Timing Recovery With Frequency Offset and Random Walk: Cramér–Rao Bound and a Phase-Locked Loop Postprocessor

Aravind R. Nayak, *Member, IEEE*, John R. Barry, *Senior Member, IEEE*, German Feyh, *Member, IEEE*, and Steven W. McLaughlin, *Fellow, IEEE*

Abstract—We consider the problem of timing recovery for bandlimited, baud-rate sampled systems with intersymbol interference and a timing offset that can be modeled as a combination of a frequency offset and a random walk. We first derive the Cramér–Rao bound (CRB), which is a lower bound on the estimation error variance for any timing estimator. Conventional timing recovery is based on a phase-locked loop (PLL). We compare the conventional timing-recovery method with the CRB for realistic timing parameters for the magnetic recording channel, and observe a 7 dB signal-to-noise ratio gap between the two. Next, we propose a PLL postprocessor based on the maximum *a posteriori* estimation principle that performs to within 1.5 dB of the CRB. This postprocessor performs time-invariant filtering and time-varying scaling of the PLL timing estimates. The refined timing estimates from the postprocessor are then used to get refined samples by interpolating the samples taken at the PLL's timing estimates. Finally, we present suboptimal implementations that allow a performance-complexity tradeoff.

Index Terms—Cramér–Rao bound (CRB), maximum *a posteriori* (MAP), phase-locked loop (PLL), random walk, timing recovery.

I. INTRODUCTION

TIMING recovery is an integral part of communication and data storage systems. Most physical communication channels are analog in nature, whereas the data to be communicated is usually digital and discrete. At the transmitter, the data to be transmitted is used to modulate an analog waveform to suit the channel characteristics. At the receiver, the analog waveform needs to be converted back into the digital domain for further processing. This is achieved with the aid of a sampling device that samples the received waveform at instants determined by a timing-recovery device. The quality of the samples, and hence, the performance of the receiver, depend strongly on the performance of the timing-recovery device.

In this paper, we consider the problem of timing recovery for bandlimited, baud-rate sampled systems with intersymbol interference (ISI). A baud-rate sampled system is one where the receiver collects one sample per transmitted symbol, as is common in high-rate communication systems and high-density magnetic recording. At the high data rates and storage densities of these systems, the sampling devices [analog-to-digital converters (ADCs)] are usually run at as high a sampling rate as possible, and the operating speed of the ADCs ends up being the limiting factor that determines the maximum operating data rate of the overall system. For a given ADC sampling rate, oversampling requirements due to the algorithms employed lead to a corresponding reduction in the maximum data rate for the overall system. In general, oversampling leads to performance improvement, when compared with baud-rate sampling [1]. However, with careful design, this performance loss for baud-rate sampled algorithms can be reduced to a minimum, while allowing significantly higher data rates than those possible with algorithms that require oversampling [1], [2].

The timing-offset model considered here consists of a frequency offset and a random walk. A popular method of timing recovery is based on the phase-locked loop (PLL) [3]. A PLL is a feedback-control system that adapts the sampling instants based on an error signal derived from the samples. A PLL consists of a sampling device, a timing-error detector (TED) (also called a phase-error detector), a loop filter, and a voltage-controlled oscillator (VCO). The samples from the sampling device are used by the TED to generate estimates of the timing error. These timing-error estimates are filtered by the loop filter, which is usually a lowpass filter, to generate the PLL timing estimates. The VCO finally uses these timing estimates to generate the sampling instants at which the sampling device is activated. The conventional timing-recovery method, for the system model considered here, involves a second-order loop filter.

The PLL is well known in the literature, and is an integral part of communication and control systems, largely due to its simplicity and good performance. However, the PLL does not necessarily lead to the lowest estimation error possible. The PLL's loss of performance can be quantified by comparing its error variance to that predicted by the Cramér–Rao bound (CRB), which is a lower bound on the error variance of any unbiased estimator [4]. The CRB is well known in the literature for constant-offset [5] and frequency-offset timing models [6]. In this paper, we derive the CRB for a general timing-offset model consisting of a frequency offset and a random walk. For a magnetic recording channel with realistic system parameters, the

PLL performs around 7 dB worse than the CRB, and this large gap motivates the search for better timing-recovery methods.

Considerable effort in the literature on improving timing-recovery methods has focused on improving components of the PLL, such as the TED [7] or the decisions used by the TED [8]. Postprocessing or feedforward techniques have been proposed in the literature for oversampled systems (with two to four samples per symbol duration) involving techniques like square non-linearities and conditional maximum-likelihood (ML) estimation [9]–[11]. In this paper, we propose a postprocessor, based on the maximum *a posteriori* (MAP) estimation principle, that processes the PLL timing estimates for the whole received block to generate refined timing estimates with reduced error variance. This method is especially attractive for baud-rate systems that employ iterative processing at the receiver, where the requirement of causality can be relaxed, and the timing estimates from the current iteration can be used in the next iteration of the receiver's processing. Examples include high-rate communication and data-storage systems that employ iterative decoding, equalization, and synchronization [8], [12]–[14].

The PLL postprocessor is derived as the MAP estimator [4] for a linearized model of the PLL outputs. The vector of timing estimates from the PLL is projected onto the underlying timing-offset model using a matrix operation on the PLL outputs. To reduce complexity, we propose an approximation that allows replacing the matrix operation by a time-invariant filtering followed by a time-varying scaling, without any perceptible performance degradation. The PLL postprocessor performs to within 1.5 dB of the CRB, a gain of around 5.5 dB. Finally, we present suboptimal implementations of the algorithm that allow a performance-complexity tradeoff.

The postprocessor developed here can be viewed in a more general context. For any system characterized by a nonlinear measurement equation, the traditional approach to estimating system parameters has been to linearize the measurement equation using Taylor series techniques, and then use the Kalman filtering approach [15], [16]. We take an alternative approach. With a suitably chosen TED, the PLL gives access to linear observations. Then the projection operator defined by the MAP estimator can be used to get the desired estimates.

The rest of the paper is organized as follows. Section II introduces the system model under consideration. In Section III, the CRB is derived. In Section IV, the conventional PLL-based timing-recovery method is reviewed, and its simulated performance is compared with the CRB. In Section V, the MAP-based PLL postprocessor is derived, and the performance improvement possible is demonstrated by simulation. In Section VI, suboptimal implementations of the PLL postprocessor are proposed, and the corresponding performance-complexity tradeoff is discussed. Finally, the conclusions are presented in Section VII.

II. SYSTEM MODEL

Consider the pulse amplitude modulated system shown in Fig. 1. The channel output waveform $y(t)$ is given by

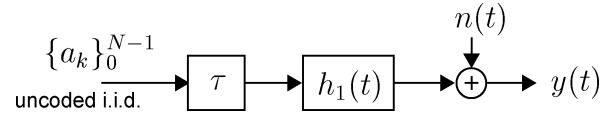$$y(t) = \sum_{l=0}^{N-1} a_l h_1(t - lT - \tau_l) + n(t) \qquad (1)$$

Fig. 1. System block diagram with timing offsets, channel distortion, and additive noise.

where T is the bit period, $a_l \in \{\pm 1\}$ are the N independent, identically distributed (i.i.d.) data symbols equally likely to be -1 or $+1$, $h_1(t)$ is the channel impulse response, $n(t)$ is additive white Gaussian noise, and τ_l is the unknown timing offset for the lth symbol. The timing-offset model is given by

$$\tau_{k+1} = \tau_k + \Delta T + w_k$$
$$= \tau_0 + (k+1)\Delta T + \sum_{l=0}^{k} w_l \qquad (2)$$

where τ_0 is the initial timing offset, ΔT is the frequency-offset parameter, and the random variables $\{w_i\}$ characterize a random walk and are i.i.d. zero-mean Gaussian random variables of variance σ_w^2. This timing model has been considered in the literature for magnetic recording applications [17], [18]. The parameter σ_w^2 determines the severity of the random walk, and represents the receiver's *a priori* knowledge regarding the random walk component. Also assumed is the *a priori* knowledge that τ_0 is a zero-mean Gaussian random variable with variance $\sigma_{\tau_0}^2$, and ΔT is a zero-mean Gaussian random variable with variance $\sigma_{\Delta T}^2$.

The channel response $h(t)$ is assumed to be bandlimited to the frequency range $[-(1/2T), 1/2T)$, with the result that baud-rate samples taken at a sampling rate of $1/T$ provide sufficient statistics. To eliminate out-of-band noise at the receiver, the received waveform $y(t)$ is filtered by a front-end filter with impulse response $f(t)$ to get the waveform $r(t)$.

III. CRB FOR THE TIMING PROBLEM

The resulting waveform $r(t)$ is sampled at instants kT to arrive at baud-rate samples $\{r_k\}$ given by

$$r_k = \sum_{l=0}^{N-1} a_l h(kT - lT - \tau_l) + n_k = x_k + n_k \qquad (3)$$

where $h(t) = h_1(t) * f(t)$, x_k is the signal component of r_k, and $\{n_k\}$ are zero-mean i.i.d. normal random variables with variance σ^2. The signal-to-noise ratio (SNR) is defined to be SNR $= E_h/2\sigma^2 T$, where $E_h = \int_{-\infty}^{\infty} |h(t)|^2 dt$ is the energy in the channel impulse response $h(t)$. A total of $N + 2M$ samples are collected and stacked in an observation vector $\boldsymbol{r} = [r_{-M} \ r_{-M+1} \ \dots \ r_{N+M-1}]^T$, where eventually $M \rightarrow \infty$. Also constructed are the vectors $\boldsymbol{x} = [x_0 \ x_1 \ \dots \ x_{N-1}]^T$, $\boldsymbol{a} = [a_0 \ a_1 \ \dots \ a_{N-1}]^T$, $\boldsymbol{\tau} = [\tau_0 \ \tau_1 \ \dots \ \tau_{N-1}]^T$, the parameter vector $\boldsymbol{\theta} = [\boldsymbol{\tau}^T \ \boldsymbol{a}^T]^T$ of size $2N \times 1$, and the vector of the estimates $\hat{\theta}_k$ that the receiver generates based on $\boldsymbol{r}$, given by $\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\tau}}^T \ \hat{\boldsymbol{a}}^T]^T$.

In the trained case, where the transmitted symbols $\{a_k\}$ are known at the receiver, the problem is simpler, and the receiver only needs to estimate $\{\tau_k\}$. However, communication systems can not be fully trained, and the receiver has no knowledge of the transmitted data in this case. The general timing-recovery problem can be phrased as one of estimating $\{a_k\}$ and $\{\tau_k\}$, given $\{r_k\}$ and knowledge of the channel response $h(t)$. The CRB for this problem is given by [4]

$$E_{\boldsymbol{r},\boldsymbol{\theta}}\left[\left(\hat{\boldsymbol{\theta}}(\boldsymbol{r}) - \boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}(\boldsymbol{r}) - \boldsymbol{\theta}\right)^T\right] \geq \boldsymbol{J}_T^{-1} \qquad (4)$$

where $\boldsymbol{J}_T$ is the total information matrix, defined by [4]

$$\boldsymbol{J}_T = E_{\boldsymbol{r},\boldsymbol{\theta}}\left\{\left[\frac{\partial}{\partial\boldsymbol{\theta}}\ln f(\boldsymbol{r},\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\boldsymbol{\theta}}\ln f(\boldsymbol{r},\boldsymbol{\theta})\right]^T\right\} \qquad (5)$$

where $f(\boldsymbol{r},\boldsymbol{\theta})$ is the joint probability density of $\boldsymbol{r}$ and $\boldsymbol{\theta}$. The estimation-error variance for the ith parameter θ_i is bounded by the ith diagonal entry of $\boldsymbol{J}_T^{-1}$.

Let $f_{r|\theta}(\boldsymbol{r}|\boldsymbol{\theta})$ be the distribution of $\boldsymbol{r}$, given $\boldsymbol{\theta}$. Also, let $f_\theta(\boldsymbol{\theta})$ be the *a priori* distribution of $\boldsymbol{\theta}$. The total information $\boldsymbol{J}_T$ can then be rewritten as [4]

$$\boldsymbol{J}_T = E_{\boldsymbol{\theta}}[\boldsymbol{J}_\theta] + \boldsymbol{J}_{ap} \qquad (6)$$

where the Fisher information matrix $\boldsymbol{J}_\theta$ is given by [4]

$$\boldsymbol{J}_\theta = E_{\boldsymbol{r}}\left\{\left[\frac{\partial}{\partial\boldsymbol{\theta}}\ln f_{r|\theta}(\boldsymbol{r}|\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\boldsymbol{\theta}}\ln f_{r|\theta}(\boldsymbol{r}|\boldsymbol{\theta})\right]^T\right\} \qquad (7)$$

and the *a priori* information matrix $\boldsymbol{J}_{ap}$ is

$$\boldsymbol{J}_{ap} = E_{\boldsymbol{\theta}}\left\{\left[\frac{\partial}{\partial\boldsymbol{\theta}}\ln f_\theta(\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\boldsymbol{\theta}}\ln f_\theta(\boldsymbol{\theta})\right]^T\right\}. \qquad (8)$$

Since the distribution $f_{r|\theta}(\boldsymbol{r}|\boldsymbol{\theta})$ is multivariate normal, given by $f_{r|\theta}(\boldsymbol{r}|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x}(\boldsymbol{\theta}), \sigma^2\boldsymbol{I})$, $\boldsymbol{J}_\theta$ can be simplified to [19]

$$\boldsymbol{J}_\theta = \frac{1}{\sigma^2}\boldsymbol{G}^T\boldsymbol{G} \qquad (9)$$

where $\boldsymbol{G}$ is the *sensitivity matrix*

$$\boldsymbol{G} = \left[\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{x}^T(\boldsymbol{\theta})\right] = [\boldsymbol{g}_0\boldsymbol{g}_1\cdots\boldsymbol{g}_{2N-1}] \qquad (10)$$

and the $N \times 1$ column vector $\boldsymbol{g}_i = \partial\boldsymbol{x}(\boldsymbol{\theta})/\partial\theta_i$ denotes the variation in the mean $\boldsymbol{x}(\boldsymbol{\theta})$ with respect to θ_i, the ith element of $\boldsymbol{\theta}$. In addition, if $\boldsymbol{G}$ is partitioned as $\boldsymbol{G} = [\boldsymbol{G}_\tau \ \boldsymbol{G}_a]$, where $\boldsymbol{G}_\tau$ corresponds to the derivatives with respect to the timing parameters τ_k, and $\boldsymbol{G}_a$ corresponds to the data parameters a_k, then $\boldsymbol{J}_\theta$ takes the following block diagonal form [20]:

$$\boldsymbol{J}_\theta = \begin{bmatrix} \boldsymbol{J}_\tau & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J}_a \end{bmatrix} \qquad (11)$$

where $\boldsymbol{J}_\tau = (1/\sigma^2)\boldsymbol{G}_\tau^T\boldsymbol{G}_\tau$ and $\boldsymbol{J}_a = (1/\sigma^2)\boldsymbol{G}_a^T\boldsymbol{G}_a$. For the specific system model considered here, $E_{\boldsymbol{\theta}}[\boldsymbol{J}_\tau]$ evaluates to [20]

$$E_{\boldsymbol{\theta}}[\boldsymbol{J}_\tau] = \frac{E_{h'}}{\sigma^2 T}\boldsymbol{I}_{N\times N} \qquad (12)$$

where we let $M \to \infty$ and $E_{h'} = \int_{-\infty}^{\infty}|(\partial/\partial t)h(t)|^2\mathrm{d}t$ is the energy in the derivative of the pulse shape $h(t)$. (See Appendix for proof.) Since the data symbols are zero-mean and independent of the timing parameters, the *a priori* information matrix $\boldsymbol{J}_{ap}$ has a block diagonal form as well, given by

$$\boldsymbol{J}_{ap} = \begin{bmatrix} \boldsymbol{J}_{ap}^\tau & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J}_{ap}^a \end{bmatrix} \qquad (13)$$

where $\boldsymbol{J}_{ap}^\tau$ and $\boldsymbol{J}_{ap}^a$ are the *a priori* information matrices corresponding to the timing and the data parameters, respectively. With *a priori* information that τ_0 is $\mathcal{N}(0, \sigma_{\tau_0}^2)$ and ΔT is $\mathcal{N}(0, \sigma_{\Delta T}^2)$, the *a priori* timing information matrix $\boldsymbol{J}_{ap}^\tau$ evaluates to [20]

$$\boldsymbol{J}_{ap}^\tau = \frac{1}{\sigma_w^2}\begin{bmatrix} \beta_1 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}_{N\times N}, \text{ where} \qquad (14)$$

$$\beta_1 = \left(\frac{N^2 - N - 1}{N^2} + \frac{\sigma_w^2}{\sigma_{\tau_0}^2} + \frac{\sigma_w^2}{N^2\sigma_{\Delta T}^2}\right). \qquad (15)$$

Combining (6), (11), and (13), the total information matrix $\boldsymbol{J}_T$ is block diagonal, as well, given by

$$\boldsymbol{J}_T = \begin{bmatrix} \boldsymbol{J}_T^\tau & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J}_T^a \end{bmatrix} \qquad (16)$$

where $\boldsymbol{J}_T^\tau$ represents the total timing information matrix. The matrix $\boldsymbol{J}_T^\tau$ needs to be inverted to arrive at the CRB. To get the CRB on timing estimation, it is sufficient to evaluate and invert $\boldsymbol{J}_T^\tau$, which is given by

$$\boldsymbol{J}_T^\tau = E_{\boldsymbol{\theta}}[\boldsymbol{J}_\tau] + \boldsymbol{J}_{ap}^\tau. \qquad (17)$$

Combining (17), (12), and (14) leads to

$$\boldsymbol{J}_T^\tau = \frac{1}{\sigma_w^2}\begin{bmatrix} \beta_2 & -1 & 0 & \cdots & 0 \\ -1 & \lambda & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -1 & \lambda & -1 \\ 0 & \cdots & 0 & -1 & \lambda-1 \end{bmatrix}, \text{ where} \qquad (18)$$

$$\lambda = 2 + E_{h'}\frac{\sigma_w^2}{\sigma^2 T} \quad \text{and} \quad \beta_2 = \beta_1 + E_{h'}\frac{\sigma_w^2}{\sigma^2 T}. \qquad (19)$$

The *a priori* information parameters $\sigma_{\tau_0}^2$ and $\sigma_{\Delta T}^2$ occur in the denominator of the fractions in β_2 [see (19) and (15)]. Both of these parameters being zero leads to perfect knowledge of the initial timing offset and the frequency offset, and the estimation problem is simply that of estimating the random walk. If $\sigma_w =$

0, $\boldsymbol{J}_T^\tau$, as written out in (18), is not invertible. Therefore, the special cases of one or more of the parameters σ_{τ_0}, $\sigma_{\Delta T}$, and σ_w being zero are dealt with separately.

Assuming that these parameters are nonzero, we arrive at the CRB by inverting $\boldsymbol{J}_T^\tau$ using the Cholesky decomposition [20], [21]. First, $\sigma_w^2 \boldsymbol{J}_T^\tau$ is factorized as $\sigma_w^2 \boldsymbol{J}_T^\tau = \boldsymbol{L}\boldsymbol{L}^T$, where $\boldsymbol{L}$ is a lower triangular matrix. Next, we solve the equation $\boldsymbol{L}\boldsymbol{C} = \boldsymbol{I}$ for $\boldsymbol{C}$. Finally, we solve $\boldsymbol{L}^T \boldsymbol{B} = \boldsymbol{C}$ for $\boldsymbol{B}$, leading to $[\boldsymbol{J}_T^\tau]^{-1} = \sigma_w^2 \boldsymbol{B}$, and this is given by

$$[\boldsymbol{J}_T^\tau]_{ij}^{-1} = \sigma_w^2 \frac{a_{\max(i,j)} n_{\min(i,j)}}{n_N - n_{N-1}} \tag{20}$$

where

$$
\begin{aligned}
a_j &= \frac{\eta^{N-j} + \eta^{-N+1+j}}{\eta + 1} \\
n_j &= \frac{(\beta_2\eta - 1)\eta^j + (\eta^2 - \beta_2\eta)\eta^{-j}}{\eta^2 - 1} \\
\eta &= \frac{\lambda + \sqrt{\lambda^2 - 4}}{2}.
\end{aligned}
\tag{21}
$$

Finally, the Cramér–Rao lower bound on the timing estimation error variance for the kth timing offset evaluates to

$$\frac{E\left[(\hat{\tau}_k - \tau_k)^2\right]}{T^2} \geq \frac{[\boldsymbol{J}_T^\tau]_{kk}^{-1}}{T^2}. \tag{22}$$

If $\sigma_{\Delta T} = 0$ and $\sigma_{\tau_0} = 0$, then this represents the case of a pure random walk. This leads to the following expression for $\boldsymbol{J}_T^\tau$:

$$\boldsymbol{J}_T^\tau = \frac{1}{\sigma_w^2}
\begin{bmatrix}
\lambda & -1 & 0 & \cdots & 0 \\
-1 & \lambda & -1 & & \vdots \\
0 & \ddots & \ddots & \ddots & 0 \\
\vdots & \ddots & -1 & \lambda & -1 \\
0 & \cdots & 0 & -1 & \lambda - 1
\end{bmatrix}, \text{ where} \tag{23}$$

$$\lambda = 2 + E_{h'} \frac{\sigma_w^2}{\sigma^2 T}. \tag{24}$$

Inverting this, following the procedure for the earlier case, leads to a similar expression [21], where

$$[\boldsymbol{J}_T^\tau]_{ij}^{-1} = \sigma_w^2 \frac{a_{\max(i,j)} n_{\min(i,j)}}{n_N - n_{N-1}} \tag{25}$$

where

$$
\begin{aligned}
a_j &= \frac{\eta^{N-j} + \eta^{-N+1+j}}{\eta + 1} \\
n_j &= \frac{\eta^{j+2} - \eta^{-j}}{\eta^2 - 1} \\
\eta &= \frac{\lambda + \sqrt{\lambda^2 - 4}}{2}.
\end{aligned}
\tag{26}
$$

In this case, further simplification is possible, and the CRB is given by

$$\frac{E\left[(\hat{\tau}_k - \tau_k)^2\right]}{T^2} \geq \frac{[\boldsymbol{J}_T^\tau]_{kk}^{-1}}{T^2} = h_{ss}f(k) \tag{27}$$

where

$$h_{ss} = \frac{\sigma_w^2}{T^2} \frac{\eta}{\eta^2 - 1} \tag{28}$$

$$f(k) = \tanh\left(\left(N + \frac{1}{2}\right)\ln\eta\right) \tag{29}$$

$$\cdot \left[1 - \frac{\sinh\left((N - 2k + \frac{1}{2})\ln\eta\right)}{\sinh\left((N + \frac{1}{2})\ln\eta\right)}\right]. \tag{30}$$

Here, h_{ss} is the steady-state value of the CRB, and $f(k)$ is the k-dependent term, and this represents the variation of the CRB with the index k. For moderately high SNR and high enough N, $f(k)$ is approximately unity in the center of the packet, and therefore, $E[(\hat{\tau}_k - \tau_k)^2]/T^2 \approx h_{ss}$ in this region. As the SNR and N increase, the steady-state value becomes more representative of the CRB for all the indices k.

The case with $\sigma_w = 0$ represents a combination of an initial timing offset and a frequency offset, and has been dealt with in the literature [5], [6]. Therefore, it is omitted here for the sake of brevity.

IV. CONVENTIONAL PLL-BASED TIMING RECOVERY

As opposed to sampling the signal at the precise time instants kT, as in the previous section, conventional timing recovery employs a PLL to arrive at the sampling instants $kT + \hat{\tau}_k$.

After front-end filtering, the channel received waveform $y(t)$ to remove out-of-band noise, the resulting continuous-time waveform $r(t)$ can be modeled as

$$r(t) = \sum_l a_l h(t - lT - \tau_l) + n_1(t) \tag{31}$$

where $n_1(t)$ is bandlimited to $[-(1/2T), 1/2T)$. The continuous-time waveform $r(t)$ is then sampled at timing instants $kT + \hat{\tau}_k$ based on the estimates $\{\hat{\tau}_k\}$ of $\{\tau_k\}$ produced by the timing-recovery system, leading to samples $\{r_k\}$. Ideally, sampling would occur at instants $kT + \tau_k$. The samples r_k are given by

$$r_k = \sum_l a_l h(kT + \hat{\tau}_k - lT - \tau_l) + n_{1k} \tag{32}$$

where $\{n_{1k}\}$ are i.i.d. zero-mean Gaussian random variables of variance σ^2.

Conventional timing recovery is based on a PLL. A first-order PLL updates its estimate of τ_k according to

$$\hat{\tau}_{k+1} = \hat{\tau}_k + \alpha\hat{\epsilon}_k \tag{33}$$

where α is the PLL gain, and where $\hat{\epsilon}_k$ is the receiver's estimate of the estimation error $\epsilon_k = \tau_k - \hat{\tau}_k$. For zero steady-state error while tracking a frequency offset, a second-order PLL is employed, which updates the estimate of τ_k according to

$$\hat{\tau}_{k+1} = \hat{\tau}_k + \alpha\hat{\epsilon}_k + \beta\sum_{l=0}^{k-1}\hat{\epsilon}_l \qquad (34)$$

where there is an additional gain parameter β.

The receiver employs a TED to arrive at timing-error estimates. The widely used Müller and Mueller (MM) TED [22] generates $\hat{\epsilon}_k$ according to

$$\hat{\epsilon}_k = r_k\hat{d}_{k-1} - r_{k-1}\hat{d}_k \qquad (35)$$

where $\hat{d}_k$ is the receiver's estimate of the noiseless, perfectly timed kth sample d_k given by

$$d_k = \sum_l a_l h(kT - lT). \qquad (36)$$

As an example, consider the following partial-response channel occurring in digital magnetic recording applications, denoted as the PR-IV channel. Full-response equalization leads to noise enhancement for longitudinal recording channels due to the spectral null around zero frequency. For this reason, partial-response equalization targets are usually prefered. Equalization is carried out in two steps, where the first step equalizes the received channel waveform to the partial-response target. The target chosen is such that it is spectrally similar to the channel response, and also has relatively few taps. The first property reduces noise enhancement, and the second property allows for ML trellis-based detection. This technique is popularly known as the partial-response maximum-likelihood (PRML) technique [23]. It is customary to treat the first stage of equalization as part of the front end of the receiver, with the result that the equivalent discrete-time overall channel, combining the magnetic recording channel and the front-end filter, is the partial-response target chosen. For the PR-IV channel that has the pulse shape [23]

$$h(t) = \text{sinc}\left(\frac{t}{T}\right) - \text{sinc}\left(\frac{t-2T}{T}\right) \qquad (37)$$

the noiseless, perfectly timed samples are given by

$$d_k = a_k - a_{k-2}. \qquad (38)$$

Performance of the MM TED can be improved by using soft estimates $\tilde{d}_k$ in place of hard estimates $\hat{d}_k$ [24]. For the PR-IV channel, a memoryless soft slicer of the form [24]

$$\tilde{d}_k = E[d_k|r_k] = \frac{2\sinh(2r_k/\sigma^2)}{\cosh(2r_k/\sigma^2) + e^{2/\sigma^2}} \qquad (39)$$
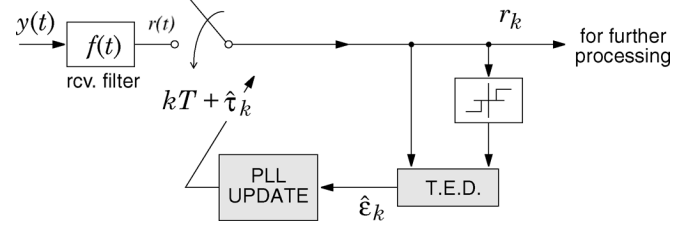

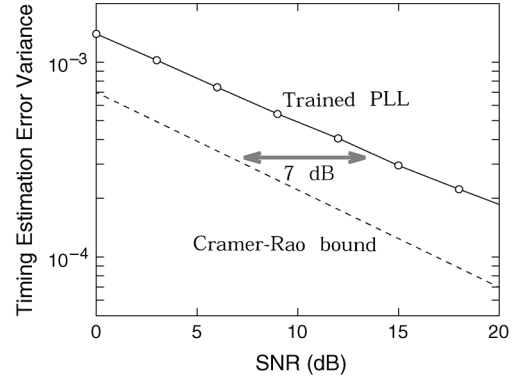Fig. 2.   Conventional timing recovery employs a PLL.

Fig. 3.   Trained PLL is 7 dB away from the CRB.

leads to reduced timing-estimation-error variance. With a hard slicer for the PR-IV channel, the estimates $\hat{d}_k$ take on one of three values, $0, \pm 2$. With the soft slicer, as $r_k \to \infty, \tilde{d}_k \to 2$, and as $r_k \to -\infty, \tilde{d}_k \to -2$. When $r_k = 0$, we have $\tilde{d}_k = 0$, as well. However, when the noise is such that r_k is near its decision thresholds, the hard quantizer is more likely to make errors. The soft slicer outputs a soft estimate of d_k that conveys the reliability information, as well, and this additional information helps in reducing the timing-estimation-error variance.

The conventional timing-recovery method is summarized in Fig. 2. Fig. 3 compares the performance of the trained PLL with the CRB for the PR-IV channel. For simplicity, the timing-offset model was chosen to be a plain random walk characterized by $\sigma_w/T = 0.33\%$, $\tau_0 = 0$, and $\Delta T = 0$. Fig. 3 plots the steady-state CRB of (27) versus the timing-estimation-error of the trained PLL averaged over all symbol positions. A first-order PLL is chosen, with the gain parameter α chosen at each SNR to minimize the timing-estimation-error variance [17]. The results are averaged over 1000 blocks of length 500 each. The PLL with ideal decisions $(\hat{d}_k = d_k)$ is about 7 dB away from the CRB, and this gap motivates the PLL postprocessor proposed in the next section.

V. PLL POSTPROCESSOR

It has been shown in [18] and [25] that the PLL can be modeled as a modified Kalman filter. For a linear Gaussian system, the Kalman filter is the optimal *causal* processing to minimize the mean-square estimation-error variance, and it has been shown [17], [18], [25] to be equivalent to a PLL with time-varying gain parameters. In the tracking mode, the gains are time-invariant, and the Kalman filter reduces to a PLL with constant gains [17]. Therefore, to outperform the constant-gain PLL in the tracking mode, it is necessary to relax

the causality constraint and to perform block processing. In the rest of this section, we develop a simplified linear model for the PLL outputs, and propose a PLL postprocessor based on MAP estimation for the linear model.

For any TED, we can write

$$\hat{\epsilon}_k = \epsilon_k + \nu_k \tag{40}$$
$$= \tau_k - \hat{\tau}_k + \nu_k \tag{41}$$

where ν_k is the TED's measurement noise, which could depend on $\{\epsilon_k\}$ is a nonlinear fashion. Constructing a new observation $y_k = \hat{\tau}_k + \hat{\epsilon}_k$ leads to the following observation model:

$$y_k = \tau_k + \nu_k. \tag{42}$$

An ideal TED would output $\hat{\epsilon}_k = \epsilon_k$. In the presence of noise, a linear TED would output $\hat{\epsilon}_k = \epsilon_k + \nu_k$, where ν_k is independent of ϵ_k. Practical TEDs are linear only in an SNR-dependent range around $\epsilon_k = 0$. However, in the tracking mode, the operating ϵ_k values are usually small enough for the linearity assumption to hold.

The PLL outputs an estimate of τ_k based on only the prior observations $\{r_l\}_0^{k-1}$. This is, therefore, an *a priori* estimate of τ_k. To this, $\hat{\epsilon}_k$, which is based on the present observation r_k, as well, is added. Therefore, y_k is an *a posteriori* estimate of τ_k, and the quality of this estimate depends on the quality of the TED output.

In vector notation, the measurement model can be written as follows:

$$\boldsymbol{y} = \boldsymbol{\tau} + \boldsymbol{\nu} \tag{43}$$

where $\boldsymbol{y} = [y_0 \ y_1 \ \ldots \ y_{N-1}]^T$, $\boldsymbol{\tau} = [\tau_0 \ \tau_1 \ \ldots \ \tau_{N-1}]^T$, and $\boldsymbol{\nu} = [\nu_0 \ \nu_1 \ \ldots \ \nu_{N-1}]^T$.

To get the linear measurement model, it is assumed that $\boldsymbol{\nu}$ is independent of $\boldsymbol{\tau}$, and also that $\boldsymbol{\nu}$ is $\mathcal{N}(\boldsymbol{0}, \sigma_n^2 \boldsymbol{I})$, where σ_n^2 is the TED error variance. To get the statistics of $\boldsymbol{\tau}$, consider the general case where the timing offsets consist of an initial phase offset τ_0, a frequency offset characterized by ΔT, and a random walk characterized by $\{w_k\}$, $k = 1, \ldots, N - 1$. Let $\boldsymbol{\phi} = [\Delta T \ \tau_0 \ w_1 \ \ldots \ w_{N-1}]^T$. Then

$$\boldsymbol{\tau} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 \\ 2 & 1 & 1 & 1 & \ddots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ N-1 & 1 & 1 & 1 & \cdots & 1 \end{bmatrix} \boldsymbol{\phi} = \boldsymbol{M}\boldsymbol{\phi}. \tag{44}$$

$\boldsymbol{\phi}$ is assumed to be $\mathcal{N}(\boldsymbol{0}, \boldsymbol{K}_\phi)$ where $\boldsymbol{K}_\phi = \mathrm{diag}(\sigma_{\Delta T}^2, \sigma_{\tau_0}^2, \sigma_w^2, \ldots, \sigma_w^2)$. Therefore, $\boldsymbol{\tau}$ is Gaussian, as well, with a covariance matrix given by [4]

$$\boldsymbol{K}_\tau = \boldsymbol{M}\boldsymbol{K}_\phi \boldsymbol{M}^T. \tag{45}$$
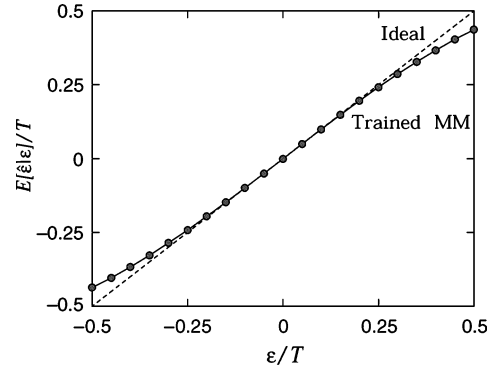

Fig. 4. S-curve for the trained MM TED on the PR-IV channel.

Evaluating this leads to

$$[\boldsymbol{K}_\tau]_{ij} = ij\sigma_{\Delta T}^2 + \sigma_{\tau_0}^2 + \min(i,j)\sigma_w^2 \tag{46}$$

for $i, j = 0, 1, \ldots, N - 1$. The MAP estimator for this linear Gaussian model is [4]

$$\hat{\boldsymbol{\tau}}_{\mathrm{map}}(\boldsymbol{y}) = \left(\boldsymbol{K}_\tau + \sigma_n^2 \boldsymbol{I}\right)^{-1} \boldsymbol{K}_\tau \boldsymbol{y}. \tag{47}$$

An alternative representation of this estimator is using the eigen-decomposition of $\boldsymbol{K}_\tau$. Let $\{\boldsymbol{W}, \boldsymbol{\Lambda}\}$ represent this decomposition such that $\boldsymbol{K}_\tau = \boldsymbol{W}^T \boldsymbol{\Lambda} \boldsymbol{W}$, where $\boldsymbol{\Lambda}$ is a diagonal matrix, with the main diagonal entries being $\{\lambda_i\}_{i=1}^N$, the N eigenvalues of $\boldsymbol{K}_\tau$, and $\boldsymbol{W}$ is a unitary matrix with the ith row being the eigenvector of $\boldsymbol{K}_\tau$ corresponding to the eigenvalue λ_i. The MAP estimator can then be rewritten as

$$\hat{\boldsymbol{\tau}}_{\mathrm{map}}(\boldsymbol{y}) = \boldsymbol{W}^T \left(\boldsymbol{\Lambda} + \sigma_n^2 \boldsymbol{I}\right)^{-1} \boldsymbol{\Lambda} \boldsymbol{W} \boldsymbol{y}. \tag{48}$$

The MAP estimator for this linear Gaussian system is efficient, i.e., it achieves the CRB for the linear model of (43) [4]. Let $\boldsymbol{\epsilon} = \boldsymbol{\tau} - \hat{\boldsymbol{\tau}}_{\mathrm{map}}(\boldsymbol{y})$ be the estimation error. The error variance for the MAP estimator is given by [4]

$$E[\boldsymbol{\epsilon}^T \boldsymbol{\epsilon}] = \sigma_n^2 \sum_{i=1}^N \frac{\lambda_i}{\lambda_i + \sigma_n^2}. \tag{49}$$

The error variance of the PLL postprocessor is a function of the TED noise variance σ_n^2. To evaluate σ_n^2, the so-called S-curve is used [22], [23], which is the plot of the average timing-error estimate $(E[\hat{\epsilon}|\epsilon])$ as a function of the actual timing error (ϵ). Ideally, this is a straight line of unit slope passing through the origin.

The normalized timing function (i.e., $E[\hat{\epsilon}|\epsilon]/T$ versus ϵ/T) (labeled "Ideal" and "Trained MM," corresponding to the ideal timing function and that of the trained MM TED, respectively) for the PR-IV channel is plotted in Fig. 4 for SNR = 10 dB. The timing function for the MM TED defined by (35) does not have unit slope at origin, as desired. Therefore, it is normalized
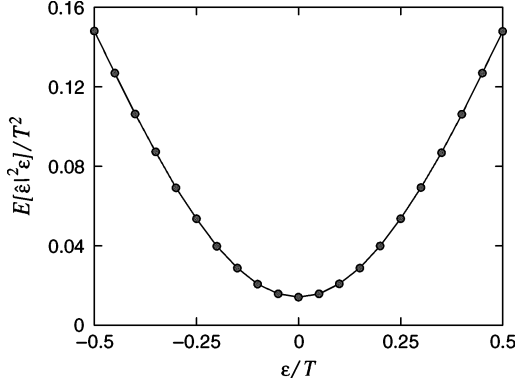
Fig. 5. Second moment of timing error for the trained MM TED on the PR-IV channel.
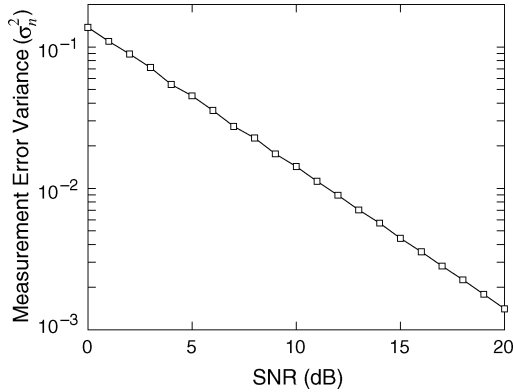

Fig. 6. MM TED measurement error variance.

Fig. 7. Postprocessor timing estimates can be used to refine samples.
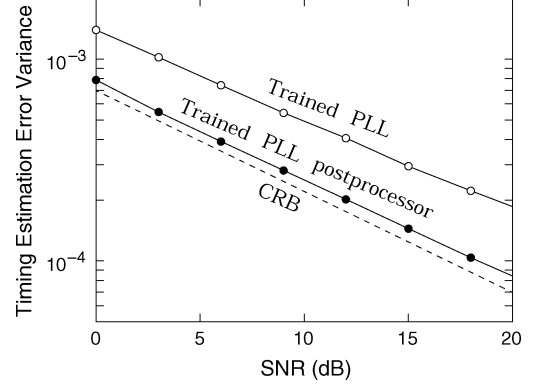
Fig. 8. PLL postprocessor 1.5 dB away from the steady-state CRB.

by a multiplicative factor. For the PR-IV channel, this factor is $3T/16$, and the TED equation now is

$$\hat{\epsilon}_k = \frac{3T}{16}(r_k \hat{d}_{k-1} - r_{k-1} \hat{d}_k). \tag{50}$$

The timing function plotted in Fig. 4 is for this normalized MM TED, and this is close to the ideal timing function for a wide range of ϵ/T.

In Fig. 5, the second moment of the trained MM TED timing error is plotted as a function of the normalized timing error ϵ/T for SNR = 10 dB for the PR-IV system. The second moment is not independent of ϵ. Also, from Fig. 4, it is clear that the estimator is not unbiased over the entire range of ϵ. To simplify the analysis, unbiasedness of the TED is assumed. Also, the value of the timing-error second moment at the origin is picked as σ_n^2, i.e., $\sigma_n^2 = E[\hat{\epsilon}^2|\epsilon]_{\epsilon=0}$. This is a good approximation to the actual TED measurement noise in the tracking mode, where ϵ/T is small, and also gives a lower bound on the actual TED measurement noise, since $E[\hat{\epsilon}^2|\epsilon]_{\epsilon=0} < E[\hat{\epsilon}^2|\epsilon]_{\epsilon \neq 0}$.

Fig. 6 plots the measurement-error variance σ_n^2 for the trained MM TED for the PR-IV channel as a function of SNR. The measurement-error variance monotonically decreases with SNR, as expected. These values of σ_n^2 are finally used in (48) to completely define the postprocessor. The refined timing estimates
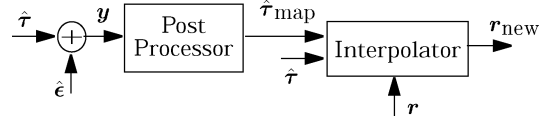
$\hat{\tau}_{\mathrm{map}}$ from the postprocessor can then be use to get new samples r_{new} by interpolating based on samples r taken at the PLL's timing estimates $\hat{\tau}$, as shown in Fig. 7.

Fig. 8 compares the performance of the trained PLL postprocessor with the trained PLL performance and the steady-state CRB for a random walk timing-offset model with $\sigma_w/T = 0.33\%$, averaged over 1000 blocks of length $N = 500$. The channel model is the same as before, namely, the PR-IV channel. The postprocessor performs to within 1.5 dB of the CRB, and is around 5.5 dB better than the trained PLL. This 1.5 dB gap has to be put in perspective by the fact that the CRB is not attainable for the overall nonlinear system with a PR-IV channel and a random walk timing model [21], and therefore, the performance loss of the postprocessor, compared with the minimum-error-variance estimator is less than 1.5 dB. An additional contributor to the performance gap is the inaccuracy of the linear model, due to the limited linear range of the TED.

VI. SUBOPTIMAL LOW-COMPLEXITY IMPLEMENTATIONS

The postprocessor described above involves matrix operations, and becomes infeasible, or at least computationally burdensome, for reasonable block lengths of around 5000, which are common in the magnetic recording industry. In this section, suboptimal implementations of the postprocessor are proposed that allow a tradeoff between performance and complexity.

First consider the MAP estimator error-covariance matrix $K_{\epsilon} = E[\epsilon\epsilon^T]$. In terms of the eigendecomposition, this evaluates to

$$K_{\epsilon} = \sigma_n^2 W^T \left(\Lambda + \sigma_n^2 I\right)^{-1} \Lambda W. \tag{51}$$

From (48) and (51)

$$\hat{\tau}_{\mathrm{map}}(y) = \frac{1}{\sigma_n^2} K_{\epsilon} y. \tag{52}$$
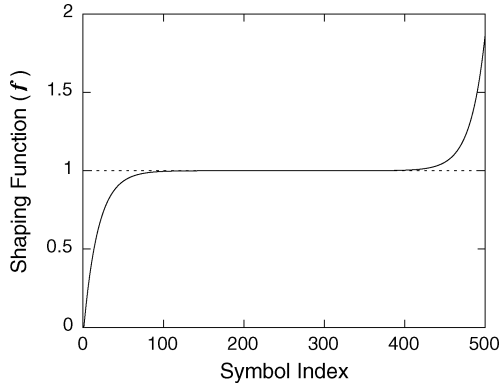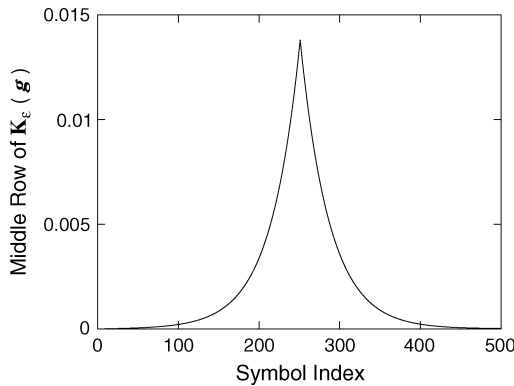
Fig. 9. Shaping function has steady-state value of unity.

Fig. 10. Except for shifting and scaling, this is a typical row of $\boldsymbol{K_\epsilon}$.

Let $\boldsymbol{f_1}$ be the $N \times 1$ vector containing the elements of the main diagonal of $\boldsymbol{K_\epsilon}$, and let $\boldsymbol{f} = \boldsymbol{f_1}/\boldsymbol{f_1}(N/2)$. (It is assumed that N is even. For odd N, pick $\lfloor N/2 \rfloor$.) Call $\boldsymbol{f}$ the shaping function. Fig. 9 shows the shaping function for the PR-IV channel with a random walk timing offset characterized by $\sigma_w/T = 0.33\%$, SNR = 10 dB, and $N = 500$.

Fig. 10 plots row 250 of $\boldsymbol{K_\epsilon}$ with same parameters as above. In general, let $\boldsymbol{g} = [g_1 \; g_2 \; \cdots \; g_N]$ represent the $(N/2)$th row of $\boldsymbol{K_\epsilon}$. (It is again assumed that N is even. For odd N, pick $\lfloor N/2 \rfloor$.) The first approximation is to replace the other rows of $\boldsymbol{K_\epsilon}$ by suitably zero-padded, shifted, and scaled versions of $\boldsymbol{g}$. Specifically, denote the ith row of $\boldsymbol{K_\epsilon}$ by row_i, and let $i < N/2$. Let $\mathrm{shift}(\boldsymbol{x}, k, \mathrm{dir})$ denote the operator whose output is the vector formed by suitably zero-padding the vector $\boldsymbol{x}$ and shifting it by k units to the left or right, depending on whether $\mathrm{dir} = \mathrm{left}$ or $\mathrm{dir} = \mathrm{right}$, respectively. Then

$$\mathrm{row}_i \approx f(i) \times \mathrm{shift}\left(\boldsymbol{g}, \frac{N}{2} - i, \mathrm{left}\right). \qquad (53)$$

The case for $i > N/2$ is similar.

Essentially, the postprocessor has been approximated by the following structure:

$$\hat{\tau}_{\mathrm{map}}(\boldsymbol{y}) \approx \boldsymbol{A_1}\boldsymbol{A_2}\boldsymbol{y} \qquad (54)$$
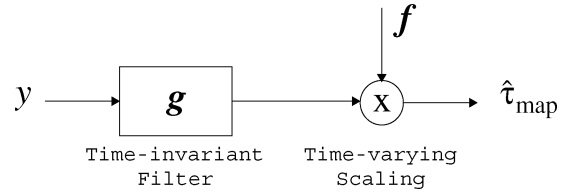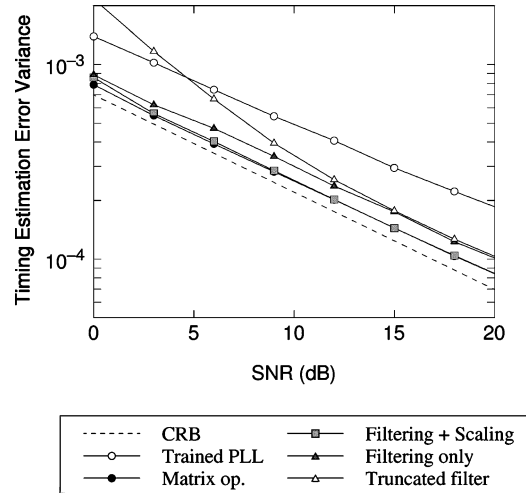

Fig. 11. Approximate PLL postprocessor.

Fig. 12. Various suboptimal strategies.

where $\boldsymbol{A_1}$ is a diagonal matrix with the ith main diagonal entry being $f(i)$, and $\boldsymbol{A_2}$ is the matrix whose rows are the shifted rows defined by (53) neglecting the multiplicative factor $f(i)$. This simplifies the implementation greatly, since $\boldsymbol{A_2}$ represents a convolution matrix, and can be implemented as a time-invariant filter whose impulse response is $\boldsymbol{g}$. $\boldsymbol{A_1}$ can be implemented as time-varying scaling of the filter output. The approximate postprocessor is shown in Fig. 11.

To reduce complexity further, the factor $\boldsymbol{A_1}$ can be neglected altogether, and also the filter can be truncated to a manageable length. The losses associated with these simplifications are shown in Fig. 12 for the PR-IV channel with a random walk timing-offset model with $\sigma_w/T = 0.33\%$, and $N = 500$. The CRB shown is the steady-state value h_{ss}. The error variances plotted for the different methods being compared are arrived at by a time average of all the error variances corresponding to the N timing offsets.

The approximation of (54) performs as well as the matrix operation of (48) leading to significant reduction in the memory requirement. As opposed to the N^2 memory elements needed for the matrix, now only $2N$ elements are needed. Neglecting the shaping function and implementing only the filtering reduces the memory requirement to N, but, more importantly, allows a time-invariant filter implementation. This approximation leads to a 1.5 dB loss. Further reduction in complexity by truncating the filter to a total of 100 terms, as opposed to 500, leads to a further loss of less than 0.5 dB at $\sigma_\epsilon^2/T^2 = 2 \times 10^{-4}$, still around 4 dB better than the PLL. Also, this approximation shows a crossover with respect to the performance of the PLL for SNR around 5 dB, i.e., for SNR < 5 dB, this performs worse than the PLL. This crossover threshold varies with the number of filter coefficients used.

VII. Conclusion

The problem of timing recovery for a baud-rate sampled, bandlimited channel that suffers from ISI and timing offsets is considered. For a general timing-offset model consisting of a frequency offset and a random walk, the CRB is derived. This is a lower bound on the timing-estimation-error variance for any estimator. The conventional PLL performs about 7 dB away from the CRB in simulations with realistic system parameters for a magnetic recording channel. Next, a PLL postprocessor is proposed as a method of gaining better performance when compared with conventional PLL-based timing recovery. This postprocessor is based on MAP estimation for a linear model of the PLL outputs, and takes the form of a matrix operation on timing estimates from the PLL. The postprocessor is then simplified to take the form of a time-invariant filtering, followed by a time-varying scaling. Finally, suboptimal implementations are presented that allow a performance-complexity tradeoff.

Appendix I
Proof of (12)

We start by defining the accumulation process where the timing offsets are as follows:

$$\tau_{k+1} = \tau_k + \delta_k = \sum_{j=0}^{k} \delta_j \qquad (55)$$

where the sequence $\{\tau_k\}$ is formed by accumulating $\{\delta_k\}$. If we assume that the elements of $\{\delta_k\}$ are i.i.d. zero-mean normal random variables, we get the random walk model. For the model of (2), we set

$$\delta_k = \begin{cases} \tau_0 + w_0 : & k = 0 \\ w_k + \Delta T : & k \neq 0. \end{cases} \qquad (56)$$

For the present, we assume that we have no *a priori* information about the variables $\{\delta_k\}$. The channel model with uniform sampling is

$$\begin{aligned} r_k &= \sum_{l=0}^{N-1} a_l h(kT - lT - \tau_l) + n_k \\ &= \sum_{l=0}^{N-1} a_l h\left(kT - lT - \sum_{j=0}^{l-1}\delta_j\right) + n_k \\ &= x_k + n_k \end{aligned} \qquad (57)$$

where x_k is the noiseless received value. The parameter to be estimated is $\boldsymbol{\tau} = [\tau_1 \ \tau_2 \ \ldots \ \tau_{N-1}]^T$. Instead of directly computing the Fisher information matrix $\boldsymbol{J}_\tau$, we define the parameter $\boldsymbol{\delta} = [\delta_0 \ \delta_1 \ \ldots \ \delta_{N-2}]^T$, compute $\boldsymbol{J}_\delta$, and then use the linear transformation relating $\boldsymbol{\tau}$ and $\boldsymbol{\delta}$ to get $\boldsymbol{J}_\tau$.

We proceed with evaluating the *sensitivity* matrix $\boldsymbol{G}$ of (10) for the parameter $\boldsymbol{\delta}$. We have

$$g_{ki} = \frac{\partial x_k}{\partial \delta_i} = \sum_{l=0}^{N-1} a_l \frac{\partial}{\partial \delta_i} h\left(kT - lT - \sum_{j=0}^{l-1}\delta_j\right). \qquad (58)$$

The derivative of $h(kT - lT - \sum_{j=0}^{l-1}\delta_j)$ with respect to δ_i is zero if δ_i does not occur in the summation inside the $h(\cdot)$ term. Therefore

$$\begin{aligned} g_{ki} &= - \sum_{l=i+1}^{N-1} a_l h'\left(kT - lT - \sum_{j=0}^{l-1}\delta_j\right) \\ &= - \sum_{l=i+1}^{N-1} a_l h'(kT - lT - \tau_l). \end{aligned} \qquad (59)$$

Next, $\boldsymbol{G}^T\boldsymbol{G}$ is given by

$$[\boldsymbol{G}^T\boldsymbol{G}]_{i_1 i_2} = \sum_{k=-M}^{N+M-1} g_{ki_1} g_{ki_2}. \qquad (60)$$

To get the averaged Fisher information, we need to take the expectation of $\boldsymbol{G}^T\boldsymbol{G}$ over all data sequences $\boldsymbol{a}$. Assuming uncorrelated data symbols in (60) and simplifying, we get

$$E[\boldsymbol{G}^T\boldsymbol{G}]_{i_1 i_2} = \sum_{j=\max(i_1,i_2)+1}^{N-1} \left\{ \sum_{k=-M}^{N+M-1} [h'(kT - jT - \tau_j)]^2 \right\} \qquad (61)$$

where the expectation is over $\boldsymbol{a}$. Letting $M \to \infty$, the bracketed term is the energy in $h'(t)$ normalized by the symbol duration T, denoted by $E_{h'}/T$. Therefore

$$E[\boldsymbol{J}_\delta]_{i_1 i_2} = \frac{E_{h'}}{\sigma^2 T} \left((N-1) - \max(i_1,i_2)\right). \qquad (62)$$

To get $\boldsymbol{J}_\tau$ from $\boldsymbol{J}_\delta$, we use the relationship between $\boldsymbol{\tau}$ and $\boldsymbol{\delta}$. Recall that $\tau_k = \sum_{j=0}^{k-1} \delta_j$. In vector notation

$$\boldsymbol{\tau} = \boldsymbol{T}\boldsymbol{\delta} \qquad (63)$$

where

$$T_{ij} = \begin{cases} 1 : & i \leq j \\ 0 : & i > j. \end{cases} \qquad (64)$$

Using this relationship between $\boldsymbol{\tau}$ and $\boldsymbol{\delta}$, we can write [19]

$$\boldsymbol{J}_\tau = \boldsymbol{T}^{-T} \boldsymbol{J}_\delta \boldsymbol{T}^{-1}. \qquad (65)$$

To compute $\boldsymbol{T}^{-1}$, recognize the fact that we need now the inverse mapping of $\tau_k = \sum_{j=0}^{k-1} \delta_j$, which is $\delta_k = \tau_{k+1} - \tau_k$. Therefore

$$T_{ij}^{-1} = \begin{cases} 1 : & i = j \\ -1 : & i = j+1 \\ 0 : & \text{else}. \end{cases} \qquad (66)$$

$\boldsymbol{T}^{-1}$ is a lower triangular Toeplitz matrix with only two nonzero diagonals, the main diagonal and an adjacent one. Using (62), (65), and (66), we get

$$E[\boldsymbol{J}_\tau] = \frac{E_{h'}}{\sigma^2 T} \boldsymbol{I}_{N \times N} \qquad (67)$$

which is the same as (12).

ACKNOWLEDGMENT

REFERENCES

[1] P. Ciblat, Y. Wu, and E. Serpedin, "On a blind fractionally sampling based carrier frequency offset estimator," *IEEE Signal Process. Lett.*, vol. 10, no. 4, pp. 89–92, Apr. 2003.

[2] F. Musa and A. Carusone, "Clock recovery in high-speed multilevel serial links," in *Proc. Int. Symp. Circuits Syst.*, May 2003, vol. 5, pp. V449–V452.

[3] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, 1st ed. New York: Wiley, 1971, vol. 2, ch. 3, pp. 37–84.

[4] ——, *Detection, Detection, Estimation, and Modulation Theory*, 1st ed. New York: Wiley, 1968, vol. 1, ch. 2, pp. 72–85.

[5] H. Meyr, M. Moeneclaey, and S. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation and Signal Processing*. New York: Wiley, 1997.

[6] D. C. Rife and R. R. Boorstyn, "Single tone parameter estimation from discrete time observations," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 5, pp. 591–598, Sep. 1974.

[7] J. Riba, J. Sala, and G. Vásquez, "Conditional maximum likelihood timing recovery: Estimators and bounds," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 835–850, Apr. 2001.

[8] A. Nayak, J. Barry, and S. McLaughlin, "Joint timing recovery and turbo equalization for coded partial response channels," *IEEE Trans. Magn.*, vol. 38, no. 5, pp. 2295–2297, Sep. 2002.

[9] M. Oerder and H. Meyr, "Digital filter and square timing recovery," *IEEE Trans. Commun.*, vol. 36, no. 5, pp. 605–612, May 1988.

[10] S. Lee, "A new non-data-aided feedforward symbol timing estimator using two samples per symbol," *IEEE Commun. Lett.*, vol. 6, no. 5, pp. 205–207, May 2002.

[11] Y. Wu and E. Serpedin, "Low-complexity feedforward symbol timing estimator using conditional maximum-likelihood principle," *IEEE Commun. Lett.*, vol. 8, no. 3, pp. 168–170, Mar. 2004.

[12] R. Gallager, "Low-density parity-check codes," *IRE Trans. Inf. Theory*, vol. IT-8, no. 1, pp. 21–28, Jan. 1962.

[13] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes," in *Proc. IEEE Int. Conf. Commun.*, May 1993, vol. 2, pp. 1064–1070.

[14] D. Raphaeli and Y. Zarai, "Combined turbo equalization and turbo decoding," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 1997, vol. 2, pp. 639–643.

[15] F. L. Lewis, *Optimal Estimation*. New York: Wiley, 1986.

[16] S. J. Julier and J. K. Uhlmann, "The scaled unscented transformation," in *Proc. IEEE Amer. Control Conf.*, Anchorage, AK, May 2002, pp. 4555–4559.

[17] A. Patapoutian, "On phase-locked loops and Kalman filters," *IEEE Trans. Commun.*, vol. 47, no. 5, pp. 670–672, May 1999.

[18] P. F. Driessen, "DPLL bit synchronizer with rapid acquisition using adaptive Kalman filtering techniques," *IEEE Trans. Commun.*, vol. 42, no. 9, pp. 2673–2675, Sep. 1994.

[19] L. L. Scharf and L. McWhorter, "Geometry of the Cramér–Rao bound," *Signal Process.*, vol. 31, no. 3, pp. 1–11, Apr. 1993.

[20] A. Nayak, "Iterative timing recovery for magnetic recording channels with low signal-to-noise ratio," Ph.D. dissertation, School Elect. Comput. Eng., Georgia Inst. Technol., Atlanta, GA, Aug. 2004.

[21] A. Nayak, J. Barry, and S. McLaughlin, "Lower bounds for the performance of iterative timing recovery at low SNR," in *Proc. 15th Int. Symp. Math. Theory Netw. Syst.*, Aug. 2002, pp. WM2–WM5.

[22] K. Mueller and M. Müller, "Timing recovery for digital synchronous data receivers," *IEEE Trans. Commun.*, vol. COM-24, no. 5, pp. 516–531, May 1976.

[23] R. Cideciyan, F. Dolivo, R. Hermann, W. Hirt, and W. Schott, "A PRML system for digital magnetic recording," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 1, pp. 38–56, Jan. 1992.

[24] A. Nayak, J. Barry, and S. McLaughlin, "Timing recovery at low SNR," Inf. Storage Ind. Consortium (INSIC), Mar. 2000, Tech. Rep..

[25] G. S. Christiansen, "Modeling of a PRML timing loop as a Kalman filter," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 1994, vol. 2, pp. 1157–1161.

Aravind R. Nayak (S'01–M'04) received the B. Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, India, in 1999, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, in 2000 and 2004, respectively.

His research interests are in the areas of communications and signal processing. He is currently a member of the Magnetic Read Channel Architecture Team with Agere Systems, Longmont, CO.

John R. Barry (S'87–M'92–SM'04) received the B.S. degree in electrical engineering from the State University of New York, Buffalo, in 1986, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1987 and 1992, respectively.

Since 1992, he has been with the Georgia Institute of Technology, Atlanta, where he is a Professor with the School of Electrical and Computer Engineering. His research interests include wireless communications, equalization, and multiuser communications. He is a coauthor, with E. A. Lee and D. G. Messerschmitt, of *Digital Communication* (Norwell, MA: Kluwer, 2004, 3rd ed.), and the author of *Wireless Infrared Communications* (Norwell, MA: Kluwer, 1994).

German Feyh (S'85–M'89) received the Diplom Ingenieur der Elektrotechnik from the Friedrich Alexander University of Erlangen-Nuremberg, Erlangen, Germany, and the M.S. and Ph.D. degrees in electrical engineering from the University of Colorado, Boulder.

He has worked in telephone communications for TeKaDe, Nuremberg, Germany, in signal surveillance for DASA, Ulm, Germany, and architecting magnetic and optical read channels for Cirrus Logic and Infineon. He is currently the Manager of the Magnetic Read Channel Team at Agere Systems, Longmont, CO.

Steven W. McLaughlin (M'88–SM'00–F'04) received the B.S. degree from Northwestern University, Evanston, IL, in 1985, the M.S.E. degree from Princeton University, Princeton, NJ, in 1986, and the Ph.D. degree from the University of Michigan, Ann Arbor, in 1992, all in electrical engineering.

From 1992–1996, he was with the Electrical Engineering faculty at the Rochester Institute of Technology, Rochester, NY. He joined the School of Electrical and Computer Engineering (ECE) at Georgia Institute of Technology (Georgia Tech), Atlanta, in September 1996, where he is now the Ken Byers Professor of ECE. He is also Deputy Director of Georgia Tech Lorraine, the European campus of Georgia Tech, Metz, France. His research interests are in the general area of communications and information theory. His research group has on-going projects in the areas of turbo, LDPC, and constrained codes for magnetic and optical recording; FEC and equalization for wireless and optical networks, quantum key distribution and data security; and theory of error-control coding. He has published more than 200 papers in refereed journals and conferences and holds 26 U.S. patents in these areas.

In 2005, Dr. McLaughlin was President of the IEEE Information Theory Society. He received the Presidential Early Career Award for Scientists and Engineers (PECASE) in 1997, where he was cited by President Clinton "for leadership in the development of high-capacity, nonbinary optical recording formats." He also received the National Science Foundation CAREER award for this work. He received (with Dr. David Warland at UC Davis) the Information Storage Industries Consortium Technical Achievement Award in 2002 for "pioneering work in the development of multilevel optical disk storage technology." From 1999–2003, he was also a Principal Scientist for Calimetrics where this work was commercialized. He received the Friend of the Graduate Student Award in 2002 from the GT Graduate Student Association. He served as the Publications Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY from 1995–1999, and co-edited (with Sergio Verdú) *Information Theory: 50 Years of Discovery* (New York: Wiley/IEEE Press, 1999).