IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 7, 2017, accepted May 23, 2017, date of publication June 1, 2017, date of current version June 28, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2709325

Channel Estimation and Data Detection in the Presence of Phase Noise in MIMO-OFDM Systems With Independent Oscillators

TAE-JUN LEE[1], (Student Member, IEEE), AND YOUNG-CHAI KO[2], (Senior Member, IEEE)
[1]Department of IT Convergence, Korea University, Seoul 136-713, South Korea
[2]School of Electrical Engineering, Korea University, Seoul 136-713, South Korea

Corresponding author: Young-Chai Ko (koyc@korea.ac.kr)

This work was supported by ICT R&D Program of MSIP/IITP, Next Generation WLAN System with High Efficient Performance, under Grant 2014- 044-006-004.

ABSTRACT In this paper, we propose the mitigating scheme to reduce the effects of phase noise in multi-input multi-output-orthogonal frequency division multiplexing system with independent oscillator in each RF chain. Our proposed schemes consist of two stages; channel estimation stage and data decoding stage. In the first stage, we propose the channel estimation algorithm based on maximum *a posteriori* (MAP) estimator and a method of selecting the training sequence for channel estimation and we provide the mathematical analysis of our proposed scheme. In the second stage, MAP estimators are used to jointly estimate phase noise at TX and RX and detect data symbols. For analysis of mean square error (MSE) performances, we derive Bayesian Cramér-Rao bound for multi-parameter estimation problem in each stage. At the end of this paper, we demonstrate from our simulation results that our mathematical analysis is accurate and the proposed algorithm improves the bit-error-rate performance and MSE performance compared with existing schemes.

INDEX TERMS MIMO, OFDM, phase noise, RF impairment, mmWave, oscillators, channel estimation, common phase error (CPE), inter-carrier interference (ICI).

I. INTRODUCTION

Multi input multi output (MIMO) orthogonal frequency division multiplexing (OFDM) system is adopted in many wireless communication systems such as IEEE 802.11 ac/ad wireless local area networks (WLAN) [1], [2] and 3GPP LTE [3] not only to achieve high spectral efficiency using spatial multiplexing, but also to be robust to frequency selective channels. However, OFDM systems suffer from phase noise which is a multiplicative phase distortion and generated by non-ideal property of the imperfect oscillators during up-conversion and down-conversion [4], [5]. While the single-carrier modulated signals are affected by the phase error in a symbol unit, OFDM transmits data symbols over many low-rate subcarriers and their phase noise is convolved with data symbols. This makes it more difficult to estimate and track phase noise. There are two effects of phase noise on the received OFDM symbols; common phase error (CPE) and inter-carrier interference (ICI). CPE is a common phase rotation of all the subcarriers in an OFDM symbol. On the other hand, ICI violates the orthogonality between the subcarriers and behaves like Gaussian noise. Those effects are greatly detrimental to synchronization and deteriorate signal-to-interference-plus-noise ratio (SINR).

In the traditional low frequency bands (lower than 10 GHz), phase noise is small enough to be ignored [6]. Hence both 3GPP LTE [3] and IEEE 802.11 ac WLAN standardization documents [1] do not specify options associated with phase noise. However, note that the variance of phase noise increases quadratically versus the carrier frequency generated by oscillators [7]. Therefore, in mmWave wireless systems such as 60GHz WLAN standards [2] and some candidate bands for the 5th generation cellular systems, the large phase noise compared to relatively low frequency band under 10GHz is one of the critical issues to be solved for the successful deployments. For this reason, although there are many synchronization problems such as timing offset, frequency offset and IQ imbalance in OFDM systems, we focus on the phase noise problem in this paper.

Phase noise mitigation algorithms in MIMO-OFDM systems have been studied in many papers [8]–[12]. All of

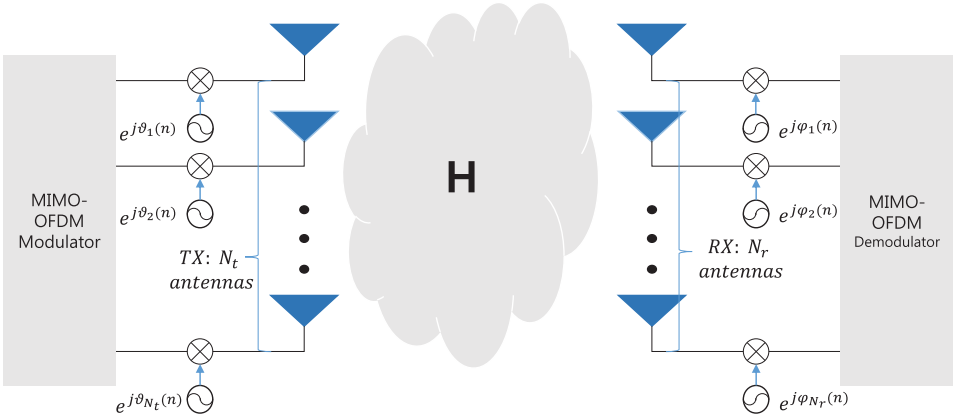

FIGURE 1. $N_t \times N_r$ point to point MIMO-OFDM systems with independent oscillators.

the phase noise cancellation algorithms mentioned above assume that transmit (TX) and receive (RX) antennas share a common oscillator, so there is only one phase noise process at TX and RX sides, respectively. In this case, phase noise mitigation algorithms in single input single output (SISO) OFDM systems can be extended and modified easily to MIMO-OFDM systems. However, in many cases of MIMO systems, independent oscillator may be equipped on each of RF chains both at TX and RX transceivers, respectively, as depicted in Fig. 1 [13]. In the case of line-of-sight (LOS) MIMO environments, which are often experienced in mmWave systems, as antennas need to be placed far apart from one another to create full rank channel, all TX or RX antennas cannot be shared by a single oscillator [14], [15]. Also, in multiuser MIMO and space division multiple access (SDMA) systems, each user has a separate user equipment (UE) and transmits their signals to a common receiver [13]. For another example, as in [16], the implementation of massive MIMO in circuit level can be simplified with the resilience against the phase noise when the independent oscillators are equipped.

On the other hand, although equipping the independent oscillators in the RF transceiver becomes more general in mmWave systems and the effect of phase noise becomes very serious compared to the relatively low bands under 10GHz, the studies on solving the phase noise problems in the systems are scarce. In [17], the effects of phase noise at both TX and RX sides in MIMO-OFDM systems are analyzed. In [13], joint estimation of channel and phase noises, and Cramėr-Rao Bound (CRLB) analysis have been proposed. In [18], authors present expectation and maximization (EM) based on phase noise estimator. However, all the above papers made the mathematical analysis or phase noise estimators in the single carrier systems. In [19], authors propose channel and CPE estimation algorithm in MIMO-OFDM systems by using least square (LS) approach. However this algorithm cannot remove ICI which gives a disastrous effect specifically in case of independent oscillators. In [20], the authors consider independent oscillators equipped only at RX end but a common oscillator at TX end, which is not a general implementation.

In this paper, our main challenge is to improve the system performances in MIMO-OFDM systems, where each TX and RX end is equipped with independent oscillators. It is general consideration on phase noise problem in OFDM systems compared with existing literature such as [8], [10], and [21]–[23]. We propose the mitigation schemes of phase noise in MIMO-OFDM systems, where each TX and RX end will encounter different phase noise. The schemes can be divided into two stage; channel estimation and data decoding. First, for channel estimation in MIMO-OFDM systems, designing training sequences is a critical issue to reduce mean squared error (MSE) of channel impulse response (CIR). In [24] and [25], when using LS channel estimation, the authors have shown that optimal pilot tone allocations under absence of phase noise require orthogonality among TX ends in time, frequency or code domain. However, neglecting the effects of phase noise in design of the preamble may lead to a significant loss in accuracy of the channel estimation. In [26], optimal training symbols are presented in the presence of frequency offset and phase noise. However, the authors in [26] consider the system that TX and RX antennas share a common oscillator, and only the case of LS channel estimation. Thus, in our paper, the effective training sequence design in our systems will be proposed through mathematical analysis of MSE. Also, we propose maximum a posteriori (MAP) estimator of CIR from mathematically derived covariance of ICI. In data decoding stage, we iteratively estimate multiple phase noises and detect data symbols using MAP estimation.

The rest of our paper is organized as follows. In Section II, the MIMO system model and phase noise model used throughout the paper are outlined. Section III introduces the techniques to estimate channel coefficients and select training sequences. Also Bayesian CRLB (BCRLB) for unknown parameter and mathematical analysis are derived.

In Section IV, our proposed algorithm in data decoding is presented and BCRLB of phase noise is derived. For the proof of validation of analysis and comparison of the schemes given in Section III and IV, Section V shows the bit-error-rate (BER) and MSE performances from numerical and simulation results. Finally, Section VII summarizes the paper and concludes our contribution.

Throughout this paper, we adopt the following notations: First Bold face small letters and large letters represent vectors and matrices like $\{\mathbf{x}, \mathbf{X}\}$. $\mathbf{1}_N$ and $\mathbf{0}_N$ represent all one and zero $N \times 1$ vector, and $\mathbf{1}_{N\times M}$ and $\mathbf{0}_{N\times M}$ represent all one and zero $N \times M$ matrix. $(\cdot)_D$ and $(\cdot)_{Cir}$ represent a diagonal matrix and circulant matrix of vector which are represented by same alphabet, respectively. And $(\mathbf{X})_{a,b}$ is the (a, b)-th entry of $\mathbf{X}$. In operators, $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$ denote transpose, conjugate and conjugate transpose operator, respectively. $\mathbf{E}[\cdot]$, $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote expectation, real, imaginary operator, respectively. Notations of $*$, $\otimes$, $\circ$ and $\circledast$ denote convolution, circular convolution, Hardarmard product and Kronecker product operator, respectively. Finally, diag $(\cdot)$ is an operator to transform an input vector into a diagonal matrix or an input matrix into a vector using only diagonal elements in the matrix and circ $(\cdot)$ is an operator to transform an input vector into a circulant matrix. **diag** $(\cdot)$ denotes an block diagonal operator to convert matrices with same size to an block diagonal matrix. In this paper, j has two meaning. When denoted by a letter subscript such as $\mathbf{y}_j$, it means the index of receive antenna. While denoted by a letter such as j, it represent an imaginary part.

II. SYSTEM MODEL

A. MIMO-OFDM

Typically up to four transmit antennas can be handled with one RF chain. But those antennas usually are used to achieve diversity gain by beamforming. In our paper, we assume that each antenna means one baseband port and our system uses multi antenna to achieve multiplexing gain. We consider a $N_t \times N_r$ point-to-point MIMO-OFDM system where TX end transmits M data streams ($M \leq \min(N_t, N_r)$) to RX end. As seen in Fig. 1, each antenna is connected to the independent local oscillator.

Let $s_i(k)$ be the data symbol at the k-th frequency domain which is transmitted from the i-th transmitter. After normalized inverse discrete Fourier transform (IDFT), the transmitted signal in discrete time domain can be written as

$$x_i(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} s_i(k) e^{\frac{j2\pi nk}{N}}, \tag{1}$$

where $x_i(n)$ is the n-th transmitted signal in the time domain of the i-th TX antenna and N is the number of subcarriers in one OFDM symbol.

The baseband signal is converted into the analog signal and upcoverted by the local oscillator, which generates phase noise by difference between carrier signal and the local oscillator. After experiencing multipath channels $\mathbf{g}_{ji} = \left[g_{ji}(0), g_{ji}(1), \cdots, g_{ji}(L-1)\right] \in \mathbb{C}^{L\times 1}$, the channel impulse response (CIR) vector with L length between the ith TX and the jth RX, the downconverted baseband signal at the jth RX antenna is given by

$$y_j(n) = \sum_{i=1}^{N_t} x_i(n) e^{j\theta_i(n)} \otimes g_{ji}(n) e^{j\phi_j(n)} + w_j(n), \tag{2}$$

where $y_j(n)$ is the n-th received signal in the time domain of the j-th RX antenna, and $\theta_i(n)$ and $\phi_j(n)$ represent TX phase noise of the i-th TX antenna and RX phase noise of the j-th RX antenna, respectively. Also, in (2), $w_j(n)$ is the n-th additive white Gaussian noise (AWGN) of the j-th RX antenna modeled as the complex Gaussian random variable $w_j(n) \sim \mathcal{CN}\left(0, \sigma_w^2\right)$.

To simplify the received signals in the j-th RX ends, (2) can be rewritten in vector form as

$$\mathbf{y}_j = \mathbf{P}_{\phi_j,D} \mathbf{F}^H \mathbf{H}_j \mathbf{F}_{N_t,D} \mathbf{P}_{\theta,D} \mathbf{x} + \mathbf{w}_j, \tag{3}$$

and the total received signal vector is given by

$$\mathbf{y} = \mathbf{P}_{\phi,D} \mathbf{F}_{N_r,D}^H \mathbf{H} \mathbf{F}_{N_t,D} \mathbf{P}_{\theta,D} \mathbf{x} + \mathbf{w}, \tag{4}$$

where $\mathbf{y}_j = \left[y_j(0), y_j(1), \cdots, y_j(N-1)\right]^T \in \mathbb{C}^{N\times 1}$ and $\mathbf{y} = \left[\mathbf{y}_1^T, \mathbf{y}_2^T, ..., \mathbf{y}_{N_r}^T\right]^T \in \mathbb{C}^{N_r N\times 1}$ denote the received signal vector of the j-th RX antenna and the total received signal vector in the time domain, respectively, and $\mathbf{x}_i = [x_i(0), x_i(1), \cdots, x_i(N-1)]^T \in \mathbb{C}^{N\times 1}$ and $\mathbf{x} = \left[\mathbf{x}_1^T, \mathbf{x}_2^T, ..., \mathbf{x}_{N_t}^T\right]^T \in \mathbb{C}^{N_t N\times 1}$ denote the transmitted signal vector of the i-th TX antenna and the total transmitted signal vector in the time domain, respectively, and $\mathbf{w}_j = \left[w_j(0), w_j(1), \cdots, w_j(N-1)\right]^T \in \mathbb{C}^{N\times 1}$ and $\mathbf{w} = \left[\mathbf{w}_1^T, \mathbf{w}_2^T, ..., \mathbf{w}_{N_r}^T\right]^T \in \mathbb{C}^{N_t N\times 1}$ denote the noise vector of the j-th RX antenna and the total noise vector in the time domain, respectively. Note that in (3) and (4), $\boldsymbol{\phi}_j = \left[\phi_j(0), \phi_j(1), \cdots, \phi_j(N-1)\right]^T \in \mathbb{R}^{N\times 1}$ and $\boldsymbol{\phi} = \left[\boldsymbol{\phi}_1^T, \boldsymbol{\phi}_2^T, ..., \boldsymbol{\phi}_{N_r}^T\right]^T \in \mathbb{R}^{N_r N\times 1}$ are the phase noise vector of the j-th RX antenna and the total phase noise vector, respectively. Denoting $\mathbf{p}_{\phi_j}$ and $\mathbf{p}_{\phi}$ as the phase error vector, $\exp(j\boldsymbol{\phi}_j)$ and $\exp(j\boldsymbol{\phi})$, respectively, we denote $\mathbf{P}_{\phi_j,D}$ and $\mathbf{P}_{\phi,D}$ as the diagonal phase error matrices, diag $\left(\mathbf{p}_{\phi_j}\right)$ and diag $\left(\mathbf{p}_{\phi}\right)$, respectively. TX phase noise vectors, $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_i$, have the same structure with RX phase noise vectors , $\boldsymbol{\phi}$ and $\boldsymbol{\phi}_j$. $\mathbf{F}_{N_r,D}$ and $\mathbf{H}_j$ represent the block discrete Fourier transform (DFT) matrix and the channel matrix in j-th RX end given by

$$\mathbf{F}_{N_r,D} = \mathbf{diag}\Big(\underbrace{[\mathbf{F}, \mathbf{F}, \cdots, \mathbf{F}]}_{N_r}\Big),$$

$$\mathbf{H}_j = \left[\mathbf{H}_{j1,D}, \mathbf{H}_{j2,D}, \cdots, \mathbf{H}_{jN_t,D}\right],$$

where is $\mathbf{F}$ a nomalized DFT matrix, $\mathbf{h}_{ji} = \sqrt{N}\mathbf{F}\mathbf{g}_{ji} \in \mathbb{C}^{N\times 1}$ is the channel frequency response (CFR) with between the i-th TX and the j-th RX antennas, and $\mathbf{H}_{ji,D}$ is the diagonal matrix of the vector $\mathbf{h}_{ji}$, diag$(\mathbf{h}_{ji})$. Finally, the total channel

matrix, $\mathbf{H}$, is given by

$$\mathbf{H} = \left[\mathbf{H}_1^T, \mathbf{H}_2^T, \cdots, \mathbf{H}_{N_r}^T\right]^T.$$

B. PHASE NOISE MODEL

In this paper, we consider the oscillator without a phase locked loop (PLL), which is often called as free-running oscillator [21]. In this case, the phase noise is modeled as Wiener process, given as

$$\phi_j(n) = \phi_j(n-1) + \zeta(n), \tag{5}$$

where $\zeta(n)$ is a independent and identically distributed (i.i.d.) Gaussian random variable following $\zeta(n) \sim \mathcal{N}\left(0, \sigma_\zeta^2\right)$. Its variance is given by $\sigma_\zeta^2 = \frac{2\pi\beta T_s}{N}$, where β denotes the two-sided 3-dB bandwidth of Lorentzian spectrum of the oscillator [22]. From (5), we can figure out that $\boldsymbol{\phi}_j$ is modeled as a Gaussian random vector $\mathcal{N}\left(0, \Phi_j\right)$, where the covariance matrix Φ_j is shown to be given as [27]

$$\left(\Phi_j\right)_{k,l} = \sigma_\zeta^2 \min(k, l). \tag{6}$$

As the phase noise with independent and identical distribution occurs at each antenna, the distribution of $\boldsymbol{\phi}$ is given by

$$\boldsymbol{\phi} \sim \mathcal{N}(0, \Phi), \Phi = \mathbf{diag}\left(\left[\Phi_1, \Phi_2, \cdots, \Phi_{N_r}\right]\right). \tag{7}$$

Also, in the case of TX phase noise, $\boldsymbol{\theta}$ follows the same distribution as RX phase noise given in (7) and thus we have

$$\boldsymbol{\theta} \sim \mathcal{N}(0, \Theta), \Theta = \mathbf{diag}\left(\left[\Theta_1, \Theta_2, \cdots, \Theta_{N_t}\right]\right). \tag{8}$$

However, other systems may implement the oscillators controlled by PLL, where phase noise process can be approximately as a zero-mean colored Gaussian process [23]. As the covariance matrix of the Gaussian process, Φ and Θ, can be calculated from specification of the oscillators, our proposed algorithm can be applied by substituting the covariance matrices of phase noise.

C. FRAME STRUCTURE AND TWO STAGE OF ESTIMATION

In this paper, we assume that transmitted signal experiences a block fading channel that does not vary within a frame. Hence, the entire system procedure is based on one frame, which consists of two different types of symbols; the block-type pilot symbol and the comb-type pilot symbol, as illustrated in Fig. 2. Many papers [12], [19], [21], [23], [25], [26] and standards of OFDM systems [1], [3] adopt the similar frame structure depicted in Fig. 2.

We investigate the schemes mitigating phase noise in both channel estimation and data decoding. In channel estimation stage, we consider the block-type pilot symbols as the training symbols which are transmitted in the beginning of a frame. Due to the fact that phase noise is usually varying much faster than the channel, only one OFDM symbol in time domain is used to estimate the channel because the estimated channel can be differently rotated at each symbol by time varying phase noise. Thus, the number of subcarriers, N, is larger

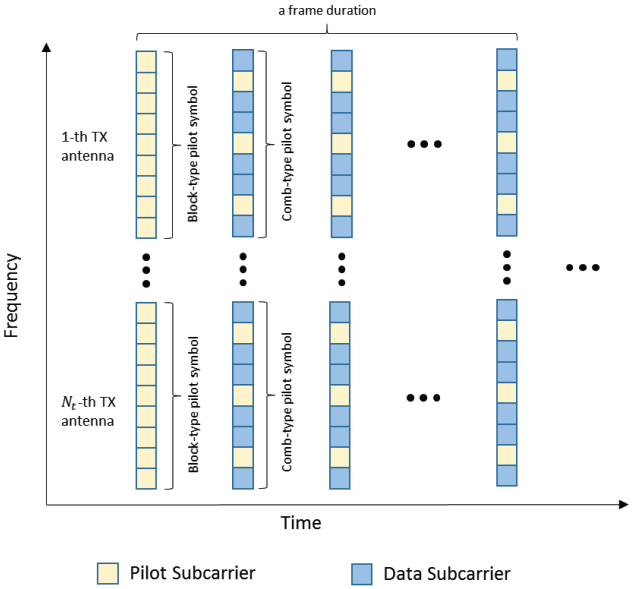

FIGURE 2. Frame structure of proposed shemes.

than the total taps of CIR to be estimated at each RX end, $L \times N_t$. In data decoding stage, the pilot subcarriers, which are inserted in the intervals of the comb-type pilot symbol, are utilized to estimate TX and RX phase noises.

III. CHANNEL ESTIMATION

When estimating the channels in the presence of phase noise, we take two steps to reduce the effects of phase noise. The first step is to estimate the channel based on Bayesian approach and the other step is to select the combination of training symbols to be robust against phase noise.

A. BCRLB OF CIR

Prior to proposing algorithms, it is important to find out the lower bound of MSE performance in the estimation problem. As we estimate random parameters using Bayesian approach in Section III-B, we derive BCRLB of parameters to be estimated in this section.

Let us define the unknown parameter vector $\boldsymbol{\eta}_1$ as

$$\boldsymbol{\eta}_1 \triangleq \left[\boldsymbol{\phi}_j^T, \boldsymbol{\theta}^T, \mathbf{g}_{j,re}^T, \mathbf{g}_{j,im}^T\right]^T \tag{9}$$

where $\mathbf{g}_j = \left[\mathbf{g}_{j1}^T, \mathbf{g}_{j2}^T, \cdots, \mathbf{g}_{jN_t}^T\right]^T$ is CIR of the j-th RX end, and $\mathbf{g}_{j,re}$ and $\mathbf{g}_{j,im}$ denote $\Re\left\{\mathbf{g}_j\right\}$ and $\Im\left\{\mathbf{g}_j\right\}$, respectively. In [28], Bayesian information matrix (BIM) is defined as

$$\mathbf{B}_1 = \mathbf{E}_{\boldsymbol{\eta}_1}\left[\Gamma_1\right] + \mathbf{E}_{\boldsymbol{\eta}_1}\left[-\Delta_{\boldsymbol{\eta}_1}^{\boldsymbol{\eta}_1} \ln p\left(\boldsymbol{\eta}_1\right)\right], \tag{10}$$

where

$$\Gamma_1 = \mathbf{E}_{\mathbf{y}_j|\boldsymbol{\eta}_1}\left[-\Delta_{\boldsymbol{\eta}_1}^{\boldsymbol{\eta}_1} \ln p\left(\mathbf{y}_j \mid \boldsymbol{\eta}_1\right)\right] \tag{11}$$

is Fisher information matrix (FIM) and $\Delta_{\boldsymbol{\eta}_1}^{\boldsymbol{\eta}_1} f \triangleq \frac{\partial f}{\partial \boldsymbol{\eta}_1}\left[\frac{\partial f}{\partial \boldsymbol{\eta}_1}\right]^T$ denotes the second order partial derivative of function f with

respect to the vector $\boldsymbol{\eta}_1$. In order to easily partial differentiate negative log-likelyhood function in (11) with respect to each parameter, we need to make signal term of (3) be divided into each parameter vector of $\boldsymbol{\eta}_1$ and remainder matrix. Hence, (3) can be rewritten as

$$\mathbf{y}_j = \Xi_1 \mathbf{p}_\theta + \mathbf{w}_j = \Lambda_1 \mathbf{p}_{\phi_j} + \mathbf{w}_j \tag{12}$$

where

$$\Xi_1 = \mathbf{P}_{\phi_j,D}\mathbf{F}^H\mathbf{H}_j\mathbf{F}_{N_t,D}\mathbf{X}_D,$$
$$\Lambda_1 = \mathrm{diag}\left(\mathbf{F}^H\mathbf{H}_j\mathbf{F}_{N_t,D}\mathbf{X}_D\mathbf{p}_\theta\right)$$

and $\mathbf{X}_D$ denotes diag $(\mathbf{x})$. Also, the signal part of (3) is divided into $\mathbf{g}_{ji}$ and the rest as

$$\mathbf{y}_j = \mathbf{P}_{\phi_{j,D}}\mathbf{F}^H \sum_{i=1}^{N_t} \mathrm{diag}\left(\mathbf{V}_{\theta_i}\mathbf{s}_i\right)\mathbf{F}_L\mathbf{g}_{ji} + \mathbf{w}_j, \tag{13}$$

where is $\mathbf{F}_L$ a $N \times L$ partial DFT matrix, and $\mathbf{v}_{\theta_i}$ and $\mathbf{V}_{\theta_i}$ denote $\sqrt{1/N}\mathbf{F}\mathbf{p}_{\theta_i}$ and circ $\left(\mathbf{v}_{\theta_i}\right)$, respectively. Also in (13), $\mathbf{s}_i = [s_i(0), s_i(1), \cdots, s_i(N-1)] \in \mathbb{C}^{N\times 1}$ and $\mathbf{s} = \left[\mathbf{s}_1^T, \mathbf{s}_2^T, ..., \mathbf{s}_{N_t}^T\right]^T \in \mathbb{C}^{N_tN\times 1}$ denote the transmitted signal vector of the i-th TX antenna and the total transmitted signal vector in the frequency domain, respectively. To remove summation term in (13), we can rewrite (13) as

$$\mathbf{y}_j = \mathbf{P}_{\phi_{j,D}}\mathbf{F}^H\tilde{\mathbf{S}}\mathbf{F}_{L,D}\mathbf{g}_j + \mathbf{w}_j = \mathbf{P}_{\phi_{j,D}}\tilde{\mathbf{Q}}\mathbf{g}_j + \mathbf{w}_j, \tag{14}$$

where

$$\tilde{\mathbf{S}} = \left[\mathrm{diag}\left(\mathbf{V}_{\theta_1}\mathbf{s}_1\right), \mathrm{diag}\left(\mathbf{V}_{\theta_2}\mathbf{s}_2\right), \cdots, \mathrm{diag}\left(\mathbf{V}_{\theta_{N_t}}\mathbf{s}_{N_t}\right)\right],$$
$$\mathbf{F}_{L,D} = \mathbf{diag}\Big(\underbrace{[\mathbf{F}_L, \cdots, \mathbf{F}_L]}_{N_t}\Big).$$

In the existence of AWGN, Γ_1 is given by [29]

$$\Gamma_1 = \frac{2}{\sigma_w^2}\Re\left[\frac{\partial\boldsymbol{\mu}_1^H}{\partial\boldsymbol{\eta}_1}\frac{\partial\boldsymbol{\mu}_1}{\partial\boldsymbol{\eta}_1^T}\right], \tag{15}$$

where $\boldsymbol{\mu}_1$ denotes $\mathbf{P}_{\phi_{j,D}}\mathbf{F}^H\mathbf{H}_j\mathbf{F}_{N_t,D}\mathbf{P}_{\theta,D}\mathbf{x}$. The partial derivative of μ with respect to each parameter can to be approximately calculated as

$$\frac{\partial\boldsymbol{\mu}_1}{\partial\boldsymbol{\phi}_j^T} \approx \frac{\partial\Lambda_1\left(\mathbf{1}_N + \boldsymbol{\phi}_j\right)}{\partial\boldsymbol{\phi}_j} = j\Lambda_1,$$
$$\frac{\partial\boldsymbol{\mu}_1}{\partial\boldsymbol{\theta}^T} \approx \frac{\partial\Xi_1\left(\mathbf{1}_{NN_t} + \boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}^T} = j\Xi_1,$$
$$\frac{\partial\boldsymbol{\mu}_1}{\partial\mathbf{g}_{j,re}^T} = \mathbf{P}_{\phi_j,D}\tilde{\mathbf{Q}},\ \frac{\partial\boldsymbol{\mu}_1}{\partial\mathbf{g}_{j,im}^T} = j\mathbf{P}_{\phi_j,D}\tilde{\mathbf{Q}}, \tag{16}$$

where phase noise is assumed to be approximated by $e^{\phi_j(n)} \approx 1 + j\phi_j(n)$ by Taylor series expansion of exponential function. Using (16), we can obtain Γ_1 as given in (17), as shown at the bottom of this page.

Then, with the phase noise model given in (7) and (8) and power delay profile of the channel, $\mathbf{c}_g \in \mathbb{R}^{L\times 1}$, we can obtain BIM given in (18), as shown at the bottom of this page, which is derived in the Appendix A in detail. In (18), $\mathbf{C}_{g,D}$,$\tilde{\mathbf{C}}_{g_{cir}}^{N_t}$ and $\mathbf{C}_x$ are respectively defined as

$$\mathbf{C}_{g,D} = \mathrm{diag}\Big(\underbrace{\left[\mathbf{c}_g^T, \cdots, \mathbf{c}_g^T\right]^T}_{N_t}\Big) \in \mathbb{R}^{LN_t\times LN_t},$$
$$\tilde{\mathbf{C}}_{g_{cir}}^{N_t} = \underbrace{\left[\tilde{\mathbf{C}}_{g,cir}, \cdots, \tilde{\mathbf{C}}_{g,cir}\right]}_{N_t} \in \mathbb{R}^{N\times NN_t},$$
$$\mathbf{C}_x = \underbrace{[\mathbf{c}_x, \cdots, \mathbf{c}_x]^T}_{N} \in \mathbb{R}^{N\times NN_t}, \tag{19}$$

where $\tilde{\mathbf{c}}_g = \left[\mathbf{c}_g^T, \mathbf{0}_{N-L}^T\right]^T \in \mathbb{R}^{N\times 1}$ is a zero-pedded power delay profile, $\tilde{\mathbf{C}}_{g_{cir}} = \mathrm{circ}\left(\tilde{\mathbf{c}}_g\right) \in \mathbb{R}^{N\times N}$ denotes the circulant matrix of $\tilde{\mathbf{c}}_g$, and $\mathbf{c}_x = \mathrm{diag}\left(\mathbf{I}\circ\left(\mathbf{x}\mathbf{x}^H\right)\right) \in \mathbb{R}^{NN_t\times 1}$ represents the power vector of the transmitted signal. In (18), $\mathbf{E}_\eta\left[\tilde{\mathbf{Q}}^H\tilde{\mathbf{Q}}\right]$ can be derived as

$$\mathbf{E}_\eta\left[\tilde{\mathbf{Q}}^H\tilde{\mathbf{Q}}\right] = \mathbf{F}_{L,D}^H\mathbf{E}_\eta\left[\tilde{\mathbf{S}}^H\tilde{\mathbf{S}}\right]\mathbf{F}_{L,D}, \tag{20}$$

$$\Gamma_1 = \frac{1}{\sigma_w^2}\Re\begin{bmatrix} \Lambda_1^H\Lambda_1 & \Lambda_1^H\Xi_1 & \Lambda_1^H\mathbf{P}_{\phi_j,D}\tilde{\mathbf{Q}} & j\Lambda_1^H\mathbf{P}_{\phi_j,D}\tilde{\mathbf{Q}} \\ \Xi_1^H\Lambda_1 & \Xi_1^H\Xi_1 & \Xi_1^H\mathbf{P}_{\phi_j,D}\tilde{\mathbf{Q}} & j\Xi_1^H\mathbf{P}_{\phi_j,D}\tilde{\mathbf{Q}} \\ \tilde{\mathbf{Q}}^H\mathbf{P}_{\phi_j,D}^H\Lambda_1 & \tilde{\mathbf{Q}}^H\mathbf{P}_{\phi_j,D}^H\Xi_1 & \tilde{\mathbf{Q}}^H\tilde{\mathbf{Q}} & j\tilde{\mathbf{Q}}^H\tilde{\mathbf{Q}} \\ -j\tilde{\mathbf{Q}}^H\mathbf{P}_{\phi_j,D}^H\Lambda_1 & -j\tilde{\mathbf{Q}}^H\mathbf{P}_{\phi_j,D}^H\Xi_1 & -j\tilde{\mathbf{Q}}^H\tilde{\mathbf{Q}} & \tilde{\mathbf{Q}}^H\tilde{\mathbf{Q}} \end{bmatrix} \tag{17}$$

$$\mathbf{B}_1 = \frac{1}{\sigma_w^2}\Re\begin{bmatrix} \mathrm{diag}\left\{\tilde{\mathbf{C}}_{g_{cir}}^{N_t}\mathbf{c}_x + \tilde{\mathbf{C}}_{g_{cir}}^{N_t}(\mathbf{I}\circ\boldsymbol{\Theta})\mathbf{c}_x\right\} + \boldsymbol{\Phi}_j & \tilde{\mathbf{C}}_{g_{cir}}^{N_t}\circ\mathbf{C}_x & \mathbf{0}_{N\times LN_t} & \mathbf{0}_{N\times LN_t} \\ \left(\tilde{\mathbf{C}}_{g_{cir}}^{N_t}\circ\mathbf{C}_x\right)^T & \frac{1}{N_t}\mathbf{X}_D^H\mathbf{X}_D + \boldsymbol{\Theta} & \mathbf{0}_{NN_t\times LN_t} & \mathbf{0}_{NN_t\times LN_t} \\ \mathbf{0}_{LN_t\times N} & \mathbf{0}_{LN_t\times NN_t} & \mathbf{E}_\eta\left\{\mathbf{Q}^H\mathbf{Q}\right\} + \frac{1}{2}\mathbf{C}_{g,D} & j\mathbf{E}_\eta\left\{\mathbf{Q}^H\mathbf{Q}\right\} \\ \mathbf{0}_{LN_t\times N} & \mathbf{0}_{LN_t\times NN_t} & -j\mathbf{E}_\eta\left\{\mathbf{Q}^H\mathbf{Q}\right\} & \mathbf{E}_\eta\left\{\mathbf{Q}^H\mathbf{Q}\right\} + \frac{1}{2}\mathbf{C}_{g,D} \end{bmatrix} \tag{18}$$

where

$$\mathbf{E}\left[\tilde{\mathbf{S}}^H\tilde{\mathbf{S}}\right] = \mathbf{E}\begin{bmatrix} \mathrm{diag}\left(\mathbf{V}^*_{\theta_1}\mathbf{s}^*_1\circ\mathbf{V}_{\theta_1}\mathbf{s}_1\right)\cdot\mathrm{diag}\left(\mathbf{V}^*_{\theta_1}\mathbf{s}^*_1\circ\mathbf{V}_{\theta_{N_t}}\mathbf{s}_{N_t}\right) & \cdots & \vdots \\ \vdots & \ddots & \\ \mathrm{diag}\left(\mathbf{V}^*_{\theta_{N_t}}\mathbf{s}^*_{N_t}\circ\mathbf{V}_{\theta_1}\mathbf{s}_1\right)\cdot\mathrm{diag}\left(\mathbf{V}^*_{\theta_{N_t}}\mathbf{s}^*_{N_t}\circ\mathbf{V}_{\theta_{N_t}}\mathbf{s}_{N_t}\right) & \cdots & \end{bmatrix}, \tag{21}$$

$$\mathbf{E}_{\boldsymbol{\eta}}\left[\mathrm{diag}\left(\mathbf{V}^*_i\mathbf{s}^*_i\circ\mathbf{V}_j\mathbf{s}_j\right)\right] \approx \begin{cases} \mathrm{diag}\left(\mathbf{s}^*_i\mathbf{s}_i\right)+\mathbf{I}\circ\left\{\mathbf{F}\mathrm{diag}\left(\mathbf{x}_i\right)\boldsymbol{\Theta}_j\mathrm{diag}\left(\mathbf{x}^*_i\right)\mathbf{F}^H\right\}, & \text{if } i=j \\ \mathrm{diag}\left(\mathbf{s}^*_i\mathbf{s}_j\right), & \text{if } i\neq j \end{cases}. \tag{22}$$

Finally, by calculating the inverse matrix of BIM matrix, BCRLB for $\eta_1(k)$ can be obtained as

$$\mathrm{BCRLB}\left(\eta_1(k)\right) = \left(\mathbf{B_1}^{-1}\right)_{k,k} \tag{23}$$

Hence, utilizing explicit statistical knowledge of channel and phase noise, we can derive the BCRLB of unknown parameters.

B. PROPOSED CHANNEL ESTIMATION ALGORITHM

It is difficult to jointly estimate the phase noise and the channel impulse response in the channel estimation stage. This is because there are more parameters to jointly be estimated than the number of received signals at the j-th RX end. Hence, the solution involves an analysis of ICI caused by phase noise and incorporating its result in channel estimator. The mathematical analysis of ICI will be presented in detail in Section III-C.

In (3), the effect of TX phase noise can be expressed as

$$\begin{aligned} \mathbf{y}_j &= \mathbf{P}_{\phi_j,D}\mathbf{F}^H\sum_{i=1}^{N_t}\mathbf{H}_{ji,D}\mathbf{F}\mathbf{P}_{\theta_i,D}\mathbf{F}^H\mathbf{s}_i+\mathbf{w}_j \\ &= \mathbf{P}_{\phi_j,D}\mathbf{F}^H\sum_{i=1}^{N_t}\mathbf{H}_{ji,D}\mathbf{V}_{\theta_i}\mathbf{s}_i+\mathbf{w}_j \\ &= \mathbf{P}_{\phi_j,D}\mathbf{F}^H\sum_{i=1}^{N_t}\mathbf{H}_{ji,D}\left(\alpha_{\theta_i}\mathbf{I}+\tilde{\mathbf{V}}_{\theta_i}\right)\mathbf{s}_i+\mathbf{w}_j \\ &= \mathbf{P}_{\phi_j,D}\mathbf{F}^H\sum_{i=1}^{N_t}\alpha_{\theta_i}\mathbf{H}_{ji,D}\mathbf{s}_i \\ &\quad+\mathbf{P}_{\phi_j,D}\mathbf{F}^H\sum_{i=1}^{N_t}\mathbf{H}_{ji,D}\tilde{\mathbf{V}}_{\theta_i}\mathbf{s}_i+\mathbf{w}_j, \end{aligned} \tag{24}$$

where α_{θ_i} is the CPE of θ_i, $\mathbf{v}_{\theta_i}[0]$, and $\tilde{\mathbf{V}}_{\theta_i}$ denotes the matrix that the diagonal element is removed at $\mathbf{V}_{\theta_i}$. The second term of (24) is the ICI which is caused by the TX phase noise. Also, the effect of RX phase noise can be written as

$$\begin{aligned} \mathbf{y}_j &= \alpha_{\phi_j}\mathbf{F}^H\sum_{i=1}^{N_t}\alpha_{\theta_i}\mathbf{H}_{ji,D}\mathbf{s}_i+\mathbf{F}^H\tilde{\mathbf{V}}_{\phi_j}\sum_{i=1}^{N_t}\alpha_{\theta_i}\mathbf{H}_{ji,D}\mathbf{s}_i \\ &\quad+\mathbf{P}_{\phi_j,D}\mathbf{F}^H\sum_{i=1}^{N_t}\mathbf{H}_{ji,D}\tilde{\mathbf{V}}_{\theta_i}\mathbf{s}_i+\mathbf{w}_j, \end{aligned} \tag{25}$$

where the second term is the ICI of RX phase noise. (25) can be simply rewritten as

$$\begin{aligned} \mathbf{y}_j &= \alpha_{\phi_j}\mathbf{F}^H\sum_{i=1}^{N_t}\alpha_{\theta_i}\mathbf{H}_{ji,D}\mathbf{s}_i+\boldsymbol{\epsilon}_{rx}+\boldsymbol{\epsilon}_{tx}+\mathbf{w}_j \\ &= \mathbf{F}^H\mathbf{S}\mathbf{F}_{L,D}\tilde{\mathbf{g}}_j+\boldsymbol{\epsilon}_{rx}+\boldsymbol{\epsilon}_{tx}+\mathbf{w}_j \\ &= \mathbf{Q}\tilde{\mathbf{g}}_j+\boldsymbol{\epsilon}_j+\mathbf{w}_j, \end{aligned} \tag{26}$$

where

$$\mathbf{S} = \left[\mathrm{diag}\left(\mathbf{s}_1\right),\mathrm{diag}\left(\mathbf{s}_s\right),\cdots,\mathrm{diag}\left(\mathbf{s}_{N_t}\right)\right]$$

and $\tilde{\mathbf{g}}_j$ is the corrupt CIR at the j-th RX end, which is rotated by α_{θ_i} and α_{ϕ_j}. Also, $\boldsymbol{\epsilon}_{tx}$ and $\boldsymbol{\epsilon}_{rx}$ represent ICI of TX and RX phase noise, and $\boldsymbol{\epsilon}_j$ is total ICI which occurs at the j-th RX end. Here we need to estimate $\tilde{\mathbf{g}}_j$ instead of $\mathbf{g}_j$, because we cannot disentangle CIR from the CPE and the rotation of the channel can be sufficiently compensated by the CPE estimation in data decoding stage, which is described in detail in Section IV-C.

In general, the channel is estimated by least square (LS) algorithm [12], [19], [24]–[26] as

$$\hat{\mathbf{g}}_{j,LS} = \left(\mathbf{Q}^H\mathbf{Q}\right)^{-1}\mathbf{Q}^H\mathbf{y}_j. \tag{27}$$

However, as the power of ICI caused by phase noise increases in proportional to the signal power, the performance of LS algorithm can be degraded, in particular, at high SNR.

Therefore, for more accurate estimation of the channel, we propose the channel estimation technique based on MAP estimation. From the fact that ICI of TX and RX phase noise consist of the summation of the CFRs and data subcarriers in a OFDM symbol, we assume that ICIs follow the Gaussian distribution by the central limit theorem. Let $\tilde{\mathbf{w}}_j$ denote $\mathbf{w}_j+\boldsymbol{\epsilon}_j$, and we can introduce from the assumption that $\tilde{\mathbf{w}}_j$ follows Gaussian distribution $\mathcal{N}\left(\mathbf{0},\sigma_w^2\mathbf{I}+\mathbf{C}_\epsilon\right)$, where $\mathbf{C}_\epsilon$ is the covariance matrix of $\boldsymbol{\epsilon}_j$ and is derived from mathematical analysis in Section III-C.

The log likelihood function (LLF) of the pdf $p\left(\mathbf{y}_j,\tilde{\mathbf{g}}_j\right)$ is given as

$$\begin{aligned} L\left(\tilde{\mathbf{g}}_j\right) &= -\ln p\left(\mathbf{y}_j\mid\tilde{\mathbf{g}}_j\right)-\ln p\left(\tilde{\mathbf{g}}_j\right) \\ &= \left(\mathbf{y}_j-\mathbf{Q}\tilde{\mathbf{g}}_j\right)^H\mathbf{C}^{-1}_{\tilde{w}}\left(\mathbf{y}_j-\mathbf{Q}\tilde{\mathbf{g}}_j\right)+\tilde{\mathbf{g}}_j^H\mathbf{C}_{\mathbf{g}}^{-1}\tilde{\mathbf{g}}_j, \end{aligned} \tag{28}$$

where $\mathbf{C}_{\tilde{w}} = \sigma_w^2\mathbf{I}+\mathbf{C}_\epsilon$. Then, MAP estimation of the channel can be found by maximizing (28), which is equal to taking the partial derivative of (28) with respect to $\tilde{\mathbf{g}}_j$ and setting it to zero, as

$$\hat{\mathbf{g}}_{j,MAP} = \left(\mathbf{Q}^H\mathbf{C}^{-1}_{\tilde{w}}\mathbf{Q}+\mathbf{C}_{\mathbf{g}}^{-1}\right)^{-1}\mathbf{Q}^H\mathbf{C}^{-1}_{\tilde{w}}\mathbf{y}_j. \tag{29}$$

C. MATHEMATICAL ANALYSIS

In this section, the covariance matrix of ICI and MSE of proposed algorithms is mathematically derived. The covariance of ICI caused by phase noise is analyzed for proposing the MAP estimation in (29). Also, mathematical analysis of MSE can be used to select the combination of the training symbols, which minimize the MSE performances in a certain group of sequences.

1) COVARIANCE MATRIX OF ICI

The ICI to be incurred by phase noise is composed of $\boldsymbol{\epsilon}_{tx}$ and $\boldsymbol{\epsilon}_{rx}$. As $\boldsymbol{\epsilon}_{tx}$ and $\boldsymbol{\epsilon}_{rx}$ occur by TX and RX phase noise, respectively, the cross-correlation can be assumed to be zero. Hence, we only need to derive covariance matrix of $\boldsymbol{\epsilon}_{tx}$ and $\boldsymbol{\epsilon}_{rx}$, respectively, and the summation of them is the covariance of total ICI, $\mathbf{C}_{\epsilon}$. Let us define the covariance of $\boldsymbol{\epsilon}_{tx}$ as

$$\mathbf{C}_{\epsilon_{tx}} = \mathbf{E}\left[\boldsymbol{\epsilon}_{tx}\boldsymbol{\epsilon}_{tx}^{H}\right]. \tag{30}$$

In (30), substituting $\boldsymbol{\epsilon}_{tx}$ with the second term of (24), $\mathbf{C}_{\epsilon_{tx}}$ is given by

$$\begin{aligned}
&\mathbf{E}\left[\boldsymbol{\epsilon}_{tx}\boldsymbol{\epsilon}_{tx}^{H}\right] \\
&= \sum_{l=1}^{N_t}\sum_{m=1}^{N_t} \mathbf{E}\left[\mathbf{P}_{\phi_j,D}\mathbf{F}^{H}\mathbf{H}_{jl,D}\tilde{\mathbf{V}}_{\theta_l}\mathbf{s}_l\mathbf{s}_m^{H}\tilde{\mathbf{V}}_{\theta_m}^{H}\mathbf{H}_{jm,D}^{H}\mathbf{F}\mathbf{P}_{\phi_j,D}^{H}\right] \\
&= \sum_{l=1}^{N_t} \mathbf{E}\left[\mathbf{P}_{\phi_j,D}\mathbf{F}^{H}\mathbf{H}_{jl,D}\tilde{\mathbf{V}}_{\theta_l}\mathbf{s}_l\mathbf{s}_l^{H}\tilde{\mathbf{V}}_{\theta_l}^{H}\mathbf{H}_{jl,D}^{H}\mathbf{F}\mathbf{P}_{\phi_j,D}^{H}\right] \\
&= \sum_{l=1}^{N_t} \mathbf{E}\Big[\mathbf{P}_{\phi_j,D}\mathbf{F}^{H}\mathbf{H}_{jl,D}\mathbf{F}\tilde{\mathbf{P}}_{\theta_l,D}\mathbf{x}_l\mathbf{x}_l^{H}\tilde{\mathbf{P}}_{\theta_l,D}^{H}\times\mathbf{F}^{H}\mathbf{H}_{jl,D}^{H}\mathbf{F}\mathbf{P}_{\phi_j,D}^{H}\Big] \\
&= \sum_{l=1}^{N_t} \mathbf{E}_{\{\boldsymbol{\phi_j},\mathbf{h}_{jl}|\boldsymbol{\theta}\}}\Big[\mathbf{P}_{\phi_j,D}\mathbf{F}^{H}\mathbf{H}_{jl,D}\mathbf{F}\left(\mathbf{x}_l\mathbf{x}_l^{H}\circ\mathbf{E}_{\{\boldsymbol{\theta}_l\}}\left[\tilde{\mathbf{p}}_{\theta_l}\tilde{\mathbf{p}}_{\theta_l}^{H}\right]\right) \\
&\qquad\times\mathbf{F}^{H}\mathbf{H}_{jl,D}^{H}\mathbf{F}\mathbf{P}_{\phi_j,D}^{H}\Big],
\end{aligned} \tag{31}$$

where $\tilde{\mathbf{p}}_{\theta_l}$ denotes the phase error vector, that is, the mean value subtracted from $\mathbf{p}_{\theta_l}$, and $\tilde{\mathbf{P}}_{\theta_l,D}$ denotes diag$\left(\tilde{\mathbf{p}}_{\theta_l}\right)$. In (31), $\mathbf{E}\left[\tilde{\mathbf{p}}_{\theta_l}\tilde{\mathbf{p}}_{\theta_l}^{H}\right]$, which is derived in the Appendix B using the prior statistical information of phase noise, can be approximated as

$$\begin{aligned}
&\mathbf{E}\left[\tilde{\mathbf{p}}_{\theta_l}\tilde{\mathbf{p}}_{\theta_l}^{H}\right]_{p,q} \\
&\approx \frac{\sigma_{\zeta}^{2}}{6N}\left[3b^{2}-3\left(2N+1\right)b+3a^{2}-3a+2N^{2}+3N+1\right],
\end{aligned} \tag{32}$$

where a and b are $\min\left(p,q\right)$ and $\max\left(p,q\right)$, respectively. By substituting $\mathbf{D}$ for $\mathbf{F}\left(\mathbf{x}_l\mathbf{x}_l^{H}\circ\mathbf{E}\left[\tilde{\mathbf{p}}_{\theta_l}\tilde{\mathbf{p}}_{\theta_l}^{H}\right]\right)\mathbf{F}^{H}$, (31) can be rewritten as

$$\begin{aligned}
&\mathbf{E}\left[\boldsymbol{\epsilon}_{tx}\boldsymbol{\epsilon}_{tx}^{H}\right] \\
&= \sum_{l=1}^{N_t}\mathbf{E}_{\{\boldsymbol{\phi}_j,\mathbf{h}_{jl}|\boldsymbol{\theta}\}}\left[\mathbf{P}_{\phi_j,D}\mathbf{F}^{H}\mathbf{H}_{ji,D}\mathbf{D}\mathbf{H}_{ji,D}^{H}\mathbf{F}\mathbf{P}_{\phi_j,D}^{H}\right] \\
&= \sum_{l=1}^{N_t}\mathbf{E}_{\{\boldsymbol{\phi}_j|\mathbf{h}_{jl}\}}\left[\mathbf{P}_{\phi_j,D}\mathbf{F}^{H}\left(\mathbf{D}\circ\mathbf{E}_{\{\mathbf{h}_{jl}\}}\left[\mathbf{h}_{jl}\mathbf{h}_{jl}^{H}\right]\right)\mathbf{F}\mathbf{P}_{\phi_j,D}^{H}\right] \\
&= \sum_{l=1}^{N_t}\mathbf{E}_{\{\boldsymbol{\phi}_j|\mathbf{h}_{jl}\}}\left[\mathbf{P}_{\phi_j,D}\mathbf{D}_1\mathbf{P}_{\phi_j,D}^{H}\right],
\end{aligned} \tag{33}$$

where

$$\mathbf{D}_1 = \mathbf{F}^{H}\left(\mathbf{D}\circ\mathbf{E}_{\{\mathbf{h}_{jl}\}}\left[\mathbf{h}_{jl}\mathbf{h}_{jl}^{H}\right]\right)\mathbf{F}.$$

Assuming that $\mathbf{p}_{\phi_j}\approx\mathbf{1}+j\boldsymbol{\phi}_j$, (33) can be determined as

$$\mathbf{E}\left[\boldsymbol{\epsilon}_{tx}\boldsymbol{\epsilon}_{tx}^{H}\right] \approx \sum_{l=1}^{N_t}\mathbf{D}_1\circ\left(\mathbf{1}_{N\times N}+\boldsymbol{\Phi}_j\right). \tag{34}$$

The covariance matrix of $\boldsymbol{\epsilon}_{rx}$, $\mathbf{C}_{\epsilon_{rx}}$, can be calculated in the similar process of above equations and can be expressed by

$$\begin{aligned}
&\mathbf{E}\left[\boldsymbol{\epsilon}_{rx}\boldsymbol{\epsilon}_{rx}^{H}\right] \\
&= \sum_{l=1}^{N_t}\sum_{m=1}^{N_t}\mathbf{E}\left[\mathbf{F}^{H}\tilde{\mathbf{V}}_{\phi_j}\mathbf{H}_{jl,D}\alpha_{\theta_l}\mathbf{s}_l\mathbf{s}_m^{H}\alpha_{\theta_m}^{*}\mathbf{H}_{jm,D}^{H}\tilde{\mathbf{V}}_{\phi_j}^{H}\mathbf{F}\right] \\
&= \sum_{l=1}^{N_t}\mathbf{E}\left[\mathbf{F}^{H}\tilde{\mathbf{V}}_{\phi_j}\mathbf{H}_{jl,D}\mathbf{s}_l\mathbf{s}_l^{H}\mathbf{H}_{jl,D}^{H}\tilde{\mathbf{V}}_{\phi_j}^{H}\mathbf{F}\right] \\
&= \sum_{l=1}^{N_t}\mathbf{E}\left[\tilde{\mathbf{P}}_{\phi_j,D}\mathbf{F}^{H}\mathbf{H}_{jl,D}\mathbf{s}_l\mathbf{s}_l^{H}\mathbf{H}_{jl,D}^{H}\mathbf{F}\tilde{\mathbf{P}}_{\phi_j,D}^{H}\right] \\
&= \sum_{l=1}^{N_t}\mathbf{E}_{\{\boldsymbol{\phi}_j|\mathbf{h}_{jl}\}}\Big[\tilde{\mathbf{P}}_{\phi_j,D}\mathbf{F}^{H}\left(\mathbf{s}_l\mathbf{s}_l^{H}\circ\mathbf{E}_{\{\mathbf{h}_{jl}\}}\left[\mathbf{h}_{jl}\mathbf{h}_{jl}^{H}\right]\right) \\
&\qquad\times\mathbf{F}\tilde{\mathbf{P}}_{\phi_j,D}^{H}\Big] \\
&= \sum_{l=1}^{N_t}\mathbf{D}_2\circ\mathbf{E}_{\{\boldsymbol{\phi}_j\}}\left[\tilde{\mathbf{p}}_{\phi_j}\tilde{\mathbf{p}}_{\phi_j}^{H}\right],
\end{aligned} \tag{35}$$

where

$$\mathbf{D}_2 = \mathbf{F}^{H}\left(\mathbf{s}_l\mathbf{s}_l^{H}\circ\mathbf{E}\left[\mathbf{h}_{jl}\mathbf{h}_{jl}^{H}\right]\right)\mathbf{F}.$$

As we assume that the statistical information of $\boldsymbol{\phi}_j$ is equal to that of $\boldsymbol{\theta}_i$, $\mathbf{E}_{\{\boldsymbol{\phi}_j\}}\left[\tilde{\mathbf{p}}_{\phi_j}\tilde{\mathbf{p}}_{\phi_j}^{H}\right]$ can be approximately expressed as (32) likewise. Therefore, we can calculate the covariance of ICI, $\mathbf{C}_{\epsilon}$, using the statistical information of channel and phase noise. With (33) and (35), we can estimate the channel using the proposed algorithm in (29).

2) MSE OF PROPOSED ALGORITHM

In this section, we derive MSE of estimated CIR mathematically. When using LS algorithm given in (27), MSE can be easily calculated as

$$\begin{aligned} MSE_{LS} &= \mathbf{E}\left[\left(\hat{\mathbf{g}}_{j,LS}-\tilde{\mathbf{g}}_j\right)^H\left(\hat{\mathbf{g}}_{j,LS}-\tilde{\mathbf{g}}_j\right)\right] \\ &= \mathbf{E}\left[\tilde{\mathbf{w}}_j^H\mathbf{Q}\left(\mathbf{Q}^H\mathbf{Q}\right)^{-1}\left(\mathbf{Q}^H\mathbf{Q}\right)^{-1}\mathbf{Q}^H\tilde{\mathbf{w}}_j\right] \\ &= \mathbf{E}\left[\mathbf{tr}\left\{\left(\mathbf{Q}^H\mathbf{Q}\right)^{-1}\mathbf{Q}^H\tilde{\mathbf{w}}_j\tilde{\mathbf{w}}_j^H\mathbf{Q}\left(\mathbf{Q}^H\mathbf{Q}\right)^{-1}\right\}\right] \\ &= \mathbf{tr}\left\{\left(\mathbf{Q}^H\mathbf{Q}\right)^{-1}\mathbf{Q}^H\mathbf{C}_{\tilde{w}}\mathbf{Q}\left(\mathbf{Q}^H\mathbf{Q}\right)^{-1}\right\}. \end{aligned} \tag{36}$$

In (36), $\mathbf{C}_{\tilde{w}}$ have already been derived in (33) and (35), and $\mathbf{Q}$ can be found out by using training symbols.

Also, MSE of the channel using proposed algorithm can be derived through similar procedure of (36). Before deriving the MSE of estimated CIR using proposed algorithm, we can rewritten (29) as

$$\begin{aligned} \hat{\mathbf{g}}_{j,MAP} &= \left(\mathbf{Q}^H\mathbf{C}_{\tilde{w}}^{-1}\mathbf{Q}+\mathbf{C}_{g,D}^{-1}\right)^{-1}\mathbf{Q}^H\mathbf{C}_{g,D}^{-1}\mathbf{Q}\tilde{\mathbf{g}}_j \\ &\quad+\left(\mathbf{Q}^H\mathbf{C}_{\tilde{w}}^{-1}\mathbf{Q}+\mathbf{C}_{g,D}^{-1}\right)^{-1}\mathbf{Q}^H\mathbf{C}_{g,D}^{-1}\tilde{\mathbf{w}}_j. \end{aligned} \tag{37}$$

By the matrix inverse lemma [30], (37) can be expressed by

$$\begin{aligned} &\hat{\mathbf{g}}_{j,MAP} \\ &= \left\{\mathbf{I}-\left(\mathbf{Q}^H\mathbf{C}_{\tilde{w}}^{-1}\mathbf{Q}\right)^{-1}\left(\mathbf{C}_{g,D}+\left(\mathbf{Q}^H\mathbf{C}_{\tilde{w}}^{-1}\mathbf{Q}\right)^{-1}\right)\right\}\tilde{\mathbf{g}}_j \\ &\quad+\left(\mathbf{Q}^H\mathbf{C}_{\tilde{w}}^{-1}\mathbf{Q}+\mathbf{C}_{g,D}^{-1}\right)^{-1}\mathbf{Q}^H\mathbf{C}_{g,D}^{-1}\tilde{\mathbf{w}}_j \\ &= \tilde{\mathbf{g}}_j-\mathbf{T}_1\tilde{\mathbf{g}}_j+\mathbf{T}_2\tilde{\mathbf{w}}_j, \end{aligned} \tag{38}$$

where

$$\mathbf{T}_1=\left(\mathbf{Q}^H\mathbf{C}_{\tilde{w}}^{-1}\mathbf{Q}\right)^{-1}\left(\mathbf{C}_{g,D}+\left(\mathbf{Q}^H\mathbf{C}_{\tilde{w}}^{-1}\mathbf{Q}\right)^{-1}\right)^{-1}$$

$$\mathbf{T}_2=\left(\mathbf{Q}^H\mathbf{C}_{\tilde{w}}^{-1}\mathbf{Q}+\mathbf{C}_{g,D}^{-1}\right)^{-1}\mathbf{Q}^H\mathbf{C}_{g,D}^{-1}.$$

Using (38), MSE of estimated CIR based on MAP algorithm is given by

$$\begin{aligned} MSE_{MAP} &= \mathbf{E}\left[\left(\hat{\mathbf{g}}_{j,MAP}-\tilde{\mathbf{g}}_j\right)^H\left(\hat{\mathbf{g}}_{j,MAP}-\tilde{\mathbf{g}}_j\right)\right] \\ &= \mathbf{E}\left[\left(\mathbf{T}_1\tilde{\mathbf{g}}_j-\mathbf{T}_2\tilde{\mathbf{w}}_j\right)^H\left(\mathbf{T}_1\tilde{\mathbf{g}}_j-\mathbf{T}_2\tilde{\mathbf{w}}_j\right)\right] \\ &= \mathbf{E}\left[\mathbf{tr}\left\{\left(\mathbf{T}_1\tilde{\mathbf{g}}_j-\mathbf{T}_2\tilde{\mathbf{w}}_j\right)\left(\mathbf{T}_1\tilde{\mathbf{g}}_j-\mathbf{T}_2\tilde{\mathbf{w}}_j\right)\right\}^H\right] \\ &= \mathbf{tr}\left\{\mathbf{T}_1\mathbf{C}_{g,D}\mathbf{T}_1^H\right\}+\mathbf{tr}\left\{\mathbf{T}_2\mathbf{C}_w\mathbf{T}_2^H\right\} \\ &\quad-2\Re\left\{\mathbf{tr}\left\{\mathbf{T}_2\mathbf{E}\left[\tilde{\mathbf{w}}_j\tilde{\mathbf{g}}_j^H\right]\mathbf{T}_1^H\right\}\right\}. \end{aligned} \tag{39}$$

Note that in (39), $\mathbf{E}\left[\tilde{\mathbf{w}}_j\tilde{\mathbf{g}}_j^H\right]$ is not calculated appeared in the third term, which can be represented by

$$\begin{aligned} \mathbf{E}\left[\tilde{\mathbf{w}}_j\tilde{\mathbf{g}}_j^H\right] &= \mathbf{E}\left[\boldsymbol{\epsilon}_{rx}\tilde{\mathbf{g}}_j^H\right]+\mathbf{E}\left[\boldsymbol{\epsilon}_{rx}\tilde{\mathbf{g}}_j^H\right] \\ &= \mathbf{E}\left[\tilde{\mathbf{P}}_{\phi_j,D}\mathbf{F}^H\sum_{i=1}^{N_t}\alpha_{\theta_i}\mathbf{H}_{ji,D}\mathbf{s}_i\tilde{\mathbf{g}}_j^H\right] \\ &\quad+\mathbf{E}\left[\mathbf{P}_{\phi_j,D}\mathbf{F}^H\sum_{i=1}^{N_t}\mathbf{H}_{ji,D}\mathbf{F}\tilde{\mathbf{P}}_{\theta_i,D}\mathbf{x}_i\tilde{\mathbf{g}}_j^H\right] \\ &= \mathbf{E}\left[\alpha_{\phi_j}^*\tilde{\mathbf{P}}_{\phi_j,D}\mathbf{F}^H\sum_{i=1}^{N_t}\mathbf{H}_{ji,D}\mathbf{s}_i\mathbf{g}_j^H\right] \\ &\quad+\mathbf{E}\left[\alpha_{\phi_j}^*\mathbf{P}_{\phi_j,D}\mathbf{F}^H\sum_{i=1}^{N_t}\mathbf{H}_{ji,D}\mathbf{F}\alpha_{\theta_j}^*\tilde{\mathbf{P}}_{\theta_i,D}\mathbf{x}_i\mathbf{g}_j^H\right]. \end{aligned} \tag{40}$$

Assuming that $\alpha_{\phi_j}\approx\exp\left(j\sum_{n=1}^{N}\phi_j(n)\right)$, $\mathbf{E}\left[\alpha_{\phi_j}^*\tilde{\mathbf{P}}_{\phi_j,D}\right]$ can be calculated as

$$\begin{aligned} \left(\mathbf{E}\left[\alpha_{\phi_j}^*\tilde{\mathbf{P}}_{\phi_j,D}\right]\right)_{i,i} &= \mathbf{E}\left[\left(e^{j\boldsymbol{\phi}_j(i)}-\alpha_{\phi_j}\right)\alpha_{\phi_j}^*\right] \\ &\approx \mathbf{E}\left[e^{j\left(\boldsymbol{\phi}_j(i)-\sum_{n=1}^{N}\boldsymbol{\phi}_j(n)\right)}-1\right] \\ &\approx 0. \end{aligned} \tag{41}$$

Hence, we can get rid of the third term of (39) in calculating the MSE of proposed algorithm, and (39) is approximated as

$$\begin{aligned} MSE_{MAP} &= \mathbf{E}\left[\left(\hat{\mathbf{g}}_{j,MAP}-\tilde{\mathbf{g}}_j\right)^H\left(\hat{\mathbf{g}}_{j,MAP}-\tilde{\mathbf{g}}_j\right)\right] \\ &\approx \mathbf{tr}\left\{\mathbf{T}_1\mathbf{C}_{g,D}\mathbf{T}_1^H\right\}+\mathbf{tr}\left\{\mathbf{T}_2\mathbf{C}_{\tilde{w}}\mathbf{T}_2^H\right\}. \end{aligned} \tag{42}$$

Through the simulation results in Section V, it will be proved whether the assumption is valid and the mathematical analysis can be close to real value.

D. SELECTION OF TRAINING SEQUENCE

In order to reduce the effect of phase noise in channel estimation, it is important to design the training sequences to be robust against phase noise. In Section III-C, we have analyzed the MSE of CIR when using LS algorithm and proposed algorithm. If we can solve an optimization problem, where the solution is equal to minimizing mathematical MSE of channel, we can find out the optimal training sequence design with regard to MSE. However, solving the optimization problem is so difficult that we propose an alternative method.

As the design of training sequences proceeds offline, the complexity of designing algorithm is not considered important. Hence, we propose the exhaustive searching in certain group of sequences considering a low MSE as the criterion. In this paper, Hadamard sequences and frequency orthogonal sequences are considered as the training sequence.

IV. PHASE NOISE MITIGATION

From now on, we introduce the schemes mitigating phase noise in data decoding. The pilot subcarriers of the comb-type pilot symbols are used to jointly estimate TX and RX phase noise. Generally, the remaining synchronization errors on the payload can be processed after channel equalization. However, to estimate phase noise easily in our paper, channel equalization is processed after estimating phase noise like existing literatures, which researched phase noise problem, such as [19], [21], and [22].

A. BCRLB OF TX AND RX PHASE NOISE

In this section, we introduce the BIM and BCRLB for joint estimation of TX and RX phase noises, which is derived in [31], assuming the perfect estimation of CIR. Let us define the parameter vector $\boldsymbol{\eta}$ to be estimated as

$$\boldsymbol{\eta} \triangleq \left[\boldsymbol{\phi}^T, \boldsymbol{\theta}^T\right]^T. \tag{43}$$

BIM of $\boldsymbol{\eta}$ is defined as

$$\mathbf{B} = \mathbf{E}_{\boldsymbol{\eta}}\left[\Gamma\right] + \mathbf{E}_{\boldsymbol{\eta}}\left[-\Delta_{\boldsymbol{\eta}}^{\boldsymbol{\eta}} \ln p\left(\boldsymbol{\eta}\right)\right], \tag{44}$$

where

$$\Gamma = \mathbf{E}_{\mathbf{y}|\boldsymbol{\eta}}\left[-\Delta_{\boldsymbol{\eta}}^{\boldsymbol{\eta}} \ln p\left(\mathbf{y} \mid \boldsymbol{\eta}\right)\right] \tag{45}$$

is FIM of $\boldsymbol{\eta}$. From (4), the received vector can be rewritten as

$$\mathbf{y} = \Xi \mathbf{p}_\theta + \mathbf{w} = \Lambda \mathbf{p}_\phi + \mathbf{w}, \tag{46}$$

where

$$\Upsilon = \mathbf{F}_{N_r,D}^H \mathbf{H} \mathbf{F}_{N_t,D} \mathbf{X}_D,$$
$$\Xi = \mathbf{P}_{\phi,D} \Upsilon, \ \Lambda = \mathrm{diag}\left(\Upsilon \mathbf{p}_\theta\right).$$

Note that $\boldsymbol{\mu}$ denotes $\mathbf{P}_{\phi,D}\mathbf{F}_{N_r,D}^H \mathbf{H} \mathbf{F}_{N_t,D} \mathbf{P}_{\theta,D} \mathbf{x}$. Using (15), we can express Γ as

$$\Gamma = \frac{2}{\sigma_w^2} \Re \left[\begin{array}{cc} \frac{\partial \boldsymbol{\mu}^H}{\partial \boldsymbol{\phi}} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\phi}^T} & \frac{\partial \boldsymbol{\mu}^H}{\partial \boldsymbol{\phi}} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}^T} \\ \frac{\partial \boldsymbol{\mu}^H}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\phi}^T} & \frac{\partial \boldsymbol{\mu}^H}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}^T}. \end{array} \right],$$

where

$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\phi}^T} \approx \frac{\partial \Lambda \left(\mathbf{1}_{NN_t} + j\boldsymbol{\phi}\right)}{\partial \boldsymbol{\phi}^T} = j\Lambda \tag{47}$$

$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}^T} \approx \frac{\partial \Xi \left(\mathbf{1}_{NN_t} + j\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}^T} = j\Xi. \tag{48}$$

Hence, we can derive FIM given by (49), as shown at the bottom of this page. As real operation can be changed to expectation, the expectation of Γ is given by

$$\begin{aligned}
&\mathbf{E}_\eta\left[\frac{\partial \boldsymbol{\mu}^H}{\partial \boldsymbol{\phi}} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\phi}^T}\right] = \mathbf{E}_\eta\left[\Lambda^H \Lambda\right] \\
&\approx \mathbf{E}_\eta\left[\mathrm{diag}\left(\Upsilon^*\left(\mathbf{1}_{N_tN} - j\boldsymbol{\theta}\right)\right) \mathrm{diag}\left(\Upsilon\left(\mathbf{1}_{N_tN} + j\boldsymbol{\theta}\right)\right)\right] \\
&= \mathbf{E}_\eta\left[\mathrm{diag}\left(\left(\Upsilon^* \mathbf{1}_{N_tN} - j\Upsilon^*\boldsymbol{\theta}\right) \circ \left(\Upsilon \mathbf{1}_{N_tN} + j\Upsilon\boldsymbol{\theta}\right)\right)\right] \\
&= \mathrm{diag}\left(\left(\Upsilon^* \mathbf{1}_{N_tN}\right) \circ \left(\Upsilon^* \mathbf{1}_{N_tN}\right)\right) \\
&\quad + \mathbf{E}_\eta\left[\mathrm{diag}\left(\left(\Upsilon^*\boldsymbol{\theta}\right) \circ \left(\Upsilon^*\boldsymbol{\theta}\right)\right)\right] \\
&= \mathrm{diag}\left(\left(\Upsilon^* \mathbf{1}_{N_tN}\right) \circ \left(\Upsilon^* \mathbf{1}_{N_tN}\right)\right) + \mathrm{diag}\left(\mathrm{diag}\left(\Upsilon \Theta \Upsilon^H\right)\right) \\
&= \mathrm{diag}\left(\left(\Upsilon^* \mathbf{1}_{N_tN}\right) \circ \left(\Upsilon^* \mathbf{1}_{N_tN}\right)\right) + \mathbf{I}_{N_tN} \circ \left(\Upsilon \Theta \Upsilon^{\mathbf{H}}\right)
\end{aligned} \tag{50}$$

and

$$\begin{aligned}
\mathbf{E}_\eta\left[\frac{\partial \boldsymbol{\mu}^H}{\partial \boldsymbol{\phi}} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}^T}\right] &\approx \mathbf{E}_\eta\left[\mathrm{diag}\left(\Upsilon^*\left(\mathbf{1}_{N_tN} - j\boldsymbol{\theta}\right)\right)\left(\mathbf{I} + j\mathrm{diag}\left(\boldsymbol{\phi}\right)\right)\Upsilon\right] \\
&= \mathrm{diag}\left(\Upsilon^* \mathbf{1}_{N_tN}\right)\Upsilon, \\
\mathbf{E}_\eta\left[\frac{\partial \boldsymbol{\mu}^H}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\phi}^T}\right] &\approx \Upsilon^H \mathrm{diag}\left(\Upsilon \mathbf{1}_{N_tN}\right), \\
\mathbf{E}_\eta\left[\frac{\partial \boldsymbol{\mu}^H}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}^T}\right] &= \mathbf{X}_D^H \mathbf{F}_{N_t,D}^H \mathbf{H}^H \mathbf{H} \mathbf{F}_{N_r,D} \mathbf{X}_D.
\end{aligned} \tag{51}$$

Using (50), (51) and the second order statistic of phase noise, we can rewrite (44) as (52) shown in the bottom of this page. Then, by inverse matrix of (52), BCRLB for $\boldsymbol{\eta}$ can be calculated as (23).

B. MATHEMATICAL ANALYSIS OF CHANNEL ERROR

Generally, channel estimation error is often assumed to be slight enough not to be considered in data decoding stage. However, as the influence of the channel error increases under the presence of phase noise, neglecting the channel error will lead a significant loss in data decoding. The solution proposed in our paper is to analyze the power of the error through mathematical derivations and incorporate the result in the phase noise estimator.

In order to analyze the effect of channel error, (24) can be transformed into

$$\begin{aligned}
\mathbf{y}_j &= \mathbf{P}_{\phi,D} \mathbf{F}^H \sum_{i=1}^{N_t} \mathrm{diag}\left(\mathbf{V}_{\theta_i} \mathbf{s}_i\right) \mathbf{F}_L \mathbf{g}_{ji} + \mathbf{w}_j \\
&= \mathbf{P}_{\phi,D} \mathbf{F}^H \sum_{i=1}^{N_t} \alpha_{\theta_i}^* \alpha_{\theta_j}^* \mathrm{diag}\left(\mathbf{V}_{\theta_i} \mathbf{s}_i\right) \mathbf{F}_L \hat{\mathbf{g}}_{ji,MAP} \\
&\quad + \mathbf{e}_g + \mathbf{w}_j,
\end{aligned} \tag{53}$$

where

$$\mathbf{e}_g = \mathbf{P}_{\phi,D} \mathbf{F}^H \sum_{i=1}^{N_t} \alpha_{\theta_i}^* \alpha_{\theta_j}^* \mathrm{diag}\left(\mathbf{V}_{\theta_i} \mathbf{s}_i\right) \mathbf{F}_L \Delta_g,$$
$$\Delta_g = \tilde{\mathbf{g}}_{ji} - \hat{\mathbf{g}}_{ji,MAP}.$$

Here, $\mathbf{e}_g$ is the term that arises because of the channel estimation error. If we denote the power of $\mathbf{e}_g$ as

$$\sigma_e^2 = \frac{1}{N} \mathbf{E}\left[\mathbf{e}_g^H \mathbf{e}_g\right], \tag{54}$$

$$\Gamma = \frac{2}{\sigma_w^2} \begin{bmatrix} \Re\left\{\Lambda^H \Lambda\right\} & \Re\left\{\Lambda^H \Xi\right\} \\ \Re\left\{\Xi^H \Lambda\right\} & \Re\left\{\mathbf{X}_D^H \mathbf{F}_{N_t,D}^H \mathbf{H}^H \mathbf{H} \mathbf{F}_{N_r,D} \mathbf{X}_D\right\} \end{bmatrix} \tag{49}$$

$$\mathbf{B} = \frac{2}{\sigma_w^2} \begin{bmatrix} \Re\left\{\mathrm{diag}\left(\Upsilon^* \mathbf{1}_{N_tN} \circ \Upsilon \mathbf{1}_{N_tN}\right) + \mathbf{I}_{N_tN} \circ \left(\Upsilon \Phi \Upsilon^H\right)\right\} + \Phi^{-1} & \Re\left\{\mathrm{diag}\left(\Upsilon^* \mathbf{1}_{N_tN}\right)\Upsilon\right\} \\ \Re\left\{\Upsilon^H \mathrm{diag}\left(\Upsilon^* \mathbf{1}_{N_tN}\right)\right\} & \Re\left\{\mathbf{X}_D^H \mathbf{F}_{N_t,D}^H \mathbf{H}^H \mathbf{H} \mathbf{F}_{N_r,D} \mathbf{X}_D\right\} + \Theta^{-1} \end{bmatrix} \tag{52}$$

we can derive σ_e^2 as

$$\begin{aligned}\sigma_e^2 &= \frac{1}{N}\sum_{l=1}^{N_t}\mathbf{E}\left[\Delta_g^H\mathbf{F}_L^H\mathrm{diag}\left(\mathbf{V}_{\theta_l}^*\mathbf{s}_l^*\right)\mathrm{diag}\left(\mathbf{V}_{\theta_l}\mathbf{s}_l\right)\mathbf{F}_L\Delta_g\right]\\ &= \frac{1}{N}\sum_{l=1}^{N_t}\mathbf{E}\left[\Delta_g^H\mathbf{F}_L^H\mathbf{E}\left[\mathrm{diag}\left(\mathbf{V}_{\theta_l}^*\mathbf{s}_l^*\circ\mathbf{V}_{\theta_l}\mathbf{s}_l\right)\right]\mathbf{F}_L\Delta_g\right].\end{aligned} \tag{55}$$

To simplify (55), $\mathbf{E}\left[\mathrm{diag}\left(\mathbf{V}_{\theta_l}^*\mathbf{s}_l^*\circ\mathbf{V}_{\theta_l}\mathbf{s}_l\right)\right]$ can be approximated by Taylor series expansion as

$$\begin{aligned}&\mathbf{E}\left[\mathrm{diag}\left(\mathbf{V}_{\theta_l}^*\mathbf{s}_l^*\circ\mathbf{V}_{\theta_l}\mathbf{s}_l\right)\right]\\ &\approx \mathbf{E}\left[\mathrm{diag}\left(\left(\mathbf{F}\mathbf{X}_{l,D}\right)^*\left(\mathbf{1}_{N_t}-j\boldsymbol{\theta}_l\right)\circ\mathbf{F}\mathbf{X}_{l,D}\left(\mathbf{1}_{N_t}+j\boldsymbol{\theta}_l\right)\right)\right]\\ &= \mathbf{E}\left[\mathrm{diag}\left(\mathbf{s}_l^*\circ\mathbf{s}_l\right)\right]+\mathbf{E}\left[\mathrm{diag}\left(\left(\mathbf{F}\mathbf{X}_{l,D}\right)^*\boldsymbol{\theta}_l\circ\mathbf{F}\mathbf{X}_{l,D}\boldsymbol{\theta}_l\right)\right]\\ &= E_s\mathbf{I}+E_s\sum_{i=1}^{N}\mathbf{E}\left[\theta_l(i)^2\right]\mathbf{I}\\ &= E_s\left(1+\sigma_\epsilon^2\frac{N(N+1)}{2}\right)\mathbf{I},\end{aligned} \tag{56}$$

where E_s represents the signal energy and is considered as unit energy. Finally, through approximation of (56), σ_e^2 can be approximately calculated as

$$\sigma_e^2 \approx E_s\left(1+\sigma_\epsilon^2\frac{N(N+1)}{2}\right)MSE_{MAP}. \tag{57}$$

We assume that $\tilde{\mathbf{w}}_e = \mathbf{w}+\mathbf{e}_g$ is effective noise which follows i.i.d Gaussian noise with variance, $\sigma_{\tilde{w}}^2 = \sigma_w^2+\sigma_e^2$. This assumption is reasonable because $\mathbf{e}_g$ behaves like the Gaussian noise.

C. PROPOSED ALGORITHM

In data decoding stage, it is difficult to detect data symbol and estimate TX and RX phase noise at the same time. Hence, we preferentially correct CPE [19] and find out initial $\mathbf{x}^{(0)}$, because CPE affects the system performance more critically than ICI. Also, in channel estimation stage, since the rotated channel impulse response by CPE, $\tilde{\mathbf{g}}_{ji}$, has been estimated, the CPEs which have arisen in the channel estimation stage need to be corrected simultaneously. In this section, we denote the CPEs to arise in channel estimation $\alpha_{\phi_j,pre}$ and $\alpha_{\theta_i,pre}$ to distinguish them from the CPEs in data decoding, α_{ϕ_j} and α_{θ_i}. The entire process of our proposed algorithm is similar with the algorithm presented in [31], but we additionally consider problems incurred by channel estimation.

1) CPE CORRECTION

In order to easily estimate the CPEs, we use the signals in the frequency domain. After normalized DFT operation, (2) can be transformed into

$$r_j(k) = \sum_{i=1}^{N_t}s_i(k)\otimes v_{\theta_i}(k)\,h_{ji}(k)\otimes v_{\phi_j}(k)+n_j(k), \tag{58}$$

where $r_j(k)$ and $n_j(k)$ are the k-th normalized DFT output of $\mathbf{y}_j$ and $\mathbf{w}_j$, respectively. As only knowing the rotated channel, we can rewritten (58) as

$$\begin{aligned}r_j(k) = &\sum_{i=1}^{N_t}s_i(k)\otimes v_{\theta_i}(k)\,\alpha_{\theta_i,pre}^*\tilde{h}_{ji}(k)\,\alpha_{\phi_j,pre}^*\otimes v_{\phi_j}(k)\\ &+n_j(k),\end{aligned} \tag{59}$$

where

$$\tilde{\mathbf{h}}_{ji} = \mathbf{F}_L\tilde{\mathbf{g}}_{ji} = \left[\tilde{h}_{ji}(0),\tilde{h}_{ji}(1),\cdots,\tilde{h}_{ji}(N-1)\right]^T.$$

We can also represent (59) in matrix form as

$$\begin{aligned}\mathbf{r}_k &= \sum_{l=0}^{N-1}\sum_{m=0}^{N-1}\mathbf{A}_{\phi,k-l}\tilde{\mathbf{H}}_l\mathbf{A}_{\theta,l-m}\mathbf{s}_m+\mathbf{n}_k\\ &= \mathbf{A}_{\phi,0}\tilde{\mathbf{H}}_k\mathbf{A}_{\theta,0}\mathbf{s}_k+\sum_{l=0,l\neq k}^{N-1}\mathbf{A}_{\phi,k-l}\tilde{\mathbf{H}}_l\mathbf{A}_{\theta,l-k}\mathbf{s}_k\\ &\quad+\sum_{l=0}^{N-1}\sum_{m=0,m\neq k}^{N-1}\mathbf{A}_{\phi,k-l}\tilde{\mathbf{H}}_l\mathbf{A}_{\theta,l-m}\mathbf{s}_m+\mathbf{n}_k,\end{aligned} \tag{60}$$

where $\mathbf{r}_k = \left[r_1(k),r_2(k),\cdots,r_{N_r}(k)\right]^T \in \mathbb{C}^{N_r\times 1}$ is the received vector , $\mathbf{s}_k = \left[s_1(k),s_2(k),\cdots,s_{N_t}(k)\right]^T \in \mathbb{C}^{N_t\times 1}$ is the data vector, and $\mathbf{n}_k = [n_1(k),n_2(k),\cdots,n_{N_r}(k)]^T \in \mathbb{C}^{N_r\times 1}$ is the AWGN vector. Note that in (59) both $\mathbf{A}_{\theta,k}\in\mathbb{C}^{N_t\times N_t}$ and $\mathbf{A}_{\phi,k}\in\mathbb{C}^{N_r\times N_r}$ are the diagoanl matrix with the diagonal elements given as $\left(\mathbf{A}_{\theta,k}\right)_{i,i} = \alpha_{\theta_i,pre}^*v_{\theta_i}(k)$ and $\left(\mathbf{A}_{\phi,k}\right)_{i,i} = \alpha_{\phi_i,pre}^*v_{\phi_i}(k)$, respectively. Also, $\tilde{\mathbf{H}}_k\in\mathbb{C}^{N_r\times N_t}$ in (59) is CFR matrix with the element given as $\left(\tilde{\mathbf{H}}_k\right)_{i,j} = \tilde{h}_{ij}(k)$. Both the second and third terms in (60) are ICI terms which can be considered as AWGN. Then we can rewrite (60) simply as

$$\mathbf{r}_k = \mathbf{A}_{\phi,0}\tilde{\mathbf{H}}_k\mathbf{A}_{\theta,0}\mathbf{s}_k+\tilde{\mathbf{n}}_k = \boldsymbol{\alpha}\circ\tilde{\mathbf{H}}_k\mathbf{s}_k+\tilde{\mathbf{n}}_k, \tag{61}$$

where

$$\begin{aligned}\boldsymbol{\alpha} &= \begin{bmatrix}\alpha_{\phi_1}\alpha_{\phi_1,pre}^*\alpha_{\theta_1}\alpha_{\theta_1,pre}^*&\cdots&\alpha_{\phi_1}\alpha_{\phi_1,pre}^*\alpha_{\theta_{N_t}}\alpha_{\theta_{N_t},pre}^*\\ \vdots&\ddots&\vdots\\ \alpha_{\phi_{N_r}}\alpha_{\phi_{N_r},pre}^*\alpha_{\theta_1}\alpha_{\theta_1,pre}^*&\cdots&\alpha_{\phi_{N_r}}\alpha_{\phi_{N_r},pre}^*\alpha_{\theta_{N_t}}\alpha_{\theta_{N_t},pre}^*\end{bmatrix}\\ &= \left[\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_{N_r}\right]^T.\end{aligned}$$

Note that S_p is the set of pilot subcarriers given as $S_p = \left[q_1,q_2,...,q_{N_p}\right]$, where N_p is the number of pilot subcarriers. Applying least square using S_p, we can estimate the CPE matrix, $\boldsymbol{\alpha}$, as [19]

$$\hat{\boldsymbol{\alpha}}_j = \left(\tilde{\mathbf{H}}_{pilot,j}^H\tilde{\mathbf{H}}_{pilot,j}\right)^{-1}\tilde{\mathbf{H}}_{pilot,j}^H\mathbf{r}_{pilot,j}, \tag{62}$$

where

$$\tilde{\mathbf{H}}_{pilot,j} = \begin{bmatrix}\tilde{h}_{j1}(q_1)s_1(q_1)&\cdots&\tilde{h}_{jN_t}(q_1)s_{N_t}(q_1)\\ \vdots&\ddots&\vdots\\ \tilde{h}_{j1}(q_{N_p})s_1(q_{N_p})&\cdots&\tilde{h}_{jN_t}(q_{N_p})s_{N_p}(q_{N_p})\end{bmatrix}$$

and $\mathbf{r}_{pilot,j} = \left[r_j(q_1), r_j(q_2), \cdots, r_j\left(q_{N_p}\right)\right]^T$ is the received signal of the j-th RX antenna from all the pilot subcarrier. After equalization using CFR and $\hat{\boldsymbol{\alpha}}_j$ estimated from (62), we can decide the initial value $\tilde{\mathbf{x}}^{(0)}$ of the data.

2) JOINT ESTIMATION BASED ON MAP ESTIMATOR

Although the CPEs can be easily estimated and compensated using pilot subcarriers, residual phase noises still remain. Hence, we propose the iterative algorithm to estimate the residual phase noise based on MAP estimation.

First, using the effective noise derived in Section IV-B and the CPE matrix estimated in (62), (4) can be transformed into

$$\begin{aligned}\mathbf{y} = {} & \boldsymbol{\alpha}_{\phi}^{*} \circledast \mathbf{1}_N \mathbf{P}_{\phi,D} \mathbf{F}_{N_r,D}^{H} \left(\hat{\boldsymbol{\alpha}} \circledast \mathbf{I}_N \circ \hat{\mathbf{H}}\right) \\ & \times \mathbf{F}_{N_t,D} \boldsymbol{\alpha}_{\theta}^{*} \circledast \mathbf{1}_N \mathbf{P}_{\theta,D} \mathbf{x} + \tilde{\mathbf{w}}_e, \end{aligned} \tag{63}$$

where

$$\begin{aligned}\boldsymbol{\alpha}_{\phi} &= \left[\alpha_{\phi_1}, \alpha_{\phi_2}, \cdots, \alpha_{\phi_{N_r}}\right]^T, \\ \boldsymbol{\alpha}_{\theta} &= \left[\alpha_{\theta_1}, \alpha_{\theta_2}, \cdots, \alpha_{\theta_{N_t}}\right]^T \end{aligned}$$

and $\hat{\mathbf{H}}$ has the same structure with $\mathbf{H}$ but is composed of $\hat{\mathbf{g}}_{ji,MAP}$ instead of $\mathbf{g}_{ji}$. Also, we assume that $\boldsymbol{\alpha}$ is perfectly estimated. The assumption can be sufficiently vaild by using the enough number of pilot subcarriers. From the fact that the phase noise can be slight enough for CPE to be approximated as $\alpha_{\phi_j} \approx \exp\left(j \sum_{n=1}^{N} \phi_j(n)\right)$, we can approximate (63) as

$$\begin{aligned}\mathbf{y} &\approx \mathbf{P}_{\tilde{\phi},D} \mathbf{F}_{N_r,D}^{H} \left(\hat{\boldsymbol{\alpha}} \circledast \mathbf{I}_N \circ \hat{\mathbf{H}}\right) \mathbf{F}_{N_t,D} \mathbf{P}_{\tilde{\theta},D} \mathbf{x} + \tilde{\mathbf{w}}_e \\ &= \tilde{\Xi} \mathbf{p}_{\tilde{\theta}} + \tilde{\mathbf{w}}_e = \tilde{\Lambda} \mathbf{p}_{\tilde{\phi}} + \tilde{\mathbf{w}}_e, \end{aligned} \tag{64}$$

where

$$\begin{aligned}\tilde{\Xi} &= \mathbf{P}_{\tilde{\phi},D} \mathbf{F}_{N_r,D}^{H} \hat{\boldsymbol{\alpha}} \circledast \mathbf{I}_N \hat{\mathbf{H}} \mathbf{F}_{N_t,D} \mathbf{X}_D \\ \tilde{\Lambda} &= \mathrm{diag}\left(\mathbf{P}_{\tilde{\phi},D} \mathbf{F}_{N_r,D}^{H} \left(\hat{\boldsymbol{\alpha}} \circledast \mathbf{I}_N \circ \hat{\mathbf{H}}\right) \mathbf{F}_{N_t,D} \mathbf{P}_{\tilde{\theta},D} \mathbf{x}\right), \end{aligned}$$

both $\tilde{\phi}$ and $\tilde{\theta}$ are residual phase noises at TX and RX, respectively, that is, the mean value subtracted from $\boldsymbol{\phi}_j$ and $\boldsymbol{\theta}_i$ at each antenna.

The log likelihood function (LLF) of the pdf $p\left(\mathbf{y}, \tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\theta}}\right)$ is given as

$$L\left(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\theta}}\right) = -\ln p\left(\mathbf{y} \mid \tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\theta}}\right) - \ln p\left(\tilde{\boldsymbol{\phi}}\right) - \ln p\left(\tilde{\boldsymbol{\theta}}\right). \tag{65}$$

Both $\tilde{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\theta}}$ follow zero mean Gaussian distribution, and their covariance matrices, $\mathbf{C}_{\tilde{\phi}}$ and $\mathbf{C}_{\tilde{\theta}}$, are derived in (32). Note that from (64), we can find LLF in (65) as

$$\begin{aligned}L\left(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\theta}}\right) &= \frac{1}{\sigma_{\tilde{w}_e}^2} \left(\mathbf{y} - \tilde{\Xi} \mathbf{p}_{\tilde{\theta}}\right)^H \left(\mathbf{y} - \tilde{\Xi} \mathbf{p}_{\tilde{\theta}}\right) \\ &\quad + \frac{1}{2} \tilde{\boldsymbol{\phi}}^T \mathbf{C}_{\tilde{\phi}}^{-1} \tilde{\boldsymbol{\phi}} + \frac{1}{2} \tilde{\boldsymbol{\theta}}^T \mathbf{C}_{\tilde{\theta}}^{-1} \tilde{\boldsymbol{\theta}} \\ &= \frac{1}{\sigma_{\tilde{w}_e}^2} \left(\mathbf{y}^H \mathbf{y} - 2\Re\left\{\mathbf{p}_{\tilde{\theta}} \tilde{\Xi}^H \mathbf{y}\right\} + \mathbf{p}_{\tilde{\theta}}^H \tilde{\Xi}^H \tilde{\Xi} \mathbf{p}_{\tilde{\theta}}\right) \\ &\quad + \frac{1}{2} \tilde{\boldsymbol{\phi}}^T \mathbf{C}_{\tilde{\phi}}^{-1} \tilde{\boldsymbol{\phi}} + \frac{1}{2} \tilde{\boldsymbol{\theta}}^T \mathbf{C}_{\tilde{\theta}}^{-1} \tilde{\boldsymbol{\theta}}. \end{aligned} \tag{66}$$

Assuming $\mathbf{p}_{\theta} \approx \mathbf{1}_{N_t N} + j\boldsymbol{\theta}$ by Taylor's series expansion, we can rewrite (66) as

$$\begin{aligned}L\left(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\theta}}\right) = {} & \frac{1}{\sigma_{\tilde{w}_e}^2} \mathbf{y}^H \mathbf{y} - \frac{2}{\sigma_{\tilde{w}_e}^2} \Re\left\{\left(\mathbf{1}_{N_t N}^T - j\tilde{\boldsymbol{\theta}}^T\right) \tilde{\Xi}^H \mathbf{y}\right\} \\ & + \frac{1}{\sigma_{\tilde{w}_e}^2} \left(\mathbf{1}_{N_t N}^T - j\tilde{\boldsymbol{\theta}}^T\right) \tilde{\Xi}^H \tilde{\Xi} \left(\mathbf{1}_{N_t N} + j\tilde{\boldsymbol{\theta}}\right) \\ & + \frac{1}{2} \tilde{\boldsymbol{\phi}}^T \mathbf{C}_{\tilde{\phi}}^{-1} \tilde{\boldsymbol{\phi}} + \frac{1}{2} \tilde{\boldsymbol{\theta}}^T \mathbf{C}_{\tilde{\theta}}^{-1} \tilde{\boldsymbol{\theta}}. \end{aligned} \tag{67}$$

Taking the partial derivative of (67) with respect to $\boldsymbol{\theta}$ and setting it to zero, we obtain $\hat{\boldsymbol{\theta}}$ as

$$\hat{\boldsymbol{\theta}} = \left(\Re\left\{\tilde{\Xi}^H \tilde{\Xi}\right\} + \frac{\sigma_{\tilde{w}_e}^2}{2} \mathbf{C}_{\tilde{\theta}}^{-1}\right)^{-1} \left(\Im\left\{\tilde{\Xi}^H \mathbf{y}\right\} - \Im\left\{\tilde{\Xi}^H \tilde{\Xi}\right\} \mathbf{1}_{N_t N}\right). \tag{68}$$

Following the similar procedures as the case of TX phase noise, we can obtain $\hat{\boldsymbol{\phi}}$ as

$$\hat{\boldsymbol{\phi}} = \left(\Re\left\{\tilde{\Lambda}^H \tilde{\Lambda}\right\} + \frac{\sigma_{\tilde{w}_e}^2}{2} \mathbf{C}_{\tilde{\phi}}^{-1}\right)^{-1} \left(\Im\left\{\tilde{\Lambda}^H \mathbf{y}\right\} - \Im\left\{\tilde{\Lambda}^H \tilde{\Lambda}\right\} \mathbf{1}_{N_t N}\right). \tag{69}$$

However, $\mathbf{x}$, $\tilde{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\theta}}$ are needed for us to calculate (68) and (69), respectively. Therefore, our proposed algorithm is to iteratively estimate the k-th values such as $\mathbf{x}^{(k)}$, $\hat{\boldsymbol{\phi}}^{(k)}$ and $\hat{\boldsymbol{\theta}}^{(k)}$ by using previous values, $\mathbf{x}^{(k-1)}$, $\hat{\boldsymbol{\phi}}^{(k-1)}$ and $\hat{\boldsymbol{\theta}}^{(k-1)}$. In iterative algorithms, first of all, it is important to set initial value, $\mathbf{x}^{(0)}$, $\hat{\boldsymbol{\phi}}^{(0)}$ and $\hat{\boldsymbol{\theta}}^{(0)}$. First, $\mathbf{x}^{(0)}$ is calculated by equalizing with estimated CFR and $\hat{\boldsymbol{\alpha}}$ and detecting the symbol. In case of $\hat{\boldsymbol{\theta}}^{(0)}$, as we do not know the initial value of residual RX phase noise, we need to modify the MAP estimator of $\hat{\boldsymbol{\theta}}$ given in (68). Considering $\boldsymbol{\epsilon}_{rx}$ incurred by $\tilde{\phi}$, as Gaussian noise, we can derive the covariance of $\boldsymbol{\epsilon}_{rx}$ by converting $\mathbf{s}_l \mathbf{s}_l^H$ into $E_s \mathbf{I}_N$ in (35), and set the initial value, $\hat{\boldsymbol{\theta}}^{(0)}$, as

$$\begin{aligned}\hat{\boldsymbol{\theta}}^{(0)} = {} & \left(\Re\left\{\tilde{\Xi}^H \tilde{\mathbf{C}}_{\tilde{w}_e}^{-1} \tilde{\Xi}\right\} + \frac{1}{2} \mathbf{C}_{\tilde{\theta}}^{-1}\right)^{-1} \\ & \times \left(\Im\left\{\tilde{\Xi}^H \tilde{\mathbf{C}}_{\tilde{w}_e}^{-1} \mathbf{y}\right\} - \Im\left\{\tilde{\Xi}^H \tilde{\mathbf{C}}_{\tilde{w}_e}^{-1} \tilde{\Xi}\right\} \mathbf{1}_{N_t N}\right), \end{aligned} \tag{70}$$

where $\tilde{\mathbf{C}}_{\tilde{w}_e} = \mathbf{C}_{\epsilon_{rx}} + \sigma_{\tilde{w}_e}^2 \mathbf{I}_{N_t N}$. Then, substituting $\hat{\boldsymbol{\theta}}^{(0)}$ and $\mathbf{x}^{(0)}$ into (69), we can find out initial value of residual RX phase noise, $\hat{\boldsymbol{\phi}}^{(0)}$. Defining that $\mathbf{e}^{(k)}$ is the k-th sum of the squared error between the demodulated data, $\hat{\mathbf{s}}_i = \mathbf{F}\hat{\mathbf{x}}_i$, in S_p and the pilot symbols, we iterate our algorithm until the sum of square error increases compared to the previous value. Our proposed algorithm is summarized in Table 1.

V. SIMULATION RESULTS

For our simulations, we consider OFDM system with 64 subcarriers, that is, $N = 64$, and set the cyclic prefix length as 16, which is one of the often adopted systems such as in [1]. We also consider two different types of symbol, block-type pilot symbols and comb-type pilot symbols. The block-type pilot symbols are the first symbol of one frame and consist

TABLE 1. Joint data detection and phase noise estimation.

Initialize $\tilde{\mathbf{x}}^{(0)}$, $\tilde{\boldsymbol{\theta}}^{(0)}$, $\tilde{\boldsymbol{\phi}}^{(0)}$, $k = 0$ and $e^{(0)} = \infty$
Repeat
1. Find out $\tilde{\mathbf{x}}^{(k+1)}$
   - Multiply $\mathbf{y}$ by $\tilde{\boldsymbol{\phi}}^{(k)*}$.
   - After DFT operation, Equalize CFR.
   - Multiplying equlized datas by circulant matrix of $\mathrm{DFT}\left\{(\tilde{\boldsymbol{\theta}}_i)^*\right\}$ each TX ends, find out $\tilde{\mathbf{s}}_i^{(k+1)}$ and $\tilde{\mathbf{x}}^{(k+1)}$.
2. Using (68) with $\tilde{\mathbf{x}}^{(k+1)}$ and $\tilde{\boldsymbol{\phi}}^{(k)}$, update $\tilde{\boldsymbol{\theta}}^{(k+1)}$.
3. Using (69) with $\tilde{\mathbf{x}}^{(k+1)}$ and $\tilde{\boldsymbol{\theta}}^{(k+1)}$, update $\tilde{\boldsymbol{\phi}}^{(k+1)}$.
4. $e^{(k+1)} = \sum_{q\in S_p}\left(\tilde{\mathbf{s}}_q^{(k+1)} - \mathbf{s}_q\right)^H\left(\tilde{\mathbf{s}}_q^{(k+1)} - \mathbf{s}_q\right)$.
5. $k = k + 1$.

Until $e^{(k+1)} > e^{(k)}$, and return $\tilde{\mathbf{s}}^{(k)}$

of only 64 pilot subcarrier. Otherwise, the comb-type pilot symbols are followed by the block-type pilot symbol and have 16 pilot subcarrier, so there are 48 data subcarriers. The data modulation is given by 16-QAM, while the pilot modulation is BPSK. The channel has 6 taps with each tap independently Rayleigh distributed. The power profile of the channel is specified by 3-dB decay per tap and the total power is normalized to have unit energy. Also, the channels among all the TX and RX antennas are assumed to be independent. Note that we consider the same number of antennas at TX and RX ($N_t = N_r = L$) and L data streams are spatially multiplexed. The phase noise model is Wiener process with the variance, given as $2\pi\beta T_s = 5 \times 10^{-3}(\mathrm{rad})^2$. Also, in our paper, we consider Hadamard sequences as only training sequences, and Hadamard n denotes the n-th column vector of Hadamard matrix, whose size is 64×64. Note that the optimal training sequences under the absence of phase noise [25] are used to training sequences for comparison, where frequency division multiplexing (FDM) is only considered in this paper.

First, we introduce the simulation results that validate our mathematical analysis and MSE performances in channel estimation stage. Figs. 3, 4 and 5 compare the MSE from simulation with the mathematical analysis in (36) and (42) versus transmit SNR when using LS algorithm (27) in 2x2 MIMO-OFDM, MAP algorithm in (29) for 2x2 MIMO-OFDM and for 4x4 MIMO-OFDM, respectively. Also, Fig. 6 illustrates the comparison between MSE of simulation result and mathematical derivation versus phase noise variance. From above figures, the mathematical results are almost the same with actual MSE of estimated CIR regardless of SNR, phase noise variance and the number of TX antennas. Hence, we can conclude that the mathematical analysis can be utilized when selecting the combination of training sequence and estimating the residual phase noise in proposed algorithm.

Due to good auto and cross correlation characteristics of Hadamard type codes, it is often adopted in estimating the wireless channel and requiring orthogonality [32]. Therefore, in our paper, we choose the combination of training sequences in the group of Hadamard sequence, which has the same length with OFDM size of 64, and consists of total 64 codes. Using exhaustive searching with the mathematical

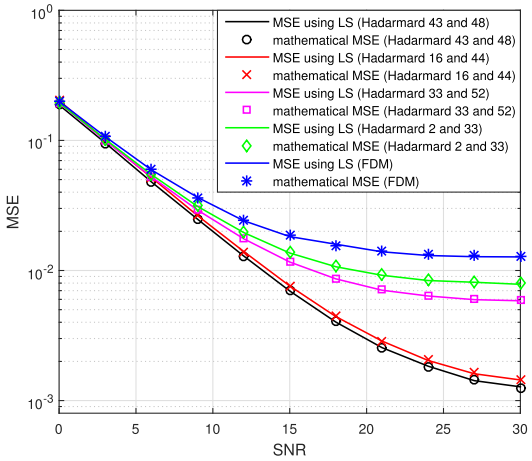

FIGURE 3. Comparison between MSE of mathematical analysis and simulation result using LS algorithm in 2x2 MIMO.

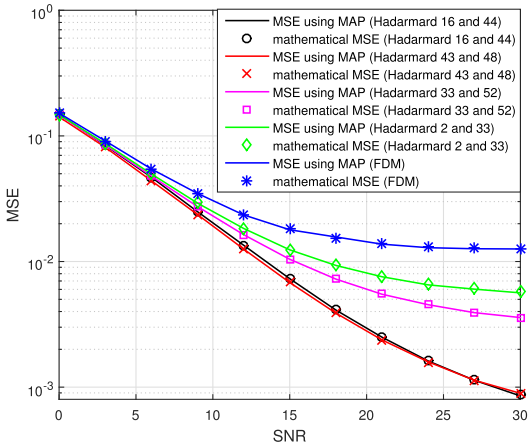

FIGURE 4. Comparison between channel MSE of mathematical analysis and simulation result using MAP algorithm in 2x2 MIMO.

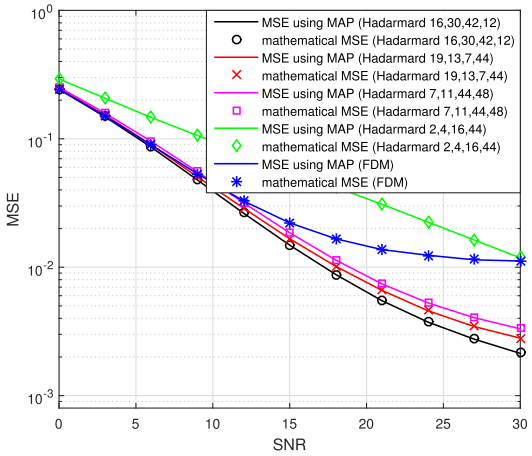

FIGURE 5. Comparison between channel MSE of mathematical analysis and simulation result using MAP algorithm in 4x4 MIMO.

MSE, we select the training sequences minimizing the MSE of channel in Hadamard group. As designing preambles proceeds off-line, the complexity and processing time of

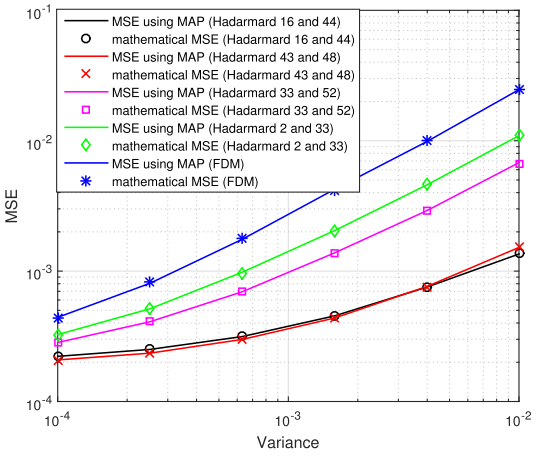

FIGURE 6. Comparison between channel MSE of mathematical analysis and simulation result using MAP algorithm in 2x2 MIMO.

TABLE 2. The optimal combination of training sequences.

N_t	Index	channel estimation scheme	
		LS algorithm	MAP algorithm
2	1	Hadamard 43	Hadamard 16
	2	Hadamard 48	Hadamard 48
3	1	Hadamard 38	Hadamard 19
	2	Hadamard 63	Hadamard 13
	3	Hadamard 60	Hadamard 7
4	1	Hadamard 45	Hadamard 16
	2	Hadamard 56	Hadamard 30
	3	Hadamard 51	Hadamard 42
	4	Hadamard 42	Hadamard 12

algorithm to find out solution of optimization problem are not considered. In Table 2, there are the optimal training sequence in Hadamard sequences each TX antenna depending on estimation algorithms and the number of TX antennas. Also, we can find out the optimal training sequences in different group of sequences if the channel statistic and phase noise model will be given.

In Fig. 7, we compare the MSE performance of our proposed scheme with one of LS channel estimation as well as BCRLB for different combination of training sequences. Note that there is significant performance gain with proposed scheme compared with the LS algorithm, and we can clearly see that the MSE performance can be more closed to BCRLB by using our proposed preamble design than FDM which is optimal sequence in the absence of phase noise. As you can see in Fig. 7, there is about 3 dB gap between MSE and BCRLB at low SNR. We think that the reason is that there are much more unknown parameters such as channel impulse response, TX and RX phase noise than the number of training sequences.

After channel estimation, in date decoding stage, we show MSE of phase noise and BER performances to represent the performance gain of our scheme in Table 1. Fig. 8 compares the MSE performance of proposed phase noise estimation scheme with one of the scheme presented in [19] as well as BCRLB and no correction case. We consider for this simulation that there exists the phase noise only at RX end for

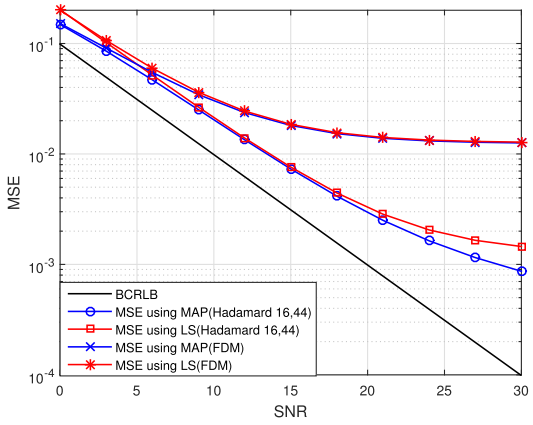

FIGURE 7. Comparison between MSE and BCRLB of channel in 2x2 MIMO-OFDM system.

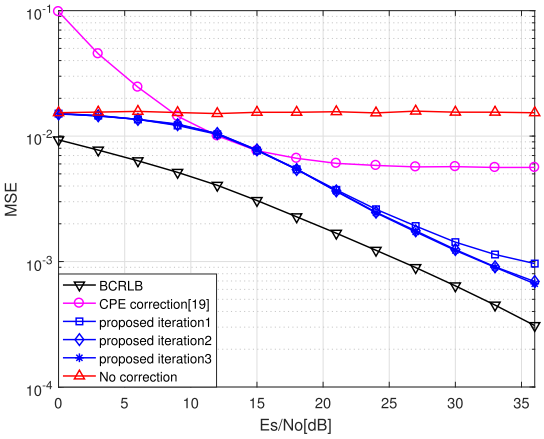

FIGURE 8. Comparison between MSE and BCRLB of RX phase noise in 2x2 MIMO-OFDM system.

comparison with the CPE correction scheme given in [19]. This is because the scheme in [19] can only estimate the CPE matrix, $\boldsymbol{\alpha}$, where its element is the product of CPEs at TX and RX phase noise. We observe from Fig. 8 that the proposed $\boldsymbol{\phi}$ estimator is closer to BCRLB than other cases. Also, the MSE performance of the proposed scheme improves by increasing the number of iterations, and saturates at a certain iteration 3. Like Fig. 7, there are gaps between MSE performances and BCRLB in Fig. 8. We think that in deriving BCRLB we assume that data and pilot subcarriers have already known, while in our proposed algorithm we only know pilot subcarriers to estimate phase noise. Hence, it is impossible to be close to theoretical lower bound.

In Fig. 9, the BER performances of proposed scheme and CPE correction are compared depending on whether channel is perfectly known. The case of no correction and without phase noise are used as upper bound and lower bound, respectively, for criterion of comparison. We observe from Fig. 9 that significant BER performance improvement can be achieved by using our proposed algorithm, compared to the scheme in [19] and no correction case. Although our proposed algorithm has high complexity because of requiring

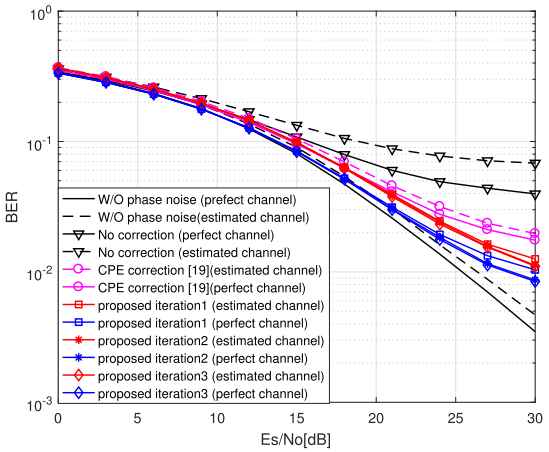

FIGURE 9. Comparison of BER performances in 2x2 MIMO-OFDM system.

inverse matrix operation, our algorithm makes about 7dB gain compared with CPE correction algorithm [19] at SNR 30dB, which is a significant gain to make the success of communication. Also, considering that the phase noise variance in simulation condition is significantly large compared to phase noise at low carrier frequency, the performance of proposed algorithm is sufficiently close to the lower bound. In addition, although the estimated channel using MAP estimator and proposed combination of training sequence in Table 2 are used instead of perfect channel, there is only at most 3dB loss at SNR 30dB for both CPE correction and proposed algorithm. Finally, we confirm that the performance saturates only in 3 iterations for most of time in our simulations. Hence, our iterative algorithm does not require the large number of iterations.

VI. COMPLEXITY ANALYSIS

In this section, we analyze the complexity of our algorithms proposed in Section III and IV. First, in channel estimation stage, we propose the MAP estimator given in (29). In (29), the complexity is mainly determined by a $N_tL \times N_tL$ matrix inversion, whose complexity is $O\left({N_t}^3 L^3\right)$, because the other matrix can be calculated in advance from the statistical information of phase noise and channel. Also, as we need to estimate the channel impulse response at all receive ends, the total computational complexity is approximately $O\left(N_r {N_t}^3 L^3\right)$ in channel estimation.

In data detection stage, our proposed algorithm is summarized in Table 1. As seen in Table 1, many operation and process are required to mitigate the effect of phase noise. However, the factor to mainly determine the complexity of our algorithm is (68) and (69), because (68) and (69) require high rank matrix inversion. Therefore, the complexity of (68) and (68) are approximately $O\left({N_t}^3 N^3\right)$ and $O\left({N_r}^3 N^3\right)$ respectively. Assuming that $N_t = N_r$ and the number of iteration is 3, which saturates the system performance in Fig. 9, the total complexity required in data decoding stage is approximately $O\left(6{N_t}^3 N^3\right)$.

VII. CONCLUSION

In this paper, we proposed the scheme mitigating the effects of phase noise in MIMO-OFDM with independent oscillators. The proposed scheme consists of channel estimation stage and data decoding stage. In channel estimation stage, we proposed channel estimation algorithm based on MAP estimator and the optimal training sequences in a certain group of sequences through mathematical analysis. Then, in data decoding stage, we proposed the iterative algorithm with MAP estimator of phase noises and the estimated channel. In addition, we derived BCRLB, which is important in estimation problem because of presenting the lower bound of estimators, for each stage. From simulation results, we confirmed that the proposed scheme can improve the system performances in terms of MSE and BER compared with the existing schemes.

APPENDIX

A. DERIVATION OF BCRLB

To derive the BCRLB of $\boldsymbol{\eta}_1$, we first find out BIM in (10). Since the second term of (10) can be calculated by the statistical information of channel and phase noise, expectation of (17) need to be found out, which is given by

$$\mathbf{E}_{\eta}\begin{bmatrix} \frac{\partial \boldsymbol{\mu}_1^H}{\partial \boldsymbol{\phi}_j}\frac{\partial \boldsymbol{\mu}_1}{\partial \boldsymbol{\phi}_j^T} & \frac{\partial \boldsymbol{\mu}_1^H}{\partial \boldsymbol{\phi}_j}\frac{\partial \boldsymbol{\mu}_1}{\partial \boldsymbol{\theta}^T} & \frac{\partial \boldsymbol{\mu}_1^H}{\partial \boldsymbol{\phi}_j}\frac{\partial \boldsymbol{\mu}_1}{\partial \mathbf{g}_{j,re}^T} & \frac{\partial \boldsymbol{\mu}_1^H}{\partial \boldsymbol{\phi}_j}\frac{\partial \boldsymbol{\mu}_1}{\partial \mathbf{g}_{j,im}^T} \\ \frac{\partial \boldsymbol{\mu}_1^H}{\partial \boldsymbol{\theta}}\frac{\partial \boldsymbol{\mu}_1}{\partial \boldsymbol{\phi}_j^T} & \frac{\partial \boldsymbol{\mu}_1^H}{\partial \boldsymbol{\theta}}\frac{\partial \boldsymbol{\mu}_1}{\partial \boldsymbol{\theta}^T} & \frac{\partial \boldsymbol{\mu}_1^H}{\partial \boldsymbol{\theta}}\frac{\partial \boldsymbol{\mu}_1}{\partial \mathbf{g}_{j,re}^T} & \frac{\partial \boldsymbol{\mu}_1^H}{\partial \boldsymbol{\theta}}\frac{\partial \boldsymbol{\mu}_1}{\partial \mathbf{g}_{j,im}^T} \\ \frac{\partial \boldsymbol{\mu}_1^H}{\partial \mathbf{g}_{j,re}}\frac{\partial \boldsymbol{\mu}_1}{\partial \boldsymbol{\phi}_j^T} & \frac{\partial \boldsymbol{\mu}_1^H}{\partial \mathbf{g}_{j,re}}\frac{\partial \boldsymbol{\mu}_1}{\partial \boldsymbol{\theta}^T} & \frac{\partial \boldsymbol{\mu}_1^H}{\partial \mathbf{g}_{j,re}}\frac{\partial \boldsymbol{\mu}_1}{\partial \mathbf{g}_{j,re}^T} & \frac{\partial \boldsymbol{\mu}_1^H}{\partial \mathbf{g}_{j,re}}\frac{\partial \boldsymbol{\mu}_1}{\partial \mathbf{g}_{j,im}^T} \\ \frac{\partial \boldsymbol{\mu}_1^H}{\partial \mathbf{g}_{j,im}}\frac{\partial \boldsymbol{\mu}_1}{\partial \boldsymbol{\phi}_j^T} & \frac{\partial \boldsymbol{\mu}_1^H}{\partial \mathbf{g}_{j,im}}\frac{\partial \boldsymbol{\mu}_1}{\partial \boldsymbol{\theta}^T} & \frac{\partial \boldsymbol{\mu}_1^H}{\partial \mathbf{g}_{j,im}}\frac{\partial \boldsymbol{\mu}_1}{\partial \mathbf{g}_{j,re}^T} & \frac{\partial \boldsymbol{\mu}_1^H}{\partial \mathbf{g}_{j,im}}\frac{\partial \boldsymbol{\mu}_1}{\partial \mathbf{g}_{j,im}^T} \end{bmatrix}. \tag{A.1}$$

The terms related to only phase noise in (A.1) can be calculated by (16) below as

$$\begin{aligned}
&\mathbf{E}\left[\frac{\partial \boldsymbol{\mu}_1^H}{\partial \boldsymbol{\phi}_j}\frac{\partial \boldsymbol{\mu}_1}{\partial \boldsymbol{\phi}_j^T}\right] \approx \mathbf{E}\left[\Lambda_1^H \Lambda_1\right] \\
&\approx \mathbf{E}\Big[\mathrm{diag}\Big(\left(\mathbf{F}^H\mathbf{H}_j\mathbf{F}_{N_t,D}\mathbf{X}_D\right)^* (\mathbf{1}-j\boldsymbol{\theta})\Big) \\
&\quad \times \mathrm{diag}\Big(\mathbf{F}^H\mathbf{H}_j\mathbf{F}_{N_t,D}\mathbf{X}_D\,(\mathbf{1}+j\boldsymbol{\theta})\Big)\Big] \\
&= \mathbf{E}\left[\mathrm{diag}\left(\left(\mathbf{F}^H\mathbf{H}_j\mathbf{F}_{N_t,D}\mathbf{x}\right)^* \circ \left(\mathbf{F}^H\mathbf{H}_j\mathbf{F}_{N_t,D}\mathbf{x}\right)\right)\right] \\
&\quad + \mathbf{E}\left[\mathrm{diag}\left(\left(\mathbf{F}^H\mathbf{H}_j\mathbf{F}_{N_t,D}\mathbf{X}_D\boldsymbol{\theta}\right)^* \circ \left(\mathbf{F}^H\mathbf{H}_j\mathbf{F}_{N_t,D}\mathbf{X}_D\boldsymbol{\theta}\right)\right)\right] \\
&= \mathrm{diag}\left(\tilde{\mathbf{C}}_{g_{cir}}^{N_t}\mathbf{c}_x + \tilde{\mathbf{C}}_{g_{cir}}^{N_t}(\mathbf{I}\circ\boldsymbol{\Theta})\,\mathbf{c}_x\right)
\end{aligned} \tag{A.2}$$

$$\begin{aligned}
&\mathbf{E}\left[\frac{\partial \boldsymbol{\mu}_1^H}{\partial \boldsymbol{\phi}_j}\frac{\partial \boldsymbol{\mu}_1}{\partial \boldsymbol{\theta}^T}\right] \approx \mathbf{E}\left[\Lambda_1^H \Xi_1\right] \\
&\approx \mathbf{E}\Big[\mathrm{diag}\left(\left(\mathbf{F}^H\mathbf{H}_j\mathbf{F}_{N_t,D}\mathbf{x}\right)^* (\mathbf{1}-j\boldsymbol{\theta})\right) \\
&\quad \times \left(\mathbf{I}+j\mathrm{diag}\left(\boldsymbol{\phi}_j\right)\right)\mathbf{F}^H\mathbf{H}_j\mathbf{F}_{N_t,D}\mathbf{X}_D\Big] \\
&= \mathbf{E}\left[\mathrm{diag}\left(\left(\mathbf{F}^H\mathbf{H}_j\mathbf{F}_{N_t,D}\mathbf{x}\right)^*\right)\mathbf{F}^H\mathbf{H}_j\mathbf{F}_{N_t,D}\mathbf{X}_D\right] \\
&= \tilde{\mathbf{C}}_{g_{cir}}^{N_t} \circ \mathbf{C}_x
\end{aligned} \tag{A.3}$$

$$\frac{1}{N^2}\mathbf{E}\left[\left(\sum_{n=2}^{p}(n-1)\,\zeta\,(n)-\sum_{m=1}^{N-p}(N-p+1-m)\,\zeta\,(p+m)\right)\left(\sum_{n=2}^{q}(n-1)\,\zeta\,(n)-\sum_{m=1}^{N-q}(N-q+1-m)\,\zeta\,(q+m)\right)\right] \tag{A.8}$$

$$\mathbf{E}\left[\frac{\partial\boldsymbol{\mu}_1^H}{\partial\boldsymbol{\theta}}\frac{\partial\boldsymbol{\mu}_1}{\partial\boldsymbol{\theta}^T}\right]\approx\mathbf{E}\left[\Xi^H\Xi\right]$$
$$=\mathbf{X}_D^H\mathbf{F}_{N_t,D}^H\mathbf{E}\left[\mathbf{H}^H\mathbf{H}\right]\mathbf{F}_{N_t,D}\mathbf{X}_D=\frac{1}{N_t}\mathbf{X}_D^H\mathbf{X}_D. \tag{A.4}$$

We can derive the Fisher information between channel and phase noise in (A.1) as

$$\mathbf{E}\left[\frac{\partial\boldsymbol{\mu}_1^H}{\partial\boldsymbol{\phi}_j}\frac{\partial\boldsymbol{\mu}_1}{\partial\mathbf{g}_{j,re}^T}\right]\approx-j\mathbf{E}\left[\Lambda^H\mathbf{P}_{\phi_j,D}\tilde{\mathbf{Q}}\right]$$
$$=-j\mathbf{E}\left[\mathrm{diag}\left(\left(\mathbf{F}^H\mathbf{H}_j\mathbf{F}_{N_t,D}\mathbf{x}\right)^*(\mathbf{1}-j\boldsymbol{\theta})\right)\mathbf{P}_{\phi_j,D}\tilde{\mathbf{Q}}\right]$$
$$=\mathbf{0}_{N\times LN_t}, \tag{A.5}$$

because the correlations between channel and phase noise is zero. The rest terms regarding $\mathbf{g}_j$, $\boldsymbol{\phi}_j$ and $\boldsymbol{\theta}$ in (A.1) also can be approximated as zero.

Lastly, the Fisher information about only channel is given by

$$\mathbf{E}\left[\frac{\partial\boldsymbol{\mu}_1^H}{\partial\mathbf{g}_{j,re}}\frac{\partial\boldsymbol{\mu}_1}{\partial\mathbf{g}_{j,re}^T}\right]=\mathbf{E}\left[\tilde{\mathbf{Q}}^H\mathbf{P}_{\phi_j,D}^H\mathbf{P}_{\phi_j,D}\tilde{\mathbf{Q}}\right]=\mathbf{E}\left[\tilde{\mathbf{Q}}^H\tilde{\mathbf{Q}}\right],$$
$$\mathbf{E}\left[\frac{\partial\boldsymbol{\mu}_1^H}{\partial\mathbf{g}_{j,re}}\frac{\partial\boldsymbol{\mu}_1}{\partial\mathbf{g}_{j,im}^T}\right]=j\mathbf{E}\left[\tilde{\mathbf{Q}}^H\tilde{\mathbf{Q}}\right],$$
$$\mathbf{E}\left[\frac{\partial\boldsymbol{\mu}_1^H}{\partial\mathbf{g}_{j,im}}\frac{\partial\boldsymbol{\mu}_1}{\partial\mathbf{g}_{j,im}^T}\right]=\mathbf{E}\left[\tilde{\mathbf{Q}}^H\tilde{\mathbf{Q}}\right]. \tag{A.6}$$

As, in (10), the second term can be obtained from the second order statistic of channel and phase noise, BIM of $\boldsymbol{\eta}_1$ given in (18) can be obtained.

B. DERIVATION OF COVARIANCE

Here, the covariance of residual phase noise ,$\tilde{\mathbf{p}}_{\theta_l}$, is derived. By Taylor's series expansion, the covariance can be approximated as

$$\left(\mathbf{E}\left[\tilde{\mathbf{p}}_{\theta_l}\tilde{\mathbf{p}}_{\theta_l}^H\right]\right)_{p,g}$$
$$=\mathbf{E}\left[\left(e^{j\boldsymbol{\theta}_l(p)}-\frac{1}{N}\sum_{n=1}^{N}e^{j\boldsymbol{\theta}_l(n)}\right)\left(e^{j\boldsymbol{\theta}_l(q)}-\frac{1}{N}\sum_{m=1}^{N}e^{j\boldsymbol{\theta}_l(m)}\right)^*\right]$$
$$\approx\frac{1}{N^2}\mathbf{E}\left[\left(\sum_{n=1}^{N}\theta_l\,(p)-\theta_l\,(n)\right)\left(\sum_{m=1}^{N}\theta_l\,(q)-\theta_l\,(m)\right)\right]. \tag{A.7}$$

Assuming that phase noises follow the Wiener process such as $\theta_j\,(n)=\theta_j\,(n-1)+\zeta\,(n)$, (A.7) can be rewritten as (A.8), as shown at the top of this page. As $\zeta\,(n)$ follows independent Gaussian distribution, (A.8) can be finally written as

$$\frac{1}{N^2}\mathbf{E}\left[\left(\sum_{n=1}^{N}\theta_l\,(p)-\theta_l\,(n)\right)\left(\sum_{m=1}^{N}\theta_l\,(q)-\theta_l\,(m)\right)\right]$$
$$=\frac{1}{N^2}\sum_{n=2}^{a}(n-1)^2\,\sigma_\zeta^2+\frac{1}{N^2}\sum_{l=1}^{N-b}(N-b+1-l)\,\sigma_\zeta^2$$
$$-\frac{1}{N^2}\sum_{m=a+1}^{b}(N-m+1)\,(m-1)\,\sigma_\zeta^2$$
$$=\frac{\sigma_\zeta^2\,(N-b)\,(N-b+1)\,(2N-2b+1)}{6N^2}$$
$$+\frac{\sigma_\zeta^2\,(N+1)\,(b-a)}{N^2}+\frac{\sigma_\zeta^2 b\,(b+1)\,(2b+1)}{6N^2}-\frac{\sigma_\zeta^2 a^2}{N^2}$$
$$-\frac{\sigma_\zeta^2\,(N+2)\,(b\,(b+1)-a\,(a+1))}{2N^2}$$
$$=\frac{\sigma_\zeta^2}{6N}\left[3b^2-3\,(2N+1)\,b+3a^2-3a+2N^2+3N+1\right], \tag{A.9}$$

Acknowledgment

This paper was presented at the IEEE Conference on Standards for Communications and Networking, Berlin, Germany, Oct. 31–Nov. 2, 2016.

REFERENCES

[1] *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 4: Enhancements for Very High Throughput for Operation in Bands below 6 GHz*, IEEE Standard 802.11ac-2013, 2013.

[2] *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band*, IEEE Standard 802.11ad-2012, 2012.

[3] *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN): Overall Description*, document TS36.300, 3GPP, 2010.

[4] A. Chorti and M. Brookes, "A spectral model for RF oscillators with power-law phase noise," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 53, no. 9, pp. 1989–1999, Sep. 2006.

[5] L. Tomba, "On the effect of Wiener phase noise in OFDM systems," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 580–583, May 1998.

[6] P. Papadimitriou, T. Ihalainen, H. Berg, and K. Hugl, "Link-level performance of an LTE ue receiver in synchronous and asynchronous networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Sep. 2013, pp. 3861–3866.

[7] D. Petrovic, 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.

[8] T. C. W. Schenk, X.-J. Tao, P. F. M. Smulders, and E. R. Fledderus, "Influence and suppression of phase noise in multi-antenna OFDM," in *Proc. IEEE 60th Veh. Technol. Conf. VTC2004-Fall*, vol. 2. Sep. 2004, pp. 1443–1447.

[9] P. Liu, S. Wu, and Y. Bar-Ness, "A phase noise mitigation scheme for MIMO WLANs with spatially correlated and imperfectly estimated channels," *IEEE Commun. Lett.*, vol. 10, no. 3, pp. 141–143, Mar. 2006.
[10] Y. Zhang and H. Liu, "MIMO-OFDM systems in the presence of phase noise and doubly selective fading," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 2277–2285, Jul. 2007.
[11] S. Bittner, E. Zimmermann, and G. Fettweis, "Exploiting phase noise properties in the design of MIMO-OFDM receivers," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2008, pp. 940–945.
[12] S. Bittner, E. Zimmermann, and G. Fettweis, "Iterative phase noise mitigation in MIMO-OFDM systems with pilot aided channel estimation," in *Proc. IEEE 66th Veh. Technol. Conf.*, Sep. 2007, pp. 1087–1091.
[13] 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, Sep. 2012.
[14] N. Hadaschik *et al.*, "Improving MIMO phase noise estimation by exploiting spatial correlations," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 3. Mar. 2005, pp. iii-833–iii-836.
[15] F. Bohagen, P. Orten, and G. E. Oien, "Design of optimal high-rank line-of-sight MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1420–1425, Apr. 2007.
[16] E. Björnson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and Circuit-aware design," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4353–4368, Aug. 2015.
[17] P. Xu, Y. Xiao, S. Zhou, and M. Zhao, "ICI analysis of MIMO-OFDM systems with independent phase noise at both transmit and receive antennas," in *Proc. 5th Int. Conf. Wireless Commun. Netw. Mobile Comput.*, Sep. 2009, pp. 1–4.
[18] A. Tarable, G. Montorsi, S. Benedetto, and S. Chinnici, "An EM-based phase-noise estimator for MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 3215–3219.
[19] H. Huang, W. G. J. Wang, and J. He, "Phase noise and frequency offset compensation in high frequency MIMO-OFDM system," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 1280–1285.
[20] A. Ishaque and G. Ascheid, "Efficient MAP-based estimation and compensation of phase noise in MIMO-OFDM receivers," *AEU-Int. J. Electron. Commun.*, vol. 67, no. 12, pp. 1096–1106, 2013.
[21] Q. Zou, A. Tarighat, and A. H. Sayed, "Compensation of phase noise in OFDM wireless systems," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5407–5424, Nov. 2007.
[22] S. Wu, P. Liu, and Y. Bar-Ness, "Phase noise estimation and mitigation for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3616–3625, Dec. 2006.
[23] D. D. Lin, R. A. 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, Sep. 2006.
[24] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615–1624, Jun. 2003.
[25] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158–1168, May 2006.
[26] H. Minn, N. Al-Dhahir, and Y. Li, "Optimal training signals for MIMO OFDM channel estimation in the presence of frequency offset and phase noise," *IEEE Trans. Commun.*, vol. 54, no. 10, pp. 1754–1759, Oct. 2006.
[27] R. Wang, H. Mehrpouyan, M. Tao, and Y. Hua, "Channel estimation, carrier recovery, and data detection in the presence of phase noise in OFDM relay systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1186–1205, Feb. 2016.
[28] J. Yang, B. Geller, and S. Bay, "Bayesian and hybrid Cramér–Rao bounds for the carrier recovery under dynamic phase uncertain channels," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 667–680, Feb. 2011.
[29] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
[30] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed. Philadelphia, PA, USA: SIAM, 2002. [Online]. Available: http://dx.doi.org/10.1137/1.9780898718027
[31] T. J. Lee and Y.-C. Ko, "Phase noise mitigation in MIMO-OFDM systems with independent oscillators," presented at the IEEE Conf. Standards Commun. Netw. (CSCN), Oct./Nov. 2016.
[32] M. K. Ozdemir and H. Arslan, "Channel estimation for wireless ofdm systems," *IEEE Commun. Surveys Tuts.*, vol. 9, no. 2, pp. 18–48, 2nd Quart., 2007.

TAE-JUN LEE received the B.S. degree in electrical engineering from Korea University, Seoul, South Korea, in 2015. His current research interests include channel estimation and phase noise cancellation techniques for mm-wave communication systems.

YOUNG-CHAI KO received the B.Sc. degree in electrical and telecommunication engineering from Hanyang University, Seoul, South Korea, and the M.S.E.E. and Ph.D. degrees in electrical engineering from the University of Minnesota, Minneapolis, MN, in 1999 and 2001, respectively. He was with Novatel Wireless as a Research Scientist in 2001. In 2001, he joined the Texas Instruments, Inc., Wireless Center, San Diego, CA, as a Senior Engineer. He is currently a Professor with the School of Electrical Engineering, Korea University. His current research interests are the performance analysis and the design of wireless communication systems.