On the Performance Limits of Data-Aided Synchronization

Yimin Jiang, *Member, IEEE*, Feng-Wen Sun, *Member, IEEE*, and John S. Baras, *Fellow, IEEE*

Abstract—This paper addresses data-aided (DA) synchronization, in which the reference parameter acquisition is aided by a training sequence known to the receiver. The Cramer–Rao lower bound (CRB) for the DA timing and/or carrier phase recovery is presented. For DA parameter estimation, the CRB typically varies with the training sequence. This indicates that different training sequences offer fundamentally different performance. In the literature, the widely cited closed-form CRB for timing and carrier phase recovery was derived under the assumption that the training sequence is independent and identically distributed (i.i.d.) and sufficiently long. In this paper, we derive a closed-form CRB for timing and carrier phase recovery with respect to an arbitrary training sequence and pulse shaping function for the over and under sampling cases. It turns out that the CRB is a weighted summation of the aperiodic correlation of the training sequence and the weighting factor is determined by the pulse shaping filter. Therefore, this paper reveals the fundamental link between a training sequence and its corresponding performance limit.

Index Terms—Cramer–Rao bound (CRB), data-aided (DA) synchronization, phase estimation, timing, synchronizations.

I. INTRODUCTION

THE Cramer–Rao bound (CRB) is a lower bound on the minimum mean squared error (MMSE) of any *unbiased* estimator [1]. Since the CRB serves as a benchmark for the performance of an actual estimator, it has received considerable attention both in theory and practice. In practical systems, synchronization parameters, such as timing and carrier phase offsets, are often acquired with the aid of a training sequence known to the receiver, which is called the data-aided (DA) estimation. In the DA cases, the CRB varies with the training sequence, which implies that different training sequences offer fundamentally different performance. Therefore, it is important to understand the performance limit as given by the CRB for any particular training sequence. Due to its practical significance, the CRB for DA synchronization has been studied extensively in the literature [2]–[9]. However, in [2]–[8] the closed-form CRB for the DA timing and/or carrier phase recovery for an arbitrary training sequence is unavailable. The authors in [2] gave a thorough summary on the CRBs for carrier frequency, phase, and

timing offsets estimation. The CRB for joint timing and carrier phase recovery was first introduced by Moeneclaey in [4], [5]; it was further discussed in his publications [3] and [6]. It was considered mathematically intractable to derive the bound for an arbitrary training sequence. Moeneclaey simplified the issue by adopting the strong law of large numbers and the assumption that the training sequence is zero mean, independent and identically distributed (i.i.d.), and sufficiently long. This method reduces the computation dramatically and provides some insights for communication receiver design. Unfortunately, it hides the interaction between a training sequence and its resultant estimation performance. In order to deal with the estimation problem in the presence of nuisance parameters, D'Andrea *et al.* proposed the *modified* CRB (MCRB) in [7]. It is pointed out in [3] that the CRBs derived previously in [4], [6] are actually MCRBs. In [9], the author took a frequency-domain interpretation of the CRB. Similar to earlier works, it assumes pseudorandom training data.

In principle, it is possible to use a brute-force numerical approach to compute the CRB for any given training sequence. However, the brute-force approach does not provide any insight on the interaction between a training sequence and its resultant CRB. One of the major difficulties in computing the CRB arises from the fact that the computation involves the inverse of an autocovariance matrix that is Toeplitz. It is well known that the inverse of a Toeplitz matrix is no longer Toeplitz, which makes it difficult to analyze the bound analytically. One technique to tackle this problem is to exploit the relation between Toeplitz matrices and their associated circulant matrices. In fact, it is well known that the inverse of a Toeplitz matrix does converge to a circulant matrix in the weak sense under certain conditions [10]. Unfortunately, the weak convergence is in the mean sense and only useful for computing the mean of some quantities such as the mean of a quadratic form associated with a random process. However, for an arbitrary training sequence, the CRB involves the evaluation of a quadratic form of the inverse matrix. Thus, the weak convergence theorem cannot be applied. It is this fact that motivated the research in [11].

By observing that communication receiver design often needs to seek optimality in regard to a data sequence transmitted within finite duration, we defined the *finite-term strong convergence* regarding two families of matrices in [11]. We further presented a condition under which the inverse of a Toeplitz matrix converges in the strong sense to a circulant matrix for finite-term quadratic forms. The finite-term strong convergence separates the length of the transmission window from that of the observation window. By increasing the observation window, the receiver design approaches the optimal solution when the noise incurred in the system is correlated. Therefore, we can

obtain a closed-form formula for the optimal receiver design through substituting the Toeplitz matrix with its associated circulant matrix when the condition of the finite-term strong convergence is met.

Based on the results presented in [11], a closed-form CRB for the DA joint timing and carrier phase offsets estimation is derived with respect to any given training sequence and pulse-shaping function. The only assumption is that the derivative of the pulse-shaping function exists, i.e., it is sufficiently smooth. The bound uncovers the close relation between a training sequence and its resultant performance limit on timing and carrier phase recovery. We present the bound both in the frequency and time domains. In particular, the time-domain expression clearly exhibits the interaction between the pulse-shaping function and the training sequence. Under the same framework, this paper reveals the tradeoff between the sampling rate and the estimation performance. This provides guidance for high-speed modem design where it is critical to make the sampling rate as low as possible.

The rest of the paper is organized as follows. Section II briefly reviews the relevant mathematical tools presented in [11]. Section III derives the CRB. Section IV further evaluates the bound for various practical scenarios and offers comparison with the state of the art in the literature. Section V concludes the paper.

II. ON THE INVERSE OF TOEPLITZ MATRICES

Toeplitz matrices and their inverses were studied in [12], [10], [13]–[20]. Grenander and Szego's book systematically documents major results on Toeplitz matrices [12]. A literature survey is given in [11] on the works related to the inverses of Toeplitz matrices.

A family of Toeplitz matrices T_n is defined by a sequence of complex numbers $\{t_i; i = \ldots, -1, 0, 1, \ldots\}$ such that the entry of T_n at the ith row and jth column is equal to t_{i-j}, i.e., $T_n = \{t_{i-j}\}$. Furthermore, we restrict our discussion to the case that $t_{-i} = t_i^*$, where t_i^* is the complex conjugate of t_i. With this restriction, T_n becomes Hermitian. Toeplitz Hermitian matrices play a pivotal role in signal processing. In fact, what is really relevant is the inverse of such a matrix rather than the matrix itself for many applications. For instance, if t_i represents the autocorrelation of a stationary random process, the inverse of T_n is associated with the joint probability density function of n consecutive samples of the random process. In *filtering* problems, such an inverse appears in the Wiener–Hopf equation [21], [22].

As mentioned earlier, one of the difficulties in analyzing the inverse matrices arises from the fact that the inverse of a Toeplitz matrix is no longer Toeplitz. A technique to tackle the problem is to exploit the relation between Toeplitz matrices and their associated circulant matrices. An $n \times n$ matrix is called a *circulant* matrix if its (i, j)th entry is only a function of $(i - j) \bmod n$. In particular, for the family of Toeplitz matrices defined by the sequence $\{t_i\}$, a family of their associated circulant matrices can be defined through the discrete-time Fourier transform (DTFT) of the sequence $\{t_i\}$. Let $\mathcal{F}(\omega)$ denote the DTFT of $\{t_i\}$, i.e.,

$$\mathcal{F}(\omega) = \sum_{k=-\infty}^{\infty} t_k e^{-jk\omega}.$$

Note that $\mathcal{F}(\omega)$ is real due to the Hermitian constraint. Let U_n denote the unitary matrix defined as

$$U_n = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{-j(2\pi/n)} & \cdots & e^{-j(2\pi(n-1)/n)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j(2\pi(n-1)/n)} & \cdots & e^{-j(2\pi(n-1)(n-1)/n)} \end{bmatrix} \tag{1}$$

and D_n denote the diagonal matrix with the ith diagonal entry equal to $\mu_{i,n} = \mathcal{F}(2\pi i/n)$, i.e.,

$$D_n = \mathrm{diag}\{\mu_{0,n}, \mu_{1,n}, \ldots, \mu_{n-1,n}\}. \tag{2}$$

The matrix

$$C_n = U_n^H D_n U_n \tag{3}$$

is a circulant matrix [12], [10].

It has been observed that in many applications substituting T_n with C_n often leads to very useful and dramatic simplification to the problems at hand. This is due to the following facts.

- The inverse of a circulant matrix is still circulant.
- The eigendecomposition of a circulant matrix is equivalent to the discrete Fourier transform (DFT), providing additional insight in the frequency domain.
- The eigenvalues of C_n is the samples of the spectrum of $\{t_i\}$.

Apparently, in order to make such a substitution meaningful, the inverses of Toeplitz matrices need to converge to their associated circulant matrices.

The most well-known convergence is the weak convergence, which is based on the weak norm defined for an $n \times n$ matrix $A = \{a_{ij}\}$ as

$$\sqrt{n^{-1} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} |a_{ij}|^2}. \tag{4}$$

It can be shown that the Toeplitz matrix T_n converges to C_n in the weak sense as long as $|\mathcal{F}(\omega)|$ is bounded [12], [10]. Note that T_n converging to C_n may not necessarily mean that T_n^{-1} converges to C_n^{-1} even if T_n^{-1} does exist. A sufficient condition for the weak convergence of the inverses is that the strong norm of T_n^{-1} and C_n^{-1} is uniformly bounded [10]. The strong norm for a Hermitian matrix A can be defined as

$$\|A\| = \max_x \left| \frac{x^H A x}{x^H x} \right|$$

where the maximum is over all the vectors of the same dimension as A.

Examining the definition of the weak norm in (4), we can see that the weak convergence is in the mean sense due to the division factor $1/n$. Indeed, several successful applications of the weak approximation theory relate to the evaluation of the mean of some quantities, such as source coding and filtering problems based on the MMSE criterion, or computing the mean of a quadratic form associated with a random process [10], [19], [18], [20].

However, the usefulness of the weak convergence theorem is severely limited due to the fact that many applications actually
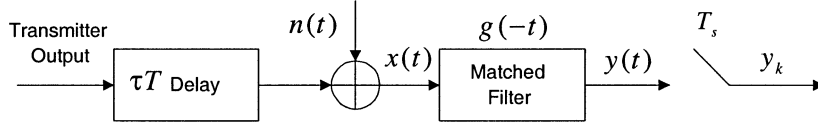
Fig. 1. Communication system model.

involve the quadratic form of T_n^{-1}, e.g., the problem addressed in this paper. Even if T_n^{-1} converges to C_n^{-1} in the weak sense, substituting T_n^{-1} with C_n^{-1} may not yield correct results since the convergence of a quadratic form can only be guaranteed if the convergence is in the strong sense.

Based on the results in [13], [14], it can be shown that T_n^{-1} can converge to a circulant matrix in the strong sense only when each T_n is an identity matrix. However, following the observation that the central part of T_n^{-1} does converge to that of C_n^{-1} under certain conditions, we can define the following *finite-term strong convergence* for two families of matrices.

Definition 1: For two families of Hermitian matrices A_n and B_n, consider the quadratic form

$$\max_{x} \frac{\|(A_n - B_n)x\|}{\|x\|} \tag{5}$$

where $\|x\| \triangleq \sqrt{x^H x}$ is the vector norm for a vector x, the maximum is over all the n-dimension vectors of the form

$$x = (0, \ldots, 0, x_{-W}, \ldots, x_0, \ldots, x_W, 0, \ldots, 0). \tag{6}$$

If (5) converges to zero for any given W as $n \to \infty$, we say that A_n converges to B_n in the finite-term strong sense. A quadratic form associated with x is called a finite-term quadratic form.

If x corresponds to a data transmission contained within the window $[-W, W]$ (where W does not increase with n), we are able to replace A_n with B_n asymptotically in evaluating the quadratic forms. Many practical applications fall into this category. The following result is established in [11].

Theorem 1: Let T_n be a family of Toeplitz Hermitian matrices associated with the sequence $\{t_i\}$, and $\mathcal{F}(\omega)$ be the DTFT of $\{t_i\}$. If $|\mathcal{F}(\omega)| \neq 0$ for $\omega \in [0, 2\pi]$ and $\sum_{k=-\infty}^{\infty} |k t_k| < \infty$, T_n^{-1} converges to C_n^{-1} in the finite-term strong sense.

Furthermore, for a vector x with the form defined in (6), the quadratic form is bounded by

$$\frac{\left\|\left(T_n^{-1} - C_n^{-1}\right)x\right\|}{\|x\|} \leq O\left(1/\sqrt{n}\right). \tag{7}$$

The theorem provides a simple way to diagonalize the inverses of a Toeplitz matrix. The condition on $\{t_i\}$ (i.e., $\sum_{k=-\infty}^{\infty} |k t_k| < \infty$) is sufficient for the convergence but not necessary. The finite-term quadratic form may converge much faster than $O(1/\sqrt{n})$.

III. DERIVATION OF THE CRAMER–RAO LOWER BOUND

We start with the mathematical formulation of the DA synchronization to establish the notation used in the presentation.

A. Problem Formulation

The baseband received signal is modeled as

$$x(t) = \sqrt{E_s} \sum_{m=-N/2}^{N/2-1} a_m g(t - mT - \tau T)e^{j\phi} + n(t) \tag{8}$$

where $g(t) = g_T(t) \otimes c(t) \otimes f(t)$ (without loss of generality, let us assume $g(t)$ is real), $g_T(t)$ is the transmitter pulse-shaping function, $c(t)$ is the channel response, $f(t)$ is the prefilter, $n(t)$ is the additive white Gaussian noise (AWGN) with two-sided power spectral density equal to $N_0/2$, T is the symbol interval, $\{a_m\}$, $m \in \mathcal{Z}$ ($\mathcal{Z}$ the set of integers) is the training sequence drawn from the complex plane with $E[|a_m|^2] = 1$. Variable ϕ models the carrier phase offset. The delay jitter τT models the absence of symbol synchronization between transmitter and receiver. It is assumed that $\phi \in [-\pi, \pi)$ and $\tau \in [-0.5, 0.5)$. The received signal $x(t)$ is passed through a matched filter with response $g(-t)$ as shown in Fig. 1. We assume that the channel impulse response and the prefilter are perfect, i.e., $g(t)$ is equal to the transmitter pulse-shaping function $g_T(t)$. The output $y(t)$ of the matched filter is sampled every T_s seconds, i.e., $T = LT_s$ where L is the sampling rate in samples per symbol period. In the DA case, the training sequence $\{a_m\}$ ($m = -N/2, \ldots, N/2 - 1$) is known to the receiver. The phase offset ϕ and timing offset τ are assumed to remain fixed over the duration of the observation.

The output of the matched filter $y(t)$ is

$$y(t) = \sqrt{E_s} \sum_{m=-N/2}^{N/2-1} a_m r(t - mT - \tau T)e^{j\phi} + N(t) \tag{9}$$

where

$$r(t) = \int_{-\infty}^{\infty} g(t + u)g(u)\, du$$

$$N(t) = \int_{-\infty}^{\infty} n(t + u)g(u)\, du.$$

Therefore, the samples of $y(t)$ are expressed as

$$y_k = \sqrt{E_s} \sum_{m=-N/2}^{N/2-1} a_m r(kT_s - mT - \tau T)e^{j\phi} + N_k \tag{10}$$

with $N_k = N(kT_s)$ that is a sequence of Gaussian random variables with zero mean and the autocorrelation function

$$R_y[k - l] = E[N_k N_l^*] = \frac{N_o}{2} r((k - l)T_s). \tag{11}$$

We rewrite (10) in terms of a matrix and vector product. First, define the following vectors:

$$y = \begin{bmatrix} y_{-K/2} & \cdots & y_0 & \cdots & y_{K/2-1} \end{bmatrix}^T$$

$$a = \begin{bmatrix} a_{-N/2} & \cdots & a_0 & \cdots & a_{N/2-1} \end{bmatrix}^T$$

$$N = \begin{bmatrix} N_{-K/2} & \cdots & N_0 & \cdots & N_{K/2-1} \end{bmatrix}^T \tag{12}$$

where the observation window length K is typically longer than LN, the training signal length, in order to capture the signal $y(t)$

beyond the training sequence for the optimal receiver design with correlated noise. Second, define a $K \times N$ matrix $\boldsymbol{R}(\tau)$ with the $\{m, n\}$th entry equal to $r((m - K/2)T_s - (n - N/2)T - \tau T)$, for $m = 0, 1, \ldots, K - 1$ and $n = 0, 1, \ldots, N - 1$. With this notation, (10) can be rewritten as

$$\boldsymbol{y} = \sqrt{E_s} \, \boldsymbol{R}(\tau) \boldsymbol{a} e^{j\phi} + \boldsymbol{N}. \tag{13}$$

The mean of $\boldsymbol{y}$ given $\boldsymbol{a}$, ϕ, and τ is

$$\boldsymbol{m}_y(\boldsymbol{a}, \phi, \tau) = E[\boldsymbol{y}|\boldsymbol{a}, \phi, \tau] = \sqrt{E_s} \, \boldsymbol{R}(\tau) \boldsymbol{a} e^{j\phi}. \tag{14}$$

The autocovariance matrix of $\boldsymbol{y}$ is

$$\mathrm{cov}[\boldsymbol{y}|\boldsymbol{a}, \phi, \tau] = \frac{N_o}{2} \boldsymbol{\Lambda} \tag{15}$$

where $\boldsymbol{\Lambda}$ is a $K \times K$ Toeplitz matrix defined as

$$\boldsymbol{\Lambda} = \begin{bmatrix} r(0T_s) & r(-T_s) & \cdots & r(-(K-1)T_s) \\ r(T_s) & r(0T_s) & \cdots & r(-(K-2)T_s) \\ \vdots & \vdots & \ddots & \vdots \\ r((K-1)T_s) & r((K-2)T_s) & \cdots & r(0T_s) \end{bmatrix} \tag{16}$$

with the $\{k, m\}$th entry of $\boldsymbol{\Lambda}$ equal to $r_{km} = r[(k - m)T_s]$. The likelihood function of ϕ and τ given $\boldsymbol{a}$ is

$$f(\boldsymbol{y}|\boldsymbol{a}, \phi, \tau) = \frac{\exp\left\{ -\frac{1}{2} (\boldsymbol{y} - \boldsymbol{m}_y)^H \left(\frac{N_o}{2} \boldsymbol{\Lambda} \right)^{-1} (\boldsymbol{y} - \boldsymbol{m}_y) \right\}}{(2\pi)^{K/2} \left| \frac{N_o}{2} \boldsymbol{\Lambda} \right|^{1/2}}. \tag{17}$$

The log-likelihood function is given by

$$\begin{aligned} l(\boldsymbol{y}|\boldsymbol{a}, \phi, \tau) &= \log(f(\boldsymbol{y}|\boldsymbol{a}, \phi, \tau)) \\ &= -\frac{1}{N_o} \left[-\boldsymbol{y}^H \boldsymbol{Q} \boldsymbol{m}_y - \boldsymbol{m}_y^H \boldsymbol{Q} \boldsymbol{y} + \boldsymbol{m}_y^H \boldsymbol{Q} \boldsymbol{m}_y \right] \\ &\quad - \left(\frac{1}{N_o} \boldsymbol{y}^H \boldsymbol{Q} \boldsymbol{y} + \log\left[(2\pi)^{K/2} \left| \frac{N_o}{2} \boldsymbol{\Lambda} \right|^{1/2} \right] \right) \end{aligned} \tag{18}$$

where $\boldsymbol{Q}$ is the inverse matrix of $\boldsymbol{\Lambda}$. Note that $\boldsymbol{\Lambda}$ is positive definite as long as the inverse exists. The likelihood function involves quadratic forms of the inverse matrix $\boldsymbol{Q}$. The $\mathrm{CRB}_{\mathrm{DA}}$'s (i.e., the CRB for DA estimation) are the diagonal entries of the inverse of the Fisher information matrix $\boldsymbol{J}$ for the joint estimation of $\{\phi, \tau\}$ [1], where $\boldsymbol{J}$ is defined as

$$\boldsymbol{J} = \begin{bmatrix} J_{\phi\phi} & J_{\phi\tau} \\ J_{\tau\phi} & J_{\tau\tau} \end{bmatrix} \tag{19}$$

whose entries are given by (let $\boldsymbol{\theta} = [\theta_1 \ \theta_2]$ with $\theta_1 = \phi$ and $\theta_2 = \tau$)

$$J_{\theta_i \theta_j} = E \left[-\frac{\partial^2 l(\boldsymbol{y}|\boldsymbol{a}, \phi, \tau)}{\partial \theta_i \partial \theta_j} \right]. \tag{20}$$

The expectation is with respect to $\boldsymbol{y}$, τ, and ϕ if ϕ and τ are random, or just $\boldsymbol{y}$ if ϕ and τ are deterministic [1]. Let us average $J_{\theta_i \theta_j}$ with respect to $\boldsymbol{y}$ first. If the result depends on ϕ and

τ, we can compute them further based on examining the condition whether ϕ and τ are deterministic or not. It can be readily verified that $J_{\theta_i \theta_j}$ is equal to

$$\begin{aligned} J_{\theta_i \theta_j} &= \frac{1}{N_0} \left[\frac{\partial \boldsymbol{m}_y^H}{\partial \theta_i} \boldsymbol{Q} \frac{\partial \boldsymbol{m}_y}{\partial \theta_j} + \frac{\partial \boldsymbol{m}_y^H}{\partial \theta_j} \boldsymbol{Q} \frac{\partial \boldsymbol{m}_y}{\partial \theta_i} \right] \\ &= \frac{2}{N_0} \Re \left[\frac{\partial \boldsymbol{m}_y^H}{\partial \theta_i} \boldsymbol{Q} \frac{\partial \boldsymbol{m}_y}{\partial \theta_j} \right] \end{aligned} \tag{21}$$

where $\Re(\cdot)$ represents the real part of a complex number. From (14), it follows that

$$\frac{\partial \boldsymbol{m}_y}{\partial \phi} = j \sqrt{E_s} \, \boldsymbol{R}(\tau) \boldsymbol{a} e^{j\phi}$$

$$\frac{\partial \boldsymbol{m}_y}{\partial \tau} = \sqrt{E_s} \frac{\partial \boldsymbol{R}(\tau)}{\partial \tau} \boldsymbol{a} e^{j\phi}.$$

Therefore, the entries of the Fisher information matrix (20) can be rewritten as

$$J_{\phi\phi} = \frac{2E_s}{N_0} \boldsymbol{a}^H \boldsymbol{R}(\tau)^H \boldsymbol{Q} \boldsymbol{R}(\tau) \boldsymbol{a} \tag{22}$$

$$J_{\phi\tau} = \frac{2E_s}{N_0} \Re \left[(-j) \boldsymbol{a}^H \boldsymbol{R}(\tau)^H \boldsymbol{Q} \frac{\partial \boldsymbol{R}(\tau)}{\partial \tau} \boldsymbol{a} \right] \tag{23}$$

$$J_{\tau\phi} = J_{\phi\tau} \tag{24}$$

$$J_{\tau\tau} = \frac{2E_s}{N_0} \boldsymbol{a}^H \frac{\partial \boldsymbol{R}(\tau)^H}{\partial \tau} \boldsymbol{Q} \frac{\partial \boldsymbol{R}(\tau)}{\partial \tau} \boldsymbol{a}. \tag{25}$$

The $\mathrm{CRB}_{\mathrm{DA}}$'s for the DA joint estimation of carrier phase and timing offsets are expressed as

$$E \left[(\phi - \hat{\phi})^2 \right] \geq \mathrm{CRB}_{\mathrm{DA}}(\phi) \triangleq \frac{J_{\tau\tau}}{J_{\phi\phi} J_{\tau\tau} - J_{\phi\tau}^2} \tag{26}$$

$$E \left[(\tau - \hat{\tau})^2 \right] \geq \mathrm{CRB}_{\mathrm{DA}}(\tau) \triangleq \frac{J_{\phi\phi}}{J_{\phi\phi} J_{\tau\tau} - J_{\phi\tau}^2}. \tag{27}$$

B. CRB in the Frequency Domain

Previous research was unable to derive a closed-form CRB for an arbitrary training sequence $\boldsymbol{a}$. In order to simplify the computation, $J_{\theta_i \theta_j}$ was approximated by being averaged over $\boldsymbol{a}$ based on the assumption that $\boldsymbol{a}$ is zero mean, i.i.d., and long enough [2]–[4], [6], [7].

Note that in our formulation, we have decoupled N, the length of the training sequence, and K, the length of the observation window. When the observation window increases, the estimator approaches the optimal solution for colored noise. Therefore, the resultant CRB is the ultimate performance lower bound. This decoupling enables us to apply the finite-term strong convergence theorem presented in Section II to derive the exact CRB for a training sequence of arbitrary length. We intend to asymptotically replace $\boldsymbol{Q}$ with $\boldsymbol{U}_K^H \boldsymbol{D}_K^{-1} \boldsymbol{U}_K$. Before applying the theorem, let us examine the conditions.

- The combined pulse-shaping function $r(t)$ is negligible for large t. This means that only a finite number of rows of $\boldsymbol{R}(\tau)$ have nonnegligible values. Thus, $\boldsymbol{m}_y$ can be considered to have a finite number of nonzero terms. This condition is equivalent to the requirement that the modulated

training sequence is completely transmitted within a finite time interval.

- The sequence $\{r(kT_s)\}$ defining $\boldsymbol{\Lambda}$ is the autocorrelation function of the noise process $\{N(kT_s)\}$. For a positive-definite $\boldsymbol{\Lambda}$, the DTFT of $\{r(kT_s)\}$ is always positive. However, oversampling typically degenerates the correlation matrix such that the inverse does not exist. This is a classical issue related to the likelihood function evaluation. In order to guarantee that the operation is meaningful, $\boldsymbol{\Lambda}$ should be positive definite. One technique to overcome the artifact in the literature is to assume that there is a small AWGN in N_k [1, p. 289].

- In regard to the condition that $\sum_k |kt_k| < \infty$, the pulse-shaping function $\{r(kT_s)\}$ usually converges to zero faster than $O(1/|k|^2)$ in practice. For instance, the magnitude of the raised cosine pulse converges to zero at a speed faster than $O(1/|k|^3)$.

In summary, all the conditions for applying the finite-term strong convergence theorem are satisfied. Thus, the inverse matrix $\boldsymbol{Q}$ can be substituted by the circulant matrix $\boldsymbol{U}_K^H \boldsymbol{D}_K^{-1} \boldsymbol{U}_K$ as $K \to \infty$. Note that the unitary matrix $\boldsymbol{U}_K$ is defined in (1) with $n = K$. The diagonal matrix $\boldsymbol{D}_K$ has the ith diagonal entry equal to $\mathcal{F}(2\pi i/K)$. Recall that the DTFT $\mathcal{F}(\omega)$ of $r(kT_s)$ $(k = \ldots, -1, 0, 1, \ldots)$ is defined as

$$\mathcal{F}(\omega) = \sum_{k=-\infty}^{\infty} r(kT_s) e^{-j\omega k}. \tag{28}$$

First, consider $\boldsymbol{U}_K \boldsymbol{R}(\tau)\boldsymbol{a}$. Its mth entry is equal to

$$\frac{1}{\sqrt{K}} \sum_{k=-K/2}^{K/2-1} \sum_{n=-N/2}^{N/2-1} r(kT_s - nT - \tau T) a_n e^{-j2\pi m(k+K/2)/K}. \tag{29}$$

For sufficiently large K, the summation with respect to k in (29) can be extended to $-\infty$ and ∞, since $r(kT_s)$ is negligible for large $|k|$. Equation (29) becomes

$$\frac{1}{\sqrt{K}} \sum_{n=-N/2}^{N/2-1} \sum_{k=-\infty}^{\infty} r(kT_s - nT - \tau T) a_n e^{-j2\pi m(k+K/2)/K}. \tag{30}$$

Let $\mathcal{R}(\omega)$ be the Fourier transform of $r(t)$, i.e.,

$$\mathcal{R}(\omega) = \int_{-\infty}^{\infty} r(t) e^{-j\omega t} \, dt. \tag{31}$$

Furthermore, $\mathcal{R}(\omega)$ is the power spectrum density of the noise process $N(t)$ and nonnegative. Define $\hat{\mathcal{R}}_{n,\tau}(\omega)$, the DTFT of $r(kT_s - nT - \tau T)$, $k = \ldots, -1, 0, 1, \ldots$, as

$$\hat{\mathcal{R}}_{n,\tau}(\omega) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} \mathcal{R}\left(\frac{\omega}{T_s} - \frac{2\pi k}{T_s}\right)$$
$$\cdot e^{-j(\tau+n)T(\omega/T_s - 2\pi k/T_s)}. \tag{32}$$

Clearly, $\hat{\mathcal{R}}_{0,0}(\omega) = \mathcal{F}(\omega)$ in (28). Let $\mathcal{A}(\omega)$ be the DTFT of $\boldsymbol{a}$, i.e.,

$$\mathcal{A}(\omega) = \sum_{n=-N/2}^{N/2-1} a_n e^{-j\omega n}. \tag{33}$$

Substituting (32) and (33) into (30), (30) becomes

$$\frac{1}{\sqrt{K}} \sum_{n=-N/2}^{N/2-1} \hat{\mathcal{R}}_{n,\tau}\left(\frac{2\pi m}{K}\right) a_n e^{-j\pi m}$$
$$= \frac{1}{T_s\sqrt{K}} \sum_{k=-\infty}^{\infty} \mathcal{R}\left(\frac{2\pi m}{KT_s} - \frac{2\pi k}{T_s}\right) \mathcal{A}\left(\frac{2\pi mL}{K} - 2\pi kL\right)$$
$$\cdot e^{-j(\tau(2\pi mL/K - 2\pi kL) + \pi m)}. \tag{34}$$

It is easy to verify that the mth entry of $\boldsymbol{U}_K \boldsymbol{R}(\tau)\boldsymbol{a}$ given by (34) is a periodic function of m with period K, which follows the property of DTFT. Based on the following equation:

$$\boldsymbol{U}_K \frac{\partial \boldsymbol{R}(\tau)}{\partial \tau} \boldsymbol{a} = \frac{\partial}{\partial \tau} \boldsymbol{U}_K \boldsymbol{R}(\tau)\boldsymbol{a} \tag{35}$$

the mth entry of (35) is also a periodic function of m with the same period. Note that the DTFT of $\partial r(kT_s - nT - \tau T)/\partial \tau$ is

$$\hat{\mathcal{R}}'_{n,\tau}(\omega) = \frac{-T}{T_s} \sum_{k=-\infty}^{\infty} j\left(\frac{\omega}{T_s} - \frac{2\pi k}{T_s}\right) \mathcal{R}\left(\frac{\omega}{T_s} - \frac{2\pi k}{T_s}\right)$$
$$\cdot e^{-j(\tau+n)T(\omega/T_s - 2\pi k/T_s)}. \tag{36}$$

Introduce the following shorthand notation:

$$\mathcal{RA}(m, k) \triangleq \frac{1}{T_s} \mathcal{R}\left(\frac{2\pi m}{KT_s} - \frac{2\pi k}{T_s}\right) \mathcal{A}\left(\frac{2\pi mL}{K} - 2\pi kL\right). \tag{37}$$

With this notation, (22) becomes

$$J_{\phi\phi} = \frac{2E_s}{N_0 K} \sum_{m=0}^{K-1} \frac{1}{\mathcal{F}(2\pi m/K)}$$
$$\cdot \left[\sum_{k,l=-\infty}^{\infty} \mathcal{RA}(m, k)\mathcal{RA}(m, l)^* e^{j2\pi\tau(k-l)L} \right], \tag{38}$$

(23) becomes

$$J_{\phi\tau} = -\frac{2E_s T}{N_0 K} \Re\left\{ \sum_{m=0}^{K-1} \frac{1}{\mathcal{F}(2\pi m/K)} \left[\sum_{k,l=-\infty}^{\infty} \mathcal{RA}(m, k) \right. \right.$$
$$\left. \left. \cdot \mathcal{RA}(m, l)^* \left(\frac{2\pi m}{KT_s} - \frac{2\pi k}{T_s}\right) e^{j2\pi\tau(k-l)L} \right] \right\} \tag{39}$$

and (25) becomes

$$J_{\tau\tau} = \frac{2E_s T^2}{N_0 K} \sum_{m=0}^{K-1} \frac{1}{\mathcal{F}(2\pi m/K)}$$
$$\cdot \left[\sum_{k,l=-\infty}^{\infty} \mathcal{RA}(m, k)\mathcal{RA}(m, l)^* \left(\frac{2\pi m}{KT_s} - \frac{2\pi k}{T_s}\right) \right.$$
$$\left. \cdot \left(\frac{2\pi m}{KT_s} - \frac{2\pi l}{T_s}\right) e^{j2\pi\tau(k-l)L} \right]. \tag{40}$$

Although the summation over k, l is from $-\infty$ to ∞, in all the cases of interest, we shall see that there are at most three nonzero terms due to the fact that the pulse-shaping function $r(t)$ is band-limited.

1) CRB for Oversampling: In the case of oversampling, $1/T_s \geq 2B$ with B the single-sided bandwidth of $r(t)$, i.e.,

$$\mathcal{R}(\omega) = 0, \ |\omega| > \frac{\pi}{T_s}.$$

There is no aliasing of $\mathcal{F}(\omega)$ in the frequency domain. Because the entries of $\boldsymbol{U}_K \boldsymbol{R}(\tau) \boldsymbol{a}$ or $\boldsymbol{U}_K \frac{\partial \boldsymbol{R}(\tau)}{\partial \tau} \boldsymbol{a}$ are periodic with period K, the summations in (38)–(40) can range from $[-K/2, K/2-1]$ instead of $[0, K-1]$. In this case, $\mathcal{R}\mathcal{A}(m, k)$ is nonzero only for $k = 0$ when $m \in [-K/2, K/2-1]$. Also note that in the case of oversampling, $\mathcal{F}(2\pi m/K) = \mathcal{R}(2\pi m/KT_s)/T_s$. Therefore, (38)–(40) can be reduced to the following simple forms:

$$J_{\phi\phi} = \frac{2E_s}{N_0 K} \sum_{m=-K/2}^{K/2-1} \frac{1}{T_s} \mathcal{R}\left(\frac{2\pi m}{KT_s}\right) \left|\mathcal{A}\left(\frac{2\pi mL}{K}\right)\right|^2 \quad (41)$$

$$J_{\phi\tau} = -\frac{2E_s}{N_0 K} \sum_{m=-K/2}^{K/2-1} \left(\frac{2\pi mL}{K}\right) \frac{1}{T_s} \mathcal{R}\left(\frac{2\pi m}{KT_s}\right)$$
$$\cdot \left|\mathcal{A}\left(\frac{2\pi mL}{K}\right)\right|^2 \quad (42)$$

$$J_{\tau\tau} = \frac{2E_s}{N_0 K} \sum_{m=-K/2}^{K/2-1} \left(\frac{2\pi mL}{K}\right)^2 \frac{1}{T_s} \mathcal{R}\left(\frac{2\pi m}{KT_s}\right)$$
$$\cdot \left|\mathcal{A}\left(\frac{2\pi mL}{K}\right)\right|^2. \quad (43)$$

These values are independent of ϕ and τ. This implies that the CRBs for timing and carrier phase recovery are the same for both random and deterministic ϕ and τ.

As the observation window $K \to \infty$, we obtain the following entries of the Fisher information matrix:

$$J_{\phi\phi} = \frac{1}{\pi T} \frac{E_s}{N_0} \int_{-\infty}^{\infty} \mathcal{R}\left(\frac{\omega}{T}\right) |\mathcal{A}(\omega)|^2 \, d\omega \quad (44)$$

$$J_{\phi\tau} = -\frac{1}{\pi T} \frac{E_s}{N_0} \int_{-\infty}^{\infty} \omega \mathcal{R}\left(\frac{\omega}{T}\right) |\mathcal{A}(\omega)|^2 \, d\omega \quad (45)$$

$$J_{\tau\tau} = \frac{1}{\pi T} \frac{E_s}{N_0} \int_{-\infty}^{\infty} \omega^2 \mathcal{R}\left(\frac{\omega}{T}\right) |\mathcal{A}(\omega)|^2 \, d\omega. \quad (46)$$

2) CRB for Undersampling: In the case of undersampling, $1/T_s < 2B$. There is aliasing in the frequency domain. In this scenario, $J_{\phi\phi}$, $J_{\phi\tau}$, and $J_{\tau\tau}$ in (38)–(40) are independent of ϕ but dependent on τ, and they should be averaged with respect to τ. In practice, it is reasonable to model τ as a uniformly distributed random variable in the receiver front end. A case of particular interest is one sample per symbol. The following equation is true for arbitrary integers k and l

$$\int_{-1/2}^{1/2} e^{j2\pi\tau(k-l)} \, d\tau = \delta[k-l]$$

where $\delta[n] = 1$, if $n = 0$; $\delta[n] = 0$, otherwise. Therefore, $J_{\phi\phi}$, $J_{\phi\tau}$, and $J_{\tau\tau}$ become

$$J_{\phi\phi} = \frac{2E_s}{N_0 K T_s^2} \sum_{m=-K/2}^{K/2-1} \frac{\left|\mathcal{A}\left(\frac{2\pi mL}{K}\right)\right|^2}{\mathcal{F}(2\pi m/K)}$$
$$\cdot \left[\sum_{k=-\infty}^{\infty} \mathcal{R}\left(\frac{2\pi m}{KT_s} - \frac{2\pi k}{T_s}\right)^2\right] \quad (47)$$

$$J_{\phi\tau} = -\frac{2E_s}{N_0 K T_s^2} \sum_{m=-K/2}^{K/2-1} \frac{\left|\mathcal{A}\left(\frac{2\pi mL}{K}\right)\right|^2}{\mathcal{F}(2\pi m/K)}$$
$$\cdot \left[\sum_{k=-\infty}^{\infty} \left(\frac{2\pi mL}{K} - 2\pi Lk\right) \mathcal{R}\left(\frac{2\pi m}{KT_s} - \frac{2\pi k}{T_s}\right)^2\right] \quad (48)$$

$$J_{\tau\tau} = \frac{2E_s}{N_0 K T_s^2} \sum_{m=-K/2}^{K/2-1} \frac{\left|\mathcal{A}\left(\frac{2\pi mL}{K}\right)\right|^2}{\mathcal{F}(2\pi m/K)}$$
$$\cdot \left[\sum_{k=-\infty}^{\infty} \left(\frac{2\pi mL}{K} - 2\pi Lk\right)^2 \mathcal{R}\left(\frac{2\pi m}{KT_s} - \frac{2\pi k}{T_s}\right)^2\right]. \quad (49)$$

Equations (47)–(49) follow the fact that $\mathcal{A}(\omega - 2\pi k) = \mathcal{A}(\omega)$ for an integer k. This allows separating $\mathcal{A}(\omega)$ from the aliased $\mathcal{R}(\omega)^2$. In practice, a pulse-shaping function is always band-limited. Its effective single-sided bandwidth B typically limits to $1/T$. This means that the nonzero terms in (32) are only those with $k = -1, 0, 1$. As in the case of oversampling, we obtain the following entries of the Fisher information matrix in integral forms by increasing the observation window $K \to \infty$:

$$J_{\phi\phi} = \frac{1}{\pi T} \frac{E_s}{N_0} \int_{-\pi}^{\pi} \frac{\sum_{k=-\infty}^{\infty} \mathcal{R}\left(\frac{\omega-2\pi k}{T}\right)^2 |\mathcal{A}(\omega)|^2}{\sum_{k=-\infty}^{\infty} \mathcal{R}\left(\frac{\omega-2\pi k}{T}\right)} \, d\omega \quad (50)$$

$$J_{\phi\tau} = -\frac{1}{\pi T} \frac{E_s}{N_0} \int_{-\pi}^{\pi}$$
$$\frac{\sum_{k=-\infty}^{\infty} (\omega - 2\pi k) \mathcal{R}\left(\frac{\omega-2\pi k}{T}\right)^2 |\mathcal{A}(\omega)|^2}{\sum_{k=-\infty}^{\infty} \mathcal{R}\left(\frac{\omega-2\pi k}{T}\right)} \, d\omega \quad (51)$$

$$J_{\tau\tau} = \frac{1}{\pi T} \frac{E_s}{N_0} \int_{-\pi}^{\pi}$$
$$\frac{\sum_{k=-\infty}^{\infty} (\omega - 2\pi k)^2 \mathcal{R}\left(\frac{\omega-2\pi k}{T}\right)^2 |\mathcal{A}(\omega)|^2}{\sum_{k=-\infty}^{\infty} \mathcal{R}\left(\frac{\omega-2\pi k}{T}\right)} \, d\omega. \quad (52)$$

Combining (44)–(46) and (50)–(52), we can recapitulate the results for both the over- and undersampling cases formally as follows.

Theorem 2: For a DA joint timing and carrier phase offsets estimator given an arbitrary training sequence $\boldsymbol{a}$ of length N, the mean-squared estimation error for timing offset τ is lower-bounded by the following CRB:

$$E\left[(\tau - \hat{\tau})^2\right] \geq \mathrm{CRB}_{\mathrm{DA}}(\tau) = \frac{J_{\phi\phi}}{J_{\phi\phi}J_{\tau\tau} - J_{\phi\tau}^2} \quad (53)$$

and the mean-squared estimation error for carrier phase offset ϕ is lower-bounded by the following CRB:

$$E\left[(\phi - \hat{\phi})^2\right] \geq \mathrm{CRB}_{\mathrm{DA}}(\phi) = \frac{J_{\tau\tau}}{J_{\phi\phi}J_{\tau\tau} - J_{\phi\tau}^2}. \quad (54)$$

When the sampling rate is larger than or equal to the Nyquist sampling rate, $J_{\phi\phi}$, $J_{\phi\tau}$, and $J_{\tau\tau}$ are given by (44)–(46). When

the sampling rate is one sample per symbol, $J_{\phi\phi}$, $J_{\phi\tau}$, and $J_{\tau\tau}$ are given by (50)–(52). In these equations, $\mathcal{R}(\omega)$ is the Fourier transform of the cascaded response of the pulse-shaping filter and the matched filter as defined in (31), and $\mathcal{A}(\omega)$ is the discrete-time Fourier transform of the training sequence as defined in (33).

For the Nyquist pulse-shaping function with one sample per symbol, the entries of the Fisher information matrix can be simplified as follows.

Corollary 1: For the Nyquist pulse-shaping function and the one-sample-per-symbol sampling rate, we have

$$J_{\phi\phi} = \frac{1}{\pi T^2} \frac{E_s}{N_0} \int_{-\infty}^{\infty} \mathcal{R}\left(\frac{\omega}{T}\right)^2 |\mathcal{A}(\omega)|^2 \, d\omega \qquad (55)$$

$$J_{\phi\tau} = -\frac{1}{\pi T^2} \frac{E_s}{N_0} \int_{-\infty}^{\infty} \omega \mathcal{R}\left(\frac{\omega}{T}\right)^2 |\mathcal{A}(\omega)|^2 \, d\omega \qquad (56)$$

$$J_{\tau\tau} = \frac{1}{\pi T^2} \frac{E_s}{N_0} \int_{-\infty}^{\infty} \omega^2 \mathcal{R}\left(\frac{\omega}{T}\right)^2 |\mathcal{A}(\omega)|^2 \, d\omega. \qquad (57)$$

Proof: With the Nyquist assumption, we have

$$\sum_k \mathcal{R}(\omega - 2\pi k/T) = T.$$

Therefore, the denominator of the integrand in (50)–(52) is equal to a constant. The summation of the nominator simply extends the integral from $[-\pi, \pi]$ to $(-\infty, \infty)$. This proves the assertion. $\qquad\square$

When compared with the oversampling case, the undersampling bound with the Nyquist pulse-shaping function simply replaces $\mathcal{R}(\cdot)$ with $\mathcal{R}(\cdot)^2$. In both cases, the frequency-domain multiplication of the pulse-shaping function and the training sequence appears in $J_{\theta_i\theta_j}$. This is the reason that we call them the *frequency-domain* CRBs.

The next subsection presents the CRB in the time domain. It will become clear that the time-domain representations are rather simple to compute and no integral is necessary.

C. CRB in the Time Domain

For convenience, we perform zero padding to the training sequence, i.e., define $a_i = 0$ for $|i| > N/2$. Note that the "zero" means physical zero rather than logical zero. With this notation, $|\mathcal{A}(\omega)|^2$ is equal to the Fourier transform of the following continuous time function:

$$\sum_{n=-\infty}^{\infty} \sum_{i=-N}^{N} a_i a_{i-n}^* \delta(t-n) \qquad (58)$$

where $\delta(t)$ is the Dirac function such that

$$\int_{-\infty}^{\infty} \delta(t) = 1.$$

The Fourier transform of $r(tT)$ is equal to

$$\int_{-\infty}^{\infty} r(tT) e^{-j\omega t} = \frac{1}{T} \mathcal{R}(\omega/T). \qquad (59)$$

Using Parseval's relation we can substitute the terms in (44)–(46) and (55)–(57) with (58) and (59), which leads to the time-domain representation of $J_{\phi\phi}$, $J_{\phi\tau}$, and $J_{\tau\tau}$.

Theorem 3: When the sampling rate is equal to or above the Nyquist sampling rate, we have

$$J_{\phi\phi} = \frac{2E_s}{N_0} \sum_{n=-N}^{N} r(nT) \sum_{i=-N}^{N} a_i a_{i-n}^* \qquad (60)$$

$$J_{\phi\tau} = j \frac{2E_s}{N_0} \sum_{n=-N}^{N} \left.\frac{dr(tT)}{dt}\right|_{t=n} \sum_{i=-N}^{N} a_i a_{i-n}^* \qquad (61)$$

$$J_{\tau\tau} = -\frac{2E_s}{N_0} \sum_{n=-N}^{N} \left.\frac{d^2 r(tT)}{dt^2}\right|_{t=n} \sum_{i=-N}^{N} a_i a_{i-n}^*. \qquad (62)$$

When the sampling rate is one sample per symbol and the combined pulse shaping function is Nyquist, we have

$$J_{\phi\phi} = \frac{2E_s}{N_0} \sum_{n=-N}^{N} u(nT) \sum_{i=-N}^{N} a_i a_{i-n}^* \qquad (63)$$

$$J_{\phi\tau} = j \frac{2E_s}{N_0} \sum_{n=-N}^{N} \left.\frac{du(tT)}{dt}\right|_{t=n} \sum_{i=-N}^{N} a_i a_{i-n}^* \qquad (64)$$

$$J_{\tau\tau} = -\frac{2E_s}{N_0} \sum_{n=-N}^{N} \left.\frac{d^2 u(tT)}{dt^2}\right|_{t=n} \sum_{i=-N}^{N} a_i a_{i-n}^* \qquad (65)$$

where $u(tT)$ is the convolution of $r(tT)$ and $r(-tT)$, i.e., $u(tT) \triangleq r(tT) \otimes r(-tT)$.

The time-domain expressions do not require an integral, which is an advantage. Furthermore, the derivative of $r(tT)$ or $u(tT)$ can be computed independently from the training sequence since they are only functions of the shaping pulse. For a given pulse-shaping function, these derivatives can be precomputed and used for the design of training sequences. Now it is clear that the dependency of the CRB on a training sequence is only related to the aperiodic correlation of the sequence.

Remark: In the literature, the CRB has been investigated with the assumption that the noise with multiple samples per symbol is white. This assumption holds when the antialiasing filter is rectangular and the bound is evaluated before the matched filter. The validity of the bound for the samples after the matched filter is due to that any sufficient statistics give precisely the same CRB. We could have applied that argument to simplify the derivation in this manuscript. However, we feel that the current approach is advantageous. For instance, if the noise before the matched filter is colored rather than white, previous methodology cannot be directly applied. In order to adopt the white noise assumption, prewhitening has to be applied. However, it is not immediately clear how the whitening filter should be accounted for in the final CRB. Using the approach presented here, we can simply replace $\mathcal{F}(\cdot)$ in (38)–(40) with the power spectral density of the noise process to account for any colored Gaussian noise. Essentially, we have proved that the DFT itself decorrelates the noise for computing the CRB.

IV. ANALYSIS OF THE CRB AND EXAMPLES

In this section, we analyze the CRB obtained and compare it with the known bound in the literature. Furthermore, we present

a few interesting examples to illustrate the performance limits offered by different training sequences.

A. $J_{\phi\tau}$: Cost of Two Unknown Parameters

Since $J_{\phi\tau}^2 \geq 0$, the following can be obtained from (26) and (27):

$$\frac{J_{\tau\tau}}{J_{\phi\phi}J_{\tau\tau} - J_{\phi\tau}^2} \geq \frac{J_{\tau\tau}}{J_{\phi\phi}J_{\tau\tau}} = \frac{1}{J_{\phi\phi}} \qquad (66)$$

$$\frac{J_{\phi\phi}}{J_{\phi\phi}J_{\tau\tau} - J_{\phi\tau}^2} \geq \frac{J_{\phi\phi}}{J_{\phi\phi}J_{\tau\tau}} = \frac{1}{J_{\tau\tau}}. \qquad (67)$$

It is straightforward to verify that the CRB_{DA} for timing/carrier phase estimation with known carrier phase/timing offset is equal to $1/J_{\tau\tau}$ $(1/J_{\phi\phi})$, respectively. Therefore, $J_{\phi\tau}$ is the *cost* in terms of larger estimation variance when the other parameter is unknown. It was observed in [2] that $J_{\phi\tau}$ reduces to zero for pseudorandom training sequences and real shaping pulse. We shall show that this observation can be easily extended to any real training sequence.

- In the oversampling case, $J_{\phi\tau}$ is given by (45). Given the assumption that $r(t)$ is real, its Fourier transform $\mathcal{R}(\omega)$ is an even function of ω. With ω an odd function, an even $|\mathcal{A}(\omega)|$ will guarantee $J_{\phi\tau} = 0$. In the under sampling case, $J_{\phi\tau}$ is given by (51). For real $r(t)$ and even $|\mathcal{A}(\omega)|$, we have

$$\begin{aligned} &\frac{\sum_{k=-\infty}^{\infty} (-\omega - 2\pi k)\mathcal{R}\left(\frac{-\omega - 2\pi k}{T}\right)|\mathcal{A}(-\omega)|^2}{\sum_{k=-\infty}^{\infty} \mathcal{R}\left(\frac{-\omega - 2\pi k}{T}\right)} \\ &= \frac{\sum_{k=-\infty}^{\infty} (-\omega + 2\pi k)\mathcal{R}\left(\frac{-\omega + 2\pi k}{T}\right)|\mathcal{A}(-\omega)|^2}{\sum_{k=-\infty}^{\infty} \mathcal{R}\left(\frac{-\omega + 2\pi k}{T}\right)} \\ &= -\frac{\sum_{k=-\infty}^{\infty} (\omega - 2\pi k)\mathcal{R}\left(\frac{\omega - 2\pi k}{T}\right)|\mathcal{A}(\omega)|^2}{\sum_{k=-\infty}^{\infty} \mathcal{R}\left(\frac{\omega - 2\pi k}{T}\right)} \end{aligned} \qquad (68)$$

  where the first equality follows from substituting k with $-k$, the second equality is due to the fact that both $\mathcal{R}(\omega)$ and $|\mathcal{A}(\omega)|$ are even functions. Therefore, we have shown that the integrand of (51) is an odd function of ω. Therefore, as long as $|\mathcal{A}(\omega)|$ is an even function of ω, $J_{\phi\tau}$ is equal to zero.

- A sufficient condition to make $J_{\phi\tau} = 0$ is that the training sequence $\boldsymbol{a}$ is real, which implies that $|\mathcal{A}(\omega)|$ is an even function.

Previous discussion shows that under very realistic conditions we are able to separate the estimation of carrier phase and timing. In the sequel, we focus on the bound with $J_{\phi\tau} = 0$.

B. Comparison With the Known Bound

First, for $J_{\phi\phi}$, (60) reduces to the known results [2]. As observed in [2], for the Nyquist pulse-shaping function, the CRB for phase estimation becomes

$$J_{\phi\phi} = 2NE_s/N_0, \qquad (69)$$

with the assumption that $E[|a_m|^2] = 1$ and $r(0) = 1$.

For timing estimation, in the literature [2]–[6], the CRB is only available for the case that the training sequence is i.i.d. pseudorandom data and very long. When the sampling rate $L \geq 2$, the CRB (denoted as $\text{CRB}_{\text{RD}}(\tau)$) is given by

$$\text{CRB}_{\text{RD}}(\tau) = \frac{1}{T^2}\left\{\frac{2E_s}{N_0}N\int_{-\infty}^{\infty} 4\pi^2 f^2 \mathcal{R}(2\pi f)\,df\right\}^{-1}. \qquad (70)$$

When $L = 1$, for the Nyquist pulse-shaping function, $\text{CRB}_{\text{RD}}(\tau)$ is given by

$$\text{CRB}_{\text{RD}}(\tau) = \frac{1}{T^2}\left\{\frac{2E_s}{N_0}N\frac{1}{T}\int_{-\infty}^{\infty} \dot{r}^2(t)\,dt\right\}^{-1}. \qquad (71)$$

In the sequel, we show that (70) and (71) are the special cases of (46) and (57) under the same assumptions. For zero mean, i.i.d. pseudorandom training sequence $\boldsymbol{a}$, as N goes to ∞, the power spectral density of $\boldsymbol{a}$ becomes white, i.e.,

$$|\mathcal{A}(\cdot)|^2 \approx N. \qquad (72)$$

Substitute $|\mathcal{A}(\omega)|^2$ with N in (46) and (57), let $\omega/T = 2\pi f$, the equivalence of (46) to the known bound (70) under this condition follows. The equivalence for the undersampling case can be established by further applying Parseval's relation

$$\int_{-\infty}^{\infty} \dot{r}(t)^2\,dt = \int_{-\infty}^{\infty} (2\pi f \mathcal{R}(2\pi f))^2\,df. \qquad (73)$$

From the representations of the CRB in the time domain, it is clear that the difference between the i.i.d. pseudorandom sequence and any fixed training sequence depends on $d^2r(tT)/dt^2|_{t=n}$ for $n \neq 0$. If $\{|d^2r(tT)/dt^2|_{t=n}; n \neq 0\}$ are very small compared with $d^2r(tT)/dt^2|_{t=0}$, the new CRB would be of minor interest. The following results show that $\{|d^2r(tT)/dt^2|_{t=n}; n \neq 0\}$ are indeed significant compared with $d^2r(tT)/dt^2|_{t=0}$.

Corollary 2: If the signal $r(t)$ is band-limited by $2\pi/T$, i.e.,

$$\mathcal{R}(\omega) = 0, \ |\omega| > \frac{2\pi}{T}$$

we have

$$\sum_{n=-\infty}^{\infty} \frac{dr(tT)}{dt}\bigg|_{t=n} = 0$$

and

$$\sum_{n=-\infty}^{\infty} \frac{d^2r(tT)}{dt^2}\bigg|_{t=n} = 0.$$

Proof: The DTFT of $\{dr(tT)/dt|_{t=n}; n\}$ is equal to

$$\frac{1}{T}\sum_{n=-\infty}^{\infty} \frac{\omega - 2\pi n}{T}\mathcal{R}\left(\frac{\omega - 2\pi n}{T}\right).$$

For $\omega = 0$, the summation is equal to zero, which means that there is no DC component. In the time domain, this implies that

$$\sum_{n=-\infty}^{\infty} \frac{dr(tT)}{dt}\bigg|_{t=n} = 0.$$

Similarly, we can show that

$$\sum_{n=-\infty}^{\infty} \left.\frac{d^2 r(tT)}{dt^2}\right|_{t=n} = 0. \qquad \square$$

This simple result shows that the combined effect of all the terms other than $d^2 r(tT)/dt^2|_{t=0}$ in (61) and (62) is as significant as $d^2 r(tT)/dt^2|_{t=0}$. Corollary 2 reveals that by properly designing the training sequence, significant gain can be obtained over the pseudorandom sequence.

C. Examples

In the sequel, we use a few typical training data patterns to illustrate the bound. The raised cosine shaping function is assumed to be the cascaded filter, whose Fourier transform is given by

$$\mathcal{R}(\omega) = \begin{cases} T, & 0 \le |\omega| \le \pi(1-\alpha)/T \\ \frac{T}{2}\left[1 - \sin\left(\frac{T|\omega|-\pi}{2\alpha}\right)\right], & \pi(1-\alpha)/T \le |\omega| \\ & \qquad \le \pi(1+\alpha)/T \\ 0, & \text{otherwise.} \end{cases} \tag{74}$$

When the rolloff factor α ranges from 0 to 1, its single-sided bandwidth ranges from $1/2T$ to $1/T$.

Example 1—Continuous Wave : The continuous wave (CW) sequence is the training sequence with data pattern

$$a_m = (1+j)/\sqrt{2} \qquad \text{(for } m = -N/2 - 1, \ldots, N/2).$$

It is widely adopted in burst preambles in time-division multiple access (TDMA) networks for carrier acquisition. We consider two different scenarios in the oversampling case: infinite-length CW and finite-length CW. For the infinite-length CW, we have

$$\mathcal{A}(\omega) = \frac{1+j}{\sqrt{2}} \sum_{n=-\infty}^{\infty} e^{-j\omega n}$$

$$= \frac{1+j}{\sqrt{2}} \sum_{n=-\infty}^{\infty} 2\pi\delta(\omega - 2\pi n). \tag{75}$$

Substituting (75) into (46), $J_{\tau\tau} = 0$ follows from the fact that $\mathcal{R}(\omega/T) = 0$ for $|\omega| \ge 2\pi$. Therefore, the infinite-length CW provides no timing information as heuristically explained in [2, p. 336]. For a finite-length CW, the autocorrelation of the training sequence is given by

$$\sum_{i=-N}^{N} a_i a_{i-n}^* = \begin{cases} N - |n|, & |n| \le N \\ 0, & \text{otherwise.} \end{cases} \tag{76}$$

Substituting (76) into (62), we obtain

$$J_{\tau\tau} =$$
$$\frac{2E_s}{N_0}\left(\sum_{n=-N}^{N} |n| \left.\frac{d^2 r(tT)}{dt^2}\right|_{t=n} - N \sum_{n=-N}^{N} \left.\frac{d^2 r(tT)}{dt^2}\right|_{t=n} \right).$$

By using the following result shown in the previous section:

$$\sum_{n=-\infty}^{\infty} \left.\frac{d^2 r(tT)}{dt^2}\right|_{t=n} = 0$$

TABLE I
$J_{\tau\tau}/(2E_s/N_0)$ OF CW GIVEN DIFFERENT SEQUENCE LENGTH N AND ROLLOFF FACTOR α, WHERE THE SAMPLING RATE IS TWO SAMPLES PER SYMBOL ($L = 2$)

α	$N = 6$	$N = 10$	$N = 20$	$N = 50$	$N = 100$
0.25	2.8189	2.8364	2.8328	2.8330	2.8330
0.50	3.0331	3.0403	3.0417	3.0416	3.0416
0.75	3.5435	3.5286	3.5326	3.5329	3.5329
1.00	5.5314	5.5402	5.5439	5.5450	5.5450

and the fact that $d^2 r(tT)/dt^2|_{t=n}$ becomes virtually zero for large n, we have

$$\lim_{N\to\infty} \sum_{n=-N}^{N} N \left.\frac{d^2 r(tT)}{dt^2}\right|_{t=n} = -\lim_{N\to\infty} \sum_{|n|>N} N \left.\frac{d^2 r(tT)}{dt^2}\right|_{t=n}$$
$$= 0.$$

This shows that for a long but finite-length CW, $J_{\tau\tau}$ converges to

$$\frac{2E_s}{N_0} \sum_{n=-\infty}^{\infty} |n| \left.\frac{d^2 r(tT)}{dt^2}\right|_{t=n} \ne 0.$$

Table I illustrates $J_{\tau\tau}/(2E_s/N_o)$ of CW sequences given different rolloff factors and sequence length. It shows that $J_{\tau\tau}$ converges to a nonzero constant as N increases. This constant is related to the pulse-shaping function but independent of N. The nonzero $J_{\tau\tau}$ captures the impact of the ramping up and ramping down of a finite-length CW. No matter how long a training sequence is, the ramping effect will not disappear. This explains why the CRB of a finite-length CW cannot converge to that of the infinite-length CW. Although CW itself has little value for timing synchronization, it is a very good pedagogical example. It shows that the bound can capture the detailed impact of a training sequence on timing estimation performance. For other sequences such as pseudorandom sequences, as shown previously, the bound increases with N. Therefore, when it is normalized by the length of the sequence, the boundary ramping effect will not play a visible role.

Example 2—Finite-Length Pseudo Random Sequence:

When we assume that the training data is i.i.d. but of finite length, $J_{\tau\tau}$ in (62) reduces to

$$\frac{J_{\tau\tau}}{N} = -\frac{2E_s}{N_0} \left.\frac{d^2 r(tT)}{dt^2}\right|_{t=0}.$$

Thus, a finite-length pseudorandom sequence has the same normalized timing bound as an infinite-length one does. This is relevant when applying the bound to the tracking mode operation of a synchronization circuit. In that case, for fixed tracking loop bandwidth, the training sequence can be assumed to be i.i.d. but of fixed length. Therefore, the result in this paper shows that the bound for an infinite-length pseudorandom sequence applies to that of a fixed-length sequence. Another interesting case is the bound of a one-bit training sequence. We can easily see that the CRB for a one-bit training sequence gives the exact CRB of a pseudorandom sequence normalized by N.
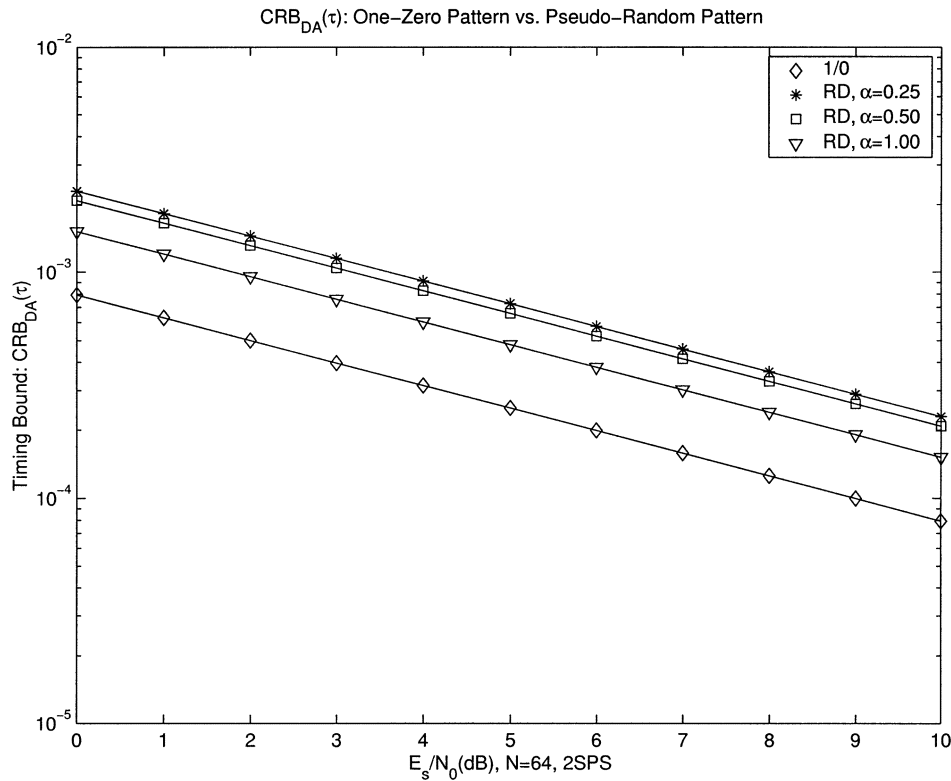
Fig. 2.   Timing bound $\mathrm{CRB}_{\mathrm{DA}}(\tau)$ for alternating one–zero and pseudorandom sequences in the oversampling case, where $N = 64$, $L = 2$, "1/0" denotes the alternating one–zero sequence, and "RD" denotes the pseudorandom sequence.

Example 3—Alternating One–Zero Sequence: The alternating one–zero pattern is the training sequence with data pattern $a_m = \sqrt{2}/2(1 + j)(-1)^m$. It is widely used as preamble in TDMA networks for timing recovery. Its autocorrelation is

$$\sum_{i=-N}^{N} a_i a_{i-n}^* = \begin{cases} (-1)^n(N - |n|), & |n| \leq N \\ 0, & \text{otherwise.} \end{cases}$$

As the sequence length $N \to \infty$, substituting this into (62), we obtain

$$\lim_{N \to \infty} \frac{J_{\tau\tau}}{N} = -\frac{2E_s}{N_0} \sum_{n=-\infty}^{\infty} (-1)^n \left. \frac{d^2 r(tT)}{dt^2} \right|_{t=n}. \qquad (77)$$

The Fourier transform of the sequence $\{(-1)^n\}$ is

$$\sum_{n=-\infty}^{\infty} (-1)^n e^{-j\omega n} = \sum_{n=-\infty}^{\infty} 2\pi\delta(\omega - \pi - 2\pi n).$$

Applying Parseval's relation, we have

$$-\frac{2E_s}{N_0} \sum_{n=-\infty}^{\infty} (-1)^n \left. \frac{d^2 r(tT)}{dt^2} \right|_{t=n}$$

$$= \frac{E_s}{\pi T N_0} \int_{-\pi}^{\pi} \omega^2 \mathcal{R}(\omega/T) 2\pi \sum_{n=-\infty}^{\infty} \delta(\omega - \pi - 2\pi n)\, d\omega$$

$$= \frac{2E_s}{N_0} \left(\pi^2 \mathcal{R}(\pi/T)/T + \pi^2 \mathcal{R}(-\pi/T)/T \right).$$

For a Nyquist band-limited pulse-shaping function, $\mathcal{R}(\omega) = 0$ for $|\omega| \geq 2\pi/T$, we have $\mathcal{R}(\pi/T)/T = \mathcal{R}(-\pi/T)/T = 1/2$. This shows that

$$-\frac{2E_s}{N_0} \sum_{n=-\infty}^{\infty} (-1)^n \left. \frac{d^2 r(tT)}{dt^2} \right|_{t=n} = \frac{2\pi^2 E_s}{N_0}.$$

Therefore, the $\mathrm{CRB}_{\mathrm{DA}}(\tau)$ for a long alternating one–zero sequence is independent of the pulse-shaping function as long as it is Nyquist. In contrast, for a pseudorandom sequence, the normalized bound is equal to

$$-\frac{2E_s}{N_0} \left. \frac{d^2 r(tT)}{dt^2} \right|_{t=0}$$

with $d^2 r(tT)/dt^2|_{t=0}$ dependent on the pulse-shaping function. For the raised cosine pulse, it is easy to show that $-d^2 r(tT)/dt^2|_{t=0} = (\pi^2 + 3\alpha^2(\pi^2 - 8))/3$.[1] Thus, an alternating one–zero sequence offers a performance improvement over a pseudorandom sequence of the same length by a factor of

$$\frac{3\pi^2}{\pi^2 + 3\alpha^2(\pi - 8)}.$$

Fig. 2 illustrates the comparison of the timing bound between an alternating one–zero sequence and a pseudorandom sequence given $N = 64$, $L = 2$, and different rolloff factors. For small α, the gain approaches $10 \log 3 \approx 4.7$ dB. For $\alpha = 1$, the

---

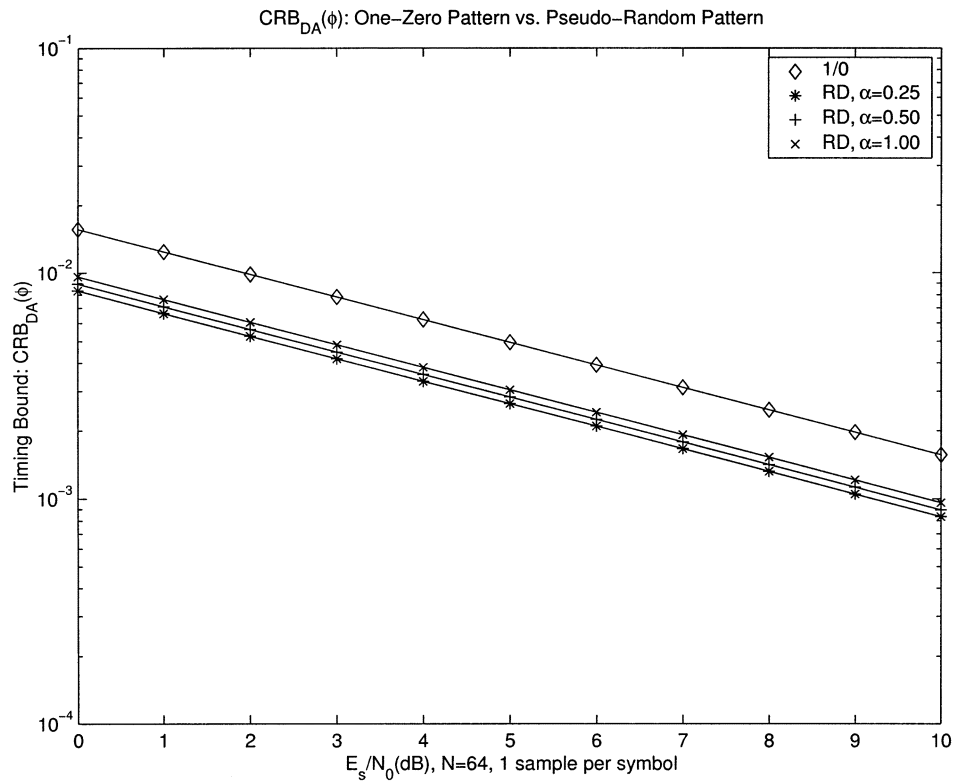[1]This expression was also derived in [6] based on (70).

Fig. 3. Phase bound $\text{CRB}_{\text{DA}}(\phi)$ for alternating one–zero and pseudorandom sequences in the one-sample-per-symbol case, where $N = 64$, $L = 1$, "1/0" denotes the alternating one–zero sequence, and "RD" denotes the pseudorandom sequence.

gain reduces to 2.8 dB. Although the normalized CRB for a finite-length alternating one–zero sequence is slightly different from that of the infinite length, the difference is negligible when N is large enough (e.g., $N = 64$). This shows that practical sequences can offer much better performance than a pseudorandom sequence. Note that an alternating one–zero sequence is not good for frame synchronization. In practice, when an alternating one–zero sequence is employed, other frame synchronization techniques may have to be used, such as embedding frame synchronization data pattern into error-correction-coded bits. In that case, the frame synchronization pattern is designed in such a way that only its successful recovery results in the correct decoding of the error correction code.

Example 4—One Sample Per Symbol: In this example, we compare the one-sample-per-symbol case with the oversampling case for both alternating one–zero and pseudorandom sequences.

In the oversampling case, the $\text{CRB}_{\text{DA}}(\phi)$ is given by (69), which is independent of the training sequence and pulse-shaping function. However, in the undersampling case, it (given by (55) and (60)) depends on both the training sequence and the pulse-shaping function. For a long alternating one–zero sequence, similar to the oversampling case, it is easy to show that the normalized $J_{\phi\phi}$ is equal to

$$\lim_{N\to\infty} \frac{J_{\phi\phi}}{N} = \frac{2E_s}{N_0}\left(\mathcal{R}(\pi/T)^2/T^2 + \mathcal{R}(-\pi/T)^2/T^2\right) = \frac{E_s}{N_0} \tag{78}$$

by using the fact that $\mathcal{R}(\pi/T)/T = 1/2$. For a pseudorandom sequence, we have

$$\frac{J_{\phi\phi}}{N} = \frac{2E_s}{N_0}\left(1 - \frac{\alpha}{4}\right). \tag{79}$$

Therefore, for the carrier phase estimation with one sample per symbol, a pseudorandom sequence outperforms an alternating one–zero sequence of the same length by a factor of

$$2 - \frac{\alpha}{2}.$$

Fig. 3 shows the comparison of the phase bound between a pseudorandom sequence and an alternating one–zero sequence in the one-sample-per-symbol scenario given $N = 64$ and different rolloff factors. For small α, the gain obtained by the pseudorandom sequence is around $10\log 2 = 3$ dB, whereas for $\alpha = 1$, the gain reduces to 2.4 dB.

For a long alternating one–zero sequence similar to the oversampling case, it is easy to show that the normalized $J_{\tau\tau}$ is equal to

$$\lim_{N\to\infty} \frac{J_{\tau\tau}}{N} = \frac{2E_s}{N_0 T^2}\left(\pi^2\mathcal{R}^2(\pi/T) + \pi^2\mathcal{R}^2(-\pi/T)\right)$$

$$= \frac{\pi^2 E_s}{N_0}.$$

This shows that the CRB given one sample per symbol is exactly 3 dB worse than that of the oversampling case for the alternating one–zero sequence. In comparison, for a pseudorandom sequence, $J_{\tau\tau}$ can be obtained by using (71) as in [6] or
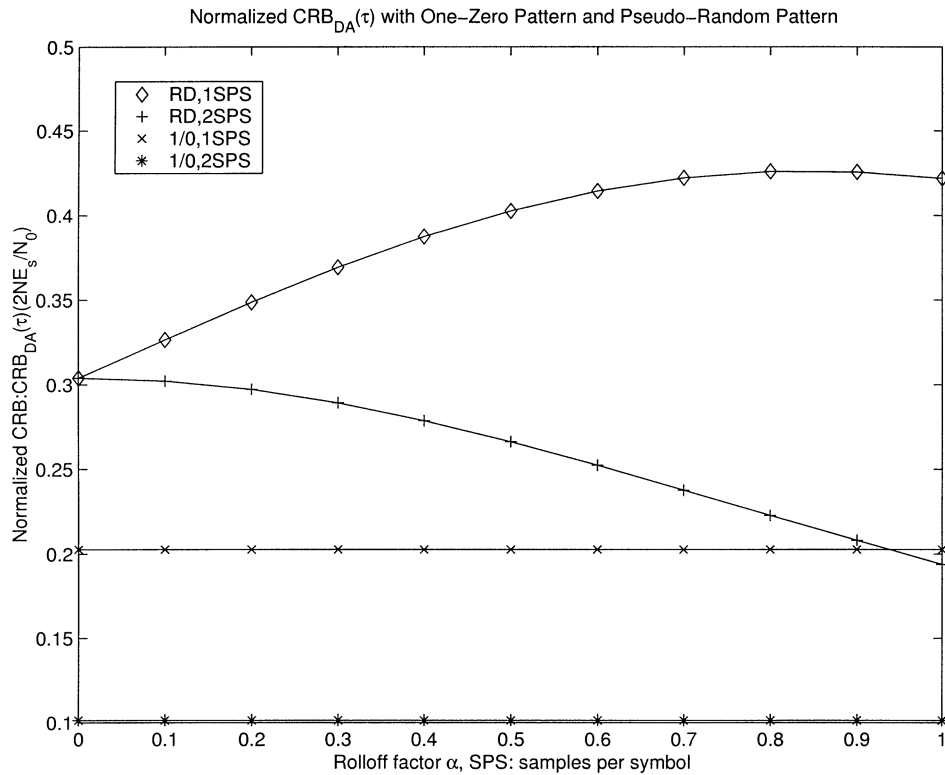
Fig. 4.   Normalized timing bound $\mathrm{CRB}_{\mathrm{DA}}(\tau) \cdot 2NE_s/N_0$ for alternating one–zero and pseudorandom sequences in both the over- and undersampling cases, where "1/0" denotes the alternating one–zero sequence, "RD" denotes the pseudorandom sequence, and "SPS" denotes "samples per symbol."

by computing the second derivative of the convolution $u(t) = r(t) \otimes r(-t)$. We obtain the following expression[2]

$$\frac{J_{\tau\tau}}{N} = -\frac{2E_s}{N_0} \left. \frac{d^2 u(tT)}{dt^2} \right|_{t=0}$$

$$= \frac{2E_s}{N_0} \left[\left(\frac{1}{2} - \frac{\pi^2}{12} \right) \alpha^3 + (\pi^2 - 8)\alpha^2 - \frac{\pi^2}{4} \alpha + \frac{\pi^2}{3} \right]. \tag{80}$$

Fig. 4 plots the normalized timing bound as a function of the rolloff factor α for both alternating one–zero and pseudorandom sequences with different sampling rates.

In summary, an alternating one–zero sequence outperforms a pseudorandom sequence for timing recovery in both the over- and undersampling cases. For phase recovery with a Nyquist pulse, all the sequences perform equally in the oversampling case. However, in the undersampling case, the alternating one–zero sequence is inferior to the pseudorandom sequence. This shows that different design tradeoffs are needed for different systems. The bound presented in this paper allows such tradeoffs to be examined readily.

V. CONCLUSION

In this paper, the closed-form CRB for the DA joint timing and carrier phase offsets estimation was derived given an arbitrary training sequence. This new bound explicitly illus-

trates the dependence of the performance limits on the training sequence. It turns out that such dependence only relates to the aperiodic correlation of the training sequence. The previously known bound for i.i.d. and infinitely long sequence is actually a special case of the bound derived in this paper. We further showed that different training sequences can offer significantly different performance. In particular, there exist sequences providing much better performance than the ideal pseudorandom sequences. By using CW sequences as an example, we have shown that the bound derived here can capture very detailed information such as the ramping up and down of a finite-length sequence. The alternating one–zero sequence is used as an example to show that practical sequences can be designed to achieve significantly better performance than a pseudorandom sequence. In a subsequent publication, the bound is further exploited for sequence design.

Strictly speaking, the CRB is associated with the receiver observation window. This means that for a specific length of the observation window, there is a CRB associated with it even for the same training sequence. It is intuitive and easy to show analytically that the longer the receiver observation window, the lower the CRB. For infinite observation window, the bound reveals the fundamental limits in the estimation based on the information conveyed by the training sequence. This is part of the reason that this paper is titled "Performance Limits ..." rather than the "Cramer–Rao Bound."

---

[2]A different expression was given in [6]. We have verified the current expression by directly performing numerical integral on the frequency-domain formula.

REFERENCES

[1] H. L. Van Trees, *Detection, Estimation and Modulation Theory, Part I.*   New York: Wiley, 1968.

[2] H. Meyr, M. Moeneclaey, and S. Fechtel, *Digital Communication Receivers, Synchronization, Channel Estimation, and Signal Processing*. New York: Wiley, 1998.

[3] M. Moeneclaey, "On the true and the modified Cramer–Rao bounds for the estimation of a scalar parameter in the presence of nuisance parameters," *IEEE Trans. Commun.*, vol. 46, pp. 1536–1544, Nov. 1998.

[4] ——, "A fundamental lower bound on the performance of practical joint carrier and bit synchronizers," *IEEE Trans. Commun.*, vol. COM-32, pp. 1007–1012, Sept. 1984.

[5] ——, "A simple lower bound on the linearized performance of practical symbol synchronizers," *IEEE Trans. Commun.*, vol. COM-31, pp. 1029–1032, Sept. 1983.

[6] C. Georghiades and M. Moeneclaey, "Sequence estimation and synchronization from nonsynchronized samples," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1649–1657, Nov. 1991.

[7] A. N. D'Andrea, U. Mengali, and R. Reggiannini, "The modified Cramer–Rao bound and its application to synchronization parameters," *IEEE Trans. Commun.*, vol. 42, pp. 1391–1399, Feb./Mar./Apr. 1994.

[8] S. White and N. Beaulieu, "On the application of the Cramer–Rao and detection theory bounds to mean square error of symbol timing recovery," *IEEE Trans. Commun.*, vol. 40, pp. 1635–1643, Oct. 1992.

[9] T. Alberty, "Frequency domain interpretation of the Cramer–Rao bound for carrier and clock synchronization," *IEEE Trans. Commun.*, vol. 43, pp. 1185–1191, Feb./Mar./Apr. 1995.

[10] R. M. Gray, "Toeplitz and circulant matrices: A review," Information Systems Lab., Center for Systems Research, Stanford Univ., Stanford, CA, Tech. Rep., 2000.

[11] F.-W. Sun, Y. Jiang, and J. S. Baras, "On the convergence of the inverses of Toeplitz matrices and their applications," *IEEE Trans. Inform. Theory*, vol. 49, pp. 180–190, Jan. 2003.

[12] U. Grenander and G. Szego, *Toeplitz Forms and Their Applications*. Berkeley, CA: Univ. Calif. Press, 1958.

[13] G. Baxter, "An asymptotic results for the finite predictor," *Math. Scand.*, vol. 10, pp. 137–144, 1962.

[14] ——, "A norm inequality for a finite section Wiener–Hopf equation," *Illinois J. Math.*, pp. 97–103, 1962.

[15] H. Kobayashi, "Simultaneous adaptive estimation and decision algorithms for carrier modulated data transmission systems," *IEEE Trans. Commun.*, vol. COM-19, pp. 268–280, June 1971.

[16] T. Kailath, A. Vieira, and M. Morf, "Inverses of Toeplitz operators, innovations, and orthogonal polynominals," *SIAM Rev.*, vol. 20, no. 1, pp. 106–119, Jan. 1978.

[17] T. Kailath, B. Levy, L. Ljung, and M. Morf, "Fast time-invariant implementations of Gaussian signal detectors," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 469–476, July 1978.

[18] J. Pearl, "Basis-restricted transformations and performance measures for spectral representation," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 751–752, Nov. 1971.

[19] ——, "On coding and filtering stationary signals by discrete Fourier transforms," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 229–232, Mar. 1973.

[20] P. J. Sherman, "Circulant approximations of the inverses of Toeplitz matrices and related quantities with applications to stationary random processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1630–1632, Dec. 1985.

[21] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[22] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1994.