



# CHALMERS

## Chalmers Publication Library

### **Stochastic Digital Backpropagation**

This document has been downloaded from Chalmers Publication Library (CPL). It is the author's version of a work that was accepted for publication in:

**IEEE Transactions on Communications (ISSN: 0090-6778)**

Citation for the published paper:

Irukulapati, N. ; Wymeersch, H. ; Johannisson, P. (2014) "Stochastic Digital Backpropagation". IEEE Transactions on Communications

Downloaded from: <http://publications.lib.chalmers.se/publication/205073>

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source. Please note that access to the published version might require a subscription.

Chalmers Publication Library (CPL) offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all types of publications: articles, dissertations, licentiate theses, masters theses, conference papers, reports etc. Since 2006 it is the official tool for Chalmers official publication statistics. To ensure that Chalmers research results are disseminated as widely as possible, an Open Access Policy has been adopted. The CPL service is administrated and maintained by Chalmers Library.

(article starts on next page)

# Stochastic Digital Backpropagation

Naga V. Irukulapati, *Student Member, IEEE*, Henk Wymeersch, *Member, IEEE*,  
Pontus Johannisson, Erik Agrell, *Senior Member, IEEE*

**Abstract**—In this paper, we propose a novel detector for single-channel long-haul coherent optical communications, termed *stochastic digital backpropagation (SDBP)*, which takes into account noise from the optical amplifiers in addition to handling deterministic linear and nonlinear impairments. We discuss the design approach behind this detector, which is based on the maximum a posteriori (MAP) principle. As closed-form expressions of the MAP detector are not tractable for coherent optical transmission, we employ the framework of Bayesian graphical models, which allows a numerical evaluation of the proposed detector. Through simulations, we observe that by accounting for nonlinear signal–noise interactions, we achieve a significant improvement in system reach with SDBP over digital backpropagation (DBP) for systems with periodic inline optical dispersion compensation. In uncompensated links with high symbol rates, the performance difference in terms of system reach for SDBP over DBP is small. In the absence of noise, the proposed detector is equivalent to the well-known DBP detector.

**Index Terms**—Digital backpropagation, factor graphs, near-MAP detector, nonlinear compensation, optical communications.

## I. INTRODUCTION

RECENT estimates suggest that there will be a ten-fold increase in the internet traffic by 2022 [1]. This ever increasing demand for data traffic continuously pushes the need for more effective and efficient optical transmission techniques. However, high-rate transmission poses several technical challenges to the system designer. In particular, the transceiver algorithms must cope with increased susceptibility to linear and nonlinear fiber impairments.

The two main impairments in the fiber-optic link are chromatic dispersion (CD) and the nonlinear Kerr effect. Compensation for CD can be performed in the optical domain, e.g., using inline dispersion-compensating fiber (DCF) or fiber Bragg gratings (FBG), or in the electrical domain through digital signal processing (DSP). Compensation of the CD in combination with the nonlinear Kerr effect can be done in DSP through the inversion of the Manakov model (or the nonlinear Schrödinger equation for a single polarization) using the split-step Fourier method (SSFM) with appropriate inverted channel parameters. The resulting algorithm is called digital backpropagation (DBP) [2], [3], which in spite of the high computational complexity, has been proposed as a

universal technique for jointly compensating for the linear and nonlinear impairments. Hence, DBP is often used to benchmark schemes proposed in the literature [4]–[16]. The assumed optimality of DBP has spurred intense research in low-complexity variations, including iterative symmetric SSFM [3], [7], noniterative asymmetric SSFM [2], weighted DBP [4], perturbation DBP [5], [8], and filtered DBP [9], [16]. The robustness of DBP has also been studied extensively for various scenarios in [10]–[15].

While DBP has received a great deal of attention, it only deals with deterministic linear and nonlinear impairments and inherently does not consider noise. It is known that the transmission performance of a fiber-optical system will ultimately be limited by non-deterministic nonlinear effects, such as *nonlinear signal–noise interaction (NSNI)* between the transmitted signal and the amplified spontaneous emission (ASE) noise [17]–[19]. The impact of NSNI has been studied extensively through numerical simulations and through experiments [20], [21]. In these works, ASE noise is loaded at the receiver as a single additive white Gaussian noise (AWGN) process. This system is then compared with a system where noise is distributed [20] and ASE noise is added as a single AWGN at the transmitter [21]. Both these studies revealed that there is a significant penalty due to NSNI for dispersion-managed links and the severity of the NSNI is dependent on the symbol rate and modulation formats used in the system. None of the DBP methods mentioned above account for NSNI, as it is often argued that one cannot compensate for NSNI in DSP due to the non-deterministic nature of ASE noise [21]. To deal with stochastic disturbances, Bayesian detection theory can be used to formulate maximum a posteriori (MAP) detectors, which are provably optimal in terms of error probability. MAP detectors have been proposed for the discrete memoryless channel [22] assuming perfect CD compensation by inline optical dispersion compensators, and utilizing a look-up table detector that can mitigate data pattern dependent nonlinear impairments [23]. A low-complexity Viterbi detector is suggested as an alternative or complementing DBP for combating fiber nonlinearities [24]. However, for computational reasons, this detector is suitable for systems with very low dispersion, hence limiting its applicability.

In this paper, we propose a near-MAP detector for single-channel<sup>1</sup> long-haul coherent optical communications that compensates not only for deterministic linear and nonlinear effects but also takes the ASE noise into account. As a consequence, NSNI is accounted for, which allows us to (i) get closer to the fundamental performance limits of the fiber-optical channel

The authors are associated with the FORCE Research Centre at Chalmers University of Technology, 41296 Gothenburg Sweden. *E-mail*: {vnaga, henkw, pontus.johannisson, agrell}@chalmers.se.

This research was supported by the Swedish Research Council (VR) under grant 2010 – 4236. The simulations were performed in part on resources provided by the Swedish National Infrastructure for Computing (SNIC) at C3SE.

Part of this work was presented in 39th European conference and exhibition on optical communication in London, UK, Sep. 2013.

<sup>1</sup>The extension of the proposed detector to wavelength-division multiplexed system is out of the scope of the current contribution and is considered for future work.

and (ii) identify regimes where DBP is suboptimal. Since DBP can be interpreted as a special case of the proposed near-MAP detector, we call the proposed method *stochastic digital backpropagation (SDBP)*. Our specific contributions are:

- Developing a novel detector that accounts for NSNI by explicitly dealing with ASE noise;
- Identifying regimes where DBP is suboptimal; and
- A novel receiver design based on factor graphs (FGs), wherein the variables are waveforms.

The remainder of this paper is structured as follows. In Section II, we describe the model for a fiber-optical transmission system, including both dispersion-managed (DM) and non-dispersion-managed (NDM) links. The derivation of the MAP detector will be described in Section III, also containing a brief overview of the theory of FGs and message passing. The implementation of the near-MAP detector is described in Section IV. In Section V, interpretation of SDBP and connection to DBP is given for a simplified model. Simulation results and discussions are presented in Section VI, followed by conclusions in Section VII.

*Notation:* Lower case bold letters (e.g.,  $\mathbf{x}$ ) are used for vector representation of the continuous-time signals and underlined lower case bold letters (e.g.,  $\underline{\mathbf{y}}$ ) for a vector of discrete-time symbols. Note that even though both these classes of signals are represented by vectors, they are quite distinct. The sequence  $\underline{\mathbf{y}}$  denotes the data that is spaced a symbol period apart and  $\underline{\mathbf{x}}$  is the oversampled continuous-time optical signal  $\mathbf{x}(t)$  (or any equivalent representation in a suitable basis). With a slight abuse of notation,  $\mathbf{x}(t)$  is used to represent a vector of dual-polarization single-wavelength continuous-time signal and  $\mathbf{x}$  is used to represent samples of  $\mathbf{x}(t)$ , where the data for each polarization is appropriately combined. Hermitian conjugation of the vector  $\mathbf{v}$  is denoted by  $\mathbf{v}^H$ . A multivariate Gaussian probability density function (PDF) of a variable  $\mathbf{r}$  with mean  $\mathbf{z}$  and covariance matrix  $\Sigma$  is denoted by  $\mathcal{N}(\mathbf{r}; \mathbf{z}, \Sigma)$ . The indicator function with proposition  $\mathcal{P}$  is given by  $\mathbb{1}(\mathcal{P})$  and the Dirac delta function is denoted by  $\delta(\cdot)$ .

## II. SYSTEM MODEL

### A. High-Level Description

The system model, shown in Fig. 1(a), comprises a dual-polarization transmitter, a pulse shaper, a fiber-optical link with  $N$  spans, and a receiver with a compensation algorithm followed by a decision unit. Each span of the fiber-optic link consists of a standard single-mode fiber (SMF) followed by an optional dispersion-compensating module (DCM) for DM links.<sup>2</sup> In this work, we have considered either a DCF or an FBG as a DCM. In between fiber spans, there are erbium-doped fiber amplifiers (EDFAs).

A sequence of  $K$  symbols  $\underline{\mathbf{s}} = [\underline{\mathbf{s}}[1], \underline{\mathbf{s}}[2], \dots, \underline{\mathbf{s}}[K]] \in \Omega^K$  is transmitted at a symbol rate  $1/T$  with a pulse shaping filter  $g(t)$ , where  $\Omega$  is the set of possible symbols in the four-dimensional constellation. Each  $\underline{\mathbf{s}}[k]$  is a complex length-2

vector consisting of data from the  $x$  and  $y$  polarizations. The transmitted signal, denoted by<sup>3</sup>  $\mathbf{v}(t, \ell_{1,1}) = \sum_{k=1}^K \underline{\mathbf{s}}[k]g(t - kT)$ , is up-converted and transmitted over a DM or NDM link. The equivalent complex baseband received signal is denoted by  $\mathbf{r}(t)$  or the equivalent vector representation  $\mathbf{r}$ . The overall goal of the receiver is to optimally recover  $\underline{\mathbf{s}}$  from  $\mathbf{r}$ . While different optimality criteria can be considered, we will aim to minimize the error probability, leading to a MAP receiver, in which the estimate of  $\underline{\mathbf{s}}$  is

$$\hat{\underline{\mathbf{s}}} = \arg \max_{\underline{\mathbf{s}} \in \Omega^K} p(\underline{\mathbf{s}}|\mathbf{r}), \quad (1)$$

where  $p(\underline{\mathbf{s}}|\mathbf{r})$  is the a posteriori distribution of  $\underline{\mathbf{s}}$  given the received signal  $\mathbf{r}$ .

### B. Signal Propagation in Optical Fibers

For the receiver to determine  $p(\underline{\mathbf{s}}|\mathbf{r})$ , an accurate propagation model is required. The propagation of light in an optical fiber is modeled by the Manakov model (with loss included) as [25, eq. (2.3.45)]

$$\frac{\partial \mathbf{v}(t, z)}{\partial z} = -j \frac{\beta_2}{2} \frac{\partial^2 \mathbf{v}(t, z)}{\partial t^2} + j \gamma (\mathbf{v}(t, z)^H \mathbf{v}(t, z)) \mathbf{v}(t, z) - \frac{\alpha}{2} \mathbf{v}(t, z), \quad (2)$$

where  $\mathbf{v}$  is the electric field for the polarization-multiplexed (PM) signal and  $\mathbf{v}^H \mathbf{v} = \|\mathbf{v}\|^2$  is the signal power summed over the two polarizations. The fiber nonlinearity coefficient is denoted by  $\gamma$ ,  $\alpha$  is the power attenuation coefficient,  $\beta_2$  is the group velocity dispersion, and  $z$  is the propagation distance. We note that (2) has been obtained by averaging over the fast polarization rotations, making the analysis applicable to fibers with low polarization mode dispersion (PMD). Hence, PMD is ignored in the rest of the paper, except in Sec. VI-C, where the influence of PMD on DBP and SDBP is analyzed.

The Manakov model (2) generally does not have an analytic solution, except for specific cases. Numerical approaches, such as the *split-step Fourier method* (SSFM) [25, eq. (2.4.10)], are typically needed to describe the signal evolution in dispersive and nonlinear media. The Manakov equation (2) can be solved using a symmetric SSFM [3], [25] or an asymmetric SSFM [2]. To derive the near-MAP detector, we choose the asymmetric SSFM as it is easy to comprehend the principles behind the proposed detector using this approach.<sup>4</sup> In the SSFM approach, the fiber span of length  $L \in \{L_{SMF}, L_{DCM}\}$  is split into  $M \in \{M_{SMF}, M_{DCM}\}$  segments, where the length of each segment  $\Delta z$  is chosen small enough that the nonlinear (Kerr nonlinearity) and linear (chromatic dispersion) effects can be modeled as acting independently. In Fig. 1(b), the asymmetric SSFM view of the fiber (SMF/DCF) for span  $n \in \{1, \dots, N\}$  is shown. Focusing on the segment  $m \in \{1, \dots, M\}$  of span  $n$ , and introducing  $\ell_{m,n} = (m-1)\Delta z + (n-1)L$ , the input

<sup>3</sup>The notation  $\ell_{1,1}$  refers to the input of the first segment of the first span of the fiber, to be clarified later in the paper.

<sup>4</sup>As will be seen in later sections, principles suggested in this paper do not change even if we had considered a symmetric SSFM. However, in the simulations symmetric SSFM is used, as it performs better at the same computational complexity [25, Ch. 2].

<sup>2</sup>The system shown in Fig. 1(a) should be considered as an example, as the principles presented in later sections of this paper can be applied directly to systems with different configurations (e.g., comprising an EDFA after both the SMF and the DCM). Alternative configurations will be presented in Sec. VI

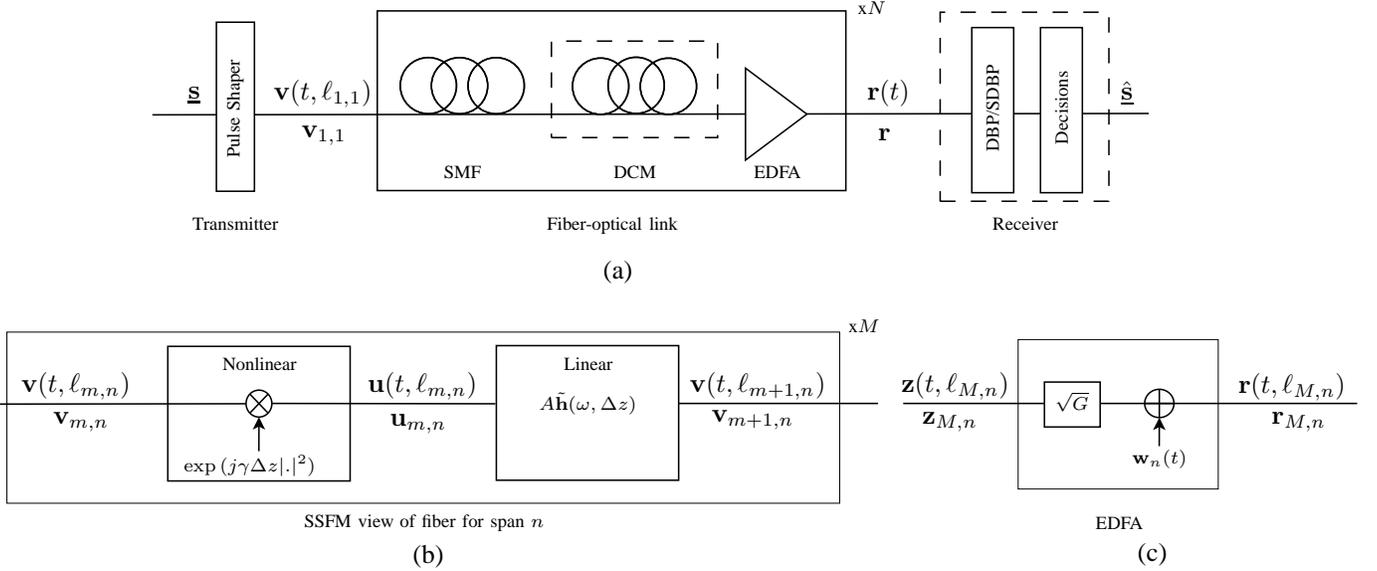


Fig. 1. (a) A fiber link with  $N$  spans where each span consists of an SMF, a DCM (for DM links), and an EDFA. The transmitter consists of a pulse shaper and the receiver consists of a compensation algorithm (DBP/SDBP) and a decision block. (b) Approximate model for the fiber (SMF/DCF) using SSFM with nonlinear and linear segments, where  $\tilde{h}(\omega, \Delta z) = \exp(j\beta_2\omega^2\Delta z/2)$  and  $A = \exp(-\alpha\Delta z/2)$ . (c) EDFA with gain  $G = \exp(\alpha L)$  and noise.

signal will be denoted by  $\mathbf{v}(t, \ell_{m,n})$  and the output signal by  $\mathbf{v}(t, \ell_{m+1,n})$ . The nonlinear effect is then given by

$$\mathbf{u}(t, \ell_{m,n}) = \mathbf{v}(t, \ell_{m,n}) \exp(j\gamma \|\mathbf{v}(t, \ell_{m,n})\|^2 \Delta z), \quad (3)$$

where  $\gamma \in \{\gamma_{\text{SMF}}, \gamma_{\text{DCM}}\}$  is the fiber nonlinear coefficient per segment, with  $\gamma_{\text{SMF}}$  and  $\gamma_{\text{DCM}}$  denoting nonlinearity coefficient in the SMF and DCM, respectively. The linear effect is described by

$$\mathbf{v}(t, \ell_{m+1,n}) = A \mathbf{u}(t, \ell_{m,n}) * \mathbf{h}(t, \Delta z), \quad (4)$$

where  $A \triangleq \exp(-\alpha\Delta z/2)$ ,  $\alpha \in \{\alpha_{\text{SMF}}, \alpha_{\text{DCM}}\}$  is the power attenuation coefficient per segment, and  $*$  denotes convolution. The CD impulse response is given by  $\mathbf{h}(t, z) = \exp(-jt^2/(2\beta_2 z)) / \sqrt{j2\pi\beta_2 z}$  [26] with corresponding frequency response  $\mathbf{h}(\omega, z) = \exp(j\beta_2\omega^2 z/2)$ , where  $\beta_2 \in \{\beta_{2,\text{SMF}}, \beta_{2,\text{DCM}}\}$  is the group velocity dispersion per segment.

Since a fiber is a lossy medium, periodic amplification of the signal is needed. Fig. 1(c) shows the operation of the optical amplifier, consisting of gain and addition of white Gaussian noise (AWGN). We assume that each EDFA compensates for the attenuation in the preceding fiber span and adds circular white complex Gaussian ASE noise  $\mathbf{w}_n(t)$  in each span, with power spectral density  $GF_n h\nu_{\text{opt}}/2$  in each polarization [27, eq. (8.1.15)] and  $h\nu_{\text{opt}}$  is the photon energy. Each EDFA is assumed to have a bandwidth  $B$ . The required gain to compensate for the attenuation in each span is  $G = \exp(\alpha L)$ . The noise figure [27, eq. (8.1.19)] of the EDFA is  $F_n = 2n_{\text{sp}}(1 - G^{-1})$ , in which  $n_{\text{sp}}$  is the spontaneous emission factor. The signal before and after the EDFA of span  $n$  will be denoted by  $\mathbf{z}(t, \ell_{M,n})$  and  $\mathbf{r}(t, \ell_{M,n})$ , respectively. The vector representation of the continuous-time signals can also be seen in Fig. 1. For example, the vector representation of the signals  $\mathbf{v}(t, \ell_{m,n})$ ,  $\mathbf{z}(t, \ell_{M,n})$ , and  $\mathbf{r}(t)$  is denoted by  $\mathbf{v}_{m,n}$ ,  $\mathbf{z}_{M,n}$ , and  $\mathbf{r}$ , respectively. When the fiber and EDFA are concatenated,

at the end of the DCM (or the SMF for the NDM link),  $\mathbf{z}_{M,n} = \mathbf{v}_{m+1,n}$  when  $m = M$ . At the end of each span,  $\mathbf{r}_{M,n \neq N} = \mathbf{v}_{1,n+1}$ , while at the end of the link  $\mathbf{r}_{M,N} = \mathbf{r}$ .

### III. MAP DETECTION

Based on the transmission system in Fig. 1, our aim is to determine  $p(\underline{\mathbf{s}}|\mathbf{r})$  through the marginalization of a joint distribution  $p(\underline{\mathbf{s}}, \mathcal{H}, \mathbf{r})$ , where  $\mathcal{H}$  encapsulates the hidden (or unobserved) variables in the system.

We utilize FGs and the sum-product algorithm (SPA) to efficiently perform this marginalization process [28]. For a detailed description of FGs, the reader is referred to [29], [30]. For our purposes, it suffices to mention that an FG is generated based on a factorization of a joint distribution and that the SPA is a message passing algorithm on the FG, with the aim of computing marginal posterior distributions.

#### A. Factorization of the Joint Distribution

We introduce as hidden variables the internal states of our transmission system (Fig. 1)

$$\mathcal{H} = \{ \{ \mathbf{v}_{m,n}, \mathbf{u}_{m,n} \}_{1 \leq m \leq M}, \mathbf{z}_{M,n}, \mathbf{r}_{M,n \neq N} \}_{1 \leq n \leq N},$$

where we note that the observation  $\mathbf{r}_{M,N} = \mathbf{r}$  is not part of  $\mathcal{H}$ . Harnessing conditional independence<sup>5</sup> among these internal

<sup>5</sup>Our system forms a Markov chain, so that the probability of an output given all previous inputs will just depend on the input to the current block.

states, the joint distribution  $p(\underline{s}, \mathcal{H}, \mathbf{r})$  is factorized as

$$\begin{aligned} p(\underline{s}, \{\{\mathbf{v}_{m,n}, \mathbf{u}_{m,n}\}_{1 \leq m \leq M}, \mathbf{z}_{M,n}, \mathbf{r}_{M,n \neq N}\}_{1 \leq n \leq N}, \mathbf{r}) \\ = p(\underline{s}) p(\mathbf{v}_{1,1} | \underline{s}) \prod_{n=1}^N p(\mathbf{r}_{M,n} | \mathbf{z}_{M,n}) p(\mathbf{z}_{M,n} | \mathbf{u}_{M,n}) \cdots \\ \prod_{m=2}^M p(\mathbf{u}_{m,n} | \mathbf{v}_{m,n}) \prod_{m=1}^{M-1} p(\mathbf{v}_{m+1,n} | \mathbf{u}_{m,n}) \cdots \\ p(\mathbf{u}_{1,n} | \mathbf{r}_{M,n-1 \neq 0}) p(\mathbf{u}_{1,1} | \mathbf{v}_{1,1}). \end{aligned} \quad (5)$$

According to this factorization, an FG is drawn<sup>6</sup> as shown in Fig. 2 for the system model of Fig. 1(a). To keep the FG framework compact, we define  $M$  in Fig. 2 to be the sum of the number of segments needed for SMF and DCM in the case of compensated links.<sup>7</sup> The FG in Fig. 2 contains the factors

$$f_A : p(\mathbf{r}_{M,n} | \mathbf{z}_{M,n}) = \mathcal{N}(\mathbf{r}_{M,n}; \sqrt{G} \mathbf{z}_{M,n}, \Sigma), \quad (6)$$

$$f_B : p(\mathbf{v}_{m+1,n} | \mathbf{u}_{m,n}) = \delta(\mathbf{v}_{m+1,n} - A \mathbf{u}_{m,n} * \mathbf{h}), \quad (7)$$

$$\begin{aligned} f_C : p(\mathbf{u}_{m,n} | \mathbf{v}_{m,n}) \\ = \delta(\mathbf{u}_{m,n} - \mathbf{v}_{m,n} \exp(j\gamma \Delta z | \mathbf{v}_{m,n}|^2)), \end{aligned} \quad (8)$$

$$f_D : p(\mathbf{v}_{1,1} | \underline{s}) = \delta(\mathbf{v}_{1,1} - \underline{s} * \mathbf{g}), \quad (9)$$

where (6) is due to the fact that  $\mathbf{r}_{M,n}$  is obtained by scaling  $\mathbf{z}_{M,n}$  with  $\sqrt{G}$  and adding Gaussian noise of zero mean and covariance matrix<sup>8</sup>  $\Sigma$ . The pulse shape is represented by  $\mathbf{g}$  in (9) and the transmitted signal is represented by  $\mathbf{v}(t, \ell_{1,1}) = \sum_{k=1}^K \underline{s}[k] \mathbf{g}(t - kT)$ , where  $T$  is the symbol period. With a slight abuse of notation in (9), we set  $\mathbf{v}_{1,1}$  to equal the convolution of  $\underline{s}$  and  $\mathbf{g}$ .

## B. Message Passing: Theory

On the generated FG, we can now perform message passing according to the SPA. In SPA, messages passed over an edge are a function of the corresponding variable. We will denote a message associated with edge/variable  $\mathbf{Q}$ , evaluated in  $\mathbf{q}$  by  $\mu_{\mathbf{Q}}(\mathbf{q})$  (or  $\mu(\mathbf{q})$  when the variable is clear from the context). The direction of the message will be represented by arrows:  $\overleftarrow{\mu}_{\mathbf{Q}}(\mathbf{q})$  and  $\overrightarrow{\mu}_{\mathbf{Q}}(\mathbf{q})$ . We note that messages can be normalized without affecting the normalized marginal. These normalized messages can be interpreted as distributions: probability mass functions and probability density functions for discrete and continuous variables, respectively.

Considering a generic function  $f(\mathbf{q}_1, \mathbf{q}_2)$  with a corresponding vertex in an FG as shown in Fig. 3, and incoming messages  $\overleftarrow{\mu}_{\mathbf{Q}_2}(\mathbf{q}_2)$  and  $\overrightarrow{\mu}_{\mathbf{Q}_1}(\mathbf{q}_1)$ , the outgoing messages are computed

<sup>6</sup>A vertex (drawn as a square) is created for every factor and an edge (drawn as a line) for every variable. A certain edge is attached to a certain node when the corresponding variable appears in the corresponding factor.

<sup>7</sup>When  $M = M_{\text{SMF}} + M_{\text{DCM}}$ , there are  $M_{\text{SMF}}$  SMF segments in the FG of Fig. 2, followed by  $M_{\text{DCM}}$  DCM segments. For uncompensated links,  $M_{\text{DCM}} = 0$ .

<sup>8</sup>Note that the dimension of  $\Sigma$  is determined by the length of the vector representation of the corresponding waveforms. The entries in  $\Sigma$  are determined by the basis used in the vector representation of the continuous-time signals (e.g., when we sample at a rate  $B$ , we find that  $\Sigma = GF_n h \nu_{\text{opt}} B / 2 \mathbf{I}$ ).

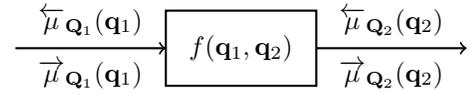


Fig. 3. Factor graph for a case where we have two variables  $\mathbf{q}_1, \mathbf{q}_2$  and a factor  $f(\mathbf{q}_1, \mathbf{q}_2)$  associated with those variables.

as [29]

$$\overleftarrow{\mu}_{\mathbf{Q}_1}(\mathbf{q}_1) \propto \int f(\mathbf{q}_1, \mathbf{q}_2) \overleftarrow{\mu}_{\mathbf{Q}_2}(\mathbf{q}_2) d\mathbf{q}_2 \quad (10)$$

$$\overrightarrow{\mu}_{\mathbf{Q}_2}(\mathbf{q}_2) \propto \int f(\mathbf{q}_1, \mathbf{q}_2) \overrightarrow{\mu}_{\mathbf{Q}_1}(\mathbf{q}_1) d\mathbf{q}_1, \quad (11)$$

where the integral should be replaced with a summation when the relevant variables are discrete, and  $\propto$  is used to normalize the messages. Note that the two messages are computed in parallel. To compute the marginal posterior of a variable, say  $\mathbf{q}_1$ , we multiply leftward (also called backward) and rightward (also called forward) messages [29]

$$p(\mathbf{q}_1) \propto \overrightarrow{\mu}_{\mathbf{Q}_1}(\mathbf{q}_1) \overleftarrow{\mu}_{\mathbf{Q}_1}(\mathbf{q}_1). \quad (12)$$

As we are ultimately interested in  $p(\underline{s} | \mathbf{r})$ , we mainly compute leftward messages in the FG in Fig. 2.

Note that many of the factors in (5) correspond to functions of the form  $f(\mathbf{q}_1, \mathbf{q}_2) = \delta(\mathbf{q}_2 - \phi(\mathbf{q}_1))$ , where  $\phi$  is bijective. This knowledge can be exploited to simplify the message computation in (10) as

$$\begin{aligned} \overleftarrow{\mu}_{\mathbf{Q}_1}(\mathbf{q}_1) &\propto \int \delta(\mathbf{q}_2 - \phi(\mathbf{q}_1)) \overleftarrow{\mu}_{\mathbf{Q}_2}(\mathbf{q}_2) d\mathbf{q}_2 \\ &= \overleftarrow{\mu}_{\mathbf{Q}_2}(\phi(\mathbf{q}_1)). \end{aligned} \quad (13)$$

Note that (13) can also be expressed as

$$\overleftarrow{\mu}_{\mathbf{Q}_1}(\phi^{-1}(\mathbf{q}_2)) = \overleftarrow{\mu}_{\mathbf{Q}_2}(\mathbf{q}_2). \quad (14)$$

For discrete variables,<sup>9</sup> we additionally have that  $\overrightarrow{\mu}_{\mathbf{Q}_2}(\mathbf{q}_2) = \overrightarrow{\mu}_{\mathbf{Q}_1}(\phi^{-1}(\mathbf{q}_2))$  (in which case we have that  $f(\mathbf{q}_1, \mathbf{q}_2) = \mathbb{1}(\mathbf{q}_2 = \phi(\mathbf{q}_1))$ ).

The specific problem considered here requires additional considerations not present in conventional FGs. In particular, we point out the following.

- *Variables correspond to waveforms:* To the best of our knowledge, this is the first application of FG where the variables themselves are entire waveforms. While these can be represented with high-dimensional vectors, it is not trivial to represent the messages associated with these variables, or to compute the messages according to the SPA.
- *Type:* Variables also have an inherent domain or type. For example when a variable is nonnegative, we write  $\text{type}(\mathbf{Q}_k) = \{\mathbf{q}_k \mid \mathbf{q}_k \geq 0\}$ . In the case of signals, type can also cover temporal and spectral characteristics. For example, when a signal is bandlimited to  $[-B, B]$ , we can write  $\text{type}(\mathbf{Q}_k) = \{\mathbf{q}_k \mid \text{FT}(\mathbf{q}_k) = 0, \forall f \notin [-B, B]\}$ , where FT is the Fourier transform. Messages only need to be evaluated within the type, e.g.,  $\mathbf{q}_1 \in \text{type}(\mathbf{Q}_1)$  for

<sup>9</sup>For continuous variables, the forward message,  $\overrightarrow{\mu}_{\mathbf{Q}_2}(\mathbf{q}_2)$  requires additional care to account for the Jacobian.

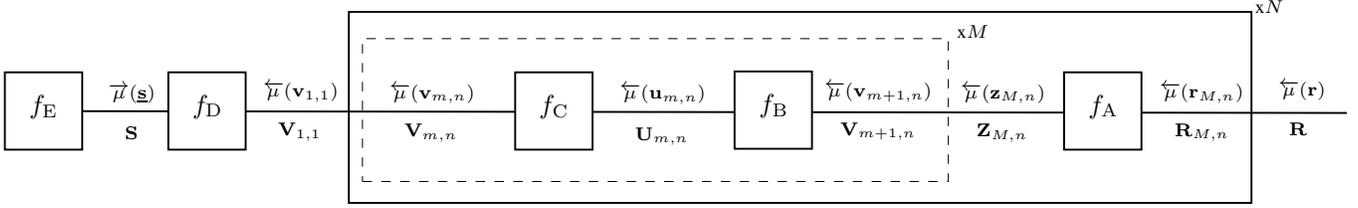


Fig. 2. Factor graph for the system model in Fig. 1(a) where  $f_A$  corresponds to the EDFA block,  $f_B$  corresponds to the CD block (for either an SMF or for a DCF depending on  $m$ ),  $f_C$  corresponds to nonlinear block (either for SMF or DCM depending on  $m$ ),  $f_D$  corresponds to the pulse shape and  $f_E$  represents prior knowledge about the symbols. We denote the variables in capitals (e.g.,  $\mathbf{V}_{m,n}$ ) to distinguish from their realizations (e.g.,  $\mathbf{v}_{m,n}$ ).

(10). However, it is sometimes convenient to evaluate a message in a larger domain, whereby the resulting message can be projected onto the type and renormalized. For example, as will be seen in the next section, the knowledge that  $\mathbf{v}_{1,1}$  has a known structure (of the form  $\underline{\mathbf{s}} * \mathbf{g}$ ) is not utilized while computing the message for  $\mathbf{v}_{1,1}$ ; rather we let  $\mathbf{v}_{1,1}$  be of any type.

### C. Message Passing in SDBP

Using (10) and (13) and the factors for each of the blocks as given in (6)–(9), the messages for each of the blocks of the model of Fig. 1 can be derived.

- *Message for the EDFA block:* Given the message  $\overleftarrow{\mu}(\mathbf{r}_{M,n})$ , the message  $\overleftarrow{\mu}(\mathbf{z}_{M,n})$  can be computed by applying (10) and (6):

$$\overleftarrow{\mu}(\mathbf{z}_{M,n}) \propto \int \mathcal{N}(\mathbf{r}_{M,n}; \sqrt{G}\mathbf{z}_{M,n}, \Sigma) \times \overleftarrow{\mu}(\mathbf{r}_{M,n}) d\mathbf{r}_{M,n}. \quad (15)$$

Observe that in general,  $\overleftarrow{\mu}(\mathbf{z}_{M,n})$  has no closed-form expression.

- *Message for the dispersion (linear) block:* We compute the message  $\overleftarrow{\mu}(\mathbf{u}_{m,n})$  at the input of the linear block given the message  $\overleftarrow{\mu}(\mathbf{v}_{m+1,n})$  at the output of the linear block. We recall that  $\mathbf{v}_{m+1,n} = \mathbf{z}_{M,n}$  when  $m = M$ , that is, at the end of the last segment, and therefore  $\overleftarrow{\mu}(\mathbf{v}_{m+1,n}) = \overleftarrow{\mu}(\mathbf{z}_{M,n})$  for  $m = M$ . To compute the message  $\overleftarrow{\mu}(\mathbf{u}_{m,n})$ , we use (13)–(14) with  $\mathbf{q}_1 = \mathbf{u}_{m,n}$  and  $\mathbf{q}_2 = \mathbf{v}_{m+1,n}$  to get

$$\overleftarrow{\mu}_{\mathbf{U}_{m,n}}(\mathbf{u}_{m,n}) = \overleftarrow{\mu}_{\mathbf{V}_{m+1,n}}(\phi(\mathbf{u}_{m,n})), \quad (16)$$

$$\overleftarrow{\mu}_{\mathbf{U}_{m,n}}(\phi^{-1}(\mathbf{v}_{m+1,n})) = \overleftarrow{\mu}_{\mathbf{V}_{m+1,n}}(\mathbf{v}_{m+1,n}), \quad (17)$$

where  $\phi(\mathbf{u}_{m,n}) = \mathbf{A}\mathbf{u}_{m,n} * \mathbf{h}$ . Hence, the outgoing message is obtained by applying a transformation to the variable of the incoming message. The form (17) has an operational meaning: if the inverse filter  $\phi^{-1}(\cdot)$  is applied to the signal  $\mathbf{v}_{m+1,n}$ , then the resulting signal  $\mathbf{u}_{m,n}$  will have the same probability density as  $\mathbf{v}_{m+1,n}$ .

- *Message for the nonlinear block:* Based on the function defined in (8) for the nonlinear operator, we find the message  $\overleftarrow{\mu}(\mathbf{v}_{m,n})$  from  $\overleftarrow{\mu}(\mathbf{u}_{m,n})$  by substituting  $\mathbf{q}_1 = \mathbf{v}_{m,n}$

and  $\mathbf{q}_2 = \mathbf{u}_{m,n}$  in (13)–(14) to get

$$\phi(\mathbf{v}_{m,n}) = \mathbf{v}_{m,n} \exp(j\gamma\Delta z|\mathbf{v}_{m,n}|^2), \quad (18)$$

$$\phi^{-1}(\mathbf{u}_{m,n}) = \mathbf{u}_{m,n} \exp(-j\gamma\Delta z|\mathbf{u}_{m,n}|^2), \quad (19)$$

$$\overleftarrow{\mu}_{\mathbf{V}_{m,n}}(\phi^{-1}(\mathbf{u}_{m,n})) = \overleftarrow{\mu}_{\mathbf{U}_{m,n}}(\mathbf{u}_{m,n}), \quad (20)$$

where (19) follows because  $|\mathbf{u}_{m,n}|^2 = |\mathbf{v}_{m,n}|^2$ . Hence, we apply the inverse nonlinearity to a signal  $\mathbf{u}_{m,n}$ , yielding a signal  $\mathbf{v}_{m,n}$  with the same probability density.

Recursive application of the rules above leads to  $\overleftarrow{\mu}(\mathbf{v}_{1,1})$ . Since  $\mathbf{v}_{1,1}$  and  $\underline{\mathbf{s}}$  are related by a one-to-one mapping, determining  $p(\underline{\mathbf{s}}|\mathbf{r})$  can equivalently be accomplished by determining  $p(\mathbf{v}_{1,1}|\mathbf{r})$ . This requires  $\overleftarrow{\mu}(\mathbf{v}_{1,1})$  and  $\overrightarrow{\mu}(\mathbf{v}_{1,1})$ . The latter message is computed by applying (11) and (9) to get

$$\begin{aligned} \overrightarrow{\mu}(\mathbf{v}_{1,1}) &\propto \sum_{\underline{\mathbf{s}}} \overrightarrow{\mu}(\underline{\mathbf{s}})p(\mathbf{v}_{1,1}|\underline{\mathbf{s}}) \\ &= \sum_{\underline{\mathbf{s}}} \sum_{k=1}^{|\Omega|^K} \mathbb{1}(\underline{\mathbf{s}} = \underline{\mathbf{s}}_k) \delta(\mathbf{v}_{1,1} - \underline{\mathbf{s}} * \mathbf{g}) \\ &= \sum_{k=1}^{|\Omega|^K} \delta(\mathbf{v}_{1,1} - \underline{\mathbf{s}}_k * \mathbf{g}), \end{aligned} \quad (21)$$

where we have utilized the fact that

$$\overrightarrow{\mu}(\underline{\mathbf{s}}) = \frac{1}{|\Omega|^K} \sum_{k=1}^{|\Omega|^K} \mathbb{1}(\underline{\mathbf{s}} = \underline{\mathbf{s}}_k), \quad (22)$$

in which  $\underline{\mathbf{s}}_k = [\underline{\mathbf{s}}_k[1], \underline{\mathbf{s}}_k[2], \dots, \underline{\mathbf{s}}_k[K]]$  is a unique realization of the  $K$  transmitted symbols. In other words, the message  $\overrightarrow{\mu}(\underline{\mathbf{s}})$  is a sum of indicator functions of all possible  $|\Omega|^K$  combinations with  $|\Omega|$  being the number of possible symbols in the constellation.

Finally, the multiplication of  $\overrightarrow{\mu}(\mathbf{v}_{1,1})$  with  $\overleftarrow{\mu}(\mathbf{v}_{1,1})$  is accomplished by evaluating each distinct  $\mathbf{v}_{1,1} = \underline{\mathbf{s}}_k * \mathbf{g}$  in the function  $\overleftarrow{\mu}(\mathbf{v}_{1,1})$ . This immediately yields the corresponding  $p(\underline{\mathbf{s}}_k|\mathbf{r})$ .

There are some challenges in implementing a receiver as suggested above.

- *Representation:* A common problem in the application of the SPA is the representation of messages. When variables are discrete, messages can be represented by vectors tabulating the value of the message for each possible value of the variable. When variables are continuous, approximate representations may be required, including grid approximations, parametric approximations,

and nonparametric approximations [28]. In our case, the problem is further exacerbated by the high-dimensional nature of the variables, as they correspond to vector representations of waveforms.

- *Computation:* Even with a practical representation, the integral (15) is generally not tractable. Secondly, when the input message in (17), (20) falls within the chosen representation, the output message may not be representable. Moreover, the forward message  $\vec{\mu}(\mathbf{v}_{1,1})$  is a sum of  $|\Omega|^K$  combinations, which is highly complex for reasonable values of  $K$ .

Both of these issues will be dealt with in the following section.

#### IV. NEAR-MAP DETECTOR IMPLEMENTATION

In this section, we discuss an implementation of the receiver outlined in the previous section. The representation problem will be tackled using a particle representation (PR) [28, Ch. 3], which may be seen as a high-dimensional variation of a particle filter [31]. The same PR will allow us to evaluate integrals of the form (15) and also to maintain a coherent representation to compute (17), (20). The exponential complexity related to the forward message  $\vec{\mu}(\mathbf{v}_{1,1})$  will be treated using a matched filter approach. Combining these techniques (PR and matched filter) leads to a receiver that bears a remarkable similarity to the well-known digital back propagation receiver (for a discussion, see Section V).

##### A. Particle Representations

Given a PDF  $p_X(x)$ , a PR, denoted by  $\text{PR}\{p_X\}$ , is a list of values<sup>10</sup>  $\{x^{(k)}\}_{k=1}^{N_p}$ , with the property that for any integrable function  $f(x)$

$$\frac{1}{N_p} \sum_{k=1}^{N_p} f(x^{(k)}) \rightarrow \int f(x) p_X(x) dx, \quad N_p \rightarrow +\infty. \quad (23)$$

One way to obtain a PR is to draw  $N_p$  i.i.d. samples from  $p_X(x)$ , though many other methods exist [28, Ch. 3]. A PR can be interpreted as follows: in the context of (23),  $p_X(x)$  can be approximated as  $p_X(x) \approx 1/N_p \sum_{k=1}^{N_p} \delta(x - x^{(k)})$ . In other words, a  $\text{PR}\{p_X\}$  can be considered as a uniform probability mass function, so that  $X$  is considered to be a uniform discrete random variable that takes on values in the set  $\{x^{(k)}\}_{k=1}^{N_p}$ .

With this PR, the messages for each of the blocks of our system model can be written, as detailed below.

- *Message for EDFA block:* Given message  $\overleftarrow{\mu}(\mathbf{r}_{M,n}) \approx 1/N_p \sum_{k=1}^{N_p} \delta(\mathbf{r}_{M,n} - \mathbf{r}_{M,n}^{(k)})$ , the message for  $\overleftarrow{\mu}(\mathbf{z}_{M,n})$

is computed as<sup>11</sup>

$$\begin{aligned} \overleftarrow{\mu}(\mathbf{z}_{M,n}) &= \frac{1}{N_p} \int \mathcal{N}(\mathbf{r}_{M,n}; \sqrt{G} \mathbf{z}_{M,n}, \Sigma) \\ &\quad \times \sum_{k=1}^{N_p} \delta(\mathbf{r}_{M,n} - \mathbf{r}_{M,n}^{(k)}) d\mathbf{r}_{M,n} \\ &= \frac{1}{N_p} \sum_{k=1}^{N_p} \mathcal{N}(\mathbf{r}_{M,n}^{(k)}; \sqrt{G} \mathbf{z}_{M,n}, \Sigma), \end{aligned} \quad (24)$$

which can be interpreted as a Gaussian mixture in  $\mathbf{z}_{M,n}$ , comprising  $N_p$  mixture components with means  $\mathbf{r}_{M,n}^{(k)}/\sqrt{G}$  and equal covariance  $\Sigma/G$ . A PR of this mixture can be obtained by drawing one sample from each mixture component,  $\mathbf{z}_{M,n}^{(k)} \sim \mathcal{N}(\mathbf{z}_{M,n}; \mathbf{r}_{M,n}^{(k)}/\sqrt{G}, \Sigma/G)$ , and thus

$$\text{PR}\{\overleftarrow{\mu} \mathbf{z}_{M,n}\} = \{\mathbf{z}_{M,n}^{(1)}, \dots, \mathbf{z}_{M,n}^{(N_p)}\}. \quad (25)$$

As mentioned in Section III-B, variables have an inherent type. When  $\text{type}(\mathbf{Z}_{M,n})$  is such that it is bandlimited to a bandwidth  $W$ , the noise samples should be added with a variance proportional to  $\min(W, B)$ .

- *Message for the dispersion (linear) block:* Starting from  $\text{PR}\{\overleftarrow{\mu} \mathbf{v}_{m+1,n}\} = \{\mathbf{v}_{m+1,n}^{(k)}\}_{k=1}^{N_p}$ , our aim is to determine  $\text{PR}\{\overleftarrow{\mu} \mathbf{u}_{m,n}\}$ . From Sec. II,  $\mathbf{z}_{M,n} = \mathbf{v}_{m+1,n}$  when  $m = M$ , so  $\text{PR}\{\overleftarrow{\mu} \mathbf{z}_{M,n}\} = \text{PR}\{\overleftarrow{\mu} \mathbf{v}_{m+1,n}\}$ . Following (16)–(17), the messages  $\overleftarrow{\mu} \mathbf{v}_{m+1,n}(\mathbf{v}_{m+1,n})$  and  $\overleftarrow{\mu} \mathbf{u}_{m,n}(\mathbf{u}_{m,n})$  are related through a transformation of the arguments. Following the probability mass function interpretation of PRs,  $\text{PR}\{\overleftarrow{\mu} \mathbf{u}_{m,n}\} = \{\mathbf{u}_{m,n}^{(k)}\}_{k=1}^{N_p}$  in which  $\mathbf{u}_{m,n}^{(k)} = \phi^{-1}(\mathbf{v}_{m+1,n}^{(k)})$ , where  $\phi^{-1}(\cdot)$  is the inverse filter of  $\mathbf{A}\mathbf{h}(t, \Delta z)$ . For computational efficiency, the message passing can be performed in the frequency domain, by introducing additional variables corresponding to the Fourier transforms of  $\mathbf{u}_{m,n}^{(k)}$  and  $\mathbf{v}_{m+1,n}^{(k)}$ . These additional variables lead to additional factors and edges in the FG, as depicted in Fig. 4. The additional factors are of the form  $p(\tilde{\mathbf{u}}_{m,n} | \mathbf{u}_{m,n}) = \delta(\tilde{\mathbf{u}}_{m,n} - \text{FFT}(\mathbf{u}_{m,n}))$ ,  $p(\tilde{\mathbf{u}}_{m+1,n} | \tilde{\mathbf{u}}_{m,n}) = \delta(\tilde{\mathbf{u}}_{m+1,n} - \tilde{\mathbf{u}}_{m,n} \mathbf{A}\mathbf{h}(\omega, \Delta z))$ , and  $p(\mathbf{v}_{m+1,n} | \tilde{\mathbf{u}}_{m+1,n}) = \delta(\mathbf{v}_{m+1,n} - \text{IFFT}(\tilde{\mathbf{u}}_{m+1,n}))$ . As  $\text{FFT}(\cdot)$  and  $\text{IFFT}(\cdot)$  are bijective operators on vectors, the PRs are simply transformations of the particles:

$$\text{PR}\{\overleftarrow{\mu} \tilde{\mathbf{u}}_{m+1,n}\} = \{\text{FFT}(\mathbf{v}_{m+1,n}^{(k)})\}_{k=1}^{N_p}, \quad (26)$$

$$\text{PR}\{\overleftarrow{\mu} \tilde{\mathbf{u}}_{m,n}\} = \{\mathbf{A}^{-1} \tilde{\mathbf{h}}^{-1}(\omega, \Delta z) \tilde{\mathbf{u}}_{m+1,n}^{(k)}\}_{k=1}^{N_p}, \quad (27)$$

$$\text{PR}\{\overleftarrow{\mu} \mathbf{u}_{m,n}\} = \{\text{IFFT}(\tilde{\mathbf{u}}_{m,n}^{(k)})\}_{k=1}^{N_p}. \quad (28)$$

- *Message for the nonlinear block:* To get the message  $\text{PR}\{\overleftarrow{\mu} \mathbf{v}_{m,n}\}$  from  $\text{PR}\{\overleftarrow{\mu} \mathbf{u}_{m,n}\} = \{\mathbf{u}_{m,n}^{(k)}\}_{k=1}^{N_p}$ , probability mass function interpretation of the PRs is used. We thus find that  $\text{PR}\{\overleftarrow{\mu} \mathbf{v}_{m,n}\} = \{\mathbf{v}_{m,n}^{(k)}\}_{k=1}^{N_p}$ , where  $\mathbf{v}_{m,n}^{(k)} = \mathbf{u}_{m,n}^{(k)} \exp(-j\gamma\Delta z |\mathbf{u}_{m,n}^{(k)}|^2)$ . In other words, we de-rotate each of the  $N_p$  sampled waveforms  $\mathbf{u}_{m,n}^{(k)}$ , where

<sup>10</sup>Variations exist where the values are weighted as obtained by importance sampling. We do not apply weighting in this context.

<sup>11</sup>Note that for  $n = N$ , the PR is initialized with  $N_p$  copies of the received signal  $\mathbf{r}$ .

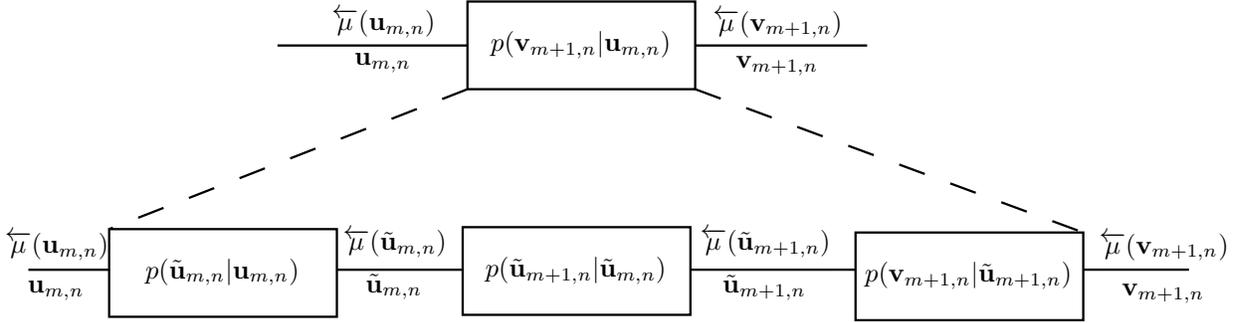


Fig. 4. Factor graph corresponding to the linear block of Fig. 1(b) with  $p(\tilde{\mathbf{u}}_{m,n}|\mathbf{u}_{m,n}) = \delta(\tilde{\mathbf{u}}_{m,n} - \text{FFT}(\mathbf{u}_{m,n}))$ ,  $p(\tilde{\mathbf{u}}_{m+1,n}|\tilde{\mathbf{u}}_{m,n}) = \delta(\tilde{\mathbf{u}}_{m+1,n} - \tilde{\mathbf{u}}_{m,n} \cdot \mathbf{A}\tilde{\mathbf{h}}(\omega, \Delta z))$ , and  $p(\mathbf{v}_{m+1,n}|\tilde{\mathbf{u}}_{m+1,n}) = \delta(\mathbf{v}_{m+1,n} - \text{IFFT}(\tilde{\mathbf{u}}_{m+1,n}))$ .

the rotation per waveform is proportional to the instantaneous power of the respective time-domain samples.

- *Posterior computation:* The message  $\text{PR}\{\tilde{\mu}_{\mathbf{v}_{1,1}}\}$  can be computed once the messages for all the segments and spans are found. Multiplication of  $\tilde{\mu}(\mathbf{v}_{1,1})$  and  $\tilde{\mu}(\mathbf{v}_{1,1})$  is required to compute the marginal distribution  $p(\mathbf{v}_{1,1}|\mathbf{r})$ . The message  $\tilde{\mu}(\mathbf{v}_{1,1})$  is a probability mass function over a high-dimensional random variable, and  $\tilde{\mu}(\mathbf{v}_{1,1})$  is available only in particle format. As the probability that any of the particles coincides with any of the values where  $\tilde{\mu}(\mathbf{v}_{1,1})$  has mass is zero almost surely,  $\tilde{\mu}(\mathbf{v}_{1,1})$  need to be smoothened to obtain a PDF that can be evaluated for any  $\mathbf{v}_{1,1}$  where  $\tilde{\mu}(\mathbf{v}_{1,1})$  has mass. Smoothing can be done, for example, using a Gaussian process [32] or through kernel density estimation [33], [34, Ch. 6]. Following this smoothing process,  $p(\mathbf{v}_{1,1}|\mathbf{r})$  can be evaluated for any  $\mathbf{v}_{1,1}$  of the form  $\mathbf{v}_{1,1} = \mathbf{s}_k * \mathbf{g}$ . This immediately yields the corresponding  $p(\mathbf{s}_k|\mathbf{r})$  and using this posterior distribution, the optimization is done according to  $\hat{\mathbf{s}} = \arg \max_{\mathbf{s}_k} p(\mathbf{s}_k|\mathbf{r})$ . Note that this process still faces the exponential complexity problem, as there are  $|\Omega|^K$  combinations of  $\mathbf{v}_{1,1}$  to consider. The posterior computation employed in this paper is described in the next section.

### B. Symbol-by-Symbol Detector

The high computational complexity of the sequence-based MAP detector described above can be avoided by considering a symbol-by-symbol detector. We assume  $\text{PR}\{\tilde{\mu}_{\mathbf{v}_{1,1}}\}$  as  $N_p$  outputs from a virtual channel, where each of the outputs is equal to  $\mathbf{s} * \mathbf{g}$ , affected by zero-mean additive noise with an unknown distribution. In order to maximize the signal-to-noise ratio, we employ a filter matched to the pulse shape (MF) followed by a sampler [35, Ch. 4], [36, Ch. 10], which can be implemented as follows. Each particle  $\mathbf{v}_{1,1}^{(k)}$  from  $\text{PR}\{\tilde{\mu}_{\mathbf{v}_{1,1}}\}$  is passed through the MF and sampled at the symbol rate at the optimal sampling times.<sup>12</sup> For each symbol  $\mathbf{s}[n]$ ,  $N_p$  samples are collected into a vector  $\{\mathbf{r}_{\text{MF}}^{(k)}[n]\}_{k=1}^{N_p}$ , where  $\mathbf{r}_{\text{MF}}[n]$  is the MF output, for the  $n$ th symbol slot. We note that now each  $\mathbf{r}_{\text{MF}}^{(k)}[n]$  is of the same dimensionality as  $\mathbf{s}[n]$  (i.e., two-dimensional real for single-polarization transmission

and four-dimensional real for dual-polarization transmission). Note that now  $\{\mathbf{r}_{\text{MF}}^{(k)}[n]\}_{k=1}^{N_p}$  can be considered as a PR of  $p(\mathbf{r}_{\text{MF}}[n]|\mathbf{s}[n])$ , where the MF output for the  $n$ th symbol slot  $\mathbf{r}_{\text{MF}}[n]$  is of the form  $\mathbf{r}_{\text{MF}}[n] = \mathbf{s}[n] + \mathbf{n}[n]$ . Here  $\mathbf{n}[n]$  is a (two- or four-dimensional) random variable, which is approximated as Gaussian with mean zero and empirical covariance matrix

$$\Sigma[n] = \frac{1}{N_p} \sum_{k=1}^{N_p} (\mathbf{r}_{\text{MF}}^{(k)}[n] - \underline{\boldsymbol{\mu}}[n])(\mathbf{r}_{\text{MF}}^{(k)}[n] - \underline{\boldsymbol{\mu}}[n])^T, \quad (29)$$

in which  $\underline{\boldsymbol{\mu}}[n] = 1/N_p \sum_k \mathbf{r}_{\text{MF}}^{(k)}[n]$ . Finally, under the assumption of uniform a priori  $p(\mathbf{s}[n])$

$$p(\mathbf{s}[n]|\mathbf{r}) \approx \exp\left(-\frac{1}{2}(\mathbf{s}[n] - \underline{\boldsymbol{\mu}}[n])^T \Sigma[n]^{-1}(\mathbf{s}[n] - \underline{\boldsymbol{\mu}}[n])\right). \quad (30)$$

The complexity of the above symbol-by-symbol detector scales as  $K \times |\Omega|$  instead of  $|\Omega|^K$  for the optimal sequence detector, and it thus amenable to an implementation. Additional performance gains may be reaped by performing joint detection over successive symbols, at a cost in terms of complexity.

### C. Remarks

- *Connection to DBP:* When there is no noise in the transmission system, PRs collapse to (high-dimensional) points, since noise samples in (24)–(25) will have variance zero. The proposed near-MAP detector is equivalent to the conventional DBP method. Due to this relation, we coin our method stochastic DBP.
- *Interaction with decoder:* The marginal posteriors  $p(\mathbf{s}[n]|\mathbf{r})$  can be used to make decisions on the transmitted symbols. They can also be used as soft information to send to a soft demodulator and a soft decoder [37]. This makes SDBP more suitable than DBP when soft decoding is employed.
- *Weakly nonlinear regime:* When the nonlinearity is negligible, the transmission system can be transformed to a concatenation of an all-pass filter, an attenuator, and an AWGN source. The optimal receiver in that case is well known and comprises an inverse all-pass filter and an MF, followed by symbol rate sampling. Both DBP

<sup>12</sup>As in [2]–[5], we assume perfect clock synchronization.

and SDBP implement this MF. Moreover, the covariance matrices  $\underline{\Sigma}[n]$  in (29) will be scaled diagonal matrices. When the nonlinearity increases, these covariance matrices will exhibit nondiagonal components, giving SDBP an advantage over DBP.

- *Strongly nonlinear regime:* When the nonlinearity is strong, the Gaussian assumptions made in (30) no longer holds. There are two ways to handle this situation. One way is to smooth  $\text{PR}\{\overline{\mu}_{\mathbf{v}_{1,1}}\}$  through a kernel density estimator [33], [34, Ch. 6], [38]. A second, easier way, is to pass  $\text{PR}\{\overline{\mu}_{\mathbf{v}_{1,1}}\}$  through an MF followed by a sampler and then perform the smoothing on the symbol-rate samples using kernel density estimation. This allows  $p(\underline{s}[n]|\mathbf{r})$  of (30) to take on any shape.<sup>13</sup>

## V. ILLUSTRATION OF SDBP FOR A SIMPLIFIED MODEL

To illustrate the operation of SDBP and its relation to DBP, we consider a simple system, comprising one SMF followed by an FBG and an EDFA, shown in Fig. 5. The top row of waveforms depicts the waveforms corresponding to the intermediate states  $\mathbf{x}$ ,  $\mathbf{y}$ ,  $\mathbf{z}$  as well as the received waveform  $\mathbf{r}$ . Note that the  $\mathbf{y}$  is affected by dispersion and nonlinearity and that the compound effect is only partially compensated in  $\mathbf{z}$ . The bottom row shows the  $N_p$  particles associated with each of these hidden waveforms, depicted in gray and the DBP waveform is shown in thick blue lines. In SDBP, the uncertainty is propagated starting from the received waveform  $\mathbf{r}$ , backwards all the way to the transmitted sequence  $\mathbf{s}$ . The  $N_p$  waveforms are propagated through the inverse of each of the blocks in the forward system. As an example, in Fig. 5 we start with the known received waveform  $\mathbf{r}$  (which exhibits no uncertainty) and pass it through the inverse of the EDFA block. Since the EDFA block adds AWGN noise, the waveforms representing the uncertainty in  $\mathbf{z}$  are generated from  $\mathbf{r}$  by adding suitable AWGN processes. The process is continued in the next block and the effects of the FBG are undone to get the uncertainty of the hidden waveform  $\mathbf{y}$ . In the same way, by passing  $N_p$  waveforms through the inverse SSFM description of SMF, we get the  $N_p$  waveforms associated with  $\mathbf{x}$ .

## VI. NUMERICAL SIMULATIONS AND DISCUSSIONS

DBP and the proposed SDBP detector with  $N_p = 500$  particles<sup>14</sup> are applied to a system with SMF and inline dispersion compensation done either with DCFs or with FBGs.<sup>15</sup> The parameters used for the SMF and DCF are given in Table I. We set number of symbols transmitted in one block to  $K = 2048$ . The span length used for SMF is  $L_{\text{SMF}} = 80$  km and the length for the DCF,  $L_{\text{DCF}}$ , is calculated such that  $D_{\text{SMF}}L_{\text{SMF}} + D_{\text{DCF}}L_{\text{DCF}} = 0$ , where  $D$  is the dispersion

<sup>13</sup>A non-exhaustive simulation analysis indicated that the use of kernel density estimation did not yield significant performance gains.

<sup>14</sup>The system has been simulated with more than 500 particles, but no significant improvement was seen in the results. The number of particles can be reduced to some extent without performance loss, but this optimization is not performed in this paper. More details are given in Sec. VI-D1.

<sup>15</sup>We note that similar results hold for a setup when there are additional EDFA's between SMF and DCM, where the DCM can be either a DCF or an FBG.

TABLE I  
CHANNEL PARAMETERS

|                   | SMF | DCF  |
|-------------------|-----|------|
| $D$ (ps/nm/km)    | 16  | -120 |
| $\gamma$ (1/W/km) | 1.3 | 5.2  |
| $\alpha$ (dB/km)  | 0.2 | 0.6  |

parameter of the fiber. FBG with an insertion loss of 3 dB and perfect dispersion compensation for the preceding SMF is used. The EDFA noise figure is 5 dB. A root raised cosine pulse in the time domain is used with a roll-off factor of 0.25 and truncation length of 16 symbol periods. The simulations are performed for dual-polarization transmission with either 16-QAM or QPSK as modulation format. The receiver is assumed to have perfect knowledge of the polarization state, as well as the carrier phase and the symbol timing. ASE noise with bandwidth equal to the used sampling frequency, which is twice the baud rate, is added in each span. The symbol error rate (SER) is used as a performance metric.

The SSFM is simulated with a segment length [39] of  $\Delta = (\epsilon L_N L_D^2)^{1/3}$ , where  $\epsilon = 10^{-4}$ ,  $L_N = 1/(\gamma P)$  is the nonlinear length,  $L_D = T^2 2\pi c / (|D|\lambda^2)$  is the dispersion length,  $\lambda$  is the wavelength,  $c$  is the speed of the light, and  $P$  is the input power to each fiber span. The number of segments per span is  $M = \lceil L/\Delta \rceil$ , where  $\lceil p \rceil$  is the smallest integer not less than  $p$ . We used the same segment length in the backward and forward system.

### A. SER and Reach Analysis

Fig. 6(a) shows the SER as a function of the input power for two different symbol rates (14 and 28 Gbaud) for a link with FBG dispersion management (called *FBG link* from here on) for DBP and SDBP. We observe that SDBP performs significantly better than DBP for both symbol rates. Using SDBP, a different optimal power is obtained and also for a given input power, lower SER is obtained in the nonlinear regime. This means using SDBP, the system is more tolerant to nonlinear effects and therefore we can have a longer reach.

In Fig. 6(b), we show the SER for a system with NDM link. We see that for both symbol rates, SDBP outperforms DBP to some extent, and the gains are smaller for the higher symbol rate. This indicates that for systems without inline dispersion compensation, the loss of performance when using DBP is smaller. The results presented in this paper corroborate the result from [21] by quantifying the gains in handling NSNI for both DM and NDM links.

In Fig. 7, we plot the SER as a function of the system reach for PM 16-QAM and PM QPSK and for different symbol rates. Each point in this plot is for an optimal input power, i.e., for different configurations, the input power corresponding to the lowest SER is selected. In Fig. 7(a), for PM 16-QAM with FBG<sup>16</sup> at 7 Gbaud, we see that there is around 17% increase in the system reach for SDBP compared to DBP. This gain remains more or less the same up to 28 Gbaud and then

<sup>16</sup>Results are not shown explicitly for the system with PM 16-QAM and DCF as DCM, as the results were similar to the system with FBG as DCM.

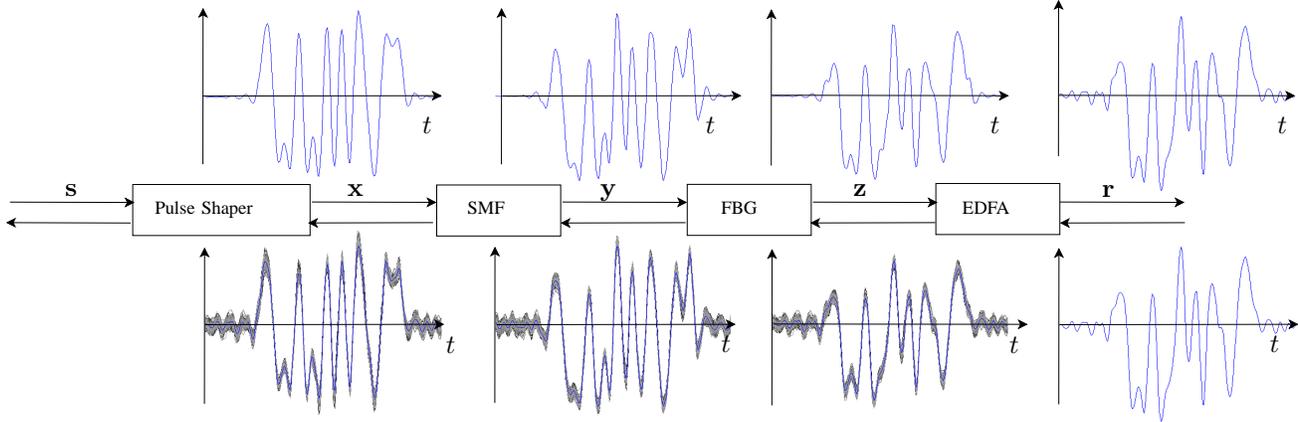


Fig. 5. A simplified system model with the unknown symbol sequence  $s$  passing forward through different blocks of the channel with  $N = 1$  spans, and the received waveform  $r$ . The top row depicts the waveforms corresponding to the hidden states  $x, y, z$  as well as the received waveform  $r$ . In the bottom row, collections of waveforms representing the uncertainty of the hidden states are shown in gray and the DBP waveform is shown in thick blue lines.

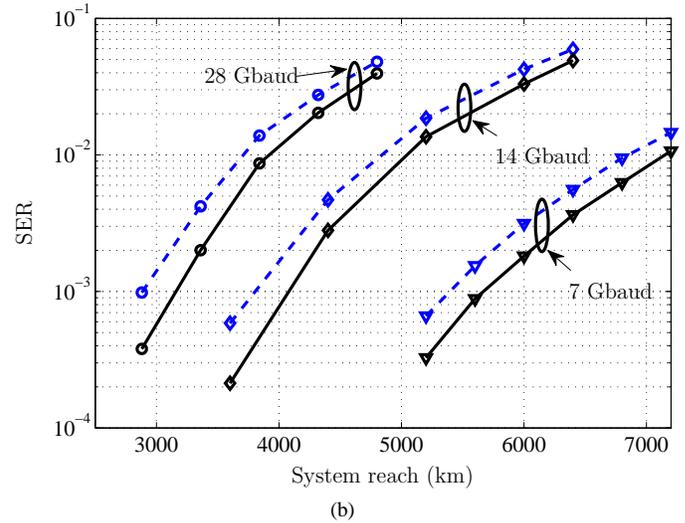
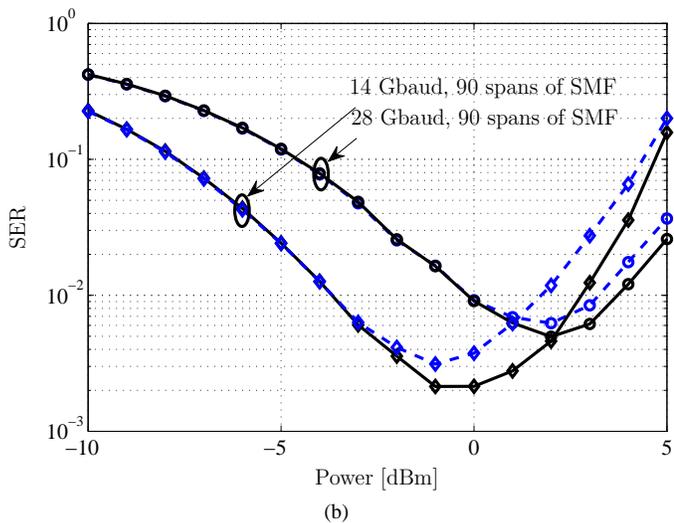
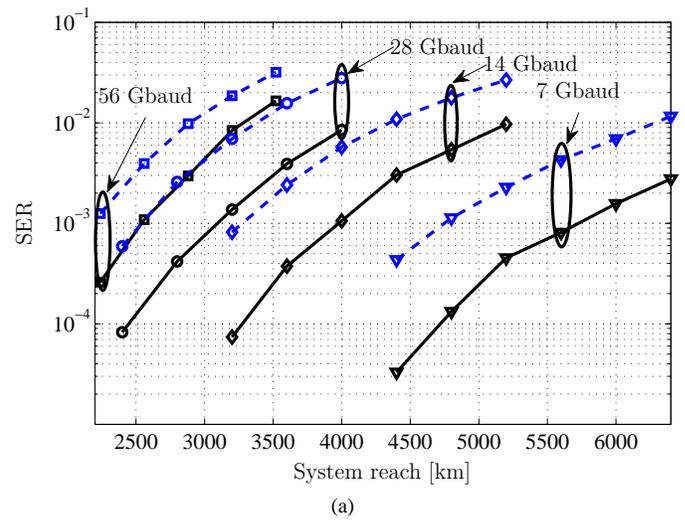
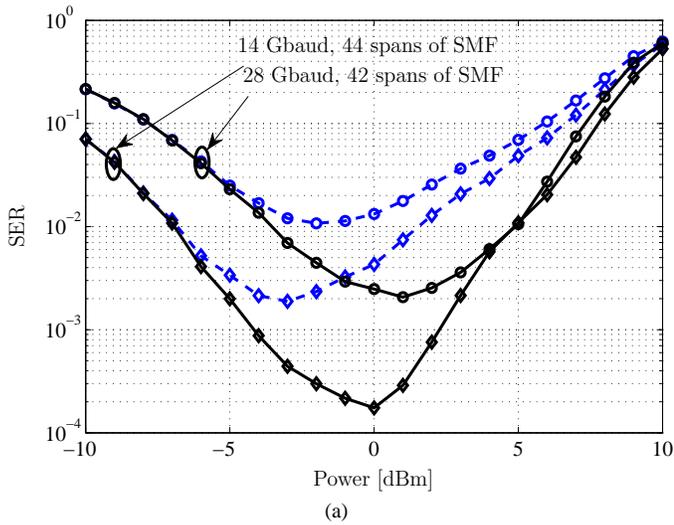


Fig. 6. SER as a function of input power for DBP (dashed, blue) and SDBP (solid, black) for PM 16-QAM at 14 and 28 Gbaud for the (a) FBG link (b) NDM link.

Fig. 7. SER as a function of system reach for DBP (dashed, blue) and SDBP (solid, black) for different symbol rates for (a) PM 16-QAM and FBG (b) PM QPSK and DCF.

decreases slightly at 56 Gbaud (at a SER of  $10^{-2}$ , we see 17% increase for 14 and 28 Gbaud, and 13% at 56 Gbaud). In Fig. 7(b), the system reach for PM QPSK is presented where DCF is used as a DCM. By comparing Fig. 7(a) and Fig. 7(b), we observe that for a given symbol rate, the gains of SDBP over DBP are higher for PM 16-QAM compared to PM QPSK.

The gains of SDBP can be explained as follows. The larger the deviation of the particle clouds, given by  $\{\mathbf{x}_{\text{MF}}^{(k)}[n]\}_{k=1}^{N_p}$  (PR $\{\hat{\mu}_{\mathbf{v},1,1}\}$  passed through an MF and a sampler), from a circular symmetric Gaussian, the higher are the expected gains in SDBP compared to DBP. For a DM link, we have observed that the particle clouds are less circularly Gaussian and hence SDBP performs better than DBP.

### B. Estimated Distributions

PDFs associated with the PRs  $\{\mathbf{x}_{\text{MF}}^{(k)}[n]\}_{k=1}^{N_p}$  for different  $n$ , for 28 Gbaud, 44 spans of 80 km SMF each, 16-QAM, and FBG link are provided in Fig. 8. In blue, we show the PDF obtained on histograms using 5000 particles, while in red, we show the Gaussian approximation, determined from the computed sample mean and covariance. It can be seen that a multivariate Gaussian distribution is often a good approximation. It can also be seen in Fig. 8(b) that for some symbols (e.g., at constellation point  $3+j$ ), the histogram-based and Gaussian PDFs do not fit, which means that some other distribution than multivariate Gaussian is needed to achieve the optimal performance. Finally, we note that the figure is not symmetric, due to the effect of the preceding and succeeding symbols, as well as the specific noise realization.

### C. Influence of PMD on DBP and SDBP

In SDBP, ASE is the only non-deterministic impairment that is currently compensated for. However, other non-deterministic impairments, in particular PMD, may deteriorate the performance of both DBP and SDBP. This deterioration becomes significant when the total differential group delay (DGD) of the system approaches the symbol period [40], [41]. It was reported in [40] that the power penalty is small for realistic transmission distances and data rates, as the residual PMD can be compensated for in a post-DBP linear equalizer. To quantify the performance degradation of DBP and SDBP in the presence of PMD, we performed a simulation by including generalized wave-plates in each span between the SMF and the FBG. The wave-plate introduces a random polarization change and DGD. This PMD simulation is according to the hinge model [42], which assumes that PMD is introduced between each span of SMF. As a part of this, each span introduces independent phase rotations, but the amount of DGD between each span will be assumed to be constant and given by  $\tau = \Delta\tau\sqrt{8N/3\pi}$  [43] where  $\tau$  is the total DGD and  $\Delta\tau$  is the DGD in each wave-plate. Note that when  $\Delta\tau = 0$ , there is no PMD and the system reverts to the one in previous sections.

A full integration of PMD compensation with SDBP would constitute an entire research project in itself. Instead, two simple strategies are considered in the simulations: one where no PMD compensation is performed, and one where the

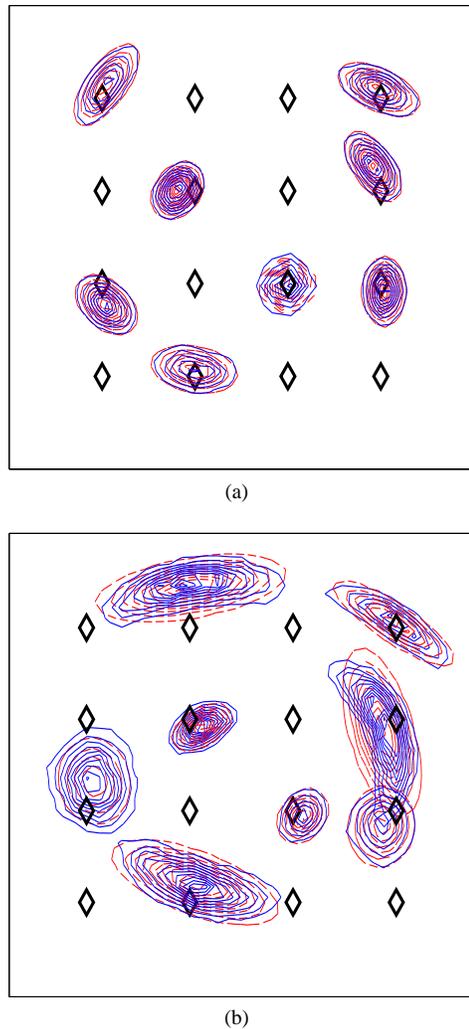


Fig. 8. PDF contours of particle representations,  $\{\mathbf{x}_{\text{MF}}^{(k)}[n]\}_{k=1}^{N_p}$  for different  $n$ , given a specific history for each symbol, for SMF of 80 km, 16-QAM modulation, FBG link at power of (a) 28 Gbaud, 42 spans at 0 dBm and (b) 28 Gbaud, 42 spans at 5 dBm.

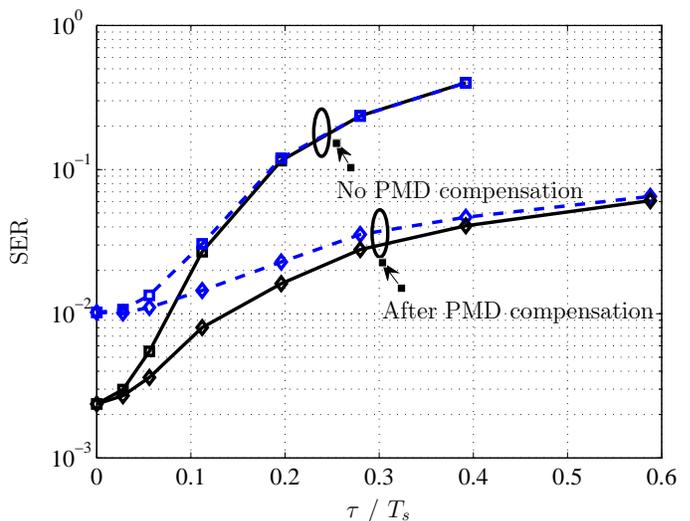


Fig. 9. SER as a function of  $\tau/T_s$ , the normalized DGD of the link, for 28 Gbaud, 42 spans of SMF with 80 km each, PM 16-QAM, and an FBG link. Dashed blue and solid black curves correspond to DBP and SDBP respectively.

receiver is assumed to operate under perfect knowledge<sup>17</sup> of the rotation matrices and the amount of DGD. In the latter case, the PMD is compensated after (S)DBP.

In Fig. 9, the SER as a function of normalized total DGD is presented for a link with 28 Gbaud, 42 spans of 80 km SMF, FBG as DCM, and 16-QAM. For each DGD value, the system is simulated with 40 realizations and an average SER is obtained in each case. Each point in this curve is simulated for the optimal power corresponding to that DGD value. Diamonds and squares represent scenarios where the PMD is compensated and not compensated for, respectively. Performance deterioration can be seen for both DBP and SDBP with increasing  $\Delta\tau$  but SDBP maintains a performance gain over DBP which decreases as DGD value increases.

#### D. Complexity Analysis of SDBP

As pointed out earlier, the objective of this paper is not to develop a low-complexity detector suitable for immediate implementation, but rather to show that DBP is not optimal and to derive an optimal detector for a single-channel fiber-optical link. For the sake of completeness, in this section we quantify the complexity of SDBP.

The complexity of SDBP scales as  $(N_p \times N \times M \times C_{\text{DBP},M} + C_{\text{dec}})$ , where  $C_{\text{DBP},M}$  is the complexity of the DBP algorithm per segment of a fiber span, and  $C_{\text{dec}}$  is the complexity associated with detection, including computation of (29)–(30). It is readily verified that  $C_{\text{dec}}$  scales as  $K \times N_p + K \times |\Omega|$ . The complexity of SDBP can be reduced by reducing  $N_p$ ,  $M$ ,  $C_{\text{DBP},M}$ . Below, we present the performance as a function of  $N_p$  and  $M$ .

##### 1) Performance as a function of number of particles:

In Fig. 10, the SER as a function of number of particles,  $N_p$ , is shown for 28 Gbaud, 42 spans of 80 km SMF, FBG link, and 16-QAM at input power of  $-1$  dBm. DBP is also shown for reference. One can see that when  $N_p < 10$ , DBP performs better than SDBP as 10 particles are not sufficient to accurately represent the statistical distribution of the signal. When  $N_p > 100$ , the performance is almost constant irrespective of the increase in the number of particles. The number of particles needed to approximate the PDFs capturing the uncertainty of variables will vary depending on the chosen set of parameters and hence Fig. 10 will be different for another set of parameters. However, we made sure that for the scenarios considered,  $N_p = 500$  was more than sufficient, while no further optimization of  $N_p$  was performed.

2) *Performance as a function of number of segments per span of SMF:* In Fig. 11, the SER as a function of the number of segments per span,  $M$ , in DBP and SDBP is shown for 28 Gbaud, 42 spans of 80 km each SMF, 16-QAM, and an FBG link. For these parameters, SDBP and DBP have their optimal input powers at 1 and  $-2$  dBm respectively. In the forward simulation using SSFM,  $M$  was set to 50, which is significantly larger than  $\lceil L/\Delta \rceil$ . It can be seen that  $M$  can be reduced in SDBP without sacrificing the gains obtained over DBP.

<sup>17</sup>Several algorithms such as the constant modulus algorithm [44] exist to compensate for PMD.

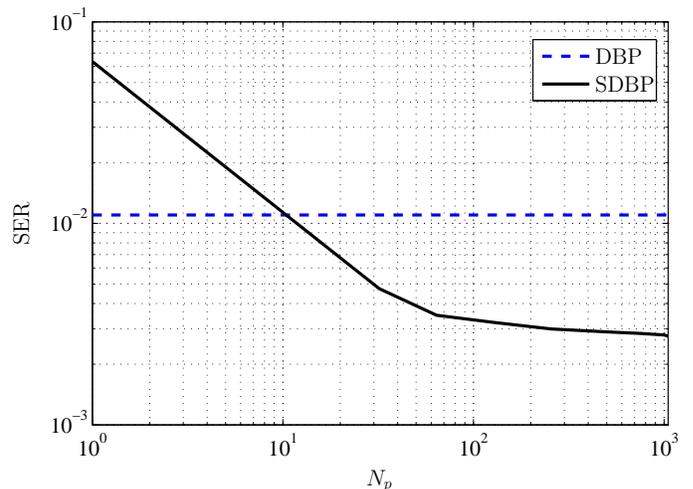


Fig. 10. SER as a function of number of particles,  $N_p$ , for 28 Gbaud, 42 spans of 80 km SMF, FBG link, and 16-QAM at an input power of  $-1$  dBm.

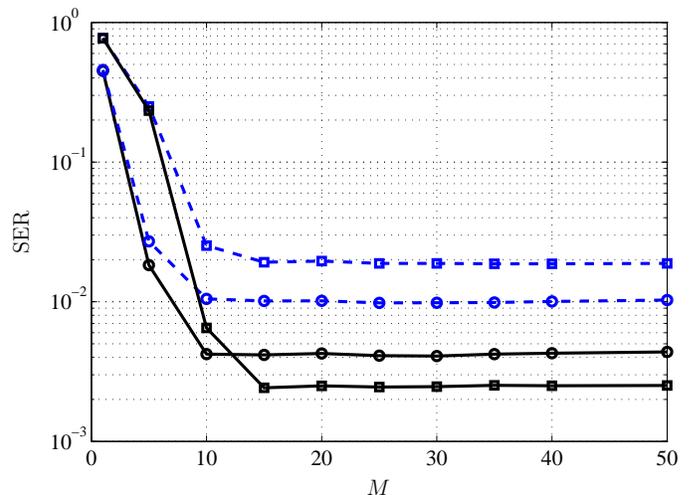


Fig. 11. SER as a function of number of segments per span,  $M$ , in DBP (dashed blue) and SDBP (solid black) for 28 Gbaud, 42 spans of 80 km each SMF, 16-QAM, and FBG link at input powers of  $-2$  dBm (circles) and  $1$  dBm (squares).

## VII. CONCLUSIONS

In this paper, we derived a MAP detector and provided an implementation of a near-MAP detector that compensates for not only deterministic linear and nonlinear effects but also accounts for the noise from the optical amplifiers. We have shown that by taking signal statistics of the end-to-end fiber channel into account, SDBP, the proposed near-MAP detector, can outperform DBP. Through simulations, we show that for uncompensated links, SDBP and DBP exhibit similar performance. For dispersion-managed links, SDBP shows a significant increase in the system reach compared to DBP.

## ACKNOWLEDGMENT

The authors would like to thank Dr. L. Beygi, Dr. D. Sen, Prof. M. Karlsson, D. Marsella, and other members of FORCE for their helpful comments and many helpful discussions.

## REFERENCES

- [1] Cisco, "VNI forecast highlights." [Online]. Available: [http://www.cisco.com/web/solutions/sp/vni/vni\\_forecast\\_highlights/index.htm](http://www.cisco.com/web/solutions/sp/vni/vni_forecast_highlights/index.htm)
- [2] E. M. Ip and J. M. Kahn, "Fiber impairment compensation using coherent detection and digital signal processing," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 502–519, Feb. 2010.
- [3] X. Li, X. Chen, G. Goldfarb, E. Mateo, I. Kim, F. Yaman, and G. Li, "Electronic post-compensation of WDM transmission impairments using coherent detection and digital signal processing," *Optics Express*, vol. 16, no. 2, pp. 880–888, Jan. 2008.
- [4] D. Rafique, M. Mussolin, M. Forzati, J. Mårtensson, M. N. Chugtai, and A. D. Ellis, "Compensation of intra-channel nonlinear fibre impairments using simplified digital back-propagation algorithm," *Optics Express*, vol. 19, no. 10, pp. 9453–9460, May 2011.
- [5] T. Hoshida, L. Dou, W. Yan, L. Li, and Z. Tao, "Advanced and feasible signal processing algorithm for nonlinear mitigation," in *Proc. Optical Fiber Communication Conference (OFC)*, 2013, p. OTh3C.3.
- [6] T. Koike-Akino, C. Duan, K. Parsons, K. Kojima, T. Yoshida, T. Sugihara, and T. Mizuochi, "High-order statistical equalizer for nonlinearity compensation in dispersion-managed coherent optical communications," *Optics Express*, vol. 20, no. 14, p. 15769, Jun. 2012.
- [7] E. Ip, E. Mateo, and T. Wang, "Reduced-complexity nonlinear compensation based on equivalent-span digital backpropagation," in *International Conference on Optical Internet (COIN)*, 2012, pp. 28–29.
- [8] W. Yan, Z. Tao, L. Dou, and L. Li, "Low complexity digital perturbation back-propagation," in *Proc. European Conference on Optical Communication (ECOC)*, 2011, p. Tu.3.A.2.
- [9] L. B. Du and A. J. Lowery, "Improved single channel backpropagation for intra-channel fiber nonlinearity compensation in long-haul optical communication systems," *Optics Express*, vol. 18, no. 16, pp. 17075–17088, Aug. 2010.
- [10] A. Ghazisaeidi and L. A. Rusch, "On the efficiency of digital back-propagation for mitigating SOA-induced nonlinear impairments," *J. Lightw. Technol.*, vol. 29, no. 21, pp. 3331–3339, Nov. 2011.
- [11] D. Rafique, J. Zhao, and A. D. Ellis, "Impact of dispersion map management on the performance of back-propagation for nonlinear WDM transmissions," in *OptoElectronics and Communications Conference (OECC)*, 2010, pp. 760–761.
- [12] D. S. Millar, S. Makovejs, C. Behrens, S. Hellerbrand, R. I. Killay, P. Bayvel, and S. J. Savory, "Mitigation of fiber nonlinearity using a digital coherent receiver," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 5, pp. 1217–1226, Sep. 2010.
- [13] D. Rafique and A. D. Ellis, "Impact of longitudinal power budget in coherent transmission systems employing digital back-propagation," in *Proc. European Conference on Optical Communication (ECOC)*, 2011, p. We.10.P1.54.
- [14] A. Dochhan, R. Rath, C. Hebebrand, J. Leibrich, and W. Rosenkranz, "Evaluation of digital back-propagation performance dependent on step-size and ADC sampling rate for coherent NRZ- and RZ-DQPSK experimental data," in *Proc. European Conference on Optical Communication (ECOC)*, 2011, p. We.10.P1.79.
- [15] G. Gao, X. Chen, and W. Shieh, "Limitation of fiber nonlinearity compensation using digital back propagation in the presence of PMD," in *Proc. Optical Fiber Communication Conference (OFC)*, 2012, p. OM3A.5.
- [16] Y. Gao, J. H. Ke, J. C. Cartledge, K. P. Zhong, and S. S.-H. Yam, "Implication of parameter values on low-pass filter assisted digital back propagation for DP 16-QAM," *IEEE Photon. Technol. Lett.*, vol. 25, no. 10, pp. 917–920, May 2013.
- [17] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, 2010.
- [18] L. Beygi, N. V. Irukulapati, E. Agrell, P. Johannisson, M. Karlsson, H. Wymeersch, P. Serena, and A. Bononi, "On nonlinearly-induced noise in single-channel optical links with digital backpropagation," *Optics Express*, vol. 21, no. 22, pp. 26376–26386, Oct. 2013.
- [19] L. B. Du, D. Rafique, A. Napoli, B. Spinnler, A. D. Ellis, M. Kuschnerov, and A. J. Lowery, "Digital fiber nonlinearity compensation: toward 1-Tb/s transport," *IEEE Signal Processing Magazine*, vol. 31, no. 2, pp. 46–56, Mar. 2014.
- [20] A. Bononi, P. Serena, and N. Rossi, "Nonlinear signal-noise interactions in dispersion-managed links with various modulation formats," *Optical Fiber Technology*, vol. 16, no. 2, pp. 73–85, Mar. 2010.
- [21] D. Foursa, O. Sinkin, A. Lucero, J.-X. Cai, G. Mohs, and A. Pilipetskii, "Nonlinear interaction between signal and amplified spontaneous emission in coherent systems," in *Proc. Optical Fiber Communication Conference (OFC)*, 2013, p. JTh2A.35.
- [22] N. Jiang, Y. Gong, J. Karout, H. Wymeersch, P. Johannisson, M. Karlsson, E. Agrell, and P. Andrekson, "Stochastic backpropagation for coherent optical communications," in *Proc. European Conference on Optical Communication (ECOC)*, 2011, p. We.10.P1.81.
- [23] Y. Cai, "MAP detection for linear and nonlinear ISI mitigation in long-haul coherent detection systems," in *IEEE Photonics Society Summer Topicals*, vol. 1, Jul. 2010, pp. 42–43.
- [24] D. Marsella, M. Secondini, and E. Forestieri, "Maximum likelihood sequence detection for mitigating nonlinear effects," *J. Lightw. Technol.*, vol. 32, no. 5, pp. 908–916, Mar. 2014.
- [25] G. P. Agrawal, *Nonlinear Fiber Optics*, 4th ed. Academic Press, 2006.
- [26] S. J. Savory, "Digital filters for coherent optical receivers," *Optics Express*, vol. 16, no. 2, pp. 804–817, Jan. 2008.
- [27] G. P. Agrawal, *Fiber-Optic Communications Systems*, 3rd ed. Wiley, 2002.
- [28] H. Wymeersch, *Iterative Receiver Design*. Cambridge University Press, 2007.
- [29] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, 2001.
- [30] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 28–41, 2004.
- [31] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, 2002.
- [32] C. E. Rasmussen, *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- [33] S. J. Sheather, "Density estimation," *Statistical Science*, vol. 19, no. 4, pp. 588–597, Nov. 2004.
- [34] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. Springer, 2009.
- [35] S. Haykin, *Communication Systems*, 4th ed. Wiley, 2001.
- [36] A. Lapidoth, *A Foundation in Digital Communication*. Cambridge University Press, 2009.
- [37] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes," in *International Conference on Communications (ICC)*, vol. 2, 1993, pp. 1064–1070.
- [38] 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," in *International Conference on Communications (ICC)*, Jun. 2011, pp. 1–5.
- [39] Q. Zhang and M. I. Hayee, "Symmetrized split-step Fourier scheme to control global simulation accuracy in fiber-optic communication systems," *J. Lightw. Technol.*, vol. 26, no. 2, pp. 302–316, 2008.
- [40] E. Ip, "Nonlinear compensation using backpropagation for polarization-multiplexed transmission," *J. Lightw. Technol.*, vol. 28, no. 6, pp. 939–951, Mar. 2010.
- [41] G. Gao, X. Chen, and W. Shieh, "Influence of PMD on fiber nonlinearity compensation using digital back propagation," *Optics Express*, vol. 20, no. 13, pp. 14406–14418, Jun. 2012.
- [42] H. Kogelnik, P. J. Winzer, L. E. Nelson, R. M. Jopson, M. Boroditsky, and M. Brodsky, "First-order PMD outage for the hinge model," *IEEE Photon. Technol. Lett.*, vol. 17, no. 6, pp. 1208–1210, Jun. 2005.
- [43] M. Karlsson, "Probability density functions of the differential group delay in optical fiber communication systems," *J. Lightw. Technol.*, vol. 19, no. 3, pp. 324–331, Mar. 2001.
- [44] P. Johannisson, M. Sjödin, M. Karlsson, H. Wymeersch, E. Agrell, and P. A. Andrekson, "Modified constant modulus algorithm for polarization-switched QPSK," *Optics Express*, vol. 19, no. 8, pp. 7734–41, Apr. 2011.