# Maximum Likelihood Sequence Estimation in the Presence of Timing Misalignment and Polarization Mode Dispersion in Optically Amplified Return-to-Zero Systems

Jian Zhao, Lian-Kuan Chen, Chun-Kit Chan
Department of Information Engineering
The Chinese University of Hong Kong
Shatin, N. T., Hong Kong
Email: jzhao2@ie.cuhk.edu.hk

***Abstract*—We investigate the performance of maximum likelihood sequence estimation (MLSE) receiver in the presence of the impairments from both the pulse carver-data modulator timing misalignment (TM) and polarization mode dispersion (PMD) in optically amplified return-to-zero (RZ) systems. RZ modulation format is commonly used in long-haul wavelength-division multiplexing transmission systems and the dominating noise source in those systems is amplified spontaneous emission (ASE) noise, which is signal dependent and requires special study when direct detection is employed. In this paper, based on the bit-to-bit error probability estimation using Karhunen-Loeve (KL) expansion and decorrelation of noise components, we use the steepest descent method to obtain sequence-to-sequence error probability and achieve the performance evaluation of MLSE receiver with arbitrary input signal pulse shape, optical filtering and electrical filtering taken into consideration. Monte Carlo simulations of a 10-Gb/s RZ system are demonstrated and agree with the theory well. The results show that the power penalty induced by TM and PMD can be effectively reduced by MLSE receiver, which thus validates its capability to enhance tolerance to both TM and PMD with shared electrical devices.**

***Keywords- Equalizers, optical fiber communication, modulation***

## I. INTRODUCTION

Return-to-zero (RZ) modulation format is extensively employed in long-haul wavelength-division multiplexing transmission systems. Compared to non-return-to-zero (NRZ) modulation format, it shows several-decibel improvement in receiver sensitivity and promises better tolerance against polarization mode dispersion (PMD) [1], [2]. The generation of such format can be implemented by using dual Mach-Zehnder modulator (MZM) configuration [3]. For proper operation of the scheme, it is essential to locate the pulse peak in the middle of the data bit slot. However, the relative time delay of the optical and electrical devices drifts over time due to temperature variation and device aging, leading to timing misalignment (TM) between the pulse carver and the data modulator. Such misalignment was experimentally demonstrated to significantly degrade the system performance [4]. To resolve the problem, several timing alignment techniques were proposed [3], [4], in which an additional monitoring stage was used for alignment controlling. Alternatively, we showed that, characterized as intersymbol interference (ISI), maximum likelihood sequence estimation (MLSE) also had the capability to combat such impairment [5].

PMD is one of the most important obstacles for high-capacity long-haul optical communication systems. The typical manifestation of PMD is that the signal is split into two orthogonal polarization modes which propagate in the fiber at different velocities, therefore causes ISI. A lot of effort for PMD compensation has been performed [6], [7], among which electronic techniques such as feed-forward equalizer(FFE), decision-feedback equalizer (DFE) and MLSE have attracted much attention for their flexibility, adaptation, and cost-effective. Up to now, the implementations of 40-Gb/s FFE and DFE, and 10-Gb/s MLSE receiver have been reported.

As a general post-detection solution to ISI, electronic equalization is not specific to a certain kind of ISI and can simultaneously combat impairments from different optical and electrical distortions [8]. Hence, it reduces the number of the required compensation components. In this paper, we will investigate the performance of MLSE receiver in the presence of both TM and PMD in optically amplified RZ systems where amplified spontaneous emission (ASE) noise dominates. Based on bit-to-bit error probability estimation using Karhunen-Loeve (KL) expansion, decorrelation of noise components, and saddlepoint approximation, we employ the steepest descent method to obtain sequence-to-sequence error probability and achieve bit error rate (BER) evaluation of MLSE receiver taking arbitrary input signal pulse shape, optical filtering and electrical filtering into consideration.

This paper is organized as follows. In Section II, we describe the system model and the operation principle of MLSE. In Section III, the impairment from TM and PMD is investigated. Bit-to-bit error probability is obtained in Section IV. In Section V, the performance of MLSE receiver in the presence of TM and PMD is evaluated. Simulations are demonstrated in Section VI and agree with the theory well. Finally, Section VII summarizes the results.



*This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.*

## II. System Model and Operation Principle of MLSE

Fig. 1 depicts the system model. The modulated signal $E_s(t)$ is obtained by using dual MZM configuration where continuous-wave light is first carved by driving an MZM with a sinusoidal voltage at half of the bit rate and then modulated by NRZ data in the second MZM [3]. $\varphi(t)$ and $\phi(t)$ are the phase changes in the pulse carver and the data modulator, respectively. The misaligned time $t_{TM}$ is emulated by an optical delay line. The input data $V_{NRZ}(t)$ is raised cosine shaped with $\alpha$ controlling the edge sharpness [5]. Fiber transmission link is modeled as a single-input, two-output setup [6]. $E_s(t)$ is split into two orthogonal polarization modes with $\gamma$ being the relative power in the fast principle state of polarization. $h_x(\gamma^{1/2}E_s(t))$ and $h_y((1-\gamma)^{1/2}E_s(t))$ denote the channel mapping of the two polarization modes and include the sources for signal degradation, e. g. PMD. Optical noises from optical amplifiers, $n_{xopt}(t)$ and $n_{yopt}(t)$, are modeled as independent additive white Gaussian noises (AWGN) with zero mean and the power spectral density of $N_0/2$ for each polarization's in-phase and quadrature components [9]. An optical bandpass filter (OBPF) with the impulse response of $h_o(t)$ is then employed to suppress the optical noise and yield the outputs of the transmission fiber, $E_{xout}(t)$ and $E_{yout}(t)$:

$$\begin{aligned} E_{xout}(t) &= (h_x(\gamma^{1/2}E_s(t)) + n_{xopt}(t)) \otimes h_o(t) = E_{s,x}(t) + n_x(t) \\ E_{yout}(t) &= (h_y((1-\gamma)^{1/2}E_s(t)) + n_{yopt}(t)) \otimes h_o(t) = E_{s,y}(t) + n_y(t) \end{aligned} \quad (1)$$

where $\otimes$ stands for the convolution operation. At the receiver, $E_{xout}(t)$ and $E_{yout}(t)$ are square-law detected and summed up to obtain the photo-current, $I_0(t)$, of the photo-detector:

$$I_0(t) = R(|E_{xout}(t)|^2 + |E_{yout}(t)|^2) \quad (2)$$

where $R$ is responsivity of the photo-detector. Finally, $I_0(t)$ is filtered by an electrical filter (EF) with the impulse response of $h_e(t)$ before it is sampled. Assume that the sampling time for the $n^{th}$ bit is $t_n$, the sampled discrete-time sequence can be written as $(I(t_0)\ I(t_1)\ldots\ I(t_{n-1})\ I(t_n)\ldots)$ with:

$$I(t_n) = I(t)\big|_{t=t_n} = (I_0(t) \otimes h_e(t))\big|_{t=t_n} = I_{ave}(t_n) + n(t_n) \quad (3)$$

where $I_{ave}(t_n)$ is the mean value of $I(t_n)$. The sampled signal is analog-to-digital converted and is decoded by MLSE receiver. The analog-to-digital converter (ADC) would introduce quantization noise, which, however, is negligible for the ADC with more than 4-bit resolution [6]. The operation of MLSE receiver realizes the optimal estimation of the input data sequence $(a_0\ a_1\ldots a_{n-1}\ a_n)$, which requires the finding of a sequence $(b_0\ b_1\ldots\ b_{n-1}\ b_n)$ that minimizes the metric of:

$$\begin{aligned} PM(I(t_n)) &= -\log(p(I(t_0), I(t_1), ..., I(t_{n-1}), I(t_n) \mid b_0, b_1, ..., b_{n-1}, b_n)) \\ &\approx PM(I(t_{n-1})) - \log(p(I(t_n) \mid b_0, b_1, ..., b_{n-1}, b_n)) \end{aligned} \quad (4)$$

where '≈' is used instead of '=' because the assumption that $I(t_p)$, $0 \le p \le n$, are uncorrelated may not be satisfied in optical systems, where EF is only a noise limiting low-pass filter instead of optimal whitened matched filter in order to reduce the complexity of MLSE receiver front end [10]. However, for RZ format, (4) is a near-optimal approximation because the bandwidth of EF is typically larger than the bit rate, leading to weak correlation among the values of $I(t_p)$, $0 \le p \le n$.

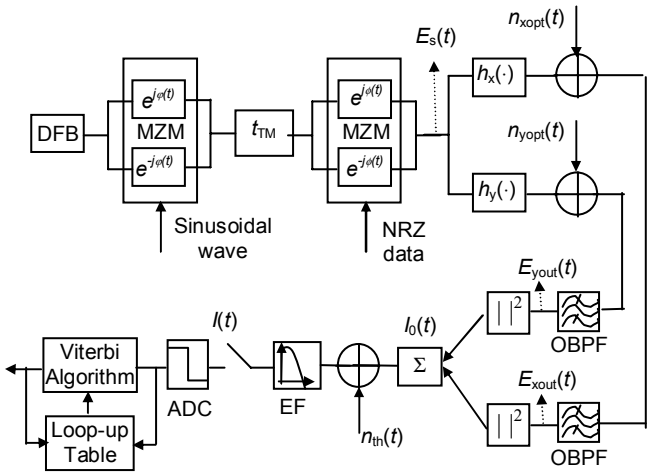


Figure 1. Block diagram of the system.

In practical systems, it is reasonable to assume that ISI affects a finite number of symbols, $m$. Therefore, $-\log(p(I(t_n)|\ b_0,\ b_1,\ldots,\ b_{n-1},\ b_n)) = -\log(p(I(t_n)|\ b_{n-m},\ b_{n-m+1},\ldots,\ b_{n-1},\ b_n))$. MLSE receiver can be modeled as a $2^m$-state machine with state $S(b_n) = [b_{n-m}\ b_{n-m+1}\ldots\ b_{n-1}]$. The calculation of metric (4) is performed by employing the Viterbi algorithm, with the initial metric for different states in the look-up table obtained by using non-parametric histogram method.

## III. Impairments from TM and PMD

From [5], it was shown that the impairment from TM could be characterized as ISI with $m$=1. Furthermore, for most practical systems, the bandwidth of the OBPF is typically larger than the spectral bandwidth of the optical signal. In such case, TM-induced ISI is linear with $I(t_n;\ a_{n-1},\ a_n)$ for TM written as $I(t_n;\ a_{n-1},\ a_n) = f_{TM}(g_{-1}a_{n-1} + g_0 a_n)$, where $g_{-1}+g_0=1$. On the other hand, PMD-induced ISI is also linear [6]. $I(t_n;\ a_{n-1},\ a_n)$ for PMD in RZ systems has the form of:

$$\begin{aligned} I(t_n; a_{n-1}, a_n) &= f^s_{a_{n-1}} a_{n-1} + (f^f_{a_n} + f^s_{a_n}) a_n \\ &= f_{PMD}(r_{-1}a_{n-1} + r_0 a_n), \qquad r_{-1} + r_0 = 1 \end{aligned} \quad (5)$$

where $f^s_{a_{n-1}}$, $f^f_{a_n}$, and $f^s_{a_n}$ are the electric fields sampled from the slow mode (SM) of $a_{n-1}$, the fast mode (FM) of $a_n$, and the SM of $a_n$ given $a_{n-1}=a_n=1$, respectively. In RZ systems, when the differential group delay (DGD) is small and the SM of $a_{n-1}$ does not interfere with the FM of $a_n$, $f^s_{a_{n-1}}=0$; On the other hand, when the DGD is large and the SM of $a_n$ separates from the FM of $a_n$, $f^s_{a_n}=0$. The combined impairments from both PMD and TM are linear ISI with $I(t_n;\ a_{n-2,}\ a_{n-1},\ a_n)$ being:

$$\begin{aligned} I(t_n; a_{n-2}, a_{n-1}, a_n) &= f^s_{a_{n-1}}(g_{a_{n-1},-1}{}^s a_{n-2} + g_{a_{n-1},0}{}^s a_{n-1}) \\ &+ f^f_{a_n}(g_{a_n,-1}{}^f a_{n-1} + g_{a_n,0}{}^f a_n) \\ &+ f^s_{a_n}(g_{a_n,-1}{}^s a_{n-1} + g_{a_n,0}{}^s a_n) \\ &= f_{both}(f_{-2}a_{n-2} + f_{-1}a_{n-1} + f_0 a_n) \end{aligned} \quad (6)$$

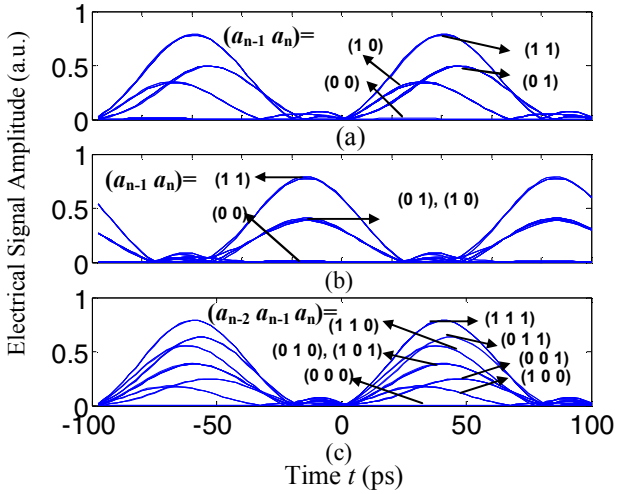


Figure 2. Eye diagrams of received signals in the presence of (a) TM with $t_{TM}$= -45 ps, (b) PMD with DGD =100 ps, and (c) TM and PMD with $t_{TM}$=-45 ps and DGD =100 ps. OBPF, α, and $\gamma$ in the figures are Gaussian shaped with 50-GHz bandwidth, 0.8, and 0.5, respectively.

where $f_{-2}+f_{-1}+f_0=1$. Note that for a fixed sampling point in time, the relative sampling phases with respect to the SM of $a_{n-1}$, the FM of $a_n$, and the SM of $a_n$ are different, leading to different values of TM coefficients, $g_{-1}$ and $g_0$. In some special cases, for example DGD=$T$ with $T$ being the bit period, we can simplify TM coefficients as $f_{a_n}^{s}=0$, $g_{a_{n-1},-1}{}^{s}=g_{a_n,-1}{}^{f}=g_{-1}$ and $g_{a_{n-1},0}{}^{s}=g_{a_n,0}{}^{f}=g_0$ because only the SM of $a_{n-1}$ and the FM of $a_n$ contribute to the sampled value and their relative sampling phase are the same. In such case, $f_{-2}$, $f_{-1}$, and $f_0$ are derived as $f_{-2}=g_{-1}r_{-1}$, $f_{-1}=g_{-1}r_0+g_0r_{-1}$, and $f_0=g_0r_0$. Fig. 2 shows the eye-diagrams of the received signal for a 10Gb/s system in the presence of (a) TM with $t_{TM}$= -45 ps, (b) PMD with DGD=100 ps, (c) both TM and PMD with $t_{TM}$= -45 ps and DGD=100 ps, respectively. OBPF, α, and $\gamma$ are 50-GHz Gaussian filter, 0.8, and 0.5, respectively. In Fig. 2(b), $f_{a_n}^{s}=0$ due to large DGD value. For a fixed sampling point in time, such as the center of the eye, the relative sampling phases with respect to the SM of $a_{n-1}$ and the FM of $a_n$ are the same, thus, $g_{a_{n-1},-1}{}^{s}=g_{a_n,-1}{}^{f}=g_{-1}$ and $g_{a_{n-1},0}{}^{s}=g_{a_n,0}{}^{f}=g_0$. In addition, notice $r_{-1}=r_0=0.5$, therefore, $f_{-2}=0.5g_{-1}$, $f_{-1}=0.5$, and $f_0=0.5g_0$, resulting in $I(t_n;1,0,1)=I(t_n;0,1,0)=I(t_n;1,1,1)/2$ which can be verified by Fig. 2(c).

## IV. Bit-to-bit Error Probability

To evaluate the performance of MLSE receiver, the distribution as well as the mean and the variance of $I(t_n; a_{n-m},\ldots, a_n)$ is required. Therefore, we write KL expansion for $n_{xopt}(t)$ and $n_{yopt}(t)$, $t\in(t_n\text{-}T_0\ t_n)$, where $T_0$ is the overall impulse response duration of the optical and the electrical filters [11]. From Appendix I, we can obtain the mean and the variance of $I(t_n; a_{n-m},\ldots, a_n)$ as:

$$I_{ave}(t_n;a_{n-m},\ldots,a_n)=I_{sig}(t_n;a_{n-m},\ldots,a_n)+2(N_0/T_0)\cdot\sum_{p=1}^{2M+1}\lambda_p \quad (7)$$

$$\begin{aligned}\sigma^2(t_n;a_{n-m},\ldots,a_n)&=2(N_0/T_0)\sum_{p=1}^{2M+1}\left|b_{x,p}(t_n;a_{n-m},\ldots,a_n)\right|^2\\&+2(N_0/T_0)\sum_{p=1}^{2M+1}\left|b_{y,p}(t_n;a_{n-m},\ldots,a_n)\right|^2\\&+2(N_0/T_0)^2\sum_{p=1}^{2M+1}\lambda_p{}^2\end{aligned} \quad (8)$$

where $M$, $\lambda_p$, $b_{x,\,p}(t_n; a_{n-m},\ldots, a_n)$, $b_{y,\,p}(t_n; a_{n-m},\ldots, a_n)$, and $I_{sig}(t_n;a_{n-m},\ldots a_n)$ are defined in Appendix I. The moment generation function (MGF) of $I(t_n; a_{n-m},\ldots, a_n)$ is:

$$\begin{aligned}&M(s;a_{n-m},\ldots,a_n)=\exp(-sI_{sig}(t_n;a_{n-m},\ldots,a_n))\cdot\\&\prod_{p=1}^{2M+1}\frac{\exp\left(\frac{s^2\left(\left|b_{x,p}(t_n;a_{n-m},\ldots,a_n)\right|^2+\left|b_{y,p}(t_n;a_{n-m},\ldots,a_n)\right|^2\right)N_0}{T_0(1+s\lambda_pN_0/T_0)}\right)}{(1+s\lambda_pN_0/T_0)^2}\end{aligned} \quad (9)$$

The distribution of $I(t_n; a_{n-m},\ldots, a_n)$ is the inverse Laplace transform of (9) [12]. Bit-to-bit error probability can be calculated directly from the MGF by using saddlepoint approximation [13], which is adopted in this paper to evaluate TM-induced power penalty with conventional detection:

$$\begin{aligned}P_e&=0.5\times E_\delta\{P(I(t_n;a_{n-m},\ldots,a_n)>\alpha)\big|_{a_n=0}\\&\quad+P(I(t_n;a_{n-m},\ldots,a_n)<\alpha)\big|_{a_n=1}\}\\&=\frac{1}{2}E_\delta\left\{\frac{\exp(\psi(s_0;a_{n-m},\ldots,a_n))}{\sqrt{2\pi\psi''(s_0;a_{n-m},\ldots,a_n)}}+\frac{\exp(\psi(s_1;a_{n-m},\ldots,a_n))}{\sqrt{2\pi\psi''(s_1;a_{n-m},\ldots,a_n)}}\right\}\end{aligned} \quad (10)$$

where $E_\delta$ is the ensemble average with $\delta$ being the set of all possible $[a_{n-m}\ldots\ a_{n-1}]$. $\psi(s; a_{n-m},\ldots, a_n)=\ln(M(s; a_{n-m},\ldots, a_n))+s\alpha-\ln|s|$. $s_0$ and $s_1$ are the negative and positive saddlepoints, respectively [13]. The optimal threshold $\alpha$ is determined numerically in practical operation.

## V. BER Evaluation of MLSE Receiver

Performance evaluation of MLSE receiver requires sequence-to-sequence error probability, which, however, is difficult to calculate due to the complexity of the distribution of $I(t_n; a_{n-m},\ldots, a_n)$. Some previous works employed the approximated closed-form expressions [8], [14]. In this paper, we use Gaussian approximation with the signal-dependent mean and variance shown in (7) and (8). Assume that in the sequence estimation using Viterbi algorithm, the estimated path $(b_0\ b_1\ldots\ b_{n-1}\ b_n)$ diverges from the correct path $(a_0\ a_1\ldots\ a_{n-1}\ a_n)$ at state $k$ and remerges with the correct path at state $k+L$, i. e. $a_k\neq b_k$ and $a_{k+L-m-1}\neq b_{k+L-m-1}$, but $a_p=b_p$ for $k\text{-}m\le p\le k\text{-}1$ and $k+L\text{-}m\le p\le k+L\text{-}1$. Define two vectors to evaluate error event as $\boldsymbol{\varepsilon}_c=[a_{k-m}\ a_{k-m+1}\ldots\ a_{k+L-2}\ a_{k+L-1}]$ and $\boldsymbol{\varepsilon}_e=[b_{k-m}\ b_{k-m+1}\ldots\ b_{k+L-2}\ b_{k+L-1}]$. BER of MLSE receiver is written as:

$$P_e\approx\sum_{\boldsymbol{\varepsilon}_c\neq\boldsymbol{\varepsilon}_e}P(\boldsymbol{\varepsilon}_c\rightarrow\boldsymbol{\varepsilon}_e)w(\boldsymbol{\varepsilon}_c,\boldsymbol{\varepsilon}_e)\left(\frac{1}{2}\right)^{L+m} \quad (11)$$

where $P(\boldsymbol{\varepsilon}_c\rightarrow\boldsymbol{\varepsilon}_e)$ is the probability of the error event $\boldsymbol{\varepsilon}_c\rightarrow\boldsymbol{\varepsilon}_e$. $w(\boldsymbol{\varepsilon}_c, \boldsymbol{\varepsilon}_e)$ is the number of nonzero components in the vector of $[(b_k\text{-}a_k)\ (b_{k+1}\text{-}a_{k+1})\ldots\ (b_{k+L-m-1}\text{-}a_{k+L-m-1})]^T$. $P_e$ is dominated by the terms involving large $P(\boldsymbol{\varepsilon}_c\rightarrow\boldsymbol{\varepsilon}_e)$, which is used to simplify (11).

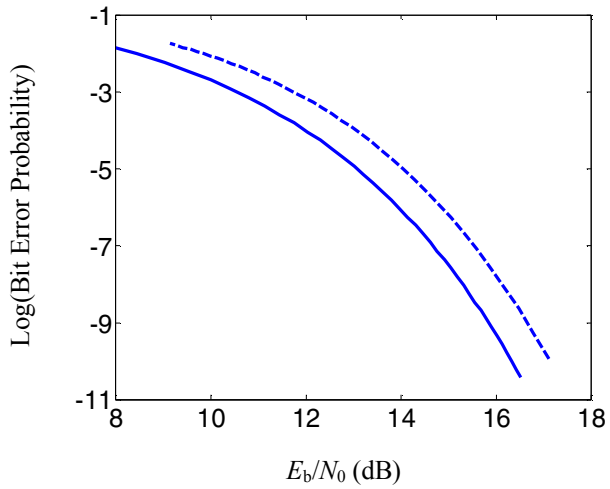


Figure 3. BER versus $E_b/N_0$ for the matched filter (solid line) and the adopted system (dashed line).

### A. Sequence-to-Sequence Error Probability

The main step to estimate $P_e$ is to calculate $P(\boldsymbol{\varepsilon}_c \rightarrow \boldsymbol{\varepsilon}_e)$. Given that $I(t_p;\ a_{p-m},\ldots,\ a_p)$ is Gaussian distributed with signal-dependent mean and variance, $k \le p \le k+L-1$, we can obtain $P(\boldsymbol{\varepsilon}_c \rightarrow \boldsymbol{\varepsilon}_e)$ by using the steepest descent method as (see Appendix II):

$$P(\boldsymbol{\varepsilon}_c \rightarrow \boldsymbol{\varepsilon}_e) = Q\left(\frac{\boldsymbol{\eta}^T \boldsymbol{k}}{\sqrt{2}}\right) \tag{12}$$

where $\boldsymbol{\eta}$ is a column vector with $L$ components of $2^{1/2} \cdot (I_{\min}(t_{p+k-1};\ \boldsymbol{\varepsilon}_c,\ \boldsymbol{\varepsilon}_e) - I_{ave}(t_{p+k-1};\ a_{p+k-m-1},\ldots,\ a_{p+k-1})) \ /\ \sigma(t_{p+k-1};\ a_{p+k-m-1},\ldots,\ a_{p+k-1})$, $1 \le p \le L$. $\boldsymbol{k}$ satisfies (A12) with $\boldsymbol{h}$ being a column vector with $L$ components of:

$$\begin{aligned} &\frac{\sqrt{2}\sigma(t_{p+k-1};a_{p+k-m-1},\ldots,a_{p+k-1})}{\sigma^2(t_{p+k-1};a_{p+k-m-1},\ldots,a_{p+k-1})} \\ &\cdot(I_{\min}(t_{p+k-1};\boldsymbol{\varepsilon}_c,\boldsymbol{\varepsilon}_e) - I_{ave}(t_{p+k-1};a_{p+k-m-1},\ldots,a_{p+k-1})) \\ &-\frac{\sqrt{2}\sigma(t_{p+k-1};a_{p+k-m-1},\ldots,a_{p+k-1})}{\sigma^2(t_{p+k-1};b_{p+k-m-1},\ldots,b_{p+k-1})} \\ &\cdot(I_{\min}(t_{p+k-1};\boldsymbol{\varepsilon}_c,\boldsymbol{\varepsilon}_e) - I_{ave}(t_{p+k-1};b_{p+k-m-1},\ldots,b_{p+k-1})) \end{aligned} \tag{13}$$

$1 \le p \le L$, where $I_{\min}(t_{p+k-1};\ \boldsymbol{\varepsilon}_c,\ \boldsymbol{\varepsilon}_e)$ is defined in Appendix II.

### B. Dominating Terms Selection

When $P(\boldsymbol{\varepsilon}_c \rightarrow \boldsymbol{\varepsilon}_e)$ is obtained, $P_e$ can be estimated by the terms involving large $P(\boldsymbol{\varepsilon}_c \rightarrow \boldsymbol{\varepsilon}_e)$ in (11). After thorough searching for large $P(\boldsymbol{\varepsilon}_c \rightarrow \boldsymbol{\varepsilon}_e)$, we give the dominating terms as follows: (1) one-bit error event, i. e. $\boldsymbol{\varepsilon}_c = [a_{k-2}\ a_{k-1}\ a_k\ a_{k+1}\ a_{k+2}]$ and $\boldsymbol{\varepsilon}_e = [b_{k-2}\ b_{k-1}\ b_k\ b_{k+1}\ b_{k+2}]$, $b_p \in \{0\ 1\}$, $k-2 \le p \le k+2$, $a_{k-2} = b_{k-2}$, $a_{k-1} = b_{k-1}$, $a_k \ne b_k$, $a_{k+1} = b_{k+1}$, and $a_{k+2} = b_{k+2}$; (2) two-bit error event and $a_k \ne a_{k+1}$, i. e. $\boldsymbol{\varepsilon}_c = [a_{k-2}\ a_{k-1}\ a_k\ a_{k+1}\ a_{k+2}\ a_{k+3}]$ and $\boldsymbol{\varepsilon}_e = [b_{k-2}\ b_{k-1}\ b_k\ b_{k+1}\ b_{k+2}\ b_{k+3}]$, $b_p \in \{0\ 1\}$, $k-2 \le p \le k+3$, $a_{k-2} = b_{k-2}$, $a_{k-1} = b_{k-1}$, $a_k \ne b_k$, $a_{k+1} \ne b_{k+1}$, $a_{k+2} = b_{k+2}$, $a_{k+3} = b_{k+3}$, and $a_k \ne a_{k+1}$. (3) $\boldsymbol{\varepsilon}_c$ and $\boldsymbol{\varepsilon}_e$ with $L > 3$ which satisfy i) $a_p \ne b_p$, $k \le p \le k+L-3$; and ii) the adjacent $a_p$, $k \le p \le k+L-3$, is different.

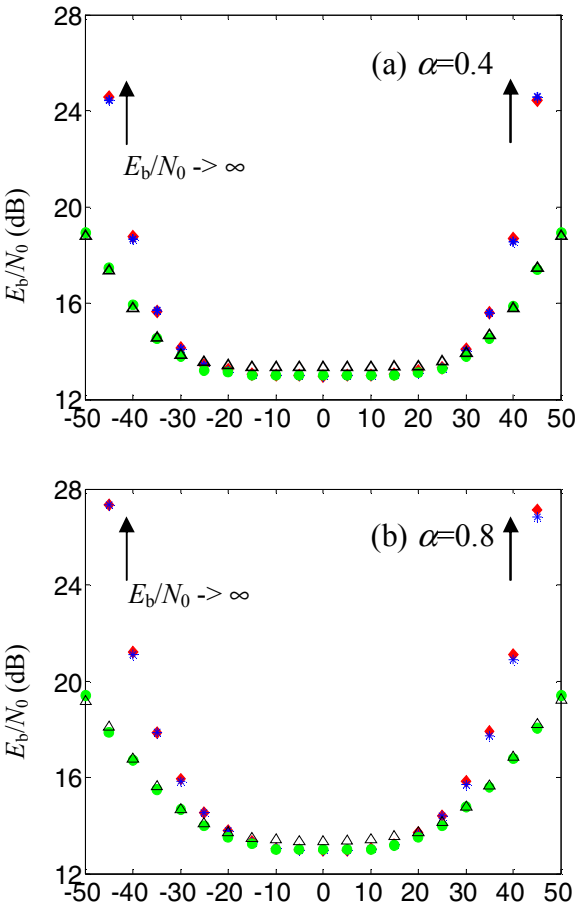


Figure 4. $E_b/N_0$ versus $t_{TM}$ for the system without PMD and with different input data pulse shape. Diamonds, asterisks, circles, and triangle-ups stand for the results from simulation without MLSE receiver, from calculation without MLSE receiver, from simulation with MLSE receiver, and from calculation with MLSE receiver, respectively.

## VI. Calculation and Simulation Results

In this section, besides the theoretical calculation, Monte Carlo simulations in a 10-Gb/s RZ system were performed. An optical RZ pulse train, consisting of 500,000 bits with 40 samples per bit, was modulated and launched into the optical fiber. $h_x(\gamma^{1/2}E_s(t))$ and $h_y((1-\gamma)^{1/2}E_s(t))$ in Fig. 1 emulated the effect of PMD with variable $\gamma$ and DGD. The OBPF was Gaussian shaped with the bandwidth of 50 GHz. The EF was a 4$^{th}$-order Butterworth filter with the optimized bandwidth in the absence of PMD and TM. The performance was evaluated in terms of $E_b/N_0$ (dB) at the BER of $10^{-4}$, where $E_b$ was the average power in one bit slot. Fig. 3 depicts the back-to-back (dashed line) performance of the system, which is compared to the performance bound (solid line) by using the matched filter. The figure shows that the $E_b/N_0$ penalty of the system is around 0.8 dB, which can be lowered close to the bound by further optimizing the bandwidths of the OBPF and the EF [15]. The ADC resolution was 5 bit to make quantization noise negligible. The metric of (4) for different states in the look-up table was obtained using non-parametric histogram method by a 200,000-bit training sequence. Fig. 4 depicts

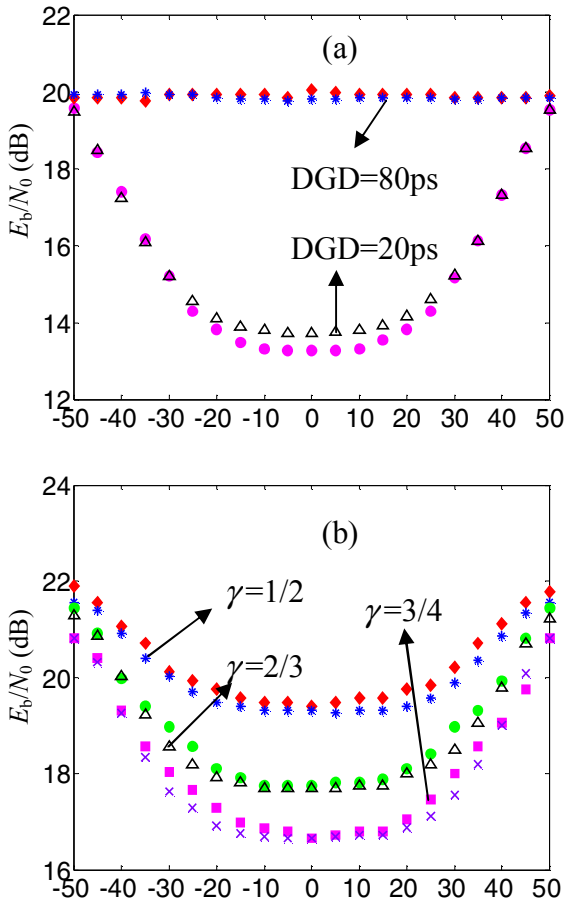


Figure 5. $E_b/N_0$ for (a) α=0.8, $\gamma$=1/2 and variable DGD; (b) α=0.8, DGD=100 ps and variable $\gamma$, when MLSE receiver is employed. Conventional detection is not shown in the figure, because in those cases, the eye is highly distorted and conventional detection leads to the power penalty -> ∞. In the figure, diamonds, circles, and squares stand for the simulated results. Asterisks, triangle-ups and crosses stand for the calculated results.

$E_b/N_0$ versus $t_{TM}$ for the system without PMD and with different input data pulse shape when the sampling phase is at the center of eye. Diamonds, asterisks, circles, and triangle-ups stand for the results from simulation without MLSE receiver, from calculation without MLSE receiver, from simulation with MLSE receiver, and from calculation with MLSE receiver, respectively. From Fig. 4, we can find that in the case of conventional detection, $E_b/N_0$ increases rapidly when $t_{TM}$ exceeds 25 ps, with the penalty profile depending on the α parameter [5]. When MLSE receiver is employed, the power penalty is lowered significantly for $|t_{TM}|$>30 ps. Thus, the tolerance to the impairment from TM is enhanced. The $E_b/N_0$ penalty for the worst TM is limited around 6 dB. We have also shown the performance of MLSE receiver to simultaneously combat PMD and TM. Fig. 5 depicts $E_b/N_0$ for (a) α=0.8, $\gamma$=1/2 and variable DGD, and (b) α=0.8, DGD=100 ps and variable $\gamma$, when MLSE receiver is employed. Diamonds, circles, and squares stand for the simulated results. Asterisks, triangle-ups and crosses stand for the calculated results. From the figure, it is shown that in the worst case of both PMD and TM where the eye is completely closed, i. e. DGD=100 ps, $t_{TM}$=-50 ps, the $E_b/N_0$ penalty of MLSE receiver is limited to around 9 dB.

## VII. Conclusion

We investigate the performance of MLSE receiver in the presence of both TM and PMD in optically amplified RZ systems. Based on the bit-to-bit error probability estimation techniques, including KL expansion, decorrelation of noise components, and saddlepoint approximation, we employ the steepest decent method to achieve the sequence-to-sequence error probability and evaluate BER of MLSE receiver with arbitrary input signal pulse shape, optical filtering and electrical filtering taken into consideration. Monte Carle simulations are performed and agree with the theory well. The results show that the power penalty for the worst TM, where the eye is completely closed, is limited by MLSE receiver to 6 dB in the absence of PMD and 9 dB in the presence of the worst PMD. The investigation validates the effectiveness of MLSE receiver for combating both TM and PMD with shared electrical devices, which, hence, relaxes the requirement for the number of compensation components.

## Appendix I

Because $n_{xopt}(t)$ and $n_{yopt}(t)$ are both AWGN, Fourier orthonormal bases are used for KL expansion. Thus, $n_{xopt}(t)$ and $n_{yopt}(t)$ are written as:

$$\begin{aligned} n_{xopt}(t) &= \sum_{p=-\infty}^{\infty} n_{xopt,p} \cdot \exp(2\pi jpt/T_0) \\ n_{yopt}(t) &= \sum_{p=-\infty}^{\infty} n_{yopt,p} \cdot \exp(2\pi jpt/T_0) \end{aligned}, \quad t\in(t_n\text{-}T_0\ t_n) \quad \text{(A1)}$$

where $n_{xopt,p}$ and $n_{yopt,p}$ are independent Gaussian variables with zero mean and the variance of their in-phase and quadrature components being $N_0/(2T_0)$. After the OBPF, $n_x(t)$ and $n_y(t)$ are:

$$\begin{aligned} n_x(t) &= \sum_{p=-M}^{M} H_o(p/T_0) n_{xopt,p} \cdot \exp(2\pi jpt/T_0) \\ n_y(t) &= \sum_{p=-M}^{M} H_o(p/T_0) n_{yopt,p} \cdot \exp(2\pi jpt/T_0) \end{aligned}, \quad t\in(t_n\text{-}T_0\ t_n) \quad \text{(A2)}$$

where $H_0(f)$ is the transfer function of the OBPF with the bandwidth evaluated by the parameter $M$ [11]. From (1), (2), (3) and (A2), $I(t_n; a_{n-m},\ldots, a_n)$ has the matrix notation as:

$$\begin{aligned} I(t_n;a_{n-m},\ldots,a_n) &= \left(R\left|E_{s,x}(t;a_{n-m},\ldots,a_n)\right|^2 \otimes h_e(t)\right)\Big|_{t=t_n} \\ &+\left(R\left|E_{s,y}(t;a_{n-m},\ldots,a_n)\right|^2 \otimes h_e(t)\right)\Big|_{t=t_n} \\ &+\boldsymbol{n}_{xopt}^{T*}\boldsymbol{v}_x(t_n;a_{n-m},\ldots,a_n)+\boldsymbol{n}_{yopt}^{T*}\boldsymbol{v}_y(t_n;a_{n-m},\ldots,a_n) \\ &+\boldsymbol{v}_x^{T*}(t_n;a_{n-m},\ldots,a_n)\boldsymbol{n}_{xopt}+\boldsymbol{n}_{xopt}^{T}\boldsymbol{Q}\boldsymbol{n}_{xopt}^{*} \\ &+\boldsymbol{v}_y^{T*}(t_n;a_{n-m},\ldots,a_n)\boldsymbol{n}_{yopt}+\boldsymbol{n}_{yopt}^{T}\boldsymbol{Q}\boldsymbol{n}_{yopt}^{*} \end{aligned} \quad \text{(A3)}$$

where * stands for the conjugate. $\boldsymbol{n}_{xopt}$ (or $\boldsymbol{n}_{yopt}$) is a column vector whose $2M$+1 components are $n_{xopt,p\text{-}M\text{-}1}$ (or $n_{yopt,p\text{-}M\text{-}1}$), $1\le p\le 2M+1$. $\boldsymbol{v}_x(t_n; a_{n\text{-}m},\ldots, a_n)$ and $\boldsymbol{v}_y(t_n; a_{n\text{-}m},\ldots, a_n)$ are column vectors whose $2M$+1 components are

$$RH_0 * ((p-M-1)/T_0) \cdot ((E_{s,x}(t;a_{n-m},...,a_n)\exp(-2\pi j(p-M-1)t/T_0)) \otimes h_e(t))\big|_{t=t_n}$$

$$RH_0 * ((p-M-1)/T_0) \cdot ((E_{s,y}(t;a_{n-m},...,a_n)\exp(-2\pi j(p-M-1)t/T_0)) \otimes h_e(t))\big|_{t=t_n} \quad \text{(A4)}$$

respectively, where $1 \le p \le 2M+1$. $\boldsymbol{Q}$ is a $(2M+1)\times(2M+1)$ matrix whose $p^{th}$-row, $q^{th}$-column element is:

$$RH_0((p-M-1)/T_0)H_0 * ((q-M-1)/T_0) \cdot (\exp(2\pi j(p-q)t/T_0) \otimes h_e(t))\big|_{t=t_n} \quad \text{(A5)}$$

where $1 \le p,q \le 2M+1$. Notice that $\boldsymbol{Q}$ is Hermitian symmetric, the eigenvalues $\lambda_p$, $1 \le p \le 2M+1$, are real and the eigenvectors are orthogonal, i. e. $\boldsymbol{Q}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{T^*}$ with $\boldsymbol{\Lambda}=\mathrm{diag}\{\lambda_p\}$ and $\boldsymbol{U}$ being an orthogonal matrix. Therefore, (A3) can be written as:

$$\begin{aligned} I(t_n;a_{n-m},...,a_n) &= I_{sig}(t_n;a_{n-m},...,a_n) + \boldsymbol{z}_{xopt}^{T^*}\boldsymbol{b}_x(t_n;a_{n-m},...,a_n) \\ &+ \boldsymbol{z}_{yopt}^{T^*}\boldsymbol{b}_y(t_n;a_{n-m},...,a_n) + \boldsymbol{b}_x^{T^*}(t_n;a_{n-m},...,a_n)\boldsymbol{z}_{xopt} \\ &+ \boldsymbol{b}_y^{T^*}(t_n;a_{n-m},...,a_n)\boldsymbol{z}_{yopt} + \boldsymbol{z}_{xopt}^{T}\boldsymbol{\Lambda}\boldsymbol{z}_{xopt}^{\ *} + \boldsymbol{z}_{yopt}^{T}\boldsymbol{\Lambda}\boldsymbol{z}_{yopt}^{\ *} \end{aligned} \quad \text{(A6)}$$

where $\boldsymbol{z}_{\mathrm{xopt}}$ (or $\boldsymbol{z}_{\mathrm{yopt}}$) is $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{n}_{\mathrm{xopt}}$ (or $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{n}_{\mathrm{yopt}}$). $\boldsymbol{b}_{\mathrm{x}}$ ($t_{\mathrm{n}}$; $a_{\mathrm{n\text{-}m}},\ldots,$ $a_{\mathrm{n}}$) (or $\boldsymbol{b}_{\mathrm{y}}$($t_{\mathrm{n}}$; $a_{\mathrm{n\text{-}m}},\ldots,$ $a_{\mathrm{n}}$)) is $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{v}_{\mathrm{x}}$($t_{\mathrm{n}}$; $a_{\mathrm{n\text{-}m}},\ldots,$ $a_{\mathrm{n}}$) (or $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{v}_{\mathrm{y}}$($t_{\mathrm{n}}$; $a_{\mathrm{n\text{-}m}},\ldots,$ $a_{\mathrm{n}}$)). $I_{\mathrm{sig}}$($t_{\mathrm{n}}$;$a_{\mathrm{n\text{-}m}},\ldots a_{\mathrm{n}}$) represents the received deterministic signal component in the absence of noise. As $\boldsymbol{U}$ is an orthogonal matrix, the components of $\boldsymbol{z}_{\mathrm{xopt}}$ (or $\boldsymbol{z}_{\mathrm{yopt}}$) are Gaussian variables with zero mean and the variance of their in-phase and quadrature components being $N_0/(2T_0)$. Let $b_{\mathrm{x,p}}$($t_{\mathrm{n}}$; $a_{\mathrm{n\text{-}m}},\ldots,$ $a_{\mathrm{n}}$) (or $b_{\mathrm{y,p}}$($t_{\mathrm{n}}$; $a_{\mathrm{n\text{-}m}},\ldots,$ $a_{\mathrm{n}}$)) be the $p^{\mathrm{th}}$ component of $\boldsymbol{b}_{\mathrm{x}}$($t_{\mathrm{n}}$; $a_{\mathrm{n\text{-}m}},\ldots,$ $a_{\mathrm{n}}$) (or $\boldsymbol{b}_{\mathrm{y}}$($t_{\mathrm{n}}$; $a_{\mathrm{n\text{-}m}},\ldots,$ $a_{\mathrm{n}}$)), $1 \le p \le 2M+1$, the mean, the variance, and the moment generation function of $I$($t_{\mathrm{n}}$; $a_{\mathrm{n\text{-}m}},\ldots,$ $a_{\mathrm{n}}$) are derived from (A6) as (7), (8) and (9), respectively.

## APPENDIX II

MLSE receiver chooses the error path if:

$$\sum_{p=k}^{k+L-1} -\ln(p(I(t_p)\,|\,a_{p-m},...,a_p)) > \sum_{p=k}^{k+L-1} -\ln(p(I(t_p)\,|\,b_{p-m},...,b_p)) \quad \text{(A7)}$$

Let $\boldsymbol{I}=[I(t_{\mathrm{k}})\ldots I(t_{\mathrm{k+L\text{-}1}})]$ and

$$\begin{aligned} F(\boldsymbol{I};\boldsymbol{\varepsilon}_c) &= \sum_{p=k}^{k+L-1} -\ln(p(I(t_p)\,|\,a_{p-m},...,a_p), \\ F(\boldsymbol{I};\boldsymbol{\varepsilon}_e) &= \sum_{p=k}^{k+L-1} -\ln(p(I(t_p)\,|\,b_{p-m},...,b_p) \end{aligned} \quad \text{(A8)}$$

which are the functions with $L$-dimension variables. Define $B(\boldsymbol{I};\boldsymbol{\varepsilon}_{\mathrm{c}}, \boldsymbol{\varepsilon}_{\mathrm{e}})$ be the locus of all points in $L$-dimension space such that $F(\boldsymbol{I};\ \boldsymbol{\varepsilon}_{\mathrm{c}})= F(\boldsymbol{I};\ \boldsymbol{\varepsilon}_{\mathrm{e}})$. Let $\boldsymbol{I}_{\min}(\boldsymbol{\varepsilon}_{\mathrm{c}},\ \boldsymbol{\varepsilon}_{\mathrm{e}}) =[I_{\min}(t_{\mathrm{k}};\ \boldsymbol{\varepsilon}_{\mathrm{c}},\ \boldsymbol{\varepsilon}_{\mathrm{e}})\ldots I_{\min}(t_{\mathrm{k+L\text{-}1}};\ \boldsymbol{\varepsilon}_{\mathrm{c}},\ \boldsymbol{\varepsilon}_{\mathrm{e}})]$ be the vector in $B(\boldsymbol{I};\ \boldsymbol{\varepsilon}_{\mathrm{c}},\ \boldsymbol{\varepsilon}_{\mathrm{e}})$ that minimizes $F(\boldsymbol{I};\ \boldsymbol{\varepsilon}_{\mathrm{c}})$, $P(\boldsymbol{\varepsilon}_{\mathrm{c}} \to \boldsymbol{\varepsilon}_{\mathrm{e}})$ can be expressed as [8]:

$$P(\boldsymbol{\varepsilon}_c \to \boldsymbol{\varepsilon}_e) = \exp\left(\frac{\boldsymbol{\eta}^{T^*}\boldsymbol{\eta}}{4} - F(\boldsymbol{I}_{\min}(\boldsymbol{\varepsilon}_e,\boldsymbol{\varepsilon}_c);\boldsymbol{\varepsilon}_c)\right) \cdot Q\left(\frac{\boldsymbol{\eta}^T\boldsymbol{k}}{\sqrt{2}}\right)\prod_{p=k}^{k+L-1}\left(\frac{\pi}{u_p}\right)^{1/2} \quad \text{(A9)}$$

where

$$u_p = \frac{1}{2}\frac{\partial^2 F(\boldsymbol{I};\boldsymbol{\varepsilon}_c)}{\partial I^2(t_p)}\bigg|_{\boldsymbol{I}=\boldsymbol{I}_{\min}(\boldsymbol{\varepsilon}_c,\boldsymbol{\varepsilon}_e)} \quad \text{(A10)}$$

$\boldsymbol{\eta}$ is a column vector with $L$ components of

$$\eta_p = \frac{1}{\sqrt{u_{p+k-1}}}\frac{\partial F(\boldsymbol{I};\boldsymbol{\varepsilon}_c)}{\partial I(t_{p+k-1})}\bigg|_{\boldsymbol{I}=\boldsymbol{I}_{\min}(\boldsymbol{\varepsilon}_c,\boldsymbol{\varepsilon}_e)}, 1 \le p \le L \quad \text{(A11)}$$

and $\boldsymbol{k}$ is

$$\boldsymbol{k} = \frac{\boldsymbol{h}}{\sqrt{\boldsymbol{h}^{T^*}\boldsymbol{h}}}, \boldsymbol{h} = \boldsymbol{u}^T \cdot \nabla(F(\boldsymbol{I};\boldsymbol{\varepsilon}_c) - F(\boldsymbol{I};\boldsymbol{\varepsilon}_e))\big|_{\boldsymbol{I}=\boldsymbol{I}_{\min}(\boldsymbol{\varepsilon}_c,\boldsymbol{\varepsilon}_e)} \quad \text{(A12)}$$

where $\boldsymbol{u}=[u_{\mathrm{k}}^{-1/2}\ \ u_{\mathrm{k+1}}^{-1/2}\ldots\ u_{\mathrm{k+L\text{-}1}}^{-1/2}]$. Given the Gaussian distribution with signal-dependent mean and variance of $I$($t_{\mathrm{n}}$; $a_{\mathrm{n\text{-}m}},\ldots,$ $a_{\mathrm{n}}$), (12) is obtained from (A9).

## REFERENCES


[1]. M. Pauer, P. J. Winzer, and W. R. Leeb, "Bit error probability reduction in direct detection optical receivers using RZ coding", *IEEE J. Lightwave Technol.*, vol. 19, pp. 1255-1262, 2001.

[2]. H. Sunnerud, M. Karlsson, and P. A. Andrekson, "A comparison between NRZ and RZ data formats with respect to PMD-induced system degradation", *IEEE Photon. Technol. Lett.*, vol. 13, pp. 448-450, 2001.

[3]. J. H. Sinsky, "High-speed data and pulse carver alignment in dual Mach-Zehnder modulator optical transmitter using microwave signal processing", *IEEE J. Lightwave Technol.,* vol. 21, pp. 412-423, 2003.

[4]. I. Kang, L. Mollenauer, B. Greene, and A. Grant, "A novel method for synchronizing the pulse carver and electroabsortion data modulator for ultralong-hual DWDM transmission", *IEEE Photon. Technol. Lett.*, vol. 14, pp. 1357-1359, 2002.

[5]. J. Zhao, L. K. Chen, C. K. Chan, "Mitigation of Timing Misalignment Induced Distortion Using Electronic Equalizer in RZ/CSRZ Systems", *IEEE Photon. Technol. Lett.*, vol. 17, pp. 1106-1108, 2005.

[6]. H. F. Haunstein W. Sauer-Greff, A. Dittrich, K. Sticht, and R. Urbansky, "Principles for electronic equalization of polarization-mode dispersion", *IEEE J. Lightwave Technol.,* vol. 22, pp. 1169-1182, 2004.

[7]. F. Buchali, H. Bulow, "Adaptive PMD compensation by electrical and optical techniques", *IEEE J. Lightwave Technol.,* vol. 22, pp. 1116-1126, 2004.

[8]. O. E. Agazzi, M. R. Hueda, H. S. Carrer, and D. E. Crivelli, "Maximum likelihood sequence estimation in dispersive optical channels", *IEEE J. Lightwave Technol.,* vol. 23, pp.749-763, 2005.

[9]. C. Lawetz and J. C. Cartledge, "Performance of optically preamplified receivers with Fabry-Perot optical filters", *IEEE J. Lightwave Technol.,* vol. 14, pp. 2467-2474, 1996.

[10]. F. Buchali, H. Bulow, "Correlation sensitive Viterbi equalization of 10 Gb/s signals in bandwidth limited receivers", in *Proc. Optical Fiber Communication Conference* (OFC), OFO2, 2005.

[11]. E. Forestieri, "Evaluating the error probability in lightwave systems with chromatic dispersion, arbitrary pulse shape and pre- and post-detection filtering", *IEEE J. Lightwave Technol.*, vol. 18, pp. 1493-1503, 2000.

[12]. N. Alic, G. C. Papen, R. E. Saperstein, L. B. Milstein, and Y. Fainman, "Signal statistics and maximum likelihood sequence estimation in intensity modulated fiber optic links containing a single optical amplifier", *Optics Express*, vol. 13, pp. 4568-4579, 2005.

[13]. C. W. Helstrom, "Distribution of the filtered output of a quadratic rectifier computed by numerical contour integration", *IEEE Trans. Inf. Theory,* vol. 32, pp. 450-463, 1986

[14]. A. J. Weiss, "On the performance of electrical equalization in optical fiber transmission systems", *IEEE Photon. Technol. Lett.*, vol. 15, pp. 1225-1227, 2003.

[15]. P. J. Winzer, M. Pfennigbauer, M. M. Strasser, and W. R. Leeb, "Optimum bandwidths for optically preamplfied NRZ receivers", *IEEE J. Lightwave Technol.*, vol. 19, pp. 1263-1273, 2000.