# Channel Estimation Algorithms for MLSD in Optical Communication Systems

Tommaso Foggi, Giulio Colavolpe, Enrico Forestieri, *Member, IEEE*, and Giancarlo Prati, *Fellow, IEEE*

*Abstract*—**Maximum likelihood sequence detection represents the most efficient technique in the electrical domain to combat fiber impairments such as polarization-mode dispersion and group-velocity dispersion. In order to successfully apply this technique, it is mandatory to estimate some key channel parameters needed by the Viterbi processor. We propose a simple and effective solution based on the least-mean-square algorithm to perform such an estimation.**

*Index Terms*—**Maximum likelihood sequence detection (MLSD), parameter estimation, Viterbi algorithm (VA).**

## I. INTRODUCTION

THE EVER increasing trasmission rate in optical systems determines a more severe impact of polarization-mode dispersion (PMD) and group-velocity dispersion (GVD) on system performance. Optical compensation techniques have been proposed in the past but, although they are very effective, the need for advanced and high-cost technologies limits their development and use. Hence, a great interest in electrical equalization schemes, and in particular in schemes based on maximum likelihood sequence detection[1] (MLSD) [1], has arisen because of the possibility of exploiting simple, low-cost, and well-known solutions [2]–[9]. In order to implement an MLSD receiver, the knowledge of the statistics of the received samples is necessary. Today, optical systems often envisage the presence of optical amplifiers, and, as a consequence, the signal in the fiber is impaired by a noise that, in the linear regime, can be modeled as additive white Gaussian noise (AWGN). Since a square-law detector is used at the receive end, postdetection noise statistics change [1], [10], [11], and can no longer be considered Gaussian. A suitable closed-form expression of the branch metrics of the Viterbi algorithm (VA) implementing the MLSD strategy is provided in [7] and [8]. However, some parameters of the distribution of the received samples depend not only on the transmitted pulse shape, filter types, and bandwidths, but also on the channel impairments and on the noise power spectral density (PSD) [8]. Due to the square-law nature of a photodiode, the estimation of the required channel

[1]We prefer the term *detection* instead of the commonly used *estimation* since the estimation theory refers to continuous parameters whereas here we are considering discrete sequences.
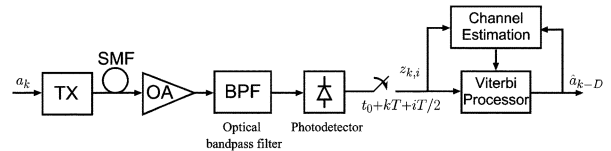


Fig. 1.  System model.

parameters is not straightforward. We propose an effective and simple algorithm which, compared to the ideal case of perfect channel knowledge, leads to a negligible performance loss.

## II. SYSTEM MODEL AND MLSD RECEIVER

The considered system model is shown in Fig. 1. A standard nonreturn-to-zero (NRZ) ON–OFF keying signal is launched into a single-mode fiber (SMF) introducing PMD and GVD, and is then optically amplified and filtered at the receive end. It is assumed that the amplified spontaneous emission (ASE) noise is dominant over thermal and shot noise. The ASE noise is modeled as an AWGN whose complex envelope is $\mathbf{w}(t) = [w_1(t), w_2(t)]^T$, where $w_1(t)$, $w_2(t)$ are independent complex noise components accounting for ASE on two orthogonal states of polarization (SOPs), each with two-sided PSD equal to $N_0$. At the optical filter output, the components of the two-dimensional complex vectors $\mathbf{s}(t) = [s_1(t), s_2(t)]^T$ and $\mathbf{n}(t) = [n_1(t), n_2(t)]^T$ represent the complex envelope of the useful signal and noise components in each SOP, respectively. The noise components are Gaussian but not white, since they are obtained by filtering the AWGN $\mathbf{w}(t)$.

After photodetection, the signal samples are processed through the VA, effectively implementing the MLSD strategy. To illustrate the proposed technique for channel and noise PSD estimation, we consider the MLSD receiver working with two samples per bit time as described in [8], since it practically represents the optimal electronic processing.[2] In this case, a postdetection electrical filter is not necessary [8]. Hence, the photodetected signal can be expressed as

$$z(t) = |s_1(t)+n_1(t)|^2 + |s_2(t)+n_2(t)|^2 = s_z(t)+n_z(t) \quad (1)$$

having defined the useful signal $s_z(t)$ and noise term $n_z(t)$ as

$$s_z(t) \triangleq |s_1(t)|^2 + |s_2(t)|^2 \quad (2)$$

$$n_z(t) \triangleq |n_1(t)|^2 + |n_2(t)|^2$$
$$+ 2\Re[s_1(t)n_1^*(t) + s_2(t)n_2^*(t)] \quad (3)$$

[2]In any case, the proposed techniques can be straightforwardly extended to the MLSD receiver working with one sample per bit time also described in [8].

where $\Re(\cdot)$ returns the real part of its argument. The signal-dependent and colored noise $n_z(t)$ is neither Gaussian nor zero-mean. Indeed, it can be shown that its mean value is $\eta \triangleq E\{n_z(t)\} = 2N_0 B_o$, where $B_o$ is the optical filter noise equivalent bandwidth. Denoting by $t_0$ a proper time offset and by $T$ the bit time, the couple of samples used by the receiver in the $k$th bit interval can be expressed as

$$\left\{ z_{k,i} \triangleq z\left(t_0 + kT + \frac{iT}{2}\right), \; i = 0,1 \right\}. \tag{4}$$

In order to compute all the VA branch metrics at discrete time $k$, the receiver uses these two samples but also needs to estimate both the noise PSD $N_0$ (or $\eta$, equivalently) and all possible values of the useful components $s_z(t_0 + kT + iT/2)$, $i = 0, 1$, of the received samples [8]. Denoting by $L$ the channel dispersion length, these useful components are functions of the current bit $a_k$ and of the $L$ previous bits $\mu_k \triangleq (a_{k-1}, a_{k-2}, \ldots, a_{k-L})$ defining the trellis state [8]. Assuming for the moment that the channel is time-invariant,[3] these functions will be denoted to as $f_i(a_k, \mu_k)$. In order to compute the VA branch metrics, the receiver needs to estimate $N_0$ (or $\eta$) and the two values $f_i(a_k, \mu_k)$, $i = 0, 1$, for each branch defined by the couple $(a_k, \mu_k)$.

## III. ESTIMATION OF THE CHANNEL PARAMETERS

We tried to estimate the optical intersymbol interference channel coefficients whose nonlinear transformation due to the square-law detection gives the values $f_i(a_k, \mu_k)$, but we did not obtain satisfactory results. Hence, we decided to directly estimate the $2 \times 2^{L+1}$ useful signal components $f_i(a_k, \mu_k)$ to be employed in the expression of the branch metrics.

We also investigated different strategies to perform the channel noise PSD estimation. Among others, we considered a solution derived from the generalized likelihood strategy [12], consisting of the maximization of the conditional joint probability density functions of the received samples given the transmitted bits and the channel parameters. However, we discovered that, due to the presence of different local maxima, the convergence of the resulting algorithm is very difficult. For this reason, we resorted to the classical minimization of the mean square error (MSE) between the received samples and the useful signal components, modified in order to take into account the nonzero-mean value of the noise component (3). Hence, the updating law for the useful signal components is based on the least-mean-square (LMS) algorithm which minimizes the mean square value

$$\sum_{i=0}^{1} E\{[z_{k,i} - \hat{\eta} - \hat{f}_i(a_k, \mu_k)]^2\} \tag{5}$$

where $\hat{\eta}$ and $\hat{f}_i(a_k, \mu_k)$ are the estimates of $\eta$ and $f_i(a_k, \mu_k)$, respectively. Notice that for the application of the LMS method, it

[3]Tracking of a time-varying channel will be discussed in Section III.

is irrelevant that the noise $\eta_z(t)$ is not Gaussian, signal-dependent, and colored. On the other hand, the presence of the mean value $\eta$ needs to be properly taken into account. For the moment, we assume that the transmitted bits are perfectly known to the receiver. The sequence of transmitted bits will be denoted to as $\{\overline{a}_k\}$ and the corresponding sequence of trellis states as $\{\overline{\mu}_k\}$. Denoting by $\hat{\eta}^{(k)}$ and $\hat{f}_i^{(k)}(a_k, \mu_k)$ the values assumed by the estimates of $\eta$ and of the useful signal components at discrete time $k$, following the LMS algorithm [1] the channel coefficients are recursively updated, as shown in (6) at the bottom of the page, where $\alpha > 0$ is a scale factor controlling the amount of adjustment (the so-called "step-size"). Clearly, at every discrete time $k$, we can update only the estimates $\hat{f}_i(\overline{a}_k, \overline{\mu}_k)$ for the actual couple $(\overline{a}_k, \overline{\mu}_k)$, that is those matching the transmitted sequence. Hence, in practice, the update of a given coefficient is not performed at a fixed rate but at a random rate, depending on the transmitted pattern. The estimates in (6) are initialized with the value in the absence of PMD and GVD.

The update of $\hat{\eta}$ cannot be done using the same method. In fact, if we consider the MSE (5), it has different minima. For example, a minimum can be obtain when $\hat{\eta} = \eta$ and $\hat{f}_i(a_k, \mu_k) = f_i(a_k, \mu_k)$, but also when $\hat{\eta} = 0$ and $\hat{f}_i(a_k, \mu_k) = f_i(a_k, \mu_k) + \eta$. Hence, we must resort to a different approach. From (1), (2), and (3), it can be observed that, neglecting the terms depending on $\eta^2$, it is

$$E\left\{[z_{k,i} - f_i(\overline{a}_k, \overline{\mu}_k) - \eta]^2\right\} \simeq f_i(\overline{a}_k, \overline{\mu}_k)\eta. \tag{7}$$

We use this property to refine the estimates of $\eta$ by substituting the expectation with a temporal mean. The update of $\eta$ is performed every $N$ bits (that is when $k$ is a multiple of $N$), where $N$ is a parameter that can be tuned to reduce fast estimate fluctuations ($N \simeq 100$ is sufficient). Furthermore, the temporal mean is computed only when $\overline{a}_k = 1$, to avoid numerical problems with quantities close to zero. Hence, at time $k$, with $k$ a multiple of $N$, we estimate

$$\hat{\eta}^{(k+1)} = \frac{1}{2|\mathcal{N}|}$$
$$\times \sum_{\substack{i=0,1 \\ n \in \mathcal{N}}} \frac{\left[z_{k-n,i} - \hat{\eta}^{(k-n)} - \hat{f}_i^{(k-n)}(\overline{a}_{k-n}, \overline{\mu}_{k-n})\right]^2}{\hat{f}_i^{(k-n)}(\overline{a}_k, \overline{\mu}_k)} \tag{8}$$

where $\mathcal{N} = \{n = 0, 1, \ldots, N - 1 | \overline{a}_{k-n} = 1\}$ and $|\mathcal{N}|$ denotes the cardinality of this set. Note that in (8) the estimates $\hat{\eta}^{(k-n)}$, $n \in \mathcal{N}$, take the same value since the estimate of $\eta$ is updated every $N$ bits. From the estimate of $\eta$, we can then simply derive the estimate of $N_0$ to be used in the branch metrics.

Up to now, we have considered the case of a time-invariant channel. However, when the channel is (slowly) time-varying, the algorithm will be able to track the channel variations, due to its adaptive nature.

$$\begin{cases} \hat{f}_i^{(k+1)}(a_k, \mu_k) = \hat{f}_i^{(k)}(a_k, \mu_k), & \text{for } (a_k, \mu_k) \neq (\overline{a}_k, \overline{\mu}_k) \\ \hat{f}_i^{(k+1)}(a_k, \mu_k) = \hat{f}_i^{(k)}(a_k, \mu_k) + \alpha \left[z_{k,i} - \hat{\eta}^{(k)} - \hat{f}_i^{(k)}(a_k, \mu_k)\right], & \text{for } (a_k, \mu_k) = (\overline{a}_k, \overline{\mu}_k) \end{cases} \tag{6}$$
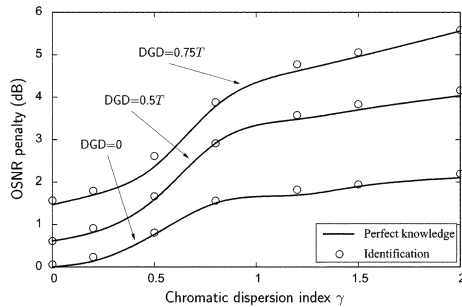
Fig. 2. OSNR penalty curves at BER $= 10^{-4}$.



Fig. 3. MSE transient when the estimation algorithm is turned on.

In the operating conditions, the transmitted bits are not available. Hence, we resort to a decision directed approach, that is we use the decisions of the VA entailing no performance loss at least when the bit-error rate (BER) is lower than $10^{-2}$. The decisions of the VA are available with a delay of $D$ symbols which can compromise the capability of the algorithm in tracking the fast channel variations. However, this is not the case of the considered channel since the dispersion we are facing is a slowly varying phenomenon. For the same reason, it is not necessary to update the channel parameters at each bit interval but they can be updated less frequently, depending on the considered channel.

Considering now the initial acquisition of the algorithm, although a training sequence can be adopted to reduce the acquisition time, we verified by computer simulations that the algorithm is able to work in a completely blind manner, as discussed in Section IV.

## IV. NUMERICAL RESULTS

We consider an optical channel affected by first-order PMD (with a worst-case power splitting equal to 0.5), and GVD, quantified through the chromatic dispersion index $\gamma = \lambda^2 R^2 D_r / (\pi c)$, where $\lambda$ is the wavelength, $R = 1/T$ is the bit rate, $D_r$ is the residual chromatic dispersion, and $c$ is the light speed [8]. Hence, the same value of $\gamma$ corresponds to different residual dispersions depending on the considered bit rate (e.g., given $\gamma = 0.1$ the residual dispersion is equal to 392 ps/nm at $R = 10$ Gb/s and to 24 ps/nm at $R = 40$ Gb/s). The optical filter is modeled as a fourth-order Butterworth type with bandpass bandwidth equal to $1.1/T$ and the MLSD receiver is based on a VA working on, at most, a 128-state trellis [8]. In Fig. 2, we show the optical signal-to-noise ratio (OSNR) penalty, obtained by computer simulations, for increasing values of $\gamma$, and for different values of the PMD differential group delay. The solid lines refer to the ideal receiver and are obtained under the hypothesis of perfect knowledge of the channel parameters, whereas the circles represent the performance of the receiver when using the discribed estimation algorithm. It can be seen that no loss in performance is entailed.

We now show that the algorithm is able to work in the decision directed mode in a completely blind manner. We consider a particular case of bad channel impairment, where the chromatic dispersion index $\gamma$ is equal to one. At the beginning and for the first 10 000 bits, the algorithm is turned OFF. In this case, for an OSNR (referred to a bandwidth equal to the bit rate) of 13 dB we have a BER of 0.25. Hence, we do not have reliable decisions. Nonetheless, when the estimation algorithm is turned
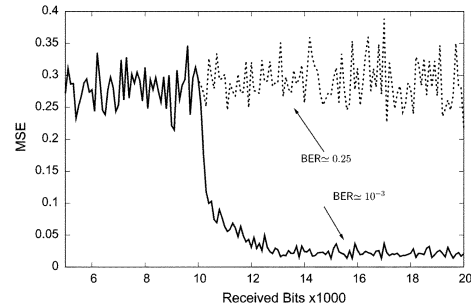
ON, it is still able to converge to a BER of $10^{-3}$ (the value corresponding to perfect knowledge of the channel parameters) in about $2000 \div 2500$ bits, as can be seen from Fig. 3 showing the time evolution of the MSE.

## V. CONCLUSION

In the literature, the benefits of the application of MLSD to optical communication systems have been widely demonstrated. In this letter, we have shown that it is possible to adaptively estimate the channel parameters, i.e., the noise-free sampled signal and the noise PSD, to be used in the computation of the branch metrics of the VA implementing the MLSD strategy. The performance of the proposed algorithm, in terms of convergence time and estimation accuracy, has been investigated. It has also been demonstrated that the performance of the proposed adaptive receiver is the same as that of the ideal receiver with perfect channel knowledge.

## REFERENCES

[1] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1996.
[2] 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.
[3] J. H. Winters and S. Kasturia, "Adaptive nonlinear cancellation for high-speed fiber-optic systems," *J. Lightw. Technol.*, vol. 10, no. 7, pp. 971–977, Jul. 1992.
[4] H. Bulow and G. Thielecke, "Electronic PMD mitigation—From linear equalization to maximum-likelihood detection," in *Proc. OFC'01*, 2001, vol. 3, pp. WDD34-1–WDD34-3.
[5] 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.
[6] H. F. Haunstein, W. Sauer-Greff, A. Dittrich, K. Sticht, and R. Urbansky, "Principles for electronic equalization of polarization-mode dispersion," *J. Lightw. Technol.*, vol. 22, no. 4, pp. 1169–1182, Apr. 2004.
[7] O. E. Agazzi, M. R. Hueda, H. S. Carrer, and D. E. Crivelli, "Maximum-likelihood sequence estimation in dispersive optical channels," *J. Lightw. Technol.*, vol. 23, no. 2, pp. 749–763, Feb. 2005.
[8] T. Foggi, E. Forestieri, G. Colavolpe, and G. Prati, "Maximum likelihood sequence detection with closed-form metrics in OOK optical systems impaired by GVD e PMD," *J. Lightw. Technol.*, vol. 24, no. 8, pp. 3073–3087, Aug. 2006.
[9] G. Bosco and P. Poggiolini, "Long-distance effectiveness of MLSE IMDD receivers," *IEEE Photon. Technol. Lett.*, vol. 18, no. 9, pp. 1037–1039, May 1, 2006.
[10] D. Marcuse, "Derivation of analytical expression for the bit-error probability in lightwave systems with optical amplifiers," *J. Lightw. Technol.*, vol. 8, no. 12, pp. 1816–1823, Dec. 1990.
[11] P. A. Humblet and M. Azizoğlu, "On the bit error rate of lightwave systems with optical amplifiers," *J. Lightw. Technol.*, vol. 9, no. 11, pp. 1576–1582, Nov. 1991.
[12] H. L. Van Trees, *Detection, Estimation, and Modulation Theory—Part I.* New York: Wiley, 1968.