Pilot-Assisted Maximum-Likelihood Frequency-Offset Estimation for OFDM Systems

Jiun H. Yu and Yu T. Su, *Member, IEEE*

Abstract—For orthogonal frequency-division multiplexing (OFDM) signals that suffer from frequency-selective fading, we derive the maximum-likelihood (ML) pilot-assisted carrier frequency offset (CFO) estimate and show that most proposals based on repetitive pilot symbols did not use the complete set of sufficient statistics. We convert the problem of obtaining the ML solution from searching exhaustively over the entire uncertainty range to that of solving a spectrum polynomial, thereby greatly reducing the computational load. By properly truncating the polynomial, we obtain a closed-form expression for the corresponding zeros so that the root-searching procedure is greatly simplified. The complexity of locating the desired root is further reduced at almost no expense of performance degradation by an alternate algorithm that uses the fact that the solution is related to the root of a special factor of the polynomial. This alternate method is very attractive for its simplicity and excellent performance that, even at low signal-to-noise ratios (SNRs), is very close to the corresponding Cramér–Rao lower bound. A detailed analysis of the mean-squared error performance is presented and the analysis is validated by simulations.

Index Terms—Frequency estimation, orthogonal frequency-division multiplexing (OFDM).

I. INTRODUCTION

ORTHOGONAL frequency-division multiplexing (OFDM) is an effective antifading modulation scheme for broad-band wireless communications. It has been adopted by several standardization groups for various applications; see [1] and the references therein. A shortcoming of OFDM systems is the sensitivity to the carrier frequency offset (CFO). The presence of a CFO causes reduction of amplitude of the desired subcarrier and induces intercarrier interference (ICI) because the desired subcarrier is no long sampled at the zero-crossings of its adjacent carriers' spectrum. Due to the inherent characteristics of OFDM signals, the tolerable frequency offset range is very limited [1].

J. H. Yu was with the Department of Communication Engineering, National Chiao Tung University, Hsinchu 30056, Taiwan. He is now with the Research and Development Center, Realtek Semiconductor Corporation, Hsinchu 30056, Taiwan (e-mail: jhyu@realtek.com.tw).

Y. T. Su is with the Department of Communication Engineering, National Chiao Tung University, Hsinchu 30056, Taiwan (e-mail: ytsu@mail.nctu.edu.tw).

There have been many CFO estimation schemes for OFDM signals [2]–[15]. These schemes can be conveniently categorized into blind and pilot-assisted schemes. Pilot-assisted schemes use well-designed pilot symbols to estimate CFO and, because these schemes are capable of achieving rapid and reliable frequency synchronization, are often used by packet-oriented systems. Moose [2] proposed a correlation-based technique that uses two consecutive identical pilot symbols to estimate CFO. Although Moose's algorithm is a maximum-likelihood (ML) estimate, its maximum frequency acquisition range is only $\pm1/2$ subcarrier spacing. Following Moose's proposal, subsequent techniques use multiple identical pilot symbols with a smaller symbol period to increase the estimation range of CFO.

Let Δf be the frequency offset and $f_o = \Delta f T = \ell + \Delta f_o$ be the normalized CFO with respect to the subcarrier spacing $1/T$, where T is the OFDM symbol period, $\ell = \lfloor f_o \rfloor$ is the integer part of f_o, while Δf_o is the fractional part. Schmidl and Cox (SC) [5] used two identical half-period symbols to estimate the fractional part of the CFO and a second full-period symbol that has a special correlation relation with the first pilot symbol to estimate ℓ. Lim [11] also proposed a similar method and exploited only two identical half-period symbols to estimate both f_o and ℓ. Morelli and Mengali (MM) [6] increased the acquisition range to $\pm L/2$ subcarrier spacing by dividing a symbol into $L > 2$ repetitive parts, as shown in Fig. 1(a). Their algorithm has been proved to be better than the SC estimate for yielding a smaller minimum mean-square error (MMSE). In addition, their algorithm needs only one symbol period for computing the CFO. The MM algorithm was further improved by the two-stage method of Minn, Tarasak, and Bhargava (MTB) [10]. Song [7] exploited the same pilot symbol structure and suggested a multistage correlation method to acquire CFO, but performance results were the same as those of Schmidl's. A simplified version of Song's estimation method that requires only two correlation steps was proposed by Patel [8].

The performance of these methods, except for the MM and MTB algorithms, depends on the correlation of two half-period identical blocks. MM and MTB used differential phases of the correlations between different pairs of adjacent fractional-period blocks to form an improved CFO estimate.

The blind schemes, on the other hand, exploit the structural and statistical properties of the transmitted OFDM signals such as cyclic prefix [3], virtual subcarrier [4], or constant-modulus [15]. Since no training symbols are required, blind methods are
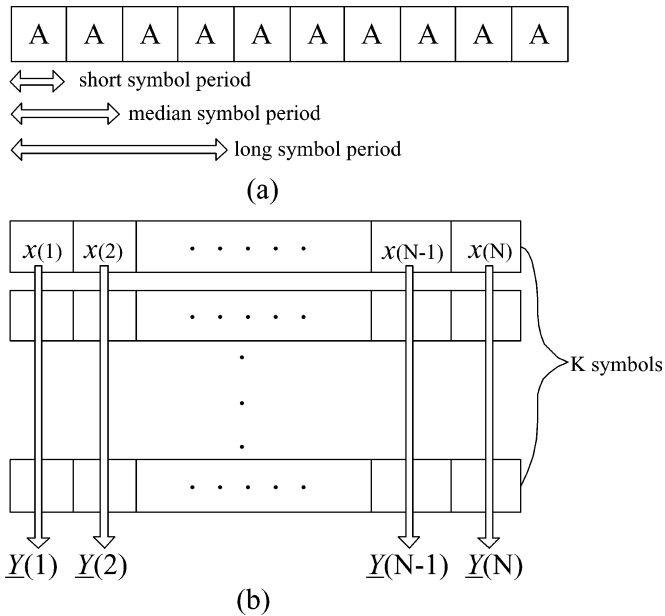
Fig. 1. (a) Short pilot symbol sequence of the IEEE 802.11a. (b) Symbol arrangement and definitions of the proposed ML estimate.

attractive for saving bandwidth and having higher throughput and are more suitable for circuit-switched transmissions.

In this paper, we present ML CFO estimates that use an arbitrary number (K) of identical fractional-period OFDM blocks. An efficient algorithm is provided to solve the associated highly nonlinear ML equation. Instead of searching within the candidate CFO range exhaustively, we only need to solve a polynomial of degree $2(K-1)$. Two methods to further reduce the complexity of extracting the desired root are presented. Numerical results indicate that even at low SNR the performance of the proposed methods still approach the corresponding Cramér–Rao lower bound.

The rest of this paper is organized as follows. Section II describes our signal model and defines related parameters. We derive the corresponding optimal frequency-offset estimate and its simplified version in Section III and Appendix I. A detailed performance analysis is presented in Section IV and Appendix II. We discuss the simulation results in Section V and, finally, in Section VI, we summarize our main results.

II. SIGNAL MODEL AND PARAMETERS

Parallel transmission of a block of $2N_u + 1$ data symbols $\{S_k\}$ drawn from a quadrature-amplitude modulation (QAM) or phase-shift keying (PSK) constellation is efficiently implemented by an M-point inverse discrete Fourier transform (IDFT). The transformed block of M (time-domain) samples

$$p(n) = \frac{1}{\sqrt{M}} \sum_{k=-N_u}^{N_u} S_k e^{j\frac{2\pi kn}{M}}, \quad 0 \le n \le M-1$$

forms a long OFDM symbol, where the equally spaced $2N_u + 1 (\le M)$ data-bearing subcarriers are mutually orthogonal over a symbol interval of $T = M\Delta t$ seconds, where Δt is the IDFT output sample interval. Oftentimes, an OFDM symbol is pre-

ceded by a cyclic prefix longer than the maximum channel delay spread to form an "extended" symbol so that ISI can be eliminated at the receiving end by simply discarding the prefix part. One can also have a short OFDM symbol whose duration is a fraction of T by placing data symbols only at the subcarriers $e^{j2\pi iLt/T}$, i.e., data sequence is transmitted at the frequencies which are multiples of L/T while zeros are inserted at the remaining frequencies [6].

Let $s(m, n)$ be the nth sample of the mth (time-domain) short pilot symbol and assume that the preamble part of a transmitted package consists of $K + 1$ identical short pilot symbols with a total preamble duration of $(K + 1)T/L$ seconds. We thus have the relation $s(m, n) = p(n) = p(n + N)$ for $m = 1, \cdots, L$ and $n = 1, \cdots, N = M/L$. Shown in Fig. 1(a) is the IEEE 802.11a standard that uses $L = 4$, $K + 1 = 10$, and $M = 64$ to form a training sequence of ten identical short symbols.

Consider a frequency-selective channel with a maximum delay spread shorter than a short symbol duration. Assuming that the combined frequency response of the prefilters is flat within the range $|f| \le B + \Delta f_{\max}$, where B is the signal bandwidth and $\Delta f_{\max}$ is the maximum frequency offset, the received baseband waveform is matched-filtered and sampled at a rate of M/T samples/s. After discarding the first symbol, the remaining received K pilot symbols $y(m, n)$ can be represented as

$$y(m,n) = e^{\frac{j2\pi f_o((m-1)N+n)}{M}} x(m,n) + w(m,n) \qquad (1)$$

for $m = 1, \cdots, K$, and $n = 1, \cdots, N = M/L$, where $w(m, n)$ are uncorrelated circularly symmetric Gaussian random variables (rvs) with zero mean and variance $\sigma_w^2 = E\{|w(m,n)|^2\}$. $x(m, n)$ is the channel output corresponding to the transmitted pilot symbol $s(m, n)$. Due to the assumption that the channel delay spread is shorter than the length of one short symbol and the channel impulse response remains the same during the preamble period, the remaining samples $x(m, n)$ are periodic. Note that the above signal model implies that the maximum CFO one can recover is $f_o = L/2$ subcarrier spacings.

Define the two vectors

$$\underline{Y}(n) = [y(1, n), \cdots, y(K, n)]^T \qquad (2)$$

and

$$\underline{A}(f) = \left[1, e^{\frac{j2\pi fN}{M}}, \cdots, e^{\frac{j2\pi f(K-1)N}{M}}\right]^T \qquad (3)$$

where $(\cdot)^T$ denotes the matrix transpose. Then, as shown in Fig. 1(b), we have

$$\underline{Y}(n) = \underline{A}(f_o) e^{\frac{j2\pi f_o n}{M}} x(1, n) + \underline{W}(n)$$
$$\stackrel{\text{def}}{=} \underline{A}(f_o) x(n) + \underline{W}(n), \quad n = 1, \cdots, N \qquad (4)$$

where $x(n) \stackrel{\text{def}}{=} e^{j2\pi f_o n/M} x(1, n)$, and $\underline{W}(n) = [w(1, n), \cdots, w(K, n)]^T$. The received samples can thus be expressed compactly as

$$\mathbf{Y} = \underline{A}(f_o)\underline{X} + \mathbf{W} \qquad (5)$$

where $\mathbf{Y} = [\underline{Y}(1), \cdots, \underline{Y}(N)]$, $\underline{X} = [x(1), \cdots, x(N)]$, and $\mathbf{W} = [\underline{W}(1), \cdots, \underline{W}(N)]$. Hence, given the received sample

vectors $\mathbf{Y}$, we have to estimate f_o through the deterministic vector $\underline{A}(f_o)$. For notational simplicity, we shall drop the argument f in $\underline{A}(f)$ in the subsequent discussion.

The above signal model (5) assumes that perfect symbol timing has been established prior to frequency synchronization. However, it is still valid even if timing error does exist provided that the selected K received pilot symbols are within the range of the preamble and the number of identical short pilot symbols is larger than K (excluding the first discarded received short symbol). Therefore, the beginning position where the K received pilot symbols are selected to form the signal is very flexible. More specifically, even if we do not have the symbol timing, we still can use the above signal model provided that the K selected short symbols are located within the legitimate interval that spans from the start of the second received short symbol to the last sample of the last transmitted short pilot symbol. Thus, the CFO estimators derived from (5) are expected to be insensitive to timing error.

III. ML ESTIMATE OF CFO

Since the noise is temporally white Gaussian, $\underline{Y}(n)$ is a multivariate Gaussian distributed random vector with covariance matrix $\sigma_w^2 I$, where I is the identity matrix. The joint ML estimates of $\underline{A}$ and $\underline{X}$, treating $\underline{X}$ as a deterministic unknown vector, are obtained by minimizing the joint probability density function

$$f(\mathbf{Y}|\underline{A},\underline{X}) = \prod_{n=1}^{N} f\left(\underline{Y}(n)|\underline{A},x(n)\right)$$
$$\propto \exp\left[-\frac{1\sum_{n=1}^{N}\|\underline{Y}(n)-\underline{A}x(n)\|^2}{\sigma_w^2}\right]. \quad (6)$$

The corresponding log-likelihood function, after dropping constant and unrelated terms, is given by

$$\Lambda(\underline{A},x(n)) = \sum_{n=1}^{N}\|\underline{Y}(n)-\underline{A}x(n)\|^2. \quad (7)$$

For a given $\underline{A}$, setting $\nabla_{x(n)}\|\underline{Y}(n)-\underline{A}x(n)\|^2 = 0$, where $\nabla_{x(n)}$ denotes complex gradient operation with respect to $x(n)$, we obtain the conditional ML estimate $\hat{x}(n) = x_{LS}(n) = \underline{A}^+\underline{Y}(n)$, where $\underline{A}^+ = \underline{A}^H/K$ and H denotes the Hermitian operation. Substituting the least-square solution $x_{LS}(n)$ into (7), we obtain

$$\Lambda(\underline{A}) = \sum_{n=1}^{N}\left\|\underline{Y}(n)-\underline{A}\,\underline{A}^+\underline{Y}(n)\right\|^2$$
$$= \sum_{n=1}^{N}\left\|\mathbf{P}_A^\perp\underline{Y}(n)\right\|^2$$
$$= N\mathrm{tr}\left(\mathbf{P}_A^\perp\hat{\mathbf{R}}_{YY}\right) \quad (8)$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix [18], $\hat{\mathbf{R}}_{YY} \overset{\text{def}}{=} (1/N)\sum_{n=1}^{N}\underline{Y}(n)\underline{Y}^H(n)$, and $\mathbf{P}_A^\perp \overset{\text{def}}{=} \mathbf{I} - \underline{A}\,\underline{A}^+$. Note that the (i,j)th entry of the matrix $\hat{\mathbf{R}}_{YY}$, $\hat{R}_{YY}(i,j)$, is the correlation value of ith and jth received symbols, i.e., $\hat{R}_{YY}(i,j) = N^{-1}\sum_{n=1}^{N} y(i,n)y^*(j,n)$. As $\hat{\mathbf{R}}_{YY}$ is the (time-averaged) autocorrelation matrix of the received

sample vectors $\underline{Y}(n)$, it is a Hermitian matrix such that $\hat{R}_{YY}(i,j) = \hat{R}_{YY}^*(j,i)$, where $(*)$ denotes the complex conjugate. The desired CFO estimate is then given by

$$\hat{f} = \arg\left\{\min_{f}\,\mathrm{tr}\left(\mathbf{P}_A^\perp\hat{\mathbf{R}}_{YY}\right)\right\}$$
$$= \arg\left\{\max_{f}\,\mathrm{tr}(\mathbf{P}_A\hat{\mathbf{R}}_{YY})\right\}$$
$$= \arg\left\{\max_{f}\,\underline{A}^H\hat{\mathbf{R}}_{YY}\underline{A}\right\}. \quad (9)$$

It can be proved that the above ML solution is the same as [9, eq. (9)] whose computing load, however, is much heavier. Although (9) gives a compact representation of the ML CFO estimate, it requires an exhaustive search over the entire uncertainty range. The resulting complexity may make its implementation infeasible.

We observe, however, that $\underline{A}$ has a special structure that can be of use to reduce the complexity of searching the desired CFO solution of (9). Invoking an approach similar to that used by the MUSIC algorithm [16], we set $z = e^{j2\pi fN/M}$ and define the parametric vector

$$\underline{A}(z) = [1, z, z^2, \cdots, z^{K-1}]^T \quad (10)$$

so that the log-likelihood $\Lambda = \underline{A}^H\hat{\mathbf{R}}_{YY}\underline{A}$ can be expressed as a polynomial of order $2K-1$ as follows:

$$\Lambda(z) = \underline{A}(z)^H\hat{R}_{YY}\underline{A}(z) = \sum_{n=-(K-1)}^{K-1} s(n)z^n \quad (11)$$

where $s(n) = \sum_{i,j}\hat{R}_{YY}(i,j)$, for $n = j - i$, and $n = -K+1, \cdots, K-1$. To highlight the usefulness of this important observation, we restate it in the form of the following proposition.

Proposition 1: The log-likelihood function for a candidate CFO f is given by

$$\Lambda(f) = \sum_{n=-(K-1)}^{K-1} s(n)e^{\frac{j2\pi nfN}{M}}. \quad (12)$$

Some remarks about this proposition are in order.

Remarks:

R1. $s(n)$ is the summation of diagonal entries of $\hat{\mathbf{R}}_{YY}$ and is also equivalent to the aperiodic autocorrelation value of the waveform $\{y(i,j), 1 \le i \le K, 1 \le j \le N\}$ at time difference $nN(T/M)$ seconds, i.e., $s(n) = \sum_{m=1}^{K-n}\sum_{j=1}^{N} y^*(m+n,j)y(m,j)$.

R2.   It can be shown that, in the absence of noise

$$s(n) = \hat{\sigma}_x^2\,(K-|n|)\,e^{-jn\theta_o} \overset{\text{def}}{=} \Delta(n) \quad (13)$$

where $\theta_o = 2\pi f_o N/M$, and $\hat{\sigma}_x^2 = (1/N)\sum_{n=1}^{N}|x(n)|^2$. When noise is present, the mean value of $s(n)$ is the same as its noiseless value except for $n = 0$; more specifically, $E[s(n)] = \Delta(n) + K\sigma_w^2\delta(n)$, where $\delta(\cdot)$ is the Kronecker delta function. Evaluating (11) at the unit circle $z = e^{j2\pi fN/M} \overset{\text{def}}{=} e^{j\theta}$, we obtain the discrete-time Fourier transform of the sequence $s(n)$,

which has an envelope similar to $\mathrm{sinc}^2(f)$ whose maximum value is at the correct "modified" frequency $f'_o = f_o(N/M)$.

R3. Due to the Hermitian nature of $\hat{\mathbf{R}}_{YY}$, $s(n)$ is a conjugate symmetric sequence of length $2K - 1$. The symmetric property of $s(n)$ guarantees that its Fourier transform $\Lambda(e^{j2\pi f})$ is real and nonnegative. This also follows from the semi-positive definiteness of the quadratic form $\underline{A}^H \hat{\mathbf{R}}_{YY} \underline{A}$. Because $s(n)$ and the log-likelihood function constitute a Fourier transform pair, we will henceforth refer to $\Lambda(e^{j2\pi f})$ as the log-likelihood spectrum or spectrum, for short, and the polynomial defined by (11) the spectrum polynomial.

R4. $\{s(n), n = 0, 1, \cdots, K - 1\} = \mathcal{C}$ constitutes a set of sufficient statistic for estimating f_o. Almost all previous correlation-based algorithms use only a subset of $\mathcal{C}$, e.g., For $K = 8$, the Moose's algorithm [2] uses $s(4)$, the GM algorithm [7] uses $s(1)$. The MM and MTB algorithms use only the phases of elements of $\mathcal{C}$. Furthermore, the former achieves its best performance when it uses only half of the phases [10]. It is expected that an algorithm that uses the sufficient statistic would outperform those that use only a part of the sufficient statistic.

R5. Computing the desired CFO estimate through (11) is equivalent to searching for the peak of the candidate spectrum $\Lambda(e^{j2\pi f})$. Hence, the spectrum can be computed using a discrete Fourier transform (DFT), but the resolution of the CFO estimate $\hat{f}_o$ depends on the size of the DFT. Padding more zeros in the sequence $s(n)$ results in higher resolution at the expense of inducing higher computation complexity.

As the spectrum is a real smooth function of θ, taking a derivative of $\Lambda(e^{j\theta})$ with respect to θ and setting $\partial\Lambda(e^{j\theta})/\partial\theta \overset{\text{def}}{=} \dot{\Lambda}(\theta) = 0$, we obtain

$$F(z) = F^*(z) \tag{14}$$

where $F(z) = \sum_{n=1}^{K-1} n s(n) z^n$ is a polynomial of order $K-1$. As mentioned before, in a noiseless environment, $\Lambda(e^{j2\pi f})$, the Fourier transform of $\{s(n)\}$, is a scaled version of the function $\mathrm{sinc}^2(f)$, and all roots of $\dot{\Lambda}(z) \overset{\text{def}}{=} \dot{\Lambda}(\theta)|_{e^{j\theta}=z} = j(F(z) - F^*(z))$ are on the unit circle. The presence of noise and multipath will modify the Fourier transform and move some roots away from the unit circle so that the solutions of $\dot{\Lambda}(\theta) = 0$ become a proper subset of those of $\dot{\Lambda}(z) = 0$. In that case, we have $\dot{\Lambda}(z) = cH_1(z)H_2(z)H_2(1/z^*)$, where c is a complex constant, $H_1(z) = \prod_k(z - \gamma_k)$, $|\gamma_k| = 1$, and $H_2(z) = \prod_i(z - \alpha_i)$, $|\alpha_i| < 1$. Although $\dot{\Lambda}(z)$ has a fixed number of roots, the distribution of these roots among $H_i(z)$ or, equivalently, the degrees of $H_1(z)$ and $H_2(z)$ depend upon SNR for the existence of nonunit-amplitude roots due to the merging of neighboring sidelobes of $\Lambda(e^{j\theta})$, which, in turn, results from large noise perturbation.

We can either restrict our search to those roots that are on the unit circle or normalize those nonunit-amplitude roots. Several reasons convince us that both approaches will most likely give the same estimate. First, the desired root is associated with the
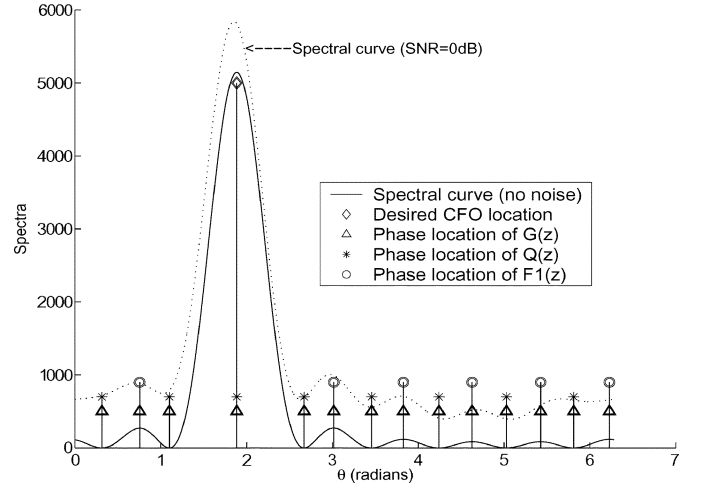

Fig. 2. Normalized log-likelihood spectrum and the associated root distribution where the spectrum is normalized by the noiseless mainlobe peak value; CFO $= 1.2$ subcarrier spacings.

peak of $\Lambda(e^{j\theta})$ and the mainlobe-peak-to-sidelobe-peak ratio is greater than 25 dB. As the relative height difference between any two neighboring sidelobe peaks is far smaller than that between the mainlobe and its two neighboring sidelobes, it is much more likely for the merging of neighboring sidelobes than that of the mainlobe and one of its neighboring sidelobes and, even if the latter merge occurs, the associated peak will most probably be the peak of the spectrum. In other words, the desired root is likely to stay at the unit circle with a very high probability. Second, our simulation has shown that, even in the presence of strong noise and severe multipath, the resulting $\Lambda(e^{j2\pi f})$ still bears a close resemblance to a scaled version of $\mathrm{sinc}^2(f)$ and the roots on the unit circle are within small neighborhoods of their noiseless locations; see Fig. 2. Finally, $\Lambda(e^{j2\pi f})$ is a linear transform of $\{s(n)\}$ and, as (B8) indicates (see Appendix II), $s(n)$ can be decomposed into two deterministic terms and two zero-mean complex perturbation terms that are uncorrelated with the deterministic part.

For simplicity, we shall use the first approach, i.e., the desired estimate is to be obtained by

$$\hat{f}_o = \frac{M}{j2\pi N} \ln \hat{z} \tag{15}$$

where

$$\hat{z} = \arg\left\{ \max_{z_i \in \Omega} \Lambda(z) \right\}, \quad \Omega = \left\{ z | \dot{\Lambda}(z) = 0, |z| = 1 \right\}. \tag{16}$$

Note that we have converted the exhaustive search problem of (9) to a root-finding problem, reducing the candidate solution number from infinity to at most $2(K - 1)$.

A. A ML CFO Estimation Algorithm

We summarize the procedure leading to (16) as follows.

1) Collect K received symbols and construct the sample correlation matrix $\hat{\mathbf{R}}_{YY}$.
2) Calculate the coefficients of $F(z)$ based on $\hat{\mathbf{R}}_{YY}$.
3) Find the nonzero unit-magnitude roots of (14).
4) Obtain the CFO estimate from (15) and (16).

We will refer to the above procedure as Algorithm A. It can be shown that, when $K = 2$, the resulting estimate is equivalent to the Moose estimate [2].

The above algorithm needs to locate all of the roots of $\dot{\Lambda}(z)$. We can reduce the order of $\dot{\Lambda}(z)$ by truncating the length of the sequence $\{s(n)\}$, i.e., using a lesser number of autocorrelation values. When we use the truncated version, $\Lambda(z) = \sum_{n=-2}^{2} s(n)z^n$, to carry out Algorithm A, the resulting algorithm is referred to as Algorithm A'. The complexity of Algorithm A' is much less than that of Algorithm A since the associated $\dot{\Lambda}(z)$ has closed-form solutions. We do expect some performance degradation as less autocorrelation values are used. In the following subsection, we present a method to further reduce the complexity of solving (14) in step 3) with little or no performance degradation.

B. A Simplified CFO Estimate

We notice that the solutions of (14) are the nonzero roots of the polynomial

$$G(z) = z^{K-1}\left(F(z) - F^*(z)\right). \tag{17}$$

On the other hand, (14) implies that the roots of $\dot{\Lambda}(z)$ satisfy the equation $\text{Im}\{F(z)\} = 0$, where $\text{Im}\{F(z)\}$ is the imaginary part of $F(z)$. This observation indicates that the nonzero roots of $F_1(z) = z^{-1}F(z) = 0$ (the root $z = 0$ of $F(z)$ is undesired) are a subset of the roots of $\text{Im}\{F(z)\} = 0$. When $F(z)$ is an arbitrary polynomial, its roots are not necessarily a subset of those of the corresponding $G(z)$ defined by (17). However, in our case, as shown in Appendix I, both $F(z)$ and $F_1(z)$ do divide $G(z)$ when there is no noise in the received signal vector $\underline{Y}(n)$ or the ensemble average $E\{\hat{\mathbf{R}}_{YY}\} = \mathbf{R}_{YY}$ is used to construct $F(z)$ and $G(z)$, i.e., the roots of $F_1(z)$ are indeed a subset of those of $G(z)$. Moreover, we can show (see Appendix I) that the following proposition holds.

Proposition 2: In the absence of noise, the polynomial defined by (17), $G(z)$, can be decomposed into

$$G(z) = F_1(z)Q(z) \tag{18}$$

where the desired CFO estimate is one of the roots of $Q(z)$ defined by

$$Q(z) = z^K - \frac{s(-1)}{(K-1)s(K-1)} = z^K - d \tag{19}$$

where $d = s(-1)/[(K-1)s(K-1)]$.

When noise is present, the above equality becomes an approximation only. Nevertheless, the desired CFO estimate can still be derived immediately from taking the Kth root of d. Fig. 2 highlights the locations of the normalized roots of $G(e^{j\theta})$, $F_1(e^{j\theta})$ and $Q(e^{j\theta})$ for $K = 8$, i.e., various local extremes of $\Lambda(z)$. The global maximum that colocated with a root of $Q(z)$ corresponds to the desired CFO estimate while the remaining roots of $Q(z)$ locate at a local minimum (null) of the spectrum. On the other hand, the roots of $F_1(z)$ are at the local sidelobe peaks of the spectrum. It is clear that the union of the roots of $Q(z)$ and $F_1(z)$ is the set of the roots of $G(z)$. Hence, the complexity of extracting the roots is significantly reduced, for we only have to solve the equation $Q(z) = 0$, which happens to

have a closed-form expression for its roots. We will show later via simulations and analysis (see Appendix II) that (18) is a valid approximation that incurs only negligible performance loss even when the system is operating at an SNR as low as 0 dB. The above discussion suggests the following simplified CFO estimate algorithm.

1) Follow 1) of Algorithm A.
2) Compute the coefficients d based on two correlation values $s(-1)$ and $s(K-1)$.
3) Solve $Q(z) = 0$ for the K unit-magnitude roots of d, $\{z_i\}$.
4) Find the estimate from (15) and (16).

The above algorithm will be referred to as Algorithm B. Note that the definition of $s(n)$ [see (13)] indicates that $d = e^{jK\theta_o}$ and suggests that steps 3) and 4) of Algorithm B can be replaced by using the estimate $(M/2\pi NK)\arg\{d\}$, where $\arg\{z\}$ is the principal value of the argument of z. The disadvantage of this estimate is that it cannot be applied to the situation when the CFO is such that $K\theta_o > 2\pi$. On the other hand, (13) also implies that the CFO can be estimated by $\tilde{f}_o(\ell) = (M/2\pi N\ell)\arg\{s(\ell)\}$, $\ell \in [-(K-1), (K-1)]$. Amongst these candidate estimates, $\hat{f}_o(\pm 1)$ is the only one that does not violate the no-phase-ambiguity requirement and, in fact, is the same as the generalized Moose (GM) algorithm that uses averaged autocorrelation values [7].

IV. PERFORMANCE ANALYSIS

Note that the CFO estimate derived from Algorithm A is identical to (9). This estimate is unbiased, for as $N \to \infty$, $\hat{\mathbf{R}}_{YY} \to \mathbf{R}_{YY} = \sigma_x^2 \underline{A}\,\underline{A}^H + \sigma_w^2\mathbf{I}$, with $\sigma_x^2 = E\{|x(n)|^2\} = \sigma_s^2/M \sum_i E\{|H_{iL}|^2\}$, $\sigma_s^2 = E\{|S_k|^2\}$, and H_{iL} is the channel response at the subcarrier $e^{j2\pi iLt/T}$, hence $E[\hat{f}_o] = f_o$. In Appendix II, we show that the variance of the ML estimate $\hat{f}_o$ is given by

$$\text{var}(\hat{f}_o) = \frac{3L^3\left[5K\gamma_b^{-2} + (6K^2 - 4)\gamma_b^{-1}\right]}{10\pi^2 MK^3(K^2 - 1)} \tag{20}$$

where $\gamma_b \overset{\text{def}}{=} \sigma_x^2/\sigma_w^2 \overset{\text{def}}{=} \overline{\text{SNR}}$. Following the analysis that leads to [16, eq. (4.1)] and upon substituting the parameter, $\theta = (2\pi Nf/M) = (2\pi f/L)$, we obtain the corresponding Cramer–Rao bound (CRB)

$$\text{CRB}(\hat{f}_o) = \frac{3L^3\gamma_b^{-1}}{2\pi^2 MK(K^2 - 1)}. \tag{21}$$

Equation (20) indicates that $\text{var}(\hat{f}_o)$ is a decreasing function of K. Therefore, replacing one long pilot symbol with several identical short symbols and using the proposed ML method $(K > 2)$ would yield a performance superior to that resulting from using the correlation-based method that correlates two identical half-period symbols $(K = 2$, i.e., Moose algorithm), even though both use the same number of data samples.

For Algorithm B, the associated variance can be approximated by (22), shown at the bottom of the next page (see Appendix II). Equation (22) reveals that, even at low SNRs, Algorithm B still gives a satisfactory performance. For Algorithm A', we expect its performance to be between those of

Algorithms A and B, since the former uses all of the available autocorrelation values while one of the two autocorrelation values the latter uses is $s(K-1)$, which involves a smaller number of time-correlation samples than $s(2)$ and is, therefore, less reliable. The simulation results shown in Section V do confirm this conjecture.

It is also clear that, as long as the true CFO does not incur phase ambiguity (i.e., $|\theta_o| < \pi$), then the Kth roots of $Q(z)$ are given by $z_i = e^{j(\arg\{d\}/K + 2\pi i/K)}$, $i = 0, \cdots, K-1$. Although in a noiseless environment $\arg\{d\}/K = \arg\{s(1)\} = \theta_o$ (assuming, without loss of generality, the desired root is $\hat{z} = \arg\{d\}/K$), the variance analysis presented in Appendix II shows that $\arg\{d\}/K$ yields a smaller variance. As a result, Algorithm B yields a performance superior to that of the generalized Moose (GM) estimate obtained by averaging over two consecutive overlapped median symbols [see Fig. 1(a)] $K-1$ times. Later simulation results indeed reveal the superiority of Algorithm B. Furthermore, it also suggests that step 4) can be replaced by picking up the Kth root whose principal argument is closest to $\arg\{s(1)\}$.

V. SIMULATION RESULTS AND DISCUSSION

Numerical examples are provided in this section to examine the behavior of the proposed CFO estimation technique. As shown in Fig. 1(a), eight short training symbols which are the same as those used in the IEEE 802.11a preamble are used in our simulation. Results reported in Figs. 3–6 assume a static frequency-selective fading channel with ten paths whose complex amplitudes are independent identically distributed (i.i.d.) complex Gaussian random variables. CFO is normalized by subcarrier spacing and the mean values and mean-squared errors (MSE) of various estimates are computed by 10^4 independent trials.

For comparison purposes, the corresponding CRBs and the behaviors of four other estimates are provided as well. These estimates are the MTB estimate (which outperforms the MM estimate [10]), the Moose estimate [2], the GM estimate, and the Patel–Song [7], [8] (PS) estimate. The last estimate uses the GM estimate as the initial estimate and selects the final estimate from the family of candidate estimates $\{\hat{f}_{\text{Moose}} \pm n, n \in S\}$, where $\hat{f}_{\text{Moose}}$ is the Moose estimate and S is the set of integer multiples of twice the acquisition range of $\hat{f}_{\text{Moose}}$ that is the closest to the initial estimate. The GM estimate is a special case of our proposals, corresponding to one that uses the truncated version of the spectrum polynomial, $\Lambda(z) = \sum_{n=-1}^{1} s(n)z^n$. The GM estimate has a frequency acquisition range larger than that of the Moose estimate but renders a less accurate estimation when CFO is within the latter's acquisition range. The PS estimate is designed to retain the advantages of both algorithms,
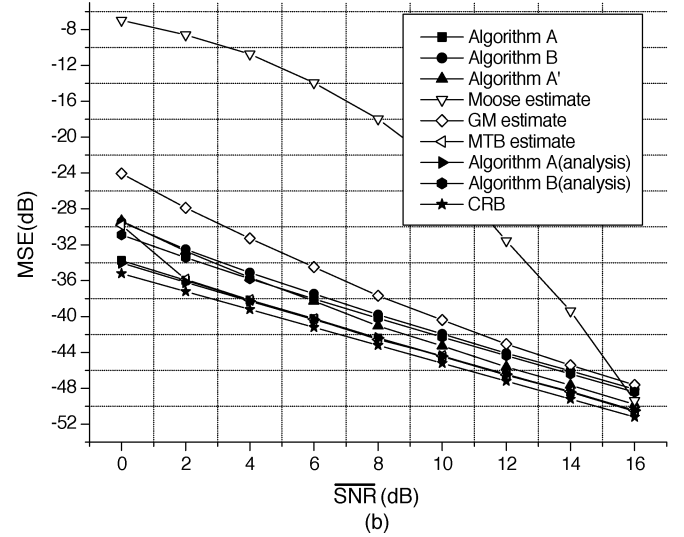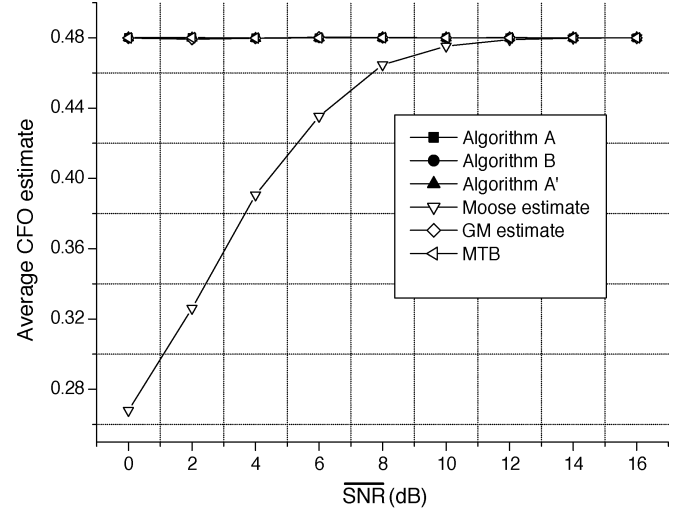

Fig. 3. (a) Averaged estimation values for various CFO estimates. (b) MSE performance of CFO estimates; $K = 8$, and true CFO $= 0.48$ subcarrier spacings.

i.e., having an acquisition range same as that of the GM estimate while achieving the performance of the Moose estimate. The coefficients used by the MTB estimate to optimally combine the differential phases of the elements in the set $\mathcal{C}$ depend on the operating SNR. In our simulation, a design $\gamma_b = 20$ dB is assumed. The MM estimate is not compared for it best performance is achieved when only $L/2$ correlations are used [10]. For the purpose of fair comparison, all algorithms whose performance is shown in the same figure use the same number of samples.

Fig. 3(a) and (b) depicts the mean and MSE of various CFO estimates as a function of SNR. Since the maximum CFO that can be corrected by the Moose estimate is $\pm 1/(2T)$, the

$$\text{var}(\hat{f}_o) \cong \frac{2ML^3(K-1)(K+2)\gamma_b^{-1} + \left[ML^3K(K-1) + L^4(4K+20) + 8L^5M^{-1}\right]\gamma_b^{-2}}{8\pi^2M^2K^2(K-1)^2}$$

$$+ \frac{L^4\left[4(K-1)\gamma_b^{-3} + (K-1)\gamma_b^{-4}\right]}{8\pi^2M^2K^2(K-1)^2} \quad (22)$$
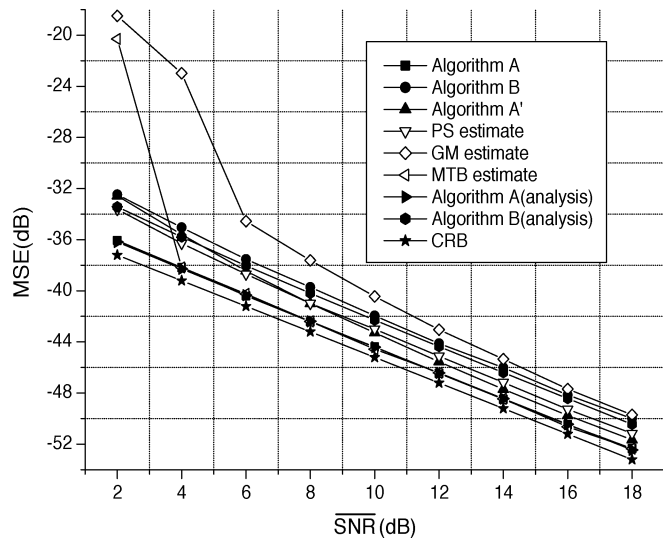
Fig. 4. MSE performance of CFO estimates; $K = 8$, and true CFO $= 1.8$ subcarrier spacings.
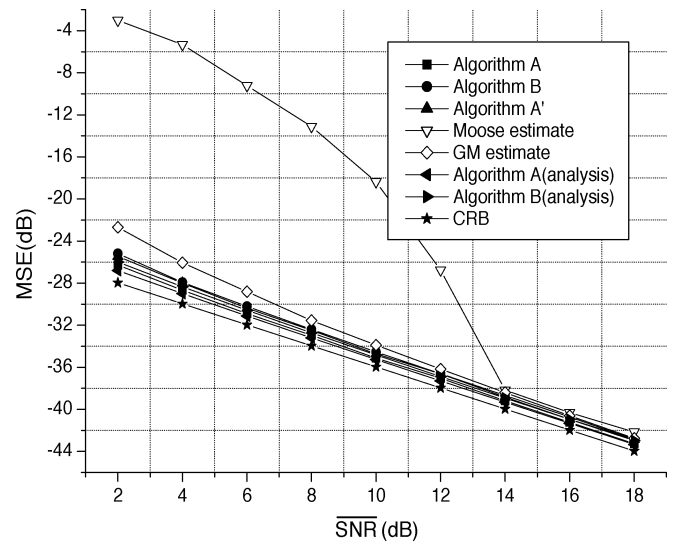

Fig. 5. MSE performance of CFO estimates; $K = 4$, and true CFO $= 0.94$ subcarrier spacings.

corresponding performance degrades rapidly at low SNR when the true CFO approaches the maximum correctable value. For Algorithms A, A', and B and the MTB and GM estimates, however, the maximum correctable CFO is $\pm 2/T$, hence they give a much better MSE performance with Algorithm A yielding a performance almost the same as that of the CRB. Both Algorithms A and A' (corresponding to $K = 8$) outperform the Moose (corresponding to $K = 2$) and GM estimates. For 802.11a systems, using eight short training symbols not only allows a larger CFO offset range but also yields better performance. At CFO $= 0.48/T$, as shown in Fig. 3, the Moose estimate is worse than the GM estimate at low SNRs. Algorithm B is simpler but its performance is a little worse than those of Algorithms A and A', though it is still better than the GM estimate at low-to-medium SNRs. The performance of the MTB estimate is almost the same as that of Algorithm A except at low SNR.

Fig. 4 plots the MSE performance for CFO $= 1.8/T$. Algorithms A and A' and the MTB estimate still maintain superior performance in this case. For the MTB estimate, however, the effects of SNR mismatch and initial CFO estimation error become apparent at low SNRs. Fig. 5 plots the MSE performance when CFO $= 0.94/T$ using four short symbols. The Moose estimate uses two identical symbols: each lasts for $K = 2$ short symbol durations so that the maximum offset range it can correct becomes $\pm 1/T$. Fig. 6 compares the MSE performance of our algorithms and the GM and PS estimates when CFO $= 1.6/T$. Figs. 5 and 6 clearly indicate that the PS (or Moose) estimate is not necessarily better than the GM estimate but that the proposed methods still provide the best performance although the improvement is less impressive. This confirms our analysis in Section IV where it is shown that the performance of Algorithm A improves as K increases. Finally, we examine the performance in a time-varying channel composed of ten uncorrelated fading paths generated by the modified Jakes' model [17]. The maximum path delay is equal to ten data sample intervals, assuming a sampling rate of 20 MHz. The performance of our CFO estimates is independent of the path numbers so long as
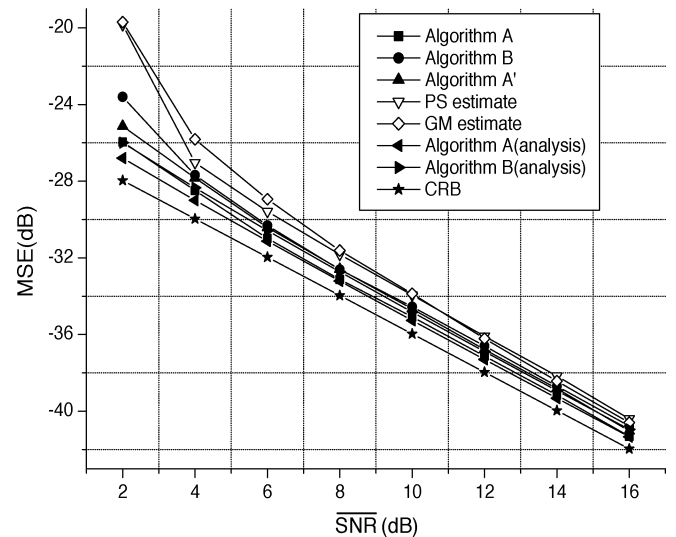

Fig. 6. MSE performance of CFO estimates; $K = 4$, and true CFO $= 1.6$ subcarrier spacings.

the channel's maximum delay spread is shorter than the length of the cyclic prefix. Assuming a mobile speed of 100 km/h, corresponding to a Doppler frequency of approximately 463 Hz when the carrier frequency is 5 GHz, we plot the corresponding MSE performance in Figs. 7 and 8. As our derivations assume a quasi-static channel that remain unchanged during the preamble period, the estimation performance is degraded due to the fact that the received signal model (4) is no longer valid. In summary, Algorithms A and A' and the MTB estimate render the best performance, followed by Algorithm B, and then the other correlation-based algorithms. When K is small, Algorithms A, A', and B yield almost the same MSE performance. The proposed methods can be used when an arbitrary number (K) of identical pilot symbols are available.

As for the computational complexity issue, we first noticed that the basic requirement is the computing of $s(n)$. Different methods call for the computation of different subsets of the sufficient statistic $\mathcal{C}$. The detailed computational complexity
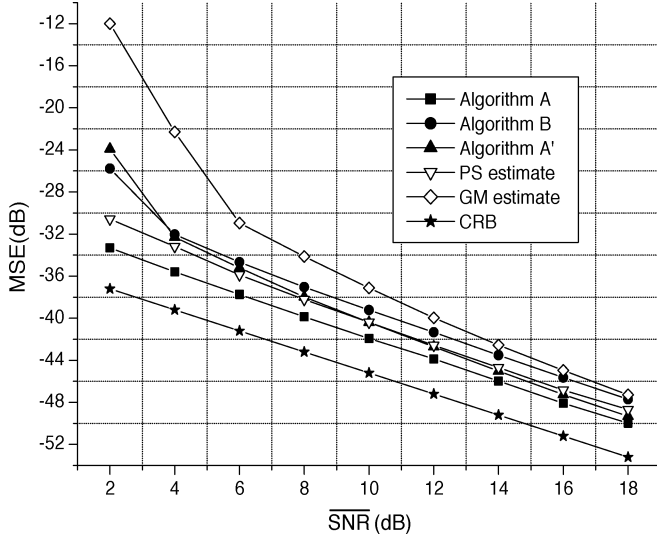
Fig. 7. MSE performance of CFO estimates in a time-varying frequency-selective Rayleigh fading channel; $K = 8$, and true CFO $= 1.8$ subcarrier spacings.
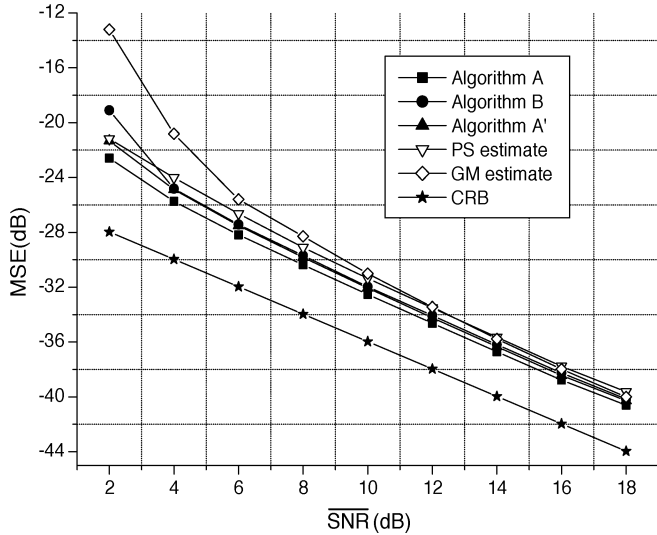

Fig. 8. MSE performance of CFO estimates in a time-varying frequency-selective Rayleigh fading channel; $K = 4$, and true CFO $= 1.6$ subcarrier spacings.

and storage requirement for the CFO estimation algorithms are listed in Table I, assuming that there are K short pilot symbols available with each symbol of N-sample duration. Clearly, Algorithm A and the MTB estimate require the highest complexity while Algorithm B is the simplest among the three proposed algorithms. The MTB algorithm needs to perform down-converting (multiplying exponentials) that require table loop-ups. The storage requirement for building the table is not included in Table I. The computing effort for finding the roots is also not included. Algorithm B needs one additional real division only. Algorithm A' has to solve a polynomial of order four that renders a closed-form expression while Algorithm A has to solve a polynomial of order $2(K-1)$, and the associated complexity is relatively high. A method to significantly reduce this complexity can be found if one uses the phase of $s(1)$ as

the initial phase value for searching for the roots. An elementary root-finding algorithm like Newton's method can then be applied to find the first unit-magnitude root of $G(z)$ in just a few iterations, as the desired root has a phase very close to that of $s(1)$.

VI. CONCLUSION

The optimal ML CFO estimate that uses several identical pilot symbols has been derived. By transforming the log-likelihood function into a spectrum polynomial, we reduce the ML estimate's complexity from that of an exhaustive search over a continuum to that of solving a polynomial. Besides the ML estimate (Algorithm A), we propose two simplified versions (Algorithms A' and B) and show that some of the previous correlation-based algorithms are special cases of our proposals. The MSE performance of the proposed algorithms are analyzed in detail, and our analysis does match the simulation results. Both the analysis and simulations indicate that the performance of Algorithm A is very close to the corresponding CRB. Algorithms A' and B require much less computational load but they do not incur a noticeable loss of performance. Numerical results also demonstrate that, in a time-varying Rayleigh fading channel, the performance of the proposed methods suffers from minor performance degradation only.

APPENDIX I
DERIVATION OF (18) AND THE ROOT DISTRIBUTION OF $G(z)$

Dividing $G(z)$ by $F_1(z)$, we obtain the quotient polynomial $Q(z)$ defined by (19) and the remainder polynomial

$$
\begin{aligned}
q(z) &= [(K-2)s(K-2)d - 2s(-2)]z^{(K-3)} \\
&\quad + [(K-3)s(K-3)d - 3s(-3)]z^{(K-4)} + \cdots \\
&\quad + [(K-(K-2))s(K-(K-2))d \\
&\qquad - (K-2)s(-(K-2))]z^{(K-(K-1))} \\
&\quad + [(K-(K-1))s(K-(K-1))d \\
&\qquad - (K-1)s(-(K-1))]z^{(K-K)} \\
&\overset{\text{def}}{=} \sum_{n=1}^{K-2} q_n z^{n-1}
\end{aligned}
\tag{A1}
$$

where $d = s(-1)/[(K-1)s(K-1)] = e^{jK\theta_o}$, $q_n = ns(n)d - (K-n)s(-(K-n))$. In the absence of noise, we have $\hat{\mathbf{R}}_{YY} = \hat{\sigma}_x^2 \underline{A}(f_o)[\underline{A}(f_o)]^H$ with $\underline{A}(f_o) = [1, e^{j2\pi f_o N/M}, \cdots, e^{j2\pi f_o(K-1)N/M}]^T$. Invoking the definition $\theta_o = 2\pi f_o N/M$ and substituting the resulting alternative expressions $d = e^{jK\theta_o}$, $\hat{\mathbf{R}}_{YY} = \hat{\sigma}_x^2 \underline{A}(\theta_o)\underline{A}(\theta_o)^H$, and $s(n) = \hat{\sigma}_x^2(K-|n|)e^{-jn\theta_o}$, for $n = -K+1, \cdots, K-1$, we find that $q_n = 0$ for $n = 1, \cdots, K-2$. Therefore, $q(z) = 0$ and the polynomial $F_1(z)$ are indeed a factor of $G(z)$. Similarly, it is easy to see that, if the ensemble average $E\{\hat{\mathbf{R}}_{YY}\} = \mathbf{R}_{YY}$ is used to construct $F_1(z)$ and $G(z)$, we still have the factorization (18).

Obviously, the roots of the quotient polynomial $Q(z) = z^K - e^{jK\theta_o}$ are the set $\{z_i = e^{j(\theta_o + 2\pi i/K)}, i = 0, \cdots, K-1\}$, which has the desired root $e^{j\theta_o}$ as one of its members. Next we prove that the other roots of $Q(z)$ are located at the sidelobes

TABLE I
COMPLEXITY COMPARISON OF VARIOUS CFO ESTIMATION ALGORITHMS

Algorithm	Correlation values required	No. of real multiplications	No. of real additions	Memory required	Phase comp. $\arg\{\cdot\}$
A	$s(1) - s(K-1)$	$2K(K-1)N$	$(2NK-2)(K-1)$	$(K-1)N$	1
A'	$s(1), s(2)$	$4(2K-3)N$	$(8K-12)N-4$	$2N$	1
B	$s(1), s(K-1)$	$4KN$	$4(NK-1)$	$2N$	2
Moose	$s(K/2)$	$2KN$	$2(NK-1)$	$KN/2$	1
GM	$s(1)$	$4(K-1)N$	$4N(K-1)-2$	N	1
PS	$s(1), s(K/2)$	$(6K-4)N$	$4N(3/2K-1)-4$	$KN/2$	2
MTB	$s(1) - s(K-1)$	$2K(K+1)N$	$(2NK-2)K+2$	KN	$K-1$

of the log-likelihood spectrum $\Lambda(e^{j2\pi f})$. Recall that $\Lambda(z) = \sum_{n=-(K-1)}^{K-1} s(n) z^n$. Assuming K is even, we obtain

$$\Lambda(z) = \sum_{n=1}^{K-1} \left(s(n) z^n + s(-n) z^{-n} \right) + K\hat{\sigma}_x^2$$
$$= \sum_{n=1}^{K-1} \left\{ 2\hat{\sigma}_x^2 (K - |n|) \cos\left(n(\theta - \theta_o) \right) \right\} + K\hat{\sigma}_x^2. \quad \text{(A2)}$$

It is obvious that $\Lambda(z) = 0$ if $K(\theta - \theta_o)/2 = i\pi$, for $i = 1, \cdots, K - 1$. Hence, $\Lambda(z) = 0$ when $z_i = e^{j(\theta_o + 2\pi i/K)}$, for $i = 1, \cdots, K - 1$, which are the remaining roots of $Q(z)$. The case K is odd can be similarly proved. Note that the roots of $G(z)$ are the extreme values of the spectrum of $\Lambda(z)$. Since the roots of $Q(z)$ are located at either the global maximum or the local minima of $\Lambda(z)$, the roots of $F_1(z)$ should sit at the local maxima, i.e., sidelobe peaks of the spectrum, as shown in Fig. 2.

The above discussion assumes a noiseless environment. When noise is present, we can easily show that the factorization of $G(z)$ still holds in the mean sense. Furthermore, the analysis of $s(n)$ presented in Appendix II convinces us that factorization (18) will remain valid with a probability close to one unless SNR is very small (say, <0 dB).

APPENDIX II
MSE PERFORMANCE ANALYSIS

A. Decomposition of the Correlation Coefficients $s(n)$

The performance of the proposed algorithms depends on the behavior of $\Lambda(z)$ whose coefficients $s(n)$, as shown in R2 of the main text, are functions of the autocorrelation matrix $\hat{\mathbf{R}}_{YY} = (1/N)\sum_{n=1}^{N} \underline{Y}(n)\, \underline{Y}^H(n)$. We thus begin our analysis by examining

$$\hat{\mathbf{R}}_{YY} = \hat{\sigma}_x^2 \underline{A}\, \underline{A}^H + \frac{1}{N}\sum_{n=1}^{N} \underline{W}(n)\underline{W}^H(n)$$
$$+ \frac{1}{N}\sum_{n=1}^{N} \left[\underline{A}x(n)\underline{W}^H(n) + \underline{W}(n)\underline{A}^H x^*(n) \right] \quad \text{(B1)}$$

where $\hat{\sigma}_x^2 \overset{\text{def}}{=} (1/N)\sum_{n=1}^{N}|x(n)|^2$. Let $\tilde{\mathbf{R}}_{YY}$ be the discrepancy matrix between $\hat{\mathbf{R}}_{YY}$ and its ensemble average, $\mathbf{R}_{YY} = E\{\hat{\mathbf{R}}_{YY}\}$, i.e., $\tilde{\mathbf{R}}_{YY} = \hat{\mathbf{R}}_{YY} - \mathbf{R}_{YY}$. It can be shown that $\mathbf{R}_{YY} = \sigma_x^2 \underline{A}\, \underline{A}^H + \sigma_w^2 \mathbf{I}$, where

$\sigma_x^2 = E[|x(n)|^2] = \sigma_s^2/M \sum_i E[|H_{iL}|^2]$, and the discrepancy matrix consists of three components

$$\tilde{\mathbf{R}}_{YY} = \left(\hat{\sigma}_x^2 - \sigma_x^2 \right) \underline{A}\, \underline{A}^H + \tilde{\mathbf{R}}_{WW} + \tilde{\mathbf{R}}_{YW} \quad \text{(B2)}$$

where $\tilde{\mathbf{R}}_{WW} = (1/N)\sum_{n=1}^{N}[\underline{W}(n)\underline{W}^H(n) - \sigma_w^2\mathbf{I}]$, $\tilde{\mathbf{R}}_{YW} = (1/N)\sum_{n=1}^{N}[\underline{A}x(n)\underline{W}^H(n) + \underline{W}(n)\underline{A}^H x^*(n)]$. The first component, $(\hat{\sigma}_x^2 - \sigma_x^2)\underline{A}\, \underline{A}^H$, is independent of SNR and the CFO estimation algorithm. Using the definition $\hat{\sigma}_w^2(k) = (1/N)\sum_{n=1}^{N}|w(k,n)|^2$, we express the second component as the sum of two matrices $\tilde{\mathbf{R}}_{WW} = \{\text{diag}[\hat{\sigma}_w^2(1), \hat{\sigma}_w^2(2), \cdots, \hat{\sigma}_w^2(K)] - \sigma_w^2\mathbf{I}\} + \hat{\mathbf{R}}_{WW}$, where the entries of the second matrix are given by $\hat{R}_{WW}(i,j) = [(1/N)\sum_{n=1}^{N} w(i,n)w^*(j,n)][1 - \delta(i-j)]$ while the first matrix is a diagonal matrix. The third component, $\tilde{\mathbf{R}}_{YW}$, represents the cross correlation of the received samples and noise. We will show that only $\hat{\mathbf{R}}_{WW}$ and $\tilde{\mathbf{R}}_{YW}$ affect the mean-squared performance of our CFO estimates.

We first note that the assumption that $\{w(m,n)\}$ are i.i.d. zero-mean Gaussian rvs with variance σ_w^2 leads immediately to $E\{\hat{R}_{WW}(i,j)\} = (1/N)\sum_{n=1}^{N} E\{w(i,n)w^*(j,n)\} = 0$ for $i \neq j$, and then the equations

$$E\left\{ \left|\hat{R}_{WW}(i,j)\right|^2 \right\} = \frac{1}{N^2}\sum_{n=1}^{N} E\left\{ |w(i,n)|^2 \right\} E\left\{ |w(j,n)|^2 \right\}$$
$$= \frac{1}{N}\sigma_w^4 \quad \text{(B3)}$$

for $i \neq j$

$$E\left\{ \hat{R}_{WW}(i,j)\hat{R}_{WW}^*(j,k) \right\}$$
$$= \frac{1}{N^2}\sum_{n=1}^{N} E\left\{ w(i,n)w(k,n)\left(w^*(j,n)\right)^2 \right\}$$
$$= 0 \quad \text{(B4)}$$

for $i \neq k$, and $E\{\hat{R}_{WW}(i,j)\hat{R}_{WW}^*(i,k)\} = 0$ for $j \neq k$. Furthermore, it can be shown that all of the elements in the upper or lower triangular part of $\hat{R}_{WW}$ are zero mean uncorrelated rvs with identical variance σ_w^4/N and $\hat{R}_{WW}(i,j) = \hat{R}_{WW}^*(j,i)$. Therefore, the rvs

$$u(n) \overset{\text{def}}{=} \sum_{\substack{i,j \\ j-i=n}} \hat{R}_{WW}(i,j), \quad n = 0, \pm 1, \cdots, \pm(K-1) \quad \text{(B5)}$$

have zero mean and variances $((K - |n|)/N)\sigma_w^4$ for $n \neq 0$. Moreover, $u(0) = 0$, $u(n) = u^*(-n)$, and $\{u(n), n = 1, \cdots, K - 1\}$ is a sequence of uncorrelated rvs. Next, let us examine the statistical properties of $\tilde{R}_{YW}$.

Recall that $\tilde{\mathbf{R}}_{YW} = (1/N)\sum_{n=1}^{N}[\underline{A}x(n)\underline{W}^H(n) + \underline{W}(n)\underline{A}^H x^*(n)]$, whose entry is given by

$$\tilde{R}_{YW}(i,j) = \frac{1}{N}\sum_{n=1}^{N}[x(n)a(i)w^*(j,n) + x^*(n)a^*(j)w(i,n)] \quad (B6)$$

with $a(i) = e^{j2\pi f_o(i-1)N/M}$. It follows immediately that $E[\tilde{R}_{YW}(i,j)] = 0$, for $x(n)$ is assumed to be uncorrelated with $w(i,n)$, for all i. Equation (B6) thus gives $E\{|\tilde{R}_{YW}(i,j)|^2\} = (2/N)\sigma_x^2\sigma_w^2$, for $i \neq j$. Similarly, we can show that the entries of the cross-correlation matrix $\tilde{R}_{YW}$ are uncorrelated unless they belong to the same column or row in the upper (or lower) triangular part of $\tilde{R}_{YW}(i,j)$. In other words, $E\{\tilde{R}_{YW}(i,j)\tilde{R}_{YW}^*(j,k)\} = 0$, for $k \neq i$, and $E\{\tilde{R}_{YW}(i,j)\tilde{R}_{YW}^*(p,q)\} = 0$, for $i \neq j \neq p \neq q$, but $E\{\tilde{R}_{YW}(i,j)\tilde{R}_{YW}^*(k,j)\} = (1/N)a(i)a^*(k)\sigma_x^2\sigma_w^2$, for $k \neq i$, and $E\{\tilde{R}_{YW}(i,j)\tilde{R}_{YW}^*(i,k)\} = (1/N)a^*(j)a(k)\sigma_x^2\sigma_w^2$, for $k \neq j$. The new zero-mean rvs

$$v(n) \overset{\text{def}}{=} \sum_{\substack{i,j \\ j-i=n}} \tilde{R}_{YW}(i,j), \quad n = 0, \pm 1, \cdots, \pm(K-1) \quad (B7)$$

have variances $(2(K-|n|)/N)\sigma_x^2\sigma_w^2$ for $n \neq 0$ and the Hermitian property $v(n) = v^*(-n)$ but they are not independent.

Now we can express the sequence $\{s(n)\}$ as

$$s(n) = \Delta(n) + K\sigma_w^2\delta(n) + u(n) + v(n), \quad n = 0, \pm 1, \cdots, \pm(K-1). \quad (B8)$$

Note that $s(n) = s^*(-n)$, $u(n)$, and $v(n)$ are uncorrelated.

B. Performance of Algorithm A

As Algorithm A calls for locating the point $\hat{\theta}$ (or $\hat{z}$) that maximizes $\Lambda(e^{j\theta}) \overset{\text{def}}{=} \Lambda(\theta)$ (or $\Lambda(z)$), it follows that $\dot{\Lambda}(\hat{\theta}) = 0$, where $\dot{\Lambda}(\theta) = d\Lambda(\theta)/d\theta = \sum_{n=-K+1}^{K-1} s(n)jne^{jn\theta}$. Assuming small estimation errors $\Delta\theta \ll 1$, we obtain $\dot{\Lambda}(\hat{\theta}) \cong \dot{\Lambda}(\theta_o) + \Upsilon(\theta_o)\Delta\theta$, where $\hat{\theta} = \theta_o + \Delta\theta$ and $\Upsilon(\theta_o) = (d\dot{\Lambda}(\theta)/d\theta)|_{\theta=\theta_o}$. The fact that $\dot{\Lambda}(\hat{\theta}) = 0$ implies $\Delta\theta \cong -\dot{\Lambda}(\theta_o)/\Upsilon(\theta_o)$, and thus the estimation error can be approximated by

$$\Delta\theta \cong \frac{\sum_{n=-K+1}^{K-1} jn[u(n)+v(n)]e^{jn\theta_o}}{\sum_{n=-K+1}^{K-1} n^2[\hat{\sigma}_x^2(K-|n|)e^{-jn\theta_o}+u(n)+v(n)]e^{jn\theta_o}}. \quad (B9)$$

The statistical properties of $u(n)$ and $v(n)$ derived in the previous section convince us that, with high probability, $\sum_{n=-K+1}^{K-1} n^2[u(n) + v(n)]e^{jn\theta_o} \ll \sum_{n=-K+1}^{K-1} n^2\hat{\sigma}_x^2(K - |n|) = \hat{\sigma}_x^2K^2(K^2 - 1)/6$, if $\gamma_b \geq 0$ dB. Ignoring the perturbations due to $u(n)$ and $v(n)$ in the denominator, we can further simplify (B9) to

$$\Delta\theta \cong \frac{6\xi}{\hat{\sigma}_x^2K^2(K^2-1)} \quad (B10)$$

where $\xi = \sum_{n=-K+1}^{K-1} jn[u(n)+v(n)]e^{jn\theta_o}$ is a zero-mean rv. Thus, the MSE of the estimate can be obtained as soon as the second moment of ξ is known.

Invoking the alternate expression $\xi = 2\sum_{n=1}^{K-1} n\text{Re}\{(u(n)+v(n))e^{j(n\theta_o+\pi/2)}\}$ and the facts that $u(n)$ and $v(n)$ are uncorrelated and $\{u(n)\}$ is white but not $\{v(n)\}$, we obtain

$$\text{var}(\xi) = \text{var}\left(\sum_{n=1}^{K-1} 2n\text{Re}\left\{(u(n))e^{j\left(n\theta_o+\frac{\pi}{2}\right)}\right\}\right)$$
$$+ \text{var}\left(\sum_{n=1}^{K-1} 2n\text{Re}\left\{(v(n))e^{j\left(n\theta_o+\frac{\pi}{2}\right)}\right\}\right)$$
$$= 4\sum_{n=1}^{K-1} n^2\left[\text{var}\left(\text{Re}\{u(n)\}\right) + \text{var}\left(\text{Re}\{v(n)\}\right)\right]$$
$$+ \Omega(K) \quad (B11)$$

where $\Omega(K)$ is the sum of the cross-correlation value of correlated elements $v(n)$'s, i.e.,

$$\Omega(K) = 4\sum_{m=1}^{K-1}\sum_{n=1}^{K-1} mnE\left[\text{Re}\left\{v(m)e^{j\left(m\theta_o+\frac{\pi}{2}\right)}\right\}\right.$$
$$\left.\cdot\text{Re}\left\{v(n)e^{j\left(n\theta_o+\frac{\pi}{2}\right)}\right\}\right](1-\delta(m-n)). \quad (B12)$$

As mentioned in Appendix II-A, $\text{var}(\text{Re}\{u(n)\}) = ((K-|n|)/2N)\sigma_w^4$ and $\text{var}(\text{Re}\{v(n)\}) = ((K-|n|)/N)\sigma_x^2\sigma_w^2$ for $n \neq 0$, after some algebra, we can show that the first term on the right-hand side of (B11) is equal to

$$4\sum_{n=1}^{K-1} n^2\left[\text{var}\left(\text{Re}\{u(n)\}\right) + \text{var}\left(\text{Re}\{v(n)\}\right)\right]$$
$$= \frac{K^2(K^2-1)\left(\sigma_w^4 + 2\sigma_x^2\sigma_w^2\right)}{6N}. \quad (B13)$$

The value of $\Omega(K)$ remains to be determined. For convenience, we define the auxiliary rv

$$O(m,n) = (n-m)\left[\tilde{R}_{YW}(m,n)e^{j\left((n-m)\theta_o+\frac{\pi}{2}\right)}\right.$$
$$\left.+\tilde{R}_{YW}(n,m)e^{-j\left((n-m)\theta_o+\frac{\pi}{2}\right)}\right] \quad (B14)$$

for $1 \leq m < K, 1 < n \leq K$, and $n > m$. If $n \leq m$, the associated rv $O(m,n)$ is zero with probability 1. It follows that entries of the upper triangular matrix $\Psi = [O(m,n)]$ are uncorrelated unless they belong to the same row or column. In other words, $E\{O(i,j)O(j,k)\} = 0$ for $k \neq i$ and $E\{O(i,j)O(p,q)\} = 0$ for $i \neq j \neq p \neq q$, but $E\{O(i,j)O(k,j)\} = (2/N)(j-i)(j-k)\sigma_x^2\sigma_w^2$ for $k \neq i$ and $E\{O(i,j)O(i,k)\} = (2/N)(j-i)(k-i)\sigma_x^2\sigma_w^2$ for $k \neq j$. By using (B7), (B12), and (B14), we can prove that $\Omega(K)$ is equivalent to the sum of cross correlations of the entries of Ψ, i.e.,

$$\Omega(K) = \sum_{i,j,p,q} E\{O(i,j)O(p,q)\}[\delta(p-i) \oplus \delta(q-j)]$$
$$= \frac{2\sigma_x^2\sigma_w^2}{N}\sum_{i,j,p,q}(j-i)(q-p)[\delta(p-i) \oplus \delta(q-j)],$$
$$1 \leq i,j,p,q \leq K \quad (B15)$$

$$\text{var}(c) = \frac{2N(K-1)(K+2)\gamma_b^{-1} + \left(NK(K-1) + 4K + 8N^{-1} + 20\right)\gamma_b^{-2} + 4(K-1)\gamma_b^{-3} + (K-1)\gamma_b^{-4}}{N^2(K-1)^2} \quad \text{(B22)}$$

where $\oplus$ denotes the "exclusive or" operator. As $\Omega(2) = 0$, $\Omega(3) = 8(2\sigma_x^2\sigma_w^2/N)$, and the recursive relation

$$\begin{aligned}
\Omega(K) &= \Omega(K-2) + 2\sum_{j,p,q} E\left\{O(1,j)O(p,q)\right\} \\
&\quad \times [\delta(p-1) \oplus \delta(q-j)] \\
&= \Omega(K-2) + \frac{(K-1)(K-2)(3K^2 - 7K + 6)}{3} \\
&\quad \times \frac{2\sigma_x^2\sigma_w^2}{N}
\end{aligned} \quad \text{(B16)}$$

holds for $K \geq 4$, we obtain

$$\Omega(K) = \frac{K(K^2-1)(K-2)(3K+1)\sigma_x^2\sigma_w^2}{15N}. \quad \text{(B17)}$$

This equation, along with (B10), (B11), and (B13), yields

$$\text{var}(\Delta\theta) = \frac{6\sigma_w^2\left[5K\sigma_w^2 + (6K^2 - 4)\sigma_x^2\right]}{5N\sigma_x^4K^3(K^2-1)}. \quad \text{(B18)}$$

The relation $\theta_o = (2\pi f_o N/M) = (2\pi f_o/L)$ then leads to

$$\text{var}(\hat{f}_o) = \frac{3L^3\left[5K\gamma_b^{-2} + (6K^2 - 4)\gamma_b^{-1}\right]}{10\pi^2MK^3(K^2-1)}. \quad \text{(B19)}$$

C. Performance of Algorithm B

Algorithm B calls for solving $Q(z) = 0$ or, equivalently, the evaluation of the corresponding roots $z_i = e^{j(\arg\{d\}/K + 2\pi/K)}$, $i = 0, \cdots, K-1$. Since $d = s(-1)/[(K-1)s(K-1)]$, $s(-1) = \hat{\sigma}_x^2(K-1)e^{j\theta_o} + v(-1) + c(-1)$, and $s(K-1) = \hat{\sigma}_x^2 e^{-j(K-1)\theta_o} + v(K-1) + c(K-1)$, it is convenient to define intermediate rvs $a = (v(-1) + c(-1))/(K-1)$ and $b = v(K-1) + c(K-1)$ (whose variances are given by $(\sigma_w^4 + 2\sigma_x^2\sigma_w^2)/(N(K-1))$ and $(\sigma_w^4 + 2\sigma_x^2\sigma_w^2)/N$, respectively) so that we can write

$$d = \frac{\hat{\sigma}_x^2 e^{j\theta_o} + a}{\hat{\sigma}_x^2 e^{-j(K-1)\theta_o} + b} = \frac{e^{j\theta_o} + \alpha}{e^{-j(K-1)\theta_o} + \beta} \quad \text{(B20)}$$

where $\alpha = a/\hat{\sigma}_x^2$ and $\beta = b/\hat{\sigma}_x^2$. Note that α and β^* are correlated but their correlation is a decreasing function of K and their variances are $(\gamma_b^{-2} + 2\gamma_b^{-1})/(N(K-1))$ and $(\gamma_b^{-2} + 2\gamma_b^{-1})/N$, respectively. After some algebra, we obtain

$$d = \frac{e^{jK\theta_o} + c}{\left|e^{-j(K-1)\theta_o} + \beta\right|^2} \quad \text{(B21)}$$

where $c = \alpha e^{j(K-1)\theta_o} + \beta^* e^{j\theta_o} + \alpha\beta^*$. It can be shown that $\text{var}(c)$, the variance of c, is given by (B22), shown at the top of the page. However, if we assume that α and β^* are uncorrelated,

the corresponding variance of c, denoted by $\text{var}_{ap}(c)$ is simply $\text{var}_{ap}(c) = \text{var}(\alpha) + \text{var}(\beta^*) + \text{var}(\alpha\beta^*)$, given by

$$\text{var}_{ap}(c) = \frac{\left(NK + \gamma_b^{-2} + 2\gamma_b^{-1}\right)\left(\gamma_b^{-2} + 2\gamma_b^{-1}\right)}{N^2(K-1)}. \quad \text{(B23)}$$

Numerical results show that the difference between $\text{var}(c)$ and $\text{var}_{ap}(c)$ is about 1 dB within the range of interest.

Equation (B21) implies $\arg\{d\} = \arg\{e^{jK\theta_o} + c\}$, and the polar coordinate representation $c = re^{j(K\theta_o + \phi)}$ gives

$$\begin{aligned}
\arg\{d\} &= \arg\left\{e^{jK\theta_o} + re^{j(K\theta_o + \phi)}\right\} \\
&= \arg\{e^{jK\theta_o}r'e^{j\psi}\}
\end{aligned} \quad \text{(B24)}$$

where r' and ψ are the amplitude and phase of $(1 + re^{j\phi})$, respectively. When $\text{var}(c) \ll 1$, i.e., $r\sin\phi \ll 1$ and $r\cos\phi \ll 1$, we have

$$\psi = \tan^{-1}\frac{r\sin\phi}{1 + r\cos\phi} \cong \frac{r\sin\phi}{1 + r\cos\phi} \cong r\sin\phi \quad \text{(B25)}$$

where we have used the approximation $\tan^{-1}x \approx x$, for $x \ll 1$. Hence, the variance of ψ can be obtained from $\text{var}(\psi) \cong \text{var}(r\sin\phi) = (1/2)\text{var}(c)$. Without loss of generality, we assume that the phase of the desired root of $Q(z)$ in (19) is $p_o = \arg\{d\}/K = \theta_o + \psi/K$, then the CFO estimate becomes $\hat{f}_o = (M/2\pi N)p_o = (M/2\pi N)(\theta_o + (\psi/K)) = f_o + (M/2\pi NK)\psi$, and, as a result, by substituting $L = M/N$, we obtain the variance of the estimate (22) as

$$\text{var}(\hat{f}_o) = \left(\frac{M}{2\pi NK}\right)^2\text{var}(\psi) \cong \frac{L^2}{8\pi^2K^2}\text{var}(c). \quad \text{(B26)}$$

REFERENCES

[1] J. Terry and J. Heiskala, *OFDM Wireless LANs: A Theoretical and Practical Guide*. Indianapolis, IN: Sams, 2001.

[2] P. H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Trans. Commun.*, vol. 42, pp. 2908–2914, Oct. 1994.

[3] F. Daffara and O. Adami, "A novel carrier recovery technique for orthogonal multicarrier systems," *Eur. Trans. Telecommun.*, vol. 7, pp. 323–334, July/Aug. 1996.

[4] H. Liu and U. Tureli, "An high-efficiency carrier estimator for OFDM communation," *IEEE Commun. Lett.*, vol. 2, pp. 104–106, Apr. 1998.

[5] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Trans. Commun.*, vol. 45, pp. 1613–1621, Dec. 1997.

[6] M. Morelli and U. Mengali, "An improved frequency offset estimator for OFDM applications," *IEEE Commun. Lett.*, vol. 3, pp. 75–77, Mar. 1999.

[7] H.-K. Song, Y.-H. You, J.-H. Paik, and Y.-S. Cho, "Frequency-offset synchronization and channel estimation for OFDM-based transmission," *IEEE Commun. Lett.*, vol. 4, pp. 95–97, Mar. 2000.

[8] S. Patel, L. S. Cimini, and B. McNair, "Comparison of frequency offset estimation techniques for burst OFDM," in *Proc. 55th IEEE Vehicular Technology Conf.*, Birmingham, Al, May 2002, pp. 772–776.

[9] J. Li, G. Liu, and G. B. Giannakis, "Carrier frequency offset estimation for OFDM based WLANs," *IEEE Signal Processing Lett.*, vol. 8, pp. 80–82, Mar. 2001.

[10] H. Minn, P. Tarasak, and V. K. Bhargava, "OFDM frequency offset estimation based on BLUE principle," in *Proc. IEEE Vehicular Technology Conf.*, Vancouver, BC, Canada, Sept. 2002, pp. 1230–1234.

[11] Y. S. Lim and J. H. Lee, "An efficient carrier frequency offset estimation scheme for an OFDM system," in *Proc. 52nd IEEE Vehicular Technology Conf.*, Sept. 2000, pp. 2453–2457.

[12] A. J. Coulson, "Maximum likelihood synchronization for OFDM using a pilot symbol: part 1: algorithms, part 2: analysis," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 2486–2503, Dec. 2001.

[13] T. Keller, L. Piazzo, P. Mandarini, and L. Hanzo, "Orthogonal frequency division multiplex synchronization techniques for frequency-selective fading channels," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 999–1008, June 2001.

[14] L. Piazzo, T. Keller, A. Falaschi, and P. Mandarini, "Time and frequency synchronization in DQPSK-OFDM based high speed wireless local area networks," *Eur. Trans. Telecommun.*, vol. 13, no. 3, pp. 279–284, May/June 2002.

[15] M. Ghogho and A. Swami, "Blind frequency-offset estimator for OFDM systems transmitting constant-modulus symbols," *IEEE Commun. Lett.*, vol. 6, pp. 343–345, Aug. 2002.

[16] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 720–741, May 1989.

[17] Y. Li and Y. L. Guan, "Modified Jakes' model for simulating multiple uncorrelated fading waveforms," in *Proc. 51st IEEE Vehicular Technology Conf.*, Tokyo, Japan, May 2000, pp. 1819–1822.

[18] G. H. Golub and C. F. VanLoan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

Jiun H. Yu was born in Nantou, Taiwan, on March 16, 1979. He received the B.S. and M.S. degrees in communication engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2001, and 2003, respectively.

He is currently a System Design Engineer with the Research and Development Center, Realtek Semiconductor Corporation, Hsinchu. His research interests include digital transmission systems, detection and estimation theory, and communication signal processing.

Yu T. Su (S'81–M'83) received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1983.

From 1983 to 1989, he was with LinCom Corporation, Los Angeles, where he was a Corporate Scientist, where his work involved the design of various measurement and digital satellite communication systems. Since September 1989, he has been with National Chiao Tung University, Hsinchu, Taiwan, where he was head of the Communication Engineering Department between 2001 and 2003. He is also affiliated with the Microelectronics and Information Systems Research Center of the same university and served as a Deputy Director from 1997 to 2000. His main research interests include communication theory and statistical signal processing.