CHALMERS
UNIVERSITY OF TECHNOLOGY

Iterative Detection and Phase-Noise Compensation for Coded Multichannel Optical Transmission

Downloaded from: https://research.chalmers.se, 2026-11-14 02:29 UTC

Citation for the original published paper (version of record):
Alfredsson, A., Agrell, E., Wymeersch, H. (2019). Iterative Detection and Phase-Noise Compensation for Coded Multichannel Optical Transmission. IEEE Transactions on Communications, 67(8): 5532-5543. http://dx.doi.org/10.1109/TCOMM.2019.2912569

N.B. When citing this work, cite the original published paper.

© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, or reuse of any copyrighted component of this work in other works.

(article starts on next page)

Iterative Detection and Phase-Noise Compensation for Coded Multichannel Optical Transmission

Arni F. Alfredsson, *Student Member, IEEE*, Erik Agrell, *Fellow, IEEE*, and Henk Wymeersch, *Member, IEEE*

***Abstract*—The problem of phase-noise compensation for correlated phase noise in coded multichannel optical transmission is investigated. To that end, a simple multichannel phase-noise model is considered and the maximum *a posteriori* detector for this model is approximated using two frameworks, namely factor graphs (FGs) combined with the sum–product algorithm (SPA), and a variational Bayesian (VB) inference method. The resulting pilot-aided algorithms perform iterative phase-noise compensation in cooperation with a decoder, using extended Kalman smoothing to estimate the *a posteriori* phase-noise distribution jointly for all channels. The system model and the proposed algorithms are verified using experimental data obtained from space-division multiplexed multicore-fiber transmission. Through Monte Carlo simulations, the algorithms are further evaluated in terms of phase-noise tolerance for coded transmission. It is observed that they significantly outperform the conventional approach to phase-noise compensation in the optical literature. Moreover, the FG/SPA framework performs similarly or better than the VB framework in terms of phase-noise tolerance of the resulting algorithms, for a slightly higher computational complexity.**

***Index Terms*—Detection, factor graph, phase noise, sum–product algorithm, variational Bayesian inference.**

I. Introduction

Phase noise is an inherent problem in optical communications. This comes due to the nonzero linewidth of light sources and local oscillators (LOs) [1], which degrades the system performance severely if not handled properly. This is particularly relevant since the research focus has shifted in recent years towards higher-order quadrature amplitude modulation (QAM) or more advanced multilevel modulation formats [2]. In general, systems become more sensitive to phase noise as the modulation order grows, and hence, effective phase-noise compensation becomes crucial. Traditionally, phase-noise compensation methods in optical communications have been blind, i.e., they do not use pilot symbols to assist with the estimation, and thus, spectral efficiency is not sacrificed. Well-known examples are the Viterbi–Viterbi algorithm [3] and blind phase search (BPS) [4]. However, blind methods suffer from ambiguity in the phase-noise estimation and are therefore susceptible to cycle slips, which result in burst errors [5] that can not be corrected with a code. This can be handled with differential encoding, which has the downside of increasing the average bit error rate (BER). Alternatively, one can resort to pilot-aided phase-noise estimation [6]–[8].

Recently, space-division multiplexing (SDM) has been a topic of particular interest. It involves the integration of various system components, such as optical hardware and digital signal processing (DSP) algorithms, as well as the utilization of multicore and multimode fibers, combinations thereof, or bundles of single-mode fibers [9]. These systems enable the joint-channel compensation of various signal impairments, in particular phase noise, as light sources and LOs can be shared between different cores or modes [10]. This gives rise to spatial correlation in the phase noise, which can be exploited to relax hardware requirements [11] and reduce receiver complexity [12]. However, the phase noise is not perfectly correlated, as environmental effects and system imperfections introduce phase drifts between cores and polarizations [10], [13]. In addition, correlated phase noise is present in wavelength-division multiplexed (WDM) systems utilizing frequency combs, as the combs enable phase locking between the different frequency lines. However, similarly to SDM systems, the phase noise is not perfectly correlated among the spectral channels, due to imperfections in the comb generation [14]. Joint-channel phase-noise compensation has also been demonstrated in systems utilizing electrically generated subcarriers [15]. Clearly, various types of optical systems are amenable to joint-channel phase-noise compensation, which is the target of this investigation.

Multiple solutions that exploit the phase-noise correlation between channels have been proposed for SDM and comb-based WDM systems. The majority has focused on schemes that facilitate DSP complexity reduction [10], [14]–[16], e.g., through the use of a master–slave strategy, in which one channel is used to produce phase-noise estimates that are used for compensation in all channels. However, the phase-noise correlation can also be used to improve performance in terms of increased phase-noise tolerance [17] or, in the case of pilot-aided schemes, lower the pilot rate while maintaining the same phase-noise tolerance.

Wireless communication systems face a similar problem with oscillator phase noise, which has given rise to a myriad of solutions that approximate the maximum *a posteriori* (MAP) detector for phase-noise channels. In particular, multiple algorithms that perform iterative phase-noise compensation and data detection have been proposed. Pertaining to single-channel transmission, blind and pilot-aided iterative solutions have been developed based on factor graphs (FGs) and the sum–product algorithm (SPA) [6], [8], [18]–[20], variational Bayesian (VB) inference [7], and the expectation-

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

This work was supported by the Swedish Research Council (VR) under Grants 2013-5642 and 2014-6138.

maximization algorithm [21], [22]. Moreover, in the context of multiple-input multiple-output (MIMO) systems for wireless transmission, various methods have been proposed for joint-channel phase-noise compensation [23]–[25]. In [26], [27], several algorithms were proposed for joint phase-noise estimation and data detection using the aforementioned frameworks, and in [28], [29], joint channel and phase-noise estimation for MIMO systems was proposed. The majority of the work has focused on multichannel models entailing oscillator phase noise that is either identical or independent between antennas in the MIMO system, in addition to channel mixing.

In this paper, we consider coded multichannel optical transmission in the presence of correlated phase noise and propose algorithms for iterative phase-noise estimation and data detection, which has yet to be addressed in the context of optical communications. The contributions can be summarized as follows: (i) We consider a simple multichannel system model that entails arbitrarily correlated phase noise and approximates the received signal after it has been processed by a typical DSP chain prior to phase-noise compensation, which has not been considered for certain multichannel optical transmission scenarios such as multicore-fiber systems, with the purpose of facilitating algorithm development for joint detection and phase-noise compensation. Using this model, we use two frameworks that have proven to be effective for similar system models in wireless communications [6], [7], [26], namely FG/SPA [30] and VB inference [31], to develop algorithms that perform iterative detection and phase-noise estimation in fiber-optical transmission. These frameworks yield different interfaces between the phase-noise estimation and the decoding. A trade-off in terms of performance and computational complexity is observed between the resulting algorithms; (ii) The phase-noise estimation consists of the extended Kalman filter (EKF) and the Rauch–Tung–Striebel smoother (RTSS). We show that the standard EKF equations can be simplified for the system model, and that the linearization imposed by the EKF on the system model does not penalize the phase-noise estimation performance for practical baud rates and laser linewidths; (iii) Finally, we demonstrate the effectiveness of the system model and the proposed algorithms using experimental data from SDM transmission through a weakly-coupled, homogeneous, multicore fiber. We then further evaluate the algorithms through Monte Carlo simulations of coded transmission in terms of phase-noise tolerance, and show that they significantly outperform the approach that is typically used in the optical literature: Per-channel phase-noise compensation using BPS and symbol detection based on minimum Euclidean distance.

Notation: The real part, imaginary part, complex conjugate, and angle of a complex number are denoted with $\Re\{\cdot\}$, $\Im\{\cdot\}$, $(\cdot)^*$, and $\angle(\cdot)$, resp. Random variables and their realizations are denoted with X and x, resp. The expectation of a random variable X with respect to a distribution P is written as $\mathsf{E}_P[X]$. Probability mass functions (PMFs) are written as $P(x)$, whereas probability density functions (PDFs) and mixed discrete–continuous distributions are denoted with $p(x)$. In particular, a multivariate real Gaussian PDF with mean $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$, and argument $\boldsymbol{x}$ is denoted with $\mathcal{N}_{\boldsymbol{x}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and its complex counterpart with argument $\boldsymbol{z}$ is written as $\mathcal{CN}_{\boldsymbol{z}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Scalars, vectors, and matrices are typeset as x, $\boldsymbol{x}$, and $\mathbf{X}$, resp. A diagonal matrix is denoted with $\mathrm{diag}(\cdot)$, whereas the identity matrix of size D is written as $\mathbf{I}_D$. Finally, the vector transpose is denoted with $(\cdot)^T$.

II. System Model

Transmission over D parallel complex-valued channels is considered. The transmitted frame in each channel is modeled as a vector of N random variables that take on values from a set $\mathcal{X}$ of constellation points. The constellation is normalized such that the mean of the constellation points is zero and the average symbol energy is E_s. The received signal is assumed to have undergone standard DSP steps such as resampling, electronic dispersion compensation, orthonormalization, timing recovery, adaptive equalization, and frequency-offset estimation. Furthermore, the adaptive equalization is assumed to have been carried out in a phase-immune fashion using, e.g., a radially-directed equalizer [32]. With these assumptions in place, the received signal in each channel, after further resampling to one sample per symbol, is approximated as the output of a system entailing additive white Gaussian noise (AWGN) and phase noise. In single-channel optical systems, this is also a typical assumption of the processed signal prior to phase-noise compensation [4], [5], [33]. The discrete-time complex baseband model is therefore written as

$$r_{i,k} = s_{i,k} e^{j\theta_{i,k}} + n_{i,k}, \tag{1}$$

where $k = 1, \ldots, N$ is a time index and $i = 1, \ldots, D$ is a channel index. The received samples, transmitted symbols, phase noise, and complex AWGN realizations are denoted with $r_{i,k}$, $s_{i,k}$, $\theta_{i,k}$, and $n_{i,k}$, resp., where the complex AWGN on channel i has variance σ_i^2 per real dimension. The vector $\boldsymbol{r}_k = [r_{1,k}, \ldots, r_{D,k}]^T \in \mathbb{C}^D$ contains the received samples in all channels at time k, and the vectors $\boldsymbol{s}_k$, $\boldsymbol{\theta}_k$, and $\boldsymbol{n}_k$ are defined analogously. Finally, let $\boldsymbol{r}$ contain all received samples, and $\boldsymbol{s}$, $\boldsymbol{\theta}$, and $\boldsymbol{n}$ be defined similarly.

It has been experimentally demonstrated that for optical multichannel transmission through, e.g., optical frequency combs [34] or multicore fibers [10] with shared lasers, the phase noise is highly correlated among the channels, albeit not fully owing to per-channel phase drifts. These drifts are caused by environmental effects as well as imperfections in hardware and DSP and are much slower than the laser phase noise, which is typically modeled as a Gaussian random walk. To describe the laser phase noise in addition to slower channel-specific drifts, we opt for a simplistic phase-noise model to facilitate the development of high-performance algorithms based on detection and estimation theory[1]. Thus, the phase noise across all channels is approximated with a multidimensional Gaussian random walk, as

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + \Delta\boldsymbol{\theta}_k, \tag{2}$$

[1]Whenever simplistic models are used for algorithm design, it is imperative to assess the efficacy of the model and the resulting algorithms with either more realistic models or experimental data. In Sec. IV-B, the latter approach is taken.

where $\boldsymbol{\theta}_1$ is uniformly distributed on $[0, 2\pi)^D$ and $\Delta\boldsymbol{\theta}_k$ is a multivariate zero-mean Gaussian random variable with covariance matrix $\mathbf{Q}$. This covariance matrix depends on the system and has to be known or estimated. Furthermore, the phase noise is assumed to be independent of the transmitted symbols and AWGN, and unknown to the transmitter and receiver. Finally, $\mathbf{Q}$ and $\boldsymbol{\sigma}^2 = [\sigma_1^2, \ldots, \sigma_D^2]$ are assumed known to the receiver.

The transmitted symbol sequence $\boldsymbol{s}$ is assumed to be obtained through an encoding function $f(\cdot)$ that maps an equiprobable information bit sequence $\boldsymbol{b}$ of length N_b to a data symbol sequence of length ND. Additional pilot symbols that are used for the phase-noise compensation are considered as a part of the channel code. The pilot distribution is assumed known to the transmitter and receiver.

III. Derivation of Algorithms

The MAP detector yields the lowest possible BER out of all detectors [35, Ch. 1.4]. It performs detection on a bit-by-bit basis according to

$$\hat{b}_l = \underset{b_l \in \{0,1\}}{\operatorname{argmax}} P(b_l | \boldsymbol{r}), \tag{3}$$

for $l = 1, \ldots, N_\mathrm{b}$, where $P(b_l|\boldsymbol{r})$ is the *a posteriori* PMF of b_l. However, the PMF in (3) is hard to compute exactly for the system model in (1), and thus, approximations are needed. To that end, note that the desired PMF in (3) can be obtained through the marginalization

$$P(b_l|\boldsymbol{r}) = \int_{\mathbb{R}^{D\times N}} \sum_{\boldsymbol{s}\in\mathcal{X}^{D\times N}} \sum_{\boldsymbol{b}\in\mathcal{U}_l(b_l)} p(\boldsymbol{b}, \boldsymbol{s}, \boldsymbol{\theta}|\boldsymbol{r}) \mathrm{d}\boldsymbol{\theta}, \tag{4}$$

where $\mathcal{U}_l(b_l) = \left\{\boldsymbol{b}' \in \{0,1\}^{N_\mathrm{b}} : b'_l = b_l\right\}$. To carry out this marginalization approximately but efficiently, we make use of two frameworks, namely FG/SPA and VB inference, motivated by the fact that they have been found effective in earlier work pertaining to phase-noise compensation in wireless applications [6]–[8], [26]. Conceptually, the resulting receiver algorithms comprise three blocks: (i) A standard iterative decoder, which has as input probabilities on bits (assuming the use of a binary code); (ii) A scheme that estimates the phase-noise distribution and has as input symbol probabilities; (iii) An interface, developed using the aforementioned frameworks, that connects the decoder and the phase-noise estimation scheme.

As the AWGN is assumed to be independent between channels, moving from single-channel to multichannel transmission does not complicate the computation of bit-wise probabilities that are fed to the decoder, given that the phase-noise distribution is known at each symbol. However, assuming the phase noise to be arbitrarily correlated across the channels in multichannel transmission prevents trivial extensions of various existing single-channel phase-noise estimation methods, such as the well-known Tikhonov algorithm in [6, Sec. IV-B]. In order to estimate the phase-noise distribution in the multichannel case, one can resort to, e.g., an extended Kalman smoother (EKS), an unscented Kalman smoother, or a particle smoother. In this work, we opt for the EKS, which is naturally suited to estimate the marginal *a posteriori* PDFs at each time k of a multidimensional Gaussian random walk. This approach proved to be successful in wireless communications [29], where one of the proposed algorithms included a soft-input EKS capable of estimating spatially correlated phase-noise processes. In this work, we apply a soft-input EKS in multichannel fiber-optical communications, and accordingly use a somewhat different system model as well as a different method to derive the soft inputs to the EKS compared with [29]. The marginal phase-noise distributions are approximated as multivariate Gaussian PDFs by the EKS; hence, they can be further marginalized in a low-complexity manner to yield the phase-noise distribution at each symbol. This also facilitates the derivations of the interfaces connecting the decoder and the phase-noise estimation, as will be shown in the following subsections.

The EKS consists of recursive equations that are used to approximately estimate the marginal *a posteriori* phase-noise PDFs, $p(\boldsymbol{\theta}_k|\boldsymbol{r})$, for $k = 1, \ldots, N$. This is accomplished through two recursive passes; a forward pass with an EKF [36, Ch. 5.2] and a backward pass with an RTSS [36, Ch. 8.2]. The resulting recursive equations are used by both frameworks to yield the final algorithms. A further justification for utilizing an EKF is that the linearization imposed by the EKF on the system model in (1) works effectively, provided that the phase noise does not vary too rapidly. For practical parameters, this is indeed the case, as illustrated in Section IV-A.

A. Phase-Noise Estimation

The EKF estimates the marginal PDFs $p(\boldsymbol{\theta}_k|\boldsymbol{r}_1, \ldots, \boldsymbol{r}_k)$ for $k = 2, \ldots, N$. However, instead of applying the standard equations [36, Ch. 5.2] directly, it is shown in Appendix A[2] that they can be reduced to

$$\mathbf{M}^\mathrm{f}_{k|k-1} = \mathbf{M}^\mathrm{f}_{k-1} + \mathbf{Q}, \tag{5}$$

$$\mathbf{M}^\mathrm{f}_{k} = (\mathbf{I}_D + \mathbf{M}^\mathrm{f}_{k|k-1}\mathbf{V}_k)^{-1}\mathbf{M}^\mathrm{f}_{k|k-1}, \tag{6}$$

$$\hat{\boldsymbol{\theta}}^\mathrm{f}_{k} = \hat{\boldsymbol{\theta}}^\mathrm{f}_{k-1} + \mathbf{M}^\mathrm{f}_{k}\boldsymbol{h}_k, \tag{7}$$

with initial conditions

$$\hat{\boldsymbol{\theta}}^\mathrm{f}_1 = \left[\angle(r_{1,1}s^*_{1,1}), \ldots, \angle(r_{D,1}s^*_{D,1})\right]^T, \tag{8}$$

$$\mathbf{M}^\mathrm{f}_1 = \operatorname{diag}\left(\frac{\sigma_1^2}{E_\mathrm{s}}, \ldots, \frac{\sigma_D^2}{E_\mathrm{s}}\right), \tag{9}$$

where $\mathbf{V}_k$ and each element of $\boldsymbol{h}_k = [h_{1,k}, \ldots, h_{D,k}]^T$ in (7) are computed as

$$\mathbf{V}_k = \operatorname{diag}\left(\frac{|s_{1,k}|^2}{\sigma_1^2}, \ldots, \frac{|s_{D,k}|^2}{\sigma_D^2}\right), \tag{10}$$

$$h_{i,k} = \frac{1}{\sigma_i^2}\Im\left\{r_{i,k}s^*_{i,k}e^{-j\hat{\theta}^\mathrm{f}_{i,k-1}}\right\}. \tag{11}$$

The rationale behind the initialization of $\mathbf{M}^\mathrm{f}_1$ is the fact that for practical signal-to-noise ratios (SNRs), the variance of the elements of $\hat{\boldsymbol{\theta}}^\mathrm{f}_1$ will be approximately half the variance of the complex AWGN. Moreover, $\mathbf{Q}$ is the covariance matrix of

[2]See also [25] for an alternative derivation of the EKF equations for a random-walk phase noise with complex-valued observations in the context of wireless MIMO transmission with independent oscillators.

$\Delta\boldsymbol{\theta}_k$ in (2), whereas $\hat{\boldsymbol{\theta}}^{\mathrm{f}}_{k-1} = [\hat{\theta}^{\mathrm{f}}_{1,k-1}, \ldots, \hat{\theta}^{\mathrm{f}}_{D,k-1}]^T$ denotes an estimate of $\boldsymbol{\theta}_{k-1}$ based on all received samples up to and including time $k-1$. Note that $s_{i,k}$ is assumed to be known for all i and k in (5)–(10). This is not true for (i,k) corresponding to data symbols and will be handled in the following subsections.

To estimate $p(\boldsymbol{\theta}_k|\boldsymbol{r}) = p(\boldsymbol{\theta}_k|\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)$, the RTSS is used. The resulting recursive equations are given by [36, Ch. 8.2]

$$\mathbf{A}_k = \mathbf{M}^{\mathrm{f}}_k (\mathbf{M}^{\mathrm{f}}_{k+1|k})^{-1}, \tag{12}$$

$$\hat{\boldsymbol{\theta}}^{\mathrm{s}}_k = \hat{\boldsymbol{\theta}}^{\mathrm{f}}_k + \mathbf{A}_k (\hat{\boldsymbol{\theta}}^{\mathrm{s}}_{k+1} - \hat{\boldsymbol{\theta}}^{\mathrm{f}}_k), \tag{13}$$

$$\mathbf{M}^{\mathrm{s}}_k = \mathbf{M}^{\mathrm{f}}_k + \mathbf{A}_k (\mathbf{M}^{\mathrm{s}}_{k+1} - \mathbf{M}^{\mathrm{f}}_{k+1|k}) \mathbf{A}^T_k, \tag{14}$$

for $k = N-1, N-2, \ldots, 1$, with initial conditions

$$\hat{\boldsymbol{\theta}}^{\mathrm{s}}_N = \hat{\boldsymbol{\theta}}^{\mathrm{f}}_N, \tag{15}$$

$$\mathbf{M}^{\mathrm{s}}_N = \mathbf{M}^{\mathrm{f}}_N. \tag{16}$$

Thus, $\hat{\boldsymbol{\theta}}^{\mathrm{s}}_k = [\theta^{\mathrm{s}}_{1,k}, \ldots, \theta^{\mathrm{s}}_{D,k}]^T$ and $\mathbf{M}^{\mathrm{s}}_k$ represent the mean vector and covariance matrix of the Gaussian approximation of $p(\boldsymbol{\theta}_k|\boldsymbol{r})$, i.e., $p(\boldsymbol{\theta}_k|\boldsymbol{r}) \approx \mathcal{N}_{\boldsymbol{\theta}_k}(\hat{\boldsymbol{\theta}}^{\mathrm{s}}_k, \mathbf{M}^{\mathrm{s}}_k)$. The EKF and RTSS equations are summarized in Algorithm 1.

Algorithm 1 EKS

Input: $\boldsymbol{r}$, $\boldsymbol{s}$, D, N, $\mathbf{Q}$, $\boldsymbol{\sigma}^2$
Output: $(\hat{\boldsymbol{\theta}}^{\mathrm{s}}_k, \mathbf{M}^{\mathrm{s}}_k)\ \forall\ k$

$\hat{\boldsymbol{\theta}}^{\mathrm{f}}_1 = \left[\angle(r_{1,1}s^*_{1,1}), \ldots, \angle(r_{D,1}s^*_{D,1})\right]^T$
$\mathbf{M}^{\mathrm{f}}_1 = \mathrm{diag}(\sigma^2_1/E_{\mathrm{s}}, \ldots, \sigma^2_D/E_{\mathrm{s}})$
for $k = 2, \ldots, N$ **do**
  for $i = 1, \ldots, D$ **do**
    $h_{i,k} = \Im\{r_{i,k}s^*_{i,k}e^{-j\hat{\theta}^{\mathrm{f}}_{i,k-1}}\}/\sigma^2_i$
  end for
  $\mathbf{V}_k = \mathrm{diag}(|s_{1,k}|^2/\sigma^2_1, \ldots, |s_{D,k}|^2/\sigma^2_D)$
  $\mathbf{M}^{\mathrm{f}}_{k|k-1} = \mathbf{M}^{\mathrm{f}}_{k-1} + \mathbf{Q}$
  $\mathbf{M}^{\mathrm{f}}_k = (\mathbf{I}_D + \mathbf{M}^{\mathrm{f}}_{k|k-1}\mathbf{V}_k)^{-1}\mathbf{M}^{\mathrm{f}}_{k|k-1}$
  $\hat{\boldsymbol{\theta}}^{\mathrm{f}}_k = \hat{\boldsymbol{\theta}}^{\mathrm{f}}_{k-1} + \mathbf{M}^{\mathrm{f}}_k \boldsymbol{h}_k$
end for
$(\hat{\boldsymbol{\theta}}^{\mathrm{s}}_N, \mathbf{M}^{\mathrm{s}}_N) = (\hat{\boldsymbol{\theta}}^{\mathrm{f}}_N, \mathbf{M}^{\mathrm{f}}_N)$
for $k = N-1, N-2, \ldots, 1$ **do**
  $\mathbf{A}_k = \mathbf{M}^{\mathrm{f}}_k (\mathbf{M}^{\mathrm{f}}_{k+1|k})^{-1}$
  $\hat{\boldsymbol{\theta}}^{\mathrm{s}}_k = \hat{\boldsymbol{\theta}}^{\mathrm{f}}_k + \mathbf{A}_k (\hat{\boldsymbol{\theta}}^{\mathrm{s}}_{k+1} - \hat{\boldsymbol{\theta}}^{\mathrm{f}}_k)$
  $\mathbf{M}^{\mathrm{s}}_k = \mathbf{M}^{\mathrm{f}}_k + \mathbf{A}_k (\mathbf{M}^{\mathrm{s}}_{k+1} - \mathbf{M}^{\mathrm{f}}_{k+1|k}) \mathbf{A}^T_k$
end for

B. FG/SPA-Based Algorithm

The FG and SPA framework carries out the marginalization of joint distributions in an efficient manner by exploiting how they can be factorized into simpler functions. For a thorough introduction to this framework, refer to [30]. Moreover, in [6], the framework was applied to the problem of data detection for single-channel satellite transmission in the presence of phase noise. In this paper, we use the framework in a similar manner as in [6], but note that the phase-noise estimation is carried out using a different technique.

The joint distribution $p(\boldsymbol{b}, \boldsymbol{s}, \boldsymbol{\theta}|\boldsymbol{r})$ in (4) factorizes as

$$\begin{aligned} p(\boldsymbol{b}, \boldsymbol{s}, \boldsymbol{\theta}|\boldsymbol{r}) &\propto p(\boldsymbol{b}, \boldsymbol{s}, \boldsymbol{\theta}) p(\boldsymbol{r}|\boldsymbol{b}, \boldsymbol{s}, \boldsymbol{\theta}) \\ &= P(\boldsymbol{s}|\boldsymbol{b}) P(\boldsymbol{b}) p(\boldsymbol{\theta}) p(\boldsymbol{r}|\boldsymbol{s}, \boldsymbol{\theta}) \\ &= P(\boldsymbol{s}|\boldsymbol{b}) P(\boldsymbol{b}) p(\boldsymbol{\theta}_1) \prod_{k=2}^{N} p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1}) \prod_{i,k} p(r_{i,k}|s_{i,k}, \theta_{i,k}) \\ &\propto P(\boldsymbol{s}|\boldsymbol{b}) \prod_{k=2}^{N} p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1}) \prod_{i,k} p(r_{i,k}|s_{i,k}, \theta_{i,k}), \end{aligned} \tag{17}$$

for $i = 1, \ldots, D$ and $k = 1, \ldots, N$, where $\propto$ denotes proportionality with respect to $\boldsymbol{b}$, $\boldsymbol{s}$, and $\boldsymbol{\theta}$. Moreover, (17) is obtained using the fact that $\boldsymbol{\theta}$ is a random walk, $P(\boldsymbol{b})$ and $p(\boldsymbol{\theta}_1)$ are uniform distributions, and the received samples are independent of each other given $\boldsymbol{s}$ and $\boldsymbol{\theta}$. Finally, $P(\boldsymbol{s}|\boldsymbol{b})$ represents the code and mapping constraints, i.e., $P(\boldsymbol{s}|\boldsymbol{b})$ equals 1 if $\boldsymbol{s} = f(\boldsymbol{b})$ and 0 otherwise, $p(r_{i,k}|s_{i,k}, \theta_{i,k}) = \mathcal{CN}_{r_{i,k}}\left(s_{i,k}e^{j\theta_{i,k}}, 2\sigma^2_i\right)$ is the likelihood function of $s_{i,k}$ and $\theta_{i,k}$, and $p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1}) = \mathcal{N}_{\boldsymbol{\theta}_k}(\boldsymbol{\theta}_{k-1}, \mathbf{Q})$.

The FG associated with (17) is shown partially in Fig. 1. In the case of coded transmission where an iterative decoder is utilized, e.g., for low-density parity-check (LDPC) codes and turbo codes, applying the SPA to this FG will yield an iterative phase-noise estimation and decoding algorithm, where in each iteration, the message $P_{\mathrm{d}}(s_{i,k})$ corresponds to the extrinsic information about $s_{i,k}$ given by the decoder. Moreover, $P_{\mathrm{u}}(s_{i,k})$ can be regarded as the likelihood function of $s_{i,k}$ from a virtual memoryless phase-noise compensated channel, and is fed to the decoder as bit-wise log-likelihood ratios (LLRs) [35, Ch. 5.4]. The decoder then either outputs the detected information bits or extrinsic coded-bit LLRs that are converted to $P_{\mathrm{d}}(s_{i,k})$[3]. By applying the SPA, it can be shown that

$$P_{\mathrm{u}}(s_{i,k}) \propto \int_{\mathbb{R}^D} p(r_{i,k}|s_{i,k}, \theta_{i,k}) \frac{p_{\mathrm{APP}}(\boldsymbol{\theta}_k)}{p_{\mathrm{d}}(\theta_{i,k})} \mathrm{d}\boldsymbol{\theta}_k, \tag{18}$$

where

$$p_{\mathrm{APP}}(\boldsymbol{\theta}_k) \propto \int_{\mathbb{R}^{D\times(N-1)}} \prod_{k=2}^{N} p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1}) \prod_{i,k} p_{\mathrm{d}}(\theta_{i,k}) d\bar{\boldsymbol{\theta}}_k \tag{19}$$

represents the *a posteriori* PDF of $\boldsymbol{\theta}_k$, $p(\boldsymbol{\theta}_k|\boldsymbol{r})$, and $\bar{\boldsymbol{\theta}}_k$ contains all elements of $\boldsymbol{\theta}$ except $\boldsymbol{\theta}_k$.

1) Interface Towards Phase-Noise Estimation: The message $p_{\mathrm{APP}}(\boldsymbol{\theta}_k)$ is approximated using Algorithm 1. However, the utilization of the EKF requires the likelihood $p(r_{i,k}|\theta_{i,k})$, represented by $p_{\mathrm{d}}(\theta_{i,k})$, to be a complex Gaussian PDF in $r_{i,k}$. For pilot symbols, this is indeed the case since their values are known. For data symbols, however, $p_{\mathrm{d}}(\theta_{i,k})$ is a mixture of

[3]For uncoded transmission, $P_{\mathrm{d}}(s_{i,k})$ is simply the *a priori* PMF of $s_{i,k}$, and since the FG corresponding to (17) does not contain any cycles in the absence of a code [26], the algorithm that results from applying the SPA will not be iterative. As the information bits are assumed equiprobable, $P_{\mathrm{u}}(s_{i,k})$ is then used to detect the symbols, followed by symbol-to-bit demapping.

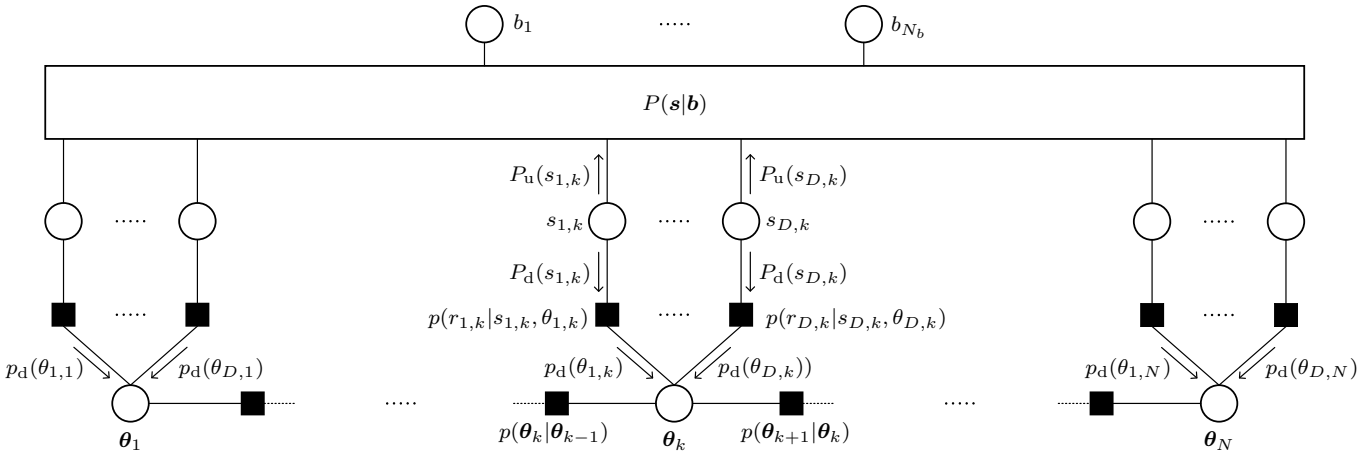

Fig. 1. A part of the FG corresponding to (17).

complex Gaussian PDFs since

$$p_{\mathrm{d}}(\theta_{i,k}) = \sum_{s_{i,k}\in\mathcal{X}} P_{\mathrm{d}}(s_{i,k}) p(r_{i,k}|s_{i,k},\theta_{i,k}) \tag{20}$$

$$= \sum_{s_{i,k}\in\mathcal{X}} P_{\mathrm{d}}(s_{i,k}) \mathcal{CN}_{r_{i,k}}\big(s_{i,k}e^{j\theta_{i,k}}, 2\sigma_i^2\big). \tag{21}$$

To solve this, the same approach is taken as in [6], i.e., $p_{\mathrm{d}}(\theta_{i,k})$ is approximated as a Gaussian PDF with mean and variance chosen such that the Kullback–Leibler (KL) divergence [37, Ch. 8.5] between the Gaussian mixture and the single Gaussian is minimized. This yields [38]

$$p_{\mathrm{d}}(\theta_{i,k}) \approx \mathcal{CN}_{r_{i,k}}\big(\bar{s}_{i,k}e^{j\theta_{i,k}}, 2\bar{\sigma}_{i,k}^2\big), \tag{22}$$

where

$$\bar{s}_{i,k} \triangleq \mathsf{E}_{P_{\mathrm{d}}}[S_{i,k}] = \sum_{s_{i,k}\in\mathcal{X}} s_{i,k} P_{\mathrm{d}}(s_{i,k}), \tag{23}$$

$$\bar{\sigma}_{i,k}^2 \triangleq \sigma_i^2 + \frac{1}{2}\mathsf{Var}(S_{i,k})$$

$$= \sigma_i^2 + \frac{1}{2}\sum_{s_{i,k}\in\mathcal{X}} |s_{i,k} - \bar{s}_{i,k}|^2 P_{\mathrm{d}}(s_{i,k}), \tag{24}$$

are soft inputs used by the EKF. For pilot symbols, $\bar{s}_{i,k} = s_{i,k}$ and $\bar{\sigma}_{i,k}^2 = \sigma_i^2$, whereas for data symbols, $\bar{s}_{i,k}$ and $\bar{\sigma}_{i,k}^2$ are initialized as $\bar{s}_{i,k} = 0$ and $\bar{\sigma}_{i,k}^2 = \sigma_i^2 + E_{\mathrm{s}}/2$. Due to (22), $\mathbf{V}_k$ in (6) and each component of $\boldsymbol{h}_k$ in (7) are now computed as

$$\mathbf{V}_k = \mathrm{diag}\left(\frac{|\bar{s}_{1,k}|^2}{\bar{\sigma}_{1,k}^2}, \ldots, \frac{|\bar{s}_{D,k}|^2}{\bar{\sigma}_{D,k}^2}\right), \tag{25}$$

$$h_{i,k} = \frac{1}{\bar{\sigma}_{i,k}^2}\Im\Big\{r_{i,k}\bar{s}_{i,k}^* e^{-j\hat{\theta}_{i,k-1}^{\mathrm{f}}}\Big\}, \tag{26}$$

and the EKF equations are initialized with

$$\hat{\boldsymbol{\theta}}_1^{\mathrm{f}} = \big[\angle(r_{1,1}\bar{s}_{1,1}^*), \ldots, \angle(r_{D,1}\bar{s}_{D,1}^*)\big]^T, \tag{27}$$

$$\mathbf{M}_1^{\mathrm{f}} = \mathrm{diag}\left(\frac{\bar{\sigma}_{1,k}^2}{E_{\mathrm{s}}}, \ldots, \frac{\bar{\sigma}_{D,k}^2}{E_{\mathrm{s}}}\right). \tag{28}$$

Hence,

$$p_{\mathrm{APP}}(\boldsymbol{\theta}_k) \approx \mathcal{N}_{\theta_{i,k}}\Big(\hat{\theta}_{i,k}^{\mathrm{s}}, M_{i,k}^{\mathrm{s}}\Big). \tag{29}$$

2) Interface Towards Decoder: Using (29), it is shown in Appendix B that $P_{\mathrm{u}}(s_{i,k})$ can be described approximately in closed form as

$$P_{\mathrm{u}}(s_{i,k}) \propto \frac{1}{\sqrt{|\xi_{i,k}(s_{i,k})|}} \exp\left(|\xi_{i,k}(s_{i,k})| - \frac{|s_{i,k}|^2}{2\sigma_i^2}\right), \tag{30}$$

with

$$\xi_{i,k}(s_{i,k}) \triangleq \frac{e^{j\hat{\theta}_{i,k}^{\mathrm{s}}}}{M_{i,k}^{\mathrm{s}}} + \frac{r_{i,k}s_{i,k}^*}{\sigma_i^2} - \frac{r_{i,k}\bar{s}_{i,k}^*}{\bar{\sigma}_{i,k}^2}, \tag{31}$$

where $M_{i,k}^{\mathrm{s}}$ denotes the ith element on the diagonal line of $\mathbf{M}_k^{\mathrm{s}}$. Note that the expression in (30) only describes $P_{\mathrm{u}}(s_{i,k})$ up to a constant and needs to be normalized. Furthermore, it is numerically unstable, and thus, the logarithm of (30) can be computed instead, yielding

$$f_{i,k}(s_{i,k}) \triangleq \ln P_{\mathrm{u}}(s_{i,k})$$

$$\approx |\xi_{i,k}(s_{i,k})| - \frac{|s_{i,k}|^2}{2\sigma_i^2} - \frac{1}{2}\ln|\xi_{i,k}(s_{i,k})|. \tag{32}$$

Finally, $P_{\mathrm{u}}(s_{i,k})$ is computed from $f_{i,k}(s_{i,k})$ as

$$P_{\mathrm{u}}(s_{i,k}) = \frac{\exp\Big(f_{i,k}(s_{i,k}) - f_{i,k}^{\max}\Big)}{\sum_{s'_{i,k}\in\mathcal{X}}\exp\Big(f_{i,k}(s'_{i,k}) - f_{i,k}^{\max}\Big)}, \tag{33}$$

where $f_{i,k}^{\max} \triangleq \max_{s_{i,k}\in\mathcal{X}} f_{i,k}(s_{i,k})$. This algorithm will from now on be referred to as FG/SPA-based phase-noise compensation (FG-PNC) and is summarized in Algorithm 2 for a single iteration.

C. VB-Based Algorithm

The VB inference framework has been used to develop effective algorithms that perform iterative phase-noise estimation and data detection for single-channel transmission [7] and wireless MIMO transmission [26]. Here, this framework is exploited in similar way to efficiently solve the marginalization problem in (4). It is worth noting, however, that in [7, Sec. IV-C], the VB algorithm is defined for a linearized system model, whereas in this work, we define the algorithm for the system model in (1). To that end, $p(\boldsymbol{b}, \boldsymbol{s}, \boldsymbol{\theta}|\boldsymbol{r})$ in (4) is approximated with a family of factorized distributions, $q_{\boldsymbol{B},\boldsymbol{S}}(\boldsymbol{b}, \boldsymbol{s}) q_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$,

Algorithm 2 FG-PNC (1 iteration)

Input: $\boldsymbol{r}, D, N, \mathbf{Q}, \boldsymbol{\sigma}^2, \mathcal{X}, P_{\mathrm{d}}(s_{i,k})\ \forall\ i,k$
Output: $P_{\mathrm{u}}(s_{i,k})\ \forall\ i,k$

for all i,k **do**
  $\bar{s}_{i,k} = \sum_{s_{i,k}\in\mathcal{X}} s_{i,k} P_{\mathrm{d}}(s_{i,k})$
  $\bar{\sigma}^2_{i,k} = \sigma^2_i + \frac{1}{2}\sum_{s_{i,k}\in\mathcal{X}} |s_{i,k} - \bar{s}_{i,k}|^2 P_{\mathrm{d}}(s_{i,k})$
end for
Compute $(\boldsymbol{\theta}^{\mathrm{s}}_k, \mathbf{M}^{\mathrm{s}}_k)\ \forall\ k$ with Alg. 1 using (25)–(28)
for all i,k **do**
  for all $s_{i,k}\in\mathcal{X}$ **do**
    $\xi_{i,k}(s_{i,k}) = \frac{e^{j\hat{\theta}^{\mathrm{s}}_{i,k}}}{M^{\mathrm{s}}_{i,k}} + \frac{r_{i,k}s^*_{i,k}}{\sigma^2_i} - \frac{r_{i,k}\bar{s}^*_{i,k}}{\bar{\sigma}^2_{i,k}}$
    $f_{i,k}(s_{i,k}) = |\xi_{i,k}(s_{i,k})| - \frac{|s_{i,k}|^2}{2\sigma^2_i} - \frac{1}{2}\ln|\xi_{i,k}(s_{i,k})|$
  end for
  $P_{\mathrm{u}}(s_{i,k}) = \exp(f_{i,k}(s_{i,k}) - \max_{s'_{i,k}\in\mathcal{X}} f_{i,k}(s'_{i,k}))$
  Normalize $P_{\mathrm{u}}(s_{i,k})$ s.t. $\sum_{s_{i,k}\in\mathcal{X}} P_{\mathrm{u}}(s_{i,k}) = 1$
end for

where $q_{\boldsymbol{B},\boldsymbol{S}}(\boldsymbol{b},\boldsymbol{s})$ and $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ represent $P(\boldsymbol{b},\boldsymbol{s}|\boldsymbol{r})$ and $p(\boldsymbol{\theta}|\boldsymbol{r})$, resp. The objective is then to minimize the KL divergence between $q_{\boldsymbol{B},\boldsymbol{S}}(\boldsymbol{b},\boldsymbol{s})q_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ and $p(\boldsymbol{b},\boldsymbol{s},\boldsymbol{\theta}|\boldsymbol{r})$, i.e.,

$$\min_{q_{\boldsymbol{B},\boldsymbol{S}}(\boldsymbol{b},\boldsymbol{s}),q_{\boldsymbol{\Theta}}(\boldsymbol{\theta})} D(q_{\boldsymbol{B},\boldsymbol{S}}(\boldsymbol{b},\boldsymbol{s})q_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\,||\,p(\boldsymbol{b},\boldsymbol{s},\boldsymbol{\theta}|\boldsymbol{r})). \tag{34}$$

This minimization is carried out by iteratively updating $q_{\boldsymbol{B},\boldsymbol{S}}(\boldsymbol{b},\boldsymbol{s})$ and $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$, and it can be shown that the update equations at the lth iteration are expressed as [31, Ch. 2.3]

$$\begin{aligned} q^{(l)}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) &\propto \exp\Big(\mathsf{E}_{q^{(l-1)}_{\boldsymbol{B},\boldsymbol{S}}}[\log p(\boldsymbol{\theta},\boldsymbol{r}|\boldsymbol{B},\boldsymbol{S})]\Big) \\ &= p(\boldsymbol{\theta})\exp\Big(\mathsf{E}_{q^{(l-1)}_{\boldsymbol{S}}}[\log p(\boldsymbol{r}|\boldsymbol{S},\boldsymbol{\theta})]\Big), \qquad (35) \\ q^{(l)}_{\boldsymbol{B},\boldsymbol{S}}(\boldsymbol{b},\boldsymbol{s}) &\propto P(\boldsymbol{b},\boldsymbol{s})\exp\Big(\mathsf{E}_{q^{(l)}_{\boldsymbol{\Theta}}}[\log p(\boldsymbol{\Theta},\boldsymbol{r}|\boldsymbol{b},\boldsymbol{s})]\Big) \\ &\propto P(\boldsymbol{s}|\boldsymbol{b})\exp\Big(\mathsf{E}_{q^{(l)}_{\boldsymbol{\Theta}}}[\log p(\boldsymbol{r}|\boldsymbol{s},\boldsymbol{\Theta})]\Big), \qquad (36) \end{aligned}$$

for $l = 1,\ldots,T$, where T is the number of iterations and $p(\boldsymbol{r}|\boldsymbol{s},\boldsymbol{\theta})$ is the likelihood function of $\boldsymbol{s}$ and $\boldsymbol{\theta}$. These updates give rise to an algorithm that performs iterative decoding and phase-noise estimation. For more details on this framework[4] and on the results in (35) and (36), refer to [7], [31]. The rest of the subsection details the computation of $q^{(l)}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ and $q^{(l)}_{\boldsymbol{B},\boldsymbol{S}}(\boldsymbol{b},\boldsymbol{s})$ for one iteration, and therefore, the superscript indicating the iteration number will be skipped hereafter for notational convenience.

Since all received samples are statistically independent of each other given $\boldsymbol{s}$ and $\boldsymbol{\theta}$, (35) and (36) can be rewritten as

$$\begin{aligned} q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) &\propto p(\boldsymbol{\theta}) \\ &\quad\cdot\prod_{i,k}\exp\bigg(-\frac{1}{2\sigma^2_i}\mathsf{E}_{q_{S_{i,k}}}\Big[\big|r_{i,k} - S_{i,k}e^{j\theta_{i,k}}\big|^2\Big]\bigg), \end{aligned} \tag{37}$$

$$q_{\boldsymbol{B},\boldsymbol{S}}(\boldsymbol{b},\boldsymbol{s}) \propto P(\boldsymbol{s}|\boldsymbol{b})$$

$$\cdot\prod_{i,k}\exp\bigg(-\frac{1}{2\sigma^2_i}\mathsf{E}_{q_{\Theta_{i,k}}}\Big[\big|r_{i,k} - s_{i,k}e^{j\Theta_{i,k}}\big|^2\Big]\bigg), \tag{38}$$

for $i = 1,\ldots,D$ and $k = 1,\ldots,N$. It can be seen from (37) that $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ relies on $q_{S_{i,k}}(s_{i,k})$, the marginals of $q_{\boldsymbol{B},\boldsymbol{S}}(\boldsymbol{b},\boldsymbol{s})$. Likewise, (38) shows that $q_{\boldsymbol{S}}(\boldsymbol{s})$ relies on $q_{\Theta_{i,k}}(\theta_{i,k})$, the marginals of $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$. Thus, computing (35) and (36) involves the marginalization of $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ and $q_{\boldsymbol{S}}(\boldsymbol{s})$.

1) Interface Towards Phase-Noise Estimation: The expression in (37) can be simplified by noting that

$$\begin{aligned} &\mathsf{E}_{q_{S_{i,k}}}\Big[\big|r_{i,k} - S_{i,k}e^{j\theta_{i,k}}\big|^2\Big] \\ &= \Big|r_{i,k} - \mathsf{E}_{q_{S_{i,k}}}[S_{i,k}]e^{j\theta_{i,k}}\Big|^2 + \mathsf{Var}(S_{i,k}) \qquad (39) \\ &= \big|r_{i,k} - \bar{\bar{s}}_{i,k}e^{j\theta_{i,k}}\big|^2 + \mathsf{Var}(S_{i,k}), \qquad (40) \end{aligned}$$

where

$$\bar{\bar{s}}_{i,k} \triangleq \mathsf{E}_{q_{S_{i,k}}}[S_{i,k}] = \sum_{s_{i,k}\in\mathcal{X}} s_{i,k}q_{S_{i,k}}(s_{i,k}), \tag{41}$$

and (39) uses the fact that $\mathsf{E}\big[|\cdot|^2\big] = \mathsf{Var}(\cdot) + |\mathsf{E}[\cdot]|^2$. Similarly to (23), $\bar{\bar{s}}_{i,k}$ represents a soft input that is utilized in the phase-noise estimation. Thus,

$$\begin{aligned} q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) &\propto p(\boldsymbol{\theta}) \\ &\quad\cdot\prod_{i,k}\exp\bigg(-\frac{1}{2\sigma^2_i}\Big(\big|r_{i,k} - \bar{\bar{s}}_{i,k}e^{j\theta_{i,k}}\big|^2 + \mathsf{Var}(S_{i,k})\Big)\bigg) \\ &\propto p(\boldsymbol{\theta})\prod_{i,k}\exp\bigg(-\frac{1}{2\sigma^2_i}\big|r_{i,k} - \bar{\bar{s}}_{i,k}e^{j\theta_{i,k}}\big|^2\bigg) \\ &\propto \prod_{k=2}^{N} p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})\prod_{i,k}\mathcal{CN}_{r_{i,k}}\big(\bar{\bar{s}}_{i,k}e^{j\theta_{i,k}}, 2\sigma^2_i\big). \end{aligned} \tag{42}$$

Analogously to FG-PNC, $\bar{\bar{s}}_{i,k} = s_{i,k}$ for pilot symbols during each iteration, and for data symbols, $\bar{\bar{s}}_{i,k}$ is initialized as 0. The structure of $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ in (42) allows for its marginals, $q_{\boldsymbol{\Theta}_k}(\boldsymbol{\theta}_k)$, to be approximated using Algorithm 1. The means and covariances of these marginals are computed with (5)–(7) and (12)–(14). Due to (42), $\mathbf{V}_k$ in (6) and the components of $\boldsymbol{h}_k$ in (7) are computed as

$$\mathbf{V}_k = \mathrm{diag}\bigg(\frac{|\bar{\bar{s}}_{1,k}|^2}{\sigma^2_1},\ldots,\frac{|\bar{\bar{s}}_{D,k}|^2}{\sigma^2_D}\bigg), \tag{43}$$

$$h_{i,k} = \frac{1}{\sigma^2_i}\Im\Big\{r_{i,k}\bar{\bar{s}}^*_{i,k}e^{-j\hat{\theta}^{\mathrm{f}}_{i,k-1}}\Big\}, \tag{44}$$

and the EKF equations are initialized with

$$\hat{\boldsymbol{\theta}}^{\mathrm{f}}_1 = \big[\angle(r_{1,1}\bar{\bar{s}}^*_{1,1}),\ldots,\angle(r_{D,1}\bar{\bar{s}}^*_{D,1})\big]^T, \tag{45}$$

$$\mathbf{M}^{\mathrm{f}}_1 = \mathrm{diag}\bigg(\frac{\sigma^2_1}{E_{\mathrm{s}}},\ldots,\frac{\sigma^2_D}{E_{\mathrm{s}}}\bigg). \tag{46}$$

Further marginalizing $q_{\boldsymbol{\Theta}_k}(\boldsymbol{\theta}_k)$ to obtain $q_{\Theta_{i,k}}(\theta_{i,k})$ is trivial since $q_{\boldsymbol{\Theta}_k}(\boldsymbol{\theta}_k)$ is approximated as a Gaussian PDF [41, Ch. 8], yielding

$$q_{\Theta_{i,k}}(\theta_{i,k}) \approx \mathcal{N}_{\theta_{i,k}}\Big(\hat{\theta}^{\mathrm{s}}_{i,k}, M^{\mathrm{s}}_{i,k}\Big). \tag{47}$$

[4]VB inference can also be regarded as a specific type of message passing on FGs, often referred to as variational message passing in that context. See [39] for more details and [40] for an example that uses this concept for joint channel estimation and decoding in wireless MIMO transmission.

2) Interface Towards Decoder: The expectation in (38) can be expressed as

$$\begin{aligned}&\mathsf{E}_{q_{\Theta_{i,k}}}\left[\left|r_{i,k}-s_{i,k}e^{j\Theta_{i,k}}\right|^2\right]\\&=|r_{i,k}|^2-2\Re\{r_{i,k}s_{i,k}^*\alpha_{i,k}^*\}+|s_{i,k}|^2,\end{aligned}\tag{48}$$

where $\alpha_{i,k}\triangleq\mathsf{E}_{q_{\Theta_{i,k}}}\left[e^{j\Theta_{i,k}}\right]$. To compute $\alpha_{i,k}$ in closed form, note that the characteristic function of a real Gaussian random variable X with mean μ and variance σ^2 is $\Phi_X(\omega)\triangleq\mathsf{E}_X\left[e^{j\omega X}\right]=\exp(j\mu\omega-\omega^2\sigma^2/2)$ [42, Ch. 4.7]. Using this fact as well as (47), $\alpha_{i,k}$ can be described as $\Phi_{\Theta_{i,k}}(1)=\exp(j\hat{\theta}^{\mathrm{s}}_{i,k}-M^{\mathrm{s}}_{i,k}/2)$. Moreover, (48) leads to

$$\begin{aligned}q_{\boldsymbol{B},\boldsymbol{S}}(\boldsymbol{b},\boldsymbol{s})&\propto P(\boldsymbol{s}|\boldsymbol{b})\prod_{i,k}\exp\left(\frac{\Re\{r_{i,k}s_{i,k}^*\alpha_{i,k}^*\}}{\sigma_i^2}-\frac{|s_{i,k}|^2}{2\sigma_i^2}\right)\\&\propto P(\boldsymbol{s}|\boldsymbol{b})\prod_{i,k}g_{i,k}(s_{i,k}),\end{aligned}\tag{49}$$

where $g_{i,k}(s_{i,k})$ corresponds to the likelihood function of $s_{i,k}$ from a virtual memoryless phase-noise compensated channel, analogously to $P_{\mathrm{u}}(s_{i,k})$ in (30). To obtain $q_{S_{i,k}}(s_{i,k})$, which is used to compute $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ in (37) during a consecutive iteration, $q_{\boldsymbol{B},\boldsymbol{S}}(\boldsymbol{b},\boldsymbol{s})$ is marginalized over all bits and all symbols except $s_{i,k}$. The marginalization of $q_{\boldsymbol{B},\boldsymbol{S}}(\boldsymbol{b},\boldsymbol{s})$ is performed in the decoder, where the function $g_{i,k}(s_{i,k})$ in (49) is converted to bit-wise LLRs that are fed to the decoder. The decoder then either outputs the detected information bits or *a posteriori* coded-bit LLRs that are converted to $q_{S_{i,k}}(s_{i,k})$.[5] This algorithm will be referred to as VB-based phase-noise compensation (VB-PNC) and is summarized in Algorithm 3.

Algorithm 3 VB-PNC (1 iteration)

Input: $\boldsymbol{r}, D, N, \mathbf{Q}, \boldsymbol{\sigma}^2, \mathcal{X}, q^{(l-1)}_{S_{i,k}}(s_{i,k})\ \forall\ i,k$

Output: $g^{(l)}_{i,k}(s_{i,k})\ \forall\ i,k$

1: **for all** i,k **do**
2: $\quad\bar{\bar{s}}_{i,k}=\sum_{s_{i,k}\in\mathcal{X}}s_{i,k}q^{(l-1)}_{S_{i,k}}(s_{i,k})$
3: **end for**
4: Compute $(\boldsymbol{\theta}^{\mathrm{s}}_k,\mathbf{M}^{\mathrm{s}}_k)\ \forall\ k$ with Alg. 1 using (43)–(46)
5: **for all** i,k **do**
6: $\quad\alpha_{i,k}=\exp(j\hat{\theta}^{\mathrm{s}}_{i,k}-M^{\mathrm{s}}_{i,k}/2)$
7: $\quad$**for all** $s_{i,k}\in\mathcal{X}$ **do**
8: $\quad\quad g_{i,k}(s_{i,k})=\exp\left(\frac{\Re\{r_{i,k}s_{i,k}^*\alpha_{i,k}^*\}}{\sigma_i^2}-\frac{|s_{i,k}|^2}{2\sigma_i^2}\right)$
9: $\quad$**end for**
10: $\quad$Normalize $g_{i,k}(s_{i,k})$ s.t. $\sum_{s_{i,k}\in\mathcal{X}}g_{i,k}(s_{i,k})=1$
11: **end for**

Remark 1: It can be seen from (23) and (41) that in the absence of pilot symbols, FG-PNC and VB-PNC will not bootstrap properly. However, it is shown in [18] that trellis-based demodulation can be used to bootstrap the Tikhonov algorithm in [6, Sec. IV-B] without pilot symbols. We conjecture that the same principles could be used for FG-PNC and VB-PNC; however, such a study is out of scope for this paper and is suggested for future work.

Remark 2: FG-PNC takes into account the symbol uncertainty when interfacing with the phase-noise estimation, in contrast to VB-PNC. This can be seen from the approximate likelihoods in (22) and (42). However, both algorithms convey the phase-noise uncertainty to the decoder, which can be seen from (30) and (49). Moreover, VB-PNC can perform iterative phase-noise estimation and data detection even in the case of uncoded transmission. As mentioned in Section III-B, the FG/SPA framework does not yield an iterative algorithm in the absence of an iterative decoder. However, multiple iterations can be performed with FG-PNC for uncoded transmission by using $P_{\mathrm{u}}(s_{i,k})$ instead of $P_{\mathrm{d}}(s_{i,k})$ to compute (23) and (24) in a consecutive iteration.

D. Distribution of Pilot Symbols

The optimal distribution of the pilot symbols for multi-channel transmission, subject to an average pilot rate across the channels, in nontrivial in general. For single-channel transmission, equispaced pilot symbols typically give good performance. However, an effective pilot distribution for multichannel transmission cannot be easily inferred from studying the single-channel case, as the optimal distribution will depend on the phase-noise correlation across channels, as well as the SNR. In [43], we investigated this problem for SDM transmission and found that arranging the pilot symbols on a wrapped diagonal in the space–time grid performs well in general. This type of distribution will be used for the remainder of this paper.

E. Conversion Between PMFs and LLRs

As detailed in the previous subsections, the inputs and outputs to FG-PNC and VB-PNC in each iteration are in the form of symbol PMFs. However, iterative decoders for binary LDPC codes and turbo codes are typically implemented in the logarithm domain [35, Ch. 5], and thus, have bit-wise LLRs as inputs and outputs.

The computation of bit-wise input LLRs is done as follows. For each transmitted symbol $s_{i,k}$, denote with $c^j_{i,k}$ the jth coded bit in the binary labeling of the constellation points. The LLR for $c^j_{i,k}$ is computed from the likelihood of $s_{i,k}$ as

$$\begin{aligned}L(c_i)&\triangleq\log\left(\frac{p(r_{i,k}|c^j_{i,k}=0)}{p(r_{i,k}|c^j_{i,k}=1)}\right)\\&=\log\left(\frac{\sum_{s_{i,k}\in\mathcal{B}_0}p(r_{i,k}|s_{i,k})}{\sum_{s_{i,k}\in\mathcal{B}_1}p(r_{i,k}|s_{i,k})}\right),\end{aligned}\tag{50}$$

for $j=1,\ldots,R_{\mathrm{m}}$, where $R_{\mathrm{m}}\triangleq\log_2|\mathcal{X}|$, and $\mathcal{B}_\nu$ is the set of constellation points that have the jth bit in the binary labeling as $\nu\in\{0,1\}$. In relation to the derived algorithms, $p(r_{i,k}|s_{i,k})$ corresponds to $P_{\mathrm{u}}(s_{i,k})$ for FG-PNC and $g_{i,k}(s_{i,k})$ for VB-PNC.

The *a posteriori* output LLRs from the decoder are defined as

$$L(c^j_{i,k}|\boldsymbol{r})\triangleq\log\left(\frac{P(c^j_{i,k}=0|\boldsymbol{r})}{P(c^j_{i,k}=1|\boldsymbol{r})}\right),\tag{51}$$

[5]For uncoded transmission, this marginalization is trivial since the transmitted symbols are independent of each other. Thus, $q_{S_{i,k}}(s_{i,k})\propto P(s_{i,k})g_{i,k}(s_{i,k})$, where $P(s_{i,k})$ is a uniform PMF for data symbols and a degenerate distribution for pilot symbols, i.e., equal to 1 if $s_{i,k}$ is equal to the pilot and 0 otherwise.

To compute $q_{S_{i,k}}(s_{i,k})$ for VB-PNC, the output LLRs in (51) are first converted to *a posteriori* coded-bit probabilities as

$$P(c_{i,k}^j = 0|\boldsymbol{r}) = \frac{e^{L(c_{i,k}^j|\boldsymbol{r})}}{1+e^{L(c_{i,k}^j|\boldsymbol{r})}}. \tag{52}$$

Then, with a slight abuse of notation, denote the probability of $s_{i,k}$ being a constellation point with binary labeling $(\nu_1,\ldots,\nu_{R_\mathrm{m}}) \in \{0,1\}^{R_\mathrm{m}}$ as $P(s_{i,k}=(\nu_1,\ldots,\nu_{R_\mathrm{m}}))$, which corresponds to $q_{S_{i,k}}(s_{i,k})$ and is computed as

$$P(s_{i,k}=(\nu_1,\ldots,\nu_{R_\mathrm{m}})) = \prod_{j=1}^{R_\mathrm{m}} P(c_{i,k}^j = \nu_j|\boldsymbol{r}). \tag{53}$$

For FG-PNC, $P_\mathrm{d}(s_{i,k})$ is found in the same way as $q_{S_{i,k}}(s_{i,k})$, except the coded-bit probabilities in (53) are computed from the extrinsic LLRs $L_\mathrm{e}(c_{i,k}^j) \triangleq L(c_{i,k}^j|\boldsymbol{r}) - L(c_{i,k}^j)$.

F. Computational Complexity

The main difference in computational complexity (in terms of the number of real additions and multiplications per outer iteration between the phase-noise compensation and decoding) of the two algorithms comes from FG-PNC computing the symbol uncertainties in (24), which VB-PNC does not. This computation scales as $\mathcal{O}(DN|\mathcal{X}|)$. Both algorithms use the same method to estimate the marginal phase-noise PDFs, which scales as $\mathcal{O}(D^3N)$ due to matrix inversions, and the likelihood computations in (33) and (49) both scale as $\mathcal{O}(DN|\mathcal{X}|)$. It is important to mention that this method of quantifying complexity only gives an approximate view of the actual hardware requirements needed to implement the algorithms. Hence, it serves as a starting point for a more detailed analysis, which is out of scope for this paper.

IV. Performance Results

In this section, we first justify the utilization of the EKF by showing that the linearization of the system model in (1) does not yield any significant penalties to the phase-noise estimation for practical baud rates and laser linewidths. Thereafter, we experimentally validate the system model in (1) and the proposed algorithms. We then further assess their performance using Monte Carlo simulations.

A. Impact of EKF Linearization

The linearization of the system model, which is imposed by the EKF, can negatively impact the phase-noise estimation [25], [28], [29] if the phase noise varies fast enough. In [29], this effect is investigated for oscillator phase noise in wireless MIMO systems and an error floor is attributed to linearization penalties. However, for all parameters of interest in fiber-optical systems, the linearization has negligible effects on the phase-noise estimation performance. To show this, the mean squared error (MSE) is computed from 10^6 realizations of a phase-noise estimate that is obtained at time k from a received sample using (63), assuming perfect knowledge of the transmitted symbol and the phase noise at time $k-1$. For 20 GBaud uncoded transmission, the resulting MSE is shown in Fig. 2 as a function of laser linewidth, for different SNRs per information bit, defined as

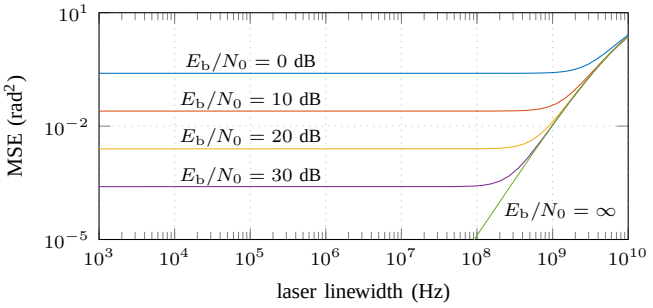

Fig. 2. MSE of a phase-noise estimate, obtained using a linearized system model, versus laser linewidth.

$$\frac{E_\mathrm{b}}{N_0} \triangleq \frac{E_\mathrm{s}}{2\sigma^2 R_\mathrm{c} R_\mathrm{m}(1-R_\mathrm{p})}, \tag{54}$$

where σ^2 is the variance per real dimension of the complex AWGN, and R_p, R_c, and R_m are the pilot rate, code rate, and bits per symbol, resp. From Fig. 2 it can be seen that for laser linewidths under 10^8 Hz, the MSE in the absence of AWGN (and thus, due to the linearization) is below 10^{-5}, while in the presence of AWGN, the MSE is orders of magnitude larger, even at high E_b/N_0. Hence, for practical laser linewidths (on the order of MHz and below) and SNRs, the total MSE is virtually only due to the AWGN; in other words, the linearization yields negligible estimation error. The choice of modulation format and pilot rate has a marginal effect on the results.

B. Experimental Verification

To assess the efficacy of the system model in (1) and the proposed algorithms, a recirculating-loop experiment involving SDM transmission of 64QAM at 20 GBaud through 3 cores of a 7-core weakly-coupled, homogeneous, multicore fiber was carried out. The experimental setup and detailed results are given in a separate article [44]. In this section, additional results are presented based on the same data set. The DSP chain prior to phase-noise compensation consisted of standard steps, applied on a per-core or per-channel basis. The results in Fig. 3 were obtained for two pilot rates using a symbol-detection algorithm analogous to FG-PNC, performing both per-channel and joint-channel processing. As can be seen, jointly processing the channels leads to a reduction in BER relative to what is attained with per-channel processing. The amount of BER reduction is greater for lower transmission distances, and also increases when the pilot rate is decreased.

Per-channel and joint-channel processing is also performed on realizations of (1) using the estimated SNR values from the experimental data, as well as a particular configuration of $\mathbf{Q}$ that was used when processing the experimental data. The results are shown in Fig. 3. A strong agreement is observed overall between the results based on experiments and simulations, which suggests that the system model in (1) and the proposed algorithms are relevant in this application. It is, however, important to note that other types of multichannel

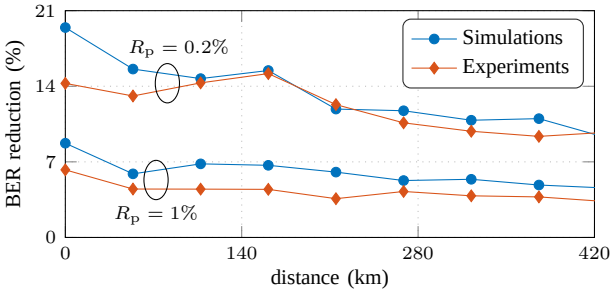

Fig. 3. Reduction in BER by performing joint-channel processing instead of per-channel processing for different distances of uncoded 64QAM transmission, comparing the results using experimental data and simulations based on the system model in (1) for two pilot rates.

optical systems could have drastically different behavior, in which case other models and algorithms may be more suitable than what has been investigated in this paper.

C. Simulation Results

The proposed algorithms are further assessed in terms of phase-noise tolerance through Monte Carlo simulations. In order to put the results into perspective, their performance is compared to a strategy entailing per-channel phase-noise compensation using BPS and symbol detection based on minimum Euclidean distance (BPS-EDD). Coded transmission of 16QAM, 64QAM, and 256QAM over 20 channels is considered, using $R_\mathrm{p} = 1\%$ for the proposed algorithms. An off-the-shelf rate-4/5 LDPC code from the DVB-S.2 standard is utilized to encode information bits independently for each channel, yielding codewords of length 64800 bits. Furthermore, the variance of the AWGN is kept identical for all channels and the phase noise has a high degree of correlation across the channels. More specifically, the covariance matrix of $\Delta\boldsymbol{\theta}_k$ in (2), $\mathbf{Q}$, is constructed such that the elements on the diagonal are equal to $\sigma_\phi^2 + \sigma_\delta^2$, with $\sigma_\delta^2 \ll \sigma_\phi^2$, while all other elements are equal to σ_ϕ^2, where $\sigma_\phi^2 = 2\pi\Delta\nu T_\mathrm{s}$ is the laser phase noise variance and is a function of the laser linewidth, $\Delta\nu$, and the symbol duration, T_s. This corresponds to a single dominant laser phase noise component that is common to all channels, in addition to phase drifts that are independent between channels. As these drifts are typically orders of magnitude slower than the laser phase noise [10], we set $\sigma_\delta^2 = \sigma_\phi^2/1000$. The laser linewidth and symbol duration product, $\Delta\nu T_\mathrm{s}$, is fixed at $5 \cdot 10^{-5}$ for 16QAM and 64QAM, and $5 \cdot 10^{-6}$ for 256QAM. Assuming a 20 GBaud symbol rate, this corresponds to linewidths of 1 MHz and 100kHz, resp. Perfect knowledge of $\mathbf{Q}$ at the receiver is assumed. BER estimates are obtained for different E_b/N_0 values by counting a minimum of 100 frame errors. Furthermore, the BER performance of coded transmission with ideal phase-noise compensation in the absence of pilots is included as a benchmark.

The scheduling for the proposed algorithms is as follows. A total of 2 outer iterations are performed, where in each outer iteration, the decoder is run for 50 decoding iterations. The information bits are then detected following the second outer iteration. Due to limitations in the decoder implementation, the decoder state is reset between each outer iteration. This incurs a performance penalty for FG-PNC as the message passing that results from applying the SPA to the FG in Fig. 1 does not assume resets in the decoder state.

When performing phase-noise compensation using BPS, 32 test phases are used for transmission of 16QAM, whereas 64 test phases are used for 64QAM and 256QAM. A filter length between 70 and 90 is used. No differential encoding is used, but instead, the initial value of the phase noise is assumed to be known. Moreover, only a single outer iteration is run since BPS does not exploit the symbol statistics. Thus, following BPS and symbol detection, the decoder performs 50 decoding iterations and outputs the detected information bits.

Figs. 4 (a), (b), and (c) show results for 16QAM, 64QAM, and 256QAM, resp., for FG-PNC, VB-PNC, and BPS-EDD. In all cases, FG-PNC and VB-PNC perform similarly and outperform BPS-EDD by a wide margin. Compared to the ideal performance, they result in a negligible SNR penalty of 0.1 dB, 0.2 dB, and 0.1 dB for 16QAM, 64QAM, and 256QAM, resp., at a BER of 10^{-4}. In addition, Fig. 4 (a) includes results for FG-PNC and VB-PNC running 1 only outer iteration, showing how the cooperation between the decoder and the phase-noise estimation improves BER performance. Analogous results are also obtained for 64QAM and 256QAM; however, for clarity, they are not shown in the plots. Fig. 4 (b) shows additional results for FG-PNC and VB-PNC using $R_\mathrm{p} = 0.1\%$. In this case, FG-PNC demonstrates superior performance to VB-PNC, which is found to be mainly due to the different likelihood approximations in (30) and (49). Similar results are also found for 16QAM and 256QAM. Finally, Fig. 4 (c) contains results when the proposed algorithms are used for per-channel processing. As expected, this yields worse performance than joint processing of all channels. Again, comparable results are found for 16QAM and 64QAM.

The strong performance of the proposed algorithms can be attributed to the following: (i) Due to the phase-noise correlation, the pilot symbol distribution allows for a more effective use of the pilots in joint-channel processing compared to per-channel processing. In addition, resistance to AWGN is improved through joint-channel processing; (ii) Furthermore, the algorithms make use of the phase-noise statistics when computing the symbol likelihoods, which has been shown to be a superior strategy to separating the phase-noise compensation and symbol detection [45]; (iii) Finally, the iterative cooperation with the decoder improves the phase-noise compensation performance, and thus, the resulting BER. As expected, however, the amount of performance improvement diminishes with increasing number of outer iterations.

V. Discussion and Conclusions

This work was motivated by the fact that in certain multi-channel optical systems, the received signal may be approximated by a simplistic multidimensional phase-noise model after being processed by standard DSP steps prior to phase-noise compensation. The MAP bit detector for this model was approximated using two frameworks, and the resulting pilot-aided algorithms perform iterative phase-noise estimation in

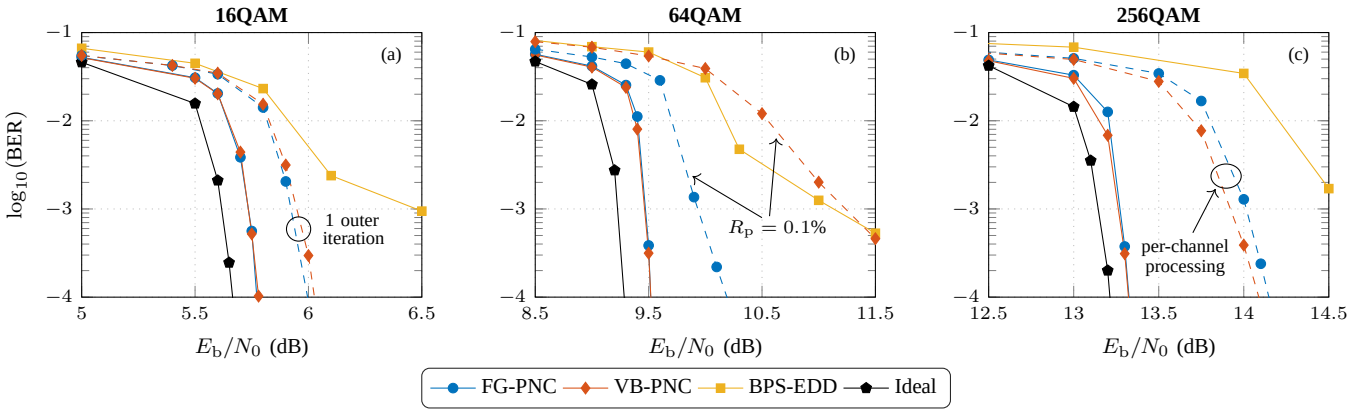

Fig. 4. BER versus SNR per information bit for 16QAM and 1 MHz laser linewidth (a), 64QAM and 1 MHz laser linewidth (b), and 256QAM and 100 kHz laser linewidth (c), assuming a symbol rate of 20 GBaud and $R_p = 1\%$.

cooperation with an iterative decoder. By exploiting the phase-noise correlation across the channels, a more capable phase-noise compensation is achieved than what is possible with per-channel processing. The system model and the proposed algorithms were verified through comparison of results based on experimental data and Monte Carlo simulations. Moreover, the algorithms were further assessed using simulations in terms of phase-noise tolerance for different modulation formats, amounts of phase noise, and pilot rates. The results in Fig. 4 show that the proposed algorithms significantly outperform the typical phase-noise compensation approach in the optical literature, giving rise to a marginal SNR penalty of 0.2 dB or less at a BER of 10^{-4} with respect to pilot-free transmission in the absence of phase noise. However, aside from the increased transmission reach shown in Fig. 3, practical implications have not been addressed in this paper. For example, can joint-channel processing increase power and spectral efficiency, or can hardware requirements be relaxed? These are important questions that we have addressed in [44].

Appendix A
Derivation of EKF Equations

The EKF equations in Sec. III-A can be derived from the general recursive Bayesian filtering equations [36, Ch. 4.2]

$$p(\boldsymbol{\theta}_k|\boldsymbol{r}_1,\ldots,\boldsymbol{r}_k) \propto p(\boldsymbol{r}_k|\boldsymbol{\theta}_k)\int_{\mathbb{R}^D} p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})p(\boldsymbol{\theta}_{k-1}|\boldsymbol{r}_1,\ldots,\boldsymbol{r}_{k-1})d\boldsymbol{\theta}_{k-1}, \tag{55}$$

for $k = 2,\ldots,N$. Conforming to the EKF, the system model in (1) at time k is linearized using a first-order Taylor expansion around an estimate of $\boldsymbol{\theta}_{k-1}$, denoted with $\hat{\boldsymbol{\theta}}^{\mathrm{f}}_{k-1}$, yielding

$$r_{i,k} \approx s_{i,k}e^{j\hat{\theta}^{\mathrm{f}}_{i,k-1}}(1+j(\theta_{i,k}-\hat{\theta}^{\mathrm{f}}_{i,k-1})) + n_{i,k}, \tag{56}$$

where $\hat{\theta}^{\mathrm{f}}_{i,k-1}$ is the ith element of $\hat{\boldsymbol{\theta}}^{\mathrm{f}}_{k-1}$. Hence,

$$\begin{aligned}
&p(\boldsymbol{r}_k|\boldsymbol{\theta}_k)\\
&= \prod_{i=1}^{D} p(r_{i,k}|\theta_{i,k})\\
&= \prod_{i=1}^{D} \mathcal{CN}_{r_{i,k}}\left(s_{i,k}e^{j\theta_{i,k}}, 2\sigma_i^2\right) &(57)\\
&\approx \prod_{i=1}^{D} \mathcal{CN}_{r_{i,k}}\left(s_{i,k}e^{j\hat{\theta}^{\mathrm{f}}_{i,k-1}}(1+j(\theta_{i,k}-\hat{\theta}^{\mathrm{f}}_{i,k-1})), 2\sigma_i^2\right)\\
&\propto \prod_{i=1}^{D} \exp\left(-\frac{|s_{i,k}|^2}{2\sigma_i^2}\left|j+\hat{\theta}^{\mathrm{f}}_{i,k-1}+\eta_{i,k}-\theta_{i,k}\right|^2\right) &(58)\\
&\propto \prod_{i=1}^{D} \exp\left(-\frac{|s_{i,k}|^2}{2\sigma_i^2}\left(\hat{\theta}^{\mathrm{f}}_{i,k-1}+\Re\{\eta_{i,k}\}-\theta_{i,k}\right)^2\right) &(59)\\
&\propto \prod_{i=1}^{D} \mathcal{N}_{\theta_{i,k}}\left(\hat{\theta}^{\mathrm{f}}_{i,k-1}+\Im\left\{\frac{r_{i,k}e^{-j\hat{\theta}^{\mathrm{f}}_{i,k-1}}}{s_{i,k}}\right\}, \frac{\sigma_i^2}{|s_{i,k}|^2}\right), &(60)
\end{aligned}$$

where $\eta_{i,k} \triangleq r_{i,k}e^{-j\hat{\theta}^{\mathrm{f}}_{i,k-1}}/(js_{i,k})$. Furthermore, (57) follows as the knowledge of $s_{i,k}$ for all i and k is assumed, and (59) is obtained by using $|z|^2 = \Re\{z\}^2 + \Im\{z\}^2$, for $z \in \mathbb{C}$, as

$$\begin{aligned}
&\left|j+\hat{\theta}^{\mathrm{f}}_{i,k-1}+\eta_{i,k}-\theta_{i,k}\right|^2\\
&= (1+\Im\{\eta_{i,k}\})^2 + \left(\hat{\theta}^{\mathrm{f}}_{i,k-1}+\Re\{\eta_{i,k}\}-\theta_{i,k}\right)^2. &(61)
\end{aligned}$$

The first term in (61) is constant with respect to $\theta_{i,k}$ and can thus be discarded. Finally, (60) is obtained since $\Re\{z/j\} = \Im\{z\}$ for $z \in \mathbb{C}$, and therefore, $p(\boldsymbol{r}_k|\boldsymbol{\theta}_k)$ can be expressed as

$$p(\boldsymbol{r}_k|\boldsymbol{\theta}_k) \approx \mathcal{N}_{\boldsymbol{\theta}_k}\left(\hat{\boldsymbol{\theta}}^{\mathrm{f}}_{k-1}+\tilde{\boldsymbol{h}}_k, \mathbf{V}_k^{-1}\right), \tag{62}$$

where $\mathbf{V}_k$ is defined in (10) and $\tilde{\boldsymbol{h}}_k \triangleq [\tilde{h}_{1,k},\ldots,\tilde{h}_{D,k}]^T$ represents the estimated difference between $\boldsymbol{\theta}_k$ and $\boldsymbol{\theta}_{k-1}$, with each of its elements computed as

$$\tilde{h}_{i,k} = \Im\left\{\frac{r_{i,k}e^{-j\hat{\theta}^{\mathrm{f}}_{i,k-1}}}{s_{i,k}}\right\}. \tag{63}$$

Finally, using the following identity for the product of two Gaussians [41, Ch. A.2],

$$\mathcal{N}_{\boldsymbol{x}}(\boldsymbol{a},\mathbf{A})\mathcal{N}_{\boldsymbol{x}}(\boldsymbol{b},\mathbf{B}) = \mathcal{N}_{\boldsymbol{b}}(\boldsymbol{a},\mathbf{A}+\mathbf{B})\mathcal{N}_{\boldsymbol{x}}(\boldsymbol{c},\mathbf{C}), \tag{64}$$

where $\mathbf{T} = (\mathbf{I}+\mathbf{A}\mathbf{B}^{-1})^{-1}$, $\mathbf{C} = \mathbf{T}\mathbf{A}$, and $\boldsymbol{c} = \mathbf{T}\boldsymbol{a} + \mathbf{T}\mathbf{A}\mathbf{B}^{-1}\mathbf{b}$, allows reducing (55) to

$$\begin{aligned}
&p(\boldsymbol{\theta}_k|\boldsymbol{r}_1,\ldots,\boldsymbol{r}_k)\\
&\approx \mathcal{N}_{\boldsymbol{\theta}_k}\Big(\hat{\boldsymbol{\theta}}^{\mathrm{f}}_{k-1}+\tilde{\boldsymbol{h}}_k,\mathbf{V}_k^{-1}\Big)\\
&\quad\cdot\int_{\mathbb{R}^D}\mathcal{N}_{\boldsymbol{\theta}_{k-1}}(\boldsymbol{\theta}_k,\mathbf{Q})\mathcal{N}_{\boldsymbol{\theta}_{k-1}}\Big(\hat{\boldsymbol{\theta}}^{\mathrm{f}}_{k-1},\mathbf{M}^{\mathrm{f}}_{k-1}\Big)d\boldsymbol{\theta}_{k-1}\\
&= \mathcal{N}_{\boldsymbol{\theta}_k}\Big(\hat{\boldsymbol{\theta}}^{\mathrm{f}}_{k-1}+\tilde{\boldsymbol{h}}_k,\mathbf{V}_k^{-1}\Big)\mathcal{N}_{\boldsymbol{\theta}_k}\Big(\hat{\boldsymbol{\theta}}^{\mathrm{f}}_{k-1},\mathbf{M}^{\mathrm{f}}_{k|k-1}\Big) \qquad (65)\\
&\propto \mathcal{N}_{\boldsymbol{\theta}_k}\Big(\mathbf{T}_k\hat{\boldsymbol{\theta}}^{\mathrm{f}}_{k-1}+\mathbf{T}_k\mathbf{M}^{\mathrm{f}}_{k|k-1}\mathbf{V}_k\Big(\hat{\boldsymbol{\theta}}^{\mathrm{f}}_{k-1}+\tilde{\boldsymbol{h}}_k\Big),\mathbf{T}_k\mathbf{M}^{\mathrm{f}}_{k|k-1}\Big)\\
&= \mathcal{N}_{\boldsymbol{\theta}_k}\Big(\mathbf{T}_k\Big(\mathbf{I}_D+\mathbf{M}^{\mathrm{f}}_{k|k-1}\mathbf{V}_k\Big)\hat{\boldsymbol{\theta}}^{\mathrm{f}}_{k-1}\\
&\qquad +\mathbf{T}_k\mathbf{M}^{\mathrm{f}}_{k|k-1}\mathbf{V}_k\tilde{\boldsymbol{h}}_k,\mathbf{T}_k\mathbf{M}^{\mathrm{f}}_{k|k-1}\Big)\\
&= \mathcal{N}_{\boldsymbol{\theta}_k}\Big(\hat{\boldsymbol{\theta}}^{\mathrm{f}}_{k-1}+\mathbf{M}^{\mathrm{f}}_k\boldsymbol{h}_k,\mathbf{M}^{\mathrm{f}}_k\Big), \qquad (66)
\end{aligned}$$

where (65) is obtained since the integral of a PDF is one, $\mathbf{T}_k \triangleq (\mathbf{I}_D+\mathbf{M}^{\mathrm{f}}_{k|k-1}\mathbf{V}_k)^{-1}$, and $\boldsymbol{h}_k \triangleq \mathbf{V}_k\tilde{\boldsymbol{h}}_k$. This leads to the recursive equations in (5)–(7).

Appendix B
Derivation of FG messages

Using (22) and (29), $P_{\mathrm{u}}(s_{i,k})$ can be computed as

$$\begin{aligned}
P_{\mathrm{u}}(s_{i,k}) &\propto \int_{\mathbb{R}^D} p(r_{i,k}|s_{i,k},\theta_{i,k})\frac{p(\boldsymbol{\theta}_k|\boldsymbol{r})}{p(r_{i,k}|\theta_{i,k})}d\boldsymbol{\theta}_k\\
&= \int_{\mathbb{R}}\frac{\mathcal{CN}_{r_{i,k}}\big(s_{i,k}e^{j\theta_{i,k}},2\sigma_i^2\big)}{p(r_{i,k}|\theta_{i,k})}\left[\int_{\mathbb{R}^{D-1}}p(\boldsymbol{\theta}_k|\boldsymbol{r})d\bar{\boldsymbol{\theta}}_{i,k}\right]d\theta_{i,k} \qquad (67)\\
&\approx \int_{\mathbb{R}}\frac{\mathcal{CN}_{r_{i,k}}\big(s_{i,k}e^{j\theta_{i,k}},2\sigma_i^2\big)}{\mathcal{CN}_{r_{i,k}}\Big(\bar{s}_{i,k}e^{j\theta_{i,k}},2\bar{\sigma}^2_{i,k}\Big)}\\
&\qquad\cdot\left[\int_{\mathbb{R}^{D-1}}\mathcal{N}_{\boldsymbol{\theta}_k}\Big(\hat{\boldsymbol{\theta}}^{\mathrm{s}}_k,\mathbf{M}^{\mathrm{s}}_k\Big)d\bar{\boldsymbol{\theta}}_{i,k}\right]d\theta_{i,k} \qquad (68)\\
&= \int_{\mathbb{R}}\frac{\mathcal{CN}_{r_{i,k}}\big(s_{i,k}e^{j\theta_{i,k}},2\sigma_i^2\big)}{\mathcal{CN}_{r_{i,k}}\Big(\bar{s}_{i,k}e^{j\theta_{i,k}},2\bar{\sigma}^2_{i,k}\Big)}\mathcal{N}_{\theta_{i,k}}\Big(\hat{\theta}^{\mathrm{s}}_{i,k},M^{\mathrm{s}}_{i,k}\Big)d\theta_{i,k} \qquad (69)\\
&\propto e^{-\frac{|s_{i,k}|^2}{2\sigma_i^2}}\int_{\mathbb{R}}\exp\left(\Re\left\{\left(\frac{r_{i,k}s^*_{i,k}}{\sigma_i^2}-\frac{r_{i,k}\bar{s}^*_{i,k}}{\bar{\sigma}^2_{i,k}}\right)e^{-j\theta_{i,k}}\right\}\right)\\
&\qquad\cdot\mathcal{N}_{\theta_{i,k}}\Big(\hat{\theta}^{\mathrm{s}}_{i,k},M^{\mathrm{s}}_{i,k}\Big)d\theta_{i,k} \qquad (70)\\
&\approx e^{-\frac{|s_{i,k}|^2}{2\sigma_i^2}}\int_{\hat{\theta}^{\mathrm{s}}_{i,k}-\pi}^{\hat{\theta}^{\mathrm{s}}_{i,k}+\pi}\exp\left(\Re\left\{\left(\frac{r_{i,k}s^*_{i,k}}{\sigma_i^2}-\frac{r_{i,k}\bar{s}^*_{i,k}}{\bar{\sigma}^2_{i,k}}\right)e^{-j\theta_{i,k}}\right\}\right)\\
&\qquad\cdot\mathcal{T}_{\theta_{i,k}}\left(\frac{e^{j\hat{\theta}^{\mathrm{s}}_{i,k}}}{M^{\mathrm{s}}_{i,k}}\right)d\theta_{i,k}, \qquad (71)
\end{aligned}$$

where $\mathcal{T}_z(\kappa)$ denotes a Tikhonov PDF with a complex parameter κ and argument z, $\bar{\boldsymbol{\theta}}_{i,k}$ contains all elements of $\boldsymbol{\theta}_k$ except $\theta_{i,k}$, (69) exploits the fact that a multivariate Gaussian is trivially marginalized [41, Ch. 8], and (71) uses the approximation $\mathcal{N}_x\big(\mu,\sigma^2\big) \approx \mathcal{T}_x\big(e^{j\mu}/\sigma^2\big)$, which is accurate for small σ^2 [6]. Using the definition of $\xi_{i,k}(s_{i,k})$ in (31) gives

$$\begin{aligned}
P_{\mathrm{u}}(s_{i,k}) &\approx e^{-\frac{|s_{i,k}|^2}{2\sigma_i^2}}\int_{\hat{\theta}^{\mathrm{s}}_{i,k}-\pi}^{\hat{\theta}^{\mathrm{s}}_{i,k}+\pi}\exp\big(\Re\big\{\xi_{i,k}(s_{i,k})e^{-j\theta_{i,k}}\big\}\big)d\theta_{i,k}\\
&\propto e^{-\frac{|s_{i,k}|^2}{2\sigma_i^2}}I_0(|\xi_{i,k}(s_{i,k})|)\int_{\hat{\theta}^{\mathrm{s}}_{i,k}-\pi}^{\hat{\theta}^{\mathrm{s}}_{i,k}+\pi}\mathcal{T}_{\theta_{i,k}}(\xi_{i,k}(s_{i,k}))d\theta_{i,k}\\
&= e^{-\frac{|s_{i,k}|^2}{2\sigma_i^2}}I_0(|\xi_{i,k}(s_{i,k})|)\\
&\approx \frac{1}{\sqrt{2\pi|\xi_{i,k}(s_{i,k})|}}\exp\left(|\xi_{i,k}(s_{i,k})|-\frac{|s_{i,k}|^2}{2\sigma_i^2}\right), \qquad (72)
\end{aligned}$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind and zeroth order. From (72), (30)–(31) follow.

Acknowledgment

The authors would like to express their gratitude to Benjamin J. Puttnam, Georg Rademacher, and Ruben S. Luís for providing experimental data that proved to be very beneficial for this work.

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] M. Karlsson and E. Agrell, "Multidimensional modulation and coding in optical transport," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 876–884, Feb. 2017.

[3] A. J. Viterbi and A. M. 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] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] J. M. Kahn and D. A. B. Miller, "Communications expands its space," *Nat. Photon.*, vol. 11, no. 1, pp. 5–8, Jan. 2017.

[10] M. D. Feuer, L. E. Nelson, X. Zhou, S. L. Woodward, R. Isaac, B. Zhu, T. F. Taunay, M. Fishteyn, J. M. Fini, and M. F. Yan, "Joint digital signal processing receivers for spatial superchannels," *IEEE Photon. Technol. Lett.*, vol. 24, no. 21, pp. 1957–1960, Nov. 2012.

[11] A. F. Alfredsson, E. Agrell, H. Wymeersch, and M. Karlsson, "Phase-noise compensation for spatial-division multiplexed transmission," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2017, p. Th4C.7.

[12] B. J. Puttnam, R. S. Luís, J.-M. Delgado Mendinueta, J. Sakaguchi, W. Klaus, Y. Awaji, N. Wada, A. Kanno, and T. Kawanishi, "Long distance transmission in a multi-core fiber with self-homodyne detection," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2015, p. Th1D.5.

[13] R. S. Luís, B. J. Puttnam, J.-M. Delgado Mendinueta, W. Klaus, Y. Awaji, and N. Wada, "Comparing inter-core skew fluctuations in multi-core and single-core fibers," in *Proc. Conf. Lasers and Electro-Opt.*, May 2015, p. SM2L.5.

[14] L. Lundberg, M. Mazur, A. Lorences-Riesgo, M. Karlsson, and P. A. Andrekson, "Joint carrier recovery for DSP complexity reduction in frequency comb-based superchannel transceivers," in *Proc. European Conf. Opt. Commun.*, Sep. 2017, p. Th.1.D.3.

[15] D. V. Souto, B.-E. Olsson, C. Larsson, and D. A. A. Mello, "Joint-polarization and joint-subchannel carrier phase estimation for 16-QAM optical systems," *J. Lightw. Technol.*, vol. 30, no. 20, pp. 3185–3191, Oct. 2012.

[16] R. G. H. van Uden, C. M. Okonkwo, V. A. J. M. Sleiffer, M. Kuschnerov, H. de Waardt, and A. M. J. Koonen, "Single DPLL joint carrier phase compensation for few-mode fiber transmission," *IEEE Photon. Technol. Lett.*, vol. 25, no. 14, pp. 1381–1384, Jul. 2013.

[17] M. P. Yankov, L. Barletta, and D. Zibar, "Phase noise compensation for nonlinearity-tolerant digital subcarrier systems with high-order QAM," *IEEE Photon. J.*, vol. 9, no. 5, Oct. 2017.

[18] S. Pecorino, S. Mandelli, L. Barletta, M. Magarini, and A. Spalvieri, "Bootstrapping iterative demodulation and decoding without pilot symbols," *J. Lightw. Technol.*, vol. 33, no. 17, pp. 3613–3622, Sep. 2015.

[19] J. Dauwels and H.-A. Loeliger, "Joint decoding and phase estimation: an exercise in factor graphs," in *Proc. IEEE International Symposium Inf. Theory*, Jun. 2003, p. 231.

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

[21] N. Noels, V. Lottici, A. Dejonghe, H. Steendam, M. Moeneclaey, M. Luise, and L. Vandendorpe, "A theoretical framework for soft-information-based synchronization in iterative (turbo) receivers," *EURASIP J. Wireless Commun. Netw.*, vol. 2005, no. 2, pp. 117–129, Apr. 2005.

[22] T. S. Shehata and M. El-Tanany, "Joint iterative detection and phase noise estimation algorithms using Kalman filtering," in *Proc. Canadian Workshop Inf. Theory*, May 2009, pp. 165–168.

[23] T. C. W. Schenk, X.-J. Tao, P. F. M. Smulders, and E. R. Fledderus, "Influence and suppression of phase noise in multi-antenna OFDM," in *Proc. Vehicular Technol. Conf.*, vol. 2, Sep. 2004, pp. 1443–1447.

[24] A. Tarable, G. Montorsi, S. Benedetto, and S. Chinnici, "An EM-based phase-noise estimator for MIMO systems," in *Proc. IEEE International Conf. Commun.*, Jun. 2013, pp. 3215–3219.

[25] L. Reggiani, L. Dossi, L. Barletta, and A. Spalvieri, "Extended Kalman filter for MIMO phase noise channels with independent oscillators," *IEEE Commun. Lett.*, vol. 22, no. 6, pp. 1200–1203, Jun. 2018.

[26] R. Krishnan, G. Colavolpe, A. Graell 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.

[27] A. O. Isikman, H. Mehrpouyan, A. A. Nasir, A. Graell i Amat, and R. A. Kennedy, "Joint phase noise estimation and data detection in coded multi-input–multi-output systems," *IET Commun.*, vol. 8, no. 7, pp. 981–989, May 2014.

[28] H. Mehrpouyan, A. A. Nasir, S. D. Blostein, T. Eriksson, G. K. Karagiannidis, and T. Svensson, "Joint estimation of channel and oscillator phase noise in MIMO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4790–4807, Sep. 2012.

[29] A. A. Nasir, H. Mehrpouyan, R. Schober, and Y. Hua, "Phase noise in MIMO systems: Bayesian Cramér–Rao bounds and soft-input estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2675–2692, May 2013.

[30] 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.

[31] M. J. Beal, "Variational algorithms for approximate Bayesian inference," Ph.D. dissertation, University College London, 2003.

[32] I. Fatadin, D. Ives, and S. J. Savory, "Blind equalization and carrier phase recovery in a 16-QAM optical coherent system," *J. Lightw. Technol.*, vol. 27, no. 15, pp. 3042–3049, Aug. 2009.

[33] ——, "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.

[34] K. Zanette, J. C. Cartledge, and M. O'Sullivan, "Correlation properties of the phase noise between pairs of lines in a quantum-dot optical frequency comb source," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2017, p. Th3I.6.

[35] W. Ryan and S. Lin, *Channel Codes: Classical and Modern*. Cambridge University Press, 2009.

[36] S. Särkkä, *Bayesian Filtering and Smoothing*, 1st ed. Cambridge, UK: Cambridge University Press, 2013.

[37] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.

[38] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Proc. Conf. Uncertainty Artificial Intelligence*, Aug. 2001, pp. 362–369.

[39] J. Dauwels, "On variational message passing on factor graphs," in *Proc. IEEE International Symposium Inf. Theory*, Jun. 2007, pp. 2546–2550.

[40] G. E. Kirkelund, C. N. Manchón, L. P. B. Christensen, E. Riegler, and B. H. Fleury, "Variational message-passing for joint channel estimation and decoding in MIMO-OFDM," in *Proc. Global Telecommun. Conf.*, Dec. 2010.

[41] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[42] S. L. Miller and D. Childers, *Probability and Random Processes: With Applications to Signal Processing and Communications*, 2nd ed. Cambridge, MA, USA: Academic Press, 2012.

[43] A. F. Alfredsson, E. Agrell, H. Wymeersch, and M. Karlsson, "Pilot distributions for phase tracking in space-division multiplexed systems," in *Proc. European Conf. Opt. Commun.*, Sep. 2017, p. P1.SC3.48.

[44] A. F. Alfredsson, E. Agrell, H. Wymeersch, B. J. Puttnam, G. Rademacher, R. S. Luís, and M. Karlsson, "Pilot-aided joint-channel carrier-phase estimation in space-division multiplexed multicore fiber transmission," *J. Lightw. Technol.*, vol. 37, no. 4, pp. 1133–1142, Feb. 2019.

[45] 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.