

OFDM Joint Data Detection and Phase Noise Cancellation Based on Minimum Mean Square Prediction Error

Yu Gong and Xia Hong

*School of Systems Engineering
The University of Reading, Reading
RG6 6AY, UK*

E-mail: {x.hong, y.gong}@reading.ac.uk

Abstract

This paper proposes a new iterative algorithm for OFDM joint data detection and phase noise (PHN) cancellation based on minimum mean square prediction error. We particularly highlight the relatively less studied problem of “overfitting” such that the iterative approach may converge to a trivial solution. Specifically, we apply a hard decision procedure at every iterative step to overcome the overfitting. Moreover, compared with existing algorithms, a more accurate Pade approximation is used to represent the phase noise, and finally a more robust and compact fast process based on Givens rotation is proposed to reduce the complexity to a practical level. Numerical simulations are also given to verify the proposed algorithm.

Keywords: Orthogonal frequency division multiplexing (OFDM), phase noise, prediction error.

1. Introduction

Orthogonal frequency division multiplexing (OFDM) becomes an ever popular scheme for high speed wireless communications. Modulating the information data onto multiple orthogonal bands in frequency with a simple pair of Fast Fourier Transform (FFT) and

Inverse-FFT (IFFT), the OFDM system can effectively combat the multipath phenomena, a major degradation factor that harms the wireless communications. Against this background, however, the success of OFDM scheme is likely to be impaired by the detrimental factors including the imperfect channel estimation, symbol offset, phase noise (PHN) etc [1, 2]. In this paper, we consider the PHN as it is relatively less well studied.

The PHN arises from the imperfections at the receiver's oscillator, damaging the orthogonality among the subcarriers in OFDM symbols [3, 4]. A typical PHN consists of two parts, namely the common PHN and the random PHN. While the common PHN is an averaging effect over one OFDM system and thus the same for all subcarriers, the random PHN varies from one subcarrier to another, resulting in intercarrier interference. Although there exist various PHN suppression algorithms in the literature, most of them mainly considered the common PHN which can be mitigated with the help of pilot symbols (e.g.[5, 6, 7]). In a recent paper [8], a family of algorithms for joint data detection and PHN cancellation has been proposed, where the common and random phase noises were treated simultaneously under a uniform framework. These proposed algorithms were based on a probabilistic approach called variational inference, with the Gaussian assumption being applied to both the channel noise and PHN.

The joint data detection and PHN cancellation can be regarded as a parameter estimation problem in statistical learning with both the data and PHN being regarded as unknown parameters. To be specific, based on suitable signal and phase models, the data and PHN are estimated using a finite number of received samples as the target of learning. This joint estimate usually comes with an iterative approach similar to the expectation-

maximization (EM) algorithm. It is known that such iterative approaches do not guarantee the convergence to the global minimum. Particularly, due to the noise in received samples, “overfitting” may occur so that, rather than produce a good fit between the data estimates and the true transmitted data, the derived estimates are too close to the received samples and fit into the noise. A common technique to overcome the overfitting problem is the parameter regularization which has a connection to Bayesian theory [9]. We note that, although the overfitting problem was not explicitly identified, the algorithms in [8] are in fact equivalent to Bayesian regularization utilizing the Gaussian distributions as the priors. As will be shown in this paper, however, the parameter regularization alone may not be effective enough to overcome the overfitting which, we believe, is a major problem in this joint approach. Other techniques are thus desirable.

In this paper, based on an objective function that minimizes the prediction errors from both the PHN model and the channel model, a set of normal equations are derived to jointly detect the data and cancel the PHN. Unlike the variational inference approaches in [8], however, the proposed algorithm does not require Gaussian assumption on the channel and PHN, providing an alternative insight into the PHN cancellation problem. Particularly, a hard-decision procedure is introduced at every iteration to overcome the overfitting problem by filtering the noise out of the symbol estimate with the a priori information from the constellation of the transmission symbols.

Another contribution from this paper is the use of the Pade approximation to represent the PHN. Existing PHN cancellation algorithms usually use the first order Taylor series to approximate the PHN (e.g. [4, 8]). Although this greatly simplifies the resultant proce-

dures, it is not as sufficiently accurate. The Pade approximation, on the contrary, offers almost a perfect approximation of the PHN even for very high noise values. This provides the proposed algorithm a potential of combating higher PHN than existent approaches. We note that, due to its non-linear form, the Pade approximation or the higher order Taylor approximation, is very difficult, if not impossible, to be used to derive variational inference approaches as those in [8]. Finally in this paper, a new fast process based on Givens rotations is designed for the proposed algorithm. This Givens rotation based approach not only has lower complexity but also is more robust than the conjugate gradient (CG) algorithm used in [8].

2. The System Model

Without losing generality, we consider an OFDM system with N subcarriers modulated by M-QAM. The transmitted complex baseband OFDM signal is written as:

$$s(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} S_k e^{j2\pi kt/T}, \quad 0 \leq t \leq T \quad (1)$$

where S_k is the data symbol at the k th subcarrier, T is the period of an OFDM symbol. The OFDM preambles are ignored in (1) as we assume they are long enough to avoid intersymbol interference and can be perfectly removed at the receiver.

We also assume the channel is slow fading so that the channel coefficients remain unchanged within an OFDM symbol period. With the presence of the PHN, the received

signal sequence is given by:

$$r_m = \frac{1}{\sqrt{N}} e^{j\theta_m} \sum_{k=0}^{N-1} h_k S_k e^{j2\pi km/N} + \eta_m, \quad m = 0, \dots, N-1 \quad (2)$$

where h_k is the channel frequency response at subcarrier k which is assumed to be known in this paper, η_m is the complex white noise with mean zero and variance σ^2 , and θ_m is the m th sample of the PHN. The task is to recover the transmission symbols S_k without knowing the PHN θ_m , given the received signal r_m for $m = 0, \dots, N-1$.

There are two types of PHN: the Wiener PHN and the Gaussian PHN [4, 10, 8]. For the Wiener PHN, θ_m forms a random walk process as:

$$\theta_m = \theta_{m-1} + \varepsilon_m, \quad m = 0, \dots, N-1 \quad (3)$$

where ε_m is a random noise with zero mean and variance σ_ε^2 .

For the Gaussian PHN, on the other hand, θ_m is stationary which is assumed to follow the autoregressive (AR) model of order size K :

$$\theta_m = \sum_{i=1}^K a_i \theta_{m-i} + \varepsilon_m. \quad m = 0, \dots, N-1 \quad (4)$$

As will be shown in Section 4.2, the coefficients a_i and the order size K can be estimated by applying the least square (LS) and the Akaike's Information Criterion (AIC) [11] on the PHN samples θ_m . The AIC method is a well known method for determining the order of the autoregressive model in statistics. Information theoretic metrics of a model's gener-

alization capability are of great importance in statistical learning. A fundamental concept in the evaluation of model generalization capability is that of cross validation [12] which is often used to derive the information theoretic metrics. Various model selection criteria have been introduced based on cross validation including the AIC method. Alternatively, if the correlation matrix of the PHN is known as was in [8], a_i can be obtained via the Yule Walker equation $\mathcal{P}\mathbf{a} = \boldsymbol{\rho}$, with $\mathbf{a} = [a_1, \dots, a_K]^T$, $\boldsymbol{\rho} = [\rho_1, \dots, \rho_K]^T$ and:

$$\mathcal{P} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{K-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{K-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{K-1} & \cdots & \cdots & \rho_1 & 1 \end{bmatrix}, \quad (5)$$

which is a circulant matrix, where $\rho_i = E[\theta_m \theta_{m-i}] / E[\theta_m^2]$ and E denotes expectation.

3. Joint Data Detection & PHN Cancellation

3.1. Minimum Mean Square Prediction Error

Assuming that (2) and (3) (or (4)) are the true representation of the underlying system, a set of consistent model parameter estimates for θ_m and S_k can be obtained from:

$$\begin{cases} r_m = \frac{1}{\sqrt{N}} e^{j\hat{\theta}_m} \sum_{k=0}^{N-1} h_k \hat{S}_k e^{j2\pi km/N} + \xi_m, & m = 0, \dots, N-1 \\ \hat{\theta}_m = \hat{\theta}_{m-1} + \omega_m, & m = 0, \dots, N-1 \end{cases} \quad (6)$$

for the Wiener PHN, or:

$$\begin{cases} r_m &= \frac{1}{\sqrt{N}} e^{j\hat{\theta}_m} \sum_{k=0}^{N-1} h_k \hat{S}_k e^{j2\pi km/N} + \xi_m, & m = 0, \dots, N-1 \\ \hat{\theta}_m &= \sum_{i=1}^K a_i \hat{\theta}_{m-i} + \omega_m, & m = 0, \dots, N-1 \end{cases} \quad (7)$$

for the Gaussian PHN, where $\hat{\theta}_m$, \hat{S}_k are the estimates of θ_m and S_k respectively, ξ_m and ω_m are referred to the prediction errors of the channel model and the phase model respectively.

Letting $\mathbf{s} = [\hat{S}_0, \dots, \hat{S}_{N-1}]^T$, $\mathbf{r} = [r_0, \dots, r_{N-1}]^T$ and the channel model prediction error vector be $\boldsymbol{\xi} = [\xi_0, \dots, \xi_{N-1}]^T$, we obtain a vector representation of the first equation of (6) as:

$$\mathbf{r} = \text{diag}\{e^{j\hat{\theta}_1}, \dots, e^{j\hat{\theta}_{N-1}}\} \times \mathbf{P} \times \text{diag}\{h_1, \dots, h_{N-1}\} \times \mathbf{s} + \boldsymbol{\xi} \quad (8)$$

where \mathbf{P} is an N by N IFFT matrix with its element at n th row and k th column being given by $p_{m,k} = \frac{1}{\sqrt{N}} e^{j(2\pi(m-1)(k-1))/N}$. It is known that $\mathbf{P}^H \mathbf{P} = \mathbf{I}$, where the superscript H denotes Hermitian transpose, and \mathbf{I} is an identity matrix with appropriate dimension.

Next we use the Pade approximation:

$$e^{j\hat{\theta}_m} \approx \frac{2 + j\hat{\theta}_m}{2 - j\hat{\theta}_m} \quad (9)$$

to replace the term $e^{j\hat{\theta}_m}$. We note that (9) only holds for sufficiently small $\hat{\theta}_m$. As shown in Table 1, the Pade approximation provides an approximation capability far superior to

Table 1: A comparison between the Pade and first order Taylor approximations

$\ \theta\ $ (rad)	0.02	0.04	0.06	0.08	0.1
$\ e^{j\theta} - \frac{2+j\theta}{2-j\theta}\ $	6.7×10^{-7}	5.33×10^{-6}	1.80×10^{-5}	4.26×10^{-5}	8.32×10^{-5}
$\ e^{j\theta} - (1 + j\theta)\ $	0.0002	0.0008	0.0018	0.0032	0.0050
$\arg[e^{j\theta} - \frac{2+j\theta}{2-j\theta}]$ (rad)	1.5908	1.6106	1.6308	1.6508	1.6708
$\arg[e^{j\theta} - (1 + j\theta)]$ (rad)	3.1349	3.1283	3.1216	3.1149	3.1883

the first order Taylor approximation of $e^{j\hat{\theta}_m} \approx 1 + j\hat{\theta}_m$. In practice, the approximation errors due to Pade approximation can be ignored, so that we have:

$$\xi_m = \frac{(2 - j\hat{\theta}_m)r_m}{2} - \frac{1}{2\sqrt{N}}(2 + j\hat{\theta}_m) \sum_{k=0}^{N-1} h_k \hat{S}_k e^{j2\pi km/N} + j\hat{\theta}_m \xi_m / 2. \quad (10)$$

Further denoting $\boldsymbol{\theta} = [\hat{\theta}_0, \dots, \hat{\theta}_{N-1}]^T$ and letting the phase model prediction error vector be $\boldsymbol{\omega} = [\omega_0, \dots, \omega_{N-1}]^T$, we represent (10) and the second equation of (6) or (7) in vector forms as:

$$\begin{cases} \boldsymbol{\xi} = \mathbf{z} - \mathbf{Q}\boldsymbol{\theta} - \mathbf{Q}_\xi \boldsymbol{\theta} \\ \boldsymbol{\omega} = \boldsymbol{\Phi}\boldsymbol{\theta} \end{cases} \quad (11)$$

respectively, where

$$\mathbf{z} = \begin{bmatrix} r_0 - \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} h_k \hat{S}_k \\ \vdots \\ r_{N-1} - \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} h_k \hat{S}_k e^{j2\pi k(N-1)/N} \end{bmatrix}, \quad (12)$$

$\mathbf{Q}_\xi = \text{diag}\{(-j\xi_1)/2, \dots, (-j\xi_{N-1})/2\}$ which includes the terms of the channel model

prediction error, $\mathbf{Q} = \text{diag}\{q_0, \dots, q_{N-1}\}$ which is the prediction errors free matrix with

$$q_m = \frac{j}{2} \left[\sum_{k=0}^{N-1} h_k \hat{S}_k e^{j2\pi km/N} / \sqrt{N} + r_m \right], \quad m = 0, \dots, N-1 \quad (13)$$

where:

$$\Phi = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}, \quad (14)$$

for the Wiener PHN, and:

$$\Phi = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ -a_1 & 1 & 0 & 0 & 0 \\ -a_2 & -a_1 & 1 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & -a_K & \dots & -a_1 & 1 \end{bmatrix}, \quad (15)$$

for the Gaussian PHN.

In order to jointly estimate the data symbol and the PHN, we construct a cost function based on minimum mean square prediction error (MMSPE) as:

$$\begin{aligned} J &= E[\xi_m^H \xi_m + (\sigma^2 / \sigma_\varepsilon^2) \omega_m^H \omega_m] \\ &\approx \frac{1}{N} [\boldsymbol{\xi}^H \boldsymbol{\xi} + (\sigma^2 / \sigma_\varepsilon^2) \boldsymbol{\omega}^H \boldsymbol{\omega}], \end{aligned} \quad (16)$$

where the approximation comes from the “expectation” being replaced by the “average” in practice. It is clear from (16) that the normalized errors from both the channel model and the PHN model are included. It is aimed that minimizing (16) gives $\hat{S}_k \rightarrow S_k$ and $\hat{\theta}_m \rightarrow \theta_m$.

3.2. An Iterative Approach

Analyzing (8) and (11) suggests that J is quadratic with respect to \mathbf{s} for a fixed $\boldsymbol{\theta}$, and vice versa. This means that the cost function J can be reduced by setting $\frac{\partial}{\partial \mathbf{s}} J = \mathbf{0}$ and $\frac{\partial}{\partial \boldsymbol{\theta}} J = \mathbf{0}$ alternatively.

First, setting $\frac{\partial}{\partial \boldsymbol{\theta}} J = \mathbf{0}$ and making use of (11) yields:

$$(\mathbf{Q}^H + \mathbf{Q}_\xi^H)(\mathbf{z} - \mathbf{Q}\boldsymbol{\theta} - \mathbf{Q}_\xi\boldsymbol{\theta}) - \frac{\sigma^2}{\sigma_\xi^2} \boldsymbol{\Phi}^H \boldsymbol{\Phi} \boldsymbol{\theta} = \mathbf{0}. \quad (17)$$

Taking the expected value in terms of channel noise ξ_m on (17) gives:

$$(\mathbf{Q}^H \mathbf{Q} + \frac{\sigma^2}{\sigma_\xi^2} \boldsymbol{\Phi}^H \boldsymbol{\Phi}) \boldsymbol{\theta} = \mathbf{Q}^H \mathbf{z} - \frac{j\sigma^2}{2} \mathbf{1}, \quad (18)$$

where $\mathbf{1}$ is a vector with all elements as 1, and it is assumed that at optimality $E(\mathbf{Q}_\xi^H \mathbf{Q}) = \mathbf{0}$, $E(\mathbf{Q}_\xi^H \mathbf{Q}_\xi) = \sigma^2 \mathbf{I}/4$ and $E(\mathbf{Q}_\xi^H \mathbf{z}) = E(\mathbf{Q}_\xi^H \mathbf{Q}_\xi) \boldsymbol{\theta} + E(\mathbf{Q}_\xi^H \boldsymbol{\xi}) = \sigma^2 \boldsymbol{\theta}/4 - j\sigma^2 \mathbf{1}/2$.

Similarly letting $\frac{\partial}{\partial \mathbf{s}} J = \mathbf{0}$ and making use of (8) gives:

$$\text{diag}\{h_0^H, \dots, h_{N-1}^H\} \times \mathbf{P}^H \times \text{diag}\{e^{-j\hat{\theta}_0}, \dots, e^{-j\hat{\theta}_{N-1}}\} \boldsymbol{\xi} = \mathbf{0}, \quad (19)$$

or equivalently:

$$\text{diag}\{h_0^2, \dots, h_{N-1}^2\} \mathbf{s} = \text{diag}\{h_0^H, \dots, h_{N-1}^H\} \times \mathbf{P}^H \times \text{diag}\{e^{-j\hat{\theta}_0}, \dots, e^{-j\hat{\theta}_{N-1}}\} \mathbf{r}. \quad (20)$$

The solutions of (18) and (20) are then substituted into each other to form an iterative procedure.

3.3. The overfitting

Although in theory it always decreases the cost function J to some extent, the above iterative approach may converge to a trivial solution due to the problem of the overfitting. The overfitting is a common issue in many parameter estimation problems, and can be serious in the case of the joint data detection and PHN cancellation. In general, the overfitting may be analyzed mathematically in terms of the high variance of the parameter estimates as a result of several possible factors, or a combination of them. There are three major factors that inflate the variance of the parameter estimates: (i) the small number of data samples, e.g. N ; (ii) the high condition number of the associated regression matrix, e.g. $C = (\max\{h_k^H h_k, \forall k\}) / (\min\{h_k^H h_k, \forall k\})$; (iii) the high variance of noise, e.g. σ^2 .

The joint data detection and PHN cancellation problem is prone to overfitting due to the fact that this is not a well-posed one, especially when the signal-to-noise-ratio (SNR) is low and/or the channel is harsh. To be specific, for an OFDM symbol with N subcarriers, there are N data symbols and N phase errors to be determined from $2N$ number of normal equations including N from the receiving data model and another N from the PHN model. Since the PHN varies from one OFDM symbol to another even for slow

fading channels, the PHN must be mitigated at every symbol. This describes a very special case of parameter estimation problem: unlike the standard parameter estimation in a linear regression model whereby the variance of parameter estimates can be improved by increasing the number of data samples, here the number of unknown parameters always equals the number of the “observation” samples. So there exists a dilemma between (i) and (ii) above, i.e., as N increases C increases too. A direct consequence is the so-called overfitting, making the estimated data symbol sufficiently far away from the true symbols.

Our solution is to apply the hard decision, i.e. map the estimated symbol to the nearest symbol in the M-QAM constellation, at the end of each iteration. The hard decision can effectively filter the noise out of the symbol estimates and remove the associated uncertainties due to the overfitting which will be otherwise carried forward over the iterations. It is true the effectiveness of the hard-decision procedure requires most estimated symbols be within the “correct” regions in the constellation. Fortunately this condition can normally be satisfied since the iterative approach should begin with the traditional OFDM symbol detection ignoring the PHN so that the BER performance at the initialization is already at a fair good level. We note that if the traditional OFDM detection cannot provide a good enough BER performance, it will be very difficult, if not impossible, to remove the PHN.

The use of the hard decision at every iteration here describes a major difference to the approaches in [8] where it was declared the no use of the hard decision was one of its advantages. In fact, we found that, although it was not explicitly stated, the hard decision must be applied at least at the initialization stage in the approaches proposed in [8], as

otherwise the iteration procedure may not work at all.

3.4. The Algorithm

In summary, the proposed minimum mean square prediction error (MMSPE) is shown below, where we assume the total number of iterations is N_s , $\boldsymbol{\theta}^{(l)}$ and $\mathbf{s}^{(l)}$ are the estimates of $\boldsymbol{\theta}$ and \mathbf{s} at the iteration step l respectively, and accordingly the resultant \mathbf{Q} and \mathbf{z} are denoted by $\mathbf{Q}(\mathbf{s}^{(l-1)})$ and $\mathbf{z}(\mathbf{s}^{(l-1)})$ respectively.

Initialization

$$\boldsymbol{\theta}^{(0)} = \mathbf{0}$$

$$\mathbf{s}^{(0)} = \text{diag}\{h_0^{-1}, \dots, h_{N-1}^{-1}\} \mathbf{P}^H \mathbf{r}$$

Replace $\mathbf{s}^{(0)}$ by its hard decision.

For $l = 1, 2, \dots, N_s$

$$\boldsymbol{\theta}^{(l)} = \left([\mathbf{Q}(\mathbf{s}^{(l-1)})]^H \mathbf{Q}(\mathbf{s}^{(l-1)}) + \frac{\sigma^2}{\sigma_\varepsilon^2} \boldsymbol{\Phi}^H \boldsymbol{\Phi} \right)^{-1} \cdot \text{Re}\{[\mathbf{Q}(\mathbf{s}^{(l-1)})]^H \mathbf{z}(\mathbf{s}^{(l-1)})\} \quad (21)$$

$$\mathbf{s}^{(l)} = \text{diag}\{h_0^{-1}, \dots, h_{N-1}^{-1}\} \mathbf{P}^H \cdot \text{diag}\{e^{-j\hat{\theta}_0^{(l-1)}}, \dots, e^{-j\hat{\theta}_{N-1}^{(l-1)}}\} \mathbf{r} \quad (22)$$

Replace $\mathbf{s}^{(l)}$ by its hard decision.

End

4. Complexity Reduction

The main complexity of the proposed approach comes from the matrix inversion in (21). Below, we show that a fast algorithm can be developed by considering the spe-

cial structure of the matrixes in this case. For simplicity of expression, we let $\mathbf{A} = [\mathbf{Q}(\mathbf{s}^{(l-1)})]^\text{H} \mathbf{Q}(\mathbf{s}^{(l-1)}) + \frac{\sigma_z^2}{\sigma_e^2} \mathbf{\Phi}^\text{H} \mathbf{\Phi}$, and $\mathbf{b} = \text{Re}\{[\mathbf{Q}(\mathbf{s}^{(l-1)})]^\text{H} \mathbf{z}(\mathbf{s}^{(l-1)})\}$. Then (21) can be rewritten in a standard linear equation form as:

$$\mathbf{A}\boldsymbol{\theta} = \mathbf{b}, \quad (23)$$

where we ignored the iteration index l for simplicity.

4.1. Wiener PHN

For the Wiener PHN, the matrix \mathbf{A} in (23) has a special form of:

$$\mathbf{A} = \begin{bmatrix} d_1 & u_1 & \cdots & 0 \\ l_1 & d_2 & \ddots & \vdots \\ & \ddots & \ddots & \ddots \\ \vdots & & \ddots & \ddots & u_{N-1} \\ 0 & \cdots & & l_{N-1} & d_N \end{bmatrix} \in \mathbf{Re}^{N \times N}. \quad (24)$$

The fast algorithm is then constructed as two stages: $N - 1$ number of successive Givens rotations followed by backward substitutions.

First, denoting $\tilde{\mathbf{A}} = [\mathbf{A} \ \mathbf{b}]$, we apply $N - 1$ number of successive Givens rotations in turn on the i th and $(i + 1)$ th rows of $\tilde{\mathbf{A}}$ to zero the elements in the lower triangular part of

\mathbf{A} for $i = 1, \dots, N - 1$. To be specific, after the $(i - 1)$ th Givens rotation, we have:

$$\tilde{\mathbf{A}}(i) = \begin{bmatrix} d_i & u_i & 0 & b_i \\ l_i & d_{i+1} & u_{i+1} & b_{i+1} \end{bmatrix}, \quad i = 1, \dots, N - 2, \quad (25)$$

which consists of row i to $i + 1$ and column i to $i + 2$ as well as column $(N + 1)$ of \mathbf{A} .

For the last two rows, we have:

$$\tilde{\mathbf{A}}(N - 1) = \begin{bmatrix} d_{N-1} & u_{N-1} & b_{N-1} \\ l_{N-1} & d_N & b_N \end{bmatrix} \quad (26)$$

A sequence of Givens rotations, $\mathbf{G}(\beta_i)$, is applied to (25) and (26) so that:

$$\tilde{\mathbf{A}}(i) \leftarrow \mathbf{G}(\beta_i)\tilde{\mathbf{A}}(i), \quad i = 1, \dots, N - 1 \quad (27)$$

where

$$\mathbf{G}(\beta_i) = \begin{bmatrix} c(\beta_i) & -s(\beta_i) \\ s(\beta_i) & c(\beta_i) \end{bmatrix} \quad (28)$$

with $c(\beta_i) = \cos(\beta_i) = d_i/\sqrt{d_i^2 + l_i^2}$ and $s(\beta_i) = \sin(\beta_i) = -l_i/\sqrt{d_i^2 + l_i^2}$. After $N - 1$ number of Givens rotations, \mathbf{A} in (23) becomes an upper triangular matrix so that $\tilde{\mathbf{A}}$ has

the form of

$$\tilde{\mathbf{A}} = \begin{bmatrix} d_1 & u_1 & u'_1 & \cdots & 0 & b_1 \\ 0 & d_2 & u_2 & u'_2 & \vdots & b_2 \\ & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & & \ddots & d_{N-1} & u_{N-1} & b_{N-1} \\ 0 & \cdots & & 0 & d_N & b_N \end{bmatrix} \in \mathbf{Re}^{N \times (N+1)} \quad (29)$$

in which we maintain the notations d_i , u_i and b_i for the simplicity of expression.

Finally, initializing $\theta_N = b_N/d_N$ and $\theta_{N-1} = (b_{N-1} - u_{N-1}\theta_N)/d_{N-1}$, the backward substitution procedure is used to solve (23) as:

$$\theta_{N-i} = (b_{N-i} - u_{N-i}\theta_{N-i+1} - u'_{N-i}\theta_{N-i+2})/d_{N-i}, \quad (30)$$

for $i = 2, \dots, N-1$. The proposed solver has a low computational cost at the level of $\mathcal{O}(N)$, compared to a standard linear square equation solver of $\mathcal{O}(N^3)$. Specifically the required costs are that of $16N-1$ flops and $N-1$ square root for Givens rotations, adding $(5N-6)$ flops for the backward substitution.

In [8], the conjugate gradient (CG) method was used to simplify the matrix inversion. As an iterative approach, the CG method requires N number of iterations to converge with the complexity being at $\mathcal{O}(N^2)$. Although in practice i ($i < N$) number of iterations may be accurate enough for the CG method so that the complexity is reduced to $\mathcal{O}(iN)$ (e.g. i is chosen as 8 in [8]), the solution is an approximate. On the contrary, the Givens rotation

based approach not only gives exact solution with complexity being fixed at $\mathcal{O}(N)$, but also is more numerical stable and has a nature link to parallel processing for the IC design.

4.2. Gaussian PHN

For the Gaussian PHN, the matrix \mathbf{A} in (23) is a central banded matrix with the non-zeros elements being only at the $2K + 1$ number of diagonals centering at the main diagonal. Similarly, we can also apply the Givens rotation but with complexity being at $\mathcal{O}(K^2N)$ or $\mathcal{O}(N)$ for $K \ll N$.

It has been shown ([8, 13]) that, as $N \rightarrow \infty$, θ_m can be approximated by a first-order autoregressive process, i.e. $K = 1$. Then we can have a Givens rotation based fast process exactly the same as that for the Wiener PHN derived above. In many systems, however, $K = 1$ may not be sufficient accurate, and a larger K is necessary. A more flexible way is to apply the Akaike's Information Criterion (AIC) to determine the order size K [11]. This is very useful since K may vary with different systems. Specifically, applying AIC on the Gaussian PHN gives:

$$\text{AIC}(K) = \log \left(\frac{|\boldsymbol{\theta}_M - \boldsymbol{\Theta}_K \cdot \mathbf{a}_K|^2}{M} \right) + \frac{2K}{M}, \quad (31)$$

where K is the assumed order of the autoregressive process, M is the total number of PHN samples observed, $\boldsymbol{\Theta}_K$ is an M by K regression matrix whose i th row is given by $[\theta_{i-1}, \dots, \theta_{i-K}]$, and:

$$\mathbf{a}_K = (\boldsymbol{\Theta}_K^T \boldsymbol{\Theta}_K)^{-1} \boldsymbol{\Theta}_K^T \boldsymbol{\theta}_M, \quad (32)$$

which is the LS solution of the coefficient vector assuming the order size is K with $\boldsymbol{\theta}_M =$

$[\theta_1, \dots, \theta_M]^T$. We note that, since using the AIC to determine the order size K and the coefficient vector \mathbf{a}_K can be done off-line due to the stationarity of the Gaussian PHN, it imposes little complexity on the system.

As an illustration, we generate a Gaussian PHN according to the Matlab code recommended for the IEEE 902.11g [13]. Similar to that in [8], the generated Gaussian PHN has a standard deviation of 3° , and passes through a single pole Butterworth filter with 3dB bandwidth being 100KHz. The $AIC(K)$ defined in (31) is plotted in Fig. 1, where it is clearly shown that $AIC(K)$ is minimized at $K = 11$ so that the autoregressive process with order 11 can well model this Gaussian PHN. Obviously in this case, the first order autoregressive process is correct to model the PHN.

5. Simulation

In the simulation below, we compare the proposed MMSPE algorithm with the ICM algorithm in [8]. The ICM algorithm was shown to have significantly better performance than the classic approach where only the common PHN is considered [6], which again justifies the complexity in terms of the benefits from the joint approach of data detection and PHN cancellation. Note that the ICM algorithm is chosen as a comparison simply because all algorithms proposed in [8] have close performance. Moreover, since the comparison results based on the Wiener PHN and Gaussian PHN are similar, only the results for the Wiener PHN are reported here.

To be specific, the parameters of the simulation system similar are set as follows:

1. An OFDM symbol size of 64 subcarriers;

2. A Rayleigh mutlipath fading channel with a delay of 64 taps and an exponentially decreasing power delay profile having a decay constant of 40 tap;
3. The Wiener PHN generated by random walk as was described in (3) with $\sigma_\varepsilon^2 = 0.5^\circ$.

Without losing generality, we assume the “random walk” process is applied on the data symbols of an OFDM symbol from one sub-carrier to another sub-carrier, and is reset for every OFDM symbol. This assumption, which was also used in the simulation in [8], is reasonable in practice as the oscillators can always be reset for every OFDM symbol.

In Fig. 2, we compares the BER performance for different approaches with 64-QAM modulation, where there are 4 iterations for both the ICM and the proposed MMSPE algorithms. It is shown that both the ICM and the MMSPE algorithms have significantly better performance than that without any PHN cancellation. In fact, there is an error floor in the BER performance for the no PHN cancellation approach, because when the SNR is sufficient large, the PHN will dominate the BER performance. It is also shown that the MMSPE algorithm can effectively further improve the performance compared to the ICM approach. We highlight that the hard decision has also been taken at the initial stage for the ICM algorithm as otherwise it will converge to a trivial solution. This clearly indicates that the hard decision process can effectively combat the “overfitting” problem which is a key point in the OFDM joint data detection and PHN cancellation. It also suggest a direction for future research such that the PHN cancellation performance may be further improved by better solving the “overfitting” problem.

Fig. 3 compares the PHN tracking performance between the ICM and MMSPE algo-

rithms by showing their PHN mean square errors defined as:

$$\text{MSE}(\boldsymbol{\theta}) = \frac{|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}|^2}{|\boldsymbol{\theta}|^2} \quad (33)$$

It is clearly shown that the proposed MMSPE approach estimates the PHN significantly more accurately.

For simpler modulations such as 32/16/8-QAM, the system is less sensitive to the the PHN and we expect less performance benefit from the proposed MMSPE algorithm. As an illustration, Fig. 4 shows the BER performance for the 16-QAM modulation, where the performance improvement of the proposed MMSPE algorithm to the ICM algorithm is obviously less significant than that with the 64-QAM.

6. Conclusions

This paper proposed an algorithm to jointly detect the OFDM data symbol and cancel the PHN based on minimum mean square prediction error. We particularly highlighted the problem of the overfitting and showed that the hard decision procedure is effective in combatting it. Compared with existing approaches, the proposed algorithm uses a more accurate Pade approximation to represent the PHN, and apply the Givens rotation to simply the complexity. The resultant approach is not only more accurate but also has less complexity than existing approaches. This result from this paper also suggests an interesting research topic in the future that the PHN cancellation performance can be future improved by better handling the overfitting problem.

Acknowledgement

The authors would like to thank Dr D. D. Lin and Professor T. J. Lim from the University of Toronto for their kind help in setting up the the simulation parameters in this article.

- [1] M. Speth, S. A. Fechtel, G. Fock, and H. Meyr, “Optimum receiver design for wireless broadband systems using OFDM–Part I,” *IEEE Transactions on Communications*, vol. 47, no. 11, pp. 1668–1677, 1999.
- [2] —, “Optimum receiver design for OFDM-based broadband transmission –Part II: a case study,” *IEEE Transactions on Communications*, vol. 49, no. 4, pp. 571–578, 2001.
- [3] T. Pollet, M. V. Bladel, and M. Moeneclaey, “BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise,” *IEEE Trans. on Commun.*, vol. 43, no. 2, pp. 191–193, Feb. 1995.
- [4] T. Tomba, “On the effect of wiener phase noise in OFDM systems,” *IEEE Transactions on Communications*, vol. 46, no. 5, pp. 580–583, 1998.
- [5] K. Nikitopoulos and A. Polydoros, “Phase-impairment effects and compensation algorithms for OFDM systems,” *IEEE Trans. on Commun.*, vol. 53, no. 4, pp. 698–707, April 2005.
- [6] S. Wu and Y. Bar-Ness, “OFDM systems in the presence of phase noise: conse-

- quences and solutions,” *IEEE Transactions on Communications*, vol. 52, no. 11, pp. 1988–1996, 2004.
- [7] S. Wu, P. Liu, and Y. Bar-Ness, “Phase noise estimation and mitigation for ofdm systems,” *IEEE Trans. on Wireless Communications*, vol. 5, no. 12, pp. 3616–3625, Dec. 2006.
- [8] D. D. Lin and T. J. Lim, “The variational inference approach to joint data detection and phase noise estimation,” *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1862–1874, 2007.
- [9] D. J. C. MacKay, “Bayesian interpolation,” *Neural Computation*, vol. 4, no. 3, pp. 415–447, 1992.
- [10] L. Piazzo and P. Mandarini, “Analysis of phase noise effects in OFDM modems,” *IEEE Transactions on Communications*, vol. 50, no. 10, pp. 1696–1705, 2002.
- [11] H. Akaike, “A new look at the statistical model identification,” *IEEE Trans Automat. Contr*, vol. AC-19, pp. 716 – 723, Dec. 1974.
- [12] M. Stone, “Cross validatory choice and assessment of statistical predictions,” *Journal of the Royal Statistical Society, Series B*, vol. 36, pp. 117–147, 1974.
- [13] “IEEE P802.11-Task Group G., 2000, phase noise matlab model,” http://www.grouper.ieee.org/groups/802/11/Reports/tgg_update.htm.

List of Figures

Figure 1: AIC of the generated Gaussian PHN, where K is the order size of the AR model for the Gaussian PHN.

Figure 2: BER performance for 64-QAM modulation.

Figure 3: PHN tracking performance for 64-QAM modulation.

Figure 4: BER performance for 16-QAM modulation.