Joint Frequency Offset and Channel Estimation for OFDM Systems Using Pilot Symbols and Virtual Carriers

Tao Cui, *Student Member, IEEE*, and Chintha Tellambura, *Senior Member, IEEE*

***Abstract*— We consider joint estimation of carrier frequency offset and channel impulse response (CIR) for orthogonal frequency division multiplexing (OFDM) systems with pilot symbols and virtual subcarriers (VCs). We derive the receive-signal correlation structure due to the pilots and VCs, give the evidence of joint multivariate Gaussian distribution of the received samples, and derive an approximate maximum likelihood (ML) frequency offset estimator. We also derive the asymptotic mean-square error (MSE) and an approximate Cramér-Rao bound (CRB) and establish the asymptotic unbiasedness. Without pilots, in high signal-to-noise ratio, our estimator is equivalent to Liu and Tureli's estimator with N_v virtual carriers. When the pilot number (N_p) is greater than the channel length L, our estimator acts as a subspace-based estimator with $N_v + N_p - L$ virtual carriers. A decision-directed joint ML estimator is derived to iteratively update the estimates of frequency offset, data symbols and CIR. The optimal pilot and virtual carrier placement strategies are also discussed. The resulting decision-directed joint estimator performs within 0.8 dB of the ideal case even when the frequency offset is as large as 20%.**

***Index Terms*— Channel estimation, frequency offset, OFDM, maximum likelihood.**

I. Introduction

ORTHOGONAL frequency division multiplexing (OFDM) has been used in European digital audio broadcasting (DAB), high performance radio local area network (HIPERLAN) and 802.11a wireless LAN standards [1]. Many existing OFDM systems use pilot symbols for channel estimation. However, a carrier frequency offset destroys subcarrier orthogonality, degrading the quality of channel estimates. Although additional pilots may be transmitted specifically for frequency offset estimation, there are alternative blind techniques, which exploit the properties of OFDM signalling: cyclic prefix, constant-modulus signalling or virtual subcarriers (VCs).

Several frequency-offset estimators based on VCs have thus been developed. A blind frequency-offset estimator that exploits the VC subspace structure is developed by Liu and Tureli [2], [3], which requires the minimization of a polynomial cost function and avoids the computational overhead of typical subspace decomposition. Both the constant-modulus property and the VC subspace are exploited in the semi-blind frequency-offset estimator of [4], resulting in a generalization of Liu and Tureli's estimator. Their estimator is also extended to multiantenna OFDM (receiver diversity) [5]. In [6], VCs are exploited to derive a frequency offset estimator, which is found to be identical to Liu and Tureli's estimator [2], [3]. The estimator [2], [3], [6] is in fact a generalized likelihood ratio test (GLRT) estimator, not a maximum likelihood (ML) one [7]. Approximate ML frequency-offset estimators are proposed in [8]–[10]. The intrinsic OFDM signal structure induces a correlation structure among the received samples; the frequency offset estimator [8] exploits this correlation. In [9], the marginal likelihood function is averaged over data and channel statistics, and the resulting estimator embodies several approximations. An ML estimator is developed in [10]. A nonlinear joint ML estimator for the channel impulse response and frequency offset is derived in [11].

Although all the estimators [2], [3], [6], [8] exploit VCs, they do not exploit *pilot symbols reserved for channel estimation*. An estimator that uses both VCs and channel estimation pilot symbols[1] is expected to perform better (indeed the semi-blind estimator [4] exploits both VCs and pilots). Motivated by this fact, we derive the receive-signal correlation structure due to the pilots and VCs, give the evidence of joint multivariate Gaussian-ness of the received samples, and derive an approximate ML frequency offset estimator. The likelihood function in this case is a function of pilot and VC locations, noise variance and channel correlation. Since the resulting frequency-offset estimator requires minimizing a polynomial along the unit circle, we provide a discrete fourier transform (DFT)-based algorithm. For OFDM systems employing VCs only, our estimator reduces to that of [2] in high signal-to-noise ratio (SNR). We also derive the relevant Cramér-Rao bound (CRB) and establish the asymptotic unbiasedness. Moreover, we investigate the impact of parameter mismatch via simulation. The frequency-offset estimate can be used to compensate the pre-DFT samples, and the channel is initially estimated by standard least-squares or minimum mean squares error (MMSE) estimators. A decision-directed joint estimator is derived to enhance the estimation accuracy of channel impulse response, frequency offset and data symbols simultaneously. Optimal pilot and virtual carrier placement strategies are also discussed. The resulting decision-directed joint estimator performs within 0.8 dB of the ideal case even

[1]Note that we are not suggesting additional pilots for frequency offset estimation.

Manuscript received July 26, 2005; revised June 30, 2006; accepted October 24, 2006. The associate editor coordinating the review of this paper and approving it for publication was R. Fantacci. This work has been supported in part by the Natural Sciences and Engineering Research Council of Canada, Informatics Circle of Research Excellence and Alberta Ingenuity Fund. This paper has been presented in part at the IEEE Vehicular Technology Conference 2004-Fall, Los Angeles, CA, USA, September 2004.

T. Cui was with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada. He is now with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: taocui@caltech.edu).

C. Tellambura is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada T6G 2V4 (e-mail: chintha@ece.ualberta.ca).

Digital Object Identifier 10.1109/TWC.2007.05524.

when the frequency offset is up to 20%.

This paper is organized as follows. Section II reviews the OFDM model and conventional data detection for OFDM. Section III introduces the approximate ML frequency-offset estimator. Section IV derives the joint channel-impulse-response and frequency-offset estimator and discusses the placement strategies for pilot symbols and VCs. Simulation results and conclusions are given in Sections V and VI, respectively.

Notation: Operators $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^\dagger$ denote transpose, conjugate transpose and Moore-Penrose pseudo-inverse respectively. The trace of matrix $\mathbf{A}$ is denoted by $\mathrm{tr}(\mathbf{A})$, and $\jmath = \sqrt{-1}$. A circularly complex Gaussian random variable (CGRV) with mean μ and variance σ^2 is denoted by $z \sim \mathcal{CN}(\mu, \sigma^2)$. The $N \times N$ discrete Fourier transform (DFT) matrix is $[\mathbf{F}]_{i,j} = \frac{1}{\sqrt{N}} e^{-\jmath \frac{2\pi}{N}(i-1)(j-1)}$. The diagonal matrix $\mathbf{\Gamma}(x) = \mathrm{diag}[1, \xi, \cdots, \xi^{N-1}]$ where $\xi = \exp(\jmath 2\pi x/N)$. The cardinality of set A is $|A|$.

II. System model

In an OFDM system, source data are grouped and mapped into d_k, which are selected from a complex signal constellation $\mathcal{Q}$ with unit energy ($E\{|d_k|^2\} = 1$). Typically, N_d complex constellation points are modulated by the inverse discrete Fourier transform (IDFT) on to N parallel subcarriers, where $N_d \leq N$. The resulting N samples during the m-th frame interval (for brevity we omit m) are given by

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{\jmath(2\pi kn/N)}, \quad n = 0, 1, \cdots, N-1, \quad (1)$$

where

$$X_k = \begin{cases} d_k & k \in I_d \\ p_k & k \in I_p, \\ 0 & k \in I_v \end{cases} \quad (2)$$

I_d is the index set of data subcarriers with $|I_d| = N_d$ elements, I_p is the index set of subcarriers reserved for pilot symbols with $|I_p| = N_p$ elements, and I_v is the index set of subcarriers reserved for VCs with $|I_v| = N_v$ elements. We have $N_d + N_p + N_v = N$. We refer to x_n ($n = 0, \ldots, N-1$) as an OFDM **symbol** or **frame**. The input symbol interval and frame interval are T_s and NT_s.

The time-domain signal is transmitted over a frequency selective fading channel with frequency offset. The received signal samples in this case may be given by

$$y_n = e^{\jmath 2\pi \epsilon \frac{n}{N}} \sum_{l=0}^{L-1} h_l x_{n-l} + w_n, \qquad n = 0, \ldots, N-1, \quad (3)$$

where $w_n \sim \mathcal{CN}(0, \sigma_n^2)$ is an additive white Gaussian noise (AWGN) sample. Channel taps $h_l \sim \mathcal{CN}(0, \sigma_l^2)$, $l = 0, \ldots, L-1$ represent the sampled overall channel impulse response (which comprises the transmit/receive filters and the physical channel $h(\tau)$). The actual frequency offset normalized by the subcarrier separation $(1/(NT_s))$ is ϵ. In (3), we assume that the channel remains constant within each OFDM symbol, and that the channel taps are identically independently distributed (iid). More generally, one should consider the case of correlated channel taps; while our frequency-offset estimator can be generalized to this case, we only treat the independent case for brevity and clarity of exposition.

For convenience, the samples in (3) can be written in vector form as

$$\mathbf{y} = \mathbf{\Gamma}(\epsilon)\mathbf{F}^H \mathbf{X} \mathbf{F}_L \mathbf{h} + \mathbf{w}, \quad (4)$$

where $\mathbf{y} = [y_0, \cdots, y_{N-1}]^T$, $\mathbf{h} = [h_{I_h(0)}, \cdots, h_{I_h(L-1)}]^T$, and $\mathbf{w} = [w_0, \cdots, w_{N-1}]^T$ denote received vector, channel vector and additive noise respectively, cut where I_h is the index set of each channel path. $\mathbf{F}_L$ is the relevant $N \times L$ submatrix of $\mathbf{F}$ corresponding to $\mathbf{h}$. The diagonal matrix $\mathbf{X} = \mathrm{diag}[X_0, \cdots, X_{N-1}]$. Using Eq. (2), we write $\mathbf{X}$ as the sum of two diagonal matrices:

$$\mathbf{X} = \mathbf{X}_d + \mathbf{X}_p, \quad (5)$$

where

$$\mathbf{X}_d = \mathrm{diag}[s_1, s_2, \cdots, s_{N-1}], s_k = \begin{cases} d_k & k \in I_d \\ 0 & \text{otherwise} \end{cases}, \quad (6)$$

and

$$\mathbf{X}_p = \mathrm{diag}[t_1, t_2, \cdots, t_{N-1}], t_k = \begin{cases} p_k & k \in I_p \\ 0 & \text{otherwise} \end{cases}. \quad (7)$$

Symbol detection is possible when an estimate $\hat{\epsilon}$ of the frequency offset ϵ and an estimate $\hat{\mathbf{h}}$ of the channel impulse response $\mathbf{h}$ are available. These unknown parameters are assumed to be constant for K OFDM frames[2]. Therefore, estimates and pilot symbols may be required once every K frames, which reduces bandwidth loss (in our simulations we typically set $K = 1$). The frequency offset is compensated by pre-multiplying $\mathbf{y}$ in (4) with $\mathbf{\Gamma}(\hat{\epsilon})^H$, and DFT demodulation yields $\mathbf{Y} = [Y_0, \ldots, Y_{N-1}]^T = \mathbf{F}\mathbf{\Gamma}(\hat{\epsilon})^H \mathbf{y}$ which is given by

$$\mathbf{Y} = \mathbf{F}\mathbf{\Gamma}(v)\mathbf{F}^H \mathbf{X} \mathbf{F}_L \mathbf{h} + \mathbf{n}, \quad (8)$$

where $v = \epsilon - \hat{\epsilon}$ is the frequency-offset estimation error or residual frequency offset. The additive noise is $\mathbf{n} = \mathbf{F}\mathbf{\Gamma}(\hat{\epsilon})^H \mathbf{w}$; $\mathbf{n}$ is statistically identical to $\mathbf{w}$. If there is perfect frequency offset estimation, (8) reduces to $\mathbf{Y} = \mathbf{X}\mathbf{H} + \mathbf{n}$, where $\mathbf{H} = \mathbf{F}_L \mathbf{h} = [H_0, \ldots, H_{N-1}]^T$, and H_k, a CGRV, is the frequency response of the channel at the k-th subcarrier. Given estimated $\hat{\mathbf{H}} = [\hat{H}_0, \ldots, \hat{H}_{N-1}]^T$, the transmitted data symbols X_k can be recovered using

$$\hat{X}_k = \frac{Y_k}{\hat{H}_k}, \; k \in I_d. \quad (9)$$

This operation is customarily known as one-tap equalization.

III. Robust Carrier Frequency Offset Estimation

A. Impact of frequency offset on channel estimation

If v is the residual frequency offset between the transmitter and the receiver, the received post-DFT samples are given by [11]

$$Y_k = \frac{\sin \pi v}{N \sin \frac{\pi v}{N}} X_k H_k e^{\pi \frac{(N-1)v}{N}} + \mathrm{ICI}_k + n_k, \quad (10)$$

[2]For example, at 200km/hr, the coherence time is more than 100 times NT_s. Therefore, K may be as large as 100 [12].

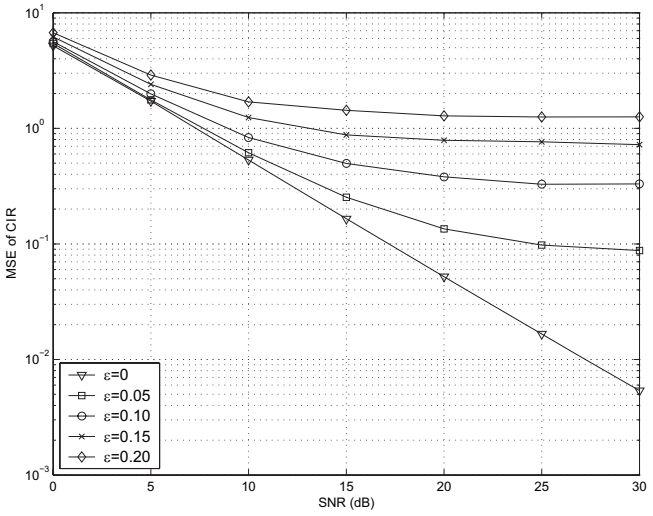

Fig. 1. MSE of channel impulse response versus SNR of estimator (12) with frequency offset =0, 0.05 0.1, 0.15, 0.2, in a QPSK OFDM system with $N = 64$, $N_p = 6$ and $N_v = 12$.

where the complex Gaussian noise $n_k \sim \mathcal{CN}(0, \sigma_n^2)$. The intercarrier interference (ICI) term is given by

$$\text{ICI}_k = \frac{1}{N} \sum_{i=0}^{N-1} \sum_{m \neq i} X_m H_m e^{j2\pi \frac{i(m-k+v)}{N}}. \tag{11}$$

The ICI term (11), which is nonzero unless $v = 0$, can severely degrade the system performance. In particular, for pilot-assisted channel estimation [13], the least-squares channel impulse response estimator becomes

$$\hat{\mathbf{h}} = \left(\mathbf{F}_L^H \mathbf{X}_p^H \mathbf{X}_p \mathbf{F}_L\right)^{-1} \mathbf{F}_L^H \mathbf{X}_p^H \mathbf{Y}, \tag{12}$$

where X_p is defined in (7). Substituting (8) into (12), we have

$$\begin{aligned} \hat{\mathbf{h}} = &\left(\mathbf{F}_L^H \mathbf{X}_p^H \mathbf{X}_p \mathbf{F}_L\right)^{-1} \mathbf{F}_L^H \mathbf{X}_p^H \mathbf{F} \mathbf{\Gamma}(v) \mathbf{F}^H \mathbf{X} \mathbf{F}_L \mathbf{h} \\ &+ \left(\mathbf{F}_L^H \mathbf{X}_p^H \mathbf{X}_p \mathbf{F}_L\right)^{-1} \mathbf{F}_L^H \mathbf{X}_p^H \mathbf{n}. \end{aligned} \tag{13}$$

If $v = 0$, the first term in (13) reduces to $\mathbf{h}$. However, when $v \neq 0$, the phase and amplitude of the channel estimate is distorted due to frequency offset. In high SNR, the channel estimate is severely degraded by the ICI term (11), which causes an error floor in the mean square error (MSE) of the channel estimate. Fig. 1 shows the MSE of the channel estimate defined as

$$\text{MSE}_h = \frac{E\{\|\mathbf{h} - \hat{\mathbf{h}}\|^2\}}{E\{\|\mathbf{h}\|^2\}}, \tag{14}$$

where a 6-ray TU channel is considered (see Section V). At a 5% frequency offset, the MSE performance degrades by 7.5 dB at MSE= 0.1. As the frequency offset increases beyond 5%, the MSE too increases notably, resulting in an error floor.

B. Approximate Maximum likelihood frequency offset estimation

We now use the ML principle to derive a novel frequency-offset estimator. When the number of subcarriers N is large, transmit samples x_n ($n = 0, \ldots, N-1$) in (1) can be modelled as complex Gaussian via the central limit theorem (CLT) [14]. When the number of paths is large, the received OFDM signal $\mathbf{y}$ may also be modelled as a multivariate complex Gaussian with zero-mean and correlation matrix (19). Note that this is a much stronger claim than the standard univariate distribution claim via the CLT. We justify this assumption in Appendix A. A similar approach has previously been used in [8] (without a rigorous justification) for an OFDM systems without pilots.

Consequently, we derive the autocorrelation of the received signal as a function of frequency offset, pilots, VCs and channel correlation. It can be readily verified that $E\{\mathbf{y}\} = 0$. The autocorrelation matrix of the received signal is given by

$$\begin{aligned} \mathbf{R}_y =& E\{\mathbf{y}\mathbf{y}^H\} = E\{\mathbf{\Gamma}(\epsilon)\mathbf{F}^H \mathbf{X} \mathbf{F}_L \mathbf{h}\mathbf{h}^H \mathbf{F}_L^H \mathbf{X}^H \mathbf{F} \mathbf{\Gamma}^H(\epsilon)\} + \sigma_n^2 \mathbf{I} \\ =& E\{\mathbf{\Gamma}(\epsilon)\mathbf{F}^H (\mathbf{X}_d + \mathbf{X}_p) \mathbf{F}_L \mathbf{R}_h \mathbf{F}_L^H (\mathbf{X}_d + \mathbf{X}_p)^H \mathbf{F} \mathbf{\Gamma}^H(\epsilon)\} + \sigma_n^2 \mathbf{I}, \end{aligned} \tag{15}$$

where the expectation is taken over both $\mathbf{h}$ and $\mathbf{X}$, and $\mathbf{R}_h = E\{\mathbf{h}\mathbf{h}^H\}$ is the autocorrelation matrix of $\mathbf{h}$. The third equality follows because $\mathbf{h}$ and $\mathbf{X}$ are statistically independent. Specifically, if tap weights $h(l)$ are statistically independent[3], $\mathbf{R}_h$ becomes

$$\mathbf{R}_h = \text{diag}[\sigma_0^2, \sigma_1^2, \ldots, \sigma_{L-1}^2]. \tag{16}$$

We can show that

$$\begin{aligned} &E\{\mathbf{X}_d\} = 0, \; E\{\mathbf{X}_d \mathbf{F}_L \mathbf{R}_h \mathbf{F}_L^H \mathbf{X}_d^H\} = rE\{\mathbf{X}_d \mathbf{X}_d^H\} = r\mathbf{D}, \\ &E\{\mathbf{X}_d \mathbf{F}_L \mathbf{R}_h \mathbf{F}_L^H \mathbf{X}_p^H\} = 0, \quad E\{\mathbf{X}_p \mathbf{F}_L \mathbf{R}_h \mathbf{F}_L^H \mathbf{X}_d^H\} = 0, \end{aligned} \tag{17}$$

where $r = \sum_{l=0}^{L-1} \sigma_l^2 / N$, and

$$\mathbf{D} = \text{diag}[d_0, d_2, \cdots, d_{N-1}], d_k = \begin{cases} 1 & k \notin I_p, I_v \\ 0 & \text{otherwise} \end{cases}. \tag{18}$$

Hence, we have

$$\begin{aligned} \mathbf{R}_y =& E\{\mathbf{\Gamma}(\epsilon)\mathbf{F}^H (\mathbf{X}_d \mathbf{F}_L \mathbf{R}_h \mathbf{F}_L^H \mathbf{X}_d^H + \mathbf{X}_p \mathbf{F}_L \mathbf{R}_h \mathbf{F}_L^H \mathbf{X}_p^H) \\ &\times \mathbf{F} \mathbf{\Gamma}^H(\epsilon)\} + \sigma_n^2 \mathbf{I} \\ =& \mathbf{\Gamma}(\epsilon) \left[\mathbf{F}^H (r\mathbf{D} + \mathbf{X}_p \mathbf{F}_L \mathbf{R}_h \mathbf{F}_L^H \mathbf{X}_p^H) \mathbf{F} + \sigma_n^2 \mathbf{I}\right] \mathbf{\Gamma}^H(\epsilon) \\ =& \mathbf{\Gamma}(\epsilon) \mathbf{G} \mathbf{\Gamma}^H(\epsilon), \end{aligned} \tag{19}$$

where $\mathbf{G} = \mathbf{F}^H (r\mathbf{D} + \mathbf{X}_p \mathbf{F}_L \mathbf{R}_h \mathbf{F}_L^H \mathbf{X}_p^H) \mathbf{F} + \sigma_n^2 \mathbf{I}$ and is Hermitian. The probability density function (pdf) of $\mathbf{y}$ is therefore

$$p(\mathbf{y}|\epsilon) = (\pi^N \det(\mathbf{R}_y))^{-1} \exp(-\mathbf{y}^H \mathbf{R}_y^{-1} \mathbf{y}). \tag{20}$$

$\det(\mathbf{R}_y) = \det(\mathbf{\Gamma}(\epsilon)\mathbf{G}\mathbf{\Gamma}^H(\epsilon)) = \det(\mathbf{G}\mathbf{\Gamma}^H(\epsilon))\mathbf{\Gamma}(\epsilon) = \det(\mathbf{G})$. Therefore, the determinant of $\mathbf{R}_y$ is independent of ϵ. We drop the terms in (20) that are independent of ϵ and derive the log-likelihood function as

$$\begin{aligned} \Lambda(\mathbf{y}|\epsilon) &= -\mathbf{y}^H \mathbf{R}_y^{-1} \mathbf{y} = -\mathbf{y}^H \left[\mathbf{\Gamma}(\epsilon)\mathbf{G}\mathbf{\Gamma}^H(\epsilon)\right]^{-1} \mathbf{y} \\ &= -\mathbf{y}^H \mathbf{\Gamma}(\epsilon) \mathbf{G}^{-1} \mathbf{\Gamma}^H(\epsilon) \mathbf{y} = -\gamma(\epsilon)^T \mathbf{y}_D^* \mathbf{G}^{-1} \mathbf{y}_D^T \gamma(\epsilon)^*, \end{aligned} \tag{21}$$

where $\gamma(\epsilon) = [1, \xi, \cdots, \xi^{N-1}]^T$, and $\mathbf{y}_D = \text{diag}[y_0, y_1, \ldots, y_{N-1}]$.

[3]Note that our estimator can also be used for correlated $h(l)$.

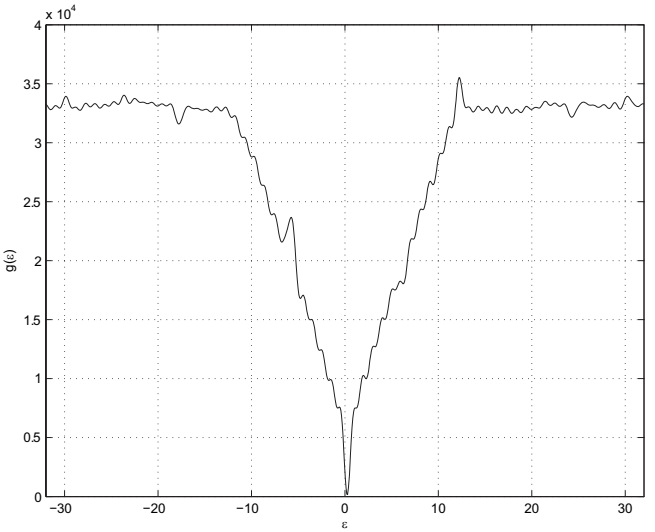

Fig. 2. The curve of function $g(\epsilon)$ averaged over 100 OFDM frames in a QPSK system with $N_p = 8$ and $N_v = 12$, the carrier frequency offset normalized by subcarrier spacing.

Maximizing the log likelihood function is equivalent to

$$\hat{\epsilon} = \arg\min_{\epsilon} g(\epsilon), \tag{22}$$

where $g(\epsilon) = \mathbf{y}^H \mathbf{\Gamma}(\epsilon)\mathbf{G}^{-1}\mathbf{\Gamma}^H(\epsilon)\mathbf{y}$.

Fig. 2 shows $g(\epsilon)$ averaged over 100 OFDM frames in a QPSK system with $N_p = 8$ and $N_v = 12$. The curve is similar to the characteristic function shown in [3, Fig. 3].

If the delay is tolerable, the received signals during the K frames can be combined for frequency offset estimation. Let $\mathbf{y}_k$ denote the received signal in the kth frame. Since the $\mathbf{y}_k$'s for $k = 1, 2, \ldots, K$ are independent, the probability density function of $(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_K)$ is

$$\begin{aligned} p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_K|\epsilon) &= \prod_{k=1}^{K} p(\mathbf{y}_k|\epsilon) \\ &= \frac{1}{\pi^{KN}\prod_{k=1}^{K}\det(\mathbf{R}_{y_k})} \exp\left(-\sum_{k=1}^{K} \mathbf{y}_k^H \mathbf{\Gamma}(\epsilon)\mathbf{G}^{-1}\mathbf{\Gamma}^H(\epsilon)\mathbf{y}_k\right). \end{aligned} \tag{23}$$

The cost function (22) is therefore

$$\begin{aligned} \hat{\epsilon} &= \arg\min_{\epsilon} \sum_{k=1}^{K} \mathbf{y}_k^H \mathbf{\Gamma}(\epsilon)\mathbf{G}^{-1}\mathbf{\Gamma}^H(\epsilon)\mathbf{y}_k \\ &= \arg\min_{\epsilon} \sum_{k=1}^{K} \gamma(\epsilon)^T (\mathbf{y}_k)_D^* \mathbf{G}^{-1} (\mathbf{y}_k)_D^T \gamma(\epsilon)^* \\ &= \arg\min_{\epsilon} \gamma(\epsilon)^T \mathbf{B}\gamma(\epsilon)^*, \end{aligned} \tag{24}$$

where and $\mathbf{B} = \sum_{k=1}^{K} (\mathbf{y}_i)_D^* \mathbf{G}^{-1} (\mathbf{y}_i)_D^T$.

Remarks:

1) Due to the Gaussian approximation, our estimator is not exactly ML. However, when the number of channel taps L increases, the estimator approaches the exact ML estimator, which is given in [9], [10]. In [9], [10], the marginal likelihood function conditioned on the channel and data symbols is averaged over the distribution of $\mathbf{h}$ and $\mathbf{X}$. Our approach in this paper carries out the average implicitly.
2) From (22) and (24), $\hat{\epsilon}$ can be calculated and subsequently used for estimating $\mathbf{h}$. This estimate $\hat{\epsilon}$ is used to initialize the joint estimator in Section IV.
3) The inverse matrix $\mathbf{G}^{-1}$ in (22) and (24) can be precomputed. The main computational cost is to find $\hat{\epsilon}$ in (22) and (24). Eq. (24) can be written as

$$g(\epsilon) = \sum_{i=0}^{N-1}\sum_{k=0}^{N-1} b_{i,k} e^{\jmath 2\pi(i-k)\epsilon/N}, \tag{25}$$

which can viewed as a polynomial, where $b_{i,k}$ is the (i,k)th entry of $\mathbf{B}$. Many frequency-offset estimators [2], [3], [5] can be simplified to (25). In the range $[-N/2, N/2]$, the cost function has many local minima, which may be obtained by solving for the roots of the polynomial. Alternatively, from [15], the cost function can be minimized in two steps. The first step is coarse search; the cost function $g(\epsilon)$ is computed over a grid of $\hat{\epsilon}$ values, say $\{\hat{\epsilon}_n\}$. $\hat{\epsilon}_n^*$ with the minimum cost $g(\hat{\epsilon}_n^*)$ is selected. Since, in a small neighborhood around $\hat{\epsilon}_n^*$, $g(\epsilon)$ is convex, the traditional Newton-Raphson search or bisection search can be applied. This approach can also be used for single carrier transmission over frequency-selective channels [16].

Since $\mathbf{G}$ is Hermitian, $g(\epsilon)$ can be written as

$$g(\epsilon) = 2\mathrm{Re}\left\{\sum_{i=0}^{N-1} a_i z^i\right\}, \tag{26}$$

where $z = \exp(-\jmath 2\pi\epsilon/N)$, and

$$a_i = \begin{cases} \sum_{j=0}^{N-1-k} b_{j,j+i} & i \neq 0 \\ \frac{1}{2}\sum_{j=0}^{N-1} b_{j,j} & i = 0 \end{cases}. \tag{27}$$

We add $(m-1)N$ zeros to the end of sequence $[a_0, a_1, \ldots, a_{N-1}]$ and perform the mN-point DFT, which yields

$$A(k) = \frac{1}{mN}\sum_{i=0}^{N-1} a_i e^{-\jmath\frac{2\pi i}{N}\frac{k}{m}}, \quad k = 0, 1, \ldots, mN-1. \tag{28}$$

Let the index of the minimum $\mathrm{Re}\{A(k)\}$ denote $\tilde{k}$. Hence, the ϵ that minimizes (25) can be approximated as $\hat{\epsilon} = \tilde{k}/m$. To further improve the estimation accuracy, second order interpolation can be used around $\tilde{k}$. Details are omitted for brevity.

Compared with the algorithm in [15], the complexity of our frequency-offset estimator can be reduced by properly choosing m. The larger the m, the better the estimation but the higher the complexity. The complexity of the FFT is $5\beta mN\log_2(mN)]$ in Flops, where $\beta = 1 - [\log_2 m + 2(1/m - 1)]/\log_2(mN)$ is due to computational saving by skipping the operations on the zeros in the FFT. The complexity of search for the minima is mN. Hence the total complexity of this algorithm is roughly $mN(5\beta\log_2(mN) + 1)$. The complexity that of the initial estimator in [11] is $6NL + 4N - 2 + 5N\log_2 N$. Hence with appropriate

m, our estimator's complexity may be less than that of [11]. Moreover, the estimator in [4] can also use our fast frequency offset estimator.

4) Since our frequency-offset estimator also makes use of the pilot symbols and VCs, which will introduce a special correlation structure into the pre-DFT samples, knowledge of the constant pilot patterns is thus exploited. However, since channel correlation and channel noise variance are not known perfectly in practice, mismatch conditions arise. We later investigate the robustness of our estimator against the parameter mismatch.
5) When no pilot symbols exist but VCs exist, $\mathbf{G}$ in (19) can be rewritten as

$$\mathbf{G} = \mathbf{F}^H(r\mathbf{D} + \sigma_n^2\mathbf{I})\mathbf{F}. \tag{29}$$

Correspondingly, $g(\epsilon)$ can be modified as

$$g(\epsilon) = \mathbf{y}^H\mathbf{\Gamma}(\epsilon)\mathbf{F}^H\mathbf{B}\mathbf{F}\mathbf{\Gamma}^H(\epsilon)\mathbf{y}, \tag{30}$$

where $\mathbf{B} = (r\mathbf{D}+\sigma_n^2\mathbf{I})^{-1} = \mathrm{diag}[b_0, b_1, \ldots, b_{N-1}]$ with

$$b_k = \begin{cases} \frac{1}{r+\sigma_n^2} & k \notin I_v \\ \frac{1}{\sigma_n^2} & k \in I_v \end{cases}. \tag{31}$$

In high SNR, $1/(r+\sigma_n^2) \ll 1/\sigma_n^2$. Hence, we may let $b_k = 0$, $k \notin I_v$. Let $\mathbf{w}_k$ denote the kth column of the IDFT matrix $\mathbf{F}^H$. Eq. (30) then becomes

$$\begin{aligned} g(\epsilon) =& \mathbf{y}^H\mathbf{\Gamma}(\epsilon)[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_N]\mathbf{B}[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_N]^H\mathbf{\Gamma}^H(\epsilon)\mathbf{y} \\ =& \frac{1}{\sigma_n^2}\sum_{k\in I_v}\mathbf{y}^H\mathbf{\Gamma}(\epsilon)\mathbf{w}_k\mathbf{w}_k^H\mathbf{\Gamma}^H(\epsilon)\mathbf{y}. \end{aligned} \tag{32}$$

Therefore, frequency offset can be estimated by

$$\begin{aligned} \hat{\epsilon} =& \arg\min_{\epsilon}\sum_{k\in I_v}\mathbf{y}^H\mathbf{\Gamma}(\epsilon)\mathbf{w}_k\mathbf{w}_k^H\mathbf{\Gamma}^H(\epsilon)\mathbf{y} \\ =& \arg\min_{\epsilon}\sum_{k\in I_v}\|\mathbf{y}^H\mathbf{\Gamma}(\epsilon)\mathbf{w}_k\|^2 \\ =& \arg\min_{\epsilon}\sum_{k\in I_v}\mathbf{w}_k^H\mathbf{\Gamma}(\epsilon)\mathbf{y}\mathbf{y}^H\mathbf{\Gamma}^H(\epsilon)\mathbf{w}_k, \end{aligned} \tag{33}$$

which is the same as the cost function given in [2]. Therefore, when no pilots exist, our proposed frequency-offset estimator is equivalent to that in [2] in high SNR. The frequency-offset estimator in [2] is in fact asymptotically optimal as opposed to the optimal claim using a GLRT detection in [6].

6) When pilots exist, $\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H$ is a circulant matrix of rank L, and any L columns or rows are independent. If $N_p \le L$, $\mathbf{X}_p\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H\mathbf{X}_p^H$ has N_p nonzero columns and rows, the N_p columns are independent. Since $r\mathbf{D}$ is a diagonal matrix, $\mathbf{A} = r\mathbf{D}+\mathbf{X}_p\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H\mathbf{X}_p^H$ has rank N_d+N_p. Let the singular value decomposition (SVD) of $\mathbf{A}$ be denoted as $\mathbf{A} = \mathbf{U}^H\mathbf{\Lambda}\mathbf{U}$, where the first N_d+N_p diagonal entries of $\mathbf{\Lambda}$ are nonzero. We rewrite $g(\epsilon)$ as

$$g(\epsilon) = \mathbf{y}^H\mathbf{\Gamma}(\epsilon)\mathbf{V}^H\mathbf{C}\mathbf{V}\mathbf{\Gamma}^H(\epsilon)\mathbf{y}, \tag{34}$$

where $\mathbf{V} = \mathbf{U}\mathbf{F}$ and $\mathbf{C} = (\mathbf{\Lambda} + \sigma_n^2\mathbf{I})^{-1} = \mathrm{diag}\{c_0, c_1, \ldots, c_{N-1}\}$ with

$$c_k = \begin{cases} \frac{1}{\lambda_{k,k}+\sigma_n^2} & k = 0, \ldots, N_d+N_p-1 \\ \frac{1}{\sigma_n^2} & \text{otherwise} \end{cases}. \tag{35}$$

Similar to (32), in high SNR $1/(\lambda_{k,k}+\sigma_n^2) \ll 1/\sigma_n^2$. Let $\mathbf{v}_k$ denote the kth column of matrix $\mathbf{V}^H$. Eq. (34) becomes

$$g(\epsilon) = \frac{1}{\sigma_n^2}\sum_{k=N_d+N_p}^{N-1}\mathbf{y}^H\mathbf{\Gamma}(\epsilon)\mathbf{v}_k\mathbf{v}_k^H\mathbf{\Gamma}^H(\epsilon)\mathbf{y}. \tag{36}$$

Since there are $N - N_d - N_p = N_v$ terms in (36), it acts as a subspace-based frequency-offset estimator with N_v virtual carriers. Therefore, when $N_p \le L$, the pilots cannot improve the performance of the frequency-offset estimator in high SNR. However, improvement is possible in low SNR. When $N_p > L$, following the same approach as in (34)-(36), the frequency-offset estimator reduces to the estimator in [2] with $N_v + N_p - L$ virtual carriers. The pilots determine $\lambda_{k,k}$, $\mathbf{v}_k$ and the performance of the estimator; thus, the structure in $\mathbf{y}$ and $\mathbf{v}_k$ due to pilot symbols improves performance. Pilots enable channel estimation, provided that $N_p \ge L$. Therefore, the frequency-offset estimate is improved due to the use of pilots available for channel estimation.

7) From (22) and (24), the cost function $g(\epsilon)$ is periodic with a period of N, which means that the range of frequency offset is wider and not limited to half of the frequency separation between adjacent subcarriers, $|\epsilon| < 0.5$. Our estimator does not divide the frequency offset into an integer part and a fractional part, and it performs coarse acquisition and fine acquisition separately, which can reduce the system complexity.
8) Receiver diversity can improve the performance of frequency offset estimation [5]. The extension of our frequency-offset estimator to receiver with diversity is straightforward by following (15)-(22). The resulting estimator may be considered as an ML extension of the GLRT estimator in [5].

C. Performance Quality Measures

We now assess the performance of our proposed frequency-offset estimator. Reference [17] shows that the expectation and MSE of the ML estimate in high SNR are approximated as

$$E\{\hat{\epsilon}\} \doteq \epsilon - \frac{E\{\dot{g}(\epsilon)\}}{E\{\ddot{g}(\epsilon)\}}, \quad \mathrm{MSE}_\epsilon = E\{\|\epsilon-\hat{\epsilon}\|^2\} \doteq \frac{E\{[\dot{g}(\epsilon)]^2\}}{[E\{\ddot{g}(\epsilon)\}]^2}, \tag{37}$$

where $g(\epsilon)$ denotes the cost function of the estimator given by (22) and (24), and $\dot{g}(\epsilon)$ and $\ddot{g}(\epsilon)$ are the first and second derivatives of $g(\epsilon)$. We prove in Appendix B that

$$E\{\hat{\epsilon}\} \doteq \epsilon, \quad \mathrm{MSE}_\epsilon = \frac{N^2}{4\pi^2}\frac{1}{\mathrm{tr}\{\mathbf{M}\mathbf{G}^{-1}\mathbf{M}\mathbf{G}-\mathbf{M}^2\}}. \tag{38}$$

Therefore, our proposed frequency-offset estimator is asymptotically unbiased.

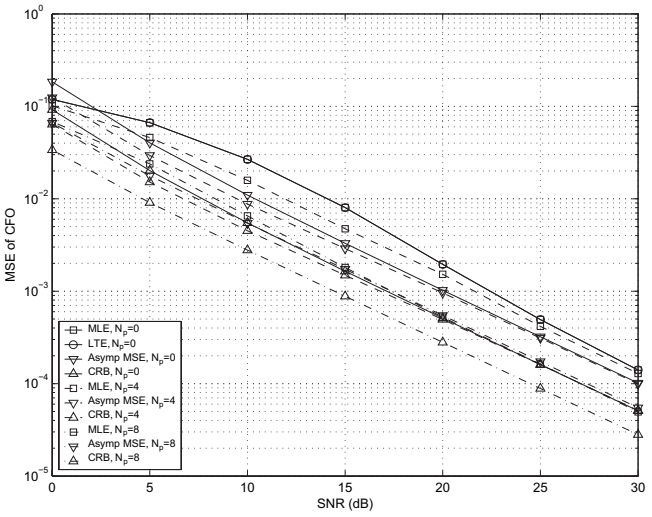

Fig. 3. Comparison of MSE of frequency offset estimation for different N_p in a QPSK OFDM system with $\epsilon = 0.25$, $N = 64$ and $N_v = 12$.

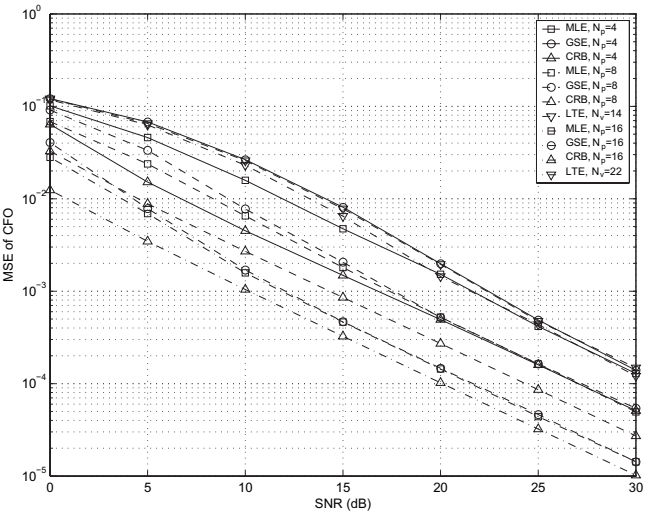

Fig. 4. MSE of frequency offset versus SNR for $N_p = 4, 8, 16$ in a QPSK OFDM system with $\epsilon = 0.25$, $N = 64$ and $N_v = 12$.

In Appendix B, we also derive the CRB [18] for the estimation of ϵ using a single OFDM frame. The CRB is given by

$$\mathrm{CRB} = \frac{N^2}{8\pi^2} \frac{1}{\mathrm{tr}\{\mathbf{M}\mathbf{G}^{-1}\mathbf{M}\mathbf{G} - \mathbf{M}^2\}}. \tag{39}$$

Moreover, assuming independent OFDM frames, the CRB for (24) using K OFDM frames can be readily obtained as $1/K$ times of (39). Similar results for the CRB have also been given in [4], [5]. However, as commented in our frequency-offset estimator, the received signal is only approximately Gaussian. Therefore, the CRB (39) derived based on the pdf (20) is also an approximation of the true CRB. The MSE (38) and the CRB (39) differ by a factor of two, which violates the asymptotic efficiency property of the ML estimator [19]. Our estimator, together with other GLRT based frequency-offset estimators, is thus not optimal. The ML frequency-offset estimator is given in [10]. However, by increasing N_p, the received signals become "more Gaussian", and hence, our frequency-offset estimator performs closer to the CRB (see simulation results).

D. Comparative Performance

For quadrature phase shift keying (QPSK), and for the six ray TU channel (Section V), Fig. 3 compares the CRB and the MSE of our estimator (denoted as "MLE") with those of Liu and Tureli's estimator [2], denoted as "LTE". The total number of subcarriers and VCs are $N = 64$ and $N_v = 12$, respectively. Without pilots, our estimator performs similar to Liu and Tureli's estimator in all SNRs (Remark #5, Page 1197). At an MSE of 2×10^{-4}, both estimators perform 1.6 dB and 4.6 dB off from the asymptotic MSE and CRB, respectively. The true MSE approaches the asymptotic MSE bound in high SNR. For 4 pilot symbols ($N_p = 4$), the true MSE, asymptotic MSE and CRB all approach those of [2] without pilots in high SNR (Remark #6, Page 1197). However, when N_p increases to 8, our estimator has a 4.5-dB gain over Liu and Tureli's estimator at an MSE of 2×10^{-4}. The gap between the true MSE and the CRB reduces to 2.5 dB, and the performance approaches the asymptotic MSE in high SNR. The integration of pilot symbols and VCs thus clearly improves the performance.

The performance of our frequency-offset estimator in relation to those of Liu and Tureli's estimator and Ghogho and Swami's estimator [4] as a function of the number of pilot symbols and SNR is of interest. Fig. 4 plots the MSE as a function of SNR for different N_p. The performance of Liu and Tureli's estimtor with the same subspace rank and the performance of the semiblind estimator ("GSE") in [4] are also shown. For increasing N_p, the MSE is greatly reduced for both our estimator and Ghogho and Swami's estimator; the former performs better than the latter in low SNR, but both perform identically in high SNR. At an MSE of 10^{-4}, the performance of our estimator increases by 5.4 dB when N_p increases from 8 to 16. With the increase of N_p, the gap between MLE and CRB decreases from 4.6 dB to 1.6 dB at an MSE of 2×10^{-4}. For $N_p = 8$, the equivalent Liu and Tureli estimator has $N_v + N_p - L = 14$ VCs. But it performs only marginally better when N_v increases from 12 to 14. Our estimator with $N_p = 8$ performs much better than Liu and Tureli's with $N_v = 12$. Similar observations also hold for $N_p = 16$. Increasing N_p improves the performance of our estimator, but increasing N_v alone does not.

Fig. 5 shows the effect of the number of blocks on the MSE with $N_p = 6$ and $N_v = 12$. The use of $K = 8$ blocks yields a 8-dB gain over the use of $K = 2$ blocks and a 4-dB gain over the use of $K = 4$ blocks at an MSE of 10^{-4}. This agrees with that the CRB (39) decreases inversely with the increase of K. The gap between the MSE and the CRB also decreases with the increase of K. The gap is 2 dB when $K = 2$ and 1 dB when $K = 4$ but 0.5 dB when $K = 8$. However, a regime of diminishing returns sets in when the number of blocks gets large ($K > 10$). The complexity and delay also increase with K. Fig. 6 shows the S-curves [20] of the MLE and GSE for an SNR of 10 dB as a function of the actual CFO (ϵ).

E. Parameter Mismatch

Our proposed estimator requires the knowledge of the noise power, power delay profile (PDP), and pilot and VC locations.

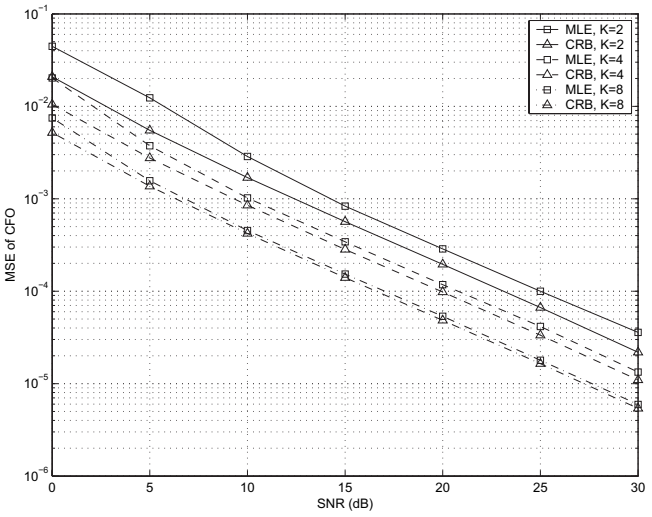

Fig. 5. MSE of frequency offset versus SNR with $K = 2, 4, 8$ in a QPSK OFDM system with $\epsilon = 0.25$, $N = 64$, $N_p = 6$ and $N_v = 12$.

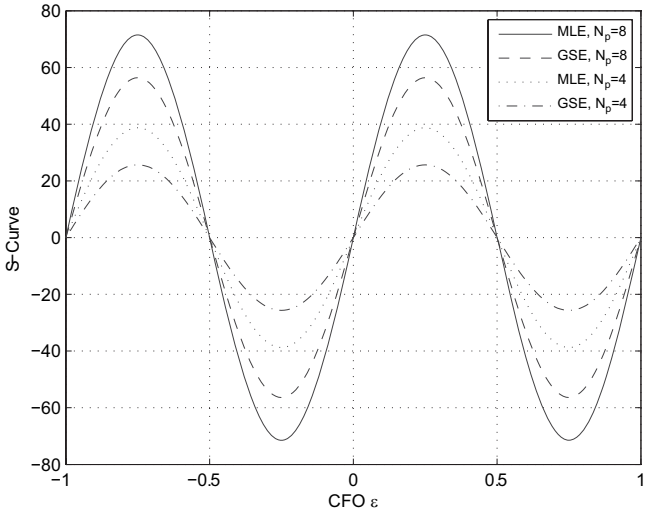

Fig. 6. S-curve for a QPSK OFDM system with $N = 64$, $N_v = 12$ and SNR$=$ 10 dB.

Since pilots and VCs are predefined, they are completely known to the receiver. However, the noise power and PDP must be estimated. The channel correlation matrix and the noise variance may be estimated as $\hat{\mathbf{R}}_h = \sum_{k=1}^{K} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H / K$ and $\sigma_n^2 = \sum_{k=1}^{K} \|\mathbf{Y}_k - \hat{\mathbf{X}}_k \mathbf{F}_L \hat{\mathbf{h}}_k\|^2 / (KN)$, where $\hat{\mathbf{h}}_k$ and $\hat{\mathbf{X}}_k$ are the estimated channel impulse response and data symbols in the k-th frame respectively.

We now test the robustness of our estimator against parameter mismatch for the six ray TU channel (Section V). We randomly generate 50 instances of PDP (all with normalized unity power). In each PDP realization, 1000 simulations are performed. The receiver, however, does not know the exact PDP or the noise variance, so in the estimator (22) we set $\mathbf{R}_h$ according to a uniform PDP, and σ_n^2 is selected for an SNR of 30 dB. This choice can be intuitively explained by the fact that a frequency-offset error is concealed in AWGN at low SNR, whereas it tends to dominate AWGN at high SNR where the noise is low. Hence, it is important to keep the frequency offset error low at the high SNR. The system uses

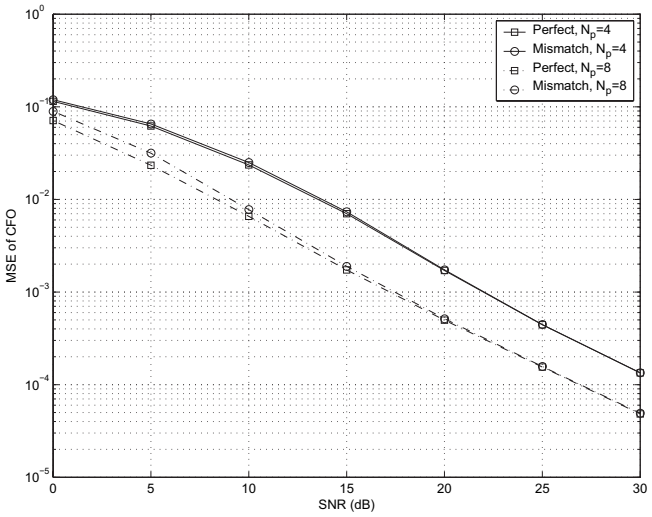

Fig. 7. MSE of frequency offset under parameters mismatch in a QPSK OFDM system with $\epsilon = 0.25$, $N = 64$ and $N_v = 12$. The PDP is randomly generated 50 times with normalized unity power. In each PDP realization, 1000 simulations are performed. $\mathbf{R}_h$ is chosen for a uniform PDP and σ_n^2 is selected for an SNR of 30 dB.

$N = 64$ subcarriers and $N_v = 12$ VCs [1]. The performance degradation due to the mismatch is negligible, especially in high SNR (Fig. 7). This would confirm the robustness of the frequency-offset estimator.

IV. Joint ML Estimation Algorithm

A. Iterative joint ML estimation

Given the estimate $\hat{\epsilon}$, the frequency offset can be compensated by using (8). The channel impulse response $\hat{\mathbf{h}}$ can be obtained using the least-squares estimator (13). Using the channel estimate and frequency offset estimate as the initial values, we derive a decision-directed joint ML estimator (DD-JMLE), following a similar idea from [11]. The joint ML estimator for $[\mathbf{h}, \epsilon, \mathbf{X}_d]$ is given by

$$[\hat{\mathbf{h}}, \hat{\epsilon}, \hat{\mathbf{X}}_d] = \arg\min_{\mathbf{h}, \epsilon, \mathbf{X}_d} \left\| \mathbf{y} - \mathbf{\Gamma}(\epsilon) \mathbf{F}^H \left(\mathbf{X}_d + \mathbf{X}_p \right) \mathbf{F}_L \mathbf{h} \right\|^2. \tag{40}$$

The DD-JMLE is iterative and uses feedback. For example, in the i-th iteration, the estimates are denoted by $\hat{\mathbf{h}}^i$, $\hat{\epsilon}^i$ and $\hat{\mathbf{X}}_d^i$, respectively, and $\mathbf{h}^0$ and ϵ^0 are the initial estimates. Fixing $\hat{\epsilon}^i$ and $\hat{\mathbf{X}}_d^i$, we get the least-squares channel estimate as

$$\hat{\mathbf{h}}^i = \left[(\mathbf{X}^{i-1} \mathbf{F}_L)^H \mathbf{X}^{i-1} \mathbf{F}_L \right]^{-1} (\mathbf{X}^{i-1} \mathbf{F}_L)^H \mathbf{F} \mathbf{\Gamma}^H (\hat{\epsilon}^{i-1}) \mathbf{y}, \tag{41}$$

where $\mathbf{X}^{i-1} = \mathbf{X}_d^{i-1} + \mathbf{X}_P$. Data symbols are detected by simply using division and hard decisions as

$$\hat{\mathbf{X}}_d^i = \operatorname{diag} \left\{ \mathbf{F} \mathbf{\Gamma}^H (\hat{\epsilon}^{i-1}) \mathbf{y} ./ \mathbf{F}_L \hat{\mathbf{h}}^i \right\}, \tag{42}$$

where $./$ denotes component-wise division of two vectors. Finally, ϵ^i can be obtained as

$$\hat{\epsilon}^i = \arg\min_{\epsilon} \left\| \mathbf{y} - \mathbf{\Gamma}(\epsilon) \mathbf{F}^H \mathbf{X}^i \mathbf{F}_L \hat{\mathbf{h}}^i \right\|^2. \tag{43}$$

This iterative procedure is repeated until convergence is achieved. Simulation results attest that few iterations are

often sufficient. Note that the initialization is not restricted to any particular offset estimator; for example, the semi-blind frequency-offset estimator [4] can also be used.

B. Pilot symbols and virtual carriers placement

Pilot symbols should be optimized for joint estimation of frequency offset and channel impulse response. For a system with no frequency offset, the placement and power allocation may be optimized by minimizing the symbol error probability with channel estimation error [21]. If we assume that the frequency offset estimation is perfect, the optimal pilot symbols for least-squares channel estimation are equispaced when $J = N/N_p$ is an integer. If J is not integeral, the pilot symbols are placed with two spacings S, $S+1$, and $S = \lfloor N/N_p \rfloor$ or quasi-equispaced. $N_1 = N - N_p S$ pilot spacings are equal to S and $N_2 = N_p - N_1$ pilot spacings are equal to $S+1$, where $\lfloor x \rfloor$ denotes the largest integer less than x.

For frequency offset estimation, the pilot symbols may be optimally designed to minimize the MSE (38), which is to maximize $\mathrm{tr}\{\mathbf{M}\mathbf{G}^{-1}\mathbf{M}\mathbf{G}\}$. Since $\mathbf{G}$ depends on $\mathbf{R}_h$ and σ_n^2, an explicit design criterion is elusive. Since the pilot symbols must also be optimized for channel estimation, we suggest they be equispaced or quasi-equispaced. To simplify the system design, they are also chosen from $\mathcal{Q}$ (the data symbol constellation) by maximizing $\mathrm{tr}\{\mathbf{M}\mathbf{G}^{-1}\mathbf{M}\mathbf{G}\}$. For example, for the 6-ray COST 207 TU model and an SNR of 20 dB, the OFDM system with $N_p = 4$ and QPSK has the optimal pilots $\sqrt{2}/2 + j\sqrt{2}/2$, $-\sqrt{2}/2 + j\sqrt{2}/2$, $-\sqrt{2}/2 - j\sqrt{2}/2$, $\sqrt{2}/2 - j\sqrt{2}/2$.

When there are only VCs, our proposed estimator is equivalent to that in [2] in high SNR. In practical systems, low and high frequency subcarriers are VCs, since they are often used for transmit filtering. We thus use the same VC placement strategy regardless of the placement of pilot symbols. Since the performance improvement due to large N_v is small [2], the number of VCs in standards as IEEE 802.11a is small [1].

V. Numerical Results

We now present numerical results to illustrate the effectiveness of the proposed joint estimator for a practical OFDM system. We assume the following system specifications:

1) Both the data and pilot symbols are chosen from the QPSK constellation, $\mathcal{Q}$.
2) The carrier frequency of the OFDM system is 5 GHz and the data bandwidth is 2 MHz. The guard interval is $N_g = 16$.
3) We consider the 6-ray COST 207 TU model with the PDP $[0.189, 0.379, 0.239, 0.095, 0.061, 0.037]$ and delay profile $[0.0, 0.2, 0.5, 1.6, 2.3, 5.0]\mu s$ [22]. The channel remains constant for each OFDM frame but varies from one to another. The PDP is known at the receiver.
4) We consider a normalized frequency offset of 0.2 and the number of frames $K = 1$.
5) The number of subcarriers is $N = 64$, the number of VCs is $N_v = 12$, and the number of pilot symbols is $N_p = 8$. The pilot symbols are equispaced and randomly chosen.

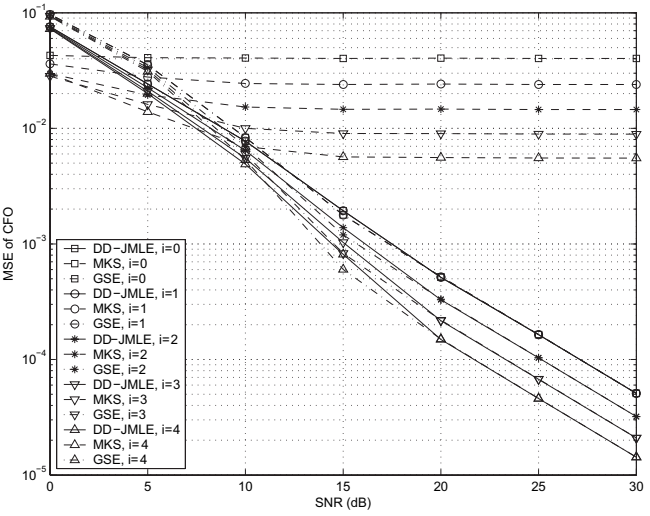

Fig. 8. Comparison of frequency offset MSE for different joint estimators in a QPSK OFDM system with $\epsilon = 0.2$, $N = 64$, $N_p = 8$ and $N_v = 12$.

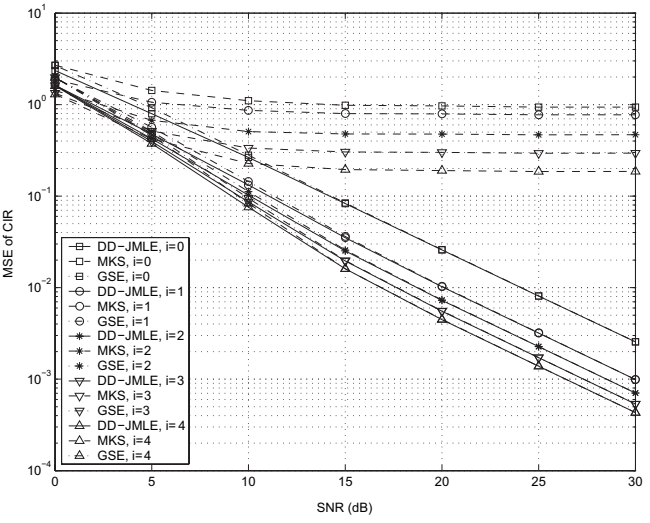

Fig. 9. Comparison of channel impulse response MSE for different joint estimators in a QPSK OFDM system with $\epsilon = 0.2$, $N = 64$, $N_p = 8$ and $N_v = 12$.

We compare the complexity and performance of DD-JMLE with that of the joint estimator [11] (denoted by "MKS") and that of the semi-blind frequency-offset estimator [4] (denoted by "GSE"). The iteration number is denoted by i.

Table I compares the complexity of DD-JMLE, MKS and GSE, where $\beta = 1 - [\log_2 m + 2(1/m - 1)]/\log_2(mN)$. GSE uses the FFT frequency offset estimation algorithm in Section III B., and the iterative algorithm in Section IV A. Therefore, the complexity of DD-JMLE and GSE are same. In general, the complexity of DD-JMLE and GSE are higher than that of MKS. But the former two achieve better performance than the latter as shown below.

Figs. 8 and 9 show the MSE of frequency offset and channel impulse response, respectively. The performance of all the estimators improves with the number of iterations (Fig. 8). Although the DD-JMLE performs better than GSE in low SNR, they both perform identically in high SNR. However, MKS exhibits a large error floor, possibly caused by the poor

TABLE I
COMPLEXITY COMPARISION OF DD-JMLE, MKS AND GSE.

Algorithm	Real products	Real additions
DD-JMLE	$2N[2N+2+m\beta\log_2(mN)]$ $+i[2N\log_2 N+4N(2N+L+5)+4L]$ $+2i\beta(mN)\log_2(mN)$	$N[3N+1+3m\beta\log 2(mN)]$ $+i[3N\log_2 N+4.5N^2+(3L+10.5)N+3L-3]$ $+3i\beta(mN)\log_2(mN)$
MKS	$2N(2L+2+\log_2 N)$ $+i\,[3(4L+5)(N-1)+2\log_2 N+8L+9]$	$2L(N-1)+3N\log_2 N$ $+i\,[(6L+7)(N-1)+3\log_2 N+4L]$
GSE	$2N[2N+2+m\beta\log_2(mN)]$ $+i[2N\log_2 N+4N(2N+L+5)+4L]$ $+2i\beta(mN)\log_2(mN)$	$N[3N+1+3m\beta\log 2(mN)]$ $+i[3N\log_2 N+4.5N^2+(3L+10.5)N+3L-3]$ $+3i\beta(mN)\log_2(mN)$

quality of the initial estimate. Interestingly, MKS performs better than both DD-JMLE and GSE in low SNR. For DD-JMLE, the second iteration has about a 2-dB gain over the first iteration at an MSE of 10^{-4}, while the gain increases to 2.8 dB and 2.4 dB for the third and fourth iterations, respectively. Similarly, for channel estimation, (Fig. 9), MKS still exhibits an error floor caused by the residual frequency offset. The DD-JMLE again performs better than GSE in low SNR. The first iteration offers the largest improvement. At an MSE of 3×10^{-3}, the performance gain after the first iteration is 4 dB due to the use of all the transmitted symbols. The second iteration has another 1.5-dB gain. In the third iteration, the gain reduces to 1.2 dB. Therefore, the performance improvement by increasing the number of iterations diminishes with the increase of i.

The BER of different data detectors is shown in Fig. 10. A reference receiver with perfect knowledge of channel impulse response and frequency offset is used as the benchmark. As before, the MKS exhibits an error floor in high SNR, which can be reduced by increasing the number of iterations. As with the frequency offset and channel impulse response results, DD-JMLE performs better than GSE in low SNR, but both perform similarly in high SNR. After the first iteration, DD-JMLE has a 2.1-dB gain over that with initial estimates at a BER of 10^{-3}. The performance gain by further increasing i is in the order of 0.1 dB. At a BER of 10^{-1}, one iteration of DD-JMLE has a 2.5-dB loss over the benchmark, while this gap reduces to 0.8 dB at a BER of 10^{-3}. However, this gap seems to remain constant with the increase in SNR and i.

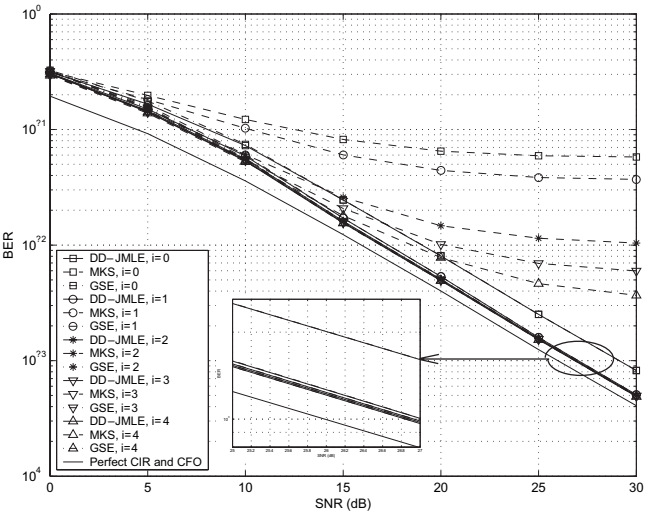

Fig. 10. Comparison of BER for different joint estimators in a QPSK OFDM system with $\epsilon=0.2$, $N=64$, $N_p=8$ and $N_v=12$.

VI. CONCLUSION

We have investigated joint estimation of channel impulse response and frequency offset for OFDM systems. A high resolution frequency-offset estimator that uses both pilot symbols and VCs has been derived. Our estimator reduces to Liu and Tureli's estimator [2] in high SNR for systems that do not employ pilot symbols. When $N_p \le L$, we find our frequency-offset estimator cannot improve the estimator performance in high SNR. If $N_p > L$, our estimator acts as a subspace-based estimator with N_v+N_p-L virtual carriers, but it performs channel estimation simultaneously. We have established the asymptotic unbiasedness and derived the asymptotic MSE and the approximate CRB. A decision-directed joint ML estimator has been derived to improve the estimates of frequency offset, data symbols and channel impulse response iteratively, and is initialized using the frequency-offset estimator (22) and the least-squares channel estimator. Pilots and VCs design rules have also been discussed. The results show that our proposed joint estimator performs within 0.8 dB of the ideal case, even when the frequency offset is as large as 20%.

APPENDIX A

We would like to test the hypothesis that $\mathbf{y}$ in (4) is multivariate Gaussian with zero mean and correlation (19). Many approaches to testing multivariate normality have been proposed in the literature. Mardia [23] has developed two tests based on multivariate generalization of kurtosis and skewness. Using these tests, despite extensive trials, we find the rejection rate is less than 1%. Thus, the evidence of non-normality is quite weak.

APPENDIX B

In this appendix, we derive the MSE and the unbiasedness of the proposed frequency-offset estimator. We also derive the Cramér-Rao bound for the estimation of ϵ. Taking the derivatives of $g(\epsilon)$ in (22) and (24), we have

$$\begin{aligned}\dot{g}(\epsilon)&=\jmath\frac{2\pi}{N}\mathbf{y}^H\mathbf{\Gamma}(\epsilon)\mathbf{B}\mathbf{\Gamma}^H(\epsilon)\mathbf{y},\\ \ddot{g}(\epsilon)&=\left(\frac{2\pi}{N}\right)^2\mathbf{y}^H\mathbf{\Gamma}(\epsilon)\mathbf{D}\mathbf{\Gamma}^H(\epsilon)\mathbf{y},\end{aligned}\tag{44}$$

where $\mathbf{B} = \mathbf{M}\mathbf{G}^{-1} - \mathbf{G}^{-1}\mathbf{M}$, $\mathbf{D} = 2\mathbf{M}\mathbf{G}^{-1}\mathbf{M} - \mathbf{M}^2\mathbf{G}^{-1} - \mathbf{G}^{-1}\mathbf{M}^2$, $\mathbf{M} = \text{diag}\{0, 1, 2, \cdots, N-1\}$, and $\mathbf{G}$ is defined in (19).

If $\mathbf{x}$ is a complex Gaussian vector with mean $\mathbf{m}$ and covariance matrix $\mathbf{S}$, and $\mathbf{A}$ is a matrix, we have [24] $E(\mathbf{x}^H\mathbf{A}\mathbf{x}) = \text{tr}(\mathbf{A}\mathbf{S}) + \mathbf{m}^H\mathbf{A}\mathbf{m}$ and

$$\begin{aligned} E(\mathbf{x}^H\mathbf{A}\mathbf{x}\mathbf{x}^H\mathbf{A}\mathbf{x}) =& \text{tr}(\mathbf{A}\mathbf{S}(\mathbf{A}+\mathbf{A}^H)\mathbf{S}) \\ &+ \mathbf{m}^H(\mathbf{A}+\mathbf{A}^H)\mathbf{S}(\mathbf{A}+\mathbf{A}^H)\mathbf{m} \\ &+ [\text{tr}(\mathbf{A}\mathbf{S}) + \mathbf{m}^H\mathbf{A}\mathbf{m}]^2. \end{aligned} \quad (45)$$

From (19), $\mathbf{R}_y = \mathbf{\Gamma}(\epsilon)\mathbf{G}\mathbf{\Gamma}^H(\epsilon)$ and $E(\mathbf{y}) = \mathbf{0}$. We thus have

$$\begin{aligned} E\{\dot{g}(\epsilon)\} &= E\left\{\jmath\frac{2\pi}{N}\mathbf{y}^H\mathbf{\Gamma}(\epsilon)\mathbf{B}\mathbf{\Gamma}^H(\epsilon)\mathbf{y}\right\} \\ &= \jmath\frac{2\pi}{N}\text{tr}\left[\mathbf{\Gamma}(\epsilon)\mathbf{B}\mathbf{\Gamma}^H(\epsilon)\mathbf{R}_y\right] \\ &= \jmath\frac{2\pi}{N}\text{tr}\left[\mathbf{\Gamma}(\epsilon)\mathbf{B}\mathbf{G}\mathbf{\Gamma}^H(\epsilon)\right]. \end{aligned} \quad (46)$$

Using the trace property $\text{tr}(\mathbf{B}\mathbf{C}) = \text{tr}(\mathbf{C}\mathbf{B})$ ($\mathbf{B}$ and $\mathbf{C}$ are matrices), we have

$$\begin{aligned} \text{tr}\left[\mathbf{\Gamma}(\epsilon)\mathbf{B}\mathbf{G}\mathbf{\Gamma}^H(\epsilon)\right] &= \text{tr}\left[(\mathbf{M}\mathbf{G}^{-1} - \mathbf{G}^{-1}\mathbf{M})\mathbf{G}\right] \\ &= \text{tr}(\mathbf{M}) - \text{tr}(\mathbf{G}^{-1}\mathbf{M}\mathbf{G}) = 0. \end{aligned} \quad (47)$$

Therefore, we have $E\{\hat{\epsilon}\} \doteq \epsilon$. Hence, our proposed frequency-offset estimator is unbiased.

Next, we derive the MSE of the frequency offset estimate. From (37), we need to compute $E\{[\dot{g}(\epsilon)]^2\}$ and $E\{\ddot{g}(\epsilon)\}$. We can show that

$$E\{\ddot{g}(\epsilon)\} = \frac{8\pi^2}{N^2}\text{tr}\left(\mathbf{M}\mathbf{G}^{-1}\mathbf{M}\mathbf{G} - \mathbf{M}^2\right). \quad (48)$$

Considering (44) and (45), we can obtain

$$E\{[\dot{g}(\epsilon)]^2\} = -2\left(\frac{2\pi}{N}\right)^2\left\{\text{tr}(\mathbf{B}\mathbf{G}\mathbf{B}\mathbf{G}) + [\text{tr}(\mathbf{B}\mathbf{G})]^2\right\}. \quad (49)$$

From (47), we have $\text{tr}(\mathbf{B}\mathbf{G}) = 0$. We thus get

$$E\{[\dot{g}(\epsilon)]^2\} = \frac{16\pi^2}{N^2}\text{tr}\left(\mathbf{M}\mathbf{G}^{-1}\mathbf{M}\mathbf{G} - \mathbf{M}^2\right). \quad (50)$$

Substituting (49) and (50) into (37) yields (38).

The Cramér-Rao bound for the estimation of ϵ is

$$\text{CRB} = -\frac{1}{E\left\{\dfrac{\partial^2\Lambda(\mathbf{y}|\epsilon)}{\partial\epsilon^2}\right\}} = \frac{1}{E\{\ddot{g}(\epsilon)\}}, \quad (51)$$

where the log-likelihood function is defined in (21). Using (48) and (51), we obtain (39).

REFERENCES

[1] "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: high-speed physical layer in the 5 GHz band," in IEEE Std 802.11a, Sept. 1999.

[2] H. Liu and U. Tureli, "A high-efficiency carrier estimator for OFDM communications," *IEEE Commun. Lett.*, vol. 2, no. 4, pp. 104–106, April 1998.

[3] U. Tureli, D. Kivanc, and H. Liu, "Experimental and analytical studies on a high-resolution OFDM carrier frequency offset estimator," *IEEE Trans. Veh. Technol.*, vol. 50, no. 2, pp. 629–643, March 2001.

[4] M. Ghogho and A. Swami, "Semi-blind frequency offset synchronization for OFDM," in *Proc. ICASSP 2002*, pp. 2333–2336.

[5] U. Tureli, P. Honan, and H. Liu, "Low-complexity nonlinear least squares carrier offset estimator for OFDM: identifiability, diversity and performance," *IEEE Trans. Signal Processing*, vol. 52, no. 9, pp. 2441–2452, Sept. 2004.

[6] B. Chen, "Maximum likelihood estimation of OFDM carrier frequency offset," *IEEE Signal Processing Lett.*, vol. 9, no. 4, pp. 123–126, April 2002.

[7] D. Warrier and U. Madhow, "Spectrally efficient noncoherent communication," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 651–668, March 2002.

[8] Y.-S. Choi, P. Voltz, and F. Cassara, "ML estimation of carrier frequency offset for multicarrier signals in Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 50, no. 2, pp. 644–655, March 2001.

[9] E. Chiavaccini and G. M. Vitetta, "Maximum-likelihood frequency recovery for OFDM signals transmitted over multipath fading channels," *IEEE Trans. Commun.*, vol. 52, no. 2, pp. 244–251, Feb. 2004.

[10] T. Cui and C. Tellambura, "Maximum-likelihood carrier frequency offset estimation for OFDM systems over frequency-selective fading channels," in *Proc. ICC 2005*.

[11] X. Ma, H. Kobayashi, and S. C. Schwartz, "Joint frequency offset and channel estimation for OFDM," in *Proc. GLOBECOM 2003*, pp. 15–19.

[12] C. Gamier, M. Loosvelt, V. L. Thuc, Y. Delignon, and L. Clavier, "Performance of an OFDM-SDMA based system in a time-varying multi-path channel," in *Proc. VTC 2001*, pp. 1686–1690.

[13] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 4, pp. 1207–1215, July 2000.

[14] S. Wei, D. Goeckel, and P. Kelly, "A modern extreme value theory approach to calculating the distribution of the peak-to-average power ratio in OFDM systems," in *Proc. ICC 2002*, pp. 1686–1690.

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

[16] M. Morelli and U. Mengali, "Carrier-frequency estimation for transmissions over selective channels," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1580–1589, Sept. 2000.

[17] M. H. Meyrs and L. Franks, "Joint carrier phase and symbol timing recovery for PAM systems," *IEEE Trans. Commun.*, vol. 28, no. 8, pp. 1121–1129, Aug. 1980.

[18] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis, First Edition*, R. Roberts, Ed. Addison Wesley, 1991.

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

[20] U. Mengali and A. D'Andrea, *Synchronization Techniques for Digital Receivers, First Edition*. Springer, 1997.

[21] X. Cai and G. B. Giannakis, "Error probability minimizing pilots for OFDM with M-PSK modulation over Rayleigh-fading channels," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 146–155, Jan. 2004.

[22] G. L. Stuber, *Principles of Mobile Communication, Second Edition*. Norwell, MA: Kluwer Academic, 2001.

[23] K. V. Mardia, "Measures of multivariate skewness and kurtosis with applications," *Biometrika*, vol. 57, no. 3, pp. 519–530, 1970.

[24] M. Brookes, "Matrix reference manual," [online] available: http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/intro.html.