2-1-2016

# Channel Estimation, Carrier Recovery, and Data Detection in the Presence of Phase Noise in OFDM Relay Systems

Rui Wang
*Tongji University*

Hani Mehrpouyan
*Boise State University*

Meixia Tao
*Shanghai Jiao Tong University*

Yingbo Hua
*University of California*

 doi: 10.1109/TWC.2015.2487268

# Channel Estimation, Carrier Recovery, and Data Detection in the Presence of Phase Noise in OFDM Relay Systems

Rui Wang, Hani Mehrpouyan, *Member, IEEE*, Meixia Tao, *Senior Member, IEEE*, and Yingbo Hua, *Fellow, IEEE*

***Abstract*—Due to its time-varying nature, oscillator phase noise can significantly degrade the performance of channel estimation, carrier recovery, and data detection blocks in high-speed wireless communication systems. In this paper, we analyze joint channel, *carrier frequency offset (CFO)*, and phase noise estimation plus data detection in *orthogonal frequency division multiplexing (OFDM)* relay systems. To achieve this goal, a detailed transmission framework involving both training and data symbols is presented. In the data transmission phase, a comb-type OFDM symbol consisting of both pilots and data symbols is proposed to track phase noise over an OFDM frame. Next, a novel algorithm that applies the training symbols to jointly estimate the channel responses, CFO, and phase noise based on the maximum a posteriori criterion is proposed. Additionally, a new *hybrid Cramér-Rao lower bound* for evaluating the performance of channel estimation and carrier recovery algorithms in OFDM relay networks is derived. Finally, an iterative receiver for joint phase noise estimation and data detection at the destination node is derived. Extensive simulations demonstrate that the application of the proposed estimation and receiver blocks significantly improves the performance of OFDM relay networks in the presence of phase noise.**

***Index Terms*—Relay networks, amplify-and-forward (AF), hybrid Cramér-Rao lower bound (HCRLB), orthogonal frequency division multiplexing (OFDM), channel estimation, carrier frequency offset, phase noise, receiver design.**

## I. Introduction

### A. Motivation and Literature Survey

Application of relaying has been identified as a suitable approach for combating long-distance channel distortion and small-scale fading in wireless communication systems [1]. Various physical layer techniques, such as distributed space-time block coding [2], precoding [3], etc., for relay systems have been extensively studied in the past decade. From these works, it can be deduced that to deliver the advantages of relay networks, the network's *channel state information (CSI)* needs to be accurately obtained [4]–[8], while the negative impact of impairments such as *carrier frequency offset (CFO)* and *phase noise (PN)* caused by Doppler shifts and oscillator imperfections needs to be mitigated [9].

In single carrier communication systems, CFO and PN are multiplicative and result in a rotation of the signal constellation from symbol to symbol and erroneous data detection [10], [11]. On the other hand, in the case of *orthogonal frequency division multiplexing (OFDM)* systems, CFO and PN are convolved with the data symbols, resulting in the rotation of the signal constellation and *inter-carrier interference (ICI)*, which can significantly deteriorate the overall performance of an OFDM system [12]–[14]. Thus, extensive research has been recently carried out to find carrier recovery schemes that complement traditional approaches, e.g., those based on the *phase-locked loop (PLL)*. More importantly, as demonstrated in [12], [15], to accurately obtain the channel, CFO, and PN parameters in communications systems, these parameters need to be jointly estimated. However, the prior art on channel and CFO estimation in relay networks has not taken into consideration the detrimental impact of PN.

Due to the presence of multiple hops between source and destination, channel estimation in relay systems is quite different from traditional point-to-point systems. For the amplify-and-forward (AF) relaying strategy, one approach is to only estimate source to destination channels [4], [5]. However, to further enhance cooperative system performance by enabling relay precoding/beamforming or relay resource allocation, the channel response of each hop needs to be separately estimated [5]–[8]. Furthermore, since the channel response from relay to destination affects the destination noise covariance matrix, estimating individual channel responses is generally required for more accurate signal detection at the destination. It is worth noting that the contributions in [4]–[8] only focus on channel estimation while ignoring the effect of CFO and PN.

Joint estimation of the channel responses and CFO in single carrier relay systems has been considered in [15], [16]. In [16], the Gauss-Hermite integration and approximate Rao-Blackwellization based joint CFO and channel estimators are proposed, while in [15] joint CFO and channel estimation via the MUSIC algorithm is analyzed. However, the works in [15], [16] ignore the effect of PN. In fact, although both CFO and PN result in an unknown rotation of signal constellation, PN is a time-varying parameter compared to the CFO and can be more difficult to estimate [11], [12]. More importantly, the negative impact of CFO and PN may be greater in the case of OFDM systems compared to single carrier systems [17], [18].

Due to its capability of combating frequency selectivity in the wireless channel, OFDM techniques have been extensively adopted in the latest wireless communication standards, e.g., Long Term Evolution, IEEE 802.11n, Bluetooth, etc. The deteriorating effect of PN on the performance of point-to-point OFDM systems is analyzed in [17], [18]. Undoubtedly, this effect can also be observed in OFDM based corporative relay systems. Hence, conducting accurate channel and CFO estimation in the presence of PN is important for maintaining the quality of service in high-speed OFDM relay networks. Joint estimation of CFO and channel in OFDM relay systems

R. Wang and M. Tao are with the Department of Electronic Engineering at Shanghai Jiao Tong University, Shanghai, P. R. China. Emails:{liouxingrui, mxtao}@sjtu.edu.cn. H. Mehrpouyan is with the Department of Computer and Electrical Engineering and Computer Science at California State University, Bakersfield, CA, USA. Email: hani.mehr@ieee.org. Y. Hua is with the Department of Electrical Engineering at the University of California, Riverside, CA, USA. Email: hua@ee.ucr.edu.

is considered in [19], [20]. In particular, a two-time-slot cooperative estimation protocol has been proposed in [19] for OFDM relay systems, while in [20] the authors studied the maximum likelihood (ML) based, and the least squares based, joint CFO and channel estimation algorithms. However, none of the approaches in [19], [20] consider the effect of PN on channel and CFO estimation or the overall relaying performance. While ignoring the effect of CFO, joint channel and PN estimation in OFDM relay networks is analyzed in [21]. Although the approach in [21] can be applied to AF relaying systems, it requires the relay to remove the cyclic prefix (CP) corresponding to the source-to-relay link and add a new CP before forwarding the OFDM symbol. Such an approach can result in significant additional overhead at the relay. Moreover, none of the approaches in [19]–[21] consider the effect of PN on joint channel and CFO estimation.

### B. Contributions

In this paper, different from [19], [20], the problem of joint CFO, PN, and channel estimation in OFDM relay systems is considered. Although joint CFO, PN, and channel estimation has been studied for point-to-point OFDM systems [12], [22], [23], to the best of the authors' knowledge, this problem has not been considered in the context of relay systems. The contributions of this paper can be summarized as follows:

- A training and data transmission framework for OFDM relay networks is proposed that enables joint estimation of channel, CFO, and PN parameters at the destination.
- A new *hybrid Cramer-Rao lower bound (HCRLB)* for analyzing the performance of joint channel, CFO, and PN estimators in OFDM relay networks is derived.
- An iterative joint channel, CFO, and PN estimator based on the *maximum a posteriori (MAP)* criterion is proposed that exploits the correlation between PN parameters to significantly reduce estimation overhead.[1] Moreover, the estimator's mean square error (MSE) performance is shown to be close to the derived HCRLB at moderate signal-to-noise ratios (SNRs).
- A comb-type OFDM symbol containing both pilots and data symbols is proposed to track the time-varying PN parameters during the data transmission interval. Next, a novel iterative receiver that applies the proposed OFDM symbol to perform joint data detection and PN tracking at the destination node is derived.
- Extensive simulations are carried out to investigate the performance of an OFDM relay system in the presence of CFO and PN. The results show that the combination of the proposed joint estimator and iterative receiver greatly enhances the bit error rate (BER) performance of OFDM relay systems with imperfect knowledge of channels, CFO, and PN.

### C. Organization

Section II presents the system model and assumptions in this paper. The joint estimation algorithm is presented in Section III. In Section IV, the HCRLB for the proposed joint estimation problem is derived. The proposed iterative receiver for joint data detection and PN tracking is present in Section V. Extensive simulation results are illustrated in Section VI. Finally, we conclude the paper in Section VII.

[1]The approach proposed here can be also applied to point-to-point systems to reduce PN estimation and carrier recovery overhead.

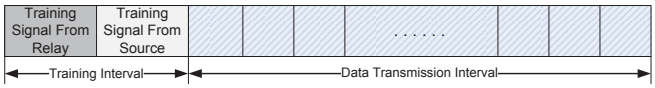


Fig. 1. The proposed timing diagram of the OFDM relay system.

### D. Notations

Small italic letters, e.g., $x$ are for scalars, bold face small letters, e.g., $\mathbf{x}$, are used for vectors, and bold face capital alphabets, e.g., $\mathbf{X}$, are used for matrices. $\hat{x}$ is used to denote the estimate of $x$. $\mathbb{E}(\cdot)$ denotes the expectation of its argument. $\odot$, $\star$, and $*$ denote the Hadamard product, linear, and circular convolutions, respectively. $\mathrm{Tr}(\mathbf{A})$, $\mathbf{A}^{-1}$, and $\det(\mathbf{A})$ denote the trace, inverse, and determinant of matrix $\mathbf{A}$, respectively. $\mathrm{Diag}(\mathbf{a})$ denotes a diagonal matrix with $\mathbf{a}$ being its diagonal entries. $\mathrm{Blkdiag}(\mathbf{A}_0, \mathbf{A}_1, \cdots, \mathbf{A}_{N-1})$ denotes a block diagonal matrix with $\mathbf{A}_0, \mathbf{A}_1, \cdots, \mathbf{A}_{N-1}$ as its diagonal matrices. $\mathbf{A}(N:M,:)$ and $\mathbf{A}(:,N:M)$ denote a submatrix containing the $N$-th to $M$-th rows of $\mathbf{A}$ and a submatrix containing the $N$-th to $M$-th columns of $\mathbf{A}$, respectively. Superscripts $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$ denote the transpose, conjugate, and conjugate transpose, respectively. $\mathbf{0}_{N\times M}$, $\mathbf{I}_N$, and $\mathbf{1}_N$ denote the $N\times M$ zero matrix, $N\times N$ identity matrix, and $N\times 1$ vector of ones, respectively. $\Re(z)$ and $\Im(z)$ denote the real and imaginary operators. $\mathbb{C}^{x\times y}$ and $\mathbb{R}^{x\times y}$ denote spaces of $x\times y$ matrices with complex and real entries, respectively. $\triangle^{\mathbf{x}}_{\mathbf{x}} f(\cdot) \triangleq \frac{\partial f}{\partial \mathbf{x}}[\frac{\partial f}{\partial \mathbf{x}} f(\cdot)]^T$ denotes the second order partial derivative of function $f(\cdot)$ with respect to vector $\mathbf{x}$. Finally, $\mathcal{CN}(\mathbf{x}, \mathbf{\Sigma})$ and $\mathcal{N}(\mathbf{x}, \mathbf{\Sigma})$ denote real and complex Gaussian distributions, respectively, with mean $\boldsymbol{\mu}$ and covariance $\mathbf{\Sigma}$.

## II. System Model

An AF relaying OFDM system is considered, where a source node transmits its signal to a destination node through a relay. Unlike the work in [21], it is assumed that the relay node simply forwards the received signal without removing the CP corresponding to the source-to-relay link and appending a new CP for the relay-to-destination link. This assumption ensures a considerably simpler relaying structure. $N$ subcarriers are used for OFDM transmission. Similar to prior work in this field, e.g., [12], [13], quasi static fading channels are considered, i.e., the CSI is assumed to be constant over the duration of a single packet. Each packet consists of two OFDM training symbols, which are followed by multiple data symbols as shown in Fig. 1. The two training symbols are used to separately estimate the channel responses and CFO in the presence of unknown PN for both the source to relay and relay to destination hops (Fig. 1).

The proposed signal model can be applied to both full-duplex and half-duplex relaying networks based on the following system setups and assumptions:

- *Full-duplex relaying*: In this setup, the proposed signal model is applicable to relaying networks that utilize

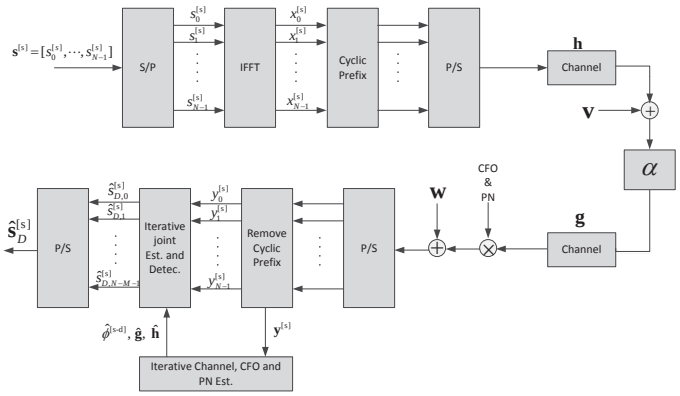


Fig. 2. Illustration of transceiver structure of the OFDM relay system.

highly directional transmit and receive antennas with large antenna gains at the relay, e.g., microwave and millimeter-wave systems [24], [25]. This approach minimizes or eliminates the effect of self-interference at the relay.[2] Moreover, it is assumed that the relay forwards its signal to the destination in passband without converting it to baseband. This assumption is practical since there are various radio frequency (RF) amplifiers that can operate at high carrier frequencies and can be utilized in full-duplex relaying networks, e.g., Mini-Circuits AVA-24+ with a frequency range of 5–20 GHz [27].

- *Half-duplex relaying*: In this setup, it is assumed that the relay forwards its received signal on a different carrier frequency and does not convert it to baseband, i.e., the relay applies on-frequency/on-channel RF relaying [28]. Moreover, the difference between the receive and transmit carrier frequencies are assumed to be small to enable the application of a low PN oscillator at the relay. An example of such an oscillator is ROS-209-319+ ultra low noise voltage controlled oscillator that has a very small PN factor of $-133$ dBc/Hz at an offset frequency of 10 KHz [29]. As such, in this setup, it is assumed that the signal forwarded from the relay is not affected by PN.

### A. Signal Transmission from Source to Destination

The overall transmission and reception structure of each OFDM symbol from the source to the destination node is illustrated in Fig. 2. Let $\mathbf{s}^{[\mathrm{s}]} \triangleq \left[s_0^{[\mathrm{s}]}, s_1^{[\mathrm{s}]}, \cdots, s_{N-1}^{[\mathrm{s}]}\right]^T$ denote the frequency domain modulated training or data signal sequence at the source node, which is then transformed into a set of parallel symbols $s_k^{[\mathrm{s}]}$, for $k = 0, \cdots, N-1$. By conducting an inverse fast Fourier transform (IFFT), we obtain the time domain signal vector $\mathbf{x}^{[\mathrm{s}]}$ as $\mathbf{x}^{[\mathrm{s}]} = \mathbf{F}^H \mathbf{s}^{[\mathrm{s}]}$, where $\mathbf{x}^{[\mathrm{s}]} \triangleq \left[x_0^{[\mathrm{s}]}, x_1^{[\mathrm{s}]}, \cdots, x_{N-1}^{[\mathrm{s}]}\right]^T$, and $\mathbf{F}$ is the normalized discrete Fourier transform (DFT) matrix with $\mathbf{F}_{n,k} = \frac{1}{\sqrt{N}} \exp\left(-j\frac{2\pi(n-1)(k-1)}{N}\right)$. After adding the CP, the parallel signal vector is transformed into a time domain sequence denoted by $x^{[\mathrm{s}]}(n)$, for $n = -L, \cdots, N-1$. Subsequently, the transmitted baseband continuous signal from the source, $\tilde{x}^{[\mathrm{s}]}(t)$, can be written as

$$\tilde{x}^{[\mathrm{s}]}(t) = \sum_{n=-L}^{N-1} x^{[\mathrm{s}]}(n) q(t - nT_s), \qquad 0 \le t \le T + T_{CP} \quad (1)$$

where $T_s = T/N$ with $T$ denoting an OFDM symbol duration, $q(t)$ is the pulse shaping filter, $T_{CP}$ is the duration of the CP, and $x^{[\mathrm{s}]}(-n) = x^{[\mathrm{s}]}(N-n)$, for $n = -L, \cdots, -1$, is the added CP symbol.

At the destination, the baseband received signal, $y^{[\mathrm{s}]}(t)$, is given by

$$\begin{aligned} y^{[\mathrm{s}]}(t) =& \alpha e^{j\theta^{[\mathrm{s\text{-}d}]}(t)} e^{j\phi^{[\mathrm{s\text{-}d}]}(t)} \big[ g(t) \star h(t) \star \tilde{x}^{[\mathrm{s}]}(t) \\ &+ g(t) \star v(t) \big] + w(t), \end{aligned} \quad (2)$$

where $\alpha$ is the constant and scalar amplification factor at the relay, $h(t)$ and $g(t)$ are the frequency-selective fading channels from source to relay and the relay to destination, respectively, and $v(t)$ and $w(t)$ are the additive noises at the relay and at the destination, respectively. Using a similar approach to point-to-point systems [12], [22], [23], in (2), $\theta^{[\mathrm{s\text{-}d}]}(t)$ is the PN corresponding to source-relay-destination link, while $\phi^{[\mathrm{s\text{-}d}]}(t) \triangleq 2\pi \triangle f^{[\mathrm{s\text{-}d}]} t$ is the CFO caused by the unmatched source and destination carrier frequencies.

After sampling at a sampling rate of $1/T_s$ and removing the CP, the received signal at the destination is determined as

$$\begin{aligned} y^{[\mathrm{s}]}(nT_s) =& \alpha e^{j\theta^{[\mathrm{s\text{-}d}]}(nT_s)} e^{j2\pi\Delta f^{[\mathrm{s\text{-}d}]} nT_s} \big[ \underbrace{g(nT_s) \star h(nT_s)}_{\triangleq c(nT_s)} \\ &* \tilde{x}^{[\mathrm{s}]}(nT_s) + g(nT_s) \star v(nT_s) \big] + w(nT_s), \end{aligned} \quad (3)$$

where circular convolution appears in (3) due to the added CP at the source node. Note that to avoid ICI, the length of CP, denoted by $N_{CP} = T_{CP}/T_s$, should be larger than $L = L_h + L_g - 1$ with $L_h$ and $L_g$ being the number of channel taps of $h(t)$ and $g(t)$, respectively. Eq. (3) can be written in vector form as[3]

$$\mathbf{y}^{[\mathrm{s}]} = \alpha \mathbf{\Lambda}_{\boldsymbol{\theta}^{[\mathrm{s\text{-}d}]}} \mathbf{\Lambda}_{\phi^{[\mathrm{s\text{-}d}]}} \left[ \mathbf{C}\mathbf{x}^{[\mathrm{s}]} + \mathbf{G}\mathbf{v} \right] + \mathbf{w}, \quad (4)$$

where

- $\mathbf{y}^{[\mathrm{s}]} \triangleq [y^{[\mathrm{s}]}(0), y^{[\mathrm{s}]}(1), \cdots, y^{[\mathrm{s}]}(N-1)]^T$,
- $\mathbf{\Lambda}_{\boldsymbol{\theta}^{[\mathrm{s\text{-}d}]}} \triangleq \mathrm{Diag}\big[e^{j\theta^{[\mathrm{s\text{-}d}]}(0)}, e^{j\theta^{[\mathrm{s\text{-}d}]}(1)}, \cdots, e^{j\theta^{[\mathrm{s\text{-}d}]}(N-1)}\big]$,
- $\mathbf{\Lambda}_{\phi^{[\mathrm{s\text{-}d}]}} \triangleq \mathrm{Diag}\big[1, e^{j2\pi\phi^{[\mathrm{s\text{-}d}]}/N}, \cdots, e^{j2\pi\phi^{[\mathrm{s\text{-}d}]}(N-1)/N}\big]$, $\phi^{[\mathrm{s\text{-}d}]} \triangleq \Delta f^{[\mathrm{s\text{-}d}]} T$ is the normalized CFO,
- $\mathbf{C} \triangleq \mathbf{F}^H \mathbf{\Lambda}_{\tilde{\mathbf{c}}} \mathbf{F}$, $\mathbf{\Lambda}_{\tilde{\mathbf{c}}} \triangleq \mathrm{Diag}(\tilde{\mathbf{c}})$ with $\tilde{\mathbf{c}} \triangleq \sqrt{N}\mathbf{F}[\mathbf{c}^T, \mathbf{0}_{N-L,1}^T]^T$ and $\mathbf{c} \triangleq [c(0), c(1), \cdots, c(L-1)]^T$,
- $\mathbf{v} \triangleq [v(-L_g+1), \cdots, v(0), \cdots, v(N-1)]^T$ and $\mathbf{w}_1 \triangleq [w(0), w(1), \cdots, w(N-1)]^T$ are the sampled additive noise at the relay and destination nodes, respectively, and

$$\mathbf{G} = \begin{bmatrix} g(L_g-1) & g(L_g-2) & \cdots & 0 \\ 0 & g(L_g-1) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g(0) \end{bmatrix} \quad (5)$$

is an $N \times (N + L_g - 1)$ matrix. The additive noise at the relay and destination are distributed as $\mathbf{v} \sim \mathcal{CN}(0, \sigma_R^2 \mathbf{I}_{N+N_{CP}})$ and $\mathbf{w} \sim \mathcal{CN}(0, \sigma_D^2 \mathbf{I}_N)$, respectively. Finally, although $\mathbf{C}$ is

[2] Application of sophisticated transceivers has also been shown to minimize or eliminate the impact of self-interference at the relay [26].

[3] For notational convenience, we discard the term $T_s$ in (4).

an $N \times N$ circulant matrix, $\mathbf{G}$ is a regular $N \times (N + L_g - 1)$ matrix, since no CP is added at the relay node.

### *B. Training Signal Transmission from Relay to Destination*

Recall that the second OFDM training symbol is transmitted from the relay to separately estimate the relay-to-destination channel. Following similar steps as above, the vector of received training signal at the destination node from the relay, $\mathbf{y}^{[\text{r}]} \triangleq [y^{[\text{r}]}(0), y^{[\text{r}]}(1), \cdots, y^{[\text{r}]}(N-1)]^T$, is given by

$$\mathbf{y}^{[\text{r}]} = \mathbf{\Lambda}_{\boldsymbol{\theta}^{[\text{r-d}]}} \mathbf{\Lambda}_{\phi^{[\text{r-d}]}} \bar{\mathbf{G}} \mathbf{x}^{[\text{r}]} + \mathbf{w}, \tag{6}$$

where

- $\mathbf{x}^{[\text{r}]} \triangleq [x^{[\text{r}]}(0), x^{[\text{r}]}(1), \cdots, x^{[\text{r}]}(N-1)]^T = \mathbf{F}^H \mathbf{s}^{[\text{r}]}$, $\mathbf{s}^{[\text{r}]}$ is the frequency domain relay training signal,
- $\mathbf{\Lambda}_{\boldsymbol{\theta}^{[\text{r-d}]}} \triangleq \text{Diag}\left[e^{j\theta^{[\text{r-d}]}(0)}, e^{j\theta^{[\text{r-d}]}(1)}, \cdots, e^{j\theta^{[\text{r-d}]}(N-1)}\right]$, $\theta^{[\text{r-d}]}(n)$ is the $n$-th PN sample corresponding to relay-destination link,
- $\mathbf{\Lambda}_{\phi^{[\text{r-d}]}} \triangleq \text{Diag}\left[1, e^{j2\pi\phi^{[\text{r-d}]}/N}, \cdots, e^{j2\pi\phi^{[\text{r-d}]}(N-1)/N}\right]$,
- $\phi^{[\text{r-d}]}$ is the normalized CFO generated by the mismatch between the relay and destination carrier frequencies,
- $\bar{\mathbf{G}}$ is a circulant channel matrix given by $\bar{\mathbf{G}} \triangleq \mathbf{F}^H \mathbf{\Lambda}_{\tilde{\mathbf{g}}} \mathbf{F}$, with $\mathbf{\Lambda}_{\mathbf{g}} = \text{Diag}(\tilde{\mathbf{g}})$, $\tilde{\mathbf{g}} \triangleq \sqrt{N}\mathbf{F}[\mathbf{g}^T, \mathbf{0}_{N-L_g,1}^T]^T$, and $\mathbf{g} \triangleq [g(0), g(1), \cdots, g(L_g-1)]^T$.

### *C. Statistical Model of Phase Noise*

Similar to [12] and based on the properties of PN in practical oscillators, PN is modeled by a Wiener process, i.e.,

$$\theta^{[i]}(n) = \theta^{[i]}(n-1) + \Delta^{[i]}(n), \qquad i = [\text{s-d}], [\text{r-d}] \tag{7}$$

where $\Delta^{[i]}(n-1)$ is a real Gaussian variable following $\Delta^{[i]}(n) \sim \mathcal{N}(0, \sigma^2_{\Delta^{[i]}})$. Here $\sigma^2_{\Delta^{[i]}} = 2\pi\beta^{[i]}T_s$ with $\beta^{[i]}$ denoting the one-sided 3-dB bandwidth of the Lorentzian spectrum of the oscillator [30], [31]. As in [12], [13], it is assumed that $\theta^{[i]}(-1) = 0$ since the residual PN at the start of the frame is estimated as part of the channel parameters. From (7), it can be concluded that the PN vector, $\boldsymbol{\theta}^{[i]} \triangleq [\theta^{[i]}(0), \theta^{[i]}(1), \cdots, \theta^{[i]}(N-1)]^T$, follows a Gaussian distribution, i.e., $\boldsymbol{\theta}^{[i]} \sim \mathcal{N}(0, \mathbf{\Psi}^{[i]})$, where the covariance matrix $\mathbf{\Psi}^{[i]}$ is given by

$$\mathbf{\Psi}^{[i]} = \sigma^2_{\Delta^{[i]}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 2 & 2 & \cdots & 2 & 2 \\ 1 & 2 & 3 & 3 & \cdots & 3 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & 3 & \cdots & N-1 & N \end{bmatrix}. \tag{8}$$

In obtaining the covariance matrix in (8), similar to prior results in this field [13], it is assumed that the PN variances are small enough such that $\theta^{[i]}(n)$ does not reach its maximum value of $\pi$. This assumption is justifiable since practical oscillators have a very small PN variance as shown in [32].

Based on the signal model in (4) and (6), it can be observed that a large number of channel, CFO, and PN parameters need to be jointly estimated, which increases the computational complexity of the receiver at the destination. Thus, to reduce estimation overhead, we take advantage of the correlation amongst the PN parameters to reduce the number of unknown parameters that need to be estimated. The eigenvalues of the covariance matrix, $\mathbf{\Psi}^{[i]}$ are illustrated in Fig. 3. From this figure it can be deduced that most eigenvalues of the matrix $\mathbf{\Psi}^{[i]}$ are close to zero. Thus, the PN vector, $\boldsymbol{\theta}^{[i]}$, can be represented as

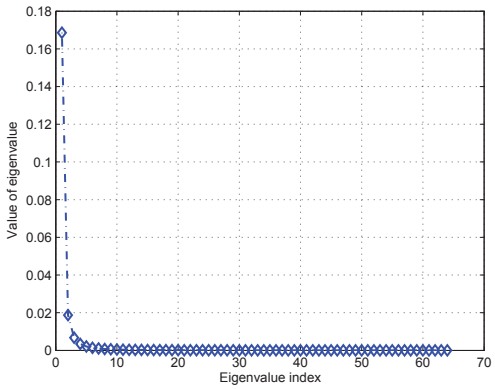


Fig. 3. Illustration of the eigenvalues of $\mathbf{\Psi}^{[i]}$ at $N = 64$ and $\sigma^2_{\Delta^{[i]}} = 10^{-4}$.

$$\boldsymbol{\theta}^{[i]} = \mathbf{\Pi}^{[i]} \boldsymbol{\eta}^{[i]}, \qquad i = [\text{s-d}], [\text{r-d}] \tag{9}$$

where $\boldsymbol{\eta}^{[i]} \sim \mathcal{N}(0, \mathbf{I}_M) \in \mathbb{C}^{M \times 1}$ is the shortened unknown PN vector containing $M \leq N$ PN parameters, while $\mathbf{\Pi}^{[i]} \in \mathbb{C}^{N \times M}$ is the transformation matrix that allows for obtaining $\boldsymbol{\eta}^{[i]}$ from $\boldsymbol{\theta}^{[i]}$. Moreover, the singular value decomposition of $\mathbf{\Psi}^{[i]}$ is given by $\mathbf{\Psi}^{[i]} = \mathbf{U}^{[i]}\mathbf{D}^{[i]}\left(\mathbf{U}^{[i]}\right)^T$, where $\mathbf{U}^{[i]}$ is the $N \times N$ eigenvector matrix of $\mathbf{\Psi}^{[i]}$ and $\mathbf{D}^{[i]} = \text{Diag}(\boldsymbol{\nu}^{[i]})$. Here, $\boldsymbol{\nu}^{[i]} \triangleq [\nu_{i,0}, \nu_{i,1}, \cdots, \nu_{i,N-1}]^T$ is the vector of the eigenvalues of $\mathbf{\Psi}^{[i]}$ arranged in decreasing order. Subsequently, the matrix $\mathbf{\Pi}^{[i]}$ in (9) can be selected as $\mathbf{\Pi}^{[i]} = \tilde{\mathbf{U}}^{[i]}\tilde{\mathbf{D}}^{[i]}$, where $\tilde{\mathbf{U}}^{[i]} = \mathbf{U}^{[i]}(:, 0 : M-1)$ and $\tilde{\mathbf{D}}^{[i]} = \text{Diag}(\tilde{\boldsymbol{\nu}}^{[i]})$ with $\tilde{\boldsymbol{\nu}}^{[i]} \triangleq \left[\sqrt{\nu_{i,0}}, \sqrt{\nu_{i,1}}, \cdots, \sqrt{\nu_{i,M-1}}\right]^T$. In the subsequent sections, $\boldsymbol{\eta}^{[i]}$, for $i = [\text{s-d}], [\text{r-d}]$ is estimated instead of $\boldsymbol{\theta}^{[i]}$. A suitable choice of $M$ that allows for accurate PN tracking is presented in Section VI.

## III. Proposed Joint Channel, CFO and Phase Noise Estimation

In order to avoid the negative impact of ICI caused by CFO and PN, in this work, the joint estimation of channel parameters, CFO, and PN is performed in the time domain. To proceed, we reformulate (4) as

$$\mathbf{y}^{[\text{s}]} = \alpha \mathbf{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}} \mathbf{\Lambda}_{\phi^{[\text{s-d}]}} \left(\mathbf{F}^H \mathbf{\Lambda}_{s^{[\text{s}]}} \mathbf{F}_{[L]} \mathbf{c} + \mathbf{G}\mathbf{v}\right) + \mathbf{w}, \tag{10}$$

where $\mathbf{s}^{[\text{s}]}$ denotes the training symbol transmitted from source such that $\mathbb{E}\left(\mathbf{s}^{[\text{s}]}\left[\mathbf{s}^{[\text{s}]}\right]^H\right) = P_{\text{T}}^{[\text{s}]}\mathbf{I}_N$, $P_{\text{T}}^{[\text{s}]}$ is the transmit power from the source, $\mathbf{\Lambda}_{s^{[\text{s}]}} \triangleq \text{Diag}(\mathbf{s}^{[\text{s}]})$, and $\mathbf{F}_{[L]} \triangleq \sqrt{N}\mathbf{F}(:, 0 : L-1)$. Similarly, the received signal $\mathbf{y}^{[\text{r}]}$ in (6) can be rewritten as

$$\mathbf{y}^{[\text{r}]} = \mathbf{\Lambda}_{\boldsymbol{\theta}^{[\text{r-d}]}} \mathbf{\Lambda}_{\phi^{[\text{r-d}]}} \mathbf{F}^H \mathbf{\Lambda}_{s^{[\text{r}]}} \mathbf{F}_{[L_g]} \mathbf{g} + \mathbf{w}, \tag{11}$$

where $\mathbf{\Lambda}_{s^{[\text{r}]}} \triangleq \text{Diag}(\mathbf{s}^{[\text{r}]})$ denotes the training symbol transmitted from relay such that $\mathbb{E}(\mathbf{s}^{[\text{r}]}\left[\mathbf{s}^{[\text{r}]}\right]^H) = P_{\text{T}}^{[\text{r}]}\mathbf{I}_N$, $P_{\text{T}}^{[\text{r}]}$ is the transmit power from the relay, and $\mathbf{F}_{[L_g]} \triangleq \sqrt{N}\mathbf{F}(:, 0 : L_g - 1)$. As in [12], it is assumed that $\mathbf{s}^{[\text{s}]}$ and $\mathbf{s}^{[\text{r}]}$ are known constant-modulus training symbols.

From the detection point of view, it may appear that one only needs to estimate the CFO, $\phi^{[\text{s-d}]}$, and the *combined* source-relay-destination channel, $\mathbf{c}$, in the presence of PN, $\boldsymbol{\theta}^{[\text{s-d}]}$. However, as shown in (10), the relay-to-destination channel, $\mathbf{g}$, affects the statistic of the additive noise at the destination node. Hence, it needs to be known to develop a joint PN estimation and data detection algorithm based on the MAP criterion. Consequently, here, the parameters of interest are: the CFO, $\phi^{[\text{s-d}]}$, the channel from source to relay, $\mathbf{h} \triangleq [h(0), h(1), \cdots, h(L_h-1)]^T$ and the relay to destination channel, $\mathbf{g}$. Moreover, in addition to the parameters of interest, there are also unknown nuisance parameters, e.g., the CFO and PN from relay to destination, $\phi^{[\text{r-d}]}$ and $\boldsymbol{\theta}^{[\text{r-d}]}$, respectively, that also need to be jointly estimated. Using the approach in [12] and the received training signal from the relay node, $\mathbf{y}^{[\text{r}]}$, the MAP estimates of the CFO from relay to destination, $\phi^{[\text{r-d}]}$, can be obtained as

$$\begin{aligned}\hat{\phi}^{[\text{r-d}]} =& \arg\min_{\phi^{[\text{r-d}]}} -\mathbf{1}_N^T \Im(\boldsymbol{\Lambda}_{\phi^{[\text{r-d}]}} \mathbf{A}\mathbf{A}^H \boldsymbol{\Lambda}_{\phi^{[\text{r-d}]}}^H)^T \\ &\times \left[\Re(\boldsymbol{\Lambda}_{\phi^{[\text{r-d}]}} \mathbf{A}\mathbf{A}^H \boldsymbol{\Lambda}_{\phi^{[\text{r-d}]}}^H) + \frac{\sigma^2 P_{\text{T}}^{[\text{r}]}}{2} [\boldsymbol{\Psi}^{[\text{r-d}]}]^{-1}\right]^{-1} \\ &\times \Im(\boldsymbol{\Lambda}_{\phi^{[\text{r-d}]}} \mathbf{A}\mathbf{A}^H \boldsymbol{\Lambda}_{\phi^{[\text{r-d}]}}^H) \mathbf{1}_N + \mathbf{1}_N^T \boldsymbol{\Lambda}_{\phi^{[\text{r-d}]}} \mathbf{A}\mathbf{A}^H \boldsymbol{\Lambda}_{\phi^{[\text{r-d}]}}^H \mathbf{1}_N,\end{aligned} \tag{12}$$

where $\mathbf{A} \triangleq [\mathbf{Y}^{[\text{r}]}]^H \mathbf{F}^H \boldsymbol{\Lambda}_{s^{[\text{r}]}} \mathbf{V}$, $\mathbf{Y}^{[\text{r}]} \triangleq \text{Diag}(\mathbf{y}^{[\text{r}]})$, and $\mathbf{V} \triangleq \mathbf{F}(:, L_g : N-1)$. Using the estimated CFO from relay to destination, $\hat{\phi}^{[\text{r-d}]}$, the PN vector $\boldsymbol{\theta}^{[\text{r-d}]}$ is estimated as

$$\begin{aligned}\hat{\boldsymbol{\theta}}^{[\text{r-d}]} =& \boldsymbol{\Pi}^{[\text{r-d}]} \Big[ \big[\boldsymbol{\Pi}^{[\text{r-d}]}\big]^T \Re\big(\hat{\boldsymbol{\Lambda}}_{\phi^{[\text{r-d}]}} \mathbf{A}\mathbf{A}^H \hat{\boldsymbol{\Lambda}}_{\phi^{[\text{r-d}]}}^H\big) \boldsymbol{\Pi}^{[\text{r-d}]} \\ &+ \frac{\sigma^2 P_{\text{T}}^{[\text{r}]}}{2} \mathbf{I}_M \Big]^{-1} \big[\boldsymbol{\Pi}^{[\text{r-d}]}\big]^T \Im\big(\hat{\boldsymbol{\Lambda}}_{\phi^{[\text{r-d}]}} \mathbf{A}\mathbf{A}^H \hat{\boldsymbol{\Lambda}}_{\phi^{[\text{r-d}]}}^H\big) \mathbf{1}_N,\end{aligned} \tag{13}$$

where $[\hat{\boldsymbol{\Lambda}}_{\phi^{[\text{r-d}]}}]_{m,m} = \exp\big(\frac{j2\pi(m-1)\hat{\phi}^{[\text{r-d}]}}{N}\big)$. Unlike, the approach in [12], in (13), the shortened PN vector $\boldsymbol{\eta}^{[\text{r-d}]}$ is estimated first which reduces the complexity of the estimator by requiring the calculation of a considerably smaller matrix inverse. Based on the estimated $\hat{\phi}^{[\text{r-d}]}$ and $\hat{\boldsymbol{\theta}}^{[\text{r-d}]}$, the remaining parameters of interest can be estimated via $\mathbf{y}^{[\text{s}]}$ and $\mathbf{y}^{[\text{r}]}$.

From (10) and (11), it can be observed that the joint estimation of channel response, CFO, and PN is a hybrid estimation problem consisting of both deterministic parameters, $\phi^{[\text{s-d}]}$, $\mathbf{h}$, $\mathbf{g}$, and random parameters, $\boldsymbol{\eta}^{[\text{s-d}]}$. The posterior distribution of the parameters of interests given the received signals, $\mathbf{y}^{[\text{s}]}$ and $\mathbf{y}^{[\text{r}]}$, can be written as

$$\begin{aligned}p\big(\phi^{[\text{s-d}]}, \boldsymbol{\eta}^{[\text{s-d}]}, \mathbf{h}, \mathbf{g} | \mathbf{y}^{[\text{s}]}, \mathbf{y}^{[\text{r}]}\big) =& p\big(\mathbf{y}^{[\text{s}]}, \mathbf{y}^{[\text{r}]} | \phi^{[\text{s-d}]}, \boldsymbol{\eta}^{[\text{s-d}]}, \mathbf{h}, \mathbf{g}\big) \\ &\times p\big(\boldsymbol{\eta}^{[\text{s-d}]}\big) / p\big(\mathbf{y}^{[\text{s}]}, \mathbf{y}^{[\text{r}]}\big).\end{aligned} \tag{14}$$

Maximizing the posterior distribution in (14) is equivalent to minimizing the negative log-likelihood function (LLF) $\mathcal{L}(\phi^{[\text{s-d}]}, \boldsymbol{\eta}^{[\text{s-d}]}, \mathbf{h}, \mathbf{g}) = -\log\big(p(\mathbf{y}^{[\text{s}]}, \mathbf{y}^{[\text{r}]} | \phi^{[\text{s-d}]}, \boldsymbol{\eta}^{[\text{s-d}]}, \mathbf{h}, \mathbf{g})\big) - \log(p(\boldsymbol{\eta}^{[\text{s-d}]}))$. Our objective is to find the joint estimates of $\phi^{[\text{s-d}]}$, $\boldsymbol{\eta}^{[\text{s-d}]}$, $\mathbf{h}$, and $\mathbf{g}$ by optimizing the following unconstrained function

$$\{\hat{\phi}^{[\text{s-d}]}, \hat{\boldsymbol{\eta}}^{[\text{s-d}]}, \hat{\mathbf{h}}, \hat{\mathbf{g}}\} \propto \arg \min_{\phi^{[\text{s-d}]}, \boldsymbol{\eta}^{[\text{s-d}]}, \mathbf{h}, \mathbf{g}} \mathcal{L}(\phi^{[\text{s-d}]}, \boldsymbol{\eta}^{[\text{s-d}]}, \mathbf{h}, \mathbf{g}) \tag{15}$$

where $\mathcal{L}(\phi^{[\text{s-d}]}, \boldsymbol{\eta}^{[\text{s-d}]}, \mathbf{h}, \mathbf{g}) = \log\det(\boldsymbol{\Sigma}) + (\mathbf{y}-\boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{y}-\boldsymbol{\mu}) + \frac{1}{2}[\boldsymbol{\eta}^{[\text{s-d}]}]^T \boldsymbol{\eta}^{[\text{s-d}]}$, $\mathbf{y} \triangleq \big[[\mathbf{y}^{[\text{s}]}]^T, [\mathbf{y}^{[\text{r}]}]^T\big]^T$, $\boldsymbol{\mu} \triangleq \Big[(\alpha \boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}} \boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}} \mathbf{F}^H \boldsymbol{\Lambda}_{s^{[\text{s}]}} \mathbf{F}_{[L]} \mathbf{c})^T, (\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{r-d}]}} \boldsymbol{\Lambda}_{\phi^{[\text{r-d}]}} \mathbf{F}^H \boldsymbol{\Lambda}_{s^{[\text{r}]}} \mathbf{F}_{[L_g]} \mathbf{g})^T\Big]^T$, $\boldsymbol{\Sigma} \triangleq \text{Blkdiag}\big(\boldsymbol{\Sigma}^{[\text{r}]}, \boldsymbol{\Sigma}^{[\text{d}]}\big)$, $\boldsymbol{\Sigma}^{[\text{r}]} = \alpha^2 \sigma_R^2 \boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}} \boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}} \mathbf{G}\mathbf{G}^H \boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}^H \boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}^H + \sigma_D^2 \mathbf{I}_N$, and $\boldsymbol{\Sigma}^{[\text{d}]} = \sigma_D^2 \mathbf{I}_N$. Although the CFO, $\phi^{[\text{s-d}]}$, and PN vector, $\boldsymbol{\theta}^{[\text{s-d}]}$, are only contained in the received signal, $\mathbf{y}^{[\text{s}]}$, the backward substitution method proposed in [12] cannot be exploited here to solve (15) due to the unknown noise covariance matrix $\boldsymbol{\Sigma}^{[\text{r}]}$. Moreover, since all the parameters of interest are coupled with each other, the optimization problem in (15) is a non-convex problem. To make (15) tractable, in the following subsections, we propose to decouple (15) into several subproblems that can be each solved separately in an iterative approach.

### *A. Phase Noise Estimation*

In the first subproblem, we intend to obtain an estimate of the PN vector $\boldsymbol{\eta}^{[\text{s-d}]}$ at the $(k+1)$-th iteration, $[\hat{\boldsymbol{\eta}}^{[\text{s-d}]}]^{[k+1]}$, via the estimates of $[\boldsymbol{\phi}^{[\text{s-d}]}]$, $\mathbf{h}$, and $\mathbf{g}$ from the $k$-th iteration, $[\hat{\boldsymbol{\phi}}^{[\text{s-d}]}]^{[k]}$, $\hat{\mathbf{h}}^{[k]}$ and $\hat{\mathbf{g}}^{[k]}$, respectively, according to

$$[\hat{\boldsymbol{\eta}}^{[\text{s-d}]}]^{[k+1]} \propto \arg\min_{\boldsymbol{\eta}^{[\text{s-d}]}} \mathcal{L}_{\boldsymbol{\eta}^{[\text{s-d}]}} \tag{16}$$

where $\mathcal{L}_{\boldsymbol{\eta}^{[\text{s-d}]}} = \log\det(\boldsymbol{\Sigma}^{[\text{r}]}) + (\mathbf{y}^{[\text{s}]} - \boldsymbol{\mu}^{[\text{s-d}]})^H [\boldsymbol{\Sigma}^{[\text{r}]}]^{-1} (\mathbf{y}^{[\text{s}]} - \boldsymbol{\mu}^{[\text{s-d}]}) + \frac{1}{2}[\boldsymbol{\eta}^{[\text{s-d}]}]^T \boldsymbol{\eta}^{[\text{s-d}]}$ with $\boldsymbol{\mu}^{[\text{s-d}]} \triangleq \alpha \boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}} \hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]} \mathbf{F}^H \boldsymbol{\Lambda}_{s^{[\text{s}]}} \mathbf{F}_{[L]} \hat{\mathbf{c}}^{[k]}$, $[\hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]}]_{m,m} = \exp(\frac{j2\pi(m-1)[\hat{\phi}^{[\text{s-d}]}]^{[k]}}{N})$, $\hat{\mathbf{c}}^{[k]} \triangleq \hat{\mathbf{h}}^{[k]} \star \hat{\mathbf{g}}^{[k]}$, $\boldsymbol{\Sigma}^{[\text{r}]} = \alpha^2 \sigma_R^2 \boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}} \hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]} \hat{\mathbf{G}}^{[k]} [\hat{\mathbf{G}}^{[k]}]^H [\hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]}]^H \boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}^H + \sigma_D^2 \mathbf{I}_N$, and $\hat{\mathbf{G}}^{[k]}$ is constructed from $\hat{\mathbf{g}}^{[k]}$ as shown in (5). As shown in Appendix A, a closed-form solution for the PN estimate at the $(k+1)$-th iteration, $[\boldsymbol{\eta}^{[\text{s-d}]}]^{[k+1]}$, can be found as

$$\begin{aligned}[\hat{\boldsymbol{\eta}}^{[\text{s-d}]}]^{[k+1]} =& \big[\Re(\mathbf{B}^H \Big[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}\Big]^{-1} \mathbf{B}) + \frac{1}{2}\mathbf{I}_M\big]^{-1} \\ &\times \Re(\mathbf{B}^H \Big[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}\Big]^{-1} \bar{\mathbf{y}}^{[\text{s}]}),\end{aligned} \tag{17}$$

where $[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}$ is the estimate of the noise covariance matrix at the $k$-th iteration. Using (17), the un-shortened PN estimates at the $(k+1)$-th iteration, $[\hat{\boldsymbol{\theta}}^{[\text{s-d}]}]^{[k+1]}$, can be determined as $[\hat{\boldsymbol{\theta}}^{[\text{s-d}]}]^{[k+1]} = \boldsymbol{\Pi}^{[\text{s-d}]} [\hat{\boldsymbol{\eta}}^{[\text{s-d}]}]^{[k+1]}$ (see Section II-C). Finally, the noise covariance matrix, $[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}$, is updated via $[\hat{\boldsymbol{\theta}}^{[\text{s-d}]}]^{[k+1]}$.

### *B. Relay to Destination Channel Estimation*

In the second subproblem, the channel response $\mathbf{g}$ is updated by applying the estimated CFO, source-to-relay channel, and PN vector, $[\hat{\phi}^{[\text{s-d}]}]^{[k]}$, $\hat{\mathbf{h}}^{[k]}$ and $[\hat{\boldsymbol{\theta}}^{[\text{s-d}]}]^{[k+1]}$, respectively. To proceed, the combined channel $\mathbf{c}$ is first rewritten as

$$\mathbf{c} = \tilde{\mathbf{G}}\mathbf{h} = \tilde{\mathbf{H}}\mathbf{g}, \tag{18}$$

where $\tilde{\mathbf{G}} \in \mathbb{C}^{L \times L_h}$ is denoted as

$$\tilde{\mathbf{G}} = \begin{bmatrix} g(0) & 0 & \cdots & 0 & 0 \\ g(1) & g(0) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & g(0) & 0 \\ 0 & 0 & \cdots & g(1) & g(0) \end{bmatrix}, \tag{19}$$

and $\tilde{\mathbf{H}} \in \mathbb{C}^{L \times L_g}$ has a similar form as $\tilde{\mathbf{G}}$. Subsequently, the optimization problem for updating the relay-to-destination channel, $\mathbf{g}$, is given by

$$\hat{\mathbf{g}}^{[k+1]} \propto \arg\min_{\mathbf{g}} \mathcal{L}_{\mathbf{g}} \propto \arg\min_{\mathbf{g}} \log\det(\boldsymbol{\Sigma}) + (\mathbf{y}-\mathbf{C}\mathbf{g})^H \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{C}\mathbf{g}), \tag{20}$$

where $\mathbf{C} \triangleq \Big[\big(\alpha \hat{\boldsymbol{\Lambda}}_{\boldsymbol{\theta}^{[\text{s-d}]}}^{[k+1]} \hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]} \mathbf{F}^H \boldsymbol{\Lambda}_{s^{[\text{s}]}} \mathbf{F}_{[L]} \hat{\tilde{\mathbf{H}}}^{[k]}\big)^T, \big(\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\theta}^{[\text{r-d}]}} \hat{\boldsymbol{\Lambda}}_{\boldsymbol{\phi}^{[\text{r-d}]}} \mathbf{F}^H \boldsymbol{\Lambda}_{s^{[\text{r}]}} \mathbf{F}_{[L_g]}\big)^T\Big]^T$ with $\hat{\tilde{\mathbf{H}}}^{[k]}$ being formed by using the estimate of the source-to-relay channel in the $k$-th iteration $\hat{\mathbf{h}}^{[k]}$ according to (18), and $\boldsymbol{\Sigma} \triangleq \text{Blkdiag}\big(\boldsymbol{\Sigma}^{[\text{r}]}, \sigma_D^2 \mathbf{I}_N\big)$ with $\boldsymbol{\Sigma}^{[\text{r}]} = \alpha^2 \sigma_R^2 \hat{\boldsymbol{\Lambda}}_{\boldsymbol{\theta}^{[\text{s-d}]}}^{[k+1]} \hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]} \mathbf{G}\mathbf{G}^H [\hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]}]^H [\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\theta}^{[\text{s-d}]}}^{[k+1]}]^H + \sigma_D^2 \mathbf{I}_N$. Since the covariance matrix $\boldsymbol{\Sigma}$ is dependent on the channel response $\mathbf{g}$ as shown in (15), it is impossible to find a closed-form solution for $\mathbf{g}$ based on (20). Thus, we propose to use the channel covariance matrix at the $k$-th (previous iteration), $\hat{\boldsymbol{\Sigma}}^{[k]}$, to obtain an estimate of $\mathbf{g}$ at the $(k+1)$-th iteration. Using this approach and by equating the gradient of $\mathcal{L}_{\mathbf{g}}$ in (20) to zero, a closed-form solution for the relay-to-destination channel at the $(k+1)$-th iteration, $\hat{\mathbf{g}}^{[k+1]}$, can be derived as

$$\hat{\mathbf{g}}^{[k+1]} = \big(\mathbf{C}^H \big[\hat{\boldsymbol{\Sigma}}^{[k]}\big]^{-1} \mathbf{C}\big)^{-1} \mathbf{C}^H [\hat{\boldsymbol{\Sigma}}^{[k]}]^{-1} \mathbf{y}. \tag{21}$$

Subsequently, using $\hat{\mathbf{g}}^{[k+1]}$, the noise covariance $[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}$ is updated.

### C. Source to Relay Channel Estimation

In the third subproblem, we intend to update the estimate of the source to relay channel based on the estimates $[\phi^{[\text{s-d}]}]^{[k]}$, $[\boldsymbol{\theta}^{[\text{s-d}]}]^{[k+1]}$, and $\mathbf{g}^{[k+1]}$ via the following optimization problem

$$\hat{\mathbf{h}}^{[k+1]} \propto \arg\min_{\mathbf{h}} \mathcal{L}_{\mathbf{h}} \propto \arg\min_{\mathbf{g}} \ (\mathbf{y}^{[\text{s}]} - \mathbf{D}\mathbf{h})^H \Big[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}\Big]^{-1} (\mathbf{y}^{[\text{s}]} - \mathbf{D}\mathbf{h}), \tag{22}$$

where $\mathbf{D} \triangleq \alpha \hat{\boldsymbol{\Lambda}}_{\boldsymbol{\theta}^{[\text{s-d}]}}^{[k+1]} \hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]} \mathbf{F}^H \boldsymbol{\Lambda}_{s^{[\text{s}]}} \mathbf{F}_{[L]} \hat{\tilde{\mathbf{G}}}^{[k+1]}$. In (22), $\hat{\tilde{\mathbf{G}}}^{[k+1]}$ is formed as indicated in (19) by using $\hat{\mathbf{g}}^{[k+1]}$. Similar to the relay to destination channel, $\mathbf{g}$, the closed-form solution of $\mathbf{h}$ in (22) can be obtained as

$$\hat{\mathbf{h}}^{[k+1]} = \big(\mathbf{D}^H \big[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}\big]^{-1} \mathbf{D}\big)^{-1} \mathbf{D}^H \big[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}\big]^{-1} \mathbf{y}^{[\text{s}]}. \tag{23}$$

### D. CFO Estimation

In order to find an estimate of the source-destination CFO at the $(k+1)$-th iteration, $[\hat{\phi}^{[\text{s-d}]}]^{[k+1]}$, similar to the steps in (16), we approximate the covariance matrix, $\boldsymbol{\Sigma}^{[\text{r}]}$ with $[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}$ and solve the unconstrained problem

$$[\hat{\phi}^{[\text{s-d}]}]^{[k+1]} \propto \arg\min_{\phi^{[\text{s-d}]}} \ \mathcal{L}_{\phi^{[\text{s-d}]}} \propto \arg\min_{\phi^{[\text{s-d}]}} \ (\mathbf{y}^{[\text{s}]} - \boldsymbol{\mu}_{\phi^{[\text{s-d}]}})^H \big[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}\big]^{-1} \times (\mathbf{y}^{[\text{s}]} - \boldsymbol{\mu}_{\phi^{[\text{s-d}]}}), \tag{24}$$

where $\boldsymbol{\mu}_{\phi^{[\text{s-d}]}} \triangleq \alpha \boldsymbol{\Lambda}_{\boldsymbol{\phi}^{[\text{s-d}]}} \hat{\boldsymbol{\Lambda}}_{\boldsymbol{\theta}^{[\text{s-d}]}}^{[k+1]} \mathbf{F}^H \boldsymbol{\Lambda}_{s^{[\text{s}]}} \mathbf{F}_{[L]} \hat{\mathbf{c}}^{[k+1]}$. To make the problem in (24) more tractable and find a closed-form solution, a Taylor series approximation similar to that in (16) is applied here. Accordingly, $e^{\frac{j2\pi m \phi^{[\text{s-d}]}}{N}}$ can be approximated as

$$e^{\frac{j2\pi m \phi^{[\text{s-d}]}}{N}} \approx e^{\frac{j2\pi m [\hat{\phi}^{[\text{s-d}]}]^{[k]}}{N}} + \big(\phi^{[\text{s-d}]} - [\hat{\phi}^{[\text{s-d}]}]^{[k]}\big) \frac{j2\pi m}{N} e^{\frac{j2\pi m [\hat{\phi}^{[\text{s-d}]}]^{[k]}}{N}}, \tag{25}$$

where $[\hat{\phi}^{[\text{s-d}]}]^{[k]}$ is the estimated CFO at the $k$-th iteration. Using (25), $\mathcal{L}_{\phi^{[\text{s-d}]}}$ in (24) can be approximated as

$$\mathcal{L}_{\phi^{[\text{s-d}]}} \approx \Big(\mathbf{y}^{[\text{s}]} - [\hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]} + (\phi^{[\text{s-d}]} - [\hat{\phi}^{[\text{s-d}]}]^{[k]}) \tilde{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]}] \mathbf{d}^{[\text{s-d}]}\Big)^H \times \Big[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}\Big]^{-1} \Big(\mathbf{y}^{[\text{s}]} - [\hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]} + \big(\phi^{[\text{s-d}]} - [\hat{\phi}^{[\text{s-d}]}]^{[k]}\big) \tilde{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]}] \mathbf{d}^{[\text{s-d}]}\Big), \tag{26}$$

where $\mathbf{d}^{[\text{s-d}]} \triangleq \alpha \hat{\boldsymbol{\Lambda}}_{\boldsymbol{\theta}^{[\text{s-d}]}}^{[k+1]} \mathbf{F}^H \boldsymbol{\Lambda}_{s^{[\text{s}]}} \mathbf{F}_{[L]} \hat{\mathbf{c}}^{[k+1]}$ and $\tilde{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]}$ is a diagonal matrix where its $m$-th diagonal element is given by $[\tilde{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}]_{m,m} = \frac{j2\pi(m-1)}{N} e^{\frac{j2\pi [\hat{\phi}^{[\text{s-d}]}]^{[k]}(m-1)}{N}}$. By setting $\frac{\partial \mathcal{L}_{\phi^{[\text{s-d}]}}}{\partial \phi^{[\text{s-d}]}} = 0$ and solving for $\phi^{[\text{s-d}]}$, a closed-form solution for the CFO estimate at the $(k+1)$-th iteration, $[\phi^{[\text{s-d}]}]^{k+1}$, can be found as

$$[\hat{\phi}^{[\text{s-d}]}]^{k+1} = [\hat{\phi}^{[\text{s-d}]}]^{k} + \frac{\Re\big((\mathbf{y}^{[\text{s}]} - \hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]} \mathbf{d}^{[\text{s-d}]})^H \Big[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}\Big]^{-1} \tilde{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]} \mathbf{d}^{[\text{s-d}]}\big)}{[\mathbf{d}^{[\text{s-d}]}]^H [\tilde{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]}]^H \Big[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}\Big]^{-1} \tilde{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]} \mathbf{d}^{[\text{s-d}]}}. \tag{27}$$

Finally, the noise covariance matrices $\boldsymbol{\Sigma}^{[\text{r}]}$, and $\hat{\boldsymbol{\Sigma}}^{[k]}$ are updated using $[\phi^{[\text{s-d}]}]^{k+1}$ as $[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k+1]}$ and $\hat{\boldsymbol{\Sigma}}^{[k+1]}$.
The overall iterative joint estimation algorithm can be summarized as follows:

**Algorithm 1**

- **Solve** $\hat{\phi}^{[\text{r-d}]}$ and $\hat{\boldsymbol{\theta}}^{[\text{r-d}]}$ using (12) and (13) and initialize $\phi^{[\text{s-d}]}$, $\mathbf{g}$, $\mathbf{h}$, $\boldsymbol{\Sigma}^{[\text{r}]}$.
- **Repeat**
  - Update $[\hat{\boldsymbol{\theta}}^{[\text{s-d}]}]^{[k+1]}$ with $[\hat{\phi}^{[\text{s-d}]}]^{[k]}$, $\hat{\mathbf{h}}^{[k]}$ and $\hat{\mathbf{g}}^{[k]}$ being fixed by using (17) and then update $[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}$;
  - Update $\hat{\mathbf{g}}^{[k+1]}$ with $[\hat{\phi}^{[\text{s-d}]}]^{[k]}$, $\hat{\mathbf{h}}^{[k]}$ and $[\hat{\boldsymbol{\theta}}^{[\text{s-d}]}]^{[k+1]}$ being fixed by using (21) and then update $[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}$;
  - Update $\hat{\mathbf{h}}^{[k+1]}$ with $[\hat{\phi}^{[\text{s-d}]}]^{[k]}$, $\hat{\mathbf{g}}^{[k+1]}$ and $[\hat{\boldsymbol{\theta}}^{[\text{s-d}]}]^{[k+1]}$ being fixed by using (23);
  - Update $[\hat{\phi}^{[\text{s-d}]}]^{[k+1]}$ with $\hat{\mathbf{h}}^{[k+1]}$, $\hat{\mathbf{g}}^{[k+1]}$ and $[\hat{\boldsymbol{\theta}}^{[\text{s-d}]}]^{[k+1]}$ being fixed by using (27) and then update $[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}$ as $[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k+1]}$;
- **Until** $e(n+1) - e(n) \leq \varepsilon$ where $e(n)$ denotes the obtained value of objective function in (47) after the $n$-th iteration and $\varepsilon$ is a pre-set convergence accuracy.

### E. Initialization of the Proposed Iterative Algorithm

In *Algorithm 1*, initial estimates of the CFO, relay-to-declination channel, source-to-relay channel, and $\hat{\boldsymbol{\Sigma}}^{[\text{r}]}$, which are denoted by $[\hat{\phi}^{[\text{s-d}]}]^{[0]}$, $\hat{\mathbf{g}}^{[0]}$, $\hat{\mathbf{h}}^{[0]}$ and $[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[0]}$, respectively, are required. Thus, we present the initialization steps for the proposed iterative estimator. Simulations in Section VI show that the proposed estimator converges to the true values of the parameters of interest for this choice of initialization.

Since the relay-to-destination CFO and PN parameters, $\hat{\phi}^{[\text{r-d}]}$ and $\hat{\boldsymbol{\theta}}^{[\text{r-d}]}$, respectively, are estimated via (12) and (13), respectively, the initial relay-to-destination channel estimates, $\hat{\mathbf{g}}^{[0]}$, can be obtained from the received signal $\mathbf{y}^{[\text{r}]}$ via $\hat{\mathbf{g}}^{[0]} = \frac{1}{N P_{\text{T}}^{[\text{r}]}} \mathbf{F}_{[L_g]}^H \boldsymbol{\Lambda}_{s^{[\text{r}]}}^H \mathbf{F} \hat{\boldsymbol{\Lambda}}_{\theta^{[\text{r-d}]}}^H \hat{\boldsymbol{\Lambda}}_{\phi^{[\text{r-d}]}}^H \mathbf{y}^{[\text{r}]}$ [12]. Next, we seek to obtain the initial estimates of the source-to-destination CFO, $[\hat{\phi}^{[\text{s-d}]}]^{[0]}$, and source-to-relay channel, $\hat{\mathbf{h}}^{[0]}$. By ignoring the PN terms, (10) can be approximated as

$$\mathbf{y}^{[\text{s}]} \approx \alpha \boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}} \mathbf{F}^H \boldsymbol{\Lambda}_{s^{[\text{s}]}} \mathbf{F}_{[L]} \hat{\tilde{\mathbf{G}}}^{[0]} \mathbf{h} + \alpha \boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}} \hat{\mathbf{G}}^{[0]} \mathbf{v} + \mathbf{w},$$

where $\hat{\tilde{\mathbf{G}}}^{[0]}$ and $\hat{\mathbf{G}}^{[0]}$ are formed via $\hat{\mathbf{g}}^{[0]}$ according to (19) and (5), respectively. Subsequently, using the ML criterion the

initial estimates of the CFO, $[\hat{\phi}^{[\text{s-d}]}]^{[0]}$, and channel, $\hat{\mathbf{h}}^{[0]}$, can be obtained by minimizing

$$\begin{aligned}\{\hat{\mathbf{h}}^{[0]},[\hat{\phi}^{[\text{s-d}]}]^{[0]}\} = \min_{\mathbf{h},\phi^{[\text{s-d}]}} \; &(\mathbf{y}^{[\text{s}]}-\alpha\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}\mathbf{F}^H\boldsymbol{\Lambda}_{s^{[\text{s}]}}\mathbf{F}_{[L]}\hat{\bar{\mathbf{G}}}^{[0]}\mathbf{h})^H\\ &\times\left[\boldsymbol{\Sigma}^{[\text{r}]}\right]^{-1}(\mathbf{y}^{[\text{s}]}-\alpha\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}\mathbf{F}^H\boldsymbol{\Lambda}_{s^{[\text{s}]}}\mathbf{F}_{[L]}\hat{\bar{\mathbf{G}}}^{[0]}\mathbf{h})\\ &+\log\det(\boldsymbol{\Sigma}^{[\text{r}]}),\end{aligned}$$

where $\boldsymbol{\Sigma}^{[\text{r}]}=\alpha^2\sigma_R^2\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}\hat{\mathbf{G}}^{[0]}[\hat{\mathbf{G}}^{[0]}]^H\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}^H+\sigma_D^2\mathbf{I}_N$. Accordingly, $[\hat{\phi}^{[\text{s-d}]}]^{[0]}$ and $\hat{\mathbf{h}}^{[0]}$ can be determined as [33]

$$\begin{aligned}\hat{\mathbf{h}}^{[0]} &= \Big(\alpha[\hat{\bar{\mathbf{G}}}^{[0]}]^H\mathbf{F}_{[L]}^H\boldsymbol{\Lambda}_{s^{[\text{s}]}}^H\mathbf{F}\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}^H\left[\boldsymbol{\Sigma}^{[\text{r}]}\right]^{-1}\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}\mathbf{F}^H\boldsymbol{\Lambda}_{s^{[\text{s}]}}\\ &\quad\times\mathbf{F}_{[L]}\hat{\bar{\mathbf{G}}}^{[0]}\Big)^{-1}[\hat{\bar{\mathbf{G}}}^{[0]}]^H\mathbf{F}_{[L]}^H\boldsymbol{\Lambda}_{s^{[\text{s}]}}^H\mathbf{F}\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}^H\left[\boldsymbol{\Sigma}^{[\text{r}]}\right]^{-1}\mathbf{y}^{[\text{s}]},\\ [\hat{\phi}^{[\text{s-d}]}]^{[0]} &= \min_{\phi^{[\text{s-d}]}}\;(\mathbf{y}^{[\text{s}]}-\alpha\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}\mathbf{F}^H\boldsymbol{\Lambda}_{s^{[\text{s}]}}\mathbf{F}_{[L]}\hat{\bar{\mathbf{G}}}^{[0]}\hat{\mathbf{h}}^{[0]})^H\\ &\quad\times\left[\boldsymbol{\Sigma}^{[\text{r}]}\right]^{-1}(\mathbf{y}^{[\text{s}]}-\alpha\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}\mathbf{F}^H\boldsymbol{\Lambda}_{s^{[\text{s}]}}\mathbf{F}_{[L]}\hat{\bar{\mathbf{G}}}^{[0]}\hat{\mathbf{h}}^{[0]}),\end{aligned}\tag{28}$$

where the minimization in (28) is carried out through a one-dimensional exhaustive search. Although this process can be computationally intensive, it is only required to be carried out at the initial setup, since for subsequent OFDM packets, the previous CFO estimates can be applied to initialize the proposed iterative estimator. As for the additive noise covariance matrix, $[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[0]}$, using the Taylor approximation in Section III-A, we have

$$\begin{aligned}[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[0]} &= \alpha^2\sigma_R^2\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}\hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}\hat{\mathbf{G}}^{[0]}[\hat{\mathbf{G}}^{[0]}]^H\hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^H\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}^H+\sigma_D^2\mathbf{I}_N\\ &\approx\boldsymbol{\Omega}+\boldsymbol{\Omega}\odot\left(\boldsymbol{\theta}^{[\text{s-d}]}[\boldsymbol{\theta}^{[\text{s-d}]}]^H\right)+\sigma_D^2\mathbf{I}_N\\ &\approx\boldsymbol{\Omega}+\boldsymbol{\Omega}\odot\boldsymbol{\Psi}^{[\text{s-d}]}+\sigma_D^2\mathbf{I}_N,\end{aligned}\tag{29}$$

where $\boldsymbol{\Omega}\triangleq\alpha^2\sigma_R^2\hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}\hat{\mathbf{G}}^{[0]}[\hat{\mathbf{G}}^{[0]}]^H\hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^H$. In (29), since $\boldsymbol{\theta}^{[\text{s-d}]}$ is not known, we use the expectation $\mathbb{E}(\boldsymbol{\theta}^{[\text{s-d}]}[\boldsymbol{\theta}^{[\text{s-d}]}]^H)=\boldsymbol{\Psi}^{[\text{s-d}]}$ instead of the term $\boldsymbol{\theta}^{[\text{s-d}]}[\boldsymbol{\theta}^{[\text{s-d}]}]^H$. This allows for a closed-form expression for obtaining the source-to-relay channel estimates.

*Remark 1:* Similar to point-to-point systems [13], [22], [23], while jointly estimating the channel, CFO, and PN parameters in OFDM relay systems, a residual ambiguity may exist amongst these parameters. In what follows, we demonstrate the impact of this ambiguity on evaluating the performance of the proposed estimators.

The negative LLF in (15) can be rewritten as

$$\begin{aligned}\{\hat{\phi}^{[\text{s-d}]},\hat{\boldsymbol{\theta}}^{[\text{s-d}]},\hat{\phi}^{[\text{r-d}]},\hat{\boldsymbol{\theta}}^{[\text{r-d}]},\hat{\mathbf{h}},\hat{\mathbf{g}}\}\propto\arg\min\;&\log\det(\boldsymbol{\Sigma})\\ &+(\mathbf{y}-\boldsymbol{\mu})^H\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\\ &+\frac{1}{2}[\boldsymbol{\theta}^{[\text{s-d}]}]^T[\boldsymbol{\Psi}^{[\text{s-d}]}]^{-1}\boldsymbol{\theta}^{[\text{s-d}]}\\ &+\frac{1}{2}[\boldsymbol{\theta}^{[\text{r-d}]}]^T[\boldsymbol{\Psi}^{[\text{r-d}]}]^{-1}\boldsymbol{\theta}^{[\text{r-d}]}.\end{aligned}\tag{30}$$

Eq. (30) is similar to (15) with the exception that $\phi^{[\text{r-d}]}$ and $\boldsymbol{\theta}^{[\text{r-d}]}$ are also treated as parameters of interest and $\boldsymbol{\eta}^{[i]}$ is replaced with $\boldsymbol{\theta}^{[i]}$. At very high SNR, i.e., $\sigma_R^2\to 0$ and $\sigma_D^2\to 0$, (30) can be further simplified as

$$\begin{aligned}\{\hat{\phi}^{[\text{s-d}]},\hat{\boldsymbol{\theta}}^{[\text{s-d}]},\hat{\phi}^{[\text{r-d}]},\hat{\boldsymbol{\theta}}^{[\text{sr-d}]},\hat{\mathbf{h}},\hat{\mathbf{g}}\}\propto\arg\min\;&\log\det(\boldsymbol{\Sigma})\\ &+(\mathbf{y}-\boldsymbol{\mu})^H\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu}).\end{aligned}\tag{31}$$

From (31) it can be concluded that the metric for estimation of parameters of interest is solely dependent on the received signal instead of the prior information at high SNR [33]. Moreover, it can be straightforwardly shown that the received training symbols, e.g., $\mathbf{y}^{[\text{r}]}$, are not altered under a common phase rotation, $\varphi_g$, between the channel response, $\hat{\mathbf{g}}$, and PN parameters $\hat{\boldsymbol{\theta}}^{[\text{r-d}]}$, i.e.,

$$\hat{\mathbf{g}}\to\exp(-j\varphi_g)\mathbf{g},\quad\hat{\boldsymbol{\theta}}^{[\text{r-d}]}\to\boldsymbol{\theta}^{[\text{r-d}]}+\varphi_g\mathbf{1}_N.\tag{32}$$

Thus, the common phase rotation, $\varphi_g$, can be considered as a phase ambiguity amongst the channel and PN parameters that cannot be estimated. Using a similar approach, it can also be shown that there exists a phase ambiguity between the estimate of the source-to-relay channel, $\hat{\mathbf{h}}$, and the estimate of the source-to-destination PN parameter, $\hat{\boldsymbol{\theta}}^{[\text{s-d}]}$ given by

$$\hat{\mathbf{h}}\to\exp(-j\varphi_h)\mathbf{h},\quad\hat{\boldsymbol{\theta}}^{[\text{s-d}]}\to\boldsymbol{\theta}^{[\text{s-d}]}+(\varphi_h+\varphi_g)\mathbf{1}_N,\tag{33}$$

where $\varphi_h$ is the phase ambiguity associated with channel $\mathbf{h}$. In addition to the ambiguity between channel and PN, a phase ambiguity may also exist between the PN and CFO as:

$$\begin{aligned}\hat{\phi}^{[\text{s-d}]}&\to\phi^{[\text{s-d}]}-\epsilon^{[\text{s-d}]},\qquad\hat{\phi}^{[\text{r-d}]}\to\phi^{[\text{r-d}]}-\epsilon^{[\text{r-d}]},\\ \hat{\boldsymbol{\theta}}^{[\text{s-d}]}&\to\boldsymbol{\theta}^{[\text{s-d}]}+\boldsymbol{\epsilon}^{[\text{s-d}]},\qquad\hat{\boldsymbol{\theta}}^{[\text{r-d}]}\to\boldsymbol{\theta}^{[\text{r-d}]}+\boldsymbol{\epsilon}^{[\text{r-d}]},\end{aligned}\tag{34}$$

where $[\boldsymbol{\epsilon}^{[\text{s-d}]}]_m=\frac{2\pi(m-1)\epsilon^{[\text{s-d}]}}{N}$ and $[\boldsymbol{\epsilon}^{[\text{r-d}]}]_m=\frac{2\pi(m-1)\epsilon^{[\text{r-d}]}}{N}$. These ambiguities make it difficult to assess the estimation accuracy of the proposed iterative estimator. Thus, here, a new approach for determining the MSE of the estimated parameters is proposed. The MSE of the channel responses $\hat{\mathbf{h}}$ and $\hat{\mathbf{g}}$, can be computed as

$$\text{MSE}_{\mathbf{g}}=||\underline{\hat{\mathbf{g}}}-\underline{\mathbf{g}}||_2^2,\quad\text{MSE}_{\mathbf{h}}=||\underline{\hat{\mathbf{h}}}-\underline{\mathbf{h}}||_2^2,\tag{35}$$

where $\underline{\hat{\mathbf{g}}}\triangleq\exp(-j\angle\hat{g}(0))\hat{\mathbf{g}}$, $\underline{\hat{\mathbf{h}}}\triangleq\exp(-j\angle\hat{h}(0))\hat{\mathbf{h}}$, $\underline{\mathbf{g}}\triangleq\exp(-j\angle g(0))\mathbf{g}$ and $\underline{\mathbf{h}}\triangleq\exp(-j\angle h(0))\mathbf{h}$. Using this approach, the phase ambiguity between the PN and channels, does not affect the MSE of channel estimation. Similarly, for the CFO and PN, the overall MSE is calculated as

$$\text{MSE}_{\phi^{[\text{s-d}]},\boldsymbol{\theta}^{[\text{s-d}]}}=||\underline{\boldsymbol{\delta}}-\underline{\hat{\boldsymbol{\delta}}}||_2^2,\tag{36}$$

where $\underline{\boldsymbol{\delta}}=\boldsymbol{\delta}-\delta_0\mathbf{1}$, $\underline{\hat{\boldsymbol{\delta}}}=\hat{\boldsymbol{\delta}}-\hat{\delta}_0\mathbf{1}$, $\boldsymbol{\delta}=[\delta_0,\delta_0,\cdots,\delta_{N-1}]^T$ with $\delta_m=\theta^{[\text{s-d}]}(m)+\frac{2\pi(m-1)\phi^{[\text{s-d}]}}{N}$, and $\hat{\boldsymbol{\delta}}=[\hat{\delta}_0,\hat{\delta}_1,\cdots,\hat{\delta}_{N-1}]^T$ with $\hat{\delta}_m=\hat{\theta}^{[\text{s-d}]}(m)+\frac{2\pi(m-1)\hat{\phi}^{[\text{s-d}]}}{N}$.

## IV. The Hybrid Cramér-Rao Lower Bound

In this section, a the HCRLB for joint estimation of channel, CFO, and PN in OFDM relay networks is derived.

As stated in *Remark 1*, due to the ambiguities between the estimation of channel responses, CFO, and PN, (10) and (11) are first rewritten as

$$\begin{aligned}\mathbf{y}^{[\text{s}]}&=\alpha\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}\left(\mathbf{F}^H\underline{\boldsymbol{\Lambda}}_{s^{[\text{s}]}}\mathbf{F}_{[L]}\underline{\mathbf{c}}+\underline{\mathbf{G}}\mathbf{v}\right)+\mathbf{w},\\ \mathbf{y}^{[\text{r-d}]}&=\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{r-d}]}}\boldsymbol{\Lambda}_{\phi^{[\text{r-d}]}}\mathbf{F}^H\underline{\boldsymbol{\Lambda}}_{s^{[\text{r}]}}\mathbf{F}_{[L_g]}\underline{\mathbf{g}}+\mathbf{w},\end{aligned}\tag{37}$$

where $\underline{\mathbf{c}}\triangleq\underline{\mathbf{h}}\star\underline{\mathbf{g}}$ with $\underline{\mathbf{h}}$ and $\underline{\mathbf{g}}$ defined in (35), $[\underline{\boldsymbol{\Lambda}}_{s^{[\text{s}]}}]_{m,m}\triangleq s_{m-1}^{[\text{s}]}\exp(j\angle c(0))$, $[\underline{\boldsymbol{\Lambda}}_{s^{[\text{r}]}}]_{m,m}\triangleq s_{m-1}^{[\text{r}]}\exp(j\angle g(0))$ are known diagonal training signal matrices that are rotated by the phases of the first elements of the channels, $\mathbf{c}$ and $\mathbf{g}$, respectively, and matrix $\underline{\mathbf{G}}$ is constructed using $\underline{\mathbf{g}}$ similar to (5). Accordingly, the HCRLB for the estimation problem is given by [34]

$$\mathbb{E}_{\mathbf{y},\boldsymbol{\theta}^{[\text{s-d}]},\boldsymbol{\theta}^{[\text{r-d}]}|\phi^{[\text{s-d}]},\phi^{[\text{r-d}]},\underline{\mathbf{g}},\underline{\mathbf{h}}}\left[(\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})(\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})^T\right] \succeq \mathbf{B}^{-1},$$

where $\boldsymbol{\lambda} \triangleq \Big[\phi^{[\text{s-d}]}, (\boldsymbol{\theta}^{[\text{s-d}]})^T, \phi^{[\text{r-d}]}, (\boldsymbol{\theta}^{[\text{r-d}]})^T, \underline{g}_0, \Re(\underline{\tilde{\mathbf{g}}})^T, \Im(\underline{\tilde{\mathbf{g}}})^T, \underline{h}_0, \Re(\underline{\tilde{\mathbf{h}}})^T, \Im(\underline{\tilde{\mathbf{h}}})^T\Big]^T$ denotes the vector of parameters of interest, $\underline{\tilde{\mathbf{g}}} \triangleq \underline{\mathbf{g}}(1:L_g-1)$, $\underline{\tilde{\mathbf{h}}} \triangleq \underline{\mathbf{h}}(1:L_h-1)$, and $\mathbf{B}$ is the Bayesian information matrix (BIM) that is given by

$$\begin{aligned}\mathbf{B} =&\mathbb{E}_{\boldsymbol{\theta}^{[\text{s-d}]},\boldsymbol{\theta}^{[\text{r-d}]}}\left[\mathbf{FIM}(\mathbf{y};\boldsymbol{\lambda})\right] + \mathbb{E}_{\boldsymbol{\theta}^{[\text{s-d}]},\boldsymbol{\theta}^{[\text{r-d}]}}\left[-\triangle^{\boldsymbol{\lambda}}_{\boldsymbol{\lambda}}\log p(\boldsymbol{\theta}^{[\text{s-d}]})\right]\\ &+ \mathbb{E}_{\boldsymbol{\theta}^{[\text{s-d}]},\boldsymbol{\theta}^{[\text{r-d}]}}\left[-\triangle^{\boldsymbol{\lambda}}_{\boldsymbol{\lambda}}\log p(\boldsymbol{\theta}^{[\text{r-d}]})\right].\end{aligned} \tag{38}$$

In (38), $\mathbf{FIM}(\mathbf{y};\boldsymbol{\lambda}) = \mathbb{E}_{\mathbf{y}}\left[-\triangle^{\boldsymbol{\lambda}}_{\boldsymbol{\lambda}}\log p(\mathbf{y};\boldsymbol{\lambda})\right]$ denotes the Fisher's information matrix (FIM). In the following subsection the BIM in (38) is derived in detail.

### *A. Derivation of* $\mathbb{E}_{\boldsymbol{\theta}^{[s\text{-}d]},\boldsymbol{\theta}^{[r\text{-}d]}}\left[\mathbf{FIM}(\mathbf{y};\boldsymbol{\lambda})\right]$

In order to derive $\mathbb{E}_{\boldsymbol{\theta}^{[\text{s-d}]},\boldsymbol{\theta}^{[\text{r-d}]}}\left[\mathbf{FIM}(\mathbf{y};\boldsymbol{\lambda})\right]$, we first derive the FIM for the parameters of interest $\boldsymbol{\lambda}$.

**Theorem** 1: The $Q \times Q$ Fisher's information matrix $\mathbf{FIM}(\mathbf{y};\boldsymbol{\lambda})$ with $Q = 2(N+L_g+L_h)$ for the joint estimation problem is given by

$$\mathbf{FIM}(\mathbf{y};\boldsymbol{\lambda}) = \begin{bmatrix} \text{FIM}_{1,1}+\Upsilon_{1,1} & \cdots & \text{FIM}_{1,Q}+\Upsilon_{1,Q} \\ \vdots & \ddots & \vdots \\ \text{FIM}_{Q,1}+\Upsilon_{Q,1} & \cdots & \text{FIM}_{Q,Q}+\Upsilon_{Q,Q}\end{bmatrix}. \tag{39}$$

In (39), $\text{FIM}_{i,j}$, for $i,j = 1,2,\cdots,Q$, is determined as

$$\text{FIM}_{i,j} = 2\Re\left(\boldsymbol{\rho}_i^H \underline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\rho}_j\right), \tag{40}$$

where $\underline{\boldsymbol{\Sigma}} = \text{Blkdiag}(\underline{\boldsymbol{\Sigma}}^{[\text{r}]}, \sigma_D^2\mathbf{I}_N)$ with $\underline{\boldsymbol{\Sigma}}^{[\text{r}]} = \alpha^2\sigma_R^2\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}\underline{\mathbf{G}}\underline{\mathbf{G}}^H\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}^H\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}^H + \sigma_D^2\mathbf{I}_N$, and $\boldsymbol{\rho}_i$ is given by

- $i = 1$
  $\boldsymbol{\rho}_i \triangleq \left[(\alpha\boldsymbol{\Lambda}\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}\mathbf{F}^H\underline{\boldsymbol{\Lambda}}_{s^{[\text{s}]}}\mathbf{F}_{[L]}\underline{\mathbf{c}})^T, \mathbf{0}_{N\times 1}^T\right]^T$,
  where $\boldsymbol{\Lambda}$ is a diagonal matrix with $[\boldsymbol{\Lambda}]_{m,m} = \frac{j2\pi(m-1)}{N}$;
- $i = 2,3,\cdots,N+1$
  $\boldsymbol{\rho}_i \triangleq \left[(\text{Diag}\left(\alpha\boldsymbol{\Lambda}_{\boldsymbol{\phi}^{[\text{s-d}]}}\mathbf{F}^H\underline{\boldsymbol{\Lambda}}_{s^{[\text{s}]}}\mathbf{F}_{[L]}\underline{\mathbf{c}}\right)\mathbf{a}_{i-2})^T, \mathbf{0}_{N\times 1}^T\right]^T$,
  where $\mathbf{a}_m \triangleq [\mathbf{0}_{1,m-1}, j\exp(j\theta^{[\text{s-d}]}(m)), \mathbf{0}_{1,N-m}]^T$;
- $i = N+2$
  $\boldsymbol{\rho}_i \triangleq \left[\mathbf{0}_{N\times 1}^T, (\boldsymbol{\Lambda}\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{r-d}]}}\boldsymbol{\Lambda}_{\phi^{[\text{r-d}]}}\mathbf{F}^H\underline{\boldsymbol{\Lambda}}_{s^{[\text{r}]}}\mathbf{F}_{[L_g]}\underline{\mathbf{g}})^T\right]^T$;
- $i = N+3, N+3,\cdots,2N+2$
  $\boldsymbol{\rho}_i \triangleq \left[\mathbf{0}_{N\times 1}^T, (\text{Diag}\left(\boldsymbol{\Lambda}_{\phi^{[\text{r-d}]}}\mathbf{F}^H\underline{\boldsymbol{\Lambda}}_{s^{[\text{r}]}}\mathbf{F}_{[L_g]}\underline{\mathbf{g}}\right)\mathbf{b}_{i-N-3})^T\right]^T$,
  where $\mathbf{b}_m \triangleq [\mathbf{0}_{1,m-1}, j\exp(j\theta^{[\text{r-d}]}(m)), \mathbf{0}_{1,N-m}]^T$;
- $i = 2N+3$
  $\boldsymbol{\rho}_i \triangleq \mathbf{E}(:,1)$;
- $i = 2N+4, 2N+5,\cdots,2N+L_h+2$
  $\boldsymbol{\rho}_i \triangleq \mathbf{E}(:,i-2N-2)$;
- $i = 2N+L_h+3, 2N+L_h+4,\cdots,2N+2L_h+1$
  $\boldsymbol{\rho}_i \triangleq j\mathbf{E}(:,i-2N-L_h-1)$,
  where
  $$\mathbf{E} \triangleq \begin{bmatrix}\alpha\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}\mathbf{F}^H\underline{\boldsymbol{\Lambda}}_{s^{[\text{s}]}}\mathbf{F}_{[L]}\underline{\tilde{\mathbf{H}}} \\ \boldsymbol{\Lambda}_{\phi^{[\text{r-d}]}}\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{r-d}]}}\mathbf{F}^H\underline{\boldsymbol{\Lambda}}_{s^{[\text{r}]}}\mathbf{F}_{[L_g]}\end{bmatrix} \tag{41}$$
  and $\underline{\tilde{\mathbf{H}}}$ is constructed via $\underline{\mathbf{h}}$ as in (19);
- $i = 2N+2L_h+2$
  $\boldsymbol{\rho}_i \triangleq \mathbf{K}(:,1)$; (42)
- $i = 2N+2L_h+3,\cdots,2N+2L_h+L_g+1$
  $\boldsymbol{\rho}_i \triangleq \mathbf{K}(:,i-2N-2L_h-1)$; (43)
- $i = 2N+2L_h+L_g+2,\cdots,Q$
  $\boldsymbol{\rho}_i \triangleq j\mathbf{K}(:,i-2N-2L_h-2L_g)$; (44)
  where
  $$\mathbf{K} \triangleq \left[(\alpha\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}\mathbf{F}^H\underline{\boldsymbol{\Lambda}}_{s^{[\text{s}]}}\mathbf{F}_{[L]}\underline{\tilde{\mathbf{G}}})^T, \mathbf{0}_{N\times 1}^T\right]^T, \tag{45}$$
  and $\underline{\tilde{\mathbf{G}}}$ is constructed using $\underline{\mathbf{g}}$ as in (18).

Moreover, in (39), for $i = 1,2,\cdots,N+1$ and $j = 2N+3, 2N+4,\cdots,2N+2L_g+1$, $\Upsilon_{i,j} = \text{Tr}\left[\underline{\boldsymbol{\Sigma}}^{-1}\mathbf{Q}_i\underline{\boldsymbol{\Sigma}}^{-1}\mathbf{Q}_j\right]$ and $\Upsilon_{i,j} = 0$ for all other $i$ and $j$. Note that $\mathbf{Q}_i \triangleq \text{Blkdiag}\left(\mathbf{W}_i, \mathbf{0}_{N\times N}\right)$ is given by

- $i = 1$,
  $\mathbf{W}_i \triangleq \left(\alpha^2\sigma_R^2\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}\underline{\mathbf{G}}\underline{\mathbf{G}}^H\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}^H\right) \odot \left(\boldsymbol{\Lambda}\boldsymbol{\vartheta}\boldsymbol{\vartheta}^H + \boldsymbol{\vartheta}(\boldsymbol{\Lambda}\boldsymbol{\vartheta})^H\right)$;
- $i = 2,\cdots,N+1$
  $\mathbf{W}_i \triangleq \left(\alpha^2\sigma_R^2\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}\underline{\mathbf{G}}\underline{\mathbf{G}}^H\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}^H\right)$
  $\odot \left(\mathbf{a}_{i-2}[\boldsymbol{\theta}^{[\text{s-d}]}]^H + \boldsymbol{\theta}^{[\text{s-d}]}\mathbf{a}_{i-2}^H\right)$;
- $i = 2N+3$
  $\mathbf{W}_i \triangleq \alpha^2\sigma_R^2\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}\left(\mathbf{D}_0\underline{\mathbf{G}}^H + \underline{\mathbf{G}}\mathbf{D}_0^H\right)\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}^H\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}^H$;
- $i = 2N+4,\cdots,2N+L_g+2$
  $\mathbf{W}_i \triangleq \alpha^2\sigma_R^2\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}$
  $\times \left(\mathbf{D}_{i-2N-3}\underline{\mathbf{G}}^H + \underline{\mathbf{G}}\mathbf{D}_{i-2N-3}^H\right)\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}^H\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}^H$;
- $i = 2N+L_g+3,\cdots,2N+2L_g+1$
  $\mathbf{W}_i \triangleq j\alpha^2\sigma_R^2\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}$
  $\times \left(\mathbf{D}_{i-L_g-2N-2}\underline{\mathbf{G}}^H - \underline{\mathbf{G}}\mathbf{D}_{i-2N-L_g-2}^H\right)$
  $\times \boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}^H\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}^H$;

In the above, $\boldsymbol{\vartheta} \triangleq \left[1, e^{\frac{j2\pi\phi^{[\text{s-d}]}}{N}}, \cdots, e^{\frac{j2\pi(N-1)\phi^{[\text{s-d}]}}{N}}\right]$ and $\mathbf{D}_m \triangleq \left[\mathbf{0}_{N\times(L_g-m-1)}, \mathbf{I}_N, \mathbf{0}_{N\times m}\right]$, $\forall m$.

*Proof:* See Appendix B. ■

Although the FIM can be obtained in closed-form, a closed-form expression for $\mathbb{E}_{\boldsymbol{\theta}^{[\text{s-d}]},\boldsymbol{\theta}^{[\text{r-d}]}}\left[\mathbf{FIM}(\mathbf{y};\boldsymbol{\lambda})\right]$ cannot be obtained due to the presence of a complex multidimensional integration. Hence, here, $\mathbb{E}_{\boldsymbol{\theta}^{[\text{s-d}]},\boldsymbol{\theta}^{[\text{r-d}]}}\left[\mathbf{FIM}(\mathbf{y};\boldsymbol{\lambda})\right]$ is numerically evaluated.

### *B. Derivation of* $\mathbb{E}_{\boldsymbol{\theta}^{[s\text{-}d]},\boldsymbol{\theta}^{[r\text{-}d]}}\left[-\triangle^{\boldsymbol{\lambda}}_{\boldsymbol{\lambda}}\log p(\boldsymbol{\theta}^{[s\text{-}d]})\right]$ *and* $\mathbb{E}_{\boldsymbol{\theta}^{[s\text{-}d]},\boldsymbol{\theta}^{[r\text{-}d]}}\left[-\triangle^{\boldsymbol{\lambda}}_{\boldsymbol{\lambda}}\log p(\boldsymbol{\theta}^{[r\text{-}d]})\right]$

Since $p\left(\boldsymbol{\theta}^{[\text{s-d}]}\right)$ and $p\left(\boldsymbol{\theta}^{[\text{r-d}]}\right)$ are independent of $\phi^{[\text{s-d}]}$, $\phi^{[\text{r-d}]}$, $\underline{\mathbf{g}}$, and $\underline{\mathbf{h}}$, we can straightforwardly obtain

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{\theta}^{[\text{s-d}]},\boldsymbol{\theta}^{[\text{r-d}]}}\left[-\triangle^{\boldsymbol{\lambda}}_{\boldsymbol{\lambda}}\log p(\boldsymbol{\theta}^{[\text{s-d}]})\right] = \text{Blkdiag}\Big(0, [\boldsymbol{\Psi}^{[\text{s-d}]}]^{-1}, 0,\\ &\mathbf{0}_{N\times N}, \mathbf{0}_{(2L_g-1)\times(2L_g-1)}, \mathbf{0}_{(2L_h-1)\times(2L_h-1)}\Big),\\ &\mathbb{E}_{\boldsymbol{\theta}^{[\text{s-d}]},\boldsymbol{\theta}^{[\text{r-d}]}}\left[-\triangle^{\boldsymbol{\lambda}}_{\boldsymbol{\lambda}}\log p(\boldsymbol{\theta}^{[\text{r-d}]})\right] = \text{Blkdiag}\Big(0, \mathbf{0}_{N\times N}, 0,\\ &[\boldsymbol{\Psi}^{[\text{r-d}]}]^{-1}, \mathbf{0}_{(2L_g-1)\times(2L_g-1)}, \mathbf{0}_{(2L_h-1)\times(2L_h-1)}\Big).\end{aligned}$$

Finally, the BIM in (38) can be calculated using the results in Sections IV-A and IV-B.

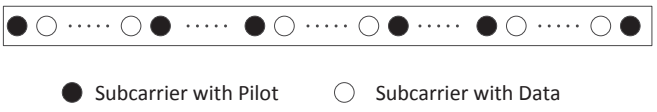


Fig. 4. Illustration of the comb-type data symbol.

### C. Derivation of the Transformed HCRLB

As shown in *Remark 1*, due to the ambiguities in the estimation of parameters of interest, the MSE of the CFO and PN is computed jointly as shown in (36). Consequently, the parameters of interests, $\boldsymbol{\lambda}$ need to be transformed to $\boldsymbol{\lambda}_{\text{mod}}=\Big[\underline{\boldsymbol{\delta}}^T, \phi^{[\text{r-d}]}, (\boldsymbol{\theta}^{[\text{r-d}]})^T, \underline{g}_0, \Re(\underline{\tilde{\mathbf{g}}})^T, \Im(\underline{\tilde{\mathbf{g}}})^T, \underline{h}_0, \Re(\underline{\tilde{\mathbf{h}}})^T, \Im(\underline{\tilde{\mathbf{h}}})^T\Big]^T$. Since $\delta_m = \theta^{[\text{s-d}]}(m) + \frac{2\pi(m-1)\phi^{[\text{s-d}]}}{N}$, this transformation can be written in matrix form as

$$\boldsymbol{\lambda}_{\text{mod}} = \boldsymbol{\Xi}\boldsymbol{\lambda},$$

where $\boldsymbol{\Xi} \triangleq \boldsymbol{\Xi}_2\boldsymbol{\Xi}_1$, $\boldsymbol{\Xi}_1 \triangleq \text{Blkdiag}\big(0, \tilde{\boldsymbol{\Xi}}_1, 1, \mathbf{I}_{N\times N}, \mathbf{I}_{(2L_g-1)\times(2L_g-1)}, \mathbf{I}_{(2L_h-1)\times(2L_h-1)}\big)$, $\boldsymbol{\Xi}_2 \triangleq \text{Blkdiag}(\tilde{\boldsymbol{\Xi}}_2, 1, \mathbf{I}_{N\times N}, \mathbf{I}_{(2L_g-1)\times(2L_g-1)}, \mathbf{I}_{(2L_h-1)\times(2L_h-1)})$, and

$$\tilde{\boldsymbol{\Xi}}_1 \triangleq \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{N\times N},$$

$$\tilde{\boldsymbol{\Xi}}_2 \triangleq \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ \frac{2\pi}{N} & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{2\pi(N-1)}{N} & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{N\times(N+1)}.$$

Thus, the HCRLB for the transformed parameters of interest, $\boldsymbol{\lambda}_{\text{mod}}$, $\mathbf{HCRLB}_{\text{mod}}$, is given by $\mathbf{HCRLB}_{\text{mod}} = \boldsymbol{\Xi}\mathbf{B}^{-1}\boldsymbol{\Xi}^T$ [33].

## V. Data Detection in Presence of Phase Noise

In this section, a receiver structure for data detection at the destination in the presence of PN is proposed. Since the PN parameters vary over an OFDM symbol, they need to be accurately tracked over the length of each symbol. Hence, we propose the transmission of comb-type data symbols from the source node, i.e., each transmitted symbol consists of both pilot and data subcarriers (see Fig. 4). As discussed in Section II-C, in each OFDM data symbol, it is sufficient to estimate the shortened PN vector of length $M$. Thus, the number of subcarriers utilized for data transmission should be less than $N-M$. The received signal at the destination during the data transmission interval is given by

$$\begin{aligned}\mathbf{y}^{[\text{s}]} &= \alpha\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}\left(\mathbf{F}\boldsymbol{\Lambda}_{\tilde{\mathbf{c}}}\mathbf{s}^{[\text{s}]} + \mathbf{G}\mathbf{v}\right) + \mathbf{w} \\ &= \mathbf{T}_{\text{T}}\mathbf{s}_{\text{T}}^{[\text{s}]} + \mathbf{T}_{\text{D}}\mathbf{s}_{\text{D}}^{[\text{s}]} + \alpha\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}\mathbf{G}\mathbf{v} + \mathbf{w},\end{aligned} \tag{46}$$

where $\mathbf{s}^{[\text{s}]}$ denotes the comb-type signal transmitted during the data transmission interval with $\mathbb{E}(\mathbf{s}^{[\text{s}]}\left[\mathbf{s}^{[\text{s}]}\right]^H) = P_{\text{T}}^{[\text{s}]}\mathbf{I}_N$, $\mathbf{s}_{\text{T}}^{[\text{s}]}$ and $\mathbf{s}_{\text{D}}^{[\text{s}]}$ denote the pilot and data vector contained in $\mathbf{s}^{[\text{s}]}$, respectively, and $\mathbf{T}_{\text{T}}$ and $\mathbf{T}_{\text{D}}$ are the associated sub-matrices of the combined channel, $\alpha\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}\boldsymbol{\Lambda}_{\boldsymbol{\phi}^{[\text{s-d}]}}\mathbf{F}\boldsymbol{\Lambda}_{\tilde{\mathbf{c}}}$, corresponding to $\mathbf{s}_{\text{T}}^{[\text{s}]}$ and $\mathbf{s}_{\text{D}}^{[\text{s}]}$, respectively. Since in (46), the unknown PN vector $\boldsymbol{\theta}^{[\text{s-d}]}$ and data vector $\mathbf{s}_{\text{D}}^{[\text{s}]}$ are coupled with each other, similar to the estimation part, an iterative method is applied here. By using the MAP criterion as in (14), the joint estimation of PN parameters and data can be formulated as

$$\begin{aligned}\{\hat{\boldsymbol{\theta}}^{[\text{s-d}]}, \mathbf{s}_{\text{D}}^{[\text{s}]}\} = \arg\min_{\boldsymbol{\theta}^{[\text{s-d}]}, \mathbf{s}_{\text{D}}^{[\text{s}]}} &\log\det(\boldsymbol{\Sigma}^{[\text{r}]}) + \left(\mathbf{y}^{[\text{s}]} - \boldsymbol{\mu}\right)^H \\ &\times \left[\boldsymbol{\Sigma}^{[\text{r}]}\right]^{-1}\left(\mathbf{y}^{[\text{s}]} - \boldsymbol{\mu}\right) + \frac{1}{2}\left[\boldsymbol{\eta}^{[\text{s-d}]}\right]^T\boldsymbol{\eta}^{[\text{s-d}]},\end{aligned} \tag{47}$$

where $\boldsymbol{\mu} \triangleq \alpha\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}\boldsymbol{\Lambda}_{\hat{\phi}^{[\text{s-d}]}}\mathbf{F}\boldsymbol{\Lambda}_{\hat{\mathbf{c}}}\mathbf{s}^{[\text{s}]}$ with $\boldsymbol{\Lambda}_{\hat{\phi}^{[\text{s-d}]}}$ and $\boldsymbol{\Lambda}_{\hat{\mathbf{c}}}$ are determined base on the estimated CFO and channels, $\hat{\boldsymbol{\phi}}^{[\text{s-d}]}$ and $\hat{\mathbf{c}}$, respectively, and $\boldsymbol{\Sigma}^{[\text{r}]} = \alpha^2\sigma_R^2\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}\boldsymbol{\Lambda}_{\hat{\phi}^{[\text{s-d}]}}\hat{\mathbf{G}}\hat{\mathbf{G}}^H\boldsymbol{\Lambda}_{\hat{\phi}^{[\text{s-d}]}}^H\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}^H + \sigma_D^2\mathbf{I}_N$ is the noise covariance matrix that is calculated via the estimated channels, $\hat{\mathbf{g}}$, and CFO, $\hat{\phi}^{[\text{s-d}]}$. First, the data symbols at the $k$-th iteration, $[\mathbf{s}_{\text{D}}^{[\text{s}]}]^{[k]}$, are used to estimate the PN at the $(k+1)$-th iteration, $[\boldsymbol{\theta}^{[\text{s-d}]}]^{[k+1]}$. To obtain a closed-form solution, as in Section III-A, (46) is approximated by

$$\begin{aligned}\mathbf{y}^{[\text{s}]} \approx &\alpha\boldsymbol{\Lambda}_{\hat{\phi}^{[\text{s-d}]}}\mathbf{F}\boldsymbol{\Lambda}_{\hat{\mathbf{c}}}[\mathbf{s}^{[\text{s}]}]^{[k]} + \text{Diag}\big(j\alpha\boldsymbol{\Lambda}_{\hat{\phi}^{[\text{s-d}]}}\mathbf{F}\boldsymbol{\Lambda}_{\hat{\mathbf{c}}}[\mathbf{s}^{[\text{s}]}]^{[k]}\big) \\ &\times \boldsymbol{\Pi}^{[\text{s-d}]}\boldsymbol{\eta}^{[\text{s-d}]} + \alpha\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\theta}^{[\text{s-d}]}}^{[k]}\boldsymbol{\Lambda}_{\hat{\phi}^{[\text{s-d}]}}\hat{\mathbf{G}}\mathbf{v} + \mathbf{w},\end{aligned}$$

where $\boldsymbol{\eta}^{[\text{s-d}]}$ denotes the shorten PN vector. By equating the gradient of (47) to zero, $[\hat{\boldsymbol{\theta}}^{[\text{s-d}]}]^{[k+1]}$ can be determined as

$$\begin{aligned}[\hat{\boldsymbol{\theta}}^{[\text{s-d}]}]^{[k+1]} =&\boldsymbol{\Pi}^{[\text{s-d}]}\big(\Re(\mathbf{M}^H\left[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}\right]^{-1}\mathbf{M}) + \frac{1}{2}\mathbf{I}_M\big)^{-1} \\ &\times \Re\big(\mathbf{M}^H\left[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}\right]^{-1} \\ &\times (\mathbf{y}^{[\text{s}]} - \alpha\boldsymbol{\Lambda}_{\hat{\phi}^{[\text{s-d}]}}\mathbf{F}\boldsymbol{\Lambda}_{\hat{\mathbf{c}}}[\mathbf{s}^{[\text{s}]}]^{[k]})\big),\end{aligned} \tag{48}$$

where $\mathbf{M} \triangleq \text{Diag}\left(j\alpha\boldsymbol{\Lambda}_{\hat{\phi}^{[\text{s-d}]}}\mathbf{F}\boldsymbol{\Lambda}_{\hat{\mathbf{c}}}[\mathbf{s}^{[\text{s-d}]}]^{[k]}\right)\boldsymbol{\Pi}^{[\text{s}]}$ and $[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]} = \alpha^2\sigma_R^2\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\theta}^{[\text{s-d}]}}^{[k]}\boldsymbol{\Lambda}_{\hat{\phi}^{[\text{s-d}]}}\hat{\mathbf{G}}\hat{\mathbf{G}}^H\boldsymbol{\Lambda}_{\hat{\phi}^{[\text{s-d}]}}^H[\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\theta}^{[\text{s-d}]}}^{[k]}]^H + \sigma_D^2\mathbf{I}_N$. Secondly, using $[\hat{\boldsymbol{\theta}}^{[\text{s-d}]}]^{[k+1]}$ and the noise covariance matrix at the $k+1$-th iteration, $[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k+1]}$, an estimate of the transmitted symbols at the $(k+1)$-th iteration can be obtained as

$$\begin{aligned}[\mathbf{s}_{\text{D}}^{[\text{s}]}]^{[k+1]} =&\big([\hat{\mathbf{T}}_{\text{D}}^{[k+1]}]^H\left[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k+1]}\right]^{-1}\hat{\mathbf{T}}_{\text{D}}^{[k+1]}\big)^{-1} \\ &\times [\hat{\mathbf{T}}_{\text{D}}^{[k+1]}]^H\left[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k+1]}\right]^{-1}(\mathbf{y}^{[\text{s}]} - \hat{\mathbf{T}}_{\text{P}}^{[k+1]}\mathbf{s}_{\text{P}}^{[\text{s}]}).\end{aligned} \tag{49}$$

In (49), although $\hat{\mathbf{T}}_{\text{P}}^{[k+1]}$ and $\hat{\mathbf{T}}_{\text{D}}^{[k+1]}$ are defined similar to $\mathbf{T}_{\text{P}}$ and $\mathbf{T}_{\text{D}}$ in (46), they are obtained via the estimates $[\hat{\boldsymbol{\theta}}^{[\text{s-d}]}]^{[k+1]}$, $\hat{\phi}^{[\text{s-d}]}$, and $\hat{\mathbf{c}}$. The overall iterative detector is given below.

**Algorithm 2**

- **Initialize** $\mathbf{s}_{\text{D}}^{[\text{s}]}$ and $\boldsymbol{\Sigma}^{[\text{r}]}$
- **Repeat**
  - Update $[\boldsymbol{\theta}^{[\text{s-d}]}]^{[k+1]}$ with the estimated $[\mathbf{s}_{\text{D}}^{[\text{s}]}]^{[k]}$ by using (48) and then update $[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}$ as $[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k+1]}$;
  - Update $[\mathbf{s}_{\text{D}}^{[\text{s}]}]^{[k+1]}$ with the estimated $[\boldsymbol{\theta}^{[\text{s-d}]}]^{[k+1]}$ by using (49);
- **Until** $q(n+1) - q(n) \leq \epsilon$ where $q(n)$ denotes the obtained value of objective function in (47) after the $n$-th iteration and $\epsilon$ is a pre-set convergence accuracy.

In Algorithm 2, initial estimates of $\mathbf{s}_{\text{D}}^{[\text{s}]}$ and $\boldsymbol{\Sigma}^{[\text{r}]}$ are obtained similar to that of the training interval.

*Remark 2:* As indicate here, the ambiguities associated with calculating the MSE for channel response, CFO, and PN parameters do not affect the data transmission interval. Let us denote the ambiguities of the channels and CFO in the training phase as $\hat{\mathbf{g}} \to \exp(-j\varphi_g)\mathbf{g}$, $\hat{\mathbf{c}} \to \exp(-j(\varphi_h+\varphi_g))\mathbf{c}$ and $\hat{\phi}^{[\text{s-d}]} \to \phi^{[\text{s-d}]} - \epsilon^{[\text{s-d}]}$. These ambiguities can be combined during the data transmission phase in the overall estimate of

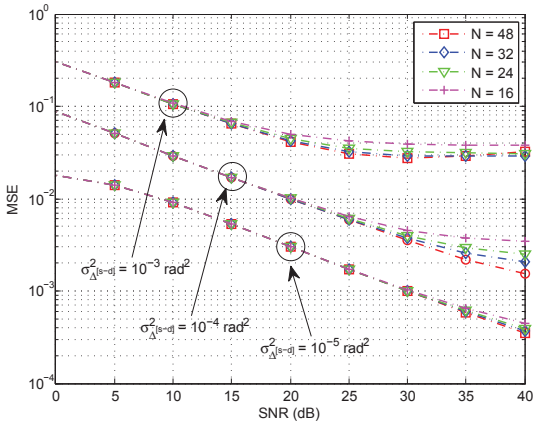


Fig. 5. The MSE of phase noise estimation with different $M$.

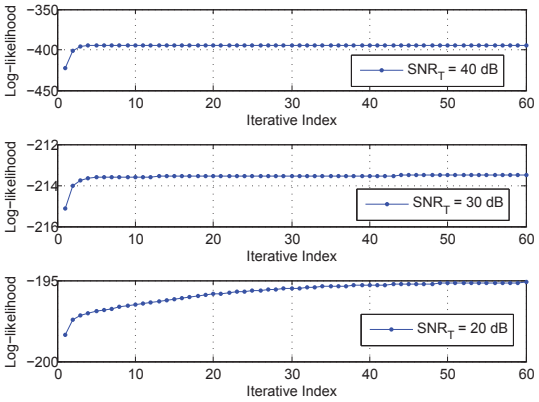


Fig. 6. The convergence of the proposed joint estimation algorithm at different SNR ($\sigma^2_{\Delta^{[s-d]}} = 10^{-4}$ rad$^2$).

the PN parameters $\boldsymbol{\theta}^{[s-d]}$ in (46), which can be written as $\hat{\boldsymbol{\theta}}^{[s-d]} \rightarrow \boldsymbol{\theta}^{[s-d]} + (\varphi_g + \varphi_h)\mathbf{1} + \boldsymbol{\epsilon}^{[s-d]}$ ($\boldsymbol{\epsilon}^{[s-d]}$ is defined in (34)). It can be clearly observed that these ambiguities do not affect the overall channel response, $\alpha \boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[s-d]}} \boldsymbol{\Lambda}_{\phi^{[s-d]}} \mathbf{F} \boldsymbol{\Lambda}_{\tilde{\mathbf{c}}}$, and the received signal in (46).

## VI. Simulation Results

In this section, extensive simulations are carried out to evaluate the performance of the proposed algorithms. In all the simulations, it is assumed that the multi-path channels exhibit unit-variance Rayleigh fading characteristics. Without loss of generality, it is assumed that the noise powers at relay and destination nodes are the same, i.e., $\sigma_R^2 = \sigma_D^2 = 1$. Moreover, the following simulation parameters are considered:

1) The multipath fading channels from relay-to-destination and source-to-relay, $\mathbf{g}$ and $\mathbf{h}$, respectively, are assumed to consist of 6 taps, i.e., $L_g = L_h = 6$,
2) $N = 64$ subcarriers are used in each OFDM symbol and all the subcarriers are modulated in quadrature phase shift keying (QPSK) format for both training and data transmission phases,
3) The normalized CFOs, $\phi^{[s-d]}$ and $\phi^{[r-d]}$, are uniformly drawn from $[-0.4, 0.4]$ and $[-0.2, 0.2]$, respectively, and
4) The PN innovation variances for, $\boldsymbol{\theta}^{[s-d]}$ and $\boldsymbol{\theta}^{[r-d]}$ are assumed to be the same, i.e., $\sigma^2_{\Delta^{[s-d]}} = \sigma^2_{\Delta^{[r-d]}} = \sigma^2_{\Delta}$.

Let us outline the choice of the scaling factor at the relay here. After removing the CP, the received signal vector at the relay in the frequency domain, $\mathbf{z}$, is given by

$$\mathbf{z} \triangleq \tilde{\mathbf{h}} \odot \mathbf{s} + \mathbf{n} \in \mathbb{C}^{N \times 1},$$

where $\tilde{\mathbf{h}} \triangleq [\tilde{h}_1, \tilde{h}_2, \cdots, \tilde{h}_N]^T$ with $\tilde{h}_k = \sum_{n=0}^{L_h-1} \exp \frac{-j2\pi kn}{K} h(n)$, for $k = 0, \cdots, N-1$. In addition, it is assumed that $\tilde{h}_k \sim \mathcal{CN}(0, L_h \sigma_h^2)$ with $\sigma_h^2$ denoting the variance of $h(n)$. Hence, the received signal power, $P_{\mathbf{z}}$, is given by $P_{\mathbf{z}} = \mathbb{E}_{\mathbf{h},\mathbf{s}}(||\mathbf{z}||_2^2) = N L_h \sigma_h^2 P_{[\mathrm{T}]}^{[s]} + N\sigma_R^2$. By considering the added CP at the source, the total power of the received signal at the relay node can be approximated as $\bar{P}_{\mathbf{z}} = P_{\mathbf{z}} \frac{N_{CP}+N}{N}$. Thus, the relay scaling factor, $\alpha$, can be determined as $\alpha = \sqrt{P_{\mathrm{T}}^{[r]} / \bar{P}_{\mathbf{z}}}$. Without loss of generality, in the remainder of this section, $\alpha = 1$ by letting $P_{\mathrm{T}}^{[r]} = \bar{P}_{\mathbf{z}}$. Moreover, it is assumed that $P_{\mathrm{T}}^{[s]} = P_{\mathrm{T}}^{[r]} = P_{\mathrm{T}}$ =SNR.[4]

Fig. 5 depicts the MSE of PN estimation, when estimating the shortened phase vector $\boldsymbol{\eta}^{[s-d]}$ for different values of $M$ (see Section II-C). For ease of comparison and to isolate the effect of CFO and channel estimation, it is assumed that the channel response, $\mathbf{c}$, and the CFO, $\phi^{[s-d]}$, are perfectly known. From the plots in Fig. 5 it can be concluded that when the PN innovation variance $\sigma^2_{\Delta^{[s-d]}}$ is small, i.e., $\sigma^2_{\Delta^{[s-d]}} = 10^{-5}$ rad$^2$, PN parameters can be accurately estimated using $M = 16$ compared to $M = 64$. Such an approach greatly reduces the PN estimation overhead. For scenarios with higher innovation variances, i.e., $\sigma^2_{\Delta^{[s-d]}} = 10^{-4}$ rad$^2$ and $\sigma^2_{\Delta^{[s-d]}} = 10^{-3}$ rad$^2$, it can be deduced that a larger value of $M$ is needed to ensure accurate PN estimation, e.g., $M = 32$. However, even for these larger PN variances, using the proposed scheme, the number of PN parameters that need to be tracked is reduced by one half. Accordingly, in the remainder of this section, $M = 32$.

In Fig. 6, the convergence of the proposed joint estimation algorithm is plotted for different SNRs. It can be observed that on average less than 50 iterations are needed for the proposed algorithm to coverage to the true estimates for a wide range of SNR values. More importantly, the result in Fig. 6 show that as the SNR increases the proposed algorithm converges more quickly, e.g., for SNR= 30 dB less than 10 iterations are needed for the proposed estimator to converge.

Figs. 7 and 8 illustrate the estimation MSE of relay-to-destination channel, $\underline{\mathbf{g}}$ (defined in *Remark 1*), for PN variances, $\sigma^2_{\Delta} = 10^{-4}$ rad$^2$ and $\sigma^2_{\Delta} = 10^{-3}$ rad$^2$, respectively, while the estimation MSE of the source-to-relay channel, $\underline{\mathbf{h}}$ (defined in *Remark 1*), is presented in Figs. 9 and 10. As a comparison, the channel estimation performance while ignoring the effect of PN on the received signal is also presented in these figures. Finally, the proposed estimation algorithms performance is benchmarked using the derived HCRLB in Section IV. Figs. 7–10 indicate that by including the PN parameters in the joint estimation problem, channel estimation performance in relay networks can be significantly enhanced. At moderate SNR, Figs. 7–10 also show that the proposed algorithm has a

[4]Due to lack of any prior art on the impact of PN on relaying networks, the performance of the proposed estimator and receiver structure cannot be compared with any existing algorithms.

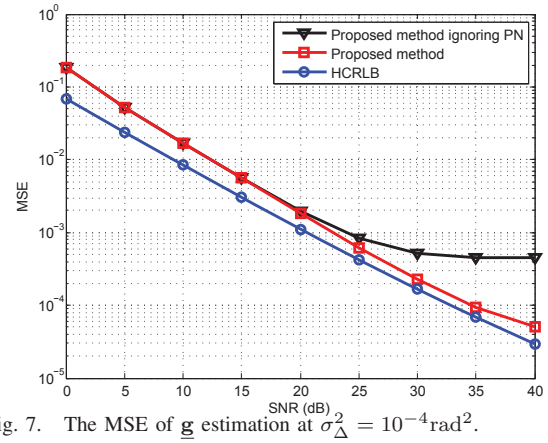


Fig. 7. The MSE of $\underline{\mathbf{g}}$ estimation at $\sigma^2_{\Delta} = 10^{-4}\text{rad}^2$.

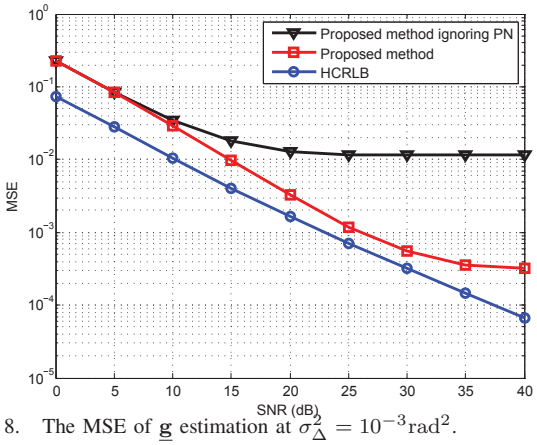


Fig. 8. The MSE of $\underline{\mathbf{g}}$ estimation at $\sigma^2_{\Delta} = 10^{-3}\text{rad}^2$.

constant performance gap with respect to the derived HCRLB bound for both PN innovation variances of $\sigma^2_{\Delta} = 10^{-4}$ rad$^2$ and $\sigma^2_{\Delta} = 10^{-3}$ rad$^2$. This is due to the inherent structure of the HCRLB, which is not necessarily a very tight bound as stated in [35]. Nevertheless, the performance of the proposed estimator is close to the derived HCRLB for moderate SNR. Finally, the results in Figs. 7–10 indicate that for large PN innovation variances, e.g., $\sigma^2_{\Delta^{[\text{s-d}]}} = 10^{-3}$ rad$^2$, the channel estimation performance suffers from an MSE error-floor at high SNR. This error-floor is caused by the time-varying PN parameters that cannot be perfectly estimated. Hence, at low SNR the overall estimation performance of the estimator is limited by the additive noise at the destination node, while at high SNR the algorithm's estimation performance is limited by the PN.

Fig. 11 illustrates the MSE for estimation of combined CFO and PN, $\underline{\boldsymbol{\delta}}$ for different PN variances. Similar to the results for channel estimation, the overall estimation performance suffers from an error floor for large PN variances, e.g., $\sigma^2_{\Delta} = 10^{-3}$ rad$^2$. This phenomenon can be similarly justified due to the imperfect estimation of PN parameters. Moreover, there is a 5 dB gap between the CFO and PN estimation MSE and the derived HCRLB at medium SNRs.

Fig. 12 illustrates the end-to-end BER of an uncoded OFDM relay network when applying the combination of the proposed iterative estimator and detector at $\sigma^2_{\Delta} = 10^{-4}\text{rad}^2$. It is observed that significant performance gains can be achieved by using the proposed joint data detection and PN estimation algorithm compared to a scheme that ignores the impact of PN. However, compared to the case with perfect channel, CFO, and PN, the proposed data detection algorithm still suffers from an error-floor at high SNR regime. This can be again attributed to imperfect PN estimation, where at high SNR, the overall BER of the OFDM relay system is dominated by PN and not the additive noise. This result indicates the importance of considering the impact of PN when determining the link budget, throughput, and coverage of wireless relay networks.

## VII. Conclusions

In this paper, joint channel, CFO, and PN estimation and data detection in OFDM relay networks is analyzed. Due to its time-varying nature, new algorithms for tracking the PN parameters in both the training and data transmission intervals are proposed. During the training interval, a new joint CFO, channel, and PN estimation algorithm that iteratively estimates these impairments is derived. To reduce estimation overhead, the proposed algorithm applies the correlation amongst the PN parameters to reduce the dimensionality of the estimation problem. Simulations show that the proposed estimator significantly enhances channel estimation performance in presence of PN, converges quickly, and performs close to the derived HCRLB at medium SNRs. Moreover, an iterative joint PN estimation and data detection receiver based on the MAP criterion at the destination node is proposed. The combination of the proposed estimation and data detection algorithms is shown to result in 5–10 dB performance gains over schemes that ignore the deteriorating effect of PN.

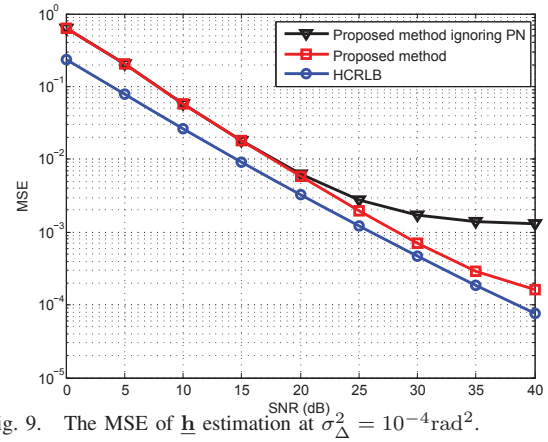


Fig. 9. The MSE of $\underline{\mathbf{h}}$ estimation at $\sigma^2_{\Delta} = 10^{-4}\text{rad}^2$.

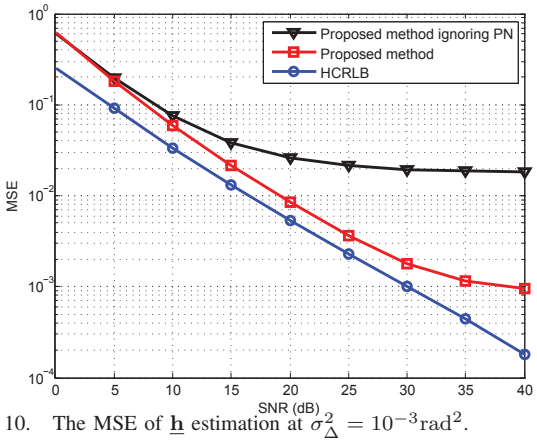


Fig. 10. The MSE of $\underline{\mathbf{h}}$ estimation at $\sigma^2_{\Delta} = 10^{-3}\text{rad}^2$.

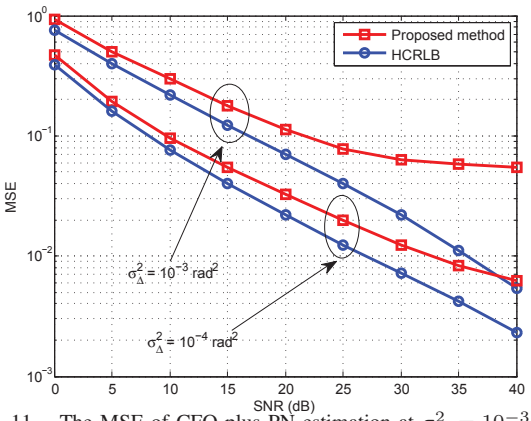


Fig. 11. The MSE of CFO plus PN estimation at $\sigma_\Delta^2 = 10^{-3}\text{rad}^2$ and $\sigma_\Delta^2 = 10^{-4}\text{rad}^2$.

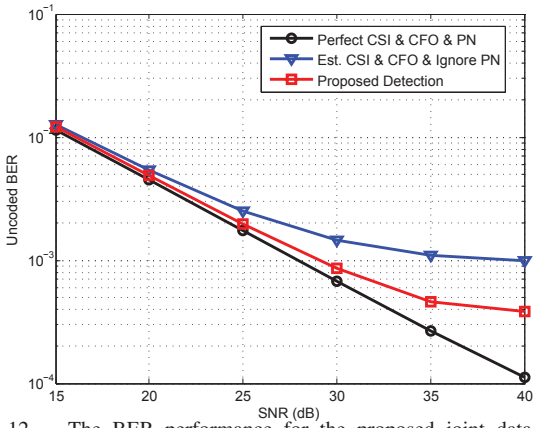


Fig. 12. The BER performance for the proposed joint data detection algorithm at $\sigma_\Delta^2 = 10^{-4}\text{rad}^2$.

## APPENDIX A
## DERIVATION OF (17)

In this section an expression for the optimization in (16) is derived. It is straightforward to determine that the optimization in (16) is a nonlinear and non-convex problem. Thus, the solution of $\boldsymbol{\eta}^{[\text{s-d}]}$ in (16) should be in general obtained through exhaustive search. To simplify the problem and obtain a closed-form solution, we first approximate the covariance matrix $\boldsymbol{\Sigma}^{[\text{r}]}$ as $[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]} = \alpha^2\sigma_R^2\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\theta}^{[\text{s-d}]}}^{[k]}\hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]}\hat{\mathbf{G}}^{[k]}[\hat{\mathbf{G}}^{[k]}]^H[\hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]}]^H[\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\theta}^{[\text{s-d}]}}^{[k]}]^H + \sigma_D^2\mathbf{I}_N$, where $[\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\theta}^{[\text{s-d}]}}^{[k]}]_{m,m} = e^{j[\hat{\theta}^{[\text{s-d}]}(m)]^{[k]}}$ is obtained from the previous iteration. Moreover, since the PN innovation variance of practical oscillators is usually small, the elements in $\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}$ can be approximated by a Taylor series expansion as $e^{j\theta^{[\text{s-d}]}(n)} \approx 1 + j\theta^{[\text{s-d}]}(n)$. This small angle approximation has also been used in [12], [22], [32] for PN estimation. Hence, the PN matrix, $\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}$, can be approximated as $\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}} \approx \mathbf{I}_N + j\text{Diag}(\boldsymbol{\theta}^{[\text{s-d}]})$ and $\mathcal{L}_{\boldsymbol{\eta}^{[\text{s-d}]}}$ in (16) can be rewritten as

$$\begin{aligned}\mathcal{L}_{\boldsymbol{\eta}^{[\text{s-d}]}} \approx& (\bar{\mathbf{y}}^{[\text{s}]} - \mathbf{B}\boldsymbol{\eta}^{[\text{s-d}]})^H \left[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}\right]^{-1}(\bar{\mathbf{y}}^{[\text{s}]} - \mathbf{B}\boldsymbol{\eta}^{[\text{s-d}]}) \\ &+ \frac{1}{2}[\boldsymbol{\eta}^{[\text{s-d}]}]^T\boldsymbol{\eta}^{[\text{s-d}]} \\ \approx& [\bar{\mathbf{y}}^{[\text{s}]}]^H\left[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}\right]^{-1}\bar{\mathbf{y}}^{[\text{s}]} - 2\Re([\bar{\mathbf{y}}^{[\text{s}]}]^H\left[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}\right]^{-1}\mathbf{B}) \\ &\times \boldsymbol{\eta}^{[\text{s-d}]} + [\boldsymbol{\eta}^{[\text{s-d}]}]^T\Re(\mathbf{B}^H\left[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}\right]^{-1}\mathbf{B})\boldsymbol{\eta}^{[\text{s-d}]} \\ &+ \frac{1}{2}[\boldsymbol{\eta}^{[\text{s-d}]}]^T\boldsymbol{\eta}^{[\text{s-d}]},\end{aligned} \tag{A.1}$$

where $\bar{\mathbf{y}}^{[\text{s}]} \triangleq \mathbf{y}^{[\text{s}]} - \alpha\hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]}\mathbf{F}^H\boldsymbol{\Lambda}_{s^{[\text{s}]}}\mathbf{F}_{[L]}\hat{\mathbf{c}}^{[k]}$ and $\mathbf{B} \triangleq j\text{Diag}(\alpha\hat{\boldsymbol{\Lambda}}_{\phi^{[\text{s-d}]}}^{[k]}\mathbf{F}^H\boldsymbol{\Lambda}_{s^{[\text{s}]}}\mathbf{F}_{[L]}\hat{\mathbf{c}}^{[k]})\boldsymbol{\Pi}^{[\text{s-d}]}$. Next, by equating the gradient of (A.1) to zero, i.e.,

$$\begin{aligned}\frac{\partial\mathcal{L}_{\boldsymbol{\eta}^{[\text{s-d}]}}}{\partial\boldsymbol{\eta}^{[\text{s-d}]}} =& -2\Re(\mathbf{B}^H\left[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}\right]^{-1}\bar{\mathbf{y}}^{[\text{s}]}) \\ &+ 2\Re(\mathbf{B}^H\left[[\hat{\boldsymbol{\Sigma}}^{[\text{r}]}]^{[k]}\right]^{-1}\mathbf{B})\boldsymbol{\eta}^{[\text{s-d}]} + \boldsymbol{\eta}^{[\text{s-d}]} \\ =& \mathbf{0}_{M\times 1},\end{aligned}$$

Then we obtain (17).

## APPENDIX B
## DERIVATION OF FIM

In this section, the FIM for joint estimation of channels, CFO, and PN parameters, i.e., $\mathbf{FIM}(\mathbf{y};\boldsymbol{\lambda})$, is derived. First, note that the combined received signal vector at the destination node in (37), $\mathbf{y} \triangleq \left[[\mathbf{y}^{[\text{s}]}]^T, [\mathbf{y}^{[\text{r}]}]^T\right]^T$ is a multivariate Gaussian random variable, i.e., $\mathbf{y} \sim \mathcal{N}(\underline{\boldsymbol{\mu}}, \underline{\boldsymbol{\Sigma}})$ with mean $\underline{\boldsymbol{\mu}} = \left[(\alpha\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}\mathbf{F}^H\underline{\boldsymbol{\Lambda}}_{s^{[\text{s}]}}\mathbf{F}_{[L]}\underline{\mathbf{c}})^T, (\boldsymbol{\Lambda}_{\theta^{[\text{r-d}]}}\boldsymbol{\Lambda}_{\phi^{[\text{r-d}]}}\mathbf{F}^H\underline{\boldsymbol{\Lambda}}_{s^{[\text{r}]}}\mathbf{F}_{[L_g]}\underline{\mathbf{g}})^T\right]^T$ and covariance $\underline{\boldsymbol{\Sigma}} = \text{Blkdiag}(\underline{\boldsymbol{\Sigma}}^{[\text{r}]}, \sigma_D^2\mathbf{I}_N)$. As a result, the $(i,j)$-th element of $\mathbf{FIM}(\mathbf{y};\boldsymbol{\lambda})$ can be determined as [33]

$$\begin{aligned}[\mathbf{FIM}(\mathbf{y};\boldsymbol{\lambda})]_{i,j} =& 2\text{Re}\Big[\frac{\partial\underline{\boldsymbol{\mu}}^H}{\partial\lambda_i}\underline{\boldsymbol{\Sigma}}^{-1}\frac{\partial\underline{\boldsymbol{\mu}}}{\partial\lambda_j}\Big] \\ &+ \text{Tr}\Big[\underline{\boldsymbol{\Sigma}}^{-1}\frac{\partial\underline{\boldsymbol{\Sigma}}}{\partial\lambda_i}\underline{\boldsymbol{\Sigma}}^{-1}\frac{\partial\underline{\boldsymbol{\Sigma}}}{\partial\lambda_j}\Big].\end{aligned} \tag{B.1}$$

To obtain (B.1), the following derivatives are evaluated as

$$\begin{aligned}\frac{\partial\underline{\boldsymbol{\mu}}}{\partial\phi^{[\text{s-d}]}} &= \left[(\alpha\boldsymbol{\Lambda}\boldsymbol{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}\boldsymbol{\Lambda}_{\boldsymbol{\phi}^{[\text{s-d}]}}\mathbf{F}^H\underline{\boldsymbol{\Lambda}}_{s^{[\text{s}]}}\mathbf{F}_{[L]}\underline{\mathbf{c}})^T, \mathbf{0}_{N\times 1}^T\right]^T, \\ \frac{\partial\underline{\boldsymbol{\mu}}}{\partial\phi^{[\text{r-d}]}} &= \left[\mathbf{0}_{N\times 1}^T, (\boldsymbol{\Lambda}\boldsymbol{\Lambda}_{\theta^{[\text{r-d}]}}\boldsymbol{\Lambda}_{\phi^{[\text{r-d}]}}\mathbf{F}^H\underline{\boldsymbol{\Lambda}}_{s^{[\text{r}]}}\mathbf{F}_{[L_g]}\underline{\mathbf{g}})^T\right]^T, \\ \frac{\partial\underline{\boldsymbol{\mu}}}{\partial\theta^{[\text{s-d}]}(m)} &= \left[(\text{Diag}\left(\alpha\boldsymbol{\Lambda}_{\phi^{[\text{s-d}]}}\mathbf{F}^H\underline{\boldsymbol{\Lambda}}_{s^{[\text{s}]}}\mathbf{F}_{[L]}\underline{\mathbf{c}}\right)\mathbf{a}_m)^T, \mathbf{0}_{N\times 1}^T\right]^T, \\ \frac{\partial\underline{\boldsymbol{\mu}}}{\partial\theta^{[\text{r-d}]}(m)} &= \left[\mathbf{0}_{N\times 1}^T, (\text{Diag}\left(\boldsymbol{\Lambda}_{\phi^{[\text{r-d}]}}\mathbf{F}^H\underline{\boldsymbol{\Lambda}}_{s^{[\text{r}]}}\mathbf{F}_{[L_g]}\underline{\mathbf{g}}\right)\mathbf{b}_m)^T\right]^T,\end{aligned} \tag{B.2}$$

where $\mathbf{a}_m$ and $\mathbf{b}_m$ are defined below (40). Moreover, for channel responses $\underline{\mathbf{g}}$ and $\underline{\mathbf{h}}$, for $m = 1, 2, \cdots, L_g$, we have $\frac{\partial\underline{\boldsymbol{\mu}}}{\partial(g(0))} = \mathbf{E}(:,1)$ and

$$\frac{\partial\underline{\boldsymbol{\mu}}}{\partial\Re(g(m))} = \mathbf{E}(:,m), \quad \frac{\partial\underline{\boldsymbol{\mu}}}{\partial\Im(g(m))} = j\mathbf{E}(:,m), \tag{B.3}$$

and, for $m = 1, 2, \cdots, L_h$, we have $\frac{\partial\underline{\boldsymbol{\mu}}}{\partial(h(0))} = \mathbf{K}(:,1)$ and

$$\frac{\partial\underline{\boldsymbol{\mu}}}{\partial\Re(h(m))} = \mathbf{K}(:,m), \quad \frac{\partial\underline{\boldsymbol{\mu}}}{\partial\Im(h(m))} = j\mathbf{K}(:,m), \tag{B.4}$$

where $\mathbf{E}$ and $\mathbf{K}$ are defined as in (44) and (45), respectively. Since $\phi^{[\text{r-d}]}$, $\boldsymbol{\theta}^{[\text{r-d}]}$, $\underline{\mathbf{h}}$ are irrelevant to the noise covariance matrix $\boldsymbol{\Sigma}$, it is straightforward to determine that $\frac{\partial\boldsymbol{\Sigma}}{\partial\phi^{[\text{r-d}]}} =$

$\frac{\partial \underline{\mathbf{\Sigma}}}{\partial \theta^{[\text{r-d}]}(m)} = \frac{\partial \underline{\mathbf{\Sigma}}}{\partial \underline{h}(0)} = \frac{\partial \underline{\mathbf{\Sigma}}}{\partial \Re(\underline{h}(m))} = \frac{\partial \underline{\mathbf{\Sigma}}}{\partial \Im(\underline{h}(m))} = \mathbf{0}$, $\forall m$. Moreover, for the CFO and PN parameters, $\phi^{[\text{s-d}]}$ and $\boldsymbol{\theta}^{[\text{s-d}]}$, we can obtain that $\frac{\partial \underline{\mathbf{\Sigma}}}{\partial \phi^{[\text{s-d}]}} = \text{Blkdiag}\big(\frac{\partial \underline{\mathbf{\Sigma}}^{[\text{r}]}}{\partial \phi^{[\text{s-d}]}}, \mathbf{0}_{N\times N}\big)$, where $\frac{\partial \underline{\mathbf{\Sigma}}^{[\text{r}]}}{\partial \phi^{[\text{s-d}]}} = \big(\alpha^2\sigma_R^2 \mathbf{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}} \underline{\mathbf{G}}\underline{\mathbf{G}}^H \mathbf{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}^H\big) \odot \big(\mathbf{\Lambda}\boldsymbol{\vartheta}\boldsymbol{\vartheta}^H + \boldsymbol{\vartheta}(\mathbf{\Lambda}\boldsymbol{\vartheta})^H\big)$, and

$$\begin{aligned}\frac{\partial \underline{\mathbf{\Sigma}}^{[\text{r}]}}{\partial \theta^{[\text{s-d}]}(m)} =& \big(\alpha^2\sigma_R^2 \mathbf{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}} \underline{\mathbf{G}}\underline{\mathbf{G}}^H \mathbf{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}^H\big) \\ &\odot \big(\mathbf{a}_m[\boldsymbol{\theta}^{[\text{s-d}]}]^H + \boldsymbol{\theta}^{[\text{s-d}]}\mathbf{a}_m^H\big),\ \forall m\end{aligned} \tag{B.5}$$

where $\boldsymbol{\vartheta} \triangleq \big[1, e^{\frac{j2\pi\phi^{[\text{s-d}]}}{N}}, \cdots, e^{\frac{j2\pi(N-1)\phi^{[\text{s-d}]}}{N}}\big]$. For channel response $\underline{\mathbf{g}}$, based on the structure of $\mathbf{G}$ as shown in (5), we have

$$\begin{aligned}\frac{\partial \underline{\mathbf{\Sigma}}^{[\text{r}]}}{\partial g(0)} =& \alpha^2\sigma_R^2 \mathbf{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}} \mathbf{\Lambda}_{\phi^{[\text{s-d}]}} \big(\mathbf{D}_0 \underline{\mathbf{G}}^H + \underline{\mathbf{G}}\mathbf{D}_0^H\big) \mathbf{\Lambda}_{\phi^{[\text{s-d}]}}^H \mathbf{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}^H, \\ \frac{\partial \underline{\mathbf{\Sigma}}^{[\text{r}]}}{\partial \Re(g(m))} =& \alpha^2\sigma_R^2 \mathbf{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}} \mathbf{\Lambda}_{\phi^{[\text{s-d}]}} \big(\mathbf{D}_m \underline{\mathbf{G}}^H + \underline{\mathbf{G}}\mathbf{D}_m^H\big) \\ &\times \mathbf{\Lambda}_{\phi^{[\text{s-d}]}}^H \mathbf{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}^H,\ m = 1, 2, \cdots, L_g - 1, \\ \frac{\partial \underline{\mathbf{\Sigma}}^{[\text{r}]}}{\partial \Im(g(m))} =& j\alpha^2\sigma_R^2 \mathbf{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}} \mathbf{\Lambda}_{\phi^{[\text{s-d}]}} \big(\mathbf{D}_m \underline{\mathbf{G}}^H - \underline{\mathbf{G}}\mathbf{D}_m^H\big) \\ &\times \mathbf{\Lambda}_{\phi^{[\text{s-d}]}}^H \mathbf{\Lambda}_{\boldsymbol{\theta}^{[\text{s-d}]}}^H,\ m = 1, 2, \cdots, L_g - 1,\end{aligned} \tag{B.6}$$

where $\mathbf{D}_m \triangleq \big[\mathbf{0}_{N\times(L_g-m-1)}, \mathbf{I}_N, \mathbf{0}_{N\times m}\big]$. Subsequently, the derivatives of the covariance matrix with respect to the relay and imaginary parts of the relay-to-destination channel parameters are given by $\frac{\partial \underline{\mathbf{\Sigma}}}{\partial g(0)} = \text{Blkdiag}\big(\frac{\partial \underline{\mathbf{\Sigma}}^{[\text{r}]}}{\partial g(0)}, \mathbf{0}_{N\times N}\big)$ and $\frac{\partial \underline{\mathbf{\Sigma}}}{\partial \Im(g(m))} = \text{Blkdiag}\big(\frac{\partial \underline{\mathbf{\Sigma}}^{[\text{r}]}}{\partial \Im(g(m))}, \mathbf{0}_{N\times N}\big)$, respectively. By combing (B.2)-(B.6) together, the results in Theorem 1 are derived.

## REFERENCES


[1] Editors: Y. Hua, D. W. Bliss, S. Gazor, Y. Rong, and Y. Sung, "Theories and methods for advanced wireless relays: Issue I," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, Sep. 2012.

[2] S. Yiu, R. Schober, and L. Lampe, "Distributed space time block coding," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1195–1206, Jul. 2006.

[3] R. Wang and M. Tao, "Joint source and relay precoding designs for MIMO two-way relaying based on MSE criterion," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1352–1365, 2012.

[4] F. Gao, T. Cui, and A. Nallanathan, "On channel estimation and optimal training design for amplify and forward relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1907–1916, May 2008.

[5] Y. Jing and X. Yu, "ML-based channel estimations for non-regenerative relay networks with multiple transmit and receive antennas," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1428–1439, Sep. 2012.

[6] P. Lioliou, M. Viberg, and M. Matthaiou, "Bayesian approach to channel estimation for AF MIMO relaying systems," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1440–1451, Sep. 2012.

[7] J. Ma, P. Orlik, J. Zhang, and G. Li, "Pilot matrix design for estimating cascaded channels in two-hop MIMO amplify-and-forward relay systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1956–1965, Jun. 2011.

[8] Y. Rong, M. Khandaker, and Y. Xiang, "Channel estimation of dual-hop MIMO relay system via parallel factor analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2224–2233, Jun. 2012.

[9] M. Dohler, R. W. Heath, A. Lozano, C. B. Papadias, and R. A. Valenzuela, "Is the PHY layer dead?" *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 159–165, Apr. 2011.

[10] S. Bay and C. Herzet, J.-M. Brossier, J.-P. Barbot, and B. Geller, "Analytic and asymptotic analysis of Bayesian Cramér-Rao bound for dynamical phase offset estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 61–70, Jan. 2008.

[11] N. Noels, H. Steendam, M. Moeneclaey, and H. Bruneel, "Carrier phase and frequency estimation for pilot-symbol assisted transmission: Bounds and algorithms," *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4578–4587, Dec. 2005.

[12] D. D. Lin, R. Pacheco, T. J. Lim, and D. Hatzinakos, "Joint estimation of channel response, frequency offset, and phase noise in OFDM," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3542–3554, Sep. 2006.

[13] D. D. Lin and T. J. Lim, "The variational inference approach to joint data detection and phase noise estimation in OFDM," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1862–1874, May 2007.

[14] T. C. W. Schenk, X.-J. Tao, P. F. M. Smulders, and E. R. Fledderus, "On the influence of phase noise induced ICI in MIMO OFDM systems," *IEEE Commun. Lett.*, vol. 9, no. 8, pp. 682–684, Aug. 2005.

[15] H. Mehrpouyan and S. Blostein, "Bounds and algorithms for multiple frequency offset estimation in cooperative networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1300–1311, Apr. 2011.

[16] K. J. Kim, R. Iltis, and H. Poor, "Frequency offset and channel estimation in cooperative relay networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 3142–3155, Sep. 2011.

[17] T. Pollet, M. Van Bladel, and M. Moeneclaey, "BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise," *IEEE Trans. Commun.*, vol. 43, no. 234, pp. 191–193, Feb./Mar./Apr. 1995.

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

[19] Z. Zhang, W. Zhang, and C. Tellambura, "Cooperative OFDM channel estimation in the presence of frequency offsets," *IEEE Trans. Veh. Technol.*, vol. 58, no. 7, pp. 3447–3459, Sep. 2009.

[20] L. Thiagarajan, S. Sun, and T. Quek, "Joint carrier frequency offset and channel estimation in OFDM based non-regenerative wireless relay networks," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP)*, Apr. 2009, pp. 2569–2572.

[21] P. Rabiei, W. Namgoong, and N. Al-Dhahir, "On the performance of ofdm-based amplify-and-forward relay networks in the presence of phase noise," *Communications, IEEE Transactions on*, vol. 59, no. 5, pp. 1458–1466, 2011.

[22] J. Tao, J. Wu, and C. Xiao, "Estimation of channel transfer function and carrier frequency offset for OFDM systems with phase noise," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4380–4387, Oct. 2009.

[23] F. Septier, Y. Delignon, A. Menhaj-Rivenq, and C. Garnier, "Monte carlo methods for channel, phase noise, and frequency offset estimation with unknown noise variances in OFDM systems," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3613–3626, Aug. 2008.

[24] J. Wells, *Multi-Gigabit Microwave and Millimeter-Wave Wireless Communications*. First Edition. Artech House, 2010.

[25] K.-C. Huang and D. J. Edwards, *Millimetre Wave Antennas for Gigabit Wireless Communications: A Practical Guide to Design and Analysis in a System Context*. John Wiley and Sons, Ltd., 2008.

[26] Y. Lee, J. Ha, and J. Choi, "Design of a wideband indoor repeater antenna with high isolation for 3g systems," *IEEE Antennas and Wireless Propagation Letters*, vol. 9, pp. 697–700, 2010.

[27] Mini-Circuits, "Wideband microwave amplifier ava-24+." [Online]. Available: http://www.minicircuits.com/pdfs/AVA-24+.pdf

[28] K. Salehian, M. Guillet, B. Caron, and A. Kennedy, "On-channel repeater for digital television broadcasting service," *IEEE Trans. on Broadcasting*, vol. 48, no. 2, pp. 97–102, 2002.

[29] Mini-Circuits, "Ultra low noise voltage conrtolled ocillator ros-209-319+." [Online]. Available: http://www.minicircuits.com/pdfs/ROS-209-319+.pdf

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

[31] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: a unifying theory and numerical methods for characterization," *IEEE Trans. Circuits and Systems I: Fund. Theory and Appl.*, vol. 47, no. 5, pp. 655–674, May 2000.

[32] H. Mehrpouyan *et al.*, "Joint estimation of channel and oscillator phase noise in MIMO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4790–4807, Sep. 2012.

[33] S. M. Kay, *Fundamentals of statistical signal processing: Estimation theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[34] H. L. V. Trees, *Detection, Estimation, and Modulation Theory*. John Wiley and Sons Inc., 2001.

[35] H. L. V. Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. John Wiley and Sons Inc., 2007.