Suppression of Fiber Nonlinearities and PMD in Coded-Modulation Schemes With Coherent Detection by Using Turbo Equalization

Ivan B. Djordjevic, Lyubomir L. Minkov, Lei Xu, and Ting Wang

***Abstract*—We propose a maximum *a posteriori* probability (MAP) turbo equalizer based on the sliding-window *multilevel* Bahl–Cocke–Jelinek–Raviv algorithm. This scheme is suitable for *simultaneous* nonlinear and linear impairment mitigation in multilevel coded-modulation schemes with coherent detection. The proposed scheme employs large-girth quasi-cyclic LDPC codes as channel codes. We demonstrate the efficiency of this method in dealing with fiber nonlinearities by performing Monte Carlo simulations. In addition, we provide the experimental results that demonstrate the efficiency of this method in dealing with polarization mode dispersion. We also study the ultimate channel capacity limits, assuming an independent identically distributed source.**

***Index Terms*—Coherent detection; Fiber-optics communications; Fiber nonlinearities; LDPC codes; Multilevel coded modulation; Multilevel MAP detection; Turbo equalization**.

I. Introduction

In order to adapt to the ever-increasing demands of telecommunication needs, network operators already consider 100 Gb/s per dense wavelength division multiplexing (DWDM) channel transmission. At those data rates, the performance of fiber-optic communication systems is degraded significantly due to intrachannel and interchannel fiber nonlinearities, polarization-mode dispersion (PMD), and chromatic dispersion [1–8]. To deal with those channel impairments, novel advanced techniques in modulation and detection and coding and signal processing should be developed. To deal with chromatic dispersion and PMD, a number of channel equalization techniques have been proposed recently including the digital filtering approach [1], maximum likelihood sequence detection (MLSD) [2], turbo equalization (see [3] and references therein), and the backpropagation method [4,5]. To simultaneously suppress chromatic dispersion and the PMD, orthogonal frequency division multiplexing (OFDM) has been proposed [6,7]. On the other hand, to deal with intrachannel nonlinearities one may use constrained coding [8], turbo equalization [3,8], and backpropagation [4,5]. Moreover, it has been shown in [3] that fiber nonlinearities, chromatic dispersion, and PMD can be compensated for simultaneously by using low-density parity-check (LDPC) coded turbo equalization. This scheme so far has been studied only for *binary* transmission with direct detection (see [3] and references therein).

In this paper, we propose the sliding-window *multilevel* ($M>2$) maximum *a posteriori* probability (MAP) turbo equalization scheme based on the multilevel Bahl–Cocke–Jelinek–Raviv (BCJR) algorithm-based equalizer (called here the multilevel BCJR equalizer). When used in combination with large-girth LDPC codes [3,9] as channel codes, this scheme represents a universal equalizer scheme for *simultaneous* suppression of fiber nonlinearities, for chromatic dispersion compensation, and for PMD compensation. The BCJR algorithm [10,11] is the MAP algorithm that can be used not only to decode different convolutional and block codes, but also as a MAP detector. On the other hand, the use of large-girth LDPC codes is essential because the large girth increases the minimum distance and decorrelates the extrinsic info in the LDPC decoding process. To further improve the overall bit-error ratio (BER) performance, we perform the iteration of extrinsic log-likelihood ratios (LLRs) between the LDPC decoder and the multilevel BCJR equalizer. We use the extrinsic information transfer (EXIT) chart approach attributed to ten Brink [12,13] to

Manuscript received June 26, 2009; revised September 25, 2009; accepted September 30, 2009; published October 28, 2009 (Doc. ID 113360).

I. B. Djordjevic (e-mail: ivan@ece.arizona.edu) and L. L. Minkov are with the Department of Electrical and Computer Engineering of the University of Arizona, Tucson, Arizona 85721, USA.

L. Xu and T. Wang are with NEC Laboratories America, Princeton, New Jersey 08540, USA.

Digital Object Identifier 10.1364/JOCN.1.000555

1943-0620/09/060555-10/$15.00

match the LDPC decoders, for large-girth quasi-cyclic LDPC codes, and the multilevel BCJR equalizer. We further show how to combine this scheme with multilevel coded-modulation schemes with coherent detection.

Given the fact that an LDPC-coded turbo equalizer, based on the multilevel BCJR algorithm, is an excellent nonlinear intersymbol interference (ISI) equalizer candidate, the question about fundamental limits on the channel capacity of coded-modulation schemes naturally arises [3,4,14–19]. For completeness of presentation we also provide the independent identically distributed (IID) channel capacity study.

The contributions of the paper can be summarized as follows: (i) the first demonstration to our knowledge of the efficiency of turbo equalization for *nonbinary* modulation schemes with coherent detection in the context of fiber-optic communication, (ii) demonstration of the possibility of fiber nonlinearity mitigation by simulation, (iii) demonstration of PMD compensation by experiments, (iv) the IID information capacity evaluation, and (v) demonstration of the possibility for 100 Gb/s upgrade using the existing 10 Gb/s infrastructure.

The paper is organized as follows. The LDPC-coded turbo equalization scheme, based on the multilevel BCJR algorithm, is described in Section II. The design of large-girth LDPC codes used in the turbo equalizer is given in Section III. In Section IV we study the efficiency of the LDPC-coded turbo equalizer in suppression of fiber nonlinearities. In Section V we provide a channel capacity study. In Section VI we study the efficiency of this method for PMD compensation and provide experimental validation. Finally, in Section VII, some important concluding remarks are given.

II. Multilevel BCJR Equalizer and LDPC-Coded Turbo Equalizer Description

As mentioned in Section I, we propose a multilevel BCJR equalizer suitable for simultaneous suppression of both nonlinear and linear impairments in multilevel coded-modulation schemes with coherent detection.

The sliding-window multilevel BCJR equalizer operates on a discrete dynamical trellis description of the optical channel. This dynamical trellis is uniquely defined by the following triplet: the previous state, the next state, and the channel output. The state in the trellis is defined as $\mathbf{s}_j = (x_{j-m}, x_{j-m+1}, \ldots, x_j, x_{j+1}, \ldots, x_{j+m}) = \mathbf{x}[j-m, j+m]$, where x_k denotes the index of the symbol from the set of possible indices $\mathbf{X}=\{0,1,\ldots,M-1\}$, with M being the number of points in the corresponding M-ary signal constellation such as M-ary phase-shift keying (PSK), M-ary quadrature-amplitude modulation (QAM), or M-ary polarization-shift keying (Po1SK). Every symbol carries $l=\log_2 M$ bits, using the appropriate mapping rule (natural, Gray, anti-Gray, etc.) For example, for QPSK and Gray mapping, the sequences of bits 00, 01, 11, and 10 are mapped to the signal constellation points QPSK(0)=(1,0), QPSK(1)=(0,1), QPSK(2)=(−1,0), and QPSK(3)=(0,−1), respectively. The memory of the state is equal to $2m+1$, with $2m$ being the number of *symbols* that influence the observed symbol from both sides. An example trellis of memory $2m+1=3$ for 4-ary modulation formats (such as QPSK) is shown in Fig. 1. The trellis has $M^{2m+1}=64$ states $(\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_{63})$, each of which corresponds to a different 3-symbol pattern (configuration). The state index is determined by considering $(2m+1)$ symbols as digits in the numerical system with the base M. For example, in Fig. 1, the quaternary numerical system (with the base 4) is used. [In this system, 18 is represented by $(102)_4$.] The left column in the dynamic trellis represents the current states and the right column denotes the terminal states. The branches are labeled by two symbols, the input symbol (denoted as the blue symbol) and the output symbol, which is the central symbol of the terminal state (the red symbol). The multilevel BCJR equalizer operates on a trellis with a depth of 128 symbols. The window of 128 symbols includes m symbols from the previous window and m symbols from the incoming window. For the complete description of the dynamical trellis, the transition probability density functions (PDFs) $p(y_j|x_j)=p(y_j|\mathbf{s})$, $\mathbf{s}\in\mathbf{S}$ are needed, where $\mathbf{S}$ is the set of states in the trellis and y_j is the complex number (corresponding to the transmitted symbol index x_j) with its real part being the in-phase channel sample and its imaginary part being the quadrature

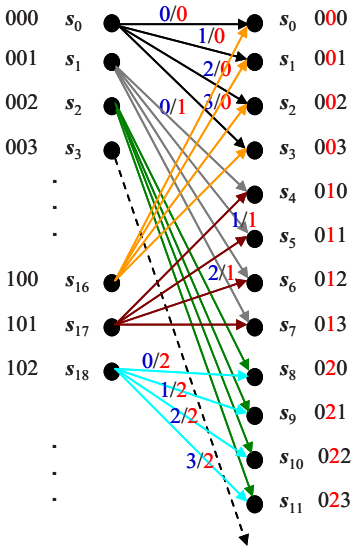

Fig. 1. (Color online) A portion of the trellis for the four-level BCJR equalizer with memory $2m+1=3$.

channel sample. The conditional PDFs can be determined from *collected histograms* or by using the *instanton–Edgeworth expansion* method [20]. The number of edges originating in any of the left-column states is M, and the number of merging edges in the arbitrary terminal state is also M. Notice that samples y_j $(j=1,\ldots,n)$ are correlated, and to decorrelate them we perform symbol-level interleaving as shown in Fig. 2.

The transmitter configuration, for multilevel coding (MLC) with parallel independent decoding (PID) [21], is shown in Fig. 2(a). The PID is chosen because it performs comparable with multistage decoding as shown in [21], while having lower complexity and latency. The bit streams originating from l different information sources are encoded using different (n,k_i) LDPC codes of code rate $R_i=k_i/n$. k_i denotes the number of information bits of the ith $(i=1,2,\ldots,l)$ component LDPC code, and n denotes the codeword length, which is the same for all LDPC codes. The outputs of l LDPC encoders are written row-wise into a block-interleaver block. The mapper accepts l bits at time instance j from the $(l\times n)$ interleaver column-wise and determines the corresponding M-ary $(M=2^l)$ signal constellation point $\mathbf{s}_j=(\phi_{1,j},\phi_{2,j})=|\mathbf{s}_j|\exp(jf_j)$. The coordinates correspond to in-phase and quadrature components of M-ary QAM constellation, after appropriate mapping. The bit-interleaved coded modulation (BICM) scheme [22], as pointed out in [21], can be considered as a special MLC scheme in which all of the component codes are of the same rate. Notice that in conventional BICM only one decoder is used, with the operating speed of encoders/decoders being $R_s\log_2 M$, where R_s is the symbol rate and M is the constellation size. In our coded-modulation scheme, however, the operating speed of encoders/decoders is R_s, and as such is more suitable for high-speed implementation. The receiver configuration is shown in Fig. 2(b). The received electrical field at the ith transmission interval is denoted by $S_i=|S_i|\exp(j\varphi_{Si})$, $\varphi_{Si}=\varphi_i+\varphi_{S,PN}$, where φ_i denotes the data phasor and $\varphi_{S,PN}$ denotes the laser phase noise process of the transmitting laser. The local laser electrical field is denoted by $L=|L|\exp(j\varphi_L)$, where φ_L denotes the laser phase noise process of the local laser. The outputs of upper- and lower-balanced branches, proportional to $\mathrm{Re}\{S_iL^*\}$ and $\mathrm{Im}\{S_iL^*\}$, respectively, are used as inputs of the multilevel BCJR equalizer.

Before we explain how the symbol LLRs are calculated in the multilevel BCJR equalizer block, let us introduce the following notation, which is adopted from an excellent tutorial paper by Ryan [11] (see also [13]). Notice that in [11] the BCJR decoder for *binary* turbo codes is described. Since in most textbooks (e.g., [13]) the multilevel turbo equalization is not considered at all, for the completeness of presentation we provide its description here.

The *forward metric* is defined as $\alpha_j(\mathbf{s})=\log\{p(\mathbf{s}_j=\mathbf{s},\mathbf{y}[1,j])\}$ $(j=1,2,\ldots,n)$, the *backward metric* is defined as $\beta_j(\mathbf{s})=\log\{p(\mathbf{y}[j+1,n]|\mathbf{s}_j=\mathbf{s})\}$, and the *branch metric* is defined as $\gamma_j(\mathbf{s}',\mathbf{s})=\log[p(\mathbf{s}_j=\mathbf{s},y_j,\mathbf{s}_{j-1}=\mathbf{s}')]$. The corresponding metrics can be calculated iteratively as follows:

$$\alpha_j(\mathbf{s})=\max_{\mathbf{s}'}{}^*[\alpha_{j-1}(\mathbf{s}')+\gamma_j(\mathbf{s}',\mathbf{s})],$$

$$\beta_{j-1}(\mathbf{s}')=\max_{\mathbf{s}}{}^*[\beta_j(\mathbf{s})+\gamma_j(\mathbf{s}',\mathbf{s})],$$

$$\gamma_j(\mathbf{s}',\mathbf{s})=\log[p(y_j|\mathbf{x}[j-m,j+m])P(x_j)]. \quad (1)$$

The $\max^*$ operator is defined by $\max^*(x,y)=\log(e^x+e^y)$, and it is efficiently calculated by $\max^*(x,y)=\max(x,y)+c_f(x,y)$, where $c_f(x,y)$ is the correction factor, defined as $c_f(x,y)=\log[1+\exp(-|x-y|)]$, which is commonly approximated or implemented using a lookup table. $p(y_j|\mathbf{x}[j-m,j+m])$ is obtained, as already explained above, by either collecting the histograms or by the instanton–Edgeworth expansion

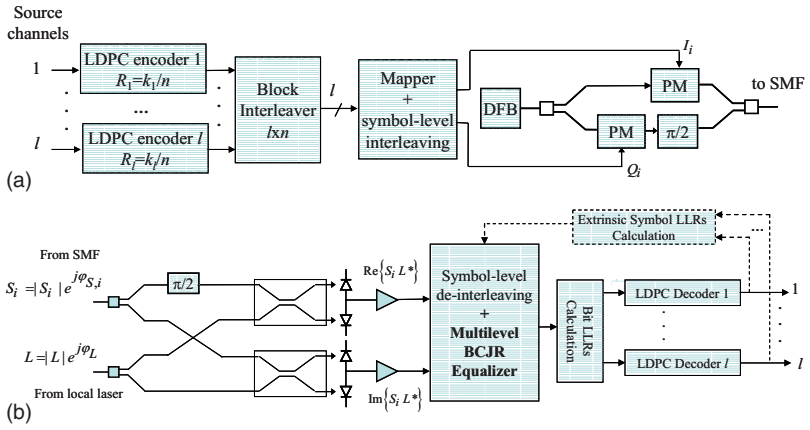

Fig. 2. (Color online) Proposed coded-modulation and turbo equalization schemes: (a) transmitter configuration and (b) receiver configuration. PM, phase modulator; DFB, distributed-feedback laser.

method, and $P(x_j)$ represents the *a priori* probability of transmitted symbol x_j. In the first outer iteration, $P(x_j)$ is set to either $1/M$ (because equally probable transmission is observed) for an existing transition from the trellis given in Fig. 1 or to zero for a nonexisting transition.

The outer iteration is defined as the calculation of symbol LLRs in the multilevel BCJR equalizer block, the calculation of corresponding bit LLRs needed for LDPC decoding, the LDPC decoding, and the calculation of extrinsic symbol LLRs needed for the next iteration. The iterations within the LDPC decoder, which are based on the min-sum-with-correction-term algorithm [23], are called inner iterations. The initial forward and backward metrics values are set to

$$\alpha_0(\mathbf{s}) = \begin{cases} 0, & \mathbf{s} = \mathbf{s}_0 \\ -\infty, & \mathbf{s} \neq \mathbf{s}_0 \end{cases} \quad \text{and} \quad \beta_n(\mathbf{s}) = \begin{cases} 0, & \mathbf{s} = \mathbf{s}_0 \\ -\infty, & \mathbf{s} \neq \mathbf{s}_0 \end{cases}, \tag{2}$$

where $\mathbf{s}_0$ is an initial state.

Let $\mathbf{s}' = \mathbf{x}[j-m-1, j+m-1]$ represent the previous state, $\mathbf{s} = \mathbf{x}[j-m, j+m]$ represent the present state, $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ represent the transmitted word of symbols, and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ represent the received sequence of samples. The LLR, denoting the reliability, of symbol $x_j = \delta$ $(j = 1, 2, \ldots, n)$ can be calculated by

$$\Lambda(x_j = \delta) = \max_{(\mathbf{s}', \mathbf{s}): x_j = \delta}{}^{*} [\alpha_{j-1}(\mathbf{s}') + \gamma_j(\mathbf{s}', \mathbf{s}) + \beta_j(\mathbf{s})] - \max_{(\mathbf{s}', \mathbf{s}): x_j = \delta_0}{}^{*} [\alpha_{j-1}(\mathbf{s}') + \gamma_j(\mathbf{s}', \mathbf{s}) + \beta_j(\mathbf{s})], \tag{3}$$

where δ represents the observed symbol ($\delta \in \{0, 1, \ldots, M-1\} - \{\delta_0\}$) and δ_0 is the reference symbol. The forward and backward metric is calculated using Eq. (1). The forward and backward recursion steps of the four-level BCJR MAP detector are illustrated in Figs. 3(a) and 3(b), respectively. In Fig. 3(a) $\mathbf{s}$ denotes an arbitrary terminal state, which has $M = 4$ edges originating from corresponding initial states, denoted as $\mathbf{s}'_1$, $\mathbf{s}'_2$, $\mathbf{s}'_3$, and $\mathbf{s}'_4$. Notice that the first term in the branch metric [see Eq. (1)] is calculated only once, before the detection/decoding takes place, and stored. The second term, $\log[P(x_j)]$, is recalculated in every outer iteration. The forward metric of state $\mathbf{s}$ in the jth step $(j = 1, 2, \ldots, n)$ is updated by preserving the maximum term (in max* sense) $\alpha_{j-1}(\mathbf{s}'_k) + \gamma_j(\mathbf{s}, \mathbf{s}'_k)$ $(k = 1, 2, 3, 4)$. The procedure is repeated for every state in the column of terminal states of the jth step. A similar procedure is used to calculate the backward metric of state $\mathbf{s}'$, $\beta_{j-1}(\mathbf{s}')$, [in the $(j-1)$th step], as shown in Fig. 3(b), but now proceeding backward $(j = n, n-1, \ldots, 1)$.

Let the c_k denote the kth bit carried by the symbol denoted by index x_j. The bit LLRs c_k $(k = 1, 2, \ldots, l)$ are

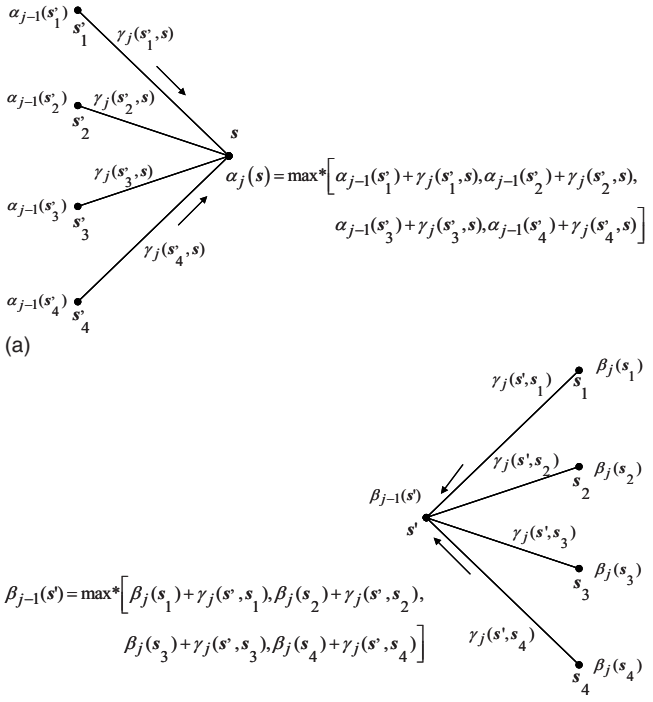

Fig. 3. Forward/backward recursion steps for $M = 4$-level BCJR equalizer: (a) the forward recursion step, (b) the backward recursion step.

determined from the symbol LLRs of Eq. (3), in the bit LLR calculation block, as follows:

$$L(\hat{c}_k) = \log \frac{\sum_{x_j: c_k = 0} \exp[\Lambda(x_j)]}{\sum_{x_j: c_k = 1} \exp[\Lambda(x_j)]}, \tag{4}$$

where the summation in the numerator is performed over all symbol indices x_j having 0 at position k, while the summation in the denominator is performed over all symbol indices x_j having 1 at the same position.

The bit LLR calculation block forwards the bit LLRs from Eq. (4) to the soft-decoding LDPC decoder implemented based on the sum-product-with-correction-term algorithm, as mentioned above. To improve the overall performance of the LDPC-coded turbo equalizer we perform the iteration of *extrinsic* LLRs between the LDPC decoder and the multilevel BCJR equalizer. The extrinsic *bit* LLRs at the input of the extrinsic symbol calculation block, in tth iteration, are determined by

$$L_{\text{LDPC},e}[c_k^{(t)}] = L_{\text{LDPC}}(c_k^{(t)}) - L_{\text{LDPC}}(c_k^{(t-1)}), \tag{5}$$

where $L_{\text{LDPC}}(c_k)$ is the corresponding LLR of bit c_k, while indices t and $t-1$ are used to denote the current and previous iterations. The extrinsic bit LLRs from Eq. (5) are used to calculate the extrinsic *symbol* LLRs [in the extrinsic symbol calculation block of Fig. 2(b)], which are used as *a priori* symbol LLRs, in the third line of Eq. (1) by

$$L_{\mathrm{BCJR},a}(x_j) = \log[P(x_j)] = \sum_{k=0}^{l-1} (1 - c_k) L_{D,e}(c_k). \quad (6)$$

The use of large-girth LDPC codes is essential because the large girth in addition to increasing the minimum distance also decorrelates the extrinsic bit LLRs. To facilitate the implementation at high speed, we prefer the use of quasi-cyclic codes [3,9] rather than random LDPC codes; the corresponding design is briefly given in Section III (please refer to [9] for more details on large-girth LDPC codes). To optimally match the multilevel BCJR equalizer and LDPC decoder, quasi-cyclic LDPC codes are selected using the concept of EXIT charts as explained in [12].

Notice that the complexity of the dynamic trellis grows exponentially because the number of states is determined by M^{2m+1}, so that the increase in signal constellation leads to an increase of the base, while the increase in the state memory assumption $(2m+1)$ leads to an increase of exponent. We will show in the case of QPSK transmission that even a small state memory assumption $(2m+1=3)$ leads to significant performance improvement with respect to the case $m=0$. For larger constellations and/or larger memories the reduced-state BCJR algorithm [24] is to be used instead. Moreover, we have shown in [25] that approximating $\max^*(x,y)$ by $\max(x,y)$ in the regime when intrachannel nonlinearities dominate leads to insignificant performance degradation, while the forward and the backward steps in the BCJR equalizer become the forward and the backward Viterbi equalizers, respectively, which are already implemented at 10 Gb/s. Another interesting approach to reduce the channel memory, introduced recently [4,5], is the backpropagation method. Namely, in point-to-point links, the receiver knows the dispersion map configuration and can propagate the received signal through the dispersion map with fiber parameters [group velocity dispersion (GVD), second-order GVD, and nonlinearity coefficient] of opposite signs to that used in the original map. However, the nonlinear interaction of the amplified spontaneous emission (ASE) noise and Kerr nonlinearities cannot be compensated for. Moreover, the complexity of this approach is about 2 orders of magnitude higher than that of the linear equalizer, as shown in [5]. The best strategy would be to use the coarse backpropagation (with a reasonably small number of coefficients) to reduce the channel memory and compensate for the remaining channel distortions by the turbo equalization scheme introduced in this section. Another advantage of turbo equalization is that it can operate even when the conditional estimates obtained from the instanton–Edgeworth expansion method [20] are not perfect because we iterate extrinsic information between soft decoders and the BCJR equalizer until the valid codewords are obtained (or a predetermined number of iterations has been reached), which improves the system performance.

III. Large-Girth Quasi-Cyclic LDPC Codes

The parity-check matrix H of quasi-cyclic LDPC codes considered in this paper can be represented by

$$H = \begin{bmatrix} I & I & I & \cdots & I \\ I & P^{S[1]} & P^{S[2]} & \cdots & P^{S[c-1]} \\ I & P^{2S[1]} & P^{2S[2]} & \cdots & P^{2S[c-1]} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ I & P^{(r-1)S[1]} & P^{(r-1)S[2]} & \cdots & P^{(r-1)S[c-1]} \end{bmatrix}, \quad (7)$$

where I is a $p \times p$ (p is a prime number) identity matrix, P is a $p \times p$ permutation matrix ($p_{i,i+1}=p_{p,1}=1$, $i=1,2,\ldots,p-1$; other elements of P are zeros), and r and c represent the number of rows and columns in Eq. (7), respectively. The set of integers S are to be carefully chosen from the set $\{0,1,\ldots,p-1\}$ so that the cycles of short length, in the corresponding Tanner (bipartite) graph representation of Eq. (7), are avoided. (The Tanner graph of a code is created according to the following rule: check (function) node c is connected to variable (bit) node v whenever element h_{cv} in the corresponding parity-check matrix H is a 1.) We have shown in [3] that large-girth (the shortest cycle in the bipartite graph), $g \geqslant 10$, LDPC codes provide excellent improvement in coding gain over the corresponding turbo-product codes (TPCs). At the same time, the complexity of the LDPC decoders described in [23] is lower than that of TPCs, selecting LDPC codes as excellent candidates for application to systems for 40 and 100 Gb/s transmission. For example, by selecting $p=1123$ and $S=\{0,2,5,13,20,37,58,91,135,160,220,292,354,712,830\}$, an LDPC code of rate 0.8, girth $g=10$, column weight 3, and length $N=16{,}845$ is obtained, which is used later in Section IV.

IV. Mitigation of Intrachannel Nonlinearities via LDPC-Coded Turbo Equalization Based on the Multilevel BCJR Algorithm

We are turning our attention to the evaluation of the proposed scheme in suppression of intrachannel nonlinearities, which are the most challenging impairments to mitigate. For this purpose we developed a realistic fiber-optic communication system model based on the nonlinear Schrödinger equation that was solved using the split-step Fourier method. This model takes into account Kerr nonlinearities, nonlinear phase noise, stimulated Raman scattering, dispersion effects, ASE noise, linear filtering effects, intersymbol interference, and linear cross-talk effects.

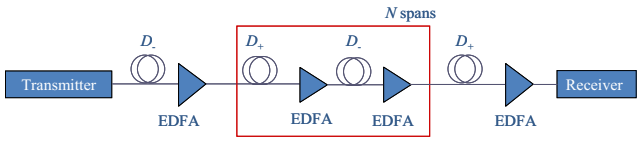

Fig. 4. (Color online) Dispersion map under study.

The submarinelike dispersion map under study is shown in Fig. 4, which is chosen in such a way that the presence of intrachannel nonlinearities is predominant. The span length is set to $L=120$ km, and each span consists of $2L/3$ km of D_+ fiber followed by $L/3$ km of D_- fiber. Precompensation of −1600 ps/nm and the corresponding postcompensation are also applied. The parameters of D_+ and D_- fibers, used in simulations, are given in Table I. The QPSK modulation format is observed with return-to-zero (RZ) pulses having a duty cycle of 33% ($\sqrt{(P_p)}\cos\{(\pi/2)\sin[\pi R_s(t-0.5/R_s)]\}$, R_s is the symbol rate, and P_p is the peak power), and the averaged launched power being set to 0 dBm. Erbium-doped fiber amplifiers (EDFAs) with a noise figure of 5 dB are deployed after every fiber section, the bandwidth of the optical filter is set to $3R_l$ and the bandwidth of the electrical filter to $0.7R_l$, with R_l being the line rate that is defined as the symbol rate of 50 gigasymbols/s (the notation gigasymbol/s is shortened in the rest of the paper to GS/s) divided by a code rate. The line rate is appropriately chosen so that the effective aggregate information rate is 100 Gb/s. The scheme is therefore evaluated for use in 100 Gb/s transmission and 100 G Ethernet.

The results of Monte Carlo simulations (performed by parallel computing on several Dual Intel Quad-Core Xeon CPUs) for a single-channel single-polarization optical QPSK transmission system with Gray mapping operating at 50 GS/s, with the dispersion map described above, are shown in Fig. 5 for a BCJR detector depth of 128 symbols. The number of spans was changed from 4 to 84, the uncoded BER at 50 GS/s and BER after iterative decoding at line rate $R_l=R_s/R$ ($R_s=50$ GS/s, R is the code rate) were calculated and are given in Fig. 5 against the total transmission length. The number of inner (LDPC decoder) iterations was set to 25, and the number of outer (multilevel BCJR–LDPC decoder) iterations was set

TABLE I
FIBER PARAMETERS

Parameters	D_+ Fiber	D_- Fiber
Dispersion [ps/(nm km)]	20	−40
Dispersion slope [ps/(nm^2 km)]	0.06	−0.12
Effective cross-sectional area [μm^2]	110	50
Nonlinear refractive index [m^2/W]	2.6×10^{-20}	2.6×10^{-20}
Attenuation coefficient [dB/km]	0.19	0.25

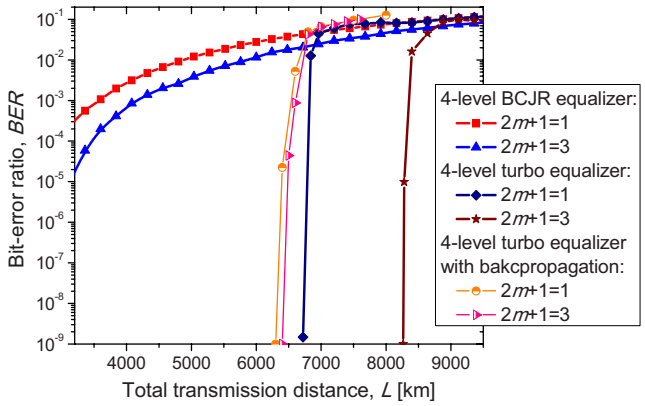

Fig. 5. (Color online) BER performance of turbo equalizer based on 4-level BCJR equalizer for QPSK with Gray mapping. The dispersion map used for backpropagation is different from that shown in Fig 4. It is based on only single-mode fiber with EDFAs deployed every 100 km.

to 3. We can see that for 22 spans the 4-level BCJR equalizer with memory $2m+1=3$ provides more than 1 order in magnitude improvement in the BER over the memoryless case ($2m+1=1$). For the turbo equalization scheme based on the four-level BCJR equalizer of memory $2m+1=1$ and the LDPC(16935,13550) code of girth 10 and column weight 3 (described in Section III), we achieve transmission over 55 spans (6600 km) without any error. On the other hand, for the turbo equalization scheme based on the four-level BCJR equalizer of memory $2m+1=3$ and the same LDPC code, we are able to achieve even 8160 km of error-free transmission at an aggregate rate of 100 Gb/s. Because we are concerned with 100 Gb/s per wavelength transmission, only QPSK was studied.

In Fig. 5 we also report the results obtained by turbo equalization with backpropagation. The dispersion map was composed of standard single-mode fiber (SMF) only with EDFAs of noise figure 6 dB being deployed every 100 km. We can see that backpropagation can indeed reduce the channel memory because the improvement for $m=1$ is small compared with the case of $m=0$.

V. IID INFORMATION CAPACITY

In this section we address the problem of calculating the channel capacity for an IID information source, in the literature also known as the achievable information rate (see [3,14,15] and references therein). The IID channel capacity represents a lower bound on channel capacity. To calculate the IID channel capacity, similarly we model the whole transmission system as the nonlinear ISI channel, in which the m previous symbols and the next m symbols influence the observed symbol. The optical communication system is characterized by the conditional PDF of the

output complex vector of samples $\mathbf{y}=(y_1,\ldots,y_n,\ldots)$, where $y_i=(\mathrm{Re}\{y_i\},\mathrm{Im}\{y_i\})\in\mathbf{Y}$ ($\mathbf{Y}$ represents the set of all possible output samples), given the source sequence of M-ary symbols $\mathbf{x}=(x_1,\ldots,x_n,\ldots)$, $x_i\in\mathbf{X}=\{0,1,\ldots,M-1\}$. The $\mathrm{Re}\{y_i\}$ corresponds to the in-phase channel sample, and the $\mathrm{Im}\{y_i\}$ represents the quadrature channel sample. The information rate, expressed in bits/channel use, can be calculated by [26,27]:

$$I(\mathbf{Y};\mathbf{X})=H(\mathbf{Y})-H(\mathbf{Y}|\mathbf{X}), \tag{8}$$

where $H(\mathbf{U})=E[\log_2 P(\mathbf{U})]$ denotes the entropy of a random variable $\mathbf{U}$ and $E(\cdot)$ denotes the mathematical expectation operator. By using the Shannon–McMillan–Brieman theorem that states[1] [26]:

$$E[\log_2 P(\mathbf{Y})]=\lim_{n\to\infty}(1/n)\log_2 P(\mathbf{y}[1,n]), \tag{9}$$

the information rate can be determined by calculating $\log_2[P(\mathbf{y}[1,n])]$, by propagating the sufficiently long source sequence. By substituting Eq. (9) into Eq. (8) we obtain the following expression suitable for practical calculation of IID information capacity:

$$I(\mathbf{Y};\mathbf{X})=\lim_{n\to\infty}\frac{1}{n}\left[\sum_{i=1}^{n}\log_2 P(y_i|\mathbf{y}[1,i-1],\mathbf{x}[1,n])-\sum_{i=1}^{n}\log_2 P(y_i|\mathbf{y}[1,i-1])\right]. \tag{10}$$

Based on Section II, the conditional PDF $P(y_i|\mathbf{y}[1,i-1],\mathbf{x}[1,n])$ is related to $P(y_i|\mathbf{x}[i-m,i+m])=P(y_i|\mathbf{s})$. Therefore, the first term in Eq. (10) can be straightforwardly estimated from transition PDFs $P(y_i|\mathbf{s})$. To calculate $\log_2 P(y_i|\mathbf{y}[1,i-1])$ we use the forward recursion of the multilevel BCJR algorithm we described in Section II, wherein the forward metric and backward metric are modified as follows:

$$\alpha_j(\mathbf{s})=\max_{\mathbf{s}'}{}^{*}[\alpha_{j-1}(\mathbf{s}')+\gamma_j(\mathbf{s}',\mathbf{s})-\log_2 M],$$

$$\gamma_j(\mathbf{s}',\mathbf{s})=\log[p(y_j|\mathbf{x}[j-m,j+m])]. \tag{11}$$

The ith term $\log_2 P(y_i|\mathbf{y}[1,i-1])$ can be calculated iteratively:

$$\log_2 P(y_i|\mathbf{y}[1,i-1])=\max_{\mathbf{s}}{}^{*}\,\alpha_i(\mathbf{s}), \tag{12}$$

where the max* operator was applied for all $\mathbf{s}\in\mathbf{S}$ ($\mathbf{S}$ denotes the set of states in the trellis shown in Fig. 1).

This method is applicable to both memoryless channels and for channels with memory. In Fig. 6 we report the information capacities for different signal constellation sizes and two types of QAM constellations, square QAM and star QAM [28], by observing a linear channel model. These results are obtained by using Eq. (10) and are consistent with the digital communication literature [29], and they are provided as verification of the method. We can see that information capacity can be closely approached even with the IID source providing that constellation size is sufficiently large. It is interesting to notice that the star QAM outperforms the corresponding square QAM for low and medium signal-to-noise ratios (SNRs), whereas for high SNRs the square QAM outperforms the star QAM.

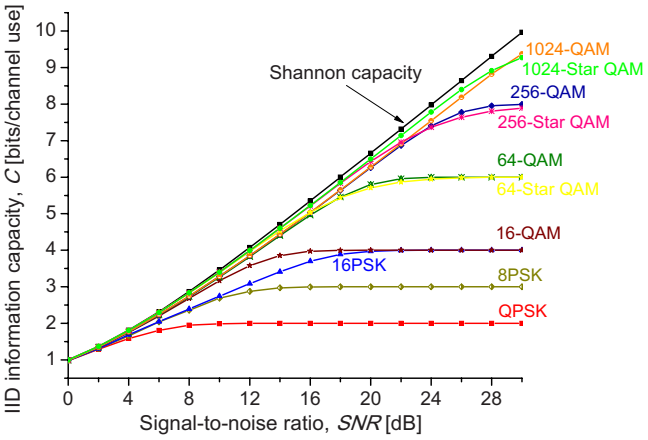

Fig. 6. (Color online) IID information capacities for linear channel model and different signal constellation sizes. (64-star QAM contains 8 rings with 8 points each, 256-star QAM contains 16 rings with 16 points, and 1024-star QAM contains 16 rings with 64 points.) The SNR is defined as E_s/N_0, where E_s is the symbol energy and N_0 is the power spectral density.

In Fig. 7 we show the IID channel capacity against the total transmission length (obtained by Monte

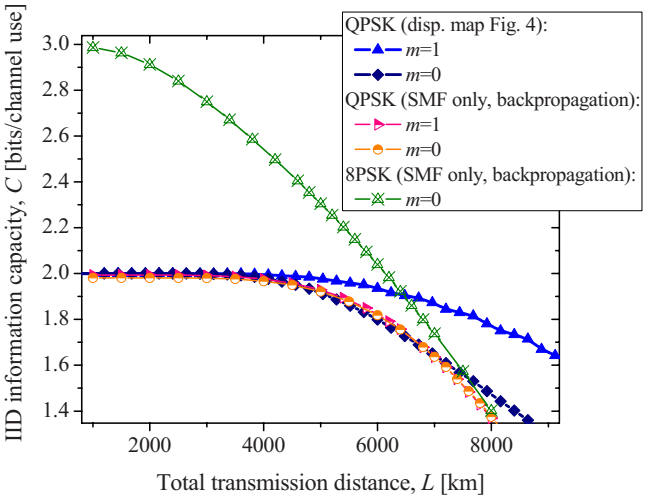

Fig. 7. (Color online) IID information capacity for QPSK and 8PSK with a symbol rate of 50 GS/s against the transmission distance.

[1]Our model is a particular instance of McMillan's discrete stationary channel model [27]. The fiber-optics channel model described in this paper can be considered stationary for sufficiently long information sequences and sufficiently long channel memory assumptions. Notice that other papers on channel capacity [4,16–19] also consider the channel as stationary because all calculations of channel capacity assume that the corresponding transition PDFs are independent of the initial moment.

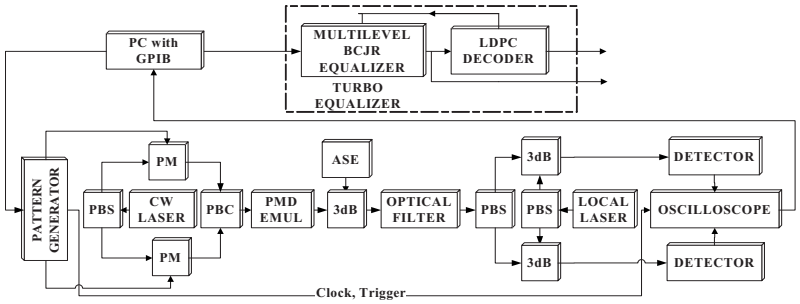

Fig. 8. Experimental setup for the polarization multiplexed BPSK study. CW Laser, continuous-wave laser; PM, phase modulator; ASE, amplified spontaneous emission noise source; 3 dB, 3 dB coupler.

Carlo simulations), for the dispersion map shown in Fig. 4 and the QPSK modulation format with an aggregate data rate of 100 Gb/s for two different memories in the multilevel BCJR equalizer. We see that by using an LDPC code (with a rate of $R=0.8$) longer than that used in Fig. 5, we are able to extend the transmission distance by 600 km for a BCJR equalizer memory of $m=0$ and even by 1440 km (resulting in a total transmission length 9600 km) for a BCJR equalizer memory of $m=1$. The transmission distance can further be increased by observing the larger memory channel assumptions, which requires higher computational complexity. On the other hand, we can use the backpropagation approach [4,5] to keep the channel memory reasonably low and then apply the method described in this section. In Fig. 7 we also show the IID information capacity, when the backpropagation method is used, for a dispersion map composed of standard SMF only with EDFAs with a noise figure of 6 dB being deployed every 100 km.

VI. PMD Compensation by Multilevel Turbo Equalization

Figure 8 shows the experimental setup for PMD compensation study in polarization multiplexed schemes with coherent detection. In this example, we jointly perform detection and decoding of symbols transmitted in two orthogonal polarizations. The two orthogonal polarizations of a continuous-wave laser source are separated by a polarization beam splitter and are modulated by two phase modulators (Covega) driven at 10 Gb/s (Anritsu MP1763C). (The symbol rate was determined by available equipment.) A precoded test pattern was loaded into the pattern generator via personal computer with a general purpose interface bus (GPIB). A polarization beam combiner was used to combine the two modulated signals, followed by a PMD emulator (JDSU PE3), which introduced a controlled amount of differential group delay (DGD) to the signal. Then the signal distorted by PMD was mixed with a controlled amount of ASE noise with a 3 dB coupler. The modulated signal level was maintained at 0 dB while the ASE power level was changed to obtain different optical signal-to-noise ratios (OSNRs). Next, the optical signal was preamplified, filtered (JSDU 2 nm bandpass filter) and coherently detected. The coherent detection is performed by mixing the received signal with a signal from a local laser with a 3 dB coupler. The resulting signal is detected with a detector (Agilent 11982A) and an oscilloscope (Agilent DCA 86105A), triggered by the data pattern that was used to acquire the samples. To maintain a constant power of −6 dBm at the detector, a variable attenuator was used. Data was transferred via the GPIB back to the PC. The PC also served as a multilevel turbo equalizer with offline processing. To avoid any imbalance of two independent symbols transmitted in two polarizations, we detect both symbols simultaneously. Because the symbols transmitted in both polarizations are considered one supersymbol, the BER performance of the turbo equalizer is independent of the power splitting ratio between the principle states of polarization.

The experimental results for the BER performance of the proposed multilevel turbo equalizer are summarized in Fig. 9. For the experiment, a quasi-cyclic LDPC(16935, 13550) code of girth 10 and column weight 3 (introduced in Section III) was used as the

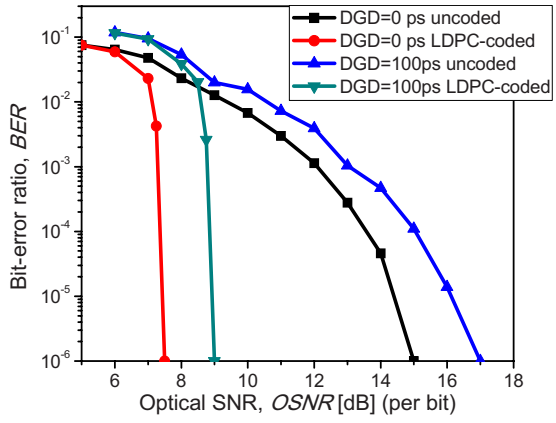

Fig. 9. (Color online) BER performance of the multilevel turbo for PMD cpmpensation.

channel code. The number of extrinsic iterations between the LDPC decoder and the BCJR equalizer was set to 3, and the number of the intrinsic LDPC decoder iterations was set to 25. The state memory of $2m+1=3$ was sufficient for the compensation of the first-order PMD with a DGD of 100 ps. The OSNR penalty for 100 ps of DGD is 1.5 dB at a BER of 10^{-6}. The coding gain for a DGD of 0 ps is 7.5 dB at a BER of 10^{-6}, and the coding gain for a DGD of 100 ps is 8 dB. In the experiment, synchronized external-cavity lasers were used as transmitting and local lasers, so that PMD is a dominant effect.

VII. Conclusion

In this paper we have proposed the turbo equalization scheme based on a maximum *a posteriori* probability (MAP) turbo equalizer implemented using the sliding-window *multilevel* BJCR algorithm. The proposed scheme is an excellent equalization candidate for use in a multilevel coded-modulation scheme with coherent detection. This scheme is suitable for *simultaneous* nonlinear and linear impairment mitigation. The proposed scheme employs large-girth quasi-cyclic LDPC codes as channel codes. We show that even a transmission distance of 8160 km with an aggregate rate of 100 Gb/s, based on QPSK modulation, can be achieved without any countable error, when a BJCR equalizer of memory $m=1$ is employed and LDPC(16935,13550) code is used as the channel code. With polarization multiplexing, even two channels of aggregate rate 100 Gb/s can be transmitted using the same wavelength channel. The IID information capacity results indicate that by using longer LDPC codes of the same code rate ($R=0.8$) and the same BCJR equalizer (of memory $m=1$) the total transmission distance can be extended for an additional 1440 km, resulting in a total of 9600 km. Notice that similar dispersion maps are in use in optical transmission systems operating at 10 Gb/s. Therefore, the proposed multilevel turbo equalizer can be used for 100 Gb/s upgrade using the installed 10 Gb/s optical transmission systems. By using the coded-modulation scheme proposed here not only transmission but also all signal processing related to detection and decoding are effectively done at lower symbol rates (e.g., 50 GS/s), where dealing with nonlinear effects and PMD is more manageable, while keeping the aggregate rate at 100 Gb/s or above. Notice that the complexity of this equalizer grows exponentially as the state memory and signal constellation sizes increase. An alternative approach to perform a nonlinear equalization is based on backpropagation. This method, however, has complexity that is about 2 orders of magnitude higher than that of a linear equalizer as shown by Ip and Kahn [5] and cannot account for the nonlinear ASE noise–Kerr nonlinearities interaction. To simultaneously solve both problems, we proposed in Section II the use of coarse backpropagation (with a reasonably small number of coefficients) to reduce the channel memory and compensate for the remaining channel distortions by the turbo equalization proposed in this paper. We have also shown in Section VI that this scheme can compensate for any imbalance in polarization and phase. Finally, we described a method to evaluate IID information capacities, which is applicable to both memoryless channels and channels with memory.

Acknowledgments

This paper was supported in part by the National Science Foundation (NSF) under grant IHCS-0725405 and in part by NEC Laboratories America, Princeton, New Jersey, USA.

References

[1] S. J. Savory, "Digital filters for coherent optical receivers," *Opt. Express*, vol. 16, pp. 804–817, Jan. 2008.

[2] N. Alic, G. C. Papen, R. E. Saperstein, R. Jiang, C. Marki, Y. Fainman, and S. Radic, "Experimental demonstration of 10 Gb/s NRZ extended dispersion-limited reach over 600 km-SMF link without optical dispersion compensation," in *Optical Fiber Communication Conf. and Expo. and the Nat. Fiber Optic Engineers Conf.*, 5–10 March 2006, paper OWB7.

[3] 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.

[4] R.-J. Essiambre, G. J. Foschini, G. Kramer, and P. J. Winzer, "Capacity limits of information transport in fiber-optic networks," *Phys. Rev. Lett.*, vol. 101, paper 163901, Oct. 2008.

[5] E. Ip and J. M. Kahn, "Nonlinear impairment compensation using backpropagation," in *Optical Fibre, New Developments*. Vienna, Austria: In-Tech, to be published.

[6] W. Shieh, X. Yi, Y. Ma, and Y. Tang, "Theoretical and experimental study on PMD-supported transmission using polarization diversity in coherent optical OFDM systems," *Opt. Express*, vol. 15, pp. 9936–9947, 2007.

[7] I. B. Djordjevic, L. Xu, and T. Wang, "Simultaneous chromatic dispersion and PMD compensation by using coded-OFDM and girth-10 LCPC codes," *Opt. Express*, vol. 16, no. 14, pp. 10269–10278, July 2008.

[8] I. B. Djordjevic, "Suppression of intrachannel nonlinearities in high-speed WDM systems," in *Advanced Technologies for High-Speed Optical Communications*, L. Xu, ed. Trivandrum-Kerala, India: Research Signpost, 2007, pp. 247–277.

[9] M. P. C. Fossorier, "Quasi-cyclic low-density parity-check codes from circulant permutation matricies," *JETP*, vol. 50, pp. 1788–1794, Aug. 2004.

[10] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. IT-20, pp. 284–287, Mar. 1974.

[11] W. E. Ryan, "Concatenated convolutional codes and iterative decoding," in *Wiley Encyclopedia of Telecommunications*, J. G. Proakis, ed. New York: Wiley, 2002.

[12] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 40, pp. 1727–1737, Oct. 2001.

[13] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications*. Upper Sadle River: Pearson Prentice-Hall, 2004.

[14] I. B. Djordjevic, B. Vasic, M. Ivkovic, and I. Gabitov, "Achievable information rates for high-speed long-haul optical transmission," *J. Lightwave Technol.*, vol. 23, pp. 3755–3763, Nov. 2005.

[15] M. Ivkovic, I. B. Djordjevic, and B. Vasic, "Calculation of achievable information rates of long-haul optical transmission systems using instanton approach," *J. Lightwave Technol.*, vol. 25, pp. 1163–1168, May 2007.

[16] J. M. Kahn and K.-P. Ho, "Spectral efficiency limits and modulation/detection techniques for DWDM systems," *IEEE J. Sel. Top. Quantum Electron.*, vol. 10, no. 2, pp. 259–272, Mar./Apr. 2004.

[17] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fiber communications," *Nature*, vol. 411, no. 6841, pp. 1027–1030, June 2001.

[18] K. S. Turitsyn, S. A. Derevyanko, I. V. Yurkevich, and S. K. Turitsyn, "Information capacity of optical fiber channels with zero average dispersion," *Phys. Rev. Lett.*, vol. 91, no. 20, paper 203901, Nov. 2003.

[19] J. Tang, "The channel capacity of a multispan DWDM system employing dispersive nonlinear optical fibers and an ideal coherent optical receiver," *J. Lightwave Technol.*, vol. 20, no. 7, pp. 1095–1101, July 2002.

[20] M. Ivkovic, I. Djordjevic, P. Rajkovic, and B. Vasic, "Pulse energy probability density functions for long-haul optical fiber transmission systems by using instantons and Edgeworth expansion," *IEEE Photon. Technol. Lett.*, vol. 19, no. 20, pp. 1604–1606, Oct. 2007.

[21] J. Hou, P. H. Siegel, L. B. Milstein, and H. D. Pfitser, "Capacity approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2141–2155, Sept. 2003.

[22] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, pp. 927–946, May 1998.

[23] H. Xiao-Yu, E. Eleftheriou, D.-M. Arnold, and A. Dholakia, "Efficient implementations of the sum-product algorithm for decoding of LDPC codes," in *Proc. IEEE GLOBECOM*, vol. 2, Nov. 2001, pp. 1036–1036E.

[24] G. Colavolpe, G. Ferrari, and R. Raheli, "Reduced-sate BCJR-type algorithms," *IEEE J. Sel. Areas Commun.*, vol. 19, pp. 848–858, May 2001.

[25] I. B. Djordjevic and B. Vasic, "Nonlinear BCJR equalizer for suppression of intrachannel nonlinearities in 40 Gb/s optical communication systems," *Opt. Express*, vol. 14, pp. 4625–4635, 2006.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[27] F. M. Reza, *An Introduction to Information Theory*. New York: McGraw-Hill, 1961.

[28] W. T. Webb, and R. Steele, "Variable rate QAM for mobile radio," *IEEE Trans. Commun.*, vol. 43, pp. 2223–2230, July 1995.

[29] J. G. Proakis, *Digital Communications*. Boston: McGraw Hill, 2001.