Chalmers Publication Library

(article starts on next page)

Stochastic Digital Backpropagation with Residual Memory Compensation

Naga V. Irukulapati, *Student Member, IEEE*, Domenico Marsella, Pontus Johannisson,
Erik Agrell, *Senior Member, IEEE*, Marco Secondini, *Member, IEEE*, and Henk Wymeersch, *Member, IEEE*

(Invited Paper)

Abstract—**Stochastic digital backpropagation (SDBP) is an extension of digital backpropagation (DBP) and is based on the maximum a posteriori principle. SDBP takes into account noise from the optical amplifiers in addition to handling deterministic linear and nonlinear impairments. The decisions in SDBP are taken on a symbol-by-symbol (SBS) basis, ignoring any residual memory, which may be present due to nonoptimal processing in SDBP. In this paper, we extend SDBP to account for memory between symbols. In particular, two different methods are proposed: a Viterbi algorithm (VA) and a decision directed approach. Symbol error rate (SER) for memory-based SDBP is significantly lower than the previously proposed SBS-SDBP. For inline dispersion-managed links, the VA-SDBP has up to 10 and 14 times lower SER than DBP for QPSK and 16-QAM, respectively.**

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

I. INTRODUCTION

DIGITAL backpropagation (DBP) has been proposed as a universal technique for jointly compensating for the intra-channel linear and nonlinear impairments in the coherent fiber-optic system [1]–[5]. As a result, the DBP has been used to benchmark schemes proposed in the literature [6]–[10]. The assumed optimality of DBP has spurred intense research in low-complexity variations, including weighted DBP, perturbation DBP, and filtered DBP [10], [11]. While the focus of the current paper is on single-channel systems, for wavelength division multiplexing (WDM) systems, DBP is typically employed for the center channel, thereby accounting only for the intra-channel effects. Inter-channel nonlinear effects in WDM systems can be modeled by taking the advantage of the temporal correlations of the nonlinear phase noise using a time-varying system with inter-symbol interference (ISI) and thereby compensating for these inter-channel nonlinear effects [12]–[14]. 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 nondeterministic nonlinear effects, such as *nonlinear signal–noise interaction* (NSNI) between the transmitted signal and the amplified spontaneous emission (ASE) noise, limit the transmission performance of a fiber-optic system [10], [15], [16]. Studies on the impact of NSNI reveal that there is a significant penalty due to NSNI for inline optical dispersion-managed (DM) links, and the severity of the NSNI is dependent on modulation formats and the symbol rate used in the system [17], [18]. It is often argued that NSNI cannot be compensated for in digital signal processing (DSP) due to the nondeterministic nature of ASE noise [18] and as a result, none of the DBP methods account for NSNI. To deal with stochastic disturbances, Bayesian detection theory can be used to formulate maximum a posteriori probability (MAP) detectors, which are provably optimal in terms of minimizing the error probability. MAP detectors have been proposed for the discrete memoryless channel [19] assuming perfect chromatic dispersion (CD) compensation in a DM link, and a look-up table detector that can mitigate data-pattern-dependent nonlinear impairments [20]. A low-complexity Viterbi detector is suggested as an alternative or to complement DBP for combating fiber nonlinearities [21]. In [22], the stochastic digital backpropagation (SDBP) algorithm was proposed to compensate not only for deterministic linear and nonlinear effects, but also to account for the ASE noise. However, the decisions were taken on a symbol-by-symbol (SBS) basis after applying a matched filter. This approach was later shown to be suboptimal [23].

In this paper, we extend [22] to address the sub-optimality in SBS-SDBP by explicitly accounting for residual memory, which may be present due to nonoptimal processing in SDBP. In particular, we propose two different methods based on the Viterbi algorithm (VA) and a decision-directed (DD) approach. The VA approach is similar to [21], but does not rely on sending long training sequences to learn the VA branch metrics. The DD approach uses previously decoded symbols when taking the decisions for the current symbol and as a result is computationally less complex than the VA approach. Extensive simulation results indicate significant performance improvements over DBP and SBS-SDBP, in particular for DM links. While the proposed algorithm is computationally complex, we believe this receiver can serve as an inspiration to design low-complexity approaches that still significantly outperform DBP.

The remainder of this paper is organized as follows. We first describe the underlying mathematical framework on which SDBP is built, namely factor graphs (FGs) and message-passing algorithms, in Sec. II. In Sec. III, the system model is detailed. Sec. IV is devoted to the description of the VA and DD approaches, as well as how these are incorporated into SDBP framework. We present numerical results in Sec. V, followed by our conclusions in Sec. VI.

Notation: Lower case bold letters (e.g., $\mathbf{x}$) are used to denote vectors, including sequences of symbols and vector representations of continuous-time signals (e.g., through over-sampling). The transpose of the vector $\mathbf{v}$ is denoted by $\mathbf{v}^{\mathrm{T}}$. A multivariate Gaussian probability density function (PDF) of a real variable $\mathbf{r}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathcal{N}(\mathbf{r}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

II. FACTOR GRAPHS FOR RECEIVER DESIGN

Optimal receiver design for digital communications in terms of minimizing the error probability is based on a MAP criterion. However, MAP detectors can be computationally intractable (often with exponential complexity in the dimensionality of the unknown variable), except for certain communication systems. For this reason, much research effort has been devoted in developing *near-MAP detectors*, which can balance near-optimal performance with reasonable computational complexity. A practical framework that has emerged since the early 2000s as a general and automated way to develop near-MAP detectors is that of FGs [24], [25]. An FG is a graph that describes the statistical relation between the variable of interest (i.e., the unknown transmitted data) and the observation (i.e., the received waveform). By performing a message-passing algorithm on such an FG, it is possible to determine the MAP estimate, or an approximation thereof.

FGs have been widely used in wireless communication as they provide a methodology to design receivers in a systematic and near-automated way [26], [27]. Some applications include message-passing decoders for low-density parity-check (LDPC) and turbo codes [28], iterative demodulation and decoding for bit-interleaved coded modulation [29], joint equalization and decoding [30], channel estimation [31], timing synchronization [32], and phase-noise recovery [33]. It should be noted that while FGs generally lead to the most powerful known receiver algorithms with polynomial complexity, they are often too complex to be implemented as is. For that reason, FGs are often used as a first approach from which practical algorithms can be developed with lower complexity [30].

In the context of coherent fiber-optic communication, FGs have only seen limited utilization, mainly due to their high computational complexity. Examples include demodulation [34], decoding [35], equalization [36], and computation of information rates [37]. Nevertheless, FGs can serve as a good basis to develop low-complexity receivers. A key application is the design of a near-MAP detector in the nonlinear regime, as proposed in [22]. The resulting receiver, coined SDBP, showed significant performance gains over DBP. In SDBP, the posterior distribution is obtained by marginalizing the joint distribution of input, all intermediate, unobserved variables in
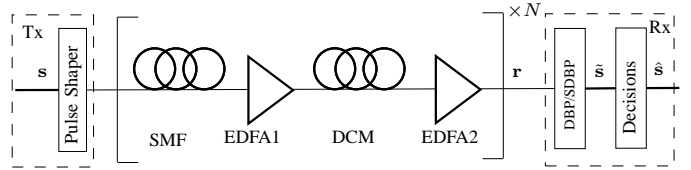

Fig. 1. A fiber link with N spans where each span consists of an SMF, a DCM module (for DM links), and EDFAs.

the channel, and the received signal, over the unobserved variables. The statistical relationship between all these variables can be described with a FG (in this case a Markov chain), and the marginalization is performed by message passing. However, the FG-based SDBP receiver proposed in [22] is not a true MAP receiver due to a number of heuristic design choices that were made: (i) a matched filter followed by symbol-rate sampling was employed, similar to DBP, which is suboptimal and may not give rise to sufficient statistics and may exhibit residual memory; (ii) decisions were made on a SBS basis, ignoring any residual memory; and (iii) for each SBS decision, a Gaussian approximation of messages was introduced. The first issue was addressed in [38], while a possible solution to the third issue was discussed in [23], considering alternative distributions to a Gaussian in Cartesian coordinates. An alternative approach would be to use a non-parametric approach with a kernel, where the kernel bandwidth is a free parameter that should be tuned [39]. The second issue will be addressed in this work.

III. SYSTEM MODEL

The system that will be considered is a single-channel fiber-optic system as shown in Fig. 1, comprising a dual-polarization transmitter block (Tx), including a pulse shaper, a fiber-optic link with N spans, and a receiver block (Rx) that implements a compensation algorithm followed by a decision unit. Each span of the fiber-optic link consists of a transmission fiber, which is a standard single-mode fiber (SMF) and an optional dispersion-compensating module (DCM) for DM links. In between fiber spans, there are erbium-doped fiber amplifiers (EDFAs) that compensate for the losses in the previous fiber. As indicated in Fig. 1, the transmitted data is denoted by $\mathbf{s}$, the decoded data by $\hat{\mathbf{s}}$, and the received signal by $\mathbf{r}$. The noise and gain of the EDFAs are resp., denoted by $\mathbf{w}_{ni}$ and G_i, where $i \in \{1, 2\}$ corresponds to EDFA1 and EDFA2.

A sequence of K four-dimensional symbols $\mathbf{s} = [s_1, s_2, \ldots, s_K]^{\mathrm{T}} \in \Omega^K$ is transmitted at a symbol rate $1/T_s$ with a pulse-shaping filter $g(t)$, where $\Omega \subset \mathbb{R}^4$ is the set of symbols in the four-dimensional constellation, consisting of in-phase and quadrature data from the x and y polarizations. The overall goal of the receiver is to optimally recover $\mathbf{s}$ from $\mathbf{r}$. While different optimality criteria can be considered, we aim to minimize the error probability, leading to a MAP receiver, in which the estimate of $\mathbf{s}$ is

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

where $p(\mathbf{s}|\mathbf{r})$ is the a posteriori probability distribution of $\mathbf{s}$ given the received signal $\mathbf{r}$. Note that in all derivations, we will consider all signals and vectors to be real.

IV. SDBP AND PROPOSED APPROACHES

As indicated in Sec. I, message passing on an FG relies on local dependencies. In a fiber, the lowest level of local deterministic dependencies that one can exploit are (i) the linear and nonlinear operation from the split-step Fourier method (SSFM) and (ii) the statistical dependency between input and output of the EDFAs, including ASE noise. Considering the signals after each linear and nonlinear step of each segment of each span of the SSFM [22, Fig. 1], and the signals after each EDFA, all as *unobserved variables*, one can factorize the joint distribution of the transmitted data $\mathbf{s}$, the received waveform $\mathbf{r}$, and these unobserved variables. The posterior distribution in (1) can thus be interpreted as a marginalization of the joint distribution $p(\mathbf{s}, \text{unobserved variables}|\mathbf{r})$. This marginalization from the joint distribution to $p(\mathbf{s}|\mathbf{r})$ is done using the framework of FGs and a message-passing algorithm called sum-product algorithm (SPA) [28].

Definition 1 (Particle representation of a distribution). *A list of particles (or samples)* $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$, ..., $\mathbf{x}^{(N_p)}$, *denoted by* $\{\mathbf{x}^{(n)}\}_{n=1}^{N_p}$, *form a particle representation of a distribution* $p(\mathbf{x})$ *when* $p(\mathbf{x}) \approx 1/N_p \sum_{n=1}^{N_p} \delta(\mathbf{x} - \mathbf{x}^{(n)})$.

Definition 2 (Distributions associated with particles). *With a list of particles* $\{\mathbf{x}^{(n)}\}_{n=1}^{N_p}$ *defined over* $\mathbb{R}^{4M}$, *we associate two distributions:* $q_c(\mathbf{x})$ *is a distribution (obtained, e.g., through a parametric approximation) defined over* $\mathbb{R}^{4M}$ *for which the particles form a sample representation, while* $q_d(\mathbf{x})$ *is a distribution defined only over* Ω^M, *with* $q_d(\mathbf{x}) \propto q_c(\mathbf{x})$.

A. Stochastic Digital Backpropagation

The main idea of SDBP is to marginalize out the unobserved variables through computing messages, which describe statistically (i.e., in the form of a distribution) the uncertainty of the corresponding variable. This allows us to obtain a description of $p(\mathbf{s}|\mathbf{r})$. The messages are computed backwards, starting with the received signal $\mathbf{r}$ at span N of the fiber-optic link of Fig. 1 until the transmitter is reached.

For the fiber-optic channel, closed-form expressions of the distributions are not possible to derive except for some specific scenarios, so the message/distribution is represented with a list of N_p particles. These particles are propagated at each stage of the fiber-optic link in Fig. 1 starting from $\mathbf{r}$ as described below.

We start with the known received waveform $\mathbf{r}$ in Fig. 1, which exhibits no uncertainty and is thus represented by N_p identical particles (line 2 in Algorithm 1). These particles are passed through the inverse of the EDFA2 block of the last span to get a collection of particles, which describe the uncertainty regarding the variable before EDFA2 (line 4 in Algorithm 1). The particles are then back propagated through the inverse of the SSFM of the DCM (line 5 in Algorithm 1), where $\text{SSFM}_1^{-1}(\mathbf{r}^{(n)})$ implements the inverse SSFM for entire fiber span. The particles are then back propagated through EDFA1

Algorithm 1 Pseudo-code for implementation of SDBP

1: **procedure** SDBP($\mathbf{r}$)
2:      $\mathbf{r}^{(n)} \leftarrow \mathbf{r} \quad \forall n$ ▷ create N_p replicas of $\mathbf{r}$
3:      **for** $i = N$ to 1 **do**   ▷ Iteration over spans
4:          $\mathbf{r}^{(n)} \leftarrow (\mathbf{r}^{(n)} + \mathbf{w}_{n2}^{(n)})/\sqrt{G_2} \quad \forall n$ ▷ EDFA2
5:          $\mathbf{r}^{(n)} \leftarrow \text{SSFM}_1^{-1}(\mathbf{r}^{(n)}) \quad \forall n$ ▷ DCM
6:          $\mathbf{r}^{(n)} \leftarrow (\mathbf{r}^{(n)} + \mathbf{w}_{n1}^{(n)})/\sqrt{G_1} \quad \forall n$ ▷ EDFA1
7:          $\mathbf{r}^{(n)} \leftarrow \text{SSFM}_2^{-1}(\mathbf{r}^{(n)}) \quad \forall n$ ▷ SMF
8:      **end for**
9:      $\tilde{\mathbf{s}}^{(n)} \leftarrow \text{MF}(\mathbf{r}^{(n)}) \quad \forall n$ ▷ MF followed by sampling at symbol rate
10: **end procedure**

(line 6 in Algorithm 1) and through the inverse of the SSFM of the SMF (line 7 in Algorithm 1). This process is repeated for all N spans. Note that when $N_p = 1$ and $\mathbf{w}_{n2}^{(n)} = \mathbf{w}_{n1}^{(n)} = 0$ for all n, these steps are identical to DBP.

As a final step (line 9 in Algorithm 1), SDBP must compute the message related to the transmitted data $\mathbf{s}$, based on the message describing the waveform after pulse shaping. A heuristic approach has been used in [22], where each particle waveform is passed through a matched filter (MF), matched to the pulse shape, and sampled at the symbol rate[1] at the optimal sampling times leading to N_p particles, $\{\tilde{\mathbf{s}}^{(n)}\}_{n=1}^{N_p}$, with $\tilde{\mathbf{s}}^{(n)} \in \mathbb{R}^{4K}$. The particles $\{\tilde{\mathbf{s}}^{(n)}\}_{n=1}^{N_p}$ can be viewed as samples from a distribution $q_c(\mathbf{s})$ (defined for $\mathbf{s} \in \mathbb{R}^{4K}$), for which $q_d(\mathbf{s})$ (defined for $\mathbf{s} \in \Omega^K$) provides an approximation of $p(\mathbf{s}|\mathbf{r})$. It is important to note that $q_d(\mathbf{s})$ is only an approximation of $p(\mathbf{s}|\mathbf{r})$ and need not be identical to $p(\mathbf{s}|\mathbf{r})$, as the use of a MF followed by sampling at the symbol rate is a heuristic. Hence, performing SBS decisions on the marginals of $q_c(\mathbf{s})$ as in [22] may not lead to optimal performance (in terms of minimizing the probability of error, either symbol-wise or sequence-wise). In fact, alternatives to a MF were explored in [38], indicating performance improvements. In this paper, we propose to exploit residual memory present due to nonoptimal processing, by making a decision regarding $\mathbf{s}$ based on the entire distribution $q_c(\mathbf{s})$, rather than its marginals, leading to the following detector

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s} \in \Omega^K} q_d(\mathbf{s}), \qquad (2)$$

where again $q_d(\mathbf{s}) \propto q_c(\mathbf{s})$. Solving (2) is hard for two reasons: (i) the number of possible sequences, Ω^K, is exponential in K, making the optimization infeasible for large values of K, and (ii) for any specific sequence in Ω^K, $q_d(\mathbf{s})$ is hard to determine since we only have particles $\{\tilde{\mathbf{s}}^{(n)}\}_{n=1}^{N_p}$, representing $q_c(\mathbf{s})$. In order to address the first issue, we impose a Markov structure onto $q_d(\mathbf{s})$. To solve the second issue, the set of particles is smoothed with a distribution, which will be discussed in Sec. IV-B. We now present two approaches that use variations of (2) to make decisions on $\mathbf{s}$.

---

[1]A matched filter maximizes the signal-to-noise ratio for a signal affected by AWGN noise [40, Ch. 10].

B. Viterbi Algorithm-Stochastic Digital Backpropagation

Assuming that $q_{\mathrm{d}}(\mathbf{s})$ follows a Markov structure with memory[2] $L \geq 0$, we define $\mathbf{x}_k = [s_{k-1}, s_{k-2}, \ldots, s_{k-L}]^{\mathrm{T}}$ and $\mathbf{y}_k = [s_k \ \mathbf{x}_k^{\mathrm{T}}]^{\mathrm{T}}$. Then $q_{\mathrm{d}}(\mathbf{s})$ can be factorized as

$$q_{\mathrm{d}}(\mathbf{s}) = \prod_{k=L+1}^{K} q_{\mathrm{d}}(s_k | s_{k-1}, s_{k-2}, \ldots, s_{k-L})$$
$$= \prod_{k=L+1}^{K} q_{\mathrm{d}}(s_k | \mathbf{x}_k). \quad (3)$$

Using Bayes' rule, $q_{\mathrm{d}}(s_k|\mathbf{x}_k)$ can be written as

$$q_{\mathrm{d}}(s_k|\mathbf{x}_k) = \frac{q_{\mathrm{d}}(s_k, \mathbf{x}_k)}{q_{\mathrm{d}}(\mathbf{x}_k)} = \frac{q_{\mathrm{d}}(\mathbf{y}_k)}{q_{\mathrm{d}}(\mathbf{x}_k)}. \quad (4)$$

Using (3) and (4), (2) can be rewritten as

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s} \in \Omega^K} \{\ln q_{\mathrm{d}}(\mathbf{s})\}$$
$$= \arg \max_{\mathbf{s} \in \Omega^K} \left\{ \sum_{k=1}^{K} [\ln q_{\mathrm{d}}(\mathbf{y}_k) - \ln q_{\mathrm{d}}(\mathbf{x}_k)] \right\}$$
$$= \arg \min_{\mathbf{s} \in \Omega^K} \sum_{k=1}^{K} \psi_k(s_k, \mathbf{x}_k), \quad (5)$$

where ψ_k is the branch metric and $\mathbf{x}_k$ is the state used in the VA.

The values of $q_{\mathrm{d}}(\mathbf{y}_k)$ and $q_{\mathrm{d}}(\mathbf{x}_k)$ can be computed by marginalizing $q_{\mathrm{d}}(\mathbf{s})$. However, we have access only to $q_{\mathrm{c}}(\mathbf{s})$ through the particles $\{\tilde{\mathbf{s}}^{(n)}\}_{n=1}^{N_p}$. Denoting the appropriate subsequences from $\tilde{\mathbf{s}}^{(n)}$ by $\mathbf{y}_k^{(n)}$ and $\mathbf{x}_k^{(n)}$, we find that $q_{\mathrm{c}}(\mathbf{y}_k) \approx 1/N_p \sum_{n=1}^{N_p} \delta(\mathbf{y}_k - \mathbf{y}_k^{(n)})$ and $q_{\mathrm{c}}(\mathbf{x}_k) \approx 1/N_p \sum_{n=1}^{N_p} \delta(\mathbf{x}_k - \mathbf{x}_k^{(n)})$. In order to evaluate $q_{\mathrm{d}}(\mathbf{y}_k)$ and $q_{\mathrm{d}}(\mathbf{x}_k)$, we can impose a parametric approximation for $q_{\mathrm{c}}(\mathbf{y}_k)$ and $q_{\mathrm{c}}(\mathbf{x}_k)$ for which the logarithm is easy to compute. The Gaussian distribution is such a parametric approximation.[3] Hence $q_{\mathrm{c}}(\mathbf{y}_k) = \mathcal{N}(\mathbf{y}_k; \boldsymbol{\mu}_k^y, \boldsymbol{\Sigma}_k^y)$ and $q_{\mathrm{c}}(\mathbf{x}_k) = \mathcal{N}(\mathbf{x}_k; \boldsymbol{\mu}_k^x, \boldsymbol{\Sigma}_k^x)$. The means $\boldsymbol{\mu}_k^y$, $\boldsymbol{\mu}_k^x$ and covariances $\boldsymbol{\Sigma}_k^y$, $\boldsymbol{\Sigma}_k^x$ are estimated as

$$\boldsymbol{\mu}_k^y = \frac{1}{N_p} \sum_{n=1}^{N_p} \mathbf{y}_k^{(n)}, \quad \boldsymbol{\mu}_k^x = \frac{1}{N_p} \sum_{n=1}^{N_p} \mathbf{x}_k^{(n)},$$
$$\boldsymbol{\Sigma}_k^y = \frac{1}{N_p - 1} \sum_{n=1}^{N_p} (\mathbf{y}_k^{(n)} - \boldsymbol{\mu}_k^y)(\mathbf{y}_k^{(n)} - \boldsymbol{\mu}_k^y)^{\mathrm{T}},$$
$$\boldsymbol{\Sigma}_k^x = \frac{1}{N_p - 1} \sum_{n=1}^{N_p} (\mathbf{x}_k^{(n)} - \boldsymbol{\mu}_k^x)(\mathbf{x}_k^{(n)} - \boldsymbol{\mu}_k^x)^{\mathrm{T}}. \quad (6)$$

The factors in (4) can be written as

$$q_{\mathrm{d}}(\mathbf{y}_k) \propto \exp \left\{ -\frac{1}{2} (\mathbf{y}_k - \boldsymbol{\mu}_k^y)^{\mathrm{T}} (\boldsymbol{\Sigma}_k^y)^{-1} (\mathbf{y}_k - \boldsymbol{\mu}_k^y) \right\}, \quad (7)$$
$$q_{\mathrm{d}}(\mathbf{x}_k) \propto \exp \left\{ -\frac{1}{2} (\mathbf{x}_k - \boldsymbol{\mu}_k^x)^{\mathrm{T}} (\boldsymbol{\Sigma}_k^x)^{-1} (\mathbf{x}_k - \boldsymbol{\mu}_k^x) \right\}. \quad (8)$$

---

[2]The memory L is a tuning parameter, where larger L will lead to higher complexity and better performance.

[3]A non-parametric approach with a kernel can be used as an alternative [39], where the kernel bandwidth is a free parameter that should be tuned.

As a result, ψ_k can be simplified as

$$\psi_k(s_k, \mathbf{x}_k) \propto (\mathbf{y}_k - \boldsymbol{\mu}_k^y)^{\mathrm{T}} (\boldsymbol{\Sigma}_k^y)^{-1} (\mathbf{y}_k - \boldsymbol{\mu}_k^y)$$
$$- (\mathbf{x}_k - \boldsymbol{\mu}_k^x)^{\mathrm{T}} (\boldsymbol{\Sigma}_k^x)^{-1} (\mathbf{x}_k - \boldsymbol{\mu}_k^x). \quad (9)$$

To find an estimate $\hat{\mathbf{s}}$ using (5), a VA[4] is used with the current state as $\mathbf{x}_k$ and the current symbol as s_k with branch metric as in (9) for the kth symbol slot. Observe that since the search space for $\mathbf{s}$ is Ω^K, the search space for $\mathbf{y}_k$ is Ω^{L+1} and $\mathbf{x}_k \in \Omega^L$. When $L = 0$, VA-SDBP reverts back to SBS-SDBP from [22].

C. Decision Directed-Stochastic Digital Backpropagation

The second approach, DD-SDBP, combines the idea of exploiting memory, as in VA-SDBP, with taking decisions on a SBS basis, as in SBS-SDBP. In this approach, the previously decoded symbols, $[\hat{s}_{k-1}, \hat{s}_{k-2}, \ldots, \hat{s}_{k-L}]^{\mathrm{T}}$, are used while taking decisions for the current symbol $\mathbf{s}_k$ and as a result $\mathbf{x}_k$ in (9) can be interpreted as a constant that does not affect the optimization in (9). Thus, decisions on s_k in DD-SDBP are taken as

$$\hat{s}_k = \arg \max_{s_k \in \Omega} q_{\mathrm{d}}(s_k | \hat{s}_{k-1}, \hat{s}_{k-2}, \ldots, \hat{s}_{k-L})$$
$$= \arg \max_{s_k \in \Omega} \psi(s_k, \hat{\mathbf{x}}_k), \quad (10)$$

which can be solved recursively starting[5] from $k = L+1$ using a similar Gaussian approximation as in VA-SDBP. However, in contrast to VA-SDBP, DD-SDBP can only account for causal memory effects. Note that the search space for $\mathbf{y}_k$ of (7) is Ω instead of Ω^{L+1} as the decisions have to be taken only for s_k. When $L = 0$, DD-SDBP also reverts to SBS-SDBP.

V. NUMERICAL SIMULATIONS AND DISCUSSION

A. Simulation Setup and Performance Metrics

The simulation setup is shown in Fig. 1. The pulse shape used at the transmitter is a root-raised-cosine pulse with a roll-off factor of 0.25 and truncation length of 16 symbol periods. The simulations are performed for a polarization-multiplexed signal, with no polarization mode dispersion[6], either with 16-QAM or QPSK as modulation format and symbol rates R_s of 14 Gbaud, 28 Gbaud, and 56 Gbaud. In each polarization, $K = 4096$ symbols[7] are transmitted in each block of the Monte Carlo simulation. This signal is input to the channel with N spans. The parameters used for the SMF are $D = 16$ ps/(nm km), $\gamma = 1.3$ (W km)$^{-1}$, $\alpha = 0.2$ dB/km, which are according to the ITU-T G.652. We have considered a fiber Bragg grating (FBG) as a DCM.[8] Propagation in the fibers is simulated using the SSFM with a segment length [41] of $\Delta = (\epsilon L_{\mathrm{N}} L_{\mathrm{D}}^2)^{1/3}$, where $\epsilon = 10^{-4}$, $L_{\mathrm{N}} = 1/(\gamma P)$ is the nonlinear length, $L_{\mathrm{D}} = T^2 2\pi c/(|D|\lambda^2)$ is the dispersion length, λ is the wavelength,

---

[4]We assume a uniform a priori distribution over all the states which implies the symbols at the start of trellis are unknown.

[5]For $k = 1, \ldots, L$, decisions on s_k are taken on an SBS basis.

[6]Effect of PMD on DBP and SBS-SDBP has been reported in [22].

[7]Except for 56 GBd NDM links, where we have simulated with $K = 8192$ symbols to properly account for the memory in the system.

[8]However, a dispersion compensating fiber (DCF), simulated according to G.655 specifications, exhibited similar trends as the FBG.

TABLE I
NUMBER OF SPANS, N, USED IN DM AND NDM LINKS

	R_s [GBd]	DM	NDM
QPSK	14	50	110
	28	35	110
	56	35	110
16-QAM	14	50	110
	28	40	110
	56	40	110

c is the speed of the light, and P is the average input power to each fiber span. We used the same segment lengths for simulating the channel and for both DBP and SDBP.

The number of spans, N, used in each of the scenarios is summarized in Table I. We also considered a non-DM (NDM) system, wherein there are no DCMs. The span length used for SMF, L_{SMF}, is 80 km for 16-QAM and 120 km for QPSK.[9] FBG with an insertion loss of 3 dB and perfect dispersion compensation for the preceding SMF is used. The launch power into the DCM is 4 dB below that of the transmission fiber, which is compensated for by the EDFA after DCM. The noise figure is 5 dB for each of the amplifiers. Ideal low-pass filters with one-sided bandwidth of R_s are used in EDFAs and in the beginning of the receiver. The filtered signal is sent to DBP and three different SDBP detectors:

1) SBS-SDBP from [22];
2) DD-SDBP for $L \in \{1, 2\}$, as proposed in Sec. IV-C; and
3) VA-SDBP for $L \in \{1, 2\}$, as proposed in Sec. IV-B.

In all SDBP detectors, $N_p = 500$ particles were used to generate the results, but we verified that even with $N_p = 1500$, similar performance was obtained. The receiver is assumed to have perfect knowledge of the polarization state, as well as the carrier phase and the symbol timing.

We consider two performance metrics. To capture the absolute performance of each detector, we determine the symbol error rate (SER). To capture the relative performance gain of SDBP over DBP, we introduce $G_X = SER_{DBP}/SER_{X\text{-SDBP}}$, where $X \in \{SBS, DD, VA\}$, and in which SER_{DBP} and $SER_{X\text{-SDBP}}$ are *lowest* SERs obtained for the respective algorithms.

B. Results and Discussion

The SER as a function of input power is shown in Fig. 2 for the 56 Gbaud system with FBG as DCM for (a) QPSK and (b) 16-QAM. Due to complexity reasons, VA-SDBP is simulated only with $L = 1$ for 16-QAM. Comparing the SER of SBS-SDBP with DD-SDBP/VA-SDBP, we can conclude that by taking the residual memory into account, the SER is significantly reduced. One can also see that for both VA-SDBP and DD-SDBP, increasing L leads to a decrease in the SER. We expect this gain to saturate as L increases. We also see

[9]The number of spans and span lengths are selected such that the symbol error rate for DBP is around 0.01.

that the optimal power (i.e., corresponding to the lowest SER) varies for each detector: compared to DBP, the optimal power for SDBP is up to 2 dB larger for QPSK and up to 4 dB larger for 16-QAM.
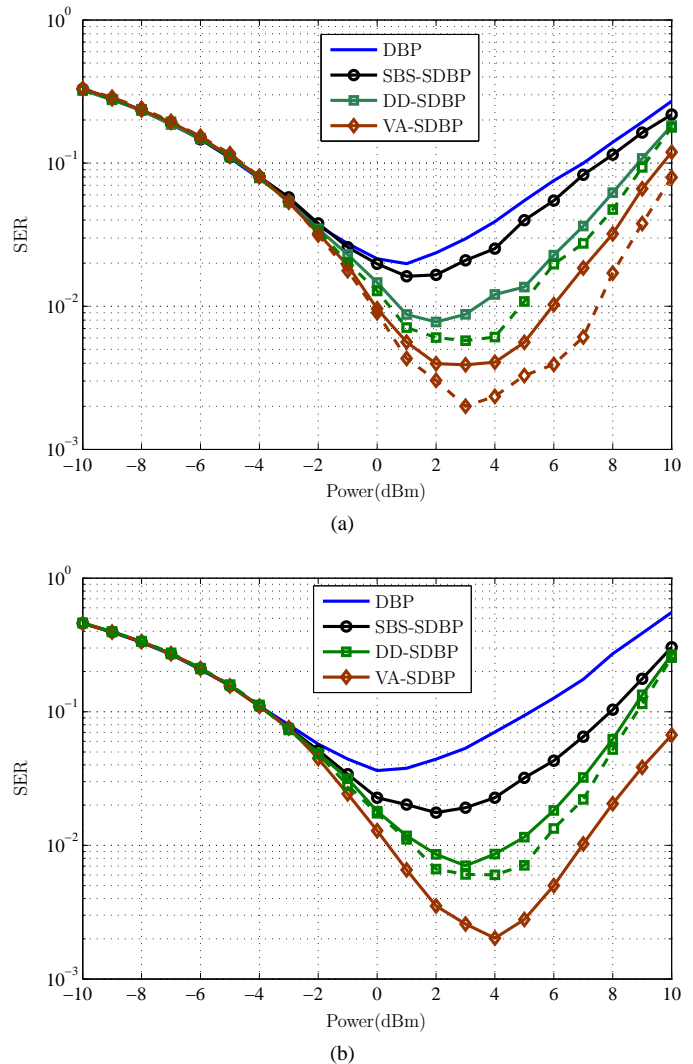

(a)

(b)

Fig. 2. SER as a function of input power for 56 Gbaud, FBG as DCM, for (a) QPSK and (b) 16-QAM. Solid (resp. dashed) lines in VA-SDBP and DD-SDBP represent cases when $L = 1$ (resp. $L = 2$).

Similar behavior is observed for other symbol rates, although we do not show all results. Instead, a summary of the performance gains is presented in Fig. 3 for QPSK with $L = 2$ and for 16-QAM with $L = 1$. As the complexity of VA-SDBP grows exponentially with L, the same value of L was used for both DD-SDBP and VA-SDBP to have a fair comparison. From Fig. 3, irrespective of the symbol rate, DM or NDM links, we observe a clear trend: $SER_{VA\text{-SDBP}} < SER_{DD\text{-SDBP}} < SER_{SBS\text{-SDBP}} < SER_{DBP}$. The VA-SDBP can account for both causal and non-causal effect, giving it a performance benefit over DD-SDBP. SBS-SDBP ignores both causal and non-causal memory and thus exhibits the worst performance among the three selected approaches. The decreasing gains with increasing symbol rate for SBS-SDBP can be explained as follows. The larger the deviation of the particle clouds,
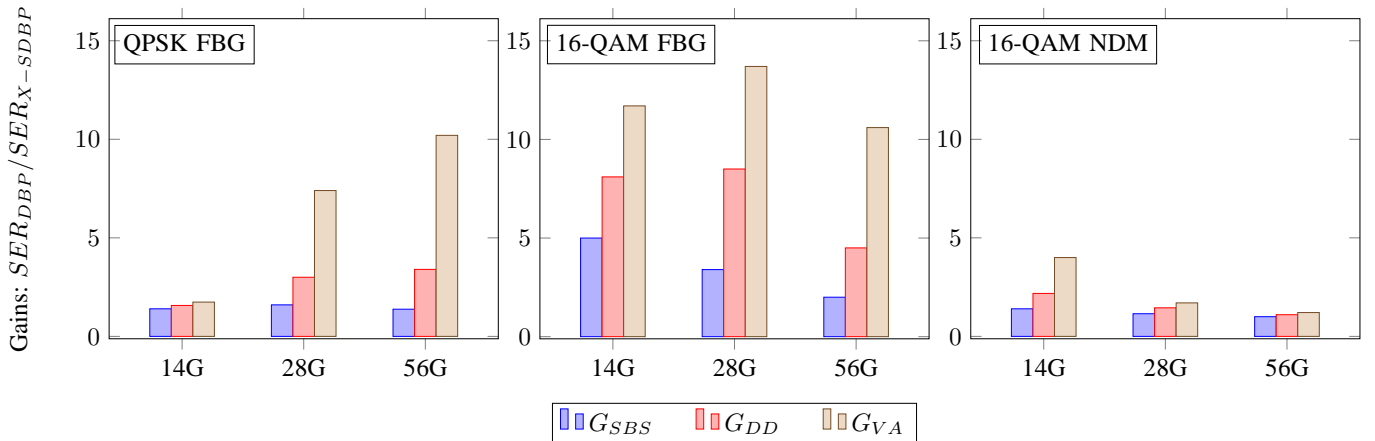
Fig. 3. Gains in SER for the proposed algorithms compared to DBP.

given by $\{\tilde{\mathbf{s}}^{(n)}\}_{n=1}^{N_p}$, from a circular symmetric Gaussian, the higher are the expected gains in SDBP compared to DBP. As the symbol rate increases, the particle clouds in SBS-SDBP tend to become more circular Gaussian and hence the gains decrease. Also for a DM link, we have observed that the particle clouds are less circular Gaussian and hence gains are higher for SDBP in DM links compared to NDM links.

The gains in VA-SDBP for QPSK increase with increasing symbol rate whereas for 16-QAM, the gains increase from 14 GBd to 28 GBd and then the gains decrease. This maybe due to the use of $L = 1$ for 16-QAM, which is not sufficient to account for the residual memory, especially at high symbol rates. The main drawback of VA-SDBP is its complexity, which grows exponentially with the memory L. So, an interesting case would be to test a low-complexity version of VA for 16-QAM with higher memory. DD-SDBP is a tradeoff between SBS-SDBP and VA-SDBP, in terms of complexity and performance. Irrespective of which SDBP approach is used, there is always an improvement in terms of SER compared to the traditional DBP algorithm (i.e., $G_X > 1$). This means that NSNI plays an important role in the systems under consideration and one can gain significantly by taking these interactions into account. The NSNI is more important in the DM systems than the NDM systems, so that gains are lower in NDM systems. An additional observation that can be made from Fig. 3 is that the gains are in general higher for 16-QAM than for QPSK as 16-QAM has more nonlinearities than QPSK and hence more signal-noise interactions, and thereby more gains of SDBP approach compared to DBP. Gains in QPSK NDM links (not shown here) turn out to be lower than corresponding gains of 16-QAM NDM case.

VI. CONCLUSIONS

We have extended the SDBP algorithm to account for residual memory that may be present due to nonoptimal processing in SDBP. Specifically, we proposed DD-SDBP and VA-SDBP to account for this memory, at an increased cost in terms of complexity. Extensive simulations were performed to evaluate these methods for 16-QAM and QPSK, and for DM and NDM links. Results suggest a significant improvement by

the proposed detectors for DM links with up to 10 times lower SER for QPSK and up to 14 times lower SER for 16-QAM, compared to DBP.

The VA-SDBP can provide optimal decisions on the transmitted sequence, but does so at a high computational cost. Alternatives to consider are low-complexity variations of the VA, as well as algorithms that provide symbol-wise optimal decisions, such as the Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm [42].

Further gains over the proposed algorithms may be possible and remains the topic of ongoing and future research. Lower SER can be expected by increasing the memory in VA-SDBP until the SER gains saturate. In addition, the use of a matched filter is not necessarily optimal. Initial results in this direction can be found in [38]. Finally, in SBS-SDBP, DD-SDBP, and VA-SDBP, the particles after matched filtering and sampling are approximated with a multivariate Gaussian distribution, which need not be a good approximation, especially at high input powers. Other types of distributions should be considered.

REFERENCES

[1] R. J. Essiambre and P. J. Winzer, "Fibre nonlinearities in electronically pre-distorted transmission," in *Proc. European Conference on Optical Communication (ECOC)*, 2005.

[2] K. Roberts, C. Li, L. Strawczynski, M. O'Sullivan, and I. Hardcastle, "Electronic precompensation of optical nonlinearity," *IEEE Photon. Technol. Lett.*, vol. 18, no. 2, pp. 403–405, 2006.

[3] 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.

[4] 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.

[5] T. Fehenberger and N. Hanik, "Digital back-propagation of a superchannel: achievable rates and adaption of the GN model," in *Proc. European Conference on Optical Communication (ECOC)*, 2014, pp. 1–3.

[6] T. Koike-Akino, K. Parsons, C. Kojima, C. Duan, T. Yoshida, T. Sugihara, and T. Mizuochi, "Fractionally-spaced statistical equalizer for fiber nonlinearity mitigation in digital coherent optical systems," in *Proc. Optical Fiber Communication Conference (OFC)*, 2012, p. OTh3C.3.

[7] D. S. Millar, S. Makovejs, C. Behrens, S. Hellerbrand, R. I. Killey, 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.

[8] 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.

[9] M. Secondini, D. Marsella, and E. Forestieri, "Enhanced split-step Fourier method for digital backpropagation," in *Proc. European Conference on Optical Communication (ECOC)*, 2014.

[10] 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.

[11] A. Napoli, Z. Maalej, V. A. J. M. Sleiffer, M. Kuschnerov, D. Rafique, E. Timmers, B. Spinnler, T. Rahman, L. D. Coelho, and N. Hanik, "Reduced complexity digital back-propagation methods for optical communication systems," *J. Lightw. Technol.*, vol. 32, no. 7, pp. 1351–1362, 2014.

[12] M. Secondini and E. Forestieri, "On XPM mitigation in WDM fiber-optic systems," *IEEE Photon. Technol. Lett.*, vol. 26, no. 22, pp. 2252–2255, 2014.

[13] P. Serena and A. Bononi, "A time-domain extended Gaussian noise model," *J. Lightw. Technol.*, vol. 33, no. 7, pp. 1459–1472, 2015.

[14] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Inter-channel nonlinear interference noise in WDM systems: modeling and mitigation," *J. Lightw. Technol.*, vol. 33, no. 5, pp. 1044–1053, 2015.

[15] 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.

[16] 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. 26 376–26 386, Oct. 2013.

[17] 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.

[18] 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.

[19] 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.

[20] Y. Cai, "MAP detection for linear and nonlinear ISI mitigation in long-haul coherent detection systems," in *Proc. IEEE Photonics Society Summer Topicals*, vol. 1, Jul. 2010, pp. 42–43.

[21] 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.

[22] N. V. Irukulapati, H. Wymeersch, P. Johannisson, and E. Agrell, "Stochastic digital backpropagation," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 3956–3968, 2014.

[23] N. V. Irukulapati, D. Marsella, P. Johannisson, M. Secondini, H. Wymeersch, E. Agrell, and E. Forestieri, "On maximum likelihood sequence detectors for single-channel coherent optical communications," in *Proc. European Conference on Optical Communication (ECOC)*, 2014.

[24] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and

[24] 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.

[25] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1295–1322, Jun. 2007.

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

[27] A. P. Worthen and W. E. Stark, "Unified design of iterative receivers using factor graphs," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 843–849, 2001.

[28] F. R. Kschischang and B. J. Frey, "Iterative decoding of compound codes by probability propagation in graphical models," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 219–230, 1998.

[29] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, 1998.

[30] M. Tuchler and A. C. Singer, "Turbo equalization: An overview," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 920–952, 2011.

[31] Q. Guo, L. Ping, and D. Huang, "A low-complexity iterative channel estimation and detection technique for doubly selective channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4340–4349, 2009.

[32] C. Herzet, N. Noels, V. Lottici, H. Wymeersch, M. Luise, M. Moeneclaey, and L. Vandendorpe, "Code-aided turbo synchronization," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1255–1271, 2007.

[33] G. Colavolpe, A. Barbieri, and G. Caire, "Algorithms for iterative decoding in the presence of strong phase noise," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 9, pp. 1748–1757, 2005.

[34] F. Yu, N. Stojanovic, F. N. Hauske, D. Chang, Z. Xiao, G. Bauch, D. Pflueger, C. Xie, Y. Zhao, L. Jin, Y. Li, L. Li, X. Xu, and Q. Xiong, "Soft-decision LDPC turbo decoding for DQPSK modulation in coherent optical receivers," in *Proc. European Conference on Optical Communication (ECOC)*, 2011, p. We.10.P1.70.

[35] B. P. Smith and F. R. Kschischang, "Future prospects for FEC in fiber-optic communications," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 5, pp. 1245–1257, 2010.

[36] I. B. Djordjevic, L. L. Minkov, and H. G. Batshon, "Mitigation of linear and nonlinear impairments in high-speed optical networks by using LDPC-coded turbo equalization," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 6, pp. 73–83, Aug. 2008.

[37] I. B. Djordjevic, B. Vasic, M. Ivkovic, and I. Gabitov, "Achievable information rates for high-speed long-haul optical transmission," *J. Lightw. Technol.*, vol. 23, no. 11, pp. 3755–3763, 2005.

[38] H. Wymeersch, N. V. Irukulapati, I. A. Sackey, P. Johannisson, and E. Agrell, "Backward particle message passing," in *Proc. Signal Processing Advances in Wireless Communications (SPAWC)*, 2015.

[39] 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 *Proc. International Conference on Communications (ICC)*, Jun. 2011.

[40] A. Lapidoth, *A Foundation in Digital Communication*. Cambridge University Press, 2009.

[41] 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.

[42] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 284–287, 1974.