Carrier-Frequency Estimation for Transmissions over Selective Channels

Michele Morelli and Umberto Mengali, *Fellow, IEEE*

***Abstract*—This paper deals with carrier-frequency estimation for burst transmissions over frequency-selective channels. Three estimation schemes are proposed, all based on the use of known training sequences. The first scheme employs an arbitrary sequence and provides joint maximum-likelihood (ML) estimates of the carrier frequency and the channel response. Its implementation complexity is relatively high but its accuracy achieves the Cramer–Rao bound. The second scheme is still based on the ML criterion, but the training sequence is periodic, which helps to reduce the computational load. The third scheme also employs periodic sequences, but its structure comes from heuristic reasoning. Theoretical analysis and simulations are employed to assess the performance of the three schemes.**

***Index Terms*—Frequency estimation, frequency-selective fading.**

I. Introduction

MANY digital communication systems operating over frequency-selective fading channels employ a signaling format consisting of frames of data, each preceded by a preamble of known symbols (training sequence). Preambles serve to estimate the channel response and to allow for fast start-up equalization and/or maximum-likelihood data sequence detection. Depending on the fading rate, the channel estimate may or may not be updated during the data sequence, perhaps in a decision-directed fashion. Channel estimation through training sequences (TSs) is an important issue in time-division multiple-access radio systems and HF digital transmissions. Papers [1]–[4] and references therein give a representative sample of the results in this area.

In dealing with channel estimation, most investigators assume zero frequency offset between the carrier and the local reference at the receiver. In practice, this means that the offset is so small that the demodulated signal incurs only negligible phase rotations during the preamble duration. Using stable oscillators is not a viable route to meet such conditions for, in general, the stability requirements would be too stringent. Furthermore, even ideal oscillators would be inadequate in a mobile communication environment experiencing significant Doppler shifts. The only solution is to measure the frequency offset accurately.

This problem has been investigated in [5] and [6] for the additive white Gaussian noise (AWGN) channel and in [7] for flat fading. The case of frequency-selective fading has been addressed by Hebley and Taylor in [8]. These authors assume that the channel is unknown and derive a low-SNR approximation to the maximum-likelihood (ML) offset estimator. As an averaging over the channel realizations is involved in their scheme, they assume that the channel statistics are known or can be measured in some way.

In the present paper, we return to the problem discussed by Hebley and Taylor, but we aim at the joint ML estimation of the channel response and the frequency offset. As we shall see, the solution consists of two separate step: a frequency-offset estimator (that we call MLE#1) and a channel estimator. The latter is the traditional scheme brought out in [2] for zero offset, save that the frequency error is first compensated in the input signal.

As we shall see, the MLE#1 is a powerful estimator. Its accuracy achieves the Cramer–Rao bound (CRB) and its estimation range is as large as $\pm 50\%$ of the symbol rate. However, as it is rather complex to implement, we also consider other solutions. In particular, we show that if the TS is composed of L identical blocks, then the computational load of the ML estimator can be cut down by a factor L. The price to be paid is a narrowing of the estimation range by the same factor. This scheme is denoted MLE#2.

As a further simplification, a third estimation method is derived following *ad hoc* reasoning. Like MLE#2, it operates on periodic TSs and achieves the CRB at high signal-to-noise ratio (SNR).

The rest of the paper is organized as follows. In the next section, we describe the signal model and introduce basic notations. Section III deals with the joint ML estimation of the channel response and the frequency offset. The performance of the MLE#1 is investigated in Section IV. The MLE#2 is derived in Section V, while Section VI deals with the *ad hoc* estimator. Simulation results are discussed in Section VII, and some conclusions are offered in Section VIII.

II. Signal Model

We assume a linear modulation (e.g., PSK or QAM) and a frequency-selective channel with a slow evolution in time relative to the signaling interval T_s. Under such conditions, the received signal samples, taken at symbol rate, are expressed by

$$x(n) = e^{j2\pi n\nu} s(n) + w(n), \qquad n = 0, 1, \ldots, N-1 \quad (1)$$

where ν is the carrier-frequency offset normalized to $1/T_s$, $w(n)$ is Gaussian noise, and $s(n)$ is given by

$$s(n) = \sum_{k=0}^{L-1} h(k) a_{n-k}, \qquad n = 0, 1, \ldots, N-1. \quad (2)$$

Paper approved by S. Roy, the Editor for Communication Theory/Systems of the IEEE Communications Society. Manuscript received February 23, 1999; revised November 25, 1999.

The authors are with the Department of Information Engineering, University of Pisa, 56100 Pisa, Italy (e-mail: morelli@iet.unipi.it).

Publisher Item Identifier S 0090-6778(00)07527-9.

Here, L represents the channel memory, $\boldsymbol{h} = [h(0), h(1), \ldots, h(L-1)]^{\mathrm{T}}$ is a vector containing the T_s-spaced samples of the channel response (the superscript $(\cdot)^{\mathrm{T}}$ indicates vector transpose) and $\{a_n;\ -L+1 \le n \le N-1\}$ are the training symbols. The $L-1$ symbols at the start of the preamble, prior to the first observation at $n=0$, are the *precursors*.

Equation (1) can be written in matrix form as

$$\boldsymbol{x} = \boldsymbol{\Gamma}(\nu)\boldsymbol{A}\boldsymbol{h} + \boldsymbol{w} \tag{3}$$

where $\boldsymbol{x} = [x(0), x(1), \ldots, x(N-1)]^{\mathrm{T}}$, $\boldsymbol{\Gamma}(\nu)$ is a diagonal matrix

$$\boldsymbol{\Gamma}(\nu) = \mathrm{diag}\{1, e^{j2\pi\nu}, e^{j4\pi\nu}, \ldots, e^{j2\pi(N-1)\nu}\} \tag{4}$$

and $\boldsymbol{A}$ is a $N \times L$ matrix with entries

$$[\boldsymbol{A}]_{i,j} = a_{i-j}, \qquad 0 \le i \le N-1,\ \ 0 \le j \le L-1. \tag{5}$$

Finally, $\boldsymbol{w} = [w(0), w(1), \ldots, w(N-1)]^{\mathrm{T}}$ is a zero-mean Gaussian vector with covariance matrix

$$\boldsymbol{C}_{\boldsymbol{w}} = \mathrm{E}\left\{\boldsymbol{w}\boldsymbol{w}^H\right\} = \sigma_n^2 \boldsymbol{I}_N \tag{6}$$

where $\boldsymbol{I}_N$ is the $N \times N$ identity matrix and the superscript $(\cdot)^H$ means "Hermitian transpose."

The SNR is defined as

$$\mathrm{SNR} = \sigma_s^2/\sigma_n^2 \tag{7}$$

with

$$\sigma_s^2 = \frac{1}{N}\sum_{n=0}^{N-1} |s(n)|^2. \tag{8}$$

III. ML Frequency Estimation

For a given pair $(\boldsymbol{h}, \nu)$, the vector $\boldsymbol{x}$ is Gaussian with mean $\boldsymbol{\Gamma}(\nu)\boldsymbol{A}\boldsymbol{h}$ and covariance matrix $\sigma_n^2\boldsymbol{I}_N$. Thus, the likelihood function for the parameters $(\boldsymbol{h}, \nu)$ takes the form

$$\Lambda(\boldsymbol{x}; \tilde{\boldsymbol{h}}, \tilde{\nu}) = \frac{1}{(\pi\sigma_n^2)^N} \cdot \exp\left\{-\frac{1}{\sigma_n^2}\left[\boldsymbol{x} - \boldsymbol{\Gamma}(\tilde{\nu})\boldsymbol{A}\tilde{\boldsymbol{h}}\right]^H \left[\boldsymbol{x} - \boldsymbol{\Gamma}(\tilde{\nu})\boldsymbol{A}\tilde{\boldsymbol{h}}\right]\right\} \tag{9}$$

where $\tilde{\boldsymbol{h}}$ and $\tilde{\nu}$ are trial values of $\boldsymbol{h}$ and ν. As indicated in [9], two possible approaches may be taken to estimate ν according to the ML criterion. One is the Bayesian approach adopted in [8], which consists of modeling $\tilde{\boldsymbol{h}}$ as a random vector with some probability density $p(\tilde{\boldsymbol{h}})$ and computing the average of $\Lambda(\boldsymbol{x}; \tilde{\boldsymbol{h}}, \tilde{\nu})$ with respect to $p(\tilde{\boldsymbol{h}})$. This gives the *marginal* likelihood of ν, from which the desired estimate of ν is obtained looking for the location of the maximum. The alternative approach, the one we follow here, aims at maximizing $\Lambda(\boldsymbol{x}; \tilde{\boldsymbol{h}}, \tilde{\nu})$ over $\tilde{\boldsymbol{h}}$ and $\tilde{\nu}$. The location of the maximum gives the *joint* ML estimates of $\boldsymbol{h}$ and ν.

To proceed, we keep $\tilde{\nu}$ fixed in (9) and let $\tilde{\boldsymbol{h}}$ vary in the $2L$-dimensional space C^L. In these conditions, $\Lambda(\boldsymbol{x}; \tilde{\boldsymbol{h}}, \tilde{\nu})$ achieves a maximum for

$$\hat{\boldsymbol{h}}(\tilde{\nu}) = (\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H\boldsymbol{\Gamma}^H(\tilde{\nu})\boldsymbol{x}. \tag{10}$$

Next, substituting (10) into (9) and varying $\tilde{\nu}$, it is found that maximizing (9) is equivalent to maximizing

$$g(\tilde{\nu}) = \boldsymbol{x}^H\boldsymbol{\Gamma}(\tilde{\nu})\boldsymbol{B}\boldsymbol{\Gamma}^H(\tilde{\nu})\boldsymbol{x} \tag{11}$$

where $\boldsymbol{B}$ is the *projection matrix*

$$\boldsymbol{B} = \boldsymbol{A}(\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H. \tag{12}$$

In summary, the ν-estimator reads

$$\hat{\nu} = \arg\max_{\tilde{\nu}}\{g(\tilde{\nu})\}. \tag{13}$$

In the following, this is called the ML estimator 1 (MLE#1).

Notice that (11) may be put in the form

$$g(\tilde{\nu}) = -\rho(0) + 2\mathrm{Re}\left\{\sum_{m=0}^{N-1}\rho(m)e^{-j2\pi m\tilde{\nu}}\right\} \tag{14}$$

where $\mathrm{Re}\{\cdot\}$ denotes the real part of the enclosed quantity, $\rho(m)$ is a weighted correlation of the data

$$\rho(m) = \sum_{k=m}^{N-1}[\boldsymbol{B}]_{k-m,k}\,x(k)x^*(k-m) \tag{15}$$

and $[\boldsymbol{B}]_{i,j}$ is the (i,j)-entry of $\boldsymbol{B}$.

Some remarks are of interest as follows.

1) From (10), (11), and (13), it appears that the estimates of ν and $\boldsymbol{h}$ are *decoupled*, meaning that the former can be computed first and then exploited to get the latter. Actually, the estimate of $\boldsymbol{h}$ is obtained by setting $\tilde{\nu} = \hat{\nu}$ into (10)

$$\hat{\boldsymbol{h}} = (\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H\boldsymbol{\Gamma}^H(\hat{\nu})\boldsymbol{x}. \tag{16}$$

This result coincides with the classical channel estimate [2] obtained for $\nu = 0$, save that the signal samples $x(n)$ are counter-rotated to compensate for the frequency offset (in fact the vector $\boldsymbol{\Gamma}^H(\hat{\nu})\boldsymbol{x}$ in (16) has components $\{x(n)e^{-j2\pi n\hat{\nu}}\}$).

2) If $\boldsymbol{A}$ is square $(N = L)$ and nonsingular, then from (12) one sees that $\boldsymbol{B} = \boldsymbol{I}_N$ and the right-hand side (RHS) of (11) becomes independent of $\tilde{\nu}$. In these conditions, maximizing $g(\tilde{\nu})$ is meaningless and the MLE#1 cannot be computed. Physically, this means that the L data $\boldsymbol{x}$ are insufficient to estimate the $L+1$ parameters $(\boldsymbol{h}, \nu)$.

3) For $L = 1$, the channel has no intersymbol interference and the matrix $\boldsymbol{A}$ reduces to the vector $\boldsymbol{A} = [a_0, a_1, \ldots, a_{N-1}]^{\mathrm{T}}$. Then, (11) becomes

$$g(\tilde{\nu}) = \frac{1}{\sum_{n=0}^{N-1}|a_n|^2}\left|\boldsymbol{A}^H\boldsymbol{\Gamma}^H(\tilde{\nu})\boldsymbol{x}\right|^2 \tag{17}$$

TABLE I
COMPUTATIONAL LOAD

Algorithm	Real products	Real additions
MLE#1	$2N[2N+2+K\eta\log_2(KN)]$	$N[3N+1+3K\eta\log_2(KN)]$
MLE#2	$2N[N+L+K\eta\log_2(KN/L)]/L$	$N[2N+2L-2+3K\eta\log_2(KN/L)]/L$
AHE	$2M(2N-ML-L)+M$	$2M(2N-ML-L)-2$
HTE	$3H(2N-H-1)+1$	$2H(2N-H-1)-2$

and the MLE#1 takes the form

$$\hat{\nu}=\arg\max_{\tilde{\nu}}\left\{\left|\sum_{n=0}^{N-1}x(n)a_n^*e^{-j2\pi n\tilde{\nu}}\right|^2\right\}. \tag{18}$$

With an unmodulated carrier ($a_n=1$, $0\le n\le N-1$), this equation reduces to the ML estimator proposed by Rife and Boorstyn [5] for the AWGN channel.

4) As discussed in [5], the maximization of the RHS of (14) requires a two-step procedure. The first one (*coarse search*) computes $g(\tilde{\nu})$ over a grid of $\tilde{\nu}$-values, say $\{\tilde{\nu}_n\}$, and determines the location $\tilde{\nu}_M$ of the maximum. In the second step (*fine search*), the $g(\tilde{\nu}_n)$-values are interpolated and the local maximum nearest to $\tilde{\nu}_M$ is found. Note that, occasionally, the function $g(\tilde{\nu})$ is so distorted by the noise that its maximum is far from the true ν. When this occurs, the MLE#1 makes large errors (*outliers*). The SNR below which the outliers start to occur is called the *threshold* of the estimator.
5) Because of the very form of (14), the $g(\tilde{\nu}_n)$ values can be efficiently computed through fast Fourier transform (FFT) techniques. To this end, the N correlations $\rho(m)$ are first calculated from (15). Next, the sequence

$$\rho'(m)=\begin{cases}\rho(m), & 0\le m\le N-1\\ 0, & N\le m\le NK-1\end{cases} \tag{19}$$

is formed, K being a design parameter (*pruning factor*) [10]. Finally, the FFT of $\rho'(m)$ is computed for

$$\tilde{\nu}_n=\frac{n}{KN}\qquad -\frac{KN}{2}\le n<\frac{KN}{2}. \tag{20}$$

This yields the quantities in brackets in (14) and, ultimately, the set $\{g(\tilde{\nu}_n)\}$.
6) From (14), it is seen that $g(\tilde{\nu})$ is periodic of unit period and, in consequence, its maxima occur at a (normalized) distance of 1 from each other. Thus, the MLE#1 gives ambiguous estimates unless the true ν is confined within the interval $|\nu|\le 1/2$. This is the *estimation range* of the MLE#1. It should be noted that $|\nu|\le 1/2$ is the maximum estimation range that can be expected for any frequency estimator operating on baud rate samples.

IV. PERFORMANCE OF THE MLE#1

The performance of the MLE#1 can be assessed with the methods indicated in [11]. Assuming $\mathrm{SNR}\gg 1$ and an observation window extending over many symbol intervals, the expectation and variance of the MLE#1 are approximated by

$$\mathrm{E}\{\hat{\nu}\}\cong\nu-\frac{\mathrm{E}\{\dot{g}(\nu)\}}{\mathrm{E}\{\ddot{g}(\nu)\}} \tag{21}$$

$$\mathrm{var}\{\hat{\nu}\}\cong\frac{\mathrm{E}\left\{[\dot{g}(\nu)]^2\right\}}{\left[\mathrm{E}\{\ddot{g}(\nu)\}\right]^2} \tag{22}$$

where $\dot{g}(\nu)$ and $\ddot{g}(\nu)$ represent the first and second derivatives of $g(\tilde{\nu})$ at $\tilde{\nu}=\nu$. Exploiting these relations, in Appendix A, it is found that

$$\mathrm{E}\{\hat{\nu}\}=\nu \tag{23}$$

$$\mathrm{var}\{\hat{\nu}\}=\frac{1}{2N}\,\frac{\boldsymbol{s}^H\boldsymbol{s}}{\boldsymbol{y}^H(\boldsymbol{I}_N-\boldsymbol{B})\boldsymbol{y}}\,(\mathrm{SNR})^{-1} \tag{24}$$

where $\boldsymbol{s}=\boldsymbol{A}\boldsymbol{h}$ and $\boldsymbol{y}$ is a vector with components

$$y(n)=2\pi n\cdot s(n),\qquad n=0,1,\ldots,N-1. \tag{25}$$

Equation (23) says that the MLE#1 is unbiased for any channel. Furthermore, in Appendix B, it is shown that its variance coincides with the CRB. These results are in keeping with the asymptotic efficiency property of the ML estimator [9, p. 167].

From (24), it is seen that the accuracy of the MLE#1 depends on the channel response (through $\boldsymbol{s}$ and $\boldsymbol{y}$), the SNR, and the TS (through $\boldsymbol{s}$, $\boldsymbol{y}$ and $\boldsymbol{B}$). The TS also affects the variance of the channel estimates. Optimal binary TSs for channel estimation are found by computer search in [2] and [3]. However, best TSs for *joint* frequency and channel estimation are not explicitly known and need further investigation.

The computational complexity of the MLE#1 can be assessed as follows. Assume that the entries of the matrix $\boldsymbol{B}$ have been precomputed and stored. Then, the correlations in (15) require a total of $N(N+1)$ complex products and $N(N-1)/2$ complex additions. Also, the FFT needs $(KN/2)\log_2(KN)$ complex products and $(KN)\log_2(KN)$ complex additions. The overall operations are summarized in the first row of Table I. In writing these figures, we have born in mind that a complex product amounts to four real products plus two real additions, while a complex addition is equivalent to two real additions. The coefficient η in Table I is defined as

$$\eta=1-\frac{\log_2 K+2(1/K-1)}{\log_2(KN)} \tag{26}$$

and accounts for the computational saving achievable by skipping the operations on the zeros in the FFT [10]. Notice that η

equals 1 for $K = 2$, which means that no saving is possible unless $K > 2$. It should be stressed that the operations in Table I take into account only the coarse search; the fine search is comparatively easier and can be neglected.

V. ML Estimation with Periodic Training Sequences

Imposing some structure on the TS can reduce the complexity of the ML estimator. In the following, we consider periodic sequences, meaning that they result from the repetition of a fixed block of symbols. For convenience, we choose a block of length L (the duration of the channel response) and an observation interval of $N = LP$ symbols. As we shall see, in this way the complexity of the estimator can be cut down by a factor L.

To proceed, we observe that, in the present conditions, the matrix $\boldsymbol{A}$ in (5) may be written as

$$\boldsymbol{A} = [\boldsymbol{C}, \boldsymbol{C}, \ldots \boldsymbol{C}]^T \tag{27}$$

where $\boldsymbol{C}$ is an $L \times L$ circulant matrix with elements

$$[\boldsymbol{C}]_{i,j} = a_{|i-j|_L}, \qquad 0 \le i, j \le L-1 \tag{28}$$

and $|i-j|_L$ means "$i-j$ modulo L." Then, substituting (27) into (12), it is found that $\boldsymbol{B}$ becomes an array of P^2 elements

$$\boldsymbol{B} = \frac{1}{P}\begin{bmatrix} \boldsymbol{I}_L & \boldsymbol{I}_L & \cdots & \boldsymbol{I}_L \\ \boldsymbol{I}_L & \boldsymbol{I}_L & \cdots & \boldsymbol{I}_L \\ \vdots & \vdots & & \vdots \\ \boldsymbol{I}_L & \boldsymbol{I}_L & \cdots & \boldsymbol{I}_L \end{bmatrix} \tag{29}$$

where $\boldsymbol{I}_L$ denotes the $L \times L$ identity matrix. In other words, the elements of $\boldsymbol{B}$ are given by

$$[\boldsymbol{B}]_{k-m,\,k} = \begin{cases} 1/P, & \text{for } m = iL, \;\; 0 \le i \le P-1 \\ 0, & \text{otherwise} \end{cases} \tag{30}$$

and, in consequence, $g(\tilde{\nu})$ in (14) takes the form

$$g(\tilde{\nu}) = -\xi(0) + 2\mathrm{Re}\left\{\sum_{m=0}^{P-1} \xi(m) e^{-j2\pi mL\tilde{\nu}}\right\} \tag{31}$$

with

$$\xi(m) = \frac{1}{P}\sum_{k=mL}^{N-1} x(k)x^*(k-mL), \qquad m = 0, 1, \ldots, P-1. \tag{32}$$

As usual, the ML estimate of ν is obtained by maximizing $g(\tilde{\nu})$. This estimator is referred to as MLE#2.

The following remarks are of interest.

1) The complexity reduction afforded by a periodic TS is clear by comparing (31) and (32) with (14) and (15). Only P correlations $\xi(m)$ are involved in (31), whereas N correlations $\rho(m)$ appear in (14). Furthermore, the FFT is now performed over KP points, not KN. The second row in Table I shows the computational load involved in the coarse search of the MLE#2. The number of operations is approximately reduced by a factor L with respect to the MLE#1.
2) From (31), it is seen that $g(\tilde{\nu})$ is periodic of period $1/L$. This means that the MLE#2 gives ambiguous estimates unless the true ν is confined within $|\nu| \le 1/(2L)$. Thus, the estimation range of MLE#2 is L times narrower than that of MLE#1.
3) The estimation variance of the MLE#2 is obtained by substituting (29) into (24). After some algebra, it is found

$$\mathrm{var}\{\hat{\nu}\} = \frac{3(\mathrm{SNR})^{-1}}{2\pi^2 N(N^2 - L^2)}. \tag{33}$$

Note that the RHS of (33) is still the CRB (for periodic sequences). It appears that the estimation accuracy increases as L decreases and achieves a maximum for $L = 1$. Correspondingly, the CRB equals the result obtained in [5] with an unmodulated sinusoid and AWGN channel.
4) For $P = 2$, the location of the maximum of (31) is found explicitly as

$$\hat{\nu} = \frac{1}{\pi N}\arg\left\{\sum_{k=N/2}^{N-1} x(k)x^*(k-N/2)\right\}. \tag{34}$$

As we shall see in the next section, the complexity of (34) is significantly smaller than that of the standard MLE#2 (i.e., with $P \neq 2$).

VI. *Ad Hoc* Frequency Estimation

Although the MLE#2 is simpler than the MLE#1, for $P \neq 2$ it still requires a grid search. Here, we propose a third scheme that does not need any search and still achieves the CRB at high SNR values. Our approach is *ad hoc* and can be explained as follows.

We make the same assumptions as with the MLE#2 and, in particular, we take a periodic sequence of period L and an observation interval of length $N = LP$, a multiple of L. The proposed method exploits the correlations

$$R(m) = \frac{1}{N-mL}\sum_{k=mL}^{N-1} x(k)x^*(k-mL), \qquad m = 0, 1, \ldots, M \tag{35}$$

where M is a design parameter not greater than $P-1$. To see how the $R(m)$ can be used, let us write (1) in the form

$$x(n) = e^{j2\pi n\nu}[s(n) + \tilde{w}(n)] \tag{36}$$

with $\tilde{w}(n) = w(n)e^{-j2\pi n\nu}$. Then, inserting into (35) yields

$$R(m) = \sigma_s^2 e^{j2\pi mL\nu}[1 + \gamma(m)] \tag{37}$$

with

$$\gamma(m) = \frac{1}{\sigma_s^2(N-mL)}\sum_{k=mL}^{N-1} \cdot [s(k)\tilde{w}^*(k-mL) + s^*(k)\tilde{w}(k) + \tilde{w}(k)\tilde{w}^*(k-mL)]. \tag{38}$$

In deriving these equations, we have used the fact that $s(n)$ is periodic of period L [as is clear from (2)]. For the same reason, we rewrite (8) as

$$\sigma_s^2 = \frac{1}{L}\sum_{n=0}^{L-1}|s(n)|^2. \tag{39}$$

Next, defining the angle

$$\varphi(m) = \arg\{R(m)R^*(m-1)\} \tag{40}$$

and denoting by $\gamma_I(m)$ the imaginary part of $\gamma(m)$, it can be shown that the following approximation holds [12]:

$$\varphi(m) \cong 2\pi\nu L + \gamma_I(m) - \gamma_I(m-1), \qquad m = 1, 2, \ldots, M \tag{41}$$

provided that L satisfies the condition $|2\pi\nu L| < \pi$ and the SNR is sufficiently high to render $\gamma_I(m)$ and $\gamma_I(m-1)$ small in amplitude compared to unity.

Equation (41) indicates that $\varphi(m)$ is the sum of a deterministic component (proportional to ν) plus a zero-mean random disturbance. Thus, the set $\{\varphi(m)\}$ can be used to compute the best linear unbiased estimator of ν. This estimator has the form [9, p. 136]

$$\hat{\nu} = \frac{1}{2\pi L}\sum_{m=1}^{M}\alpha(m)\varphi(m) \tag{42}$$

where $\alpha(m)$ are the components of

$$\boldsymbol{\alpha} = \frac{\boldsymbol{C}_\varphi^{-1}\mathbf{1}}{\mathbf{1}^{\mathrm{T}}\boldsymbol{C}_\varphi^{-1}\mathbf{1}} \tag{43}$$

$\boldsymbol{C}_\varphi$ is the covariance matrix of $\boldsymbol{\varphi} = [\varphi(1), \varphi(2), \ldots, \varphi(M)]^{\mathrm{T}}$ and $\mathbf{1} = [1, 1, \ldots, 1]^{\mathrm{T}}$ is a vector of M ones. The variance of (42) is found to be [9]

$$\mathrm{var}\{\hat{\nu}\} = \frac{1}{(2\pi L)^2}\,\frac{1}{\mathbf{1}^{\mathrm{T}}\boldsymbol{C}_\varphi^{-1}\mathbf{1}}. \tag{44}$$

In the sequel, we call (42) the *ad hoc* estimator (AHE).

The computation $\boldsymbol{C}_\varphi^{-1}$ is carried out in Appendix C. It turns out that $\boldsymbol{C}_\varphi$ is singular for $M > P/2$. In fact, the angles $\varphi(m)$ with $m > P/2$ are linearly dependent on those with $1 \le m \le P/2$. Physically, this means that the information borne by $\boldsymbol{\varphi}$ is entirely contained in its first $P/2$ components. For this reason, in the following we assume $1 \le M \le P/2$.

Using the results in Appendix C, (43) and (44) become

$$\alpha(m) = 3\frac{(P-m)(P-m+1) - M(P-M)}{M(4M^2 - 6PM + 3P^2 - 1)}, \qquad 1 \le m \le M \tag{45}$$

$$\mathrm{var}\{\hat{\nu}\} = \frac{3\times(\mathrm{SNR})^{-1}}{4\pi^2 ML^3(4M^2 - 6PM + 3P^2 - 1)}. \tag{46}$$

Note that, as M varies, (46) achieves a minimum for $M = P/2$ which coincides with the CRB as expressed by the RHS of (33).

The following remarks about the AHE are in order.

1) Equation (45) says that the weights $\alpha(m)$ are independent of the channel. Thus, no channel information is needed to implement the AHE.
2) As (42) provides an explicit expression for $\hat{\nu}$, no grid search is needed and the computational load is reduced in comparison with MLE#1 and MLE#2. This is seen from the third row in Table I, which indicates that the number of operations with AHE increases only linearly with N while it increases as N^2 with MLE#1 and MLE#2.
3) Bearing in mind that $1 \le M \le P/2$, we have $M = 1$ for $P = 2$ (equivalently, for $N = 2L$). In these conditions, (42) reduces to (34), which means that AHE and MLE#2 coincide.
4) From the discussion leading to (42), it is clear that the AHE gives correct results as long as the approximation (41) holds, which occurs for $|2\pi\nu L| < \pi$ (and high SNR values). Otherwise, the relation between $\varphi(m)$ and ν is highly nonlinear and the estimates become inaccurate. In summary, the estimation range of AHE is limited to $|\nu| \le 1/2L$, exactly as with the MLE#2.
5) At this point, it is interesting to introduce the Hebley and Taylor estimator (HTE) [8]

$$\hat{\nu} = \frac{1}{(H+1)\pi}\arg\left\{\sum_{m=1}^{H}\psi(m)\right\}. \tag{47}$$

In this equation, H is a design parameter, $\psi(m)$ is defined as

$$\psi(m) = \frac{1}{N-m}\sum_{k=m}^{N-1}x(k)x^*(k-m) \cdot \sum_{l_1=0}^{L-1}\sum_{l_2=0}^{L-1}\mu(l_1, l_2)a_{k-l_1}^* a_{k-m-l_2} \tag{48}$$

and $\mu(l_1, l_2) = \mathrm{E}\{h^*(l_1)h(l_2)\}$ is the channel correlation function. It turns out that increasing H improves the estimation accuracy and decreases the estimation range, which is given by $|\nu| \le 1/(H+1)$. The computational load associated with (47) is shown in the fourth row of Table I [the figures do not account for the computation of $\mu(l_1, l_2)$]. It should be stressed that the HTE operates with arbitrary (*aperiodic*) TSs.
6) Comparisons between AHE and HTE are not simple. Estimation range, estimation accuracy, and implementation complexity in each scheme depend on various design parameters that can be traded off in various ways to meet assigned performance requirements. In the next section, we return on this issue in a specific case.

VII. Simulation Results

Computer simulations have been run to check and extend the analytical results of the previous sections. In a first set of simulations, we have assumed a QPSK format. All the pulse shaping is performed at the transmitter through a raised-cosine rolloff filter $g_T(t)$ with a rolloff of 0.5. The receiver front end is only equipped with a noise-rejection filter. The transmission medium is modeled as the typical urban channel of the European GSM

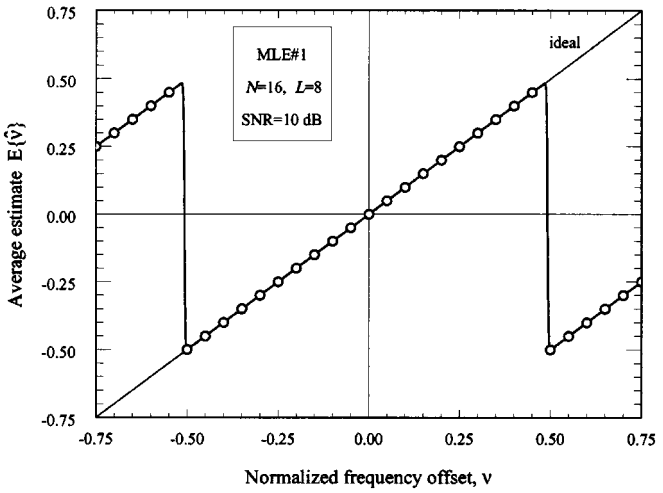


Fig. 1. Average frequency estimate versus ν for MLE#1.

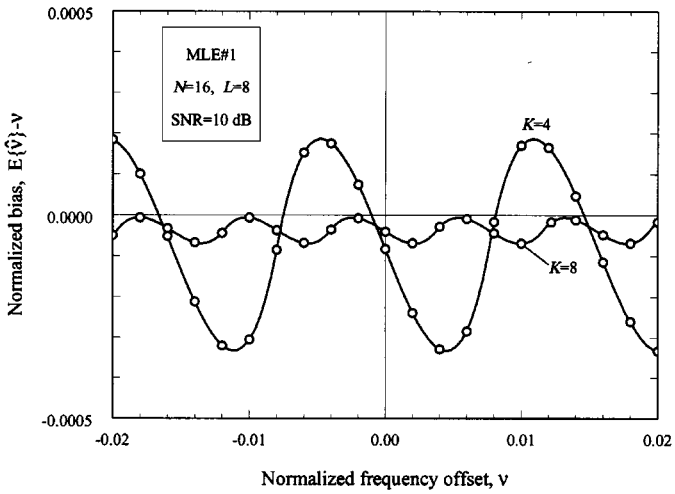


Fig. 2. Bias versus ν for MLE#1.

system [13] with six paths. The channel response is given the form

$$h(k) = \sum_{i=0}^{5} A_i g_T(kT_s - \tau_i - t_0) \tag{49}$$

where $\{A_i\}$ and $\{\tau_i\}$ are attenuations and delays of the paths and t_0 is a timing phase which is chosen equal to $3T_s$. The normalized delays $\{\tau_i/T_s\}$ are chosen equal to {0, 0.054, 0.135, 0.432, 0.621, 1.351}, while the $\{A_i\}$ are independent and Gaussian random variables with zero mean and variances (in decibels) $\{-3, 0, -2, -6, -8, -10\}$. In these conditions, the RHS of (49) takes significant values only for $0 \leq k \leq 7$, which means that the model (2) holds true with $L = 8$.

The TSs are taken from [3]. It is worth noting that they are optimal for channel estimation and perfect frequency recovery (i.e., $\hat{\nu} = \nu$) but are not guaranteed to be optimal for frequency estimation.

When dealing with MLE#1 and MLE#2, a pruning factor of 4 is normally adopted. Also, a parabolic interpolation is chosen in the implementation of the fine search in MLE#1 and MLE#2. The parameters M and H in AHE and HTE are chosen equal to $P/2$ and $N/2$, respectively. Perfect knowledge of the channel correlations $\mathrm{E}\{h^*(l_1)h(l_2)\}$ is assumed with HTE.

Fig. 1 illustrates the average estimates versus ν as obtained with MLE#1. The observation length is $N = 16$ and the TS, expressed in hexadecimal form, is CC14. The bits are transmitted as BPSK symbols using the mapping $0 \rightarrow -1, 1 \rightarrow 1$. The ideal line $\mathrm{E}\{\hat{\nu}\} = \nu$ is also shown for comparison. The estimates seem unbiased over a range as large as $|\nu| \leq 1/2$. In reality, they are affected by a micro-bias (see Fig. 2) whose strength decreases as the pruning factor K increases. The micro-bias is a consequence of the interpolation operation in the fine search.

Fig. 3 gives the mean square error (MSE) $\mathrm{E}\{[\hat{\nu} - \nu]^2\}$ versus SNR for the MLE#1 with two observation lengths, $N = 16$ and $N = 32$. The corresponding TSs are CC14 and 5230F641 [3]. In each simulation run, the true offset is taken randomly from the interval $(-0.1, 0.1]$ and the channel response is generated from (49). The CRB is computed from the RHS of (24). For a given SNR, the solid lines labeled "min/max CRB" indicate the

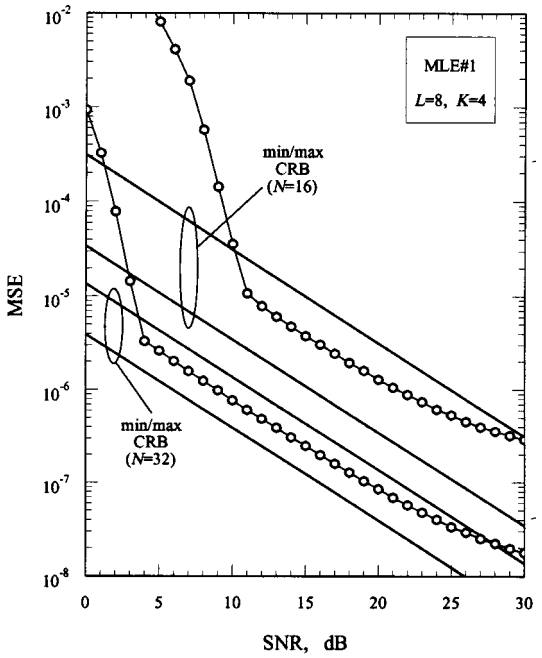


Fig. 3. Accuracy of MLE#1.

minimum and maximum CRB obtained in 10^5 simulation runs. It is seen that, at high SNR values, the MSE exhibits a floor as a consequence of the micro-bias. At intermediate SNRs, the MSE lies between the bounds. Finally, as the SNR decreases, the MSE shows an abrupt increase reflecting the occurrence of outliers. The SNR at which the increase begins establishes the estimator threshold.

Fig. 4 shows the normalized mean square channel estimation error (MSCEE) $\mathrm{E}\{\|\hat{\boldsymbol{h}} - \boldsymbol{h}\|^2\}/\sigma_s^2$ versus SNR for MLE#1 ($\|\cdot\|$ denotes the norm of the enclosed vector). The operating conditions are the same as in Fig. 3. Marks indicate simulation results, while solid lines represent theoretical values corresponding to ideal frequency recovery. Such values are computed by substituting $\hat{\nu} = \nu$ into (16). This yields [2]

$$\mathrm{MSCEE}|_{\hat{\nu}=\nu} = (\mathrm{SNR})^{-1} \times \mathrm{tr}\left[\left(\boldsymbol{A}^H \boldsymbol{A}\right)^{-1}\right] \tag{50}$$

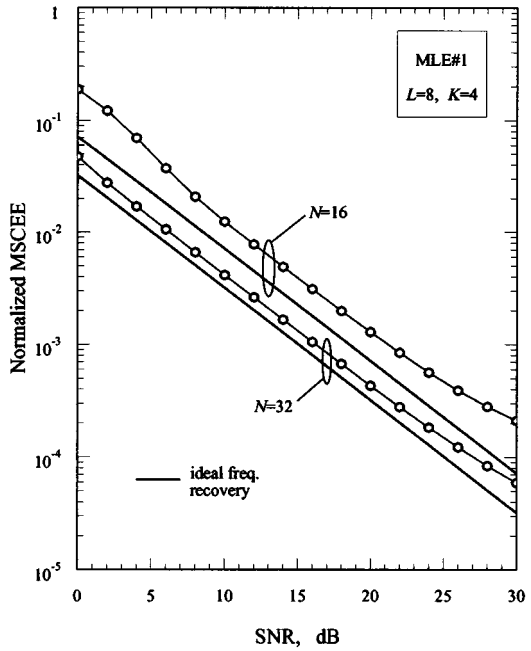


Fig. 4. Normalized MSCEE versus SNR for MLE#1.

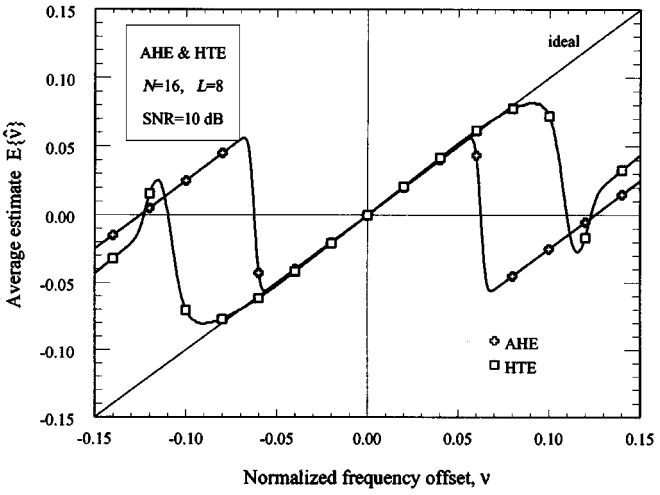


Fig. 5. Average frequency estimate versus ν with AHE and HTE.

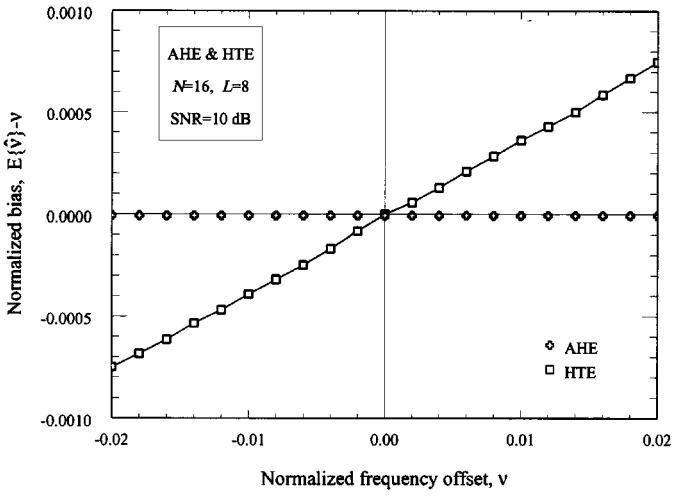


Fig. 6. Bias versus ν with AHE and HTE.

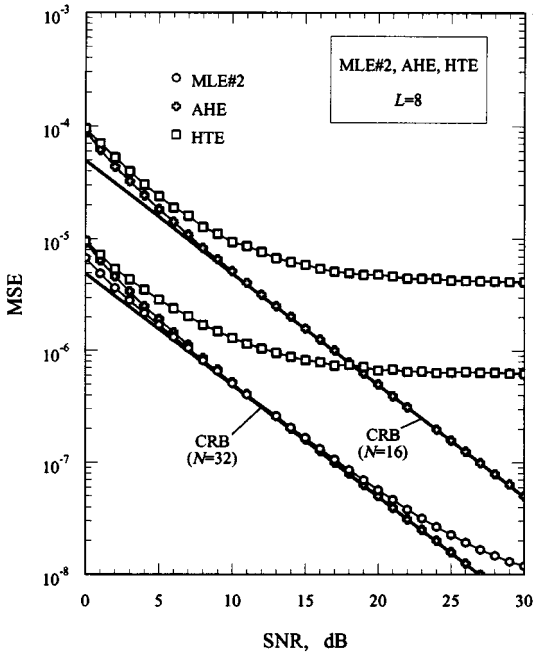


Fig. 7. Accuracy of MLE#2, AHE, and HTE.

where $\mathrm{tr}(\cdot)$ is the trace of the matrix. It is seen that the loss in accuracy due to the residual frequency error is approximately 2.5 dB with $N = 16$ and 1.2 dB with $N = 32$.

Average frequency estimates versus ν are illustrated in Fig. 5 for AHE and HTE. The TS is now periodic of period 8 and is obtained by repeating the sequence C2 twice. It appears that the AHE gives unbiased estimates for $|\nu| \leq 0.06$, which is about what we expect ($|\nu| \leq 1/2L$). At first sight, the HTE looks fine over a broader range, say $|\nu| \leq 0.08$. In reality, Fig. 6 indicates that it has a bias increasing with ν.

Fig. 7 shows the MSE versus SNR for MLE#2, AHE, and HTE. The observation length is either $N = 16$ or $N = 32$ and the TS is obtained by repeating two or four times the sequence C2. The true ν is taken randomly in the interval $(-0.05,0.05]$. The CRBs are computed from the RHS of (33). Note that the conditions expressed in 3) of Section VI hold true for $N = 16$, which means that the AHE and the MLE#2 coincide. It is seen that the AHE and MLE#2 are very accurate even at high SNRs. The HTE exhibits a floor as a consequence of the bias.

Fig. 8 shows simulation results for channel estimation. The estimates are obtained by substituting in (16) the $\hat{\nu}$ provided by either AHE or HTE. The observation length in $N = 32$ and the TS is the same as in Fig. 7. The MSCEE corresponding to perfect frequency recovery is also shown for reference. It is seen that AHE gives much better performance than HTE, particularly at high SNRs.

All the results discussed so far are concerned with linear modulation (QPSK). It is interesting to investigate whether the proposed estimators can operate even with nonlinear modulations, in particular with GMSK. The question makes sense since, as shown by Laurent [14], GMSK can be viewed as an approximate form of OQPSK. In the following simulations, the channel response is computed from (49) (for $0 \leq k \leq 7$) by modeling $g_T(t)$ as the Laurent pulse corresponding to a premodulation

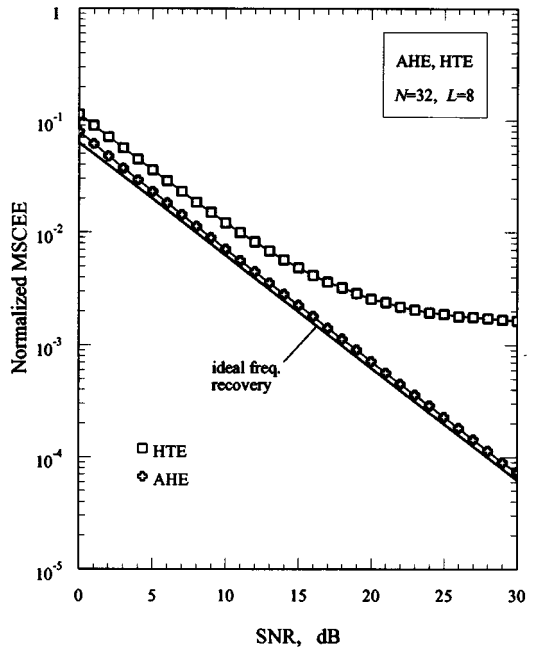


Fig. 8. Normalized MSCEE versus SNR for AHE and HTE.

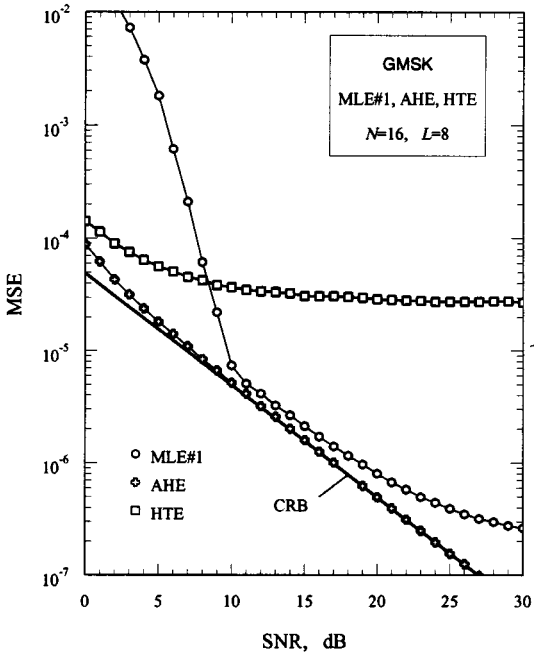


Fig. 9. Accuracy of MLE#1, AHE, and HTE with a GMSK signal.

bandwidth of $0.3/T_s$. No approximation is made in generating the signal samples in (1). Indeed, they are taken from a real GMSK waveform.

Fig. 9 illustrates the estimation accuracy with MLE#1, AHE, and HTE for $N = 16$. The TS used with MLE#1 and HTE is derived from the *midamble* of the GSM system, i.e.,

$$\{1, -j, 1, j, 1, -j, -1, -j, -1, j, -1, -j, -1, j, -1, -j\}. \quad (51)$$

With AHE the first eight symbols of (51) are repeated twice. The true ν is chosen randomly in the interval $(-0.05, 0.05]$ and

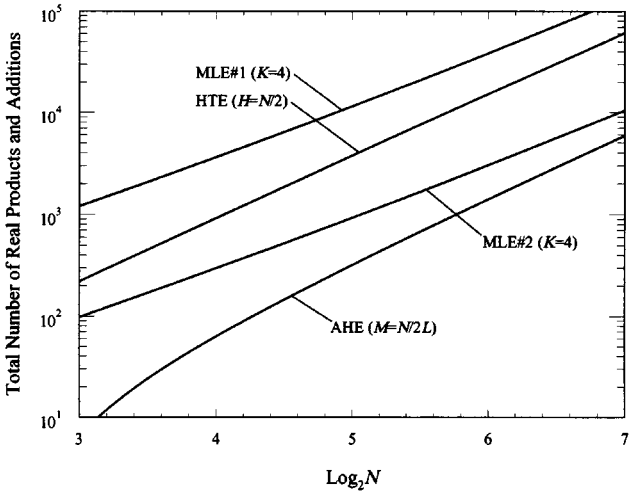


Fig. 10. Complexity of the proposed algorithms.

the CRB is computed from the RHS of (33) with $L = 8$. At first sight, it seems strange that the AHE is superior to the MLE#1. A possible explanation is that, in the present conditions, the MLE#1 is no longer ML since the signal model (2) holds only approximately. On the other hand, the arguments leading to the AHE do not require a linear modulation. Actually, they only assume that the signal samples are periodic of period L, which is true even with GMSK because of the periodicity of the TS.

Fig. 10 illustrates the total operations (additions plus multiplications) involved in the various estimators as a function of the observation length. The curves are computed from the results in Table I with $K = 4$ (for MLE#1 and MLE#2), $M = N/2L$ (for AHE), and $H = N/2$ (for HTE). It is seen that MLE#1 and AHE have the maximum and the minimum complexity, respectively. The gap between MLE#1 and HTE is overestimated since the computation of the correlations $\mathrm{E}\{h^*(l_1)h(l_2)\}$ in the HTE is not taken into account.

VIII. Conclusions

We have discussed three carrier-frequency estimation schemes for burst-mode transmissions over frequency-selective fading channels. They exploit the TS available for channel estimation. Two of them, MLE#1 and MLE#2, are realizations of the ML estimator for either arbitrary or periodic sequences. We have shown that the periodic property is useful to reduce the computational load. The third scheme, the AHE, is *ad hoc*.

The performance of these estimators has been investigated analytically and by simulation. It has been found that MLE#1 has the best performance, but it is the most complex to implement. MLE#2 and AHE are simpler but have a narrower estimation range. Comparisons have been made with the scheme proposed by Hebley and Taylor [8].

It is worth noting that all the estimators derived in this paper can be easily extended to the case of diversity reception provided that the frequency offset is the same for each diversity branch, as is assumed in [8].

The question of which scheme is better is not easily answered because of the different weights that may be given to the various performance indicators, i.e., estimation accuracy,

estimation range, computational complexity, and constraints on the TS. It is likely that the choice will depend on the specific application. For example, the AHE is the simplest and has excellent accuracy. On the other hand, it needs a periodic sequence and, as such, it cannot be used in the current GSM system. The only possible candidates for this application are the MLE#1 and the HTE.

APPENDIX A

In this appendix, we compute the average and variance of the MLE#1. To begin, we take the derivatives of $g(\nu)$ in (11)

$$\dot{g}(\nu) = j2\pi \boldsymbol{x}^H \boldsymbol{\Gamma}(\nu) \boldsymbol{D}^{(1)} \boldsymbol{\Gamma}^H(\nu) \boldsymbol{x} \tag{A1}$$

$$\ddot{g}(\nu) = 4\pi^2 \boldsymbol{x}^H \boldsymbol{\Gamma}(\nu) \boldsymbol{D}^{(2)} \boldsymbol{\Gamma}^H(\nu) \boldsymbol{x} \tag{A2}$$

with

$$\boldsymbol{D}^{(1)} = \boldsymbol{MB} - \boldsymbol{BM} \tag{A3}$$

$$\boldsymbol{D}^{(2)} = 2\boldsymbol{MBM} - \boldsymbol{BM}^2 - \boldsymbol{M}^2\boldsymbol{B}. \tag{A4}$$

In these equations, the matrix $\boldsymbol{B}$ is defined in (12) and $\boldsymbol{M}$ is given by

$$\boldsymbol{M} = \mathrm{diag}\{0, 1, 2, \ldots, N-1\}. \tag{A5}$$

Substituting (3) into (A1) and bearing in mind that $\boldsymbol{h}^H \boldsymbol{A}^H \boldsymbol{D}^{(1)} \boldsymbol{Ah} = 0$ results in

$$\dot{g}(\nu) = j2\pi \left[\boldsymbol{h}^H \boldsymbol{A}^H \boldsymbol{D}^{(1)} \tilde{\boldsymbol{w}} + \tilde{\boldsymbol{w}}^H \boldsymbol{D}^{(1)} \boldsymbol{Ah} + \tilde{\boldsymbol{w}}^H \boldsymbol{D}^{(1)} \tilde{\boldsymbol{w}} \right] \tag{A6}$$

where $\tilde{\boldsymbol{w}} = \boldsymbol{\Gamma}^H(\nu)\tilde{\boldsymbol{w}}$ is a noise vector statistically equivalent to $\boldsymbol{w}$. Clearly, the first two terms in the RHS of (A6) have zero mean. The third term has the following expectation:

$$\mathrm{E}\left\{ \tilde{\boldsymbol{w}}^H \boldsymbol{D}^{(1)} \tilde{\boldsymbol{w}} \right\} = \sum_{k=0}^{N-1} \sum_{n=0}^{N-1} (k-n) [\boldsymbol{B}]_{k,n} \mathrm{E}\left\{ \tilde{w}^*(k) \tilde{w}(n) \right\}. \tag{A7}$$

On the other hand, since $\mathrm{E}\{\tilde{\boldsymbol{w}}\tilde{\boldsymbol{w}}^H\} = \sigma_n^2 \boldsymbol{I}_N$, from (A6) and (A7), we get $\mathrm{E}\{\dot{g}(\nu)\} = 0$, and from (21), we conclude that

$$\mathrm{E}\{\hat{\nu}\} = \nu. \tag{A8}$$

Next, we concentrate on the variance of $\hat{\nu}$. From (22), it is clear that we need the expectations $\mathrm{E}\{\ddot{g}(\nu)\}$ and $\mathrm{E}\{[\dot{g}(\nu)]^2\}$. Following the same approach adopted above, it is found

$$\mathrm{E}\{\ddot{g}(\nu)\} = 4\pi^2 \boldsymbol{h}^H \boldsymbol{A}^H \boldsymbol{D}^{(2)} \boldsymbol{Ah} \tag{A9}$$

and inserting (A4) and (12) into (A9) produces

$$\mathrm{E}\{\ddot{g}(\nu)\} = 2\boldsymbol{y}^H (\boldsymbol{B} - \boldsymbol{I}_N) \boldsymbol{y} \tag{A10}$$

where $\boldsymbol{y} = 2\pi \boldsymbol{MAh}$ is an N-dimensional vector with elements $y(n) = 2\pi n \cdot s(n)$.

Now, from (A6), we have

$$\mathrm{E}\{[\dot{g}(\nu)]^2\} = -4\pi^2 \mathrm{E}\left\{ 2\boldsymbol{h}^H \boldsymbol{A}^H \boldsymbol{D}^{(1)} \tilde{\boldsymbol{w}} \tilde{\boldsymbol{w}}^H \boldsymbol{D}^{(1)} \boldsymbol{Ah} + [\tilde{\boldsymbol{w}}^H \boldsymbol{D}^{(1)} \tilde{\boldsymbol{w}}]^2 \right\}. \tag{A11}$$

Then, assuming that the SNR is sufficiently high so that the last term (noise $\times$ noise contribution) can be neglected and bearing in mind (6), it is found

$$\mathrm{E}\{[\dot{g}(\nu)]^2\} = -2\sigma_n^2 \boldsymbol{y}^H (\boldsymbol{B} - \boldsymbol{I}_N) \boldsymbol{y}. \tag{A12}$$

Finally, substituting (A10) and (A12) into (22) yields the desired result

$$\mathrm{var}\{\hat{\nu}\} = \frac{1}{2} \frac{\sigma_n^2}{\boldsymbol{y}^H (\boldsymbol{I}_N - \boldsymbol{B}) \boldsymbol{y}} \tag{A13}$$

which coincides with (24) if the SNR definition in (7) is taken into account.

APPENDIX B

In this appendix, we highlight the major steps leading to the CRB in the estimation of ν. To begin, we call $\boldsymbol{h}_R$ and $\boldsymbol{h}_I$ the real and imaginary components of the channel response $\boldsymbol{h}$ and define $\boldsymbol{\varphi} \triangleq (\boldsymbol{h}_R, \boldsymbol{h}_I, \nu)$ the set of the unknown parameters. Then, the components of the Fisher information matrix $\boldsymbol{F}$ are [9]

$$[\boldsymbol{F}]_{i,j} = -\mathrm{E}\left\{ \frac{\partial^2 \ln \Lambda(\boldsymbol{x}; \boldsymbol{\varphi})}{\partial \varphi(i) \partial \varphi(j)} \right\}, \qquad 0 \le i, j \le 2L \tag{B1}$$

where $\Lambda(\boldsymbol{x}; \boldsymbol{\varphi})$ is the probability density function of $\boldsymbol{x}$. Substituting (9) into (B1) yields

$$\boldsymbol{F} = \frac{2}{\sigma_n^2} \begin{bmatrix} \mathrm{Re}\,(\boldsymbol{A}^H\boldsymbol{A}) & -\mathrm{Im}\,(\boldsymbol{A}^H\boldsymbol{A}) & -\mathrm{Im}\,(\boldsymbol{A}^H\boldsymbol{y}) \\ \mathrm{Im}\,(\boldsymbol{A}^H\boldsymbol{A}) & \mathrm{Re}\,(\boldsymbol{A}^H\boldsymbol{A}) & \mathrm{Re}\,(\boldsymbol{A}^H\boldsymbol{y}) \\ \mathrm{Im}\,(\boldsymbol{y}^H\boldsymbol{A}) & \mathrm{Re}\,(\boldsymbol{y}^H\boldsymbol{A}) & \boldsymbol{y}^H\boldsymbol{y} \end{bmatrix} \tag{B2}$$

where $\boldsymbol{y}$ is defined in (25). Letting $\boldsymbol{F}^{-1}$ be the inverse of $\boldsymbol{F}$, the CRB for the estimation of ν is expressed as

$$\mathrm{CRB} = [\boldsymbol{F}^{-1}]_{2L,\,2L}. \tag{B3}$$

The RHS of (B3) can be computed as follows. Define the $(2L+1)$-dimensional vector

$$\boldsymbol{b} = \frac{\sigma_n^2}{2\boldsymbol{y}^H (\boldsymbol{I}_N - \boldsymbol{B}) \boldsymbol{y}} \begin{bmatrix} \mathrm{Im}(\boldsymbol{u}) & -\mathrm{Re}(\boldsymbol{u}) & 1 \end{bmatrix}^T \tag{B4}$$

where $\boldsymbol{B}$ is given in (12) and

$$\boldsymbol{u} = (\boldsymbol{A}^H\boldsymbol{A})^{-1} \boldsymbol{A}^H \boldsymbol{y}. \tag{B5}$$

Then, multiplying (B2) by $\boldsymbol{b}$ results in the vector $\boldsymbol{e}_{2L}$ with components

$$e_{2L}(i) = \begin{cases} 0, & 0 \le i \le 2L-1. \\ 1, & i = 2L \end{cases} \tag{B6}$$

This says that $\boldsymbol{b}$ coincides with the last column of $\boldsymbol{F}^{-1}$ and, in consequence, from (B3) we have

$$\mathrm{CRB} = \frac{\sigma_n^2}{2\boldsymbol{y}^H (\boldsymbol{I}_N - \boldsymbol{B}) \boldsymbol{y}}. \tag{B7}$$

APPENDIX C

In this appendix, we compute the matrix $\boldsymbol{C}_\varphi^{-1}$. From (41), we have

$$[\boldsymbol{C}_\varphi]_{m,l} = \mathrm{E}\{[\gamma_I(m) - \gamma_I(m-1)][\gamma_I(l) - \gamma_I(l-1)]\} \tag{C1}$$

or, alternately

$$[\boldsymbol{C}_\varphi]_{m,l} = \mu(m,l) + \mu(m-1,l-1) - \mu(m-1,l) - \mu(m,l-1) \tag{C2}$$

with

$$\mu(m,l) = \mathrm{E}\{\gamma_I(m)\gamma_I(l)\}. \tag{C3}$$

Next, we observe that the last term in the RHS of (38) is negligible for $\mathrm{SNR} \gg 1$ and, in consequence, we have

$$\gamma(m) \cong \frac{1}{\sigma_n^2(N-mL)} \sum_{k=mL}^{N-1} [s(k)\tilde{w}^*(k-mL) + s^*(k)\tilde{w}(k)]. \tag{C4}$$

Inserting this result into (C3) and bearing in mind (7) produces

$$\mu(m,l) = \frac{(\mathrm{SNR})^{-1}}{L(P-m)(P-l)} \cdot [P - \max(m,l) - (P-m-l)u(P-m-l)] \tag{C5}$$

where $\mathrm{u}(n)$ is the unit-step function. From (C5) and (C2), it is easily seen that $\boldsymbol{C}_\varphi$ is singular for $M > P/2$.

For $1 \le M \le P/2$, (C5) reduces to

$$\mu(m,l) = \frac{(\mathrm{SNR})^{-1}}{L(P-m)(P-l)} \min(m,l) \tag{C6}$$

and from (C2), it follows that

$$[\boldsymbol{C}_\varphi^{-1}]_{i,j} = \begin{cases} L(M-P) \times \mathrm{SNR}, & i \ne j \\ L\,[M-P+(P-i)(P-i+1)] \times \mathrm{SNR}, & i = j. \end{cases} \tag{C7}$$

REFERENCES

[1] A. Milewski, "Periodic sequences with optimal properties for channel estimation and fast start-up equalization," *IBM J. Res. Develop.*, vol. 27, pp. 426–431, 1983.

[2] S. N. Crozier, D. D. Falconer, and S. A. Mahmoud, "Least sum of squared errors (LSSE) channel estimation," *Proc. Inst. Elect. Eng.*, pt. F, vol. 138, pp. 371–378, Aug. 1991.

[3] C. Tellambura, M. G. Parker, Y. Jay Guo, S. J. Shepherd, and S. K. Barton, "Optimal sequences for channel estimation using discrete Fourier transform techniques," *IEEE Trans. Commun.*, vol. 47, pp. 230–238, Feb. 1999.

[4] H.-N. Lee and G. J. Pottie, "Fast adaptive equalization/diversity combining for time-varying dispersive channels," *IEEE Trans. Commun.*, vol. 46, pp. 1146–1162, Sept. 1998.

[5] D. C. Rife and R. R. Boorstyn, "Single-tone parameter estimation from discrete-time observations," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 591–598, Sept. 1974.

[6] M. Morelli and U. Mengali, "Feedforward frequency estimation for PSK: A tutorial review," *Eur. Trans. Telecommun.*, vol. 9, pp. 103–116, Mar./Apr. 1998.

[7] W. Y. Kuo and M. P. Fitz, "Frequency offset compensation of pilot symbol assisted modulation in frequency flat fading," *IEEE Trans. Commun.*, vol. 45, pp. 1412–1416, Nov. 1997.

[8] M. G. Hebley and D. P. Taylor, "The effect of diversity on a burst-mode carrier-frequency estimator in the frequency-selective multipath channel," *IEEE Trans. Commun.*, vol. 46, pp. 553–560, Apr. 1998.

[9] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

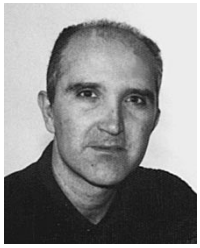
[10] J. D. Markel, "FFT Pruning," *IEEE Trans. Audio Electroacoust.*, vol. AU-19, pp. 305–311, Dec. 1971.

[11] M. H. Meyrs and L. E. Franks, "Joint carrier phase and symbol timing recovery for PAM systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 1121–1129, Aug. 1980.

[12] U. Mengali and M. Morelli, "Data-aided frequency estimation for burst digital transmission," *IEEE Trans. Commun.*, vol. 45, pp. 23–25, Jan. 1997.

[13] *Radio Transmission and Reception*, GSM Recommendation 05.05, ETSI/PT 12, V. 3.11.0, Jan. 1990.

[14] P. A. Laurent, "Exact and approximate construction of digital phase modulations by superposition of amplitude modulated pulses," *IEEE Trans. Commun.*, vol. COM-34, pp. 150–160, Feb. 1986.



Michele Morelli was born in Pisa, Italy, in 1965. He received the Laurea (*cum laude*) in electrical engineering and the "Premio di Laurea SIP" from the University of Pisa, Pisa, Italy, in 1991 and 1992, respectively. From 1992 to 1995, he was with the Department of Information Engineering of the University of Pisa, where he received the Ph.D. degree in electrical engineering.

He is currently a Research Fellow at the Centro Studi Metodi e Dispositivi per Radiotrasmissioni of the Italian National Research Council (CNR) in Pisa. His interests include digital communication theory, with emphasis on synchronization algorithms.

Umberto Mengali (M'69–SM'85–F'90) received his education in electrical engineering from the University of Pisa, Pisa, Italy. In 1971, he received the Libera Docenza in telecommunications from the Italian Education Ministry.

Since 1963, he has been with the Department of Information Engineering of the University of Pisa, Pisa, Italy, where he is a Professor of Telecommunications. In 1994, he was a Visiting Professor at the University of Canterbury, Canterbury, New Zealand, as an Erskine Fellow. His research interests include digital communications and communication theory, with emphasis on synchronization methods and modulation techniques. He has published approximately 80 journal papers and has co-authored the book *Synchronization Techniques for Digital Receivers* (New York: Plenum Press, 1997).

Prof. Mengali is a member of the Communication Theory Committee and has been an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS from 1985 to 1991. He is now Editor for Communication Theory of the *European Transactions on Telecommunications*. He is listed in *American Men and Women in Science*.