

Optimal Placement of Training Symbols for Frequency Acquisition: A Cramér-Rao Bound Approach

Aravind R. Nayak¹, John R. Barry², and Steven W. McLaughlin², *Fellow, IEEE*

¹Agere Systems, Longmont, CO 80501 USA

²Georgia Institute of Technology, Atlanta, GA 30332 USA

We consider the problem of training symbol placement for timing acquisition in digital magnetic recording systems that can be modeled as band-limited, baud-rate sampled systems with intersymbol interference and frequency offset. The conventional approach is to place known symbols at the start of the sector and use these at the detector to run a trained phase-locked loop (PLL). We introduce an additional degree of freedom by allowing arbitrary locations for the training symbols. We first consider a simplified system model where only the training symbols are written, or equivalently, the unknown data is assumed to be zero. We derive the modified Cramér-Rao bound (CRB) on the timing estimation error variance as a function of the training symbol locations and then derive the optimal training symbol placement strategy to minimize the CRB. The optimal strategy, called the split-preamble strategy, is to split the known symbols into two halves and place these at the beginning and at the end of the sector. Simulations with a proposed PLL-based method show that the split-preamble arrangement leads to a reduced frequency estimation error variance and also greatly reduces the occurrence of lost or added symbols, i.e., cycle slips. Finally, we present a simplified analysis of the problem when both known and unknown data are present, and show that the same arrangement also minimizes the CRB in this case.

Index Terms—Cramér-Rao bound, frequency offset, phase-locked loop, timing acquisition, training symbol placement.

I. INTRODUCTION

TIMING acquisition is an important component of digital magnetic storage systems. During the read process, the detector usually has nominal knowledge of the symbol duration, and no knowledge of the instant the desired sector starts in the readback waveform. During acquisition, start of sector is detected and the detector's estimate of the symbol duration is refined. A common method to detect start of sector is by using correlation detection. The goal is to have the residual timing offset within half a symbol duration. Next, known symbols at the start of the sector, also called preamble symbols, are used to operate a phase-locked loop (PLL) in a trained fashion, leading to improved timing estimates at the detector. At the end of acquisition, the detector has a better estimate of the instant the sector started and also of the symbol duration. The conventional strategy of placing all the known symbols at the start of the sector is motivated by the necessity of having good timing estimates *before* the user data begins to ensure low enough error rates (for an example, see [1]). This, however, is not necessarily the optimal training symbol placement strategy from the point of view of the mean-squared error performance of the timing estimator, especially when the causality constraint can be relaxed. An example of such a situation is when the recording system uses iterative processing [2]–[4], in which case the samples from a previous iteration can be used for off-line processing to get better timing estimates before the next iteration.

The optimality of training symbol placement has been considered in the communications literature with respect to channel estimation [5]–[9] where the optimization criteria considered are

the mean squared channel estimation error and the channel capacity assuming independent and identically distributed (i.i.d.) input symbols. Depending on the particular channel setting and the particular performance metric chosen, the optimal strategy varies. With finite impulse response fading channels, the optimal training strategy consists of arranging the known symbols in small contiguous clusters placed periodically in the data stream [5], [6]. These training symbol clusters are chosen to be as *equal* as possible, subject to the power constraint on training, and also to number of taps of the channel impulse response. In [8], channels that are both time and frequency selective are considered. It is shown that to minimize the mean squared channel estimation error, the optimal training strategy is to use equispaced and equi-powered training symbols. In [9], multiple access communication systems undergoing fading are considered. To minimize the influence of asynchronous interference on the packets of one user from those of other users, it is shown that the best placement strategy is to use two clusters of equal or quasiequal size at the ends of the packet.

In this paper, the problem of timing acquisition is considered for band-limited, baud-rate sampled channels that face inter-symbol interference and a frequency offset. To improve acquisition, the constraint of having all the training symbols at the start of the sector is relaxed, allowing arbitrary locations for the training symbols. The metric of comparison used is the mean-squared timing estimation error (MSE). First, a simplified system model is considered, where only the training symbols are assumed to be written. Equivalently, the unknown data is assumed to be zero. The modified Cramér-Rao bound (CRB) [10] on the timing estimation error variance is derived as a function of preamble locations, assuming a frequency offset timing model. Then, it is shown that the CRB and the MSE are minimized by the split-preamble arrangement, where the

training symbols are split into two halves, and these halves are placed at the beginning and at the end of the sector. The split-preamble strategy has already been proposed for timing recovery in a patent by McEwen *et al.* [11]. The contribution of this paper is to prove the optimality of the split-preamble strategy and provide further analysis of its performance.

In addition to minimizing the MSE, the split-preamble strategy also helps reduce the occurrence of cycle slips when using PLL-based timing recovery. A cycle slip is said to have occurred when the detector's timing estimator locks on to a timing estimate that is offset from the actual timing instant by an integral multiple of the symbol duration. This leads to the detector adding or dropping whole symbols. Cycle slips are catastrophic for communication and data storage systems that employ error control codes (ECC), as the added or lost symbols almost always leads to the ECC decoder failure. By simulation, it is shown that the proposed training symbol placement strategy greatly reduces the occurrence of cycle slips with a proposed PLL-based timing recovery method.

Finally, instead of assuming that only known data was written, we account for the presence of both known and unknown data. For tractable analysis, a simplified linear channel model is derived, and the CRB is derived for frequency estimation using this model. Then, it is shown that the split-preamble arrangement minimizes the CRB on timing estimation error variance and the MSE in this case.

The proofs of most results in this paper have been presented in the Appendix to improve the presentation. The rest of the paper is organized as follows. In Section II, the system model is presented in detail, and the conventional acquisition method is reviewed. In Section III, the modified CRB on the timing estimation error variance is derived as a function of the training symbol placement assuming only known data being written. In Section IV, it is shown that the split-preamble strategy minimizes the CRB. In Section V, the mean square error performance of the split-preamble placement is compared to that of the conventional preamble placement, demonstrating the possible performance improvement. In Section VI, analytical and simulation results are presented demonstrating the MSE improvements due to the split-preamble strategy when using a proposed PLL-based timing recovery method. Also shown is an improvement in the cycle slip occurrence rate for the split-preamble strategy with the proposed PLL-based method when compared to using a trained PLL with all known symbols at start of the sector. In Section VII, a simplified analysis of the case with both known and unknown data is presented, where the split-preamble placement is shown to minimize the MSE. The conclusions are presented in Section VIII.

II. BACKGROUND: SYSTEM MODEL AND CONVENTIONAL ACQUISITION

Consider the pulse-amplitude modulated (PAM) system of Fig. 1 where the readback waveform $y(t)$ is given by

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

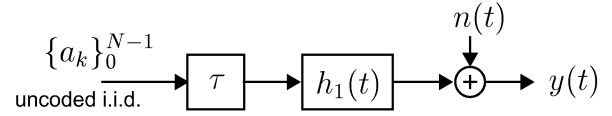


Fig. 1. System block diagram with timing offsets, channel distortion and additive noise.

where $a_l \in \{\pm 1\}$ are the N i.i.d. data symbols, $h_1(t)$ is the channel impulse response, $n(t)$ is additive white Gaussian noise, T is the receiver's estimate of the symbol duration and τ_l is the unknown timing offset for the l th symbol. A mismatch between T and the actual symbol duration T' leads to a frequency offset. For the frequency offset timing model, the timing offsets $\{\tau_k\}$ are given by $\tau_k = \tau_0 + k\Delta T$, where τ_0 is the initial timing offset, $\Delta T = T' - T$ is the symbol duration offset. The frequency offset present in the system is $\Delta f = 1/T' - 1/T$. For convenience and ease of notation, we characterize the timing offsets using ΔT , the symbol duration offset, instead of Δf , the frequency offset.

At the detector, $y(t)$ is filtered by a low-pass filter with impulse response $\frac{1}{T} \frac{\sin(\pi t/T)}{\pi t/T}$ to arrive at waveform $r(t)$. It is assumed that $h_1(t)$ is band-limited such that baud-rate samples are sufficient statistics to reconstruct $r(t)$.

In most systems, a certain number of known symbols are placed at the start of the sector to aid acquisition of the initial timing offset and the frequency offset. The first stage of acquisition deals exclusively with estimating the initial timing offset, or equivalently the start of the sector. At the detector, the read-back waveform is correlated with the waveform generated based on the known data and the channel model, and peak detection is used to detect the start of transmission. The symbols chosen for this section of the preamble have to be such that very high correlation is obtained when the waveforms match, and the correlation dies down rapidly as the waveforms get mismatched. The goal of this phase of acquisition is to get to within $[-0.5T, 0.5T)$ of the actual start of transmission, where T is the symbol duration. Depending on the channel and the noise conditions, we can do better than just get in the $[-0.5T, 0.5T)$ range and the residual timing error can be effectively modeled as a zero-mean Gaussian random variable. This stage of acquisition is common to both the conventional acquisition method and the acquisition method proposed in Section VI.

The next portion of the preamble is conventionally a periodic pattern that helps in acquiring the frequency offset. The periodic tone is spectrally compact and therefore, passes through the channel without distortion and can be detected easily at the detector. A common method is to use a trained phase-locked loop (PLL) to acquire the frequency.

The conventional PLL-based timing recovery method is shown in Fig. 2 [12]. At the detector the waveform $r(t)$ is sampled at instants $\{kT + \hat{\tau}_k\}$ according to the output of the timing recovery block to produce samples $\{r_k\}$ given by

$$r_k = \sum_{l=0}^{N-1} a_l h(kT + \hat{\tau}_k - lT - \tau_0 - l\Delta T) + n_k \quad (2)$$

where $\{n_k\}$ are i.i.d. zero-mean Gaussian random variables of variance σ^2 and $h(t) = h_1(t) * f(t)$.

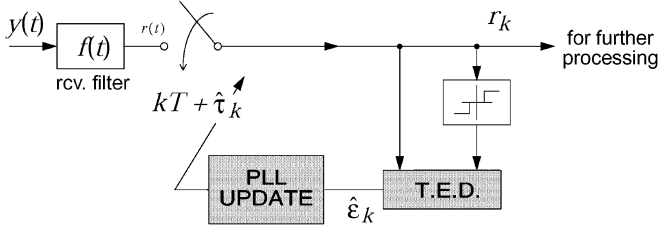


Fig. 2. Conventional timing recovery is based on the phase-locked loop.

A timing error detector (TED) operates on samples and transmitted symbols corresponding to these samples to estimate the timing error. The output of the TED is accumulated by a PLL. The TED is chosen according to the channel model and the modulation scheme, and the PLL structure is chosen according to the timing offset model. With a frequency offset, or equivalently a symbol duration offset, where the detector's estimate of the symbol duration is inaccurate, a second-order PLL is employed to have zero steady-state error [12]. A second-order PLL updates its timing estimate according to

$$\hat{\tau}_{k+1} = \hat{\tau}_k + \alpha \hat{\epsilon}_k + \beta \sum_{l=0}^{k-1} \hat{\epsilon}_l \quad (3)$$

where α and β are gain parameters, and $\hat{\epsilon}_k$ is the output of the TED, which estimates the timing error $\epsilon_k = \tau_k - \hat{\tau}_k$.

The widely used Mueller and Müller TED generates this estimate according to [13]:

$$\hat{\epsilon}_k = r_k d_{k-1} - r_{k-1} d_k \quad (4)$$

where d_k is the k th noiseless, perfectly-timed sample. With knowledge of the training symbols and also of the channel response, d_k can be constructed at the detector.

The trained PLL will operate satisfactorily whenever the residual phase offset it faces after the first stage of acquisition is small. The PLL gain is set reasonably high to allow for fast frequency acquisition. High gain increases the frequency range over which the PLL can lock in and also reduces the overhead by reducing the required length of the preamble. At the end of this second phase of acquisition, the residual frequency offset can also be modeled as a zero-mean Gaussian random variable.

III. CRAMÉR-RAO BOUND AS A FUNCTION OF THE TRAINING SYMBOL LOCATIONS

The CRB on the error variance of any unbiased estimator $\hat{\theta}(\mathbf{r})$ of an unknown deterministic parameter θ based on observations \mathbf{r} is given by

$$E[(\hat{\theta}_i(\mathbf{r}) - \theta_i)^2] \geq \mathbf{J}_\theta^{-1}(i, i) \quad (5)$$

where $\hat{\theta}_i(\mathbf{r})$ and θ_i denote the i th element of $\hat{\theta}(\mathbf{r})$ and θ respectively, the expectation is taken over \mathbf{r} , and $\mathbf{J}_\theta^{-1}(i, i)$ is the i th diagonal element of the inverse of the Fisher information matrix given by

$$\mathbf{J}_\theta = E \left\{ \left[\frac{\partial}{\partial \theta} \ln f_{\mathbf{r}|\theta}(\mathbf{r}|\theta) \right] \left[\frac{\partial}{\partial \theta} \ln f_{\mathbf{r}|\theta}(\mathbf{r}|\theta) \right]^T \right\} \quad (6)$$

where the expectation is over \mathbf{r} , and $f_{\mathbf{r}|\theta}(\mathbf{r}|\theta)$ is the probability density of \mathbf{r} conditioned on θ [14].

The specific problem considered in this section is as follows. For tractability, it is assumed that the unknown data symbols are zero, and that only the known data symbols are written. The K known data symbols are written at K arbitrarily chosen locations out of the possible N locations, where N is the sector length. This leads to the following PAM system model where the readback waveform is given by:

$$y(t) = \sum_{l \in \mathcal{I}} a_l h_1(t - lT - \tau_l) + n(t) \quad (7)$$

where \mathcal{I} , the set of K training locations, is a subset of the universal set of locations in the sector given by $\mathcal{U} = \{0, 1, \dots, N-1\}$. The other parameters are as defined for the system of (1). After low-pass filtering, uniform samples of the resulting waveform $r(t)$ are taken, and the k th uniform sample taken at instant kT is given by

$$r_k = \sum_{l \in \mathcal{I}} a_l h(kT - lT - \tau_l) + n_k \quad (8)$$

where $\{n_k\}$ are i.i.d. zero-mean Gaussian random variables of variance σ^2 . A total of $N + 2M$ samples are assumed to be taken, and stacked into an observation vector $\mathbf{r} = [r_{-M} \ r_{-M+1} \ \dots \ r_{N+M-1}]^T$, where eventually, $M \rightarrow \infty$. The problem now is to evaluate the CRB for the estimation of the timing parameter $\theta = [\Delta T \ \tau_0]^T$ given the set of observations \mathbf{r} . This is, in general, a function of the set of training symbols $\{a_l\}$. The average CRB, averaged over all training symbol patterns, is got by taking the expectation of \mathbf{J}_θ^{-1} with respect to the training symbols, i.e., $E\{\mathbf{J}_\theta^{-1}\}$. This is intractable for the system model considered here. Instead, a looser, more tractable lower bound, called the modified CRB (MCRB), is employed. The modified CRB is given by [10]

$$E[(\hat{\theta}_i(\mathbf{r}) - \theta_i)^2] \geq \{E[\mathbf{J}_\theta]\}^{-1}(i, i) \quad (9)$$

where the expectation is now over \mathbf{r} and $\{a_l\}$. For ease of notation, the MCRB is referred to as the CRB in the sequel.

The Fisher information matrix \mathbf{J}_θ has an elegant form when the density $f_{\mathbf{r}|\theta}(\mathbf{r}|\theta)$ is multivariate normal [15]. Specifically, if \mathbf{r} is normally distributed with mean $\mathbf{m}_r(\theta)$ and covariance matrix \mathbf{R} , then \mathbf{J}_θ can be simplified to

$$\mathbf{J}_\theta = \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} \quad (10)$$

where \mathbf{G} is the sensitivity matrix, given by

$$\mathbf{G} = \left[\frac{\partial}{\partial \theta} \mathbf{m}_r^T(\theta) \right] = [g_0 \ g_1 \ \dots \ g_{P-1}] \quad (11)$$

P is the size of the vector θ and the $(N+2M) \times 1$ column vector $\mathbf{g}_i = \frac{\partial \mathbf{m}_r(\theta)}{\partial \theta_i}$ denotes the variation in the mean $\mathbf{m}_r(\theta)$ with respect to θ_i . For the system model under consideration, $P = 2$ and $\mathbf{R} = \sigma^2 \mathbf{I}_{(N+2M) \times (N+2M)}$, where $\mathbf{I}_{(N+2M) \times (N+2M)}$ is the identity matrix of dimension $(N+2M)$. Substituting these

and letting $M \rightarrow \infty$, the average Fisher information evaluates to [16]

$$E[\mathbf{J}_\theta] = \frac{E_{h'}}{\sigma^2} \begin{bmatrix} \sum_{k \in \mathcal{I}} k^2 & \sum_{k \in \mathcal{I}} k \\ \sum_{k \in \mathcal{I}} k & K \end{bmatrix} \quad (12)$$

where $E_{h'}$ is the energy in the derivative of the pulse shape $h(t)$. (See Appendix for proof.) Inverting this, the CRB on estimating the timing parameters during acquisition is given by

$$\begin{aligned} E[(\hat{\Delta T}(r) - \Delta T)^2] &\geq V_{\Delta T}, \\ E[(\hat{\tau}_0(r) - \tau_0)^2] &\geq V_{\tau_0} \end{aligned} \quad (13)$$

where

$$\begin{aligned} V_{\Delta T} &= \frac{\sigma^2}{KE_{h'}} \frac{1}{\frac{1}{K} \sum_{k=0}^{K-1} x_k^2 - \left(\frac{1}{K} \sum_{k=0}^{K-1} x_k\right)^2} \\ &= \frac{\sigma^2}{KE_{h'}} \frac{1}{\text{var}(\mathcal{I})}, \\ V_{\tau_0} &= \frac{\sigma^2}{KE_{h'}} \frac{\frac{1}{K} \sum_{k=0}^{K-1} x_k^2}{\frac{1}{K} \sum_{k=0}^{K-1} x_k^2 - \left(\frac{1}{K} \sum_{k=0}^{K-1} x_k\right)^2} \\ &= \frac{\sigma^2}{KE_{h'}} \left(1 + \frac{\text{mean}^2(\mathcal{I})}{\text{var}(\mathcal{I})}\right) \end{aligned} \quad (14)$$

where the set \mathcal{I} has been redefined in terms of its K ordered elements $x_k, k = 0, 1, \dots, K-1$, such that $0 \leq x_0 < x_1 < \dots < x_{K-1} \leq (N-1)$. (See Appendix for proof.) The mean and variance of any given vector $\mathcal{S} = [s_0, s_1, \dots, s_{K-1}]$ are defined by $\text{mean}(\mathcal{S}) = (1/K) \sum s_k$ and $\text{var}(\mathcal{S}) = (1/K) \sum s_k^2 - ((1/K) \sum s_k)^2$. The problem of choosing the best training locations, therefore, is the same as choosing the set $\mathcal{I} = \{x_k\}_0^{K-1}$.

The problem statement thus far has been that of estimating ΔT and τ_0 , given the set of observations \mathbf{r} . This can be generalized to the problem of estimating ΔT and τ_m , for any m . For the generalized problem, the CRB on the estimation of ΔT is the same as before, and the CRB for τ_m is given by

$$V_{\tau_m} = \frac{\sigma^2}{KE_{h'}} \left(1 + \frac{(\text{mean}(\mathcal{I}) - m)^2}{\text{var}(\mathcal{I})}\right). \quad (15)$$

(See Appendix for proof.) Therefore, the CRB on the estimation of ΔT is independent of the particular choice of m . The CRB on the estimation of τ_m is minimized when $m = \text{mean}(\mathcal{I})$.

The relative importance of estimating the two parameters ΔT and τ_m can be quantified by the mean squared error (MSE) defined by

$$\text{MSE} = \frac{1}{N} \sum_{k=0}^{N-1} E[(\tau_k - \hat{\tau}_k)^2]. \quad (16)$$

The mean squared error can be rewritten as $\text{MSE} =$

$$\frac{1}{N} \sum_{k=0}^{N-1} E[(\tau_m + (k-m)\Delta T - \hat{\tau}_m - (k-m)\hat{\Delta T})^2] \quad (17)$$

where m is chosen as the reference index. For CRB-achieving estimates of ΔT and τ_m , the MSE is given by

$$\text{MSE} = \frac{\sigma^2}{KE_{h'}} \left\{1 + \frac{\text{var}(\mathcal{U}) + (\text{mean}(\mathcal{U}) - \text{mean}(\mathcal{I}))^2}{\text{var}(\mathcal{I})}\right\}. \quad (18)$$

(See Appendix for proof.) The MSE expression is independent of the choice of the reference index m . The ultimate goal of the timing estimation problem is to minimize the MSE. In the next section, the training symbol arrangement that minimizes the CRB on ΔT is first derived, and then it is shown that this arrangement minimizes the MSE.

IV. OPTIMAL TRAINING LOCATIONS

In this section, the optimal index set \mathcal{I} to minimize $V_{\Delta T}$, or equivalently to maximize $\text{var}(\mathcal{I})$, is arrived at. This is the split-preamble strategy. Next, the problem of minimizing V_{τ_0} is considered, and it is shown that the minimizing arrangement is in fact different in this case. The split-preamble arrangement minimizes the CRB for the estimation of τ_m , where $m = \text{mean}(\mathcal{I}) = \text{mean}(\mathcal{U}) = (N-1)/2$. Finally, it is shown that the split-preamble strategy minimizes the mean squared error. In other words, the problem of minimizing the MSE is equivalent to that of minimizing the estimation error variance of $[\Delta T \ \tau_{(N-1)/2}]$, and not of $[\Delta T \ \tau_0]$.

A. Minimizing $V_{\Delta T}$

Theorem: Let K be an even natural number. Let $\mathcal{I} = [i_0, i_1, \dots, i_{K-1}]$ be any vector of integers satisfying

$$0 \leq i_0 < i_1 < \dots < i_{K-1} \leq (N-1) \quad (19)$$

and let \mathcal{I}_o be given by

$$\mathcal{I}_o = \left[0, 1, \dots, \frac{K}{2} - 1, N - \frac{K}{2}, \dots, N - 2, N - 1\right]. \quad (20)$$

Then

$$\text{var}(\mathcal{I}_o) \geq \text{var}(\mathcal{I}) \quad (21)$$

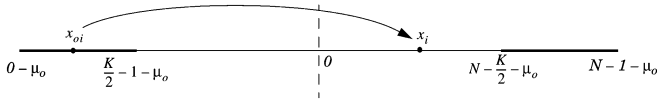
with equality iff $\mathcal{I} = \mathcal{I}_o$.

Proof: Let

$$\begin{aligned} \mathcal{I}_o &= [i_{o0}, i_{o1}, \dots, i_{o(K-1)}], \text{ i.e.,} \\ i_{ok} &= k, 0 \leq k \leq \frac{K}{2} - 1, \\ &= (N - K) + k, \frac{K}{2} \leq k \leq K - 1. \end{aligned} \quad (22)$$

The mean of \mathcal{I}_o is $\mu_o = (N-1)/2$. The variance of a vector does not change if a constant is added to or subtracted from all the elements. Therefore, two new vectors \mathcal{I}^1 and \mathcal{I}_o^1 are now formed by subtracting μ_o from \mathcal{I} and \mathcal{I}_o . In symbols

$$\begin{aligned} \mathcal{I}^1 &= [x_0, x_1, \dots, x_{K-1}], \\ \mathcal{I}_o^1 &= [x_{o0}, x_{o1}, \dots, x_{o(K-1)}], \\ x_k &= i_k - \mu_o, \\ x_{ok} &= i_{ok} - \mu_o. \end{aligned} \quad (23)$$

Fig. 3. Various constraints of ϵ .

The subtraction does not change the variance, and therefore, $\text{var}(\mathcal{I}^1) = \text{var}(\mathcal{I})$ and $\text{var}(\mathcal{I}_o^1) = \text{var}(\mathcal{I}_o)$. An alternative notation is

$$\mathcal{I}^1 = \mathcal{I}_o^1 + \epsilon \quad (24)$$

where $\epsilon = [\epsilon_0, \epsilon_1, \dots, \epsilon_{K-1}]$ and $\epsilon_i = x_i - x_{oi}$. The set ϵ is the set of perturbations needed to go from set \mathcal{I}_o^1 to the set \mathcal{I}^1 , and satisfies the following constraints:

$$\begin{aligned} 0 &\leq \epsilon_0 \leq \epsilon_1 \leq \dots \leq \epsilon_{\frac{K}{2}-1} \leq (N-K) \\ &\quad - (N-K) \leq \epsilon_{\frac{K}{2}} \leq \epsilon_{\frac{K}{2}+1} \leq \dots \leq \epsilon_{K-1} \leq 0 \\ \epsilon_{\frac{K}{2}-1} - \epsilon_{\frac{K}{2}} &\leq (N-K). \end{aligned} \quad (25)$$

Fig. 3 illustrates the constraints on ϵ . The elements of \mathcal{I}_o^1 are marked along the real line. The left-half elements are all negative and the only perturbations allowed are positive. The perturbations $\{\epsilon_i\}_{0}^{(K/2)-1}$ form an increasing sequence with the maximum being $N-K$, which happens when $i_{(K/2)-1} = N - (K/2) - 1$. Similarly, the second restriction arises from the right half. The third restriction arises from the fact that there are N slots and K distinct, ordered elements, and therefore, no two consecutive elements can be separated by more than $N-K$.

The variances under consideration, $\text{var}(\mathcal{I}_o)$ and $\text{var}(\mathcal{I})$, can now be rewritten as

$$\begin{aligned} \text{var}(\mathcal{I}_o) &= \frac{1}{K} \sum x_{ok}^2, \\ \text{var}(\mathcal{I}) &= \frac{1}{K} \sum x_k^2 - \left(\frac{1}{K} \sum \epsilon_k \right)^2. \end{aligned} \quad (26)$$

Next, consider the left half, i.e., $0 \leq k \leq (K/2) - 1$. The intuition for the following steps is as follows. For the left half, all the x_{ok} s are negative, and positive ϵ_k s will ensure that $|x_k| < |x_{ok}|$ as long as x_k does not exceed $-x_{ok}$ for all possible ϵ_k . From (22), (23), and (24)

$$\begin{aligned} 0 &\leq \epsilon_k \leq N-K, \\ \Rightarrow x_{ok} &\leq x_{ok} + \epsilon_k \leq N-K + x_{ok}, \\ \Rightarrow i_{ok} - \mu_o &\leq x_k \leq i_{ok} - \mu_o + N-K, \\ \Rightarrow |x_k| &\leq \max(|i_{ok} - \mu_o|, |i_{ok} - \mu_o + N-K|), \\ \Rightarrow |x_k| &\leq \max\left(\frac{N-1}{2} - k, N-K + \frac{N-1}{2} + k\right), \\ \Rightarrow |x_k| &\leq \frac{N-1}{2} - k, \\ \Rightarrow |x_k| &\leq |x_{ok}|. \end{aligned} \quad (27)$$

The equality condition is satisfied iff $\epsilon_k = 0$ since $\frac{N-1}{2} - k > N-K + \frac{N-1}{2} + k$ for the left half. By proceeding similarly, it

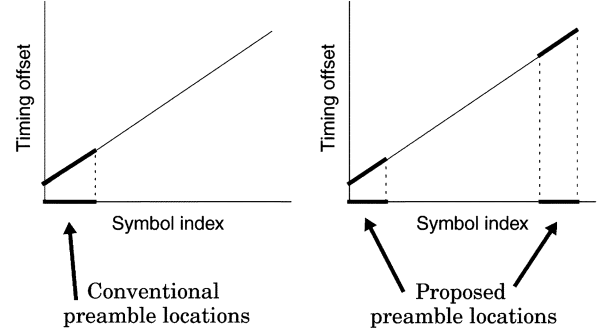


Fig. 4. Proposed training symbol placement strategy.

can be shown that the same holds for the right half. Combining the two leads to

$$|x_k| \leq |x_{ok}| \quad \text{for } 0 \leq k \leq (K-1) \quad (28)$$

with equality iff $x_k = x_{ok}$ for all k .

Combining (26), (28) and the fact that $((1/K) \sum \epsilon_k)^2 \geq 0$ with equality iff $\epsilon = \mathbf{0}$ leads to the desired result, i.e.,

$$\text{var}(\mathcal{I}_o) \geq \text{var}(\mathcal{I}) \quad (29)$$

with equality iff $\mathcal{I} = \mathcal{I}_o$.

Therefore, to minimize $V_{\Delta T}$, the optimum strategy is the split-preamble strategy, where the available training symbols are split into two equal halves and placed at the beginning and at the end of the sector, as shown in Fig. 4. The intuition behind the split-preamble strategy is that the observations need to be as far apart as possible while estimating the slope of a straight line.

B. Minimizing V_{τ_0}

Next consider V_{τ_0} . Recall from (14) that

$$V_{\tau_0} = \frac{\sigma_n^2}{E_h' K} \left(1 + \frac{\text{mean}(\mathcal{I})^2}{\text{var}(\mathcal{I})} \right). \quad (30)$$

Therefore, the optimal arrangement in this case (\mathcal{I}') is the one that satisfies

$$\frac{\text{var}(\mathcal{I}')}{\text{mean}(\mathcal{I}')^2} > \frac{\text{var}(\mathcal{I})}{\text{mean}(\mathcal{I})^2} \quad (31)$$

for all $\mathcal{I} \neq \mathcal{I}'$. This can be split into two steps: first, find the optimal arrangement for each possible value of the mean and then find the optimum over all mean values, i.e., for each value of the mean k , first a set \mathcal{I}_k is chosen such that $\text{mean}(\mathcal{I}_k) = k$ and

$$\text{var}(\mathcal{I}_k) > \text{var}(\mathcal{I}) \quad (32)$$

for all $\mathcal{I} \neq \mathcal{I}_k$ with $\text{mean}(\mathcal{I}) = k$. As the next step, a set $\mathcal{I}' = \mathcal{I}_k$ is picked, where

$$\frac{\text{var}(\mathcal{I}_k)}{k^2} > \frac{\text{var}(\mathcal{I}_j)}{j^2} \quad (33)$$

for all $j \neq k$.

With $N \gg 1$, the solution of this problem leads to the following arrangement. A total of K_1 known symbols are placed at the start and $K - K_1$ at the end of the sector where

$$K_1 = \left\lceil K \left(1 - \frac{(3N - K)K}{6N^2} \right) \right\rceil \quad (34)$$

where $\lceil \cdot \rceil$ denotes the rounding operation. (See the Appendix for details of the proof.) For $N \gg K$, $K_1 \approx K$, which means that almost all the known symbols need to be placed at the start of the sector to do the best estimation of τ_0 . The optimal arrangement for estimating ΔT performs worse by a factor of approximately 2 as far as the estimation error variance for τ_0 goes. One interpretation of this is that the known symbols at the end of the sector do not contribute to any great extent to the estimation of the initial offset τ_0 , and the error variance is roughly equivalent to that with only $K/2$ known symbols, all at the beginning.

One way to work around this mismatch of the optimal solutions is to redefine the indices so that the mean for the optimal solution for estimating ΔT is zero, i.e., subtract $(N - 1)/2$ from all the indices, as done in the proof in the previous subsection. From (30), it is evident having a zero mean is the best possible solution given K training symbols. Therefore, the optimal scheme for estimating ΔT performs optimally for τ_0 too. The difference being, in this case, the timing offset at the center of the sector is estimated instead of at one end. Assuming that the estimates $\hat{\Delta T}$ and $\hat{\tau}_0$ are used to compute the expected timing offsets for the remaining positions, this arrangement makes intuitive sense as this minimizes the maximum interpolation distance, thus minimizing the maximum absolute value of the interpolation error. This also ties in with the earlier analysis that the CRB on estimation of τ_m is minimized for $m = \text{mean}(\mathcal{I})$. This intuitive explanation is formalized in the next subsection, where the split-preamble arrangement is shown to minimize the MSE.

C. Minimizing the MSE

Recall from (18) that the MSE for the CRB-achieving timing estimator is given by

$$\text{MSE} = \frac{\sigma^2}{KE_{h'}} \left\{ 1 + \frac{\text{var}(\mathcal{U}) + (\text{mean}(\mathcal{U}) - \text{mean}(\mathcal{I}))^2}{\text{var}(\mathcal{I})} \right\}. \quad (35)$$

The index set for the split-preamble strategy, namely the set \mathcal{I}_o , has the following properties.

- It has the highest variance among all index sets \mathcal{I} .
- $\text{mean}(\mathcal{I}_o) = \text{mean}(\mathcal{U})$.

Combining these properties with the MSE expression leads to the conclusion that the split-preamble arrangement minimizes the MSE.

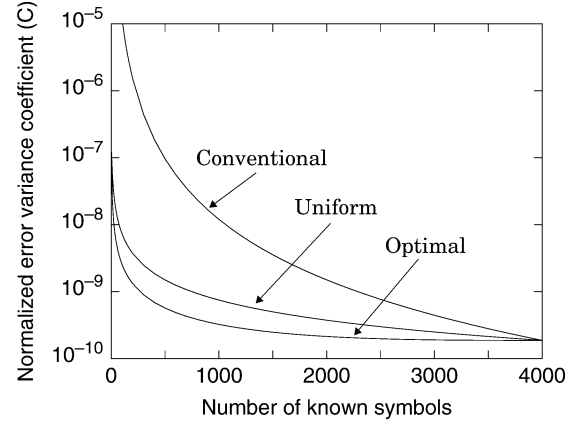


Fig. 5. The split-preamble arrangement is much better than the conventional one.

V. PERFORMANCE EVALUATION OF THE SPLIT-PREAMBLE ARRANGEMENT

In the previous section, the split-preamble strategy was shown to minimize the CRB on frequency estimation, and also the mean squared error. In this section, three schemes are compared for performance with respect to estimating ΔT :

- all the known symbols at the start of the sector (\mathcal{I}_1) with error variance V_1 ;
- the split-preamble arrangement (\mathcal{I}_2) with V_2 ;
- sprinkle the known symbols uniformly throughout the sector (\mathcal{I}_3) with V_3 .

The CRB on the timing estimation error variance for these three schemes can be evaluated using (14) to be the following:

$$\begin{aligned} C_1 &= \frac{V_1 E_{h'}}{\sigma^2} = \frac{12}{K(K^2 - 1)}, \\ C_2 &= \frac{V_2 E_{h'}}{\sigma^2} = \frac{1}{\frac{K(K^2 - 1)}{12} + \frac{KN(N - K)}{4}}, \\ C_3 &= \frac{V_3 E_{h'}}{\sigma^2} = \frac{(K - 1)^2 V_1}{(N - 1)^2 \sigma^2} \end{aligned} \quad (36)$$

where the normalized error variance coefficient C is defined above.

Fig. 5 compares the performance of the three training symbol arrangements for $N = 4000$, and K ranging from 1 to 4000. The block length $N = 4000$ was chosen to be similar to the ones currently used for digital magnetic recording. The optimal arrangement significantly outperforms the conventional one, and the uniform sprinkling scheme is not much worse off. As an example, compare the total number of symbols needed to get the normalized error variance coefficient $C = 2 \times 10^{-9}$. The optimal scheme needs 86 symbols, the uniform scheme needs 248 symbols, whereas the conventional scheme needs 1588 symbols of the 4000 to be known! The MSE performance comparison of the three schemes leads to similar conclusions, and is omitted for brevity.

VI. PLL-BASED METHOD WITH SPLIT PREAMBLE

A cycle slip is said to have occurred when the detector gains or loses whole symbols. This could happen if the detector's

timing estimates are off by whole multiples of the symbol duration. At low signal-to-noise ratio (SNR), PLL exhibits cycle slips. Even with perfect phase initialization, the PLL occasionally loses track and settles at a timing offset that is an integral multiple of the symbol duration. This leads to addition or deletion of symbols in the data stream sent to the rest of the detector. For digital communication or data storage applications that employ error correction codes (ECC) to combat channel noise, this almost always leads to the ECC decoder failure.

An added advantage with the split-preamble method is that it is well equipped to handle cycle slips. Independent estimators can be used at the beginning and at the end of the sector to locate the start and the end training symbols, thus reducing the effects of cycle slips in the tracking mode in between.

Simulation results are presented for the following simple acquisition strategy with the split-preamble arrangement. Let the two preamble sequences be denoted by S_1 (beginning) and S_2 (end), respectively.

- At the detector, correlation detection is performed and a trained PLL, trained to the sequence S_1 , is run.
- At the end of the S_1 sequence, the PLL is switched to the decision-directed mode and operated in that mode till the end of the sector, i.e., for the unknown data and for S_2 . Let the timing estimates from the PLL be denoted by $\{\hat{\tau}_k\}$ for $k = 0$ to $k = N - 1$.
- Next, out of the N decisions available, the last $K/2$ are picked and correlated with the $K/2$ symbols corresponding to S_2 for different shifts and the shift needed, say m , to get the correlation peak, is recorded.
- The symbol duration offset can then be estimated using m and $\{\hat{\tau}_k\}$. One simple method is to use just one PLL timing estimate from each end, for example, $\hat{\tau}_{(K/2)-1}$ and $\hat{\tau}_{N-1-(K/2)}$. In this case, the symbol duration offset estimate is given by

$$\Delta T = \frac{\hat{\tau}_{N-1-(K/2)} - \hat{\tau}_{(K/2)-1} - mT}{N - K}. \quad (37)$$

This is the method used for the simulation results presented in the next section. An alternative method with potentially improved performance is to use multiple timing estimates from each end of the sector and compute ΔT by a least-squares linear fit.

- The timing instants to be used for the tracking mode are then got by interpolating linearly between $\hat{\tau}_{(K/2)-1}$ and $\hat{\tau}_{N-1-(K/2)} - mT$. These new timing estimates can then be used to arrive at refined samples by interpolation.

By this method, though the exact location of the cycle slip during tracking has not been detected, the cycle slip can be corrected as long as the correlation detection of the S_2 sequence in the decision sequence was successful. To aid good start-up of the PLL, S_1 is chosen to be a periodic pattern, and to aid the correlation detection for the end of the sector, S_2 is chosen to be a pseudorandom pattern.

Fig. 6 shows the MSE for frequency estimation as a function of the SNR for the split-preamble strategy with the PLL-based method described above. The simulation parameters are chosen to be similar to those for digital magnetic recording. The channel

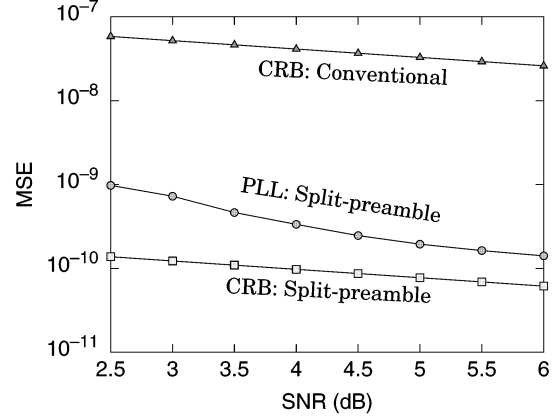


Fig. 6. Splitting the preamble leads to improved frequency estimates.

is a PR-IV channel [17], which is characterized by the impulse response $h(t)$, where

$$h(t) = \frac{\sin(\pi t/T)}{\pi t/T} - \frac{\sin(\pi(t-2T)/T)}{\pi(t-2T)/T}. \quad (38)$$

The sector length was chosen to be $N = 5000$. $K = 120$ of these symbols are known at the detector. For the conventional system, a length-120 periodic version of the $[11-1-1]$ pattern is used as the training sequence. For the proposed scheme, a length-60 sequence formed by repetitions of the $[11-1-1]$ pattern is used for the training symbols at the start of the sector. For the known sequence at the end of the sector, a pseudo-random length-60 sequence is chosen. The timing offsets used are $\tau_0 = 0$ and $\Delta T = 0.002T$. The PLL gain parameters are $\alpha = 0.04$ and $\beta = \alpha^2/4$, chosen to minimize the number of cycle slips for the conventional system at an SNR of 5 dB. Also shown in this plot are the CRB for frequency estimation with the conventional and the split-preamble training symbol placement strategies. At SNR = 5 dB, the PLL-based method with the split-preamble strategy shows more than two orders of magnitude improvement when compared to the CRB with the conventional training symbol placement strategy, and performs to within half an order of magnitude of the CRB with the split-preamble arrangement. Performance can be improved further by using PLL gains that are optimized for each SNR, and also by least-squares frequency estimation.

Fig. 7 shows the cycle slip performance as a function of SNR for the aforementioned acquisition scheme when compared to the conventional scheme where all the known symbols are placed at the beginning of the sector. For the split-preamble arrangement, the number of cycle slips that were not corrected is plotted. The split-preamble acquisition scheme reduced the number of residual cycle slips by a factor of 10 to 100 depending on the SNR. In terms of SNR, at a cycle slip rate of 10^{-3} , it leads to an improvement of around 1.5 dB.

Also simulated was the system where all the data symbols were known at the detector, i.e., the fully trained system. The fully trained system did not exhibit any cycle slips for the range of SNRs considered when the PLL-based system was used for timing recovery. This leads to the conclusion that one way to improve the timing recovery performance is to improve the

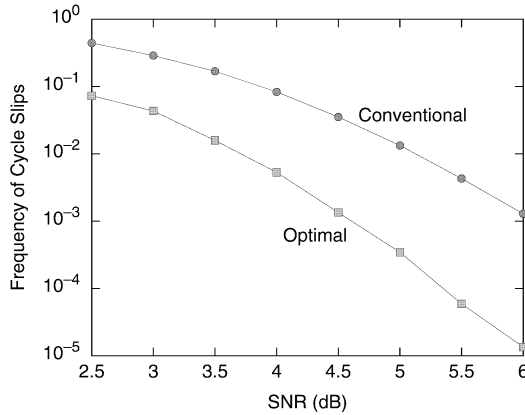


Fig. 7. Splitting the preamble reduces the occurrence of cycle slips.

quality of the decisions available to the timing recovery block, providing additional justification for iterative timing recovery methods [16].

VII. JOINT ACQUISITION AND TRACKING

In Sections III and IV, acquisition was considered in isolation, where only the training symbols were written. In this section, a simple version of the problem of joint acquisition and tracking is considered, where the unknown user data is not assumed to be zero. The timing offset model once again consists of an initial timing offset τ_0 and a frequency offset parameter ΔT . The presence of both known and unknown data makes the CRB, and also the MCRB, intractable. To facilitate analysis, a linear observation model based on the characteristics of the timing error detector (TED) is first derived. Based on this model, the CRB on estimating the timing parameters and also the MSE are computed, and finally, it is shown that the split-preamble arrangement minimizes the MSE in this case.

A. Linear Model

An ideal timing error detector would output $\hat{\epsilon}_k = \epsilon_k$. In the presence of noise, the TED would output a noisy version given by $\hat{\epsilon}_k = \epsilon_k + \nu_k$. For a linear TED, ν_k would be independent of ϵ_k . Practical TEDs are linear for 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. In the trained mode, the TED is linear over a much larger range, and therefore, the TED linearity assumption holds even with the large timing offsets faced during acquisition. For any TED, we can write

$$\hat{\epsilon}_k = \epsilon_k + \nu_k \quad (39)$$

$$= \tau_k - \hat{\tau}_k + \nu_k. \quad (40)$$

Constructing a new observation $y_k = \hat{\tau}_k + \hat{\epsilon}_k$ leads to the following observation model:

$$y_k = \tau_k + \nu_k. \quad (41)$$

Therefore, to arrive at the new model, corresponding outputs from the PLL and the TED need to be added. The PLL outputs an estimate of τ_k assuming that only observations $\{r_l\}_{l=0}^{k-1}$

are available. This is, therefore, an *a priori* estimate of τ_k . The TED's estimate $\hat{\epsilon}_k$ is based on the present observation r_k as well. Hence, y_k , which is the sum of $\hat{\tau}_k$ and $\hat{\epsilon}_k$, is an *a posteriori* estimate of τ_k and the quality of this estimate depends on the quality of the TED output. The new observation model is linear where the TED characteristics are linear.

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

$$\mathbf{y} = \boldsymbol{\tau} + \boldsymbol{\nu} \quad (42)$$

where $\mathbf{y} = [y_0 y_1 \dots y_{K-1}]^T$, $\boldsymbol{\tau} = [\tau_0 \tau_1 \dots \tau_{K-1}]^T$, and $\boldsymbol{\nu} = [\nu_0 \nu_1 \dots \nu_{K-1}]^T$. The noise variables $\{\nu_k\}$ were assumed to be independent zero-mean normal random variables. The timing recovery block assumed for this section takes advantage of the K known symbols in the training phase, leading to a measurement noise variance σ_a^2 for the observations corresponding to the known symbols (acquisition), and σ_t^2 for the observations corresponding to the data symbols (tracking) such that $\sigma_a^2 < \sigma_t^2$. An example is the PLL-based timing recovery system, where the timing estimation error variance is lesser with training than otherwise. Though the timing estimation error variance as a function of the bit position does not show the step-function type behavior assumed here, but rather varies in a gradual fashion, this observation noise model is used as an approximation.

B. CRB for the Linear Model

Let the parameter to be estimated be $\boldsymbol{\theta} = [\Delta T \tau_0]^T$, and let $\mathbf{x}(\boldsymbol{\theta})$ be the noiseless observation. For the linear model of (42), the k th component of $\mathbf{x}(\boldsymbol{\theta})$ is just $x_k = \tau_k = k\Delta T + \tau_0$. The noise variables are not i.i.d. Using (10), the Fisher information matrix can be written as

$$\mathbf{J}_\theta = \frac{1}{\sigma_a^2} \sum_{k \in \mathcal{I}_a} \begin{bmatrix} k^2 & k \\ k & 1 \end{bmatrix} + \frac{1}{\sigma_t^2} \sum_{k \in \mathcal{I}_t} \begin{bmatrix} k^2 & k \\ k & 1 \end{bmatrix} \quad (43)$$

where the ordered vectors \mathcal{I}_a and \mathcal{I}_t contain the indices corresponding to the training and the data symbols respectively. The elements of \mathcal{I}_a and \mathcal{I}_t together span all the integers from 0 to $N - 1$. Let m_a and m_t be the mean of \mathcal{I}_a and \mathcal{I}_t respectively, and let their respective variances be V_a and V_t .

The determinant of \mathbf{J}_θ can be written as

$$|\mathbf{J}_\theta| = \left\{ \frac{K}{\sigma_a^2} + \frac{N-K}{\sigma_t^2} \right\} \left\{ \frac{K}{\sigma_a^2} V_a + \frac{N-K}{\sigma_t^2} V_t \right\} \dots + \frac{K}{\sigma_a^2} \frac{N-K}{\sigma_t^2} (m_a - m_t)^2. \quad (44)$$

The CRB then evaluates to

$$V_{\Delta T} = \frac{1}{|\mathbf{J}_\theta|} \left[\frac{K}{\sigma_a^2} + \frac{N-K}{\sigma_t^2} \right], \quad V_{\tau_0} = \frac{1}{|\mathbf{J}_\theta|} \left[\frac{K}{\sigma_a^2} (V_a + m_a^2) + \frac{N-K}{\sigma_t^2} (V_t + m_t^2) \right]. \quad (45)$$

[See the Appendix for proof of (44) and (45).]

