Near-Optimal Training-Based Estimation of Frequency Offset and Channel Response in OFDM with Phase Noise

Darryl Dexu Lin, Ryan A. Pacheco, Teng Joon Lim and Dimitrios Hatzinakos
The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto
10 King's College Road, Toronto, Ontario, Canada M5S 3G4
{linde, ryan, limtj, dimitris}@comm.utoronto.ca

***Abstract*— We propose an efficient training-based OFDM channel impulse response (CIR) and carrier frequency offset (CFO) estimation algorithm that addresses the problem of phase noise (PHN), assuming that the PHN has a known prior distribution. The optimal joint estimation of CIR, PHN and CFO was described in an earlier work of ours. In this paper, we focus on the case where a training symbol consists of two identical halves in the time domain, and propose a variant to Moose's CFO estimation algorithm that accounts for PHN in CFO estimation. This is followed by an optimal joint CIR and PHN estimation scheme tailored for this "repeating training symbol" setup. It is assumed that the PHN process is Gaussian with known mean and covariance matrix. This encompasses both Wiener PHN and Gaussian PHN. It is shown through simulations that the proposed algorithm performs almost as well as the optimal JCPCE algorithm at much lower complexity. To further reduce the complexity of the proposed scheme, the conjugate gradient (CG) method is used and we show that it can be realized using the Fast Fourier Transform (FFT).**

I. Introduction

Orthogonal frequency division multiplexing (OFDM) is a well-known modulation technique 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 (DSL) 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 inter-symbol 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.

Various CFO estimation schemes for OFDM have been studied in the past without considering the presence of PHN. In [1], a MUSIC-based blind CFO estimator was proposed. It was later shown to be a maximum likelihood estimator [2]. In [3], a CFO estimator of much lower complexity was introduced which uses repeating training symbols (i.e. a training symbol with two identical halves in time domain).

To our knowledge, the optimal joint estimation of CFO and CIR in the presence of PHN was derived for the first time in [4], where the "complete likelihood function" of the CIR, CFO and PHN was maximized. This approach yields maximum a posteriori (MAP) joint estimates of CFO, PHN and CIR with uniform priors for CFO and CIR, and an informative prior for PHN which can be obtained from measurements or hardware specifications. However, the proposed optimal estimator suffers from the same drawback as [1] and [2] in that the CFO estimation stage requires a search operation over a range of candidate values, resulting in excessive complexity for practical implementation.

In this paper, our goal is to achieve near-optimal CFO and channel estimation performance with less complexity. Motivated by [3], we will assume that the training symbols for channel estimation have a repeating structure. Furthermore, instead of a single joint optimization routine, we divide the task into two steps. We first perform CFO estimation by optimally canceling the effect of PHN. With the CFO estimated and removed, we then jointly estimate the CIR and PHN. The notations used in this paper will follow those in [4].

II. Signal Model

We assume a slow fading frequency-selective channel where the CIR remains 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.

Assuming perfect timing synchronization and a cyclic prefix that is longer than the CIR, the complex baseband received signal of an OFDM symbol within the training period sampled at rate N/T can be written as a length-N vector, after the removal of the cyclic prefix:

$$\mathbf{r} = \mathbf{EPGF}^H\mathbf{d} + \mathbf{n}. \tag{1}$$

In (1), $\mathbf{E} = \text{diag}([1, e^{j2\pi\epsilon/N}, \cdots, e^{j2\pi(N-1)\epsilon/N}]^T)$ is the CFO matrix, where $\epsilon = \Delta f T$ is the normalized CFO. $\mathbf{P} = \text{diag}([e^{j\theta_0}, \cdots, e^{j\theta_{N-1}}]^T)$ is the PHN matrix, in which

1-4244-0355-3/06/$20.00 (c) 2006 IEEE

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

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

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

the discrete-time PHN process $\boldsymbol{\theta} = [\theta_0, \cdots, \theta_{N-1}]^T$ has a multivariate Gaussian distribution $p(\boldsymbol{\theta}) = \mathcal{N}(\mathbf{0}, \mathbf{\Phi})$. The covariance matrix $\mathbf{\Phi}$ can be determined *a priori* from the power spectral density (PSD) of the voltage controlled oscillator (VCO) output as discussed in [4]. In particular, the PHN process from a free running oscillator may be modeled as a nonstationary Wiener process (referred to as *Wiener PHN*). And the PHN process from a phase-locked oscillator may be modeled as a wide-sense-stationary coloured Gaussian process (referred to as *Gaussian PHN*). Using $\mathbf{g} = [g_0, \cdots, g_{L-1}]^T$ to denote the CIR, where L is the channel length, the channel convolution matrix $\mathbf{G}$ is formed by circular rotations of $\frac{1}{\sqrt{N}}\mathbf{g}$. $\mathbf{F} \in \mathbb{C}^{N\times N}$ is the DFT matrix with the (l,m)th element being $\mathbf{F}_{l,m} = \frac{1}{\sqrt{N}}e^{-j\frac{2\pi(l-1)(m-1)}{N}}$, and $\mathbf{d}$ is a data vector containing constant modulus training symbols. $\mathbf{n}$ is complex white Gaussian noise with variance σ^2 per dimension, i.e. $p(\mathbf{n}) = \mathcal{CN}(\mathbf{0}, 2\sigma^2\mathbf{I})$.

Letting $\mathbf{D} = \mathrm{diag}(\mathbf{d})$, we have $\mathbf{D}^H\mathbf{D} = 2\rho^2\mathbf{I}$, where $E_s = 2\rho^2$ is the symbol energy per subcarrier. We can now introduce the following equivalent representation of (1) for the convenience of channel estimation:

$$\mathbf{r} = \mathbf{EPF}^H\mathbf{DWg} + \mathbf{n}, \tag{2}$$

where $\mathbf{W}$ 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 $\mathbf{V} \in \mathbb{C}^{N\times(N-L)}$.

III. JOINT CFO/PHN/CIR ESTIMATOR (JCPCE)

In [4], we derived the optimal Joint CFO/PHN/CIR Estimator (JCPCE) that optimizes the "complete likelihood function" $p(\mathbf{r}, \epsilon, \boldsymbol{\theta}, \mathbf{g}) = p(\mathbf{r}|\epsilon, \boldsymbol{\theta}, \mathbf{g})p(\epsilon)p(\boldsymbol{\theta})p(\mathbf{g})$. Since we assume no prior knowledge of ϵ and $\mathbf{g}$, $p(\epsilon)$ and $p(\mathbf{g})$ are constants. But the prior distribution of $\boldsymbol{\theta}$ is available, and takes the form $p(\boldsymbol{\theta}) = \mathcal{N}(\mathbf{0}, \mathbf{\Phi})$. Therefore, the "complete negative log-likehood function" can be written as

$$\begin{aligned}\mathcal{L}(\epsilon, \boldsymbol{\theta}, \mathbf{g}) = &\ -\log p(\mathbf{r}|\epsilon, \boldsymbol{\theta}, \mathbf{g}) - \log p(\boldsymbol{\theta}) \\ = &\ \tfrac{1}{2\sigma^2}(\mathbf{r} - \mathbf{EPF}^H\mathbf{DWg})^H(\mathbf{r} - \mathbf{EPF}^H\mathbf{DWg}) \\ &+ \tfrac{1}{2}\boldsymbol{\theta}^T\mathbf{\Phi}^{-1}\boldsymbol{\theta}.\end{aligned} \tag{3}$$

Our objective is to find the optimal estimates

$$(\hat{\epsilon}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{g}}) = \arg\min_{\epsilon, \boldsymbol{\theta}, \mathbf{g}} \mathcal{L}(\epsilon, \boldsymbol{\theta}, \mathbf{g}). \tag{4}$$

Table I summarizes the JCPCE algorithm, in which the joint optimizers $\hat{\epsilon}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\mathbf{g}}$ may be found in three simple steps.

IV. CFO ESTIMATION BASED ON REPEATING TRAINING SYMBOLS

A crucial drawback of JCPCE is that $\hat{\epsilon}$ can only be found by searching over a range of candidate values, which implies complexity that scales with the inverse of the resolution required for ϵ. For cases where the system has limited computational power, it is beneficial to obtain a closed form solution for $\hat{\epsilon}$. When no PHN is present, the pioneering work of Moose [3] achieves just that by assuming the two halves of a training symbol are identical.

A. Moose's CFO Estimator

In the Moose algorithm, we transmit an OFDM symbol with two identical halves in the time domain. Such a signal is easily generated [5] by transmitting $N/2$ training symbols $d_0, \cdots, d_{N/2-1}$ on the even sub-carriers, and zeros on the odd sub-carriers. The N-point sequence in time at the receiver, with CFO and PHN distortion, can be written as

$$r_n = \frac{1}{\sqrt{N/2}}e^{j(\theta_n + 2\pi\epsilon n/N)}\sum_{k=0}^{N/2-1} h_k d_k e^{j4\pi nk/N} + \eta_n, \tag{5}$$

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

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

$$\begin{aligned}\mathbf{r}_1 &= \mathbf{x} + \mathbf{n}_1 \\ \mathbf{r}_2 &= e^{j\pi\epsilon}\mathbf{x} + \mathbf{n}_2,\end{aligned} \tag{6}$$

where $\mathbf{x} = \mathbf{EGF}^H\mathbf{d} \in \mathbb{C}^{\frac{N}{2}\times 1}$, and $\mathbf{F}^H\mathbf{d} \in \mathbb{C}^{\frac{N}{2}\times 1}$ 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 ML estimate of ϵ is

$$\begin{aligned}\hat{\epsilon} &= \arg\max_{\epsilon} p(\mathbf{r}_1, \mathbf{r}_2|\epsilon) \\ &= \arg\max_{\epsilon} p(\mathbf{r}_2|\epsilon, \mathbf{r}_1)p(\mathbf{r}_1|\epsilon),\end{aligned} \tag{7}$$

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

$$\begin{aligned}\mathbf{r}_2 &= e^{j\pi\epsilon}\mathbf{r}_1 - e^{j\pi\epsilon}\mathbf{n}_1 + \mathbf{n}_2 \\ &= e^{j\pi\epsilon}\mathbf{r}_1 + \mathbf{z},\end{aligned} \tag{8}$$

where $p(\mathbf{z}) = \mathcal{CN}(\mathbf{0}, 4\sigma^2\mathbf{I})$. 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

$$-\log p(\mathbf{r}_2|\epsilon, \mathbf{r}_1) = \frac{1}{4\sigma^2}(\mathbf{r}_2 - e^{j\pi\epsilon}\mathbf{r}_1)^H(\mathbf{r}_2 - e^{j\pi\epsilon}\mathbf{r}_1). \tag{9}$$

And it follows that

$$\hat{\epsilon} = \frac{1}{\pi}\measuredangle\mathbf{r}_1^H\mathbf{r}_2. \tag{10}$$

B. CFO Estimator with PHN Rejection

In the presence of PHN, the derivation presented above fails because (8) no longer holds. We propose, in the following, a

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

CFO estimation algorithm that optimally cancels the effect of PHN, as an alternative to Step 1 of the JCPCE algorithm in Table I.

Rewriting (6) to include the PHN distortion, we have

$$\begin{aligned} \mathbf{r}_1 &= \mathbf{P}_1\mathbf{x} + \mathbf{n}_1 \\ \mathbf{r}_2 &= e^{j\pi\epsilon}\mathbf{P}_2\mathbf{x} + \mathbf{n}_2, \end{aligned} \tag{11}$$

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

$$\begin{aligned} \hat{\epsilon} &= \arg\max_\epsilon p(\mathbf{r}_1, \mathbf{r}_2|\epsilon) \\ &= \arg\max_\epsilon \int_{\boldsymbol{\theta}_2}\int_{\boldsymbol{\theta}_1} p(\mathbf{r}_1, \mathbf{r}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\epsilon) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2 \end{aligned} \tag{12}$$

where

$$\begin{aligned} & p(\mathbf{r}_1, \mathbf{r}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\epsilon) \\ = \; & p(\mathbf{r}_1, \mathbf{r}_2|\epsilon, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\ = \; & p(\mathbf{r}_2|, \mathbf{r}_1, \epsilon, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) p(\mathbf{r}_1|\epsilon, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2). \end{aligned} \tag{13}$$

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

$$\hat{\epsilon} = \arg\max_\epsilon \int_{\boldsymbol{\theta}_2}\int_{\boldsymbol{\theta}_1} p(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2. \tag{14}$$

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

$$\begin{aligned} \mathbf{r}_2 &= e^{j\pi\epsilon}\mathbf{P}_\triangle\mathbf{r}_1 - e^{j\pi\epsilon}\mathbf{P}_\triangle\mathbf{n}_1 + \mathbf{n}_2 \\ &= e^{j\pi\epsilon}\mathbf{P}_\triangle\mathbf{r}_1 + \mathbf{z}, \end{aligned} \tag{15}$$

where $p(\mathbf{z}) = \mathcal{CN}(\mathbf{0}, 4\sigma^2\mathbf{I})$. In other words, $p(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathcal{CN}(e^{j2\pi\epsilon}\mathbf{P}_\triangle\mathbf{r}_1, 4\sigma^2\mathbf{I})$. This means that $p(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is only a function of $\boldsymbol{\theta}_\triangle$ instead of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ individually. We may therefore rewrite (14) as

$$\begin{aligned} \hat{\epsilon} &= \arg\max_\epsilon \int_{\boldsymbol{\theta}_\triangle} p(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_\triangle) p(\boldsymbol{\theta}_\triangle) d\boldsymbol{\theta}_\triangle \\ &= \arg\max_\epsilon p(\mathbf{r}_2|\mathbf{r}_1, \epsilon). \end{aligned} \tag{16}$$

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 $\frac{N}{2} \times \frac{N}{2}$ blocks:

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Omega}_1 & \boldsymbol{\Upsilon} \\ \boldsymbol{\Upsilon}^T & \boldsymbol{\Omega}_2 \end{bmatrix}, \tag{17}$$

then $\boldsymbol{\theta}_\triangle = \boldsymbol{\theta}_2 - \boldsymbol{\theta}_1 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_2 - \boldsymbol{\Upsilon} - \boldsymbol{\Upsilon}^T)$.

Proof: The result follows from the linear transformation of the Gaussian random vector $[\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T]^T$. ■

Denoting $\boldsymbol{\Phi}_\triangle \doteq \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_2 - \boldsymbol{\Upsilon} - \boldsymbol{\Upsilon}^T$, we may write $p(\boldsymbol{\theta}_\triangle) = \mathcal{N}(\mathbf{0}, \boldsymbol{\Phi}_\triangle)$. Finally, we use the following lemma to evaluate $p(\mathbf{r}_2|\mathbf{r}_1, \epsilon)$ in (16).

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

$$p(\mathbf{r}_2|\mathbf{r}_1, \epsilon) = \mathcal{CN}(e^{j\pi\epsilon}\mathbf{r}_1, \mathbf{R}_1\boldsymbol{\Phi}_\triangle\mathbf{R}_1^H + 4\sigma^2\mathbf{I}), \tag{18}$$

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

Proof: See Appendix I. ■

From (16) and Lemma 2, we see that

$$\begin{aligned} \hat{\epsilon} &= \arg\max_\epsilon \log p(\mathbf{r}_2|\mathbf{r}_1, \epsilon) \\ &= \arg\min_\epsilon (\mathbf{r}_2 - e^{j\pi\epsilon}\mathbf{r}_1)^H \\ &\quad \times (\mathbf{R}_1\boldsymbol{\Phi}_\triangle\mathbf{R}_1^H + 4\sigma^2\mathbf{I})^{-1}(\mathbf{r}_2 - e^{j\pi\epsilon}\mathbf{r}_1) \\ &= \tfrac{1}{\pi}\measuredangle\, \mathbf{r}_1^H(\mathbf{R}_1\boldsymbol{\Phi}_\triangle\mathbf{R}_1^H + 4\sigma^2\mathbf{I})^{-1}\mathbf{r}_2. \end{aligned} \tag{19}$$

This expression has a very similar correlation form compared to (10) except for a weighting matrix that accounts for the distortion caused by PHN.

C. Joint PHN and CIR Estimation

With the CFO estimated using (19), 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 re-derive the remaining steps of JCPCE.

Expressing (5) in the matrix form yields

$$\mathbf{r} = \hat{\mathbf{E}}\mathbf{P}\check{\mathbf{F}}^H\mathbf{D}\mathbf{W}\mathbf{g} + \mathbf{n}, \tag{20}$$

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; $\mathbf{D} = \mathrm{diag}(\mathbf{d}) \in \mathbb{C}^{N/2\times N/2}$ contains the length-$N/2$ training symbol; $\mathbf{g} \in \mathbb{C}^{L\times 1}$ is the channel impulse response.

Similar to (3), we obtain the "complete negative log-likelihood function":

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}, \mathbf{g}) &= -\log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{g}) - \log p(\boldsymbol{\theta}) \\ &= \tfrac{1}{2\sigma^2}(\mathbf{r} - \hat{\mathbf{E}}\mathbf{P}\check{\mathbf{F}}^H\mathbf{D}\mathbf{W}\mathbf{g})^H(\mathbf{r} - \hat{\mathbf{E}}\mathbf{P}\check{\mathbf{F}}^H\mathbf{D}\mathbf{W}\mathbf{g}) \\ &\quad + \tfrac{1}{2}\boldsymbol{\theta}^T\boldsymbol{\Phi}^{-1}\boldsymbol{\theta}. \end{aligned} \tag{21}$$

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

$$\hat{\mathbf{g}} = (4\rho^2)^{-1}\mathbf{W}^H\mathbf{D}^H\check{\mathbf{F}}\mathbf{P}^H\hat{\mathbf{E}}^H\mathbf{r}. \tag{22}$$

Noticing that

$$\begin{aligned} & (\mathbf{r} - \hat{\mathbf{E}}\mathbf{P}\check{\mathbf{F}}^H\mathbf{D}\mathbf{W}\hat{\mathbf{g}})^H(\mathbf{r} - \hat{\mathbf{E}}\mathbf{P}\check{\mathbf{F}}^H\mathbf{D}\mathbf{W}\hat{\mathbf{g}}) \\ = \; & \mathbf{r}^H\mathbf{r} - (4\rho^2)^{-1}\mathbf{r}\hat{\mathbf{E}}\mathbf{P}\check{\mathbf{F}}^H\mathbf{D}\mathbf{W}\mathbf{W}^H\mathbf{D}^H\check{\mathbf{F}}\mathbf{P}^H\hat{\mathbf{E}}^H\mathbf{r} \\ = \; & \tfrac{1}{4\rho^2}\mathbf{r}^H\hat{\mathbf{E}}\mathbf{P}\begin{bmatrix} \mathbf{F}^H\mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}^H\mathbf{D} \end{bmatrix}\begin{bmatrix} \mathbf{I} + \mathbf{V}\mathbf{V}^H & -\mathbf{W}\mathbf{W}^H \\ -\mathbf{W}\mathbf{W}^H & \mathbf{I} + \mathbf{V}\mathbf{V}^H \end{bmatrix} \\ & \times \begin{bmatrix} \mathbf{F}^H\mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}^H\mathbf{D} \end{bmatrix}^H \mathbf{P}^H\hat{\mathbf{E}}^H\mathbf{r}, \end{aligned} \tag{23}$$

and substituting (23) into (21), we have after simplification

$$\mathcal{L}(\boldsymbol{\theta}) = \tfrac{1}{8\sigma^2\rho^2}\mathbf{u}^T\hat{\mathbf{E}}\mathbf{A}\hat{\mathbf{E}}^H\mathbf{u}^* + \tfrac{1}{2}\boldsymbol{\theta}^T\boldsymbol{\Phi}^{-1}\boldsymbol{\theta} \tag{24}$$

where

$$\begin{aligned} \mathbf{A} = \; & \mathbf{R}^H\begin{bmatrix} \mathbf{F}^H\mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}^H\mathbf{D} \end{bmatrix}\begin{bmatrix} \mathbf{I} + \mathbf{V}\mathbf{V}^H & -\mathbf{W}\mathbf{W}^H \\ -\mathbf{W}\mathbf{W}^H & \mathbf{I} + \mathbf{V}\mathbf{V}^H \end{bmatrix} \\ & \times \begin{bmatrix} \mathbf{F}^H\mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}^H\mathbf{D} \end{bmatrix}^H \mathbf{R}. \end{aligned} \tag{25}$$

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

TABLE II

MODIFIED JCPCE WITH CLOSED-FORM CFO ESTIMATION.

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

and $\mathbf{R} = \mathrm{diag}(\mathbf{r})$, $\mathbf{u} = [e^{j\theta_0}, \cdots, e^{j\theta_{N-1}}]^T$. Therefore,

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &\approx \frac{1}{8\sigma^2\rho^2}(\mathbf{1}+j\boldsymbol{\theta})^T\hat{\mathbf{E}}\mathbf{A}\hat{\mathbf{E}}^H(\mathbf{1}-j\boldsymbol{\theta}) + \frac{1}{2}\boldsymbol{\theta}^T\mathbf{\Phi}^{-1}\boldsymbol{\theta} \\ &= \boldsymbol{\theta}^T\mathrm{Re}(\hat{\mathbf{E}}\mathbf{A}\hat{\mathbf{E}}^H)\boldsymbol{\theta} + 4\sigma^2\rho^2\boldsymbol{\theta}^T\mathbf{\Phi}^{-1}\boldsymbol{\theta} \\ &\quad -2\boldsymbol{\theta}^T\mathrm{Im}(\hat{\mathbf{E}}\mathbf{A}\hat{\mathbf{E}}^H)\mathbf{1}.\end{aligned} \tag{26}$$

Solving $\partial\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \mathbf{0}$ gives us the optimal estimate of $\boldsymbol{\theta}$

$$\hat{\boldsymbol{\theta}} = [\mathrm{Re}(\hat{\mathbf{E}}\mathbf{A}\hat{\mathbf{E}}^H) + 4\sigma^2\rho^2\mathbf{\Phi}^{-1}]^{-1}\mathrm{Im}(\hat{\mathbf{E}}\mathbf{A}\hat{\mathbf{E}}^H)\mathbf{1}. \tag{27}$$

2) Backward Substitution: Let $\hat{\mathbf{P}} = \mathrm{diag}(\exp(j\hat{\boldsymbol{\theta}}))$ and plug it into (22), the channel estimate after removing the CFO and PHN is therefore:

$$\hat{\mathbf{g}} = (4\rho^2)^{-1}\mathbf{W}^H\mathbf{D}^H\check{\mathbf{F}}\hat{\mathbf{P}}^H\hat{\mathbf{E}}^H\mathbf{r}. \tag{28}$$

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

V. COMPLEXITY ANALYSIS AND LOW COMPLEXITY IMPLEMENTATION

The proposed modified JCPCE algorithm leads to tremendous computational saving compared to the original one, as the estimation of the CFO now has a convenient closed form. However, further complexity reduction is still needed in order to bring the overall complexity down to the order of $\mathcal{O}(N\log N)$, to allow for practical implementation. Note that in the modified JCPCE, the main computational tasks reside in evaluating equations (19), (27) and (28). We will now investigate the complexity of each computation and seek means to reduce it.

A. Evaluation of $\hat{\mathbf{g}}$

In (28), 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.

B. Evaluation of $\hat{\boldsymbol{\theta}}$

We invoke the conjugate gradient (CG) method [6] to lower the complexity of evaluating $\hat{\boldsymbol{\theta}}$ in (27). The techniques used here is similar to those in [4], and some details will be omitted in the subsequent description.

TABLE III

CONJUGATE GRADIENT ALGORITHM FOR EVALUATING (27).

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

1) Wiener Phase Noise: The inverse of Wiener PHN covariance matrix $\mathbf{\Phi}$ has a convenient tridiagonal structure [7]. Writing $\mathbf{\Psi} = \frac{1}{4\sigma^2\rho^2}\mathbf{\Phi}$, $\mathbf{\Psi}^{-1} = 4\sigma^2\rho^2\mathbf{\Phi}^{-1}$ would also be tridiagonal. Let $\mathbf{q} = \mathrm{Im}(\hat{\mathbf{E}}\mathbf{A}\hat{\mathbf{E}}^H)\mathbf{1}$, 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. The evaluation of (27) is now equivalent to solving a linear equation $[\mathrm{Re}(\hat{\mathbf{E}}\mathbf{A}\hat{\mathbf{E}}^H) + \mathbf{\Psi}^{-1}]\hat{\boldsymbol{\theta}} = \mathbf{q}$. This problem can be easily tackled by the conjugate gradient method. The complete algorithm is presented in Table III.

Of all the operations in Table III, the dominant complexity is associated with the matrix-vector multiplication $[\mathrm{Re}(\hat{\mathbf{E}}\mathbf{A}\hat{\mathbf{E}}^H) + \mathbf{\Psi}^{-1}]\boldsymbol{\nu}_k$. Thanks to the tridiagonal form of $\mathbf{\Psi}^{-1}$, this can be performed easily. More specifically, evaluating $[\mathrm{Re}(\hat{\mathbf{E}}\mathbf{A}\hat{\mathbf{E}}^H) + \mathbf{\Psi}^{-1}]\boldsymbol{\nu}_k$ requires $6N + 2N\log N$ operations. The CG algorithm requires a maximum of N iterations to converge to the exact solution. But our simulation shows that only a small number of iterations are needed for accurate estimates. Hence, the complexity of evaluating (27) is $\mathcal{O}(iN\log N)$, where i is the number of CG iterations.

2) Gaussian Phase Noise: In the case of Gaussian PHN, we notice that $\mathbf{\Psi}$, as a Toeplitz matrix, can be approximated by a circulant matrix $\tilde{\mathbf{\Psi}}$ [8]. 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 [9]. Being a circulant matrix, the eigenvalue decomposition (EVD) 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 simplified estimator for $\boldsymbol{\theta}$ becomes

$$\hat{\boldsymbol{\theta}} = [\mathrm{Re}(\hat{\mathbf{E}}\mathbf{A}\hat{\mathbf{E}}^H) + \tilde{\mathbf{\Psi}}^{-1}]^{-1}\mathrm{Im}(\hat{\mathbf{E}}\mathbf{A}\hat{\mathbf{E}}^H)\mathbf{1}. \tag{29}$$

This problem can be treated similar to the Wiener PHN case using the CG method in Table III by replacing $\mathbf{\Psi}$ with $\tilde{\mathbf{\Psi}}$. It can be shown that its complexity is again $\mathcal{O}(iN\log N)$.

C. Evaluation of $\hat{\epsilon}$

Complexity reduction is also available for (19) by using the CG method. First we notice that $\mathbf{\Phi}_\triangle$, evaluated according to Lemma 1, is a Toeplitz matrix for both Wiener and Gaussian

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

PHN, hence a close circulant approximation $\tilde{\mathbf{\Phi}}_{\triangle}$ can be found [8]. Then we may concentrate on the matrix-vector product

$$\mathbf{x} = (\mathbf{R}_1 \tilde{\mathbf{\Phi}}_{\triangle} \mathbf{R}_1^H + 4\sigma^2 \mathbf{I})^{-1} \mathbf{r}_2. \quad (30)$$

This is equivalent to solving a linear equation $(\mathbf{R}_1 \tilde{\mathbf{\Phi}}_{\triangle} \mathbf{R}_1^H + 4\sigma^2 \mathbf{I})\mathbf{x} = \mathbf{r}_2$, which can be computed efficiently using the CG method analogous to the one described in Table III. For space limitation, we will not present the complete CG algorithm here. But it can be easily verified that the complexity of applying CG method is again $\mathcal{O}(iN \log N)$.

VI. Simulations

In the following, we simulate the performance of the proposed modified JCPCE based on algorithms presented in Table II. 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 4 taps. 2) An OFDM training symbol size of $N = 64$ subcarriers with each subcarrier modulated in QPSK format. 3) Baseband sampling rate $f_s = 20$ MHz (subcarrier spacing of 312.5 KHz). 4) The Wiener PHN is generated as a random-walk process with incremental PHN standard derivation of $\alpha_\phi = 1.0$ deg. The exact expression for $\mathbf{\Phi}$ can be found in [4]. 5) The Gaussian PHN has a standard deviation of $\theta_{rms} = 4$ deg (i.e. $R_\theta(0) = (\pi\theta_{rms}/180)^2$). It is generated, according to the Matlab code recommended for the IEEE 802.11g standard [9], as i.i.d. Gaussian samples passed through a single pole Butterworth filter of 3dB bandwidth $\Omega_o = 100$ kHz. Hence, the PHN covariance matrix $\mathbf{\Phi}$ is $\mathbf{\Phi}_{i,j} = (\pi\theta_{rms}/180)^2 e^{-\frac{2\pi\Omega_o|i-j|}{f_s}}$. This small amount of PHN can severely degrade the performance of higher-order modulation schemes, such as 64-QAM.

The achievable CFO estimation range is $|\epsilon| < 0.5$ for the original JCPCE algorithm and $|\epsilon| < 1$ for the modified JCPCE algorithm. In the simulations, the CFO term ϵ will be generated from a uniform distribution in $[-0.8, 0.8]$ corresponding to a maximum CFO of 250 KHz.

A. Unresolvable Residual Common Phase Rotation

Fig. 1 plots two instances of the PHN process (from the Wiener and Gaussian model, respectively) and their estimates via the modified JCPCE algorithm. At SNR = 30 dB, it is seen that the Wiener PHN is estimated accurately, while the estimator for the Gaussian PHN differs from the actual profile by a constant phase rotation. This constant rotation δ creates and equal but opposite rotation in the channel estimate – a phenomenon called *residual common phase rotation* (RCPR), which is analyzed comprehensively in [4]. Left untreated, the effect of RCPR is that in the data detection stage the receiver will experience an exacerbated PHN distortion. Fortunately, the statistics of the exacerbated PHN can be easily found [4], and dealt with accordingly in the data detection stage [10], [11]. Since this is outside the scope of this paper, we will simply assume from hereon that δ can be perfectly corrected to facilitate easy assessment of the quality of channel estimation.

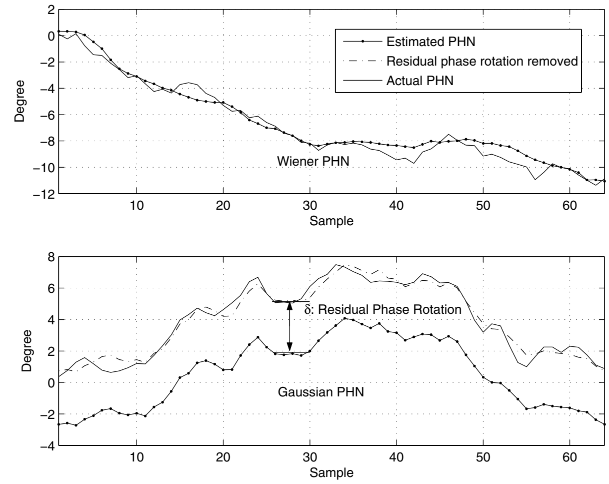

Fig. 1. Actual vs. estimated PHN profile using modified JCPCE.

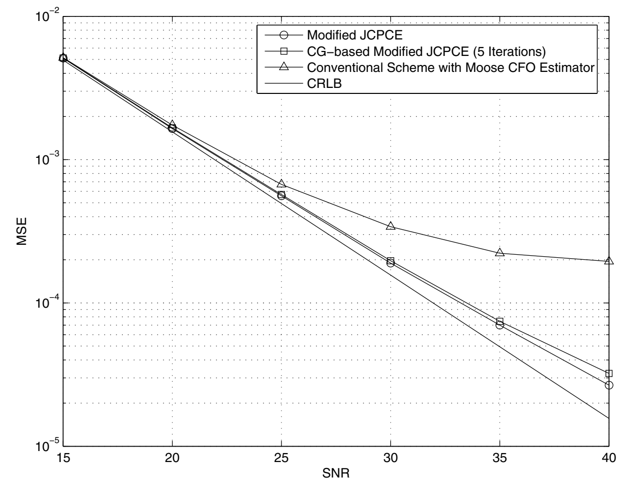

Fig. 2. MSE vs. SNR channel estimation performance with Wiener PHN.

RCPR is only minor in Wiener PHN channel due to the assumption of perfect phase synchronization at the beginning of the OFDM symbol.

B. Channel Estimation with Wiener PHN

Fig. 2 plots the channel estimation mean-squared error (MSE) as a function of the system SNR in the presence of both CFO and Wiener PHN using repeating training symbols. The proposed modified JCPCE (Table II) is plotted against the Cramér-Rao Lower Bound (CRLB) for estimating $\mathbf{g}$ in an OFDM channel free of CFO or PHN, which can be shown to be $\text{CRLB}(\mathbf{g}) = L/\text{SNR}$ with $\text{SNR} = E_s/N_o = \rho^2/\sigma^2$. It is seen that the modified JCPCE almost completely cancels the effect of CFO and PHN distortion. We also plot the low complexity implementation of the modified JCPCE using the CG method described in Section V. It is shown that even with as few iterations as $i = 5$ for the evaluation of (19) and (27), little performance degradation is introduced as a result.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

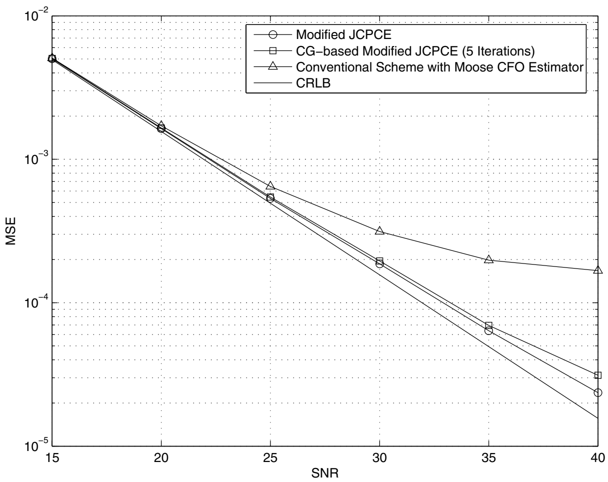

Fig. 3. MSE vs. SNR channel estimation performance with Gaussian PHN.

The top-most curve is the conventional channel estimator, which uses the Moose algorithm [3] to estimate the CFO and performs the subsequent channel estimation by ignoring the PHN. An error floor exists in the channel estimate obtained as such due to the inter-carrier interference (ICI) created by the PHN that causes a constant SNR degradation to the channel estimate. It can be checked from [4] that the original JCPCE performance curve is almost identical to the modified JCPCE performance curve obtained in this simulation, demonstrating that the modified JCPCE produces near-optimal performance while enjoying tremendous computational saving.

C. Channel Estimation with Gaussian PHN

Fig. 3 simulates the modified JCPCE for Gaussian PHN. Here we keep the same simulation settings except for the PHN statistics. It is seen again that the modified JCPCE performs close to the CRLB of the ideal distortionless channel. Therefore, it is safe to conclude that in practice the modified JCPCE is a preferred efficient alternative to the original JCPCE. In addition, we also plot the low complexity implementation of the modified JCPCE as described in Section V with $i = 5$ CG iterations for evaluating (19) and (27).

VII. Conclusions

This paper studied efficient OFDM channel and CFO estimation in the presence of PHN. We consider a training symbol structure similar to that in [3] and propose a near-optimal channel estimator which exploits the repetitive structure of the training symbol. The resultant modified JCPCE algorithm improves upon the original JCPCE algorithm as it does not require the expensive frequency offset search operation.

Furthermore, we explored ways to further reduce the complexity of the proposed estimator 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, where each iteration can be computed efficiently using FFT. This paper paves the way for the design of OFDM detectors in the presence of PHN [10], [11], where the CIR and CFO can now be safely assumed known.

Appendix I
Proof of Lemma 2

Using the Iterated Expectation Theorem [12, 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

$$\begin{aligned} \mathrm{E}(\mathbf{y}) &= \mathrm{E}_{\mathbf{x}}\mathrm{E}_{\mathbf{y}}(\mathbf{y}|\mathbf{x}) \\ \mathrm{V}(\mathbf{y}) &= \mathrm{V}_{\mathbf{x}}(\mathrm{E}_{\mathbf{y}}(\mathbf{y}|\mathbf{x})) + \mathrm{E}_{\mathbf{x}}(\mathrm{V}_{\mathbf{y}}(\mathbf{y}|\mathbf{x})). \end{aligned} \tag{31}$$

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

$$\begin{aligned} \mathrm{E}(\mathbf{r}_2|\mathbf{r}_1, \epsilon) &= \mathrm{E}_{\boldsymbol{\theta}_\triangle}(\mathrm{E}_{\mathbf{r}_2}(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_\triangle)) \\ \mathrm{V}(\mathbf{r}_2|\mathbf{r}_1, \epsilon) &= \mathrm{V}_{\boldsymbol{\theta}_\triangle}(\mathrm{E}_{\mathbf{r}_2}(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_\triangle)) \\ &\quad + \mathrm{E}_{\boldsymbol{\theta}_\triangle}(\mathrm{V}_{\mathbf{r}_2}(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_\triangle)). \end{aligned} \tag{32}$$

Also we have

$$\begin{aligned} \mathrm{E}_{\mathbf{r}_2}(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_\triangle) &= e^{j\pi\epsilon}\mathbf{P}_\triangle\mathbf{r}_1 \approx e^{j\pi\epsilon}\mathbf{R}_1(\mathbf{1} + j\boldsymbol{\theta}_\triangle); \\ \mathrm{V}_{\mathbf{r}_2}(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_\triangle) &= 4\sigma^2\mathbf{I}. \end{aligned} \tag{33}$$

Therefore, after simple matrix algebra we obtain

$$\begin{aligned} \mathrm{E}_{\boldsymbol{\theta}_\triangle}(\mathrm{E}_{\mathbf{r}_2}(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_\triangle)) &= e^{j\pi\epsilon}\mathbf{r}_1; \\ \mathrm{V}_{\boldsymbol{\theta}_\triangle}(\mathrm{E}_{\mathbf{r}_2}(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_\triangle)) &= \mathbf{R}_1\boldsymbol{\Phi}_\triangle\mathbf{R}_1^H; \\ \mathrm{E}_{\boldsymbol{\theta}_\triangle}(\mathrm{V}_{\mathbf{r}_2}(\mathbf{r}_2|\mathbf{r}_1, \epsilon, \boldsymbol{\theta}_\triangle)) &= 4\sigma^2\mathbf{I}. \end{aligned} \tag{34}$$

We then readily arrive at our final result

$$p(\mathbf{r}_2|\mathbf{r}_1, \epsilon) = \mathcal{CN}(e^{j\pi\epsilon}\mathbf{r}_1, \mathbf{R}_1\boldsymbol{\Phi}_\triangle\mathbf{R}_1^H + 4\sigma^2\mathbf{I}). \tag{35}$$

References

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

[2] B. Chen, "Maximum likelihood estimation of OFDM carrier frequency offset," *IEEE Sig. Proc. Letters*, vol. 9, no. 4, pp. 123–126, Apr. 2002.

[3] P. H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Trans. Comms.*, vol. 42, no. 10, pp. 2908–2914, Oct. 1994.

[4] D. D. Lin, R. A. Pacheco, T. J. Lim, and D. Hatzinakos, "Optimal OFDM channel estimation with carrier frequency offset and phase noise," *Proc. IEEE WCNC'06*, Apr. 2006.

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

[6] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer-Verlag, 1999.

[7] A. Kavčić and J. M. F. Moura, "Matrices with banded inverses: inversion algorithms and factorization of Gauss-Markov processes," *IEEE Trans. Info. Theory*, vol. 46, no. 4, pp. 1495–1509, July 2000.

[8] T. F. Chan, "An optimal circulant preconditioner for toeplitz systems," *SIAM J. Sci. Stat. Comput.*, vol. 9, no. 4, pp. 766–771, July 1988.

[9] IEEE P802.11-Task Group G. (2000) Phase Noise Matlab Model. [Online]. Available: http://grouper.ieee.org/groups/802/11/Reports/tgg_update.htm

[10] D. D. Lin, Y. Zhao, and T. J. Lim, "OFDM phase noise cancellation via approximate probabilistic inference," *Proc. IEEE WCNC'05*, vol. 1, pp. 27–32, Mar. 2005.

[11] D. D. Lin and T. J. Lim, "The variational inference approach to joint data detection and phase noise estimation in OFDM," *accepted to IEEE Trans. Signal Processing*.

[12] M. I. Jordan, *An Introduction to Learning in Graphical Models*. In Preparation, 2004.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.