

Copyright Notice

©2010 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

This document was downloaded from Chalmers Publication Library (<http://publications.lib.chalmers.se/>), where it is available in accordance with the IEEE PSPB Operations Manual, amended 19 Nov. 2010, Sec. 8.1.9 (<http://www.ieee.org/documents/opsmanual.pdf>)

(Article begins on next page)

An ML-based Detector for Optical Communication in the Presence of Nonlinear Phase Noise

A. Serdar Tan⁽¹⁾, Henk Wymeersch⁽¹⁾, Pontus Johannisson⁽²⁾, Erik Agrell⁽¹⁾, Peter Andrekson⁽²⁾, Magnus Karlsson⁽²⁾

⁽¹⁾Department of Signals and Systems, Chalmers University of Technology, SE-412 96 Göteborg, Sweden

⁽²⁾Department of Microtechnology and Nanoscience, Chalmers University of Technology, SE-412 96 Göteborg, Sweden
Email: ahmets@chalmers.se, henkw@chalmers.se

Abstract—We present a closed-form maximum likelihood-based data detection algorithm for long-haul optical channels with dominant nonlinear phase noise induced by self-phase modulation. The closed-form detector is evaluated in terms of symbol error rate as a function of input power, and compared with other sub-optimal detectors as well as a non-parametric detector. We show that the performance of the detector deteriorates for high input power levels, resulting in an optimal operation region. We also provide insights into the behavior of the detector in the highly nonlinear regime.

I. INTRODUCTION

Coherent optical communication systems employing polarization multiplexed multi-level quadrature amplitude modulation (M-QAM) offers a significant increase in data rate compared to traditional binary signaling, such as on-off keying. However, M-QAM systems are more sensitive to channel impairments, due to the encoding of data into the phase of the optical signal and the more densely packed constellation. One of the major limitations is the Kerr nonlinearity, which distorts the propagating signal and, for a given system, gives rise to a finite optimal input power. The impairments from the Kerr nonlinearity include both deterministic and stochastic effects [1]. For example, intrachannel four-wave mixing leads to nonlinear inter-symbol interference (ISI), but in principle this effect can be compensated for, as shown in [2]. On the contrary, the self-phase modulation (SPM)–noise interaction is interesting to study since it is a fundamental limitation that only can be partially compensated for using statistical methods.

The interaction of the signal and optical amplifier noise via the Kerr nonlinearity leads to nonlinear phase noise (NLPN) [3], which in signal space can be viewed as symbols elongated in the phase direction. An experimental investigation has showed that NLPN is a limiting factor [4] and recent results indicate that NLPN is important up to 40 Gbaud [5]. A mitigation technique for this effect was studied in [6] under the assumption of distributed amplification (see [7]) and constant amplitude binary modulation.

Maximum likelihood (ML) detection, which requires the computation of the probability density function of the received signal, is used extensively in linear wireless communications. The ML-based algorithms developed for linear channels will cause catastrophic performance degradations when used for

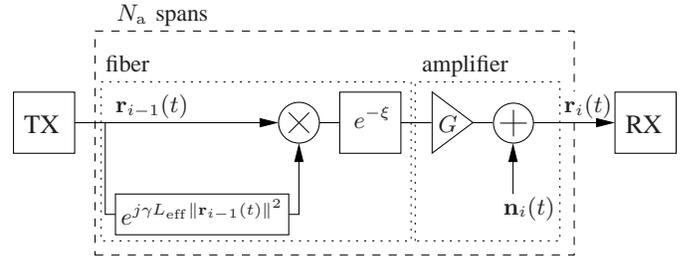


Fig. 1. Optical transmission model: the transmitted signal passes through N_a spans of fiber. Each span induces a non-linear rotation and attenuation. Every span has an amplifier that compensates for the attenuation and causes ASE noise.

nonlinear channels. A study on the performance degradation is presented in [8] for clock recovery in the presence of NLPN. In [9], ML decision boundaries are derived for an optical channel with nonlinear phase noise, only considering distributed amplification. In addition, mitigation for 16-QAM is inherently sub-optimal as it comprises two stages: ring discrimination followed by phase post-compensation. Very little work has been performed on ML detection in the presence of discrete amplifiers.

In this paper, we derive a closed-form ML-based data detector for a polarization-multiplexed M-QAM system with discrete amplification, limited by NLPN. The long-haul link is assumed to use optical dispersion compensation in order to be able to neglect the effects of dispersion and nonlinear ISI. The closed-form detector is compared with a number of sub-optimal detectors that account for varying degrees of nonlinearity, as well as a complex non-parametric detector. Although the latter performs better than the proposed closed-form detector, the closed-form detector provides the best complexity-performance trade-off.

II. SYSTEM MODEL

We consider a 16-QAM dual polarization coherent optical communication system as given in Figure 1. The transmission system in long-haul optical links consists of multiple amplifier stages to compensate for the dispersion and signal attenuation. Each amplifier stage consists of a single mode fiber (SMF) followed by a dispersion compensating fiber (DCF) with Kerr nonlinear parameters γ_{SMF} and γ_{DCF} , respectively. The DCF is assumed to ideally compensate for the chromatic

dispersion. In each fiber span, the signal power is attenuated by $e^{-\xi} = e^{-(\alpha_{\text{SMF}}L_{\text{SMF}} + \alpha_{\text{DCF}}L_{\text{DCF}})}$ where α_{SMF} and α_{DCF} are attenuation factors and L_{SMF} and L_{DCF} are the lengths of SMF and DCF, respectively. The amplifiers restore the signal power to the levels in the transmitter side by a power gain $G = e^{\xi}$. In addition, each amplification adds amplified spontaneous emission (ASE) noise, which is modeled as an additive white Gaussian noise process. The complex baseband signal fed to the optical link by the transmitter (TX) is given by

$$\mathbf{s}(t) = \sqrt{P_{\text{in}}} \sum_{n=-\infty}^{+\infty} \mathbf{a}_n p(t - nT), \quad (1)$$

where $\mathbf{a}_n = [a_n^{(X)}, a_n^{(Y)}]^T \in \Omega^2$ is the n -th data vector from a normalized constellation Ω with $\mathbb{E}[\mathbf{a}_n \mathbf{a}_n^H] = \mathbf{I}$, T is the symbol duration, P_{in} is the launch power, and $p(t)$ is a pulse with peak amplitude 1 at $t = 0$. The optical signal after the i -th amplifier is given by

$$\mathbf{r}_i(t) = \mathbf{r}_{i-1}(t) \exp(j\gamma L_{\text{eff}} \|\mathbf{r}_{i-1}(t)\|^2) + \mathbf{n}_i(t), \quad (2)$$

where $\mathbf{r}_0(t) = \mathbf{s}(t)$, $\gamma = \gamma_{\text{SMF}}$ is the nonlinearity parameter of the fiber¹, $L_{\text{eff}} = (1 - e^{-\alpha L}) / \alpha$ is the effective length of the fiber for attenuation factor $\alpha = \alpha_{\text{SMF}}$ and length $L = L_{\text{SMF}}$, the operator $\|\cdot\|$ represents the norm given by $\|\mathbf{x}\| := \sqrt{\mathbf{x}^H \mathbf{x}}$, and $\mathbf{n}_i(t)$ is ASE noise, which is zero-mean and has power spectral density per polarization $N_{0,\text{ASE}} = h\nu n_{\text{sp}}(G - 1)$ in which h is Planck's constant, ν is the optical frequency, n_{sp} is the spontaneous emission factor, and G is the amplifier gain. Prior to launching into the next span, the noise is bandlimited to a bandwidth B .

In order to simplify the system model and focus on the effect of SPM, we neglect the effects of polarization mode dispersion and local oscillator phase noise and assume that perfect timing and frequency synchronization is achieved. This implies that the performance we achieve can be interpreted as a lower bound on the symbol error rate (SER) of a practical system.

III. DATA DETECTION

A. Discrete-time Observation

The received signal after N_a consecutive amplifier stages is converted to the electrical domain, filtered at the bandwidth B and sampled at the symbol rate $1/T$. The filter is assumed to be flat within the signal bandwidth and to be a square root Nyquist filter for rate $1/T$. This gives rise to equivalent samples $\mathbf{r}_{i,k} = \mathbf{r}_i(kT)$, where

$$\mathbf{r}_{i,k} = \mathbf{r}_{i-1,k} \exp(j\gamma L_{\text{eff}} \|\mathbf{r}_{i-1,k}\|^2) + \mathbf{n}_{i,k}. \quad (3)$$

Substituting $\mathbf{r}_{i-1,k}$ recursively, we find that $\mathbf{r}_k \triangleq \mathbf{r}_{N_a,k}$ can be expressed as

$$\mathbf{r}_k = \mathbf{s}_k \exp\left(j\gamma L_{\text{eff}} \sum_{i=0}^{N_a-1} \|\mathbf{r}_{i,k}\|^2\right) + \mathbf{w}_k, \quad (4)$$

where $\mathbf{s}_k = \sqrt{P_{\text{in}}} \mathbf{a}_k$ (assuming negligible inter-symbol interference), $\mathbf{w}_k \sim \mathcal{CN}(\mathbf{0}, N_a \sigma_{\text{ASE}}^2 \mathbf{I})$, and $\sigma_{\text{ASE}}^2 = BN_{0,\text{ASE}}$. Further substitution of (3) into (4) yields

$$\mathbf{r}_k = \mathbf{s}_k \exp\left(j\gamma L_{\text{eff}} N_a \|\mathbf{s}_k\|^2 + j\psi_k + j\phi_k\right) + \mathbf{w}_k, \quad (5)$$

where, conditioned on \mathbf{s}_k , ψ_k is a zero-mean Gaussian random variable with variance

$$\sigma_{\psi}^2 = 2\gamma^2 L_{\text{eff}}^2 \|\mathbf{s}_k\|^2 \sigma_{\text{ASE}}^2 (N_a - 1) N_a (2N_a - 1) / 6, \quad (6)$$

and ϕ_k is a random variable with mean $\bar{\phi} = \gamma L_{\text{eff}} \sigma_{\text{ASE}}^2 (N_a - 1) N_a / 2$, but is independent of \mathbf{s}_k . The derivation of steps (3) to (5) and exact expressions for \mathbf{w}_k , ψ_k and ϕ_k can be found in Appendix A.

B. Approximate Likelihood Function

In the observation model given by (5), both ψ_k and ϕ_k are correlated with \mathbf{w}_k . Due to the correlation, deriving the likelihood function $p(\mathbf{r}_k | \mathbf{a}_k)$ is difficult. Eventhough simulations show that ϕ_k is negligible, we approximate $\phi_k \approx \bar{\phi}$. We further assume that ψ_k and \mathbf{w}_k are *independent*, in order to derive a tractable, near-optimal detector. The likelihood function is then given by

$$p(\mathbf{r}_k | \mathbf{a}_k) = \int_{-\infty}^{+\infty} p(\mathbf{r}_k, \psi_k | \mathbf{a}_k) d\psi_k \quad (7)$$

$$= \int_{-\infty}^{+\infty} p(\mathbf{r}_k | \mathbf{a}_k, \psi_k) p(\psi_k | \mathbf{a}_k) d\psi_k. \quad (8)$$

Here, $p(\mathbf{r}_k | \mathbf{a}_k, \psi_k)$ and $p(\psi_k | \mathbf{a}_k)$ are both Gaussian probability density functions given by

$$p(\mathbf{r}_k | \mathbf{a}_k, \psi_k) \quad (9)$$

$$\propto \exp\left(-\frac{1}{N_a \sigma_{\text{ASE}}^2} \left\| \mathbf{r}_k - \mathbf{s}_k e^{(j\gamma L_{\text{eff}} N_a \|\mathbf{s}_k\|^2 + j\psi_k + j\bar{\phi})} \right\|^2\right)$$

$$\propto \exp\left(-\frac{\|\mathbf{s}_k\|^2}{N_a \sigma_{\text{ASE}}^2}\right)$$

$$\times \exp\left(\frac{2\text{Re}\left\{\mathbf{r}_k^H \mathbf{s}_k e^{(j\gamma L_{\text{eff}} N_a \|\mathbf{s}_k\|^2 + j\bar{\phi})} e^{j\psi_k}\right\}}{N_a \sigma_{\text{ASE}}^2}\right), \quad (10)$$

and

$$p(\psi_k | \mathbf{a}_k) = \frac{1}{\sqrt{2\pi\sigma_{\psi}^2}} \exp\left(-\frac{\psi_k^2}{2\sigma_{\psi}^2}\right). \quad (11)$$

When σ_{ψ}^2 is sufficiently small, the integration over ψ_k in (8) can be carried out approximately. The complete derivation can be found in Appendix B, and leads to the following likelihood function:

$$p(\mathbf{r}_k | \mathbf{a}_k) \propto \exp\left(-\frac{\|\mathbf{s}_k\|^2}{N_a \sigma_{\text{ASE}}^2}\right) \frac{I_0(|\beta_k|)}{I_0(1/\sigma_{\psi}^2)}, \quad (12)$$

where $I_0(\cdot)$ is the zeroth order modified Bessel function of the first kind, and

$$\beta_k = \frac{2\mathbf{r}_k^H \mathbf{s}_k e^{(j\gamma L_{\text{eff}} N_a \|\mathbf{s}_k\|^2 + j\bar{\phi})}}{N_a \sigma_{\text{ASE}}^2} + \frac{1}{\sigma_{\psi}^2}. \quad (13)$$

¹Since the input power to the DCF is low, SPM is neglected in the DCF.

C. Closed-Form Data Detectors

The approximate likelihood function in (12) can be used as a data detector, which is given by

$$\hat{\mathbf{a}}_k = \arg \max_{\mathbf{a}_k \in \Omega^2} p(\mathbf{r}_k | \mathbf{a}_k) \quad (14)$$

$$= \arg \max_{\mathbf{a}_k \in \Omega^2} \exp\left(-\frac{\|\mathbf{s}_k\|^2}{N_a \sigma_{\text{ASE}}^2}\right) \frac{I_0(|\beta_k|)}{I_0(1/\sigma_\psi^2)}. \quad (15)$$

In the rest of the work, the detector in (15) will be called the closed-form detector.

We will also consider two competing detectors: if we ignore the nonlinearity completely, we have the observation model $\mathbf{r}_k = \mathbf{s}_k + \mathbf{n}_k$, leading to the regular ML (RML) detector:

$$\hat{\mathbf{a}}_k = \arg \min_{\mathbf{a}_k \in \Omega^2} \|\mathbf{r}_k - \mathbf{s}_k\|^2. \quad (16)$$

Note that in a linear channel, where $\gamma = 0$, the detector in (15) reverts to the RML detector. The second competing detector only considers the deterministic part of the nonlinearity, assuming $\mathbf{r}_k = \mathbf{s}_k \exp(j\gamma L_{\text{eff}} N_a \|\mathbf{s}_k\|^2) + \mathbf{n}_k$, leading to the regular ML detector with post-compensation (RML-PC):

$$\hat{\mathbf{a}}_k = \arg \min_{\mathbf{a}_k \in \Omega^2} \left\| \mathbf{r}_k - \mathbf{s}_k e^{j\gamma L_{\text{eff}} N_a \|\mathbf{s}_k\|^2} \right\|^2. \quad (17)$$

The complexity of the detectors scales as $\mathcal{O}(|\Omega|^2)$, where $|\Omega|$ denotes the number of elements in the constellation.

D. Non-Parametric Detector

As we cannot develop an exact ML detector, there is no baseline performance indicator. To gain insight into the approximations we made in developing the closed-form detector, we here introduce a non-parametric detector [10], which first learns the distributions $p(\mathbf{r}_k | \mathbf{a}_k)$, for every possible value of \mathbf{a}_k , as a function of \mathbf{r}_k . Once these conditional distributions are known, we can evaluate $p(\mathbf{r} | \mathbf{a})$ for a new, previously unseen observation \mathbf{r} , for every possible value of $\mathbf{a} \in \Omega^2$.

1) *Training*: To train the detector we fix the transmitted symbol vector to $\mathbf{a} \in \Omega^2$ and generate a large number of observations $\mathbf{r}_1, \dots, \mathbf{r}_N$. We then approximate

$$p(\mathbf{r} | \mathbf{a}) \approx \frac{1}{N} \sum_{i=1}^N K_\sigma(\mathbf{r} - \mathbf{r}_i), \quad (18)$$

where $K(\cdot)$ is a so-called kernel, a symmetric distribution with unit variance. We choose a two-dimensional complex Gaussian:

$$K_\sigma(\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^2} \exp\left(-\frac{\|\mathbf{x}\|^2}{2\sigma^2}\right). \quad (19)$$

The parameter σ in (18) is called the bandwidth and is selected according to the method described in [11]. This procedure is repeated for every possible value of $\mathbf{a} \in \Omega^2$. The complexity of the training procedure scales as $\mathcal{O}(N|\Omega|^2)$.

TABLE I
SYSTEM PARAMETER VALUES

γ_{SMF}	1.2 W ⁻¹ km ⁻¹	γ_{DCF}	5.2 W ⁻¹ km ⁻¹
α_{SMF}	0.20 dB/km	α_{DCF}	0.60 dB/km
L_{SMF}	80 km	L_{DCF}	11 km
B	14 GHz	$\lambda = c/\nu$	1.55 μm
n_{sp}	1.5	N_a	22

2) *Testing*: Once the detector is trained, we can commence practical data detection. For every new observation \mathbf{r}_k , we evaluate $p(\mathbf{r}_k | \mathbf{a})$, for every $\mathbf{a} \in \Omega^2$, and find the most likely value for \mathbf{a} . The complexity of the testing procedure scales as $\mathcal{O}(N|\Omega|^2)$ per observation. This makes the application of the non-parametric detector unsuitable for online processing.

IV. NUMERICAL RESULTS

A. Simulation Set-up

We present numerical results for 16-QAM with the proposed closed-form detector, and compare with the three other detectors (RML, RML-PC, and the non-parametric detector). The number of training samples for the non-parametric detector is set to $N = 2 \times 10^5$. The system we consider operates at 14 Gbaud per polarization, corresponding to an overall data rate of 112 Gbit/s, which is compatible with emerging standards for 100 Gigabit Ethernet. It has 22 spans, every span consisting of 80 km of single-mode fiber and 11 km of dispersion-compensating fiber. The values of all system parameters are given in Table I.

B. Symbol Error Rate Performance

We first present results for a single polarization in Figure 2 and then extend to dual polarization in Figure 3. From Figure 2, we see that the RML detector has the worst performance, as expected since it does not consider the phase shift due to SPM. The RML-PC detector performance is significantly better compared to the RML detector owing to its simple post-compensation. However, its performance degrades as the input power increases, because of the assumption $\psi_k = \phi_k = 0$. The closed-form detector and non-parametric detector yield significantly lower SER values compared to the other detectors. The closed-form and non-parametric detectors yield their best performance around 4.7 dBm and 5.7 dBm input power, respectively. The difference is expected due to the assumptions made in the derivation of the closed-form detector.

For a dual polarization 16-QAM system, Figure 3 indicates that the performance of the RML and RML-PC detectors degrade considerably, whereas that of the closed-form and non-parametric detector degrade slightly compared to the single polarization 16-QAM system. The degradation is caused by the increase in nonlinear phase rotation which is proportional to the total power in *both* polarizations. In case of dual polarization, the optimal operation region is around 4.2 dBm and 5.2 dBm for the closed-form and non-parametric detectors, respectively.

Overall, compared to RML and RML-PC, the closed-form detector is less sensitive to small changes in the input power

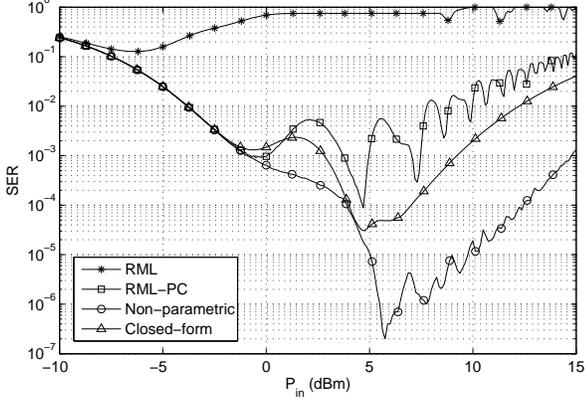


Fig. 2. SER as a function of P_{in} for single polarization 16-QAM.

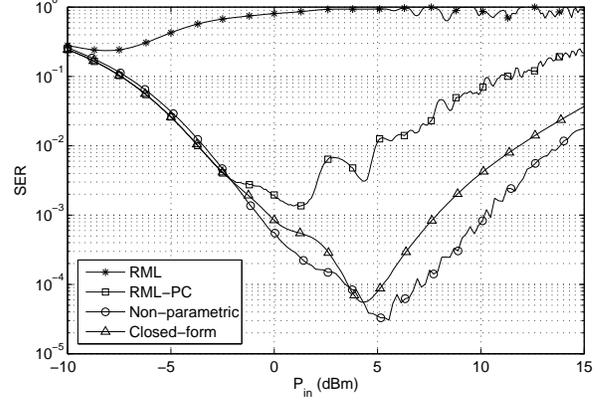


Fig. 3. SER as a function of P_{in} for dual polarization 16-QAM.

and yields better performance. The non-parametric detector performs better than the others, but the complexity is significantly higher.

C. Discussion

In both Figure 2 and Figure 3, we observe an oscillatory behavior of the SER performance. This effect is also apparent in the results from [9], but no explanation was given. This oscillatory behavior is caused by the signal power dependent nonlinear phase rotation in the received constellation. The signal power dependent phase rotation causes the signal points on a ring with greater radius to rotate faster than the other signal points. In Figure 4, the normalized received signal constellations of a single polarization 16-QAM system are given for different input power levels. The 16-QAM constellations consist of three rings. The inner ring, comprising 4 symbols, is least affected by SPM, so that increasing the input power has no significant detrimental effect. The middle and outer ring, comprising 8 and 4 symbols, respectively, are affected by SPM, due to the dependence of σ_ψ^2 on $\|\mathbf{a}_k\|$. Two effects occur: overlap of clusters within a given ring and overlap of clusters from the outer with the middle ring. The latter effect is visible at 4.00 dBm and 6.30 dBm (top left and bottom left in Figure 4). The former effect is more pronounced at 7.10 dBm (bottom right in Figure 4). The interaction of these two effects causes the oscillations in SER.

V. CONCLUSION

We have presented a closed-form maximum likelihood-based data detection algorithm for long-haul optical channels with dominant nonlinear phase noise induced by self-phase modulation. The detector is suitable for any memoryless modulation format. We have evaluated the closed-form detector in terms of symbol error rate as a function of input power, and compared with other sub-optimal detectors as well as a non-parametric detector. From our simulations, we have observed that the proposed detector yields the best complexity-performance trade-off. We have also provided a qualitative explanation of the behavior of the detector in the highly nonlinear regime.

APPENDIX A COMPLETE DISCRETE-TIME OBSERVATION

We explain the transition from (3) to (5) in detail. In the first step, substituting $\mathbf{r}_{i-1,k}$ recursively in (3), and noting $\mathbf{r}_k \triangleq \mathbf{r}_{N_a,k}$, we find that

$$\mathbf{r}_k = \mathbf{r}_{N_a-1,k} \exp(j\gamma L_{\text{eff}} \|\mathbf{r}_{N_a-1,k}\|^2) + \mathbf{n}_{N_a,k} \quad (20)$$

$$= \mathbf{r}_{N_a-2,k} \exp(j\gamma L_{\text{eff}} (\|\mathbf{r}_{N_a-2,k}\|^2 + \|\mathbf{r}_{N_a-1,k}\|^2)) + \mathbf{n}_{N_a-1,k} \exp(j\gamma L_{\text{eff}} \|\mathbf{r}_{N_a-1,k}\|^2) + \mathbf{n}_{N_a,k} \quad (21)$$

$$= \mathbf{r}_{0,k} \exp(j\gamma L_{\text{eff}} \sum_{i=0}^{N_a-1} \|\mathbf{r}_{i,k}\|^2) + \sum_{i=1}^{N_a} \mathbf{n}_{i,k} \exp(j\gamma L_{\text{eff}} \sum_{l=i}^{N_a-1} \|\mathbf{r}_{l,k}\|^2) \quad (22)$$

$$= \mathbf{s}_k \exp(j\gamma L_{\text{eff}} \sum_{i=0}^{N_a-1} \|\mathbf{r}_{i,k}\|^2) + \mathbf{w}_k, \quad (23)$$

where $\mathbf{r}_{0,k} \triangleq \mathbf{s}_k$ and \mathbf{w}_k is the second term in (22). In the next step, further recursive substitution of $\mathbf{r}_{i,k}$ yields

$$\begin{aligned} \mathbf{r}_k &= \mathbf{s}_k \exp\left(j\gamma L_{\text{eff}} \left(N_a \|\mathbf{s}_k\|^2\right.\right. \\ &+ \sum_{i=1}^{N_a} (N_a - i) 2\text{Re} \left\{ \mathbf{s}_k^H e^{-j\gamma L_{\text{eff}} \sum_{l=0}^{i-1} \|\mathbf{r}_{l,k}\|^2} \mathbf{n}_{i,k} \right\} \\ &+ \sum_{i=2}^{N_a-1} 2\text{Re} \left\{ \mathbf{n}_{i,k}^H \sum_{l=1}^{i-1} e^{-j\gamma L_{\text{eff}} \sum_{m=l}^{i-1} \|\mathbf{r}_{m,k}\|^2} \mathbf{n}_{l,k} \right\} \\ &\left.\left. + \sum_{i=1}^{N_a-1} (N_a - i) \|\mathbf{n}_{i,k}\|^2\right)\right) \quad (24) \end{aligned}$$

$$= \mathbf{s}_k \exp\left(j\gamma L_{\text{eff}} N_a \|\mathbf{s}_k\|^2 + j\psi_k + j\phi_k\right) + \mathbf{w}_k, \quad (25)$$

where ψ_k represents the second term and ϕ_k represents the sum of third and last terms inside the exponential in (24).

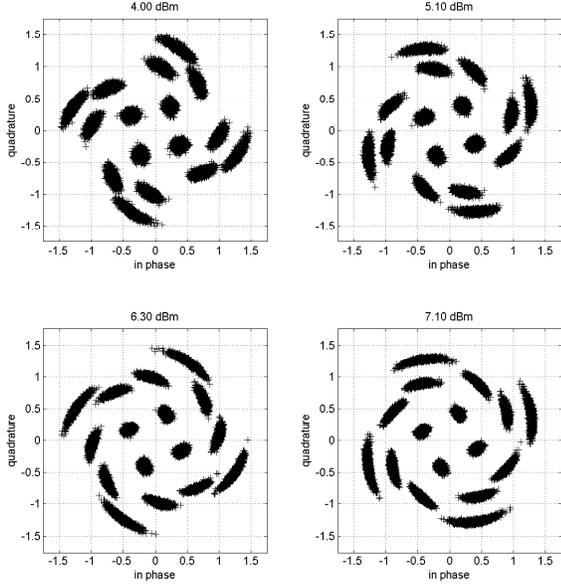


Fig. 4. Received 16-QAM constellation for different input power levels: 4 dBm (left top), 5.1 dBm (right top), 6.3 dB (left bottom), 7.1 dBm (right bottom).

Then, it is straightforward to show that

$$\begin{aligned} \mathbf{w}_k &\sim \mathcal{CN}(\mathbf{0}, N_a \sigma_{\text{ASE}}^2 \mathbf{I}), \\ \psi_k &\sim \mathcal{N}(0, 2\gamma^2 L_{\text{eff}}^2 \|\mathbf{s}_k\|^2 \sigma_{\text{ASE}}^2 (N_a - 1) N_a (2N_a - 1)/6), \\ \text{and} \\ \bar{\phi} &= \mathbb{E}[\phi_k] = \gamma L_{\text{eff}} \sigma_{\text{ASE}}^2 (N_a - 1) N_a / 2. \end{aligned}$$

APPENDIX B DERIVATION OF THE LIKELIHOOD FUNCTION

We derive the approximate likelihood function $p(\mathbf{r}_k | \mathbf{a}_k)$. Let $g(\mu, \sigma^2; x)$ denote a Gaussian density function in x with mean μ and variance σ^2 in x . The following property relates a Gaussian distribution to a von Mises distribution [12].

Property I: For $\sigma^2 \ll 1$ and x restricted to the interval $[-\pi, \pi]$

$$g(0, \sigma^2; x) \approx \frac{1}{2\pi I_0(1/\sigma^2)} \exp(\cos(x)/\sigma^2), \quad (26)$$

where $I_0(\cdot)$ is the zeroth order modified Bessel function of the first kind.

We now return to our problem of determining $p(\mathbf{r}_k | \mathbf{a}_k)$. Substituting (10) into (8) yields

$$\begin{aligned} p(\mathbf{r}_k | \mathbf{a}_k) & \\ \propto \exp\left(-\frac{\|\mathbf{s}_k\|^2}{N_a \sigma_{\text{ASE}}^2}\right) & \int_{-\infty}^{+\infty} \exp(\text{Re}\{z_k e^{j\psi_k}\}) g(0, \sigma_{\psi}^2; \psi_k) d\psi_k, \end{aligned} \quad (27)$$

where $z_k = 2\mathbf{r}_k^H \mathbf{s}_k e^{(j\gamma L_{\text{eff}} N_a \|\mathbf{s}_k\|^2 + j\bar{\phi})} / (N_a \sigma_{\text{ASE}}^2)$. We now

use Property I and find that

$$p(\mathbf{r}_k | \mathbf{a}_k) \quad (28)$$

$$\propto \exp\left(-\frac{\|\mathbf{s}_k\|^2}{N_a \sigma_{\text{ASE}}^2}\right) \frac{1}{2\pi I_0(1/\sigma_{\psi}^2)} \quad (29)$$

$$\times \int_{-\pi}^{\pi} \exp(\text{Re}\{z_k e^{j\psi_k}\}) \exp(\cos(\psi_k)/\sigma_{\psi}^2) d\psi_k,$$

$$= \exp\left(-\frac{\|\mathbf{s}_k\|^2}{N_a \sigma_{\text{ASE}}^2}\right) \frac{1}{2\pi I_0(1/\sigma_{\psi}^2)} \quad (30)$$

$$\times \int_{-\pi}^{\pi} \exp(\text{Re}\{\beta_k e^{j\psi_k}\}) d\psi_k,$$

$$= \exp\left(-\frac{\|\mathbf{s}_k\|^2}{N_a \sigma_{\text{ASE}}^2}\right) \frac{I_0(|\beta_k|)}{I_0(1/\sigma_{\psi}^2)}, \quad (31)$$

where $\beta_k = z_k + 1/\sigma_{\psi}^2$. We remind that σ_{ψ}^2 is itself a function of \mathbf{a}_k .

ACKNOWLEDGMENT

We would like to acknowledge funding from VINNOVA within the IKT grant 2007-02930.

REFERENCES

- [1] P. J. Winzer and R.-J. Essiambre, "Advanced optical modulation formats," *Proceedings of the IEEE*, vol. 94, no. 5, pp. 952–985, 2006.
- [2] E. Ip and J. M. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *Journal of Lightwave Technology*, vol. 26, no. 20, pp. 3416–3425, 2008.
- [3] J. P. Gordon and L. F. Mollenauer, "Phase noise in photonic communications systems using linear amplifiers," *Optics Letters*, vol. 15, no. 23, pp. 1351–1353, 1990.
- [4] H. Kim and A. H. Gnauck, "Experimental investigation of the performance limitation of DPSK systems due to nonlinear phase noise," *IEEE Photonics Technology Letters*, vol. 15, no. 320–322, 2003.
- [5] L. D. Coelho, O. Gaete, E. D. Schmidt, B. Spinnler, and N. Hanik, "Impact of PMD and nonlinear phase noise on the global optimization of DPSK and DQPSK systems," in *Optical Fiber Communication Conference (OFC)*, 2010, p. OWE5.
- [6] K. P. Ho and J. M. Kahn, "Electronic compensation technique to mitigate nonlinear phase noise," *Journal of Lightwave Technology*, vol. 22, no. 3, pp. 779–783, 2004.
- [7] L. F. Mollenauer, "Distributed amplification for lightwave transmission system," Oct. 22 1991, U.S. Patent 5,058,974.
- [8] A. S. Tan, H. Wymeersch, P. Johannisson, M. Sjödin, E. Agrell, P. A. Andrekson, and M. Karlsson, "The impact of self-phase modulation on digital clock recovery in coherent optical communication," in *European Conference on Optical Communication (ECOC)*, 2010, p. We.7.A.7.
- [9] A. P. Lau and J. M. Kahn, "Signal design and detection in presence of nonlinear phase noise," *Journal of Lightwave Technology*, vol. 25, no. 10, pp. 3008–3016, 2007.
- [10] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer-Verlag, 2001.
- [11] Z. Botev, "Nonparametric density estimation via diffusion mixing," The University of Queensland, Postgraduate Series, Nov. 2007, <http://espace.library.uq.edu.au/view/UQ:120006>.
- [12] D. J. Best and N. I. Fisher, "Efficient simulation of the von Mises distribution," *Applied Statistics*, vol. 28, no. 2, pp. 152–157, 1979.