Data-Aided Timing Recovery for Recording Channels With Data-Dependent Noise

J. Riani¹, J. W. M. Bergmans¹, S. van Beneden¹, and A. Immink²

¹Eindhoven University of Technology, 5600 MB, Eindhoven, The Netherlands

²Philips Research Laboratories, 5656 AA, Eindhoven, The Netherlands

In high-density data storage systems, noise becomes highly correlated and data dependent as a result of media noise, channel nonlinearities, and front-end filters. In such environments, conventional timing recovery schemes will exhibit large residual timing jitter and, especially, data-dependent timing jitter. This paper presents a new data-aided timing recovery algorithm for data storage systems with data-dependent noise. We derive a maximum-likelihood timing recovery scheme based on a data-dependent Gauss–Markov model of the noise. The timing recovery algorithm incorporates data-dependent noise prediction parameters in the form of linear prediction filters and prediction error variances. Moreover, because noise can be nonstationary in practice, we propose an adaptive algorithm to estimate and track the noise prediction parameters. Simulation results, for an idealized optical storage channel incorporating a simple model of media noise, illustrate the merits of our algorithm.

Index Terms—Data-dependent noise, digital recording, partial-response techniques, timing-recovery.

I. INTRODUCTION

TIMING recovery is one of the critical functions for reliable data detection in digital recording systems. The key problem in timing recovery is the determination of time instants at which the replay signal should be sampled for reliable data recovery. This problem has been a subject of investigation for many decades. Among the existing solutions [1], data-aided (DA) timing recovery schemes, e.g., [2]–[5], are known to be more powerful. DA schemes use the transmitted data sequence as side information to facilitate timing recovery. This information is available to the receiver either in the form of a known preamble pattern preceding the user data, or as decisions taken from the bit detector.

Timing recovery becomes more critical as storage density increases because of increasing system performance sensitivity to timing jitter on the one hand and increasing bandwidth limitations, signal-to-noise ratio (SNR) degradation, noise nonstationarity, and data dependency on the other hand. Although the problem of data detection in such noise environments has received considerable attention, e.g., see [6], [7], much less attention has been devoted to the problem of timing recovery. Conventional timing recovery schemes assume that the noise at their input is stationary and that noise statistics are independent of the transmitted data. However, in high-density recording systems, noise becomes colored and data dependent [8], [9]. This data-dependent nature of the noise significantly deteriorates the performance of timing recovery. It increases timing jitter, i.e., the difference between the ideal and the estimated sampling instants, for a given bandwidth of the timing recovery loop. Large timing jitter leads to an increased bit-error rate and possibly even to loss of lock.

A simple form of timing recovery for data-dependent noise was reported in literature for optical communication channels where noise was modeled as additive white and Gaussian

(AWG) with a noise variance dependent on the transmitted symbol [10]. This algorithm is not based on an optimal timing function but is derived as a modification of the well-known Mueller and Müller algorithm [2].

In this paper, we derive an optimal timing recovery algorithm for data-dependent correlated noise. The key to our timing recovery approach is the modeling of noise as a data-dependent finite-order Markov process [8]. Based on this model, maximum-likelihood (ML) timing recovery is addressed. The resulting structure is a timing recovery scheme with a new timing error detector (TED) that incorporates, on the one hand, data-dependent noise prediction and on the other hand a data-dependent weighing. Moreover, because in practice noise can be nonstationary, an adaptation algorithm that estimates and tracks noise model parameters is proposed. This estimation algorithm is simpler than that presented in [8].

Although this paper assumes that the transmitted data is known to the timing recovery scheme, its results can be easily extended to the case where soft information is available [11] and in the context of iterative timing recovery [12], [13] where an iterative soft decoder is used. In fact, this would boil down to simply substituting the TEDs in [11] and [12] with the one presented in this paper.

The remainder of this paper is organized as follows. Section II describes the system model and nomenclature. Section III presents the ML timing recovery for data-dependent noise. Efficiency analysis of the ML timing recovery is addressed in Section IV. Section V presents a simple sample by sample based adaptation of the data-dependent noise parameters. Simulation results for a partial response maximum-likelihood (PRML) system are presented in Section VI and show the important merits of our scheme.

II. SYSTEM MODEL AND PROBLEM DEFINITION

In Fig. 1, a zero-mean data sequence $a_k \in \{\pm 1\}$ of length N , i.e., $a_1, a_2, \dots, a_N$, of data rate $1/T$ is applied to a channel with symbol response $h(t)$, additive noise $u(t)$, and an *a priori*

Fig. 1. System model.

unknown and possibly time varying delay ϕ (in bit intervals T). Prior to detection, the receiver performs prefiltering that serves to suppress noise and may also condition intersymbol interference (ISI). The prefilter output is first sampled and then passed to a detector that produces bit decisions. For clarity of this paper, we assume that excess bandwidth at the prefilter output is negligible and consider only baud-rate sampling. The results of this paper can be easily extended to the oversampled case. The sampling instants are expressed as $t_k = (k + \psi)T$, where ψ is a sampling phase (normalized in units T). Based on the sampled sequence x_k , the receiver produces bit decisions $\hat{a}_k$ as well as a clock signal that indicates the sampling instants t_k . In order for the detector to operate properly, a timing recovery subsystem ensures that the sampling phase ψ closely approaches ϕ . The timing recovery subsystem takes the form of a phase-locked loop (PLL) with a timing-error detector (TED), loop filter (LF), and a voltage controlled oscillator (VCO). The TED produces an estimate χ_k of the sampling-phase error $\Delta = \phi - \psi$. In this paper we restrict attention to data-aided (DA) TEDs, where a_k is assumed to be available to the receiver in the form of a known preamble, or as decisions, taken from the detector, when bit-error rates are small. PLL behavior depends also on the LF and VCO. A detailed description of this dependence can be found in [14].

To simplify the forthcoming analysis we assume, first, that the loop has a sufficiently high bandwidth to enable the variations of ϕ to be tracked. This means that we can take ϕ to be fixed. Second, the sampling-phase errors Δ are restricted to a fraction of a symbol interval T (this reflects the situation when the PLL is in lock; PLL acquisition properties are beyond the scope of this paper). In this case, the equivalent discrete impulse response q_k^Δ of the system up until the detector input can be linearized as $q_k^\Delta \simeq q_k^0 + \Delta q_k'$, where q_k' is the derivative of q_k^Δ with respect to Δ at $\Delta = 0$. Both responses q_k^0 and q_k' are assumed to be known to the receiver. The detector input sequence can be written as

$$x_k \simeq (q^0 * a)_k + \Delta(q' * a)_k + n_k \quad (1)$$

where $*$ denotes linear convolution and n_k is the equivalent noise sequence at the detector input, i.e., $n_k = x_k - (q^\Delta * a)_k$. Unless specified otherwise, we assume that q_k^0 corresponds to the ideal ISI structure assumed by the detector. Any misalignment ISI (linear or nonlinear) at ideal sampling phase, i.e., due to a mismatch between q_k^0 and the ideal detector response, is embedded in the noise n_k . The noise n_k includes also channel noise that may be linearly or nonlinearly data dependent. The key to our timing recovery approach is the modeling of the noise as proposed in [8]. We recapitulate the assumptions on the properties of the noise n_k as follows.

A. Finite Correlation Length

The noise n_k is assumed to be independent of past samples before some length $L \geq 0$ (finite Markov memory length). This independence implies that

$$p(n_k | n_{k-1}, \dots, n_1, \underline{a}_1^N) = p(n_k | n_{k-1}, \dots, n_{k-L}, \underline{a}_1^N) \quad (2)$$

where $p(\cdot)$ denotes the probability density function (pdf) of n_k conditioned on the past noise samples and on the data $\underline{a}_1^N$, where $\underline{a}_{k_2}^{k_1} = [a_{k_1}, a_{k_1+1}, \dots, a_{k_2}]$ for $k_2 \geq k_1$. The conditioning on $\underline{a}_1^N$ is meant to take into account the data-dependent correlation of the noise n_k .

B. Finite Data-Dependent Span

The noise n_k depends only on its first K -neighbor symbols, i.e., $\underline{a}_{k-K_1}^{k+K_2}$, that we call symbol cluster, where $K = K_1 + K_2 + 1$. The conditioned noise pdf given in (2) becomes

$$p(n_k | n_{k-1}, \dots, n_{k-L}, \underline{a}_1^N) = p(n_k | n_{k-1}, \dots, n_{k-L}, \underline{a}_{k-L-K_1}^{k+K_2}). \quad (3)$$

C. Joint Gaussian Pdf's

The joint pdf $p(n_k, n_{k-1}, \dots, n_{k-L} | \underline{a}_{k-L-K_1}^{k+K_2})$, conditioned on the data sequence, is Gaussian with a covariance matrix $\mathbf{C}_k = \mathbf{C}(\underline{a}_{k-L-K_1}^{k+K_2})$ of size $(L+1) \times (L+1)$, i.e.,

$$p(n_k, \dots, n_{k-L} | \underline{a}_{k-L-K_1}^{k+K_2}) = \frac{\exp[-\underline{N}_k^T \mathbf{C}_k^{-1} \underline{N}_k]}{\sqrt{(2\pi)^{L+1} \det \mathbf{C}_k}} \quad (4)$$

where $[\cdot]^T$ denotes the transpose operation and the $(L+1) \times 1$ vector $\underline{N}_k = [n_k, \dots, n_{k-L}]^T$.

It is implicitly assumed here that, given the data sequence, the noise n_k has zero mean. This assumption is not entirely true in general, e.g., in the presence of channel nonlinearities [15]. In such case the vector $\underline{N}_k$ throughout the paper has to be replaced with $\underline{N}_k - E[\underline{N}_k | \underline{a}_{k-L-K_1}^{k+K_2}]$. For clarity, we omit the mean of n_k in the sequel.

III. MAXIMUM-LIKELIHOOD TIMING-ERROR DETECTOR

Data-aided ML timing recovery is optimum when no prior statistical knowledge about the phase-error Δ is available. Before developing the DA ML-TED for sample-by-sample timing recovery, let us first derive the one-shot ML estimator of the phase-error Δ based on the observation of the overall detector input sequence $x_1, \dots, x_N$. To this aim, we assume in this section that noise statistics are known and fixed during the transmission of the N symbols $\underline{a}_1^N$. The DA ML estimate of the phase-error Δ is obtained by maximizing the likelihood function, i.e.,

$$\Delta^{\text{ML}} = \arg \max_{\delta} p(x_1, \dots, x_N | \underline{a}_1^N, \Delta = \delta) \quad (5)$$

over all possible phase-errors δ , where the likelihood function $p(x_1, \dots, x_N | \underline{a}_1^N, \delta)$ is the joint probability density function of the received samples $x_1, \dots, x_N$ conditioned on the transmitted symbols $\underline{a}_1^N$ and on the phase-error $\Delta = \delta$.

In order to derive a practical criterion from (5), a few classical steps are needed. We first apply Bayes rule and obtain

$$p(x_1, \dots, x_N | \underline{a}_1^N, \delta) = \prod_{k=1}^N p(x_k | x_{k-1}, \dots, x_1, \underline{a}_1^N, \delta). \quad (6)$$

Upon invoking (1), (2), and (3) and applying Bayes rule once again, (6) can then be factorized into

$$p(x_1, \dots, x_N | \underline{a}_1^N, \delta) = \prod_{k=1}^N \frac{p(x_k, x_{k-1}, \dots, x_{k-L} | \underline{a}_{k-L-K_1}^{k+K_2}, \delta)}{p(x_{k-1}, \dots, x_{k-L} | \underline{a}_{k-L-K_1}^{k+K_2}, \delta)}. \quad (7)$$

The right-hand factors in (7) can be rewritten using (4) as

$$\frac{p(x_k, \dots, x_{k-L} | \underline{a}_{k-L-K_1}^{k+K_2}, \delta)}{p(x_{k-1}, \dots, x_{k-L} | \underline{a}_{k-L-K_1}^{k+K_2}, \delta)} \propto \frac{\exp[-(\underline{E}_k - \delta \underline{S}_k)^T \mathbf{C}_k^{-1} (\underline{E}_k - \delta \underline{S}_k)]}{\exp[-(\underline{e}_k - \delta \underline{s}_k)^T \mathbf{c}_k^{-1} (\underline{e}_k - \delta \underline{s}_k)]} \quad (8)$$

where the $L \times L$ matrix $\mathbf{c}_k$ is the lower principal submatrix of $\mathbf{C}_k$, i.e., $\mathbf{C}_k = \begin{bmatrix} \alpha_k & \underline{v}_k^T \\ \underline{v}_k & \mathbf{C}_k \end{bmatrix}$, and where the column vectors $\underline{E}_k$, $\underline{e}_k$, $\underline{S}_k$, and $\underline{s}_k$ are given, as function of the error signal $e_k = x_k - (q^0 * a)_k$ and the so-called signature signal $s_k = (q' * a)_k$, by

$$\begin{aligned} \underline{E}_k &= [e_k, \dots, e_{k-L}]^T, \\ \underline{e}_k &= [e_{k-1}, \dots, e_{k-L}]^T, \\ \underline{S}_k &= [s_k, \dots, s_{k-L}]^T \end{aligned}$$

and

$$\underline{s}_k = [s_{k-1}, \dots, s_{k-L}]^T.$$

The proportionality factor in (8) equals $\sqrt{((2\pi)^L \det \mathbf{c}_k / (2\pi)^{L+1} \det \mathbf{C}_k)}$ which is independent of δ and thus can be simply ignored. It follows, by taking the logarithm of (7) and invoking (8), that ML phase-error estimation is obtained by minimizing the following cost function:

$$\Lambda(\delta) = \sum_{k=1}^N (\underline{E}_k - \delta \underline{S}_k)^T \mathbf{C}_k^{-1} (\underline{E}_k - \delta \underline{S}_k) - (\underline{e}_k - \delta \underline{s}_k)^T \mathbf{c}_k^{-1} (\underline{e}_k - \delta \underline{s}_k). \quad (9)$$

This expression of $\Lambda(\delta)$ is still quite complex in that it involves inversions of the matrices $\mathbf{C}_k$ and $\mathbf{c}_k$ for all possible symbol clusters $\underline{a}_{k-L-K_1}^{k+K_2}$. A simplified expression of $\Lambda(\delta)$ can be derived via the matrix inversion lemma [16] and reads

$$\Lambda(\delta) = \sum_{k=1}^N \frac{1}{\sigma_k^2} (\underline{w}_k^T (\underline{E}_k - \delta \underline{S}_k))^2 \quad (10)$$

where $\underline{w}_k = \begin{bmatrix} 1 \\ -\mathbf{c}_k^{-1} \underline{v}_k \end{bmatrix}$ (of size $(L+1) \times 1$) and $\sigma_k^2 = \alpha_k - \underline{v}_k^T \mathbf{c}_k^{-1} \underline{v}_k$. The complexity to compute $\Lambda(\delta)$ is brought down to $O(N(L+1))$ in (10) instead of $O(N(L+1)^2)$ in (9). The vectors $\mathbf{c}_k^{-1} \underline{v}_k$ can be interpreted as data-dependent noise predictors and

Fig. 2. The ML-TED. The vector $\underline{w}_k$ and the variance σ_k^2 are varied as function of the symbol cluster $\underline{a}_{k-L-K_1}^{k+K_2}$.

the values σ_k^2 as noise-prediction variances. In fact, for a given symbol cluster $\underline{a}_{k-L-K_1}^{k+K_2}$, $\underline{w}_k = \underline{w}(\underline{a}_{k-L-K_1}^{k+K_2})$ acts to whiten the noise n_k by subtracting from n_k the predicted component from the past noise samples. The variance of the whitened noise, i.e., of $\underline{w}_k^T \underline{n}_k$, equals σ_k^2 .

The ML one-shot phase-error estimate Δ^{ML} can be easily derived from (10) and is given by

$$\Delta^{\text{ML}} = \frac{1}{\sum_{k=1}^N \frac{1}{\sigma_k^2} (\underline{w}_k^T \underline{S}_k)^2} \sum_{k=1}^N \frac{1}{\sigma_k^2} (\underline{w}_k^T \underline{E}_k) (\underline{w}_k^T \underline{S}_k). \quad (11)$$

The ML phase-error estimate (11) can be seen as a normalized average of an instantaneous timing error function given by $(1/\sigma_k^2)(\underline{w}_k^T \underline{E}_k)(\underline{w}_k^T \underline{S}_k)$. Because, in a PLL based timing recovery scheme, the averaging operation is ensured by the loop filter, the ML timing error detector (ML-TED) can be simply written as

$$\chi_k^{\text{ML}} = \frac{1}{\sigma_k^2} (\underline{w}_k^T \underline{E}_k) (\underline{w}_k^T \underline{S}_k) \quad (12)$$

where the vector $\underline{w}_k = \underline{w}(\underline{a}_{k-L-K_1}^{k+K_2})$ and the scalar $\sigma_k^2 = \sigma^2(\underline{a}_{k-L-K_1}^{k+K_2})$ correspond to the cluster $\underline{a}_{k-L-K_1}^{k+K_2}$.

Equation (12) presents two interesting properties. First, the division with σ_k^2 provides a weighing for every cluster of symbols $\underline{a}_{k-L-K_1}^{k+K_2}$. The weight of a given cluster is inversely proportional to σ_k^2 . More reliable symbol clusters that have smaller “unpredictable” noise variance will be attributed higher gains in the extraction of timing information than noisy clusters. Second, the “predictable” component of n_k from $n_{k-1}, \dots, n_{k-L}$ is removed via the scalar product with $\underline{w}_k$, thus allowing less noise power to be sensed by the timing recovery subsystem. These two properties together make up the strength of the proposed TED.

A block diagram of the ML-TED is shown in Fig. 2. This TED has attractive practical properties. First, from an implementation standpoint, the proposed TED is quite simple in that it requires only two additional FIR filters of length $(L+1)$ and one division. Second, the causal and minimum phase structure of $\underline{w}_k$ causes the latency of the ML-TED to be small. This limits the increase in the overall delay of the timing recovery loop due to the ML-TED. This property is very crucial in view of the impact

Fig. 3. Phase-domain model of the ML-TED. The TED gain $K_d(\underline{a})$ and additive noise $u_k(\underline{a})$ are dependent on the symbol cluster $\underline{a} = \underline{a}_{k-K_1}^{k+K_2}$.

of the overall delay of the timing-recovery loop on its stability margin and convergence speed [17].

Example 1: In the case of zero-mean additive white and data-independent noise with a variance σ^2 , we have $L = 0$, $\sigma_k^2 = \sigma^2$, and $\underline{w}_k = 1$. Equation (10) boils down to

$$\Lambda(\delta) = \frac{1}{\sigma^2} \sum_{k=1}^N (e_k - \delta(q' * a)_k)^2$$

where $e_k = x_k - (q^0 * a)_k$. The optimum TED in this case is the zero-forcing (ZF) TED [1]. Its output, multiplied by σ^2 , is given by

$$\chi_k^{\text{ZF}} = e_k(q' * a)_k. \quad (13)$$

Because the ZF-TED achieves maximum-likelihood when noise is data-independent AWGN, we consider the ZF-TED as baseline of comparison.

IV. EFFICIENCY OF DATA-DEPENDENT TIMING RECOVERY

The objective of any timing error detector is to provide an indication of the phase-error present at the detector input. The capability of the timing recovery loop to track fast timing variations depends heavily on how much timing information the TED can extract from the incoming signal, while rejecting the noise as much as possible. Good noise suppression requires the loop bandwidth to be as small as possible, whereas a wide bandwidth is required in order to track fast timing variations. In order to quantify this tradeoff, a measure of efficiency was introduced in [18]. The efficiency of a TED was defined as the amount of the timing information that the TED is able to extract from the incoming signal per unit of time and SNR. In this section, we extend the efficiency analysis of [18] to the ML-TED and show that this efficiency exceeds that of the ZF-TED.

The ML-TED (11) can be linearized as indicated in Fig. 3, where $K_d(\underline{a}) = ((\underline{w}(\underline{a})^T \underline{S}(\underline{a}))^2 / \sigma^2(\underline{a}))$ denotes the TED gain and $u_k(\underline{a})$ is the TED additive noise which induces jitter in the PLL. Both quantities are symbol cluster dependent. The TED noise and average gain can be written as

$$K_d = E_{\underline{a}} \left[\frac{(\underline{w}(\underline{a})^T \underline{S}(\underline{a}))^2}{\sigma^2(\underline{a})} \right]$$

$$u_k = \frac{(\underline{w}_k^T \underline{S}_k)(\underline{w}_k^T \underline{N}_k)}{\sigma_k^2} \quad (14)$$

where $E_{\underline{a}}[\cdot]$ denotes averaging over all possible symbol clusters $\underline{a}$ of length $K_1 + K_2 + L + 1$ and where, for clarity, the two equivalent notations $\underline{X}(\underline{a}_{k-L-K_1}^{k+K_2})$ and $\underline{X}_k$ are used.

The efficiency of a TED was defined in [18] as

$$\gamma = \frac{1}{\text{SNR}} \frac{K_d^2}{\mathcal{U}(0)}$$

where $\mathcal{U}(0)$ is the power spectral density of u_k at dc. Invoking (14) and remarking that $\underline{w}_k^T \underline{N}_k$ is white with variance σ_k^2 , the expression of the ML-TED efficiency can be simplified as

$$\gamma^{\text{ML}} = \frac{1}{\text{SNR}} E_{\underline{a}} \left[\frac{(\underline{w}(\underline{a})^T \underline{S}(\underline{a}))^2}{\sigma^2(\underline{a})} \right]. \quad (15)$$

This efficiency not only includes a measure of the high frequency spectrum of the transmitted data, i.e., $E[s_k^2] = \int (2\pi\Omega)^2 |Q(e^{j2\pi\Omega})|^2 \mathcal{A}(e^{j2\pi\Omega}) d\Omega$ where Q denotes the Fourier transform of q_k^0 and $\mathcal{A}$ is the data power spectral density, but also includes a measure of how noisy every symbol cluster is. The efficiency γ^{ML} can be seen as the average of a per-cluster efficiency $\gamma(\underline{a}) = (1/\text{SNR})(\underline{w}(\underline{a})^T \underline{S}(\underline{a}))^2 / \sigma^2(\underline{a})$. Good symbol clusters for timing recovery are clusters $\underline{a}$ for which $\gamma(\underline{a})$ is maximized. This result can be exploited to design optimal preamble patterns. This must be subject to maximizing the average per-cluster efficiency over all possible clusters in the preamble pattern, i.e., $E_{\underline{a} \in \text{preamble}}[\gamma(\underline{a})]$.

Example 2: For the sake of comparison between the ZF-TED and the ML-TED, let us consider the case when the noise n_k is white, i.e., $L = 0$, but with a data-dependent variance $\sigma^2(\underline{a})$. The ML-TED efficiency simplifies then to

$$\gamma^{\text{ML}} = \frac{1}{\text{SNR}} E_{\underline{a}} \left[\frac{s_k^2}{\sigma^2(\underline{a})} \right].$$

The ZF-TED has a gain $K_d = E_{\underline{a}}[s_k^2]$ and noise $u_k = s_k n_k$. Its efficiency in this case can be written as

$$\gamma^{\text{ZF}} = \frac{1}{\text{SNR}} \frac{E_{\underline{a}}[s_k^2]^2}{E_{\underline{a}}[s_k^2 \sigma^2(\underline{a})]}.$$

Now using the Cauchy–Schwarz inequality, it is easy to prove that

$$\gamma^{\text{ML}} \geq \gamma^{\text{ZF}}$$

with equality only when the noise variance is data-independent, i.e., $\sigma^2(\underline{a}) = C^{\text{st}}$.

For the sake of illustration, let us consider, in this example, the simplifying case where the data is uncoded and the noise variance of the symbol cluster $\underline{a}_{k-K_1}^{k+K_2}$ is only dependent on the central bit a_k , i.e., $\sigma^2(\underline{a}_{k-K_1}^{k+K_2}) = \sigma^2(a_k)$.

The efficiency of the ML-TED simplifies in this case to $\gamma^{\text{ML}} = (1/\text{SNR}) E_{\underline{a}}[s_k^2] E_{\underline{a}}[(1/\sigma^2(a_k))]$ because s_k is independent of a_k due to the fact that $q_0^0 = 0$. Similarly, the ZF-TED efficiency can be written as $\gamma^{\text{ZF}} = (1/\text{SNR})(E_{\underline{a}}[s_k^2]/E_{\underline{a}}[\sigma^2(a_k)])$. The gain in efficiency brought by the ML-TED over the ZF-TED can be expressed as function of $\beta = \sigma^2(-1)/\sigma^2(1)$ as follows:

$$\frac{\gamma^{\text{ML}}}{\gamma^{\text{ZF}}}(\beta) = E[1/\sigma^2(a_k)] E[\sigma^2(a_k)] = \frac{1}{4} \left(2 + \beta + \frac{1}{\beta} \right).$$

Fig. 4. Gain in efficiency of the ML-TED over the ZF-TED as function of $\beta = \sigma^2(-1)/\sigma^2(1)$, for the case that data is uncoded and $L = K_1 = K_2 = 0$ as considered in the illustration of Example 2.

Fig. 4 shows this gain as function of β . It shows in particular that for $\beta \gg 1$ or $\beta \ll 1$ a substantial improvement in efficiency is obtained using the ML-TED compared to the ZF-TED. For $\beta \gg 1$ we have $(\gamma^{\text{ML}}/\gamma^{\text{ZF}}) \propto \beta$ and $(\gamma^{\text{ML}}/\gamma^{\text{ZF}}) \propto (1/\beta)$ for $\beta \ll 1$.

V. ADAPTIVE DATA-DEPENDENT NOISE CHARACTERIZATION

In the previous section, we assumed that $\underline{w}(a_{k-L-K_1}^{k+K_2})$ and $\sigma^2(a_{k-L-K_1}^{k+K_2})$ are known for all symbol clusters. However, the statistics of the noise are not known in practice and need to be estimated from the received signal. Moreover, tracking these statistics adaptively is preferable in many applications because the noise may be nonstationary. An estimation algorithm of the noise model parameters was presented in [8]. This is based on first estimating the covariance matrices $\mathbf{C}(\underline{a})$ and then deriving the vectors $\underline{w}(\underline{a})$ and the variances $\sigma^2(\underline{a})$ via solving a linear equation that involves inverting the matrices $\mathbf{C}(\underline{a})$. This means that at every adaptation of one $\underline{w}(\underline{a})$ a covariance matrix needs to be inverted which can be complex especially for high values of L . A simpler alternative that does not involve inverting the covariance matrices can be proposed. In fact, as mentioned in Section III, the scalar product with $\underline{w}(a_{k-L-K_1}^{k+K_2}) = [-\rho(a_{k-L-K_1}^{k+K_2})]$ is meant to whiten the noise samples $n_k, \dots, n_{k-L}$, for the symbol cluster $a_{k-L-K_1}^{k+K_2}$, and $\sigma^2(a_{k-L-K_1}^{k+K_2})$ is the variance of the whitened noise. Thus, a scheme to estimate and track the prediction vector $\rho(a_{k-L-K_1}^{k+K_2})$ can be simply based on minimizing $E[(\underline{w}_k^T \underline{N}_k)^2]$. The overall estimation scheme is shown in Fig. 5.

At every clock cycle, one prediction vector $\rho(a_{k-L-K_1}^{k+K_2})$ and one variance $\sigma^2(a_{k-L-K_1}^{k+K_2})$ are adapted. The adaptation of the prediction vector is based on the least mean square (LMS) technique and seeks to minimize $(\underline{w}_k^T \underline{N}_k)^2$. The adaptation of $\rho(a_{k-L-K_1}^{k+K_2})$ and estimation of $\sigma^2(a_{k-L-K_1}^{k+K_2})$ are given by

$$\begin{aligned} \rho(\underline{a})^{\text{new}} &= \rho(\underline{a})^{\text{old}} + \mu_\rho (\underline{w}^{\text{old}}(\underline{a})^T \underline{N}_k) \underline{n}_k \\ \sigma^2(\underline{a})^{\text{new}} &= (1 - \mu_{\sigma^2}) \sigma^2(\underline{a})^{\text{old}} + \mu_{\sigma^2} (\underline{w}^{\text{old}}(\underline{a})^T \underline{N}_k) \underline{n}_k \end{aligned} \quad (16)$$

Fig. 5. Adaptation of $\rho(a_{k-L-K_1}^{k+K_2})$ and estimation of $\sigma^2(a_{k-L-K_1}^{k+K_2})$. The averaging $(\cdot)$ is symbol cluster dependent.

Fig. 6. Discrete-time model of a PRML system including a data-dependent noise generator.

where μ_ρ and μ_{σ^2} denote the adaptation constants for the adaptation of $\rho(a_{k-L-K_1}^{k+K_2})$ and estimation of $\sigma^2(a_{k-L-K_1}^{k+K_2})$, $\underline{n}_k = [n_{k-1}, \dots, n_{k-L}]^T$, and $\underline{a} = a_{k-L-K_1}^{k+K_2}$.

In practice, n_k is not available to the receiver and the adaptation of the prediction parameters has to be based on the error signal e_k . In this case, one would like to ensure a proper dimensioning of the timing recovery loop. In fact, in order to ensure that average TED gain is well defined, one must include in the characterization of the prediction parameters, used by the ML-TED, a constraint on the average TED gain. A simple solution to this issue is presented in the Appendix.

VI. SIMULATION RESULTS FOR A PRML SYSTEM

Receivers for PRML systems typically use a linear equalizer followed by a Viterbi detector (VD). The equalizer aims at shaping the channel response h_k to an acceptably shorter target response g_k in order to limit the implementation complexity of the detector. A discrete-time model of a PRML system is shown in Fig. 6.

By way of illustration, we consider run-length-limited data with run-length parameters $(d, k) = (1, 7)$ transmitted over an idealized optical storage channel according to the Braat-Hopkins model [19]

$$H(f) = \begin{cases} \frac{\sin(\pi f)}{\pi f} \left(\cos^{-1} \left| \frac{f}{f_c} \right| - \frac{f}{f_c} \sqrt{1 - \left(\frac{f}{f_c} \right)^2} \right), & |f| < f_c \\ 0, & |f| \geq f_c \end{cases}$$

where f_c denotes the optical cutoff frequency normalized to the baud frequency, fixed at 1/3 in the sequel. These choices reflect the system described in [20]. The channel output is corrupted by two different noise components. The first one is data-dependent noise m_k (media noise) and the second noise component is additive white Gaussian noise z_k (electronic noise) with zero mean and variance σ_z^2 . One of the most dominant sources of media noise in optical storage is caused by size variations of the pits written on the disc. In terms of replay signal, this can be equivalently seen as if the pits on the disc represent fuzzy ones, i.e., values of the form $1 + u_k$. The lands, representing -1 on the disc, are not hampered by media noise. The data-dependent

Fig. 7. Open-loop characteristics of the ZF and ML TEDs at SMNR = 12 dB and SANR = 16 dB.

noise m_k results then from a noise source u_k that is injected at the channel input, i.e., $m_k = (h * u)_k$. The noise u_k is modeled as additive white Gaussian noise with a variance that depends on the bit a_k , $\sigma_{u,-1}^2 = 0$ for $a_k = -1$ and $\sigma_{u,1}^2 \neq 0$ for $a_k = 1$. Two SNR measures are defined: a signal-to-media-noise ratio (SMNR) and a signal-to-additive-noise ratio (SANR) given by

$$\text{SMNR} = \frac{2}{\sigma_{u,1}^2}$$

and

$$\text{SANR} = \frac{\sum_k h_k^2}{\sigma_z^2}.$$

The channel output is subject to a delay ϕ as shown in Fig. 6. The channel output r_k is first filtered by the equalizer and then interpolated at a delay $-\psi$ where ψ is provided by the timing recovery subsystem. The interpolator is implemented via a six-tap Lagrange filter [21]. The equalized and interpolated signal x_k is subtracted from a reference signal $(g * a)_k$ to produce an error signal e_k where g_k denotes the detector target response. This error signal is used by the timing recovery subsystem to adjust the interpolation phase and by the noise characterization block to estimate the noise prediction parameters. The 5-tap target response $g = [0.17, 0.5, 0.67, 0.5, 0.17]$ and a 9-tap equalizer are used. Bit detection is implemented via a Viterbi detector. Because of the $d = 1$ constraint, the number of states in the Viterbi trellis reduces to 10 which is standard for current optical storage systems.

Before comparing our timing recovery algorithm with the ZF algorithm, a few steps are needed. First, the equalizer taps are trained using the LMS algorithm at $\phi = \psi = 0$ and noise characterization is achieved as explained in Section V. The noise characterization parameters are fixed to $L = 3$ and $K_1 = K_2 = 1$. Second, a calibration process is used in order to ensure that both the ZF-TED and the ML-TED have the same open-loop gain. To this aim, we fix ϕ at a given value and normalize the ML-TED and the ZF-TED such that their open-loop average output equals Δ .

The open-loop characteristics of both the ZF and the ML TEDs after calibration are shown in Fig. 7 for the case where media noise is dominant, e.g., SMNR = 12 dB and SANR =

16 dB. The left plot shows the average of the open-loop TED output χ_k where it is apparent that after calibration, both ZF and ML TEDs have the same gain especially near $\Delta = 0$. However, as shown in the right plot of Fig. 7, the variance of χ_k for the ZF-TED is always higher than that of the ML-TED. The reduction in the variance of the open-loop TED output amounts in this case to around 2.5 dB at $\Delta = 0$. One should recall that in closed-loop and when the PLL is in tracking mode, only the TED behavior for $\Delta \simeq 0$ matters. The increase in the open-loop noise variance around $\Delta = 0.5$ can be explained by the fact that the first order approximation of x_k as function of Δ given in (1) holds only for $\Delta \simeq 0$.

The gain in the open-loop TED variance depends obviously on SANR and SMNR. This gain as function of SANR and SMNR follows the same trend as the gain in timing jitter in closed-loop simulations. This is the subject of the next paragraph.

In order to assess the performance gain of the ML-TED over the ZF-TED in closed-loop as function of SMNR and SANR, we force the delay ϕ to be a step function of time, i.e., $\phi_k = 0$ for $k < k_0$ and $\phi_k = 0.1$ for $k \geq k_0$ where the timing recovery loop is closed at $k = k_0$. The loop filter parameters, i.e., natural frequency w_n and damping factor ζ , are optimized in order to achieve the best BER. This optimization was carried out at SANR = SMNR = 16 dB. The optimal BER was achieved for $w_n = 0.03$ and $\zeta = 2$. Because optimization of the loop filter parameters at different values of SANR and SMNR did not show any important performance improvement, we simply fix the loop filter parameters throughout our simulations.

A first measure to estimate the performance of a timing recovery scheme is the timing jitter defined as the interpolation phase-error variance, i.e., $\sigma_\Delta^2 = E[(\phi - \psi)^2]$. Fig. 8 shows timing jitter as function of SMNR for different values of SANR for the ZF and the ML timing recovery schemes. First, it is apparent that the ML timing recovery is always superior to the ZF timing recovery. Second, the gain of our scheme over the ZF scheme is highly dependent on the ratio of SMNR and SANR. For a given SANR, the gain is higher at low SMNR and vice versa. This gain goes from 0.3 dB at SANR = 10 dB and SMNR = 22 dB to around 2.5 dB for SANR = 16 dB and

Fig. 8. Timing jitter as function of SMNR for different values of SANR.

Fig. 9. Simulated BER as function of SMNR for different values of SANR.

SMNR = 12 dB. This is in accordance with the open-loop gain in TED output variance around $\Delta = 0$ shown in Fig. 7.

In terms of BER, Fig. 9 shows simulated BER as function of SMNR for different values of SANR. At low SANR values, e.g., SANR = 10 dB, the BER curve is not a steep function of SMNR and thus a little gain in timing jitter translates into a big gain in SMNR, e.g., 6 dB gain at BER = 10^{-3} . At high SANR values, e.g., SANR = 16 dB, the BER is mainly determined by the data-dependent media noise. For this reason, a substantial gain in timing jitter does not translate directly into a relatively big gain in BER. Still, a gain of more than 1 dB in SMNR is achieved at BER = 10^{-3} . At higher SMNR values the impact of timing jitter is much more visible and our timing recovery scheme allows a gain of 2 dB in SMNR.

VII. CONCLUSION

In this paper, a new timing recovery algorithm for recording channels with data-dependent noise was presented. Based on a Gauss–Markov correlated noise model, a maximum-likelihood timing recovery algorithm was derived and analyzed. The new algorithm incorporates, on the one hand, data-dependent noise

prediction and on the other hand a data-dependent weighing. The noise prediction aims at whitening the data-dependent noise and the weighing makes the gain of the timing error detector data dependent, i.e., smaller gain for noisier data patterns and vice versa. Moreover, because in practice noise can be non-stationary, a simple adaptation scheme is proposed to estimate and track the noise prediction parameters. Simulation results for a partial response maximum-likelihood system show that the proposed algorithm allows significant improvements in performance in the presence of data-dependent noise.

APPENDIX

A. Dimensioning of the ML Timing Recovery Loop

We described in Section III the ML-TED (12). This TED uses knowledge about noise in the form of a data-dependent whitening vector $\underline{w}_k^T = \underline{w}^T(\underline{a}_{k-L-K_1}^{k+K_2})$ and whitened noise variance $\sigma^2(\underline{a}_{k-L-K_1}^{k+K_2})$. The characterization of the whitening vector and whitened noise variance was presented in Section V. This characterization does not involve any constraint on the TED gain and thus the overall gain of the timing recovery loop is “ill-defined.” However, for proper dimensioning of the timing recovery loop, one would like to have a controlled TED gain. This Appendix describes how noise characterization (16) can be modified to include a constraint on the TED gain.

Fig. 3 describes the phase-domain model of the TED of (12). The TED gain is given by

$$K_d(\underline{a}_{k-L-K_1}^{k+K_2}) = \frac{(\underline{w}(\underline{a})^T \underline{S}(\underline{a}))^2}{\sigma^2(\underline{a})} \quad (17)$$

and the TED noise $u_k(\underline{a}_{k-L-K_1}^{k+K_2})$ is given by

$$u_k(\underline{a}_{k-L-K_1}^{k+K_2}) = \frac{(\underline{w}(\underline{a})^T \underline{N}_k)(\underline{w}(\underline{a})^T \underline{S}(\underline{a}))}{\sigma^2(\underline{a})}.$$

The average TED gain $\bar{K}_d$ is defined as the average of $K_d(\underline{a})$ over all possible symbol clusters, i.e.,

$$\begin{aligned} \bar{K}_d &= \sum_{\underline{a}} p(\underline{a}) K_d(\underline{a}) \\ &= \sum_{\underline{a}} p(\underline{a}) \frac{(\underline{w}(\underline{a})^T \underline{S}(\underline{a}))^2}{\sigma^2(\underline{a})} \end{aligned} \quad (18)$$

where $p(\underline{a})$ is the probability of occurrence of the symbol cluster $\underline{a}$. This probability depends only on the coding scheme and is assumed to be known *a priori*.

In order to constrain the average TED gain (18) to a fixed value, e.g., 1, while characterizing the data-dependent noise, the variances $\sigma^2(\underline{a})$ for all symbol clusters $\underline{a}$ must be scaled with the same value such that $\bar{K}_d = 1$. The adaptation of $\underline{\rho}(\underline{a})$ (equivalently $\underline{w}(\underline{a})$) is unchanged and is given by (16). The estimation of $\sigma^2(\underline{a})$ is modified to

$$\begin{aligned} \lambda &= \sum_{\underline{a}' \in I} p(\underline{a}') \frac{(\underline{w}^{\text{old}}(\underline{a}')^T \underline{S}(\underline{a}'))^2}{\sigma^{\text{old}}(\underline{a}')} \\ \sigma^{\text{new}}(\underline{a}) &= (1 - \mu_{\sigma^2}) \sigma^{\text{old}}(\underline{a}) + \mu_{\sigma^2} \lambda (\underline{w}^{\text{old}}(\underline{a})^T \underline{N}_k) n_k \end{aligned} \quad (19)$$

where we introduced the data-independent variable λ in order to force $\bar{K}_d$ to 1.

In order to reduce implementation complexity, the variable λ can be computed on a subset I of symbol clusters (not necessary all symbol clusters). In this case, $p(\underline{a})$ must be normalized such that $\sum_{\underline{a}' \in I} p(\underline{a}') = 1$. The variable λ needs to be updated only if the prediction parameters of one of the symbol clusters of I is updated.

In practice, the set I may be chosen to contain few symbol clusters. A particular case is $I = \{\underline{a}_0\}$, where the TED gain is then constrained to be unity for the symbol cluster $\underline{a}_0$. This is of interest when the symbol cluster $\underline{a}_0$ is often present in the transmitted data, e.g., part of a preamble sequence. In this case, λ is computed as

$$\lambda = \frac{(\underline{w}(\underline{a}_0)^T \underline{S}(\underline{a}_0))^2}{\sigma^2(\underline{a}_0)}$$

and needs to be recomputed only if $\underline{w}(\underline{a}_0)$ or $\sigma^2(\underline{a}_0)$ are changed. Note that $\underline{w}(\underline{a}_0)^T \underline{S}(\underline{a}_0)$ is computed in the TED (12) and thus does not require any extra circuitry.

ACKNOWLEDGMENT

This work was supported by the European Union under project IST-2001-34168 (TwoDOS).

REFERENCES

- [1] J. W. M. Bergmans, *Digital Baseband Transmission and Recording*. Norwell, MA: Kluwer, 1996.
- [2] K. H. Mueller and M. Muller, "Timing recovery in digital synchronous data receivers," *IEEE Trans. Commun.*, vol. COM-24, no. 5, pp. 516–531, May 1976.
- [3] S.U.H. Qureshi, "Timing recovery for equalized partial-response systems," *IEEE Trans. Commun.*, vol. COM-24, no. 5, pp. 516–531, May 1976.
- [4] R. D. Cideciyan *et al.*, "A PRML system for digital magnetic recording," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 3, pp. 38–56, Jan. 1992.
- [5] J. W. M. Bergmans and H. W. Wong-Lam, "A class of data-aided timing-recovery schemes," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1819–1827, Feb./Mar./Apr. 1995.
- [6] A. Kavcic and J.M.F. Moura, "Correlation-sensitive adaptive sequence detection," *IEEE Trans. Magn.*, vol. 34, no. 3, pp. 763–771, May 1998.
- [7] A. Kavcic and J.M.F. Moura, "The Viterbi algorithm and Markov noise memory," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 291–301, Jan. 2000.
- [8] A. Kavcic and A. Pataoutian, "A signal-dependent autoregressive channel model," *IEEE Trans. Magn.*, vol. 35, no. 5, pp. 2316–2318, Sep. 1999.
- [9] J. Slander and A. Pataoutian, "Performance of a signal-dependent autoregressive channel model," *IEEE Trans. Magn.*, vol. 36, no. 5, pp. 2197–2199, Sep. 2000.
- [10] A. Gameiro, "Baud sampling bit synchroniser for channels with data dependent noise," *Electron. Lett.*, vol. 34, no. 21, pp. 2000–2002, Oct., 15th 1998.
- [11] H. Song, J. Liu, B. Kumar, and Y. Yuan, "Application of soft output Viterbi algorithm (SOVA) in timing recovery for iterative soft decoded magnetic recording channels," presented at the IEEE Int. Magnetics Conf., Jan. 2001.
- [12] A. R. Nayak, J. R. Barry, and S. W. McLaughlin, "Joint timing recovery and turbo equalization for coded partial response channels," *IEEE Trans. Magn.*, vol. 38, no. 5, pp. 2295–2297, Sep. 2002.
- [13] J. R. Barry *et al.*, "Iterative timing recovery," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 89–102, Jan. 2004.
- [14] F. M. Gardner, *Phaselock Techniques*, 2nd ed. New York: Wiley, 1979.
- [15] H. Sun, G. Mathew, and B. Farhang-Boroujeny, "Detection techniques for high-density magnetic recording," *IEEE Trans. Magn.*, vol. 41, no. 3, pp. 1193–1199, Mar. 2005.
- [16] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*. New York: Addison-Wesley, 1991.
- [17] J.W.M. Bergmans, "Effect of loop delay on stability of discrete-time PLL," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 42, no. 4, pp. 229–231, 1995.
- [18] J. W. M. Bergmans, "Efficiency of data-aided timing recovery techniques," *IEEE Trans. Inf. Theory*, vol. 41, no. 5, pp. 1397–1408, Sep. 1995.
- [19] G. Bouwhuis *et al.*, *Principles of Optical Disc Systems*. Bristol, U.K.: Adam Hilger, 1985.
- [20] T. Narahara *et al.*, "Optical disc system for digital video recording," *Jpn. J. Appl. Phys.*, pt. 1, vol. 39, no. 2B, pp. 912–919, Feb. 2000.
- [21] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, 9th ed. New York: Dover, 1970.

Manuscript received November 1, 2005; revised July, 3 2006. Corresponding author: J. Riani (e-mail: j.riani@tue.nl).