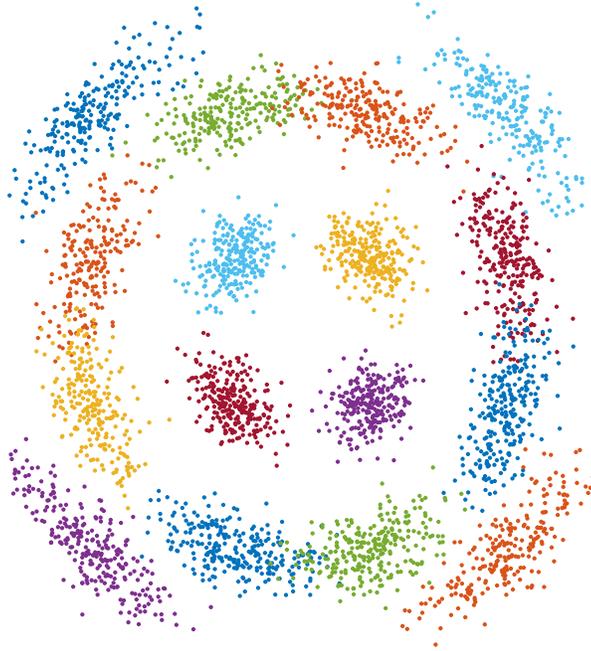




CHALMERS

UNIVERSITY OF TECHNOLOGY



Joint Phase Noise Estimation and Symbol Detection in Optical Communication Systems

Master's thesis in Communication Engineering

ARNI ALFREDSSON

Communication Systems Group
Department of Signals and Systems
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2015

EX022/2015

Joint Phase Noise Estimation and Symbol Detection in Optical Communication Systems

ARNI ALFREDSSON

Department of Signals and Systems
Technical Report No. EX022/2015

Communication Systems Group
Department of Signals and Systems
Chalmers University of Technology
SE-41296 Gothenburg, Sweden
Telephone: + 46 (0)31-772 1000
www.chalmers.se

Front cover illustration:

Received samples from a quadrature amplitude modulation transmission, affected by phase noise and additive white Gaussian noise. Colors indicate different transmitted symbols.

Copyright ©2015 ARNI ALFREDSSON.
All rights reserved.

This thesis has been prepared using L^AT_EX.

Abstract

Demands for higher transmission rates are ever increasing as a result of requirements imposed by a number of different applications. To meet these demands, worldwide research efforts for optical communications have led to tremendous growth in performance in the past few decades. Nowadays, optical communication systems find applications in various scenarios, ranging from backbone networks to data centers.

In recent years, coherent communication systems in fiber-optical channels have attracted interest in the academic and industrial communities. Coherent optical systems enable the encoding of information in both the amplitude and phase of the signal, which can result in improved performance with respect to spectral and power efficiency. Furthermore, advances in the electronic hardware have facilitated effective signal impairment compensation using digital signal processing, allowing the adoption of algorithms from wireless communications. A particular impairment is laser phase noise, which can severely limit the gains promised by coherent optical systems.

In this thesis the problem of optimal symbol detection in the presence of laser phase noise in uncoded optical communication systems is studied. We assume a single carrier transmission on two independent polarizations in a channel that includes phase noise, additive white Gaussian noise, and a random constant phase offset on each polarization component. To this end, the maximum a posteriori (MAP) symbol detector is presented, which is shown to be analytically intractable. Then, a pilot-based algorithm developed using the factor graph framework and the sum product algorithm, that jointly estimates phase and detects symbols for arbitrary quadrature amplitude modulation (QAM) constellations, is studied. Performance is evaluated using Monte Carlo simulations for quadrature phase shift keying, 16-QAM and 64-QAM constellations. Results show that this algorithm is more tolerant to phase noise compared to other algorithms found in optical literature.

Keywords: MAP detection, phase noise, estimation, optical communication, coherent detection, QAM, pilot-aided, factor graphs, sum-product algorithm.

Preface

This report is the product of a one-year 60 credit master's thesis work conducted at Chalmers University of Technology, between September 2014 and July 2015. It is an extension of A. Alfredsson, R. Krishnan, and E. Agrell, "Joint Phase Noise Estimation and Symbol Detection in Optical Communication Systems," a journal article manuscript that, at the time of writing, is being prepared for submission.

Arni Alfredsson

Gothenburg, June 2015

Acknowledgments

I would like to begin by thanking my examiner, Prof. Erik Agrell, for his support throughout the thesis work. He has given me a lot of useful advice when I have been faced with difficulties, while sharing his valuable insights and experience. I am grateful for the opportunity he has given me to further my academic efforts, by offering me a PhD position at Chalmers.

I want to thank the professors, in addition to others, who contributed to all the excellent courses that I attended at Chalmers. They played a big role in making me interested in the field of communication engineering, and got me motivated to be a part of the academic community after obtaining my master's degree. Furthermore, I would like to acknowledge Henrik Eliasson, Magnus Karlsson, and Pontus Johannisson for helpful discussions regarding the thesis.

My gratitude goes to Rajet Krishnan, who served as my supervisor during my thesis work. He was generous in giving me support and advice in every facet of the work this last winter. I am particularly grateful he reached out and offered me to work on this thesis.

I appreciate all the different people I have come to know here at Chalmers, from the master's program, the Department of Signals and Systems, as well as others. Thanks to them, these past two years in Gothenburg have passed very quickly.

Lastly, special thanks to my friends and family, and in particular to Jóhanna, for the love and support they have given me during my time at Chalmers.

Arni Alfredsson

Gothenburg, June 2015

Contents

Abstract	i
Preface	ii
Acknowledgments	iii
Acronyms	vi
1 Introduction	1
1.1 Optical Communications	1
1.2 Signal Impairments	2
1.2.1 Laser Phase Noise	3
1.2.2 Phase Noise Compensation – Prior work	4
1.3 Optimal Detection	6
1.4 Contributions	7
1.5 Organization	7
1.6 Concepts	8
1.6.1 Marginal Distributions	8
1.6.2 Factor Graphs and the Sum Product Algorithm	8
2 Optimal Symbol Detection	10
2.1 System Model	10
2.2 MAP Symbol Detection	11
3 Receiver Implementation	17
3.1 Estimation of Phase Difference	17
3.2 Phase Noise Estimation – Symbol Detection	18
3.3 Multiple Iterations	23
3.4 Numerical problems	24
4 Simulation Results	28
4.1 Linewidth Tolerance	29
4.2 Iterations	31

4.3 Pilot Density	32
4.4 Benefits of Phase Noise PDFs	33
5 Conclusion and Extensions	35
5.1 Future Work	35
Bibliography	37
A Message Derivations	41
B Variant of Proposed Algorithm	42

Acronyms

ASE	amplified spontaneous emission
AWGN	additive white Gaussian noise
BER	bit error rate
DE	differential encoding
DSP	digital signal processing
FG	factor graph
MAP	maximum a posteriori
PD	pilot density
PDF	probability density function
PMF	probability mass function
QAM	quadrature amplitude modulation
QPSK	quadrature phase shift keying
SER	symbol error rate
SNR	signal to noise ratio
SP	sensitivity penalty
SPA	sum product algorithm
SPM	self-phase modulation
VV	Viterbi-Viterbi

1 Introduction

1.1 Optical Communications

Optical communications in some shape or form have been around for centuries. The modern definition is a lightwave system that uses visible or near-infrared light to convey information, and typically guides it through an optical fiber, although unguided systems also exist. Ancient civilizations used mirrors and fire to communicate over distances, while during the medieval centuries, information was conveyed through the use of lamps and other devices [1, Ch. 1]. In the 18th century, information was signaled through hundreds of kilometers with so-called optical telegraphs. These tools were quite restricted with respect to the amount of information they could deliver at a time however. In the 19th century and the first part of the 20th century, electrical and microwave communication systems were predominant, as these technologies were capable of much higher transmission rates than the previously used optical techniques. In the 1960s, optical communications started to receive focus again with the invention of the laser along with the discovery of optical fibers. Several years later, a breakthrough in optical communications research took place, leading to world-wide research efforts. Since then, research and commercial systems have seen colossal progress in reach and capacity. The bit rate-distance product, a common measure of performance in communication systems, has grown exponentially in recent years for optical systems. Long-haul transmissions of terabits per second data rates have already been experimentally demonstrated [2, 3].

Demands for higher transmission rates are ever increasing as a result of high bandwidth requirements, imposed by a number of different applications, such as online multimedia and gaming. Optical communication systems have therefore found their use in various applications, on a worldwide scale, due to their high capacity. One of the main reasons for the large transmission rates is the fact that the carriers used in these systems have frequencies of around 200 THz. Typically, the bandwidth of the

signal can be a few percent of the carrier frequency, implying a potential for high bit rates. Additionally, high spectral efficiency can be attained using coherent detection with polarization multiplexing [4]. These factors, among others, lead to potential data rates that are unmatched by that of any other types of communication systems.

Coherent communication systems in fiber-optical channels have in particular been a topic of interest in recent years. These systems allow information to be encoded in both the amplitude and phase, on each polarization component of the lightwave. This facilitates the use of modulation formats that have the potential to deliver better performance in terms of spectral and power efficiency. Furthermore, advances in the electronic hardware have enabled effective channel impairment compensation using digital signal processing (DSP). This has spurred the application of DSP algorithms from wireless communications to coherent optical communication systems [5].

1.2 Signal Impairments

An optical fiber introduces various linear signal impairments that have to be compensated for, e.g. chromatic dispersion and polarization mode dispersion, among others [4]. More specifically, imperfections in the optical fiber lead to a random state of polarization in the electromagnetic wave at the receiver end. When the signal is separated into its two polarization components at the receiver, the random state of polarization causes the outputs to contain linear combinations of both components. The mixing of the components is typically described with a Jones matrix

$$\mathbf{A} = \frac{1}{\sqrt{uu^* + vv^*}} \begin{bmatrix} u & v \\ -v^* & u^* \end{bmatrix}, \quad (1.1)$$

where u and v are circularly symmetric complex Gaussian variables.

Additionally, transmitted optical signals are attenuated due to loss in the fiber. As a result, optical amplifiers are routinely inserted in optical channels with periodic intervals. Each section of the channel between the optical amplifiers is referred to as a span. The optical amplifiers compensate for the signal attenuation, but introduce

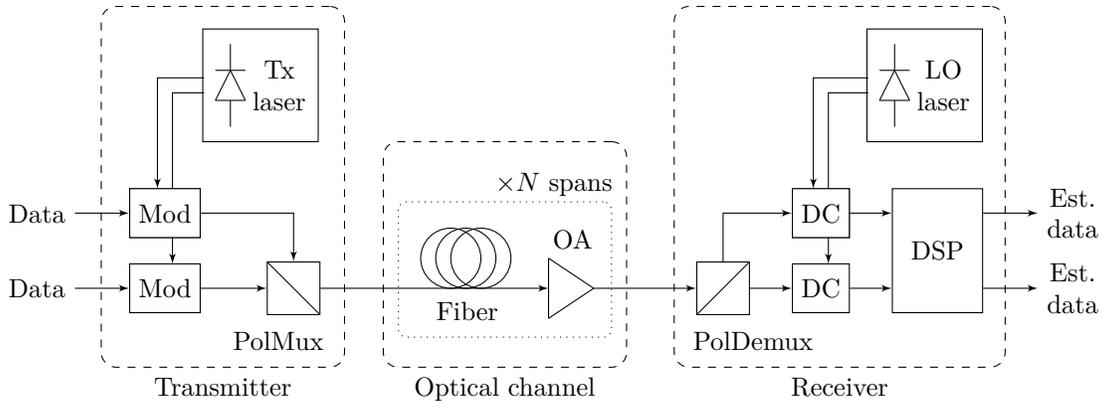


Figure 1.1: A high-level block diagram for a single-carrier polarization multiplexed transmission over an optical channel. (DC: Downconversion, LO: Local oscillator, Mod: Modulation, OA: Optical amplifier, PolDemux: Polarization demultiplexing, PolMux: Polarization multiplexing, Rx: Receiver, Tx: Transmitter.)

amplified spontaneous emission (ASE) noise to the transmission. Furthermore, the transmission quality is degraded by shot noise from the receiver’s local oscillator laser, and thermal noise from the receiver’s components. These noise contributions are commonly modeled as additive white Gaussian noise (AWGN).

In addition to the aforementioned linear impairments, an optical fiber brings about a number of nonlinear impairments to the signal. Examples of those are self-phase modulation (SPM), cross-phase modulation, and four-wave mixing. These effects occur due to the Kerr effect, which imposes an amplitude-dependent phase shift to the signal. For single-carrier transmissions, a particular type of nonlinear phase noise occurs which is often referred to as the Gordon-Mollenauer effect. It is caused by an interaction between ASE and the signal through SPM. The severity of this impairment increases with longer transmission distances.

1.2.1 Laser Phase Noise

A high-level overview of a single-carrier polarization multiplexed transmission, over an optical fiber channel, is depicted in Fig. 1.1. In particular, the transmitter laser and the local oscillator laser at the receiver are shown. Noise in these lasers causes phase drift in the generated waveforms. These drifts are typically referred to as laser

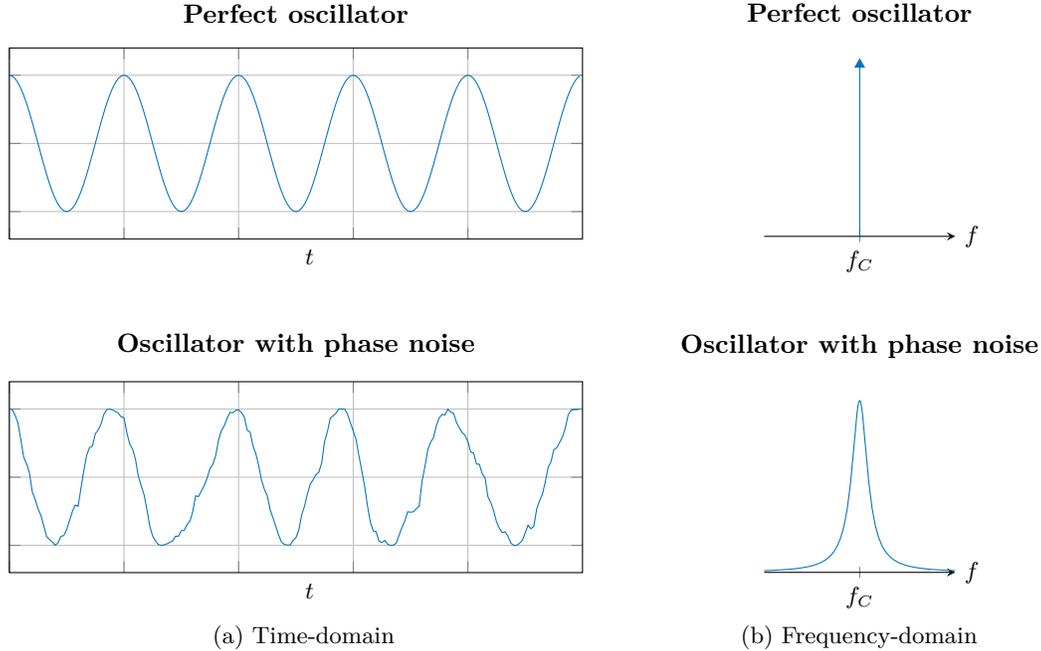


Figure 1.2: Laser phase noise impact on a single carrier, where the phase noise manifests itself as jitter in time (a), and spectral growth in frequency (b).

phase noise, and lead to spectral growth of the signal. The effects on a single carrier with a frequency f_C , and its power spectral density are depicted in Fig. 1.2. The 3-dB bandwidth of the power spectral density curve corresponding to a noisy oscillator is called the laser linewidth. The linewidth is an indication of the laser stability, and is proportional to the phase noise variance. The focus in this thesis is on scenarios where laser phase noise is dominant, thus it will be assumed to be the sole source of phase noise. The transmitted symbol phase has to be recovered at the receiver for successful detection in coherent systems. Hence, phase noise can degrade the system performance, and its severity intensifies with increasing modulation order. The effects of laser phase noise on the received samples of a quadrature amplitude modulation (QAM) transmission in the presence of AWGN is shown in Fig. 1.3. The phase noise causes the constellation to randomly rotate around its center point.

1.2.2 Phase Noise Compensation – Prior work

Substantial research has been made for blind phase noise compensation in coherent optical systems, i.e. methods that do not use any pilot symbols known at the re-

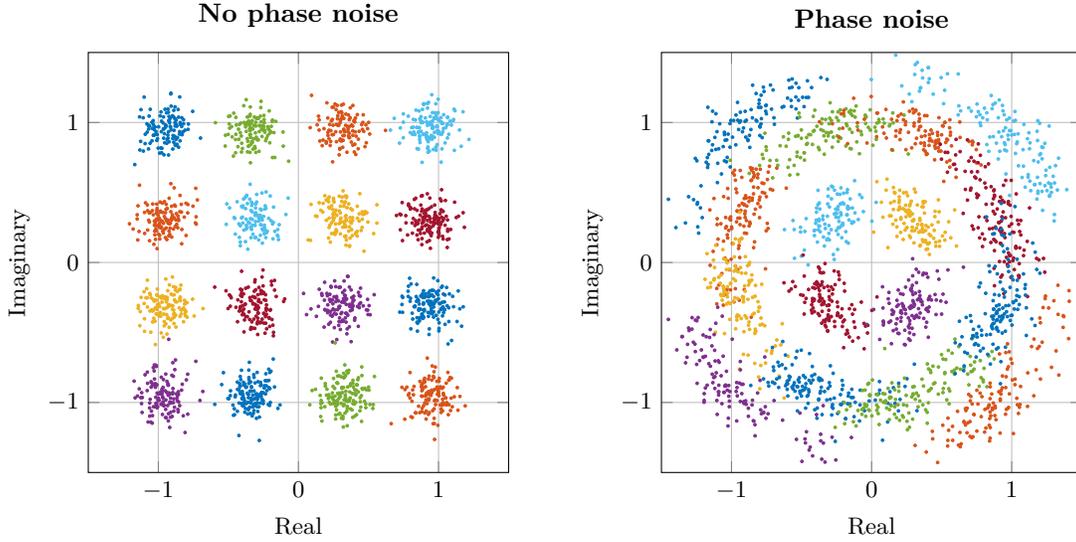


Figure 1.3: Laser phase noise impact on the received samples of a 16-QAM transmission in the presence of AWGN. The colors indicate different transmitted symbols.

ceiver. A classic feedback-based method for estimating and mitigating phase noise is a phase-locked loop [4]. However, delays in feedback loops decrease the effectiveness of previous phase estimates. Coherent optical systems utilize extensive parallelization together with pipelining to achieve the required throughput. In [6] it is shown that this leads to large amounts of feedback delay, making feedback-based methods, either hardware-based techniques or DSP algorithms, impractical for optical systems.

Various feedforward algorithms that do not require feedback loops have been proposed, e.g. the Viterbi-Viterbi (VV) [7], and its variations [8–10], which are developed for phase shift keying constellations. These algorithms will perform poorly for QAM as they assume constant symbol amplitude levels. Algorithms based on VV and quadrature phase shift keying (QPSK) partitioning [11] have been proposed that improve upon VV for QAM [12, 13]. Furthermore, among the best-known feedforward algorithms designed for QAM is the blind phase search algorithm [6]. Other approaches have been proposed for QAM [14], reporting less complexity than the blind phase search algorithm with similar performance [15].

Blind phase noise compensation algorithms have ambiguities in the phase noise estimation, due to lack of pilots giving absolute phase references. This may cause cycle slips which can further cause catastrophic failures [16]. A common way to mitigate

this is to employ differential encoding (DE), but DE has the downside that the average bit error rate (BER) will be higher [8]. Algorithms have been proposed that utilize pilots to estimate the phase noise without ambiguity [17–20], or to reduce the amount of cycle slips after blind phase noise estimation at the receiver [21]. This eliminates the need for DE, resulting in lower average BER. Another advantage of using pilots is the fact that modulation transparency is generally easier to achieve [22]. Moreover, the pilot symbols can be reused to perform other tasks in addition to phase noise estimation, such as polarization demultiplexing, frequency offset estimation, and fiber nonlinearity compensation [23]. On the other hand, pilots introduce overhead to the transmission which results in reduced data rate.

1.3 Optimal Detection

Detection algorithms in uncoded digital communication systems are designed such that the symbol error rate (SER) performance of the system is minimized. The optimal detector is the maximum a posteriori (MAP) symbol detector [24, p. 254]. In [25] it is shown that for a transmission in the presence of random phase noise and AWGN, the receiver that detects symbols with minimum SER has a detector-estimator structure. The received samples are first used to estimate the a posteriori probability density function (PDF) of the phase noise. This PDF is then utilized when performing symbol detection. In other words, the phase noise estimation does not entail obtaining point estimates of the phase noise. In optical communications, receiver algorithms generally do not conform to this structure. Relevant work can be found in [8], where symbol-by-symbol detection is performed by maximizing a particular a posteriori PDF. However, this is not the MAP symbol detector when the phase noise introduces correlation between the received samples.

The biggest challenge with realizing the MAP symbol detector is computing the a posteriori phase noise PDF, as it turns out to be an infinite dimensional problem in general [25]. Multiple attempts have been made to address this problem in the context of wireless communications, by using various approximations and frame-

works in order to derive low-complexity algorithms [26–28]. Factor graphs (FGs) are a notable framework that, when employed in conjunction with the sum product algorithm (SPA), can be utilized to derive a variety of algorithms in a wide range of applications. Specifically, it can be used to realize the MAP symbol detector by means of marginalization.

1.4 Contributions

In this thesis, motivated by the presence of laser phase noise in optical communication systems, and the optimal receiver structure in [25], we review the theory of optimal symbol detection for transmissions affected by AWGN and phase noise. We consider an optical channel where laser phase noise is dominant, and other impairments have been compensated for. We follow the approach introduced by Colavolpe *et al.* in [26], applying the FG framework to the problem of realizing the MAP symbol detector. Furthermore, we use the SPA to derive a practical pilot-based implementation of this detector. Finally, we compare the derived algorithm with conventional receiver structures, involving a blind or a pilot-based algorithm that finds point estimates of the phase noise and compensates for it using the estimates. We evaluate the performance of the different algorithms with Monte Carlo simulations. The results show that the proposed algorithm outperforms the other algorithms by a wide margin, while requiring low pilot density (PD).

1.5 Organization

The remainder of the thesis is organized as follows. In Chapter 2 the optical communication system model is presented, and a review of the optimal symbol detection for transmissions affected by AWGN and phase noise is introduced. An application of the FG framework to the problem of realizing the optimal detector is given, followed by usage of the SPA. In Chapter 3 a low-complexity algorithm, based on the FG and SPA from the preceding chapter, is derived. Finally, in Chapter 4 the simulation

results are presented, and the concluding remarks are given in Chapter 5.

1.6 Concepts

1.6.1 Marginal Distributions

Consider a joint distribution $p(x_1, x_2, \dots, x_N)$ with arguments x_1, x_2, \dots, x_N , whose domains are $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_N$, respectively. For this joint distribution, there are N associated marginal distributions $p_i(x_i)$, where $i = 1, 2, \dots, N$. The i -th marginal distribution is obtained by summing or integrating the joint distribution over the domain of all arguments x_j , where $j \neq i$. If the arguments are discrete, the i -th marginal distribution is computed as

$$p_i(x_i) = \sum_{x_1 \in \mathcal{D}_1} \cdots \sum_{x_{i-1} \in \mathcal{D}_{i-1}} \sum_{x_{i+1} \in \mathcal{D}_{i+1}} \cdots \sum_{x_N \in \mathcal{D}_N} p(x_1, x_2, \dots, x_N). \quad (1.2)$$

Similarly for continuous arguments, the marginalization is computed by integrating, as

$$p_i(x_i) = \int_{\mathcal{D}_1} \cdots \int_{\mathcal{D}_{i-1}} \int_{\mathcal{D}_{i+1}} \cdots \int_{\mathcal{D}_N} p(x_1, x_2, \dots, x_N) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_N. \quad (1.3)$$

For large N , the summations or integrals required to compute the marginal distributions may lack closed-form solutions, and directly carrying them out is impractical. However, exploiting the way a joint distribution can be factorized, the problem of finding the marginal distributions can be solved more efficiently.

1.6.2 Factor Graphs and the Sum Product Algorithm

FGs are graphical models that visually represent how multivariable functions can be factorized into marginal functions of fewer variables. They are bipartite, which by definition means their nodes can be divided into two disjoint sets. One set contains nodes representing marginal functions, while the other set contains nodes corre-

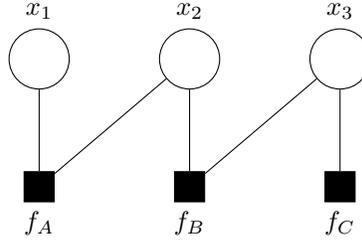


Figure 1.4: An FG corresponding to the product $f_A(x_1, x_2)f_B(x_2, x_3)f_C(x_3)$.

sponding to variables. The edges between the nodes in these two sets describe which variables are arguments of which marginal functions after the factorization. As an example, consider a multivariable function $f(x_1, x_2, x_3)$ which can be factorized as

$$f(x_1, x_2, x_3) = f_A(x_1, x_2)f_B(x_2, x_3)f_C(x_3). \quad (1.4)$$

Thus, x_1 is an argument of f_A , x_2 is an argument of f_A and f_B , and x_3 is an argument of f_B and f_C . The FG that corresponds to this factorization is depicted in Fig. 1.4. The circles and rectangles represent variable nodes and function nodes, respectively, and the lines correspond to edges. FGs can be used to describe the factorization of joint PDFs, and more specifically, the conditional independence among the random variables associated with the PDFs.

The SPA is an algorithm that works in FGs, and attempts to compute the marginals of joint distributions efficiently by exploiting their factorization. It involves message passing, where the messages describe some marginal function. Associated with every edge on the FG are two messages, one in each direction. Every message is a single-argument function of the variable whose node is connected to the edge corresponding to the message. By using the SPA, finding marginal distributions reduces to computing and propagating these messages between the nodes. FGs and the SPA are explained in more details in [29].

2 Optimal Symbol Detection

2.1 System Model

The transmission of complex modulation symbols in two polarizations is considered, over an AWGN channel with laser phase noise and a random state of polarization at the receiver end. The constant modulus algorithm [30] is assumed to have been successfully applied for polarization demultiplexing, a process where the original polarization components are retrieved from the linear combination that occurs due to (1.1). A success is defined as separating the polarization components and subjecting them to a random constant phase offset [31]. Other impairments, such as nonlinear phase noise, chromatic dispersion, and polarization mode dispersion, are assumed to have been compensated for. Considering linear modulation, perfect symbol synchronization, matched filtering, Nyquist sampling, and no intersymbol interference or frequency offset, the discrete-time baseband complex equivalent system model is written as

$$\begin{bmatrix} r_{x,k} \\ r_{y,k} \end{bmatrix} = \begin{bmatrix} s_{x,k} e^{jC_x} \\ s_{y,k} e^{jC_y} \end{bmatrix} e^{j\theta_k} + \begin{bmatrix} n_{x,k} \\ n_{y,k} \end{bmatrix}, \quad (2.1)$$

for $k = 0, 1, \dots, N - 1$. A schematic of the model is sketched in Fig. 2.1. The indices x and y denote the polarization components, and N is the number of transmitted symbols.

The vectors \mathbf{r}_x and \mathbf{r}_y are the received samples on each polarization, whereas \mathbf{s}_x and \mathbf{s}_y contain complex M -ary modulation symbols. The symbols are drawn independently

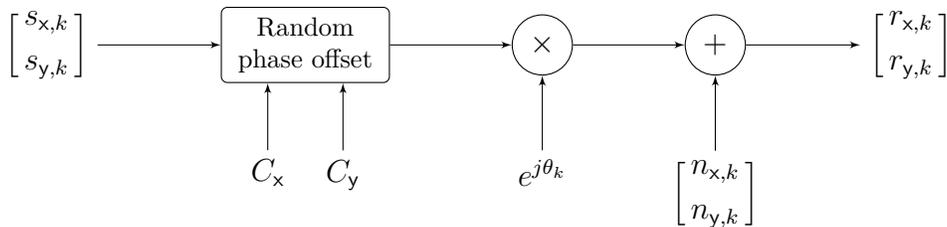


Figure 2.1: Discrete-time baseband complex equivalent model for a single-carrier polarization multiplexed transmission over an optical channel.

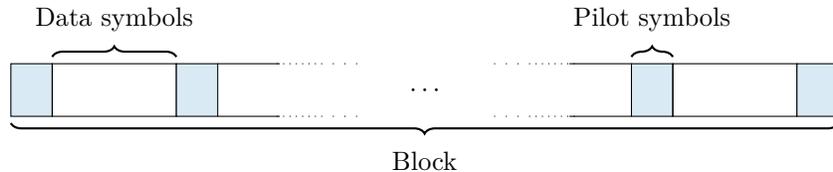


Figure 2.2: Pilot distribution in a transmitted block.

from a constellation $\mathcal{M} \subset \mathbb{C}$ with equal probability. The constellation is subjected to an energy constraint such that the average transmitted symbol energy is E_s . Pilot symbols are distributed uniformly in \mathbf{s}_x and \mathbf{s}_y , as depicted in Fig. 2.2. The noise vectors \mathbf{n}_x and \mathbf{n}_y are independent of each other, and contain independent and identically distributed circularly symmetric complex Gaussian components, $n_{x,k} \sim \mathcal{N}_{\mathbb{C}}(0, 2\sigma^2)$ and $n_{y,k} \sim \mathcal{N}_{\mathbb{C}}(0, 2\sigma^2)$. The vector $\boldsymbol{\theta}$ accounts for the accumulated phase noise from the transmitter and the local oscillator laser, and affects both polarizations equally. Moreover, it is defined modulo 2π , and is modeled as a random-walk process, expressed as

$$\theta_k = (\theta_{k-1} + \Delta_k) \bmod 2\pi, \quad (2.2)$$

with initial component $\theta_0 \sim \mathcal{U}[0, 2\pi)$. The step Δ_k is a real Gaussian random variable with zero mean and variance

$$\sigma_{\Delta}^2 = 2\pi\Delta\nu T_s, \quad (2.3)$$

where $\Delta\nu$ is the total laser linewidth and T_s is the symbol duration. The phase noise $\boldsymbol{\theta}$ is unknown both to the receiver and the transmitter, and is independent of the transmitted symbols as well as the AWGN. The constant phase offset on \mathbf{r}_x and \mathbf{r}_y is encapsulated in the random variables C_x and C_y , drawn from a uniform distribution in the interval $[0, 2\pi)$.

2.2 MAP Symbol Detection

In this section the FG framework and the SPA are used to realize the optimal symbol detector. As mentioned in Section 1.3, the detector that minimizes the SER for an uncoded transmission is the MAP symbol detector. This detector is performed

symbol-by-symbol according to

$$\hat{s}_{w,k} = \operatorname{argmax}_{s_{w,k} \in \mathcal{M}} p(s_{w,k} | \mathbf{r}_x, \mathbf{r}_y), \quad (2.4)$$

for $k = 0, 1, \dots, N - 1$ and $w \in \{\mathbf{x}, \mathbf{y}\}$. The indices \mathbf{x} and \mathbf{y} refer to the polarization components, and $p(s_{w,k} | \mathbf{r})$ is the a posteriori probability mass function (PMF) for the k th symbol on polarization w . The detection rule in (2.4) considers all the received samples when decision on each transmitted symbol is made. In [25] it is shown that in the presence of phase noise, the PMF in (2.4) can be written in the form

$$p(s_{w,k} | \mathbf{r}_x, \mathbf{r}_y) \propto \int_0^{2\pi} p(r_{w,k} | s_{w,k}, \theta_k) p(\theta_k | \bar{\mathbf{r}}_{w,k}) d\theta_k, \quad (2.5)$$

where $\bar{\mathbf{r}}_{w,k}$ is the vector of received samples from both polarizations except for $r_{w,k}$, and θ_k is the phase noise at index k . The integral in (2.5) illustrates the fact that the a posteriori phase noise PDF is used in the symbol detection. In [25], it is further shown that this PDF does not have a closed form solution in general, and computing the integral is an infinite dimensional problem.

The PMF in (2.4) can be obtained by marginalizing the joint PDF of all system variables

$$p(s_{x,k} | \mathbf{r}_x, \mathbf{r}_y) = \sum_{s_x \in \mathcal{S}_x} \sum_{s_y \in \mathcal{M}^N} \int_{\mathcal{T}} \int_0^{2\pi} \int_0^{2\pi} p(\mathbf{s}_x, \mathbf{s}_y, \boldsymbol{\theta}, C_x, C_y | \mathbf{r}_x, \mathbf{r}_y) dC_x dC_y d\boldsymbol{\theta}, \quad (2.6)$$

where $\mathcal{S}_x = \{\mathbf{s}'_x \in \mathcal{M}^N : s'_{x,k} = s_{x,k}\}$ and $\mathcal{T} = [0, 2\pi)^N$. The corresponding PMF $p(s_{y,k} | \mathbf{r}_x, \mathbf{r}_y)$ is obtained analogously by swapping \mathbf{x} and \mathbf{y} in (2.6). The computational complexity in (2.6) is unpractical, and the problem can be solved in more efficient ways. In this thesis, the PMF in (2.4) is computed using an FG which is constructed from PDF $p(\mathbf{s}_x, \mathbf{s}_y, \boldsymbol{\theta} | \mathbf{r}_x, \mathbf{r}_y)$, the joint a posteriori PDF of all variables in the system model except for C_x and C_y . Omitting these two variables reduces the complexity of the resulting FG. The justification for this omission stems from the fact that the variables are constant throughout the received sequence. Thus, the phase difference between the polarizations can be estimated and compensated for with high accuracy

in the majority of cases. A constant phase offset is still applied to both polarizations. Because the phase difference is absent, however, the phase offset can be regarded as a component in the phase noise $\boldsymbol{\theta}$.

Looking again at the joint PDF $p(\mathbf{s}, \boldsymbol{\theta} | \mathbf{r})$ and using Bayes' theorem results in¹

$$\begin{aligned} p(\mathbf{s}_x, \mathbf{s}_y, \boldsymbol{\theta} | \mathbf{r}_x, \mathbf{r}_y) &= \frac{p(\mathbf{r}_x, \mathbf{r}_y | \mathbf{s}_x, \mathbf{s}_y, \boldsymbol{\theta}) p(\mathbf{s}_x, \mathbf{s}_y, \boldsymbol{\theta})}{p(\mathbf{r}_x, \mathbf{r}_y)} \\ &\propto p(\mathbf{r}_x, \mathbf{r}_y | \mathbf{s}_x, \mathbf{s}_y, \boldsymbol{\theta}) p(\mathbf{s}_x) p(\mathbf{s}_y) p(\boldsymbol{\theta}) \end{aligned} \quad (2.7)$$

$$\propto p(\boldsymbol{\theta}) p(\mathbf{r}_x | \mathbf{s}_x, \boldsymbol{\theta}) p(\mathbf{r}_y | \mathbf{s}_y, \boldsymbol{\theta}), \quad (2.8)$$

where (2.7) follows from the fact that the transmitted symbols are independent of the phase noise. Moreover, the joint PDF of the received samples, $p(\mathbf{r}_x, \mathbf{r}_y)$, does not depend on \mathbf{s}_x or \mathbf{s}_y , and accordingly it is constant when maximizing with respect to $s_{w,k}$. All constellation points are equiprobable and the transmitted symbols are independent and identically distributed, therefore $p(\mathbf{s}_x)$ and $p(\mathbf{s}_y)$ are constant. Furthermore, when conditioned on $\boldsymbol{\theta}$, the received samples are independent of each other, resulting in (2.8).

The random-walk model for the phase noise process is essentially a first-order Markov chain. Consequently,

$$p(\theta_k | \theta_{k-1}, \dots, \theta_0) = p(\theta_k | \theta_{k-1}) = p_\Delta(\theta_k - \theta_{k-1}), \quad (2.9)$$

for $k = 0, 1, \dots, N - 1$, where $p_\Delta(\cdot)$ is defined as the PDF of the step $\Delta_k \bmod 2\pi$, a wrapped normal distribution expressed by

$$p_\Delta(\lambda) \triangleq \sum_{\tau=-\infty}^{\infty} g(0, \sigma_\Delta^2; \lambda - \tau 2\pi), \quad \lambda \in [0, 2\pi), \quad (2.10)$$

where $g(\mu, \sigma^2; z)$ is a real Gaussian PDF with mean μ , variance σ^2 , and argument z .

¹Since the marginal distributions are eventually maximized, constants in the expressions can be ignored. Therefore the proportionality relationship between expressions suffices and will be used throughout the thesis.

By using (2.9), the joint PDF of $\boldsymbol{\theta}$ can be factored as

$$p(\boldsymbol{\theta}) = p(\theta_0) \prod_{k=1}^{N-1} p_{\Delta}(\theta_k - \theta_{k-1}), \quad (2.11)$$

allowing further decomposition of the joint PDF in (2.8)

$$\begin{aligned} p(\mathbf{s}_x, \mathbf{s}_y, \boldsymbol{\theta} | \mathbf{r}_x, \mathbf{r}_y) &\propto p(\boldsymbol{\theta}) p(\mathbf{r}_x | \mathbf{s}_x, \boldsymbol{\theta}) p(\mathbf{r}_y | \mathbf{s}_y, \boldsymbol{\theta}) \\ &\propto p(\theta_0) \prod_{k=1}^{N-1} p_{\Delta}(\theta_k - \theta_{k-1}) \prod_{k=0}^{N-1} f_{x,k}(s_{x,k}, \theta_k) f_{y,k}(s_{y,k}, \theta_k), \end{aligned} \quad (2.12)$$

for $k = 0, 1, \dots, N-1$ and $w \in \{x, y\}$, where

$$\begin{aligned} f_{w,k}(s_{w,k}, \theta_k) &\triangleq \exp\left\{ \frac{1}{\sigma^2} \operatorname{Re}\{r_{w,k} s_{w,k}^* e^{-j\theta_k}\} - \frac{|s_{w,k}|^2}{2\sigma^2} \right\} \\ &\propto p(r_{w,k} | s_{w,k}, \theta_k). \end{aligned} \quad (2.13)$$

A part of the FG constructed from (2.12) can be seen in Fig. 2.3, along with the SPA messages. The graph is cycle-free, implying that the SPA, when applied to this FG, yields exact a posteriori distributions [29]. Hence, it is optimal as it realizes the exact MAP symbol detector.

For each k and w , let $\vec{P}_w(s_{w,k})$ denote the message from variable node $s_{w,k}$ to factor node $f(s_{w,k}, \theta_k)$. This message corresponds to the a priori probabilities of the transmitted symbols. The message in the reverse direction is denoted by $\overleftarrow{P}_w(s_{w,k})$, and corresponds to the a posteriori symbol probabilities. The message from factor node $f(s_{w,k}, \theta_k)$ to variable node θ_k is denoted by $p_w(\theta_k)$, and is computed as

$$p_w(\theta_k) \propto \sum_{s_{w,k} \in \mathcal{M}} \vec{P}_w(s_{w,k}) f(s_{w,k}, \theta_k) \propto p(r_{w,k} | \theta_k), \quad (2.14)$$

for $k = 0, 1, \dots, N-1$ and $w \in \{x, y\}$. The message from factor node $p_{\Delta}(\theta_k - \theta_{k-1})$ to variable node θ_k is denoted by $p_f(\theta_k)$, and the message from $p_{\Delta}(\theta_{k+1} - \theta_k)$ to variable node θ_k is denoted by $p_b(\theta_k)$. These correspond to a posteriori phase noise PDFs, and are computed in a recursive fashion. Computation of the forward recursion entails the messages $p_f(\theta_k)$ being calculated in a forward direction based on the received samples. Similarly, computing the backward recursion involves calculating

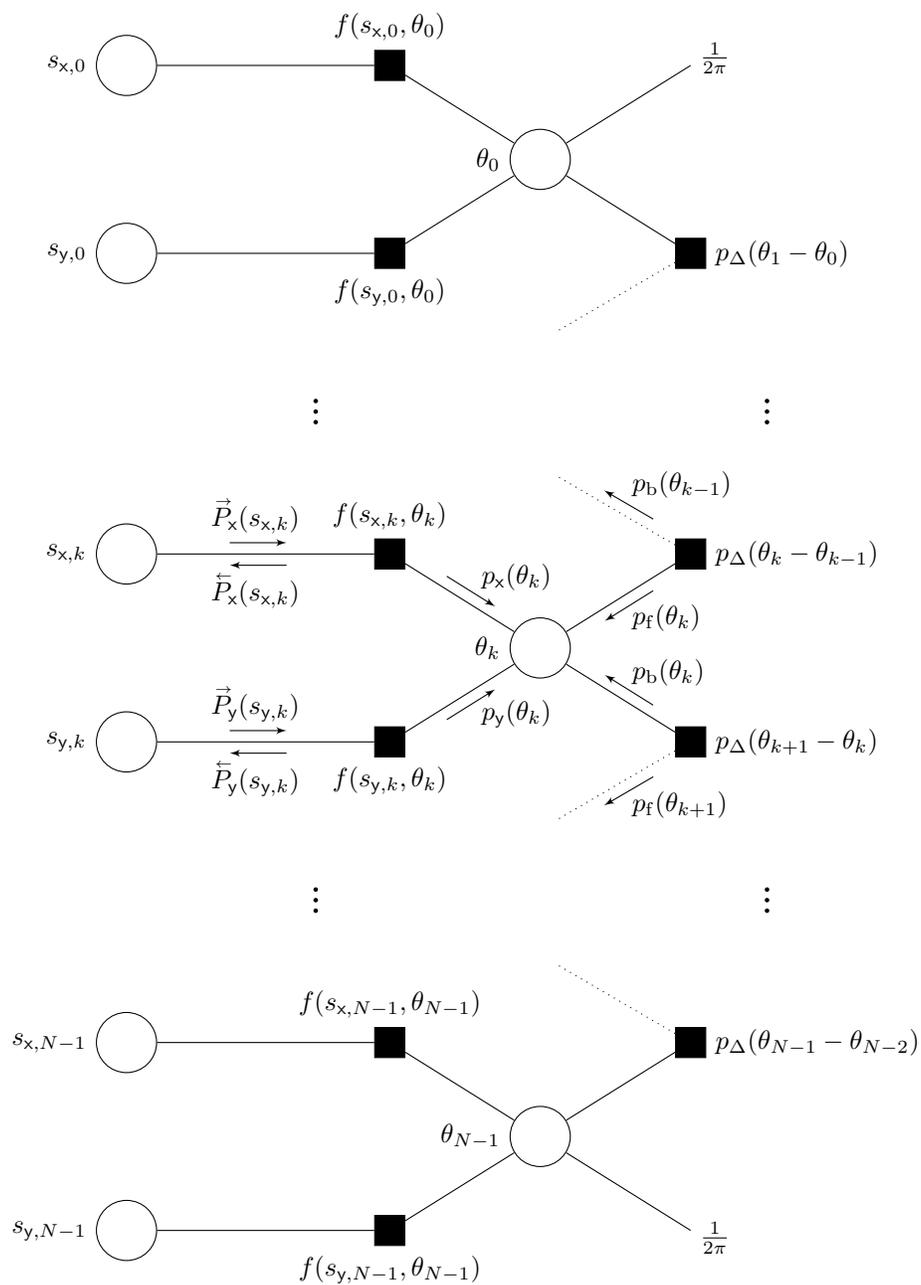


Figure 2.3: Part of FG constructed from the distribution in (2.12).

the messages $p_b(\theta_k)$ in a backward direction based on the received samples. The computation of the messages is expressed as

$$p_f(\theta_k) \propto \int_0^{2\pi} p_x(\theta_{k-1})p_y(\theta_{k-1})p_f(\theta_{k-1})p_\Delta(\theta_k - \theta_{k-1})d\theta_{k-1}, \quad (2.15)$$

$$p_b(\theta_k) \propto \int_0^{2\pi} p_x(\theta_{k+1})p_y(\theta_{k+1})p_b(\theta_{k+1})p_\Delta(\theta_{k+1} - \theta_k)d\theta_{k+1}, \quad (2.16)$$

with initial conditions

$$p_f(\theta_0) = p_b(\theta_{N-1}) = \frac{1}{2\pi}. \quad (2.17)$$

Finally, the messages $\bar{P}_x(s_{x,k})$ and $\bar{P}_y(s_{y,k})$ are computed as

$$\bar{P}_x(s_{x,k}) \propto \int_0^{2\pi} p_f(\theta_k)p_b(\theta_k)p_y(\theta_k)f(s_{x,k}, \theta_k)d\theta_k, \quad (2.18)$$

$$\bar{P}_y(s_{y,k}) \propto \int_0^{2\pi} p_f(\theta_k)p_b(\theta_k)p_x(\theta_k)f(s_{y,k}, \theta_k)d\theta_k, \quad (2.19)$$

for $k = 0, 1, \dots, N - 1$. The messages in (2.18) and (2.19) correspond to the a posteriori PMFs $p(s_{x,k}|\mathbf{r}_x, \mathbf{r}_y)$ and $p(s_{y,k}|\mathbf{r}_x, \mathbf{r}_y)$, respectively. Furthermore, it can be shown that

$$p_f(\theta_k)p_b(\theta_k)p_y(\theta_k) \propto p(\theta_k|\bar{\mathbf{r}}_{x,k}), \quad (2.20)$$

$$p_f(\theta_k)p_b(\theta_k)p_x(\theta_k) \propto p(\theta_k|\bar{\mathbf{r}}_{y,k}), \quad (2.21)$$

where $p(\theta_k|\bar{\mathbf{r}}_{w,k})$, for $w \in \{x, y\}$, is the a posteriori phase noise PDF in (2.5).

The message-passing algorithm can be summarized as follows. The received samples, along with the a priori probabilities of the transmitted symbols, are the inputs to the algorithm. These inputs are used to calculate the message in (2.14). The messages in (2.15) and (2.16), corresponding to the a posteriori phase noise PDFs, are then computed in a recursive fashion. Finally, the messages in (2.18) and (2.19), corresponding to the a posteriori symbol probabilities, are calculated. These messages are the outputs of the algorithm. Symbol detection is performed according to (2.4), by selecting the argument, i.e. the constellation point in \mathcal{M} , that maximizes the messages in (2.18) and (2.19).

3 Receiver Implementation

In the preceding chapter, the optimal symbol detector was realized using the FG framework and the SPA. The messages produced by the SPA, however, are difficult to implement in practice, as they contain integrals. As a result, approximations are required in order to obtain efficient implementations. In this chapter a practical receiver algorithm is presented. First, estimation and compensation, for the phase offset due to the polarization demultiplexing, are described. Afterwards, an efficient algorithm, is derived based on approximations of the results given by the SPA in the preceding section.

3.1 Estimation of Phase Difference

As mentioned in Chapter 2, the variables C_x and C_y represent the phase offset on each polarization components. Moreover, the FG in Fig. 2.3 is constructed from the joint PDF $p(\mathbf{s}_x, \mathbf{s}_y, \boldsymbol{\theta} | \mathbf{r}_x, \mathbf{r}_y)$, with the assumption that \mathbf{r}_x and \mathbf{r}_y are free of the relative phase difference caused by these offsets. If not compensated for, the a posteriori phase noise PDF estimation will not work properly. By using the pilot symbols, however, the relative phase difference between the two polarizations can be estimated.

Let $L \subset \{0, 1, \dots, N - 1\}$ denote the set of indices for the pilots in the transmitted block. The phase difference between the transmitted and the received pilot symbol on polarization $w \in \{x, y\}$ at index $l \in L$ is computed as

$$\begin{aligned}
 r_{w,l} s_{w,l}^* &= s_{w,l} s_{w,l}^* e^{j(\theta_l + C_w)} + n_{w,l} s_{w,l}^* \\
 &= |s_{w,l}|^2 e^{j(\theta_l + C_w)} + |s_{w,l}| n_{w,l} e^{-j \arg\{s_{w,l}\}} \\
 &\propto e^{j(\theta_l + C_w)} + \frac{n_{w,l} e^{-j \arg\{s_{w,l}\}}}{|s_{w,l}|}.
 \end{aligned} \tag{3.1}$$

Rotating the AWGN vectors \mathbf{n}_x and \mathbf{n}_y does not change their statistics as they

contain complex circularly symmetric Gaussian components. The relative phase difference between the polarizations is then

$$\begin{aligned}
r_{x,l}s_{x,l}^*(r_{y,l}s_{y,l}^*)^* &\propto \left(e^{j(\theta_l+C_x)} + \frac{n_{x,l}e^{-j\arg\{s_{x,l}\}}}{|s_{x,l}|} \right) \left(e^{j(\theta_l+C_y)} + \frac{n_{y,l}e^{-j\arg\{s_{y,l}\}}}{|s_{y,l}|} \right)^* \\
&= e^{j(\theta_l+C_x-\theta_l-C_y)} + \frac{n_{y,l}^*e^{j(\theta_l+C_x+\arg\{s_{y,l}\})}}{|s_{y,l}|} \\
&\quad + \frac{n_{x,l}e^{-j(\theta_l+C_y+\arg\{s_{x,l}\})}}{|s_{x,l}|} + \frac{n_{x,l}n_{y,l}^*e^{j(\arg\{s_{y,l}\}-\arg\{s_{x,l}\})}}{|s_{x,l}||s_{y,l}|} \\
&= \rho + m_l,
\end{aligned} \tag{3.2}$$

where $\rho = e^{jC_d} = e^{C_x-C_y}$, and m_l includes a sum of terms contributed by the noise. It can be shown that m_l is drawn from a zero-mean, non-Gaussian distribution. Therefore, ρ can be estimated by the sample mean

$$\hat{\rho} = \frac{1}{|L|} \sum_{l \in L} r_{x,l}s_{x,l}^*r_{y,l}^*s_{y,l}. \tag{3.3}$$

It is worth noting, however, that since the parameter of interest is the argument of ρ , the sample sum suffices, as scaling a complex number does not affect its argument. Rotating $r_{y,k}$ with $\arg\{\hat{\rho}\}$ gives

$$\begin{bmatrix} r_{x,k} \\ r_{y,k}e^{j\arg\{\hat{\rho}\}} \end{bmatrix} \approx \begin{bmatrix} s_{x,k} \\ s_{y,k} \end{bmatrix} e^{j\tilde{\theta}_k} + \begin{bmatrix} n_{x,k} \\ n_{y,k}e^{jC_d} \end{bmatrix}, \tag{3.4}$$

where $\tilde{\theta}_k = (\theta_k + C_x) \bmod 2\pi$, for $k = 0, 1, \dots, N-1$. As seen in (3.4), the relative phase difference has been approximately eliminated, albeit the received samples on both polarizations are still affected by a phase offset. The phase noise $\tilde{\theta}$ can be regarded as a shifted version of the original phase noise θ , with unchanged statistics.

3.2 Phase Noise Estimation – Symbol Detection

In order to derive an efficient algorithm, the approach in [26] is adopted, namely selecting parametrized canonical distributions to approximate messages involving continuous PDFs [32]. Hence, the messages $p_f(\theta_k)$ and $p_b(\theta_k)$ are constrained to be

in a family of Tikhonov PDFs. This PDF, also known as a von Mises or a circular normal distribution [33, p. 35], is defined as

$$t(\kappa; \theta) = \frac{1}{2\pi I_0(|\kappa|)} \exp\{\operatorname{Re}\{\kappa e^{-j\theta}\}\}, \quad \theta \in [0, 2\pi), \quad (3.5)$$

where $\kappa \in \mathbb{C}$, and $I_0(\cdot)$ is the modified Bessel function of the first kind and zeroth order. The argument and the modulus of κ characterize the mean and the variance, respectively, of the PDF in (3.5); the Tikhonov PDFs are completely described by a single parameter. The direct computation of the integrals in (2.15) and (2.16) thus reduces to updating the parameters describing the Tikhonov PDFs.

A closer look at the message in (2.14) reveals it to be a mixture of Gaussian PDFs. As in [26], the message is approximated by the closest Gaussian PDF in terms of the Kullback-Leibler divergence, a measure of the similarity between two distributions. The mean and variance of this Gaussian PDF are $\mathbb{E}[r_{w,k}|\theta_k]$ and $\operatorname{Var}[r_{w,k}|\theta_k]$, respectively, for $k = 0, 1, \dots, N-1$ and $w \in \{\mathbf{x}, \mathbf{y}\}$, and are expressed as

$$\begin{aligned} \mathbb{E}[r_{w,k}|\theta_k] &= \sum_{s_{w,k} \in \mathcal{M}} \mathbb{E}[r_{w,k}|s_{w,k}, \theta_k] \vec{P}_w(s_{w,k}) \\ &= \sum_{s_{w,k} \in \mathcal{M}} s_{w,k} \vec{P}_w(s_{w,k}) e^{j\theta_k} \\ &= \alpha_{w,k} e^{j\theta_k}, \end{aligned} \quad (3.6)$$

and

$$\begin{aligned} \operatorname{Var}[r_{w,k}|\theta_k] &= \mathbb{E}[|r_{w,k}|^2|\theta_k] - |\alpha_{w,k}|^2 \\ &= \sum_{s_{w,k} \in \mathcal{M}} \mathbb{E}[|r_{w,k}|^2|s_{w,k}, \theta_k] \vec{P}_w(s_{w,k}) - |\alpha_{w,k}|^2 \\ &= \sum_{s_{w,k} \in \mathcal{M}} \left(\operatorname{Var}[r_{w,k}|s_{w,k}, \theta_k] + |s_{w,k}|^2 \right) \vec{P}_w(s_{w,k}) - |\alpha_{w,k}|^2 \\ &= 2\sigma^2 + \sum_{s_{w,k} \in \mathcal{M}} |s_{w,k}|^2 \vec{P}_w(s_{w,k}) - |\alpha_{w,k}|^2, \\ &= 2\sigma^2 + \beta_{w,k} - |\alpha_{w,k}|^2, \end{aligned} \quad (3.7)$$

where $\alpha_{w,k}$ and $\beta_{w,k}$ represent the first and second moments of $s_{w,k}$, and are expressed

as

$$\alpha_{w,k} = \sum_{s_{w,k} \in \mathcal{M}} s_{w,k} \vec{P}_w(s_{w,k}), \quad (3.8)$$

$$\beta_{w,k} = \sum_{s_{w,k} \in \mathcal{M}} |s_{w,k}|^2 \vec{P}_w(s_{w,k}). \quad (3.9)$$

Using the results from (3.6) and (3.7) gives

$$\begin{aligned} p_w(\theta_k) &\approx g_{\mathbb{C}}(\mathbb{E}[r_{w,k}|\theta_k], \mathbf{Var}[r_{w,k}|\theta_k]; r_{w,k}) \\ &= g_{\mathbb{C}}(\alpha_{w,k} e^{j\theta_k}, 2\sigma^2 + \beta_{w,k} - |\alpha_{w,k}|^2; r_{w,k}) \\ &\propto \exp\left\{ \frac{2 \operatorname{Re}\{r_{w,k} \alpha_{w,k}^* e^{-j\theta_k}\}}{2\sigma^2 + \beta_{w,k} - |\alpha_{w,k}|^2} \right\} \exp\left\{ -\frac{|\alpha_{w,k}|^2 + |r_{w,k}|^2}{2\sigma^2 + \beta_{w,k} - |\alpha_{w,k}|^2} \right\} \\ &\propto t\left(\frac{2r_{w,k} \alpha_{w,k}^*}{2\sigma^2 + \beta_{w,k} - |\alpha_{w,k}|^2}; \theta_k \right), \end{aligned} \quad (3.10)$$

where $g_{\mathbb{C}}(\mu, \sigma^2; z)$ is a complex Gaussian PDF with mean μ , variance σ^2 , and argument z . Furthermore, the messages $p_f(\theta_k)$ and $p_b(\theta_k)$ are constrained to be in a family of Tikhonov PDFs, as

$$p_f(\theta_k) \propto t(a_{f,k}; \theta_k), \quad (3.11)$$

$$p_b(\theta_k) \propto t(a_{b,k}; \theta_k), \quad (3.12)$$

for $k = 0, 1, \dots, N-1$, where $a_{f,k}$ and $a_{b,k}$ denote forward and backward coefficients, respectively. These coefficients can be recursively calculated, thereby allowing for an efficient means of computing the messages in (3.11) and (3.12). To this end, by using (3.10) and (3.11), the message in (2.15) can be simplified as

$$\begin{aligned} p_f(\theta_k) &\propto \int_0^{2\pi} p_x(\theta_{k-1}) p_y(\theta_{k-1}) p_f(\theta_{k-1}) p_{\Delta}(\theta_k - \theta_{k-1}) d\theta_{k-1} \\ &\approx \int_0^{2\pi} \exp\left\{ \frac{2 \operatorname{Re}\{r_{x,k-1} \alpha_{x,k-1}^* e^{-j\theta_{k-1}}\}}{2\sigma^2 + \beta_{x,k-1} - |\alpha_{x,k-1}|^2} \right\} \exp\left\{ \frac{2 \operatorname{Re}\{r_{y,k-1} \alpha_{y,k-1}^* e^{-j\theta_{k-1}}\}}{2\sigma^2 + \beta_{y,k-1} - |\alpha_{y,k-1}|^2} \right\} \\ &\quad \cdot \exp\left\{ \operatorname{Re}\{a_{f,k-1} e^{-j\theta_{k-1}}\} \right\} p_{\Delta}(\theta_k - \theta_{k-1}) d\theta_{k-1} \\ &\approx \int_0^{2\pi} \exp\left\{ \operatorname{Re}\{z_{f,k} e^{-j\theta_{k-1}}\} \right\} p_{\Delta}(\theta_k - \theta_{k-1}) d\theta_{k-1} \\ &\propto \int_0^{2\pi} t(z_{f,k}; \theta_{k-1}) p_{\Delta}(\theta_k - \theta_{k-1}) d\theta_{k-1}, \end{aligned} \quad (3.13)$$

where

$$z_{f,k} = a_{f,k-1} + \frac{2r_{x,k-1}\alpha_{x,k-1}^*}{2\sigma^2 + \beta_{x,k-1} - |\alpha_{x,k-1}|^2} + \frac{2r_{y,k-1}\alpha_{y,k-1}^*}{2\sigma^2 + \beta_{y,k-1} - |\alpha_{y,k-1}|^2}. \quad (3.14)$$

Define the 2π -periodic functions

$$\tilde{t}(\kappa; \eta) \triangleq \frac{1}{2\pi I_0(|\kappa|)} \exp\{\operatorname{Re}\{\kappa e^{-j\eta}\}\}, \quad (3.15)$$

$$\tilde{p}_\Delta(\eta) \triangleq \sum_{\tau=-\infty}^{\infty} g(0, \sigma_\Delta^2; \eta - \tau 2\pi), \quad (3.16)$$

where $\kappa \in \mathbb{C}$ and $\eta \in \mathbb{R}$. Using (3.15) and (3.16), the integral in (3.13) can be rewritten as

$$\begin{aligned} p_f(\theta_k) &\propto \int_0^{2\pi} t(z_{f,k}; \theta_{k-1}) p_\Delta(\theta_k - \theta_{k-1}) d\theta_{k-1} \\ &= \int_0^{2\pi} \tilde{t}(z_{f,k}; \theta_{k-1}) \tilde{p}_\Delta(\theta_k - \theta_{k-1}) d\theta_{k-1} \\ &= \int_{\theta_k - \pi}^{\theta_k + \pi} \tilde{t}(z_{f,k}; \theta_{k-1}) \tilde{p}_\Delta(\theta_k - \theta_{k-1}) d\theta_{k-1}. \end{aligned} \quad (3.17)$$

For practical values of σ_Δ^2 , $\tilde{p}_\Delta(\theta_k - \theta_{k-1})$ in (3.17) is virtually zero everywhere in $[\theta_k - \pi, \theta_k + \pi)$ except in an interval much smaller than 2π , centered around θ_k . Consequently, it can be approximated as a Gaussian PDF

$$\tilde{p}_\Delta(\theta_k - \theta_{k-1}) \approx g(\theta_k, \sigma_\Delta^2; \theta_{k-1}). \quad (3.18)$$

Thus, the integral in (3.17) simplifies to

$$\begin{aligned} p_f(\theta_k) &\propto \int_{\theta_k - \pi}^{\theta_k + \pi} \tilde{t}(z_{f,k}; \theta_{k-1}) g(\theta_k, \sigma_\Delta^2; \theta_{k-1}) d\theta_{k-1} \\ &\propto \int_{\theta_k - \pi}^{\theta_k + \pi} \exp\{\operatorname{Re}\{z_{f,k} e^{-j\theta_{k-1}}\}\} g(\theta_k, \sigma_\Delta^2; \theta_{k-1}) d\theta_{k-1}. \end{aligned} \quad (3.19)$$

The integral in (3.19) can further be approximated as a Tikhonov PDF [26], resulting in

$$p_f(\theta_k) \approx t(a_{f,k}; \theta_k), \quad (3.20)$$

where the forward coefficients $a_{f,k}$, for $k = 1, \dots, N-1$, are computed recursively as

$$a_{f,k} = \frac{z_{f,k}}{1 + \sigma_{\Delta}^2 |z_{f,k}|}, \quad (3.21)$$

with initial condition $a_{f,0} = 0$. Similarly, using (3.10), (3.18), and (3.12), the messages in (2.16) may be approximated as

$$p_b(\theta_k) \approx t(a_{b,k}; \theta_k), \quad (3.22)$$

where the coefficients $a_{b,k}$, for $k = 0, \dots, N-2$, are computed recursively as

$$a_{b,k} = \frac{z_{b,k}}{1 + \sigma_{\Delta}^2 |z_{b,k}|}, \quad (3.23)$$

with initial condition $a_{b,N-1} = 0$, and

$$z_{b,k} = a_{b,k+1} + \frac{2r_{x,k+1}\alpha_{x,k+1}^*}{2\sigma^2 + \beta_{x,k+1} - |\alpha_{x,k+1}|^2} + \frac{2r_{y,k+1}\alpha_{y,k+1}^*}{2\sigma^2 + \beta_{y,k+1} - |\alpha_{y,k+1}|^2}. \quad (3.24)$$

The expressions for $\overleftarrow{P}_w(s_{w,k})$ are calculated using (2.18) and (2.19) for polarization x and y, respectively. In Appendix A it is shown that they can be approximated as

$$\overleftarrow{P}_w(s_{w,k}) \propto I_0(|\xi_{w,k}|) \exp\left\{-\frac{|s_{w,k}|^2}{2\sigma^2}\right\}, \quad (3.25)$$

where

$$\xi_{x,k} = a_{f,k} + a_{b,k} + \frac{2r_{y,k}\alpha_{y,k}^*}{2\sigma^2 + \beta_{y,k} - |\alpha_{y,k}|^2} + \frac{r_{x,k}s_{x,k}^*}{\sigma^2}, \quad (3.26)$$

$$\xi_{y,k} = a_{f,k} + a_{b,k} + \frac{2r_{x,k}\alpha_{x,k}^*}{2\sigma^2 + \beta_{x,k} - |\alpha_{x,k}|^2} + \frac{r_{y,k}s_{y,k}^*}{\sigma^2}. \quad (3.27)$$

The expression in (3.25) can be numerically problematic. A solution to this is derived in Section 3.4.

Using the above approximations, it can be shown that the message products in (2.20) and (2.21), corresponding to the a posteriori phase noise PDFs, simplify to

$$p_f(\theta_k)p_b(\theta_k)p_y(\theta_k) \propto t\left(a_{f,k} + a_{b,k} + \frac{2r_{y,k}\alpha_{y,k}^*}{2\sigma^2 + \beta_{y,k} - |\alpha_{y,k}|^2}; \theta_k\right), \quad (3.28)$$

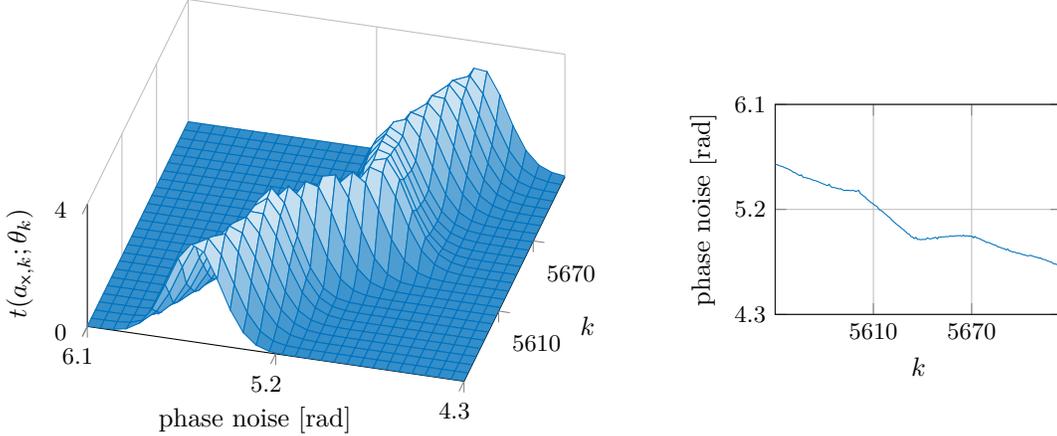


Figure 3.1: 3D plot of $t(a_{x,k}; \theta_k)$, representing the approximate a posteriori phase noise PDF $p(\theta_k | \bar{\mathbf{r}}_{x,k})$. Corresponding point estimates are depicted in the plot on the right.

$$p_f(\theta_k)p_b(\theta_k)p_x(\theta_k) \propto t\left(a_{f,k} + a_{b,k} + \frac{2r_{x,k}\alpha_{x,k}^*}{2\sigma^2 + \beta_{x,k} - |\alpha_{x,k}|^2}; \theta_k\right). \quad (3.29)$$

A graphical illustration of (3.28) can be seen in Fig. 3.1, for a small range of indices k . To this purpose, a short hand notation for (3.28) is defined as

$$t(a_{x,k}; \theta_k) \triangleq t\left(a_{f,k} + a_{b,k} + \frac{2r_{x,k}\alpha_{x,k}^*}{2\sigma^2 + \beta_{x,k} - |\alpha_{x,k}|^2}; \theta_k\right). \quad (3.30)$$

Finding the means of the PDFs at each time k , produces point estimates of the phase noise. A plot of these point estimates is shown in the upper left corner of Fig. 3.1, to emphasize the difference between PDFs and point estimates of the phase noise. As mentioned earlier, using point estimates to compensate for the phase noise does not lead to optimal symbol detection in general. The variance of the a posteriori phase noise PDFs gives information about the uncertainty of the estimates. By compensating for the phase noise using point estimates, as opposed to PDFs, this information is discarded. This has been shown to be suboptimal [25].

3.3 Multiple Iterations

A single iteration of the algorithm is defined as follows. First and second order moments of the transmitted symbols, described in (3.6) and (3.7), are first calculated,

given the a priori symbol probabilities. Afterwards, the a posteriori phase noise PDFs are estimated by way of calculating their forward and backward coefficients recursively with (3.21) and (3.23). Finally, the a posteriori symbol probabilities are calculated with (3.25).

Once the a posteriori symbol probabilities have been calculated, symbol detection can be performed according to (2.4). The first order moment, however, is zero for symbols whose a priori probabilities are equal for all the constellation points. When the first iteration is run, this is the case for all the data symbols on either polarization. If $\alpha_{x,k}$ and $\alpha_{y,k}$ are zero, (3.14) and (3.24) reduce to $z_{f,k} = a_{f,k-1}$ and $z_{b,k} = a_{b,k+1}$, respectively. Thus, the coefficients $a_{f,k}$ and $a_{b,k}$ do not get updated with new information; the data symbols do not contribute to the estimation of the a posteriori phase noise PDF in the first iteration of the algorithm.

The performance of the algorithm can be improved by running multiple iterations. After calculating the message $\overleftarrow{P}_w(s_{w,k})$ in a particular iteration, corresponding to the a posteriori symbol probabilities, it is used as the a priori symbol probabilities in the next iteration, represented by the message $\overrightarrow{P}_w(s_{w,k})$. This can be performed until a stopping criterion is fulfilled. It should be noted that $\overrightarrow{P}_w(s_{w,k})$ is assumed to be a PMF, i.e. the sum of $\overrightarrow{P}_w(s_{w,k})$ over all possible values of $s_{w,k}$ is assumed to be one. The message $\overleftarrow{P}_w(s_{w,k})$ is in general only equal to the a posteriori symbol probabilities up to a scaling factor; the sum of $\overleftarrow{P}_w(s_{w,k})$ over all possible values of $s_{w,k}$ is not one. To be used as the a priori symbol probabilities, it generally has to be scaled properly beforehand. A high level flowchart of the algorithm structure can be seen in Fig. 3.2.

3.4 Numerical problems

The expression in (3.25) can take on large values which is problematic from an implementation standpoint. The modified Bessel function $I_0(\cdot)$, however, can be

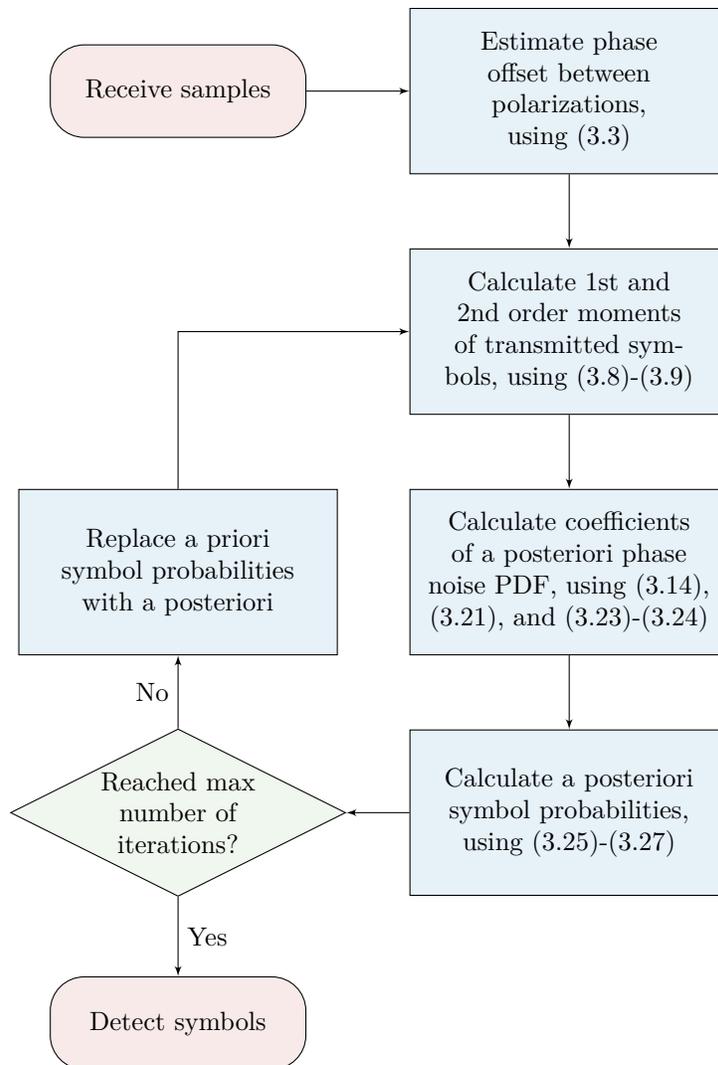


Figure 3.2: A high level flowchart of the algorithm, from receiving the samples to detecting the transmitted symbols, illustrating the iterative structure of the implementation.

approximated with an asymptotic expansion [34, p. 377]

$$I_0(z) \sim \frac{e^z}{\sqrt{2\pi z}}, \quad z \in \mathbb{C}, \quad (3.31)$$

which is a close approximation for $|z| \gg 1$ and $|\arg\{z\}| < \frac{\pi}{2}$. In practice, (3.26) and (3.27) satisfy the former condition with high probability. Moreover, as the argument of $I_0(\cdot)$ in (3.25) is the modulus of a complex number, the latter condition is satisfied.

Define a scaled version of the message in (3.25), which can be simplified as

$$\tilde{\overleftarrow{P}}_w(s_{w,k}) \triangleq A \overleftarrow{P}_w(s_{w,k}) \quad (3.32)$$

$$\approx \frac{A \exp\left\{-\frac{|s_{w,k}|^2}{2\sigma^2} + |\xi_{w,k}|\right\}}{\sqrt{|\xi_{w,k}|}}, \quad (3.33)$$

for $k = 0, 1, \dots, N-1$, $w \in \{x, y\}$, and $A > 0$. The expression in (3.33) is positive, and the natural logarithm is a monotonically increasing function for positive arguments. Hence

$$\begin{aligned} \operatorname{argmax}_{s_{w,k} \in \mathcal{M}} \tilde{\tilde{P}}_w(s_{w,k}) &= \operatorname{argmax}_{s_{w,k} \in \mathcal{M}} \tilde{P}_w(s_{w,k}) \\ &= \operatorname{argmax}_{s_{w,k} \in \mathcal{M}} \log \tilde{P}_w(s_{w,k}) \\ &\approx \operatorname{argmax}_{s_{w,k} \in \mathcal{M}} \left(-\frac{|s_{w,k}|^2}{2\sigma^2} + |\xi_{w,k}| - \frac{1}{2} \log |\xi_{w,k}| \right). \end{aligned} \quad (3.34)$$

The expression that is maximized in (3.34) is not numerically problematic. It represents the logarithm of scaled a posteriori symbol probabilities, and suffices when performing symbol detection. For it to be used as a priori symbol probabilities in a consecutive iteration, however, it has to be transformed to a PMF. This entails reverting the logarithmic transformation and normalizing. In this context, normalizing signifies scaling such that the sum of the expression, over all possible values of $s_{w,k}$, is one. This imposes the constraint

$$\sum_{s_{w,k} \in \mathcal{M}} \tilde{\tilde{P}}_w(s_{w,k}) = A \sum_{s_{w,k} \in \mathcal{M}} \tilde{P}_w(s_{w,k}) = 1. \quad (3.35)$$

Solving (3.35) for A gives

$$A = \left(\sum_{s_{w,k} \in \mathcal{M}} \tilde{P}_w(s_{w,k}) \right)^{-1}. \quad (3.36)$$

Substituting (3.36) in (3.32), and taking the natural logarithm of both sides yields

$$\log \tilde{\tilde{P}}_w(s_{w,k}) = \log \tilde{P}_w(s_{w,k}) - \log \sum_{s_{w,k} \in \mathcal{M}} \tilde{P}_w(s_{w,k}). \quad (3.37)$$

The latter term can be rewritten as

$$\log \sum_{s_{w,k} \in \mathcal{M}} \tilde{P}_w(s_{w,k}) = \log \left\{ \max_{s_{w,k} \in \mathcal{M}} \tilde{P}_w(s_{w,k}) \sum_{s_{w,k} \in \mathcal{M}} \frac{\tilde{P}_w(s_{w,k})}{\max_{s_{w,k} \in \mathcal{M}} \tilde{P}_w(s_{w,k})} \right\}$$

$$\begin{aligned}
&= \log \max_{s_w, k \in \mathcal{M}} \tilde{P}_w(s_w, k) + \log \sum_{s_w, k \in \mathcal{M}} \frac{\tilde{P}_w(s_w, k)}{\max_{s_w, k \in \mathcal{M}} \tilde{P}_w(s_w, k)} \\
&= \max_{s_w, k \in \mathcal{M}} \log \tilde{P}_w(s_w, k) + \log \sum_{s_w, k \in \mathcal{M}} \frac{\exp\{\log \tilde{P}_w(s_w, k)\}}{\exp\{\max_{s_w, k \in \mathcal{M}} \log \tilde{P}_w(s_w, k)\}},
\end{aligned} \tag{3.38}$$

where (3.38) follows from the fact that the natural logarithm is a monotonically increasing function for positive arguments. Using these results, the message in (3.32) can be expressed as a function of $\log \tilde{P}_w(s_w, k)$, and is computed as

$$\begin{aligned}
\tilde{\tilde{P}}_w(s_w, k) = \exp \left\{ \log \tilde{P}_w(s_w, k) - \max_{s_w, k \in \mathcal{M}} \log \tilde{P}_w(s_w, k) \right. \\
\left. - \log \sum_{s_w, k \in \mathcal{M}} \frac{\exp\{\log \tilde{P}_w(s_w, k)\}}{\exp\{\max_{s_w, k \in \mathcal{M}} \log \tilde{P}_w(s_w, k)\}} \right\}.
\end{aligned} \tag{3.39}$$

4 Simulation Results

In this chapter the performance of the algorithm is assessed in terms of BER for different amounts of phase noise, signal to noise ratio (SNR), PD, and number of iterations. The algorithm is evaluated using Monte Carlo simulations for QPSK, 16-QAM, and 64-QAM constellations. The transmitted blocks contain at most 10 000 data symbols, and for each result, errors for each transmission are accumulated until the total number of transmitted bits reaches 10^8 , or if the number of errors exceeds a minimum of 10 000. The bits are Gray encoded before they are mapped to symbols, and no forward error correction coding is used. Pilots are normalized such that their energy is E_s . When simulating a PD of more than 50%, single data symbols are inserted between blocks of pilots. For 50% or lower PDs, single pilots are inserted between blocks of data symbols.

When finding sensitivity penalty (SP) induced by the phase noise, BER values for different amounts of linewidth are compared with an ideal theoretical BER, when only AWGN affects the transmitted signal. This theoretical BER versus SNR per bit is expressed as [35]

$$\text{BER} \approx \frac{\sqrt{M} - 1}{\sqrt{M} \log_2 \sqrt{M}} \operatorname{erfc} \left(\sqrt{\frac{3\gamma \log_2 M}{2(M-1)}} \right) + \frac{\sqrt{M} - 2}{\sqrt{M} \log_2 \sqrt{M}} \operatorname{erfc} \left(3 \sqrt{\frac{3\gamma \log_2 M}{2(M-1)}} \right), \quad (4.1)$$

where M is the number of points in the QAM constellation, erfc is the complementary error function, and γ is the SNR per bit, defined as

$$\gamma = \frac{E_s}{2\sigma^2 \log_2 M}, \quad (4.2)$$

where $\log_2 M$ represents the number of bits per symbol. Using the expression in (4.1), the minimum required γ for a BER of 10^{-3} is approximately 6.79 dB, 10.52 dB, and 14.77 dB for QPSK, 16-QAM, and 64-QAM, respectively. The SP is the additional SNR required, when phase noise is present, to achieve the same BER as the ideal case. When simulating the blind reference algorithms, the transmission blocks con-

Table 4.1: Linewidth tolerance comparison for different algorithms and modulation formats.

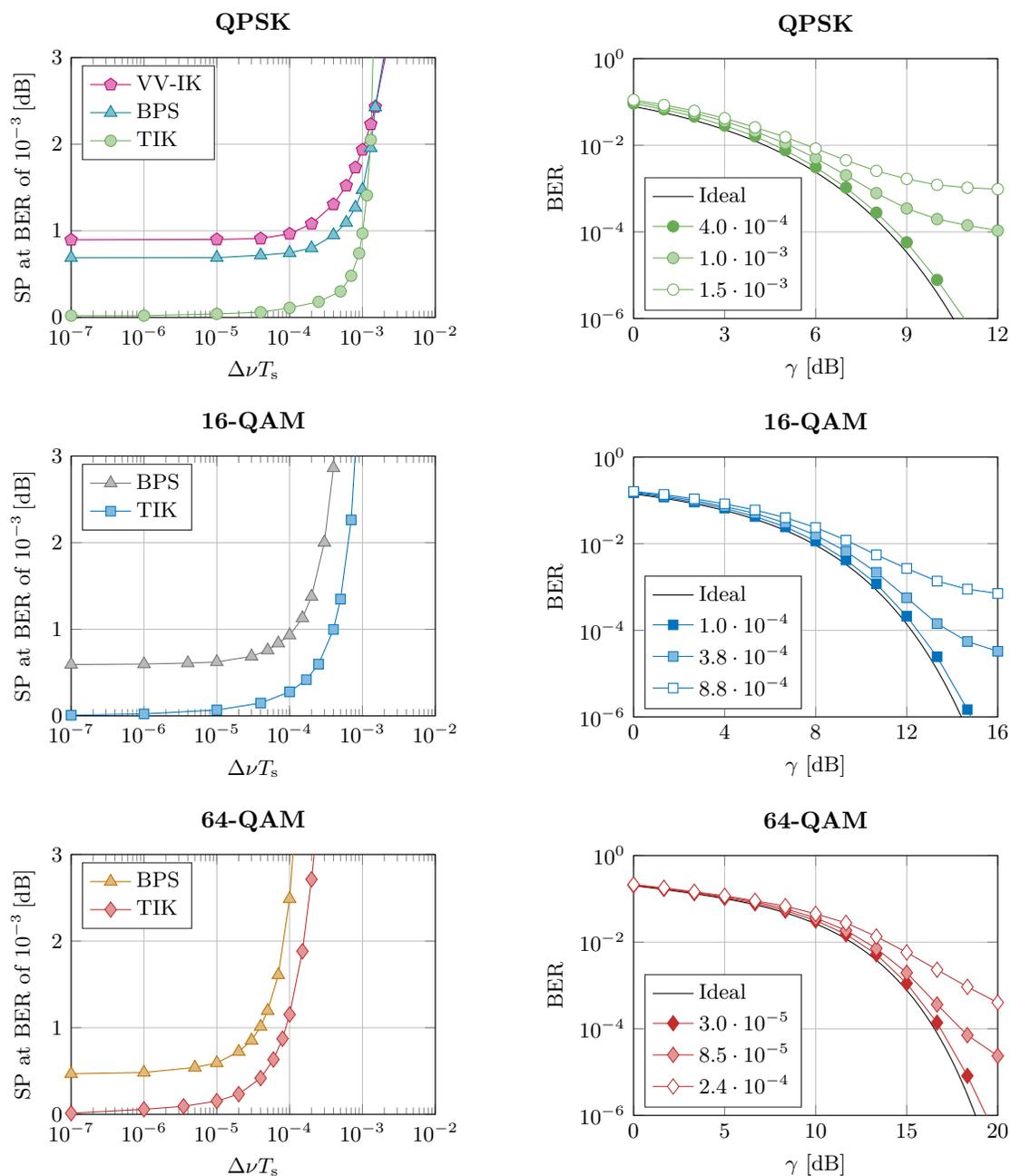
Algorithm	PD	$\Delta\nu T_s$ for 1 dB SP at a BER of 10^{-3}		
		QPSK	16-QAM	64-QAM
TIK (3 it.)	2.86%	$1.21 \cdot 10^{-3}$	$5.74 \cdot 10^{-4}$	$1.33 \cdot 10^{-4}$
TIK (2 it.)	2.86%	$1.01 \cdot 10^{-3}$	$3.98 \cdot 10^{-4}$	$9.02 \cdot 10^{-5}$
PA-Z	2.86%	$9.0 \cdot 10^{-4}$	$4.8 \cdot 10^{-4}$	$6.5 \cdot 10^{-5}$
PA-MO	1.79%	$7.5 \cdot 10^{-4}$	$1.8 \cdot 10^{-4}$	$3.5 \cdot 10^{-5}$
BPS	0%	$4.1 \cdot 10^{-4}$	$1.4 \cdot 10^{-4}$	$4.0 \cdot 10^{-5}$
NLS-CPE	0%	-	$1.5 \cdot 10^{-4}$	$2.7 \cdot 10^{-5}$
VV-IK	0%	$1.6 \cdot 10^{-4}$	-	-

tain 200 000 symbols. Additionally, DE, described in [6], is employed due to phase ambiguity. In the following subsections the proposed algorithm will be referred to as TIK, and the following algorithms are used for comparison: A VV variant (VV-IK) proposed in [8], a blind phase search algorithm (BPS) proposed in [6], a scheme based on nonlinear least squares (NLS-CPE) proposed in [15], and two pilot aided schemes, referred to as PA-MO and PA-Z, proposed in [18] and [20], respectively.

4.1 Linewidth Tolerance

The SP at a BER of 10^{-3} is shown in Fig. 4.1a for TIK running 2 iterations. It is compared to BPS for all modulation formats, and additionally to VV-IK for QPSK. For 16-QAM and 64-QAM transmissions, TIK suffers from the lowest SP when compared to the other considered algorithms, for all of the tested linewidths. For QPSK transmissions, the same is observed except when the linewidth and symbol duration product $\Delta\nu T_s$ surpasses $1.3 \cdot 10^{-3}$. As the linewidth approaches zero, TIK attains close to zero penalty while the blind reference algorithms maintain a minimum penalty. This is due to the DE the blind algorithms use, which increases the average BER.

Table 4.1 shows the maximum reported tolerable $\Delta\nu T_s$ for a 1 dB penalty at a BER of 10^{-3} , for the considered algorithms. The second column shows the required PD



(a) Linewidth tolerance of the algorithm for 2 iterations, 2.86% PD, and different modulation formats. For comparison, VV-IK and BPS are included.

(b) BER performance of the algorithm for 2 iterations, 2.86% PD, different modulation formats, and different values of $\Delta\nu T_s$.

Figure 4.1: Linewidth tolerance and BER performance of the algorithm.

to achieve the tolerance values. After running 2 iterations with a PD of 2.86%, TIK outperforms the compared algorithms for QPSK and 64-QAM transmissions. After 3 iterations, TIK has the best performance for all modulation formats. Additionally, the effect of different amounts of linewidth on the BER for TIK is shown

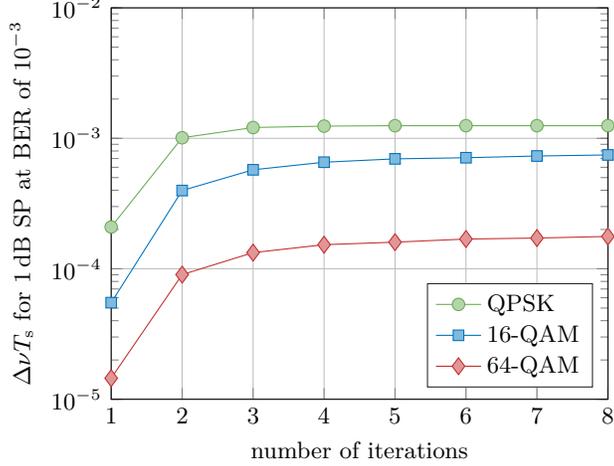


Figure 4.2: Linewidth tolerance of the algorithm for 2.86% PD and a different number of iterations, at a fixed SNR and BER.

in Fig. 4.1b. An ideal theoretical curve computed using (4.1), where only AWGN affects the transmission, is plotted for comparison. For $\Delta\nu T_s$ values of $4.0 \cdot 10^{-4}$, $1.0 \cdot 10^{-4}$, and $3.0 \cdot 10^{-5}$ for QPSK, 16-QAM, and 64-QAM, respectively, BER values close to the ideal case are achieved. For higher linewidths, the phase noise causes the BER curves to hit an error floor as the SNR increases.

4.2 Iterations

To demonstrate the linewidth tolerance improvement, the maximum tolerable $\Delta\nu T_s$ for a 1 dB SP at a BER of 10^{-3} , for a different number of iterations, is depicted in Fig. 4.2. After 8 iterations, tolerable $\Delta\nu T_s$ values of $1.25 \cdot 10^{-3}$, $7.46 \cdot 10^{-4}$, and $1.77 \cdot 10^{-4}$ are obtained for QPSK, 16-QAM, and 64-QAM, respectively. As Fig. 4.2 shows, the performance improves initially with increasing number of iterations. Beyond a particular number of iterations, however, negligible improvement in performance is observed. In the case of QPSK, the improvements are marginal after 3 iterations. For 16-QAM and 64-QAM, the tolerance increases more gradually.

The improvement of the a posteriori phase noise PDF estimation from increasing the number of iterations is shown in Fig. 4.3, by comparing the mean of (3.28) to the true phase noise. If the number of iterations is more than one, the data symbols

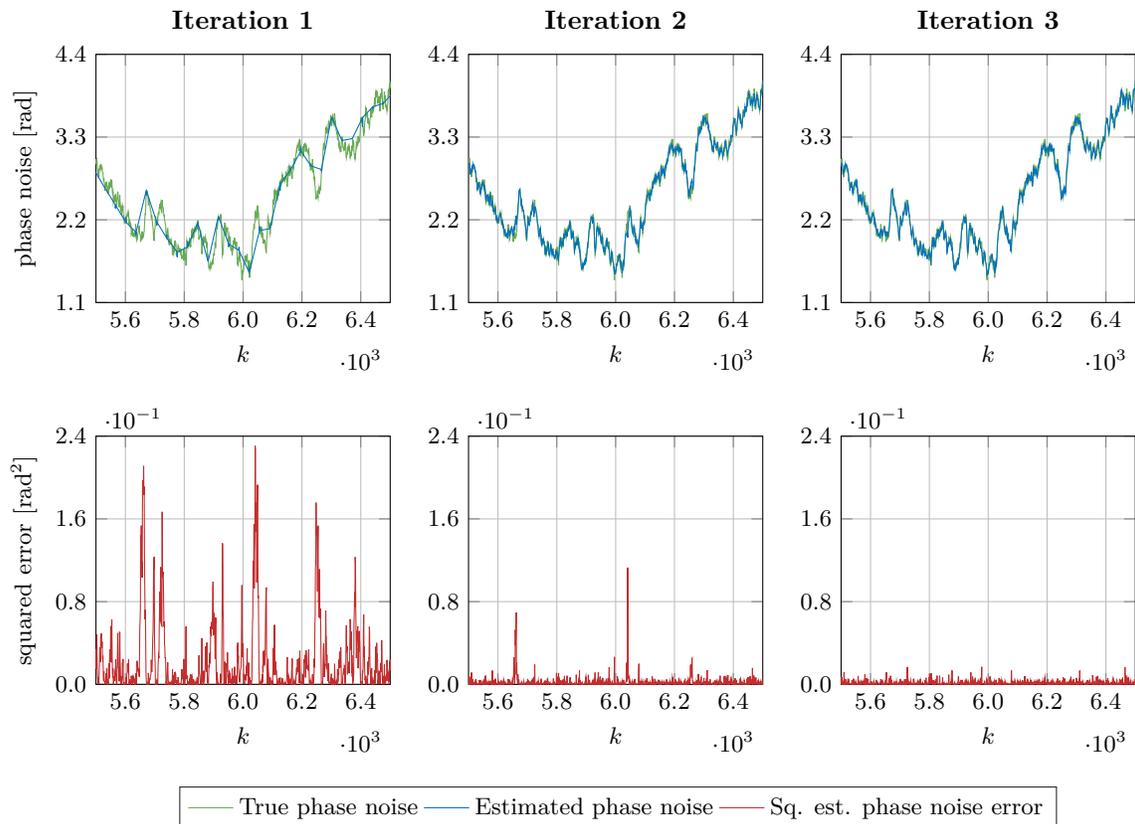


Figure 4.3: Comparison of the true phase noise versus the estimated phase noise (top figures), and the squared error of the estimated phase noise (bottom figures). The figures in the left, middle, and right columns show the results for 1, 2, and 3 iterations, respectively.

are used in the estimation of the coefficients in (3.21) and (3.23). Thus, the mean squared error between the estimated phase noise and the true phase noise decreases.

4.3 Pilot Density

The performance of the algorithm depends significantly on the PD. To get a better understanding of this dependency, the maximum $\Delta\nu T_s$ values for a 1 dB SP at a BER of 10^{-3} , for different PDs and 2 iterations, are plotted in Fig. 4.4. For a low PD the relationship between the linewidth tolerance and the PD is approximately linear. As the density approaches 100%, the tolerance improvements become negligible and reach an upper bound for tolerable $\Delta\nu T_s$ values. This bound is found to be $5.53 \cdot 10^{-3}$, $1.39 \cdot 10^{-3}$, and $3.59 \cdot 10^{-4}$ for QPSK, 16-QAM, and 64-QAM, respectively. The

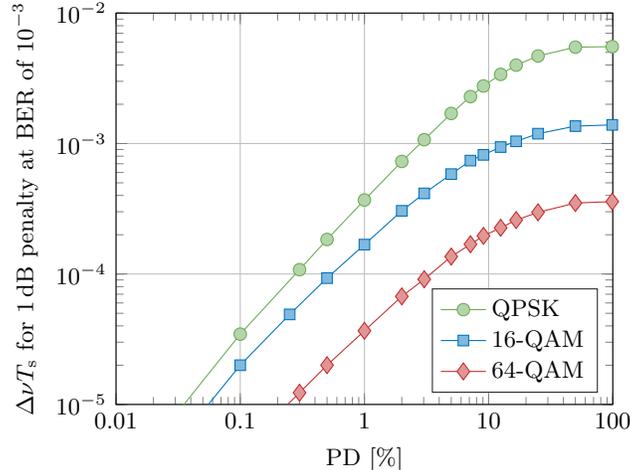


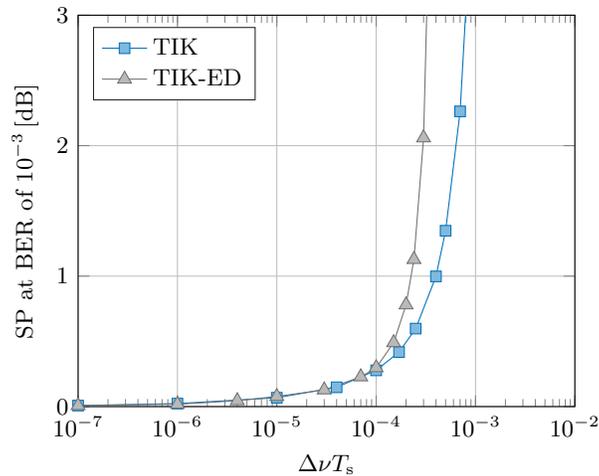
Figure 4.4: Linewidth tolerance of the algorithm for 2 iterations and different PDs, at a fixed SNR and BER.

number of iterations has negligible effect on this bound. However, the number will affect how quickly the bound is approached as the PD is increased.

4.4 Benefits of Phase Noise PDFs

In this section the performance gains of using a posteriori phase noise PDFs in the symbol detection are illustrated for a 16-QAM transmission. For comparison, a suboptimal variation of the proposed algorithm is used, where point estimates of the phase noise are utilized, as opposed to the phase noise PDFs. This variant is described and derived in Appendix B and is referred to as TIK-ED. The two algorithms, TIK and TIK-ED, are compared in Fig. 4.5, in terms of linewidth tolerance and BER performance. All results are obtained from running 2 iterations of the algorithms, with 2.86% PD.

Fig. 4.5a demonstrates the decline in linewidth tolerance when phase noise point estimates are used as opposed to PDFs. As the linewidth increases, TIK-ED is quicker to reach a point where the SP starts growing rapidly. In other words, it is less tolerant to linewidth increments. For lower linewidths, the difference in performance is negligible. Fig. 4.5b shows that as the SNR increases, both algorithms hit an error floor caused by the phase noise. The error floor is higher for TIK-ED, which highlights the



(a) Linewidth tolerance comparison of the algorithms.

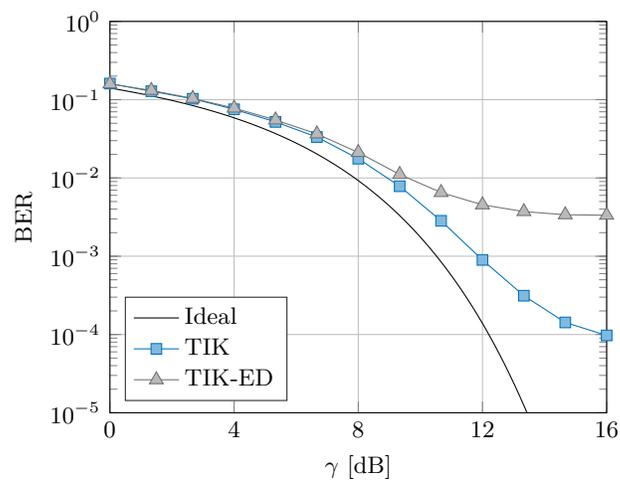
(b) BER comparison of algorithms for $\Delta\nu T_s = 5.0 \cdot 10^{-4}$.

Figure 4.5: Linewidth tolerance and BER performance of the algorithms.

fact that discarding the information contained in the PDFs results in performance penalty. For lower SNRs, the difference between the algorithms is negligible as the AWGN becomes the predominant limitation of the BER performance.

5 Conclusion and Extensions

In this work, the optimal symbol detector in the presence of phase noise was presented, which makes use of phase noise PDFs when performing symbol detection. This is in contrast to the approach most phase noise compensation algorithms in optical communications conform to, namely using point estimates of the phase noise. This detector was realized using the FG framework and the SPA, and was furthermore shown to be analytically intractable. An approach from the wireless communications literature was adopted to implement a practical algorithm. This algorithm is pilot based, and uses observations from both polarizations of the received signal to estimate the phase noise PDFs.

The algorithm's performance was assessed using Monte Carlo simulations. When 8 iterations are run using 2.86% PD, the algorithm has approximately 3.1, 5.3, and 4.4 times the laser linewidth tolerance of the tested blind algorithms, for QPSK, 16-QAM, and 64-QAM, respectively. Moreover, after running 3 iterations using 2.86% PD, the algorithm outperforms other considered pilot based algorithms that use similar PD, in terms of laser linewidth tolerance, for all the tested modulation formats.

Finally, a suboptimal variation of the algorithm was derived, where point estimates are produced from the phase noise PDFs. These point estimates are used to compensate for the phase noise, followed by symbol detection assuming no residual phase noise. Simulations confirm that using point estimates instead of PDFs of the phase noise results in a higher BER, and consequently, lower laser linewidth tolerance.

5.1 Future Work

There are many aspects of the algorithm left to explore, which gives motivation for extensions of this work. Examples of interesting problems are the following.

- Complexity analysis of the algorithm, and comparison with other algorithms from literature.
- Joint phase offset estimation, phase noise estimation and symbol detection. This entails including the variables responsible for the random phase difference between the polarizations in the FG.
- Modeling of the phase difference between the polarizations as a drifting noise process, as opposed to a static difference, and modifying the FG correspondingly.
- Algorithm evaluation using more realistic channel models that include nonlinear phase noise. Additionally, if possible, evaluate using real data, to gain a better insight into its performance in real optical communication systems.
- Investigation of a parallelized implementation of the algorithm, to determine if it is realizable in a receiver of a real optical communication system.
- Joint processing of multiple cores/modes in an optical fiber. Assuming that the same phase noise applies to the different cores/modes, an extension of the algorithm to combine more than two observations is possible. A complexity analysis in this context might be desirable, to determine the algorithm's scalability with increasing number of cores/modes.

Bibliography

- [1] G. P. Agrawal, *Fiber-optic communications systems*, 4th ed. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2010.
- [2] J. Renaudier, R. Rios-Müller, P. Tran, P. L. Schmalen, and G. Charlet, “Spectrally efficient 1-Tb/s transceivers for long-haul optical systems,” *J. Lightw. Technol.*, vol. 33, no. 7, pp. 1452–1458, Apr. 2015.
- [3] X. Liu, S. Chandrasekhar, P. Winzer, T. Lotz, J. Carlson, J. Yang, G. Cheren, and S. Zederbaum, “1.5-Tb/s guard-banded superchannel transmission over 56x 100-km (5600-km) ULAF using 30-Gbaud pilot-free OFDM-16QAM signals with 5.75-b/s/hz net spectral efficiency,” in *European Conf. Optical Communication*, Sep. 2012, paper Th.3.C.5.
- [4] E. Ip, A. P. T. Lau, D. J. F. Barros, and J. M. Kahn, “Coherent detection in optical fiber systems,” *Opt. Exp.*, vol. 16, no. 2, pp. 753–791, Jan. 2008.
- [5] M. G. Taylor, “Coherent detection method using DSP for demodulation of signal and subsequent equalization of propagation impairments,” *IEEE Photonics Technol. Lett.*, vol. 16, no. 2, pp. 674–676, Feb. 2004.
- [6] T. Pfau, S. Hoffmann, and R. Noe, “Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for M-QAM constellations,” *J. Lightw. Technol.*, vol. 27, no. 8, pp. 989–999, Apr. 2009.
- [7] A. J. Viterbi, “Nonlinear estimation of PSK-modulated carrier phase with application to burst digital transmission,” *IEEE Trans. Inf. Theory*, vol. 29, no. 4, pp. 543–551, Jul. 1983.
- [8] E. Ip and J. M. Kahn, “Feedforward carrier recovery for coherent optical communications,” *J. Lightw. Technol.*, vol. 25, no. 9, pp. 2675–2692, Sep. 2007.
- [9] Z. Tao, L. Li, L. Liu, W. Yan, H. Nakashima, T. Tanimura, S. Oda, T. Hoshida, and J. C. Rasmussen, “Improvements to digital carrier phase recovery algorithm

- for high-performance optical coherent receivers,” *IEEE J. Sel. Topics. Quantum Electron.*, vol. 16, no. 5, pp. 1201–1209, Sep. 2010.
- [10] S. Hoffmann, R. Peveling, T. Pfau, O. Adamczyk, R. Eickhoff, and R. Noe, “Multiplier-free real-time phase tracking for coherent QPSK receivers,” *IEEE Photonics Technol. Lett.*, vol. 21, no. 3, pp. 137–139, Feb. 2009.
- [11] F. Rice, M. Rice, and B. Cowley, “A new algorithm for 16QAM carrier phase estimation using QPSK partitioning,” *Dig. Signal Process.*, vol. 12, no. 1, pp. 77–86, Jan. 2002.
- [12] I. Fatadin, D. Ives, and S. J. Savory, “Laser linewidth tolerance for 16-QAM coherent optical systems using QPSK partitioning,” *IEEE Photonics Technol. Lett.*, vol. 22, no. 9, pp. 631–633, May 2010.
- [13] S. M. Bilal, A. Carena, C. Fludger, and G. Bosco, “Dual stage CPE for 64-QAM optical systems based on a modified QPSK-partitioning algorithm,” *IEEE Photonics Technol. Lett.*, vol. 26, no. 3, pp. 267–270, Feb. 2014.
- [14] F. Peng, H. Wymeersch, A. S. Tan, M. Sjodin, P. Johannisson, E. Agrell, P. Andrekson, and M. Karlsson, “Operational regime of symbol-by-symbol phase noise estimation for POLMUX 16-QAM,” in *European Conf. Optical Communication*, Sep. 2010, paper We.7.A.5.
- [15] N. Argyris, S. Dris, C. Spatharakis, and H. Avramopoulos, “High performance carrier phase recovery for coherent optical QAM,” in *Optical Fiber Communication Conf.*, Mar. 2015, paper W1E.1.
- [16] E. Ibragimov, B. Zhang, T. J. Schmidt, C. Malouin, N. Fediakine, and H. Jiang, “Cycle slip probability in 100G PM-QPSK systems,” in *Optical Fiber Communication Conf.*, Mar. 2010, paper OWE2.
- [17] X. Zhou, L. E. Nelson, P. Magill, R. Isaac, B. Zhu, D. W. Peckham, P. I. Borel, and K. Carlson, “High spectral efficiency 400 Gb/s transmission using PDM time-domain hybrid 32-64 QAM and training-assisted carrier recovery,” *J. Lightw. Technol.*, vol. 31, no. 7, pp. 999–1005, Apr. 2013.

-
- [18] M. Morsy-Osman, Q. Zhuge, L. R. Chen, and D. V. Plant, "Feedforward carrier recovery via pilot-aided transmission for single-carrier systems with arbitrary M-QAM constellations," *Opt. Exp.*, vol. 19, no. 24, pp. 24 331–24 343, Nov. 2011.
- [19] M. Magarini, L. Barletta, A. Spalvieri, F. Vacondio, T. Pfau, M. Pepe, M. Bertolini, and G. Gavioli, "Pilot-symbols-aided carrier-phase recovery for 100-G PM-QPSK digital coherent receivers," *IEEE Photonics Technol. Lett.*, vol. 24, no. 9, pp. 739–741, May. 2012.
- [20] F. Zhang, Y. Li, J. Wu, W. Li, X. Hong, and J. Lin, "Improved pilot-aided optical carrier phase recovery for coherent M-QAM," *IEEE Photonics Technol. Lett.*, vol. 24, no. 18, pp. 1577–1580, Sep. 2012.
- [21] H. Cheng, Y. Li, F. Zhang, J. Wu, J. Lu, G. Zhang, J. Xu, and J. Lin, "Pilot-symbols-aided cycle slip mitigation for DP-16QAM optical communication systems," *Opt. Exp.*, vol. 21, no. 19, pp. 22 166–22 172, Sep. 2013.
- [22] M. Morsy-Osman, Q. Zhuge, M. Chagnon, X. Xu, and D. V. Plant, "Experimental demonstration of pilot-aided polarization recovery, frequency offset and phase noise mitigation," in *Optical Fiber Communication Conf.*, Mar. 2013, paper OTu3I.6.
- [23] M. H. Morsy-Osman, L. R. Chen, and D. V. Plant, "Joint mitigation of laser phase noise and fiber nonlinearity using pilot-aided transmission for single-carrier systems," in *37th European Conf. Exposition Optical Communications*. Optical Society of America, 2011, paper Tu.3.A.3.
- [24] J. G. Proakis, *Digital communications*, 4th ed. Boston, MA, USA: McGraw-Hill, 2000.
- [25] P. Y. Kam, S. S. Ng, and T. S. Ng, "Optimum symbol-by-symbol detection of uncoded digital data over the Gaussian channel with unknown carrier phase," *IEEE Trans. Commun.*, vol. 42, no. 8, pp. 2543–2552, Aug. 1994.
- [26] G. Colavolpe, A. Barbieri, and G. Caire, "Algorithms for iterative decoding in the presence of strong phase noise," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 9, pp. 1748–1757, Sep. 2005.

-
- [27] F. Simoens, D. Duyck, H. Çırpan, E. Panayırçı, and M. Moeneclaey, “Monte carlo solutions for blind phase noise estimation,” *EURASIP J. Wireless Commun. Netw.*, vol. 2009, pp. 1–11, Jan. 2009.
- [28] M. Nissilä and S. Pasupathy, “Adaptive iterative detectors for phase-uncertain channels via variational bounding,” *IEEE Trans. Commun.*, vol. 57, no. 3, pp. 716–725, Mar. 2009.
- [29] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, “Factor graphs and the sum-product algorithm,” *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
- [30] D. N. Godard, “Self-recovering equalization and carrier tracking in two-dimensional data communication systems,” *IEEE Trans. Commun.*, vol. 28, no. 11, pp. 1867–1875, Nov. 1980.
- [31] K. Kikuchi, “Performance analyses of polarization demultiplexing based on constant-modulus algorithm in digital coherent optical receivers,” *Opt. Exp.*, vol. 19, no. 10, pp. 9868–9880, May 2011.
- [32] A. P. Worthen and W. E. Stark, “Unified design of iterative receivers using factor graphs,” *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 843–849, Feb. 2001.
- [33] S. R. Jammalamadaka and A. Sengupta, *Topics in Circular Statistics*, ser. Multivariate analysis. River Edge, NJ, USA: World Scientific, 2001.
- [34] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, 10th ed., ser. Applied mathematics. DC, USA: U.S. Government Printing Office, 1972.
- [35] K. Cho and D. Yoon, “On the general BER expression of one- and two-dimensional amplitude modulations,” *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1074–1080, Jul. 2002.

A Message Derivations

A closed form solution of the messages corresponding to the a posteriori symbol probabilities in Section 3.2 is derived. Using (2.18), the message $\overleftarrow{P}_x(s_{x,k})$, for $k = 0, 1, \dots, N-1$, can be simplified as

$$\begin{aligned}
\overleftarrow{P}_x(s_{x,k}) &\propto \int_0^{2\pi} p_f(\theta_k) p_b(\theta_k) p_y(\theta_k) f(s_{x,k}, \theta_k) d\theta_k \\
&\approx \int_0^{2\pi} t(a_{f,k}; \theta_k) t(a_{b,k}; \theta_k) t\left(\frac{2r_{y,k}\alpha_{y,k}^*}{2\sigma^2 + \beta_{y,k} - |\alpha_{y,k}|^2}; \theta_k\right) \\
&\quad \cdot \exp\left\{\frac{1}{\sigma^2} \operatorname{Re}\{r_{x,k}s_{x,k}^* e^{-j\theta_k}\} - \frac{|s_{x,k}|^2}{2\sigma^2}\right\} d\theta_k \\
&\propto \int_0^{2\pi} \exp\{\operatorname{Re}\{\xi_{x,k} e^{-j\theta_k}\}\} \exp\left\{-\frac{|s_{x,k}|^2}{2\sigma^2}\right\} d\theta_k \\
&= 2\pi I_0(|\xi_{x,k}|) \exp\left\{-\frac{|s_{x,k}|^2}{2\sigma^2}\right\} \int_0^{2\pi} t(\xi_{x,k}; \theta_k) d\theta_k \\
&\propto I_0(|\xi_{x,k}|) \exp\left\{-\frac{|s_{x,k}|^2}{2\sigma^2}\right\}, \tag{A.1}
\end{aligned}$$

where (A.1) stems from the fact that an integral of a PDF over its entire support is one, and

$$\xi_{x,k} \triangleq a_{f,k} + a_{b,k} + \frac{2r_{y,k}\alpha_{y,k}^*}{2\sigma^2 + \beta_{y,k} - |\alpha_{y,k}|^2} + \frac{r_{x,k}s_{x,k}^*}{\sigma^2}. \tag{A.2}$$

Similarly, using (2.19), the message $\overleftarrow{P}_y(s_{y,k})$, for $k = 0, 1, \dots, N-1$, can be simplified as

$$\begin{aligned}
\overleftarrow{P}_y(s_{y,k}) &\propto \int_0^{2\pi} p_f(\theta_k) p_b(\theta_k) p_x(\theta_k) f_{y,k}(s_{y,k}, \theta_k) d\theta_k \\
&\approx I_0(|\xi_{y,k}|) \exp\left\{-\frac{|s_{y,k}|^2}{2\sigma^2}\right\}, \tag{A.3}
\end{aligned}$$

where

$$\xi_{y,k} \triangleq a_{f,k} + a_{b,k} + \frac{2r_{x,k}\alpha_{x,k}^*}{2\sigma^2 + \beta_{x,k} - |\alpha_{x,k}|^2} + \frac{r_{y,k}s_{y,k}^*}{\sigma^2}. \tag{A.4}$$

B Variant of Proposed Algorithm

A suboptimal variation of the proposed algorithm is derived for evaluation in Section 4.4, where point estimates of the phase noise are used, as opposed to phase noise PDFs. These point estimates are obtained by finding the mean of the Tikhonov PDFs in (3.28) and (3.29), i.e. by finding the argument of the complex parameter, describing the PDFs, expressed as

$$\hat{\theta}_{x,k} = \arg \left\{ a_{f,k} + a_{b,k} + \frac{2r_{y,k}\alpha_{y,k}^*}{2\sigma^2 + \beta_{y,k} - |\alpha_{y,k}|^2} \right\}, \quad (\text{B.1})$$

$$\hat{\theta}_{y,k} = \arg \left\{ a_{f,k} + a_{b,k} + \frac{2r_{x,k}\alpha_{x,k}^*}{2\sigma^2 + \beta_{x,k} - |\alpha_{x,k}|^2} \right\}, \quad (\text{B.2})$$

for $k = 0, 1, \dots, N - 1$. The point estimates in (B.1) are used for the received samples on polarization \mathbf{x} , whereas the point estimates in (B.2) are used for the received samples on polarization \mathbf{y} . The estimates are treated as the true values of the phase noise, which corresponds to the case where the a posteriori phase noise PDFs in (2.5) are Dirac delta functions. A Dirac delta function is defined as

$$\delta(i) \triangleq \begin{cases} 0, & i \neq 0 \\ \infty, & i = 0 \end{cases}, \quad (\text{B.3})$$

where $i \in \mathbb{R}$. In this case, the symbol detection in (2.4) reduces to

$$\begin{aligned} \operatorname{argmax}_{s_{w,k} \in \mathcal{M}} p(s_{w,k} | \mathbf{r}_x, \mathbf{r}_y) &\approx \operatorname{argmax}_{s_{w,k} \in \mathcal{M}} \int_0^{2\pi} p(r_{w,k} | s_{w,k}, \theta_k) \delta(\theta_k - \hat{\theta}_{w,k}) d\theta_k \\ &= \operatorname{argmax}_{s_{w,k} \in \mathcal{M}} p(r_{w,k} | s_{w,k}, \hat{\theta}_{w,k}) \\ &= \operatorname{argmax}_{s_{w,k} \in \mathcal{M}} \log p(r_{w,k} | s_{w,k}, \hat{\theta}_{w,k}) \\ &= \operatorname{argmin}_{s_{w,k} \in \mathcal{M}} \left| r_{w,k} - s_{w,k} e^{j\hat{\theta}_{w,k}} \right|^2 \\ &= \operatorname{argmin}_{s_{w,k} \in \mathcal{M}} \left| e^{-j\hat{\theta}_{w,k}} \left(r_{w,k} - s_{w,k} e^{j\hat{\theta}_{w,k}} \right) \right|^2 \end{aligned} \quad (\text{B.4})$$

$$= \operatorname{argmin}_{s_{w,k} \in \mathcal{M}} \left| r_{w,k} e^{-j\hat{\theta}_{w,k}} - s_{w,k} \right|^2, \quad (\text{B.5})$$

for $k = 0, 1, \dots, N - 1$ and $w \in \{\mathbf{x}, \mathbf{y}\}$, where (B.4) comes from the fact that rotating a complex number does not affect its magnitude. Thus, the received symbols are derotated with the point estimates, followed by symbol detection. The detection is performed symbol-by-symbol by minimizing the Euclidean distance between the received sample and a transmitted symbol, assuming no residual phase noise.