

Channel, Phase Noise, and Frequency Offset in OFDM Systems: Joint Estimation, Data Detection, and Hybrid Cramér-Rao Lower Bound

Omar H. Salim, Ali A. Nasir, *Student Member, IEEE*, Hani Mehrpouyan, *Member, IEEE*, Wei Xiang, *Senior Member, IEEE*, Salman Durrani, *Senior Member, IEEE*, and Rodney A. Kennedy, *Fellow, IEEE*

Abstract

Oscillator phase noise (PHN) and carrier frequency offset (CFO) can adversely impact the performance of orthogonal frequency division multiplexing (OFDM) systems since they can result in inter carrier interference and rotation of the signal constellation. In this paper, we propose an expectation conditional maximization (ECM) based algorithm for joint estimation of channel, PHN, and CFO in OFDM systems. We outline the signal model for the estimation problem and derive the hybrid Cramér-Rao lower bound (HCRB) for the joint estimation problem. Next, we propose an iterative receiver based on an *extended Kalman filter* for joint data detection and PHN tracking. Numerical results show that compared to existing algorithms, the performance of the proposed ECM-based estimator is closer to the derived HCRB and outperforms the existing estimation algorithms at moderate to high signal-to-noise ratio (SNR). In addition, the combined estimation algorithm and iterative receiver are more computationally efficient than existing algorithms and result in improved average uncoded and coded bit error rate (BER) performance.

Index Terms

OFDM, channel estimation, phase noise, frequency offset, Bayesian, hybrid Cramér-Rao lower bound, Kalman filter, data detection.

Omar H. Salim and Wei Xiang are with the School of Mechanical and Electrical Engineering, University of Southern Queensland, Australia. Ali A. Nasir, Salman Durrani, and Rodney Kennedy are with Research School of Engineering, The Australian National University, Canberra, Australia. Hani Mehrpouyan is with the Department of Electrical and Computer Engineering, California State University, Bakersfield. Emails: OmarHazim.Salim@usq.edu.au, ali.nasir@anu.edu.au, hani.mehr@ieee.org, wei.xiang@usq.edu.au, salman.durrani@anu.edu.au, and rodney.kennedy@anu.edu.au. This research was partially supported under Australian Research Council's Discovery Projects funding scheme (project number DP140101133).

I. INTRODUCTION

A. Motivation and Literature Survey

Orthogonal frequency division multiplexing (OFDM) is a powerful multi-carrier modulation technique for increasing the bandwidth efficiency of wireless communication systems. By converting a frequency-selective channel into multiple frequency-flat subchannels, OFDM can mitigate the detrimental effects of frequency-selective fading [1, 2]. Hence, OFDM has been adopted by existing and future wireless local area network (WLAN) standards such as IEEE 802.11ac and IEEE 802.11ad [3, 4]. However, OFDM systems are much more sensitive to synchronization errors than single-carrier systems.

In practice, OFDM systems require timing offset estimation, carrier frequency offset (CFO) estimation, phase noise (PHN) tracking as well as channel estimation. Timing synchronization for OFDM systems has been well investigated for the past two decades [5, 6]. Compared to timing offsets, OFDM is very sensitive to CFO and PHN, which arise due to instabilities and the thermal noise in the local oscillator [7], respectively. CFO and time varying PHN result in a common phase error (CPE) and inter-carrier interference (ICI) at the receiver, degrading the performance of OFDM systems [8–13]. In particular, the impact of PHN in systems operating at higher carrier frequencies, e.g., V-band/60 GHz and E-band/70–80 GHz, can be even more profound [14]. Thus, as wireless communication systems and standards, e.g., IEEE 802.11ad, migrate to millimeter-wave frequencies to take advantage of the large bandwidth in this band and adopt higher order modulations and closely spaced subcarriers to achieve higher spectral efficiencies, it is increasingly important to develop efficient and accurate estimation and detection algorithms for compensating the effect of CFO and PHN in OFDM systems.

In order to jointly estimate channel, CFO, and time varying PHN, training signals are used in OFDM systems. In the context of point-to-point systems, joint channel and CFO estimation based on the expectation-maximization (EM) approach is proposed in [15]. However, in [15], the authors do not take the effect of PHN into account. In [8] and [16], a MAP estimator is used for joint estimation of channel, CFO, and PHN. However, the estimation approach in [8] and [16] is based on a small angle approximation (single-order Taylor series expansion of PHN), that adversely affects the performance of the estimation and data detection algorithms, especially for higher order modulations. In addition, as shown in this

paper, the approach in [8] and [16] can be computationally very complex. Recently, the authors in [17] proposed a joint channel, CFO, and PHN estimation algorithm based on the sequential Monte Carlo and EM approaches. However, as explained in [18], the estimation complexity of Monte Carlo based method in [17] is very high. More importantly, in [8, 16, 17], the hybrid Cramér-Rao lower bound (HCRB) for the joint estimation of channel impulse response (CIR), PHN, and CFO in OFDM systems is not derived and the performances of the proposed estimators are not benchmarked against their respective estimation performance bounds. Recently, the problem of joint estimation of channel, CFO, and PHN was considered in the context of OFDM relay networks in [19]. However, the approach in [19] is also based on the *maximum a posteriori* (MAP) criterion, which is computationally very complex.

Given the time-varying nature of PHN, it needs to be tracked not only during the training interval but also during the data transmission interval. Hence, following the training period, a receiver structure for joint data detection and PHN mitigation in the data transmission period is required. In the existing literature, joint data detection and PHN mitigation is analyzed in [13, 20–22]. However, the algorithms proposed in [13] and [21] are based on the assumption of perfect knowledge of channel and CFO. Moreover, the PHN tracking and data detection approach presented in [20] is computationally complex and suffers from performance degradation for higher order modulations. The PHN tracking in [22] requires the application of pilots throughout an OFDM symbol, which adversely affects the bandwidth efficiency of the system. In addition, our simulations show that the approach in [22] is outperformed by the receiver structure proposed here.

B. Contributions

In this paper, a computationally efficient training based approach for joint channel, CFO, and PHN estimation in OFDM systems is presented. In order to detect the data symbols in the presence of time-varying PHN, an iterative receiver is proposed. The major contributions of this paper can be summarized as follows:¹

- We propose an expectation conditional maximization (ECM) based estimator for jointly obtaining the channel, CFO, and PHN parameters in OFDM systems. The ECM based estimation is carried out in

¹This paper was in part presented at IEEE International Workshop on Signal Processing Advances in Wireless Communications [23].

two steps. In the expectation or E-step, an *extended Kalman filter* (EKF) based estimator is utilized to accurately track the PHN over the training OFDM symbol. During the maximization or M-step, the channel and CFO parameters are estimated by minimizing the derived negative log likelihood function (cf. (6)).

- We derive an expression for the HCRB for the joint estimation of the channel, CFO and PHN in OFDM systems. Simulation results show that, compared to the existing algorithms in the literature, the mean square error (MSE) of the proposed algorithm is closer to the HCRB and outperforms the existing estimation algorithms at moderate to high signal-to-noise ratio (SNR).
- We propose a new iterative algorithm based on the EKF for data detection and tracking the unknown time-varying PHN throughout the OFDM data packet.
- We show that the proposed estimation and detection algorithms are computationally efficient, compared to existing algorithms in the literature. In addition, the proposed estimation and detection algorithms outperform existing algorithms in terms of uncoded and coded bit error rate (BER) performance.

C. Organization

The remainder of the paper is organized as follows. Section II describes the system model and the assumptions used in this work. Section III derives the HCRB for joint channel, CFO and PHN estimation in OFDM systems. Section IV describes the proposed ECM based estimator while Section V presents the proposed receiver for joint data detection and PHN tracking. Section V analyzes the complexity of the proposed estimation and data detection algorithms and compares it with existing schemes. Section VI provides numerical and simulation results. Finally, Section VII concludes the paper.

D. Notations

Superscripts $(\cdot)^*$, $(\cdot)^H$, and $(\cdot)^T$ denote the conjugate, the conjugate transpose, and the transpose operators, respectively. Bold face small letters, e.g., \mathbf{x} , are used for vectors, bold face capital alphabets, e.g., \mathbf{X} , are used for matrices, and $[\mathbf{X}]_{x,y}$ represents the entry in row x and column y of \mathbf{X} . \mathbf{I}_X , $\mathbf{0}_{X \times X}$, and $\mathbf{1}_{X \times X}$ denote the $X \times X$ identity, all zero, and all 1 matrices, respectively. The notation $\mathbf{X}(n_1:n_2, m_1:m_2)$

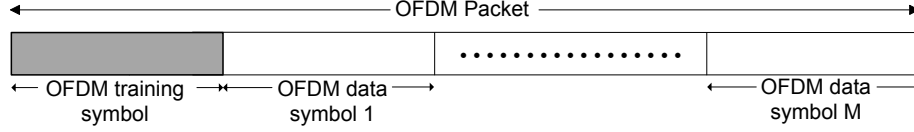


Fig. 1. Timing diagram for transmission of training and data symbols within an OFDM packet.

is used to denote a submatrix of \mathbf{X} from row n_1 to row n_2 and from column m_1 to column m_2 . $|\cdot|$ is the absolute value operator, $|\mathbf{x}|$ denotes the element-wise absolute value of a vector \mathbf{x} , and $\text{diag}(\mathbf{x})$ is used to denote a diagonal matrix, where the diagonal elements are given by vector \mathbf{x} . $\mathbb{E}_{x,y}[\cdot]$ denotes the expectation over x and y . $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote the real and imaginary parts of a complex quantity, respectively. $\nabla_{\mathbf{x}}$ and $\Delta_{\mathbf{y}}^{\mathbf{x}}$ represent the first and the second-order partial derivatives operator, i.e., $\nabla_{\mathbf{x}} = [\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_N}]^T$ and $\Delta_{\mathbf{y}}^{\mathbf{x}} = \nabla_{\mathbf{y}} \times \nabla_{\mathbf{x}}^T$. $\mathcal{N}(\mu, \sigma^2)$ and $\mathcal{CN}(\mu, \sigma^2)$ denote real and complex Gaussian distributions with mean μ and variance σ^2 , respectively. \otimes denotes circular convolution. Finally, \dot{z} denotes the Jacobian of z .

II. SYSTEM MODEL

We consider an OFDM packet of $(M + 1)$ symbols, which consists of one training symbol and M data symbols, as illustrated in Fig. 1. In this paper, the following set of assumptions are adopted:

- A1. The channel is modeled as a slow fading frequency-selective channel, i.e., the channel is assumed to be quasi-static, which is constant and *unknown* over the OFDM packet duration and changes from packet to packet following a complex Gaussian distribution.
- A2. The time-varying PHN changes from symbol to symbol and is modeled as a Wiener process, i.e., $\theta_n = \theta_{n-1} + \delta_n$, $\forall n$, where θ_n is the PHN at the n th instant, $\delta_n \sim \mathcal{N}(0, \sigma_{\delta}^2)$ is the PHN innovation and σ_{δ}^2 is the variance of the innovation process [24],[25].
- A3. The CFO is modeled as a *deterministic unknown parameter* over a packet and is assumed to change from packet to packet.
- A4. The training symbol is assumed to be *known* at the receiver.
- A5. The timing offset is assumed to be perfectly estimated. Hence it is not considered.

Note that assumptions A1, A2, A3, and A5 are in line with previous channel, CFO and PHN estimation algorithms in [8–10, 17, 20, 22, 25]. Assumption A2 is also reasonable in many practical scenarios to

describe the behavior of practical oscillators [24] [8]. Furthermore, assumption A4 is adopted in the IEEE 802.11ac/ad standards to estimate channel and CFO in [3, 4, 8, 13, 22, 26, 27].

The complex baseband OFDM signal is given by

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} d_k e^{j2\pi kn/N} \quad n = 0, 1, \dots, N-1, \quad (1)$$

where d_k , for $k = 1, \dots, N$, is the modulated training symbol, x_n is the n th sample of the transmitted OFDM symbol, N is the number of subcarriers, and k denotes the subcarrier index. At the receiver, after removing the cyclic prefix, the complex baseband received signal, r_n , is given by

$$r_n = e^{j(\bar{\theta}_n + 2\pi n\epsilon/N)} \bar{s}_n + \eta_n \quad (2a)$$

$$= e^{j(\theta_n + 2\pi n\epsilon/N)} s_n + \eta_n, \quad (2b)$$

where $\bar{s}_n \triangleq \bar{h}_n \otimes x_n$ is the received OFDM training symbol, $\{\bar{\theta}_n\}_{n=0}^{N-1}$ is the discrete-time PHN sequence, ϵ is the normalized CFO, $\{\bar{h}_l\}_{l=0}^{L-1}$ is the channel impulse response, L is the channel length, and $\bar{h}_l \sim \mathcal{CN}(\mu_{h_l}, \sigma_{h_l}^2)$. Note that (2b) is an equivalent system model representation of (2a), where $s_n \triangleq h_n \otimes x_n$, $h_n \triangleq e^{j\bar{\theta}_0} \bar{h}_n$ and $\theta_n \triangleq \bar{\theta}_n - \bar{\theta}_0$. This equivalent system model helps to distinguish between the phase disturbance caused by PHN and the channel phase for the first sample, which in turn resolves the phase ambiguity in the joint estimation problem as indicated in Section IV. In addition, $\{\eta_n\}_{n=0}^{N-1}$ is the complex *additive white Gaussian noise* (AWGN) with zero-mean and known variance σ_w^2 . The received signal, $\mathbf{r} \triangleq [r_0, r_1, \dots, r_{N-1}]^T$, in vector form is given by

$$\mathbf{r} = \mathbf{E} \mathbf{P} \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{h} + \boldsymbol{\eta}, \quad (3)$$

where

- $\mathbf{E} \triangleq \text{diag}([1, e^{j2\pi\epsilon/N}, \dots, e^{j2\pi\epsilon/N \times (N-1)}]^T)$ is the $N \times N$ CFO matrix,
- $\mathbf{P} \triangleq \text{diag}([e^{j\bar{\theta}_0}, e^{j\bar{\theta}_1}, \dots, e^{j\bar{\theta}_{N-1}}]^T)$ is the $N \times N$ PHN matrix,
- \mathbf{F} is an $N \times N$ DFT matrix, i.e., $[\mathbf{F}]_{l,n} \triangleq (1/\sqrt{N}) e^{-j(2\pi nl/N)}$ for $n, l = 0, 1, \dots, N-1$,
- $\mathbf{D} \triangleq \text{diag}(\mathbf{d})$, $\mathbf{d} \triangleq [d_0, d_1, \dots, d_{N-1}]^T$ is the modulated training vector,
- \mathbf{W} is an $N \times L$ DFT matrix, i.e., $\mathbf{W} \triangleq \mathbf{F}(1:N, 1:L)$,
- L denotes the number of channel taps,
- $\mathbf{h} \triangleq [h_0, h_1, \dots, h_{L-1}]^T$ is the channel impulse response (CIR), and

- $\boldsymbol{\eta} \triangleq [\eta_0, \eta_1, \dots, \eta_{N-1}]^T$ is the noise vector.

III. DERIVATION OF THE HYBRID CRAMÉR-RAO BOUND

In this section, the HCRB for joint estimation of CIR, PHN, and CFO parameters in OFDM systems is derived. The HCRB is a lower bound on the joint estimation of random, e.g., PHN, and deterministic, e.g., CIR and CFO parameters. Let $\boldsymbol{\lambda} = [\boldsymbol{\theta}^T \mathbf{h}^T \epsilon]^T$ be the vector of hybrid parameters of interest, where $\boldsymbol{\theta} \triangleq [\theta_0, \dots, \theta_{N-1}]^T$ is a vector of random PHN parameters and the channel vector, \mathbf{h} , and the CFO, ϵ , are modeled as deterministic parameters. The accuracy of estimating $\boldsymbol{\lambda}$ is lower bounded by the HCRB, $\boldsymbol{\Omega}$, as [28]

$$\mathbb{E}_{\mathbf{r}, \boldsymbol{\theta} | \epsilon} \left[(\hat{\boldsymbol{\lambda}}(\mathbf{r}) - \boldsymbol{\lambda})(\hat{\boldsymbol{\lambda}}(\mathbf{r}) - \boldsymbol{\lambda})^T \right] \succeq \boldsymbol{\Omega}. \quad (4)$$

Let us define $\boldsymbol{\Omega} \triangleq \mathbf{B}^{-1}$. Here, \mathbf{B} is an $(N + L + 1) \times (N + L + 1)$ hybrid information matrix (HIM), which is determined according to the following theorem.

Theorem 1: The closed-form HIM for joint estimation of CIR, PHN, and CFO is given by

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \mathbf{b}_{13} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \mathbf{b}_{23} \\ \mathbf{b}_{31} & \mathbf{b}_{32} & \mathbf{b}_{33} \end{bmatrix}, \quad (5)$$

where

- $\mathbf{B}_{11} \triangleq \boldsymbol{\Xi}_{D_{11}} + \boldsymbol{\Xi}_{P_{11}}$ is the $N \times N$ HIM for the estimation of $\boldsymbol{\theta}$,
- $\boldsymbol{\Xi}_{D_{11}} \triangleq \text{diag}(\mathbf{h}^H \mathbf{W}^H \mathbf{D}^H \mathbf{F}) \boldsymbol{\Sigma}^{-1} \text{diag}(\mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{h})$, $\boldsymbol{\Sigma} \triangleq \sigma_w^2 \mathbf{I}_N$, and $\boldsymbol{\Xi}_{P_{11}}$ is an $N \times N$ tridiagonal matrix, where its diagonal elements are given by $\frac{1}{\sigma_\delta^2} [1, 2, \dots, 2, 1]$ and its off-diagonal elements are given by $\frac{-1}{\sigma_\delta^2} [1, \dots, 1]$,
- $\mathbf{B}_{22} \triangleq \mathbf{W}^H \mathbf{D}^H \boldsymbol{\Sigma}^{-1} \mathbf{D} \mathbf{W}$ is an $L \times L$ information matrix for the estimation of \mathbf{h} ,
- $\mathbf{B}_{12} = \mathbf{B}_{21}^T \triangleq \frac{-j}{\sigma_w^2} \text{diag}(\mathbf{h}^H \mathbf{W}^H \mathbf{D}^H \mathbf{F}) \mathbf{F}^H \mathbf{D} \mathbf{W}$,
- $\mathbf{b}_{13} = \mathbf{b}_{31}^T \triangleq \frac{1}{\sigma_w^2} \text{diag}(\mathbf{h}^H \mathbf{W}^H \mathbf{D}^H \mathbf{F} \sqrt{\mathbf{M}}) \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{h}$,
- $\mathbf{b}_{23} = \mathbf{b}_{32}^T \triangleq j \mathbf{W}^H \mathbf{D}^H \mathbf{F} \boldsymbol{\Sigma}^{-1} \sqrt{\mathbf{M}} \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{h}$, and
- $\mathbf{b}_{33} \triangleq \mathbf{h}^H \mathbf{W}^H \mathbf{D}^H \mathbf{F} \boldsymbol{\Sigma}^{-1} \mathbf{M} \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{h}$ is a scalar representing the information for the estimation of CFO, ϵ .

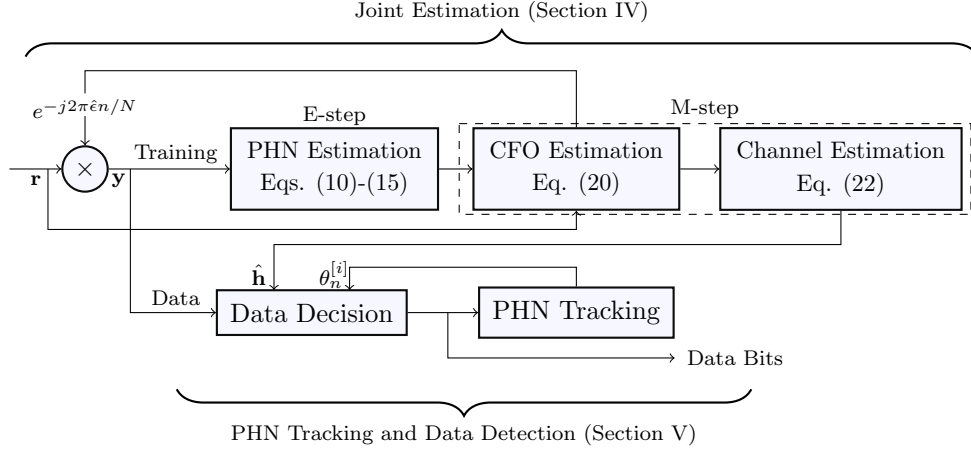


Fig. 2. Proposed estimator based on an ECM algorithm and data detection.

Proof: The proof is given in Appendix A.

Finally, the HCRB, Ω , is given by the inverse of HIM, i.e., $\Omega = \mathbf{B}^{-1}$.

IV. PROPOSED ECM BASED ESTIMATOR

In this section, an ECM based algorithm that utilizes the OFDM training symbol to jointly estimate the CIR, CFO and PHN at the receiver is derived. Joint data detection and PHN tracking during data transmission interval is analyzed in Section V.

Fig. 2 depicts the block diagram of the overall system employing the proposed ECM based estimator and the joint data detection and PHN mitigation algorithm. As illustrated in Fig. 2, the proposed ECM based estimator applies the training OFDM symbol at the beginning of each packet to estimate the CIR and CFO in the presence of PHN. Next, the data detection is performed by: 1) mitigating the impact of CFO over the length of packet by multiplying the received packet with the complex conjugate of the estimated CFO coefficients supplied by the proposed ECM estimator; 2) tracking the PHN parameters using an *iterative* algorithm that utilizes an EKF; 3) mitigating the effect of PHN over the received packet; and 4) detecting data symbols using the estimated CIR and hard decision decoding.

As shown in Fig. 2, the ECM algorithm iterates between the expectation step (E-step) and the maximization step (M-step). In the E-step, an EKF is used to update the PHN vector at the $(i+1)$ th iteration, $\theta^{[i+1]}$, using the CIR and CFO estimates, $\hat{\mathbf{h}}^{[i]}$ and $\hat{c}^{[i]}$, respectively, obtained from the previous iteration,

i.e., i th iteration. Next, in the M-step, the estimates of the CIR and CFO at the $(i + 1)$ th iteration, $\hat{\mathbf{h}}^{[i+1]}$ and $\hat{\epsilon}^{[i+1]}$, respectively are obtained.

For the given problem, the incomplete data set is given by the $N \times 1$ vector $\mathbf{s} \triangleq \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{h} = [s_0, s_1, \dots, s_{N-1}]^T$ and the received data, \mathbf{r} in (3). The complete data set is defined as $\mathbf{z} \triangleq [\mathbf{r}^T \boldsymbol{\theta}^T \mathbf{h}^T \epsilon]^T$. Following [8, 29], it is assumed that there is no prior knowledge of ϵ and \mathbf{h} and the prior distribution of θ is known. Therefore, the complete data set can be written as $\mathbf{z} \triangleq [\mathbf{r}^T \boldsymbol{\theta}^T]^T$. Moreover, the negative log likelihood function (LLF) of the complete data, $\log p(\mathbf{z}; \epsilon)$, is given by

$$\log p(\mathbf{z}; \epsilon) = C + \frac{1}{\sigma_w^2} \sum_{n=0}^{N-1} \| r_n - e^{j2\pi\epsilon n/N} e^{j\theta_n} s_n \|^2 + \log p(\theta_0) + \sum_{n=0}^{N-1} \log p(\theta_n | \theta_{n-1}), \quad (6)$$

where C is a constant. The E and M-steps for estimating the CIR, CFO, and PHN in the training interval are detailed in the following subsections.

A. E-step

In this step, the received signal r_n is first multiplied by $e^{-j2\pi\hat{\epsilon}^{[i]}n/N}$. Subsequently, the signal $y_n \triangleq e^{-j2\pi\hat{\epsilon}^{[i]}n/N} r_n$ is used to estimate the PHN vector, where $\hat{\epsilon}^{[i]}$ is the latest CFO estimate obtained from the previous iteration. We propose to use an EKF during the E-step to estimate the PHN samples $\boldsymbol{\theta}$. The intuition behind choosing the EKF will be explained shortly after (8).

The signal y_n can be written as

$$y_n = e^{-j2\pi\hat{\epsilon}^{[i]}n/N} r_n = e^{j2\pi n \Delta\hat{\epsilon}/N} e^{j\theta_n} s_n^{[i]} + \tilde{w}_n, \quad (7)$$

where $s_n^{[i]}$ is the n th symbol of the vector $\mathbf{s}^{[i]} \triangleq \mathbf{F}^H \mathbf{D} \mathbf{W} \hat{\mathbf{h}}^{[i]}$, $\Delta\hat{\epsilon} \triangleq \epsilon - \hat{\epsilon}^{[i]}$, and $\tilde{w}_n \triangleq w_n e^{-j2\pi\hat{\epsilon}^{[i]}n/N}$. The state and observation equations at time n are given by

$$\theta_n = \theta_{n-1} + \delta_n, \quad (8)$$

$$y_n = z_n + w_n = e^{j\theta_n} s_n + \tilde{w}_n, \quad (9)$$

respectively. Since the observation equation in (9) is a non-linear function of the unknown state vector $\boldsymbol{\theta}$, the EKF is used instead of the simple Kalman filtering. The EKF uses the Taylor series expansion to

linearize the non-linear observation equation in (9) about the current estimates [30]. Thus, the Jacobian of z_n , \dot{z}_n , is evaluated by computing the first order partial derivative of z_n with respect to θ_n as

$$\begin{aligned}\dot{z}_n &= \frac{\partial z(\theta_n)}{\partial \theta_n} \Big|_{\theta_n = \hat{\theta}_{n|n-1}} = jz(\hat{\theta}_{n|n-1}) \\ &= j e^{j\hat{\theta}_{n|n-1}^{[i]}} \hat{s}_n.\end{aligned}\quad (10)$$

The first and second moments of the state vector at the i th iteration denoted by $\hat{\theta}_{n|n-1}^{[i]}$ and $M_{n|n-1}^{[i]}$, respectively, are given by

$$\hat{\theta}_{n|n-1}^{[i]} = \hat{\theta}_{n-1|n-1}^{[i]}, \quad (11)$$

$$M_{n|n-1}^{[i]} = M_{n-1|n-1}^{[i]} + \sigma_\delta^2. \quad (12)$$

Given the observation y_n , the Kalman gain K_n , posteriori state estimate $\hat{\theta}_{n|n}^{[i]}$, and the filtering error covariance, $M_{n|n}^{[i]}$ are given by

$$K_n = M_{n|n-1}^{[i]} \dot{z}^*(\theta_{n|n-1}) (\dot{z}(\theta_{n|n-1}) M_{n-1|n-1}^{[i]} \dot{z}^*(\theta_{n|n-1}) + \sigma_w^2)^{-1}, \quad (13)$$

$$\hat{\theta}_{n|n}^{[i]} = \hat{\theta}_{n|n-1}^{[i]} + \Re\{K_n(y_n - e^{j\hat{\theta}_{n|n-1}^{[i]}} \hat{s}_n^{[i]})\}, \quad (14)$$

$$M_{n|n}^{[i]} = \Re\{M_{n|n-1}^{[i]} - K_n \dot{z}(\theta_{n|n-1}) M_{n|n-1}^{[i]}\}, \quad (15)$$

respectively. Before starting the EKF recursion (10)-(15), $\hat{\theta}_{1|0}^{[0]}$ and $M_{1|0}^{[0]}$ are initialized to $\hat{\theta}_{1|0}^{[0]} = 0$ and $M_{1|0}^{[0]} = \sigma_\delta^2$. The initialization choice for the PHN follows from the assumption that the complex channel parameter takes into account the PHN corresponding to the first symbol.

B. M-step

In this step, the CIR and CFO are estimated by minimizing the LLF in (6). In order to further reduce the complexity associated with the M -step, the minimization in (6) is carried out with respect to one of the parameters while keeping the remaining parameters at their most recently updated values [31], [32]. First, by using the channel estimate at the i th iteration, $\hat{\mathbf{h}}^{[i]}$, and the PHN vector estimate from the E-step, $\hat{\boldsymbol{\theta}}^{[i+1]}$, the LLF in (6) is minimized with respect to ϵ to obtain the CFO estimate for the $(i+1)$ th iteration, $\hat{\epsilon}^{[i+1]}$, as

$$\hat{\epsilon}^{[i+1]} = \arg \min_{\epsilon} \sum_{n=0}^{N-1} \|r_n - e^{j2\pi\epsilon n/N} e^{j\theta_n} s_n\|^2 \Big|_{\theta_n = \hat{\theta}_n^{[i]}, \mathbf{h} = \hat{\mathbf{h}}^{[i]}}. \quad (16)$$

After simplifying (16), we have

$$\hat{\epsilon}^{[i+1]} = \arg \max_{\epsilon} \sum_{n=0}^{N-1} \Re\{(r_n)^* \hat{S}_n^{[i]} e^{j2\pi\epsilon n/N}\}, \quad (17)$$

where $\hat{S}_n^{[i]} = e^{j\hat{\theta}_n^{[i]}} \hat{s}_n$. In order to resolve the nonlinearity in (17), we can approximate the term $e^{j2\pi\epsilon n/N}$ using a second order Taylor series expansion around the pervious CFO estimate, $\hat{\epsilon}^{[i]}$, as

$$e^{j2\pi\epsilon n/N} = e^{j2\pi\hat{\epsilon}^{[i]} n/N} + (\epsilon - \hat{\epsilon}^{[i]}) (j \frac{2\pi}{N} n) e^{j2\pi\hat{\epsilon}^{[i]} n/N} + \frac{1}{2} (\epsilon - \hat{\epsilon}^{[i]})^2 (j \frac{2\pi}{N} n)^2 e^{j2\pi\hat{\epsilon}^{[i]} n/N}. \quad (18)$$

Substituting (18) into (17), $\hat{\epsilon}^{[i+1]}$ is given by

$$\begin{aligned} \hat{\epsilon}^{[i+1]} = \arg \max_{\epsilon} \Big\{ & \sum_{n=0}^{N-1} \Re\{(r_n)^* \hat{S}_n^{[i]} e^{j2\pi\hat{\epsilon}^{[i]} n/N} \\ & + (\epsilon - \hat{\epsilon}^{[i]}) \sum_{n=0}^{N-1} \Re\{(r_n)^* \hat{S}_n^{[i]} (j \frac{2\pi}{N} n) e^{j2\pi\hat{\epsilon}^{[i]} n/N}\} \\ & + \frac{1}{2} (\epsilon - \hat{\epsilon}^{[i]})^2 \sum_{n=0}^{N-1} \Re\{(r_n)^* \hat{S}_n^{[i]} (j \frac{2\pi}{N} n)^2 e^{j2\pi\hat{\epsilon}^{[i]} n/N}\} \Big\}. \end{aligned} \quad (19)$$

Taking the derivative of (19) with respect to ϵ and equating the result to zero, the estimate of ϵ at the $(i+1)$ th iteration is given by

$$\hat{\epsilon}^{[i+1]} = \hat{\epsilon}^{[i]} + \frac{N}{2\pi} \frac{\sum_{n=0}^{N-1} n \Im\{(r_n)^* \hat{S}_n^{[i]} e^{j2\pi\hat{\epsilon}^{[i]} n/N}\}}{\sum_{n=0}^{N-1} n^2 \Re\{(r_n)^* \hat{S}_n^{[i]} e^{j2\pi\hat{\epsilon}^{[i]} n/N}\}}. \quad (20)$$

Next, by setting θ and ϵ to their latest updated values, the updated value of $\hat{\mathbf{h}}$ at the $(i+1)$ th iteration, $\hat{\mathbf{h}}^{[i+1]}$, is determined as outlined below.

Based on the vectorial form of the received signal in (3), the negative LLF, in (6), can be written as

$$\log p(\mathbf{z}; \epsilon) = \mathbf{C}1 + \|\mathbf{r} - \mathbf{E}\mathbf{P}\mathbf{\Gamma}\mathbf{h}\|^2 + \log p(\theta). \quad (21)$$

where $\mathbf{\Gamma} \triangleq \mathbf{F}^H \mathbf{D} \mathbf{W}$. Taking the derivative of (21) with respect to \mathbf{h} and equating the result to zero, the estimate of \mathbf{h} at the $(i+1)$ th iteration is given by

$$\hat{\mathbf{h}}^{[i+1]} = (\mathbf{\Gamma}^H \mathbf{\Gamma})^{-1} \mathbf{\Gamma}^H \hat{\mathbf{P}}^H \hat{\mathbf{E}}^H \mathbf{r}, \quad (22)$$

where $\hat{\mathbf{E}} \triangleq \text{diag}([e^{(j2\pi\hat{\epsilon}^{[i+1]}/N) \times 0}, e^{(j2\pi\hat{\epsilon}^{[i+1]}/N)}, \dots, e^{(j2\pi\hat{\epsilon}^{[i+1]}/N) \times (N-1)}]^T)$ and $\hat{\mathbf{P}} \triangleq \text{diag}([e^{j\hat{\theta}_0^{[i]}}, e^{j\hat{\theta}_1^{[i]}}, \dots, e^{j\hat{\theta}_{N-1}^{[i]}}]^T)$.

Note that $\hat{\epsilon}^{[i+1]}$ and $\hat{\theta}^{[i]} \triangleq [\hat{\theta}_0^{[i]}, \hat{\theta}_1^{[i]}, \dots, \hat{\theta}_{N-1}^{[i]}]^T$ are determined as in (20) and (14), respectively.

Using (14), (20), and (22), the proposed algorithm iteratively updates the PHN, CFO, and CIR estimates, respectively. The algorithm is terminated when the difference between the likelihood functions of two iterations is smaller than a threshold ζ , i.e.,

$$\left| \sum_{n=0}^{N-1} \left\| r_n - e^{j2\pi\hat{\epsilon}^{[i+1]}n/N} e^{j\hat{\theta}_n^{[i+1]}} s_n^{[i+1]} \right\|^2 - \sum_{n=0}^{N-1} \left\| r_n - e^{j2\pi\hat{\epsilon}^{[i]}n/N} e^{j\hat{\theta}_n^{[i]}} s_n^{[i]} \right\|^2 \right| \leq \zeta. \quad (23)$$

C. Initialization and Convergence

The appropriate initialization of CFO and CIR, i.e., $\hat{\epsilon}^{[0]}$ and $\hat{\mathbf{h}}^{[0]}$, respectively, is essential to ensure the global convergence of the proposed estimator [33]. The initialization process can be summarized as follows:

- The initial CFO estimate is obtained by applying an exhaustive search for the value of ϵ that minimizes the cost function, $\sum_{n=0}^{N-1} \| r_n - e^{j2\pi\epsilon n/N} \hat{s}_n \|^2$. Here, \hat{s}_n is the n th symbol of the vector $\hat{\mathbf{s}} \triangleq \mathbf{F}^H \mathbf{D} \mathbf{W} \hat{\mathbf{h}}$ with $\hat{\mathbf{h}} \triangleq (\mathbf{\Gamma}^H \mathbf{\Gamma})^{-1} \mathbf{\Gamma}^H \hat{\mathbf{E}}^H \mathbf{r}$. Note that this exhaustive search needs to be only carried out at the system start up to initialize the estimation process. Simulations in Section VI indicate that an exhaustive search with a coarse step size of 10^{-2} is sufficient for the initialization of the proposed estimator.
- Using $\hat{\epsilon}^{[0]}$, the initial channel estimate, $\hat{\mathbf{h}}^{[0]}$, is obtained by applying the relationship, $\hat{\mathbf{h}}^{[0]} = (\mathbf{\Gamma}^H \mathbf{\Gamma})^{-1} \mathbf{\Gamma}^H (\hat{\mathbf{E}}^{[0]})^H \mathbf{r}$. Here, $\hat{\mathbf{E}}^{[0]} = \hat{\mathbf{E}}|_{\hat{\epsilon}=\hat{\epsilon}^{[0]}}$.

Note that based on the equivalent system model in (2b) and the simulation results in Section VI, it can be concluded that the proposed ECM algorithm converges globally when the PHN vector $\hat{\boldsymbol{\theta}}$ is initialized as $\hat{\boldsymbol{\theta}}^{[0]} = \mathbf{0}_{N \times 1}$.

Simulation results in Section VI show that at SNRs of 20 dB or higher the proposed ECM-based estimator always converges to the true estimates in only 2 iterations.

V. JOINT DATA DETECTION AND PHN MITIGATION

In this section, we propose an iterative detector that utilizes an EKF to track the PHN parameters during the data transmission interval.

At first, using the estimated CFO value, the effect of CFO on the received data symbol, \mathbf{r} , in (3) is compensated. As shown Fig. 2, the resulting signal, $\mathbf{y} \triangleq [y_1, \dots, y_n]$, where y_n is defined in (9), passes

through an *iterative* data detection and PHN estimation block. We propose to use an EKF to track the PHN samples, θ , over the data symbols. The PHN estimation is similar to that in (10)-(15) and is not presented here to avoid repetition. However, instead of training-based PHN tracking, the PHN estimation is followed in decision-directed fashion for the received data symbols. In other words, the estimate of the data symbol in the previous iteration, $\hat{\mathbf{d}}^{[i-1]}$, is used to update the symbol's PHN estimate at the current iteration $\hat{\theta}^{[i]}$. Particularly, $\mathbf{s}^{[i]}$ in (9), is calculated as $\mathbf{s}^{[i]} = \mathbf{F}^H \hat{\mathbf{D}}^{[i-1]} \mathbf{W} \hat{\mathbf{h}}$, where $\hat{\mathbf{h}}$ is the CIR vector estimate obtained from the ECM estimator during the training interval, and $\hat{\mathbf{D}}^{[i-1]} \triangleq \text{diag}(\hat{\mathbf{d}}^{[i-1]})$. Next, the data vector estimate is updated for the i th iteration. Following [20] and based on the received signal in (3), the negative LLF for the CFO compensated signal, \mathbf{y} , can be written as

$$\log p(\mathbf{y}, \hat{\mathbf{d}}, \hat{\theta}) = C + \frac{1}{2\sigma_w^2} \|\mathbf{y} - \hat{\mathbf{P}} \mathbf{F}^H \hat{\mathbf{\Upsilon}} \hat{\mathbf{d}}\|^2 + \frac{1}{2\xi_d} \|\hat{\mathbf{d}}\|^2 + \log p(\theta), \quad (24)$$

where

- $\hat{\mathbf{\Upsilon}} \triangleq \text{diag}(\mathbf{W} \hat{\mathbf{h}})$ is the estimated channel frequency response,
- $\hat{\mathbf{d}} \triangleq [\hat{d}_0, \hat{d}_1, \dots, \hat{d}_{N-1}]^T$ is the estimate of the modulated data vector, and
- ξ_d is the average transmitted symbol power and normalized to 1,

Taking the derivative of (24) with respect to \mathbf{d} and equating the result to zero, the estimate of \mathbf{d} at the i th iteration, $\hat{\mathbf{d}}^{[i]}$ is given by

$$\hat{\mathbf{d}}^{[i]} = (\hat{\mathbf{\Upsilon}}^H \hat{\mathbf{\Upsilon}} + \frac{\sigma_w^2}{\xi_d} \mathbf{I}_N)^{-1} \hat{\mathbf{\Upsilon}}^H \hat{\mathbf{P}} \mathbf{F}^H \mathbf{y}, \quad (25)$$

where $\hat{\mathbf{P}} \triangleq \text{diag}([e^{j\hat{\theta}_0^{[i]}}, e^{j\hat{\theta}_1^{[i]}}, \dots, e^{j\hat{\theta}_{N-1}^{[i]}}]^T)$ and $\hat{\theta}^{[i]} \triangleq [\hat{\theta}_0^{[i]}, \hat{\theta}_1^{[i]}, \dots, \hat{\theta}_{N-1}^{[i]}]^T$ are obtained via the EKF based estimator.

Using the EKF set of equations (10)–(15) and (25), the proposed algorithm iteratively updates the PHN and data estimates, respectively, and stops when the difference between likelihood functions of two iterations is smaller than a threshold ζ , i.e.,

$$\left| \sum_{n=0}^{N-1} \left\| y_n - e^{j\hat{\theta}_n^{[i+1]}} \hat{s}_n^{[i+1]} \right\|^2 - \sum_{n=0}^{N-1} \left\| y_n - e^{j\hat{\theta}_n^{[i]}} \hat{s}_n^{[i]} \right\|^2 \right| \leq \zeta. \quad (26)$$

Let $\hat{\mathbf{d}}^{[0]}$ denote the initial estimate of the transmitted data vector. Appropriate initialization of $\hat{\mathbf{d}}^{[0]}$ results in the proposed iterative detector to converge quickly. In our algorithm, the initial data estimate is obtained using $\hat{\mathbf{d}}^{[0]} = (\hat{\mathbf{\Upsilon}}^H \hat{\mathbf{\Upsilon}} + \frac{\sigma_w^2}{\xi_d} \mathbf{I}_N)^{-1} \hat{\mathbf{\Upsilon}}^H \hat{\mathbf{P}} \mathbf{F}^H \mathbf{y}$, where $\hat{\mathbf{P}}_{[m-1]}$ is the PHN matrix estimate obtained

from the previous OFDM symbol. Simulation results in Section VI indicate that at SNR= 20 dB the proposed detector, on average, converges after 2 iterations. The overall estimation and detection algorithm is summarized in Algorithm V on the next page.

A. Complexity Analysis

In this subsection, the computational complexity of the proposed estimator and detector is compared with that of [8] and [20]. Throughout this section, computational complexity is defined as the number of complex additions and multiplications [34].

Let us denote the computational complexity of the proposed estimator by $C_{\text{EST}} = C_{\text{EST}}^{[M]} + C_{\text{EST}}^{[A]}$, where $C_{\text{EST}}^{[M]}$ and $C_{\text{EST}}^{[A]}$ denote the number of complex multiplications and additions used by the estimator, respectively. $C_{\text{EST}}^{[M]}$ and $C_{\text{EST}}^{[A]}$ are determined as

$$\begin{aligned}
 C_{\text{EST}}^{[M]} = & \left[\underbrace{N}_{(10)} + \underbrace{5N}_{(13)} + \underbrace{2N}_{(14)} + \underbrace{2N}_{(15)} + \underbrace{7N}_{(20)} + \underbrace{LN(2N+1)}_{(22)} + \underbrace{N(N^2 + L(N+1))}_{s_n \text{ in (14)}} \right] t_{\text{ECM}} \\
 & + \left[\underbrace{3N}_{\sum_{n=0}^{N-1} \|r_n - e^{j2\pi\hat{\epsilon}n/N} \hat{s}_n\|^2} + \underbrace{LN(2N+1)}_{\hat{\mathbf{h}} \triangleq \xi_d^{-1} \mathbf{\Gamma}^H \hat{\mathbf{E}}^H \mathbf{r}} + \underbrace{N(N^2 + L(N+1))}_{\hat{\mathbf{s}} \triangleq \mathbf{F}^H \mathbf{D} \mathbf{W} \hat{\mathbf{h}}} \right] t_{\text{initialize}} \\
 & + \underbrace{N^2(N+L)}_{\mathbf{\Gamma} \text{ in (22)}}, \tag{27}
 \end{aligned}$$

$$\begin{aligned}
 C_{\text{EST}}^{[A]} = & \left[\underbrace{N}_{(12)} + \underbrace{N}_{(13)} + \underbrace{2N}_{(14)} + \underbrace{N}_{(15)} + \underbrace{2N+1}_{(20)} + \underbrace{L(N-1)(2N+1)}_{(22)} \right. \\
 & \left. + \underbrace{N(N-1)(L+1) + N(L-1)}_{s_n \text{ in (14)}} \right] t_{\text{ECM}} \\
 & + \left[\underbrace{2N}_{\sum_{n=0}^{N-1} \|r_n - e^{j2\pi\hat{\epsilon}n/N} \hat{s}_n\|^2} + \underbrace{L(N-1)(2N+1)}_{\hat{\mathbf{h}} \triangleq \xi_d^{-1} \mathbf{\Gamma}^H \hat{\mathbf{E}}^H \mathbf{r}} + \underbrace{N(N-1)(L+1) + N(L-1)}_{\hat{\mathbf{s}} \triangleq \mathbf{F}^H \mathbf{D} \mathbf{W} \hat{\mathbf{h}}} \right] t_{\text{initialize}} \\
 & + \underbrace{N(N-1)(N+L)}_{\mathbf{\Gamma} \text{ in (22)}}, \tag{28}
 \end{aligned}$$

where t_{ECM} is the number of iterations in the ECM estimator and $t_{\text{initialize}}$ is the number of iterations required to initialize the ECM algorithm. The latter depends on the step size of the exhaustive search used to initialize the CFO estimates.

Algorithm 1 PROPOSED ECM ESTIMATOR AND DATA DETECTION ALGORITHMS

ECM ESTIMATOR
Initialization

$\hat{\theta}_{1|0}^{[0]} = 0$ and $M_{1|0}^{[0]} = \sigma_\delta^2$ and obtain $\hat{\epsilon}^{[0]}$ and $\hat{\mathbf{h}}^{[0]}$ using an exhaustive search and (22) with coarse step size i.e., 10^{-2}

while $\left| \sum_{n=0}^{N-1} \left\| r_n - e^{j2\pi\hat{\epsilon}^{[i+1]}n/N} e^{j\hat{\theta}_n^{[i+1]}} s_n^{[i+1]} \right\|^2 - \sum_{n=0}^{N-1} \left\| r_n - e^{j2\pi\hat{\epsilon}^{[i]}n/N} e^{j\hat{\theta}_n^{[i]}} s_n^{[i]} \right\|^2 \right| > \zeta$. **do**
for $n = 0, 1, \dots, N-1$ **do**

(10)–(15)

end for

for $n = 0, 1, \dots, N-1$ **do**

$$\hat{\epsilon}^{[i+1]} = \hat{\epsilon}^{[i]} + \frac{N}{2\pi} \frac{\sum_{n=0}^{N-1} n \Im \left\{ (r_n)^* \hat{S}_n^{[i]} e^{j2\pi\hat{\epsilon}^{[i]}n/N} \right\}}{\sum_{n=0}^{N-1} n^2 \Re \left\{ (r_n)^* \hat{S}_n^{[i]} e^{j2\pi\hat{\epsilon}^{[i]}n/N} \right\}}$$

end for

$$\hat{\mathbf{h}}^{[i+1]} = (\mathbf{\Gamma}^H \mathbf{\Gamma})^{-1} \mathbf{\Gamma}^H \hat{\mathbf{P}}^H \hat{\mathbf{E}}^H \mathbf{r}$$

$$\hat{\mathbf{h}}^{[i]} = \hat{\mathbf{h}}^{[i+1]}, \hat{\boldsymbol{\theta}}^{[i]} = \hat{\boldsymbol{\theta}}^{[i+1]}, \hat{\epsilon}^{[i]} = \hat{\epsilon}^{[i+1]}$$

end while

DATA DETECTION

for $m = 1, \dots, M$ **do**

Initialization

$$\text{Obtain } \hat{\mathbf{d}}^{[0]} = (\hat{\mathbf{\Upsilon}}^H \hat{\mathbf{\Upsilon}} + \frac{\sigma_w^2}{\xi_d} \mathbf{I}_N)^{-1} \hat{\mathbf{\Upsilon}}^H \mathbf{F} \hat{\mathbf{P}}_{[m-1]}^H \mathbf{y}$$

Replace $\hat{\mathbf{d}}^{[0]}$ by its hard decision.

while $\left| \sum_{n=0}^{N-1} \left\| y_n - e^{j\hat{\theta}_n^{[i+1]}} \hat{s}_n^{[i+1]} \right\|^2 - \sum_{n=0}^{N-1} \left\| y_n - e^{j\hat{\theta}_n^{[i]}} \hat{s}_n^{[i]} \right\|^2 \right| > \zeta$ **do**

Using the EKF set of equation in Section IV-A to estimate the PHN parameters,

$$\hat{\mathbf{d}}^{[i]} = (\hat{\mathbf{\Upsilon}}^H \hat{\mathbf{\Upsilon}} + \frac{\sigma_w^2}{\xi_d} \mathbf{I}_N)^{-1} \hat{\mathbf{\Upsilon}}^H \mathbf{F} \hat{\mathbf{P}}^H \mathbf{y}.$$

Replace $\hat{\mathbf{d}}^{[i]}$ by its hard decision.

$$\hat{\mathbf{d}}^{[i]} = \hat{\mathbf{d}}^{[i+1]}$$

end while

end for

Similarly, the computational complexity of the proposed detector is denoted by $C_{\text{DATA DET}}^{[M]}$ and $C_{\text{DATA DET}}^{[A]}$, where $C_{\text{DATA DET}}^{[M]}$ and $C_{\text{DATA DET}}^{[A]}$ denote the number of complex multiplications and additions used by the estimator, respectively. $C_{\text{DATA DET}}^{[M]}$ and $C_{\text{DATA DET}}^{[A]}$ are determined as

$$C_{\text{DATA DET}}^{[M]} = \left[\underbrace{N}_{(10)} + \underbrace{5N}_{(13)} + \underbrace{2N}_{(14)} + \underbrace{2N}_{(15)} + \underbrace{N(N^2 + L(N+1))}_{s_n \text{ in } (14)} + \underbrace{N^2(5N+1)}_{(25)} \right] t_{\text{DATA DET}} \\ + \underbrace{N^2(5N+1)}_{\hat{\mathbf{d}}^{[0]} = (\hat{\mathbf{\Upsilon}}^H \hat{\mathbf{\Upsilon}} + \frac{\sigma_w^2}{\xi_d} \mathbf{I}_N)^{-1} \hat{\mathbf{\Upsilon}}^H \mathbf{F} \mathbf{P}_{[m-1]}^H \mathbf{y}} + \underbrace{NL}_{\hat{\mathbf{\Upsilon}} \text{ in } (25)}, \quad (29)$$

$$C_{\text{DATA DET}}^{[A]} = \left[\underbrace{N}_{(12)} + \underbrace{N}_{(13)} + \underbrace{2N}_{(14)} + \underbrace{N}_{(15)} + \underbrace{N(N-1)(L+1) + N(L-1)}_{s_n \text{ in } (14)} \right. \\ \left. + \underbrace{N(N^2 + N(N-1)(4N+1))}_{(25)} \right] t_{\text{DATA DET}} + \underbrace{N(N^2 + N(N-1)(4N+1))}_{\hat{\mathbf{d}}^{[0]} = (\hat{\mathbf{\Upsilon}}^H \hat{\mathbf{\Upsilon}} + \frac{\sigma_w^2}{\xi_d} \mathbf{I}_N)^{-1} \hat{\mathbf{\Upsilon}}^H \mathbf{F} \mathbf{P}_{[m-1]}^H \mathbf{y}} \\ + \underbrace{N(L-1)}_{\hat{\mathbf{\Upsilon}} \text{ in } (25)}, \quad (30)$$

where $t_{\text{DATA DET}}$ is the number of iterations required by the detector in (26).

Following similar steps as in (27)-(30), we can find the computational complexity of the estimator in [8] as

$$C_{[8, \text{EST}]}^{[M]} = [N^2(11N+7) + 2N] t_{[8]} + N^2(9N+4L+1) + LN \quad (31)$$

$$C_{[8, \text{EST}]}^{[A]} = [2N^3 + (N-1)(9N^2 + 7N + 2) + 1] t_{[8]} + 2N^3 + (N-1)(N(7N+4L+1) + L) \quad (32)$$

where $t_{[8]}$ is the number of iterations required for estimating the CFO via an exhaustive search in [8]. Moreover, the notations $C_{[8, \text{EST}]}^{[M]}$ and $C_{[8, \text{EST}]}^{[A]}$ are used to denote the number of complex multiplications and additions used by the estimator in [8], respectively.

The computational complexity of the detector in [20] is given by

$$C_{[20, \text{DATA DET}]}^{[M]} = [N^2(11N+6)] t_{[20]} + N^2(6N+1) \quad (33)$$

$$C_{[20, \text{DATA DET}]}^{[A]} = [N(N-1)(9N+6) + N^2(2N+1)] t_{[20]} + N^2(6N-5) + N(N-1) \quad (34)$$

where $t_{[20]}$ is the number of iterations used by the detector in [20]. Note that since the estimation approach of [8] and [16] are similar, the complexity of estimation algorithm in [16] can be calculated as in (31)

and (32). We do not present the computational complexity of the algorithm in [22] since the approach in [22] only considers channel and PHN estimation while assuming that no CFO is present.

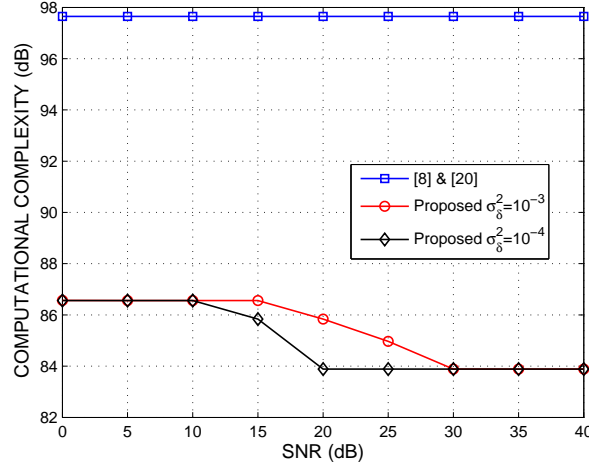


Fig. 3. Comparison of computational complexity of the proposed algorithms for PHN variance, $\sigma_\delta^2 = [10^{-3}, 10^{-4}] \text{ rad}^2$ and the algorithms in [8] & [20].

In Fig. 3, the computational complexity of the proposed algorithms for PHN variance, $\sigma_\delta^2 = [10^{-3}, 10^{-4}] \text{ rad}^2$ is compared against the existing algorithms in [8] and [20]. Simulations indicate that at low SNR, i.e., $SNR < 10 \text{ dB}$, on average, the proposed estimator and detector converge after $t_{\text{ECM}} = 3$ and $t_{\text{DATA DET}} = 5$ iterations, respectively. In addition, the number of iterations in the proposed estimator and detector is gradually decreased to $t_{\text{ECM}} = t_{\text{DATA DET}} = 2$ at $SNR \geq 20 \text{ dB}$ and $SNR \geq 30 \text{ dB}$ for $\sigma_\delta^2 = 10^{-4} \text{ rad}^2$ and $\sigma_\delta^2 = 10^{-3} \text{ rad}^2$, respectively. On the other hand, the detector in [20] requires a total of $t_{[20]} = 4$ iterations to converge for PHN variance, $\sigma_\delta^2 = [10^{-3}, 10^{-4}] \text{ rad}^2$. Since simulations indicate that the proposed ECM algorithm converges to the true estimates when the CFO estimates are initialized with a step size of 10^{-2} , here $t_{\text{initialize}}$ is set to 10^2 . Moreover, the results in Section VI indicate that to reach a similar estimation accuracy and system performance, the approach in [8] requires the step size for the exhaustive search to be set to 10^{-2} , i.e., $t_{[8]} = 10^3$ for a step size of 10^{-3} . For example, at $SNR = 20 \text{ dB}$, $\sigma_\delta^2 = 10^{-4} \text{ rad}^2$, $L = 4$ and $N = 64$ and using this system setup and by applying (27)-(34), it can be determined that the proposed estimation and data detection algorithm is computationally more efficient than that of Lin et. al. in [8] and [20] by a factor of 130.

VI. SIMULATION RESULTS AND DISCUSSIONS

In this subsection, we present simulation results to evaluate the performance of the proposed estimation and data detection algorithms. We consider an OFDM packet to consist of $m = 6$ OFDM symbols, comprising an OFDM training symbol followed by 5 data symbols. The data symbols are drawn from normalized 64, 128, or 256 quadrature amplitude modulation (QAM). The sampling rate of the OFDM signal is 20 MHz corresponding to the OFDM sampling duration of $T_s = 50$ nanoseconds. The channel impulse response (CIR) is assumed to be a Rayleigh fading multipath channel with a delay of $L = 4$ taps and an exponentially decreasing power delay profile with the average channel power = [-1.52 -6.75 -11.91 -17.08] dB. The Wiener PHN is generated with PHN variances of $\sigma_\delta^2 = [10^{-3}, 10^{-4}] \text{ rad}^2$. The unknown normalized CFO is assumed to be uniformly distributed over the range $\epsilon \in (-0.5, 0.5)$ for each simulation. Unless specified otherwise, an OFDM training symbol size of $N = 64$ subcarriers is used with each subcarrier modulated in *quadrature phase-shift keying (QPSK)* with subcarrier spacing = 312.5 kHz. The simulation results are averaged over 1×10^5 Monte Carlo simulation runs.

A. Estimation Performance

In this subsection, we compare the performance of the proposed ECM estimator with the HCRB in Theorem 1 and the MAP estimator in [8]. Figs. 4, 5, and 6 plot the HCRB and the mean square error (MSE) for estimating the channel impulse response, carrier frequency offset, and PHN, respectively. The HCRB in (6) is numerically evaluated for two different PHN variance, e.g., $\sigma_\delta^2 = [10^{-3}, 10^{-4}] \text{ rad}^2$.

The following observations can be made from the figures:

- 1) The HCRB and the estimator's MSE are dependent on the variance of the PHN process. Note that, $\sigma_\delta^2 = 10^{-3} \text{ rad}^2$, corresponds to presence of a very strong PHN [35].
- 2) Fig. 4 shows that the HCRB for the channel estimation does not suffer from an error floor. However, Figs. 5 and 6 show that the HCRB for CFO and PHN suffer from an error floor, which is directly related to the variance of the PHN process. This is due to the fact that at low SNR the performance of the system is dominated by AWGN, while at high SNR the performance of the proposed estimator is limited by PHN and the resulting ICI.

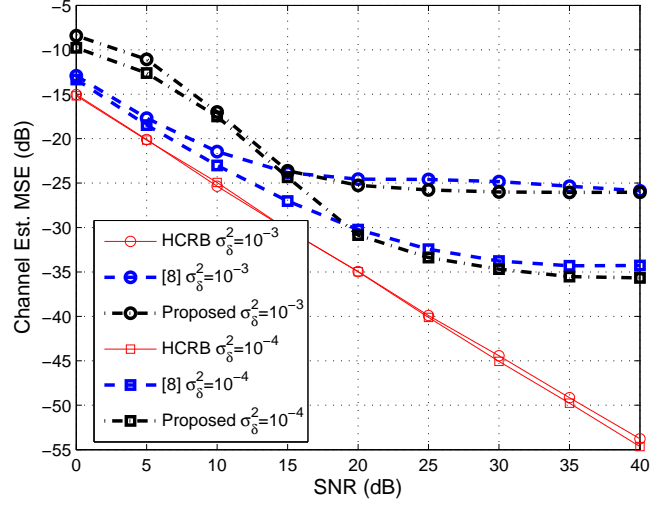


Fig. 4. Channel estimation MSE for the proposed and MAP estimators for PHN variance, $\sigma_\delta^2 = [10^{-3}, 10^{-4}] \text{ rad}^2$.

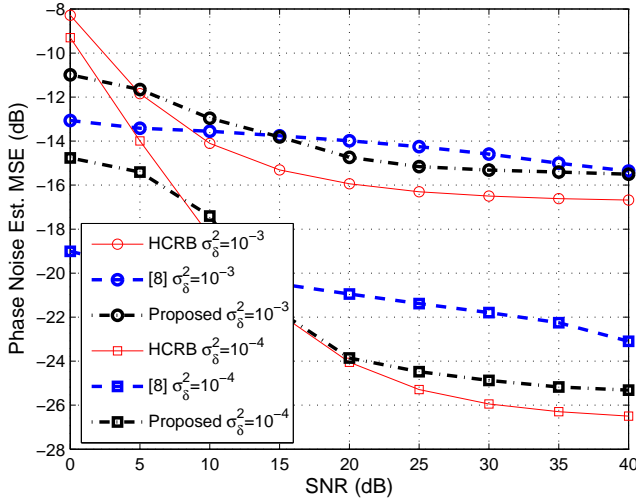


Fig. 5. PHN estimation MSE for the proposed and MAP estimators for PHN variance, $\sigma_\delta^2 = [10^{-3}, 10^{-4}] \text{ rad}^2$.

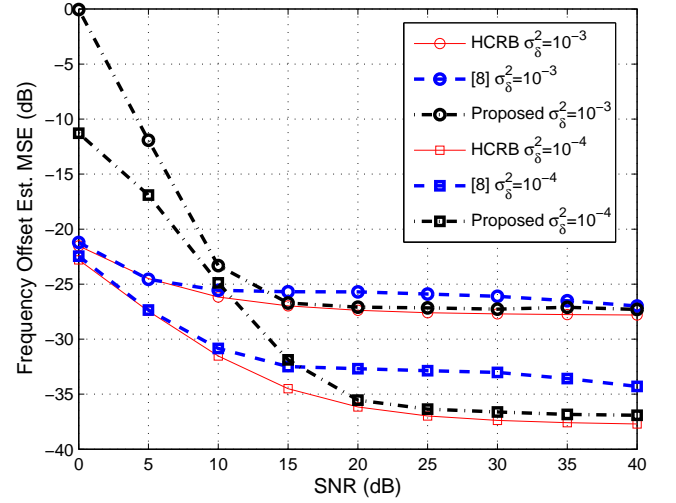


Fig. 6. CFO estimation MSE for the proposed and MAP estimators for PHN variance, $\sigma_\delta^2 = [10^{-3}, 10^{-4}] \text{ rad}^2$.

- 3) Figs. 4, 5, and 6 show that the proposed estimator outperforms the estimation algorithm in [8]. The biggest gain is achieved in the MSE of PHN estimation, followed by the MSE of CFO and channel estimation. For example, for $\sigma_\delta^2 = 10^{-4} \text{ rad}^2$ and at high SNR, the proposed estimator results in a 2–3 dB performance gain compared to that of [8] while estimating PHN or CFO. This performance gain is in addition to the lower complexity of the proposed estimator as shown in Section V-A.
- 4) Figs. 4, 5, and 6 show that at low SNR, i.e., $SNR < 15 \text{ dB}$, the proposed estimator is outperformed

by the estimation algorithm in [8]. This can be attributed to the fact that EKF is not an optimal minimum mean square filter compared to MAP estimator due to the application of a first order Taylor series approximation to linearize the observation model [30]. Therefore, at low SNR the estimation accuracy of the proposed EKF estimator is limited by the variance of the phase noise innovations and noise from AWGN. However, this performance of the proposed estimator does not affect the overall BER performance at low SNR since the performance of the system is dominated by AWGN at low SNR, while at high SNR the performance of the proposed estimator is limited by PHN and the resulting ICI only.

Note that in Fig. 5, the PHN estimation MSE of the proposed estimator and the estimator in [20] are lower than the HCRB at low SNR. This is due to the fact that the HCRB cannot be derived in closed-form while taking into account the prior knowledge of the range of CFO values, i.e., $(-0.5, 0.5)$. However, the proposed estimator and the estimator in [20] take into account this prior information while estimating the PHN, CFO, and the channel parameters.

B. Comparison with Existing Work

In the following, we examine the combined estimation and data detection performance in terms of the uncoded BER of the OFDM system. The following system setups are considered for comparison:

- (i) The proposed estimation and data detection algorithm (labelled as “proposed”).
- (ii) The estimation and data detection algorithm in [8] and [20], respectively (labelled as “[8] & [20]”).
- (iii) The data detection in [22] combined with the proposed estimation algorithm (labelled “[22, Data detection]”).
- (iv) The estimation algorithm in [8] with the proposed data detection algorithm (labelled “[8, MAP estimator]”).
- (v) As a reference, a system that applied the proposed estimation algorithm but utilizes no PHN tracking during OFDM data symbols (labelled “No CFO cancel. and PHN track.”).
- (vi) As a lower-bound on the BER performance, a system assuming perfect channel, PHN, and CFO estimation (labelled “Perf. CIR, PHN & CFO est.”).

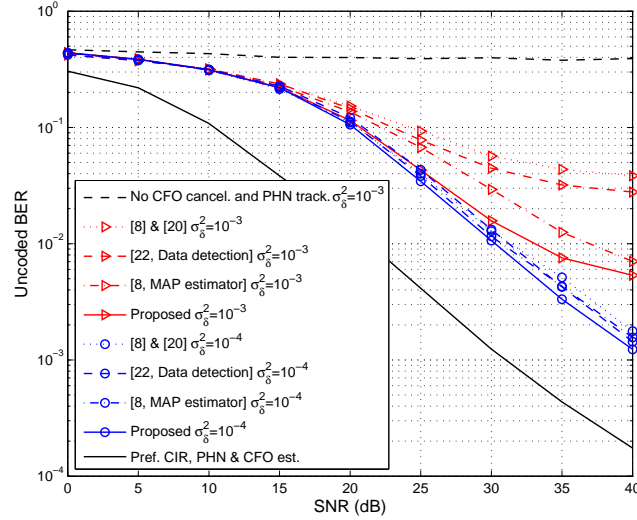


Fig. 7. Comparison of uncoded BER of the proposed algorithms for PHN variance, $\sigma_\delta^2 = [10^{-3}, 10^{-4}] \text{ rad}^2$ and 64-QAM modulation with the algorithms in [8] & [20] and [22].

Fig. 7 depicts the uncoded BER performances of the OFDM systems listed above. The following observations can be made from Fig. 7:

- 1) The results demonstrate that without phase tracking and CFO cancellation throughout the packet, the OFDM system performance deteriorates significantly. On the other hand, by combining the proposed estimation and data detection algorithms, the BER performance of an OFDM system is shown to improve immensely even in the presence of very strong PHN, e.g., $\sigma_\delta^2 = 10^{-3} \text{ rad}^2$.
- 2) Compared to existing algorithms, the BER performance of an OFDM system using the proposed algorithms is closer to the ideal case of perfect CIR, PHN, and CFO estimation (a performance gap of 10 dB at SNR = 20 dB).
- 3) It can be clearly observed that the proposed receiver structure outperforms the algorithms in [8] and [20]. This performance improvement can be attributed to the fact that the proposed ECM based joint CIR, PHN, and CFO estimator applies a single-order Taylor series approximation using the most recent estimated PHN values to obtain an updated PHN estimates unlike the approach in [8] and [20].
- 4) It is clear that the performance of the proposed data detection algorithm outperforms the algorithm in [22]. This result is anticipated, since at high PHN variance, the approximation of PHN parameters

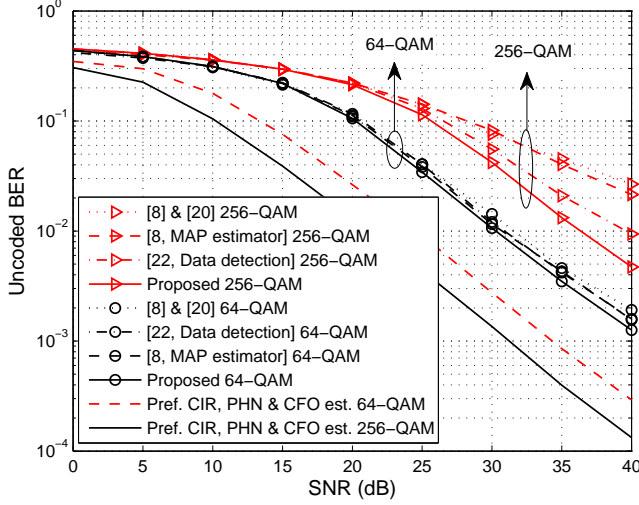


Fig. 8. Comparison of uncoded BER of the proposed algorithms for 64-QAM and 256-QAM modulations with the algorithm in [8] & [20] and [22] at PHN variance, $\sigma_\delta^2 = 10^{-4} \text{ rad}^2$.

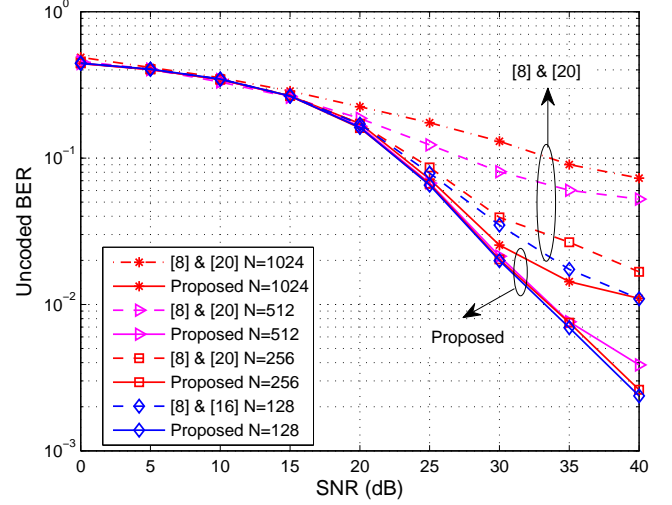


Fig. 9. Comparison of uncoded BER of the proposed algorithms with the algorithm in [8] & [20] for varying training symbol lengths $N=128, 256, 512$ and 1024 , at $\sigma_\delta^2 = 10^{-4} \text{ rad}^2$ and 128-QAM modulation.

using linear interpolation in [22] highly deviates from the true PHN parameters. Therefore, the linear interpolation approach in [22] may not be used in the presence of very strong PHN, e.g., $\sigma_\delta^2 = 10^{-3} \text{ rad}^2$.

- 5) It is clear that the performance of the proposed estimation algorithm outperforms the algorithm in [8]. This result is anticipated, since the algorithm in [8] is based on a small angle approximation and at high PHN variance, the approximation of PHN parameters using a small angle approximation in [8] highly affects the performance of the data detection algorithm.
- 6) Finally, Fig. 7 shows that in the presence of PHN, the overall BER performance of an OFDM system suffers from an error floor at high SNR, since at high SNR the performance of an OFDM system is dominated by PHN, which cannot be completely eliminated.

C. Effect of Modulation and OFDM System Parameter

Fig. 8 evaluates the uncoded BER performances of an OFDM system for higher order modulations, i.e., 256-quadrature amplitude modulation (256-QAM). The results in Fig. 8 shows that even for denser constellation, the proposed estimation and data detection algorithms significantly improve the overall system performance compared to that of [8, MAP estimator], [20] and [22]. For example, to achieve a BER of 3×10^{-2} with a PHN variance of 10^{-4} rad^2 , the proposed algorithm outperforms the algorithms

in [8, MAP estimator] and [20] by a margin of 6 dB and 7 dB, respectively. In addition, the proposed algorithm outperforms the algorithm in [22] by a margin of 3 dB at a BER of 10^{-2} with a PHN variance of 10^{-4} rad^2 . More importantly, this gap widens at high SNR values.

Fig. 9 illustrates the uncoded BER performance of an OFDM system for different number of subcarriers, e.g., $N = 128, 256, 512$ and 1024, within an OFDM symbol. Based on the results in Fig. 9, it can be concluded that the proposed algorithm is not sensitive to the subcarrier spacing at low-to-medium SNRs while at high SNRs, the BER degrades as one moves to $N = 1024$ subcarriers. This is because increasing the number of subcarriers results in more ICI that is caused by the residual PHN and CFO. More importantly, the BER performance of an OFDM system using the proposed algorithms outperforms that of [8] and [20] for any value of N . For instant, at $\text{BER} = 10^{-2}$ and $N=128$, the SNR gain for the proposed algorithms is almost 8 dB than the algorithms in [8] and [20]. Moreover, at $\text{SNR} = 35 \text{ dB}$ with $N = 512$ or $N = 1024$, the BER for the proposed algorithms is almost 10 times lower than the algorithms in [8] and [20].

It is worth mentioning that Figs. 7, 8 and 9 show that although the estimator of PHN, CIR and CFO at low SNR has lower performance than the estimator in [8], the estimation accuracy at low SNR does not impact the overall BER performance. This fact is anticipated, since at low SNR the performance of the system is dominated by AWGN, while at high SNR and even in the absence of AWGN, i.e., $\text{SNR} \rightarrow \infty$, the performance of the proposed estimator and detector are limited by PHN and CFO. Moreover, as shown in simulation results in [8, 13, 20, 25], the effects of PHN and CFO are dominant on the system performance than AWGN at moderate to high SNRs.

Finally, the coded BER performance is shown in Fig. 10. The low-density parity-check (LDPC) is employed with the channel coding rate, 1/2 and codeword bits 1296 under 64-QAM modulation. The encoding algorithm in [38] is used for encoding, while the soft-decision iterative decoding algorithm based on a sum-product algorithm in [39] is utilized for decoding the estimated data vector in (25). The results in Fig. 10 illustrates that, the proposed estimation and data detection algorithms significantly improve the overall system performance compared to the existing the algorithms in [8] and [20]. For example, to achieve a BER of 10^{-4} with a PHN variance of 10^{-4} rad^2 , the proposed algorithm outperforms the algorithms in [8, MAP estimator] and [8] & [20] by a margin of 5 dB and 10 dB, respectively. More

importantly, compared to existing algorithms, the coded BER performance of an OFDM system using the proposed algorithms is closer to the ideal case of perfect CIR, PHN, and CFO estimation (a performance gap of 2 dB at BER of 10^{-4}).

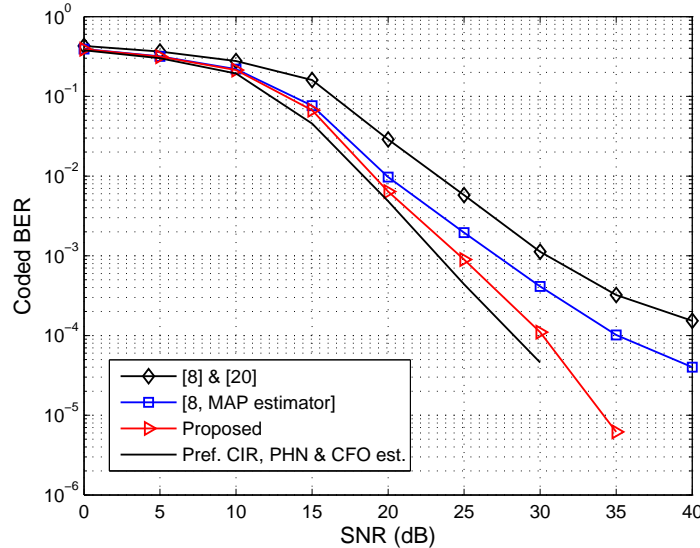


Fig. 10. Comparison of coded BER of the proposed algorithms for PHN variance, $\sigma_\delta^2 = 10^{-4} \text{ rad}^2$ and 64-QAM modulation with the algorithms in [8] & [20] and [8, MAP estimator].

VII. CONCLUSION

In this paper, an ECM based algorithm is proposed for joint estimation of channel, PHN, and CFO in OFDM systems. The signal model for the estimation problem is outlined in detail and the HCRB for the joint estimation of channel, PHN, and CFO in OFDM systems is derived. Simulation indicate that the estimation MSE of the proposed algorithm is closer to the derived HCRB and outperforms the existing estimation algorithms at moderate to high SNR. Next, an *iterative* algorithm for joint data detection and PHN mitigation is proposed for the OFDM data symbols. The proposed algorithm employs an EKF based approach to track the time-varying PHN parameters throughout the OFDM data symbols. The performance of the proposed estimation and detection algorithms have been evaluated for different PHN variances $\sigma_\delta^2 = [10^{-3}, 10^{-4}] \text{ rad}^2$, different number of subcarriers $N = [64, 128, 256, 512, 1024]$, and different modulation schemes, 64, 128, 256-QAM. Numerical results show that the proposed ECM based estimator and the iterative data detection algorithm are not only computationally more efficient

than the existing algorithms but also outperform existing algorithms in terms of uncoded and coded BER performance. For example, the uncoded BER for the proposed algorithms almost has SNR gain of 8 dB than the existing algorithms at an BER of 10^{-2} with $N = 128$. More importantly, the coded BER performance using the proposed algorithms is closer to the ideal case of perfect CIR, PHN, and CFO estimation with a performance gap of 2 dB at BER of 10^{-4} and PHN variance, $\sigma_\delta^2 = 10^{-4} \text{ rad}^2$.

APPENDIX A

DERIVATION OF THE HCRB

The hybrid information matrix \mathbf{B} can be written as [28]

$$\mathbf{B} = \mathbf{\Xi}_D + \mathbf{\Xi}_P, \quad (\text{A.1})$$

where $\mathbf{\Xi}_D \triangleq \mathbb{E}_\theta [\Psi(\boldsymbol{\theta}, \mathbf{h}, \epsilon)]$ with $\Psi(\boldsymbol{\theta}, \mathbf{h}, \epsilon) \triangleq \mathbb{E}_{\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon} [-\Delta_\lambda^\lambda \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon)|\mathbf{h}, \epsilon]$ denoting the Fisher's information matrix (FIM) and $\mathbf{\Xi}_P \triangleq \mathbb{E}_{\boldsymbol{\theta}|\mathbf{h}, \epsilon} [-\Delta_\lambda^\lambda \log p(\boldsymbol{\theta}|\mathbf{h}, \epsilon)|\mathbf{h}, \epsilon]$ is the prior information matrix with $p(\boldsymbol{\theta}|\mathbf{h}, \epsilon)$ denoting the prior distribution of PHN vector given the CIR and CFO. Thus, we first obtain expressions for matrices $\mathbf{\Xi}_D$ and $\mathbf{\Xi}_P$.

A. Computation of $\mathbf{\Xi}_D \triangleq \mathbb{E}_\theta [\Psi(\boldsymbol{\theta}, \mathbf{h}, \epsilon)]$

We partition the matrix $\mathbf{\Xi}_D$ as

$$\begin{aligned} \mathbf{\Xi}_D &= \begin{bmatrix} \mathbf{\Xi}_{D11} & \mathbf{\Xi}_{D12} & \boldsymbol{\xi}_{D13} \\ \mathbf{\Xi}_{D21} & \mathbf{\Xi}_{D22} & \boldsymbol{\xi}_{D23} \\ \boldsymbol{\xi}_{D31} & \boldsymbol{\xi}_{D32} & \xi_{D33} \end{bmatrix}, \\ &= \begin{bmatrix} \mathbb{E}_\theta[\Psi_{11}(\boldsymbol{\theta}, \mathbf{h}, \epsilon)] & \mathbb{E}_\theta[\Psi_{12}(\boldsymbol{\theta}, \mathbf{h}, \epsilon)] & \mathbb{E}_\theta[\psi_{13}(\boldsymbol{\theta}, \mathbf{h}, \epsilon)] \\ \mathbb{E}_\theta[\Psi_{21}(\boldsymbol{\theta}, \mathbf{h}, \epsilon)] & \mathbb{E}_\theta[\Psi_{22}(\boldsymbol{\theta}, \mathbf{h}, \epsilon)] & \mathbb{E}_\theta[\psi_{23}(\boldsymbol{\theta}, \mathbf{h}, \epsilon)] \\ \mathbb{E}_\theta[\psi_{31}(\boldsymbol{\theta}, \mathbf{h}, \epsilon)] & \mathbb{E}_\theta[\psi_{32}(\boldsymbol{\theta}, \mathbf{h}, \epsilon)] & \mathbb{E}_\theta[\psi_{33}(\boldsymbol{\theta}, \mathbf{h}, \epsilon)] \end{bmatrix}, \end{aligned} \quad (\text{A.2})$$

where

$$\Psi_{11}(\boldsymbol{\theta}, \mathbf{h}, \epsilon) \triangleq \mathbb{E}_{\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon} [-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon)|_{\mathbf{h}, \epsilon}], \quad (\text{A.3})$$

$$\Psi_{22}(\boldsymbol{\theta}, \mathbf{h}, \epsilon) \triangleq \mathbb{E}_{\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon} [-\Delta_{\mathbf{h}}^{\mathbf{h}} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon)|_{\mathbf{h}, \epsilon}], \quad (\text{A.4})$$

$$\psi_{33}(\boldsymbol{\theta}, \mathbf{h}, \epsilon) \triangleq \mathbb{E}_{\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon} [-\Delta_{\epsilon}^{\epsilon} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon)|_{\mathbf{h}, \epsilon}], \quad (\text{A.5})$$

$$\Psi_{12}(\boldsymbol{\theta}, \mathbf{h}, \epsilon) \triangleq \mathbb{E}_{\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon} [-\Delta_{\boldsymbol{\theta}}^{\mathbf{h}} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon)|_{\mathbf{h}, \epsilon}], \quad (\text{A.6})$$

$$\psi_{13}(\boldsymbol{\theta}, \mathbf{h}, \epsilon) \triangleq \mathbb{E}_{\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon} [-\Delta_{\boldsymbol{\theta}}^{\epsilon} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon)|_{\mathbf{h}, \epsilon}], \quad (\text{A.7})$$

$$\Psi_{21}(\boldsymbol{\theta}, \mathbf{h}, \epsilon) \triangleq \mathbb{E}_{\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon} [-\Delta_{\mathbf{h}}^{\boldsymbol{\theta}} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon)|_{\mathbf{h}, \epsilon}], \quad (\text{A.8})$$

$$\psi_{23}(\boldsymbol{\theta}, \mathbf{h}, \epsilon) \triangleq \mathbb{E}_{\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon} [-\Delta_{\mathbf{h}}^{\epsilon} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon)|_{\mathbf{h}, \epsilon}], \quad (\text{A.9})$$

$$\psi_{31}(\boldsymbol{\theta}, \mathbf{h}, \epsilon) \triangleq \mathbb{E}_{\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon} [-\Delta_{\epsilon}^{\boldsymbol{\theta}} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon)|_{\mathbf{h}, \epsilon}], \quad (\text{A.10})$$

$$\psi_{32}(\boldsymbol{\theta}, \mathbf{h}, \epsilon) \triangleq \mathbb{E}_{\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon} [-\Delta_{\epsilon}^{\mathbf{h}} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon)|_{\mathbf{h}, \epsilon}], \quad (\text{A.11})$$

1) *Computation of $\Xi_{D_{11}} \triangleq \mathbb{E}_{\boldsymbol{\theta}}[\Psi_{11}(\boldsymbol{\theta}, \mathbf{h}, \epsilon)]$* : To compute the log-likelihood function in (A.3), $p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon)$ is given by

$$p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon) = C \exp \left[\frac{-1}{\sigma_w^2} (\mathbf{r} - \boldsymbol{\mu}(\epsilon))^H (\mathbf{r} - \boldsymbol{\mu}(\epsilon)) \right], \quad (\text{A.12})$$

where $C \triangleq (\pi \sigma_w^2)^{-N}$. Given $\boldsymbol{\theta}$, \mathbf{h} , and ϵ , \mathbf{r} is a complex Gaussian vector with mean vector $\boldsymbol{\mu}(\mathbf{h}, \epsilon) = \mathbf{EPF}^H \mathbf{D} \mathbf{W} \mathbf{h}$ and covariance matrix $\sigma_w^2 \mathbf{I}_N$. Based on (A.3) and (A.12), we compute the expectation with respect to \mathbf{r} , yielding [30]

$$\mathbb{E}_{\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon} \left[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon)|_{\mathbf{h}, \epsilon} \right] = \mathbb{E}_{\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon} \left\{ \left[\frac{\partial}{\partial \boldsymbol{\theta}^*} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon) \right] \left[\frac{\partial}{\partial \boldsymbol{\theta}^*} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon) \right]^H \right\} \quad (\text{A.13})$$

which can be simplified as [36]

$$\mathbb{E}_{\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon} [-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon)|_{\mathbf{h}, \epsilon}] = \frac{\partial \boldsymbol{\mu}^H}{\partial \boldsymbol{\theta}^*} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}^T}, \quad (\text{A.14})$$

where $\boldsymbol{\mu}(\boldsymbol{\theta}, \mathbf{h}, \epsilon) = \mathbf{EPF}^H \mathbf{D} \mathbf{W} \mathbf{h}$, $\boldsymbol{\Sigma} = \mathbb{E} \{ (\mathbf{r} - \boldsymbol{\mu}(\boldsymbol{\theta}, \mathbf{h}, \epsilon))^H (\mathbf{r} - \boldsymbol{\mu}(\boldsymbol{\theta}, \mathbf{h}, \epsilon)) \} = \sigma_w^2 \mathbf{I}_N$, $\frac{\partial \boldsymbol{\mu}^H}{\partial \boldsymbol{\theta}^*}$ is an $N \times N$ matrix and its n th row is given by

$$\begin{aligned} \left[\frac{\partial \boldsymbol{\mu}^H}{\partial \boldsymbol{\theta}^*} \right]_{n,:} &= \frac{\partial}{\partial \theta_n^*} \left[\mathbf{h}^H \mathbf{W}^H \mathbf{D}^H \mathbf{F} \mathbf{P}^H \mathbf{E}^H \right] \\ &= \mathbf{h}^H \mathbf{W}^H \mathbf{D}^H \mathbf{F} \mathbf{U}_n \mathbf{E}^H \end{aligned} \quad (\text{A.15})$$

where $\mathbf{U}_n \triangleq \text{diag}([\mathbf{0}_{1 \times (n-1)}, -je^{-j\theta_n}, \mathbf{0}_{1 \times (N-n)}])$. By simplifying (A.15), we have

$$\frac{\partial \boldsymbol{\mu}^H}{\partial \boldsymbol{\theta}^*} = -j \text{diag}(\mathbf{h}^H \mathbf{W}^H \mathbf{D}^H \mathbf{F} \mathbf{P}^H \mathbf{E}^H). \quad (\text{A.16})$$

In addition, $\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}^T}$ is an $N \times N$ matrix and its n th column is given by

$$\left[\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}^*} \right]_{:,n} = \frac{\partial}{\partial \theta_n^T} [\mathbf{E} \mathbf{P} \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{h}] \quad (\text{A.17})$$

By following the simplification given above (A.16), we obtain

$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}^T} = j \text{diag}(\mathbf{E} \mathbf{P} \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{h}). \quad (\text{A.18})$$

By substituting (A.16) and (A.18) into (A.13) and simplifying, $\Xi_{D_{11}}$ is found to be an $N \times N$ matrix, such that

$$\Xi_{D_{11}} = \mathbb{E}_{\boldsymbol{\theta}} [\text{diag}(\mathbf{h}^H \mathbf{W}^H \mathbf{D}^H \mathbf{F}) \boldsymbol{\Sigma}^{-1} \text{diag}(\mathbf{F} \mathbf{D} \mathbf{W} \mathbf{h})] = \text{diag}(\mathbf{h}^H \mathbf{W}^H \mathbf{D}^H \mathbf{F}) \boldsymbol{\Sigma}^{-1} \text{diag}(\mathbf{F} \mathbf{D} \mathbf{W} \mathbf{h}). \quad (\text{A.19})$$

2) *Computation of $\Xi_{D_{22}} = \mathbb{E}_{\boldsymbol{\theta}} [\Psi_{22}(\boldsymbol{\theta}, \mathbf{h}, \epsilon)]$* : Based on (A.3) and (A.12), it follows that

$$\begin{aligned} \mathbb{E}_{\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon} [-\Delta_{\mathbf{h}}^{\mathbf{h}} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon)|_{\mathbf{h}, \epsilon}] &= \mathbb{E}_{\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon} \left\{ \left[\frac{\partial}{\partial \mathbf{h}^*} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon) \right] \left[\frac{\partial}{\partial \mathbf{h}^T} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon) \right]^H \right\} \\ &= \frac{\partial \boldsymbol{\mu}^H}{\partial \mathbf{h}^*} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \mathbf{h}^T}, \end{aligned} \quad (\text{A.20})$$

where $\frac{\partial \boldsymbol{\mu}^H}{\partial \mathbf{h}^*}$ is an $L \times N$ matrix and its l th row is given by

$$\begin{aligned} \left[\frac{\partial \boldsymbol{\mu}^H}{\partial \mathbf{h}^*} \right]_{l,:} &= \frac{\partial}{\partial h_l^*} [\mathbf{h}^H \mathbf{W}^H \mathbf{D}^H \mathbf{F} \mathbf{P}^H \mathbf{E}^H] \\ &= \mathbf{u}_l \mathbf{W}^H \mathbf{D}^H \mathbf{F} \mathbf{P}^H \mathbf{E}^H, \end{aligned} \quad (\text{A.21})$$

where $\mathbf{u}_l \triangleq \text{diag}([\mathbf{0}_{1 \times (l-1)}, 1, \mathbf{0}_{1 \times (L-l)}])$. By simplifying (A.21), we have

$$\frac{\partial \boldsymbol{\mu}^H}{\partial \mathbf{h}^*} = \mathbf{W}^H \mathbf{D}^H \mathbf{F} \mathbf{P}^H \mathbf{E}^H. \quad (\text{A.22})$$

In addition, $\frac{\partial \boldsymbol{\mu}}{\partial \mathbf{h}^T}$ is an $N \times L$ matrix and its l th column is given by:

$$\left[\frac{\partial \boldsymbol{\mu}}{\partial \mathbf{h}^T} \right]_{:,l} = \frac{\partial}{\partial h_l^T} [\mathbf{E} \mathbf{P} \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{h}]. \quad (\text{A.23})$$

By following the simplification given above (A.22), we obtain

$$\frac{\partial \boldsymbol{\mu}}{\partial \mathbf{h}^T} = \mathbf{E} \mathbf{P} \mathbf{F}^H \mathbf{D} \mathbf{W}. \quad (\text{A.24})$$

After substituting the derivatives in (A.22) and (A.24) into (A.20) and carrying out straightforward algebraic manipulations and simplifying, $\Xi_{D_{22}}$ is found to be an $L \times L$ matrix and can be obtained as

$$\Xi_{D_{22}} = \mathbb{E}_{\theta}[\mathbf{W}^H \mathbf{D}^H \Sigma^{-1} \mathbf{D} \mathbf{W}] = \mathbf{W}^H \mathbf{D}^H \Sigma^{-1} \mathbf{D} \mathbf{W}. \quad (\text{A.25})$$

3) *Computation of $\xi_{D_{33}} = \mathbb{E}_{\theta}[\psi_{33}(\boldsymbol{\theta}, \mathbf{h}, \epsilon)]$:* Based on (A.3) and (A.12), it follows that:

$$\begin{aligned} \mathbb{E}_{\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon}[-\Delta_{\epsilon}^{\epsilon} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon)|_{\mathbf{h}, \epsilon}] &= \mathbb{E}_{\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon} \left\{ \left[\frac{\partial}{\partial \epsilon^*} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon) \right] \left[\frac{\partial}{\partial \epsilon^T} \log p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{h}, \epsilon) \right]^H \right\} \\ &= \frac{\partial \boldsymbol{\mu}^H}{\partial \epsilon^*} \Sigma^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \epsilon^T}, \end{aligned} \quad (\text{A.26})$$

where $\frac{\partial \boldsymbol{\mu}^H}{\partial \epsilon^*}$ is a $1 \times N$ vector and given by

$$\begin{aligned} \frac{\partial \boldsymbol{\mu}^H}{\partial \epsilon^*} &= \frac{\partial}{\partial \epsilon^*} \left[\mathbf{h}^H \mathbf{W}^H \mathbf{D}^H \mathbf{F} \mathbf{P}^H \mathbf{E}^H \right] \\ &= \mathbf{h}^H \mathbf{W}^H \mathbf{D}^H \mathbf{F} \mathbf{P}^H \dot{\mathbf{E}}^H. \end{aligned} \quad (\text{A.27})$$

where $\dot{\mathbf{E}} \triangleq \text{diag}([0, \frac{j2\pi}{N} e^{(j2\pi\epsilon/N)}, \dots, \frac{j2\pi(N-1)}{N} e^{(j2\pi\epsilon/N) \times (N-1)}]^T)$. Similar to (A.27), we have

$$\begin{aligned} \frac{\partial \boldsymbol{\mu}^H}{\partial \epsilon^T} &= \frac{\partial}{\partial \epsilon^T} \left[\mathbf{E} \mathbf{P} \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{h} \right] \\ &= \dot{\mathbf{E}} \mathbf{P} \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{h}. \end{aligned} \quad (\text{A.28})$$

After substituting the derivatives in (A.27) and (A.28) into (A.26) and carrying out straightforward algebraic manipulations and simplifying, $\xi_{D_{33}}$ is found to be a scalar and can be obtained as

$$\xi_{D_{33}} = \mathbb{E}_{\theta}[\mathbf{h}^H \mathbf{W}^H \mathbf{D}^H \mathbf{F} \Sigma^{-1} \mathbf{M} \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{h}] = \mathbf{h}^H \mathbf{W}^H \mathbf{D}^H \mathbf{F} \Sigma^{-1} \mathbf{M} \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{h}, \quad (\text{A.29})$$

where $\mathbf{M} \triangleq \text{diag} \left([(2\pi \frac{0}{N})^2, (2\pi \frac{1}{N})^2, \dots, (2\pi \frac{N-1}{N})^2]^T \right)$.

4) *Computation of the remaining terms in (A.2):* By following similar steps as in (A.13)-(A.29), it can be found that

- $\Xi_{D_{12}} = \mathbb{E}_{\theta}[\Psi_{12}(\boldsymbol{\theta}, \mathbf{h}, \epsilon)]$ is an $N \times L$ matrix and given by:

$$\Xi_{D_{12}} = \frac{-j}{\sigma_w^2} \text{diag}(\mathbf{h}^H \mathbf{W}^H \mathbf{D}^H \mathbf{F}) \mathbf{F}^H \mathbf{D} \mathbf{W} \quad (\text{A.30})$$

and $\Xi_{D_{21}}$ can be determined as $\Xi_{D_{21}} = \Xi_{D_{12}}^T$,

- $\xi_{D_{13}} = \mathbb{E}_{\theta}[\psi_{13}(\boldsymbol{\theta}, \mathbf{h}, \epsilon)]$ is a $N \times 1$ vector and given by

$$\xi_{D_{13}} = \frac{1}{\sigma_w^2} \text{diag}(\mathbf{h}^H \mathbf{W}^H \mathbf{D}^H \mathbf{F} \sqrt{\mathbf{M}}) \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{h} \quad (\text{A.31})$$

and $\xi_{D_{31}}$ can be determined as $\xi_{D_{31}} = \xi_{D_{13}}^T$,

- $\xi_{D_{23}} = \mathbb{E}_{\theta}[\psi_{23}(\theta, \mathbf{h}, \epsilon)]$ is a $L \times 1$ vector and given by

$$\xi_{D_{23}} = j\mathbf{W}^H \mathbf{D}^H \mathbf{F} \Sigma^{-1} \sqrt{\mathbf{M}} \mathbf{F}^H \mathbf{D} \mathbf{W} \mathbf{h} \quad (\text{A.32})$$

and $\xi_{D_{32}}$ can be determined as $\xi_{D_{32}} = \xi_{D_{23}}^T$,

B. Computation of $\Xi_P \triangleq \mathbb{E}_{\theta|\mathbf{h}, \epsilon} [-\Delta_{\lambda}^{\lambda} \log p(\theta|\mathbf{h}, \epsilon)|\mathbf{h}, \epsilon]$

The second factor in HIM, defined in (A.1), can be written as:

$$\begin{aligned} \Xi_P &= \mathbb{E}_{\theta|\mathbf{h}, \epsilon} [-\Delta_{\lambda}^{\lambda} \log p(\theta|\mathbf{h}, \epsilon)|\epsilon] \triangleq \begin{bmatrix} \Xi_{P_{11}} & \Xi_{P_{12}} & \xi_{P_{13}} \\ \Xi_{P_{21}} & \Xi_{P_{22}} & \xi_{P_{23}} \\ \xi_{P_{31}} & \xi_{P_{32}} & \xi_{P_{33}} \end{bmatrix} \\ &= \begin{bmatrix} \mathbb{E}_{\theta} [-\Delta_{\theta}^{\theta} \log p(\theta)] & \mathbb{E}_{\theta} [-\Delta_{\theta}^{\mathbf{h}} \log p(\theta)] & \mathbb{E}_{\theta} [-\Delta_{\theta}^{\epsilon} \log p(\theta)] \\ \mathbb{E}_{\theta} [-\Delta_{\mathbf{h}}^{\theta} \log p(\theta)]^T & \mathbb{E}_{\theta} [-\Delta_{\mathbf{h}}^{\mathbf{h}} \log p(\theta)] & \mathbb{E}_{\theta} [-\Delta_{\mathbf{h}}^{\epsilon} \log p(\theta)] \\ \mathbb{E}_{\theta} [-\Delta_{\epsilon}^{\theta} \log p(\theta)]^T & \mathbb{E}_{\theta} [-\Delta_{\epsilon}^{\mathbf{h}} \log p(\theta)] & \mathbb{E}_{\theta} [-\Delta_{\epsilon}^{\epsilon} \log p(\theta)] \end{bmatrix}. \end{aligned} \quad (\text{A.33})$$

where $p(\theta)$ is the prior distribution of θ .

1) *Computation of $\Xi_{P_{11}} \triangleq \mathbb{E}_{\theta} [-\Delta_{\theta}^{\theta} \log p(\theta)]$* : From [37, eq.(19)], we obtain the $N \times N$ matrix $\mathbb{E}_{\theta} [-\Delta_{\theta}^{\theta} \log p(\theta)]$

as

$$\Xi_{P_{11}} = \frac{-1}{\sigma_{\delta}^2} \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & 0 & 1 & -2 & 1 \\ 0 & \cdots & 0 & 1 & -1 \end{bmatrix}. \quad (\text{A.34})$$

2) *Computation of $\Xi_{P_{12}} \triangleq \mathbb{E}_{\theta} [-\Delta_{\theta}^{\mathbf{h}} \log p(\theta)]$, $\Xi_{P_{21}} \triangleq \mathbb{E}_{\theta} [-\Delta_{\mathbf{h}}^{\theta} \log p(\theta)]^T$, $\Xi_{P_{22}} \triangleq \mathbb{E}_{\theta} [-\Delta_{\mathbf{h}}^{\mathbf{h}} \log p(\theta)]$, $\xi_{P_{13}} \triangleq \mathbb{E}_{\theta} [-\Delta_{\theta}^{\epsilon} \log p(\theta)]$, $\xi_{P_{31}} \triangleq \mathbb{E}_{\theta} [-\Delta_{\epsilon}^{\theta} \log p(\theta)]^T$, $\xi_{P_{23}} \triangleq \mathbb{E}_{\theta} [-\Delta_{\mathbf{h}}^{\epsilon} \log p(\theta)]$, $\xi_{P_{32}} \triangleq \mathbb{E}_{\theta} [-\Delta_{\epsilon}^{\mathbf{h}} \log p(\theta)]$,*

and $\xi_{P_{33}} \triangleq \mathbb{E}_{\boldsymbol{\theta}} [-\Delta_{\epsilon}^{\epsilon} \log p(\boldsymbol{\theta})]$: Since CFO is a deterministic parameter and no prior knowledge of \mathbf{h} is assumed, we have

$$\Xi_{P_{12}} = \Xi_{P_{21}}^T = \mathbf{0}_{N \times L}, \quad (\text{A.35})$$

$$\Xi_{P_{22}} = \mathbf{0}_{L \times L}, \quad (\text{A.36})$$

$$\boldsymbol{\xi}_{P_{13}} = \boldsymbol{\xi}_{P_{31}}^T = \mathbf{0}_{N \times 1}, \quad (\text{A.37})$$

$$\boldsymbol{\xi}_{P_{23}} = \boldsymbol{\xi}_{P_{32}}^T = \mathbf{0}_{L \times 1}, \quad (\text{A.38})$$

$$\xi_{P_{33}} = 0. \quad (\text{A.39})$$

Using the above results, we can evaluate the HIM in (6), since $\mathbf{B}_{11} = \Xi_{D_{11}} + \Xi_{P_{11}}$, $\mathbf{B}_{22} = \Xi_{D_{22}} + \Xi_{P_{22}} = \Xi_{D_{22}}$, $\mathbf{b}_{33} = \xi_{D_{33}} + \xi_{P_{33}} = \xi_{D_{33}}$, $\mathbf{B}_{12} = \mathbf{B}_{21}^T = \Xi_{D_{12}} + \Xi_{P_{12}} = \Xi_{D_{12}}$, $\mathbf{b}_{13} = \mathbf{b}_{31}^T = \boldsymbol{\xi}_{D_{13}} + \boldsymbol{\xi}_{P_{13}} = \boldsymbol{\xi}_{D_{13}}$, and $\mathbf{b}_{23} = \mathbf{b}_{32}^T = \boldsymbol{\xi}_{D_{23}} + \boldsymbol{\xi}_{P_{23}} = \boldsymbol{\xi}_{D_{23}}$.

REFERENCES

- [1] L. Nuaymi, *WiMAX: Technology for broadband wireless access*, New York, USA, John Wiley & Sons, 2007.
- [2] Y. Zhang and H. -H. Chen, *Mobile WIMAX: Toward broadband wireless metropolitan area networks*, New York, USA, Taylor & Francis Group, 2008.
- [3] T. Paul and T. Ogunfunmi, “Wireless LAN comes of age: Understanding the IEEE 802.11n amendment,” *IEEE Circuits Syst. Mag.*, vol. 8, no. 1, pp. 28–54, 2008.
- [4] X. Zhu, A. Doufexi, and T. Kocak, “Throughput and coverage performance for IEEE 802.11ad millimeter-wave WPANs,” in *Proc. IEEE VTC (Spring)*, Budapest, Hungary, May 2011, pp. 1–5.
- [5] P. Chevillat and D. Maiwald and G. Ungerboeck, “Rapid training of a voiceband data-modem receiver employing an equalizer with fractional-T spaced coefficients,” *IEEE Commun. Mag.*, vol. 35, pp. 869–876, Sep. 1987.
- [6] T. Schmidl and D. Cox, “Robust frequency and timing synchronization for OFDM,” *IEEE Commun. Mag.*, vol. 45, no. 12, pp. 1613–1621, Dec. 1997.
- [7] A. Chorti and M. Brookes, “A spectral model for RF oscillators with power-law phase noise,” *IEEE Trans. Circuits Syst.*, vol. 53, no. 9, pp. 1989–1999, Sep. 2006.

- [8] D. D. Lin, R. Pacheco, T. J. Lim, and D. Hatzinakos, "Joint estimation of channel response, frequency offset, and phase noise in OFDM," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3542–3554, Sept. 2006.
- [9] A. Armada and M. Calvo, "Phase noise and sub-carrier spacing effects on the performance of an OFDM communication system," *IEEE Commun. Lett.*, vol. 2, no. 1, pp. 11–13, Jun. 1998.
- [10] A. Armada and M. Calvo, "Understanding the effects of phase noise in orthogonal frequency division multiplexing (OFDM)," *IEEE Trans. Broadcast.*, vol. 47, no. 2, pp. 153–159, Jun. 2001.
- [11] D. Petrovi, W. Rave, and G. Fettweis, "Phase noise suppression in OFDM including intercarrier interference," in *Proc. Int. OFDM Workshop*, 2003, pp. 219–224.
- [12] D. Petrovi, W. Rave, and G. Fettweis, "Phase noise suppression in OFDM using Kalman filter," in *Proc. WPMC*, 2003.
- [13] D. Petrovi, W. Rave, and G. Fettweis, "Effects of phase noise on OFDM systems with and without PLL: Characterization and compensation," *IEEE Trans. Commun.*, vol. 55, no. 8, pp. 1607–1616, Aug. 2007.
- [14] H. Mehrpouyan, M. R. Khanzadi, M. Matthaiou, A. M. Sayeed, R. Schober, and Y. Hua, "Improving bandwidth efficiency in E-band communication systems," *IEEE Commun. Mag.*, in press.
- [15] J. -H. Lee, J. C. Han, and S. -C. Kim, "Joint carrier frequency synchronization and channel estimation for OFDM systems via the EM algorithm," *IEEE Trans. Veh. Technol.*, vol. 55, no. 1, pp. 167–172, Jan. 2006.
- [16] J. Tao, J. Wu, and C. Xiao, "Estimation of channel transfer function and carrier frequency offset for OFDM systems with phase noise," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4380–4387, Oct. 2009.
- [17] R. Carvajal, J. C. Aguero, B. I. Godoy, and G. C. Goodwin, "EM-based maximum-likelihood channel estimation in multicarrier systems with phase distortion," *IEEE Trans. Veh. Technol.*, vol. 62, no. 1, pp. 152–160, Jan. 2013.
- [18] F. Septier, Y. Delignon, A. Menhaj-Rivenq, and C. Garnier, "Monte Carlo methods for channel, phase noise, and frequency offset estimation with unknown noise variances in OFDM systems," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3613–3626, Aug. 2008.

- [19] R. Wang and H. Mehrpouyan and M. Tao and Y. Hua, "Channel estimation, carrier recovery, and data detection in the presence of phase noise in OFDM relay systems," *submitted to IEEE Trans. Signal Proc.* [Online]. Available: <http://arxiv.org/pdf/1309.7564.pdf>
- [20] D. D. Lin and T. J. Lim, "The variational inference approach to joint data detection and phase noise estimation in OFDM," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1862–1874, May 2007.
- [21] Y. Gong and X. Hong, "A new algorithm for OFDM joint data detection and phase noise cancellation," in *Proc. IEEE ICC*, Beijing, China, May 2008, pp. 636–640.
- [22] P. Rabiei, W. Namgoong, and N. Al-Dhahir, "A non-iterative technique for phase noise ICI mitigation in packet-based OFDM systems," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5945–5950, Nov. 2010.
- [23] O. H. Salim, A. A. Nasir, H. Mehrpouyan, and W. Xiang, "Phase noise and carrier frequency offset in OFDM systems: Joint estimation and hybrid Cramér-Rao lower bound," in *Proc. IEEE SPAWC*, Darmstadt, Germany, Jun. 2013.
- [24] A. M. A. Demir and J. Roychowdhury, "Phase noise in oscillators: A unifying theory and numerical methods for characterization," *IEEE Trans. Biomed. Circuits Syst.*, vol. 47, pp. 655–674, May 2000.
- [25] D. Lin, R. Pacheco, T. J. Lim, and D. Hatzinakos, "Optimal OFDM channel estimation with carrier frequency offset and phase noise," in *Proc. IEEE WCNC*, Apr. 2006.
- [26] S. Wu and Y. Bar-Ness, "A phase noise suppression algorithm for OFDM-based WLANs," *IEEE Commun. Lett.*, vol. 6, no. 12, pp. 535–537, Dec. 2002.
- [27] F. Munier, T. Eriksson, and A. Svensson, "An ICI reduction scheme for OFDM system with phase noise over fading channels," *IEEE Trans. Commun.*, vol. 56, no. 12, pp. 1119–1126, Jul. 2008.
- [28] H. L. van Trees and K. L. Bell, *Bayesian bounds for parameter estimation and nonlinear filtering/tracking*. New York: Wiley-Interscience; Wiley; IEEE, 2007.
- [29] E. P. Simon, L. Ros, H. Hijazi, and M. Ghogho, "Joint carrier frequency offset and channel estimation for OFDM systems via the EM algorithm in the presence of very high mobility," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 754–765, Feb. 2012.
- [30] S. M. Kay, *Fundamentals of Statistical Signal Processing, Estimation Theory*. Signal Processing Series: Prentice Hall, 1993.

- [31] X. L. Meng and D. B. Rubin, "Maximum likelihood estimation via the ECM algorithm: A general framework," *IEEE Trans. Commun., Biometrika*, vol. 8, no. 2, pp. 267–278, Jun. 1993.
- [32] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. Hoboken, NJ: Wiley, 2008.
- [33] T. A. Fesler and A. O. Hero, "Space-alternating generalized expectation maximization algorithm," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2664–2677, Oct. 1994.
- [34] H. Mehrpouyan and S. D. Blostein, "Bounds and algorithms for multiple frequency offset estimation in cooperative networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1300–1311, Apr. 2011.
- [35] H. Mehrpouyan, A. A. Nasir, S. D. Blostein, T. Eriksson, G. K. Karagiannidis, and T. Svensson, "Joint estimation of channel and oscillator phase noise in MIMO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4790–4807, Sept. 2012.
- [36] S. Zhang, F. Gao, and C.-X. Pei, "Optimal training design for individual channel estimation in two-way relay networks," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4987–4991, Sept. 2012.
- [37] S. Bay, C. Herzet, J. -M. Brossier, J. -P. Barbot, and B. Geller, "Analytic and asymptotic analysis of Bayesian Cramer-Rao bound for dynamical phase offset estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4987–4991, Sept. 2012.
- [38] T. J. Richardson and R. L. Urbanke, "Efficient encoding of low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 638 - 656, Feb. 2001.
- [39] S. J. Johnson, "Introducing low-density parity-check codes," [Online]. Available: <http://sigpromu.org/sarah/SJohnsonLDPCintro.pdf>