OFDM Phase Noise Cancellation via Approximate Probabilistic Inference

Darryl Dexu Lin, Yi Zhao and Teng Joon Lim

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, zhaoyi, limtj}@comm.utoronto.ca

Abstract— **We propose a systematic probabilistic framework to address the phase noise (PHN) problem in OFDM. In addition to deriving the optimal data detection scheme in the presence of PHN, we introduce a series of suboptimal approaches to blindly cancel the effect of PHN without the aid of pilot symbols. Not only do these algorithms provide the means to efficiently eliminate the effect of PHN in OFDM, they also open the door to much wider applications of advanced probabilistic inference algorithms in solving communications problems.**

I. Introduction

OFDM is becoming the technology of choice in fourth generation wireless communication systems. Wireless applications that use OFDM include wireless LAN IEEE 802.11a/g, fixed broadband wireless access IEEE 802.16a, terrestrial broadcast of digital television (DVB-T), HiperLAN2 in Europe, and HiSWANa in Japan. While OFDM is considered a practical scheme to combat frequency selective channel fading and to increase data rate, many practical challenges still face OFDM system designers [1]. In this paper, we consider the phase noise (PHN) problem that arises from the imperfections of a practical voltage-controlled oscillator (VCO). Improved RF circuit design could conceivably alleviate the problem but not eliminate it. Therefore, it is necessary to design digital signal processing techniques to combat residual PHN in the high-performance systems envisioned for the future, which are very sensitive to PHN. Despite the efforts of many researchers working in this area, most results are still based on over-simplified models and are sub-optimal.

The PHN problem appears similar to the channel estimation problem, but is in fact more challenging. The difficulties lie in the fact that channel impulse responses generally remain constant for multiple symbol periods, enabling the blind separation of the clean signal and channel utilizing the estimated statistics of the received signal [2]. However, PHN patterns vary from one OFDM symbol to another, and cannot be estimated by the same approach.

In [3], the effect of PHN on the system performance was studied and it was found that OFDM is orders more sensitive to PHN than a single carrier system. Tomba [4] provides a more detailed treatment on the OFDM error probability in the presence of PHN for different modulation schemes. Methods to estimate and mitigate the effect of PHN have been presented in [5], [6], [7]. These papers decompose PHN into two components: the *common phase noise* which is present in all subcarriers and the *random phase noise* which induces inter-carrier interference (ICI). While the common phase noise is measured as the average angular rotation of the constellation on the pilot subcarriers and cancelled on the data subcarriers [8], the random phase noise is simply ignored. This approach results in straightforward and easy-to-implement solutions, but is sub-optimal, as ICI can be ignored only when the phase noise power spectrum is very narrow. In this paper, we will show that the random phase noise can in fact be accurately estimated, and subsequently cancelled together with the common phase noise, as long as they can be modelled together as a coloured Gaussian process. Indeed, we will not even distinguish between the two components of PHN, and instead investigate the general PHN issue, by first developing a probabilistic model, and then deriving signal processing algorithms to solve the problem.

The mathematical foundations of our solutions are the approximate probabilistic inference algorithms. One instance of probabilistic inference, the sum-product algorithm in factor graphs [9], has been actively researched in recent years in connection with the decoding of Turbo codes and low-density parity-check (LDPC) codes. Recently, probabilistic inference has also been successfully used in image processing to perform scene analysis [10]. In this paper, we apply approximate probabilistic inference algorithms, namely variational inference and iterative conditional mode (ICM) [14], to mitigate PHN in OFDM, a novel way to approach this problem that leads to surprisingly low-complexity solutions.

Notation: $\mathbf{1}$ and $\mathbf{0}$ represent the all-one and all-zero column vectors; $\mathrm{diag}(\mathbf{x})$ is a diagonal matrix with the vector $\mathbf{x}$ on its diagonal; $\mathrm{diag}(\mathbf{X})$ is a diagonal matrix with the diagonal elements of matrix $\mathbf{X}$ on its diagonal; $\mathcal{N}(\mu, \mathbf{\Sigma})$ and $\mathcal{CN}(\mu, \mathbf{\Sigma})$ represent respectively real and circularly symmetric Gaussian random vectors with mean μ and covariance matrix $\mathbf{\Sigma}$.

II. Phase Noise Statistics

In this paper, we assume a system with a VCO controlled by a PLL, hence the PHN process may be assumed to be a zero-mean, wide sense stationary (WSS), coloured Gaussian process [11], [12]. Denoting the phase noise process at the output of the VCO by $\theta(t)$, the samples of $\theta(t)$ within the mth OFDM symbol, $\boldsymbol{\theta}_m$, has a multivariate Gaussian prior distribution: $p(\boldsymbol{\theta}_m) = \mathcal{N}(\mathbf{0}, \mathbf{\Phi})$, where the samples are taken at a rate of N/T samples per second, N is the number of

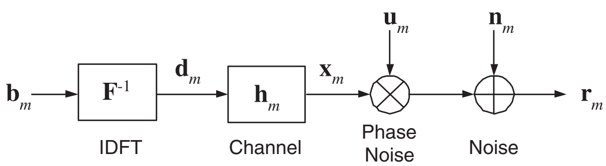

Fig. 1. Phase noise channel model.

OFDM sub-carriers, and T is the period of the OFDM symbol. For this model to be useful, however, the covariance matrix, $\mathbf{\Phi}$, must be available. Conveniently, instead of measuring $\mathbf{\Phi}$ directly, we may calculate it according to the specifications of the phase-locked VCO.

We first write the output of the VCO with PHN as:

$$\begin{aligned} s(t) &= \sin(\omega t + \theta(t)) \\ &\approx \sin(\omega t) + \cos(\omega t)\theta(t), \end{aligned} \tag{1}$$

where the approximation is tight when $\theta(t)$ is small. It follows that the power spectral density (PSD) of the output signal is

$$S_s(f) = \delta(f - \frac{\omega}{2\pi}) + S_\theta(f - \frac{\omega}{2\pi}), \tag{2}$$

where the delta function is contributed by $\sin(\omega t)$ and the frequency shift of $S_\theta(f)$ is caused by $\cos(\omega t)$. The shape of $S_s(f)$ may be measured by the spectrum analyzer or provided as part of the VCO specifications. Shifting $S_s(f)$ by the offset frequency $\frac{\omega}{2\pi}$ and removing the delta function, we then have the PSD of $\theta(t)$, $S_\theta(f)$.

The autocorrelation function $R_\theta(\tau)$ of the PHN process can be obtain 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 $\mathbf{\Phi}$ is extracted from $R_\theta(\tau)$:

$$\mathbf{\Phi}_{i,j} = R_\theta\left(\mid i - j \mid \frac{T}{N}\right), \tag{3}$$

since T/N is the sampling period.

III. Signal Model

Assuming perfect timing synchronization, the complex baseband received signal of the mth OFDM symbol, sampled at rate N/T, may be written in the time domain as:

$$\mathbf{r}_m = (\mathbf{h}_m \otimes \mathbf{d}_m) \circ \mathbf{u}_m + \mathbf{n}_m, \tag{4}$$

where $\mathbf{d}_m$ is the mth transmitted OFDM symbol of length N, i.e. the inverse DFT of the original data sequence $\mathbf{b}_m$; $\mathbf{h}_m$ is the channel impulse response of length $L_h < L_c$, L_c being the length of the cyclic prefix; $\mathbf{n}_m$ is the additive complex white Gaussian noise with variance σ^2 per dimension, and $\mathbf{u}_m = [\exp(j\theta_{m,1}), \cdots, \exp(j\theta_{m,N})]^T$ is the time-domain PHN pattern (or discrete-time PHN sequence). In (4), $\otimes$ represents circular convolution, and $\circ$ represents the Shur (or element-wise) product. Notice that although a full OFDM symbol contains $L_c + N$ time samples, in this signal model we assume the cyclic prefix has been removed and so there are only N samples per OFDM symbol. Depicting circular convolution by a circulant matrix $\mathbf{H}$ and leaving out the symbol index m, we may rewrite (4) as follows:

$$\mathbf{r} \approx \mathbf{Hd} \circ (\mathbf{1} + j\boldsymbol{\theta}) + \mathbf{n}, \tag{5}$$

where $\mathbf{u} \approx \mathbf{1} + j\boldsymbol{\theta}$ for small $\boldsymbol{\theta}$, $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_N]^T$. Assuming the channel impulse response is $\mathbf{h} = [h_0, \cdots, h_{L_h-1}]^T$, the first row of $\mathbf{H}$ is $[h_0, 0, \cdots, 0, h_{L_h-1}, \cdots, h_1]$.

We assume that the channel matrix $\mathbf{H}$ is perfectly known at the receiver. This is possible even in the presence of PHN because in a slow fading environment, the channel is estimated through training over a few OFDM symbols. The effect of PHN is thus averaged out (over time) and does not significantly affect the accuracy of channel estimation. In contrast to many papers discussing PHN, we do not assume that the channel is perfectly equalized before PHN estimation and cancellation, which would be difficult, looking at (5).

IV. Phase Noise Cancellation Algorithms

A. Conventional Schemes

In this section, we briefly describe conventional PHN cancellation schemes [5], [6]. Taking discrete Fourier transforms (DFT) on both sides of (4), we have

$$\mathbf{r}^f = [\mathbf{h}^f \circ \mathbf{d}^f] \otimes \mathbf{u}^f + \mathbf{n}^f \in \mathbb{C}^N \tag{6}$$

where the superscript f denotes the DFT of a vector. Since the PHN is a low pass process, it is assumed that only the first term in $\mathbf{u}^f$, u_1^f, is significant and therefore

$$\mathbf{r}^f \approx [\mathbf{h}^f \circ \mathbf{d}^f] \cdot u_1^f + \mathbf{n}^f. \tag{7}$$

Based on this approximation, and assuming the channel is perfectly equalized, the conventional schemes focuses on estimating the *common phase noise* term $\theta = \angle u_1^f$. These schemes differ in how the averaged common PHN is obtained, such as the use of pilot symbols in [5] and the use of tentative data symbols in [6]. The estimate of θ is

$$\hat{\theta} = \frac{1}{L}\sum_{k=1}^{L}\beta_k, \tag{8}$$

where L is the number of frequency-domain pilot symbols or the number of tentative data symbols, and β_k is the phase difference between the kth subcarrier sample and the kth pilot symbol or data symbol. In addition, if θ is assumed to be constant over multiple OFDM symbols, averaging over a few OFDM symbols yields smoother estimates.

Considering a common PHN across all subcarriers is only realistic in good quality, low-bandwidth VCO's, while the assumption that the PHN estimates do not change much over multiple OFDM symbols is not practical. In the sequel, we show that practical algorithms can be developed from more realistic models.

B. Exact Inference

We can re-write the received signal as

$$\begin{array}{rcl}\mathbf{r} & \approx & \mathbf{Hd} \circ (\mathbf{1} + j\boldsymbol{\theta}) + \mathbf{n} \\ & = & \mathbf{x} + j\mathbf{x} \circ \boldsymbol{\theta} + \mathbf{n}\end{array} \tag{9}$$

where $\mathbf{x} = \mathbf{Hd}$. To simplify the development, assume that $p(\mathbf{d}) = \mathcal{CN}(\mathbf{0}, 2\rho^2\mathbf{I})$ and that the prior pdf's can all be expressed in terms of Gaussian random vectors:

$$\begin{array}{rcl}p(\mathbf{x}) & = & \mathcal{CN}(\mathbf{0}, 2\rho^2\mathbf{HH}^H) \\ p(\boldsymbol{\theta}) & = & \mathcal{N}(\mathbf{0}, \boldsymbol{\Phi}) \\ p(\mathbf{r}|\mathbf{x}, \boldsymbol{\theta}) & = & \mathcal{CN}(\mathbf{x} + j\mathbf{x} \circ \boldsymbol{\theta}, 2\sigma^2\mathbf{I})\end{array} \tag{10}$$

where ρ^2 is the transmitted signal power per dimension.

The optimal posterior estimate of $\mathbf{x}$ is derived using classical estimation theory by treating $\boldsymbol{\theta}$ as a nuisance parameter and "integrating it out" [13] to obtain $p(\mathbf{x}|\mathbf{r}^*)$ where $\mathbf{r}^*$ denotes the value of $\mathbf{r}$ that is observed. With some effort (see Appendix I), we find that

$$\begin{array}{rl}p(\mathbf{x}|\mathbf{r}^*) \propto p(\mathbf{r}^*|\mathbf{x})p(\mathbf{x}) = & \mathcal{CN}(\mathbf{x}, \mathrm{diag}(\mathbf{x})\boldsymbol{\Phi}\mathrm{diag}(\mathbf{x})^H + 2\sigma^2\mathbf{I}) \\ & \cdot\mathcal{CN}(\mathbf{0}, 2\rho^2\mathbf{HH}^H).\end{array} \tag{11}$$

The optimal estimate of $\mathbf{x}$, and hence $\mathbf{b}$ since $\mathbf{x} = \mathbf{HF}^{-1}\mathbf{b}$ where $\mathbf{F}^{-1}$ is the IDFT matrix, is the one which maximizes the expression in (11). Unfortunately, the optimal estimate of $\mathbf{b}$ can only be found if each symbol hypothesis is tested, resulting in exponential complexity.

C. Variational Inference

Instead of finding $\mathbf{x}$ and $\boldsymbol{\theta}$ which maximize $p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{r}^*)$, the variational technique first looks for a parameterized Q-distribution, $Q(\mathbf{x}, \boldsymbol{\theta})$, which closely resembles $p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{r}^*)$, and then finds $\mathbf{x}$ and $\boldsymbol{\theta}$ that maximize $Q(\mathbf{x}, \boldsymbol{\theta})$. The simplicity of the variational technique lies in the fact that when $Q(\mathbf{x}, \boldsymbol{\theta})$ is properly selected (e.g. as a Gaussian distribution), the maximizers can be easily deduced.

To derive the variational inference algorithm, we first introduce a concept called *Variational Free Energy* (also called *Helmholtz Free Energy* or *Gibbs Free Energy*) [14] in the context of the PHN problem:

$$\mathcal{F}(Q, p) = \int_{\mathbf{x}, \boldsymbol{\theta}} Q(\mathbf{x}, \boldsymbol{\theta}) \ \log \frac{Q(\mathbf{x}, \boldsymbol{\theta})}{p(\mathbf{x}, \boldsymbol{\theta}, \mathbf{r}^*)} d\mathbf{x}d\boldsymbol{\theta}. \tag{12}$$

Here we use $p(\mathbf{x}, \boldsymbol{\theta}, \mathbf{r}^*)$ instead of $p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{r}^*)$ because they are proportional and hence equivalent in the free energy formulation. As is easily observed, this expression is exactly the Kullback-Leibler divergence [15] between $Q(\mathbf{x}, \boldsymbol{\theta})$ and $p(\mathbf{x}, \boldsymbol{\theta}, \mathbf{r}^*)$, $D(Q(\mathbf{x}, \boldsymbol{\theta})\|p(\mathbf{x}, \boldsymbol{\theta}, \mathbf{r}^*))$. By minimizing $\mathcal{F}(Q, p)$ over the parameters of $Q(\mathbf{x}, \boldsymbol{\theta})$, we obtain a Q-distribution $Q(\mathbf{x}, \boldsymbol{\theta})$ that most "resembles" $p(\mathbf{x}, \boldsymbol{\theta}, \mathbf{r}^*)$.

In cases where there are multiple arguments in the Q-function, an additional simplification can be made by factorizing $Q(\mathbf{x}, \boldsymbol{\theta})$ into a product form (also known as mean-field distribution), i.e. $Q(\mathbf{x}, \boldsymbol{\theta}) = Q_x(\mathbf{x})Q_\theta(\boldsymbol{\theta})$. As a result, the variational algorithm proceeds by minimizing the variational free energy iteratively over the parameters of $Q(\mathbf{x})$ and $Q(\boldsymbol{\theta})$ (where the subscripts of the Q-functions have been dropped for simplicity of notation).

For PHN estimation, we assume that

$$\begin{array}{rcl}Q(\mathbf{x}) & = & \mathcal{CN}(\mathbf{m}_x, \mathbf{S}_x), \\ Q(\boldsymbol{\theta}) & = & \mathcal{N}(\mathbf{m}_\theta, \mathbf{S}_\theta).\end{array} \tag{13}$$

It is worth noting that the posteriors of $\mathbf{x}$ and $\boldsymbol{\theta}$ are now parameterized by their means and variances (namely, $\mathbf{m}_x$, $\mathbf{m}_\theta$, $\mathbf{S}_x$, and $\mathbf{S}_\theta$), which then become the targets of optimization instead of the Q-functions themselves. Substituting the Q-functions from (13) into (12), we have the closed form expression of $\mathcal{F}(Q, p)$ as derived in Appendix II.

Obviously, the optimal parameters are hard to obtain analytically. The usual practice is to update each one of them in turn, while holding the others constant. After a few iterations, the algorithm is guaranteed to converge to a set of solutions at a local minimum of the free energy expression.

Algorithm 1: In the t-th iteration, the parameter update equations for the variational algorithm are:

$$\begin{array}{rcl}\mathbf{S}_\theta^{(t)} & = & \sigma^2[\sigma^2\boldsymbol{\Phi}^{-1} + \mathrm{diag}(\mathbf{S}_x^{(t-1)}) + \mathbf{M}_x^{H(t-1)}\mathbf{M}_x^{(t-1)}]^{-1}; \\ \mathbf{m}_\theta^{(t)} & = & \sigma^{-2}\mathbf{S}_\theta^{(t-1)} \cdot \mathrm{Re}[j\mathbf{M}_x^{H(t-1)}(\mathbf{m}_x^{(t-1)} - \mathbf{r}^*)]; \\ \mathbf{S}_x^{(t)} & = & 2\sigma^2[\frac{\sigma^2}{\rho^2}(\mathbf{HH}^H)^{-1} + \mathrm{diag}(\mathbf{S}_\theta^{(t-1)}) \\ & & +(j\mathbf{M}_\theta^{(t-1)} + \mathbf{I})^H(j\mathbf{M}_\theta^{(t-1)} + \mathbf{I})]^{-1}; \\ \mathbf{m}_x^{(t)} & = & \frac{1}{2}\sigma^{-2}\mathbf{S}_x^{(t-1)}(j\mathbf{M}_\theta^{(t-1)} + \mathbf{I})^H\mathbf{r}^*,\end{array} \tag{14}$$

where $\mathbf{M}_x = \mathrm{diag}(\mathbf{m}_x)$ and $\mathbf{M}_\theta = \mathrm{diag}(\mathbf{m}_\theta)$.

Proof: Differentiating the Gibbs free energy expression w.r.t. $\mathbf{S}_\theta$ and setting the result zero, we may solve the updating equation for $\mathbf{S}_\theta$ as a function of $\mathbf{m}_\theta$, $\mathbf{S}_x$, and $\mathbf{m}_x$, i.e. $\mathbf{S}_\theta^{(t)} = \arg\min_{\mathbf{S}_\theta} \mathcal{F}(\mathbf{m}_x^{(t-1)}, \mathbf{m}_\theta^{(t-1)}, \mathbf{S}_x^{(t-1)}, \mathbf{S}_\theta)$. The same procedure applies in finding the updating equations for $\mathbf{m}_\theta$, $\mathbf{S}_x$, and $\mathbf{m}_x$. ■

In each iteration of the variational algorithm, these parameters are updated in turn to generate new posterior estimates of $\mathbf{x}$ and $\boldsymbol{\theta}$ that decrease the free energy monotonically. The particular update order is chosen due to the dependence of $\mathbf{m}_\theta$ and $\mathbf{m}_x$ on $\mathbf{S}_\theta$ and $\mathbf{S}_x$. For initialization, a tentative data decision is made ignoring the effect of PHN and fed back as $\mathbf{m}_x^{(0)}$. At the end of the iterations, posterior distributions of $\mathbf{x}$ and $\boldsymbol{\theta}$ are extracted. Obviously, since $Q(\mathbf{x})$ is assumed to be Gaussian, a most reasonable final estimate of $\mathbf{x}$ (clean of PHN distortion) to be forwarded to the equalization and detection stage is the mean of $Q(\mathbf{x})$, i.e. $\mathbf{m}_x$.

The difference between the variational inference and exact inference lies in the fact that in variational inference the posterior estimates are "forced" to be independent Gaussian functions $Q(\mathbf{x})$ and $Q(\boldsymbol{\theta})$, making final symbol decisions easy to make. The price paid here is that since the posterior estimates are not exact posterior distributions, the algorithm requires a number of iterations to converge to a local minimum of the Gibbs free energy.

D. Iterative Conditional Mode

Variational inference has made a very complex problem computationally tractable, but a further simplification is possible by assuming the posteriors $Q(\mathbf{x})$ and $Q(\boldsymbol{\theta})$ to be delta functions instead of Gaussian.

In this case, the Q-functions are $\delta(\mathbf{x}, \hat{\mathbf{x}})$ and $\delta(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$, respectively. The notation $\delta(\mathbf{a}, \hat{\mathbf{a}})$ denotes a Dirac delta function with the following properties: $\int_{-\infty}^{\infty} \delta(\mathbf{a}, \hat{\mathbf{a}}) f(\mathbf{a})\ d\mathbf{a} = f(\hat{\mathbf{a}})$, and $\int_{-\infty}^{\infty} \delta(\mathbf{a}, \hat{\mathbf{a}})\ d\mathbf{a} = 1$. The minimization of Gibbs free energy over the parameters $\hat{\mathbf{x}}$ and $\hat{\boldsymbol{\theta}}$ is equivalent to maximizing $\mathcal{L}(\mathbf{x}, \boldsymbol{\theta}) = \log p(\mathbf{r}^*, \mathbf{x}, \boldsymbol{\theta})$ over $\mathbf{x}$ and $\boldsymbol{\theta}$. Simply put, the algorithm iteratively performs optimal point estimation for one of the two unknowns while holding the other fixed, hence the name Iterative Conditional Mode (ICM).

Since $p(\mathbf{r}^*, \mathbf{x}, \boldsymbol{\theta}) = p(\mathbf{r}^*|\mathbf{x}, \boldsymbol{\theta}) p(\mathbf{x}) p(\boldsymbol{\theta})$, $\mathcal{L}(\mathbf{x}, \boldsymbol{\theta})$ is evaluated to be:

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \boldsymbol{\theta}) \quad = \quad & \tfrac{1}{2\rho^2}(\mathbf{x}^H (\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{x}) + \tfrac{1}{2}\boldsymbol{\theta}^H \boldsymbol{\Phi}\boldsymbol{\theta} + \\ & \tfrac{1}{2\sigma^2}(\mathbf{r}^* - \mathbf{x} \circ (1 + j\boldsymbol{\theta}))^H (\mathbf{r}^* - \mathbf{x} \circ (1 + j\boldsymbol{\theta})). \end{aligned} \tag{15}$$

Algorithm 2: In the t-th iteration, the parameter update equations for the ICM algorithm are:

$$\begin{aligned} \boldsymbol{\theta}^{(t)} \quad &= \quad [\sigma^2 \boldsymbol{\Phi}^{-1} + \mathbf{X}^{H(t-1)} \mathbf{X}^{(t-1)}]^{-1} \\ & \quad \cdot \mathrm{Re}[j \mathbf{X}^{H(t-1)} (\mathbf{x}^{(t-1)} - \mathbf{r}^*)] \\ \mathbf{x}^{(t)} \quad &= \quad [\tfrac{\sigma^2}{\rho^2} (\mathbf{H}\mathbf{H}^H)^{-1} + (\mathbf{I} + j\mathbf{Y}^{(t-1)})^H (\mathbf{I} + j\mathbf{Y}^{(t-1)})]^{-1} \\ & \quad \cdot (\mathbf{I} + j\mathbf{Y}^{(t-1)})^H \mathbf{r}^*, \end{aligned} \tag{16}$$

where $\mathbf{X} = \mathrm{diag}(\mathbf{x})$ and $\mathbf{Y} = \mathrm{diag}(\boldsymbol{\theta})$.

Proof: Differentiating $\mathcal{L}(\mathbf{x}, \boldsymbol{\theta})$ w.r.t. $\boldsymbol{\theta}$ and setting the result to zero, we may find the new estimate for $\boldsymbol{\theta}$ as a function of $\mathbf{x}$, i.e. $\boldsymbol{\theta}^{(t)} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{x}^{(t-1)}, \boldsymbol{\theta})$. The same procedure applies in finding the updating functions for $\mathbf{x}$. ■

The saving in ICM is that we no longer require the covariance matrices $\mathbf{S}_x$ and $\mathbf{S}_\theta$ of the posterior distribution. The drawback is that point estimates do not take into account the uncertainties at each iteration, thus ICM gets stuck in local minima more easily and in general produce inferior results compared to variational inference. ICM is also a more "ad hoc" algorithm compared to exact inference and its variational counterpart. It resembles the heuristic decision-directed approach where the data symbols and PHN are detected/estimated iteratively until decisions are made for both unknowns. However, the ICM algorithm derived here differs from the heuristic one in that no detection decisions are made during the iterations.

E. Low Complexity Scheme

As can be seen in (14) and (16), the major complexity associated with the proposed algorithms is the evaluation of the inverse of $N \times N$ matrices. Such an inversion requires a complexity of $\mathcal{O}(N^3)$, while a practical OFDM system generally requires a complexity of $\mathcal{O}(N \log(N))$. Therefore, a lower complexity alternative must be devised.

A simple observation at the proposed algorithm reveals that it applies equally well to any portion of an OFDM symbol. This means that we may partition the OFDM symbol and perform PHN cancellation on each section of the partition as shown in Fig. 2. Assuming each section has K samples, $\mathbf{x}$ and $\boldsymbol{\theta}$ are both partitioned into N/K sub-vectors, i.e. $\mathbf{x} = [\mathbf{x}_1^T, \cdots, \mathbf{x}_{N/K}^T]^T$ and $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T, \cdots, \boldsymbol{\theta}_{N/K}^T]^T$. Denoting these sub-vectors as $\{\mathbf{x}_i\}_{i=1}^{N/K}$ and $\{\boldsymbol{\theta}_i\}_{i=1}^{N/K}$, we have the density functions similar to (10):

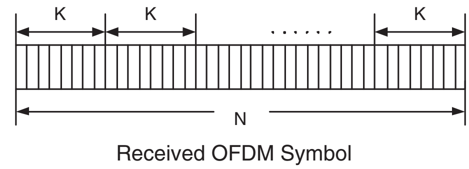

Fig. 2. OFDM symbol partitioning for low-complexity PHN removal.

$$\begin{aligned} p(\mathbf{x}_i) \quad &= \quad \mathcal{CN}(\mathbf{0}, 2\rho^2 (\mathbf{H}\mathbf{H}^H)_i) \\ p(\boldsymbol{\theta}_i) \quad &= \quad \mathcal{N}(\mathbf{0}, \boldsymbol{\Phi}_i) \\ p(\mathbf{r}_i|\mathbf{x}_i, \boldsymbol{\theta}_i) \quad &= \quad \mathcal{CN}(\mathbf{x}_i + j\mathbf{x}_i \circ \boldsymbol{\theta}_i, 2\sigma^2 \mathbf{I}) \end{aligned} \tag{17}$$

where $(\mathbf{H}\mathbf{H}^H)_i$ and $\boldsymbol{\Phi}_i$ are the ith diagonal block of $\mathbf{H}\mathbf{H}^H$ and $\boldsymbol{\Phi}$, respectively. The PHN cancellation scheme then follows exactly from the variational and ICM schemes discussed earlier, except that iterations are performed for each partitioned section. Taking the ICM technique for example, in the t-th iteration, the update equations of the iterative algorithm are:

$$\begin{array}{lcl} \text{For} & & i = 1 : N/K \\ \hline \boldsymbol{\theta}_i^{(t)} & = & [\sigma^2 \boldsymbol{\Phi}_i^{-1} + \mathbf{X}_i^{H(t-1)} \mathbf{X}_i^{(t-1)}]^{-1} \\ & & \cdot \mathrm{Re}[j \mathbf{X}_i^{H(t-1)} (\mathbf{x}_i^{(t-1)} - \mathbf{r}_i^*)] \\ \mathbf{x}_i^{(t)} & = & [\frac{\sigma^2}{\rho^2} (\mathbf{H}\mathbf{H}^H)_i^{-1} + (\mathbf{I} + j\mathbf{Y}_i^{(t-1)})^H \\ & & (\mathbf{I} + j\mathbf{Y}_i^{(t-1)})]^{-1} \cdot (\mathbf{I} + j\mathbf{Y}_i^{(t-1)})^H \mathbf{r}_i^*, \\ \hline \text{End} & & \end{array} \tag{18}$$

Since $(\mathbf{H}\mathbf{H}^H)_1 = \cdots = (\mathbf{H}\mathbf{H}^H)_{N/K}$ and $\boldsymbol{\Phi}_1 = \cdots = \boldsymbol{\Phi}_{N/K}$, their inverses need to be calculated only once per OFDM symbol. The overall complexity of the algorithm is now reduced from $\mathcal{O}(N^3)$ to $\mathcal{O}(N/K \times K^3) = \mathcal{O}(NK^2)$. The savings in implementation complexity with the simplified scheme is most significant when we consider longer OFDM symbols, such as in terrestrial broadcast of digital television (DVB-T) where N is in the range of thousands. The lower complexity does come with a slightly degraded performance, however, because we are effectively assuming $\boldsymbol{\Phi}$ and $\mathbf{H}\mathbf{H}^H$ to be block diagonal matrices.

V. Simulations

To verify the effectiveness of the proposed PHN cancellation schemes, we present a set of simulations as follows. The role of the PHN canceller in an OFDM receiver is depicted in Fig. 3. As can be seen, the received signal first goes through a tentative equalization and data detection module which equalizes the multipath channel and detects the transmitted data as if there is no PHN distortion. Such a decision is inaccurate, but is necessary to initialize the iterative PHN cancellation algorithm (so that it is not trapped in the first iteration). The remodulated OFDM symbol convolved by the channel impulse response is then fed into the PHN cancellation module as the initial estimate of $\mathbf{x}$ to start the iterative PHN cancellation process. The output of the PHN canceller module is the final estimate of $\mathbf{x}$, ideally clean of PHN distortions. The last equalization and detection stage is identical to the first one, except that this time, the data decision will be more accurate. The method for channel equalization (MMSE or zero-forcing) is a choice for the system designers.

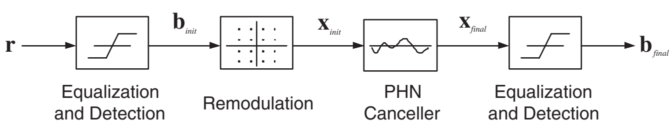

Fig. 3. OFDM receiver with PHN canceller.

In our simulations, we adopted zero-forcing equalization (i.e. inverting the channel at each subcarrier), for its low complexity and widespread use in practice. The PHN cancellation module consists of 5 iterations for the variational algorithm and 3 iterations for the ICM algorithm, since the variational method takes longer to converge. We also assumed the following system parameters: 1) A Rayleigh fading channel with three taps two samples away from each other; 2) An OFDM symbol size of 64 subcarriers with each subcarrier modulated in 64-QAM format; 3) A baseband sampling rate of 20 MHz; 4) A phase-locked VCO at the receiver with a PHN of 3 degrees RMS. The random PHN is generated, according to the Matlab code recommended for the IEEE 802.11g standard [16], as i.i.d. Gaussian samples passed through a Butterworth filter of 100 KHz 3dB bandwidth.

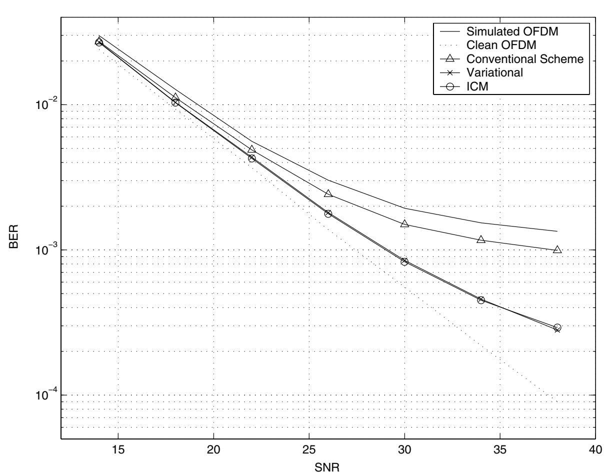

Fig. 4. Performance comparison between the conventional and proposed PHN cancellation schemes.

As illustrated in Fig. 4, we demonstrate the performance of the proposed PHN cancellation algorithms compared to the conventional scheme. The dotted line indicates the bit-error-rate (BER) of a OFDM receiver free of PHN (the ideal scenario), and the solid line indicates the BER of an OFDM receiver with PHN but without a PHN canceller (the worst case scenario). In between these two curves are the BER performance of receivers implementing the conventional PHN cancellation scheme (triangles) and the proposed PHN cancellation schemes (crosses and circles). It is obvious that both the variational and ICM techniques perform much better than the conventional one. It is somewhat surprising to notice that there is not much difference between the variational and ICM curves. We suspect that the superiority of variational method is not evident here because the local minimum of the Gibbs free energy is rather deep, and ICM suffices in such a case. Therefore, in the remaining simulations, we will leave out the variational method and use ICM alone for analysis and comparisons.

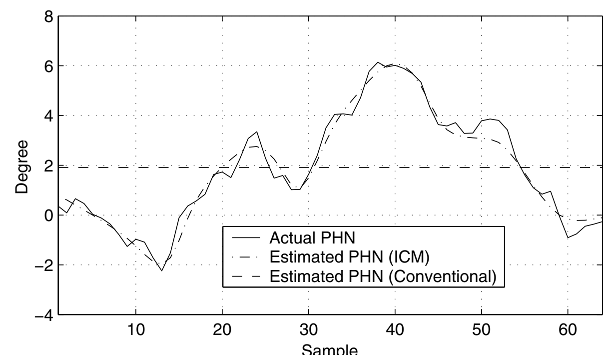

Fig. 5. Phase noise profile estimated using the conventional and proposed PHN cancellation schemes.

Fig. 5 compares the actual PHN profile with the PHN profile estimated using both the conventional scheme and the ICM technique at 30dB SNR. Apparently, the conventional model is insufficient to capture the dynamics of PHN. Through the ICM technique, however, we have very accurately estimated the phase noise profile, resulting in a much improved BER performance.

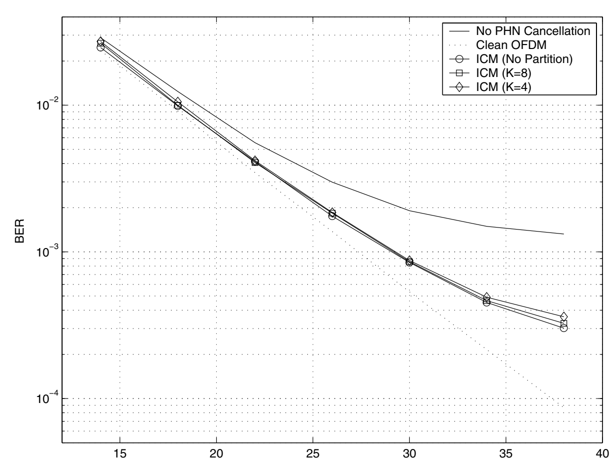

Fig. 6. Performance of the low complexity PHN cancellation schemes.

In Fig. 6 we investigate the performance of the low complexity simplified ICM technique by partitioning the OFDM symbol into blocks of size $K=8$ and $K=4$. It is seen from the BER curves that the OFDM receiver does not suffer from significant performance degradation even when K is as small as 4, implying that the PHN canceller can be implemented efficiently in a practical receiver.

VI. CONCLUSIONS

In this paper, we studied the blind cancellation of phase noise distortions at the OFDM receiver frontend. Instead of devising an optimal detection scheme for the original data symbol $\mathbf{b}$, we focused on the optimal and suboptimal estimation of the clean data sequence $\mathbf{x}$. We derived the exact posterior distribution $p(\mathbf{x}|\mathbf{r}^*)$ in Section IV.B. By simplifying the optimization objective, we also proposed the suboptimal estimation algorithms in Section IV.C and IV.D. Finally, a more practical scheme is introduce in Section IV.E by partitioning the OFDM symbol into smaller sections. By way of such a simplification, the complexity of the algorithm is further reduced.

The proposed PHN cancellation scheme is not limited to OFDM transmissions, as in the derivation we made no assumption about the particular form of $\mathbf{x}$, except that it can be modelled with a Gaussian prior distribution. Certainly, the proposed algorithms can be readily applied to single carrier or spread spectrum transmission schemes as well.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Mihnea Moldoveanu from Redline Communications for initiating this research, and Anitha Kannan from the University of Toronto for insightful discussions on probabilistic inference algorithms.

APPENDIX I
CLOSED FORM EXPRESSION OF $p(\mathbf{r}^*|\mathbf{x})$

Since $p(\mathbf{r}|\mathbf{x},\boldsymbol{\theta})$ and $p(\boldsymbol{\theta})$ are Gaussian distributed, It can be shown that the distribution of $p(\mathbf{r}|\mathbf{x})$ is also Gaussian. Denoting the mean of $\mathbf{r}|\mathbf{x}$ to be $\mathcal{E}(\mathbf{r}|\mathbf{x})$ and the variance as $\mathcal{V}(\mathbf{r}|\mathbf{x})$, we have

$$\begin{aligned}\mathcal{E}(\mathbf{r}|\mathbf{x}) &= \mathcal{E}_\theta[\mathcal{E}_r(\mathbf{r}|\mathbf{x},\boldsymbol{\theta})] \\ \mathcal{V}(\mathbf{r}|\mathbf{x}) &= \mathcal{V}_\theta[\mathcal{E}_r(\mathbf{r}|\mathbf{x},\boldsymbol{\theta})]+\mathcal{E}_\theta[\mathcal{V}_r(\mathbf{r}|\mathbf{x},\boldsymbol{\theta})].\end{aligned} \tag{19}$$

Because $p(\mathbf{r}|\mathbf{x},\boldsymbol{\theta})=\mathcal{CN}(\mathbf{x}+j\mathbf{x}\circ\boldsymbol{\theta},2\sigma^2\mathbf{I})$, it is straightforward to infer that

$$\begin{aligned}\mathcal{E}_r(\mathbf{r}|\mathbf{x},\boldsymbol{\theta}) &= \mathbf{x}+j\mathbf{x}\circ\boldsymbol{\theta} \\ \mathcal{V}_r(\mathbf{r}|\mathbf{x},\boldsymbol{\theta}) &= 2\sigma^2\mathbf{I}.\end{aligned} \tag{20}$$

Given that $p(\boldsymbol{\theta})=\mathcal{N}(\mathbf{0},\boldsymbol{\Phi})$, with some further manipulations, we obtain

$$\begin{aligned}\mathcal{E}_\theta[\mathcal{E}_r(\mathbf{r}|\mathbf{x},\boldsymbol{\theta})] &= \mathbf{x} \\ \mathcal{V}_\theta[\mathcal{E}_r(\mathbf{r}|\mathbf{x},\boldsymbol{\theta})] &= \mathrm{diag}(\mathbf{x})\boldsymbol{\Phi}\mathrm{diag}(\mathbf{x})^H \\ \mathcal{E}_\theta[\mathcal{V}_r(\mathbf{r}|\mathbf{x},\boldsymbol{\theta})] &= 2\sigma^2\mathbf{I},\end{aligned} \tag{21}$$

which implies that

$$\begin{aligned}\mathcal{E}(\mathbf{r}|\mathbf{x}) &= \mathbf{x} \\ \mathcal{V}(\mathbf{r}|\mathbf{x}) &= \mathrm{diag}(\mathbf{x})\boldsymbol{\Phi}\mathrm{diag}(\mathbf{x})^H+2\sigma^2\mathbf{I}.\end{aligned} \tag{22}$$

Therefore, $p(\mathbf{r}^*|\mathbf{x})=\mathcal{CN}(\mathbf{x},\mathrm{diag}(\mathbf{x})\boldsymbol{\Phi}\mathrm{diag}(\mathbf{x})^H+2\sigma^2\mathbf{I})$.

APPENDIX II
CLOSED FROM EXPRESSION OF $\mathcal{F}(Q,p)$

A simple expansion of the Gibbs free energy in (12) shows that

$$\begin{aligned}\mathcal{F}(Q,p) &= \int_{\mathbf{x},\boldsymbol{\theta}} Q(\mathbf{x})Q(\boldsymbol{\theta})\ \log\frac{Q(\mathbf{x})Q(\boldsymbol{\theta})}{p(\mathbf{x},\boldsymbol{\theta},\mathbf{r})}d\mathbf{x}d\boldsymbol{\theta} \\ &= \int_{\mathbf{x},\boldsymbol{\theta}} Q(\mathbf{x})Q(\boldsymbol{\theta})\ \log\frac{Q(\mathbf{x})Q(\boldsymbol{\theta})}{p(\mathbf{r}|\mathbf{x},\boldsymbol{\theta})p(\mathbf{x})p(\boldsymbol{\theta})}d\mathbf{x}d\boldsymbol{\theta} \\ &= -\int_{\mathbf{x}} Q(\mathbf{x})\log p(\mathbf{x})d\mathbf{x}-\int_{\boldsymbol{\theta}} Q(\boldsymbol{\theta})\log p(\boldsymbol{\theta})d\boldsymbol{\theta} \\ &\quad -\int_{\mathbf{x},\boldsymbol{\theta}} Q(\mathbf{x})Q(\boldsymbol{\theta})\ \log p(\mathbf{r}|\mathbf{x},\boldsymbol{\theta})d\mathbf{x}d\boldsymbol{\theta} \\ &\quad +\int_{\mathbf{x}} Q(\mathbf{x})\log Q(\mathbf{x})d\mathbf{x}+\int_{\boldsymbol{\theta}} Q(\boldsymbol{\theta})\log Q(\boldsymbol{\theta})d\boldsymbol{\theta}\end{aligned} \tag{23}$$

Substituting (10) and (13) into the expression above and applying Gaussian expectation properties, we may obtain Gibbs free energy as a function of the Q-distribution parameters (with the constant terms omitted):

$$\begin{aligned}\mathcal{F} &= \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}_x)+\mathbf{m}_x^H\boldsymbol{\Sigma}^{-1}\mathbf{m}_x+\tfrac{1}{2}\mathrm{tr}(\boldsymbol{\Phi}^{-1}\mathbf{S}_\theta) \\ &\quad +\tfrac{1}{2}\mathbf{m}_\theta^T\boldsymbol{\Phi}^{-1}\mathbf{m}_\theta-\log|\mathbf{S}_\theta|^{\frac{1}{2}}-\log|\mathbf{S}_x| \\ &\quad +\tfrac{1}{2\sigma^2}\{\mathrm{tr}[\mathrm{diag}(\mathbf{S}_\theta)\mathbf{S}_x]+\mathbf{m}_x^H\mathrm{diag}(\mathbf{S}_\theta)\mathbf{m}_x \\ &\quad +\mathrm{tr}[(j\mathbf{M}_\theta+\mathbf{I})\mathbf{S}_x(j\mathbf{M}_\theta+\mathbf{I})^H] \\ &\quad +[(j\mathbf{M}_\theta+\mathbf{I})\mathbf{m}_x-\mathbf{r}]^H[(j\mathbf{M}_\theta+\mathbf{I})\mathbf{m}_x-\mathbf{r}]\}\end{aligned}$$

where $\mathbf{M}_\theta=\mathrm{diag}(\mathbf{m}_\theta)$, $\boldsymbol{\Sigma}=2\rho^2\mathbf{H}\mathbf{H}^H$.

REFERENCES

[1] R. van Nee, *OFDM wireless multimedia communications*, Artech House, Boston, 2000.

[2] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics," *IEEE Trans. Info. Theory*, vol. 40, pp. 340–349, Mar. 1994.

[3] 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.

[4] L. Tomba, "On the effect of Wiener phase noise in OFDM systems," *IEEE Trans. Comms.*, vol. 46, pp. 580–583, May 1998.

[5] P. Robertson and S. Kaiser, "Analysis of the effects of phase noise in orthogonal frequency division multiplex (OFDM) systems," *Proc. ICC'95*, pp. 1652–1657, June 1995.

[6] K. Nikitopoulos and A. Polydoros, "Compensation schemes for phase noise and residual frequency offset in OFDM systems," *Proc. IEEE GlobeCom'01*, vol. 1, pp. 330–333, Nov. 2002.

[7] S. Wu and Y. Bar-Ness, "A phase noise suppression algorithm for OFDM-based WLANs," *IEEE Comms. Lett.*, vol. 6, no. 12, pp. 535–537, Dec. 2002.

[8] S. Lang, R. M. Rao, and B. Daneshrad, "Design and development of a 5.25 GHz software defined wireless OFDM communication platform," *IEEE Comm. Magazine*, vol. 42, no. 6, pp. 40–46, June 2004.

[9] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Info. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[10] B. J. Frey and N. Jojic, "A comparison of algorithms for inference and learning in probabilistic graphical models," *to appear on IEEE Trans. Patt. Anal. Mach. Int.*

[11] L. Zhao and W. Namgoong, "A novel adaptive phase noise compensation scheme for communication receivers," *Proc. IEEE GlobeCom'03*, vol. 4, pp. 2274–2279, Dec. 2002.

[12] A. Mehrotra, "Noise analysis of phase-locked loops," *IEEE Trans. Circ. Sys.*, vol. 49, no. 9, pp. 1309–1315, Sept. 2002.

[13] S. M. Kay, *Fundamentals of statistical signal processing - estimation theory*, chapter 10.7, pp. 328–330, Prentice Hall PTR, Upper Saddle River, New Jersey, 1993.

[14] B. J. Frey, *Graphical models for machine learning and digital communication*, MIT Press, Cambridge, Massachusetts, 1998.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, 1991.

[16] IEEE P802.11 Task Group G, *http://grouper.ieee.org/groups/802/11/Reports/tgg_update.htm.*

