Performance of Turbo Equalizers for Optical PMD Channels

Mark Jäger, Tobias Rankl, Joachim Speidel, *Member, IEEE*, Henning Bülow, and Fred Buchali

***Abstract*—This paper investigates the performance of iterative (turbo) equalization to mitigate the effects of a polarization-mode dispersion (PMD) in nonreturn-to-zero (NRZ) intensity-modulated optical-fiber transmission systems. A PMD can lead to severe distortions in the received electrical signal and is a key limiter for the development of high-bit-rate transmission over currently used fibers. In order to reduce the distortions due to a PMD, the performance of symbol-by-symbol maximum *a posteriori* (sbs-MAP) soft-in/soft-out (SISO) decoders is studied. The SISO algorithms are adapted to the noise statistics of the optical channel where the photo detector leads to a non-Gaussian signal-dependent noise at the receiver side. The modified SISO algorithms are successfully employed for turbo equalization and results show that iterative (turbo) equalization and decoding for the compensation of a PMD can lead to a tremendous reduction in the bit error ratio (BER). Moreover, it is shown that, due to the robustness of mutual information, the extrinsic information transfer (EXIT) chart can be applied for the design of iterative receivers in optical transmission systems even with a non-Gaussian noise.**

***Index Terms*—Iterative equalization, optical communications, polarization-mode dispersion (PMD), signal-dependent noise.**

I. Introduction

HIGH-BIT-RATE optical transmission systems suffer from several sources of signal distortion, resulting in intersymbol interference (ISI) at the receiver. The main limitation in cases of bit rates beyond 10 Gb/s is the polarization-mode dispersion (PMD) [1]. In contrast to static distortions, e.g., caused by a chromatic dispersion (CD), the PMD is time variant, which makes its compensation more difficult.

Several optical and electrical compensators for the PMD have been developed within the last years. A growing interest is on electrical equalizers, because they have the potential to be implemented as low-cost integrated circuits [2]. Electrical adaptive threshold equalizers and feed-forward and decision feedback equalizers have been studied for 10-Gb/s transmission systems [3]. With the increasing bit rates, advanced equalization and detection schemes such as the maximum likelihood sequence detection [4] or the iterative decoding techniques [5], [6] are of paramount interest for optical transmissions. In this paper, symbol-by-symbol maximum *a posteriori* (sbs-MAP) soft-in/soft-out (SISO) decoders, which can be used in iterative decoding schemes, are studied in detail for optical channels. At the receiver, the photodiode leads to a signal-dependent noise in the electrical signal; therefore, the SISO algorithms have to be adapted to the channel noise statistics. Moreover, the modified algorithms are used to investigate the performance of turbo equalization for optical channels.

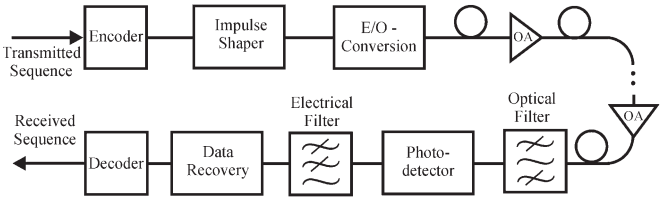

Fig. 1. Model of an optical communication system.

This paper is organized as follows. Section II introduces the optical communication system corrupted by a PMD. A low-pass equivalent model of the optical channel is used to derive the probability density function (PDF) of the noisy signal at the output of the receiver. Then, a discrete-time nonlinear state-based system model is derived to mathematically describe the impairments due to a PMD. A chromatic dispersion of the link is assumed to be fully compensated. The SISO algorithms are modified according to the channel noise statistics in Section III. Since high-bit-rate and/or low-cost applications require a reduced complexity, suboptimal sbs-MAP algorithms are also presented. The SISO algorithms are then considered for the equalization of the optical PMD channel. Section IV presents the results for the iterative equalization and decoding of optical channels corrupted by a PMD. It is shown that the extrinsic information transfer (EXIT) chart can be used to study the convergence of iterative decoding schemes, and this chart is applied to find suitable codes and receiver designs. The conclusion is drawn in Section V.

II. Optical System Model

A simplified block diagram of an optical-fiber communication system is shown in Fig. 1. At the transmitter, the encoder with a code rate $R = K/(K+1)$ adds a defined amount of redundancy to the messages by encoding K information bits into a code word of length $K+1$. Throughout this paper, nonreturn-to-zero (NRZ) impulse shapes with ON–OFF keying (OOK) and a finite optical extinction ratio (ER) ζ are assumed. Optical amplifiers (OAs) are inserted periodically along the transmission fiber to compensate for transmission losses. When

Manuscript received June 9, 2005; revised October 9, 2005.

M. Jäger, T. Rankl, and J. Speidel are with the University of Stuttgart, Institute of Telecommunications, Stuttgart 70569, Germany (e-mail: rankl@inue.uni-stuttgart.de).

H. Bülow and F. Buchali are with Alcatel Research and Innovation, Alcatel Standard Elektrik Lorenz Aktiengesellschaft (SEL AG), Stuttgart D-70499, Germany.

Digital Object Identifier 10.1109/JLT.2005.863281

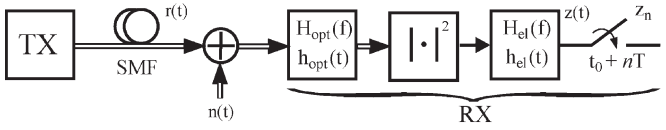

Fig. 2. Equivalent low-pass model of the optical communication system.

the optical signal passes through each OA, an amplified spontaneous emission (ASE) noise is accumulated and becomes the dominant noise.[1] At the receiver, the optical signals are sent through an optical-bandpass filter, followed by a photodiode and an electrical low-pass filter.

A. Equivalent Low-Pass Model

To derive the noise distribution for the sampled electrical output signal, the optical and electrical signals are represented by low-pass equivalents. The block diagram is shown in Fig. 2, where $H_{\rm opt}(f)$ and $H_{\rm el}(f)$ are the low-pass equivalent transfer functions of optical and electrical filters, respectively. The noise from the cascaded OA can be approximately modeled as a single complex additive white Gaussian noise (AWGN) noise source $n(t)$ with a power spectral density N_0 per polarization axis. For a binary signal with a mean power of 1/2, the optical signal-to-noise ratio (OSNR) is defined in front of the optical filter $H_{\rm opt}(f)$ as

$$\text{OSNR} = 10\log\frac{1}{8\sigma^2} - 10\log R \tag{1}$$

where $2\sigma^2 = N_0 B_{\rm ref}$ and $B_{\rm ref}$ is used as reference bandwidth. In [7], numerical methods are presented to compute the PDF of the output sequence z_n taking into account arbitrary transfer functions of the optical and electrical filters. To study the principles of optical transmission channels, a closed form for the PDF of z_n is advantageous. In the following discussion, an ideal optical bandpass of a rectangular shape with bandwidth B and an electrical low-pass integrate-and-dump filter is assumed [2], [4], [8]. For simulation purposes, the bandwidths B and $B_{\rm ref}$ are equated. Optical and electrical filters are chosen such that no interference between neighboring bits at the sampling instants $t_0 + nT$ is caused by these filters, and the output PDF of z_n obeys a noncentral chi-square distribution with N degrees of freedom (DOFs) [4]

$$f(z|y=P) = \frac{1}{2\sigma^2}\left(\frac{z}{P}\right)^{\frac{N-2}{4}} \exp\left(-\frac{z+P}{2\sigma^2}\right) I_{\frac{N}{2}-1}\left(\frac{\sqrt{z\cdot P}}{\sigma^2}\right) \tag{2}$$

where y and its realization P is the signal part of the noisy z, $I_\alpha(x)$ is the αth-order modified Bessel function of the first kind, and N is proportional to the optical bandwidth B and the bit duration T, i.e., $N = 2(BT+1)$. Note, (2) is only valid for $BT \gg 1$. To simplify the notation, we have dropped index n.

[1]Therefore, other noise sources such as a shot noise or a thermal noise, which are caused by the photodiode and the following electronic circuitry, are neglected.

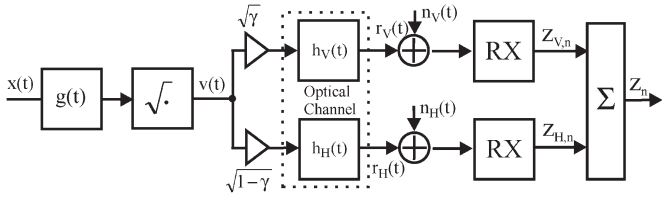

Fig. 3. Equivalent low-pass optical channel model for a PMD (RX, Fig. 2).

B. Channel Model for PMD

A fiber-based PMD is caused by an asymmetry of the fiber core in which the light polarized in one axis travels faster than the light polarized in the orthogonal axis. In any ordinary single-mode fiber (SMF), there exist two independent orthogonal propagation modes called the principle states of polarization (PSPs). The difference of the group delays of the two PSPs is referred to as a differential group delay (DGD) $\Delta\tau$. Since the first-order PMD is considered as the dominant PMD effect at 10 Gb/s [9], the effects of higher order PMD are neglected. Note that the algorithms investigated in this paper are not restricted to the first-order PMD. They can be used for the equalization of any kind of distortion like a CD higher order PMD, as well as for nonlinear distortions. Fig. 3 shows a general approach to extend the low-pass equivalent model shown in Fig. 2 to take into account the effects of PMD [4]. Let $x(t)$ denote the transmitted bit sequence with $x(t) = \sum_{n=-\infty}^{\infty} x_n \cdot \delta(t-nT)$, where $x_n \in \{0,1\}$ is the transmitted bit at a discrete-time instant n. The stream of input bits is sent through an impulse shaper with an impulse response $g(t)$. For electrooptical conversion, a linear intensity modulation is assumed. Therefore, the modulator can be modeled by taking the square root of the incoming electrical signal. The optical (field strength) signal $v(t)$ is then split according to the power distribution γ into the two PSPs. The optical channel is modeled by two different impulse responses $h_{\rm H}(t)$ and $h_{\rm V}(t)$ for the horizontal and the vertical polarization states, respectively. Complex AWGN noises $n_{\rm V}(t)$ and $n_{\rm H}(t)$ are added to each polarization component independently, and the resulting signals are passed through the receiver (RX), which was introduced in Fig. 2. The PDF of the output z_n is the sum of two statistically independent noncentral chi-square-distributed variables $z_{{\rm V},n}$ and $z_{{\rm H},n}$, each with $N' = (BT+1)$ DOF. Therefore, z_n has also a noncentral chi-square distribution with $N = 2N'$ DOF and variance $\sigma_z^2 = 2N\sigma^4 + 4\sigma^2\cdot(P_{\rm V}+P_{\rm H})$, where $P_{\rm V}$ $(P_{\rm H})$ is the signal power of the vertically (horizontally) polarized light at the output of the optical transmission channel. For the first-order PMD, the optical channel impulse responses $h_{\rm V}(t)$ and $h_{\rm H}(t)$ for the two PSPs are given by $h_{\rm V}(t) = \delta(t - (\Delta\tau/2))$ and $h_{\rm H}(t) = \delta(t + (\Delta\tau/2))$, respectively. The overall system impulse response[2] at the receiver output in Fig. 2 is assumed to be $z(t) =: \widetilde{h}(t)$

$$\widetilde{h}(t) = \begin{cases} \cos^2\left(\frac{\pi t}{2T}\right), & \text{for } -T \le t < T \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

[2]A detailed derivation of the system impulse response for an optical transmission system with a square law detector can be found in [10].

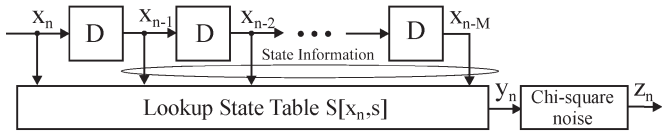

Fig. 4. Block diagram of the state-based channel model with a chi-square noise generator.

In the following discussion, a power distribution of $\gamma = 0.5$ is considered.[3] In order to detect the highest amplitude of the transmitted signal, the optimal sampling time offset in Fig. 2 is simply $t_0 = 0$ for $\gamma = 0.5$. If the condition $\Delta\tau/2 < T$ is satisfied, the bit x_n in the received signal is only corrupted by the precursor x_{n-1} and the postcursor x_{n+1}. Let $x(t) = \delta(t)$; then, the output signal of the overall sampled system in Fig. 3 is the discrete-time impulse response $z_n =: h_n$ with

$$h_n = \frac{1}{2} \cdot \sum_{\widetilde{n}=n-1}^{n+1} \widetilde{h}\left((\widetilde{n}-n)T - \frac{\Delta\tau}{2}\right) + \widetilde{h}\left((\widetilde{n}-n)T + \frac{\Delta\tau}{2}\right). \tag{4}$$

Using (3) and (4) results in

$$h_n = (\delta_{n-1} + \delta_{n+1}) \cdot \frac{1}{2}\cos^2\left(\frac{\pi}{2} + \frac{\pi\Delta\tau}{2T}\right) + \delta_n \cdot \cos^2\left(\frac{\pi\Delta\tau}{2T}\right) \tag{5}$$

where δ_k is the Kronecker delta.

C. State Transition Channel Model

The optical communication link in Fig. 3, in case of no optical noise, can also be described by the discrete-time state-based model with a memory length M in Fig. 4, which generates the output value $y_n = S[x_n, s]$ depending on the transmitted bit x_n and the state s of the shift register [2]. The noise is generated by a separate generator. The relation between the input x_n, the current state s, and the corresponding output y_n is implemented as a lookup table $S[x_n, s]$. This is a general approach, which also allows the nonlinearities of the optical fiber to be taken into account.

If $M = 2$ in Fig. 4, the channel model can also be described by the state-transition diagram shown in Fig. 5. The vertices are labeled with the input bit x_n and the corresponding output PDF $f_{x_n,s}(z_n) = f(z_n|x_n, s)$ of the decision variable z_n. With the simplifying assumptions for the electrical and optical filters, the PDF $f_{x_n,s}(z_n)$ is given by (2), where $P = S[x_n, s]$. Note that the description of an optical channel with a state-transition diagram is not limited to the filter assumptions made above. For arbitrary transfer functions, the number of states has to be increased and the PDF have to be evaluated numerically. The receivers described in the following section operate on the basis of the model in Fig. 4.

[3]For $\gamma = 0.5$, the PMD leads to the severest distortion in the detected electrical signal.

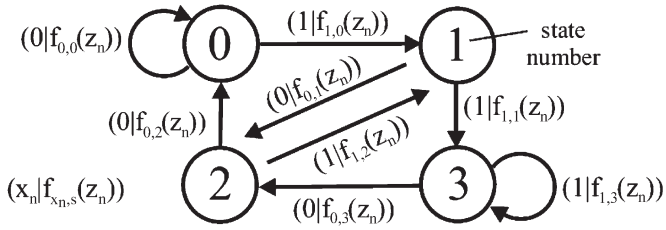

Fig. 5. State-transition diagram of the optical PMD channel.

III. SISO Detectors for Optical Transmission

To obtain an optimal estimate of the transmitted data, two principal detection rules exist.

1) The maximum likelihood sequence detection (MLSD) implemented by the Viterbi algorithm. The sequence error probability is minimized, if the *a priori* sequence probabilities are equal.
2) The sbs-MAP detection implemented by the BCJR algorithm. The bit error probability is minimized, and the *a priori* probability of each bit is required.

In this paper, we focus on equalization algorithms that are based on sbs-MAP. The advantage of sbs-MAP decoders over MLSD decoders is that they can provide reliability information [*a posteriori* probabilities (APPs)] on each transmitted bit as soft-output information (*a posteriori* log-likelihood values) and can accept *a priori* information as an additional input. These properties enable sbs-MAP decoders to be used in iterative-equalization schemes (e.g., turbo equalization), which will be considered in detail in Section IV, or they can be used as a soft-output equalizer in combination with a soft-input forward error correction (FEC) decoder.

A. Log-Domain Bahl–Cocke–Jelinek–Raviv (BCJR) Algorithm

An efficient implementation of the sbs-MAP decoding rule, which operates on the trellis structure of the code, was first introduced in [11] and is called BCJR algorithm. In this section, the BCJR scheme for the equalization of the optical channel corrupted by a PMD, as introduced in Section II, is considered. The key element of the BCJR algorithm is the calculation of the *a posteriori* log-likelihood value (L-value) of each transmitted bit x_n in the sequence $\mathbf{x}$ conditioned on the received sequence $\mathbf{z}$

$$L(x_n|\mathbf{z}) = \ln \frac{P[x_n = 1|\mathbf{z}]}{P[x_n = 0|\mathbf{z}]}. \tag{6}$$

In order to incorporate the trellis structure into the computation, (6) can be rewritten as

$$L(x_n|\mathbf{z}) = \ln \frac{\sum_{X+} f(s_{n-1} = s', s_n = s, \mathbf{z})}{\sum_{X-} f(s_{n-1} = s', s_n = s, \mathbf{z})} \tag{7}$$

where s_n is the encoder state at time n. X_+ and X_- represent the set of pairs (s', s) of the state transition $(s_{n-1} = s') \rightarrow (s_n = s)$, which corresponds to the event $x_n = 1$ and $x_n = 0$, respectively. The direct evaluation of (7) is often numerically unstable for large and even moderate sequence lengths. A stable

version can be obtained in the log domain where the L-value $L(x_n|\mathbf{z})$ is calculated as

$$L(x_n|\mathbf{z}) = \ln \sum_{X+} \exp\left(\widetilde{\alpha}_{n-1}(s') + \widetilde{\gamma}_n(s',s) + \widetilde{\beta}_n(s)\right) - \ln \sum_{X-} \exp\left(\widetilde{\alpha}_{n-1}(s') + \widetilde{\gamma}_n(s',s) + \widetilde{\beta}_n(s)\right) \quad (8)$$

where $\widetilde{\alpha}_n(s)$ is the forward metric and $\widetilde{\beta}_n(s)$ is the backward metric. $\widetilde{\gamma}_n(s',s)$ is the branch metric, which depends on the noise statistics of the channel model. The forward metric for time instant n only depends on the received bits before that time and can be iteratively calculated as

$$\widetilde{\alpha}_n(s) = \ln \sum_{s'} \exp\left(\widetilde{\alpha}_{n-1}(s') + \widetilde{\gamma}(s',s)\right) \quad (9)$$

where the sum is taken over all possible channel states s', which end in state s. The backward metric only depends on the received bits after time instant n

$$\widetilde{\beta}_{n-1}(s') = \ln \sum_{s} \exp\left(\widetilde{\beta}_n(s) + \widetilde{\gamma}(s',s)\right) \quad (10)$$

where the sum is taken over all possible encoder states s at a time instant n, which can be reached from state s' at a time instant $n-1$.

To reduce the computational amount for the summations in (8)–(10), the Jacobian algorithm Jac-Log$(x,y) =: \ln(e^x + e^y) = \max(x,y) + \ln(1+\mathrm{e}^{-|x-y|})$ can be used, where the correction term $r(|x-y|) = \ln(1+\mathrm{e}^{-|x-y|})$ is bounded by $0 < r(|x-y|) \leq \ln(2)$ and can be stored in a look-up table [12], [13]. To further speed up calculations, the correction term can be completely omitted. This approximation is then referred to as a Max-Log. To evaluate the summation in (8)–(10), the Jac-Log(.,.) and Max-Log(.,.) functions can be applied recursively.

The most important parameter for the BCJR algorithm is the branch metric $\widetilde{\gamma}(s',s)$, since all other parameters can be derived from it. $\widetilde{\gamma}_n(s',s)$ can be written as

$$\widetilde{\gamma}_n = \ln\left(P[x_n]\right) + \widetilde{f}\left(z_n|x_n,s'\right) \quad (11)$$

where the event x_n corresponds to the transition from state s' to state s, $P[x_n]$ is the *a priori* probability of the bit x_n, and $\widetilde{f}(z_n|x_n,s') = \ln(f(z_n|x_n,s'))$. Using the L-value notation, $P[x_n]$ can be expressed as [14]

$$P[x_n] = A_n \cdot \exp\left(\frac{(2x_n-1)L(x_n)}{2}\right) \quad (12)$$

where A_n is equal for all transitions from time instant $n-1$ to n. Therefore, A_n will cancel out in the APP calculation in (8), and (11) can be written as

$$\widetilde{\gamma}_n = \frac{(2x_n-1)L(x_n)}{2} + \widetilde{f}\left(z_n|x_n,s'\right). \quad (13)$$

Assuming a noncentral chi-square distribution with N DOFs and omitting terms that are equal for all transitions from time index $n-1$ to time index n, $\widetilde{f}(z_n|x_n,s')$ can be written as

$$\widetilde{f}(z_n|x_n,s') = -\frac{S[x_n,s']}{2\sigma^2} - \frac{N-2}{4}\ln\left(S[x_n,s']\right) + \ln\left(I_{\frac{N}{2}-1}\left(\frac{\sqrt{z_n \cdot S[x_n,s']}}{\sigma^2}\right)\right). \quad (14)$$

We have found that the Bessel function in (14) can be evaluated by conventional software tools without numerical problems if its argument $\chi = \sqrt{z_n \cdot S[x_n,s']}/\sigma^2$ is smaller than about 100. However, for large χ, the following consideration is feasible. The modified Bessel function of the first kind $I_\alpha(\chi)$ tends to infinity even for moderate arguments.[4] To avoid numerical problems, $I_\alpha(\chi)$ can be approximated for sufficiently large χ with [15]

$$I_\alpha(\chi) \approx \frac{\mathrm{e}^\chi}{\sqrt{2\pi\chi}}\left(1 - \frac{\mu-1}{8\chi} + \frac{(\mu-1)(\mu-9)}{2!(8\chi)^2} \pm \cdots\right) \quad (15)$$

where $\mu = 4\alpha^2$. For $\chi > 100$, it is sufficient to approximate $I_\alpha(\chi)$ by the first term $\mathrm{e}^\chi/\sqrt{2\pi\chi}$ [16]. Then, from (14), we get

$$\widetilde{f}(z_n|x_n,s') \approx -\frac{S[x_n,s']}{2\sigma^2} - \frac{N-2}{4}\ln\left(S[x_n,s']\right) + \left(\frac{\sqrt{z_n \cdot S[x_n,s']}}{\sigma^2}\right) - \ln\sqrt{2\pi\left(\frac{\sqrt{z_n \cdot S[x_n,s']}}{\sigma^2}\right)}. \quad (16)$$

B. Suboptimal Algorithms for APP Calculation

Since optical transmission systems operate at high bit rates, the implementation complexity of the algorithms is of great importance. Therefore, suboptimal decoding techniques with reduced complexities are investigated. In the following discussion, the soft-output Viterbi equalizer (SOVE) and the intrinsic L-value calculation are introduced. Both are suboptimal algorithms that provide approximations of the true APP.

1) Soft-Output Viterbi Equalizer (SOVE): The Log-BCJR algorithm can be simplified by eliminating the backward metric calculation and by using the Max-Log instead of the Jac-Log calculation. This reduced-complexity algorithm is referred to as an SOVE [17]. To ensure that a bit has passed all taps in the discrete-time channel model at a time instant n, the soft output for bit $n-M$ is computed. Thus, the SOVE has a decision delay of M. Since the backward metric is not calculated, the soft-output value can be computed as

$$L(x_{n-M}) = \max_{\{X|x_{n-M}\}_+}\left[\widetilde{\alpha}_{n-1}(s') + \widetilde{\gamma}_n(s',s)\right] - \max_{\{X|x_{n-M}\}_-}\left[\widetilde{\alpha}_{n-1}(s') + \widetilde{\gamma}_n(s',s)\right] \quad (17)$$

where $\{X|x_n - M\}_+$ and $\{X|x_n - M\}_-$ are the set of pairs (s',s) for the state transitions $(s_{n-1} = s') \rightarrow (s_n = s)$,

[4] For $\chi > 700$ and $\alpha > 0$, $I_\alpha(\chi) > 10^{300}$.

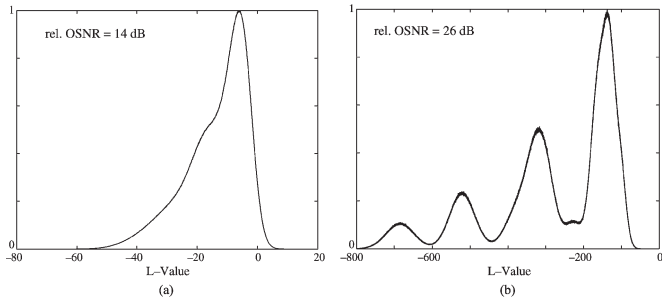

Fig. 6. Simulated shapes of *a posteriori* L-value distributions at the output of the BCJR detector (equalizer) with $\Delta\tau/T = 1.25$ and ordinate normalized to 1. Distributions for a transmitted logical 0 for an OSNR of (a) 14 and (b) 26 dB.

which result from the event $x_{n-M} = 1$ and $x_{n-M} = 0$, respectively.

2) Intrinsic L-Value Calculation (Optimal Threshold): To further reduce the computational complexity, a very simple computation method for the soft-output information is to omit the forward and backward metric calculations. The L-value computation for $M = 2$ is then reduced to

$$L(x_{n-1}) = \max_{\{X|x_{n-1}\}_+} [\widetilde{\gamma}_n(s', s)] - \max_{\{X|x_{n-1}\}_-} [\widetilde{\gamma}_n(s', s)]\,. \tag{18}$$

This method calculates the so-called intrinsic or channel L-value by observing the channel output. Sequence information is not taken into account, and the decision is made on each received value individually. The calculation is similar to the one described in [5]. The hard-decided bits are equal to the result obtained by the decision with an optimal threshold.

C. Shapes of L-Value Distributions for a Chi-Square-Distributed Noise

In Fig. 6, the simulated *a posteriori* L-value distributions at the output of the BCJR equalizer are shown for a normalized DGD of $\Delta\tau/T = 1.25$ and OSNRs of 14 and 26 dB. The histograms in Fig. 6 belong to a transmitted logical 0 ($f_D(\xi|X = 0)$). The depicted histograms show the development of several local maxima for the increasing OSNR. The number as well as the distance between the maxima increases for an increasing OSNR. The development of the maxima is more obvious for the L-value distributions of a logical 0 than for a logical 1 for the same OSNR (not shown here).

The reason why the L-value distribution depends on the OSNR is given by the correlation between the noise and the signal [16]. For the used optical channel model, there are eight different bit patterns that lead to six different output values y_n, with each output value having a different noise variance. Generally, by decreasing the noise variance of an output value y_n, the mean value of the resulting L-value distribution increases after the equalization process. The smaller the noise variance, the higher the reliability value. Therefore, the L-value distributions in Fig. 6 consist of several overlapping approximately Gaussian-shaped distributions with different means and variances. The higher the OSNR, the greater the difference of the means; thus, several maxima can be obtained. For small OSNRs, the distributions overlap and only one maximum can be seen. Since the BCJR algorithm also processes sequence information, the output L-value distribution does not only depend on the different signal values y_n, but also on the transmitted sequence $\mathbf{x}$.

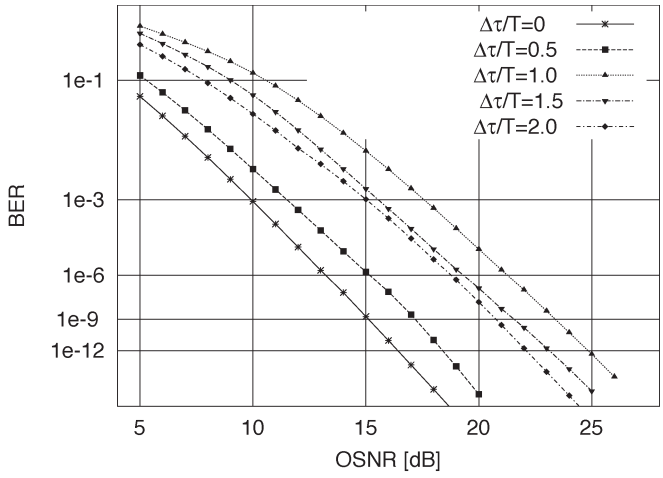

Fig. 7. BER versus OSNR for the BCJR equalizer in a logarithmic scale for various DGDs.

D. Performance Comparison of Equalization Algorithms

The performance of the above SISO equalization algorithms are compared with respect to the achievable bit error ratio (BER). For computer simulations, the optical channel is modeled by chi-square-distributed noise with $N = 12$ DOFs ($BT = 5$), an ER of $\zeta_{\mathrm{dB}} = 13$ dB, and a power splitting ratio between the PSPs of $\gamma = 0.5$.

In Fig. 7, the BER as a function of the OSNR for the BCJR equalizer is shown for various normalized DGDs. The BER curves are obtained using methods described in [16] and [18] that are based on Monte Carlo simulations and the statistics of the reliability information. In Fig. 8, the OSNR penalty to achieve a BER of 10^{-12} is shown for various SISO equalization algorithms. The following observations can be made.

1) The BCJR equalizer shows the best performance compared to all other investigated equalization schemes. The maximum penalty value of 8.9 dB occurs for a DGD of $\Delta\tau/T = 1.25$.
2) For $\Delta\tau/T < 0.75$, the performance difference between the BCJR equalizer and the SOVE is less than 0.25 dB. The maximum penalty of 12.1 dB occurs for $\Delta\tau/T = 1.25$ for the SOVE.
3) The optimal threshold decision (intrinsic L-value calculation) is able to recover the transmitted data as long as the eye is open. Since a completely closed eye occurs for $\Delta\tau/T = 1$, the penalty curve has a pole.

The results for the BCJR equalizer and the SOVE show that the OSNR penalties are almost the same, particularly for $\Delta\tau/T < 0.75$. Since the SOVE omits the backward metric computation, the L-value calculation can be performed with

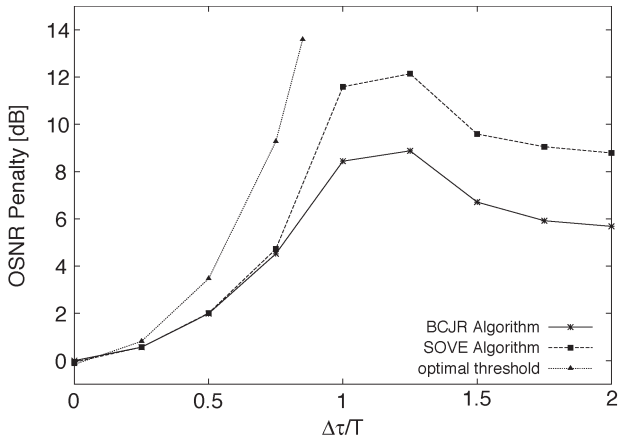

Fig. 8. OSNR penalty versus the normalized DGD for various detection algorithms.

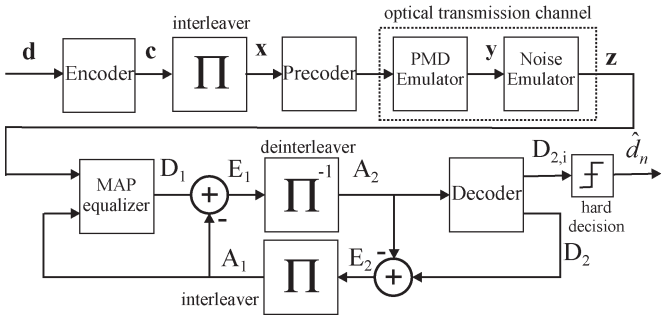

Fig. 9. Transmitter and receiver structure for the iterative equalization.

less operations and the decision delay is only M. Therefore, it is advantageous to use the SOVE instead of the BCJR algorithm. The optimal threshold decision is not an adequate alternative due to its pole for $\Delta\tau/T = 1$.

IV. Turbo Equalization of PMD

In this section, the turbo equalization of a coded data transmission over an optical channel with a PMD is considered. The block diagram of the optical transmission system with a turbo equalizer is shown in Fig. 9. The values of $\mathbf{d}$ are protected by an FEC using an encoder with a code rate R. The receiver consists of two parts: 1) the sbs-MAP equalizer as an inner code component (to correct the ISI distortion) and 2) the soft-output decoder as an outer code component. Both parts are required to achieve an iterative operation. In general, for the FEC, any coding scheme that is able to generate soft-output values and process *a priori* information can be used. In this paper, a recursive systematic convolutional (RSC) code is used in the encoder.

A. Principles and System Definitions

An overview of the turbo equalization can be found in [14] and [19]. The idea of turbo equalization is to communicate soft information (L-values) between the sbs-MAP equalizer and the sbs-MAP decoder. The combination of the optical transmission channel and the RSC encoder corresponds to a

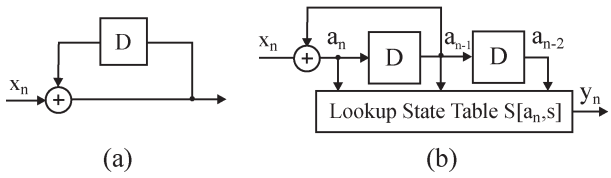

Fig. 10. (a) Rate 1 recursive precoder with memory 1 and $(G_r, G) = (03, 02)$ and (b) equivalent model of the cascade of the precoder and the optical channel.

serial concatenation of two codes. The optical channel plays the role of an inner encoder and the RSC encoder of an outer encoder. The block diagram for the principle of the transmitter and receiver structure for an iterative equalization is shown in Fig. 9. On the transmitter side, a block of data $\mathbf{d}$ is encoded by the outer encoder with code rate R yielding the coded sequence $\mathbf{c}$. The encoded bit stream $\mathbf{c}$ is obtained by inserting a coded bit after each Kth information bit of $\mathbf{d}$. As a result, the code word $\mathbf{c}$ with a code rate $R = K/(K+1)$ is obtained. An acceptable coding overhead for optical transmission systems lies between 5%–15%, which corresponds to code rates between $R = 0.87$ and 0.95. A random interleaver is placed after the encoder to guarantee an almost statistical independence of the adjacent bits in the sequence $\mathbf{x}$. To improve the system performance, a rate-1 recursive encoder (referred to as precoder) is used. The cascade of the precoder and the nonrecursive optical channel becomes recursive, and an interleaving gain can be achieved [20]. In the following discussion, the precoder in Fig. 10(a) with a memory length $\nu = 1$ is considered. The cascade of the rate-1 recursive precoder and the nonrecursive optical channel can be represented by an equivalent state-based channel model, which is shown in Fig. 10(b). Therefore, the concatenation of the precoder and optical channel can be described with a single state-transition diagram. Note that the complexity of the sbs-MAP equalizer is not increased as long as the memory ν of the precoder is smaller than the memory M of the optical channel model.

At the receiver side in Fig. 9, the sbs-MAP equalizer takes the channel observations $\mathbf{z}$ and calculates the *a posteriori* L-values D_1. From D_1, the *a priori* L-values A_1 are subtracted, yielding the extrinsic L-values $E_1 = D_1 - A_1$. E_1 are then sent through a deinterleaver to become the *a priori* knowledge A_2 for the (outer) channel decoder. The outer decoder calculates the *a posteriori* L-values D_2, which are feed back as extrinsic information $E_2 = D_2 - A_2$ to become the *a priori* knowledge A_1 for the sbs-MAP equalizer after interleaving. After a certain number of iterations, a hard decision on the *a posteriori* L-values $D_{2,i}$ of the information bits is performed to obtain the estimates $\hat{d}_n$ of the transmitted bits. In the following discussion, unless mentioned otherwise, an information bit rate of 10 Gb/s is assumed. The redundancy, which is added by the encoder with a code rate R, increases the bit rate by a factor of $1/R$.

B. EXIT-Chart Analysis

A powerful tool to analyze the design options for turbo codes is the EXIT chart. It was found by ten Brink in [12]

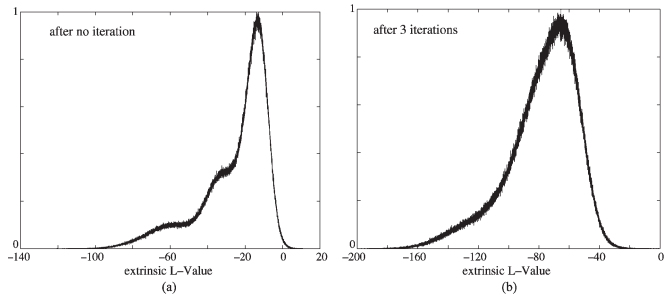

Fig. 11. Simulated shapes of extrinsic L-value distributions $f_{E1}(\xi|X=0)$ at the output of the inner decoder after various iterations, ordinate normalized to 1.

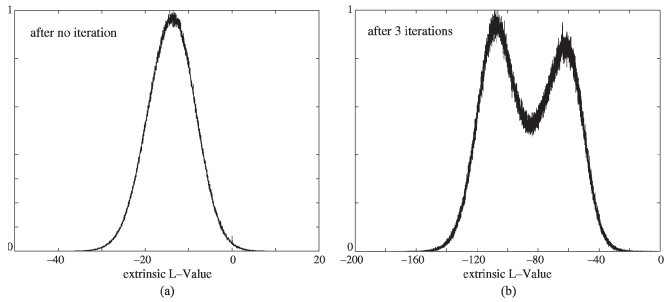

Fig. 12. Simulated shapes of extrinsic L-value distributions $f_{E2}(\xi|C=0)$ at the output of the outer decoder after various iterations, ordinate normalized to 1.

and [13]. For the EXIT-chart analysis, mutual information of the *a priori* and the extrinsic L-value distributions are used to define transfer characteristics for the inner and the outer decoders. A crucial observation of ten Brink in [12] is that the extrinsic L-value distributions, which are passed between the inner and the outer decoders, can be described approximately by a single-parameter Gaussian distribution with a variance σ_A^2

$$f(\xi|X=x)=\frac{1}{\sqrt{2\pi\sigma_A^2}}\cdot\exp\left(-\frac{\left(\xi-\frac{\sigma_A^2(2x-1)}{2}\right)^2}{2\sigma_A^2}\right) \quad (19)$$

where $x\in\{0,1\}$. Therefore, the inner and the outer decoders for serially concatenated codes can be analyzed independently from each other by modeling the *a priori* information for the decoders with (19).

In Figs. 11 and 12, the simulated extrinsic L-value distributions E_1 and E_2, which are passed between the inner and the outer decoders are presented. For the analysis, a precoded optical channel and a memory-3 outer convolutional code are used. The L-value distributions are obtained for a DGD of $\Delta\tau=108$ ps, a code rate of $R=9/10$, and an OSNR of 14.5 dB. However, the histogram measurements show that the Gaussian assumption for the extrinsic information distributions is not fulfilled for the optical PMD channel. The number of performed iterations significantly changes the shape of the PDF. The corresponding EXIT chart is shown in Fig. 13. The input *a priori* mutual information I_{A1} of the inner decoder, which is equal to the extrinsic output mutual information I_{E2}

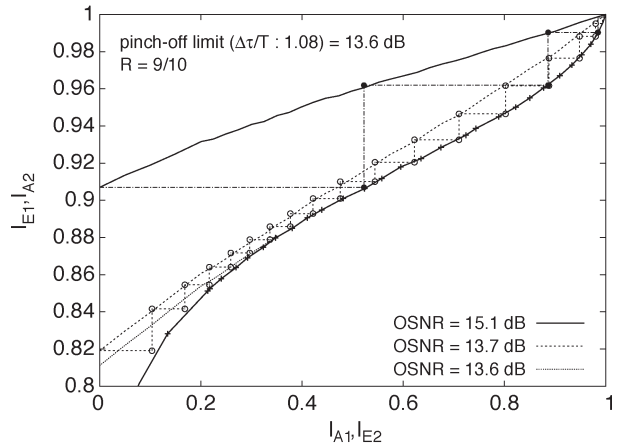

Fig. 13. Receiver EXIT chart for OSNR $=13.6$ dB, 13.7 dB, and 15.1 dB; code rate $R=9/10$; and $\Delta\tau/T=1.08$.

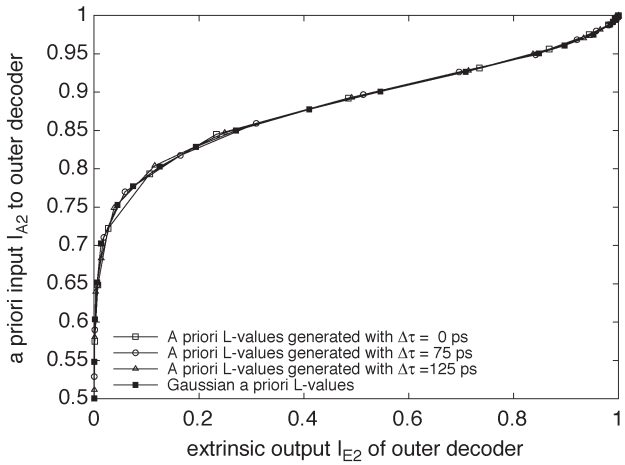

Fig. 14. Mutual information transfer characteristics for the outer decoder, computed using different shapes for the *a priori* L-value distributions.

of the outer decoder, is shown on the abscissa. The input *a priori* information I_{A2} of the outer decoder, which is equal to the extrinsic output information I_{E1} of the inner decoder, is given on the ordinate. The step curve, which is the decoding trajectory in Fig. 13, is not limited to the Gaussian assumption. It is computed by taking the real L-value distributions into account. To obtain the points $(I_{E1,k}=I_{A2,k}, I_{E2,k}=I_{A1,k+1})$ for each iteration k, the extrinsic output PDF are simulated at the respective component decoder and the corresponding mutual information is calculated. The decoder transfer characteristics match very well with the iterative decoding trajectory in Fig. 13.

The transfer characteristic of the outer decoder is computed without channel information. To investigate if the transfer characteristics change for different input *a priori* L-value distributions, the *a priori* L-value distributions are generated with the inner decoder by varying the OSNRs for various DGDs. Fig. 6 visualizes the generated *a priori* L-value distributions for $\Delta\tau=125$ ps. The obtained transfer characteristics of the outer code are shown in Fig. 14. The curves show, almost independently from the DGD, a good agreement with the transfer characteristic obtained by modeling the *a priori*

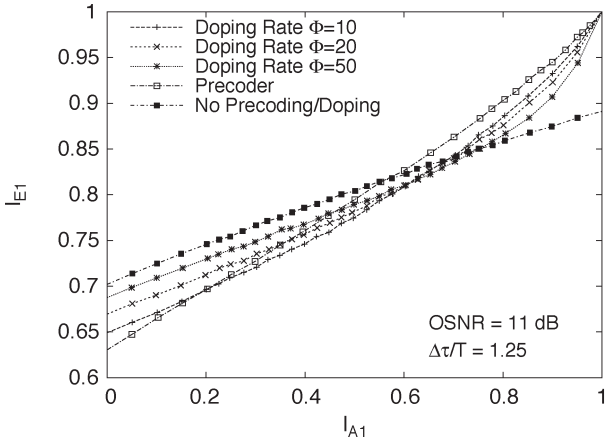

Fig. 15. Transfer characteristics for the optical channel with a normalized DGD of $\Delta\tau/T = 1.25$.

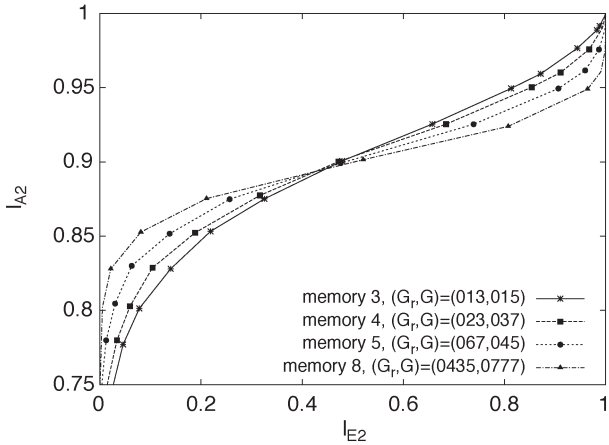

Fig. 16. Transfer characteristics of outer decoders with code rate $R = 9/10$ for convolutional codes with various code memories.

information with a Gaussian distribution. Although the real L-value distributions are non-Gaussian over all iterations (Figs. 11 and 12), the mutual information measure turns out to be robust against changes in the actual extrinsic distributions, which is in line with the observations in [12]. In [12], a generalized Gaussian distribution with different decay parameters α is used to generate the L-values and hardly any change could be obtained in the transfer characteristics. From the simulations in Fig. 14, the robustness is obviously maintained even for a chi-square-distributed channel noise.

C. Code Design Alternatives

The EXIT chart will now be used to optimize the outer and the inner transfer characteristics for the optical channel.

1) Inner Code Design: In Fig. 15, the chart for the inner transfer characteristic of an optical channel with a normalized DGD of $\Delta\tau/T = 1.25$ is shown. An almost error-free transmission is achieved if point (1, 1) in the EXIT chart is reached [12]. As can be seen from Fig. 15, this is not possible for the nonprecoded channel. Therefore, an error floor will occur after only a few iterations (Fig. 18). The advantage of precoding is that the inner transfer characteristics are bend up and can thus reach the point (1, 1). The disadvantage is that the starting point $I_{E1}(I_{A1} = 0)$ of the inner transfer characteristic decreases. A compromise can be found by code doping [21]. Instead of replacing all information bits x_n, only a fraction $1/\Phi$ is replaced by precoded bits. This is referred to as code doping with a doping rate Φ. For $n \neq m\Phi$, the encoder sends an information bit; for $n = m\Phi$, an information bit is replaced by an output bit of the precoder $(m = 0, \pm 1, \ldots; \Phi \in I\!N)$. The disadvantage of code doping is the additional sbs-MAP decoder that is required at the receiver side. By increasing the doping rate Φ, the starting points $I_{E1}(I_{A1} = 0)$ of the transfer curves go up, whereas the slope of the transfer curves decreases. The higher the doping rate, the later the transfer curve is bend up, i.e., the number of required iterations to reach point (1, 1) increases.

2) Outer Code Design: For the outer encoder, the RSCs are investigated. The transfer characteristic of the outer encoder is independent of the OSNR and the DGD, but it depends on the code rate R. Increasing R will shift the outer transfer characteristic up. The extrinsic transfer characteristics for various code memories ν and code rate $R = 9/10$ are shown in Fig. 16. Note that the *a priori* input information I_{A2} to the decoder is shown on the ordinate and the extrinsic output information I_{E2} on the abscissa. The code memory ν is the most important design parameter for the outer decoder. It has been found in [12] and [13] that different generator polynomials only have a small impact on the shape of the outer transfer characteristic. The advantage of an increased code memory is that less iterations are required to reach point (1, 1) in the EXIT chart.

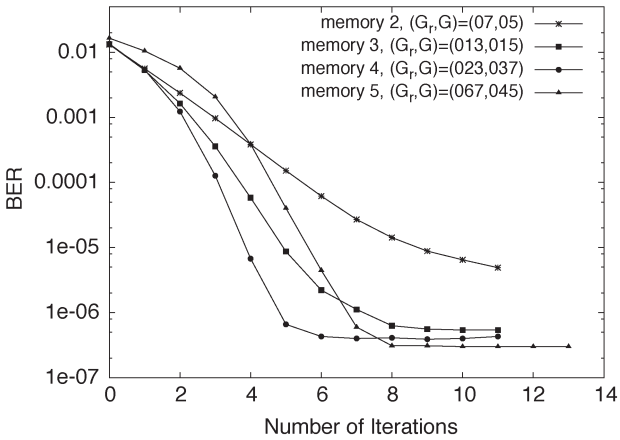

Fig. 17. BER versus the number of iterations for the turbo equalizer using outer codes with various memory lengths and $R = 19/20$.

D. Error Floor

In Fig. 17, the improvements in the BER versus the number of iterations are shown for a code rate of $R = 19/20$ and various memory lengths of the RSC encoder. A serious problem for iterative-equalization schemes in optical transmission are the potential error floors, which are dearly visible in Fig. 17. Traditional coding schemes, like Reed–Solomon (RS) codes, perform worse than turbo codes at low OSNR values but could have a better performance for high OSNR values, e.g., the

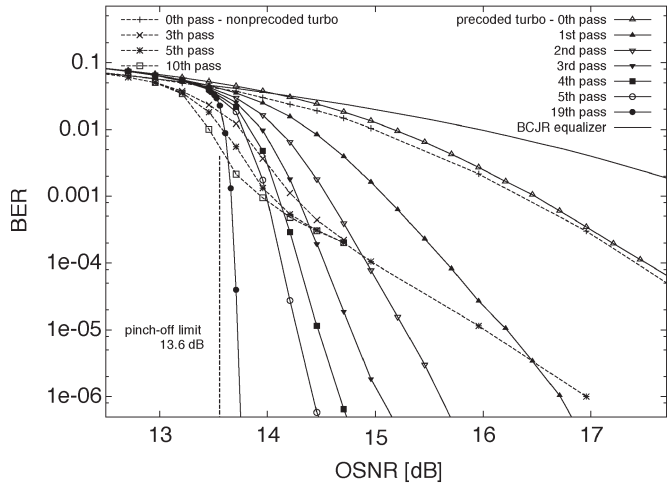

Fig. 18. BER for an optical channel with $\Delta\tau = 108$ ps, $R = 9/10$, and the memory-3 convolutional encoder.

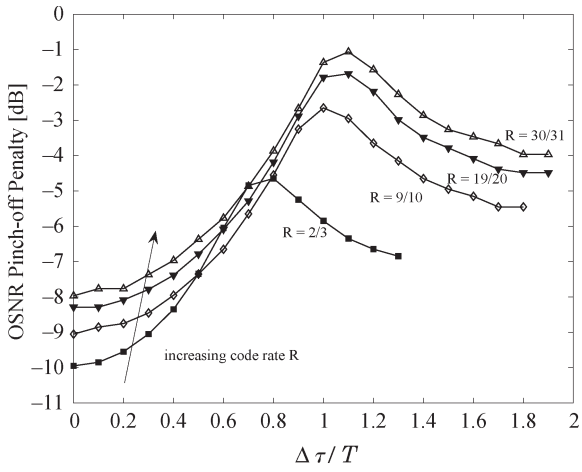

Fig. 19. OSNR pinch-off penalty versus the DGD for the turbo equalizer with code rates $R = 2/3, 9/10, 19/20,$ and $30/31$ and an information bit rate of 10 Gb/s (the ordinate is related to Fig. 8).

(255, 239) RS code with a 6.7% overhead is standardized for an FEC in optical transmission systems. This code corrects up to $t = 8$ byte errors anywhere in the code word. A common approach to eliminate error floors and to achieve an almost error-free transmission is to combine the turbo equalization with an outer t-error correcting RS code. In this application, it is important to ensure that the output of the turbo equalizer contains no more than t byte errors, which may cause the RS code to fail.

E. Penalty Curves for the Coded Transmission

As explained in Section IV-C, there are many DOFs to design the inner and the outer codes. In the following discussion, a precoded optical channel and an outer memory-3 convolutional encoder are considered if not mentioned otherwise. The codes are chosen because of their good performance and reasonable low complexity. The inner and outer decoders operate on a trellis with four and eight states, respectively. In Fig. 18, the simulated BER for an optical channel with precoding $\Delta\tau = 108$ ps and outer code rate $R = 9/10$ is shown for various numbers of iterations. An interleaver with memory of 10^5 bits is used, and a total of 10^8 bits are transmitted. Obviously, BER can be significantly reduced by performing several iterations. For comparison, the BER curves for the nonprecoded optical channel [22] are shown. Obviously, without precoding, the turbo equalizer has a lower performance and shows an early error floor.

To determine the maximum coding gain, which is achievable with turbo equalization, the EXIT chart can be used. For a given DGD and a code rate R, the required OSNR can be found for which the decoding trajectory just manages to sneak through a narrow tunnel (turbo cliff region). It is referred to as pinch-off limit and is shown in Figs. 13 and 18. The OSNR pinch-off penalty curves for optical channels with an FEC and code rates $R = 2/3, 9/10, 19/20,$ and $30/31$ are shown in Fig. 19. Obviously, the principle shape of the penalty curve for the coded system is the same as that for the uncoded system (Fig. 8). This implies that, for the electrical equalization, an increased OSNR for $\Delta\tau \neq 0$ is always required. Channel coding shifts the penalty curve down; since the effects on the output signal of a PMD depend on the bit rates, the maximum penalty occurs for a coded transmission at smaller DGD values. Fig. 19 shows that even for code rates close to 1 (e.g., $R = 30/31$), the iterative equalization can achieve a coding gain of approximately 8 dB over the uncoded equalization.

V. Conclusion

In this paper, several SISO algorithms for electrical equalization of the PMD in optical channels are presented. Sbs-MAP algorithms are designed to incorporate signal-dependent chi-square-distributed noise. A discrete-time state-based channel model is derived to describe the effects of the first-order PMD on the detected output signal. The performances of the BCJR sbs-MAP equalizer and the suboptimal SOVE as well as an optimal threshold decision are compared. The OSNR penalty curves in Fig. 8 show that, for normalized DGD values $\Delta\tau/T \leq 0.75$, the performance of the optimal BCJR and suboptimal complexity-reduced SOVE algorithm is almost the same. The optimal threshold decision is worse; however, it has the lowest complexity. The BCJR SISO algorithm is considered for the turbo equalization of a PMD as well. It is shown that an additional digital channel precoder is necessary to improve the system performance compared to a nonprecoded scheme; see Fig. 18. It was found that the L-value distributions for signal-dependent noises can exhibit several maxima, depending on the OSNR and the DGD. Although the extrinsic L-value distributions are generally non-Gaussian, it turns out that the EXIT chart can still be used to study the convergence of iterative-equalization schemes. This is due to the robustness of the mutual information against changes in the extrinsic L-value distributions. Therefore, code design techniques, which are known from AWGN ISI channels, can also be used for the optical ISI channel. The iterative (turbo) equalization is successfully applied for optical channels. It can provide an OSNR improvement of about 8–10 dB for code

rates R of 30/31–2/3 compared to uncoded transmission. The iterative equalization and decoding can reduce the BER by several orders of magnitude. However, error floors can remain as serious problems. They can be overcome by an additional outer RS code with a small overhead.

REFERENCES

[1] S. Lanne and E. Corbel, "Practical considerations for optical polarization mode dispersion compensators," *J. Lightw. Technol.*, vol. 22, no. 4, pp. 1033–1040, Apr. 2004.

[2] H. Haunstein, W. Sauer-Greff, A. Dittrich, K. Sticht, and R. Urbansky, "Principles for electronic equalization of polarization-mode dispersion," *J. Lightw. Technol.*, vol. 22, no. 4, pp. 1169–1182, Apr. 2004.

[3] F. Buchali and H. Bülow, "Adaptive PMD compensation by electrical and optical techniques," *J. Lightw. Technol.*, vol. 22, no. 4, pp. 1116–1126, Apr. 2004.

[4] A. Weiss, "On the performance of electrical equalization in optical fiber transmission systems," *IEEE Photon. Technol. Lett.*, vol. 15, no. 9, pp. 1225–1227, Sep. 2003.

[5] G. Bosco, G. Montorsi, and S. Benedetto, "Soft decoding in optical systems," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1258–1265, Aug. 2003.

[6] Y. Cai, J. Morris, T. Adali, and C. Menyuk, "On turbo code decoder performance in optical-fiber communication systems with dominating ASE noise," *IEEE Photon. Technol. Lett.*, vol. 15, no. 3, pp. 727–734, Mar. 2003.

[7] G. Bosco, A. Carena, V. Curri, R. Gaudino, P. Poggiolini, and S. Benedetto, "A novel analytical approach to the evaluation of the impact of fiber parametric gain on the bit error rate," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2154–2163, Dec. 2001.

[8] D. Marcuse, "Derivation of analytical expressions for the bit-error probability in lightwave systems with optical amplifiers," *J. Lightw. Technol.*, vol. 8, no. 12, pp. 1816–1823, Dec. 1990.

[9] A. Willner, "Chromatic dispersion and polarization mode dispersion—Managing key limitations in optical communication systems," *OPN Trends*, vol. 13, no. 3, pp. 16–20, Mar. 2002.

[10] G. Einarsson, *Principles of Lightwave Communications*. Chichester, U.K.: Wiley, 1996.

[11] L. 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, no. 2, pp. 284–287, Mar. 1974.

[12] S. ten Brink, "Design of concatenated coding schemes based on iterative decoding convergence," Dr.-Ing. dissertation (Ph.D.), Inst. Telecommun., Univ. Stuttgart, Stuttgart, Germany, 2001.

[13] ——, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[14] J. Hagenauer, "The turbo principle: Tutorial introduction and state of the art," in *Proc. 1st Int. Symp. Turbo Codes*, Brest, France, Sep. 3–5, 1997, vol. 2, pp. 1–12.

[15] E. Desurvire, D. Bayart, B. Desthieux, and S. Bigo, *Erbium Doped Fiber Amplifiers—Device and System Development*, 3rd ed. New York: Wiley, 2002.

[16] M. Jäger, "Maximum a posteriori soft-in/soft-out decoder for chi-square distributed noise," M.S. thesis, Inst. Telecommun., Univ. Stuttgart, Stuttgart, Germany, Aug. 2004.

[17] G. Bauch and V. Franz, "A comparison of soft-in/soft-out algorithms for 'turbo decoding'," in *Proc. 1st Int. Symp. Turbo Codes*, Brest, France, Sep. 1997, pp. 1–12.

[18] N. Letzepis and A. Grant, "Bit error rate estimation for turbo decoding," in *Proc. IEEE Int. Symp. Inf. Theory*, Yokohama, Japan, Jul. 2003, p. 437.

[19] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes. 1," in *Proc. IEEE Int. Conf. Communications*, Geneva, Switzerland, May 1993, pp. 1064–1070.

[20] K. Narayanan, "Effect of precoding on the convergence of turbo equalization for partial response channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 4, pp. 686–698, Apr. 2001.

[21] A. Boronka, N. Muhammad, and J. Speidel, "Removing error floor for bit interleaved coded modulation MIMO transmission with iterative detection," in *Proc. IEEE Int. Conf. Communications (ICC)*, Seoul, Korea, May 2005, pp. 2392–2396.

[22] H. Haunstein and R. Urbansky, "Application of electronic equalization and error correction in lightwave systems," presented at the Proc. Eur. Conf. Optical Communications (ECOC), Stockholm, Sweden, 2004, Paper Th1.5.1.

Mark Jäger was born in Stuttgart, Germany, in 1978. He received the M.Sc. degree in applied optics from the Rose-Hulman Institute of Technology, Terre Haute, IN, in 2003 and the Dipl.Ing. degree in electrical engineering and information technology from the University of Stuttgart, Germany, in 2004. He is currently working toward the Ph.D. degree at the Interdisciplinary Center for Scientific Computing, University of Heidelberg, Germany.

He is developing multidimensional signal processing algorithms, in cooperation with the Robert Bosch GmbH, Stuttgart. His current research interests include optical communication systems, target tracking, and classification algorithms.

Tobias Rankl was born in Stuttgart, Germany, in 1977. He received the Dipl.Ing. degree in electrical engineering and information technology in 2003 from the University of Stuttgart, where he is currently working toward the Ph.D. degree.

He is currently a Teaching and Research Assistant at the Institute of Telecommunications, University of Stuttgart. His research interests include channel coding, channel equalization, and optical communications.

Joachim Speidel (M'94) received the Dipl.Ing. and Dr.Ing. degrees (both summa cum laude) in electrical engineering and information technology from the University of Stuttgart, Germany, in 1975 and 1980, respectively.

From 1980 to 1992, he was with Philips Communications (now Lucent Technologies Bell Labs Innovations), where he worked in the broad field of digital communications. He has held various positions in research and development as a member of a Technical Staff, Laboratory Head, and finally as Vice President. Since autumn 1992, he has been a Full Professor and the Director of the Institute of Telecommunications at the University of Stuttgart. His research interests include digital communications in mobile, optical, and electrical networks, with an emphasis on transmission, modulation, source, and channel coding.

Henning Bülow received the Dipl.Ing. degree in electrical engineering from the University of Dortmund, Dortmund, Germany, in 1985 and the Ph.D. degree in electrical engineering from the University of Berlin, Berlin, Germany, in 1988 for his work on integrated optic switching matrices.

In 1990, he joined the Research Center of Alcatel, Stuttgart, Germany, where he worked on erbium-doped fiber amplifiers in optical analog TV systems, on optical and electrical time-multiplexed 40-Gb/s transmission systems, and on the assessment of data transmissions in the presence of a polarization-mode dispersion (PMD). Since 1998, he has been the Head of a research team investigating the dynamic mitigation of transmission impairments at 10, 40, and 160 Gb/s by adaptive electronic and optical signal processing.

Fred Buchali received the Diploma degree in electrical engineering and the Ph.D. degree from Humboldt University, Berlin, Germany, in 1988 and 1991, respectively. His Ph.D. research involved InGaAs-InP p-i-n photodiodes for fiber-optic communications.

In 1991, he joined Duisburg University, Duisburg, Germany, where he worked on development of InAlAs–InGaAs HFET and metal-semiconductor-metal (MSM) detectors. In 1992, he joined the Research and Innovation Department of Alcatel, Stuttgart, Germany, where he was involved with the process evaluation, package design, and characterization of optoelectronic and electronic devices. Since 1999, he has been working with the Optical Transmission Systems Group. His current research interests include the mitigation of transmission impairments in 10-, 40-, and 160-Gb/s systems.