

Optimal OFDM Channel Estimation with Carrier Frequency Offset and 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 optimal training-based OFDM channel impulse response (CIR) estimation algorithm that addresses the phase noise (PHN) and carrier frequency offset (CFO) problem. If left unattended, these combined problems severely degrade the accuracy of the channel estimate and ultimately the quality of the wireless link. The solution involves the joint optimization of a complete log-likelihood function over the unknown CIR, PHN and CFO. To reduce the complexity of the proposed algorithm, a simplification based on the conjugate gradient method is introduced, yielding an efficient realization using the Fast Fourier Transform (FFT) with only minor performance degradation.

I. INTRODUCTION

An open problem in OFDM receiver design is statistically optimal channel estimation when both phase noise (PHN) and carrier frequency offset (CFO) are present. The channel estimation problem in the presence of both CFO and PHN is a challenging one because of the complexity introduced by the coupled unknown parameters.

While the detrimental effects of CFO and PHN have been well documented [1], successful alleviation of these combined problems based on a statistically optimal receiver implementation, which must include channel estimation in the presence of CFO and PHN, have not been proposed. In [2], [3] and [4] PHN suppression methods were proposed for frequency selective channels but the channel frequency response was assumed to be known prior to PHN suppression. In [5] 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 [6] channel estimation was performed but the PHN was estimated using at least one “carrier recovery” pilot tone that required frequency guard bands on both band edges to minimize interference from data symbols. Only specific frequency selective channels were considered in the simulations, but the performance may degrade in general Rayleigh frequency selective channels since a channel null could occur in the vicinity of the pilot tone.

In this paper, our goal is to tackle the channel estimation problem when CFO and PHN are present from a maximum likelihood stand-point. Special features of the likelihood function are taken advantage of that enable a unique and elegant joint estimation scheme achieving optimal performance. We name it the *Joint CFO/PHN/CIR Estimator* (JCPCE).

Notation: Upper and lower case bold face letters indicate matrices and column vectors, respectively; $\mathbf{1}$ represents the all-one column vector; $\text{diag}(\mathbf{x})$ is a diagonal matrix with the

vector \mathbf{x} on its diagonal; $\text{diag}(\mathbf{X})$ is a diagonal matrix with the diagonal elements of square matrix \mathbf{X} on its diagonal; $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ denote the real and imaginary part of a vector or matrix; $E(\cdot)$ and $V(\cdot)$ 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}

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

II. SYSTEM DESCRIPTION

A. Prior Statistics of Phase Noise

Two different models of PHN are available in the literature [1]. The first one models a free-running oscillator and assumes the PHN process to be a Wiener process that is nonstationary, with a power that 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 coloured Gaussian process that is wide sense stationary (WSS) and has finite-power. In this paper, our solution covers both scenarios. For simplicity, we will refer to the first one as *Wiener PHN* and the second one as *Gaussian PHN*.

In both cases, denoting the phase noise process at the output of the VCO by $\theta(t)$, the samples of $\theta(t)$ within the m th 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, 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. In the rest of this section we explain how $\boldsymbol{\Phi}$ can be determined from the power spectral density (PSD) of the VCO output.

We first write the output of the VCO with PHN as¹ $s(t) = e^{j(2\pi f_o t + \theta(t))}$. Then the autocorrelation function of $s(t)$, $R_s(\tau)$, can be calculated:

$$\begin{aligned} R_s(\tau) &\doteq E\{s^*(t)s(t+\tau)\} \\ &= e^{j2\pi f_o \tau} E\{e^{j(\theta(t+\tau) - \theta(t))}\}. \end{aligned} \quad (2)$$

¹This is equivalent to considering real-valued sinusoids. See [7, pgs. 369–372] for details.

We will now briefly describe the autocorrelation functions of $s(t)$ for Wiener and Gaussian PHN respectively.

1) *Wiener Phase Noise*: The Wiener PHN process $\theta(t) = \int_0^t \phi(\tau) d\tau$, where $\phi(t)$ is zero-mean stationary Gaussian process with autocorrelation function $R_\phi(\tau) = 4\pi\beta_l\delta(\tau)$. Then it is known that [8]

$$\begin{aligned} R_s(\tau) &= e^{j2\pi f_o\tau} e^{-2\pi\beta_l|\tau|} \\ S_s(f) &= \frac{1}{\pi\beta_l} \left(\frac{1}{1+((f-f_o)/\beta_l)^2} \right) \end{aligned} \quad (3)$$

where $S_s(f) = \mathcal{F}\{R_s(\tau)\}$ denotes the PSD. So β_l is the 3dB bandwidth of the VCO output which can be easily measured using a spectrum analyzer. The discrete-time samples of $\theta(t)$ form a random-walk process $\theta_k = \theta_{k-1} + \phi_k$, $k = 0, \dots, N-1$, where $p(\phi_k) = \mathcal{N}(0, \alpha_\phi^2)$, $\alpha_\phi^2 = 4\pi\beta_l T/N$. Assuming $\theta_{-1} = 0$ due to perfect synchronization at the beginning of the OFDM symbol, the Gaussian-distributed PHN vector $\boldsymbol{\theta} = [\theta_0, \dots, \theta_{N-1}]^T$ has a covariance matrix

$$\boldsymbol{\Phi} = \alpha_\phi^2 \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & \dots & N \end{bmatrix}. \quad (4)$$

2) *Gaussian Phase Noise*: In this case $\theta(t)$ is modelled as a stationary random process with autocorrelation function $R_\theta(\tau)$. It is known that [9]

$$\begin{aligned} R_s(\tau) &= e^{j2\pi f_o\tau} e^{-R_\theta(0)} e^{R_\theta(\tau)} \\ &\approx e^{j2\pi f_o\tau} e^{-R_\theta(0)} (1 + R_\theta(\tau)). \end{aligned} \quad (5)$$

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 the PHN process in coherent receivers. Exact analysis of the PSD is found in [9] but for our purposes it is enough to use the above approximation to obtain

$$S_s(f) = e^{-R_\theta(0)} [\delta(f - f_o) + S_\theta(f - f_o)], \quad (6)$$

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 $S_\theta(f)$ and $R_\theta(f)$ can be found. Finally the value on the i th row and j th column of $\boldsymbol{\Phi}$ is

$$\boldsymbol{\Phi}_{i,j} = R_\theta \left(\left| i - j \right| \frac{T}{N} \right), \quad (7)$$

since T/N is the sampling period.

In the subsequent derivations, since both types of PHN can be sufficiently characterized by the covariance matrix $\boldsymbol{\Phi}$, we shall not distinguish between the two unless specifically stated.

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).

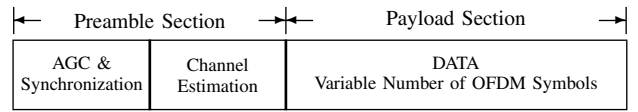


Fig. 1. OFDM packet structure.

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 an N point sequence for $n = 0, \dots, N-1$:

$$r_n = \frac{1}{\sqrt{N}} e^{j(\theta_n + 2\pi\epsilon n/N)} \sum_{k=0}^{N-1} h_k d_k e^{j2\pi nk/N} + \eta_n, \quad (8)$$

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. (8) may be written in matrix form as:

$$\mathbf{r} = \mathbf{E}\mathbf{P}\mathbf{F}^H \mathbf{H}\mathbf{d} + \mathbf{n}, \quad (9)$$

where $\mathbf{F} \in \mathbb{C}^{N \times N}$ is the DFT matrix with the (l, m) th element being $\mathbf{F}_{l,m} = \frac{1}{\sqrt{N}} e^{-j\frac{2\pi(l-1)(m-1)}{N}}$; $\mathbf{d} = [d_0, \dots, d_{N-1}]^T$ is the data vector; $\mathbf{n} = [\eta_0, \dots, \eta_{N-1}]^T$ is the noise vector; $\mathbf{P} = \text{diag}([e^{j\theta_0}, \dots, e^{j\theta_{N-1}}]^T)$ is the PHN matrix; $\mathbf{E} = \text{diag}([1, e^{j2\pi\epsilon/N}, \dots, e^{j2\pi(N-1)\epsilon/N}]^T)$ is the CFO matrix; and $\mathbf{H} = \text{diag}(\mathbf{h}) = \text{diag}([h_0, \dots, h_{N-1}]^T)$ is the channel matrix. Notice that although a full OFDM symbol contains $N_g + N$ time samples, 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. Depicting the channel response by a circulant matrix \mathbf{G} , we may rewrite (9) as

$$\mathbf{r} = \mathbf{E}\mathbf{P}\mathbf{G}\mathbf{F}^H \mathbf{d} + \mathbf{n}. \quad (10)$$

This is true because the circulant matrix \mathbf{G} can be diagonalized by the DFT matrix, i.e. $\mathbf{F}\mathbf{G}\mathbf{F}^H = \mathbf{H}$ or $\mathbf{G}\mathbf{F}^H = \mathbf{F}^H \mathbf{H}$. Using $\mathbf{g} = [g_0, \dots, g_{L-1}]^T$ to denote the CIR, where L is the channel length, \mathbf{G} is formed by circular rotations of $\frac{1}{\sqrt{N}}\mathbf{g}$. The CIR can be converted to the channel frequency response by writing $\mathbf{h} = \mathbf{W}\mathbf{g} \in \mathbb{C}^{N \times 1}$, 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 $\mathbf{V} \in \mathbb{C}^{N \times (N-L)}$ are orthogonal unitary matrices satisfying $\mathbf{W}^H \mathbf{V} = \mathbf{0}$ and $\mathbf{W}\mathbf{W}^H + \mathbf{V}\mathbf{V}^H = \mathbf{I}$.

Let $\mathbf{D} = \text{diag}(\mathbf{d})$. We can now introduce the following equivalent representation of (10) for the convenience of channel estimation:

$$\mathbf{r} = \mathbf{E}\mathbf{P}\mathbf{F}^H \mathbf{D}\mathbf{W}\mathbf{g} + \mathbf{n}. \quad (11)$$

III. CHANNEL ESTIMATION WITH CFO AND PHN

A. Ignoring CFO and PHN

To obtain the maximum likelihood (ML) estimate of the channel, we assume that the tap length of the impulse response (or equivalently, the dimension of \mathbf{g}), L , is known *a priori*. If we ignore CFO and PHN (i.e. assume $\mathbf{E}\mathbf{P} = \mathbf{I}$), from (11)

TABLE I

JOINT CFO/PHN/CIR ESTIMATOR (JCPCE).

we may write the negative log-likelihood function:

$$\begin{aligned}\mathcal{L}(\mathbf{g}) &= -\log p(\mathbf{r}|\mathbf{g}) \\ &= \frac{1}{2\sigma^2}(\mathbf{r} - \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{g})^H (\mathbf{r} - \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{g}),\end{aligned}\quad (12)$$

where

$$p(\mathbf{r}|\mathbf{g}) = \mathcal{CN}(\mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{g}, 2\sigma^2 \mathbf{I}).\quad (13)$$

Setting $\partial \mathcal{L}(\mathbf{g})/\partial \mathbf{g}^* = \mathbf{0}$ yields the ML channel estimate

$$\hat{\mathbf{g}} = (\mathbf{W}^H \mathbf{D}^H \mathbf{D} \mathbf{W})^{-1} \mathbf{W}^H \mathbf{D}^H \mathbf{F} \mathbf{r}.\quad (14)$$

It shall be assumed hereafter that constant-modulus training symbols are used, i.e. $\mathbf{D}^H \mathbf{D} = 2\rho^2 \mathbf{I}$ and the symbol energy per subcarrier $E_s = 2\rho^2$. Further simplifications lead to

$$\hat{\mathbf{g}} = (2\rho^2)^{-1} \mathbf{W}^H \mathbf{D}^H \mathbf{F} \mathbf{r}.\quad (15)$$

However, this estimator is unusable if the CFO ϵ and PHN θ are not equal to zero. In the next section we will derive a channel estimator that works even with non-zero CFO and PHN.

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

Looking at (11), it is obvious that given only \mathbf{r} , 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’’ $p(\mathbf{r}, \epsilon, \theta, \mathbf{g}) = p(\mathbf{r}|\epsilon, \theta, \mathbf{g})p(\epsilon)p(\theta)p(\mathbf{g})^2$. Since we assume no prior knowledge of ϵ and \mathbf{g} , $p(\epsilon)$ and $p(\mathbf{g})$ are constants. Also, we have assumed in Section II-A that the prior distribution of θ is

$$p(\theta) = \mathcal{N}(\mathbf{0}, \Phi),\quad (16)$$

where Φ is known. The ‘‘complete negative log-likelihood function’’ can therefore be written as

$$\begin{aligned}\mathcal{L}(\epsilon, \theta, \mathbf{g}) &= -\log p(\mathbf{r}|\epsilon, \theta, \mathbf{g}) - \log p(\theta) \\ &= \frac{1}{2\sigma^2}(\mathbf{r} - \mathbf{E} \mathbf{P} \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{g})^H (\mathbf{r} - \mathbf{E} \mathbf{P} \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{g}) \\ &\quad + \frac{1}{2} \theta^T \Phi^{-1} \theta.\end{aligned}\quad (17)$$

Our objective is to find the optimal estimates

$$(\hat{\epsilon}, \hat{\theta}, \hat{\mathbf{g}}) = \arg \min_{\epsilon, \theta, \mathbf{g}} \mathcal{L}(\epsilon, \theta, \mathbf{g}).\quad (18)$$

1) *Forward Substitution:* Solving $\partial \mathcal{L}(\epsilon, \theta, \mathbf{g})/\partial \mathbf{g}^* = \mathbf{0}$ produces the ML channel estimate³ in terms of ϵ and θ

$$\hat{\mathbf{g}} = (2\rho^2)^{-1} \mathbf{W}^H \mathbf{D}^H \mathbf{F} \mathbf{P}^H \mathbf{E}^H \mathbf{r}.\quad (19)$$

Noticing that

$$\begin{aligned}\mathbf{r} - \mathbf{E} \mathbf{P} \mathbf{F}^H \mathbf{D} \mathbf{W} \hat{\mathbf{g}} \\ &= \mathbf{r} - (2\rho^2)^{-1} \mathbf{E} \mathbf{P} \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{W}^H \mathbf{D}^H \mathbf{F} \mathbf{P}^H \mathbf{E}^H \mathbf{r} \\ &= (2\rho^2)^{-1} \mathbf{E} \mathbf{P} \mathbf{F}^H \mathbf{D} \mathbf{V} \mathbf{V}^H \mathbf{D}^H \mathbf{F} \mathbf{P}^H \mathbf{E}^H \mathbf{r},\end{aligned}\quad (20)$$

²The complete likelihood function is proportional to the *a posteriori* distribution $p(\epsilon, \theta, \mathbf{g}|\mathbf{r})$.

³Note that we are maximizing the ‘‘complete likelihood function’’ $p(\mathbf{r}, \epsilon, \theta, \mathbf{g})$ rather than the conventional likelihood function $p(\mathbf{r}|\epsilon, \theta, \mathbf{g})$.

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

and substituting (20) into (17), we have after simplification

$$\begin{aligned}\mathcal{L}(\epsilon, \theta) &= \frac{1}{4\sigma^2 \rho^2} \mathbf{u}^T \mathbf{E} \mathbf{R}^H \mathbf{F}^H \mathbf{D} \mathbf{V} \mathbf{V}^H \mathbf{D}^H \mathbf{F} \mathbf{R} \mathbf{E}^H \mathbf{u}^* \\ &\quad + \frac{1}{2} \theta^T \Phi^{-1} \theta,\end{aligned}\quad (21)$$

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

$$\begin{aligned}\mathcal{L}(\epsilon, \theta) &\approx \frac{1}{4\sigma^2 \rho^2} (\mathbf{1} + j\theta)^T \mathbf{E} \mathbf{C} \mathbf{C}^H \mathbf{E}^H (\mathbf{1} - j\theta) + \frac{1}{2} \theta^T \Phi^{-1} \theta \\ &= \frac{1}{4\sigma^2 \rho^2} \left[\theta^T \text{Re}(\mathbf{E} \mathbf{C} \mathbf{C}^H \mathbf{E}^H) \theta + 2\sigma^2 \rho^2 \theta^T \Phi^{-1} \theta \right. \\ &\quad \left. - 2\theta^T \text{Im}(\mathbf{E} \mathbf{C} \mathbf{C}^H \mathbf{E}^H) \mathbf{1} + \mathbf{1}^T \mathbf{E} \mathbf{C} \mathbf{C}^H \mathbf{E}^H \mathbf{1} \right].\end{aligned}\quad (22)$$

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

$$\hat{\theta} = [\text{Re}(\mathbf{E} \mathbf{C} \mathbf{C}^H \mathbf{E}^H) + 2\sigma^2 \rho^2 \Phi^{-1}]^{-1} \text{Im}(\mathbf{E} \mathbf{C} \mathbf{C}^H \mathbf{E}^H) \mathbf{1}.\quad (23)$$

Substituting (23) into (22) and simplifying, we have

$$\begin{aligned}\mathcal{L}(\epsilon) &\propto -\mathbf{1}^T \text{Im}(\mathbf{E} \mathbf{C} \mathbf{C}^H \mathbf{E}^H)^T [\text{Re}(\mathbf{E} \mathbf{C} \mathbf{C}^H \mathbf{E}^H) + 2\sigma^2 \rho^2 \Phi^{-1}]^{-1} \\ &\quad \times \text{Im}(\mathbf{E} \mathbf{C} \mathbf{C}^H \mathbf{E}^H) \mathbf{1} + \mathbf{1}^T \mathbf{E} \mathbf{C} \mathbf{C}^H \mathbf{E}^H \mathbf{1}.\end{aligned}\quad (24)$$

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

$$\hat{\epsilon} = \arg \min_{\epsilon} \mathcal{L}(\epsilon).\quad (25)$$

2) *Backward Substitution:* After finding $\hat{\epsilon}$ and correspondingly $\hat{\mathbf{E}}$, the values of $\hat{\theta}$ can be determined by substituting $\mathbf{E} = \hat{\mathbf{E}}$ into (23):

$$\hat{\theta} = [\text{Re}(\hat{\mathbf{E}} \mathbf{C} \mathbf{C}^H \hat{\mathbf{E}}^H) + 2\sigma^2 \rho^2 \Phi^{-1}]^{-1} \text{Im}(\hat{\mathbf{E}} \mathbf{C} \mathbf{C}^H \hat{\mathbf{E}}^H) \mathbf{1}.\quad (26)$$

Letting $\hat{\mathbf{P}} = \text{diag}(\exp(j\hat{\theta}))$ and plugging it into (19), the ML channel estimate after removing the CFO and PHN is therefore:

$$\hat{\mathbf{g}} = (2\rho^2)^{-1} \mathbf{W}^H \mathbf{D}^H \hat{\mathbf{F}} \hat{\mathbf{P}}^H \hat{\mathbf{E}}^H \mathbf{r}.\quad (27)$$

We summarize the complete JCPCE algorithm in Table I.

Note that the luxury of easily finding the jointly optimal estimates for ϵ , θ and \mathbf{g} is to a great extent owed 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 [10] to establish the optimality

TABLE II
CONJUGATE GRADIENT ALGORITHM FOR EVALUATING (26).

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

of MUSIC-based CFO estimation in OFDM.

C. Complexity Analysis and Low Complexity Implementation

In the OFDM system, computational complexity is a critical issue, since the use of FFT implies a low complexity order of $\mathcal{O}(N \log N)$. In the implementation of the JCPCE, the main computational tasks reside in evaluating equations (24), (26) and (27). We will now investigate the complexity of each computation and seek means to reduce it.

In (27), 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 (26), which involves a matrix inversion requiring in general a complexity order of $\mathcal{O}(N^3)$. However, as we will show in the following, with the help of the conjugate gradient (CG) method [11], we are able to lower the complexity to an acceptable level.

1) *Wiener Phase Noise*: The inverse of Wiener PHN covariance matrix $\boldsymbol{\Phi}$ has a convenient tridiagonal structure [12]. If we let $\boldsymbol{\Psi} = \frac{1}{2\sigma^2\rho^2}\boldsymbol{\Phi}$, $\boldsymbol{\Psi}^{-1} = 2\sigma^2\rho^2\boldsymbol{\Phi}^{-1}$ can be written as:

$$\boldsymbol{\Psi}^{-1} = \frac{2\sigma^2\rho^2}{\alpha_\phi^2} \begin{bmatrix} 2 & -1 & & \mathbf{0} \\ -1 & 2 & -1 & \\ & \ddots & \ddots & \ddots \\ \mathbf{0} & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix}. \quad (28)$$

Let $\mathbf{q} = \text{Im}(\hat{\mathbf{E}}\mathbf{C}\mathbf{C}^H\hat{\mathbf{E}}^H)\mathbf{1}$, 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 (26) is now equivalent to solving a linear equation $[\text{Re}(\hat{\mathbf{E}}\mathbf{C}\mathbf{C}^H\hat{\mathbf{E}}^H) + \boldsymbol{\Psi}^{-1}]\hat{\boldsymbol{\theta}} = \mathbf{q}$. This problem can be easily tackled by the conjugate gradient method. The complete algorithm is presented in Table II.

Of all the operations in Table II, the dominant complexity is associated with the matrix-vector multiplication $[\text{Re}(\hat{\mathbf{E}}\mathbf{C}\mathbf{C}^H\hat{\mathbf{E}}^H) + \boldsymbol{\Psi}^{-1}]\boldsymbol{\nu}_k$. Thanks to the tridiagonal form of $\boldsymbol{\Psi}^{-1}$, this can be performed easily. More specifically, evaluating $[\text{Re}(\hat{\mathbf{E}}\mathbf{C}\mathbf{C}^H\hat{\mathbf{E}}^H) + \boldsymbol{\Psi}^{-1}]\boldsymbol{\nu}_k$ 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 to the exact solution. But our simulations show that the number of iterations required for good estimation performance is much smaller than N . In conclusion, the complexity of evaluating (26) is $\mathcal{O}(iN \log N)$, where i is the number of iterations in the CG algorithm.

2) *Gaussian Phase Noise*: In the case of Gaussian PHN, we notice that $\boldsymbol{\Psi}$, as a Toeplitz matrix, can be approximated by a circulant matrix $\tilde{\boldsymbol{\Psi}}$ [13] according to this simple result:

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

$$\tilde{\psi}_i = \frac{(N-i)\psi_i + i\psi_{N-i}}{N}, \quad (29)$$

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

Proof: See [13]. ■

It can be shown that this approximation is asymptotically exact as $N \rightarrow \infty$ for an autocorrelation matrix $\boldsymbol{\Psi}$ of a first-order autoregressive process, which is a good fit to the Gaussian PHN process assumed in [14]. Being a circulant matrix, the eigenvalue decomposition (EVD) of $\boldsymbol{\Psi}$ is $\mathbf{F}\boldsymbol{\Lambda}_{\tilde{\boldsymbol{\Psi}}}\mathbf{F}^H$ and $\tilde{\boldsymbol{\Psi}}^{-1} = \mathbf{F}\boldsymbol{\Lambda}_{\tilde{\boldsymbol{\Psi}}}^{-1}\mathbf{F}^H$, where $\boldsymbol{\Lambda}_{\tilde{\boldsymbol{\Psi}}}$ is a diagonal matrix. It is well-known that $\boldsymbol{\Lambda}_{\tilde{\boldsymbol{\Psi}}} = \text{diag}(\sqrt{N}\mathbf{F}^H\tilde{\boldsymbol{\varphi}}_1)$, where $\tilde{\boldsymbol{\varphi}}_1$ is the first column of $\tilde{\boldsymbol{\Psi}}$. Replacing $\boldsymbol{\Psi}$ by $\tilde{\boldsymbol{\Psi}}$, the simplification for (26) becomes

$$\hat{\boldsymbol{\theta}} = [\text{Re}(\hat{\mathbf{E}}\mathbf{C}\mathbf{C}^H\hat{\mathbf{E}}^H) + \tilde{\boldsymbol{\Psi}}^{-1}]^{-1} \text{Im}(\hat{\mathbf{E}}\mathbf{C}\mathbf{C}^H\hat{\mathbf{E}}^H)\mathbf{1}. \quad (30)$$

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

The computation of (24) can be done almost identically as (26) using the CG algorithm, and its analysis is omitted here.

IV. SIMULATIONS

In the following, we simulate the performance of the proposed JCPCE based on the algorithm presented in Table I. 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 of $\alpha_\phi = 0.6$ deg. The covariance matrix $\boldsymbol{\Phi}$ is as depicted in (4). 5) The Gaussian PHN has a standard deviation of $\theta_{rms} = 3$ 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 [14], 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 $\boldsymbol{\Phi}$ is $\Phi_{i,j} = (\pi\theta_{rms}/180)^2 e^{-\frac{2\pi\Omega_o|i-j|}{f_s}}$.

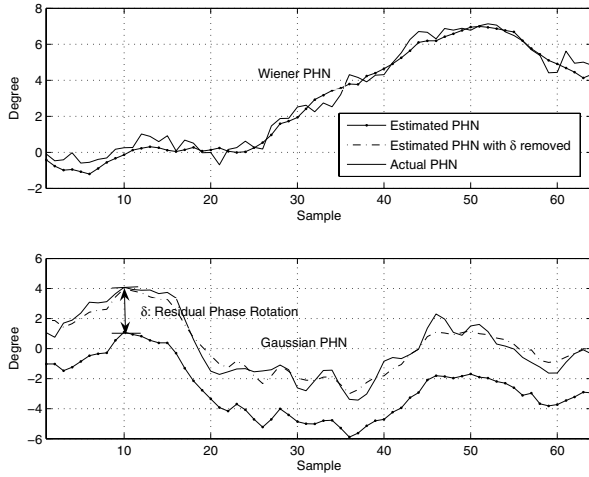


Fig. 2. Effect of residual common phase rotation in JCPCE.

A. Unresolvable Residual Common Phase Rotation

We will first perform simulations with no CFO to study the joint PHN and CIR estimation described as part of the JCPCE algorithm (Steps 2 and 3 in Table I with $\hat{\mathbf{E}} = \mathbf{I}$). Fig. 2 plots two instances of the PHN process (from the Wiener and Gaussian model, respectively) and their estimates via the 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 an equal but opposite rotation in the channel estimate – a phenomenon called *residual common phase rotation* (RCPR). The exact analysis of 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 θ_o and \mathbf{g}_o , respectively. As $\text{SNR} \rightarrow \infty$, the jointly ML estimates, $\hat{\theta}$ and $\hat{\mathbf{g}}$, calculated using the JCPCE algorithm approach

$$\hat{\theta} \rightarrow \theta_o + \delta \mathbf{1} \quad (31)$$

$$\hat{\mathbf{g}} \rightarrow e^{-j\delta} \mathbf{g}_o, \quad (32)$$

where $\delta = \arg \min_{\alpha} (\theta_o + \alpha \mathbf{1})^T \Phi^{-1} (\theta_o + \alpha \mathbf{1})$.

Proof: See Appendix I. ■

In brief, the RCPR, represented by an unknown constant δ , is introduced to shift the optimal estimate θ_o such that $\theta_o + \delta \mathbf{1}$ is closer to a zero-mean Gaussian process defined by the covariance matrix Φ . 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 making $\hat{\theta}$ most likely, $\hat{\theta} = \theta_o + \delta \mathbf{1}$ is approximately zero mean. That implies δ should be approximately the negative sample mean of θ_o : $\delta \approx -\frac{1}{N} \theta_o^T \mathbf{1}$.

Since $\theta_o \sim \mathcal{N}(\mathbf{0}, \Phi)$, it is easy to see that $\delta \sim \mathcal{N}(0, \mathbf{1}^T \Phi \mathbf{1} / N^2)$.

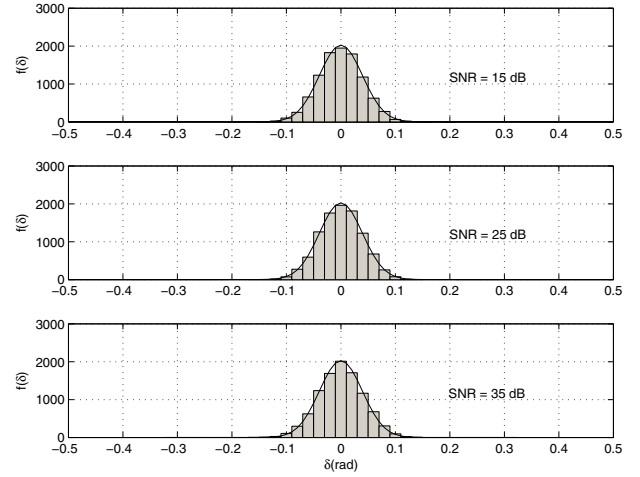


Fig. 3. The predicted pdf vs. the histogram of δ at SNR = 15, 25, 35 dB.

Fig. 3 shows the pdf of the measured δ compared to the Gaussian prediction, where δ is measured in simulation as the mean difference between $\hat{\theta}$ and θ_o . It is seen that this prediction holds very well at different SNR's. Hence, we now have a much better knowledge about the behaviour of RCPR, and know that it is not significant, as its variance is a fraction of that of PHN. In practice, we also have the option to estimate and remove δ in the data detection stage using the pilot symbols embedded in the transmitted OFDM symbols. Since this is not the subject of this paper, we will assume from hereon that δ can be perfectly corrected to facilitate easy assessment of the quality of channel estimation.

B. Channel Estimation Performance with both CFO and PHN

We now simulate the channel estimation performance in the presence of both CFO and PHN. The achievable CFO estimation range is $|\epsilon| < 0.5$ for the JCPCE. In the 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.

Fig. 4 and Fig. 5 plot the channel estimation MSE as a function of the system SNR ($\text{SNR} = E_s/N_o = \rho^2/\sigma^2$) 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 in (15) because it completely fails when $\epsilon \neq 0$.) The JCPCE algorithm is compared 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 $\text{CRLB}(\mathbf{g}) = L/\text{SNR}$. It is seen that the complete JCPCE algorithm almost completely cancels the effect of CFO and PHN distortion. The partial JCPCE, which optimally cancels CFO but ignores PHN, deviates from the CRLB at high SNR, demonstrating that PHN has a major effect in channel estimation even with optimal CFO estimation. We also plot the low complexity implementation of JCPCE using the CG method described in Section III-C. It is shown that for both types of PHN, there is only a small performance degradation

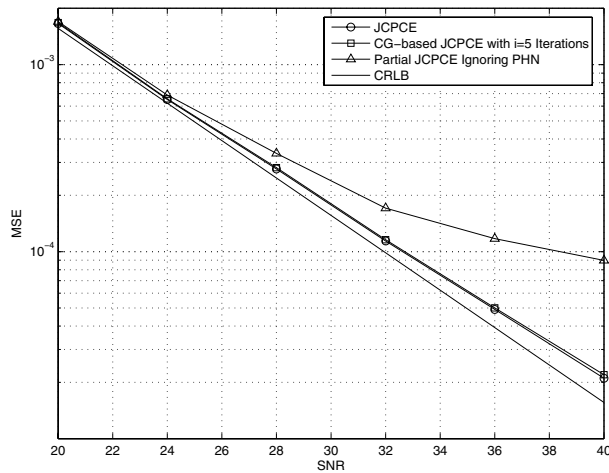


Fig. 4. MSE vs. SNR channel estimation performance (Wiener PHN).

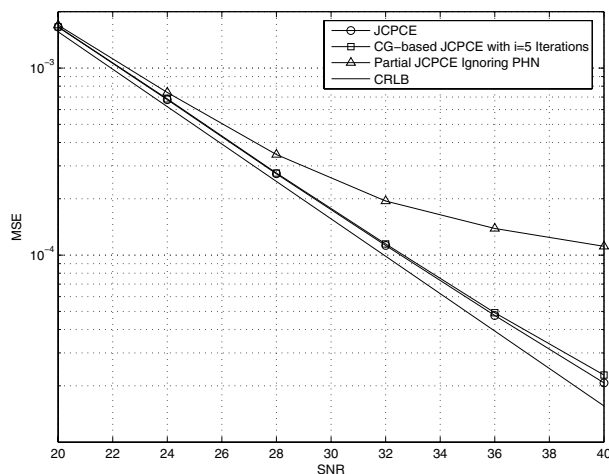


Fig. 5. MSE vs. SNR channel estimation performance (Gaussian PHN).

when using $i = 5$ CG iterations to evaluate (24) and (26).

V. CONCLUSIONS

This paper derived the optimal training-based OFDM channel estimation in the presence of CFO and PHN. The complex joint optimization problem turns out to be solvable through three elegantly decoupled optimization steps. In addition, we explored ways to reduce the complexity through the conjugate gradient method. This paper paves the way for the design of OFDM detectors in the presence of PHN (e.g. [4]), where the CIR and CFO (which are quasi-static) can now be safely assumed known.

APPENDIX I 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 $\text{SNR} \rightarrow \infty$ in relation to \mathbf{g}_o and $\boldsymbol{\theta}_o$.

Looking at (17), 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

$$\mathcal{L}_{p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{g})}(\boldsymbol{\theta}, \mathbf{g}) = \frac{1}{2\sigma^2} \|\mathbf{r} - \text{EPF}^H \text{DW}\mathbf{g}\|^2; \quad (33a)$$

$$\mathcal{L}_{p(\boldsymbol{\theta})}(\boldsymbol{\theta}) = \frac{1}{2} \boldsymbol{\theta}^T \boldsymbol{\Phi}^{-1} \boldsymbol{\theta}. \quad (33b)$$

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

$$(\boldsymbol{\theta}_o, \mathbf{g}_o) = \arg \min_{\boldsymbol{\theta}, \mathbf{g}} \mathcal{L}_{p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{g})}(\boldsymbol{\theta}, \mathbf{g}), \quad (34)$$

but the minimizer is not unique, since

$$(\boldsymbol{\theta}_o + \delta \mathbf{1}, e^{-j\delta} \mathbf{g}_o) = \arg \min_{\boldsymbol{\theta}, \mathbf{g}} \mathcal{L}_{p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{g})}(\boldsymbol{\theta}, \mathbf{g}), \quad (35)$$

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 (35), 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 $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{g}}) = \arg \min_{\boldsymbol{\theta}, \mathbf{g}} \mathcal{L}_{p(\boldsymbol{\theta})}(\boldsymbol{\theta})$ subject to $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{g}}) \in \mathcal{S}_{p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{g})}$.

REFERENCES

- [1] L. Piazzi and P. Mandarini, "Analysis of phase noise effects in OFDM modems," *IEEE Trans. Comms.*, vol. 50, no. 10, pp. 1696–1705, Oct. 2002.
- [2] S. Wu and Y. Bar-Ness, "OFDM systems in the presence of phase noise: consequences and solutions," *IEEE Trans. Comms.*, vol. 52, no. 11, pp. 1988–1996, Nov. 2004.
- [3] K. Nikitopoulos and A. Polydoros, "Phase-impairment effects and compensation algorithms for OFDM systems," *IEEE Trans. Comms.*, vol. 53, no. 4, pp. 698–707, Apr. 2005.
- [4] 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.
- [5] S. Wu and Y. Bar-Ness, "OFDM channel estimation in the presence of frequency offset and phase noise," *Proc. IEEE ICC'03*, vol. 5, pp. 3366–3370, May 2003.
- [6] M. S. El-Tanany, Y. Wu, and L. Hazy, "Analytical modeling and simulation of phase noise interference in OFDM-based digital television terrestrial broadcasting systems," *IEEE Trans. Broadcasting*, vol. 47, no. 1, pp. 20–31, Mar. 2001.
- [7] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. McGraw-Hill, 1991.
- [8] T. Pollet, M. V. Bladel, and M. Moeneclaey, "BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise," *IEEE Trans. Comms.*, vol. 43, pp. 191–193, Feb. 1995.
- [9] N. Abramson, "Bandwidth and spectra of phase-and-frequency-modulated waves," *IEEE Trans. Comms.*, pp. 407–414, Dec. 1963.
- [10] B. Chen, "Maximum likelihood estimation of OFDM carrier frequency offset," *IEEE Sig. Proc. Letters*, vol. 9, no. 4, pp. 123–126, Apr. 2002.
- [11] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer-Verlag, 1999.
- [12] A. Kavcic 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.
- [13] T. F. Chan, "An optimal circulant preconditioner for toeplitz systems," *SIAM J. Sci. Stat. Comput.*, vol. 9, no. 4, pp. 766–771, July 1988.
- [14] IEEE P802.11-Task Group G. (2000) Phase Noise Matlab Model. [Online]. Available: <http://grouper.ieee.org/groups/802/11/Reports/tgg-update.htm>