# Joint-Polarization Phase-Noise Estimation and Symbol Detection for Optical Coherent Receivers

Arni F. Alfredsson, *Student Member, IEEE*, Rajet Krishnan, and Erik Agrell, *Senior Member, IEEE*

***Abstract*—The problem of optimal symbol detection in the presence of laser phase noise is studied, for uncoded polarization-multiplexed fiber-optic transmission. To this end, the maximum *a posteriori* (MAP) symbol detector is presented. Specifically, it is emphasized that obtaining phase-noise point estimates, and treating them as the true values of the phase noise, is in general suboptimal. Furthermore, a pilot-based algorithm that approximates the MAP symbol detector is developed, using approaches adopted from the wireless literature. The algorithm performs joint-polarization phase-noise estimation and symbol detection, for arbitrary modulation formats. Through Monte Carlo simulations, the algorithm is compared to existing solutions from the optical communications literature. It is demonstrated that joint-polarization processing can significantly improve upon the single-polarization case, with respect to linewidth tolerance. Finally, it is shown that with less than 3% pilot overhead the algorithm can be used with lasers having up to 6 times larger linewidths than the most well-performing blind algorithms can tolerate.**

***Index Terms*—Coherent detection, digital signal processing, optical fiber communication, phase noise.**

## I. Introduction

PHASE noise is a critical issue for higher-order modulation format transmission in coherent optical communication systems. As the phase of the optical carrier is used to convey information in coherent transmission, phase recovery at the receiver is imperative for successful symbol detection. Phase noise can severely degrade the system performance, and in order to scale systems to higher-order modulation formats, a well-performing phase-noise estimation is of great importance.

For signal demodulation, a local oscillator (LO) laser at the receiver acts as a phase reference to the laser at the transmitter. As these lasers have finite linewidths [1], their relative phase drifts with time and results in a phenomenon commonly referred to as phase noise. Effects due to the fiber Kerr nonlinearity will contribute to the phase noise as well, but the severity of nonlinear phase noise is highly dependent on, e.g., the symbol rate, the number of wavelength-division multiplexed channels, and the channel power. The focus in this paper is on systems where laser phase noise is dominant, and nonlinear phase noise will not be considered.

Manuscript received April 14, 2016; revised June 6, 2016 and July 5, 2016; accepted July 12, 2016. Date of publication July 21, 2016; date of current version September 6, 2016.

A. F. Alfredsson and E. Agrell are with the Department of Signals and Systems, Chalmers University of Technology, Göteborg SE-41296, Sweden (e-mail: arnia@chalmers.se; agrell@chalmers.se).

R. Krishnan is with CommSig AB, Göteborg SE-41296, Sweden (e-mail: rajet.krishnan@huawei.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2016.2593981

A classic feedback-based method for estimating and mitigating phase noise is a phase-locked loop [1]. Delays in feedback loops, however, decrease the effectiveness of earlier phase-noise estimates used for feedback. Coherent optical systems utilize extensive parallelization and pipelining to achieve the required throughput. In [2] it is shown that this leads to large amounts of feedback delay, making feedback-based methods, either hardware-based techniques or digital signal processing (DSP) algorithms, impractical for these systems.

Substantial research has been done for feedforward and blind phase-noise compensation. Feedforward algorithms do not require feedback loops, and blind algorithms do not use any pilot symbols which are known to the receiver. Examples of blind feedforward algorithms are the Viterbi–Viterbi [3] algorithm and its variations [4]–[6], which are developed for phase-shift keying modulation. These algorithms will perform poorly for higher-order quadrature amplitude modulation (QAM) formats, as they assume equal phase spacing between constellation points. Algorithms based on the Viterbi–Viterbi algorithm and quadrature phase-shift keying (QPSK) partitioning [7], which improve upon the original Viterbi–Viterbi algorithm for QAM, are presented in [8], [9]. Furthermore, among the well-known feedforward algorithms designed for higher-order QAM is the blind phase search (BPS) algorithm [2]. Other approaches have been proposed for QAM, some reporting less complexity than the BPS algorithm with similar linewidth tolerance [10]–[12].

Blind phase-noise compensation algorithms have ambiguities in the phase-noise estimation, owing to rotational symmetries in constellations and lack of absolute phase references. This may cause cycle slips which can further cause catastrophic failures. This is commonly mitigated through differential encoding, but it has the downside that the average bit error rate (BER) will be higher [4]. Pilot-based solutions have been suggested, e.g., to estimate phase noise without ambiguity [13]–[16], or to reduce the probability of cycle slips occurring after blind phase-noise estimation at the receiver [17]. Using pilots eliminates the need for differential encoding, resulting in a lower average BER. In addition, modulation transparency is generally easier to achieve when pilots are utilized [18]. Moreover, pilots can be reused to perform other tasks in addition to phase-noise estimation, such as polarization demultiplexing, frequency offset estimation, and fiber nonlinearity compensation [19]. However, pilots introduce overhead to the transmission, which results in lower information data rate than in the absence of pilots.

The majority of the approaches mentioned above process each polarization-multiplexed (PM) channel separately. However, as both channels generally share the same transmitter and LO laser, they experience the same laser phase noise. It is

0733-8724 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

demonstrated in [17], [20]–[22] that performing joint-polarization processing improves the phase-noise estimation, motivating the development of algorithms that exploit both channels jointly.

Detection algorithms in uncoded digital communication systems are designed such that the symbol error rate (SER) performance of the system is minimized. The detector that achieves this is the maximum *a posteriori* (MAP) symbol detector [23, p. 254]. In [24] it is shown that for transmission in the presence of phase noise and additive white Gaussian noise (AWGN), this detector solves the joint problem of phase-noise estimation and symbol detection. Furthermore, it has a particular structure. When detecting a symbol, the received samples, corresponding to the rest of the symbols, are first used to estimate the *a posteriori* probability density function (PDF) of the phase noise. This PDF is then used when detecting the symbol. In other words, the MAP symbol detector treats the phase noise as a nuisance parameter [25, p. 328]. Hence, it does not produce explicit phase-noise point estimates. In optical communications, receiver algorithms generally do not conform to this structure, but rather perform explicit phase-noise compensation, followed by detection assuming the signal is devoid of any residual phase noise.

The biggest challenge with realizing the MAP symbol detector, however, is computing the *a posteriori* phase-noise PDF, as it turns out to be an infinite-dimensional problem in general [24]. This problem has been addressed in the wireless literature [26]–[28] as oscillator phase noise is present in wireless systems, albeit not as severe as laser phase noise in fiber-optic systems. Furthermore, in recent years, advances in the electronic hardware in optical systems have spurred the application of DSP algorithms from wireless to optical communications. A recent example is [29], where the authors borrow an algorithm from the wireless literature and demonstrate its excellent performance for single-polarization coherent fiber-optic transmission, in the presence of phase noise and frequency fluctuations. The optimal receiver structure in [24], along with the presence of laser phase noise in PM fiber-optic transmission and the existence of phase noise compensation solutions in the wireless literature, forms the main motivations for this work.

In this paper, we first present the MAP symbol detector for PM transmission over an optical phase-noise channel and show that it uses data on both polarizations to solve the joint problem of laser-phase noise estimation and symbol detection, assuming identical phase noise on the polarizations. Secondly, we introduce a pilot-aided method for estimating and compensating for a constant phase offset between the polarizations. Thirdly, we exploit the wireless literature to address the problem of realizing an efficient implementation of the MAP symbol detector. Specifically, we follow the approach proposed by Colavolpe *et al.* in [26] to derive a pilot-based implementation that approximates the detector. The resulting algorithm performs phase-noise estimation and symbol detection, jointly for both polarizations, for the first time in fiber-optical communications. Through Monte Carlo simulations we demonstrate that joint-polarization processing has significant benefits. Moreover, we compare the proposed algorithm with conventional receivers, which involve a blind or pilot-based algorithm that finds point estimates of the phase noise and compensates for it using the estimates. The simulation results show that the proposed algorithm outperforms existing solutions by a wide margin, requiring low pilot overhead to achieve the superior performance.

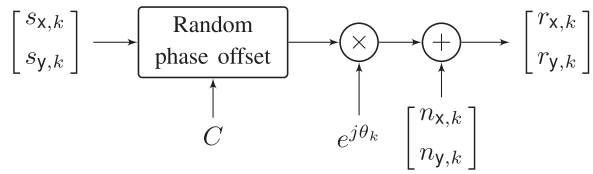


Fig. 1. Discrete-time complex baseband model.

The paper is organized as follows. In Section II, the system model is presented. Section III presents a review of the optimal symbol detection for PM transmission affected by AWGN and phase noise. Furthermore, the factor-graph (FG) framework is applied to the problem of realizing the MAP symbol detector, which allows the usage of the sum–product algorithm (SPA). In Section IV, an efficient algorithm is derived, based on the method of canonical distributions. Finally, in Section V, the simulation results are presented, and the concluding remarks are given in Section VI.

## II. System Model

PM transmission of complex modulation symbols is considered, over an AWGN channel with laser phase noise and a random state of polarization at the receiver end. Polarization demultiplexing is assumed to have been performed successfully, using a standard algorithm such as the constant modulus algorithm [30]. This entails separating the polarization components, but imposing a drifting phase difference between them [31], [32]. This drift occurs on a much longer time scale than the laser phase noise; therefore, it is assumed to be a constant offset throughout the transmitted block. Other impairments, such as chromatic dispersion, polarization mode dispersion, and frequency offset, are assumed to have been compensated for. Considering perfect symbol synchronization, matched filtering, one sample per symbol, and no nonlinear phase noise or intersymbol interference, the discrete-time complex baseband model is written as [20], [22]

$$\begin{bmatrix} r_{\mathsf{x},k} \\ r_{\mathsf{y},k} \end{bmatrix} = \begin{bmatrix} s_{\mathsf{x},k} \\ s_{\mathsf{y},k} e^{jC} \end{bmatrix} e^{j\theta_k} + \begin{bmatrix} n_{\mathsf{x},k} \\ n_{\mathsf{y},k} \end{bmatrix}, \tag{1}$$

for $k = 0, 1, \ldots, N-1$, where $\mathsf{x}$ and $\mathsf{y}$ denote the two orthogonal polarizations, and $N$ is the length of the transmitted symbol block. A schematic of the model is sketched in Fig. 1.

The vectors $\boldsymbol{r}_{\mathsf{x}} = [r_{\mathsf{x},0}, \ldots, r_{\mathsf{x},N-1}]$ and $\boldsymbol{r}_{\mathsf{y}} = [r_{\mathsf{y},0}, \ldots, r_{\mathsf{y},N-1}]$ are the received samples, where as $\boldsymbol{s}_{\mathsf{x}} = [s_{\mathsf{x},0}, \ldots, s_{\mathsf{x},N-1}]$ and $\boldsymbol{s}_{\mathsf{y}} = [s_{\mathsf{y},0}, \ldots, s_{\mathsf{y},N-1}]$ contain complex $M$-ary modulation symbols. The data symbols are drawn independently from a constellation $\mathcal{M} \in \mathbb{C}$ with equal probability. The constellation is subject to an energy constraint such that the average transmitted symbol energy is $E_{\mathrm{s}}$. Pilot symbols are distributed with regular intervals in $\boldsymbol{s}_{\mathsf{x}}$ and $\boldsymbol{s}_{\mathsf{y}}$. Moreover, the positions of the pilots in polarization $\mathsf{y}$ are shifted by half

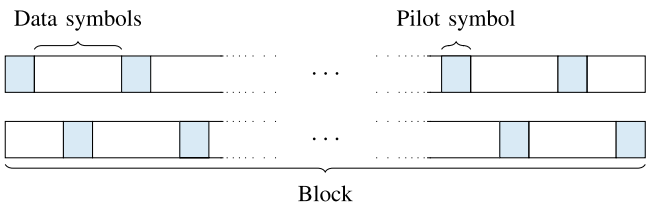


Fig. 2. Pilot distribution in a transmitted block on polarizations x (above) and y (below).

of the pilot interval. This arrangement is depicted in Fig. 2. The noise vectors $\boldsymbol{n}_{\mathsf{x}} = [n_{\mathsf{x},0}, \ldots, n_{\mathsf{x},N-1}]$ and $\boldsymbol{n}_{\mathsf{y}} = [n_{\mathsf{y},0}, \ldots, n_{\mathsf{y},N-1}]$ are independent of each other, and contain independent and identically distributed, circularly symmetric complex Gaussian components, $n_{\mathsf{x},k} \sim \mathcal{N}_{\mathbb{C}}(0, 2\sigma_{\mathrm{n}}^2)$ and $n_{\mathsf{y},k} \sim \mathcal{N}_{\mathbb{C}}(0, 2\sigma_{\mathrm{n}}^2)$. The vector $\boldsymbol{\theta}$ accounts for the accumulated phase noise from the transmitter and the LO laser, and affects both polarizations equally. Moreover, it is defined modulo $2\pi$, and is modeled as a random-walk process, where

$$\theta_k = (\theta_{k-1} + \Delta_k) \mod 2\pi, \tag{2}$$

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

$$\sigma_{\Delta}^2 = 2\pi\Delta\nu T_{\mathrm{s}}, \tag{3}$$

where $\Delta\nu$ is the total laser linewidth and $T_{\mathrm{s}}$ 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 between $\boldsymbol{r}_{\mathsf{x}}$ and $\boldsymbol{r}_{\mathsf{y}}$ is encapsulated in the random variable $C$, drawn from a uniform distribution in the interval $[0, 2\pi)$.

## III. Optimal Symbol Detection and FG Framework Application

In this section the FG framework and the SPA are used to realize the optimal symbol detector. As mentioned in the introduction, the detector that minimizes the SER for an uncoded transmission is the MAP symbol detector. It is performed symbol-by-symbol according to

$$\hat{s}_{w,k} = \underset{s_{w,k} \in \mathcal{M}}{\operatorname{argmax}}\ p(s_{w,k} | \boldsymbol{r}_{\mathsf{x}}, \boldsymbol{r}_{\mathsf{y}}), \tag{4}$$

for $k = 0, 1, \ldots, N-1$ and $w \in \{\mathsf{x}, \mathsf{y}\}$. The function $p(s_{w,k} | \boldsymbol{r}_{\mathsf{x}}, \boldsymbol{r}_{\mathsf{y}})$ is the *a posteriori* probability mass function (PMF) of the $k$th symbol on polarization $w$. The detection rule in (4) considers all the received samples when decision on each transmitted symbol is made. In [24] it is shown that in the presence of phase noise, the PMF in (4) can be written as[1]

$$p(s_{w,k} | \boldsymbol{r}_{\mathsf{x}}, \boldsymbol{r}_{\mathsf{y}}) \propto \int_0^{2\pi} p(r_{w,k} | s_{w,k}, \theta_k) p(\theta_k | \bar{\boldsymbol{r}}_{w,k}) d\theta_k, \tag{5}$$

where $\bar{\boldsymbol{r}}_{w,k}$ is the vector of received samples on both polarizations except $r_{w,k}$ and $\theta_k$ is the phase noise at the $k$th sample. Moreover, the function $p(r_{w,k} | s_{w,k}, \theta_k)$ is the likelihood function of the $k$th received sample on polarization $w$, which has a complex Gaussian PDF with mean $s_{w,k} e^{j\theta_k}$, variance $2\sigma_{\mathrm{n}}^2$, and argument $r_{w,k}$. The integral in (5) highlights the fact that the *a posteriori* phase-noise PDF is used in the symbol detection. In [24], it is further shown that this PDF does not have a closed form solution in general, and computing the integral directly is impractical.

The PMF in (4) can be obtained by marginalizing the joint PDF $p(\boldsymbol{s}_{\mathsf{x}}, \boldsymbol{s}_{\mathsf{y}}, \boldsymbol{\theta}, C | \boldsymbol{r}_{\mathsf{x}}, \boldsymbol{r}_{\mathsf{y}})$ of all the system variables,

$$p(s_{\mathsf{x},k} | \boldsymbol{r}_{\mathsf{x}}, \boldsymbol{r}_{\mathsf{y}}) = \sum_{\boldsymbol{s}_{\mathsf{x}} \in \mathcal{S}_{\mathsf{x}}} \sum_{\boldsymbol{s}_{\mathsf{y}} \in \mathcal{M}^N} \int_{\mathcal{T}} \int_0^{2\pi} p(\boldsymbol{s}_{\mathsf{x}}, \boldsymbol{s}_{\mathsf{y}}, \boldsymbol{\theta}, C | \boldsymbol{r}_{\mathsf{x}}, \boldsymbol{r}_{\mathsf{y}})\, dC\, d\boldsymbol{\theta}, \tag{6}$$

where $\mathcal{S}_{\mathsf{x}} = \{\boldsymbol{s}'_{\mathsf{x}} \in \mathcal{M}^N : s'_{\mathsf{x},k} = s_{\mathsf{x},k}\}$ and $\mathcal{T} = [0, 2\pi)^N$. The corresponding PMF $p(s_{\mathsf{y},k} | \boldsymbol{r}_{\mathsf{x}}, \boldsymbol{r}_{\mathsf{y}})$ is obtained analogously by swapping x and y in (6). The computational complexity of this approach is impractical, and the problem can be solved in more efficient ways. In this paper, the PMF in (4) is computed using an FG which is constructed from the joint PDF $p(\boldsymbol{s}_{\mathsf{x}}, \boldsymbol{s}_{\mathsf{y}}, \boldsymbol{\theta} | \boldsymbol{r}_{\mathsf{x}}, \boldsymbol{r}_{\mathsf{y}})$, i.e., the joint *a posteriori* PDF of all system variables except $C$. In other words, the phase offset between $\boldsymbol{r}_{\mathsf{x}}$ and $\boldsymbol{r}_{\mathsf{y}}$ is assumed to be absent when the FG is constructed. Omitting $C$ reduces the complexity of the resulting FG. The rationale for this omission is the fact that the variable is constant throughout the received sequence. Thus, the phase offset between the polarizations can be estimated and compensated for with high accuracy in the majority of cases. After compensation, the phase offset between the polarizations is approximately absent. The estimation and compensation methods for this phase offset will be detailed in Section IV-A.

FGs can be used to visually represent how a joint distribution may be factorized into a product of simpler marginal functions. A message-passing algorithm that operates on FGs is the SPA. It can be used to derive algorithms in various applications, e.g., iterative decoding algorithms in error-correcting codes, such as low-density parity-check and turbo codes [26]. The tutorial paper [33] offers a thorough explanation of FGs and the SPA.

Rewriting the joint PDF $p(\boldsymbol{s}_{\mathsf{x}}, \boldsymbol{s}_{\mathsf{y}}, \boldsymbol{\theta} | \boldsymbol{r}_{\mathsf{x}}, \boldsymbol{r}_{\mathsf{y}})$ using Bayes' theorem results in

$$\begin{aligned} p(\boldsymbol{s}_{\mathsf{x}}, \boldsymbol{s}_{\mathsf{y}}, \boldsymbol{\theta} | \boldsymbol{r}_{\mathsf{x}}, \boldsymbol{r}_{\mathsf{y}}) &= \frac{p(\boldsymbol{r}_{\mathsf{x}}, \boldsymbol{r}_{\mathsf{y}} | \boldsymbol{s}_{\mathsf{x}}, \boldsymbol{s}_{\mathsf{y}}, \boldsymbol{\theta}) p(\boldsymbol{s}_{\mathsf{x}}, \boldsymbol{s}_{\mathsf{y}}, \boldsymbol{\theta})}{p(\boldsymbol{r}_{\mathsf{x}}, \boldsymbol{r}_{\mathsf{y}})} \\ &\propto p(\boldsymbol{r}_{\mathsf{x}}, \boldsymbol{r}_{\mathsf{y}} | \boldsymbol{s}_{\mathsf{x}}, \boldsymbol{s}_{\mathsf{y}}, \boldsymbol{\theta}) p(\boldsymbol{s}_{\mathsf{x}}) p(\boldsymbol{s}_{\mathsf{y}}) p(\boldsymbol{\theta}) \quad (7) \\ &\propto p(\boldsymbol{\theta}) p(\boldsymbol{r}_{\mathsf{x}} | \boldsymbol{s}_{\mathsf{x}}, \boldsymbol{\theta}) p(\boldsymbol{r}_{\mathsf{y}} | \boldsymbol{s}_{\mathsf{y}}, \boldsymbol{\theta}), \quad (8) \end{aligned}$$

where (7) follows from the fact that the transmitted symbols are independent of the phase noise. Moreover, the joint PDF of the received samples, $p(\boldsymbol{r}_{\mathsf{x}}, \boldsymbol{r}_{\mathsf{y}})$, does not depend on $\boldsymbol{s}_{\mathsf{x}}$ or $\boldsymbol{s}_{\mathsf{y}}$, 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, $p(\boldsymbol{s}_{\mathsf{x}})$ and $p(\boldsymbol{s}_{\mathsf{y}})$ are constant. Furthermore, when conditioned on $\boldsymbol{\theta}$, the received samples are independent of each other, resulting in (8).

[1]Only $\boldsymbol{s}_{\mathsf{x}}$, $\boldsymbol{s}_{\mathsf{y}}$, and $\boldsymbol{\theta}$ affect the maximization in (4). Constants in expressions with respect to these parameters can thus be ignored. Therefore the proportionality relationship $\propto$ suffices and will be used throughout the paper.

The random-walk model for the phase noise has the Markov property such that

$$p\left(\theta_k|\theta_{k-1},\ldots,\theta_0\right)=p\left(\theta_k|\theta_{k-1}\right)=p_\Delta\left(\theta_{k-1};\theta_k\right), \tag{9}$$

for $k=0,1,\ldots,N-1$. The function $p_\Delta$ is a wrapped normal PDF with variance $\sigma_\Delta^2$, expressed as

$$p_\Delta\left(\mu;z\right)\triangleq\sum_{\tau=-\infty}^{\infty}g(\mu,\sigma_\Delta^2;z-\tau 2\pi),\, z\in[0,2\pi), \tag{10}$$

where $g(\mu,\sigma^2;z)$ is a real Gaussian PDF with mean $\mu$, variance $\sigma^2$, and argument $z$. Using (9), the joint PDF of $\boldsymbol{\theta}$ can be factored as

$$p\left(\boldsymbol{\theta}\right)=p\left(\theta_0\right)\prod_{k=1}^{N-1}p_\Delta\left(\theta_{k-1};\theta_k\right), \tag{11}$$

allowing further decomposition of the joint PDF in (8),

$$\begin{aligned}&p\left(\boldsymbol{s}_{\mathsf{x}},\boldsymbol{s}_{\mathsf{y}},\boldsymbol{\theta}|\boldsymbol{r}_{\mathsf{x}},\boldsymbol{r}_{\mathsf{y}}\right)\\&\propto p\left(\theta_0\right)\prod_{k=1}^{N-1}p_\Delta\left(\theta_{k-1};\theta_k\right)\prod_{k=0}^{N-1}f(s_{\mathsf{x},k},\theta_k)f(s_{\mathsf{y},k},\theta_k),\end{aligned} \tag{12}$$

where

$$\begin{aligned}f(s_{w,k},\theta_k)&\triangleq\exp\left\{\frac{1}{\sigma_{\mathrm{n}}^2}\operatorname{Re}\left\{r_{w,k}s_{w,k}^*e^{-j\theta_k}\right\}-\frac{|s_{w,k}|^2}{2\sigma_{\mathrm{n}}^2}\right\}\\&\propto g_{\mathbb{C}}(s_{w,k}e^{j\theta_k},2\sigma_{\mathrm{n}}^2;r_{w,k})\\&=p\left(r_{w,k}|s_{w,k},\theta_k\right),\end{aligned} \tag{13}$$

for $k=0,1,\ldots,N-1$ and $w\in\{\mathsf{x},\mathsf{y}\}$, where $g_{\mathbb{C}}(\mu,\sigma^2;z)$ is a complex Gaussian PDF with mean $\mu$, variance $\sigma^2$, and argument $z$. A part of the FG constructed from (12) can be seen in Fig. 3. The graph is cycle-free, implying that the SPA, when applied to this FG, yields exact *a posteriori* distributions [33]. Hence, it is optimal as it realizes the exact MAP symbol detector.

For each $k$ and $w$, let $\overrightarrow{P}_{w,k}(s_{w,k})$ 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. These probabilities, as mentioned in Section II, are all $1/M$ by default. If multiple iterations are run, however, they will be replaced by *a posteriori* probabilities from the previous iteration, see Section IV-C. The message from $f(s_{w,k},\theta_k)$ to variable node $\theta_k$ is denoted by $p_w(\theta_k)$, and is computed as

$$p_w\left(\theta_k\right)\propto\sum_{s_{w,k}\in\mathcal{M}}\overrightarrow{P}_{w,k}(s_{w,k})f(s_{w,k},\theta_k)\propto p\left(r_{w,k}|\theta_k\right), \tag{14}$$

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

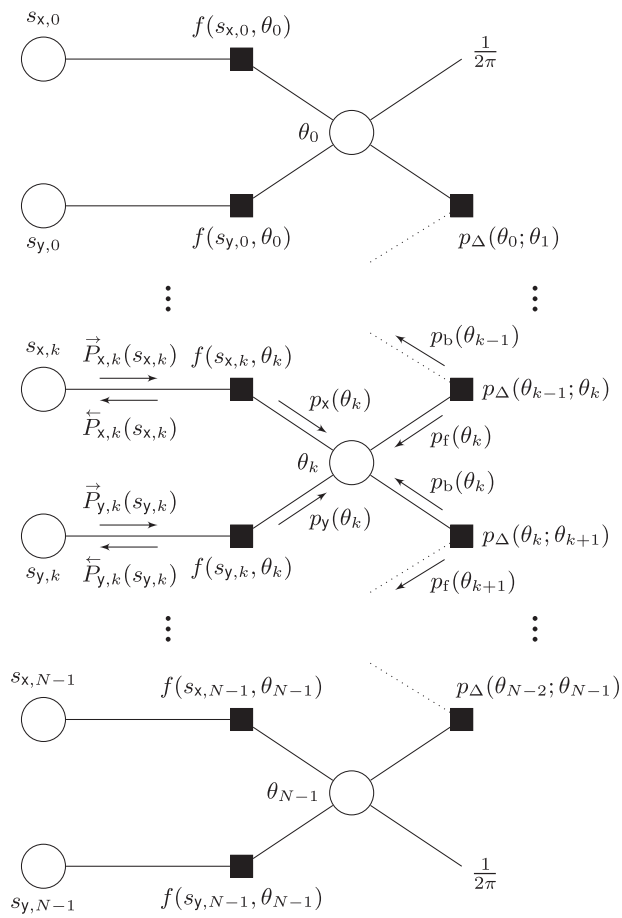


Fig. 3. Part of FG constructed from the distribution in (12).

as

$$\begin{aligned}p_{\mathrm{f}}(\theta_k)=&\int_0^{2\pi}p_{\mathsf{x}}(\theta_{k-1})p_{\mathsf{y}}(\theta_{k-1})p_{\mathrm{f}}(\theta_{k-1})\\&\cdot p_\Delta\left(\theta_{k-1};\theta_k\right)d\theta_{k-1},\end{aligned} \tag{15}$$

$$\begin{aligned}p_{\mathrm{b}}(\theta_k)=&\int_0^{2\pi}p_{\mathsf{x}}(\theta_{k+1})p_{\mathsf{y}}(\theta_{k+1})p_{\mathrm{b}}(\theta_{k+1})\\&\cdot p_\Delta\left(\theta_k;\theta_{k+1}\right)d\theta_{k+1},\end{aligned} \tag{16}$$

with initial conditions

$$p_{\mathrm{f}}(\theta_0)=p_{\mathrm{b}}(\theta_{N-1})=\frac{1}{2\pi}. \tag{17}$$

Finally, let $\overleftarrow{P}_{w,k}(s_{w,k})$ denote the message from factor node $f(s_{w,k},\theta_k)$ to variable node $s_{w,k}$. It is computed as

$$\overleftarrow{P}_{\mathsf{x},k}(s_{\mathsf{x},k})\propto\int_0^{2\pi}p_{\mathrm{f}}(\theta_k)p_{\mathrm{b}}(\theta_k)p_{\mathsf{y}}(\theta_k)f(s_{\mathsf{x},k},\theta_k)d\theta_k, \tag{18}$$

$$\overleftarrow{P}_{\mathsf{y},k}(s_{\mathsf{y},k})\propto\int_0^{2\pi}p_{\mathrm{f}}(\theta_k)p_{\mathrm{b}}(\theta_k)p_{\mathsf{x}}(\theta_k)f(s_{\mathsf{y},k},\theta_k)d\theta_k, \tag{19}$$

for polarizations $\mathsf{x}$ and $\mathsf{y}$, respectively, and $k=0,1,\ldots,N-1$. The messages in (18) and (19) correspond to the *a posteriori* PMFs $p\left(s_{\mathsf{x},k}|\boldsymbol{r}_{\mathsf{x}},\boldsymbol{r}_{\mathsf{y}}\right)$ and $p\left(s_{\mathsf{y},k}|\boldsymbol{r}_{\mathsf{x}},\boldsymbol{r}_{\mathsf{y}}\right)$, respectively.

Furthermore, it can be shown that

$$p_{\mathrm{f}}(\theta_k)p_{\mathrm{b}}(\theta_k)p_{\mathsf{y}}(\theta_k) \propto p(\theta_k|\bar{\boldsymbol{r}}_{\mathsf{x},k}), \tag{20}$$

$$p_{\mathrm{f}}(\theta_k)p_{\mathrm{b}}(\theta_k)p_{\mathsf{x}}(\theta_k) \propto p(\theta_k|\bar{\boldsymbol{r}}_{\mathsf{y},k}), \tag{21}$$

where $p(\theta_k|\bar{\boldsymbol{r}}_{w,k})$, for $w \in \{\mathsf{x},\mathsf{y}\}$, is the *a posteriori* phase-noise PDF in (5). The PDFs in (20) and (21) can be used to produce point estimates of the phase noise. However, using point estimates to compensate for the phase noise does not lead to optimal symbol detection. 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, the information about this uncertainty is discarded. This has been shown to degrade the performance of symbol detection [24].

The message-passing algorithm can be summarized as follows. The received samples, the *a priori* probabilities of the transmitted symbols, and the AWGN and phase-noise variance, are the inputs to the algorithm. The inputs are used to calculate the messages in (14). The messages in (15) and (16), corresponding to the *a posteriori* phase-noise PDFs, are then computed in a recursive fashion. Finally, the messages in (18) and (19), corresponding to the *a posteriori* symbol probabilities, are calculated. These messages are the outputs of the algorithm. Symbol detection is performed according to (4), by selecting the argument, i.e., the constellation point in $\mathcal{M}$, that maximizes the messages in (18) and (19). In coded systems that employ iterative soft-decision decoding, the message $\overrightarrow{P}_{w,k}(s_{w,k})$ corresponds to the extrinsic PMF of $s_{w,k}$ provided by the decoder. Moreover, the messages in (18) and (19) are used to compute log-likelihood ratios which are then used as channel information for the decoder.

## IV. Receiver Implementation

The messages produced by the SPA are difficult to implement in practice, as they include integrals. One approach is to approximate the integrals as weighted sums of impulses by discretizing the phase noise. This approximation approaches the performance of the exact MAP detector, but becomes impractical, as the number of discretization levels increases [26]. As a result, other approximations are required in order to obtain an efficient implementation. In this section, the estimation and compensation for the phase offset, due to the polarization demultiplexing, are described. Afterwards, a pilot-based algorithm is derived, using approximations of the results given by the SPA in the preceding section. The resulting DSP chain is depicted in Fig. 4, showing how the proposed algorithm differs from conventional setups, where phase noise estimation is performed separately for each polarization, independently of the symbol detection.

### A. Compensation of Phase Offset Between Polarizations

As mentioned in Section III, the variable $C$ embodies the phase offset between the polarizations. Moreover, the FG in Fig. 3 is constructed from the joint PDF $p(\boldsymbol{s}_{\mathsf{x}}, \boldsymbol{s}_{\mathsf{y}}, \boldsymbol{\theta}|\boldsymbol{r}_{\mathsf{x}}, \boldsymbol{r}_{\mathsf{y}})$, with the assumption that $\boldsymbol{r}_{\mathsf{x}}$ and $\boldsymbol{r}_{\mathsf{y}}$ are free of this phase offset. If not compensated for, the *a posteriori* phase-noise PDF estimation will not work properly. However, through utilization of the pilot symbols, the phase offset between the polarizations can be estimated.

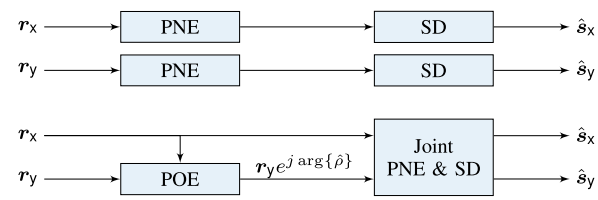


Fig. 4. A conventional DSP chain (top) and the proposed DSP chain (bottom). (POE: Phase offset estimation, PNE: Phase noise estimation, SD: Symbol detection.)

Let $L$ denote the set of pilot indices on polarization $\mathsf{x}$ in the transmitted block, $L \subset \{0, 1, \ldots, N-K-1\}$. Furthermore, let $K$ denote the number of indices by which the pilots on polarization $\mathsf{y}$ are delayed, with respect to the pilots on polarization $\mathsf{x}$. For pilot symbols, the transmitted constellation point and hence its phase are known to the receiver. The phase offset between the transmitted pilot symbol and its corresponding received sample on polarization $\mathsf{x}$ at index $l \in L$ can thus be computed as

$$\begin{aligned} r_{\mathsf{x},l}s^*_{\mathsf{x},l} &= s_{\mathsf{x},l}s^*_{\mathsf{x},l}e^{j\theta_l} + n_{\mathsf{x},l}s^*_{\mathsf{x},l} \\ &= |s_{\mathsf{x},l}|^2 e^{j\theta_l} + |s_{\mathsf{x},l}|n_{\mathsf{x},l}e^{-j\arg\{s_{\mathsf{x},l}\}} \\ &\propto e^{j\theta_l} + \frac{n_{\mathsf{x},l}e^{-j\arg\{s_{\mathsf{x},l}\}}}{|s_{\mathsf{x},l}|}. \end{aligned} \tag{22}$$

Rotating the AWGN vectors $\boldsymbol{n}_{\mathsf{x}}$ and $\boldsymbol{n}_{\mathsf{y}}$ does not change their statistics as they contain complex circularly symmetric Gaussian components. Similarly, the phase offset between the transmitted pilot symbol and its corresponding received sample on polarization $\mathsf{y}$ at index $l+K$ is computed as

$$r_{\mathsf{y},l+K}\, s^*_{\mathsf{y},l+K} \propto e^{j(\theta_{l+K}+C)} + \frac{n_{\mathsf{y},l+K}\, e^{-j\arg\{s_{\mathsf{y},l+K}\}}}{|s_{\mathsf{y},l+K}|}. \tag{23}$$

The phase-noise samples in (23) can be rewritten as

$$\theta_{l+K} = \theta_l + \sum_{k=l+1}^{l+K} \Delta_k = \theta_l + \Delta_K, \tag{24}$$

where $\Delta_K \sim \mathcal{N}(0, K\sigma^2_\Delta)$. The phase offset between the polarizations is expressed as

$$r_{\mathsf{x},l}s^*_{\mathsf{x},l}(r_{\mathsf{y},l+K}\, s^*_{\mathsf{y},l+K})^* \propto \rho e^{-j\Delta_K} + m_l \tag{25}$$

$$\approx \rho\left(1 - j\Delta_K - \frac{\Delta_K^2}{2}\right) + m_l, \tag{26}$$

where $\rho = e^{-jC}$, and $m_l$ comprises a sum of terms contributed by the noise. The exponential in (25) is approximated using a second-order Taylor series expansion, which is a close approximation for practical values of $K\sigma^2_\Delta$. Moreover, it can be shown that $m_l$ is drawn from a zero-mean, non-Gaussian distribution.

Treating $\rho$ as an unknown deterministic parameter, and taking the expectation of (25), gives

$$\mathsf{E}\left[r_{\mathsf{x},l}s^*_{\mathsf{x},l}(r_{\mathsf{y},l+K}s^*_{\mathsf{y},l+K})^*\right] \approx \rho\left(1-\frac{K\sigma^2_\Delta}{2}\right). \quad (27)$$

The right hand side of (27) can be estimated by the sample mean [25, p. 138]. Moreover, as scaling does not affect the argument of a complex number, the sample sum suffices, yielding

$$\hat{\rho} = \sum_{l\in L} r_{\mathsf{x},l}s^*_{\mathsf{x},l}r^*_{\mathsf{y},l+K}s_{\mathsf{y},l+K}. \quad (28)$$

This estimator will have the argument $-C$ on the average, for practical values of laser linewidth and pilot overhead. Rotating $r_{\mathsf{y},k}$ with the argument of (28) gives

$$\begin{bmatrix} r_{\mathsf{x},k} \\ r_{\mathsf{y},k}e^{j\arg\{\hat{\rho}\}} \end{bmatrix} \approx \begin{bmatrix} s_{\mathsf{x},k} \\ s_{\mathsf{y},k} \end{bmatrix} e^{j\theta_k} + \begin{bmatrix} n_{\mathsf{x},k} \\ \rho n_{\mathsf{y},k} \end{bmatrix}, \quad (29)$$

for $k = 0, 1, \ldots, N-1$. Comparing (29) with (1), it is evident that $\boldsymbol{r}_\mathsf{x}$ and $\boldsymbol{r}_\mathsf{y}$ are essentially devoid of any phase offset after compensation. Consequently, constructing an FG from the joint PDF $p(\boldsymbol{s}_\mathsf{x}, \boldsymbol{s}_\mathsf{y}, \boldsymbol{\theta}|\boldsymbol{r}_\mathsf{x}, \boldsymbol{r}_\mathsf{y})$, and the model in (29), is valid after following the procedure described above.

### B. Phase-Noise Estimation and Symbol Detection

In order to obtain an efficient implementation, the approach in [26] is adopted, namely selecting parametrized canonical distributions to approximate messages involving continuous PDFs [34]. It is claimed in [26] that utilizing the Tikhonov canonical distribution yields performance close to the MAP detector. Furthermore, this approach has repeatedly proven its merit in phase noise estimation under different scenarios [35]–[37]. Hence, the messages $p_\mathrm{f}(\theta_k)$ and $p_\mathrm{b}(\theta_k)$ are constrained to be in a family of Tikhonov PDFs. This PDF, also known as a von Mises distribution, is defined as [38, p. 35]

$$t(\kappa;\theta) \triangleq \frac{1}{2\pi I_0(|\kappa|)}\exp\left\{\mathrm{Re}\left\{\kappa e^{-j\theta}\right\}\right\}, \theta\in[0,2\pi), \quad (30)$$

where $\kappa \in \mathbb{C}$ and $I_0$ is the modified Bessel function of the first kind and zeroth order. The argument and modulus of $\kappa$ characterize the mean and variance, respectively, of the PDF in (30); a Tikhonov PDF is completely described by a single parameter. The direct computation of the integrals in (15) and (16) can thus be reduced to recursively updating the parameters describing the Tikhonov PDFs.

A closer look at the messages in (14) reveals that they are a mixture of Gaussian PDFs. As in [26], the messages are 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 these Gaussian PDFs are $\mathsf{E}[r_{w,k}|\theta_k]$ and $\mathsf{Var}[r_{w,k}|\theta_k]$, respectively, for $k = 0, 1, \ldots, N-1$ and $w \in \{\mathsf{x}, \mathsf{y}\}$. It can be shown that

$$\mathsf{E}[r_{w,k}|\theta_k] = \alpha_{w,k}e^{j\theta_k}, \quad (31)$$

$$\mathsf{Var}[r_{w,k}|\theta_k] = 2\sigma^2_\mathrm{n} + \beta_{w,k} - |\alpha_{w,k}|^2, \quad (32)$$

where

$$\alpha_{w,k} \triangleq \sum_{s_{w,k}\in\mathcal{M}} s_{w,k}\vec{P}_{w,k}(s_{w,k}), \quad (33)$$

$$\beta_{w,k} \triangleq \sum_{s_{w,k}\in\mathcal{M}} |s_{w,k}|^2\vec{P}_{w,k}(s_{w,k}), \quad (34)$$

represent the first and second moments of $s_{w,k}$, respectively. Using the results from (31) and (32) yields

$$\begin{aligned} p_w(\theta_k) &\approx g_\mathbb{C}(\mathsf{E}[r_{w,k}|\theta_k], \mathsf{Var}[r_{w,k}|\theta_k]; r_{w,k}) \\ &= t\left(\frac{2r_{w,k}\alpha^*_{w,k}}{2\sigma^2_\mathrm{n}+\beta_{w,k}-|\alpha_{w,k}|^2}; \theta_k\right). \end{aligned} \quad (35)$$

For pilot symbols, $\vec{P}_{w,k}(s_{w,k}) = 1$ in (33) and (34) for the constellation point $s_{w,k} \in \mathcal{M}$ used as a pilot symbol, and 0 for the points in $\mathcal{M}\setminus\{s_{w,k}\}$. This gives an unambiguous phase reference in (35) for indices $w$ and $k$ corresponding to pilot symbols.

The messages $p_\mathrm{f}(\theta_k)$ and $p_\mathrm{b}(\theta_k)$ are constrained to be in a family of Tikhonov PDFs,

$$p_\mathrm{f}(\theta_k) \propto t(a_{\mathrm{f},k};\theta_k), \quad (36)$$

$$p_\mathrm{b}(\theta_k) \propto t(a_{\mathrm{b},k};\theta_k), \quad (37)$$

for $k = 0, 1, \ldots, N-1$, where $a_{\mathrm{f},k}$ and $a_{\mathrm{b},k}$ denote the $k$th forward and backward coefficients, respectively. These coefficients can be recursively calculated, allowing for an efficient means of computing the messages in (36) and (37) for all $k$. To this end, we substitute (35) and (36) into (15), and exploit that the product of univariate Tikhonov PDFs is proportional to another univariate Tikhonov PDF. Thus,

$$p_\mathrm{f}(\theta_k) \approx \int_0^{2\pi} t(z_{\mathrm{f},k-1};\theta_{k-1})p_\Delta(\theta_{k-1};\theta_k)d\theta_{k-1}, \quad (38)$$

where

$$\begin{aligned} z_{\mathrm{f},k-1} \triangleq a_{\mathrm{f},k-1} &+ \frac{2r_{\mathsf{x},k-1}\alpha^*_{\mathsf{x},k-1}}{2\sigma^2_\mathrm{n}+\beta_{\mathsf{x},k-1}-|\alpha_{\mathsf{x},k-1}|^2} \\ &+ \frac{2r_{\mathsf{y},k-1}\alpha^*_{\mathsf{y},k-1}}{2\sigma^2_\mathrm{n}+\beta_{\mathsf{y},k-1}-|\alpha_{\mathsf{y},k-1}|^2}. \end{aligned} \quad (39)$$

Let the following $2\pi$-periodic function be defined as

$$\tilde{t}(\kappa;\eta) \triangleq \frac{1}{2\pi I_0(|\kappa|)}\exp\left\{\mathrm{Re}\left\{\kappa e^{-j\eta}\right\}\right\}, \quad (40)$$

where $\kappa \in \mathbb{C}$ and $\eta \in \mathbb{R}$. Using (40), the integral in (38) can be rewritten as

$$p_\mathrm{f}(\theta_k) \approx \int_{\theta_k-\pi}^{\theta_k+\pi} \tilde{t}(z_{\mathrm{f},k-1};\theta_{k-1})p_\Delta(\theta_{k-1};\theta_k)d\theta_{k-1}. \quad (41)$$

For practical values of $\sigma^2_\Delta$, the wrapped normal PDF $p_\Delta$ in (41) is virtually zero everywhere in $[\theta_k-\pi, \theta_k+\pi)$ except in an interval much smaller than $2\pi$, centered around $\theta_k$. Consequently, it can be approximated as a Gaussian PDF,

$$p_\Delta(\theta_{k-1};\theta_k) \approx g(\theta_{k-1}, \sigma^2_\Delta; \theta_k). \quad (42)$$

Thus, the integral in (41) simplifies to

$$p_{\mathrm{f}}(\theta_k) \approx \int_{\theta_k-\pi}^{\theta_k+\pi} \tilde{t}(z_{\mathrm{f},k-1};\theta_{k-1}) g(\theta_{k-1},\sigma_\Delta^2;\theta_k) d\theta_{k-1}$$
$$\propto \int_{\theta_k-\pi}^{\theta_k+\pi} \exp\left\{\mathrm{Re}\left\{z_{\mathrm{f},k-1} e^{-j\theta_{k-1}}\right\}\right\} \cdot g(\theta_{k-1},\sigma_\Delta^2;\theta_k) d\theta_{k-1}. \quad (43)$$

The integral in (43) can be approximately expressed [26] in the same form as (40), with the complex coefficient denoted by $a_{\mathrm{f},k}$ and computed as

$$a_{\mathrm{f},k} = \frac{z_{\mathrm{f},k-1}}{1+\sigma_\Delta^2 |z_{\mathrm{f},k-1}|}. \quad (44)$$

Therefore, the forward coefficients $a_{\mathrm{f},k}$ can be computed recursively for $k = 1,\ldots,N-1$ from (39) and (44), with the initial condition $a_{\mathrm{f},0} = 0$. Similarly, using (35), (37), and (42), the backward coefficients $a_{\mathrm{b},k}$, for $k = N-2,\ldots,0$, are computed recursively as

$$a_{\mathrm{b},k} = \frac{z_{\mathrm{b},k+1}}{1+\sigma_\Delta^2 |z_{\mathrm{b},k+1}|}, \quad (45)$$

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

$$z_{\mathrm{b},k+1} \triangleq a_{\mathrm{b},k+1} + \frac{2r_{\mathsf{x},k+1}\alpha_{\mathsf{x},k+1}^*}{2\sigma_{\mathrm{n}}^2 + \beta_{\mathsf{x},k+1} - |\alpha_{\mathsf{x},k+1}|^2} + \frac{2r_{\mathsf{y},k+1}\alpha_{\mathsf{y},k+1}^*}{2\sigma_{\mathrm{n}}^2 + \beta_{\mathsf{y},k+1} - |\alpha_{\mathsf{y},k+1}|^2}. \quad (46)$$

The expressions for $\overleftarrow{P}_{w,k}(s_{w,k})$ are calculated using (18) and (19) for polarization $\mathsf{x}$ and $\mathsf{y}$, respectively. In Appendix A it is shown that they can be approximated as

$$\overleftarrow{P}_{w,k}(s_{w,k}) \approx I_0(|\xi_{w,k}|) \exp\left\{-\frac{|s_{w,k}|^2}{2\sigma_{\mathrm{n}}^2}\right\}, \quad (47)$$

where

$$\xi_{\mathsf{x},k} \triangleq a_{\mathrm{f},k} + a_{\mathrm{b},k} + \frac{2r_{\mathsf{y},k}\alpha_{\mathsf{y},k}^*}{2\sigma_{\mathrm{n}}^2 + \beta_{\mathsf{y},k} - |\alpha_{\mathsf{y},k}|^2} + \frac{r_{\mathsf{x},k}s_{\mathsf{x},k}^*}{\sigma_{\mathrm{n}}^2}, \quad (48)$$

$$\xi_{\mathsf{y},k} \triangleq a_{\mathrm{f},k} + a_{\mathrm{b},k} + \frac{2r_{\mathsf{x},k}\alpha_{\mathsf{x},k}^*}{2\sigma_{\mathrm{n}}^2 + \beta_{\mathsf{x},k} - |\alpha_{\mathsf{x},k}|^2} + \frac{r_{\mathsf{y},k}s_{\mathsf{y},k}^*}{\sigma_{\mathrm{n}}^2}. \quad (49)$$

The expression in (47) can be numerically problematic. A solution to this is derived in Appendix B.

### C. Multiple Iterations

A single iteration of the algorithm is defined as follows. First and second order moments of the transmitted symbols, described in (33) and (34), are first calculated, given the *a priori* symbol probabilities $\overrightarrow{P}_{w,k}$. Afterwards, the *a posteriori* phase-noise PDFs are estimated by way of calculating their forward and backward coefficients recursively with (39) and (44)–(46). Finally, the *a posteriori* symbol probabilities are calculated with (47).

Once the *a posteriori* symbol probabilities have been calculated, symbol detection can be performed according to (4). The first order moment (33), however, is zero for symbols whose constellation points are equiprobable. When the first iteration is run, this is indeed the case for all the data symbols. If $\alpha_{\mathsf{x},k}$ and $\alpha_{\mathsf{y},k}$ are zero, (39) and (46) reduce to $z_{\mathrm{f},k-1} = a_{\mathrm{f},k-1}$ and $z_{\mathrm{b},k+1} = a_{\mathrm{b},k+1}$, respectively. Thus, the coefficients $a_{\mathrm{f},k}$ and $a_{\mathrm{b},k}$ 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.

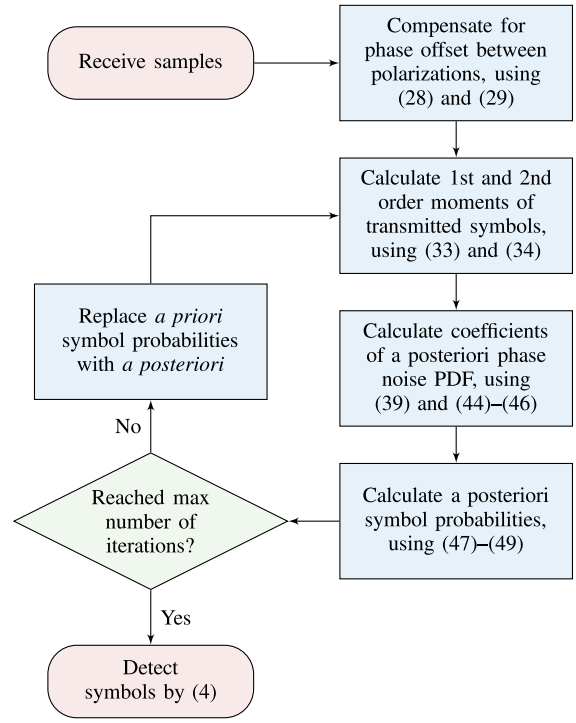


Fig. 5. A high level flowchart of the algorithm, from receiving the samples to detecting the transmitted symbols, illustrating the iterative structure of the implementation.

The performance of the algorithm can be improved by running multiple iterations. After calculating the message $\overleftarrow{P}_{w,k}(s_{w,k})$ 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 $\overrightarrow{P}_{w,k}(s_{w,k})$. This can be performed until a stopping criterion is fulfilled. It should be noted that $\overrightarrow{P}_{w,k}(s_{w,k})$ is assumed to be a PMF, i.e., the sum of $\overrightarrow{P}_{w,k}(s_{w,k})$ over all possible values of $s_{w,k}$ is one. The message $\overleftarrow{P}_{w,k}(s_{w,k})$ is in general only equal to the *a posteriori* symbol probabilities up to a scaling factor; the sum of $\overleftarrow{P}_{w,k}(s_{w,k})$ over all possible values of $s_{w,k}$ is not one. Thus, to be used as the *a priori* symbol probabilities, scaling is generally necessary beforehand. This scaling procedure can present numerical problems; Appendix B details how this is managed. A high level flowchart of the algorithm structure is depicted in Fig. 5.

### D. Complexity and Parallelization

The complexity of the algorithms is estimated by counting the number of real multiplications and additions. Assume that complex multiplications require 4 real multiplications and 2 additions, and furthermore, that nonlinear functions are computed through a lookup table. To compute (28), $12N_{\mathrm{p}}$ multiplications

and $7N_{\mathrm{p}} - 1$ additions are required in total, where $N_{\mathrm{p}}$ is the number of pilots per transmitted block per polarization. Additionally, 4 multiplications and 2 additions are required per symbol to rotate $\boldsymbol{r}_{\mathrm{y}}$ with the argument of (28). Moreover, the total number of multiplications and additions per symbol per polarization for a single iteration of the algorithm in Section IV-B and the normalization of (47), detailed in Appendix B, is found to be $17M + 13$ and $14M + 7$, respectively, where $M$ is the number of constellation points.

The parallelization of the algorithm in [26] is presented in [39], and it is furthermore shown that the parallelization of the forward-backward recursions can degrade performance. It is also detailed how the performance degradation can be reduced at the cost of increased complexity. The analysis in [39] can be utilized to parallelize the algorithm in Section IV-B.

## V. Simulation Results

In this section, the performance of the algorithm is assessed in terms of BER for different amounts of phase noise, signal-to-noise ratio (SNR), pilot overhead, and number of iterations. The algorithm is evaluated using Monte Carlo simulations, using the system model in (1), for transmission of PM-QPSK, PM-16-QAM, and PM-64-QAM constellations. The transmitted blocks contain approximately $N = 10\,000$ symbols per polarization when simulating the proposed algorithm; the exact value of $N$ depends on the pilot overhead. The value of $N$ will not affect the performance of the algorithm significantly, provided that $N \geq 10\,000$.

When simulating the constant phase offset between the polarizations according to (1), a realization of the random variable $C$ is generated for each transmitted block. Moreover, the realizations of $C$ are independent between blocks. When simulating the BPS algorithm, the filter half width is set to 9 and the transmission blocks contain 200 000 symbols. Moreover, 32 test phases values are used when transmitting PM-QPSK and PM-16-QAM, whereas 64 test phases values are used in the case of PM-64-QAM transmission. Additionally, differential encoding, described in [2], is employed due to ambiguities in the phase-noise estimation. For each result, errors in the transmission are accumulated until the total number of transmitted bits reaches $10^8$, or if the number of bit errors exceeds a minimum of 1000, regardless of the block length $N$. 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 $E_{\mathrm{s}}$.

When finding sensitivity penalty (SP) induced by phase noise and pilot overhead, BER values for different amounts of laser linewidth are compared to a theoretical ideal BER, calculated for transmission with no pilot overhead over an AWGN channel. This BER of Gray-coded $M$-QAM is [40]

$$\mathrm{BER} \approx \frac{\sqrt{M}-1}{\sqrt{M}\log_2\sqrt{M}}\,\mathrm{erfc}\left(\sqrt{\frac{3\gamma_{\mathrm{b}}\log_2 M}{2(M-1)}}\right) + \frac{\sqrt{M}-2}{\sqrt{M}\log_2\sqrt{M}}\,\mathrm{erfc}\left(3\sqrt{\frac{3\gamma_{\mathrm{b}}\log_2 M}{2(M-1)}}\right), \quad (50)$$

where erfc is the complementary error function. The SNR per information bit is

$$\gamma_{\mathrm{b}} \triangleq \frac{E_{\mathrm{s}}}{2\sigma_{\mathrm{n}}^2\, R\log_2 M}, \quad (51)$$

where $\log_2 M$ represents the number of bits per symbol and $R$ is the effective information rate per bit, defined as $R = 1/(1+\mathrm{OH})$, where OH is the pilot overhead. Using (50) and letting $R = 1$ in (51), the required $\gamma_{\mathrm{b}}$ for a BER of $10^{-3}$ is approximately 6.79 dB, 10.52 dB, and 14.77 dB for PM-QPSK, PM-16-QAM, and PM-64-QAM, respectively. The SP is the additional $\gamma_{\mathrm{b}}$ required, due to phase noise and loss in effective information rate owing to pilot symbols, to achieve the same BER as in the ideal case.

In the following sections, the proposed Tikhonov algorithm (TIK) will be compared to the original single-channel Tikhonov algorithm (TIK-S) proposed by Colavolpe *et al.* in [26] and the blind phase search algorithm in [2]. It is worth emphasizing that TIK-S and BPS perform single-polarization processing. Performance gains from using phase noise PDFs coupled with utilizing pilots can be assessed by comparing TIK-S with BPS. Furthermore, the benefits of joint-polarization processing may be quantified by comparing TIK with TIK-S.

### A. Linewidth Tolerance

The SP at a BER of $10^{-3}$ is shown in Fig. 6 for TIK running 2 iterations, BPS, and TIK-S for PM-QPSK, PM-16-QAM, and PM-64-QAM. TIK has the lowest SP for all the tested linewidths and constellations. At low linewidths, TIK and TIK-S attain close to zero penalty. The penalty does not reach zero due to the pilot overhead. Moreover, BPS maintains a minimum penalty due to the differential encoding which increases the average BER.

The maximum tolerable $\Delta\nu T_{\mathrm{s}}$ for a 1 dB penalty at a BER of $10^{-3}$ for the considered algorithms is presented in Table I. The values for BPS, which are acquired using single-polarization processing, are taken from [2]. As observed in [2, Section V-A], if data on both polarizations are used for phase noise estimation using BPS, the values in Table I corresponding to BPS can be approximately doubled. However, for the system model in (1), dual-polarization processing will not work properly in the majority of cases without compensating for $C$ beforehand [32]. Moreover, optimizing the parameters of BPS for each $\Delta\nu T_{\mathrm{s}}$ value can lead to further performance improvements for the algorithm [2]. The second column shows the required pilot overhead to achieve the tolerance values. After running 2 iterations at 2.86% pilot overhead, TIK has the best performance in all considered cases. Specifically, when compared to TIK-S, a significant increase in linewidth tolerance is observed, owing to the joint-polarization processing.

As explained in Section III, the laser phase noise variance $\sigma_{\Delta}^2$, which is a function of the laser linewidth $\Delta\nu$, is an input to the proposed algorithm. In a practical scenario, the estimated laser linewidth may be inaccurate, in which case performance will suffer. However, simulations show that for practical values of laser linewidths, performance suffers only marginally even if the estimated laser linewidth is much smaller or larger than in reality.

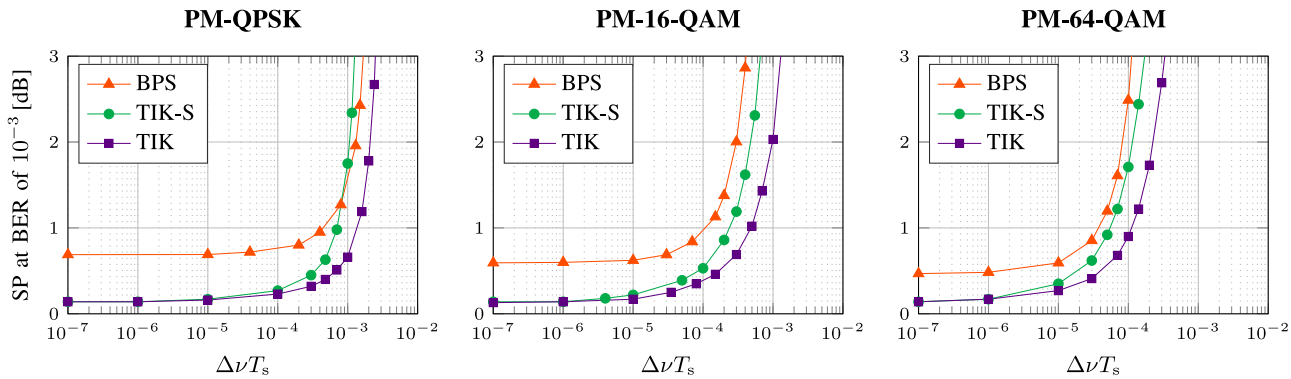


Fig. 6. SP for TIK running 2 iterations at 2.86% pilot overhead, for different modulation formats. For comparison, SP for BPS and TIK-S is also shown.

TABLE I
LINEWIDTH TOLERANCE COMPARISON FOR DIFFERENT ALGORITHMS AND MODULATION FORMATS

| Algorithm | OH | $\Delta\nu T_s$ for 1 dB penalty at a BER of $10^{-3}$ | | |
|---|---|---|---|---|
| | | PM-QPSK | PM-16-QAM | PM-64-QAM |
| TIK (9 it.) | 2.86% | $1.86 \cdot 10^{-3}$ | $8.42 \cdot 10^{-4}$ | $2.00 \cdot 10^{-4}$ |
| TIK (2 it.) | 2.86% | $1.41 \cdot 10^{-3}$ | $4.86 \cdot 10^{-4}$ | $1.11 \cdot 10^{-4}$ |
| TIK-S (9 it.) | 2.86% | $9.43 \cdot 10^{-4}$ | $4.11 \cdot 10^{-4}$ | $9.90 \cdot 10^{-5}$ |
| BPS [2] | 0% | $4.10 \cdot 10^{-4}$ | $1.40 \cdot 10^{-4}$ | $4.00 \cdot 10^{-5}$ |

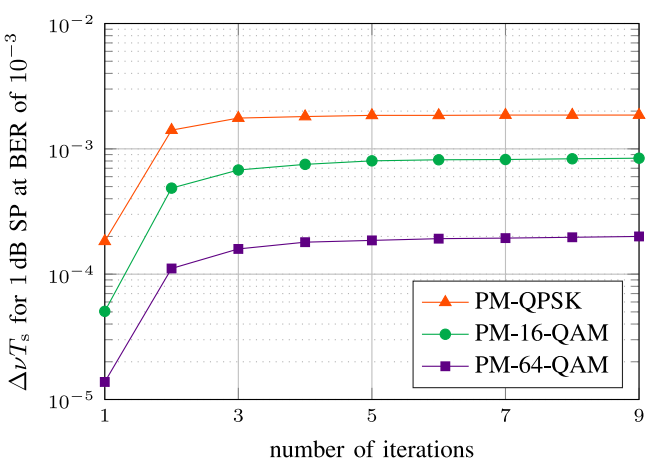


Fig. 7. SP for TIK running a different number of iterations at 2.86% pilot overhead, for a fixed $\gamma_b$ and BER.

## B. 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 different numbers of iterations, is depicted in Fig. 7. After 9 iterations, tolerable $\Delta\nu T_s$ values of $1.86 \cdot 10^{-3}$, $8.42 \cdot 10^{-4}$, and $2.00 \cdot 10^{-4}$ are obtained for PM-QPSK, PM-16-QAM, and PM-64-QAM, respectively. As Fig. 7 shows, the performance improves initially with an increasing number of iterations. However, beyond a particular number of iterations, negligible improvement in performance is observed. In the case of PM-QPSK, the improvements are marginal after 3 iterations. For PM-16-QAM and PM-64-QAM, the tolerance increases more gradually.

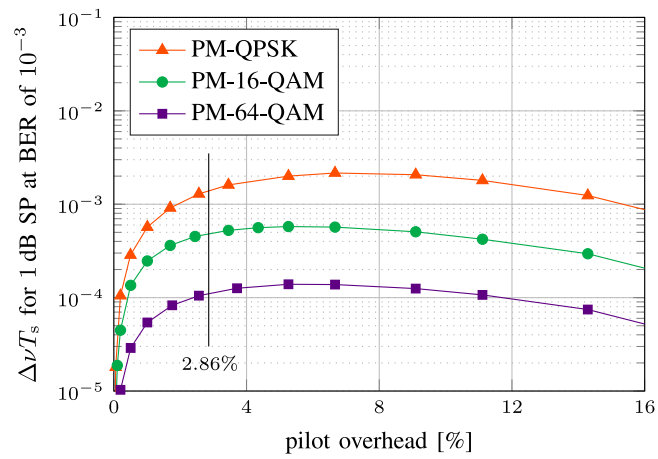


Fig. 8. Linewidth tolerance of TIK running 2 iterations at different pilot overhead, for a fixed $\gamma_b$ and BER.

## C. Pilot Overhead

The performance of the algorithm depends significantly on the pilot overhead. To gain insight into this dependency, the maximum $\Delta\nu T_s$ values for a 1 dB SP at a BER of $10^{-3}$, for different pilot overhead and 2 iterations, are plotted in Fig. 8. As expected, there is an optimal pilot overhead for each constellation. If the overhead is too low, the algorithm will only cope with low amounts of phase noise. On the other hand, as the pilot overhead increases, the SNR loss due to reduction in the transmission rate will eventually be the limiting factor. For PM-QPSK, the optimal pilot overhead is roughly 7%. For PM-16-QAM and PM-64-QAM, the optimal overhead is slightly lower.

# VI. CONCLUSION

In this work, the MAP symbol detector for PM transmission in the presence of AWGN and phase noise was presented, which performs joint-polarization processing to make use of phase-noise PDFs when performing symbol detection. This is

in contrast to what many phase-noise compensation algorithms in fiber-optic communications conform to, namely first estimating the phase noise and then compensating for it before detection. The MAP symbol detector was realized using the FG framework and the SPA. Further, an approach proposed by Colavolpe *et al.* was adopted for implementing a pilot-based algorithm that approximates the optimal detector. This algorithm performs joint-polarization phase-noise estimation and symbol detection.

The performance of the algorithm was assessed using Monte Carlo simulations. It was shown that a significant increase in linewidth tolerance is gained from performing joint-polarization processing. Moreover, after running 9 iterations at $2.86\%$ pilot overhead, the algorithm had approximately $4.5$, $6.0$, and $5.0$ times the laser linewidth tolerance of the BPS algorithm, for PM-QPSK, PM-16-QAM, and PM-64-QAM, respectively.

While this paper focuses on the optimality aspect of symbol detection in the presence of phase noise, it is crucial to consider the practical aspects of the proposed scheme, e.g., parallelization, power consumption, and performance in the presence of nonlinear phase noise, and these aspects are identified as future work. Furthermore, experimental work would be useful for verifying the performance of the algorithm in real transmission scenarios, which is also interesting for future research.

## Appendix A
## A Posteriori Symbol Probabilities

A closed form solution of the messages corresponding to the *a posteriori* symbol probabilities in Section IV-B is derived. Using (18), the message $\overleftarrow{P}_{\mathsf{x},k}(s_{\mathsf{x},k})$, for $k = 0, 1, \ldots, N-1$, can be simplified as

$$\overleftarrow{P}_{\mathsf{x},k}(s_{\mathsf{x},k}) \propto \int_0^{2\pi} p_{\mathrm{f}}(\theta_k) p_{\mathrm{b}}(\theta_k) p_{\mathsf{y}}(\theta_k) f(s_{\mathsf{x},k}, \theta_k) d\theta_k \quad (52)$$

$$\approx \int_0^{2\pi} t(a_{\mathrm{f},k}; \theta_k) t(a_{\mathrm{b},k}; \theta_k) t\left(\frac{2 r_{\mathsf{y},k} \alpha_{\mathsf{y},k}^*}{2\sigma_{\mathrm{n}}^2 + \beta_{\mathsf{y},k} - |\alpha_{\mathsf{y},k}|^2}; \theta_k\right)$$

$$\cdot \exp\left\{\frac{1}{\sigma_{\mathrm{n}}^2} \mathrm{Re}\left\{r_{\mathsf{x},k} s_{\mathsf{x},k}^* e^{-j\theta_k}\right\} - \frac{|s_{\mathsf{x},k}|^2}{2\sigma_{\mathrm{n}}^2}\right\} d\theta_k \quad (53)$$

$$\propto \int_0^{2\pi} \exp\left\{\mathrm{Re}\left\{\xi_{\mathsf{x},k} e^{-j\theta_k}\right\}\right\} \exp\left\{-\frac{|s_{\mathsf{x},k}|^2}{2\sigma_{\mathrm{n}}^2}\right\} d\theta_k \quad (54)$$

$$= 2\pi I_0(|\xi_{\mathsf{x},k}|) \exp\left\{-\frac{|s_{\mathsf{x},k}|^2}{2\sigma_{\mathrm{n}}^2}\right\} \int_0^{2\pi} t(\xi_{\mathsf{x},k}; \theta_k) d\theta_k \quad (55)$$

$$\propto I_0(|\xi_{\mathsf{x},k}|) \exp\left\{-\frac{|s_{\mathsf{x},k}|^2}{2\sigma_{\mathrm{n}}^2}\right\}, \quad (56)$$

where $f(s_{\mathsf{x},k}, \theta_k)$ in (52) is defined in (13); (53) follows from (35)–(37); (54) follows from the same reasoning as (38); (55) follows from (30); and $\xi_{\mathsf{x},k}$ is defined in (48). The message $\overleftarrow{P}_{\mathsf{y},k}(s_{\mathsf{y},k})$, for $k = 0, 1, \ldots, N-1$, is derived in an analogous fashion.

## Appendix B
## Numerical Problems

The message in (47) can take on large values, which is problematic from an implementation standpoint. To manage this, the modified Bessel function $I_0$ can be approximated with an asymptotic expansion [41, p. 377],

$$I_0(z) \sim \frac{e^z}{\sqrt{2\pi z}}, z \in \mathbb{C}, \quad (57)$$

which is a close approximation for $|z| \gg 1$ and $|\arg\{z\}| < \frac{\pi}{2}$. In practice, (48) and (49) satisfy the former condition with high probability. Moreover, as the argument to $I_0$ in (47) is the modulus of a complex number, the latter condition is satisfied. Substituting (57) into (47) gives

$$\overleftarrow{P}_{w,k}(s_{w,k}) \approx \frac{\exp\left\{-\frac{|s_{w,k}|^2}{2\sigma_{\mathrm{n}}^2} + |\xi_{w,k}|\right\}}{\sqrt{2\pi|\xi_{w,k}|}}, \quad (58)$$

for $k = 0, 1 \ldots, N-1$ and $w \in \{\mathsf{x}, \mathsf{y}\}$. Taking the natural logarithm of (58) yields an expression that is numerically tractable.

As previously detailed, (47) is proportional to *a posteriori* symbol probabilities. For it to be used as *a priori* symbol probabilities in a consecutive iteration, converting it to a PMF is required. This entails scaling the message such that the sum of the expression, over all points in $\mathcal{M}$, is one. To this end, define a scaled version of (47) as

$$\overleftarrow{\tilde{P}}_{w,k}(s_{w,k}) \triangleq A \overleftarrow{P}_{w,k}(s_{w,k}), \quad (59)$$

for $k = 0, 1 \ldots, N-1$, $w \in \{\mathsf{x}, \mathsf{y}\}$, and any $A > 0$. Converting (59) to a PMF imposes the constraint

$$\sum_{s \in \mathcal{M}} \overleftarrow{\tilde{P}}_{w,k}(s) = A \sum_{s \in \mathcal{M}} \overleftarrow{P}_{w,k}(s) = 1. \quad (60)$$

Solving (60) for $A$, substituting it in (59), and taking the natural logarithm of both sides yields

$$\log \overleftarrow{\tilde{P}}_{w,k}(s_{w,k}) = \log \overleftarrow{P}_{w,k}(s_{w,k}) - \log \sum_{s \in \mathcal{M}} \overleftarrow{P}_{w,k}(s). \quad (61)$$

The latter term can be rewritten as

$$\log \sum_{s \in \mathcal{M}} \overleftarrow{P}_{w,k}(s)$$

$$= \log\left\{\max_{m \in \mathcal{M}} \overleftarrow{P}_{w,k}(m) \sum_{s \in \mathcal{M}} \frac{\overleftarrow{P}_{w,k}(s)}{\max_{m \in \mathcal{M}} \overleftarrow{P}_{w,k}(m)}\right\}$$

$$= \log \max_{m \in \mathcal{M}} \overleftarrow{P}_{w,k}(m) + \log \sum_{s \in \mathcal{M}} \frac{\overleftarrow{P}_{w,k}(s)}{\max_{m \in \mathcal{M}} \overleftarrow{P}_{w,k}(m)}$$

$$= \Psi_{w,k} + \log \sum_{s \in \mathcal{M}} \exp\left\{\log \overleftarrow{P}_{w,k}(s) - \Psi_{w,k}\right\}, \quad (62)$$

where (62) follows from the fact that the natural logarithm is a monotonically increasing function for positive arguments, and

$$\Psi_{w,k} \triangleq \max_{s \in \mathcal{M}} \log \overleftarrow{P}_{w,k}(s). \quad (63)$$

Substituting (62) in (61) yields

$$\overset{\leftarrow}{\tilde{P}}_{w,k}(s_{w,k}) = \exp\left\{\log\overset{\leftarrow}{P}_{w,k}(s_{w,k}) - \Psi_{w,k} - \log\sum_{s\in\mathcal{M}}\exp\left\{\log\overset{\leftarrow}{P}_{w,k}(s) - \Psi_{w,k}\right\}\right\}. \quad (64)$$

When the approximation in (58) is utilized, (64) is numerically tractable.

Symbols are detected according to (4), which gives

$$\hat{s}_{w,k} = \underset{s_{w,k}\in\mathcal{M}}{\operatorname{argmax}}\ \overset{\leftarrow}{\tilde{P}}_{w,k}(s_{w,k})$$

$$= \underset{s_{w,k}\in\mathcal{M}}{\operatorname{argmax}}\ \log\overset{\leftarrow}{P}_{w,k}(s_{w,k}) \quad (65)$$

$$\approx \underset{s_{w,k}\in\mathcal{M}}{\operatorname{argmax}}\left(-\frac{|s_{w,k}|^2}{2\sigma_{\mathrm{n}}^2} + |\xi_{w,k}| - \frac{1}{2}\log|\xi_{w,k}|\right), \quad (66)$$

where (65) follows from the fact that taking the natural logarithm of (64) does not affect the maximization. Moreover, $\log\overset{\leftarrow}{P}_{w,k}(s_{w,k})$ is the only term in (64) that depends on $s_{w,k}$. Lastly, (66) follows from using the approximation in (58). Thus, converting (47) to a PMF is not necessary for detection, as scaling does not affect maximization.

## Acknowledgment

The authors would like to thank H. Wymeersch, H. Eliasson, M. Karlsson, and P. Johannisson for very helpful discussions.

## References

[1] E. Ip, A. P. T. Lau, D. J. F. Barros, and J. M. Kahn, "Coherent detection in optical fiber systems," *Opt. Express*, vol. 16, no. 2, pp. 753–791, Jan. 2008.

[2] T. Pfau, S. Hoffmann, and R. Noé, "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.

[3] 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.

[4] 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.

[5] Z. Tao *et al.*, "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.

[6] S. Hoffmann, R. Peveling, T. Pfau, O. Adamczyk, R. Eickhoff, and R. Noé, "Multiplier-free real-time phase tracking for coherent QPSK receivers," *IEEE Photon. Technol. Lett.*, vol. 21, no. 3, pp. 137–139, Feb. 2009.

[7] 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.

[8] I. Fatadin, D. Ives, and S. J. Savory, "Laser linewidth tolerance for 16-QAM coherent optical systems using QPSK partitioning," *IEEE Photon. Technol. Lett.*, vol. 22, no. 9, pp. 631–633, May 2010.

[9] 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 Photon. Technol. Lett.*, vol. 26, no. 3, pp. 267–270, Feb. 2014.

[10] N. Argyris, S. Dris, C. Spatharakis, and H. Avramopoulos, "High performance carrier phase recovery for coherent optical QAM," presented at the *Proc. Optical Fiber Communication Conf.*, Los Angeles, CA, USA, Mar. 2015, Paper W1E.1.

[11] X. Zhou, "An improved feed-forward carrier recovery algorithm for coherent receivers with M-QAM modulation format," *IEEE Photon. Technol. Lett.*, vol. 22, no. 14, pp. 1051–1053, Jul. 2010.

[12] J. H. Ke, K. P. Zhong, Y. Gao, J. C. Cartledge, A. S. Karar, and M. A. Rezania, "Linewidth-tolerant and low-complexity two-stage carrier phase estimation for dual-polarization 16-QAM coherent optical fiber communications," *J. Lightw. Technol.*, vol. 30, no. 24, pp. 3987–3992, Dec. 2012.

[13] X. Zhou "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.

[14] 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. Express*, vol. 19, no. 24, pp. 24 331–24 343, Nov. 2011.

[15] M. Magarini "Pilot-symbols-aided carrier-phase recovery for 100-G PM-QPSK digital coherent receivers," *IEEE Photon. Technol. Lett.*, vol. 24, no. 9, pp. 739–741, May. 2012.

[16] 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 Photon. Technol. Lett.*, vol. 24, no. 18, pp. 1577–1580, Sep. 2012.

[17] H. Cheng *et al.*, "Pilot-symbols-aided cycle slip mitigation for DP-16QAM optical communication systems," *Opt. Express*, vol. 21, no. 19, pp. 22 166–22 172, Sep. 2013.

[18] 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," presented at the *Proc. Optical Fiber Communication Conf.*, Anaheim, CA, USA, Mar. 2013, Paper OTu3I.6.

[19] 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," presented at the Proc. European Conf. Optical Communication, Geneva, Switzerland, 2011, Paper Tu.3.A.3.

[20] F. Peng *et al.*, "Operational regime of symbol-by-symbol phase noise estimation for POLMUX 16-QAM," presented at the Proc. European Conf. Optical Communication, Torino, Italy, Sep. 2010, Paper We.7.A.5.

[21] M. Kuschnerov *et al.*, "Joint-polarization carrier phase estimation for XPM-limited coherent polarization-multiplexed QPSK transmission with OOK-neighbors," presented at the Proc. European Conf. Optical Communication, Brussels, Belgium, Sep. 2008, Paper Mo.4.D.2.

[22] R. R. Müller and D. A. De Arruda Mello, "Phase-offset estimation for joint-polarization phase-recovery in DP-16-QAM systems," *IEEE Photon. Technol. Lett.*, vol. 22, no. 20, pp. 1515–1517, Oct. 2010.

[23] J. G. Proakis, *Digital Communications*, 4th ed. Boston, MA, USA: McGraw-Hill, 2000.

[24] 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.

[25] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[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. Panayırcı, 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] M. A. Castrillón, D. A. Morero, O. E. Agazzi, and M. R. Hueda, "On the performance of joint iterative detection and decoding in coherent optical channels with laser frequency fluctuations," *Opt. Fiber Technol.*, vol. 24, pp. 5–14, 2015.

[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. Express*, vol. 19, no. 10, pp. 9868–9880, May 2011.

[32] R. Noé, T. Pfau, M. El-Darawy, and S. Hoffmann, "Electronic polarization control algorithms for coherent optical transmission," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 5, pp. 1193–1200, Sep. 2010.

[33] 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.

[34] 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.

[35] R. Krishnan, G. Colavolpe, A. G. I. Amat, and T. Eriksson, "Algorithms for joint phase estimation and decoding for MIMO systems in the presence of phase noise and quasi-static fading channels," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3360–3375, Jul. 2015.

[36] M. P. Yankov, T. Fehenberger, L. Barletta, and N. Hanik, "Low-complexity tracking of laser and nonlinear phase noise in WDM optical fiber systems," *J. Lightw. Technol.*, vol. 33, no. 23, pp. 4975–4984, Dec. 2015.

[37] S. Shayovitz and D. Raphaeli, "Message passing algorithms for phase noise tracking using Tikhonovmixtures," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 387–401, Jan. 2016.

[38] S. R. Jammalamadaka and A. Sengupta, *Topics in Circular Statistics (Series Multivariate analysis)*. River Edge, NJ, USA: World Scientific, 2001.

[39] N. Noels, J. Bhatti, H. Bruneel, and M. Moeneclaey, "Block-processing soft-input soft-output demodulator for coded PSK using DCT-based phase noise estimation," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2939–2950, Aug. 2014.

[40] 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.

[41] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables Series Applied mathematics)* 10th ed. Washington, DC, USA: U.S. Government Printing Office, 1972.

Authors' photographs and biographies not available at the time of publication.