(Article begins on next page)

Stochastic Backpropagation for Coherent Optical Communications

Nan Jiang[*], **Yan Gong**[*], **Johnny Karout**[*], **Henk Wymeersch**[*], **Pontus Johannisson**[†], **Magnus Karlsson**[†], **Erik Agrell**[*], **Peter A. Andrekson**[†]

[*]*Communication Systems Group, Department of Signals and Systems,*
[†]*Photonics Laboratory, Department of Microtechnology and Nanocience,*
Chalmers University of Technology, SE-412 96 Gothenburg, Sweden

henkw@chalmers.se

Abstract: We present stochastic backpropagation, a novel maximum a posteriori detection method for coherent optical communications. The proposed detector is shown to outperform conventional backpropagation in a scenario where nonlinear phase noise is the dominant impairment.

OCIS codes: 060.0060 (Fiber optics and optical communications), 060.1660 (Coherent communications).

1. Introduction

Multilevel quadrature amplitude modulation formats (M-QAM) have the ability to increase the spectral efficiency of coherent fiber-optical systems. However, M-QAM requires higher input powers to maintain a fixed bit-error-rate (BER). Higher input powers lead to increased intrachannel four-wave mixing and self-phase modulation (SPM), due to the Kerr effect [1]. In combination with amplified spontaneous emission (ASE) noise, SPM leads to nonlinear phase noise (NLPN) [2]. Different methods have been considered to mitigate the effect of NLPN at the receiver. Optimal compensation for binary phase-shift keying (BPSK) and differential quadrature phase-shift keying (DQPSK) were proposed in [3], and later extended to M-ary phase-shift keying in [4]. In [5], an adaptive maximum a posteriori (MAP) detection scheme with a look-up table was investigated for long-haul transmission, which can mitigate data-pattern dependent nonlinear impairments. A closed-form suboptimal detector for M-QAM was derived in [6]. In addition to these methods, which are all based on the stochastic nature of the NLPN, [7] proposed a deterministic method to mitigate the fiber nonlinearities and dispersion jointly using digital backpropagation. In this work, we propose a novel MAP detector based on the Bayesian framework of factor graphs. The detector retains the flavor of backpropagation and simultaneously accounts for stochastic impairments. Our detector can be applied to any constellation and exhibits superior performance compared to existing techniques, in a scenario where NLPN is the dominant impairment.

2. System Model

We consider a discrete-time multi-span polarization multiplexed coherent optical communication system at a moderate symbol rate using optical dispersion compensation. For clarity, we neglect chromatic dispersion so that the dominant impairment is NLPN [3, 4, 6]. In principle, the proposed framework can be extended to account for dispersion. At every symbol period, a two-dimensional complex data vector $\mathbf{a}$, drawn uniformly from a constellation Ω^2 with average energy proportional to the input power P_{in} per polarization, is transmitted. Each optical fiber span consists of single-mode fiber (SMF), dispersion-compensating fiber, and an amplifier. The discrete-time signal at the output of the first fiber span is written as $\mathbf{r}_1 = \mathbf{a}\exp(j\gamma L_{\text{eff}}\|\mathbf{a}\|^2) + \mathbf{n}_1$, where γ is the nonlinearity parameter, L_{eff} is the effective length of the SMF, and $\mathbf{n}_1$ is the ASE noise, modeled as a zero-mean complex Gaussian random variable with covariance matrix $N_0\mathbf{I}_2$, in which $\mathbf{I}_2$ is the 2×2 identity matrix. Similarly, the signal after the i-th span is written as $\mathbf{r}_i = \mathbf{x}_i + \mathbf{n}_i$, where we have introduced $\mathbf{x}_i = \mathbf{r}_{i-1}\exp\left(j\gamma L_{\text{eff}}\|\mathbf{r}_{i-1}\|^2\right)$. The model is depicted in the left side of Fig. 1, for $N = 2$. Our objective is to recover $\mathbf{a}$ given $\mathbf{r}_N$, where N is the number of fiber spans. The optimal decision rule (in terms of minimizing the symbol error rate) is the MAP detector

$$\hat{\mathbf{a}}(\mathbf{r}_N) = \arg\max_{\mathbf{a}\in\Omega^2} p(\mathbf{a}|\mathbf{r}_N). \tag{1}$$

3. Data Detection

We apply factor graphs (FG), which are a tool to efficiently compute marginal a posteriori distributions, such as $p(\mathbf{a}|\mathbf{r}_N)$. FGs have found applications in iterative decoding [8] and wireless receiver design [9].

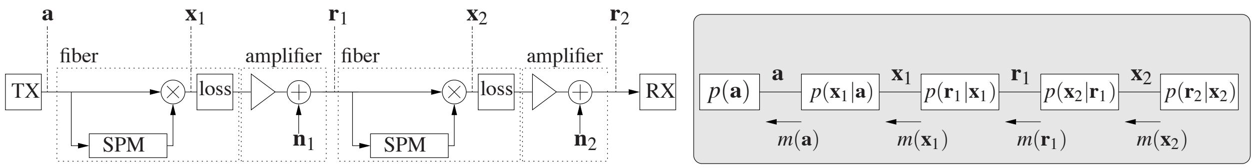

Fig. 1: *Left:* Transmission model for $N = 2$ spans with input symbol $\mathbf{a} \in \Omega^2$ and output $\mathbf{r}_2 \in \mathbb{C}^2$. Dot-dashed lines indicate names of intermediate variables. *Right:* The corresponding factor graph, with an edge for every unobserved variable, and a vertex (represented by boxes) for every factor in the factorization of $p(\mathbf{a}, \mathbf{x}_1, \mathbf{r}_1, \mathbf{x}_2 | \mathbf{r}_2)$.

3.1. Factor Graphs and the Sum-Product Algorithm

An FG is a graphical representation of a factorization of a distribution. In our case, the distribution $p(\mathbf{a}, \mathbf{x}_1, \mathbf{r}_1, \mathbf{x}_2 | \mathbf{r}_2)$ factorizes to

$$p(\mathbf{a}, \mathbf{x}_1, \mathbf{r}_1, \mathbf{x}_2 | \mathbf{r}_2) \propto p(\mathbf{a}) p(\mathbf{x}_1 | \mathbf{a}) p(\mathbf{r}_1 | \mathbf{x}_1) p(\mathbf{x}_2 | \mathbf{r}_1) p(\mathbf{r}_2 | \mathbf{x}_2), \tag{2}$$

where $\propto$ denotes equality up to a multiplicative constant. The corresponding FG is drawn in Fig. 1 on the right-hand side. For every unobserved variable, the FG contains an edge (or link). For every factor in (2), the FG contains a vertex (or node). Edges and vertices are connected when the corresponding variables appear in the corresponding factors. Note that this FG can be easily extended to a arbitrary number of fiber spans.

On an FG we can execute the sum-product algorithm (SPA), a message-passing algorithm that computes the marginal a posteriori distributions of the original distribution. In other words, the outcome of the SPA is $p(\mathbf{a}|\mathbf{r}_2)$, as well as $p(\mathbf{r}_1|\mathbf{r}_2)$, $p(\mathbf{x}_1|\mathbf{r}_2)$, and $p(\mathbf{x}_2|\mathbf{r}_2)$. Recall that $\mathbf{r}_2$ is observed and thus not considered as a variable. Messages are computed as follows: given a factor $f(\cdot)$ with variables x and y, and an incoming message $m(y)$, then the outgoing message is defined as [8]

$$m(x) = C \int f(x,y) m(y) \mathrm{d}y, \tag{3}$$

where C is a normalization constant and the integration occurs over the domain of the random variable y. When $f(\cdot)$ has only one variable (say, x), then (3) reverts to $m(x) = Cf(x)$. Some of these messages are shown in the FG in Fig. 1. Applying (3), they are given by $m(\mathbf{x}_2) \propto p(\mathbf{r}_2|\mathbf{x}_2)$, $m(\mathbf{r}_1) \propto \int p(\mathbf{x}_2|\mathbf{r}_1) m(\mathbf{x}_2) \mathrm{d}\mathbf{x}_2$, $m(\mathbf{x}_1) \propto \int p(\mathbf{r}_1|\mathbf{x}_1) m(\mathbf{r}_1) \mathrm{d}\mathbf{r}_1$, and $m(\mathbf{a}) \propto \int p(\mathbf{x}_1|\mathbf{a}) m(\mathbf{x}_1) \mathrm{d}\mathbf{x}_1$. It can be shown that $p(\mathbf{a}|\mathbf{r}_2) \propto p(\mathbf{a}) \times m(\mathbf{a})$. Hence, our problem (1) reverts to performing the message-passing rules.

3.2. Monte Carlo Integration and Detection

In practice, the integration (3) as well as the message representation may be performed through Monte Carlo techniques [9]. In this case, every message is represented by K samples. For example, we may draw samples $\mathbf{x}_1^{(1)}, \mathbf{x}_1^{(2)}, \ldots, \mathbf{x}_1^{(K)}$ from $m(\mathbf{x}_1)$, which form a sample representation of $m(\mathbf{x}_1)$. Finally, a decision regarding $\mathbf{a}$ is made as follows: the mean and covariance matrix of a four-dimensional real Gaussian distribution can be estimated from samples of $m(\mathbf{a})$. According to this Gaussian distribution, the likelihood for each point on the constellation is computed, weighed by $p(\mathbf{a})$. Eventually, the decision $\hat{\mathbf{a}}(\mathbf{r}_N)$ is made based on the largest value of the computation. We make the following observations: (i) the Monte Carlo integration can be performed offline to determine optimal decision regions, which can then be applied to the online transmission system; (ii) when we set $K = 1$ and neglect the factors $p(\mathbf{r}_i|\mathbf{x}_i)$ in the FG, we recover the well-know backpropagation algorithm for a memoryless channel. For that reason we name our proposed detector *stochastic backpropagation*.

4. Performance Analysis

We have performed computer simulations at 14 Gbaud per polarization, with $N = 22$ spans, $K = 100$, $N_0 = 4.9\ 10^{-7}$W/Hz, $\gamma = 1.25\ \mathrm{W}^{-1}\mathrm{km}^{-1}$ (corresponding to an EDFA noise figure of 4.8 dB), and $L_{\mathrm{eff}} = 17.36$ km, assuming Ω is a 16-QAM constellation. For comparison, we consider two competing detectors: (i) the first neglects NLPN completely, and makes a decision as $\hat{\mathbf{a}}(\mathbf{r}_N) = \arg\min_{\mathbf{a}\in\Omega^2} \|\mathbf{r}_N - \mathbf{a}\|^2$; (ii) the second detector performs backpropagation (in this case back-rotation), and thus only accounts for the deterministic channel effects by making a decision $\hat{\mathbf{a}}(\mathbf{r}_N) = \arg\min_{\mathbf{a}\in\Omega^2} \left\| \mathbf{r}_N - \mathbf{a} \exp(j\gamma L_{\mathrm{eff}} N \|\mathbf{a}\|^2) \right\|^2$. The symbol error rate (SER) performance of the three detectors are shown in Figs. 2–3 for single and dual polarization, respectively. We see that for both single and dual

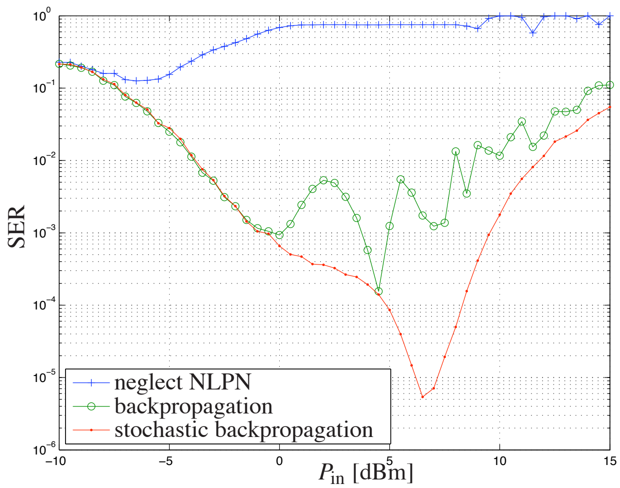

Fig. 2: SER for single polarization with 16-QAM.

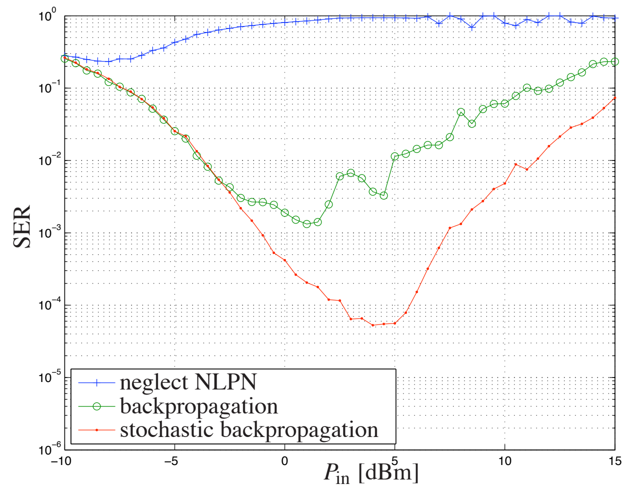

Fig. 3: SER for dual polarization with 16-QAM.

polarization, neglecting the NLPN leads to very poor performance. Backpropagation achieves good performance, but it is consistently outperformed by stochastic backpropagation. Minimal SER is achieved for input powers of 6.5 dBm and 4 dBm for single and dual polarization, respectively. The oscillations in the SER curves are due to the interaction between noise and SPM and their effect on the three different rings in the 16-QAM constellations [6].

5. Conclusion

We have presented a MAP detector for coherent optical communications at moderate baud rates, where the dominating impairment is NLPN. This new detector is based on Bayesian graphical models and can account for non-deterministic effects. Moreover, the detector can be interpreted as a generalization of backpropagation, and is thus called the stochastic backpropagation detector. The proposed approach significantly outperforms conventional backpropagation. We are currently investigating the extension to dispersive channels.

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] 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.

[3] 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.

[4] 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.

[5] Y. Cai, "MAP detection for linear and nonlinear ISI mitigation in long-haul coherent detection systems ,"*Photonics Society Summer Topical Meeting Series*, 2010, pp. 42–43.

[6] A. S. Tan, H. Wymeersch, P. Johannisson, E. Agrell, P. Andrekson, and M. Karlsson, "An ML-based detector for optical communication in the presence of nonlinear phase noise," *International Conference on Communications*, 2010.

[7] 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.

[8] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[9] H. Wymeersch, *Iterative Receiver Design*. Cambridge University Press, 2007.