Maximum-Likelihood Sequence Estimation in Dispersive Optical Channels

Oscar E. Agazzi, *Fellow, IEEE*, Mario R. Hueda, Hugo S. Carrer, and Diego E. Crivelli

***Abstract*—This paper discusses the investigation of maximum-likelihood sequence estimation (MLSE) receivers operating on intensity-modulated direct-detection optical channels. The study focuses on long-haul or metro links spanning several hundred kilometers of single-mode fiber with optical amplifiers. The structure of MLSE-based optical receivers operating in the presence of dispersion and amplified spontaneous emission (ASE), as well as shot and thermal noise, are discussed, and a theory of the error rate of these receivers is developed. Computer simulations show a close agreement between the predictions of the theory and simulation results. Some important implementation issues are also addressed. Optical channels suffer from impairments that set them apart from other channels, and therefore they need a special investigation. Among these impairments are the facts that the optical channel is nonlinear, and noise is often non-Gaussian and signal dependent. For example, in optically amplified single-mode fiber links, the dominant source of noise is ASE noise, which after photodetection is distributed according to a noncentral chi-square probability density function. In addition, optical fibers suffer from chromatic and polarization-mode dispersion (PMD). Although the use of MLSE in optical channels has been discussed in previous literature, no detailed analysis of optical receivers using this technique has been reported so far. This motivates the study reported in this paper.**

***Index Terms*—Channel estimation, chromatic dispersion, electronic dispersion compensation (EDC), equalization, maximum-likelihood sequence estimation (MLSE), non-Gaussian noise, polarization-mode dispersion (PMD).**

I. Introduction

TRADITIONAL optical receivers perform a minimal amount of signal processing. In most applications today, channel impairments either are small and left uncompensated or are compensated using optical techniques. Until recently, the operations after optical-to-electrical (O/E) conversion have been limited to clock recovery and data slicing. However, owing to advances in technology, today it is possible to perform significantly more elaborate signal processing functions at the receiver. This could be used to great advantage to compensate some of the impairments of the optical channel. Electronic compensation of channel impairments such as chromatic or polarization-mode dispersion (PMD) improves performance and brings enormous flexibility resulting from the opportunity to apply adaptive signal processing algorithms not practical in the optical domain.

Electrical signal processing techniques to compensate impairments of the optical channel have been extensively discussed in the literature. A comprehensive study of these techniques was done in [1] and [3]. Decision-feedback equalization (DFE) was proposed in [4] to compensate for modal dispersion in multimode fibers. A discussion of different equalization schemes applied to fiber optics, including a simulation-based study of the performance of maximum-likelihood sequence estimation (MLSE) was done in [2] and references thereof. Another simulation-based study of MLSE applied to optical channels was presented in [5]. Implementation of electronic equalization systems for fiber optics was described in [6] and [2], and references thereof. A comparison of the performance of the DFE and MLSE on single-mode fiber links was done in [7], where it was shown that the nonlinear nature of the single-mode fiber optic channel severely limits the performance of the former but poses no particular problems for the latter.

In this paper, a new theory of the performance of adaptive MLSE receivers operating on nonlinear channels in the presence of non-Gaussian signal-dependent noise (of which the intensity modulation/direct detection optical channel is an example) is proposed. This theory builds upon the one introduced in [8] for the special case of a receiver perfectly matched to the channel. Here, the theory is extended to the general case where the channel model available to the receiver does not match the channel exactly.

Chromatic dispersion is a relatively small impairment in optical links operating at data rates below 10 Gb/s, but it grows quickly as the data rate increases and becomes serious at 10 Gb/s and beyond. PMD, which occurs as a result of birefringence in the optical fiber, also becomes important at high data rates. Birefringence is caused by manufacturing defects and by stress, vibration, and other mechanical effects on the fiber. A typical manifestation of PMD is pulse splitting [a single transmitted pulse splits into two components that travel at different velocities and therefore arrive at the receiver at different times, causing intersymbol interference, (ISI)]. As a result of its dependence on stress and vibration, as well as on random changes in the state of polarization of the laser, PMD is nonstationary. For this reason, adaptive signal processing techniques are ideally suited to PMD compensation.

Several purely optical techniques exist to control dispersion. One of them is the use of lasers with a wavelength in the region around 1310 nm, where chromatic dispersion reaches a minimum. However, in long-haul and metro links, wavelengths

Manuscript received May 14, 2004; revised August 9, 2004. This paper was presented in part at the IEEE International Conference on Communications (ICC), Paris, France, June 2004.

O. E. Agazzi was with the Broadcom Corporation, Irvine, CA 92618 USA. He is now with ClariPhy Communications, Inc., Irvine, CA 92618 USA (e-mail: oscar.agazzi@clariphy.com).

M. R. Hueda, H. S. Carrer, and D. E. Crivelli are with the Digital Communications Research Laboratory, National University of Cordoba, 5000 Cordoba, Argentina.

Digital Object Identifier 10.1109/JLT.2004.838870

in the 1550-nm region, where the fiber attenuation is minimum and significantly lower than at 1310 nm, are generally preferred. Chromatic dispersion can also be compensated by the use of dispersion compensation fibers (DCFs) [9]. These are fibers where the slope of the delay versus wavelength curve has an opposite sign compared with normal fibers. Reels of appropriate lengths of DCF are placed at certain points in the link to compensate dispersion. Unfortunately, DCFs also cause significant attenuation, and their length has to be manually adjusted to achieve proper compensation, so link provisioning becomes expensive and time consuming. Although other optical techniques to compensate dispersion exist, in general they suffer from the problems of being costly and requiring manual adjustment. Some hybrid techniques, where the dispersion compensation is done optically but the transfer function of the optical compensation element is adjusted using an electronic error signal, have been described in the literature [10], [11]. A comparison of optical and electrical techniques to compensate PMD was done in [6]. A purely electronic solution has the advantage of higher integration and easier and faster adaptation of the compensation function. Receiver-based adaptive equalization has been applied for many years in a wide variety of nonoptical communications systems. Its application to optical receivers alleviates the need for DCFs and other costly optical dispersion compensation techniques. It also benefits from the automatic adaptation of the equalizer, thus eliminating the need for manual adjustment of optical compensation elements or the need to close the adaptation loop externally.

Because of the unique properties of the optical channel, the theory presented here departs significantly from previous work on MLSE receivers. Among the properties that require a new study are 1) the nonlinear nature of the optical channel and 2) the fact that the noise is non-Gaussian and signal dependent. Furthermore, the fact that in general the signal statistics are not known *a priori* implies that the problem of channel estimation is significantly more involved than when the noise is Gaussian and signal independent. In the latter case, channel estimation means simply estimating the signal *mean* for each branch in the trellis diagram of the receiver. If in addition the channel is linear, the signal mean for a given branch can be computed as the convolution of the channel impulse response with the sequence of symbols associated with the branch; therefore, the problem of channel estimation reduces to the estimation of the channel impulse response. In the case of interest in this paper, the entire probability density function (pdf) of the signal must be estimated for each branch in the trellis of the receiver. This problem is analyzed in considerable detail in this paper.

The paper is organized as follows. The optical channel is described in Section II, the receiver in Section III, and techniques to adaptively estimate the channel in Section IV. Techniques to predict the error rate of the optical receiver in the presence of dispersion, nonlinearity, amplified spontaneous emission (ASE), and thermal noise are developed in Section V. The techniques presented in Section V are not limited to optical channels. They can be applied to any channel where the noise is non-Gaussian and signal dependent. The effect on the bit-error rate (BER) of inaccuracies of the channel model used by the receiver, which may occur because of implementation constraints, is analyzed in Section VI. The predictions of the theory are compared with simulation results in Section VII. Single-chip integration of the MLSE receiver in complementary metal–oxide–semiconductor (CMOS) technology is briefly discussed in Section VIII. The focus is on the analog-to-digital converter (ADC), one the main challenges in the design of high-speed MLSE-based receivers. Finally, conclusions are drawn in Section IX.

II. The Optical Channel

Figs. 1 and 2 show the models of the channel and the receiver that will be used in this paper. The transmitter modulates the intensity of the transmitted signal using a binary alphabet. The optical fiber introduces chromatic and polarization-mode dispersion, as well as attenuation. To compensate the attenuation, optical amplifiers are deployed periodically over the length of the fiber. Optical amplifiers introduce ASE noise. At the receiver the signal is band limited by an optical filter, and then converted to a current by a PIN diode or avalanche photodetector. The resulting photocurrent is amplified and filtered, and the output of the filter is sampled at the symbol rate. The samples are applied to the detector. The detector is described in Section III. The samples of the received signal are represented as

$$y_n = e_n + n_n + z_n = x_n + z_n \tag{1}$$

where

$$e_n = f(a_n, a_{n-1}, \ldots, a_{n-\delta+1}) \tag{2}$$

represents the noise-free received optical signal (after O/E conversion), which is in general a nonlinear function of a group of δ consecutive transmitted bits $a_n, \ldots, a_{n-\delta+1}$, n_n are samples of the ASE noise, z_n are samples of the electrical noise, and $x_n = e_n + n_n$. In the case of interest in this paper, z_n includes shot noise (the result of the quantum nature of light), and thermal noise from the analog front end of the receiver. Nonlinearity in the optical channel stems predominantly from the photodetector. Current in photodetectors is proportional to the received optical power, which is a quadratic function of the electromagnetic-field amplitude in the fiber. Another source of nonlinearity is the intensity dependence of the index of refraction of the fiber [this effect is important, for example, in dense-wavelength-division-multiplexed (DWDM) links where the optical power is the aggregate of many channels]. It has been shown [12] that the nonlinearity of the received signal depends on the spectral width of the source. When chromatic dispersion is present, the signal at the output of the photodetector in a single-mode fiber driven by a narrow-band source is a quadratic function of the δ most recent transmitted bits [13]. An expression for $f(a_n, a_{n-1}, \ldots, a_{n-\delta+1})$ in terms of second-order Volterra kernels is discussed in [7]. Nonlinearity in optical channels can vary over a wide range. In this paper, we assume the most general case of arbitrary nonlinearity.

Thermal noise is Gaussian. Shot noise has a Poisson distribution, but for large numbers of incident photons (the case of interest here), its distribution can be closely approximated by a Gaussian. In the rest of our discussion, it will be assumed that z_n is additive white Gaussian noise with power σ_z^2. On the other

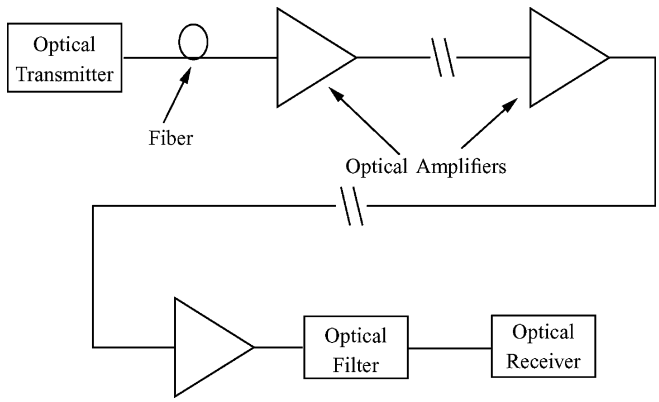

Fig. 1. Optical channel model.

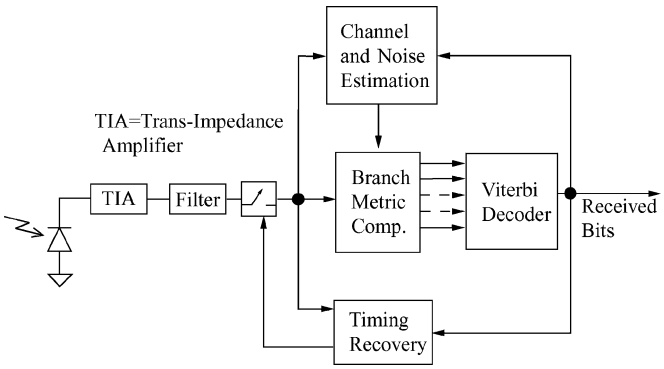

Fig. 2. Simplified optical receiver model.

hand, after photodetection, ASE noise is strongly non-Gaussian and signal dependent. It has been shown [14], [15] that the pdf of x_n is noncentral chi-square with $2M$ degrees of freedom

$$p_{\mathrm{ASE}}(x_n \mid e_n) = \frac{1}{N_0}\left(\frac{x_n}{e_n}\right)^{(M-1)/2} \exp\left(-\frac{x_n+e_n}{N_0}\right)\mathcal{I}_{M-1}\left(2\frac{\sqrt{x_n e_n}}{N_0}\right). \quad (3)$$

In (3), N_0 is related to the variance of the noise in the electromagnetic-field domain, M is the ratio of the O/E bandwidth of the front end, and $\mathcal{I}_m(\cdot)$ is the mth modified Bessel function of the first kind. For a detailed study of the statistical properties of ASE noise and a discussion of the physical meaning of the parameters of (3), the reader is referred to [15]. The pdf for the sum of the Gaussian and chi-square components of noise does not have a closed-form analytical expression, but Marcuse [14] proposes an accurate numerical approximation based on the method of steepest descent. This approximation will be used in Section VII to compute BERs for the MLSE receiver.

III. MLSE Receiver for Nonlinear Signals in the Presence of Non-Gaussian, Signal-Dependent Noise

The optimal sequence detector for signals affected by ISI and additive Gaussian noise consists of a whitened matched filter followed by a Viterbi decoder [16]. It is well known that the samples of the signal taken at the output of the whitened matched filter at the symbol rate constitute a set of sufficient statistics for the detection. In the case of non-Gaussian noise, the problem of obtaining a set of sufficient statistics by sampling a filtered version of the input signal at the symbol rate has not been solved. In this paper, we assume that the output of the photodetector is filtered and then sampled at the symbol rate, but we do not assume that the input filter is a matched filter. We assume that the samples of the signal plus noise are independent, but they are not identically distributed.

Let N be the total number of symbols transmitted. The maximum-likelihood sequence detector chooses, among the 2^N possible sequences, the one that minimizes the metric

$$m_r = \sum_{k=1}^{N} -\ln\left(p(y_k \mid e_k^{(r)})\right) \quad (4)$$

where $p(y_k \mid e_k^{(r)})$ is the conditional pdf of the received signal y_k

$$e_k^{(r)} = f(a_k^{(r)}, a_{k-1}^{(r)}, \ldots, a_{k-\delta+1}^{(r)}) \quad (5)$$

and $(a_1^{(r)}, a_2^{(r)}, \ldots, a_N^{(r)})$ $(r = 1, \ldots, 2^N)$ are the 2^N candidate sequences.

The minimization can be efficiently implemented using the Viterbi algorithm. For a channel with δ symbols of memory as in (2), the trellis has $2^{\delta-1}$ states. It is interesting to note from (3) and (4) that, unlike in the Gaussian channel where the branch metrics are simple Euclidean distances, in the optical channel computation of the branch metrics in general requires the evaluation of different functions for each branch. This is the result of the fact that the noise is signal dependent. The different functions that must be evaluated are the logs of the conditional probability density functions of the signal given the bits associated with the branches whose metrics are being computed. These are the terms in the metrics of (4). In general, the functions representing the branch metrics do not have a closed-form analytical expression.

IV. Channel Estimation

In most practical cases, the receiver does not have *a priori* knowledge of the parameters of the noise pdf, such as N_0, M, or σ_z and the nonlinear dispersion function $f(a_n, a_{n-1}, \ldots, a_{n-\delta+1})$. Therefore, they must be estimated from the received signal itself. The channel estimation methods we consider in this paper are *decision directed*; in other words, they assume that the receiver is operating normally and making decisions with a sufficiently low error rate. At the beginning of the operation, the channel estimator can be initialized with relatively crude approximations to the expected value of the signal and the pdf of the noise (for example, the former could ignore the ISI, and the latter could be initialized with a Gaussian function). Although this will result in a high initial error rate, the estimation algorithm will typically converge, and the error rate will be gradually reduced as the channel estimation improves, until convergence is completed. Convergence in decision-directed mode is not guaranteed, but in practice it is found to be very robust. Conditions for convergence of decision-directed training of various equalizer structures have been studied in [17] and references thereof. It is also possible, although less convenient, to train the channel estimator using

a special training sequence known by the receiver and sent by the transmitter during startup [18] or using blind equalization techniques [19].

Channel estimation methods may be *parametric* or *nonparametric*. Parametric methods assume that the functional form for the pdf of the signal is known, but its parameters are not, whereas nonparametric methods do not assume any knowledge of the pdf. In this paper, we consider a parametric method, the *method of moments* (MoM), and a nonparametric one, the *histogram method*. Other methods exist, but they will not be discussed in this paper.

A. Method of Moments

This method can be applied when the functional form of the pdf is known, but the values of the parameters are not. This is the case of interest in this paper, since the noise is assumed to be the sum of an ASE and a Gaussian component. From (3), it is clear that we need to estimate parameters N_0, M, and σ_z. Also the values of $f(a_n, a_{n-1}, \ldots, a_{n-\delta+1})$ are needed for all 2^δ branches in the trellis. All of these quantities can be expressed in terms of the moments of the pdf of the signal, and their values estimated by the sample moments of the input signal. Let $\mu(a_n, a_{n-1}, \ldots, a_{n-\delta+1})$ be the mean, $v(a_n, a_{n-1}, \ldots, a_{n-\delta+1})$ the variance, and $q(a_n, a_{n-1}, \ldots, a_{n-\delta+1})$ the third-order central moment of the pdf associated with branch $(a_n, a_{n-1}, \ldots, a_{n-\delta+1})$ in the trellis. Also define $\mu_0 = \mu(0,0,\ldots,0)$, $\mu_1 = \mu(1,1,\ldots,1)$, $v_0 = v(0,0,\ldots,0)$, $v_1 = v(1,1,\ldots,1)$, $E_0 = f(0,0,\ldots,0)$, and $q_0 = q(0,0,\ldots,0)$. Then, it is easy to show that

$$N_0 = \frac{v_1 - v_0}{2(\mu_1 - \mu_0)} \tag{6}$$

$$M = \frac{6\mu_0 N_0^2 - q_0}{4N_0^3} \tag{7}$$

$$E_0 = \mu_0 - MN_0 \tag{8}$$

$$\sigma_z^2 = v_0 - 2E_0N_0 - MN_0^2. \tag{9}$$

Estimates for N_0, M, E_0, and σ_z^2 can be computed from (6)–(9) by replacing the moments of the pdf by the corresponding sample moments of the signal. Notice that all parameters of the ASE and Gaussian noise are estimated based on moments associated with only two branches of the trellis, independently of the number of states of the Viterbi decoder. However, estimation of the noise-free signal values $f(a_n, a_{n-1}, \ldots, a_{n-\delta+1})$ requires that the signal mean be known for all branches. This is so because we do not constrain the nonlinearity of the channel and specifying the most general nonlinear function of δ bits requires that all its 2^δ values be specified. Equations similar to (8) can then be written to obtain estimates of $f(a_n, a_{n-1}, \ldots, a_{n-\delta+1})$ for all combinations of values of $a_n, a_{n-1}, \ldots, a_{n-\delta+1}$. This provides the complete estimation of the channel needed to implement the Viterbi algorithm as described in the previous section.

B. Histogram Method

In this method, 2^δ histograms (one for each combination of values of the receiver estimates of the δ most recent received bits $\hat{a}_n, \hat{a}_{n-1}, \ldots, \hat{a}_{n-\delta+1}$) are created.

The signal is assumed to be quantized to K bits; therefore, each histogram consists of at most 2^K bins, where K is a design parameter. Notice that each histogram can be uniquely associated with a branch in the trellis diagram of the receiver. Assuming that the number of signal samples collected is large, the histogram (normalized so that the sum of all its bins is unity) is an estimate of $p(y_n \mid e_n)$. The histogram is updated iteratively, based on the observed data. Assume that blocks of L samples are collected and that $l_k^{(n)}$ of the samples of the nth block fall into bin k of the histogram. Let $p_k^{(n)}$ be the estimate of the probability corresponding to bin k after the nth block of data is processed, and let $q_k^{(n)} = l_k^{(n)}/L$. It is shown in the Appendix ((82)) that the estimator update equation is

$$p_k^{(n)} = (1 - \lambda_k)p_k^{(n-1)} + \lambda_k q_k^{(n)} \tag{10}$$

where λ_k is a parameter that controls the speed of the update, which may be different for different bins. At any given iteration n, $p_k^{(n-1)}$ is treated as the prior estimate of the probability for bin k, and $p_k^{(n)}$ is the posterior. Thus, λ_k also controls how much weight is given to the new evidence compared with the prior. If no prior knowledge of the signal pdf is available when the receiver begins to operate, λ_k must be made equal to 1 initially so that the estimate of the pdf is based exclusively on observed samples (notice that a training sequence would be required in this case, since the receiver cannot be started in decision-directed mode without making *some* assumption about the signal pdf, on which the branch metric computations are based). In successive iterations, λ_k is decreased, thus assigning increasing weights to the observations previously accumulated. The steady-state values of λ_k are design parameters. They depend on the block size L and on how much weight it is desired to give to the new observations versus the ones previously accumulated.

The main difficulty with the histogram method is that a large number of samples is needed to obtain accurate estimates. This is particularly problematic in the tail regions of the pdf, where it may take an inordinate amount of time to obtain enough samples. In addition, samples in the tail regions could be corrupted by decision errors, even at low BERs, although this problem can be virtually eliminated using forward-error correction. On the other hand, the MoM (Section IV-A) can yield accurate estimates with a reasonable number of samples if the functional form assumed is valid. However, in certain situations, the true signal pdf may depart from the assumed functional form. One of these situations may occur, for example, in DWDM systems affected by four-wave mixing (FWM) and/or cross-phase modulation (XPM). If the number of aggressors affecting the channel under study is small, the interference could be significantly non-Gaussian. In these situations, the pdf obtained using the MoM as described in Section IV-A could be a reasonable initial approximation to the true pdf, but actual information collected from the samples of the received signal would significantly improve the accuracy of the channel estimation. To apply this method, the bins of the histogram are initialized using the best approximation to the pdf available before signal samples are collected, which we call *the prior* pdf. The prior could be obtained using the MoM or any other source of prior knowledge about the

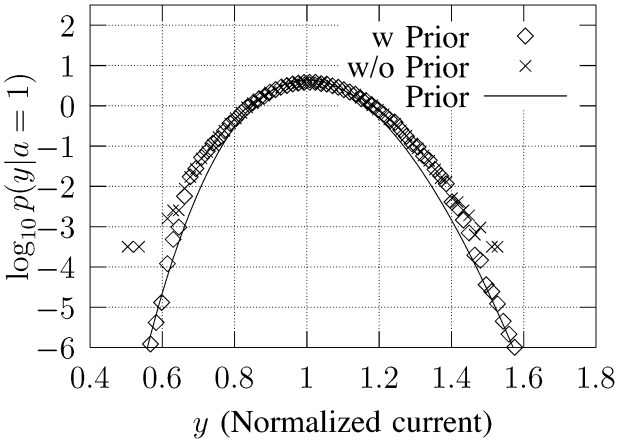

Fig. 3. Comparison between the histogram method with (w) and without (w/o) prior initialization, $L = 10^5$, and $m = 100$.

channel. The histogram is then updated iteratively, as discussed previously.

As an example, we compare results obtained from the histogram method with and without initialization. We consider transmissions over a nondispersive optical channel in the presence of ASE and thermal noise. Fig. 3 shows estimates of the signal pdf corresponding to $a_n = 1$. The figure was generated after receiving two data blocks of length $L = 10^5$ each, using $m = 100$ (see Appendix). The optical signal-to-noise ratio (OSNR) was OSNR = 16 dB (see (67)), the signal-to-Gaussian (electrical)-noise ratio (SGNR) was SGNR = 20 dB (see (68)), and the ratio of the O/E bandwidth of the front end was $M = 3$. In this example, we used $p_{\mathrm{ASE}}(x_n \mid e_n)$ given by (3) as the prior pdf. To estimate the parameters of this pdf, we applied the MoM to the first block of L samples. Then, expression (10) was used to update the pdf estimate using the rest of the data. From Fig. 3, note that the values derived from the histogram method without initialization are inaccurate in the tail regions of the pdf, owing to the small number of samples available. Initialization greatly improves the estimation in this region as a result of the use of prior knowledge. This illustrates the advantages of combining the observations with prior knowledge of the signal pdf.

In the following, we explore the effect of the selection of the prior density. Fig. 4 depicts the normalized decision threshold calculated using the histogram method with prior pdf on a dispersion-free channel. Two priors are considered: Gaussian and chi-square. For the Gaussian, the mean and variance are estimated from the first block of $L = 10^3$ received samples. For the chi-square, the parameters are estimated using the MoM on the first block of L samples. In the figure, we see the evolution of both thresholds as the number of blocks of received samples increases. Note that for the case of a Gaussian prior, around 400 data blocks were required to reach the optimum threshold, while for the case of a chi-square prior, less than 100 data blocks were necessary to reach that same threshold. Therefore, proper selection of the prior facilitates fast convergence. Note that the advantage grows in the presence of dispersion, since histograms must be computed for each one of the 2^δ branches of the trellis. This means that 2^δ times more data is needed to achieve equivalent accuracy of estimation. A good initialization is even more important in this situation.

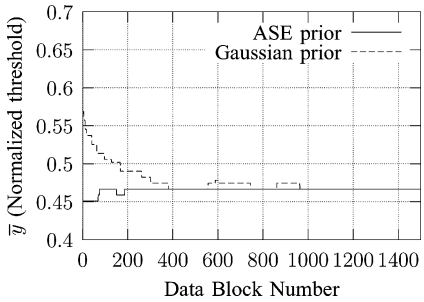

Fig. 4. Comparison between Gaussian and chi-square initialization on an ISI free channel, $L = 10^3$, $m = 100$, OSNR = 14 dB, and SGNR = 22 dB.

V. BER Computation

The probability of error of the Viterbi decoder is upper bounded by [16]

$$P_e \leq \sum_{i \neq j} W_H\left(\boldsymbol{a}^{(i)}, \boldsymbol{a}^{(j)}\right) P\left(\boldsymbol{a}^{(i)} \mid \boldsymbol{a}^{(j)}\right) P\left(\boldsymbol{a}^{(j)}\right) \quad (11)$$

where $\boldsymbol{a}^{(j)} = (a_1^{(j)}, a_2^{(j)}, \ldots, a_N^{(j)})$ represents the transmitted sequence, $\boldsymbol{a}^{(i)} = (a_1^{(i)}, a_2^{(i)}, \ldots, a_N^{(i)})$ is an erroneous sequence, N is the total number of symbols transmitted, $P(\boldsymbol{a}^{(i)} \mid \boldsymbol{a}^{(j)})$ is the probability of the error event $\boldsymbol{a}^{(j)} \rightarrow \boldsymbol{a}^{(i)}$ (the error event that occurs when the Viterbi decoder chooses sequence $\boldsymbol{a}^{(i)}$ instead of $\boldsymbol{a}^{(j)}$), and $W_H(\boldsymbol{a}^{(i)}, \boldsymbol{a}^{(j)})$ is the Hamming weight of $\boldsymbol{a}^{(i)} \wedge \boldsymbol{a}^{(j)}$ or, in other words, the number of bit errors in the error event ($\wedge$ is the exclusive OR operator). $P\left(\boldsymbol{a}^{(j)}\right)$ is the probability that the transmitter sent sequence $\boldsymbol{a}^{(j)}$. It is well known that (11) is a tight upper bound, and it can be used as an approximation to the probability of error. $P\left(\boldsymbol{a}^{(i)} \mid \boldsymbol{a}^{(j)}\right)$ can be computed based on (3) and the pdf of the Gaussian noise, but the calculation does not have a closed-form solution. Define $b(y_n, e_n) = -\ln p(y_n \mid e_n)$. Then, assuming that the erroneous sequence $\hat{\boldsymbol{a}}$ differs from the transmitted sequence $\boldsymbol{a}$ only for $n_0 \leq n \leq n_1$, the receiver will choose the erroneous path if

$$\sum_{i=n_0}^{n_1+\delta-1} b\left(y_i, f\left(a_i, a_{i-1}, \ldots, a_{i-\delta+1}\right)\right) > \sum_{i=n_0}^{n_1+\delta-1} b\left(y_i, f\left(\hat{a}_i, \hat{a}_{i-1}, \ldots, \hat{a}_{i-\delta+1}\right)\right). \quad (12)$$

Let R^N be the vector space of N-tuples of real numbers, where N is as before the total number of symbols transmitted. Given $\boldsymbol{u} = (u_1, \ldots, u_N)$, $\boldsymbol{v} = (v_1, \ldots, v_N) \in R^N$ with

$$u_n = f\left(a_n, a_{n-1}, \ldots, a_{n-\delta+1}\right) \quad (13)$$

$$v_n = f\left(\hat{a}_n, \hat{a}_{n-1}, \ldots, \hat{a}_{n-\delta+1}\right) \quad (14)$$

let $L(\boldsymbol{u}, \boldsymbol{v})$ be the locus of all points in R^N such that

$$\sum_{i=n_0}^{n_1+\delta-1} b\left(y_i, u_i\right) = \sum_{i=n_0}^{n_1+\delta-1} b\left(y_i, v_i\right). \quad (15)$$

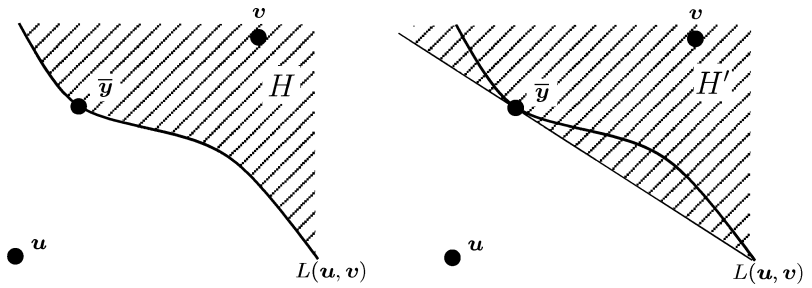

Fig. 5. Geometric interpretation of $L(\boldsymbol{u},\boldsymbol{v})$ for a 2-D case.

Notice that sequences $\boldsymbol{u}$ and $\boldsymbol{v}$ differ only between n_0 and $n_1+\delta-1$. Fig. 5 shows a geometric interpretation for a two-dimensional (2-D) case. It will be convenient in the following discussion to introduce the functions

$$F(\boldsymbol{y},\boldsymbol{u})=\sum_{i=n_0}^{n_1+\delta-1} b(y_i,u_i) \quad \text{and} \quad F(\boldsymbol{y},\boldsymbol{v})=\sum_{i=n_0}^{n_1+\delta-1} b(y_i,v_i) \tag{16}$$

where $\boldsymbol{y}=(y_1,y_2,\ldots,y_N)\in R^N$. The probability of the error event can be computed by integrating $e^{-F(\boldsymbol{y},\boldsymbol{u})}$ over the entire half-space containing $\boldsymbol{v}$. The half-space containing $\boldsymbol{v}$ will be called the *error region* for the pair $(\boldsymbol{u},\boldsymbol{v})$, and it will be denoted H (see the left side of Fig. 5). The integral is hard to compute, but an approximation can easily be found using the method of steepest descent as follows.

Define

$$G(\boldsymbol{y})=F(\boldsymbol{y},\boldsymbol{u})-F(\boldsymbol{y},\boldsymbol{v}). \tag{17}$$

Let $\overline{\boldsymbol{y}}$ be the vector in $L(\boldsymbol{u},\boldsymbol{v})$ that minimizes $F(\boldsymbol{y},\boldsymbol{u})$. Then, we can define the "distance" d between sequences $\boldsymbol{u}$ and $\boldsymbol{v}$ as

$$d=2F(\overline{\boldsymbol{y}},\boldsymbol{u}). \tag{18}$$

When the noise is Gaussian and signal independent, d is the traditional Euclidean distance between sequences $\boldsymbol{u}$ and $\boldsymbol{v}$. In the general case of non-Gaussian signal-dependent noise, d is not a distance since it does not meet the traditional conditions for the definition of distance in metric spaces. However, by analogy with the Gaussian case, we shall call it a "distance." Although in general there is no closed-form expression for d, it can be found as a solution to the constrained optimization problem consisting in minimizing $F(\boldsymbol{y},\boldsymbol{u})$ subject to the constraint $G(\boldsymbol{y})=0$. This problem can be solved numerically using commercial optimization packages. In this paper, we use the MATLAB Optimization Toolbox [20], which implements the sequential quadratic programming optimization algorithm.

To obtain the probability of the error event, we now expand $F(\boldsymbol{y},\boldsymbol{u})$ in Taylor series up to second order around the constrained minimum $\overline{\boldsymbol{y}}$. Since $F(\boldsymbol{y},\boldsymbol{u})$ is a sum of terms where each term depends on only one component of $\boldsymbol{y}$, the expansion is

$$F(\boldsymbol{y},\boldsymbol{u})\simeq F_Q(\boldsymbol{y})=F(\overline{\boldsymbol{y}},\boldsymbol{u}) + \sum_{i=n_0}^{n_1+\delta-1} p_i(y_i-\overline{y}_i)+\sum_{i=n_0}^{n_1+\delta-1} q_i(y_i-\overline{y}_i)^2 \tag{19}$$

where

$$p_i=\frac{\partial F(\boldsymbol{y},\boldsymbol{u})}{\partial y_i}\Big|_{\boldsymbol{y}=\overline{\boldsymbol{y}}} \tag{20}$$

and

$$q_i=\frac{1}{2}\frac{\partial^2 F(\boldsymbol{y},\boldsymbol{u})}{\partial y_i^2}\Big|_{\boldsymbol{y}=\overline{\boldsymbol{y}}}. \tag{21}$$

It is convenient to define the new variables

$$z_i=\sqrt{q_i}\,(y_i-\overline{y}_i) \tag{22}$$

then

$$F_Q(\boldsymbol{z})=F(\overline{\boldsymbol{y}},\boldsymbol{u})+\sum_{i=n_0}^{n_1+\delta-1}\gamma_i z_i+\sum_{i=n_0}^{n_1+\delta-1} z_i^2 \tag{23}$$

where $\gamma_i=p_i/\sqrt{q_i}$ and $\boldsymbol{z}=(z_1,\ldots,z_N)\in R^N$. Let $G(\boldsymbol{z})$ be (17) expressed as a function of $\boldsymbol{z}$, and

$$\boldsymbol{h}=\nabla G(z)|_{\boldsymbol{z}=\overline{0}} \tag{24}$$

$$\boldsymbol{k}=\frac{\boldsymbol{h}}{\|\boldsymbol{h}\|}. \tag{25}$$

It is clear that $\boldsymbol{k}$ is a unit vector normal to $L(\boldsymbol{u},\boldsymbol{v})$ at $\overline{\boldsymbol{y}}$. It is always possible to define an orthogonal transformation $\boldsymbol{V}$ such that $\boldsymbol{V}\boldsymbol{k}=\boldsymbol{e}_1=(1,0,\ldots,0)^T$. Also define

$$\boldsymbol{s}=\boldsymbol{V}\boldsymbol{z} \tag{26}$$

$$\boldsymbol{\alpha}=\boldsymbol{V}\boldsymbol{\gamma}. \tag{27}$$

Then, we can rewrite (23) in terms of variables s_i as

$$F_Q(\boldsymbol{s})=F(\overline{\boldsymbol{y}},\boldsymbol{u})+\sum_{i=n_0}^{n_1+\delta-1}\alpha_i s_i+\sum_{i=n_0}^{n_1+\delta-1} s_i^2. \tag{28}$$

By completing squares, (28) can be rewritten as

$$F_Q(\boldsymbol{s})=F(\overline{\boldsymbol{y}},\boldsymbol{u})+\sum_{i=n_0}^{n_1+\delta-1}\left(s_i+\frac{\alpha_i}{2}\right)^2-\sum_{i=n_0}^{n_1+\delta-1}\frac{\alpha_i^2}{4}. \tag{29}$$

The probability of the error event $\boldsymbol{a}\rightarrow\hat{\boldsymbol{a}}$ can be approximated as

$$P(\hat{\boldsymbol{a}}\mid\boldsymbol{a})=\int_H e^{-F(\boldsymbol{y},\boldsymbol{u})}d\boldsymbol{y}\simeq\int_{H'} e^{-F_Q(\boldsymbol{s})}d\boldsymbol{s} \tag{30}$$

where H' is the region $s_1 \geq 0$ (see the right-hand side of Fig. 5). Using (29), (30) can be expressed as

$$P(\hat{\boldsymbol{a}} \mid \boldsymbol{a}) \simeq \int_{H'} e^{-F_Q(\boldsymbol{s})} d\boldsymbol{s} = \exp\left[\frac{||\boldsymbol{\gamma}||^2}{4} - F(\overline{\boldsymbol{y}}, \boldsymbol{u})\right] Q\left(\frac{\boldsymbol{\gamma}^T \boldsymbol{k}}{\sqrt{2}}\right) \prod_{i=n_0}^{n_1+\delta-1} \left(\frac{\pi}{q_i}\right)^{1/2} \tag{31}$$

where

$$Q(x) = \frac{1}{2}\mathrm{erfc}\left(\frac{x}{\sqrt{2}}\right). \tag{32}$$

To compute an approximation to the receiver BER, (11) is used. As is common practice, the sum over error events in (11) is replaced by its largest term, whose value is approximated using (31). However, before (31) can be computed, the pairs of sequences that correspond to minimum distance error events must be identified. This can be done by exhaustive search over a set of error events of limited length.

In the remainder of this section, we consider two important special cases where the theory developed above reduces to previously known results.

A. Gaussian Case

In this section, we consider the special case of Gaussian signal-independent noise. Let σ^2 be its variance. Then

$$F(\boldsymbol{y}, \boldsymbol{u}) = \frac{1}{2\sigma^2} \sum_{i=n_0}^{n_1+\delta-1} (y_i - u_i)^2 - \frac{1}{2}(n_1 + \delta - 1 - n_0) \ln(2\pi\sigma^2). \tag{33}$$

Using (20)–(22) and (24) and (25), we get

$$\overline{\boldsymbol{y}} = \frac{1}{2}(\boldsymbol{u} + \boldsymbol{v}) \tag{34}$$

$$\boldsymbol{\gamma} = \frac{1}{\sqrt{2}\sigma}(\boldsymbol{v} - \boldsymbol{u}) \tag{35}$$

$$\boldsymbol{k} = \frac{\boldsymbol{v} - \boldsymbol{u}}{||\boldsymbol{v} - \boldsymbol{u}||} \tag{36}$$

and

$$F(\overline{\boldsymbol{y}}, \boldsymbol{u}) = \frac{||\boldsymbol{\gamma}||^2}{4} - \left(\frac{n_1 + \delta - 1 - n_0}{2}\right) \ln(2\pi\sigma^2). \tag{37}$$

Note that for Gaussian channels $F(\boldsymbol{y}, \boldsymbol{u}) = F_Q(\boldsymbol{y})$ and $H = H'$. Then, replacing in (31), we get

$$P(\hat{\boldsymbol{a}} \mid \boldsymbol{a}) = Q\left(\frac{||\boldsymbol{v} - \boldsymbol{u}||}{2\sigma}\right) \tag{38}$$

which is the usual expression for the probability of the error event $\boldsymbol{a} \to \hat{\boldsymbol{a}}$ in the Gaussian case.

B. ASE-Limited Case

Next we address the special case of ASE-limited noise. Toward this end, we use an accurate approximation of (3) given by [21, eq. (24)]

$$p_{\mathrm{ASE}}(y_n \mid e_n) \approx \frac{1}{2}\frac{1}{\sqrt{\pi N_0}}\left(\frac{e_n}{(y_n - I_{sp})^3}\right)^{1/4} \exp\left(-\frac{(\sqrt{y_n - I_{sp}} - \sqrt{e_n})^2}{N_0}\right) \tag{39}$$

where $I_{sp} = N_0 M$; then, (40), shown at the bottom of the page. Using (20), (21), and (24), and assuming high signal-to-noise ratio (SNR), we obtain

$$p_i = \frac{1 - \sqrt{u_i}\tilde{y}_i^{-1/2}}{N_0} \tag{41}$$

$$q_i = \frac{\sqrt{u_i}\tilde{y}_i^{-3/2}}{4N_0} \tag{42}$$

$$\gamma_i = \frac{2}{\sqrt{N_0}}\frac{\tilde{y}_i^{1/4}}{u_i^{1/4}}(\sqrt{\tilde{y}_i} - \sqrt{u_i}) \tag{43}$$

$$h_i = \frac{2}{\sqrt{N_0}}\frac{\tilde{y}_i^{1/4}}{u_i^{1/4}}(\sqrt{v_i} - \sqrt{u_i}) \tag{44}$$

where $\tilde{y}_i = \overline{y}_i - I_{\mathrm{sp}}$. Equations (41) through (44) can be directly applied in (31) in order to obtain the probability of the error event.

Next, we evaluate the probability of error for an ISI-free system and compare the result with that obtained in previous literature. Using (41) through (44), and taking into account that for high SNR ($|\gamma| \gg 0$) the Q function in (31) can be approximated by $Q(x) \approx (1/\sqrt{2\pi}x)e^{-x^2/2}$, we obtain from (31)

$$P(\hat{a} \mid a) \approx \frac{1}{2}\frac{\sqrt{N_0}}{\sqrt{\pi}}\frac{u^{1/4}}{\tilde{y}^{1/4}}\frac{1}{|\sqrt{u} - \sqrt{\tilde{y}}|} \exp\left[-\frac{(\sqrt{u} - \sqrt{\tilde{y}})^2}{N_0}\right]. \tag{45}$$

Let $I^{(1)}$ ($I^{(0)}$) be the mean current generated by the symbol corresponding to the bit 1 (0). For example, the probability of the error event $\mathcal{E}_{1\to 0} = \{a = 1 \to \hat{a} = 0\}$ ($\mathcal{E}_{0\to 1} = \{a = 0 \to \hat{a} = 1\}$) is given by (45) with $u = I^{(1)}$ ($u = I^{(0)}, I^{(0)} > 0$). Note that the result agrees with the one derived from [21, eq. (30)] when it is evaluated at high SNR. Similarly, it can be easily shown that in the case of $I^{(0)} = 0$ (i.e., no signal is present for the bit 0), our theory converges to [21, eq. (16)] when the latter is evaluated at high SNR.

Fig. 6 shows the probability of the error event $\mathcal{E}_{1\to 0}$ as a function of normalized (optical) signal-to-noise ratio (SNRT) defined by SNRT $= (I^{(1)}/I_{sp})M = I^{(1)}/N_0$ (see [21]). It

$$F(\boldsymbol{y}, \boldsymbol{u}) \approx \sum_{i=n_0}^{n_1+\delta-1} \left(\ln(2\sqrt{\pi N_0}) - \frac{1}{4}\ln(u_i) + \frac{3}{4}\ln(y_i - I_{\mathrm{sp}}) + \frac{(\sqrt{y_i - I_{\mathrm{sp}}} - \sqrt{u_i})^2}{N_0}\right). \tag{40}$$

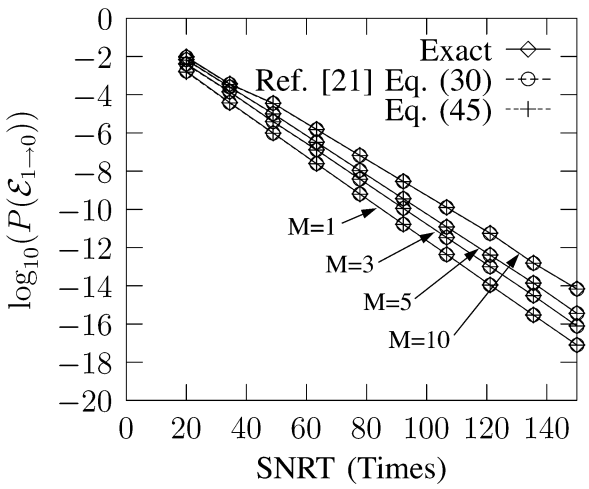

Fig. 6. Probability of the error event $P(\mathcal{E}_{1\to 0})$ as a function of SNRT.

can be seen that the exact solution (integrating the exact pdf), Marcuse's approximation ([21, eq. (30)]), and the approximation derived from the theory of this paper (45) have nearly identical behavior. It is important to mention that in this example we considered the case of high SNR only for mathematical convenience, since the theory presented in this paper is valid for low SNR as well.

VI. Mismatched Receivers

The theory developed in the previous section assumes that the receiver has perfect knowledge of the signal pdf. In many practical situations, this is not the case. In nonparametric estimation, mismatches may occur when the number of samples used in the estimation is insufficient. In parametric estimation, they may be the result of assuming a functional form for the pdf that cannot represent the channel accurately. Mismatches may also occur if the number of states of the Viterbi decoder is not sufficient to model the ISI present in the channel.

In this section, we extend the theory of Section V to mismatched receivers. Toward this end, we introduce a distinction between the conditional pdf of the signal $p(y_n \mid e_n)$ and its model assumed by the receiver, which we denote $\hat{p}(y_n \mid \hat{e}_n)$, where $\hat{e}_n = \hat{f}(a_n, a_{n-1}, \ldots, a_{n-\hat{\delta}+1})$. Note that the receiver may assume a dispersion length $\hat{\delta}$ different from the true dispersion length δ. Define

$$\hat{b}(y_n, \hat{e}_n) = -\ln \hat{p}(y_n \mid \hat{e}_n) \tag{46}$$

$$\hat{F}(\boldsymbol{y}, \hat{\boldsymbol{u}}) = \sum_{i=n_0}^{n_1+\hat{\delta}-1} \hat{b}(y_i, \hat{u}_i) \tag{47}$$

and

$$\hat{G}(\boldsymbol{y}) = \hat{F}(\boldsymbol{y}, \hat{\boldsymbol{u}}) - \hat{F}(\boldsymbol{y}, \hat{\boldsymbol{v}}) \tag{48}$$

with

$$\hat{u}_n = \hat{f}(a_n, a_{n-1}, \ldots, a_{n-\hat{\delta}+1}) \tag{49}$$

$$\hat{v}_n = \hat{f}(\hat{a}_n, \hat{a}_{n-1}, \ldots, \hat{a}_{n-\hat{\delta}+1}). \tag{50}$$

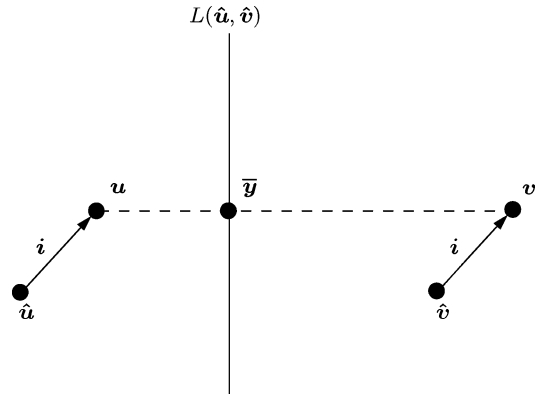

Fig. 7. Geometric interpretation for mismatched receivers.

Since the receiver makes decisions based on its own (mismatched) model of the channel, the error surface $L(\hat{\boldsymbol{u}}, \hat{\boldsymbol{v}})$ is defined by the equation

$$\hat{G}(\boldsymbol{y}) = 0 \tag{51}$$

and $\overline{\boldsymbol{y}}$ is redefined as the vector in $L(\hat{\boldsymbol{u}}, \hat{\boldsymbol{v}})$ that minimizes $F(\boldsymbol{y}, \boldsymbol{u})$. Equations (19) through (23) remain valid. Vectors $\boldsymbol{h}$ and $\boldsymbol{k}$ are replaced by

$$\hat{\boldsymbol{h}} = \nabla \hat{G}(\boldsymbol{z})|_{\boldsymbol{z}=\overline{\boldsymbol{0}}} \tag{52}$$

$$\hat{\boldsymbol{k}} = \frac{\hat{\boldsymbol{h}}}{||\hat{\boldsymbol{h}}||}. \tag{53}$$

Following steps similar to those of Section V, we finally obtain the following expression for the probability of the error event $\boldsymbol{a} \to \hat{\boldsymbol{a}}$

$$P(\hat{\boldsymbol{a}} \mid \boldsymbol{a}) \approx \exp\left[\frac{||\hat{\boldsymbol{\gamma}}||^2}{4} - F(\overline{\boldsymbol{y}}, \boldsymbol{u})\right] Q\left(\frac{\hat{\boldsymbol{\gamma}}^T \hat{\boldsymbol{k}}}{\sqrt{2}}\right) \prod_{i=n_0}^{n_1+\delta-1} \left(\frac{\pi}{q_i}\right)^{1/2}. \tag{54}$$

A. Gaussian Case

As an example of application of the above theory, consider the case of Gaussian signal-independent noise, where the receiver mismatch is caused by ISI not accounted for by the function $\hat{f}(a_n, a_{n-1}, \ldots, a_{n-\hat{\delta}+1})$ in the channel estimator. For the purpose of this example, we assume a fixed ISI pattern $\boldsymbol{i}$, whose components are given by

$$i_n = f(a_n, a_{n-1}, \ldots, a_{n-\delta+1}) - \hat{f}(a_n, a_{n-1}, \ldots, a_{n-\hat{\delta}+1}), \quad n_0 \le n \le n_1 - \delta + 1. \tag{55}$$

From Fig. 7, it is clear that

$$\overline{\boldsymbol{y}} = \frac{\hat{\boldsymbol{u}} + \hat{\boldsymbol{v}}}{2} + \boldsymbol{i} - \mu(\hat{\boldsymbol{v}} - \hat{\boldsymbol{u}}) \tag{56}$$

where

$$\mu = \frac{\boldsymbol{i} \cdot (\hat{\boldsymbol{v}} - \hat{\boldsymbol{u}})}{||\hat{\boldsymbol{v}} - \hat{\boldsymbol{u}}||^2}. \tag{57}$$

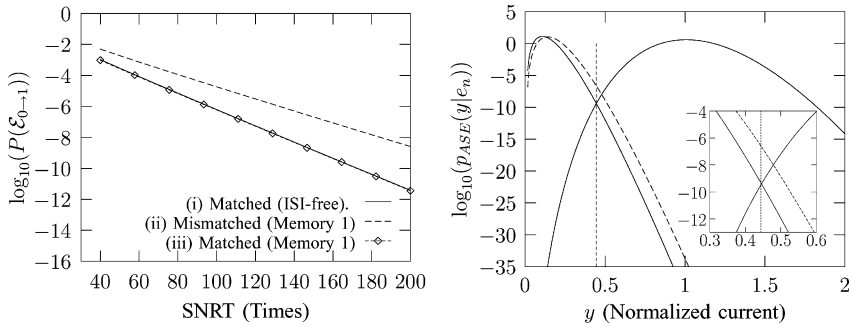

Fig. 8. Probability of the error event $\mathcal{E}_{0\to1}$ as a function of SNRT (left). Representative pdfs at SNRT $= 200$ (right).

Also from Fig. 7

$$\hat{\boldsymbol{k}} = \frac{\hat{\boldsymbol{v}} - \hat{\boldsymbol{u}}}{||\hat{\boldsymbol{v}} - \hat{\boldsymbol{u}}||}. \tag{58}$$

It is easy to verify that

$$p_i = \frac{1}{\sigma^2}\left(\frac{1}{2} - \mu\right)(\hat{v}_i - \hat{u}_i) \tag{59}$$

$$q_i = \frac{1}{2\sigma^2} \tag{60}$$

and therefore

$$\hat{\boldsymbol{\gamma}} = \frac{\sqrt{2}}{\sigma}\left(\frac{1}{2} - \mu\right)(\hat{\boldsymbol{v}} - \hat{\boldsymbol{u}}). \tag{61}$$

Also

$$F(\overline{\boldsymbol{y}}, \boldsymbol{u}) = \frac{||\hat{\boldsymbol{\gamma}}||^2}{4} - \frac{1}{2}(n_1 + \delta - 1 - n_0)\ln(2\pi\sigma^2). \tag{62}$$

Since the noise is Gaussian, note that $F(\boldsymbol{y}, \boldsymbol{u}) = F_Q(\boldsymbol{y})$ and $H = H'$. Finally, replacing all these results in (54), we get

$$P(\hat{\boldsymbol{a}} \mid \boldsymbol{a}) = Q\left(\frac{||\hat{\boldsymbol{v}} - \hat{\boldsymbol{u}}||^2 - 2\boldsymbol{i}\cdot(\hat{\boldsymbol{v}} - \hat{\boldsymbol{u}})}{2\sigma||\hat{\boldsymbol{v}} - \hat{\boldsymbol{u}}||}\right). \tag{63}$$

This is the well-known result for the probability of an error event in the presence of uncorrelated Gaussian noise and ISI (see, e.g., [22, eq. (22)]).

B. ASE-Limited Case

Based on the results of Section V-B, in the following, we analyze the case of ASE-limited noise. We consider again the case of ISI not completely accounted for by the channel estimator. Expressions (40)–(44) are still valid. The probability of error can be evaluated replacing them in (54).

As an example, we evaluate the performance of 1) a matched receiver for the case of ISI-free received signal, 2) a mismatched receiver for a channel with memory order 1, and 3) a receiver perfectly matched to the channel of Case 2).

Case 1): This system is a simple threshold detector where the threshold level is given by $\overline{y}$. Consider, for example, the calculation of the probability of the error event $\mathcal{E}_{0\to1} = \{a = 0 \to \hat{a} = 1\}$ (i.e., a transmitted bit 0 is detected as a bit 1). Assuming that $I^{(0)} > 0$, it can be obtained directly from (45)

$$P(\mathcal{E}_{0\to1}) \approx \frac{1}{2}\frac{\sqrt{N_0}}{\sqrt{\pi}}\frac{(I^{(0)})^{1/4}}{\tilde{y}^{1/4}|\sqrt{I^{(0)}} - \sqrt{\tilde{y}}|} \times \exp\left[-\frac{(\sqrt{I^{(0)}} - \sqrt{\tilde{y}})^2}{N_0}\right]. \tag{64}$$

Case 2): In this case, the receiver is also a threshold detector with the same value of $\overline{y}$ given in 1), but the channel introduces ISI. The received signal is given by

$$f(a_n, a_{n-1}) = (h_0\sqrt{I_n} + h_1\sqrt{I_{n-1}})^2 \tag{65}$$

where I_i ($I_i \in \{I^{(0)}, I^{(1)}\}$) is the mean current generated by the bit a_i. In this example, we assume that $h_0, h_1 > 0$. The value of h_0 is obtained by normalizing the channel energy (i.e., $h_0 = \sqrt{1 - h_1^2}$). In the following, we consider the particular dominant error event $\mathcal{E}_{0\to1} = \{\boldsymbol{a} = (0,1) \to \hat{\boldsymbol{a}} = (1,1)\}$ (i.e., a transmitted 0 is detected as a 1 when there is a 1 in the memory of the channel). In this case, the probability of the error event can be evaluated as

$$P(\mathcal{E}_{0\to1}) \approx \frac{1}{2}\frac{\sqrt{N_0}}{\sqrt{\pi}}\frac{(h_0\sqrt{I^{(0)}} + h_1\sqrt{I^{(1)}})^{1/2}}{\tilde{y}^{1/4}|(h_0\sqrt{I^{(0)}} + h_1\sqrt{I^{(1)}}) - \sqrt{\tilde{y}}|} \times \exp\left[-\frac{((h_0\sqrt{I^{(0)}} + h_1\sqrt{I^{(1)}}) - \sqrt{\tilde{y}})^2}{N_0}\right]. \tag{66}$$

Note that for small ISI (i.e., $h_1 \ll h_0$), the middle factor in (66) will be almost equal to the middle factor in (64). The main difference in performance between 1) and 2) is produced by the exponential function (it is easy to show that the term $(h_0\sqrt{I^{(0)}} + h_1\sqrt{I^{(1)}})$ in (66) will be larger than the term $(\sqrt{I^{(0)}})$ of (64)). This causes a degradation of the performance of the mismatched receiver.

Case 3): In this case, the receiver can be implemented by a two-state Viterbi detector (see (65)). The probability of the error event $\mathcal{E}_{0\to1}$ can be obtained using the results of Section V-B. Note that the main difference with (66) is that the matched receiver incorporates knowledge of the ISI in the channel state, therefore yielding better performance than the receiver of Case 2).

Fig. 8 (left) shows the results for the probability of the error event $\mathcal{E}_{0\to1}$ for all three cases. In this particular example, the

ISI coefficient h_1 was set to 0.05. The relatively large degradation incurred by the mismatched receiver can be understood analyzing Fig. 8 (right), where the representative pdfs of the received signal in cases 1) and 2) are depicted with solid and dashed lines, respectively (the vertical line shows the threshold used). The pdfs for the ISI-free channel are used for symbol detection in the receiver in both Case 1) and Case 2). Note that the probability of the error event $\mathcal{E}_{0\to 1}$ is given by the integral of the pdf of the symbol $I^{(0)}$ from threshold level to infinity. A significant increase in $P(\mathcal{E}_{0\to 1})$ from Case 1) to Case 2) can be observed (inset graph is a zoom of the crossing point of the pdfs). This explains the performance degradation incurred by receiver (Case 2)). Notice that Cases 1) and 2) could have been alternatively analyzed using [21, eq. (30)]. Although for clarity this is not shown, the results agree with those of Fig. 8.

VII. Simulation Results

Below we show simulation results that confirm the theoretical analysis presented in previous sections. In the following, we assume that the transmitted pulse has an unchirped envelope $\exp(-t^2/2T_0^2)$ with unit amplitude and $T_0 = 36$ ps. The extinction ratio (r_{10}), defined hereafter, was set to 10 dB. Although we consider chromatic dispersion only, the results are also valid in the presence of PMD. In order to clearly distinguish the effects on receiver performance of the chi-square-distributed ASE noise and of the Gaussian noise contributed by the receiver front end, we define two different SNRs: one that only takes into account ASE noise and another that only takes into account Gaussian thermal noise. They are defined as

$$\text{OSNR} = \frac{I^{(1)}}{2I_{\text{sp}}},\ \text{optical-signal-to-noise ratio} \tag{67}$$

$$\text{SGNR} = \frac{\zeta^2}{\sigma_z^2},\ \text{signal-to-Gaussian-noise ratio.} \tag{68}$$

Because of the nonlinear nature of the optical channel, we define ζ^2 as the signal power in the electrical domain (i.e., $\sum_{i=1}^{2^N} e_i^2/2^N$). We also define the extinction ratio as

$$r_{10} = \frac{I^{(1)}}{I^{(0)}}. \tag{69}$$

In practical implementations, the extinction ratio usually is 7 dB or greater. The amount of noise is strongly dependent on the system architecture. The OSNR depends on the transmit laser and the number of optical amplifiers used. The SGNR depends on the design of the receiver front end and the type of electrical amplifiers used. In particular, thermal noise is different for field-effect and bipolar transistors. In any case, for practical implementations, the OSNR should be greater than ~ 15 dB, and the SGNR should be greater than ~ 22 dB.

A. Performance of MLSE With an Unconstrained Complexity Viterbi Decoder

We will analyze system performance for a Viterbi decoder where exact knowledge of the channel impulse response is assumed and the number of states of the receiver is sufficient to allow an accurate representation of the nonlinear response

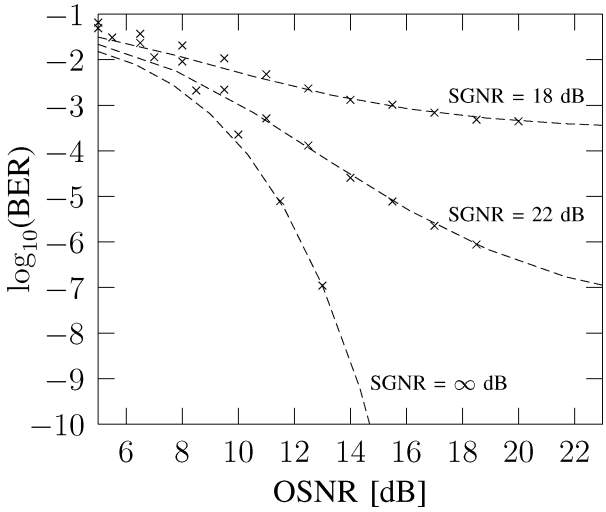

Fig. 9. BER as a function of OSNR on a dispersive (1700 ps/nm) channel for different SGNR values. Dashed curves represent the theory of this paper, and crosses are simulation results. $M = 3$, bit rate = 10 Gb/s, and $r_{10} = 10$ dB.

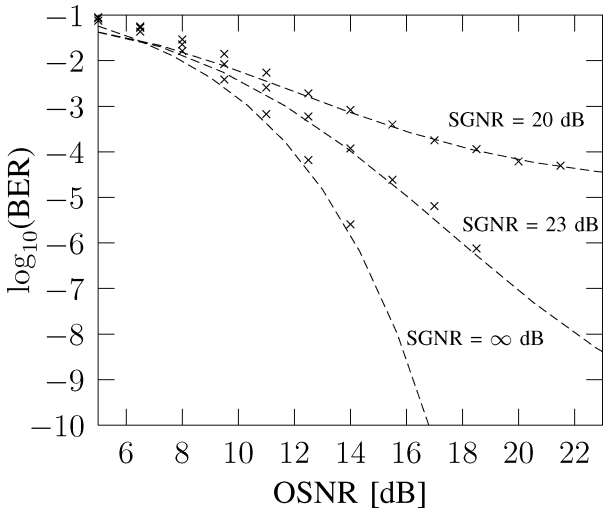

Fig. 10. BER as a function of OSNR on a dispersive (3400 ps/nm) channel for different SGNR values. Dashed curves represent the theory of this paper, and crosses are simulation results. $M = 3$, bit rate = 10 Gb/s, and $r_{10} = 10$ dB.

$f(a_n, a_{n-1}, \ldots, a_{n-\delta+1})$ (in other words, the number of states is at least $2^{\delta-1}$).

In Figs. 9–11, simulation results for the Viterbi decoder (crosses) are compared against the theory (dashed lines) for Dl products of 1700, 3400, and 5100 ps/nm, respectively, where D is the dispersion parameter and l is the fiber length. For example, they could correspond, respectively, to fiber links of 100, 200, and 300 km of standard single-mode fiber (SSMF) as specified by the International Telecommunications Union (ITU) Recommendation G.652 [23] used in the third telecommunications window (1550 nm), which leads to a dispersion parameter $D = 17$ ps/km/nm. The Viterbi decoder uses 8, 16, and 32 states, respectively, in order to model the entire channel response in all cases. The good accuracy of the analytical BER computation developed in the previous sections can be observed.

The theoretical approximation developed in this paper achieves its highest accuracy at low BERs (i.e., BER $< 10^{-3}$).

For high BERs (low SNR), the prediction of the theory may slightly depart from the simulation results. The departure is the result of two main causes. The first is the simplification of expression (11) where the sum over all error events is replaced by its largest term. In the case of high BER, the probability of some non-minimum-distance error events may be close to that of minimum-distance events, therefore the terms neglected in (11) may no longer be insignificant. Second, the quadratic approximation to the logarithm of the pdf (19) may loose some accuracy at low SNR. These factors explain the small discrepancy between fitting of theory and simulation at low SNR in Figs. 9 and 11 and 10 and 12. These inaccuracies can be corrected simply by including more error events in (11) besides those of minimum distance and increasing the order of the Taylor series expansion in (19). Nevertheless, the theory as applied in this paper proves to be extremely accurate for practical values of BER (e.g., $< 10^{-3}$), making a higher complexity unnecessary. It can be seen from the figures that the mismatch between theory and simulations in the low SNR region is less than 1 dB. Since the slope of the curves is small in this region, the discrepancy is insignificant in terms of BER.

B. Mismatched Receivers: Inaccurate Knowledge of the Signal pdf

We now illustrate the case of mismatched receivers where the mismatch is due to an inaccurate knowledge of the signal pdf. In Fig. 12, we compare three MLSE receivers using 1) Euclidean metric (EM) ("□"), 2) ASE-limited case metric (ASE-M) ("×"), and 3) the exact metric based on Marcuse's derivation ("+"). Results derived from simulation (marks) and theory (lines) are shown for SGNR $=$ 22 dB. In every case note that the values obtained from the theory accurately match simulation results. In particular, it can be seen that EM performs worse at low OSNR where ASE noise is dominant. Conversely, ASE-M performs worse in the high-OSNR region, where Gaussian noise is dominant.

C. Mismatched Receivers: Constrained Complexity Viterbi Decoder

The complexity of the Viterbi algorithm grows exponentially with the length of the channel memory. A reduction of the decoder complexity can be achieved by mismatching Viterbi memory from the channel memory (which leads to a reduction in the number of states of the receiver). In this case, the ISI coefficients not modeled by the channel estimator in the Viterbi decoder are considered as residual ISI (RISI), which modifies the statistics of the received signal. The new statistic can be computed as follows. The noise-free channel response given by (2) is truncated to the L most representative bits. Let $(\hat{a}_n, \hat{a}_{n-1}, \ldots, \hat{a}_{n-\delta+1})$ be a given group of δ consecutive transmitted bits. Assuming that the most representative bits are consecutive, the corresponding noise-free signal is

$$\hat{e}_n = f(\hat{a}_{n-L_1}, \hat{a}_{n-L_1-1}, \ldots, \hat{a}_{n-L_1-L+1}) \tag{70}$$

with L_1 integer and $L_1 + L \leqslant \delta$. The $R = \delta - L$ bits not accounted for by the channel estimator give rise to RISI. In this

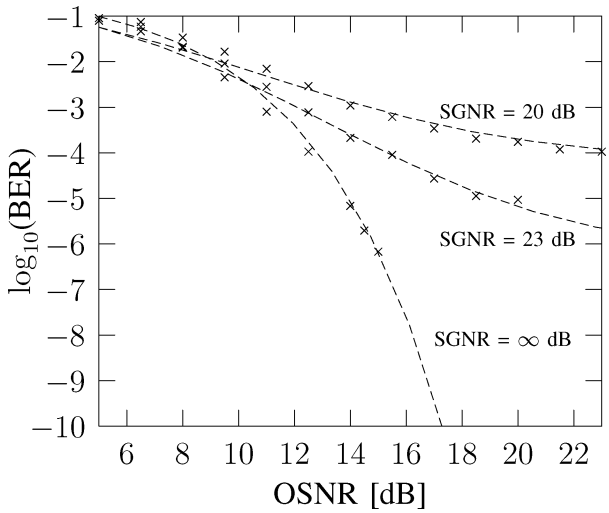

Fig. 11. BER as a function of OSNR on a dispersive (5100 ps/nm) channel for different SGNR values. Dashed curves represent the theory of this paper, and crosses are simulation results. $M = 3$, bit rate $=$ 10 Gb/s, $r_{10} = 10$ dB.

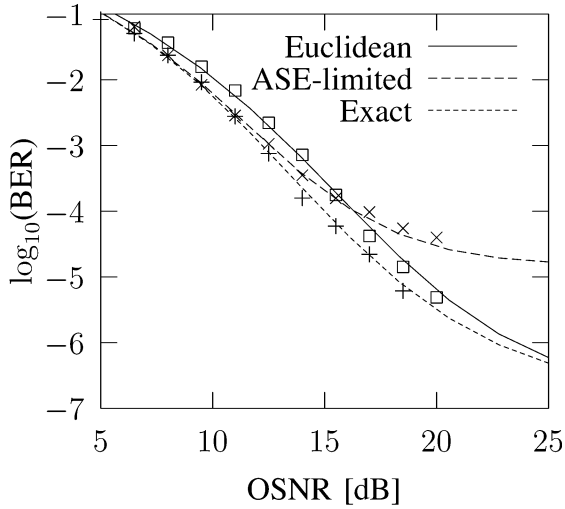

Fig. 12. BER for mismatched receivers as a function of OSNR on a dispersive (3400 ps/nm) channel for an SGNR $=$ 22 dB. Dashed curves represent the theory of this paper, and marks are simulation results. $M = 3$, bit rate $=$ 10 Gb/s, and $r_{10} = 10$ dB.

case, the conditional pdf of the received signal is evaluated averaging over the 2^R possible bit patterns, that is

$$\hat{p}(y_n \mid \hat{e}_n) = \frac{1}{2^R} \sum_{e_n \in \mathbb{E}} p(y_n \mid e_n) \tag{71}$$

where $\mathbb{E}$ denotes the set of the 2^R noise-free signals $e_n = f(a_n, a_{n-1}, \ldots, a_{n-\delta+1})$ with

$$\begin{aligned} a_{n-L_1} &= \hat{a}_{n-L_1} \\ , a_{n-L_1-1} &= \hat{a}_{n-L_1-1} \\ &\ldots \\ a_{n-L_1-L+1} &= \hat{a}_{n-L_1-L+1}. \end{aligned}$$

Fig. 13 shows the penalty incurred on a $Dl = 5100\,\mathrm{ps/km/nm}$ fiber link by 8-, 16-, and 32-state Viterbi detectors. We consider optical channel transmissions with dominant ASE noise (i.e., effects of electrical noise are neglected). For comparison, the results of the unconstrained

complexity detector can be observed in Fig. 11 (SGNR $= \infty$ dB). Equation (71) is used in (4) for metric computation. Note the good agreement of the simulation results ("$\times$") and the theory developed in this paper. In addition, from Fig. 13, it can be observed that the number of states is a critical design parameter for high-dispersion links. In particular, we see that for fiber dispersion of 5100 ps/nm (300 km of SSMF), the number of states of the Viterbi decoder needed to have a reasonable penalty (e.g., < 3 dB) is at least 16.

In the case of high RISI, function $F(\cdot,\cdot)$ is not necessarily monotonic. In this situation, the space may need to be partitioned into more than two decision regions. However, this topic will not be elaborated any further in this paper.

VIII. Implementation Considerations

Although the main purpose of this paper is to analyze the structure and performance of the maximum-likelihood sequence receiver for optical channels, it is also useful to consider some issues that so far have precluded its practical implementation. Our focus will be on receivers for data rates of 10 Gb/s or higher. The implementation that can make this technology most viable commercially is a digital monolithic integrated circuit in CMOS technology. The main challenges to the design of such a receiver are the implementation of the high-speed signal processing functions and the ADC. Typical clock frequencies allowed by current CMOS technology are in the 200-MHz range. This is far from the 10-GHz symbol rate of optical receivers of interest in this paper. Therefore, parallel processing techniques are required [24]. Application of parallel processing techniques to optical receivers to achieve a 10-Gb/s data rate with a reduced clock rate was proposed in [25]. Parallel Viterbi decoders have been described in [26] and [27], and references thereof. Application of the sliding block Viterbi algorithm [27] and other advanced digital signal processing (DSP) techniques to optical receivers have been recently discussed in [2] and [8].

In this section, we focus primarily on the ADC. An MLSE-based receiver capable of compensating the chromatic dispersion and PMD of up to 300 km of ITU G.652 single-mode fiber requires an ADC with a resolution of about 6 b and symbol rate sampling. To enable a commercially attractive solution, the ADC must be integrated with the DSP in CMOS technology. Although a 6-b, 10-GHz ADC is well beyond the state of the art in CMOS technology, the required throughput can be achieved with an interleaved array of ADCs sampling at a lower rate. The use of interleaving to increase the effective rate of ADCs has been extensively considered in the literature [28]–[30]. However, one of the difficulties encountered is that differences in gain, offset, or sampling phase among the constituent ADCs cause *fixed pattern noise* [29]. Fixed pattern noise has traditionally limited the applicability of interleaved ADCs. Calibration techniques have been developed [30], [31] to allow mismatches to be dynamically identified and compensated. These techniques are well suited to communications receiver applications, because the calibration can easily be incorporated into the adaptive loops used for timing recovery, automatic gain control (AGC), offset compensation, and equalization.

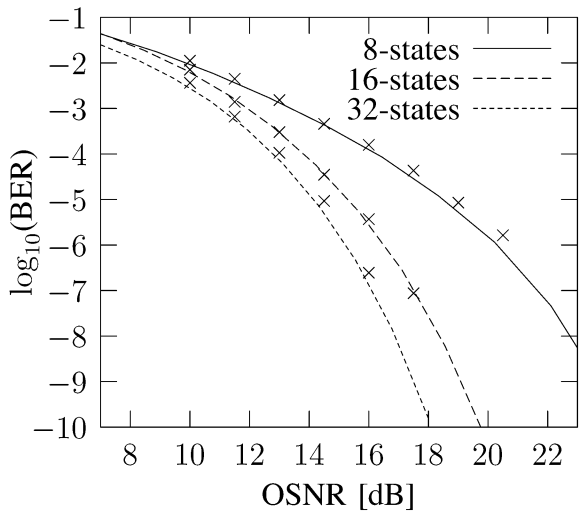

Fig. 13. Penalty incurred by constrained complexity Viterbi decoders.

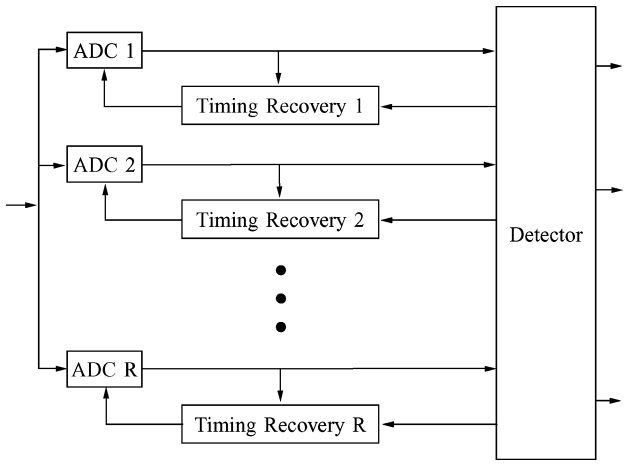

Fig. 14. ADC phase error compensation through independent timing-recovery loops.

Fig. 14 shows how phase errors among the interleaves of an ADC array can be compensated by breaking the timing recovery loop into a set of interleaved loops acting independently over each ADC [32], [33]. A similar concept can be applied to the AGC. By breaking a single AGC into interleaved loops, the gain errors of the constituent ADCs can be compensated. ADCs with a resolution of about 6 b and sampling rates of > 1 GHz have already been demonstrated in CMOS technology. An interleaved array of 8–10 of these converters could achieve an effective sampling rate of 10 GHz without loss of performance using the digital calibration techniques just described. Additional details about these techniques can be found in [32] and [33].

IX. Conclusion

In this paper, the structure and performance of MLSE-based optical receivers were analyzed. MLSE is useful to compensate dispersion in the optical fiber, which becomes a significant limitation at 10 Gb/s or higher rates. A theory of the error rate of these receivers, which are characterized by the fact that the channel is nonlinear and the noise is non-Gaussian and signal dependent, was developed. The theory has been tested by computer simulations and by analysis of limit cases where it reduces

to previously known results. The applicability of the theory developed in this paper is not limited to optical channels. It can be applied to other channels where the noise is non-Gaussian and signal dependent.

APPENDIX
HISTOGRAM UPDATE EQUATION

In this Appendix, we derive the update equation (10). We estimate the probability in each bin of the signal histogram independently. For simplicity of notation, we drop the subscript k used in Section IV-B to identify the bins. We also drop the time index n.

Let p be the probability value associated with a certain bin of the histogram. p is a random variable, and we wish to estimate its mean value and its pdf. First, we are going to assume that p is estimated in an entirely nonparametric fashion using the histogram method, but instead of doing the estimation only once based on all the data collected, we do the estimation iteratively, collecting blocks of L signal samples and using them to update previous estimates of p. This approach is necessary in a channel estimation application, since data is received continuously instead of being available in a single block when the receiver is initialized. At the beginning, since we do not have any prior knowledge of the pdf of p, we assume a uniform distribution. We conduct experiments, and we find that in L trials, the signal falls in the bin under study a total of l times. The initial prior density of p is

$$\Pi_0(p) = 1. \tag{72}$$

The joint pdf of l and p is

$$f(l,p) = f(l \mid p)\Pi_0(p) = \frac{L!}{l!(L-l)!}p^l(1-p)^{L-l} \tag{73}$$

and the posterior probability is

$$\begin{aligned}\Pi_1(p) = f(p \mid l) &= \frac{f(l,p)}{f(l)} = \frac{f(l,p)}{\int_0^1 f(l,p)dp} \\ &= \frac{(L+1)!}{l!(L-l)!}p^l(1-p)^{L-l} \\ &= \frac{\Gamma(L+2)}{\Gamma(l+1)\Gamma(L-l+1)}p^l(1-p)^{L-l}. \end{aligned}\tag{74}$$

Note that the latter expression corresponds to the Beta distribution

$$\Pi_1(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1} \tag{75}$$

with $\alpha = l+1$ and $\beta = L-l+1$.

We now iterate the procedure collecting L_1 more signal samples, of which l_1 fall in the bin under consideration. Using $\Pi_1(p)$ as the prior, we get

$$\Pi_2(p) = \frac{(L+L_1+1)!}{(l+l_1)!(L+L_1-l-l_1)!}p^{l+l_1}(1-p)^{L+L_1-l-l_1} \tag{76}$$

which corresponds to a Beta distribution function with parameters

$$\alpha = l + l_1 + 1, \tag{77}$$

$$\beta = L + L_1 - l - l_1 + 1. \tag{78}$$

In general, iterating the procedure, we get successive priors, all in the form of Beta densities, with values of α and β reflecting all samples taken in the past (in all measurements). Notice that $\Pi_0(p)$ could be regarded as a Beta density with $\alpha = \beta = 1$. In a practical application, however, we may want to limit the *memory* associated with the prior so that it does not become increasingly more difficult to adapt it in the presence of new data. This would be the case if the amount of data taken previously were very large. Therefore, for the purpose of practical adaptive noise estimation, we assume that the number $(\alpha-1)$ of previous hits in a certain bin is fixed. Then, based on the mean value of p for that bin, we compute a total equivalent number of trials in past experiments.

Let M_T be the total number of past trials and m the number of hits in those past trials falling in the bin under study. We want to fix m to a certain "reasonable" number and estimate M_T as

$$M_T = \frac{m}{E\{p\}} = \frac{m}{\overline{p}} \tag{79}$$

where $\overline{p}$ is the mean value of p based on the prior density. It can be shown that the posterior value of $\overline{p}$ is

$$E(p \mid l) = \frac{m+l}{M_T+L} \tag{80}$$

where L is the total number of new trials and l is the number of hits in the chosen bin in the L trials. Then, $\overline{p}$ can be recursively adjusted by

$$\overline{p} \leftarrow \frac{\overline{p}M_T + l}{M_T + L}.$$

For practical applications, we define

$$\lambda = \frac{L}{\hat{M}_T + L}, \quad \hat{M}_T = \frac{m}{\overline{p}}, \quad \text{and} \quad \overline{q} = \frac{l}{L} \tag{81}$$

obtaining the following recursive estimator for $\overline{p}$:

$$\overline{p} \leftarrow (1-\lambda)\overline{p} + \lambda\overline{q}. \tag{82}$$

From (81), it can be seen that for small values of $\overline{p}$ (e.g., $\overline{p} \ll 10^{-4}$) λ is essentially equal to $\lambda \approx L(\overline{p}/m)$. For large values of $\overline{p}$, $\lambda \to 1$, meaning that we give much more weight to the new evidence than to the prior knowledge. Introducing a subscript k to identify the histogram bins and a superscript n to number the successive blocks of signal samples collected, and dropping the bars, we obtain (10).

In a practical application of this method, $\overline{p}$ is not estimated solely based on measured data. Instead, it is initialized with some *a priori* pdf known to be a good approximation to the true pdf, and then it is updated iteratively using observations

and (10). In this form, the *a priori* knowledge of the pdf is combined with observations to make the best possible use of all the information available.

REFERENCES

[1] J. H. Winters and R. D. Gitlin, "Electrical signal processing techniques in long-haul fiber-optic systems," *IEEE Trans. Commun.*, vol. 38, no. 9, pp. 1439–1453, Sep. 1990.

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

[3] J. H. Winters, R. D. Gitlin, and S. Kasturia, "Reducing the effects of transmission impairments in digital fiber optic systems," *IEEE Commun. Mag.*, vol. 31, no. 6, pp. 68–76, Jun. 1993.

[4] B. L. Kasper, "Equalization of multimode optical fiber systems," *Bell Syst. Tech. J.*, vol. 61, no. 7, p. 1367, Sep. 1982.

[5] A. J. 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.

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

[7] O. E. Agazzi and V. Gopinathan. The impact of nonlinearity on electronic dispersion compensation of optical channels. presented at Optical Fiber Communication Conf. Exhibit (OFC). [CD-ROM] Optical Society of America, Washington, DC, presentation TuG6.

[8] O. E. Agazzi, D. E. Crivelli, and H. S. Carrer, "Maximum likelihood sequence estimation in the presence of chromatic and polarization mode dispersion in intensity modulation/direct detection optical channels," in *Proc. IEEE Int. Conf. Communications (ICC)*, Jun. 2004, pp. 2787–2793.

[9] R. J. Nuyts, Y. K. Park, and P. Gallion, "Dispersion equalization of a 10 Gb/s repeatered transmission system using dispersion compensation fibers," *J. Lightw. Technol.*, vol. 15, no. 1, pp. 31–42, Jan. 1997.

[10] M. Secondini, E. Forestieri, and G. Prati, "Adaptive minimum MSE controlled PLC optical equalizer for chromatic dispersion compensation," *J. Lightw. Technol.*, vol. 21, no. 10, pp. 2322–2331, Oct. 2003.

[11] E. Forestieri, G. Colavolpe, and G. Prati, "Novel MSE adaptive control of optical PMD compensators," *J. Lightw. Technol.*, vol. 20, no. 12, pp. 1997–2003, Dec. 2002.

[12] S. D. Personik, "Baseband linearity and equalization in fiber optic digital communications systems," *Bell Syst. Tech. J.*, vol. 52, no. 7, pp. 1175–1194, Sep. 1973.

[13] B. E. A. Saleh and M. I. Irshid, "Coherence and intersymbol interference in digital fiber optic communication systems," *IEEE J. Quantum Electron.*, vol. QE-18, pp. 944–951, Jun. 1982.

[14] D. Marcuse, "Calculation of bit-error probability for a lightwave system with optical amplifiers and post-detection Gaussian noise," *J. Lightw. Technol.*, vol. 9, pp. 505–513, Apr. 1991.

[15] P. A. Humblet and M. Azizoglu, "On the bit error rate of lightwave systems with optical amplifiers," *J. Lightw. Technol.*, vol. 9, pp. 1576–1582, Nov. 1991.

[16] G. D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Commun.*, vol. 18, no. 3, pp. 363–378, May 1972.

[17] R. A. Kennedy, B. D. O. Anderson, and R. R. Bitmead, "Blind adaptation of decision feedback equalizers: Gross convergence properties," *Int. J. Adapt. Control Signal Process.*, vol. 7, pp. 497–523, 1993.

[18] E. A. Lee and D. G. Messerschmitt, *Digital Communication*, 2d ed. Norwell, MA: Kluwer, 1994, p. 713.

[19] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995, p. 594.

[20] MATLAB, Software Package Version 6.1, The MathWorks, Inc., Natick, MA, 2002.

[21] 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.

[22] O. E. Agazzi and N. Seshadri, "On the use of tentative decisions to cancel intersymbol interference and nonlinear distortion (with application to magnetic recording channels)," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 394–408, Mar. 1997.

[23] *Characteristics of Single-Mode Optical Fiber and Cable*, ITU-T Recommendation G.652, 2003.

[24] K. Parhi, *VLSI Digital Signal Processing Systems*. New York: Wiley, 1999.

[25] O. E. Agazzi, V. Gopinathan, K. Parhi, K. Kota, and A. Phanse. (2000) DSP Based Equalization for Optical Channels—Feasibility of a VLSI Implementation. IEEE 802.3ae Task Force, New Orleans, LA. [Online]. Available: http://grouper.ieee.org/groups/802/3/ae/public/sep00/agazzi_1_0900.pdf

[26] H. Dawid, G. Fettweis, and H. Meyr, "A CMOS IC for Gb/s Viterbi decoding: System design and VLSI implementation," *IEEE Trans. Very Large Scale (VLSI) Syst.*, vol. 4, no. 1, pp. 17–31, Mar. 1996.

[27] P. J. Black and T. H. Y. Meng, "A 1 Gb/s, four-state, sliding block Viterbi decoder," *IEEE J. Solid-State Circuits*, vol. 32, no. 6, pp. 797–805, Jun. 1997.

[28] W. C. Black and D. A. Hodges, "Time-interleaved converter arrays," *IEEE J. Solid-State Circuits*, vol. 15, no. 6, pp. 1022–1029, Dec. 1980.

[29] A. Petraglia and S. K. Mitra, "Analysis of mismatch effects among A/D converters in a time-interleaved waveform digitizer," *IEEE Trans. Instrum. Meas.*, vol. 40, no. 5, pp. 831–835, Oct. 1991.

[30] D. Fu, K. C. Dyer, S. H. Lewis, and P. J. Hurst, "A digital background calibration technique for time-interleaved analog-to-digital converters," *IEEE J. Solid-State Circuits*, vol. 33, no. 12, pp. 1904–1911, Dec. 1998.

[31] S. M. Jamal, D. Fu, M. P. Singh, P. J. Hurst, and S. H. Lewis, "Calibration of sample-time error in a two-channel time-interleaved analog-to-digital converter," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 1, pp. 130–139, Jan. 2004.

[32] O. E. Agazzi and V. Gopinathan, "Methods and systems for DSP-based receivers," U.S. Patent US2002/0 080 898 A1 (pending), Jun. 27, 2002.

[33] ——, "Methods and systems for digitally processing optical data signals," U.S. Patent US2002/0 012 152 A1 (pending), Jan. 31, 2002.

Oscar E. Agazzi (S'79–M'82–SM'01–F'02) received the Electrical and Electronic Engineer and the Licentiate in Physics degrees from the National University of Cordoba, Cordoba, Argentina, in 1974 and 1975, respectively, and the Ph.D. degree in electronic engineering from the University of California, Berkeley, in 1982.

Between 1982 and 1987, he was with the Research Center of the Armed Forces, Buenos Aires, Argentina. In 1987, he joined AT&T Bell Laboratories, Murray Hill, NJ. In 1996, he moved to Lucent Technologies Bell Laboratories, also in Murray Hill, where he became a Bell Laboratories Fellow. Between 1997 and 2004, he was with Broadcom Corporation, Irvine, CA. He is now with ClariPhy Communications, Inc., Irvine, CA. His interests are the theory and implementation of communications systems and digital signal processing algorithms. He has published more than 40 technical papers in journals and conferences and has more than 40 patents issued or pending.

Mario R. Hueda was born in Jujuy, Argentina, in 1967. He received the Electrical and Electronic Engineer degree and the Ph.D. degree from the National University of Cordoba, Cordoba, Argentina, in 1994 and 2002, respectively.

From March 1994 to 1996, he received a fellowship from CONICOR (the Scientific and Technological Research Council of Cordoba) to carry out research and development in the area of voice-band data transmission. During the summer of 1996, he was a Visiting Scholar with Lucent Technologies—Bell Laboratories, Murray Hill, NJ, where he worked on code-division multiple-access receivers. In 1997, he joined the Digital Communications Research Laboratory at the Department of Electronic Engineering of the National University of Cordoba. He is currently also with CONICET (the National Scientific and Technological Research Council of Argentina), Buenos Aires, Argentina. His research interests include digital communications and performance analysis of communication systems.

Hugo S. Carrer was born in Cordoba, Argentina, in 1976. He received the Electronic Engineer degree from the National University of Cordoba, Cordoba, Argentina, in 2001. He is currently working toward the Ph.D. degree at the same university.

Since 2002, he has been with the Digital Communications Research Laboratory at the Department of Electronic Engineering of the National University of Cordoba. His research interests include equalization and digital signal processing in high-speed lightwave systems.

Diego E. Crivelli was born in Buenos Aires, Argentina, in 1976. He received the Electronic Engineer degree from the National University of Cordoba, Cordoba, Argentina, in 2002. He is currently working toward the Ph.D. degree at the same university.

Since 2003, he has been with the Digital Communications Research Laboratory at the Department of Electronic Engineering of the National University of Cordoba. His research interests include modulation, equalization, and coding in high-speed fiber-optic communication systems