Joint Estimation of Channel Response, Frequency Offset, and Phase Noise in OFDM

Darryl Dexu Lin, *Student Member, IEEE*, Ryan A. Pacheco, *Student Member, IEEE*, Teng Joon Lim, *Senior Member, IEEE*, and Dimitrios Hatzinakos, *Senior Member, IEEE*

***Abstract*—Accurate channel estimates are needed in orthogonal frequency-division multiplexing (OFDM), and easily obtained under the assumption of perfect phase and frequency synchronization. However, the practical receiver encounters nonnegligible phase noise (PHN) and carrier frequency offset (CFO), which create substantial intercarrier interference that a conventional OFDM channel estimator cannot account for. In this paper, we introduce an optimal (maximum a posteriori) joint estimator for the channel impulse response (CIR), CFO, and PHN, utilizing prior statistical knowledge of PHN that can be obtained from measurements or data sheets. In addition, in cases where a training symbol consists of two identical halves in the time domain, we propose a variant to Moose's CFO estimation algorithm that optimally removes the effect of PHN with lower complexity than with a nonrepeating training symbol. To further reduce the complexity of the proposed algorithms, simplified implementations based on the conjugate gradient method are also introduced such that the estimators studied in this paper can be realized efficiently using the fast Fourier transform with only minor performance degradation.**

***Index Terms*—Carrier frequency offset, channel estimation, conjugate gradient, orthogonal frequency-division multiplexing (OFDM), phase noise.**

I. Introduction

ORTHOGONAL frequency-division multiplexing (OFDM) is a well-known modulation technique [1] that has become a preferred choice in high-rate wireless and wireline communication systems such as broadband wireless access (IEEE802.16), wireless local-area networks (Wi-Fi), high-speed digital subscriber lines (DSLs), and digital broadcasting (DAB and DVB). This is due to its spectral efficiency—no guard bands are needed between adjacent frequency channels—and more importantly, its implementation simplicity compared to traditional time-domain modulation methods in channels with severe intersymbol interference (ISI), encountered whenever bit rates are required to be very large.

OFDM does have its drawbacks relative to time-domain modulation, most significantly its extreme sensitivity to time-varying multiplicative effects such as fast fading, Doppler shifts, and oscillator jitter. The latter two effects lead to a mismatch between the carrier frequencies of the received signal and the local oscillator, so that a frequency offset Δf Hz is created. Oscillator jitter also creates a very damaging effect called phase noise, meaning that the phase of the locally generated sinusoid randomly changes over time.

While the detrimental effects of frequency offset and phase noise have been well documented [2]–[8], successful alleviation of these combined problems based on a statistically optimal receiver implementation, which must include channel estimation in the presence of carrier frequency offset (CFO) and phase noise (PHN), has not been proposed. In [9] and [10], PHN suppression methods were proposed for frequency-selective channels but the channel frequency response was assumed to be known prior to PHN suppression. In [11], PHN was considered in the formulation of the channel estimation problem but was not directly used in the solution and thus the method is not statistically optimal. In [12], channel estimation was performed but first the PHN was estimated using at least one "carrier recovery" pilot tone that required frequency guard bands on both sides to minimize interference from data symbols. Only specific frequency-selective channels were considered in the simulations, so the performance in more general Rayleigh frequency-selective channels is unknown but it is expected that performance will degrade if a channel null should occur in the vicinity of the pilot tone. In [13], PHN was estimated using pilot symbols based on a linearized parametric model, but it is suboptimal because the intercarrier interference (ICI) introduced by the PHN in the received signal is ignored. In [14], a precoding method was proposed that used null guard bands in the frequency domain. The method is designed for bandlimited multiplicative effects but was also tested on the Wiener phase noise model with some success. However, it requires expensive operations such as singular value decomposition and null guard bands, which reduce spectral efficiency.

In this paper, our goal is to tackle the channel estimation problem when CFO and PHN are present through the maximization of a "complete likelihood function."[1] Special features of the complete likelihood function are taken advantage of that enable a unique and elegant joint estimation scheme achieving statistically optimal performance.

The rest of this paper will be organized as follows. Section II discusses the power spectral density (PSD) of the voltage-controlled oscillator (VCO) output in connection with the prior distribution of the PHN process and presents the signal model of a CFO/PHN channel. Section III derives the *joint CFO/PHN/CIR estimator* (JCPCE), which performs accurate channel estimation even with CFO and PHN impairment. Section IV considers

Manuscript received July 19, 2005; revised November 14, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mounir Ghogho.

The authors are with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: linde@comm.utoronto.ca; ryan@comm.utoronto.ca; limtj@comm.utoronto.ca; dimitris@comm.utoronto.ca).

Digital Object Identifier 10.1109/TSP.2006.879265

[1]The complete likelihood function is the joint distribution of the observed variables (e.g., $\mathbf{r}$) and latent variables (e.g., ϵ, $\boldsymbol{\theta}$, $\mathbf{g}$) in the probabilistic model [15, ch. 11].

1053-587X/$20.00 © 2006 IEEE

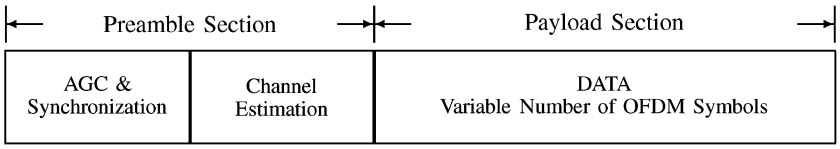


Fig. 1. OFDM packet structure.

a special case when repeating training symbols are available and introduces a variant to Moose's CFO estimation algorithm which optimally rejects the effect of PHN. The channel impulse response (CIR) is subsequently estimated by optimally cancelling the remaining PHN. Section V contains simulations of the proposed algorithms and tests their robustness in a wide range of scenarios. Section VI contains the conclusions.

Notation: Upper and lower case boldface letters indicate matrices and column vectors; $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ denote conjugation, transpose, and Hermitian transpose, respectively; $\mathbf{1}$ and $\mathbf{0}$ represent the all-one and all-zero column vector, respectively; $\mathrm{diag}(\mathbf{x})$ is a diagonal matrix with the vector $\mathbf{x}$ on its diagonal; $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ denote the real and imaginary part of a vector or matrix; $\mathrm{E}(\cdot)$ and $\mathrm{V}(\cdot)$ stand for the expected value and variance of a random variable; $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ represent, respectively, real and circularly symmetric complex Gaussian random vectors with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. In particular, for an N-dimensional circularly symmetric complex Gaussian random vector, $\mathbf{x}$

$$\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\pi^N |\boldsymbol{\Sigma}|} \exp\left\{-(\mathbf{x}-\boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right\}. \quad (1)$$

II. System Description

A. Prior Statistics of Phase Noise

Two different models of PHN are available in the literature [2]. The first one models a free-running oscillator and assumes the PHN process to be a Wiener process that is nonstationary and whose power grows with time. The second one models an oscillator controlled by a phase-locked loop (PLL) and approximates the PHN process as a zero-mean colored Gaussian process that is wide-sense stationary and has finite power. For simplicity, we will refer to the first one as *Wiener PHN* and the second one as *Gaussian PHN*, even though both assume Gaussian statistics. In this paper, we focus on Gaussian PHN, but the generalization to Wiener PHN is straightforward [16], [17] (by simply adjusting the PHN covariance matrix $\boldsymbol{\Phi}$) and is omitted here for conciseness.

Denoting the phase noise process at the output of the phase-locked VCO by $\theta(t)$, the samples of $\theta(t)$, i.e., $\theta(kT/N)$, within the mth OFDM symbol $\boldsymbol{\theta}_m$, has a multivariate Gaussian prior distribution: $p(\boldsymbol{\theta}_m) = \mathcal{N}(\mathbf{0}, \boldsymbol{\Phi})$, where the samples are taken at a rate of N/T samples per second, where N is the number of OFDM subcarriers and T is the period of the OFDM symbol. For this model to be useful, however, the covariance matrix $\boldsymbol{\Phi}$ must be available. Conveniently, instead of measuring $\boldsymbol{\Phi}$ through time averaging, we may calculate it according to the specifications of the phase-locked VCO as follows.

We first write the output of the VCO with PHN as

$$s(t) = e^{j(2\pi f_o t + \theta(t))}. \quad (2)$$

Then the autocorrelation function of $s(t)$, $R_s(\tau)$, can be calculated

$$\begin{aligned} R_s(\tau) &\doteq \mathrm{E}\{s^*(t)s(t+\tau)\} \\ &= e^{j2\pi f_o \tau} \mathrm{E}\left\{e^{j(\theta(t+\tau)-\theta(t))}\right\} \\ &= e^{j2\pi f_o \tau} e^{-R_\theta(0)} e^{R_\theta(\tau)} \\ &\approx e^{j2\pi f_o \tau} e^{-R_\theta(0)} (1 + R_\theta(\tau)) \end{aligned} \quad (3)$$

where the approximation is tight when $R_\theta(0) \ll 1$ (since $|R_\theta(\tau)| \leq R_\theta(0)$). This is a common assumption made about PHN processes [12]. The third equality is obtained from the characteristic function of $\theta(t) - \theta(t+\tau)$, which is Gaussian. It follows after taking the Fourier transform that the PSD of $s(t)$ is

$$S_s(f) = e^{-R_\theta(0)} (\delta(f - f_o) + S_\theta(f - f_o)) \quad (4)$$

where $S_\theta(f) = \mathcal{F}\{R_\theta(\tau)\}$. The shape of $S_s(f)$ may be measured by a spectrum analyzer or provided as part of the VCO specifications (phase-noise masks are commonly known) and hence $S_\theta(f)$ can be found.

The autocorrelation function $R_\theta(\tau)$ of the PHN process can be obtained from the inverse Fourier transform of $S_\theta(f)$. Since the PHN process has zero mean, this is also its autocovariance function. Finally, the value on the ith row and jth column of $\boldsymbol{\Phi}$ is extracted from $R_\theta(\tau)$

$$\boldsymbol{\Phi}_{i,j} = R_\theta\left(|i-j|\frac{T}{N}\right) \quad (5)$$

since T/N is the sampling period. The methods presented in this paper assume knowledge of $\boldsymbol{\Phi}$, but the sensitivity to inexact knowledge (bandwidth and power of the PHN) is examined through simulations in Section V-A4.

B. Signal Model

We consider a slow fading frequency-selective channel where the CIR is assumed to remain constant during each packet of transmission, which consists of multiple OFDM symbols including the initial preambles for synchronization and channel estimation as well as the variable-length payload that follows (as depicted in Fig. 1).

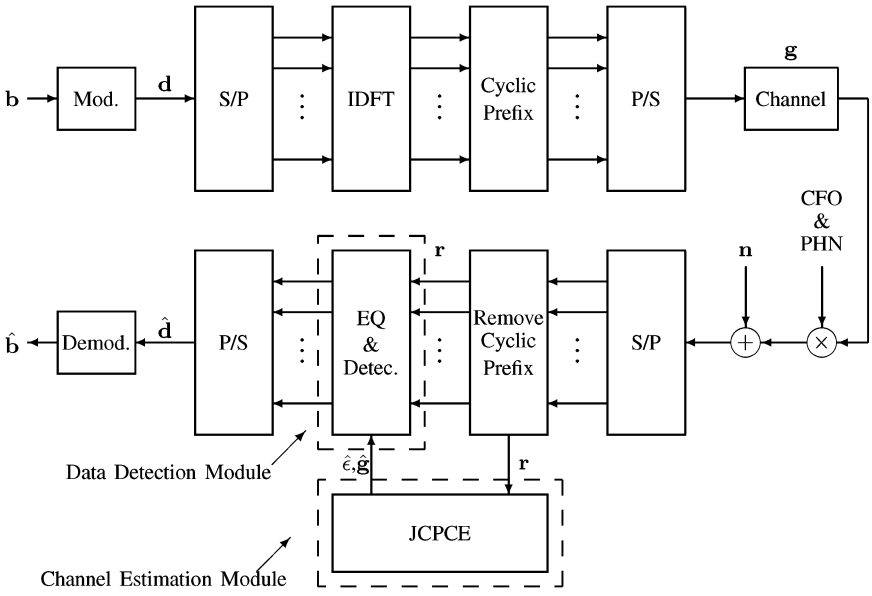


Fig. 2. OFDM transmitter/receiver structure and phase noise channel model. The channel is estimated using the proposed JCPCE.

Assuming perfect timing synchronization, the complex baseband received signal of an OFDM symbol within the training period sampled at rate N/T can be written as

$$r_n = \frac{1}{\sqrt{N}} e^{j(\theta_n + 2\pi\epsilon n/N)} \sum_{k=0}^{N-1} h_k d_k e^{j2\pi nk/N} + \eta_n, \quad n = 0, \ldots, N-1 \quad (6)$$

where $\epsilon = \Delta f T$ is the normalized CFO; $\{\theta_n\}_{n=0}^{N-1}$ is the discrete-time PHN sequence; $\{h_k\}_{k=0}^{N-1}$ is the channel frequency response at subcarriers 0 to $N-1$; $\{d_k\}_{k=0}^{N-1}$ are the transmitted data symbols belonging to an M-QAM constellation; and $\{\eta_n\}_{n=0}^{N-1}$ is complex white Gaussian noise with variance σ^2 per dimension. Equation (6) may be written in matrix form as

$$\mathbf{r} = \mathbf{EPF}^H\mathbf{Hd} + \mathbf{n} \quad (7)$$

where $\mathbf{F} \in \mathbb{C}^{N\times N}$ is the DFT matrix with the (l,m)th element being $\mathbf{F}_{l,m} = (1/\sqrt{N})e^{-j(2\pi(l-1)(m-1)/N)}$; $\mathbf{d} = [d_0, \ldots, d_{N-1}]^T$ is the data vector; $\mathbf{n} = [\eta_0, \ldots, \eta_{N-1}]^T$ is the noise vector with distribution $p(\mathbf{n}) = \mathcal{CN}(\mathbf{0}, 2\sigma^2\mathbf{I})$; $\mathbf{P} = \text{diag}([e^{j\theta_0}, \ldots, e^{j\theta_{N-1}}]^T)$ is the PHN matrix; $\mathbf{E} = \text{diag}([1, e^{j2\pi\epsilon/N}, \ldots, e^{j2\pi(N-1)\epsilon/N}]^T)$ is the CFO matrix; and $\mathbf{H} = \text{diag}(\mathbf{h}) = \text{diag}([h_0, \ldots, h_{N-1}]^T)$ is the channel matrix. Notice that although a full OFDM symbol contains $N_g + N$ time samples, with N_g being the length of the cyclic prefix, in this signal model we assume the cyclic prefix has been removed and so there are only N samples per OFDM symbol. We may rewrite (7) as

$$\mathbf{r} = \mathbf{EPGF}^H\mathbf{d} + \mathbf{n} \quad (8)$$

where $\mathbf{G}$ is defined as $\mathbf{G} = \mathbf{F}^H\mathbf{HF}$ and is a circulant matrix. Using $\mathbf{g} = [g_0, \ldots, g_{L-1}]^T$ to denote the CIR, where L is the channel length, the CIR can be converted to the channel frequency response by writing $\mathbf{h} = \mathbf{Wg} \in \mathbb{C}^{N\times 1}$. Note that $\mathbf{g}/\sqrt{N}$ is the true CIR, but for simplicity, we shall also assign the name to $\mathbf{g}$. $\mathbf{W}$ is a partition of the DFT matrix, i.e., $\mathbf{F} = [\mathbf{W}|\mathbf{V}]$, in which $\mathbf{W} \in \mathbb{C}^{N\times L}$ and $\mathbf{V} \in \mathbb{C}^{N\times(N-L)}$ are orthogonal unitary matrices satisfying $\mathbf{W}^H\mathbf{V} = \mathbf{0}$ and $\mathbf{WW}^H + \mathbf{VV}^H = \mathbf{I}$.

Let $\mathbf{D} = \text{diag}(\mathbf{d})$. We can now introduce the following equivalent representation of (7) for the convenience of channel estimation:

$$\mathbf{r} = \mathbf{EPF}^H\mathbf{DWg} + \mathbf{n}. \quad (9)$$

C. Role of Channel Estimation in Receiver Design

The transmitter/receiver structure and the channel model over a period of one OFDM symbol, taking into account the distortion caused by CFO and PHN, are illustrated in Fig. 2. In many OFDM standards, such as IEEE 802.11a and HiperLAN2, there are mainly two types of physical layer symbols: the preamble symbols and the payload symbols. Issues such as automatic gain control, frequency offset correction, and timing synchronization are resolved by the earlier portion of preamble symbols, while fine frequency tuning and channel estimation are performed by the later portion of preamble symbols. Timing synchronization is commonly performed using autocorrelation based metrics [18], [19], which are insensitive to CFO and PHN disturbances. At the receiver, when the preamble symbols used for channel estimation are received, it is assumed that perfect timing synchronization is achieved and the $\mathbf{d}$ vector is exactly known. On the other hand, when the payload is received, it can be assumed that the channel is known (except for possible fine tuning using the embedded pilot symbols) and the data symbols

are to be detected. A diagram exemplifying this packet structure is given in Fig. 1.

From (9), in the absence of the CFO term $\mathbf{E}$ and PHN term $\mathbf{P}$, it is easy to obtain the optimal estimate of $\mathbf{g}$ given the training symbols $\mathbf{d}$ and the received signal $\mathbf{r}$ in the channel estimation phase. In the data detection phase, we assume perfect knowledge of $\mathbf{g}$ and simply extract the maximum likelihood estimate of $\mathbf{d}$. Unfortunately, this simple procedure is not possible in a CFO/PHN channel, as the presence of $\mathbf{E}$ and $\mathbf{P}$ first renders the channel estimate inaccurate and later (through a different PHN sequence) impairs the performance of data detection.

This paper will focus on obtaining accurate channel and CFO estimates in the presence of PHN. Our design methodology goes as follows: we will start by introducing an optimal joint CFO/PHN/CIR estimator. Then, we seek means to decrease the complexity of the optimal estimator by transmitting repeating training symbols and also by using the conjugate gradient method. We will not be dealing with the data detection problem in this paper, because with the accurate estimate of the channel and CFO (which are quasi-static), the data detection stage only suffers from the unknown PHN distortion (which is time-varying). Interested readers are referred to [20] for detailed analysis of the challenging data detection problem in the presence of PHN.

III. Channel Estimation With CFO and PHN

A. Joint CFO/PHN/CIR Estimator (JCPCE)

An OFDM channel estimator may either estimate the channel frequency response $(\mathbf{h})$ or the CIR $(\mathbf{g})$. In this paper, we shall assume that we are always interested in the CIR $\mathbf{g}$, since its low dimensionality leads to welcome computational savings as well as lower variance for the obtained estimates. Furthermore, we assume that the tap length of the impulse response (or, equivalently, the dimension of $\mathbf{g}$) L is known a priori. Looking at (9), it is obvious that the optimal estimates for $\mathbf{E}$, $\mathbf{P}$, and $\mathbf{g}$ are coupled and in general difficult to obtain. But as the following derivation shows, we are very fortunate in this case, as the joint optimization problem can in fact be decoupled.

First, we write the "complete likelihood function" $p(\mathbf{r},\epsilon,\boldsymbol{\theta},\mathbf{g}) = p(\mathbf{r}|\epsilon,\boldsymbol{\theta},\mathbf{g})p(\epsilon)p(\boldsymbol{\theta})p(\mathbf{g})$, which is proportional to the a posteriori distribution $p(\epsilon,\boldsymbol{\theta},\mathbf{g}|\mathbf{r})$. $p(\epsilon)$ and $p(\mathbf{g})$ are constants (representing noninformative priors) as no prior knowledge of ϵ and $\mathbf{g}$ is assumed. Also, we have assumed in Section II-A that the prior distribution of $\boldsymbol{\theta}$ is $\mathcal{N}(\mathbf{0},\boldsymbol{\Phi})$, where $\boldsymbol{\Phi}$ is known. The "complete negative log-likehood function" can therefore be written as

$$\begin{aligned}\mathcal{L}(\epsilon,\boldsymbol{\theta},\mathbf{g}) &= -\log p(\mathbf{r}|\epsilon,\boldsymbol{\theta},\mathbf{g}) - \log p(\boldsymbol{\theta}) \\ &= \frac{1}{2\sigma^2}(\mathbf{r}-\mathbf{EPF}^H\mathbf{DWg})^H(\mathbf{r}-\mathbf{EPF}^H\mathbf{DWg}) \\ &\quad + \frac{1}{2}\boldsymbol{\theta}^T\boldsymbol{\Phi}^{-1}\boldsymbol{\theta}. \end{aligned} \tag{10}$$

Our objective is to find the optimal estimates

$$(\hat{\epsilon},\hat{\boldsymbol{\theta}},\hat{\mathbf{g}}) = \arg\min_{\epsilon,\boldsymbol{\theta},\mathbf{g}} \mathcal{L}(\epsilon,\boldsymbol{\theta},\mathbf{g}). \tag{11}$$

1) Forward Substitution: Solving $\partial\mathcal{L}(\epsilon,\boldsymbol{\theta},\mathbf{g})/\partial\mathbf{g}^* = \mathbf{0}$ produces the optimal channel estimate in terms of ϵ and $\boldsymbol{\theta}$

$$\hat{\mathbf{g}} = \mathbf{W}^H\mathbf{D}^{-1}\mathbf{D}^{-H}\mathbf{W}\mathbf{W}^H\mathbf{D}^H\mathbf{F}\mathbf{P}^H\mathbf{E}^H\mathbf{r}. \tag{12}$$

Note that when $\mathbf{EP} = \mathbf{I}$, this is the expression for the conventional maximum likelihood channel estimator with no PHN and CFO. It shall be assumed hereafter that constant-modulus training symbols are used, i.e., $\mathbf{D}^H\mathbf{D} = 2\rho^2\mathbf{I}$. This assumption is reasonable for a practical channel estimator, and it simplifies the expressions in subsequent derivations.[2] This assumption leads to

$$\hat{\mathbf{g}} = (2\rho^2)^{-1}\mathbf{W}^H\mathbf{D}^H\mathbf{F}\mathbf{P}^H\mathbf{E}^H\mathbf{r}. \tag{13}$$

Noticing that

$$\begin{aligned}&\mathbf{r}-\mathbf{EPF}^H\mathbf{DW}\hat{\mathbf{g}} \\ &\quad= \mathbf{r}-(2\rho^2)^{-1}\mathbf{EPF}^H\mathbf{DWW}^H\mathbf{D}^H\mathbf{F}\mathbf{P}^H\mathbf{E}^H\mathbf{r} \\ &\quad= \mathbf{r}-(2\rho^2)^{-1}\mathbf{EPF}^H\mathbf{D}(\mathbf{I}-\mathbf{VV}^H)\mathbf{D}^H\mathbf{F}\mathbf{P}^H\mathbf{E}^H\mathbf{r} \\ &\quad= \mathbf{r}-\left(\mathbf{I}-(2\rho^2)^{-1}\mathbf{EPF}^H\mathbf{DVV}^H\mathbf{D}^H\mathbf{F}\mathbf{P}^H\mathbf{E}^H\right)\mathbf{r} \\ &\quad= (2\rho^2)^{-1}\mathbf{EPF}^H\mathbf{DVV}^H\mathbf{D}^H\mathbf{F}\mathbf{P}^H\mathbf{E}^H\mathbf{r} \end{aligned} \tag{14}$$

and substituting (14) into (10), we have after simplification

$$\begin{aligned}\mathcal{L}(\epsilon,\boldsymbol{\theta}) &= \frac{1}{4\sigma^2\rho^2}\mathbf{r}^H\mathbf{EPF}^H\mathbf{DVV}^H\mathbf{D}^H\mathbf{F}\mathbf{P}^H\mathbf{E}^H\mathbf{r} \\ &\quad + \frac{1}{2}\boldsymbol{\theta}^T\boldsymbol{\Phi}^{-1}\boldsymbol{\theta} \\ &= \frac{1}{4\sigma^2\rho^2}\mathbf{u}^T\mathbf{E}\underbrace{\mathbf{R}^H\mathbf{F}^H\mathbf{DV}}_{\mathbf{C}}\underbrace{\mathbf{V}^H\mathbf{D}^H\mathbf{FR}}_{\mathbf{C}^H}\mathbf{E}^H\mathbf{u}^* \\ &\quad + \frac{1}{2}\boldsymbol{\theta}^T\boldsymbol{\Phi}^{-1}\boldsymbol{\theta} \end{aligned} \tag{15}$$

where $\mathbf{R} = \mathrm{diag}(\mathbf{r})$ and $\mathbf{u} = [e^{j\theta_0},\ldots,e^{j\theta_{N-1}}]^T$. Realizing that for small $\boldsymbol{\theta}$, $\mathbf{u} \approx \mathbf{1} + j\boldsymbol{\theta}$, and letting $\mathbf{C} = \mathbf{R}^H\mathbf{F}^H\mathbf{DV}$

$$\begin{aligned}\mathcal{L}(\epsilon,\boldsymbol{\theta}) &\approx \frac{1}{4\sigma^2\rho^2}(\mathbf{1}+j\boldsymbol{\theta})^T\mathbf{ECC}^H\mathbf{E}^H(\mathbf{1}-j\boldsymbol{\theta}) + \frac{1}{2}\boldsymbol{\theta}^T\boldsymbol{\Phi}^{-1}\boldsymbol{\theta} \\ &= \frac{1}{4\sigma^2\rho^2}\Big[\boldsymbol{\theta}^T\mathbf{ECC}^H\mathbf{E}^H\boldsymbol{\theta} + 2\sigma^2\rho^2\boldsymbol{\theta}^T\boldsymbol{\Phi}^{-1}\boldsymbol{\theta} \\ &\qquad - j\mathbf{1}^T\mathbf{ECC}^H\mathbf{E}^H\boldsymbol{\theta} + j\boldsymbol{\theta}^T\mathbf{ECC}^H\mathbf{E}^H\mathbf{1} \\ &\qquad + \mathbf{1}^T\mathbf{ECC}^H\mathbf{E}^H\mathbf{1}\Big] \\ &= \frac{1}{4\sigma^2\rho^2}\Big[\boldsymbol{\theta}^T\mathrm{Re}(\mathbf{ECC}^H\mathbf{E}^H)\boldsymbol{\theta} + 2\sigma^2\rho^2\boldsymbol{\theta}^T\boldsymbol{\Phi}^{-1}\boldsymbol{\theta} \\ &\qquad - 2\boldsymbol{\theta}^T\mathrm{Im}(\mathbf{ECC}^H\mathbf{E}^H)\mathbf{1} \\ &\qquad + \mathbf{1}^T\mathbf{ECC}^H\mathbf{E}^H\mathbf{1}\Big]. \end{aligned} \tag{16}$$

[2]The extension to non-constant-modulus training symbols simply requires the use of (12) instead of (13) in the subsequent steps.

TABLE I
JOINT CFO/PHN/CIR ESTIMATOR (JCPCE)

Step 1:	$\hat{\epsilon} = \arg\min_{\epsilon} \mathbf{1}^T\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H\mathbf{1} - \mathbf{1}^T\mathrm{Im}(\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H)^T$ $\times[\mathrm{Re}(\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H) + 2\sigma^2\rho^2\mathbf{\Phi}^{-1}]^{-1}\mathrm{Im}(\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H)\mathbf{1}$; $\hat{\mathbf{E}} = \mathrm{diag}([1, e^{j2\pi\hat{\epsilon}/N}, \cdots, e^{j2\pi(N-1)\hat{\epsilon}/N}]^T)$;
Step 2:	$\hat{\boldsymbol{\theta}} = [\mathrm{Re}(\hat{\mathbf{E}}\mathbf{C}\mathbf{C}^H\hat{\mathbf{E}}^H) + 2\sigma^2\rho^2\mathbf{\Phi}^{-1}]^{-1}\mathrm{Im}(\hat{\mathbf{E}}\mathbf{C}\mathbf{C}^H\hat{\mathbf{E}}^H)\mathbf{1}$; $\hat{\mathbf{P}} = \mathrm{diag}([e^{j\hat{\theta}_0}, \cdots, e^{j\hat{\theta}_{N-1}}]^T)$;
Step 3:	$\hat{\mathbf{g}} = (2\rho^2)^{-1}\mathbf{W}^H\mathbf{D}^H\mathbf{F}\hat{\mathbf{P}}^H\hat{\mathbf{E}}^H\mathbf{r}$.

Note that the last equality holds for real valued $\boldsymbol{\theta}$. Solving $\partial\mathcal{L}(\epsilon,\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \mathbf{0}$ gives us the optimal estimate of $\boldsymbol{\theta}$ in terms of ϵ

$$\hat{\boldsymbol{\theta}} = \left[\mathrm{Re}(\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H) + 2\sigma^2\rho^2\mathbf{\Phi}^{-1}\right]^{-1}\mathrm{Im}(\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H)\mathbf{1}. \tag{17}$$

Substituting (17) into (16) and simplifying, we have, after scaling by a constant

$$\mathcal{L}(\epsilon) = -\mathbf{1}^T\mathrm{Im}(\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H)^T\left[\mathrm{Re}(\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H) + 2\sigma^2\rho^2\mathbf{\Phi}^{-1}\right]^{-1}\mathrm{Im}(\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H)\mathbf{1} + \mathbf{1}^T\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H\mathbf{1}. \tag{18}$$

Hence by searching over a range of feasible values of ϵ, we may find the optimal estimate of ϵ

$$\hat{\epsilon} = \arg\min_{\epsilon}\mathcal{L}(\epsilon). \tag{19}$$

In the absence of PHN, i.e., with $\boldsymbol{\theta} = \mathbf{0}$ in (9), $\mathcal{L}(\epsilon) = \mathbf{1}^T\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H\mathbf{1}$ would be the metric for CFO estimation in the joint estimator for $\mathbf{g}$ and ϵ.

2) Backward Substitution: After finding $\hat{\epsilon}$ and correspondingly $\hat{\mathbf{E}}$, the values of $\hat{\boldsymbol{\theta}}$ can be determined by substituting $\mathbf{E} = \hat{\mathbf{E}}$ into (17)

$$\hat{\boldsymbol{\theta}} = \left[\mathrm{Re}(\hat{\mathbf{E}}\mathbf{C}\mathbf{C}^H\hat{\mathbf{E}}^H) + 2\sigma^2\rho^2\mathbf{\Phi}^{-1}\right]^{-1}\mathrm{Im}(\hat{\mathbf{E}}\mathbf{C}\mathbf{C}^H\hat{\mathbf{E}}^H)\mathbf{1}. \tag{20}$$

Letting $\hat{\mathbf{P}} = \mathrm{diag}(\exp(j\hat{\boldsymbol{\theta}}))$ and plugging it into (13), the optimal channel estimate after removing the CFO and PHN is therefore

$$\hat{\mathbf{g}} = (2\rho^2)^{-1}\mathbf{W}^H\mathbf{D}^H\mathbf{F}\hat{\mathbf{P}}^H\hat{\mathbf{E}}^H\mathbf{r}. \tag{21}$$

We summarize the complete JCPCE algorithm in Table I.

Note that finding closed-form expressions for the jointly optimal estimates for ϵ, $\boldsymbol{\theta}$, and $\mathbf{g}$ is due to the unitary property of CFO and PHN matrices: $\mathbf{E}^H\mathbf{E} = \mathbf{I}$ and $\mathbf{P}^H\mathbf{P} = \mathbf{I}$. This property is also utilized in [21] to establish the optimality of MUSIC-based CFO estimation in OFDM.

B. Complexity Analysis and Low Complexity Implementation

In the OFDM system, computational complexity is a critical issue. A receiver design with good performance but high complexity would conflict with the original motivation of using OFDM, which implies a low complexity order of $\mathcal{O}(N\log N)$ with the use of FFT. In the implementation of the JCPCE, the main computational tasks rest in evaluating (18), (20), and (21). We will now investigate the complexity of each computation and seek means to reduce complexity.

In (21), we see that $\mathbf{D}$, $\hat{\mathbf{P}}$, and $\hat{\mathbf{E}}$ are diagonal matrices, while $\mathbf{F}$ and $\mathbf{W}^H$ are FFT or partial FFT matrices. Thus each step of matrix-vector multiplication has a complexity order of $\mathcal{O}(N\log N)$ or less.

The more challenging task is the evaluation of (20), which involves a matrix inversion requiring in general a complexity order of $\mathcal{O}(N^3)$. However, as we will demonstrate in the following, with the help of the conjugate gradient (CG) method [22], we are able to lower the complexity to an acceptable level.

First we let $\mathbf{\Psi} = (1/2\sigma^2\rho^2)\mathbf{\Phi}$ and $\mathbf{q} = \mathrm{Im}(\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H)\mathbf{1}$, where $\mathbf{q}$ can be computed efficiently using FFT since all matrices involved in calculating $\mathbf{q}$ are either diagonal or FFT (or partial FFT) matrices. In order to proceed, we notice that $\mathbf{\Psi}$, as a Toeplitz matrix, can be approximated by a circulant matrix $\tilde{\mathbf{\Psi}}$ [23], [24] according to this simple result.

Theorem 1: The best circulant approximation to a symmetric Toeplitz matrix $\mathbf{\Psi} \in \mathbb{C}^{N\times N}$, $\tilde{\mathbf{\Psi}} \in \mathbb{C}^{N\times N}$, in the sense of minimizing the Frobenius norm $\|\mathbf{\Psi} - \tilde{\mathbf{\Psi}}\|_F$, is a circulant matrix whose first row $\tilde{\boldsymbol{\psi}}^T = [\tilde{\psi}_0, \ldots, \tilde{\psi}_{N-1}]$ has entries

$$\tilde{\psi}_i = \frac{(N-i)\psi_i + i\psi_{N-i}}{N} \tag{22}$$

where $\boldsymbol{\psi}^T = [\psi_0, \ldots, \psi_{N-1}]$ is the first row of $\mathbf{\Psi}$. This operation has complexity of $\mathcal{O}(N)$.

Proof: See [24]. ■

It can be shown that this approximation is asymptotically exact as $N \to \infty$ for an autocorrelation matrix $\mathbf{\Psi}$ of a first-order autoregressive process, which is a good fit to the PHN process assumed in [25]. Being a circulant matrix, the eigenvalue decomposition of $\tilde{\mathbf{\Psi}}$ is $\mathbf{F}\mathbf{\Lambda}_{\tilde{\Psi}}\mathbf{F}^H$ and $\tilde{\mathbf{\Psi}}^{-1} = \mathbf{F}\mathbf{\Lambda}_{\tilde{\Psi}}^{-1}\mathbf{F}^H$, where $\mathbf{\Lambda}_{\tilde{\Psi}}$ is a diagonal matrix. It is well known that $\mathbf{\Lambda}_{\tilde{\Psi}} = \mathrm{diag}(\sqrt{N}\mathbf{F}^H\tilde{\boldsymbol{\varphi}}_1)$, where $\tilde{\boldsymbol{\varphi}}_1$ is the first column of $\tilde{\mathbf{\Psi}}$. Replacing $\mathbf{\Psi}$ by $\tilde{\mathbf{\Psi}}$, the modified estimator for $\boldsymbol{\theta}$ becomes

$$\hat{\boldsymbol{\theta}} = \left[\mathrm{Re}(\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H) + \tilde{\mathbf{\Psi}}^{-1}\right]^{-1}\mathrm{Im}(\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H)\mathbf{1} \tag{23}$$

which is equivalent to solving a linear equation $[\mathrm{Re}(\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H) + \tilde{\mathbf{\Psi}}^{-1}]\boldsymbol{\theta} = \mathbf{q}$. This problem can be easily tackled by the conjugate gradient method. The complete algorithm is presented in Table II.

Of all the operations in Table II, the dominant complexity is associated with the matrix-vector multiplication $[\mathrm{Re}(\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H) + \tilde{\mathbf{\Psi}}^{-1}]\boldsymbol{\nu}_k$. Thanks to the circulant structure of $\tilde{\mathbf{\Psi}}$, this can be performed using FFT because $\tilde{\mathbf{\Psi}}^{-1} = \mathbf{F}\mathbf{\Lambda}_{\Psi}^{-1}\mathbf{F}^H$. More specifically, evaluating $[\mathrm{Re}(\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H) + \tilde{\mathbf{\Psi}}^{-1}]\boldsymbol{\nu}_k$ requires $7N + 6N\log N$ operations. Thus, the overall complexity of every iteration of the CG algorithm is $\mathcal{O}(N\log N)$. The CG algorithm requires a maximum of N iterations to converge

TABLE II
CONJUGATE GRADIENT ALGORITHM FOR EVALUATING (23)

Initialization:	
	$\boldsymbol{\theta}_0 = \mathbf{0}$
	$\boldsymbol{\gamma}_0 = [\mathrm{Re}(\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H) + \tilde{\boldsymbol{\Psi}}^{-1}]\boldsymbol{\theta}_0 - \mathbf{q} = -\mathbf{q}$
	$\boldsymbol{\nu}_0 = -\boldsymbol{\gamma}_0 = \mathbf{q}$
For	$k = 0 : i-1$
	$\alpha_k = \boldsymbol{\gamma}_k^H\boldsymbol{\gamma}_k / (\boldsymbol{\nu}_k^H[\mathrm{Re}(\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H) + \tilde{\boldsymbol{\Psi}}^{-1}]\boldsymbol{\nu}_k)$
	$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \alpha_k\boldsymbol{\nu}_k$
	$\boldsymbol{\gamma}_{k+1} = \boldsymbol{\gamma}_k + \alpha_k[\mathrm{Re}(\mathbf{E}\mathbf{C}\mathbf{C}^H\mathbf{E}^H) + \tilde{\boldsymbol{\Psi}}^{-1}]\boldsymbol{\nu}_k$
	$\beta_{k+1} = \frac{\boldsymbol{\gamma}_{k+1}^H\boldsymbol{\gamma}_{k+1}}{\boldsymbol{\gamma}_k^H\boldsymbol{\gamma}_k}$
	$\boldsymbol{\nu}_{k+1} = -\boldsymbol{\gamma}_{k+1} + \beta_{k+1}\boldsymbol{\nu}_k$
End	

to the exact solution. Yet for this particular problem, because the PHN process can be sufficiently characterized by very few signal dimensions, we only require a very small number of iterations for an accurate estimate, as will be apparent in the simulations. In conclusion, the complexity of evaluating (20) is $\mathcal{O}(iN\log N)$, where i is the number of iterations in the CG algorithm. The practical value of i is investigated by simulations in Section V-A3 and is found to be about five to ten.

The computation of (18) can be done almost identically as (20) using the circulant approximation of $\boldsymbol{\Psi}$ and the CG algorithm. However, a crucial drawback of the search method for finding the optimal $\hat{\epsilon}$ is that the complexity scales with the inverse of the resolution required for $\hat{\epsilon}$. Certainly, this problem is offset by the fact that CFO synchronization needs only be done once when the link between the transmitter and receiver is established, but it is still advantageous if further simplifications can be made.

IV. CFO ESTIMATION BASED ON REPEATED TRAINING SYMBOLS

For cases where the system has very limited computational power, it is beneficial to obtain a closed-form solution for ϵ. When no PHN is present, the pioneering work of Moose [26] achieves just that by assuming that the two halves of a training symbol are identical.

A. Moose's CFO Estimator

In the Moose algorithm [26], we transmit an OFDM symbol with two identical halves in the time domain. Such a signal is easily generated [19] by transmitting $N/2$ training symbols $d_0,\ldots,d_{N/2-1}$ on the even subcarriers and zeros on the odd subcarriers. The N-point sequence in time at the receiver, with CFO and PHN distortion, can be written as

$$r_n = \frac{1}{\sqrt{N/2}} e^{j(\theta_n + 2\pi\epsilon n/N)} \sum_{k=0}^{N/2-1} h_k d_k e^{j4\pi nk/N} + \eta_n \quad (24)$$

for $n = 0,\ldots,N-1$.

We shall first assume no PHN, i.e., $\theta_n = 0$ for $n = 0,\ldots,N-1$. Denoting $\mathbf{r}_1 = [r_0,\ldots,r_{N/2-1}]^T$ and $\mathbf{r}_2 = [r_{N/2},\ldots,r_{N-1}]^T$, we have

$$\mathbf{r}_1 = \mathbf{x} + \mathbf{n}_1; \quad \mathbf{r}_2 = e^{j\pi\epsilon}\mathbf{x} + \mathbf{n}_2 \quad (25)$$

where $\mathbf{x} = \mathbf{E}\mathbf{G}\mathbf{F}^H\mathbf{d} \in \mathbb{C}^{(N/2)\times 1}$ and $\mathbf{F}^H\mathbf{d} \in \mathbb{C}^{(N/2)\times 1}$ is the training symbol that is transmitted twice consecutively, $\mathbf{n}_1 \sim \mathcal{CN}(\mathbf{0}, 2\sigma^2\mathbf{I})$ and $\mathbf{n}_2 \sim \mathcal{CN}(\mathbf{0}, 2\sigma^2\mathbf{I})$ are independent additive noise vectors. Here the CFO matrix $\mathbf{E}$, channel circular convolution matrix $\mathbf{G}$, and DFT matrix $\mathbf{F}$ follow similar definitions as before but are only half the size.

The optimal estimate of ϵ is

$$\hat{\epsilon} = \arg\max_{\epsilon} p(\mathbf{r}_1, \mathbf{r}_2|\epsilon) = \arg\max_{\epsilon} p(\mathbf{r}_2|\epsilon, \mathbf{r}_1)p(\mathbf{r}_1|\epsilon) \quad (26)$$

which reduces to $\hat{\epsilon} = \arg\max_{\epsilon} p(\mathbf{r}_2|\epsilon, \mathbf{r}_1)$ if we assume that $p(\mathbf{r}_1|\epsilon) = p(\mathbf{r}_1)$ (this is an approximation since $\mathbf{r}_1$ and ϵ are in general not independent).

Notice that

$$\mathbf{r}_2 = e^{j\pi\epsilon}\mathbf{r}_1 - e^{j\pi\epsilon}\mathbf{n}_1 + \mathbf{n}_2 = e^{j\pi\epsilon}\mathbf{r}_1 + \mathbf{z} \quad (27)$$

where $p(\mathbf{z}) = \mathcal{CN}(\mathbf{0}, 4\sigma^2\mathbf{I})$ because the instantaneous value of $\mathbf{z}$ depends on ϵ but the statistics does not. We then have $p(\mathbf{r}_2|\epsilon, \mathbf{r}_1) = \mathcal{CN}(e^{j\pi\epsilon}\mathbf{r}_1, 4\sigma^2\mathbf{I})$. Therefore, the negative log-likelihood function becomes

$$-\log p(\mathbf{r}_2|\epsilon, \mathbf{r}_1) = \frac{1}{4\sigma^2}(\mathbf{r}_2 - e^{j\pi\epsilon}\mathbf{r}_1)^H(\mathbf{r}_2 - e^{j\pi\epsilon}\mathbf{r}_1). \quad (28)$$

And it follows that

$$\hat{\epsilon} = \frac{1}{\pi}\measuredangle\mathbf{r}_1^H\mathbf{r}_2 \quad (29)$$

where $\measuredangle x$ denotes the phase angle of a complex number x.

B. CFO Estimator With PHN Rejection

In the presence of PHN, the derivation presented above fails because (27) no longer holds. We propose, in the following, a CFO estimation algorithm that optimally accounts for PHN.

Rewriting (25) to include the PHN distortion, we have

$$\mathbf{r}_1 = \mathbf{P}_1\mathbf{x} + \mathbf{n}_1; \quad \mathbf{r}_2 = e^{j\pi\epsilon}\mathbf{P}_2\mathbf{x} + \mathbf{n}_2 \quad (30)$$

where $\mathbf{P}_1$ and $\mathbf{P}_2$ contain consequtive PHN sequences $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. The optimal estimate of ϵ is then

$$\begin{aligned}\hat{\epsilon} &= \arg\max_{\epsilon} p(\mathbf{r}_1, \mathbf{r}_2|\epsilon) \\ &= \arg\max_{\epsilon} \int_{\boldsymbol{\theta}_2}\int_{\boldsymbol{\theta}_1} p(\mathbf{r}_1, \mathbf{r}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\epsilon)d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2 \end{aligned} \quad (31)$$

where

$$\begin{aligned}p(\mathbf{r}_1, \mathbf{r}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\epsilon) &= p(\mathbf{r}_1, \mathbf{r}_2|\epsilon, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\ &= p(\mathbf{r}_2|, \mathbf{r}_1, \epsilon, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\ &\quad \times p(\mathbf{r}_1|\epsilon, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2). \end{aligned} \quad (32)$$

Assuming $p(\mathbf{r}_1|\epsilon, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = p(\mathbf{r}_1)$ as before, it follows that

$$\hat{\epsilon} = \arg\max_{\epsilon} \int_{\boldsymbol{\theta}_2}\int_{\boldsymbol{\theta}_1} p(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2. \quad (33)$$

Denoting the "differential PHN" sequence $\boldsymbol{\theta}_\Delta = \boldsymbol{\theta}_2 - \boldsymbol{\theta}_1 \in \mathbb{R}^{(N/2)\times 1}$ and $\mathbf{P}_\Delta = \text{diag}([e^{j\theta_{\Delta(0)}}, \ldots, e^{j\theta_{\Delta((N/2)-1)}}]^T)$, $\mathbf{r}_2$ can be written in terms of $\mathbf{r}_1$ as

$$\begin{aligned}\mathbf{r}_2 &= e^{j\pi\epsilon}\mathbf{P}_\Delta \mathbf{r}_1 - e^{j\pi\epsilon}\mathbf{P}_\Delta \mathbf{n}_1 + \mathbf{n}_2 \\ &= e^{j\pi\epsilon}\mathbf{P}_\Delta \mathbf{r}_1 + \mathbf{z}\end{aligned} \tag{34}$$

where $p(\mathbf{z}) = \mathcal{CN}(\mathbf{0}, 4\sigma^2\mathbf{I})$. In other words, $p(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathcal{CN}(e^{j2\pi\epsilon}\mathbf{P}_\Delta \mathbf{r}_1, 4\sigma^2\mathbf{I})$. This means that $p(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is only a function of $\boldsymbol{\theta}_\Delta$ instead of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ individually. We may therefore rewrite (33) as

$$\begin{aligned}\hat{\epsilon} &= \arg\max_\epsilon \int_{\boldsymbol{\theta}_\Delta} p(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_\Delta) p(\boldsymbol{\theta}_\Delta) d\boldsymbol{\theta}_\Delta \\ &= \arg\max_\epsilon p(\mathbf{r}_2|\mathbf{r}_1, \epsilon).\end{aligned} \tag{35}$$

Lemma 1: If $[\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T]^T \in \mathbb{R}^{N\times 1}$ is a jointly Gaussian random vector with distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Phi})$, where $\boldsymbol{\Phi} \in \mathbb{R}^{N\times N}$ can be partitioned into four $(N/2)\times(N/2)$ blocks

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_{N/2} & \boldsymbol{\Upsilon} \\ \boldsymbol{\Upsilon}^T & \boldsymbol{\Phi}_{N/2} \end{bmatrix} \tag{36}$$

then $p(\boldsymbol{\theta}_\Delta) = p(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1) = \mathcal{N}(\mathbf{0}, 2\boldsymbol{\Phi}_{N/2} - \boldsymbol{\Upsilon} - \boldsymbol{\Upsilon}^T)$.

Proof: See Appendix I. ∎

Denoting $\boldsymbol{\Phi}_\Delta \doteq 2\boldsymbol{\Phi}_{N/2} - \boldsymbol{\Upsilon} - \boldsymbol{\Upsilon}^T$, we may write $p(\boldsymbol{\theta}_\Delta) = \mathcal{N}(\mathbf{0}, \boldsymbol{\Phi}_\Delta)$. Finally, we use the following lemma to evaluate $p(\mathbf{r}_2|\mathbf{r}_1, \epsilon)$ in (35).

Lemma 2: Given $p(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_\Delta) = \mathcal{CN}(e^{j\pi\epsilon}\mathbf{P}_\Delta \mathbf{r}_1, 4\sigma^2\mathbf{I})$ and $p(\boldsymbol{\theta}_\Delta) = \mathcal{N}(\mathbf{0}, \boldsymbol{\Phi}_\Delta)$, then

$$p(\mathbf{r}_2|\mathbf{r}_1, \epsilon) = \mathcal{CN}\left(e^{j\pi\epsilon}\mathbf{r}_1, \mathbf{R}_1\boldsymbol{\Phi}_\Delta\mathbf{R}_1^H + 4\sigma^2\mathbf{I}\right) \tag{37}$$

where $\mathbf{R}_1 = \text{diag}(\mathbf{r}_1)$.

Proof: See Appendix II. ∎

From (35) and Lemma 2, we see that

$$\begin{aligned}\hat{\epsilon} &= \arg\max_\epsilon \log p(\mathbf{r}_2|\mathbf{r}_1, \epsilon) \\ &= \arg\min_\epsilon (\mathbf{r}_2 - e^{j\pi\epsilon}\mathbf{r}_1)^H \left(\mathbf{R}_1\boldsymbol{\Phi}_\Delta\mathbf{R}_1^H + 4\sigma^2\mathbf{I}\right)^{-1} \\ &\quad \times (\mathbf{r}_2 - e^{j\pi\epsilon}\mathbf{r}_1) \\ &= \frac{1}{\pi}\measuredangle \mathbf{r}_1^H \left(\mathbf{R}_1\boldsymbol{\Phi}_\Delta\mathbf{R}_1^H + 4\sigma^2\mathbf{I}\right)^{-1}\mathbf{r}_2.\end{aligned} \tag{38}$$

This expression is very similar to (29) except for a weighting matrix that accounts for the distortion caused by PHN. [Note that (38) represents a novel CFO estimation scheme in the PHN channel.]

C. Joint PHN and CIR Estimation

With the CFO estimated using (38), we now turn to the remaining channel estimation issue in the presence of PHN. Because of the special structure of the repeating training symbol, we are required to rederive the optimal joint PHN/CIR estimation algorithm.

Expressing (24) in the matrix form yields

$$\mathbf{r} = \hat{\mathbf{E}}\mathbf{P}\check{\mathbf{F}}^H\mathbf{D}\mathbf{W}\mathbf{g} + \mathbf{n} \tag{39}$$

where $\mathbf{r} = [\mathbf{r}_1^T, \mathbf{r}_2^T]^T \in \mathbb{C}^{N\times 1}$ is the time-domain received repeating training symbol; $\hat{\mathbf{E}} \in \mathbb{C}^{N\times N}$ is the CFO matrix already estimated; $\mathbf{P} \in \mathbb{C}^{N\times N}$ is the unknown PHN matrix; $\check{\mathbf{F}} = [\mathbf{F}, \mathbf{F}] \in \mathbb{C}^{N/2\times N}$ is the cascade of two DFT matrices; $\mathbf{D} = \text{diag}(\mathbf{d}) \in \mathbb{C}^{N/2\times N/2}$ contains the length-$N/2$ training symbol; and $\mathbf{g} \in \mathbb{C}^{L\times 1}$ is the channel impulse response.

Similar to (10), we obtain the "complete negative log-likelihood function"

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}, \mathbf{g}) = \frac{1}{2\sigma^2}(\mathbf{r} - \hat{\mathbf{E}}\mathbf{P}\check{\mathbf{F}}^H\mathbf{D}\mathbf{W}\mathbf{g})^H(\mathbf{r} - \hat{\mathbf{E}}\mathbf{P}\check{\mathbf{F}}^H\mathbf{D}\mathbf{W}\mathbf{g}) \\ + \frac{1}{2}\boldsymbol{\theta}^T\boldsymbol{\Phi}^{-1}\boldsymbol{\theta}.\end{aligned} \tag{40}$$

Solving $\partial\mathcal{L}(\boldsymbol{\theta}, \mathbf{g})/\partial\mathbf{g}^* = \mathbf{0}$ produces the optimal channel estimate of $\mathbf{g}$ in terms of $\boldsymbol{\theta}$

$$\hat{\mathbf{g}} = (4\rho^2)^{-1}\mathbf{W}^H\mathbf{D}^H\check{\mathbf{F}}\mathbf{P}^H\hat{\mathbf{E}}^H\mathbf{r}. \tag{41}$$

Noticing that

$$\begin{aligned}&(\mathbf{r} - \hat{\mathbf{E}}\mathbf{P}\check{\mathbf{F}}^H\mathbf{D}\mathbf{W}\hat{\mathbf{g}})^H(\mathbf{r} - \hat{\mathbf{E}}\mathbf{P}\check{\mathbf{F}}^H\mathbf{D}\mathbf{W}\hat{\mathbf{g}}) \\ &= \mathbf{r}^H\mathbf{r} - (4\rho^2)^{-1}\mathbf{r}^H\hat{\mathbf{E}}\mathbf{P}\check{\mathbf{F}}^H\mathbf{D}\mathbf{W}\mathbf{W}^H\mathbf{D}^H\check{\mathbf{F}}\mathbf{P}^H\hat{\mathbf{E}}^H\mathbf{r} \\ &= \frac{1}{4\rho^2}\mathbf{r}^H\hat{\mathbf{E}}\mathbf{P}\begin{bmatrix}\mathbf{F}^H\mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}^H\mathbf{D}\end{bmatrix}\begin{bmatrix}\mathbf{I}+\mathbf{V}\mathbf{V}^H & -\mathbf{W}\mathbf{W}^H \\ -\mathbf{W}\mathbf{W}^H & \mathbf{I}+\mathbf{V}\mathbf{V}^H\end{bmatrix} \\ &\quad \times \begin{bmatrix}\mathbf{F}^H\mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}^H\mathbf{D}\end{bmatrix}^H\mathbf{P}^H\hat{\mathbf{E}}^H\mathbf{r}\end{aligned} \tag{42}$$

and substituting (42) into (40), we have after simplification

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{8\sigma^2\rho^2}\mathbf{u}^T\hat{\mathbf{E}}\mathbf{A}\hat{\mathbf{E}}^H\mathbf{u}^* + \frac{1}{2}\boldsymbol{\theta}^T\boldsymbol{\Phi}^{-1}\boldsymbol{\theta} \tag{43}$$

where

$$\begin{aligned}\mathbf{A} = \mathbf{R}^H\begin{bmatrix}\mathbf{F}^H\mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}^H\mathbf{D}\end{bmatrix}\begin{bmatrix}\mathbf{I}+\mathbf{V}\mathbf{V}^H & -\mathbf{W}\mathbf{W}^H \\ -\mathbf{W}\mathbf{W}^H & \mathbf{I}+\mathbf{V}\mathbf{V}^H\end{bmatrix} \\ \times \begin{bmatrix}\mathbf{F}^H\mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}^H\mathbf{D}\end{bmatrix}^H\mathbf{R}\end{aligned} \tag{44}$$

and $\mathbf{R} = \text{diag}(\mathbf{r})$, $\mathbf{u} = [e^{j\theta_0}, \ldots, e^{j\theta_{N-1}}]^T$. Solving $\partial\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \mathbf{0}$ gives us, similar to (17), the optimal estimate of $\boldsymbol{\theta}$

$$\hat{\boldsymbol{\theta}} = \left[\text{Re}(\hat{\mathbf{E}}\mathbf{A}\hat{\mathbf{E}}^H) + 4\sigma^2\rho^2\boldsymbol{\Phi}^{-1}\right]^{-1}\text{Im}(\hat{\mathbf{E}}\mathbf{A}\hat{\mathbf{E}}^H)\mathbf{1}. \tag{45}$$

We summarize the modified JCPCE algorithm for the case of repeating training symbols in Table III.

TABLE III
MODIFIED JCPCE WITH CLOSED-FORM CFO ESTIMATION

Step 1:	$\hat{\epsilon} = \frac{1}{\pi}\angle\, \mathbf{r}_1^H(\mathbf{R}_1\mathbf{\Phi}_\Delta\mathbf{R}_1^H + 4\sigma^2\mathbf{I})^{-1}\mathbf{r}_2$;
	$\hat{\mathbf{E}} = \text{diag}([1, e^{j2\pi\hat{\epsilon}/N}, \cdots, e^{j2\pi(N-1)\hat{\epsilon}/N}]^T)$;
Step 2:	$\hat{\boldsymbol{\theta}} = [\text{Re}(\hat{\mathbf{E}}\mathbf{A}\hat{\mathbf{E}}^H) + 4\sigma^2\rho^2\mathbf{\Phi}^{-1}]^{-1}\text{Im}(\hat{\mathbf{E}}\mathbf{A}\hat{\mathbf{E}}^H)\mathbf{1}$;
	$\hat{\mathbf{P}} = \text{diag}([e^{j\hat{\theta}_0}, \cdots, e^{j\hat{\theta}_{N-1}}]^T)$;
Step 3:	$\hat{\mathbf{g}} = (4\rho^2)^{-1}\mathbf{W}^H\mathbf{D}^H\check{\mathbf{F}}\hat{\mathbf{P}}^H\hat{\mathbf{E}}^H\mathbf{r}$.

D. Complexity Analysis and Low Complexity Implementation

The derivation in the previous section shows that, with the help of repeating training symbols, the CFO estimation can be done in closed form with the distortion due to PHN optimally rejected. However, the final expression in (38) still requires a matrix inversion. Fortunately, complexity reduction is also available for this computation since $\mathbf{\Phi}_\Delta$ is a Toeplitz matrix with a close circulant approximation $\tilde{\mathbf{\Phi}}_\Delta$. We may concentrate on the matrix vector product

$$\mathbf{x} = \left(\mathbf{R}_1\tilde{\mathbf{\Phi}}_\Delta\mathbf{R}_1^H + 4\sigma^2\mathbf{I}\right)^{-1}\mathbf{r}_2. \tag{46}$$

This is equivalent to solving a linear equation $(\mathbf{R}_1\tilde{\mathbf{\Phi}}_\Delta\mathbf{R}_1^H + 4\sigma^2\mathbf{I})\mathbf{x} = \mathbf{r}_2$, which can be computed efficiently using the CG method analogous to the one described in Table II.

V. SIMULATIONS

In this section, we simulate the performance of JCPCE and its variants based on algorithms presented in Tables I–III. The following system parameters are assumed in our simulations unless stated otherwise:

1) a Rayleigh multipath fading channel with a delay of $L = 10$ taps and an exponentially decreasing power delay profile that has a decay constant of four taps;
2) an OFDM training symbol size of $N = 64$ subcarriers with each subcarrier modulated in quadrature phase-shift keying (QPSK) format;
3) baseband sampling rate $f_s = 20$ MHz (subcarrier spacing of 312.5 kHz);
4) a phase-locked VCO at the receiver with PHN standard deviation of $\theta_{\text{rms}} = 3$ degrees (i.e., $R_\theta(0) = (\pi\theta_{\text{rms}}/180)^2$).

The channel is normalized such that $||\mathbf{g}||^2/N = 1$. The random PHN is generated, according to the Matlab code recommended for the IEEE 802.11g standard [25], as independent identically distributed Gaussian samples passed through a single-pole Butterworth filter of 3 dB bandwidth $\Omega_o = 100$ kHz. Hence, the PHN covariance matrix $\mathbf{\Phi}$ is

$$\mathbf{\Phi}_{i,j} = \left(\frac{\pi\theta_{\text{rms}}}{180}\right)^2 e^{-\frac{2\pi\Omega_o|i-j|}{f_s}}. \tag{47}$$

A. Channel Estimation With PHN Only

In general, PHN is a more complex effect than CFO and is harder to analyze. We will first perform simulations with no CFO to study the joint PHN and CIR estimation described as part of the JCPCE algorithm (Steps 3–5 in Table I with $\hat{\mathbf{E}} = \mathbf{I}$), as well as its low-complexity variant summarized in Table II.

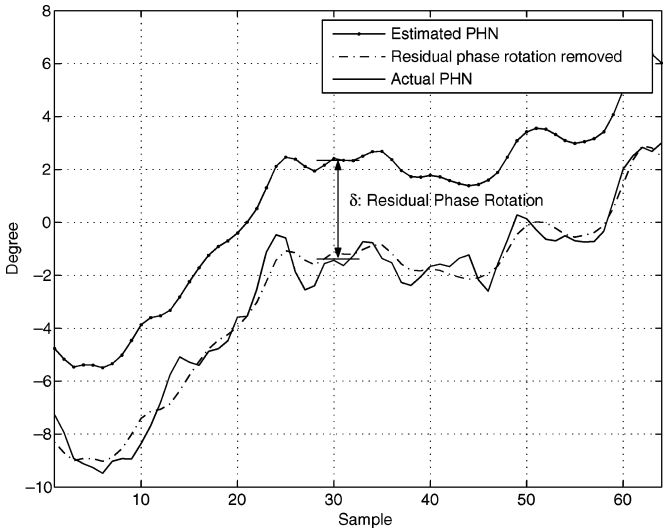


Fig. 3. Effect of residual common phase rotation in JCPCE, SNR = 30 dB.

1) Unresolvable Residual Common Phase Rotation: Fig. 3 plots an instance of the PHN process and its estimate via the JCPCE algorithm. The figure also depicts the peculiar effect of residual common phase rotation at the output of the JCPCE. At SNR = 30 dB, it can be seen that the PHN profile can be estimated accurately but differs from the actual PHN by a constant phase rotation that shifts the estimate towards the zero degree line. This constant rotation creates an equal but opposite rotation in the channel estimate (which is difficult to see graphically). The exact analysis of this *residual common phase rotation* (RCPR) is difficult, but we have a fairly good understanding of its origin, which is summarized in the following proposition and is backed up by further simulations.

Proposition 1: Assume the actual PHN process and channel impulse response are $\boldsymbol{\theta}_o$ and $\mathbf{g}_o$, respectively. As SNR $\rightarrow \infty$, the jointly optimal estimates $\hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{g}}$, calculated using the JCPCE algorithm approach

$$\hat{\boldsymbol{\theta}} \rightarrow \boldsymbol{\theta}_0 + \delta\mathbf{1} \tag{48}$$

$$\hat{\mathbf{g}} \rightarrow e^{-j\delta}\mathbf{g}_o \tag{49}$$

where

$$\delta = \arg\min_\alpha(\boldsymbol{\theta}_0 + \alpha\mathbf{1})^T\mathbf{\Phi}^{-1}(\boldsymbol{\theta}_0 + \alpha\mathbf{1}). \tag{50}$$

Proof: See Appendix III. ■

A simple justification to this proposition is provided in Appendix III. In brief, the RCPR, represented by an unknown constant δ, is introduced to shift the optimal estimate $\boldsymbol{\theta}_o$ such that $\boldsymbol{\theta}_o + \delta\mathbf{1}$ is closer to a zero-mean Gaussian process defined by the covariance matrix $\mathbf{\Phi}$. Thus although the PHN estimate we have obtained is "maximum a posteriori," it is not unbiased.

This proposition not only gives us a qualitative understanding of the phenomenon but also offers quantitative predictions. By

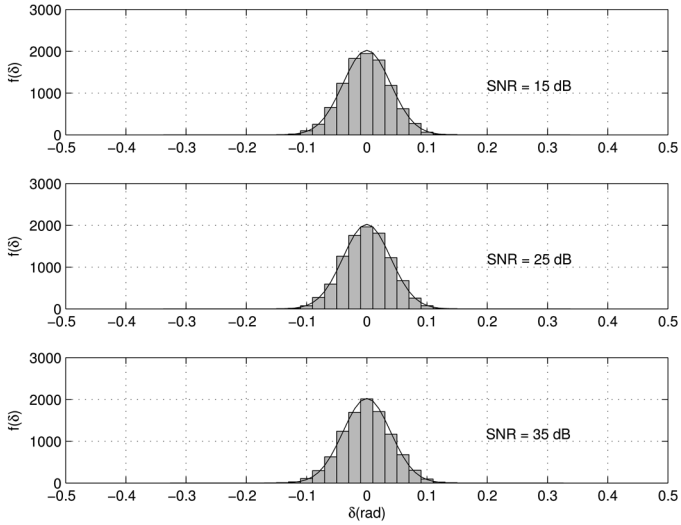


Fig. 4. The predicted pdf versus the histogram of δ at SNR $= 15, 25, 35$ dB.

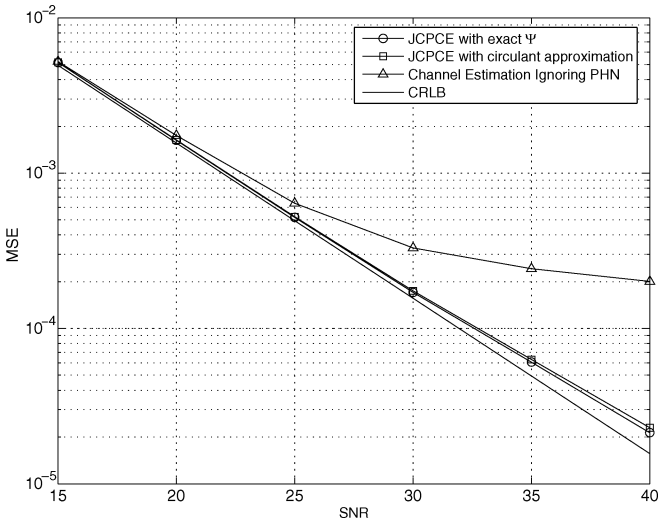


Fig. 5. MSE versus SNR channel estimation performance $(\epsilon = 0)$.

making $\hat{\boldsymbol{\theta}}$ most likely, $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_o + \delta\mathbf{1}$ is approximately zero mean. That implies that δ should be approximately the negative sample mean of $\boldsymbol{\theta}_o$

$$\delta \approx -\frac{1}{N}\boldsymbol{\theta}_o^T\mathbf{1}. \tag{51}$$

Since $p(\boldsymbol{\theta}_o) = \mathcal{N}(\mathbf{0}, \boldsymbol{\Phi})$, it is easy to see that $p(\delta) = \mathcal{N}(0, \mathbf{1}^T\boldsymbol{\Phi}\mathbf{1}/N^2)$.

Fig. 4 shows the probability density function (pdf) of the measured δ compared to the Gaussian prediction, where δ is measured in simulation as the mean difference between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_o$. It is seen that this prediction holds very well at different SNRs. Hence, we now have much better knowledge about the behavior of RCPR and know that it is not significant, as its variance is a fraction of that of PHN.

We call this RCPR "unresolvable" because it cannot be corrected in the channel estimation stage using a likelihood-based estimator. The consequence of RCPR is the rotation of the estimated channel impulse response from $\mathbf{g}_o$ by δ. Left untreated, using this biased channel estimate in the data detection stage in a PHN channel is equivalent to having a perfect channel estimate but an exacerbated PHN. In fact, the equivalent PHN process would be zero-mean Gaussian with a new covariance matrix

$$\boldsymbol{\Phi}_\delta = \boldsymbol{\Phi} + \sigma_\delta^2\mathbf{1}\mathbf{1}^T \tag{52}$$

where $\sigma_\delta^2 = \mathbf{1}^T\boldsymbol{\Phi}\mathbf{1}/N^2$. Our joint PHN/data detection algorithm [20] will then use $\boldsymbol{\Phi}_\delta$ as the PHN covariance matrix at the data detection stage.

Alternatively, δ can also be estimated in the data detection stage using pilot symbols embedded in the transmitted OFDM symbols. Since this is not the subject of this paper, we will not discuss it much further but assume from hereon that δ can be perfectly corrected to facilitate easy assessment of the quality of channel estimation. In particular, in each simulation, δ is set to be the mean difference between the actual PHN process and the estimated PHN process.

2) Channel Estimation Performance: The channel estimation performance of JCPCE is shown in Fig. 5, where the mean squared error (MSE) of estimating $\mathbf{g}/\sqrt{N}$ is plotted against the system SNR (SNR $= E_s/N_o = \rho^2/\sigma^2$). The performance of the proposed channel estimator is compared to the Cramér–Rao lower bound (CRLB) for an OFDM channel without PHN distortion. As calculated in Appendix IV, the CRLB for estimating the impulse response $\mathbf{g} \in \mathbb{C}^{L\times 1}$ is

$$\text{CRLB}(\mathbf{g}) = \frac{L}{\text{SNR}}. \tag{53}$$

It is surprising to see that with the proposed JCPCE, the channel estimate is almost as good as the optimal performance with no PHN. This implies that our channel estimator almost completely cancels the effect of PHN distortion. The channel estimation performance as a result of using $\tilde{\boldsymbol{\Psi}}$ with the circulant approximation (Section III-B) is also plotted, and it is seen that the approximation has very little effect on the performance of the channel estimator. On the other hand, the performance of the conventional channel estimator (which ignores the random PHN but removes the common phase noise) is much poorer than the proposed JCPCE, especially at high SNR. An error floor exists because the ICI created by the PHN causes a constant SNR degradation to the channel estimator.

3) Effect of Number of Iterations of CG Algorithm: The number of iterations in the conjugate gradient method for PHN estimation is a crucial factor in the overall complexity of the low-complexity implementation of JCPCE. Fig. 6 illustrates the performance of the channel estimator at SNR $= 20, 30, 40$ dB as a function of the number of CG iterations in Table II. The plots reveal that, in general, reliable channel estimates can be obtained with five to ten CG iterations. This is due to the superior convergence properties of the conjugate gradient method.

4) Sensitivity to PHN Modelling Error: It is well known that the accuracy of the prior statistics plays an important role in the performance of Bayesian estimators. The Bayesian estimation

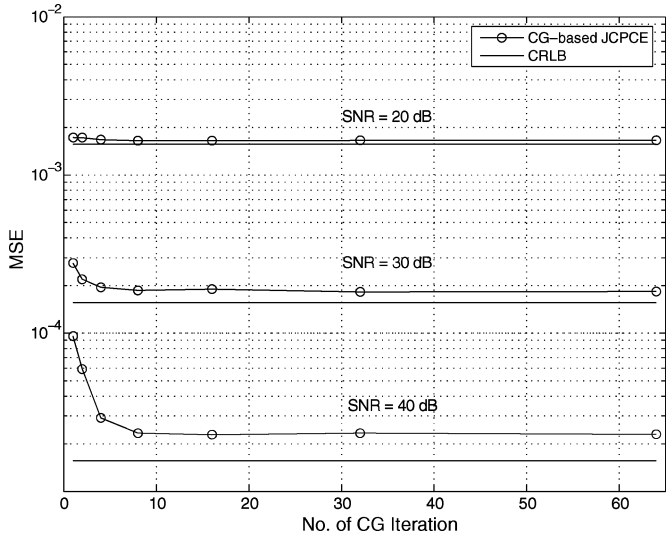


Fig. 6. Channel estimation performance for different number of CG iterations ($\epsilon = 0$).

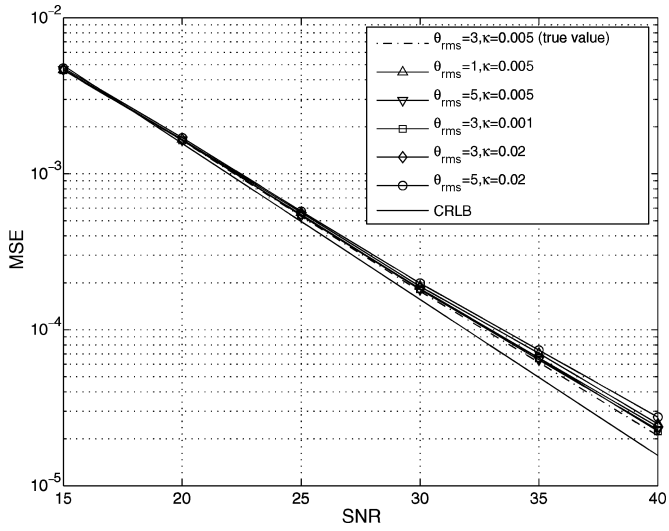


Fig. 7. MSE versus SNR channel estimation performance with PHN modelling error ($\epsilon = 0$).

of the PHN in the JCPCE algorithm is therefore influenced by the accuracy of the prior distribution $p(\boldsymbol{\theta})$ and, more specifically, the covariance matrix $\boldsymbol{\Phi}$. From (47), it is seen that modelling errors may occur in two places: the PHN standard deviation $\theta_{\rm rms}$ and the relative bandwidth $\kappa = \Omega_o/f_s$. In our simulations, the PHN is generated by setting $\theta_{\rm rms} = 3\,{\rm deg}$ and $\kappa = 0.005$. We artificially introduce erroneous PHN statistics in the channel estimator by varying $\boldsymbol{\Phi}$ in Table I over a range of values of $\theta_{\rm rms}$ and κ to test the robustness of JCPCE to inaccurate PHN statistics. Fig. 7 depicts the performance of JCPCE. The result demonstrates that even with significant errors in PHN statistics, JCPCE performs close to the CRLB.

5) Performance at Different Levels of PHN: Fig. 8 studies the channel estimation accuracy as a function of the severity of PHN distortion. We vary the parameters of the PHN generated in the simulations over different settings of $\theta_{\rm rms}$ and κ and assume perfect knowledge of these statistics at the channel estimator (no modelling error). It is seen that even in extreme cases such as

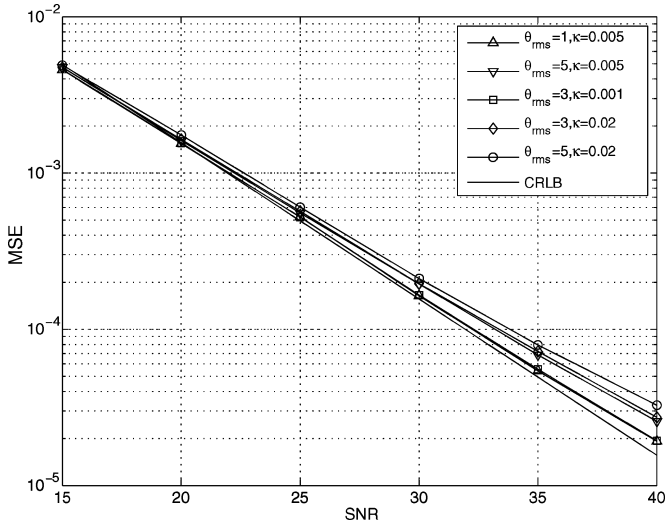


Fig. 8. MSE versus SNR channel estimation performance at different levels of PHN ($\epsilon = 0$).

$\theta_{\rm rms} = 5\,{\rm deg}$ and $\kappa = 0.02$, JCPCE is able to perform close to the CRLB, confirming that the proposed algorithm is robust to very severe PHN in the channel.

B. Channel Estimation With Both CFO and PHN

In this section, we will examine the channel estimation performance in the presence of both CFO and PHN. The allowable CFO estimation range is $|\epsilon| < 0.5$ for the JCPCE algorithm and $|\epsilon| < 1$ for the modified JCPCE algorithm. In the following simulations, the CFO term ϵ will be generated from a uniform distribution in $[-0.4, 0.4]$ corresponding to a maximum CFO of 125 kHz. The PHN generator will use the setting $\theta_{\rm rms} = 3\,{\rm deg}$ and $\kappa = 0.005$. Care should be taken, however, when simulating both CFO and PHN. Now the effective PHN seen by the PHN estimator (Step 2 in both Table I and Table III) is a combination of the residual CFO estimation error (as an additional PHN process with linearly varying phase) and the actual PHN. So δ should now be equal to the mean difference between the effective PHN process and the estimated PHN process and removed before channel estimation.

1) Channel Estimation Performance of JCPCE: Fig. 9 plots the channel estimation MSE as a function of the system SNR in the presence of both CFO and PHN. The complete JCPCE algorithm is compared to the partial JCPCE where PHN estimation is omitted. (We cannot compare with the conventional channel estimator that ignores PHN and CFO because it completely fails when $\epsilon \neq 0$.) It is seen that the complete JCPCE algorithm almost completely cancels the effect of both CFO and PHN distortion. The partial JCPCE, which optimally cancels CFO but ignores PHN, deviates from the CRLB, demonstrating that PHN has a major effect in channel estimation even with optimal CFO estimation.

2) Channel Estimation Performance of Modified JCPCE: Fig. 10 simulates the modified JCPCE (Table III) given repeating training symbols. Here we keep the same simulation settings except for letting the training symbol have a repeating structure. The performance of the modified JCPCE is compared against the CRLB and the conventional channel estimator with

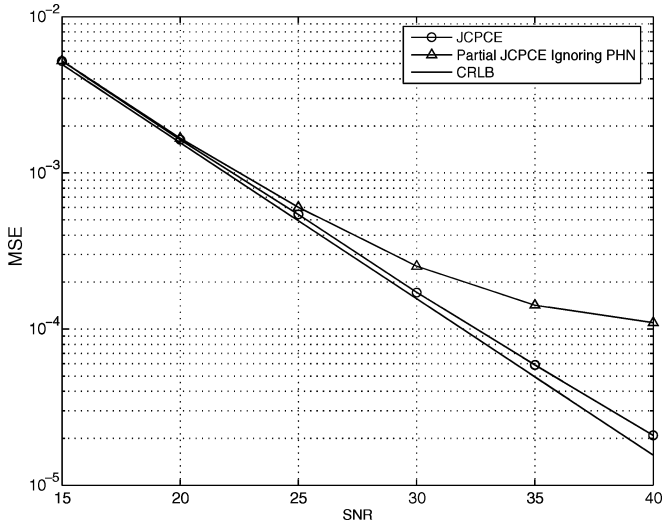


Fig. 9. MSE versus SNR channel estimation performance using JCPCE algorithm ($\epsilon \in [-0.4, 0.4]$).

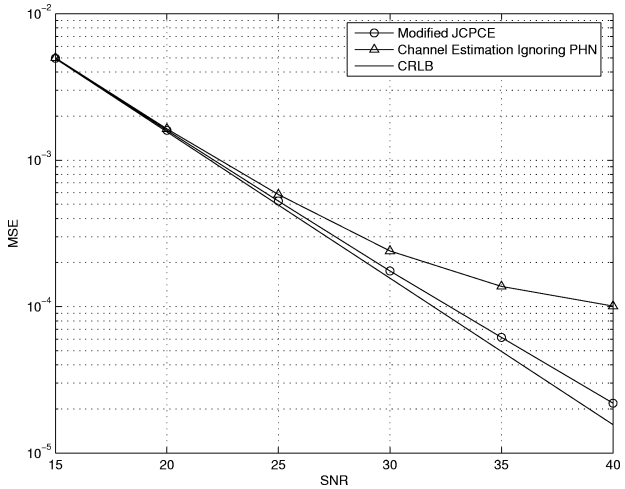


Fig. 10. MSE versus SNR channel estimation performance using modified JCPCE algorithm ($\epsilon \in [-0.4, 0.4]$).

the CFO estimated using Moose's method in (29) and the random PHN ignored. It is seen that the conventional method suffers from an error floor due to the untreated PHN. The modified JCPCE in Fig. 10 has very similar performance to the original JCPCE in Fig. 9, showing that the modified JCPCE, while having much lower complexity, does not perform worse. Therefore, in practice, the modified JCPCE is a preferred scheme over the regular JCPCE.

VI. Conclusions

The problem of channel estimation in a practical OFDM receiver that suffers from phase noise and frequency offset has not been adequately addressed to date. In this paper, we first derived the maximum a posteriori estimator of the channel response, phase noise, and frequency offset, incorporating prior knowledge of the phase noise statistics and using a constant-modulus training sequence. Next, we proposed a less complex estimator, which requires a training symbol that has two identical halves in the time domain. The lower complexity is obtained because an expensive exhaustive search over all feasible frequency offsets is no longer needed. The proposed method's CFO estimate is more accurate than the one from [26] since it is based on an accurate model of the PHN present in the signal.

Furthermore, we explored ways to reduce the complexity of the proposed estimators through the use of the conjugate gradient iteration. It is demonstrated that the channel estimators are able to perform well with a very small number of CG iterations, with each iteration efficiently computed using the FFT. It is evident that the proposed channel estimators can be readily implemented without substantial increase to the overall complexity of conventional OFDM receivers. This paper provides a firm foundation for the design of OFDM detectors in the presence of PHN [20], where the CIR and CFO can now be safely assumed known. The multiple-user (orthogonal frequency-division multiple-access) extension of this paper can be readily carried out through the expectation–maximization (EM) framework introduced in [27] and is currently being studied.

Appendix I
Proof of Lemma 1

From

$$\boldsymbol{\theta}_\Delta = \boldsymbol{\theta}_2 - \boldsymbol{\theta}_1 = [-\mathbf{I} \quad \mathbf{I}] \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix} \tag{54}$$

and

$$p\left(\begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix} \right) = \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Phi}_{N/2} & \boldsymbol{\Upsilon} \\ \boldsymbol{\Upsilon}^T & \boldsymbol{\Phi}_{N/2} \end{bmatrix} \right) \tag{55}$$

we obtain

$$\begin{aligned} p(\boldsymbol{\theta}_\Delta) &= \mathcal{N}\left([-\mathbf{I} \quad \mathbf{I}] \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, [-\mathbf{I} \quad \mathbf{I}] \begin{bmatrix} \boldsymbol{\Phi}_{N/2} & \boldsymbol{\Upsilon} \\ \boldsymbol{\Upsilon}^T & \boldsymbol{\Phi}_{N/2} \end{bmatrix} \begin{bmatrix} -\mathbf{I} \\ \mathbf{I} \end{bmatrix} \right) \\ &= \mathcal{N}(\mathbf{0}, 2\boldsymbol{\Phi}_{N/2} - \boldsymbol{\Upsilon} - \boldsymbol{\Upsilon}^T). \end{aligned} \tag{56}$$

Appendix II
Proof of Lemma 2

Using the iterated expectation theorem [15, ch. 14] and its analog in covariance, given a Gaussian distributed $\mathbf{x}$ and a Gaussian conditional distribution for $\mathbf{y}|\mathbf{x}$, the marginal distribution of $\mathbf{y}$ is also Gaussian and is related to the conditional distribution by

$$\begin{aligned} \mathrm{E}(\mathbf{y}) &= \mathrm{E}_{\mathbf{x}} \mathrm{E}_{\mathbf{y}}(\mathbf{y}|\mathbf{x}) \\ \mathrm{V}(\mathbf{y}) &= \mathrm{V}_{\mathbf{x}} \left(\mathrm{E}_{\mathbf{y}}(\mathbf{y}|\mathbf{x}) \right) + \mathrm{E}_{\mathbf{x}} \left(\mathrm{V}_{\mathbf{y}}(\mathbf{y}|\mathbf{x}) \right). \end{aligned} \tag{57}$$

Applied to the conditional distribution $p(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_\Delta)$, we have

$$\begin{aligned} \mathrm{E}(\mathbf{r}_2|\mathbf{r}_1, \epsilon) &= \mathrm{E}_{\boldsymbol{\theta}_\Delta} \left(\mathrm{E}_{\mathbf{r}_2}(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_\Delta) \right) \\ \mathrm{V}(\mathbf{r}_2|\mathbf{r}_1, \epsilon) &= \mathrm{V}_{\boldsymbol{\theta}_\Delta} \left(\mathrm{E}_{\mathbf{r}_2}(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_\Delta) \right) \\ &\quad + \mathrm{E}_{\boldsymbol{\theta}_\Delta} \left(\mathrm{V}_{\mathbf{r}_2}(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_\Delta) \right). \end{aligned} \tag{58}$$

Also we have

$$\begin{aligned}\mathrm{E}_{\mathbf{r}_2}(\mathbf{r}_2|\mathbf{r}_1,\epsilon,\boldsymbol{\theta}_\Delta) &= e^{j\pi\epsilon}\mathbf{P}_\Delta\mathbf{r}_1 \approx e^{j\pi\epsilon}\mathbf{R}_1(\mathbf{1}+j\boldsymbol{\theta}_\Delta)\\ \mathrm{V}_{\mathbf{r}_2}(\mathbf{r}_2|\mathbf{r}_1,\epsilon,\boldsymbol{\theta}_\Delta) &= 4\sigma^2\mathbf{I}.\end{aligned} \tag{59}$$

Therefore, after simple matrix algebra, we obtain

$$\begin{aligned}\mathrm{E}_{\boldsymbol{\theta}_\Delta}\left(\mathrm{E}_{\mathbf{r}_2}(\mathbf{r}_2|\mathbf{r}_1,\epsilon,\boldsymbol{\theta}_\Delta)\right) &= e^{j\pi\epsilon}\mathbf{r}_1\\ \mathrm{V}_{\boldsymbol{\theta}_\Delta}\left(\mathrm{E}_{\mathbf{r}_2}(\mathbf{r}_2|\mathbf{r}_1,\epsilon,\boldsymbol{\theta}_\Delta)\right) &= \mathbf{R}_1\boldsymbol{\Phi}_\Delta\mathbf{R}_1^H\\ \mathrm{E}_{\boldsymbol{\theta}_\Delta}\left(\mathrm{V}_{\mathbf{r}_2}(\mathbf{r}_2|\mathbf{r}_1,\epsilon,\boldsymbol{\theta}_\Delta)\right) &= 4\sigma^2\mathbf{I}.\end{aligned} \tag{60}$$

We then readily arrive at our final result

$$p(\mathbf{r}_2|\mathbf{r}_1,\epsilon) = \mathcal{CN}\left(e^{j\pi\epsilon}\mathbf{r}_1, \mathbf{R}_1\boldsymbol{\Phi}_\Delta\mathbf{R}_1^H + 4\sigma^2\mathbf{I}\right). \tag{61}$$

Appendix III
Proof of Proposition 1

Consider the minimization of the complete negative log-likelihood function $\mathcal{L}(\boldsymbol{\theta},\mathbf{g})$, where the actual values of the variables are $\mathbf{g}_o$ and $\boldsymbol{\theta}_o$. We examine the joint optimizers of $\mathcal{L}(\boldsymbol{\theta},\mathbf{g})$ as SNR $\rightarrow\infty$ in relation to $\mathbf{g}_o$ and $\boldsymbol{\theta}_o$.

Looking at (10), it is seen that $\mathcal{L}(\boldsymbol{\theta},\mathbf{g})$ has two components, associated with $p(\mathbf{r}|\boldsymbol{\theta},\mathbf{g})$ and $p(\boldsymbol{\theta})$, respectively. Denote

$$\mathcal{L}(\boldsymbol{\theta},\mathbf{g})_{p(\mathbf{r}|\boldsymbol{\theta},\mathbf{g})} = \frac{1}{2\sigma^2}\|\mathbf{r}-\mathbf{P}\mathbf{F}^H\mathbf{D}\mathbf{W}\mathbf{g}\|^2 \tag{62a}$$

$$\mathcal{L}(\boldsymbol{\theta})_{p(\boldsymbol{\theta})} = \frac{1}{2}\boldsymbol{\theta}^T\boldsymbol{\Phi}^{-1}\boldsymbol{\theta}. \tag{62b}$$

As SNR $\rightarrow\infty$ (i.e., $\sigma^2\rightarrow 0$)

$$(\boldsymbol{\theta}_o,\mathbf{g}_o) = \arg\min_{\boldsymbol{\theta},\mathbf{g}}\mathcal{L}_{p(\mathbf{r}|\boldsymbol{\theta},\mathbf{g})}(\boldsymbol{\theta},\mathbf{g}) \tag{63}$$

but the minimizer is not unique, since

$$(\boldsymbol{\theta}_o+\delta\mathbf{1}, e^{-j\delta}\mathbf{g}_o) = \arg\min_{\boldsymbol{\theta},\mathbf{g}}\mathcal{L}_{p(\mathbf{r}|\boldsymbol{\theta},\mathbf{g})}(\boldsymbol{\theta},\mathbf{g}) \tag{64}$$

for arbitrary angle δ. This is because introducing two opposite phase rotations to $\mathbf{u}$ and $\mathbf{g}$ does not alter the overall channel response, and hence the likelihood.

Assume the uniqueness of (64), i.e., $\mathcal{S}_{p(\mathbf{r}|\boldsymbol{\theta},\mathbf{g})} \equiv \{(\boldsymbol{\theta}_o+\delta\mathbf{1}, e^{-j\delta}\mathbf{g}_o)\}$ describes a complete set of optimizers for $\mathcal{L}_{p(\mathbf{r}|\boldsymbol{\theta},\mathbf{g})}(\boldsymbol{\theta},\mathbf{g})$. Notice that any variable pair $(\boldsymbol{\theta},\mathbf{g})\in\mathcal{S}_{p(\mathbf{r}|\boldsymbol{\theta},\mathbf{g})}$ makes $\mathcal{L}_{p(\mathbf{r}|\boldsymbol{\theta},\mathbf{g})}(\boldsymbol{\theta},\mathbf{g}) = 0$, or $p(\mathbf{r}|\boldsymbol{\theta},\mathbf{g}) = \infty$. It then follows that the optimizer $(\hat{\boldsymbol{\theta}},\hat{\mathbf{g}})$ of $\mathcal{L}(\boldsymbol{\theta},\mathbf{g})$ must be a subset of $\mathcal{S}_{p(\mathbf{r}|\boldsymbol{\theta},\mathbf{g})}$, as any other pair $(\boldsymbol{\theta},\mathbf{g})$ would make the complete likelihood finite. Consequently, the only task remaining is to find $(\hat{\boldsymbol{\theta}},\hat{\mathbf{g}}) = \arg\min_{\boldsymbol{\theta},\mathbf{g}}\mathcal{L}(\boldsymbol{\theta})_{p(\boldsymbol{\theta})}$ subject to $(\hat{\boldsymbol{\theta}},\hat{\mathbf{g}})\in\mathcal{S}_{p(\mathbf{r}|\boldsymbol{\theta},\mathbf{g})}$.

Appendix IV
CRLB for OFDM Channel Estimation

In the absence of CFO and PHN, the received signal in an OFDM channel can be written, similar to (9), as

$$\mathbf{r} = \mathbf{F}^H\mathbf{D}\mathbf{W}\mathbf{g} + \mathbf{n}. \tag{65}$$

Thus

$$p(\mathbf{r}|\mathbf{g}) = \mathcal{CN}(\mathbf{F}^H\mathbf{D}\mathbf{W}\mathbf{g}, 2\sigma^2\mathbf{I}) \tag{66}$$

or equivalently

$$\log p(\mathbf{r}|\mathbf{g}) = -\frac{1}{2\sigma^2}(\mathbf{F}^H\mathbf{D}\mathbf{W}\mathbf{g}-\mathbf{r})^H(\mathbf{F}^H\mathbf{D}\mathbf{W}\mathbf{g}-\mathbf{r}). \tag{67}$$

Taking the derivative with respect to $\mathbf{g}^*$, we have

$$\begin{aligned}\frac{\partial}{\partial\mathbf{g}^*}\{\log p(\mathbf{r}|\mathbf{g})\} &= -\frac{1}{2\sigma^2}\mathbf{W}^H\mathbf{D}^H\mathbf{F}(\mathbf{F}^H\mathbf{D}\mathbf{W}\mathbf{g}-\mathbf{r})\\ &= \frac{1}{2\sigma^2}\mathbf{W}^H\mathbf{D}^H\mathbf{F}\mathbf{n}.\end{aligned} \tag{68}$$

From [28], the Fisher information matrix can be evaluated

$$\begin{aligned}\mathbf{I}(\mathbf{g}) &= \mathrm{E}\left\{\left[\frac{\partial}{\partial\mathbf{g}^*}\log p(\mathbf{r}|\mathbf{g})\right]\left[\frac{\partial}{\partial\mathbf{g}^*}\log p(\mathbf{r}|\mathbf{g})\right]^H\right\}\\ &= \mathrm{E}\left\{\frac{1}{(2\sigma^2)^2}\mathbf{W}^H\mathbf{D}^H\mathbf{F}\mathbf{n}\mathbf{n}^H\mathbf{F}^H\mathbf{D}\mathbf{W}\right\}\\ &= \frac{1}{(2\sigma^2)^2}\mathbf{W}^H\mathbf{D}^H\mathbf{F}\mathrm{E}\{\mathbf{n}\mathbf{n}^H\}\mathbf{F}^H\mathbf{D}\mathbf{W}\\ &= \frac{1}{2\sigma^2}\mathbf{W}^H\mathbf{D}^H\mathbf{D}\mathbf{W}\\ &= \frac{\rho^2}{\sigma^2}\mathbf{I}\end{aligned} \tag{69}$$

where the last equality follows from the constant modulus assumption of the training data, i.e., $\mathbf{D}^H\mathbf{D} = 2\rho^2\mathbf{I}$. Specifically in this simulation, QPSK training symbols are used, i.e., $\mathbf{d} = \rho\times[\{\pm1\}+j\{\pm1\}]^{N\times1}$.

The CRLB is therefore

$$\begin{aligned}\mathrm{CRLB}(\mathbf{g}) &= \mathrm{tr}\left[\mathbf{I}^{-1}(\mathbf{g})\right]\\ &= \frac{L\sigma^2}{\rho^2}\\ &= \frac{L}{\mathrm{SNR}}\end{aligned} \tag{70}$$

where $\mathrm{SNR} = E_s/N_o = 2\rho^2/2\sigma^2$.

References

[1] R. E. Ziemer and R. L. Peterson, *Introduction to Digital Communication*. Upper Saddle River, NJ: Prentice-Hall, 2001, ch. 4.10.

[2] L. Piazzo and P. Mandarini, "Analysis of phase noise effects in OFDM modems," *IEEE Trans. Commun.*, vol. 50, pp. 1696–1705, Oct. 2002.

[3] L. Zhao and W. Namgoong, "A novel adpative phase noise compensation scheme for communication receivers," in *Proc. IEEE GLOBECOM'03*, Dec. 2003, vol. 4, pp. 2274–2279.

[4] K. Nikitopoulos and A. Polydoros, "Decision-directed compensation of phase noise and residual frequency offset in a space-time OFDM receiver," *IEEE Commun. Lett.*, vol. 8, pp. 573–575, Sep. 2004.

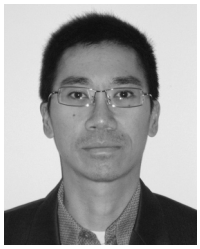
[5] G. Liu and W. Zhu, "Compensation of phase noise in OFDM systems using an ICI reduction scheme," *IEEE Trans. Broadcast.*, vol. 50, pp. 399–407, Dec. 2004.
[6] K. Nikitopoulos and A. Polydoros, "Phase-impairment effects and compensation algorithms for OFDM systems," *IEEE Trans. Commun.*, vol. 53, pp. 698–707, Apr. 2005.
[7] J. Tubbax, B. Côme, L. V. der Perre, S. Donnay, M. Engels, H. D. Man, and M. Moonen, "Compensation of IQ imbalance and phase noise in OFDM systems," *IEEE Trans. Wireless Commun.*, no. 3, pp. 872–877, May 2005.
[8] A. Mehrotra, "Noise analysis of phase-locked loops," *IEEE Trans. Circuits Syst.*, vol. 49, pp. 1309–1315, Sep. 2002.
[9] S. Wu and Y. Bar-Ness, "OFDM systems in the presence of phase noise: Consequences and solutions," *IEEE Trans. Commun.*, vol. 52, pp. 1988–1996, Nov. 2004.
[10] K. Nikitopoulos and A. Polydoros, "Compensation schemes for phase noise and residual frequency offset in OFDM systems," in *Proc. IEEE GLOBECOM'01*, Nov. 2001, vol. 1, pp. 330–333.
[11] S. Wu and Y. Bar-Ness, "OFDM channel estimation in the presence of frequency offset and phase noise," in *Proc. IEEE Int. Conf. Communications (ICC) 2003*, May 2003, vol. 5, pp. 3366–3370.
[12] M. S. El-Tanany, Y. Wu, and L. Házy, "Analytical modeling and simulation of phase noise interference in OFDM-based digital television terrestrial broadcasting systems," *IEEE Trans. Broadcast.*, vol. 47, pp. 20–31, Mar. 2001.
[13] R. A. Casas, S. L. Biracree, and A. E. Youtz, "Time domain phase noise correction for OFDM signals," *IEEE Trans. Broadcast.*, vol. 48, pp. 230–236, Sep. 2002.
[14] A. Scaglione, S. Barbarossa, and G. B. Giannakis, "Robust OFDM transmissions over frequency-selective channels with multiplicative time-selective effects," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP) 2000*, Jun. 2000, vol. 5, pp. 2677–2680.
[15] M. I. Jordan, *An Introduction to Probabilistic Graphical Models* (in preparation).
[16] D. D. Lin, R. A. Pacheco, T. J. Lim, and D. Hatzinakos, "Optimal OFDM channel estimation with carrier frequency offset and phase noise," in *Proc. IEEE Wireless Communications Networking Conf. (WCNC) 2006*, Apr. 2006.
[17] D. D. Lin, R. A. Pacheco, T. J. Lim, and D. Hatzinakos, "Near-optimal training-based estimation of frequency offset and channel response in OFDM with phase noise," in *Proc. IEEE Int. Conf. Communications (ICC) 2006*, Jun. 2005.
[18] P. Chevillat, D. Maiwald, and G. Ungerboeck, "Rapid training of a voiceband data-modem receiver employing an equalizer with fractional-T spaced coefficients," *IEEE Trans. Commun.*, vol. 35, pp. 869–876, Sep. 1987.
[19] T. Schmidl and D. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Trans. Commun.*, vol. 45, pp. 1613–1621, Dec. 1997.
[20] D. D. Lin, Y. Zhao, and T. J. Lim, "OFDM phase noise cancellation via approximate probabilistic inference," in *Proc. IEEE Wireless Communications Networking Conf. (WCNC) 2005*, Mar. 2005, vol. 1, pp. 27–32.
[21] B. Chen, "Maximum likelihood estimation of OFDM carrier frequency offset," *IEEE Signal Process. Lett.*, vol. 9, no. 4, pp. 123–126, Apr. 2002.
[22] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer-Verlag, 1999.
[23] C. R. Vogel, *Computational Methods for Inverse Problems*. Philadelphia, PA: SIAM, 2002.
[24] T. F. Chan, "An optimal circulant preconditioner for Toeplitz systems," *SIAM J. Sci. Stat. Comput.*, vol. 9, no. 4, pp. 766–771, July 1988.
[25] *Phase Noise Matlab Model*, IEEE P802.11-Task Group G, 2000 [Online]. Available: http://grouper.ieee.org/groups/802/11/Reports/tgg_update.htm
[26] P. H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Trans. Commun.*, vol. 42, pp. 2908–2914, Oct. 1994.
[27] M.-O. Pun, S.-H. Tsai, and C.-C. J. Kuo, "An EM-based joint maximum likelihood estimation of carrier frequency offset and channel for uplink OFDMA systems," in *Proc. IEEE Vehicular Technology Conf. (VTC) 2004—Fall*, Sep. 2004, vol. 1, pp. 598–602.
[28] S. M. Kay, *Fundamentals of Statistical Signal Processing—Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993, ch. 10.7, pp. 328–330.

Darryl Dexu Lin (S'01) received the B.A.Sc. degree (with honors) in engineering science and the M.A.Sc. degree in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 2001 and 2003, respectively, where he is currently pursuing the Ph.D. degree.

His research interests include statistical inference and its applications in CDMA multiuser detection and multicarrier modulation.

Ryan A. Pacheco (S'98) is currently pursuing the Ph.D. degree at the University of Toronto, Toronto, ON, Canada.

His main contributions have been in CDMA and OFDM receiver design. In CDMA, his work focused on semiblind schemes for interference suppression in multiuser frequency-selective channels. In OFDM, his work has been in three main areas: frame synchronization (802.11a WLANs), phase-noise cancellation, and power efficient transmission (using angle modulation).



Teng Joon Lim (S'92–M'95–SM'02) received the B.Eng. degree from the National University of Singapore in 1992 and the Ph.D. degree from the University of Cambridge, U.K., in 1996.

He has been an Assistant Professor at the Department of Electrical and Computer Engineering of the University of Toronto since December 2000, where he leads the Wireless Multiple Access research group. In the five years prior to that, he was a Member of Technical Staff with the Centre for Wireless Communications in Singapore, serving as the leader of the digital communications and signal processing group. His research interests are in wireless transceiver design, in particular multiuser detection, OFDM and OFDMA receiver structures, MIMO techniques, precoding in downlink channels, and cross-layer aspects of cooperative network design, and he has published widely in these areas.

Prof. Lim contributes regularly in organizing conferences, serving on technical program committees, and organizing seminars for the IEEE Toronto Communications Chapter.

Dimitrios Hatzinakos (SM'98) received the diploma degree from the University of Thessaloniki, Greece, in 1983, the M.A.Sc. degree from the University of Ottawa, Ottawa, ON, Canada, in 1986, and the Ph.D. degree from Northeastern University, Boston, MA, in 1990, all in electrical engineering.

In 1990, he joined the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, where he now is a tenured Professor. He was Chair of the Communications Group of the department during July 1999–June 2004. Since November 2004, he has held the Bell Canada Chair in Multimedia at the University of Toronto. His research interests are in the areas of multimedia signal processsing and communications. He is author/coauthor of more than 150 papers in technical journals and conference proceedings and has contributed to eight books in his areas of interest. His experience includes consulting through Electrical Engineering Consociates Ltd. and contracts with United Signals and Systems Inc., Burns and Fry Ltd., Pipetronix Ltd., Defense Research Establishment Ottawa (DREO), Vaytek Inc., Nortel Networks, Vivosonic Inc., and CANAMET Inc. He was Guest Editor for *Signal Processing* special issue on "Signal Processing Technologies for Short Burst Wireless Communications" (October 2000).

Prof. Hatzinakos is a member of EURASIP, the Professional Engineers of Ontario, and the Technical Chamber of Greece. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1998 till 2002. He was a Member of the IEEE Statistical Signal and Array Processing Technical Committee from 1992 to 1995 and Technical Program Cochair of the Fifth Workshop on Higher-Order Statistics in July 1997.