techniques provide a better tool to analyze the optical fiber communication system.

Chapter 5

Conclusions and Future Work

In this work, we note the limited ability of existing signal processing tools to understand or mitigate signal-noise nonlinear interaction effects in an optical fiber communication system.

We proposed and validated a nonlinear decision feedback equalizer utilizing a modified training scheme. We applied the designed DFE to a fiber-optic system under various conditions and modulation formats and showed that our DFE promises good performance at low computational complexity. We also studied the problem of performance prediction of the DFE both in absence and in presence of error propagation.

Next, we sought to understand the nature of the noise component at the receiver end of an optical communication channel which is not additive Gaussian due to the signal-signal, signal-noise and noise-noise interactions. We used a fundamental scenario (transmission of a single signal pulse) for our setup and performed GMM decomposition on the isolated noise and nonlinear interaction components. We validated our technique and learned that at higher levels of nonlinearity, the GMM analysis is more accurate than an additive Gaussian or a Volterra series transfer

function model.

Finally, we generalized the single pulse analysis by transmitting multiple signal pulses, both similar and dissimilar. We showed that sending 7 or 9 pulses is a good approximation to predict the analysis of a practical communication system (when an infinite number of signal pulses would be transmitted). We proved that using a 3- or higher-order GMM is necessary to fully understand the nonlinear interaction, and that an additive Gaussian noise model is an unsatisfactory tool for analysis.

As future work, we propose to study in depth the interaction between more than two pulses. This would lead to large (multi-dimensional covariance matrices) and new methods may have to be designed. With the knowledge on signal-noise nonlinear interaction gleaned from the research presented in this thesis, better receivers, and especially new signal processing algorithms can be designed.

Bibliography

[1] *The decline of the mobile web.* [Online]. Available: http://cdixon.org/2014/04/07/the-decline-of-the-mobile-web/

[2] *Zoom Avant Ndeg12.* [Online]. Available: http://www.images-et-reseaux.com/fr/blogs/2012/07/zoom-avant-ndeg12/

[3] X. Wang and M. K. Reiter, "Mitigating bandwidth-exhaustion attacks using congestion puzzles," in *Proceedings of the 11th ACM Conference on Computer and Communications Security*, ser. CCS '04. New York, NY, USA: ACM, 2004, pp. 257–267.

[4] G. Chauvel, "Dispersion in optical fibers," *Anritsu Corporation*, p. 1, 2008.

[5] G. P. Agrawal, *Fiber-Optic Communication Systems*, 3rd ed. Wiley-Interscience, 2002.

[6] X. Zhou and J. Yu, "Advanced coherent modulation formats and algorithms: Higher-order multi-level coding for high-capacity system based on 100Gbps channel," in *2010 OFC/NFOEC*, March 2010.

[7] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948. [Online]. Available: http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x

[8] P. J. Winzer, "Optical networking beyond WDM," *IEEE Photonics Journal*, vol. 4, no. 2, pp. 647–651, April 2012.

[9] W. Rosenkranz and C. Xia, "Electrical equalization for advanced optical communication systems," {*AEU*} - *International J. of Electronics and Communications*, vol. 61, no. 3, pp. 153 – 157, 2007.

[10] K. Azadet, E. F. Haratsch, H. Kim, F. Saibi, J. H. Saunders, M. Shaffer, L. Song, and M.-L. Yu, "Equalization and FEC techniques for optical transceivers," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 3, pp. 317–327, Mar 2002.

[11] E. Ip and J. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *J. Lightwave Technol.*, vol. 26, no. 20, pp. 3416–3425, Oct 2008.

[12] J. Zweck and C. R. Menyuk, "Validity of the additive white Gaussian noise model for quasi-linear long-haul return-to-zero optical fiber communications systems," *Journal of Lightwave Technology*, vol. 27, no. 16, pp. 3324–3335, Aug 2009.

[13] K. Peddanarappagari and M. Brandt-Pearce, "Volterra series transfer function of single-mode fibers," *Lightwave Technology, Journal of*, vol. 15, no. 12, pp. 2232–2241, Dec 1997.

[14] D. M. Baney, P. Gallion, and R. S. Tucker, "Theory and measurement techniques for the noise figure of optical amplifiers," *Optical Fiber Technology*, vol. 6, no. 2, pp. 122–154, Apr 2000.

[15] G. P. Agrawal, *Nonlinear Fiber Optics*, 4th ed.   Academic Press, 2006.

[16] M. Bohn, W. Rosenkranz, F. Horst, B. Offrein, G.-L. Bona, and P. Krumm-rich, "Integrated optical FIR-filters for adaptive equalization of fiber channel impairments at 40 Gbit/s," in *Optical Communication Theory and Techniques*, E. Forestieri, Ed. Springer, 2005, pp. 181–188.

[17] C. Poole, J. Wiesenfeld, D. DiGiovanni, and A. M. Vengsarkar, "Optical fiber-based dispersion compensation using higher order modes near cutoff," *J. Lightwave Technol.*, vol. 12, no. 10, pp. 1746–1758, Oct 1994.

[18] G. D. Forney, "The Viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, March 1973.

[19] J. Zhao and A. Ellis, "Performance improvement using a novel MAP detector in coherent WDM systems," in *2008 ECOC*, Sept 2008.

[20] Y. Cai, D. Foursa, C. Davidson, J.-X. Cai, O. Sinkin, M. Nissov, and A. Pilipet-skii, "Experimental demonstration of coherent MAP detection for nonlinearity mitigation in long-haul transmissions," in *2010 OFC/NFOEC*, Mar 2010.

[21] M. S. Faruk and K. Kikuchi, "Adaptive frequency-domain equalization in digital coherent optical receivers," *Opt. Express*, vol. 19, no. 13, pp. 12 789–12 798, Jun 2011.

[22] R. Kudo, T. Kobayashi, K. Ishihara, Y. Takatori, A. Sano, E. Yamada, H. Masuda, and Y. Miyamoto, "PMD compensation in optical coherent single carrier transmission using frequency-domain equalisation," *Electronics Letters*, vol. 45, no. 2, pp. 124–125, January 2009.

[23] H. Song and M. Brandt-Pearce, "Model-centric nonlinear equalizer for coherent long-haul fiber-optic communication systems," in *2013 GLOBECOM*, Dec 2013, pp. 2394–2399.

[24] Y. Gao, F. Zhang, L. Dou, Z. Chen, and A. Xu, "Intra-channel nonlinearities mitigation in pseudo-linear coherent QPSK transmission systems via nonlinear electrical equalizer," *Optics Communications*, vol. 282, no. 12, pp. 2421 – 2425, 2009.

[25] J. Pan and C.-H. Cheng, "Wiener-Hammerstein model based electrical equalizer for optical communication systems," *J. Lightwave Technol.*, vol. 29, no. 16, pp. 2454–2459, Aug 2011.

[26] P. S. R. Diniz, *Adaptive Filtering Algorithms and Practical Implementation*, 3rd ed.   Springer, 2008.

[27] M. Aharonovich and S. Arnon, "Performance improvement of optical wireless communication through fog with a decision feedback equalizer," *J. Opt. Soc. Am. A*, vol. 22, no. 8, pp. 1646–1654, Aug 2005.

[28] M. Zamani, C. Li, C. Chen, and Z. Zhang, "An adaptive decision-feedback channel estimation for coherent optical OFDM," in *2013 OFC/NFOEC*, Mar 2013.

[29] J.-Y. Lin and C.-H. Wei, "Adaptive nonlinear decision feedback equalization with channel estimation and timing recovery in digital magnetic recording systems," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Trans. on*, vol. 42, no. 3, pp. 196–206, Mar 1995.

[30] L. Agarossi, S. Bellini, F. Bregoli, and P. Migliorati, "Equalization of non-linear optical channels," in *1998 ICC*, vol. 2, Jun 1998, pp. 662–667.

[31] Y. Lu and E. Alon, "Design techniques for a 66 Gb/s 46 mW 3-tap decision feedback equalizer in 65 nm CMOS," *Solid-State Circuits, IEEE J. of*, vol. 48, no. 12, pp. 3243–3257, Dec 2013.

[32] J. Fickers, A. Ghazisaeidi, M. Salsi, G. Charlet, P. Emplit, and F. Horlin, "Decision-feedback equalization of bandwidth-constrained N-WDM coherent optical communication systems," *J. Lightwave Technol.*, vol. 31, no. 10, pp. 1529–1537, May 2013.

[33] X. Han and C.-H. Cheng, "Nonlinear filter based decision feedback equalizer for optical communication systems," *Opt. Express*, vol. 22, no. 7, pp. 8712–8719, Apr 2014.

[34] C. Vladeanu and C. Paleologu, "MMSE single user iterative receiver for asynchronous DS-CDMA systems," in *2006 ICC*, 2006, pp. 227–230.

[35] S. B. Alexander, *Optical Communication Receiver Design.* SPIE Optical Engineering Press, 1997.

[36] H. Song and M. Brandt-Pearce, "A 2-d discrete-time model of physical impairments in wavelength-division multiplexing systems," *J. Lightwave Technol.*, vol. 30, no. 5, pp. 713–726, March 2012.

[37] R. Wang, N. Jindal, T. Bruns, A. Bahai, and D. Cox, "Comparing RLS and LMS adaptive equalizers for nonstationary wireless channels in mobile ad hoc networks," in *Personal, Indoor and Mobile Radio Communications, 2002.*

The 13th IEEE International Symposium on, vol. 3, Sept 2002, pp. 1131–1135 vol.3.

[38] S. Ardalan, "Floating-point error analysis of recursive least-squares and least-mean-squares adaptive filters," *Circuits and Systems, IEEE Transactions on*, vol. 33, no. 12, pp. 1192–1208, Dec 1986.

[39] J. P. Gordon and L. F. Mollenauer, "Phase noise in photonic communications systems using linear amplifiers," *Opt. Lett.*, vol. 15, no. 23, pp. 1351–1353, Dec 1990.

[40] E. Ip, "Coherent detection and digital signal processing for fiber optic communications," p. 217, 2009.

[41] E. Ip, E. Mateo, and T. Wang, "Reduced-complexity nonlinear compensation based on equivalent-span digital backpropagation," in *2012 COIN*, May 2012, pp. 28–29.

[42] P. L. Zador, "Error probabilities in data system pulse regenerator with DC restoration," *Bell Sys. Tech. J.*, vol. 45, no. 6, pp. 979–984, 1966.

[43] D. Duttweiler, J. Mazo, and D. Messerschmitt, "An upper bound on the error probability in decision-feedback equalization," *Information Theory, IEEE Trans. on*, vol. 20, no. 4, pp. 490–497, Jul 1974.

[44] A. H. Gnauck, R. W. Tkach, A. R. Chraplyvy, and T. Li, "High-capacity optical transmission systems," *Journal of Lightwave Technology*, vol. 26, no. 9, pp. 1032–1045, May 2008.

[45] J. G. Proakis, *Intersymbol Interference in Digital Communication Systems*. John Wiley Sons, Inc., 2003.

[46] J. A. Gubner, *Probability and Random Processes for Electrical and Computer Engineers.* New York, NY, USA: Cambridge University Press, 2006.

[47] A. Mecozzi, "Limits to long-haul coherent transmission set by the kerr nonlinearity and noise of the in-line amplifiers," *Journal of Lightwave Technology*, vol. 12, no. 11, pp. 1993–2000, Nov 1994.

[48] K.-P. Ho, "Performance degradation of phase-modulated systems due to nonlinear phase noise," *IEEE Photonics Technology Letters*, vol. 15, no. 9, pp. 1213–1215, Sept 2003.

[49] L. Beygi, E. Agrell, M. Karlsson, and P. Johannisson, "Signal statistics in fiber-optical channels with polarization multiplexing and self-phase modulation," *Journal of Lightwave Technology*, vol. 29, no. 16, pp. 2379–2386, Aug 2011.

[50] A. Mecozzi, "Probability density functions of the nonlinear phase noise," *Opt. Lett.*, vol. 29, no. 7, pp. 673–675, Apr 2004.

[51] Y. Yadin, M. Shtaif, and M. Orenstein, "Nonlinear phase noise in phase-modulated WDM fiber-optic communications," *IEEE Photonics Technology Letters*, vol. 16, no. 5, pp. 1307–1309, May 2004.

[52] A. P. T. Lau and J. M. Kahn, "Signal design and detection in presence of nonlinear phase noise," *Journal of Lightwave Technology*, vol. 25, no. 10, pp. 3008–3016, Oct 2007.

[53] I. B. Djordjevic, B. Vasic, M. Ivkovic, and 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.

[54] M. H. Taghavi, G. C. Papen, and P. H. Siegel, "On the multiuser capacity of WDM in a nonlinear optical fiber: Coherent communication," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5008–5022, Nov 2006.

[55] N. J. Muga, M. C. Fugihara, M. F. S. Ferreira, and A. N. Pinto, "Non-Gaussian ASE noise in raman amplification systems," *Journal of Lightwave Technology*, vol. 27, no. 16, pp. 3389–3398, Aug 2009.

[56] A. Demir, "Nonlinear phase noise in optical-fiber-communication systems," *Journal of Lightwave Technology*, vol. 25, no. 8, pp. 2002–2032, Aug 2007.

[57] *Clustering Using Gaussian Mixture Models.* [Online]. Available: https://www.mathworks.com/help/stats/clustering-using-gaussian-mixture-models.html?requestedDomain=www.mathworks.com/

[58] M. Nilsson, H. Gustaftson, S. Andersen, and W. Kleijn, "Gaussian mixture model based mutual information estimation between frequency bands in speech," in *Proceedings of 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2002, pp. 525–528.

[59] J. Harshan and E. Viterbo, "Constellation constrained capacity of additive Gaussian mixture noise channels," in *Information Theory and its Applications (ISITA), 2014 International Symposium on*, Oct 2014, pp. 110–114.

[60] J. Ilow, D. Hatzinakos, and A. N. Venetsanopoulos, "Performance of FH SS radio networks with interference modeled as a mixture of Gaussian and alpha-stable noise," *IEEE Transactions on Communications*, vol. 46, no. 4, pp. 509–520, Apr 1998.

[61] M. Novey and T. Adali, "Complex fixed-point ICA algorithm for separation of QAM sources using Gaussian mixture model," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 2, April 2007, pp. II–445–II–448.

[62] J. T. Flåm, J. Jaldén, and S. Chatterjee, "Gaussian mixture modeling for source localization," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on.* IEEE, 2011, pp. 2604–2607.

[63] J. T. Flåm, S. Chatterjee, K. Kansanen, and T. Ekman, "On MMSE estimation: A linear model under Gaussian mixture statistics," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3840–3845, 2012.

[64] L. Fortuna, G. Nunnari, and S. Nunnari, *Nonlinear Modeling of Solar Radiation and Wind Speed Time Series.* Springer International Publishing, 2016.

[65] D. Reynolds, *Gaussian Mixture Models.* Boston, MA: Springer US, 2009, pp. 659–663. [Online]. Available: https://doi.org/10.1007/978-0-387-73003-5_196

[66] T. Moon, "The expectation-maximization algorithm," *Signal Processing Magazine, IEEE*, vol. 13, no. 6, pp. 47–60, Nov 1996.

[67] J. Rousseau and K. Mengersen, "Asymptotic behaviour of the posterior distribution in overfitted mixture models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 5, pp. 689–710, 2011.

[68] G. P. Agrawal, *Nonlinear Fiber Optics*, 5th ed. Academic Press, 2013.

[69] L. N. Binh and D. Liet, *Nonlinear Optical Systems : principles, applications, and advanced signal processing with MATLAB and Simulink models.* Boca Raton, Fla. : CRC, 2012.

[70] J.-H. Lee, D.-H. Han, and B.-Y. Choi, "Analysis of fiber nonlinearities by perturbation method," *J. Opt. Soc. Korea*, vol. 9, no. 1, pp. 6–12, Mar 2005.

[71] H. Song and M. Brandt-Pearce, "A discrete-time polynomial model of single channel long-haul fiber-optic communication systems," in *2011 IEEE International Conference on Communications (ICC)*, June 2011, pp. 1–6.

[72] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 03 1951.

[73] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York, NY, USA: Wiley-Interscience, 1991.

[74] P. J. Smith, M. Shafi, and H. Gao, "Quick simulation: A review of importance sampling techniques in communications systems," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 4, pp. 597–613, 1997.

[75] R. Rubinstein and D. Kroese, *Simulation and the Monte Carlo Method*, 2nd ed. John Wiley and Sons, 2007.

[76] *An introduction to kernel density estimation.* [Online]. Available: http://www.mvstat.net/tduong/research/seminars/seminar-2001-05/

[77] *Kernel Density Estimators.* [Online]. Available: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/AV0405/MISHRA/kde.html

[78] M. Jeruchim, "Techniques for estimating the bit error rate in the simulation of digital communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 1, pp. 153–170, Jan 1984.

[79] Y. Matsuyama, "The weighted EM algorithm and block monitoring," in *Neural Networks,1997., International Conference on*, vol. 3, Jun 1997, pp. 1936–1941 vol.3.

[80] M. K. Simon, "Extensions to the analysis of regenerative repeaters with quantized feedback," *Bell Sys. Tech. J.*, vol. 46, no. 8, pp. 1831–1851, 1967.

[81] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[82] D. Maiti and M. Brandt-Pearce, "Modified nonlinear decision feedback equalizer for long-haul fiber-optic communications," *Lightwave Technology, Journal of*, vol. 33, no. 18, pp. 3763–3772, Sept 2015.