Low-Complexity Tracking of Laser and Nonlinear Phase Noise in WDM Optical Fiber Systems

Yankov, Metodi Plamenov; Fehenberger, Tobias; Barletta, Luca; Hanik, Norbert

Published in:
Journal of Lightwave Technology

Link to article, DOI:
10.1109/JLT.2015.2493202

Publication date:
2015

Document Version
Peer reviewed version

Low-Complexity Tracking of Laser and Nonlinear Phase Noise in WDM Optical Fiber Systems

Metodi P. Yankov, *Student Member, IEEE,* Tobias Fehenberger, *Student Member, IEEE,* Luca Barletta, *Member, IEEE,* and Norbert Hanik, *Senior Member, IEEE*

***Abstract*—In this paper the wavelength division multiplexed (WDM) fiber optic channel is considered. It is shown that for ideal distributed Raman amplification (IDRA), the Wiener process model is suitable for the non-linear phase noise due to cross phase modulation from neighboring channels. Based on this model, a phase noise tracking algorithm is presented. We approximate the distribution of the phase noise at each time instant by a mixture of Tikhonov distributions, and derive a closed form expression for the posterior probabilities of the input symbols. This reduces the complexity dramatically compared to previous trellis based approaches, which require numerical integration. Further, the proposed method performs very well in low-to-moderate signal-to-noise ratio (SNR), where standard decision directed (DD) methods, especially for high order modulation, fail. The proposed algorithm does not rely on averaging, and therefore does not experience high error floors at high SNR in severe phase noise scenarios. The laser linewidth (LLW) tolerance is thereby increased for the entire SNR region compared to previous DD methods. In IDRA WDM links the algorithm is shown to effectively combat the combined effect of both laser phase noise and non-linear phase noise, which cannot be neglected in such scenarios. In a more practical lumped amplification scheme we show near-optimal performance for 16QAM, 64QAM and 256QAM with LLW up to 100kHz, and reasonable performance for LLW of 1MHz for 16QAM and 64QAM, at the moderate received SNR region. The performance in these cases is close to the information rate achieved by the above mentioned trellis processing.**

***Index Terms*—Phase noise, WDM, optical fiber communication, Wiener process, Cross phase modulation**

I. Introduction

THE non-linear phase noise (NLPN) is the main factor for the currently limited achievable rates on the wavelength division multiplexed (WDM) optical fiber channels [1]. Due to the interaction between self and cross phase modulation (SPM and XPM, resp.) effects, the amplified spontaneous emission (ASE) noise and chromatic dispersion in the fiber, the non-linear phase rotation due to SPM and XPM generally cannot be canceled completely at the receiver. Furthermore, XPM and SPM introduce memory in the channel, which makes optimal detection even more challenging. Recent works have gone into modeling this memory [2]–[4], and shown that the phase noise is highly correlated in time and frequency. A model is also proposed that allows to calculate the auto correlation function (ACF) of the phase noise. Using the time domain ACF properties, the authors in [2] separate the contribution of the noise into linear and non-linear part, and thereby are able to show potential increase in the maximum achievable rate in a point to point WDM link. A strategy for canceling the phase noise is proposed in [5], based on a frequency domain equalizer. This work is extended in [6] to a multi-carrier modulation, which is shown to be beneficial in combating XPM interference. Gaussian input is assumed in [2] [5] [6], which is not currently realizable in practice.

On top of the NLPN, fiber-optic systems generally suffer from phase noise due to imperfect lasers. The non-zero laser line width (LLW) results in time-varying carrier phase offset. The Wiener process has previously been shown to be accurate in modeling the laser phase noise, and also the NLPN for a few particular cases [7]. For QPSK constellations simple methods such as Viterbi and Viterbi are shown to be effective for carrier phase recovery. Extensions of this method to larger order constellations are possible, e.g. [8]. However, they generally require hard decision on the signal's amplitude before the phase is estimated. Other types of decision directed (DD) methods are available for higher order constellations, e.g. [10] [11] [12]. In [12], a second order DD phase-locked loop (PLL) method is employed, which is shown to be effective also in the presence of frequency fluctuations. We highlight the method from [10] as very popular among engineers, due to its simple implementation. DD methods typically require an SNR relatively high to the order of modulation (equivalently, low symbol error rate (SER)). Alternatively, the phase offset may be estimated from very long window averages, which makes the system unreliable in severe phase noise scenarios, since the phase varies significantly within the window. Another problem with the method from [10] is its vulnerability to phase slips. A modulation size independent method was derived in [14] for quadrature amplitude modulation (QAM) constellations, which aims at forcing the received symbols in each quadrant to the original square shape, and uses a PLL-like circuit to track the phase offset. This method also suffers greatly at low SNR/high SER, however it is very simple to implement. The above methods do not generally exploit the Wiener process model for the phase noise in order to improve the estimate. The authors in [13] propose an extension to the algorithm from [10], where a pilot-based coarse estimation is combined with the sub-sequent DD estimation, in order to combat the phase-slip problem.

M. Yankov is with the Department of Photonics Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark: e-mail: meya@fotonik.dtu.dk

T. Fehenberger, L. Barletta and N. Hanik are with the Institute for Communications Engineering, Technische Universität München, 80333 Munich, Germany.

Iterative decoding and laser phase-noise cancellation is proposed in [15], later extended to cover the above mentioned NLPN in WDM systems [16]. Up to 16QAM is considered there. The clear problems with iterative receivers are complexity and latency.

In [17], a Kalman filter is used that needs to be linearized since the output is not a linear function of the phase noise. The Kalman filter approach has low complexity and was shown to be near-optimal in some cases [18]. However, it suffers performance degradation for high information rates and SNRs. In a more general approach [19] the phase is discretized into bins, and trellis processing is used for phase tracking, Mutual Information (MI) calculation and demodulation. This method allows for processing of non-linear functions of the phase. Both [17] and [19] are basically a special case of the sum-product algorithm for finding marginal and posterior distributions, which can be represented with a graph. In [20] this algorithm is presented in the context of phase noise, where the graph may be constructed as a Markov chain. Several approximations are also proposed there for constant amplitude modulation, such as phase-shift keying (PSK). Of these approximations we mention the modeling of the distribution of the phase noise as a Tikhonov (also known as von Mises) distribution.

In this paper we show that the Wiener process is suitable for modeling the phase noise due to XPM in WDM systems with ideal distributed raman amplification (IDRA). Then we propose a simple approximate detector based on phase noise tracking, which is able to combat the combined effect of laser phase noise and NLPN in the fiber-optic channel. Focus is given on discrete input constellations, particularly QAM.

II. Channel Model

The fiber channel model under investigation is given in Fig. 1. Data is modulated into constellation symbols x, which are drawn from a finite size alphabet $\mathcal{X}$. The modulated symbols are then passed through a Nyquist pulse shaping filter and up-converted to the desired carrier frequency on the frequency grid. During upconversion laser phase noise is introduced, which is modeled by a Wiener process. The signal is then combined with the other channels, and sent on N_{span} spans of optical fiber. In order to compensate for attenuation, IDRA is employed [1]. At the receiver, the WDM channel is down-converted to baseband, while introducing laser phase noise identically distributed to the transmitter's. The interfering channels are filtered out, and the desired channel is sent for baseband processing. This includes digital back-propagation (DBP) of the channel of interest only, in order to remove SPM, and the subsequent phase noise tracking algorithm. The signal after DBP is denoted as y. We are interested in the MI between x and y

$$\mathcal{I}(X;Y) = \mathcal{H}(X) - \mathcal{H}(X|Y) = \mathcal{H}(X) + \lim_{K\to\infty} \frac{1}{K} \log_2(p(x_1^K|y_1^K)), \quad (1)$$

where x_1^K and y_1^K denote the input and output sequences from time 1 to K, respectively, and we have used the convergence properties of the entropy function for long sequences [21]. Evaluating the joint and/or conditional probability of the entire sequences is an exponentially complex problem, and so a typical receiver will usually approximate $p(x_1^K|y_1^K) \approx \prod_k p(x_k|y_1^K)$, or even $p(x_1^K|y_1^K) \approx \prod_k p(x_k|y_k)$. When such a mismatched receiver is employed, the result is an upper bound on the entropy $\mathcal{H}(X|Y)$ [21], that gives an achievable information rate (AIR), which is a lower bound on the MI rate in (1).

A. Simplified Channel Model

In order to design a phase noise tracking algorithm, we employ the following simplified model of a standard phase noise channel

$$y_k = x_k \exp(j(\theta_k^{tx} + \theta_k^{rx} + \theta_k^{NL})) + w_k, \quad (2)$$

where $j = \sqrt{-1}$ is the imaginary unit. The phase noise contribution from the transmitter, receiver and the non-linearity (θ^{tx}, θ^{rx} and θ^{NL}, respectively) are all modeled by a Wiener process, e.g., for the transmitter term

$$\theta_k^{tx} = \theta_{k-1}^{tx} + \Delta_{tx} v_k, \quad (3)$$

where the v_k's are standard i.i.d. Gaussian variables, and Δ_{tx}^2 is the process noise variance. If we assume independent sources, they can be combined into a single phase noise process θ with parameter $\Delta^2 = \Delta_{tx}^2 + \Delta_{rx}^2 + \Delta_{NL}^2$, and the channel model becomes

$$y_k = x_k \exp(j\theta_k) + w_k. \quad (4)$$

The term w_k is a sample of additive white Gaussian noise (AWGN) with zero mean and variance assumed to be known at the receiver. When the phase noise is generated by a local oscillator (LO) with a certain spectral width around the central frequency (in case of lasers the more popular term is linewidth), the parameter Δ is found as

$$\Delta^2 = 2\pi f_W T_s, \quad (5)$$

where f_W is the width of the spectrum in Hz at -3dB of the maximum value (also known as full-width half-maximum bandwidth), and T_s is the sampling time in seconds.

B. The Non-Linear Term

In order to validate the Wiener model for the NLPN, we simulate an IDRA link, and examine the power spectral density (PSD) of the phase noise process. For demonstration, we choose a link with 5 channels at 100 GBaud each, of total length 4000km (the other fiber parameters are standard, and are listed in Table II), *without* transmitter and receiver phase noise. The input constellation in this case is 256QAM and the input power is -4dBm, which we found to be optimal at this distance. As shown in [2] [22], the NLPN is highly correlated within a window of several tens of symbols, which can be exploited in order to estimate the actual phase noise samples as

$$\hat{\theta}_k = \angle \sum_{l=k-L/2}^{k+L/2} y_l x_l^*, \quad (6)$$

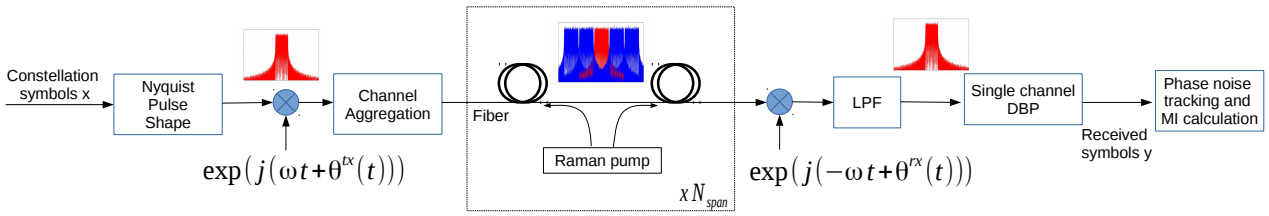

Fig. 1. Optical fiber channel model. Fiber parameters are defined in Section IV.

where $L+1$ is the window size, $(\cdot)^*$ denotes complex conjugate, and $\angle(\cdot)$ denotes the angle of a complex number. We then compare the estimated PSD of $\{\hat{\theta}_k\}$ to the theoretical PSD of a Wiener process given by the Lorentzian function with process noise variance

$$\Delta^2 = \mathrm{E}_k\left[(\hat{\theta}_k - \hat{\theta}_{k-1})^2\right]. \tag{7}$$

Example PSDs are given in Fig. 2 for different values of L. The PSD of $\{\hat{\theta}_k\}$ is obtained by Welch's method [23] with 10^6 samples, which were divided into 1999 blocks, 1000 samples each, with 50% overlap. Depending on the window size reasonable match can be obtained to the theoretical model. Long window results in relatively smooth PSD due to the better estimation in presence of noise. This would translate to better modeling of $\{\theta_k\}$ by the Wiener process. However, it leads to underestimation of Δ^2_{NL}. Decreasing L results in stronger oscillations at high frequency, and a large bias in the low frequency. In the rest of the paper the window size for estimating the samples θ_k, and thereby Δ^2, is chosen to be $L=20$, which as we see in Fig. 2 is a reasonable compromise between modeling accuracy and bias.

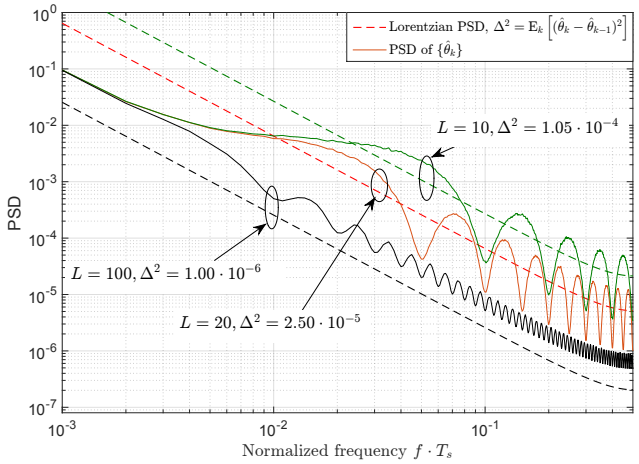

Fig. 2. Lorentzian PSD of a Wiener process (dashed lines) with process noise variance as in (7), together with the PSD of the phase noise $\{\hat{\theta}_k\}$ (solid lines), for the link given in Section II-B. Depending on the choice of the window L a good match can be found to the theoretical model.

III. Phase Noise Tracking

As mentioned in Section II, in order to compute AIRs, we need to compute the posterior probability of the input sequence $p(x_1^K|y_1^K)$, which we approximate as $\prod_k p(x_k|y_1^K)$, resulting in a lower bound on the MI rate. In this section we propose an efficient algorithm for calculating the posterior distributions at each time recursively.

Marginalizing the phase noise at time k, the posterior can be re-written as

$$\begin{aligned}
p(x_k|y_1^K) &= \int_{-\pi}^{\pi} p(x_k,\theta_k|y_1^K)d\theta_k \\
&= \int_{-\pi}^{\pi} p(x_k|\theta_k,y_1^K)p(\theta_k|y_1^K)d\theta_k \\
&= \int_{-\pi}^{\pi} \frac{p(x_k)p(y_k|x_k,\theta_k)}{p(y_k|\theta_k)}p(\theta_k|y_1^K)d\theta_k \qquad (8)\\
&\propto \int_{-\pi}^{\pi} \frac{p(x_k)p(y_k|x_k,\theta_k)}{p(y_k|\theta_k)}p(y_k|\theta_k)p(y_{k+1}^K|\theta_k)p(\theta_k|y_1^{k-1})d\theta_k \qquad (9)\\
&= \int_{-\pi}^{\pi} p(x_k)p(y_k|x_k,\theta_k)p(y_{k+1}^K|\theta_k)p(\theta_k|y_1^{k-1})d\theta_k. \qquad (10)
\end{aligned}$$

To get to (8) we used the fact that given the state θ_k, the input samples are independent of the past and future: $p(x_k|\theta_k,y_1^K) = p(x_k|\theta_k,y_k)$. To get to (10) we have used the fact that output samples are independent of the past given the phase: $p(y_{k+1}^K|\theta_k,y_1^{k-1}) = p(y_{k+1}^K|\theta_k)$, and we have removed the factor $p(y_{k+1}^K|y_1^{k-1})$ that does not depend on x_k and θ_k. The first and second term under the integral in Eq. (10) are the prior distribution of the constellation symbols and the likelihood of the output at time k, respectively. In order to derive expressions for the last two terms, we first define forward and backward recursions for the posterior distributions $p(\theta_k|y_1^k)$ and $p(y_k^K|\theta_k)$, which we model by mixtures of M and N Tikhonov distributions, respectively

$$p(\theta_k|y_1^k) = \sum_{m=1}^{M} \alpha_{m,k}\, t(w_{m,k};\theta_k), \tag{11}$$

$$p(y_k^K|\theta_k) = \sum_{n=1}^{N} \beta_{n,k}\, t(u_{n,k};\theta_k). \tag{12}$$

The terms $\alpha_{m,k}$ and $\beta_{n,k}$ are non-negative mixing coefficients, and are such that $\sum_m \alpha_{m,k} = 1$ and $\sum_n \beta_{n,k} = 1$. The Tikhonov distribution at θ with a complex parameter w is defined as

$$t(w;\theta) = \frac{\exp(\mathrm{Re}\,[w\cdot\exp(-j\theta)])}{2\pi I_0(|w|)}, \qquad \theta\in[-\pi;\pi), \tag{13}$$

and 0 elsewhere. In (13), I_0 is the zero-th order modified Bessel function of the first kind and $\mathrm{Re}\,[z]$ is the real part of z. We now derive mixture parameters for the predictive backward recursion term that appears in Eq. (10)

$$\begin{aligned} p(y_{k+1}^K|\theta_k) &= \sum_{n=1}^{N} \bar{\beta}_{n,k} t(\bar{u}_{n,k};\theta_k) \\ &= \int_{-\pi}^{\pi} p(y_{k+1}^K|\theta_{k+1},\theta_k) p(\theta_{k+1}|\theta_k) d\theta_{k+1} \\ &= \int_{-\pi}^{\pi} p(y_{k+1}^K|\theta_{k+1}) g(\theta_k,\Delta;\theta_{k+1}) d\theta_{k+1} \\ &= \sum_{n=1}^{N} \beta_{n,k+1} \int_{-\pi}^{\pi} t(u_{n,k+1};\theta_{k+1}) g(\theta_{k+1},\Delta;\theta_k) d\theta_{k+1}, \end{aligned} \tag{14}$$

where $g(\mu,\sigma;z)$ is the Gaussian probability density function (PDF) at z with mean μ and standard deviation σ. The update parameters can then be found as

$$\bar{\beta}_{n,k} = \beta_{n,k+1}, \qquad \bar{u}_{n,k} = \frac{u_{n,k+1}}{1+\Delta^2|u_{n,k+1}|}, \tag{15}$$

where we have used the fact that the convolution of a Gaussian and Tikhonov distributions is a Tikhonov with a modified complex parameter [20]. In order to complete the recursion, the updates for $\beta_{n,k}$ and $u_{n,k}$ are found from the following:

$$\begin{aligned} p(y_k^K|\theta_k) &= p(y_k|\theta_k, y_{k+1}^K) p(y_{k+1}^K|\theta_k) \\ &= \sum_{n=1}^{N} \bar{\beta}_{n,k} \sum_{x_k\in\mathcal{X}} p(x_k) p(y_k|x_k,\theta_k) t(\bar{u}_{n,k};\theta_k) \\ &\propto \sum_{n=1}^{N} \sum_{x_k\in\mathcal{X}} \mu_{n,k}(x_k) t(\bar{u}_{n,k} + 2\cdot\mathrm{SNR}\cdot y_k x_k^*;\theta_k), \end{aligned} \tag{16}$$

where the likelihood $p(y_k|x_k,\theta_k)$ is expressed as a Tikhonov approximation to the Gaussian

$$p(y_k|x_k,\theta_k) \approx \frac{2\cdot\mathrm{SNR}\cdot I_0(2\cdot\mathrm{SNR}|y_k x_k^*|) t(2\cdot\mathrm{SNR}\cdot y_k x_k^*;\theta_k)}{\exp(\mathrm{SNR}(|y_k|^2+|x_k|^2))}. \tag{17}$$

In (16) we have used the fact that the product of two Tikhonov distributions may also be expressed as a Tikhonov distribution in order to calculate the sub-component mixture coefficient

$$\mu_{n,k}(x_k) = \frac{\bar{\beta}_{n,k}\cdot p(x_k) I_0(|\bar{u}_{n,k} + 2\cdot\mathrm{SNR}\cdot y_k x_k^*|)}{I_0(|\bar{u}_{n,k}|)\exp(\mathrm{SNR}\cdot|x_k|^2)}. \tag{18}$$

Due to the discrete nature of the input signal, the number of components needed for tracking the phase noise grows exponentially with time. In order to avoid this problem, we propose an approximation to the inner sum in (16), where at each step we only take the sub-component with the largest mixing coefficient

$$\hat{x}_{n,k} = \arg\max_{x_k\in\mathcal{X}} \mu_{n,k}(x_k), \tag{19}$$

$$u_{n,k} = \bar{u}_{n,k} + 2\cdot\mathrm{SNR}\cdot y_k \hat{x}_{n,k}^*, \tag{20}$$

$$\beta_{n,k} = B\cdot\mu_{n,k}(\hat{x}_{n,k}), \tag{21}$$

where B is such that $\sum_{n=1}^{N}\beta_{n,k} = 1$.

Similarly we express the predictive forward distribution appearing in (10), as

$$p(\theta_k|y_1^{k-1}) = \sum_{m=1}^{M} \bar{\alpha}_{m,k} t(\bar{w}_{m,k};\theta_k), \tag{22}$$

where

$$\bar{\alpha}_{m,k} = \alpha_{m,k-1}, \qquad \bar{w}_{m,k} = \frac{w_{m,k-1}}{1+\Delta^2|w_{m,k-1}|}. \tag{23}$$

The update parameters are found as

$$\rho_{m,k}(x_k) = \frac{\bar{\alpha}_{m,k}\cdot p(x_k) I_0(|\bar{w}_{m,k} + 2\cdot\mathrm{SNR}\cdot y_k x_k^*|)}{I_0(|\bar{w}_{m,k}|)\exp(\mathrm{SNR}\cdot|x_k|^2)}, \tag{24}$$

$$\hat{x}_{m,k} = \arg\max_{x_k\in\mathcal{X}} \rho_{m,k}(x_k), \tag{25}$$

$$w_{m,k} = \bar{w}_{m,k} + 2\cdot\mathrm{SNR}\cdot y_k \hat{x}_{m,k}^*, \tag{26}$$

$$\alpha_{m,k} = A\cdot\rho_{m,k}(\hat{x}_{m,k}), \tag{27}$$

where A is such that $\sum_{m=1}^{M}\alpha_{m,k} = 1$. We are now ready to express the posterior distribution (10) as

$$\begin{aligned} p(x_k|y_1^K) = p(x_k) \sum_{m=1}^{M} \bar{\alpha}_{m,k} \sum_{n=1}^{N} \bar{\beta}_{n,k} \times \\ \int_{-\pi}^{\pi} p(y_k|x_k,\theta_k) t(\bar{u}_{n,k};\theta_k) t(\bar{w}_{m,k};\theta_k) d\theta_k. \end{aligned} \tag{28}$$

Using the expression for the likelihood (17), the integrand in (28) becomes a product of three Tikhonov distributions in θ_k, which is again a scaled Tikhonov of θ_k

$$\begin{aligned} p(y_k|x_k,\theta_k) t(\bar{u}_{n,k};\theta_k) t(\bar{w}_{m,k};\theta_k) \propto \\ \frac{I_0(|\bar{w}_{m,k} + \bar{u}_{n,k} + 2\cdot\mathrm{SNR}\cdot y_k x_k^*|)}{I_0(|\bar{w}_{m,k}|) I_0(|\bar{u}_{n,k}|)\exp(\mathrm{SNR}(|y_k|^2+|x_k|^2))} \times \\ t(\bar{w}_{m,k} + \bar{u}_{n,k} + 2\cdot\mathrm{SNR}\cdot y_k x_k^*;\theta_k), \end{aligned} \tag{29}$$

where we have removed the proportionality factors, independent of θ_k and x_k. The scaling factor goes out of the integral in (28), and the remaining Tikhonov distribution integrates to one. The expression for the posterior is then

$$\begin{aligned} p(x_k|y_1^K) = p(x_k) \sum_{m=1}^{M} \bar{\alpha}_{m,k} \sum_{n=1}^{N} \bar{\beta}_{n,k} \times \\ \frac{I_0(|\bar{w}_{m,k} + \bar{u}_{n,k} + 2\cdot\mathrm{SNR}\cdot y_k x_k^*|)}{I_0(|\bar{w}_{m,k}|) I_0(|\bar{u}_{n,k}|)\exp(\mathrm{SNR}(|y_k|^2+|x_k|^2))}. \end{aligned} \tag{30}$$

The expression (30) is a sum of $N\cdot M$ elements, which is very efficiently calculated in the log domain, using the large value approximation of the modified Bessel function

$$I_0(w) \approx \frac{\exp(w)}{\sqrt{2\pi w}}. \tag{31}$$

Typical values of $|\bar{w}|$ and $|\bar{u}|$ are above 200, for which the approximation in (31) is accurate.

A. Summary

The complete algorithm is summarized in the following steps:

1) **Initialize**:

$$\alpha_{m,0} = 1/M, \qquad \beta_{n,K+1} = 1/N,$$
$$w_{m,0} = \frac{1}{\Delta^2} e^{j(m\cdot\frac{2\pi}{M}-\pi)}, \qquad u_{n,K+1} = \frac{1}{\Delta^2} e^{j(n\cdot\frac{2\pi}{N}-\pi)} \tag{32}$$

2) **Forward recursion** - equations (23),(25)-(27)
3) **Backward recursion** - equations (15),(19)-(21)
4) **Posteriors calculation** - Eq. (30)

The initialization values are chosen such that each component corresponds to a Gaussian distribution of the phase with variance Δ^2, and the means of the components are uniformly spaced within $[-\pi;\pi)$.

B. Phase Noise Distribution

The phase noise PDF at each time is not explicitly calculated by the algorithm, but can be found, if needed for further processing, as

$$p(\theta_k|y_1^k) \propto p(y_{k+1}^K|\theta_k)p(\theta_k|y_1^k) = \sum_{n=1}^{N} \bar{\beta}_{n,k}\, t(\bar{u}_{n,k};\theta_k) \sum_{m=1}^{M} \alpha_{m,k}\, t(w_{m,k};\theta_k). \tag{33}$$

An example distribution update is given in the surface plot in Fig. 3 for a standard Wiener phase noise channel, simulated via (4), with parameter $\Delta = 0.05$ and 256QAM input. The distribution is given as a surface plot, where the height of the surface is described by the color bar. We evaluate the phase noise distribution for all phases within $[-\pi;\pi)$, with a resolution of $2\pi/512$ radians. We also plot the actual phase noise realizations. In this case the number of mixture components in the forward and backward recursions is $M = N = 4$. We clearly see the different components, one of which tracks the true phase noise realization. Due to the unknown initial phase and the 4-fold symmetry of the constellation, there is phase ambiguity in multiples of $\pi/2$ radians. In order to combat this problem, we insert pilot symbols at rate P, which give an absolute reference for the phase. Pilot symbols are accounted for in the model by changing the PMF $p(x_k)$ at the pilot instants to an indicator function, which is '1' for the true sent symbol, and '0' otherwise. Even though this is a sub-optimal pilot design, it suffices for our further analysis. The resulting phase noise distribution after pilot insertion is given in Fig. 4.

C. Relation to Other Algorithms

When the PDFs $p(\theta_k|y_1^K)$, $p(\theta_k|y_1^k)$ and $p(y_k^K|\theta_k)$ are modeled by a Gaussian, the solution of the recursions is the Kalman filter [9]. The difference between our approach and the Kalman filter is visible in Eq. (28). If the densities are modeled as Gaussians, the likelihood $p(y_k|\theta_k, x_k)$, which is a non-linear function of the phase θ_k, needs to be linearized in order to make the integration simple, as done here. The linearization leads to sub-optimal performance of the Kalman filter. We note that the Tikhonov distribution may be seen as a Gaussian "wrapped" around $\pm\pi$. Example of the distribution with mean

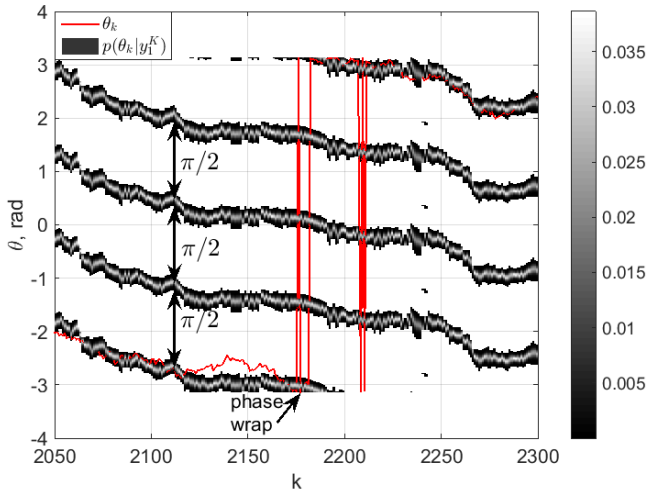

Fig. 3. Probability distribution of the phase noise at time k (given by the surface plot), together with actual phase noise samples, given by the red line (pale line on black and white printer). Brighter color represents higher probability, given by the side color bar.

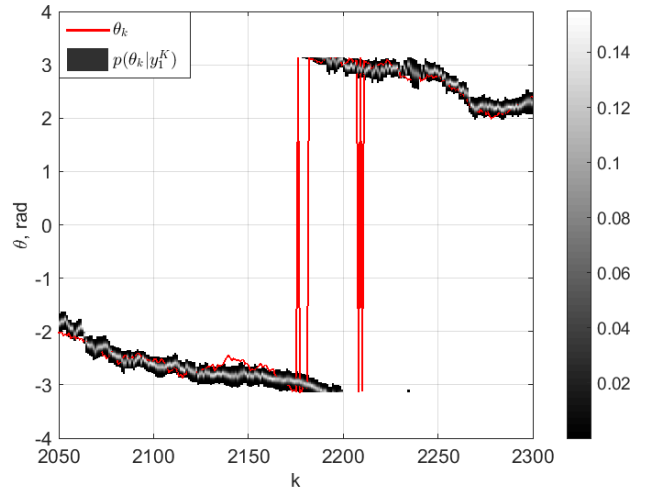

Fig. 4. Probability distribution of the phase noise at time k after insertion of $P = 0.2\%$ pilot symbols. Brighter color represents higher probability, given by the side color bar. Note the different scale w.r.t Fig. 3 due to the absence of the $\pi/2$ ambiguity.

$-\pi/4$ and different variances is given in Fig. 5, together with the Gaussian and a wrapped Gaussian, as defined in the legend. The latter distribution and the Tikhonov have support set $[-\pi;\pi)$. For small variance the three distributions coincide due to the exponentially vanishing tail of the Gaussian. For large variance we see the wrapping becoming significant and the distributions diverging from each other.

The trellis approach from [19] replaces the above distributions with histograms. This means that the integrations for the updates (Eq. (14) and its forward recursion analogue) and the posterior calculation (Eq. (28)) become sums. For large constellations and fine resolution of the phase, the complexity increases significantly.

The approach in [20] also employs Tikhonov distributions, however, only for PSK constellations. There, the information is

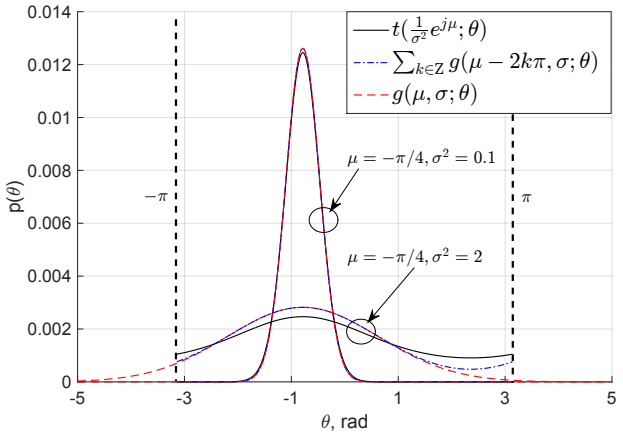

Fig. 5. Example Tikhonov vs. Gaussian and wrapped Gaussian distributions with the same parameters. For small variance the three distributions are almost identical.

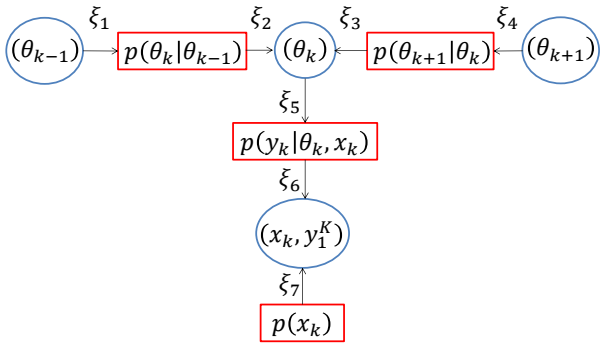

Fig. 6. One section of the graph, used for estimating the posterior distributions. Factors are given as rectangles, and variables with circles. The factors are represented by the corresponding distributions.

already included in the phase and therefore phase tracking and posterior calculation can be performed simultaneously. The focus here is on QAM constellations.

We note that all of the above approaches, including the one proposed in this paper, in principle employ the sum-product rules for estimating densities in graphs. In this particular case, the graph represents a first-order hidden Markov chain. One section of the graph for this algorithm is given in Fig. 6. The variable nodes are given as circles, and the factor nodes as rectangles. The direction of the messages ξ is given by the arrows, and the messages themselves are calculated as:

- $\xi_1 = p(\theta_{k-1}|y_1^{k-1})$
- ξ_2 - Eq. (22)
- ξ_3 - Eq. (14)
- $\xi_4 = p(y_{k+1}^K|\theta_{k+1})$
- $\xi_5 = \xi_2 \cdot \xi_3$
- $\xi_6 = \int_{-\pi}^{\pi} \xi_5 \cdot p(y_k|\theta_k, x_k) d\theta_k$
- $\xi_7 = p(x_k)$

According to the sum-product rule, the unnormalized distribution at each variable is the product of all incoming messages. Then we find $p(x_k|y_1^K) \propto p(x_k, y_1^K) = \xi_6 \cdot \xi_7$, which is exactly the expression in Eq. (10).

D. Complexity

In the following we refer to the proposed algorithm as Tikhonov mixture model (TMM) algorithm.

The complete algorithm can be cast into the log-domain with standard max-log approximations. Thus the complexity is dominated by the computation of Euclidean distances (ED), needed for calculating the largest mixing coefficients in (19) and (25). There, the complexity is linear in the constellation size and the number of mixing coefficients. The forward and backward recursions require one max-log operation at each time across M and N elements (for the normalization in (21) and (27), respectively), in order to calculate the updates in (15) and (23). The calculation of the posterior requires one max-log across $N \cdot M$ elements, bringing the total number of max-log operations to $\approx \mathcal{O}(K \cdot (M \cdot N + M + N))$. Typical values of M and N are as small as 2 or 4, so we conclude that the complexity is still dominated by the EDs calculation, which requires complex multiplication. We compare these numbers to the DD algorithm [10], where a cost function of the EDs to the closest constellation symbol is calculated for N_p candidate phases, and the minimum is taken as the phase rotation. The number of EDs (complex multiplications) needed is therefore $\mathcal{O}(K \cdot N_p \cdot |\mathcal{X}|)$. Typical number of test phases (taken directly from [10]) is $N_p \approx 16$ for $|\mathcal{X}| \leq 16$ and $N_p \approx 64$ for $|\mathcal{X}| \leq 256$. Our method requires $\mathcal{O}(K \cdot (M+N) \cdot |\mathcal{X}|)$, which for 256QAM and $M = N = 2$ Tikhonov components is more than an order smaller.

One issue with the sequential processing of our algorithm is the latency. We note that even though very long sequences are needed for the convergence $\lim_{K\to\infty} -\frac{1}{K}\log_2 p(x_1^K|y_1^K) = \mathcal{H}(X|Y)$, a real receiver does not aim at computing the entropy. The value of K may therefore be kept at a reasonable value, while keeping the posteriors in (30) accurate.

IV. Results

We examine the performance of the TMM algorithm mainly in terms of AIR in *bits/channel use*. In case of a standard Wiener phase noise channel 1 bit/channel use means 1 bit/s/Hz/complex dimension. In case of the WDM link, 1 bit/channel use means 1 bit/s/Hz/polarization/complex dimension. The input power in the latter case is defined as input power per channel. In all cases, the AIR is estimated from a block of length 10^5 symbols.

A. Standard Wiener Phase Noise Channel

We start by analyzing a standard Wiener phase noise channel, which is simulated via (4). In Fig. 7, the AIRs are given for a channel with $f_W \cdot T_s = 8 \cdot 10^{-5}$ and 256QAM input. This is beyond the capabilities of the DD algorithm [10], and our simulations also confirm that the DD algorithm fails with such parameters, giving lower bounds on the MI well below zero. We examine the performance of the TMM algorithm with $M = N = 2$ and 4 mixture components, and pilot rates of $P = 0.005, 0.01, 0.02$, and 0.05. For reference, we also plot the AIR on an AWGN channel without phase noise, and what is achieved by the trellis algorithm from [19] with pilot rate $P = 0.002$ and 128 states. The rate loss due to the pilot

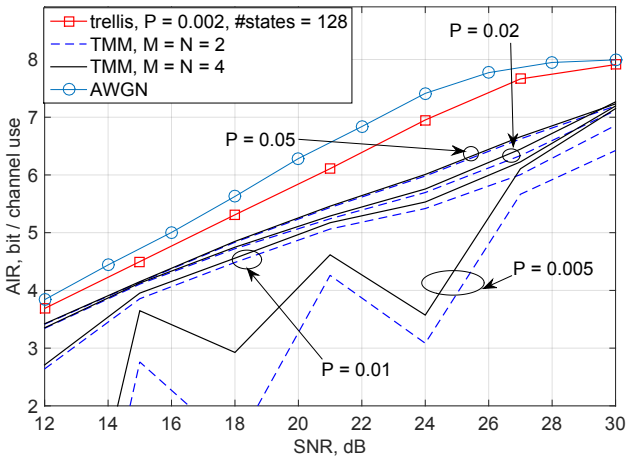

Fig. 7. Comparison of pilot rates for the TMM algorithm. From bottom curve to top: $P = 0.005, 0.01, 0.02, 0.05$. Standard Wiener phase noise channel, $f_W \cdot T_s = 8 \cdot 10^{-5}$, 256QAM input. For small pilot rates the algorithm is unstable.

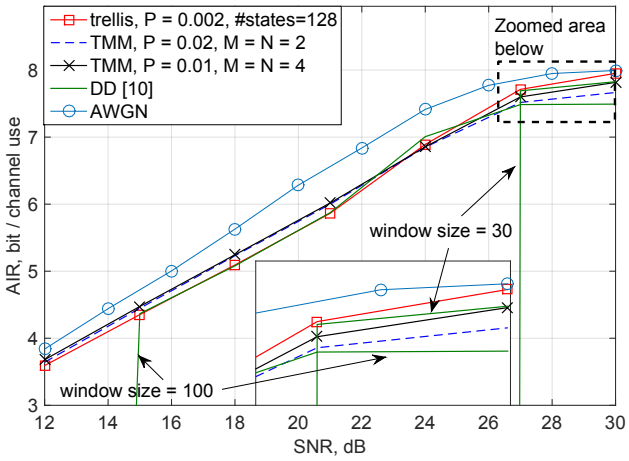

Fig. 8. Comparison with the DD algorithm. The trellis and the TMM algorithm have optimized pilot rates, given in the legend. The DD algorithm is tested with sliding window length of 30 and 100 symbols. Inset: very high rate (low BER, respectively). Standard Wiener phase noise channel, $f_W \cdot T_s = 8 \cdot 10^{-6}$, 256QAM input.

symbols is taken into account in the AIR calculation by scaling the entropy $\mathcal{H}(X)$ in (1) by $1 - P$. For such large values of Δ^2 and insufficient pilot rate the algorithm is unstable, and often switches to adjacent $\pi/2$ components. This leads to very poor lower bounds on the MI, sometimes even negative. At pilot rate above 1% the algorithm is stable, and we see only marginal improvement going from 2 to 4 mixture components. The performance is sub-optimal at high rates, however a stable behavior is observed over the entire SNR range.

In Fig. 8 we show the performance at a smaller value $f_W \cdot T_s = 8 \cdot 10^{-6}$. In this case the algorithm from [10] is able to operate with minor SNR penalty at BER $= 10^{-3}$ with an averaging window of size around 30. This window represents the length over which test phases are evaluated, and should not be confused with the window L we used in Section II-B for estimating Δ^2. As seen from Fig. 8, the achievable rate with

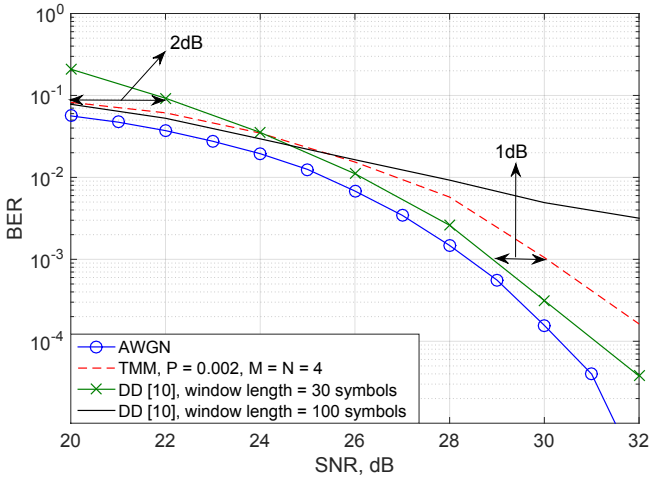

Fig. 9. BER at high SNR for the TMM algorithm and the DD algorithm [10] with 256QAM input, standard Wiener phase noise channel, $f_W \cdot T_s = 8 \cdot 10^{-6}$. For the DD algorithm, at low SNR short windows result in high BER, whereas at high SNR long windows result in high error floor. The TMM algorithm performs well for the entire region.

this window size is close to the maximum $\mathcal{H}(X) = \log_2 |\mathcal{X}|$, which corresponds to the above mentioned low BER and SER. However, when the SNR is reduced, the algorithm fails due to the higher SER. Increasing the window length helps at the lower SNRs, and at 100 symbols a stable behavior is observed down to SNR$=$ 15dB. Long averaging windows, however, result in a sub-optimal performance at high SNR, where the AIR appears to achieve a maximum value, smaller than $\mathcal{H}(X)$. This is because small-scale, fast variations of the phase noise are not captured by the long window averaging. This would correspond to an error floor after de-mapping. The TMM algorithm on the other hand achieves stable performance, close to the reference trellis based AIR at significantly reduced complexity, for the entire SNR region.

To further illustrate the problem with the error floor, we analyze the BER at high SNR. On Fig. 9, the BER is given at $f_W \cdot T_s = 8 \cdot 10^{-6}$. We assume Gray labeling of the symbols. The algorithm from [10] requires differential coded modulation, which technically will increase the BER slightly. Therefore we can argue that the comparison is fair, with a slight advantage given to the DD algorithm. At high SNR, the TMM algorithm experiences up to 1dB loss compared to the DD algorithm with averaging window of length 30 symbols. However, at low SNR (BER $\approx 10^{-1}$), such window length results in around 2dB loss compared to the TMM algorithm. Increasing the window length helps, but results in very high BER at high SNR. For completeness, in Table I the SNR penalty to the AWGN channel performance at BER$= 10^{-3}$ is given for several combinations of modulation format and LLW. The window length of the DD algorithm is 30, and we simulated SNRs up to 40dB. The TMM algorithm employs 4 components and $P = 0.002$ pilot rate. For high values of $f_W \cdot T_s$ the DD algorithm completely fails with this window length, while reception with the TMM algorithm is still possible. As mentioned above, decreasing the window size of the DD algorithm might improve the penalty and allow for

TABLE I
SNR PENALTY @ BER=10^{-3}. INFINITE PENALTY MEANS THAT THE REQUIRED BER IS NOT ACHIEVED.

$f_W \cdot T_s$	$5 \cdot 10^{-6}$	$1 \cdot 10^{-5}$	$5 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$5 \cdot 10^{-4}$
256QAM, TMM	1.1dB	1.3dB	4.9dB	7.8dB	∞
256QAM, DD [10]	0.6dB	0.9dB	∞	∞	∞
64QAM, TMM	1.0dB	1.2dB	2.2dB	4.4dB	9.8dB
64QAM, DD [10]	0.5dB	0.7dB	1.7dB	∞	∞
16QAM, TMM	0.9dB	1.1dB	1.2dB	1.3dB	3.1dB
16QAM, DD [10]	0.1dB	0.2dB	0.3dB	0.9dB	∞

transmission, but will result in increased penalty at lower SNR (higher BER, respectively).

B. IDRA WDM Optical Fiber Link

Next we evaluate the TMM algorithm in a WDM optical link. The link is simulated using the split-step Fourier method (SSFM). The fiber and transceiver parameters are given in Table II. Single polarization transmission is used and single-channel digital back-propagation is performed on the central channel, which is the channel of interest.

In Fig. 10, the AIRs are given as a function of the launch power per channel. For reference, we calculate the AIR with a pseudo-ideal phase noise removal (PIPNR). This is achieved by pre-processing the output samples as $\dot{y}_k = y_k e^{-j\dot{\theta}_k}$. The estimates of the phase noise $\dot{\theta}_k$ are obtained similar to (6), but from a window of past samples only, i.e. $\dot{\theta}_k = \angle \sum_{l=k-L-1}^{k-1} y_l x_l^*$. This is done in order for the calculated values to be AIRs. To better understand this, observe that the probability $p(x_1^K|y_1^K)$ may be expressed from the product rule as $p(x_1^K|y_1^K) \propto \prod_k p(x_k|y_1^K, x_1^{k-1})$, which means that if lower bounds on the MI are targeted, only past samples may be used to obtain mismatched probability distribution.

In Fig. 10 we also plot the AIR in the idealized case of no laser phase noise without any processing. In this case we assume memoryless channel and model the likelihoods $p(y_k|x_k)$ as Gaussian distributions with known mean and variance. To put the values of the LLW in the perspective of Section IV-A, we estimated the value of Δ_{NL}^2 at the optimal input power in the idealized case of $f_W = 0$ to be $\Delta_{NL}^2 \approx 2.5 \cdot 10^{-5}$. The term $f_W \cdot T_s$ then becomes $\approx 3.9 \cdot 10^{-6}, 4.2 \cdot 10^{-6}$ and $6 \cdot 10^{-6}$ in the 3 cases of LLW, respectively. The window size for the DD algorithm is optimized to 500 (we note that the AIR at the optimal input power increases very slowly for window sizes from 200 to 500, and then starts to decrease). The pilot rate for the trellis and the TMM algorithm is fixed to a minimal value of $P = 0.002$. The TMM algorithm has $M = N = 4$ mixture components. We see that the performance of the TMM algorithm is close to the pseudo-ideal one. The AIR in the idealized case without processing is below what is achieved in the 10kHz case when phase noise tracking is performed. This is due to the non-zero value of Δ_{NL}. The consequence of this observation is two-fold:

1) Even if ideal lasers are used, correlations in the NLPN can still be exploited by tracking it in order to improve the performance.
2) The algorithm for tracking should be able to account for the combined effect of laser phase noise and NLPN.

The latter observation suggests that even when perfect knowledge is available for the lasers, the combined phase noise process variance should always be estimated as $\Delta^2 = \Delta_{tx}^2 + \Delta_{rx}^2 + \Delta_{NL}^2$ for the desired link set-up. In case the lasers are not ideally characterized, the process noise variance may be estimated from training data via (7). The latter approach is also more robust to instabilities.

TABLE II
SYSTEM PARAMETERS, IDRA TRANSMISSION

Span length	100 km
Symbol rate	100 GBaud
Number of channels	5
Guardband	2 GHz (2% of symbol rate)
Total bandwidth	510 GHz
Oversampling factor	32
Pulse shape	$sinc$
LLW	0 kHz, 10 kHz and 100 kHz
Fiber loss	$\alpha = 0.2$ dB/km
Non-linear coefficient	$\gamma = 1.3$ (W·km)$^{-1}$
Dispersion	$D = 17$ ps/(nm·km)
Central wavelenght	$\lambda_0 = 1.55$ μm
SSFM step	0.1 km

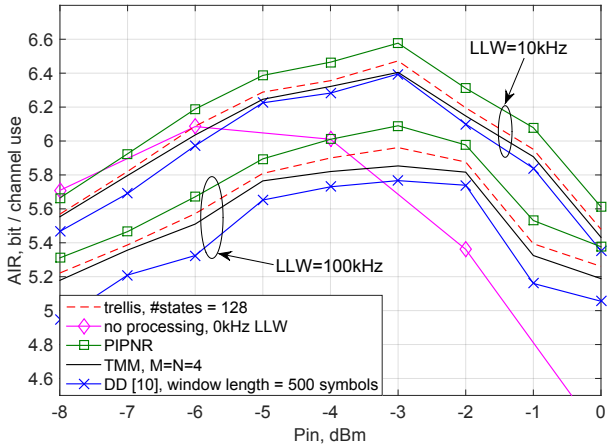

Fig. 10. AIRs after 40 spans, 256QAM, at different LLWs at transmitter and receiver, 100GBaud.

C. EDFA WDM Optical Fiber Link

In order to assess the performance of the TMM algorithm in a more practical scenario, we consider a lumped amplification scheme, where Erbium doped fiber amplifiers (EDFA) are inserted at the end of each fiber span, instead of the Raman pump. The new system parameters are given in Table III. The fiber parameters are the same as in Table II.

In this case we employ both polarizations. DBP is not performed here, but only electronic chromatic dispersion compensation in the frequency domain. Polarization mode dispersion is neglected in the SSFM.

TABLE III
SYSTEM PARAMETERS, EDFA TRANSMISSION

Symbol rate	28 GBaud
Number of channels	17
Guardband	0.56 GHz (2% of symbol rate)
EDFA noise figure	4 dB
LLW	10 kHz, 100 kHz and 1 MHz

We also study the performance of the algorithm for smaller constellations, where the LLW tolerance is generally higher. AIRs are given for 256QAM, 64QAM and 16QAM, at 10 spans, 30 spans and 50 spans, respectively, in Fig. 11 for 10kHz LLW, Fig. 12 for 100kHz LLW and Fig. 13 for 1MHz LLW. The distances and constellation sizes are chosen such that the maximum AIR at the optimal input power is smaller than $\approx \frac{3}{4}\log_2|\mathcal{X}|$, which is the desired operating point for energy and spectral efficient communications. This point can be seen as the maximum SNR, at which the slope of the MI with discrete input on a Gaussian channel is the same as for Gaussian input, i.e., where the AIR is not yet limited by the size of the constellation. For such information rates and sufficiently large values of Δ, DD methods generally perform poorly due to the high SER, regardless of the modulation format.

At 10kHz LLW the TMM algorithm achieves near-optimal performance in all cases up to the optimal input power for the respective distances and constellations. We only see an instability for very high input power and low SNR in the case of 16 QAM after 50 spans. Similar to the IDRA case, the DD algorithm from [10] requires around 500 samples for averaging out the noise. At 100kHz LLW, the phase noise cannot be considered constant for such a long period, and the DD algorithm fails. The TMM method requires increase in the pilot rate to $P = 0.005$, and achieves stable performance, close to that of the trellis method. We note that the number of states in the trellis was increased to 256, and we see that in the case of 256 QAM there is still around 0.5 bits/channel use gap to the PIPNR rate. For smaller constellations 256 states are enough to see convergence in the performance. At 1MHz LLW the TMM method becomes sub-optimal, however, still achieves reasonable performance for 16 QAM and 64 QAM. The pilot rate in this case is increased to $P = 0.05$. The gap to the trellis method is around 0.4 and 0.2 bits/channel use for 64 QAM and 16 QAM, respectively.

V. DISCUSSION AND FUTURE WORK

As shown earlier, the trellis method from [19] achieves near-PIPNR performance in nearly all cases of interest, however, it is very complex. This is particularly the case for large values of Δ, where the entire range $[-\pi;\pi)$ must be covered with very fine precision. Recently, the authors in [24] studied this problem, and proposed a low-complexity solution, which basically reduces the state space and thus the complexity of the algorithm. A comparison in terms of complexity and performance between the TMM and the solution from [24] would be of interest, but out of the scope of this paper.

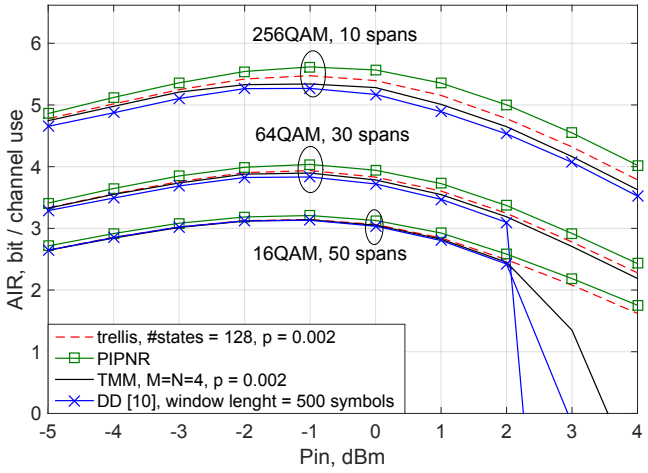

Fig. 11. AIRs with 10kHz LLW for different modulation formats at different distances. The DD method requires long averaging window. The TMM closely approaches the trellis method in all cases.

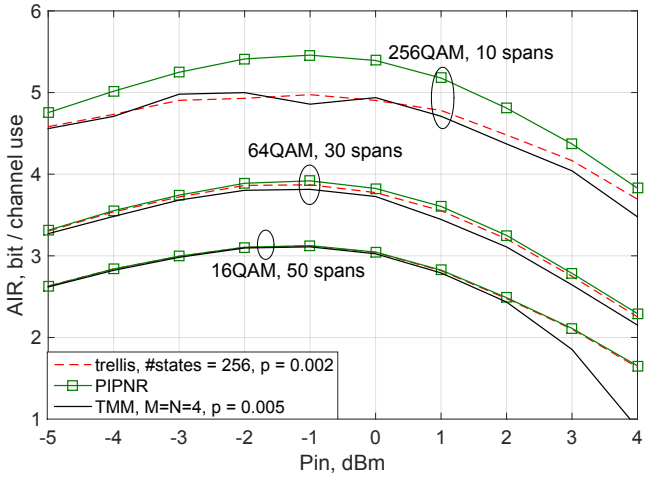

Fig. 12. AIRs with 100kHz LLW for different modulation formats at different distances. The DD method cannot be used. The trellis method requires more states than before. The TMM requires a slight increase in the pilot rate, but still achieves a performance close to that of the trellis method in all cases.

As mentioned in Section II-B, the choice of the window length L for estimating Δ^2 may improve the quality of the Wiener process model for the NLPN. We found that the AIRs in Section IV vary only slightly for $L \in [10;100]$, which means that the proposed method is generally robust to variations and instabilities in the estimation of Δ. We note that Eq. (6) may be seen as a convolution of the signal $y_k \cdot x_k^*$ with a rectangular pulse. Optimizing the pulse shape may further improve the quality of the model.

A remark on the pilot symbols assumptions in our design follows. In cases of very narrow linewidth, we have confirmed that if the initial $\pi/2$ ambiguity is avoided, the algorithm is stable in tracking the phase noise and does not require subsequent pilots. Even though the $P = 0.002$ is negligible in terms of reduced spectral efficiency, if the initial phase noise value is known, the pilots may be entirely removed.

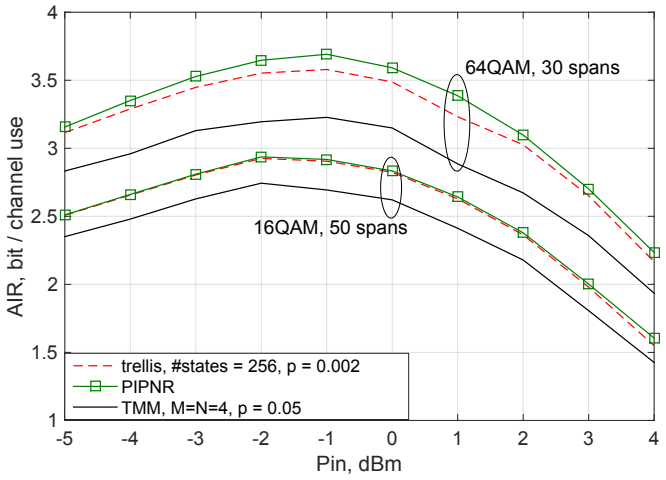

Fig. 13. AIRs with 1MHz LLW for different modulation formats at different distances. The DD method cannot be used. The TMM method requires an increased pilot rate, and achieves stable performance, with a loss compared to the trellis method of 0.4 and 0.2 bits/channel use for 64 QAM and 16 QAM, respectively.

This is a reasonable assumption when e.g. previous blocks are decoded correctly, and such information can be extracted. We note that due to the increasing order of modulation in optical fiber systems, inserting pilot symbols is becoming a more and more popular approach in the research community due to the improved equalization they provide. The same pilots that are used for equalization may generally be used for phase noise tracking.

VI. Conclusion

An algorithm was proposed for tracking the phase noise in wavelength division multiplexed optical fiber channels. It was shown that in ideal distributed raman amplification (IDRA) links, the proposed method can effectively combat the combined effect of laser phase noise and non-linear phase noise, outperforming previous decision directed methods, at significantly reduced complexity compared to previous trellis methods. Near optimal performance can be achieved for IDRA links, but also in more practical lumped amplification links with dual polarization input. Depending on the severity of the phase noise, pilot symbols may be introduced, which allow for stable performance on a wide variety of SNRs, achievable information rates and laser linewidths.

References

[1] R. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *Journal of Lightwave Technology*, vol. 28, no. 4, pp. 662-701, Feb. 15, 2010.

[2] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Properties of nonlinear noise in long, dispersion-uncompensated fiber links," *Optics Express*, vol. 21, no. 22, pp. 25685-25699, Nov. 4, 2013.

[3] R. Dar, M. Shtaif, and M. Feder, "New bounds on the capacity of the non-linear fiber-optic channel," *Optics Letters*, vol. 39, no. 2, pp. 398-401, Jan. 15, 2014.

[4] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Inter-channel nonlinear interference noise in WDM systems: modeling and mitigation," *Journal of Lightwave Technology*, vol. 33, no. 5, pp. 1044-1053, Mar. 1, 2015.

[5] M. Secondini and E. Forestieri, "On XPM mitigation in WDM fiber-optic systems," *Photonics Technology Letters*, vol. 26, no. 22, pp. 2252-2255, Nov. 15, 2014.

[6] D. Marsella, M. Secondini, E. Agrell, and E. Forestieri, "A simple strategy for mitigating XPM in nonlinear WDM optical systems," in Proceedings of Optical Fiber Communication Conference, Paper TH4D.3, 2015.

[7] M. Magarini, A. Spalvieri, F. Vacondio, M. Bertolini, M. Pepe, and G. Gavioli, "Empirical modeling and simulation of phase noise in long-haul coherent optical systems," *Optics Express*, vol. 19, no. 23, pp. 22455-22461, Nov. 7, 2011.

[8] I. Fatadin, D. Ives, and S. J. Savory, "Laser linewidth tolerance for 16-QAM coherent optical systems using QPSK partitioning," *Photonics Technology Letters*, vol. 22, no. 9, pp. 631-633, May. 1, 2010.

[9] L. Barletta, M. Magarini, and A. Spalvieri, "The information rate transferred through the discrete-time Wiener's phase noise channel," *Journal of Lightwave Technology*, vol. 30, no. 10, pp. 1480-1486, May 15, 2012.

[10] T. Pfau, S. Hoffman, and R. Noé, "Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for M-QAM constellations," *Journal of Lightwave Technology*, vol. 27, no. 8, pp. 989-999, Apr. 15, 2009.

[11] S. Zhang, P. Kam, C. Yu, and J. Chen, "Decision-aided carrier phase estimation for coherent optical communications," *Journal of Lightwave Technology*, vol. 28, no. 11, pp. 1597-1607, June 1, 2010.

[12] T. N. Huynh, A. T. Nguyen, W. Ng, L. Nguyen, L. A. Rusch, and L. P. Barry, "BER performance of coherent optical communications systems employing monolithic tunable lasers with excess phase noise," *Journal of Lightwave Technology*, vol. 32, no. 10, pp. 1973-1980, May 15, 2014.

[13] M. Magarini, L. Barletta, A. Spalvieri, F. Vacondio, T. Pfau, M. Pepe, M. Bertolini, and G. Gavioli, "Pilot-symbols-aided carrier-phase recovery for 100-G PM-QPSK digital coherent receivers," *Photonics Technology Letters*, vol. 24, no. 9, pp. 739-741, May 1, 2012.

[14] Y. Gao, A. P. T. Lau, and C. Lu, "Modulation-format-independent carrier phase estimation for square M-QAM systems," *Photonics Technology Letters*, vol. 25, no. 11, pp. 1073-1076, Jun. 1, 2013.

[15] C. Pan and F. R. Kschischang, "Coded-aided phase tracking for coherent fiber channels," *Journal of Lightwave Technology*, vol. 32, no. 6, pp. 1041-1047, Mar. 15, 2014.

[16] C. Pan, H. Bülow, W. Idler, L. Schmalen, and F. R. Kschischang, "Optical nonlinear-phase-noise compensation for 9x32 Gbaud polDM-16 QAM transmission using a code-aided expectation-maximization algorithm," *Journal of Lightwave Technology*, vol. 33, no. 17, pp. 3679-3686, Sep. 1, 2014.

[17] L. Barletta, M. Magarini, and A. Spalvieri, "A new lower bound below the information rate of Wiener phase noise channel based on Kalman carrier recovery," *Optics Express*, vol. 20, no. 23, pp. 25471-25477, Nov. 5, 2012.

[18] L. Barletta, M. Magarini, and A. Spalvieri, "Tight upper and lower bounds to the information rate of the phase noise channel," IEEE International Symposium on Information Theory (ISIT), pp. 2284-2288, July 7-12, 2013.

[19] L. Barletta, F. Bergamelli, M. Magarini, N. Carapellese, and A. Spalvieri, "Pilot-aided trellis-based demodulation," *Photonics Technology Letters*, vol. 25, no. 13, pp. 1234-1237, July 1, 2013.

[20] A. Barbieri and G. Colavolpe. "Soft-output decoding of rotationally invariant codes over channels with phase noise," *IEEE Transactions on Communications*, vol. 55, no. 11, pp. 2125-2133, Nov. 2007.

[21] D. M. Arnold, H. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Transactions on Information Theory*, vol. 52, no. 8, Aug. 2006.

[22] T. Fehenberger, M. P .Yankov, L. Barletta, and N. Hanik, "Compensation of XPM interference by blind tracking of the nonlinear phase in WDM systems with QAM input," Proceedings of European Conference on Optical Communications (ECOC), Paper P.5.8, Oct. 2015.

[23] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. AU-15, no. 2, June 1967.

[24] S. Pecorino, S. Mandelli, L. Barletta, M. Magarini, and A. Spalvieri, "Bootstrapping iterative demodulation and decoding without pilot symbols," *Journal of Lightwave Technology*, vol. 33, no. 17, pp. 3616-3622, Sep. 1, 2015.