Received 21 February 2021; revised 21 March 2021; accepted 12 April 2021. Date of publication 14 April 2021; date of current version 20 April 2021.

Digital Object Identifier 10.1109/OJCOMS.2021.3073348

Joint Channel and Phase Noise Estimation and Data Detection for GFDM

AMIRHOSSEIN MOHAMMADIAN (Graduate Student Member, IEEE), AND CHINTHA TELLAMBURA (Fellow, IEEE)

Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 1H9, Canada

CORRESPONDING AUTHOR: A. MOHAMMADIAN (e-mail: am11@ualberta.ca)

ABSTRACT Generalized frequency division multiplexing (GFDM), an enabler of beyond-5G wireless networks, can be critically impaired due to radio frequency (RF) phase noise. However, joint channel estimation and phase noise compensation for GFDM systems have not been addressed before. Hence, we tackle this problem. To this end, we propose an iterative algorithm for joint channel and phase noise estimation and two algorithms for joint data detection and phase noise compensation. These algorithms use linear and non-linear least-squares (NLS) methods and employ block-type and comb-type pilots. The complexity of these algorithms is also analyzed. Moreover, to reduce their complexity, interpolation techniques are deployed to decrease the number of unknowns. We also analyze the signal-to-interference-plus noise ratio (SINR) and sum-rate of GFDM contaminated with phase noise. Furthermore, the accuracy of the channel and phase noise estimates is established via Cramér-Rao lower bounds (CRLBs). The simulation results illustrate that the mean-squared error (MSE) performance of the proposed joint channel and phase noise estimator reaches the CRLB. Moreover, the proposed joint data symbol detection and phase noise compensation algorithms nearly eliminate the impacts of phase noise in GFDM systems.

INDEX TERMS GFDM, phase noise, non-linear least-squares (NLS), Cramér-Rao lower bound (CRLB).

I. INTRODUCTION

FIFTH generation (5G) and beyond 5G wireless communication networks must handle peak data rates of at least 1 Tb/s, over-the-air latency of 10–100 μs, user-experienced data rates of 1 Gb/s and the support of high user mobility ($\geq$ 1000 km/h) [1]–[3]. Although 4G wireless (Long Term Evolution (LTE) and other standards) widely uses orthogonal frequency division multiplexing (OFDM), its drawbacks include high peak-to-average power ratios, bandwidth loss associated with the cyclic prefix (CP), high out-of-band (OOB) emissions emanating from rectangular filters, and high synchronisation issues due to orthogonality mismatch. Due to these issues, OFDM may not answer the challenges for future wireless networks such as low power consumption for machine-to-machine communication and low latency for Tactile Internet and Internet-of-Things [4].

Hence, novel non-orthogonal multicarrier waveforms have been developed to address the limitations of OFDM, and these include filter bank multicarrier (FBMC) and generalized frequency division multiplexing (GFDM) [5]. FBMC avoids the CP, which improves the spectral efficiency. Moreover, it has low OOB emissions and is robust against doubly-dispersive fading [6]. However, FBMC requires a long-length filter, an issue for short-burst transmissions [7]. On the other hand, GFDM is a block-based multicarrier waveform, where the data is divided into many subcarriers and subsymbols, using circular pulse shaping for each subcarrier. GFDM has advantages of low peak-to-average-power ratios, low OOB emissions, flexible time-frequency structure, and high spectral efficiency due to low CP usage [8]. However, since it sacrifices orthogonality, implementation complexity can increase. Thus, several methods have been developed to handle this problem [9], [10]. Therefore, GFDM may enable future wireless standards. This potential has motivated extensive studies of GFDM for cognitive radio networks [11]–[14], space-time codes [15],

filter designs [16], [17], multiple-input and multiple-output communications [18] Internet-of-Things [19], and optical networks [20]. This list is not exhaustive.

However, non-orthogonal subcarriers of GFDM result in inter-carrier interference (ICI) and inter-symbol interference (ISI). These inherent interference levels are further exacerbated by radio frequency (RF) impairments such as phase noise, carrier frequency offset (CFO), and in-phase and quadrature (IQ) imbalance, which are critical [21]–[24]. In particular, phase noise is an acute problem, which is mainly caused by aging circuit components and the poor accuracy of local oscillators. To be precise, phase noise is the phase difference between the phase of the local oscillator and phase of the carrier signal [22], [25]. Phase noise reduces the signal-to-noise ratio (SNR), increases the bit error rate (BER), and limits the data rate [24], [26]. In addition, it destroys coherency between the channel estimate and the actual channel gain. Thus, one solution is to eliminate phase noise via hardware improvements. But that requires stringent constraints on the fabrication of RF components, increasing the cost [27]. Due to all these reasons, phase noise estimation and compensation algorithms are essential for the accurate channel estimation and data detection for GFDM.

A. RELATED WORKS

The impacts of phase noise on OFDM have been widely investigated [23], [28], [29], and phase noise estimation and compensation techniques have been developed. Thus, [30] develops an least-squares (LS) filter for phase noise compensation. In [26], [31], the phase noise process is parameterized with a sinusoidal waveform to develop an LS estimator. In contrast, linear-interpolation schemes [24], [27] improves phase noise estimation accuracy. Moreover, the estimator [32] uses a codebook, which is selected by minimizing the Euclidean distance between the known pilot symbols and the signal constellation. Additionally, the estimators in [33], [34] exploit the phase noise spectrum's geometrical structure.

However, the aforementioned OFDM works assume perfect channel state information at the receiver, an unrealistic assumption. Thus, channel estimation algorithms are also necessary and have been widely investigated. For instance, [35] investigates joint channel estimation and phase noise suppression, [36] does so with a power series for phase noise, or [37] does by using sequential Monte-Carlo and the expectation-maximization approach. Similar estimators that perform time-domain interpolation utilize data and pilot symbols [22], [25]. Finally, [38] uses the expectation conditional maximization method for joint channel, CFO and phase noise estimation.

Furthermore, the problem of FBMC channel estimation has been extensively studied [39]–[42]. For example, [39] develops an LS estimator that uses pilots but [42] deploys superimposed pilots. Moreover, phase noise impacts on FBMC systems are studied [43], showing that phase noise degrades the quality of channel estimation. Therefore, it develops an LS phase noise compensation algorithm that uses pilots.

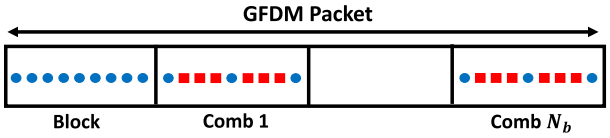


FIGURE 1. Timing of symbols within a GFDM packet.

On the other hand, the impacts of phase noise, timing offset, and CFO on the signal-to-interference ratio (SIR) of GFDM systems are studied in [44]. Phase noise is shown to exacerbate ICI and ISI terms and degrade the system's SIR performance. Moreover, GFDM full-duplex transceivers and cognitive radio networks with phase noise, CFO and IQ imbalances are studied in [13], [14], [21], [45]. Therefore, phase noise estimation and compensation are necessary to prevent the degradation of GFDM performance.

It is important to emphasize that while phase noise compensation has been widely investigated for OFDM, this is not the case for GFDM. For example, GFDM studies [11], [18], [46] address channel estimation only, without considering phase noise. Likewise, [47]–[49] consider joint channel and CFO estimation only. Similarly, [50]–[52] compensate for IQ imbalance in GFDM. Finally, [53] compensates for IQ imbalance and phase noise in multicarrier systems, which encompasses GFDM as a special case. The focus of [53] differs from our present paper. Thus, [53] does not estimate joint channel and phase noise, derive the Cramér-Rao lower bounds (CRLBs), nor analyze the signal-to-interference-plus noise ratio (SINR) and sum-rate of GFDM. To the best of our knowledge, joint channel estimation and phase noise compensation for GFDM systems have not been investigated before.

B. PROBLEM TACKLED HEREWITH AND CONTRIBUTIONS

In this paper, we address the following problem: how can joint channel estimation, phase noise compensation and data detection be performed for GFDM efficiently? This problem is essential and challenging for the reasons mentioned above. In particular, because of its non-orthogonal subcarriers, GFDM is limited by inherent ICI and ISI terms, which are boosted by phase noise. That is why phase noise compensation for GFDM is essential. This fact motivates the development of phase noise estimation and compensation algorithms.

We consider a quasi-static frequency selective channel constant during a GFDM packet and changes from packet to packet [38]. Each packet has $N_b + 1$ GFDM symbols, where N_b may be selected depending on the wireless channel's coherence time. In contrast, phase noise varies from one symbol to another, and thus symbol-wise estimation of

phase noise is needed, a necessity for accurate data detection. In contrast, a single channel estimate can be used for all symbols in a packet.

In the GFDM packet (Fig. 1), blue circles and red squares indicate pilot and data symbols, respectively. The first symbol is a GFDM pilot block, which is known at the receiver. The remaining symbols are comb-type symbols in the payload portion of the packet (data and pilot subcarriers). The estimation process has two stages focusing on 1) channel estimation, and 2) data detection. Both stages compensate for phase noise. The first stage performs channel estimation. The second stage uses these estimates to detect each subsequent symbols' data. Moreover, to reduce the complexity of phase noise estimation, we extend the interpolation techniques from [25], which were initially designed for OFDM receivers, for GFDM systems.

Specifically, this paper makes the following contributions regarding phase noise compensation for GFDM systems:

- We propose an efficient channel and phase noise estimator based on the non-linear least-squares (NLS) approach for the first stage. Moreover, for the second stage, we propose two algorithms for joint data detection and phase noise compensation: 1) iterative linear LS algorithm, and 2) closed-form LS algorithm. All algorithms utilize pilot symbols and deploy time interpolation to reduce computational complexity.
- To quantify the impacts of phase noise, we derive the effective SINR for three cases: 1) ideal compensation of phase noise, 2) no compensation for phase noise, and 3) compensation of phase noise with the proposed algorithm. Furthermore, we analyze the GFDM sum-rate for Case 3.
- To gauge the accuracy of our channel and phase noise estimators, we derive their CRLBs. The CRLB is a lower bound on the mean-squared error (MSE) of an unbiased estimator [54]. A trade-off exists between the accuracy and complexity of the estimators. Thus, we also provide their complexity analysis.
- Finally, extensive numerical and theoretical results are presented. These consider matched-filter (MF) and zero-forcing (ZF) GFDM receivers and compare their achievable sum-rates. Importantly, we show that the proposed estimators significantly reduce GFDM receivers' sensitivity to phase noise impairments.

We should emphasize that all of the proposed algorithms and performed analyses are formulated based on low-complexity GFDM parameters, including transmitter and receiver filters.

In [25], authors propose an iterative algorithm for joint channel and phase noise estimation for OFDM based on the linear LS approach. Since that algorithm does not estimate channel and phase noise simultaneously, the estimation error can be high. In contrast, we use the NLS approach, achieve a low estimation error and reach the CRLB, which proves its efficiency. Reference [25] also develops an iterative linear LS algorithm for data detection. Following the same approach, we propose an iterative data detection algorithm. However, we also offer a closed-form LS algorithm, which reduces the complexity and improves the BER performance.

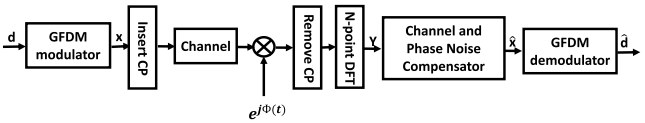


FIGURE 2. GFDM block diagram with phase noise.

This paper is organized as follows. With the system model (Section II), Section III develops the algorithms for channel and phase noise estimation and for data detection and phase noise compensation. Section IV analyzes the effective SINR, sum-rate, CRLB and computational complexity. In Section V, simulation results, theoretical results, and the MSE and BER performances of the proposed algorithms are compared and verified. Finally, Section VI presents the conclusions.

Notations: Gaussian variable $X \sim N(\mu, \sigma^2)$ has mean μ and variance σ^2. The $[n, m]$-th element of $\mathbf{A}$ is $\mathbf{A}[n, m]$. The N-point DFT (discrete fourier transform) matrix is denoted by $\mathbf{F}_N$, while $\mathbf{I}_N$ and $\mathbf{0}_N$ are is the identity and all zero matrix matrices of size $N \times N$. The superscripts $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^{-1}$ indicate the complex conjugation, transpose, Hermitian transpose and matrix inversion. Circular convolution between two vectors $\mathbf{x_1} \in \mathbb{C}^{1\times N}$ and $\mathbf{x_2} \in \mathbb{C}^{1\times N}$ is defined as $x_1[n] \circledast x_2[n] = \sum_{k=0}^{N-1} x_1[n]x_2[n-k]$. Moreover, $\mathbf{A} \circ \mathbf{B}$ indicates the Hadamard product [55] of matrices $\mathbf{A}$ and $\mathbf{B}$. The diagonal matrix formed by vector $\mathbf{x}$ is $\text{diag}(\mathbf{x})$. $\mathbb{Re}\{z\}$ and $\mathbb{Im}\{z\}$ denote the real and imaginary parts of z. $\mathbb{Tr}\{\mathbf{A}\}$ is the trace of $\mathbf{A}$, and $\mathbb{E}\{\cdot\}$ indicates the statistical expectation. The rows of circulant matrix $\mathbf{A}$ are formed by circular right shifts of the elements of vector $\mathbf{a}$ [56].

II. SYSTEM MODEL

Fig. 2 shows the GFDM block diagram with phase noise. Each GFDM symbol has M time-slots and K subcarriers and consists of $N = MK$ independent and identically distributed (i.i.d.) complex data symbols, $\mathbf{d} = [d_0, \ldots, d_{N-1}]^T \in \mathbb{C}^{N\times 1}$, with zero mean and $\mathbb{E}\{\mathbf{d}\mathbf{d}^H\} = \sigma_d^2 \mathbf{I}_N$. The data vector $\mathbf{d}$ can be decomposed into K subvectors $\mathbf{d} = [\mathbf{d}_0^T, \ldots, \mathbf{d}_K^T]^T \in \mathbb{C}^{N\times 1}$, where $\mathbf{d}_k = [d_{kM}, \ldots, d_{kM+M-1}]^T \in \mathbb{C}^{M\times 1}$, for $k = 0, 1, \ldots, K-1$, denotes the M complex data symbols for the k-th subcarrier. In the GFDM modulator, after upsampling data symbol d_{kM+m} by a factor of K, the corresponding output is circularly convoluted with the normalized prototype filter $g[n]$ and then up-converted to its corresponding subcarrier frequency. Hence, the discrete GFDM signal per symbol-time interval, $\mathbf{x} = [x[0], \ldots, x[N-1]]^T \in \mathbb{C}^{N\times 1}$, may be expressed as [10]

$$\mathbf{x} = \mathbf{F}_N^H \sum_{k=0}^{K-1} \tilde{\mathbf{P}}_k \tilde{\mathbf{\Psi}} \tilde{\mathbf{\Lambda}} \mathbf{F}_M \mathbf{d}_k, \tag{1}$$

where $\tilde{\mathbf{\Lambda}} = [\overbrace{\mathbf{I}_M, \ldots, \mathbf{I}_M}^{\mathrm{K}}] \in \mathbb{C}^{N\times M}$ is the repetition matrix, $\tilde{\mathbf{\Psi}} \in \mathbb{C}^{N\times N}$ is the diagonal frequency-domain filtering matrix that has $\sqrt{M}\mathbf{F}_N(g[n])$ as its diagonal elements, and $\tilde{\mathbf{P}}_k \in \mathbb{C}^{N\times N}$ is a permutation matrix to up-convert the k-th subcarrier to its corresponding frequency. To reduce the size of matrices in (1), the frequency response of prototype filter $g[n]$ is typically designed to be zero everywhere except for LM samples, where L is called the repetition factor [57]. This paper uses the mostly-considered choice of $L = 2$ [10], [52], [57]. Thus, the GFDM signal (1) can be simplified as [57]

$$\mathbf{x} = \mathbf{F}_N^H \sum_{k=0}^{K-1} \mathbf{P}_k \mathbf{\Psi} \mathbf{\Lambda} \mathbf{F}_M \mathbf{d}_k, \tag{2}$$

where $\mathbf{\Lambda} = [\mathbf{I}_M, \mathbf{I}_M]^T \in \mathbb{C}^{2M\times M}$, and

$$\mathbf{\Psi} = \begin{bmatrix} \mathbf{\Psi_1} & \mathbf{0}_M \\ \mathbf{0}_M & \mathbf{\Psi_2} \end{bmatrix} = \mathrm{diag}(\gamma_0, \gamma_1, \ldots, \gamma_{2M-1}) \in \mathbb{C}^{2M\times 2M}, \tag{3}$$

where $\{\gamma_0, \gamma_1, \ldots, \gamma_{2M-1}\}$ indicates the frequency response of the prototype filter. Moreover, $\mathbf{P}_k \in \mathbb{C}^{N\times 2M}$ is a subcarrier mapping, which is given by [52]

$$\mathbf{P}_k = \begin{bmatrix} \mathbf{0}_M & \cdots & \mathbf{0}_M & \underbrace{\mathbf{I}_M}_{\text{(k+1)-th}} & \cdots & \mathbf{0}_M \\ \mathbf{0}_M & \cdots & \mathbf{I}_M & \mathbf{0}_M & \cdots & \mathbf{0}_M \end{bmatrix}^T. \tag{4}$$

Moreover, the GFDM signal (2) can be re-expressed as

$$\mathbf{x} = \mathbf{\Xi}\mathbf{d}, \tag{5}$$

where $\mathbf{\Xi} = \mathbf{F}_N^H \mathbf{P} \mathrm{diag}(\mathbf{W}, .., \mathbf{W}) \in \mathbb{C}^{N\times N}$ is the GFDM modulator matrix, $\mathbf{W} = \mathbf{\Psi}\mathbf{\Lambda}\mathbf{F}_M \in \mathbb{C}^{2M\times M}$, and $\mathbf{P} = [\mathbf{P}_0, \ldots, \mathbf{P}_{K-1}] \in \mathbb{C}^{N\times 2N}$. To prevent the ISI caused by the multipath channel, a cyclic prefix with length L_{cp} is added. After digital-to-analog conversion, the signal is transmitted through the quasi-static frequency selective channel. The received baseband signal after the removal of cyclic prefix may be expressed as [58]

$$y[n] = e^{j\phi[n]}(h[n] \circledast x[n]) + w[n], \;\; n = 0, 1, \ldots, N-1, \tag{6}$$

where $\phi[n]$ is the random phase noise of the local oscillator, $h[n]$ is the equivalent discrete-time baseband channel impulse response with length L_{ch}, and $w[n]$ is additive white Gaussian noise with variance σ_w^2. The channel is modeled as a quasi-static frequency-selective fading channel [52]. The channel length is shorter than the cyclic prefix ($L_{ch} - 1 \leq L_{cp}$).

In addition, to model phase noise, the standard free-running oscillator with a Brownian motion process [44] is deployed, which is widely used for simulations and mathematical analyses [21], [22], [44], [58]. Because the difference between two samples of a Brownian motion is Gaussian distributed, $\phi[n] - \phi[n + n_0] \sim \mathcal{N}(0, 4\pi\beta n_0 T_s)$, where $\phi[n]$ is Brownian motion with 3-dB bandwidth of β and the autocorelation of $\phi[n]$ is [38], [44]

$$\mathbb{E}\left\{e^{j\phi[n_1]}e^{-j\phi[n_2]}\right\} = e^{-2|n_1-n_2|\pi\beta T_s}, \tag{7}$$

where T_s is sampling time. In practice, the variance of phase noise $4\pi\beta T_s$ is small, e.g., the estimated variance of a free running oscillator operating at 2.8 GHz is 10^{-4} rad^2 [59], [60].

Following the above details, the output samples in frequency domain can be derived by taking the DFT of the received baseband signal $y[n]$ (6) as

$$Y[l] = J[l] \circledast (H[l]X[l]) + W_n[l], \;\; l = 0, 1, \ldots, N-1, \tag{8}$$

where

- $J[l] = \frac{1}{N}\sum_{n=0}^{N-1} e^{j\phi[n]}e^{-j\frac{2\pi ln}{N}}$ is the frequency response of phase noise,
- $H[l] = \sum_{n=0}^{L_{ch}-1} h[n]e^{-j\frac{2\pi ln}{N}}$ is the channel frequency response,
- $X[l] = \sum_{n=0}^{N-1} x[n]e^{-j\frac{2\pi ln}{N}}$ is the transmitted GFDM symbols in the frequency domain,
- $W_n[l] = \sum_{n=0}^{N-1} w[n]e^{-j\frac{2\pi ln}{N}}$ is the frequency response of additive white Gaussian noise.

The channel coefficients $H[l]$ are circularly symmetric Gaussian variables with mean zero and variance $\sigma_h^2 = \mathbb{E}\{|H[l]|^2\}$ [25], [52]. To reveal the impact of phase noise, circular convolution (8) must be expanded. This expansion yields the received signal (8) as

$$Y[l] = J[0]H[l]X[l] + \sum_{r=1}^{N-1} J[(l-r)_N]H[r]X[r] + W_n[l], \tag{9}$$

where $(l-r)_N$ stands for $((l-r)\mathrm{mod}N)$. Equation (9) indicates that phase noise causes two types of distortions: 1) common phase error, $J[0]$, and 2) ICI, which is the second term of (9). By using matrix notation, (9) can be compactly represented as

$$\mathbf{y} = \mathbf{J}\mathbf{H}\mathbf{x} + \mathbf{W_n}, \tag{10}$$

where

- $\mathbf{y} = [Y[0], \ldots, Y[N-1]]^T \in \mathbb{C}^{N\times 1}$ is received signal vector,
- $\mathbf{J} = \begin{bmatrix} J[0] & J[N-1] & \cdots & J[1] \\ J[1] & J[0] & \cdots & J[2] \\ \vdots & \vdots & \ddots & \vdots \\ J[N-1] & J[N-2] & \cdots & J[0] \end{bmatrix} \in \mathbb{C}^{N\times N}$ is circulant phase noise matrix,
- $\mathbf{H} = \mathrm{diag}([H[0], \ldots, H[N-1]]) \in \mathbb{C}^{N\times N}$ is the channel frequency response matrix,
- $\mathbf{x} = \mathbf{P}\mathrm{diag}(\mathbf{W}, \ldots, \mathbf{W})\mathbf{d} = [X[0], \ldots, X[N-1]]^T \in \mathbb{C}^{N\times 1}$ is the GFDM data vector,
- $\mathbf{W_n} = [W_n[0], \ldots, W_n[N-1]]^T \in \mathbb{C}^{N\times 1}$ is the additive white Gaussian noise vector.

If there is no phase noise, $\mathbf{J}$ becomes the identity matrix.

III. JOINT CHANNEL ESTIMATION AND PHASE NOISE COMPENSATION

To detect the data vector $\mathbf{x}$ in (10), channel matrix $\mathbf{H}$ and phase noise matrix $\mathbf{J}$ are required. However, in practice, the receiver does not know them and must estimate them. This scenario is widely investigated (e.g., for OFDM systems see [61]–[63] and references therein). Thus, we next propose three algorithms. Algorithm 1 uses one block-type pilot GFDM symbol for joint channel and phase noise estimation in the first stage of the estimation process. In the second stage, Algorithm 2 or Algorithm 3 can be used for data detection. They both utilize comb-type GFDM symbols to for phase noise compensation and data detection. Note that the channel estimate from Stage 1 is used in Stage 2; and thus last two algorithms use the output from Algorithm 1. This usage is possible because wireless channels are usually slowly varying compared to phase noise, which may change significantly from one GFDM symbol to another. Thus, previous phase noise estimates cannot detect the data symbols in the payload portion of the packet and must be estimated per symbol for accurate data detection. However, the first stage's channel estimate can be used during the whole packet.

A. JOINT CHANNEL AND PHASE NOISE ESTIMATION

Herein, we propose Algorithm 1 for joint estimation of channel coefficients and phase noise. It deploys one block-type pilot symbol, i.e., all $N = MK$ subsymbols are known at the receiver. The output (10) motivates the following optimization problem for this estimation:

$$\min_{\mathbf{H},\mathbf{J}} \|\mathbf{y} - \mathbf{JHx}\|^2. \tag{11}$$

This problem (11) can be solved via the LS approach [64]. That is, $\mathbf{H}$ and $\mathbf{J}$ are estimated by minimizing the squared Euclidean distance between the observation vector $\mathbf{y}$ and the reconstructed noiseless observation $\mathbf{JHx}$ [64]. LS problems can be linear (standard) or non-linear, depending on whether the cost function is a linear function of the unknown parameters or not [65]. While the linear case has a closed-from solution, the non-linear case is solved through successive iterative algorithms and by linearizing the non-linear cost function [65]. Clearly, the problem at hand (11) is non-linear function of the entries of $\mathbf{H}$ and $\mathbf{J}$ because of the matrix product. This fact motivates an NLS estimator. To this end, the proposed algorithm exploits two ideas:

1) To save computational complexity, we first reduce the number of unknown channel and phase noise parameters, $2N-1$, by relating the frequency response of channel to its time response and by interploating phase noise.
2) We approximate the resulting NLS problem into linear one. With this idea, we develop an iterative channel and phase noise estimator.

Step 1: To this end, we rewrite the received signal in frequency domain (10) as

$$\mathbf{y} = \mathbf{J}'\mathbf{H}'\mathbf{x} + \mathbf{W_n}, \tag{12}$$

where

$$\mathbf{H}' = J[0]\mathbf{H} = \text{diag}\big([H'[0], \ldots, H'[N-1]]\big) \in \mathbb{C}^{N\times N}$$
$$\mathbf{J}' = \frac{1}{J[0]}\mathbf{J} = \begin{bmatrix} J'[0] & J'[N-1] & \cdots & J'[1] \\ J'[1] & J'[0] & \cdots & J'[2] \\ \vdots & \vdots & \ddots & \vdots \\ J'[N-1] & J'[N-2] & \cdots & J'[0] \end{bmatrix} \in \mathbb{C}^{N\times N}, \tag{13}$$

where $H'[l] = J[0]H[l]$ $l = 0, 1, \ldots, N-1$, $J'[0] = 1$ and $J'[l] = J[l]/J[0]$, $l = 1, \ldots, N-1$. Moreover, $\mathbf{J}'$ is circulant matrix which is formed by elements of $\tilde{\mathbf{J}} = [J'[0], J'[1], \ldots, J'[N-1]]^T \in \mathbb{C}^{N\times 1}$.

Note that there are N channel coefficients $H'[l]$, $l = 0, \ldots, N-1$, and $N-1$ phase noise parameters $J'[l]$, $l = 1, \ldots, N-1$. Since the number of known pilots, N, is less than the number of unknown parameters, $2N-1$, this problem is underdetermined and the solution is not unique. To avoid this conundrum, we model the channel and phase noise with fewer parameters to reduce the number of unknowns.

1) CHANNEL MODEL

Since the channel length, L_{ch}, is much smaller than the GFDM symbol size N, similar to [52], we can relate channel frequency response $H'[l]$, $l = 0, \ldots, N-1$, to its time response $h'[n]$, $n = 0, \ldots, L_{ch}-1$, through

$$\tilde{\mathbf{h}} = \mathbf{F}_h\bar{\mathbf{h}}, \tag{14}$$

where

$$\tilde{\mathbf{h}} = \big[H'[0], \ldots, H'[N-1]\big]^T \in \mathbb{C}^{N\times 1}$$
$$\bar{\mathbf{h}} = \big[h'[0], \ldots, h'[L_{ch}-1]\big]^T \in \mathbb{C}^{L_{ch}\times 1}$$
$$\mathbf{F}_h = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{-j\frac{2\pi}{N}} & \cdots & e^{-j\frac{2\pi(L-1)}{N}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j\frac{2\pi(N-1)}{N}} & \cdots & e^{-j\frac{2\pi(N-1)(L-1)}{N}} \end{bmatrix} \in \mathbb{C}^{N\times L_{ch}}, \tag{15}$$

where $h'[l] = J[0]h[l]$, $l = 0, 1, \ldots, L_{ch}-1$. Moreover, $\mathbf{F}_h$ contains the first L_{ch} columns of the DFT matrix $\mathbf{F}_N$. In this way, we estimate $\bar{\mathbf{h}}$ instead of $\tilde{\mathbf{h}}$, which reduces the number of unknown parameters into L_{ch}.

2) PHASE NOISE MODEL

To reduce the number of unknowns, phase noise is interpolated. Thus, time domain phase noise components $e^{j\phi[n]}$, $n = 0, \ldots, N-1$ can be interpolated from $e^{j\phi[q(N-1)/(Q-1)]}$, $q = 0, \ldots, Q-1 (Q < N)$ as [25]

$$\tilde{\mathbf{p}} \approx \boldsymbol{\Upsilon}\bar{\mathbf{p}}, \tag{16}$$

where

$$\tilde{\mathbf{p}} = \left[e^{j\phi[0]}, e^{j\phi[1]}, \ldots, e^{j\phi[N-1]}\right]^T \in \mathbb{C}^{N\times 1}$$
$$\bar{\mathbf{p}} = \left[e^{j\phi[0]}, e^{j\phi[(N-1)/(Q-1)]}, \ldots, e^{j\phi[N-1]}\right]^T \in \mathbb{C}^{Q\times 1}, \quad (17)$$

and $\boldsymbol{\Upsilon} \in \mathbb{C}^{N\times Q}$ is called interpolation matrix. It helps to approximate phase noise with fewer time samples. Thus, instead of estimating $J'[l],\ l = 0, \ldots, N-1$, we compute approximate phase noise samples. With phase noise component $J[l]$ and interpolation scheme (16), we find

$$\tilde{\mathbf{J}} = \frac{1}{N}\mathbf{F}_N\tilde{\mathbf{c}} \approx \mathbf{F}_c\bar{\mathbf{c}}, \quad (18)$$

where $\mathbf{F}_c = \frac{1}{N}\mathbf{F}_N\boldsymbol{\Upsilon} \in \mathbb{C}^{N\times Q}$, $\tilde{\mathbf{c}} = \frac{1}{J[0]}\tilde{\mathbf{p}}$, and

$$\tilde{\mathbf{J}} = \left[J'[0], J'[1], \ldots, J'[N-1]\right]^T \in \mathbb{C}^{N\times 1}$$
$$\bar{\mathbf{c}} = \frac{1}{J[0]}\bar{\mathbf{p}} = [c[0], \ldots, c[Q-1]]^T \in \mathbb{C}^{Q\times 1}. \quad (19)$$

Therefore, $\bar{\mathbf{c}}$ can be estimated instead of $\tilde{\mathbf{J}}$, which appears the number of unknown parameters to be Q. However, since $J'[0] = 1$, the number of parameters is in fact $Q-1$.

Remark 1 (Interpolation Matrix $\boldsymbol{\Upsilon}$): This matrix can be determined in two ways. First, if power spectral density (PSD) of phase noise is unknown, a non-optimal, interpolation matrix $\boldsymbol{\Upsilon}$ can be constructed from linear interpolation as [25]

$$\boldsymbol{\Upsilon}_L[n_1, m_1] = \begin{cases} m_1 - \frac{(n_1-1)(Q-1)}{N-1} & \frac{(m_1-1)(N-1)}{Q-1} \le n_1 - 1 \le \frac{m_1(N-1)}{Q-1} \\ \frac{(n_1-1)(Q-1)}{N-1} - (m_1-2) & \frac{(m_1-2)(N-1)}{Q-1} \le n_1 - 1 \le \frac{(m_1-1)(N-1)}{Q-1} \\ 0, & otherwise \end{cases} \quad (20)$$

where $n_1 = 1, 2, \ldots, N$ and $m_1 = 1, 2, \ldots, Q$. Second, if the PSD of phase noise is known, the optimal interpolation matrix can be derived by minimizing the MSE of interpolating $\tilde{\mathbf{p}}$ from $\bar{\mathbf{p}}$ as

$$\boldsymbol{\Upsilon}_O = \underset{\boldsymbol{\Upsilon}}{arg\min}\, \mathbb{E}\|\tilde{\mathbf{p}} - \boldsymbol{\Upsilon}\bar{\mathbf{p}}\|^2. \quad (21)$$

Thus, the solution of (21) is the optimal interpolation matrix, which is given by

$$\boldsymbol{\Upsilon}_O = \mathbf{R}_{\tilde{\mathbf{p}},\bar{\mathbf{p}}}\mathbf{R}_{\bar{\mathbf{p}}}^{-1}, \quad (22)$$

where $\mathbf{R}_{\tilde{\mathbf{p}},\bar{\mathbf{p}}} = \mathbb{E}\{\tilde{\mathbf{p}}\bar{\mathbf{p}}^H\}$ and $\mathbf{R}_{\bar{\mathbf{p}}} = \mathbb{E}\{\bar{\mathbf{p}}\bar{\mathbf{p}}^H\}$, which based on the autocorelation of $\phi[n]$ in (7) and phase noise expressions in (17) can be derived as

$$\mathbf{R}_{\tilde{\mathbf{p}},\bar{\mathbf{p}}}[n_1, m_1] = e^{-2|(n_1-1)-\frac{(m_1-1)(N-1)}{Q-1}|\pi\beta T_s}$$
$$\mathbf{R}_{\bar{\mathbf{p}}}[m_1, m_2] = e^{-2|(m_1-m_2)|\frac{N-1}{Q-1}\pi\beta T_s}, \quad (23)$$

where $n_1 = 1, 2, \ldots, N$ and $m_1, m_2 = 1, 2, \ldots, Q$. Thus, by substituting (23) in (22), we get the optimal interpolation matrix. Note that, it depends on the bandwidth, β, of phase noise. To determine the bandwidth, the receiver requires the PSD of phase noise. The PSD of the free-running oscillator typically is the Lorenzian spectrum [66], which can be estimated from two single point PSD measurements [67]. Therefore, it is not unrealistic to assume that the receiver knows β through prior measurements.

Consequently, the optimization problem in (11) is converted to joint $\bar{\mathbf{h}}$ (14) and $\bar{\mathbf{c}}$ (18) estimation problem as

$$\min_{\bar{\mathbf{h}},\bar{\mathbf{c}}} \ \left\|\mathbf{y} - \mathbf{J}'\mathbf{H}'\mathbf{x}\right\|^2 \quad (24)$$
$$s.t.\ \ J'[0] = 1, \quad (25)$$

where $\mathbf{H}'$ is related to $\bar{\mathbf{h}}$ by (14) and $\mathbf{J}'$ is related to $\bar{\mathbf{c}}$ by (18). Thus, by relating frequency response of channel to its time response in (14) and also deploying interpolation technique for approximating phase noise components in (16), the number of unknowns parameters in (11) from $2N-1$ into $L_{ch}+Q-1$ in (24). However, the resulting optimization problem (24) is non-linear with a non-convex cost function [25], [52].

Step 2: To solve the estimation problem (24), we introduce an NLS iterative estimator of $\bar{\mathbf{h}}$ and $\bar{\mathbf{c}}$. This algorithm employs local linearization with a first-order approximation. First, initialize $\bar{\mathbf{c}}$ by setting $\hat{\mathbf{c}}_0 = [1, 1, \ldots, 1]^T \in \mathbb{C}^{Q\times 1}$. Second, by substituting it in (18) and (13), and deriving the $\mathbf{J}'_0$, the optimal $\bar{\mathbf{h}}$ of (24) is given by

$$\hat{\mathbf{h}}_0 = \left(\mathbf{F}_h^H\mathbf{X}^H{\mathbf{J}'_0}^H\mathbf{J}'_0\mathbf{X}\mathbf{F}_h\right)^{-1}\mathbf{F}_h^H\mathbf{X}^H{\mathbf{J}'_0}^H\mathbf{y}, \quad (26)$$

where $\mathbf{X} = \mathrm{diag}(\mathbf{x})$. Let $\hat{\mathbf{h}}_{i-1}$ and $\hat{\mathbf{c}}_{i-1}$ be the parameter estimates at the $(i-1)$-th iteration. If $\Delta\bar{\mathbf{h}}$ and $\Delta\bar{\mathbf{c}}$ are the estimation errors, the parameter estimates in the i-th iteration are given by

$$\hat{\mathbf{h}}_i = \hat{\mathbf{h}}_{i-1} + \Delta\bar{\mathbf{h}}, \hat{\mathbf{c}}_i = \hat{\mathbf{c}}_{i-1} + \Delta\bar{\mathbf{c}}. \quad (27)$$

Furthermore, we have

$$\hat{\mathbf{H}}'_i = \hat{\mathbf{H}}'_{i-1} + \Delta\mathbf{H}',\ \hat{\mathbf{J}}'_i = \hat{\mathbf{J}}'_{i-1} + \Delta\mathbf{J}', \quad (28)$$

where estimation errors $\Delta\mathbf{H}'$ and $\Delta\mathbf{J}'$ are related to $\Delta\bar{\mathbf{h}}$ and $\Delta\bar{\mathbf{c}}$ via (14) and (18), respectively. By substituting (27) and (28) in (24) and using (14) and (18), we rewrite the optimization problem (24) to estimate $\Delta\bar{\mathbf{h}}$ and $\Delta\bar{\mathbf{c}}$ as

$$\min_{\Delta\bar{\mathbf{h}},\Delta\bar{\mathbf{c}}} \ \left\|\mathbf{y} - \left(\mathbf{J}'_{i-1} + \Delta\mathbf{J}'\right)\left(\mathbf{H}'_{i-1} + \Delta\mathbf{H}'\right)\mathbf{x}\right\|^2$$
$$\approx \min_{\Delta\bar{\mathbf{h}},\Delta\bar{\mathbf{c}}} \left\|\mathbf{z} - \mathbf{T}'_{i-1}\mathbf{F}_c\Delta\bar{\mathbf{c}} - \hat{\mathbf{J}}'_{i-1}\mathbf{X}\mathbf{F}_h\Delta\bar{\mathbf{h}}\right\|^2 \quad (29)$$
$$s.t.\ \ \mathbf{G}\Delta\bar{\mathbf{c}} = 0, \quad (30)$$

where $\mathbf{z} = \mathbf{y} - \hat{\mathbf{J}}'_{i-1}\hat{\mathbf{H}}'_{i-1}\mathbf{x} \in \mathbb{C}^{N\times 1}$, $\mathbf{T}'_{i-1} \in \mathbb{C}^{N\times N}$ is the circulant matrix formed by elements of $\hat{\mathbf{H}}'_{i-1}\mathbf{x}$, and $\mathbf{G} = [G[0], G[1], \ldots, G[Q-1]] \in \mathbb{C}^{1\times Q}$ is the first row of $\mathbf{F}_c = \frac{1}{N}\mathbf{F}_N\boldsymbol{\Upsilon}$ matrix. Note that, this approximation ignores the cross term between $\Delta\bar{\mathbf{h}}$ and $\Delta\bar{\mathbf{c}}$. Moreover, based on the constraint in (30), we can remove $\Delta\bar{c}[0]$ from parameter

estimates by substituting $\mathbf{F}'_c\Delta\bar{\mathbf{c}}'$ instead $\mathbf{F}_c\Delta\bar{\mathbf{c}}$ in (29). We have

$$\mathbf{F}_c\Delta\bar{\mathbf{c}} = \mathbf{F}'_c\Delta\bar{\mathbf{c}}', \tag{31}$$

where

$$\begin{aligned}\Delta\bar{\mathbf{c}}' &= \left[\Delta c'[1], \ldots, \Delta c'[Q-1]\right]^T \in \mathbb{C}^{Q-1\times 1}\\ \mathbf{F}'_c &= \mathbf{S}_2 - \frac{1}{G[0]}\mathbf{S}_1([G[1], G[2], \ldots, G[Q-1]])\\ \mathbf{F}_c &= [\mathbf{S}_1|\mathbf{S}_2],\ \mathbf{S}_1 \in \mathbb{C}^{N\times 1},\ \mathbf{S}_2 \in \mathbb{C}^{N\times Q-1}.\end{aligned} \tag{32}$$

In this way, by applying (31), constraint (30) is satisfied and optimization problem in (29) can be written as

$$\min_{\Delta\bar{\mathbf{h}},\Delta\bar{\mathbf{c}}'} \left\|\mathbf{z} - \mathbf{T}'_{i-1}\mathbf{F}'_c\Delta\bar{\mathbf{c}}' - \hat{\mathbf{J}}'_{i-1}\mathbf{X}\mathbf{F}_h\Delta\bar{\mathbf{h}}\right\|^2. \tag{33}$$

For solving (33), we decompose all the parameters that should be estimated into their real and imaginary parts and put them together in a vector $\boldsymbol{\Theta} \in \mathbb{C}^{2(L_{ch}+Q-1)\times 1}$ as

$$\boldsymbol{\Theta} = \left[\mathbb{R}\mathrm{e}\{\bar{\mathbf{h}}\}^T, \mathbb{I}\mathrm{m}\{\bar{\mathbf{h}}\}^T, \mathbb{R}\mathrm{e}\{\bar{\mathbf{c}}'\}^T, \mathbb{I}\mathrm{m}\{\bar{\mathbf{c}}'\}^T\right]^T. \tag{34}$$

Thus, by deploying (34), we reformat the non-linear optimization problem in (33) as a linear LS problem, given by

$$\min_{\boldsymbol{\Theta}} \|\bar{\mathbf{z}} - \boldsymbol{\Gamma}\Delta\boldsymbol{\Theta}\|^2, \tag{35}$$

where $\bar{\mathbf{z}} = [\mathbb{R}\mathrm{e}\{\mathbf{z}\}^T, \mathbb{I}\mathrm{m}\{\mathbf{z}\}^T]^T \in \mathbb{C}^{2N\times 1}$ and matrix $\boldsymbol{\Gamma} \in \mathbb{C}^{2N\times 2(L_{ch}+Q-1)}$ is derived in (72) (Appendix A). The LS solution of (35) yields the optimal estimate of unknown matrix $\Delta\boldsymbol{\Theta}$ in the i-th iteration as

$$\Delta\boldsymbol{\Theta} = \left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\bar{\mathbf{z}}. \tag{36}$$

Accordingly, from (36) and (34), $\Delta\bar{\mathbf{h}}$ and $\Delta\bar{\mathbf{c}}'$ are derived. Then, by using (31), $\Delta\bar{\mathbf{c}}$ is derived from $\Delta\bar{\mathbf{c}}'$. Finally, estimates of $\bar{\mathbf{h}}$ and $\bar{\mathbf{c}}$ will be updated by (27).

The proposed joint channel and phase noise estimation method is summarized in Algorithm 1. It recursively works until the objective function stops improving or until it reaches the maximum number of iterations.

Note that in Step 4 of Algorithm 1, near optimal incremental terms $\Delta\bar{\mathbf{h}}$ and $\Delta\bar{\mathbf{c}}$ are derived by (36) through setting $\boldsymbol{\Theta}$ in (34) as predetermined $\hat{\Theta}_{i-1}$, which results in $\|\bar{\mathbf{z}}-\boldsymbol{\Gamma}\Delta\boldsymbol{\Theta}\|^2_{\boldsymbol{\Theta}=\hat{\Theta}_{i-1}} \leq \|\bar{\mathbf{z}}-\boldsymbol{\Gamma}\Delta\boldsymbol{\Theta}\|^2_{\boldsymbol{\Theta}=\hat{\Theta}_i}$ [25], [52]. Thus, the objective function decreases with the number of iterations and converges to a local minimum. The convergence and accuracy of this algorithm is experimentally investigated in Section V. Moreover, the output of the algorithm, which is the channel estimate, is utilized in Algorithm 2 for data detection and phase noise estimation.

Algorithm 1 Joint Channel and Phase Noise Estimation

1: Set $\hat{\mathbf{c}}_0 = [1, 1, \ldots, 1]^T$ and find the initial $\hat{\mathbf{h}}_0$ by using (26)
2: Set the maximum number of iteration $I_{\max}$ and $i = 1$.
3: **do while** $|\mho_i - \mho_{i-1}| \leq \epsilon$, and $i < I_{\max}$,
4: Derive $\Delta\bar{\mathbf{h}}$ and $\Delta\bar{\mathbf{c}}$ by finding $\Delta\boldsymbol{\Theta}_o$ in (36).
5: Update the estimates on $\bar{\mathbf{h}}$ and $\bar{\mathbf{c}}$ by (27).
6: $i \leftarrow i + 1$
7: **dend do**
where $\mho_i = \|\mathbf{y} - \mathbf{J}'_i\mathbf{H}'_i\mathbf{x}\|^2$

B. JOINT DATA SYMBOL AND PHASE NOISE ESTIMATION

This algorithm uses comb-type GFDM symbols and uses the first-stage channel estimate. It performs phase noise compensation and data detection. Consider the vector $\mathbf{d} = \mathbf{d}_d + \mathbf{d}_p$, where $\mathbf{d}_d \in \mathbb{C}^{N\times 1}$ and $\mathbf{d}_p \in \mathbb{C}^{N\times 1}$ indicate data vector and pilots sequence, respectively. Furthermore, we assume one pilot subsymbol every Δk subcarriers for pilot sequence $\mathbf{d}_p$ and we keep the rest of subsymbols zeros, which are the position of data vector $\mathbf{d}_d$. Thus, the number of pilot subcarriers is $K_p = \lfloor \frac{K}{\Delta k} \rfloor$.

Let $\mathbf{v} = [v_1, v_2, \ldots, v_P]^T \in \mathbb{C}^{P\times 1}$ shows the position of pilot symbols in data vector $\mathbf{d}$, where $P = MK_p$ indicates the total number of pilot symbols. Then, we have $\mathbf{d}_p = [d^{v_1}, 0, 0, \ldots, 0, d^{v_2}, \ldots, d^{v_{P-1}}, 0, 0, \ldots, 0, d^{v_P}]$, and $\mathbf{d}_d = \mathbf{d} - \mathbf{d}_p$. Furthermore, similar to the first stage, interpolation is employed to reduce the complexity of phase noise estimation.

Two algorithms are next developed for joint data and phase noise estimation (stage 2). First, we follow a similar approach as in [25] and propose an iterative algorithm to solve the joint NLS estimation problem. This algorithm first detects GFDM symbols, passes the output through the GFDM demodulator, and finally outputs the transmitted complex data symbols. Second, we propose a closed-form LS algorithm as well, which includes the GFDM demodulator and directly detects transmitted complex data symbols.

1) ALGORITHM 2

The data detection and phase noise estimation problem can be formulated as

$$\min_{\mathbf{x}_d,\bar{\mathbf{c}}} \left\|\mathbf{y} - \mathbf{J}\mathbf{H}'\left(\mathbf{x}_p + \mathbf{x}_d\right)\right\|^2, \tag{37}$$

where $\mathbf{y}$ is the observed signal, $\mathbf{J}$ is the circulant matrix formed by the elements of $\mathbf{F}_c\bar{\mathbf{c}}$ in (18), $\mathbf{x}_d = \mathbf{P}\mathrm{diag}(\mathbf{W}, \ldots, \mathbf{W})\mathbf{d}_d \in \mathbb{C}^{N\times 1}$ and $\mathbf{x}_p = \mathbf{P}\mathrm{diag}(\mathbf{W}, \ldots, \mathbf{W})\mathbf{d}_p \in \mathbb{C}^{N\times 1}$.

As with the original channel and phase noise estimation problem (11), (37) is solved via the LS approach [64]. Thus, $\mathbf{x}_d$ and $\bar{\mathbf{c}}$ by minimizing the squared distance between the observation vector $\mathbf{y}$ and the reconstructed noiseless observation $\mathbf{J}\mathbf{H}'(\mathbf{x}_p + \mathbf{x}_d)$. Note that (37) is a non-linear function of unknown parameters of $\mathbf{x}_d$ and $\bar{\mathbf{c}}$ with a non-convex cost function. Therefore, it has no closed-form solution [65].

Algorithm 2 Joint Data and Phase Noise Estimation

1: Set $\hat{\mathbf{c}}_0 = [\hat{J}[0], \hat{J}[0], ..., \hat{J}[0]]^T \in \mathbb{C}^{Q\times 1}$
2: Set $I_{\max}$ and $i = 1$.
3: **do while** $|\varrho_i - \varrho_{i-1}| \leq \epsilon$, and $i < I_{\max}$
4: Compute optimal data $\hat{\mathbf{x}}_{i-1}$ by (38).
5: Compute phase noise $\hat{\mathbf{c}}_i$ by (39).
6: $i \leftarrow i + 1$
7: **end do**
where $\varrho_i = \|\mathbf{y} - \mathbf{J}_i\mathbf{H}'(\mathbf{x}_p + \hat{\mathbf{x}}_{d,i-1})\|^2$.

Therefore, similar to the first stage, we propose an iterative algorithm. It has two steps: 1) detecting the data symbol by assuming that phase noise is known, 2) estimating the phase noise with the detected data symbol. The algorithm runs until the predefined maximum number of iterations is reached, or the cost function reaches a plateau.

Step 1: Suppose that vector $\hat{\mathbf{c}}_{i-1}$ (estimation of $\bar{\mathbf{c}}$) and phase noise matrix $\hat{\mathbf{J}}_{i-1}$ are known in the $(i-1)$-th iteration. By solving the linear LS problem, the optimal data symbol is given by

$$\hat{\mathbf{x}}_{d,i-1} = \left(\mathbf{H}'\right)^{-1}\left(\hat{\mathbf{J}}_{i-1}^H\hat{\mathbf{J}}_{i-1}\right)^{-1}\hat{\mathbf{J}}_{i-1}^H\left(\mathbf{Y} - \hat{\mathbf{J}}_{i-1}\mathbf{H}'\mathbf{x}_p\right). \quad (38)$$

Step 2: By deploying estimated symbol in (38), we find the optimal $\hat{\mathbf{c}}_i$ by solving a linear LS problem as

$$\hat{\mathbf{c}}_i = \left(\mathbf{F}_c^H\mathbf{T}''^H_{i-1}\mathbf{T}''_{i-1}\mathbf{F}_c\right)^{-1}\mathbf{F}_c^H\mathbf{T}''^H_{i-1}\mathbf{Y}, \quad (39)$$

where $\mathbf{T}''_{i-1} \in \mathbb{C}^{N\times N}$ is the circulant matrix formed by elements of $\mathbf{H}'(\hat{\mathbf{x}}_{d,i-1} + \mathbf{x}_p)$.

The iterative joint data and phase noise estimation is summarized in Algorithm 2. Note that in order to correct the scaling ambiguity in $\mathbf{H}'$, we estimate the common phase error $J[0]$ by using LS method to minimize the cost function, given by

$$\min_{J[0]} \sum_{l=0}^{N-1} \left|Y[l] - J[0]H'[l]X_p[l]\right|^2, \quad (40)$$

where $X[l]$ is l-th element of vector $\mathbf{x}_p$. By solving the minimization problem, the estimate of common phase error is derived as

$$\hat{J}[0] = \frac{\sum_{l=0}^{N-1} H'^*[l]X_p^*[l]Y[l]}{\sum_{l=0}^{N-1}|H'[l]|^2|X_p[l]|^2}. \quad (41)$$

The proposed algorithm outputs the GFDM symbol $\hat{\mathbf{x}}$, and complex data symbols are demodulated as $\hat{\mathbf{d}} = \mathbf{B}\hat{\mathbf{x}}$, where $\mathbf{B} \in \mathbb{C}^{N\times N}$ is the demodulation matrix. The matrix depends on the type of the receiver. In this paper, we consider two common GFDM receivers; namely, MF and ZF [52]. The matrices for these are $\mathbf{B}_{\text{MF}} = \text{diag}(\mathbf{W}^H, \ldots, \mathbf{W}^H)\mathbf{P}^H$ and $\mathbf{B}_{\text{ZF}} = (\mathbf{P}\text{diag}(\mathbf{W}, \ldots, \mathbf{W}))^{-1}$.

Algorithm 3 Joint Data and Phase Noise Estimation

1: Compute phase noise unknowns $\hat{\dot{\mathbf{p}}}$ by (45).
2: Compute data symbols $\hat{\mathbf{d}}$ via $\hat{\dot{\mathbf{p}}}$ in (43).

2) ALGORITHM 3

With the transmit GFDM signal (5) and (10), the observed signal with comb-type GFDM symbols may be expressed as

$$\mathbf{y} = \mathbf{F}_N\mathbf{E}\mathbf{F}_N^H\mathbf{H}'\mathbf{F}_N\boldsymbol{\Xi}\mathbf{d} + \mathbf{W}_\mathbf{n}, \quad (42)$$

where $\boldsymbol{\Xi}$ indicates GFDM modulator (5), $\mathbf{d} = \mathbf{d}_d + \mathbf{d}_p$, and $\mathbf{E} = \text{diag}(\tilde{\mathbf{p}})$ indicates the phase noise matrix in time domain. In (42), we convert the multiplication of channel and GFDM signal into time domain, and then multiply the output with the phase noise matrix in time domain, and finally convert the result to the frequency domain. After passing the observed signal (42) through the phase noise compensator and GFDM demodulator, transmitted complex data symbols can be detected as

$$\hat{\mathbf{d}} = \mathbf{C}\dot{\mathbf{p}}, \quad (43)$$

where $\mathbf{C} = \mathbf{B}\mathbf{F}_N^H(\mathbf{H}')^{-1}\mathbf{F}_N\hat{\mathbf{y}}\boldsymbol{\Upsilon} \in \mathbb{C}^{N\times Q}$, where $\mathbf{B} \in \mathbb{C}^{N\times N}$ is the GFDM demodulation matrix and $\hat{\mathbf{y}} = \text{diag}(\mathbf{F}_N^H\mathbf{y}) \in \mathbb{C}^{N\times N}$ is the observed signal in time domain. Moreover, in (43), we approximate conjugate of phase noise $\tilde{\mathbf{p}}^*$ by using interpolation matrix $\boldsymbol{\Upsilon}$ as $\tilde{\mathbf{p}}^* = \boldsymbol{\Upsilon}\dot{\mathbf{p}}$, where $\dot{\mathbf{p}} \in \mathbb{C}^{Q\times 1}$ is the unknown vector. Let $\mathbf{d}_p^v = [d_p^{v_1}, d_p^{v_2}, \ldots, d_p^{v_P}]^T \in \mathbb{C}^{P\times 1}$ be pilot vector corresponding to pilot index set $\mathbf{v}$. Therefore, we can have

$$\mathbf{d}_p^v = \mathbf{C}_P\dot{\mathbf{p}}, \quad (44)$$

where $\mathbf{C}_P \in \mathbb{C}^{P\times Q}$ matrix corresponds to pilot index set $\mathbf{v}$, which is derived from matrix $\mathbf{C}$ in (43). Since $\mathbf{d}_p^v$ and $\mathbf{C}_P$ are known, we can estimate the unknown vector $\dot{\mathbf{p}}$ by using linear LS estimator as

$$\hat{\dot{\mathbf{p}}} = \left(\mathbf{C}_P^H\mathbf{C}_P\right)^{-1}\mathbf{C}_P^H\mathbf{d}_p^v. \quad (45)$$

By substituting the estimated $\hat{\dot{\mathbf{p}}}$ in (43), the transmitted complex data symbols can be obtained. This closed-form joint data and phase noise estimation is summarized in Algorithm 3.

IV. PERFORMANCE ANALYSIS

This section investigates the impacts of phase noise on the effective SINR and the sum-rate of GFDM. Moreover, the CRLB is derived to evaluate the MSE performance of Algorithm 1. Finally, the complexity of the proposed algorithms is analyzed.

A. EFFECTIVE SINR DERIVATION

The effective SINR can quantify the impacts of phase noise on GFDM performance. Thus, we derive this measure for three different cases. These are 1) Ideal compensation for phase noise, 2) No compensation of phase noise, and

3) Phase noise compensation with Algorithm 1. As is customary, data symbols, phase noise, the channel coefficients, and the additive white Gaussian noise terms are assumed to be independent random variables [52], [58].

1) CASE 1: IDEAL COMPENSATION OF PHASE NOISE

If both $\mathbf{J}$ and $\mathbf{H}$ matrices in (10) are perfectly known, the receiver can perform ideal compensation. Although this case is unrealizable, it yields an upper bound of the achievable effective SINR, which serves as a performance benchmark for any phase noise compensation algorithm. To derive this bound, we note that the term $\mathbf{JHx}$ in (10) is the desired component. Thus, based on $\mathbb{E}\{\mathbf{W_n}^H\mathbf{W_n}\} = \sigma_w^2$, the effective SINR can be derived as

$$\begin{aligned}\text{SINR}_{\text{ideal}} &= \frac{\mathbb{E}\{\mathbf{x}^H\mathbf{H}^H\mathbf{J}^H\mathbf{JHx}\}}{\sigma_w^2} \\ &= \frac{\sigma_h^2\sigma_d^2}{N\sigma_w^2}\sum_{l_1=0}^{N-1}\left(|\gamma_{l_1'}|^2 + |\gamma_{M+l_1'}|^2\right), \end{aligned} \tag{46}$$

where $l_1' = (l_1)_M$.

Proof: See proof of (46) in Appendix B. ■

As can be observed in the expression (46), the derived effective SINR is independent of the 3-dB phase noise bandwidth β, which makes sense since phase noise is ideally compensated.

2) CASE 2: NO COMPENSATION OF PHASE NOISE

This case yields a lower bound of achievable effective SINR, which gauges the phase-noise compensation algorithm's performance improvement. To this end, we next derive the effective SINR. To do that, we rewrite the received signal in (10) as

$$\mathbf{y} = \mathbf{Hx} + (J[0]-1)\mathbf{Hx} + (\mathbf{J} - J[0]\mathbf{I}_N)\mathbf{Hx} + \mathbf{W_n}. \tag{47}$$

Similar to the derivation in (46), by using the signal model (47) and after straightforward mathematical manipulation, the effective SINR is given by

$$\begin{aligned}&\text{SINR}_{\text{No}} \\ &= \frac{\mathbb{E}\{\mathbf{x}^H\mathbf{H}^H\mathbf{Hx}\}}{\left(\mathbb{E}\{|J[0]-1|^2\} + \sum_{l=1}^{N-1}\sigma_{J,l}^2\right)\mathbb{E}\{\mathbf{x}^H\mathbf{H}^H\mathbf{Hx}\} + \sigma_w^2} \\ &= \frac{\text{SINR}_{\text{ideal}}}{2(1-\mu)\text{SINR}_{\text{ideal}} + 1}, \end{aligned} \tag{48}$$

where $\sigma_{J,l}^2 = \mathbb{E}\{|J[l]|^2\}$ and $\mu = \mathbb{R}\text{e}\{\mathbb{E}\{J[0]\}$ (derived in (81), Appendix C). Note that the constant μ depends on phase noise bandwidth β. For example, if the bandwidth is zero ($\beta = 0$), μ will also be zero, and the SINR of Case 2 will be equal to that of the ideal case ($\text{SINR}_{\text{No}} = \text{SINR}_{\text{ideal}}$). Thus, depending on the noise bandwidth, the SINR of Case 2 is bounded as

$$\frac{\text{SINR}_{\text{ideal}}}{2\text{SINR}_{\text{ideal}} + 1} < \text{SINR}_{\text{No}} \leq \text{SINR}_{\text{ideal}}. \tag{49}$$

3) CASE 3: PROPOSED ALGORITHMS

The effective SINR achieved with the proposed phase noise compensation algorithm is derived. That SINR can be compared with the upper and lower bounds from case 1 and case 2, respectively. Thus, we can evaluate the effectiveness of the proposed algorithm. To derive the effective SINR, let $\hat{\mathbf{J}}$ be the estimate of phase noise matrix $\mathbf{J}$ by the proposed estimation algorithm. Thus, we can rewrite the system model (10) as

$$\mathbf{y} = \hat{\mathbf{J}}\mathbf{Hx} + \left(\mathbf{J} - \hat{\mathbf{J}}\right)\mathbf{Hx} + \mathbf{W_n}, \tag{50}$$

where $\mathbf{J}$ and $\hat{\mathbf{J}}$ are circulant matrices formed by the elements of $\tilde{\mathbf{j}} = \frac{1}{N}\mathbf{F}_N\tilde{\mathbf{p}}$ and $\tilde{\mathbf{j}}_{\mathbf{app}} = \mathbf{F}_c\bar{\mathbf{p}}$ in (18), respectively. Thus, we can formulate the effective SINR as

$$\begin{aligned}\text{SINR}_{\text{prop}} &= \frac{\mathbb{E}\left\{\mathbf{x}^H\mathbf{H}^H\hat{\mathbf{J}}^H\hat{\mathbf{J}}\mathbf{Hx}\right\}}{\mathbb{E}\left\{\mathbf{x}^H\mathbf{H}^H\left(\mathbf{J}-\hat{\mathbf{J}}\right)^H\left(\mathbf{J}-\hat{\mathbf{J}}\right)\mathbf{Hx}\right\} + \sigma_w^2} \\ &= \frac{\xi_1\text{SINR}_{\text{ideal}}}{(1+\xi_1 - 2\mathbb{R}\text{e}\{\xi_2\})\text{SINR}_{\text{ideal}} + 1}, \end{aligned} \tag{51}$$

where $\xi_1 = \frac{1}{N}\mathbb{T}\text{r}\{\boldsymbol{\Upsilon}\mathbf{R}_{\bar{\mathbf{p}}}\boldsymbol{\Upsilon}^H\}$ and $\xi_2 = \frac{1}{N}\mathbb{T}\text{r}\{\mathbf{R}_{\tilde{\mathbf{p}},\bar{\mathbf{p}}}\boldsymbol{\Upsilon}^{\mathbf{H}}\}$.

Proof: See proof of (51) in Appendix D. ■

Note that $\mathbf{R}_{\bar{\mathbf{p}}}$ and $\mathbf{R}_{\tilde{\mathbf{p}},\bar{\mathbf{p}}}$ are derived in (23) and $\boldsymbol{\Upsilon}_L$ is expressed in (20) for linear interpolation. Moreover, the effective SINR (51) derived for optimal interpolation matrix $\boldsymbol{\Upsilon}_O = \mathbf{R}_{\tilde{\mathbf{p}},\bar{\mathbf{p}}}\mathbf{R}_{\bar{\mathbf{p}}}^{-1}$ in (22) can be simplified as

$$\text{SINR}_{\text{prop},O} = \frac{\xi\text{SINR}_{\text{ideal}}}{(1-\xi))\text{SINR}_{\text{ideal}} + 1}, \tag{52}$$

where $\xi = \frac{1}{N}\mathbb{T}\text{r}\{\mathbf{R}_{\tilde{\mathbf{p}},\bar{\mathbf{p}}}\mathbf{R}_{\bar{\mathbf{p}}}^{-1}\mathbf{R}_{\tilde{\mathbf{p}},\bar{\mathbf{p}}}^H\}$. If the length, Q, of interpolation vector equals the number of data symbols N, we have $\tilde{\mathbf{p}} = \bar{\mathbf{p}}$ and thus $\mathbf{R}_{\tilde{\mathbf{p}},\bar{\mathbf{p}}} = \mathbf{R}_{\bar{\mathbf{p}}}$. In this case, $\xi = 1$. Moreover, $\bar{\mathbf{p}}$ is a subvector of $\tilde{\mathbf{p}}$ and so $\mathbf{R}_{\bar{\mathbf{p}}}$ is a submatrix of $\mathbf{R}_{\tilde{\mathbf{p}},\tilde{\mathbf{p}}}$. Therefore, it is true that $\mathbb{T}\text{r}\{\mathbf{R}_{\tilde{\mathbf{p}},\bar{\mathbf{p}}}\mathbf{R}_{\bar{\mathbf{p}}}^{-1}\mathbf{R}_{\tilde{\mathbf{p}},\bar{\mathbf{p}}}^H\} > \mathbb{T}\text{r}\{\mathbf{R}_{\bar{\mathbf{p}}}\mathbf{R}_{\bar{\mathbf{p}}}^{-1}\mathbf{R}_{\bar{\mathbf{p}}}^H\} = Q$. In this case, $\xi > \frac{Q}{N}$. Depending on the range of ξ, $\text{SINR}_{\text{prop},O}$ may be bounded as

$$\frac{\frac{Q}{N}\text{SINR}_{\text{ideal}}}{\left(1-\frac{Q}{N}\right)\text{SINR}_{\text{ideal}} + 1} < \text{SINR}_{\text{prop},O} \leq \text{SINR}_{\text{ideal}}. \tag{53}$$

B. SUM-RATE ANALYSIS

To gauge the proposed estimation/compensation algorithms, we analyze the sum rate of the GFDM system. The sum rate should increase with the elimination of interference terms. Thus, our analysis reveals if the proposed algorithms are effective or not in terms of eliminating them. To this end, the sum rate of GFDM over N data symbols can be formulated as [57]

$$R = \sum_{k'=0}^{K-1}\sum_{m'=0}^{M-1}\log_2\left(1 + \Phi_{k'M+m'}\right), \tag{54}$$

where $\Phi_{k'M+m'}$ indicates the SINR for the m'-th data symbol and the k'-th subcarrier. In order to derive the SINRs, the

desired signal component of the demodulator output (50) must be extracted. To do so, the signal model (50) is expressed as

$$\mathbf{y} = \mathbf{F}_N \mathbf{E}_{\mathbf{app}} \mathbf{F}_N^H \mathbf{H}\mathbf{x} + \mathbf{F}_N^H \big(\mathbf{E} - \mathbf{E}_{\mathbf{app}}\big)\mathbf{F}_N \mathbf{H}\mathbf{x} + \mathbf{W}_{\mathbf{n}}, \quad (55)$$

where $\mathbf{E}_{\mathbf{app}} = \text{diag}(\boldsymbol{\Upsilon}\tilde{\mathbf{p}})$, $\mathbf{E} = \text{diag}(\tilde{\mathbf{p}})$. Note that (55) converts the signal and channel multiplication in the frequency domain into the time domain and then multiplies the result with the phase noise matrix in the time domain converts it to the frequency domain finally. This transformation enables the derivation of the power of desired and interference components. With the demodulation of the signal (55), the m'-th data symbol on the k'-th subcarrier is detected as

$$\begin{aligned}\hat{d}_{k'M+m'} = \mathbf{F}_{M,m'}^H \boldsymbol{\Lambda}^H \boldsymbol{\Psi}_R \mathbf{P}_{k'}^H \mathbf{y} = \bar{d}_{k'M+m'} + I_{d,k'M+m'} \\ + I_{\phi,k'M+m'} + w_{k'M+m'}, \end{aligned} \quad (56)$$

where $\mathbf{F}_{M,m'}$ indicates the $m'+1$-th column of matrix $\mathbf{F}_M$ and $\boldsymbol{\Psi}_R = \text{diag}(\gamma_0^r, \gamma_1^r, \ldots, \gamma_{2M-1}^r) \in \mathbb{C}^{2M\times 2M}$, where $\{\gamma_0^r, \gamma_1^r, \ldots, \gamma_{2M-1}^r\}$ indicate the frequency response of the receiver prototype filter. In (56), $\bar{d}_{k'M+m'}$ indicates the desired symbol, $I_{d,k'M+m'}$ and $I_{\phi,k'M+m'}$ represent interference terms from other symbols and phase noise approximation, and $w_{k'M+m'}$ indicates the equivalent noise. By exploiting the GFDM signal (2), the components of (56) can be derived as

$$\begin{aligned}\bar{d}_{k'M+m'} &= \mathbf{U}_{k',m'} \mathbf{F}_N \mathbf{E}_{\mathbf{app}} \mathbf{F}_N^H \mathbf{H} \mathbf{C}_{k',m'} \mathbf{d}_{k'M+m'} \\ I_{d,k'M+m'} &= R_{d,k'M+m'} - \bar{d}_{k'M+m'} \\ R_{d,k'M+m'} &= \mathbf{U}_{k',m'} \mathbf{F}_N \mathbf{E}_{\mathbf{app}} \mathbf{F}_N^H \mathbf{H}\mathbf{x} \\ I_{\phi,k'M+m'} &= \mathbf{U}_{k',m'} \mathbf{F}_N \big(\mathbf{E} - \mathbf{E}_{\mathbf{app}}\big) \mathbf{F}_N^H \mathbf{H}\mathbf{x} \\ w_{k'M+m'} &= \mathbf{U}_{k',m'} \mathbf{W}_{\mathbf{n}}, \end{aligned} \quad (57)$$

where $\mathbf{U}_{k',m'} = \mathbf{F}_{M,m'}^H \boldsymbol{\Lambda}^H \boldsymbol{\Psi}_R \mathbf{P}_{k'}^H$ and $\mathbf{C}_{k',m'} = \mathbf{P}_{k'} \boldsymbol{\Psi} \boldsymbol{\Lambda} \mathbf{F}_{M,m'}$. Consequently, the SINR, $\Phi_{k'M+m'}$, for the m'-th data symbol on the k'-th subcarrier can be written as

$$\Phi_{k'M+m'} = \frac{P^s_{k'M+m'}}{P^{I^d}_{k'M+m'} + P^{I^\phi}_{k'M+m'} + P^w_{k'M+m'}}, \quad (58)$$

where

$$\begin{aligned} P^s_{k'M+m'} &= \mathbb{E}\Big\{\bar{d}_{k'M+m'} \bar{d}^H_{k'M+m'}\Big\} \\ P^{I^d}_{k'M+m'} &= P^{R^d}_{k'M+m'} - P^s_{k'M+m'} \\ P^{R^d}_{k'M+m'} &= \mathbb{E}\Big\{R_{d,k'M+m'} R^H_{d,k'M+m'}\Big\} \\ P^{I^\phi}_{k'M+m'} &= \mathbb{E}\Big\{I_{\phi,k'M+m'} I^H_{\phi,k'M+m'}\Big\} \\ P^w_{k'M+m'} &= \mathbb{E}\Big\{w_{k'M+m'} w^H_{k'M+m'}\Big\}. \end{aligned} \quad (59)$$

In Appendix E, we derive the powers of signal components in (59).

C. CRAMÉR-RAO LOWER BOUND (CRLB)

CRLB is a lower bound on the covariance matrix of any unbiased estimator of unknown parameters [54]. If an unbiased estimator achieves this lower bound, the estimator is said to be efficient. It thus achieves the smallest possible MSE among all unbiased methods. Thus, the CRLB is widely used to evaluate the quality of estimators. In the following, CRLBs are derived for the channel and phase noise estimates computed during the first block-type pilot of the GFDM packet (Fig. 1). We assume that the N pilot symbols are i.i.d. complex random variables with mean zero and the identical variance. They are also independent of data symbols, channel coefficients, phase noise, and additive noise.

To derive the CRLBs, the received signal (12) is expressed as

$$\mathbf{Y} = \mathbf{S}_{\boldsymbol{\Theta}} + \bar{\mathbf{W}}, \quad (60)$$

where

$$\mathbf{S}_{\boldsymbol{\Theta}} = \mathbf{J}'_{\mathbf{app}} \mathbf{H}'\mathbf{x}, \;\; \bar{\mathbf{W}} = \mathbf{W}_{\mathbf{n}} + \Big(\mathbf{J}' - \mathbf{J}'_{\mathbf{app}}\Big)\mathbf{H}'\mathbf{x}, \quad (61)$$

where $\mathbf{J}'$ and $\mathbf{J}'_{\mathbf{app}}$ are circulant matrices formed by the elements of $\tilde{\mathbf{J}} = \frac{1}{N}\mathbf{F}_N \tilde{\mathbf{c}}$ and $\boldsymbol{\Psi}_{\mathbf{app}} = \mathbf{F}_c \bar{\mathbf{c}}$ in (18). Note that $\mathbf{J}' - \mathbf{J}'_{\mathbf{app}}$ indicates the approximation error in (16). Similar to Appendix D, after straightforward derivations, the covariance matrix of $\bar{\mathbf{W}}$ is equal to $\mathbb{E}\{\bar{\mathbf{W}}\bar{\mathbf{W}}^{\mathbf{H}}\} = \sigma_{\bar{w}}^2 \mathbf{I}_N$, where $\sigma_{\bar{w}}^2$ can be derived as

$$\begin{aligned}\sigma_{\bar{w}}^2 &= \mathbb{E}\Big\{\mathbf{x}^H \mathbf{H}^H \Big(\mathbf{J}' - \mathbf{J}'_{\mathbf{app}}\Big)^H \Big(\mathbf{J}' - \mathbf{J}'_{\mathbf{app}}\Big)\mathbf{H}\mathbf{x}\Big\} + \sigma_w^2 \\ &= \mathbb{E}\Big\{\Big\|\tilde{\mathbf{J}} - \boldsymbol{\Psi}_{\mathbf{app}}\Big\|^2\Big\} \sigma_h^2 \sigma_d^2 \sum_{l_1=0}^{N-1} \Big(|\gamma_{l'_1}|^2 + |\gamma_{M+l'_1}|^2\Big) + \sigma_w^2 \\ &= \frac{\sigma_h^2 \sigma_d^2}{N} \sum_{l_1=0}^{N-1} \Big(|\gamma_{l'_1}|^2 + |\gamma_{M+l'_1}|^2\Big) \\ &\quad \times \left(1 + \frac{1}{N}\mathbb{T}\mathrm{r}\Big\{\boldsymbol{\Upsilon}\mathbf{R}_{\tilde{\mathbf{p}}}\boldsymbol{\Upsilon}^H\Big\} - 2\mathbb{R}\mathrm{e}\left\{\frac{1}{N}\mathbb{T}\mathrm{r}\Big\{\mathbf{R}_{\tilde{\mathbf{p}},\bar{\mathbf{p}}}\boldsymbol{\Upsilon}^{\mathbf{H}}\Big\}\right\}\right) \\ &\quad + \sigma_w^2, \end{aligned} \quad (62)$$

Since $\mathbf{R}_{\tilde{\mathbf{p}}}$ and $\mathbf{R}_{\tilde{\mathbf{p}},\bar{\mathbf{p}}}$ in (23) depend on phase noise bandwidth β, $\sigma_{\bar{w}}^2$ is a function of β. According to the CRLB, for any unbiased estimator $\hat{\Theta}$ of Θ, the covariance matrix, $\mathbf{R}_{\hat{\Theta}} = \mathbb{E}\{(\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)^H\}$, should satisfy

$$\mathbf{R}_{\hat{\Theta}} \geq \boldsymbol{\Omega}_{\hat{\Theta}}^{-1}, \quad (63)$$

where $\boldsymbol{\Omega}_{\hat{\Theta}} \in \mathbb{C}^{2(L_{ch}+Q-1)\times 2(L_{ch}+Q-1)}$ is the Fisher information matrix, which is given by

$$\boldsymbol{\Omega}_{\hat{\Theta}} = \frac{2}{\sigma_{\bar{w}}^2} \mathbb{R}\mathrm{e}\left\{\left(\frac{\partial \mathbf{S}_{\boldsymbol{\Theta}}}{\partial \boldsymbol{\Theta}}\right)^T \left(\frac{\partial \mathbf{S}_{\boldsymbol{\Theta}}}{\partial \boldsymbol{\Theta}}\right)^*\right\}. \quad (64)$$

Note that (63) implies that $\mathbf{R}_{\hat{\Theta}} - \boldsymbol{\Omega}_{\hat{\Theta}}^{-1}$ is positive semidefinite. However, $\partial \mathbf{S}_{\boldsymbol{\Theta}} / \partial \boldsymbol{\Theta}$ is derived in (89) element by element (see Appendix F). Therefore, by substituting (89) in Fisher information matrix (64), we have

$$\begin{aligned}\mathbb{E}\Big\{|h'[n_1] - \hat{h}'[n_1]|^2\Big\} \geq \boldsymbol{\Omega}_{\hat{\Theta}}^{-1}[n_1+1, n_1+1] \\ + \boldsymbol{\Omega}_{\hat{\Theta}}^{-1}[n_1+L_{ch}+1, n_1+L_{ch}+1] \\ n_1 = 0, 1, \ldots, L_{ch}. \end{aligned} \quad (65)$$

TABLE 1. Complexity growth of Algorithm 1.

Step number	Multiplications
Step 4	$\mathcal{O}(N^3 + 16N^2(L_{ch}+Q) + 8N(L_{ch}+Q)^2)$
Step 5	0

TABLE 2. Complexity growth of Algorithm 2.

Step number	Multiplications
Step 4	$\mathcal{O}(3N^3 + N^2Q + NQ)$
Step 5	$\mathcal{O}(N^3 + 2N^2Q)$

$$\mathbb{E}\left\{|c'[n_2] - \hat{c}'[n_2]|^2\right\} \geq \mathbf{\Omega}^{-1}_{\hat{\Theta}}[n_2 + 2L_{ch}, n_2 + 2L_{ch}] + \mathbf{\Omega}^{-1}_{\hat{\Theta}}[n_2 + 2L_{ch} + Q - 1, n_2 + 2L_{ch} + Q - 1] \quad n_2 = 1, \ldots, Q-1. \tag{66}$$

D. COMPUTATIONAL COMPLEXITY ANALYSIS

We analyze the number of real-valued multiplications required for joint channel and phase noise estimation in Algorithm 1 and joint data symbol and phase noise compensation in Algorithm 2.

In Algorithm 1, the iteration process contains Step 4 and Step 5. In Step 4, unknown matrix $\Delta\mathbf{\Theta}$ is optimally estimated via (36). This operation requires three matrix multiplications and one matrix inverse. Moreover, this step also constructs Γ matrix based on the expressions derived in Appendix A. Note that Step 5 avoids multiplications. Due to these reasons, the complexity growth is summarized in Table 1. Since the maximum value of Q is equal to N and also the channel length, L_{ch}, and the number of time-slots M is relatively small, the computational complexity of our proposed NLS joint channel and phase noise estimator is roughly in the order of K^3.

On the other hand, the iteration process of Algorithm 2 contains steps four and five. In the former one, the optimal data symbol is derived by (38). This operation requires six matrix multiplications and two matrix inverses. Moreover, in the latter one, phase noise estimation is obtained by (39), which requires six matrix multiplications and one matrix inverse. The complexity growth of the steps in each iteration is summarized in Table 2. Similar to Algorithm 1, since the maximum value of Q is equal to N and the number of time-slots M is relatively small, the computational complexity of our estimator is roughly in the order of K^3.

Finally, the proposed closed-form algorithm in Section III-B2, Algorithm 3, requires nine matrix multiplications and two matrix inversions, according to (43) and (45). The complexity of Algorithm 3 is $\mathcal{O}(N^3 + 5N^2Q + NQ^2)$. Similar to Algorithm 2, since the complexity is in order of N^3, $N = MK$, and the number of time-slots M is relatively small, the computational complexity of this estimator is roughly in the order of K^3. Note that this complexity of Algorithm 3 is less than that of Algorithm 2 because complexity per iteration is higher in the latter.

TABLE 3. Simulation parameters.

Parameter	Value
Number of subcarriers (K)	32
Number of time slots (M)	5
Modulation	16 – QAM
Prototype Filter	Root Raised-Cosine
Roll-off factor (r_f)	0.1
Sampling time (T_s) [58]	100 ns
Channel length (L_{ch})	26
Channel model [68]	ITU outdoor channel A
Power delay profile [68]	0 dB, -1 dB, -9 dB, -10 dB, -15 dB and -20 dB for delays of 0, 3, 7, 11, 17 and 25 samples
CP length (L_{cp})	26
Length of interpolation vector	$Q_1 = 10$; $Q_2 = 32$
3-dB phase noise bandwidth	$\beta_1 = 100$ Hz; $\beta_2 = 1$ kHz
SNR values (dB)	SNR_1=20; SNR_2=30; SNR_3=40
GFDM packet length ($N_b + 1$)	6
Pilot subcarrier spacing Δ_k	4
Maximum iterations, I_{max}	3

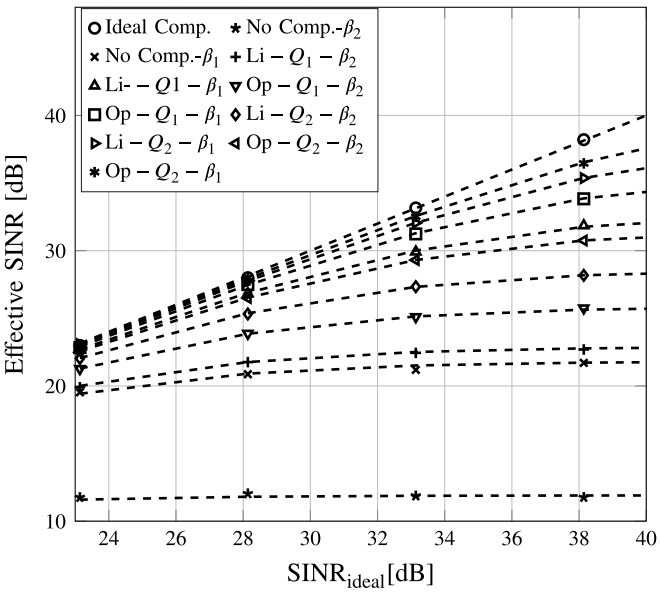


FIGURE 3. Effective SINRs versus SNR for Case 1 (legend "Ideal comp."), Case 2 (legend "No comp."), and Case 3, legends "Li" and "Op" represent linear and optimal interpolation.

V. SIMULATION RESULTS

This section measures the impacts of phase noise. First, the derived sum-rate and effective SINR expressions are verified via simulation results. Next, the MSE of the joint channel and phase noise estimator (Algorithm 1) and the CRLB are evaluated and compared. Finally, the BER of the joint data and phase noise estimator (Algorithm 2) is examined.

The simulation parameters are listed in Table 3. Each simulation result is averaged over 1000 independent trials. Pilot symbols are randomly generated complex 16-QAM (quadrature amplitude modulation) symbols. However, we have not tried to optimize the pilot symbols. Optimal pilot design for the considered system is an important and interesting topic and is left as future work.

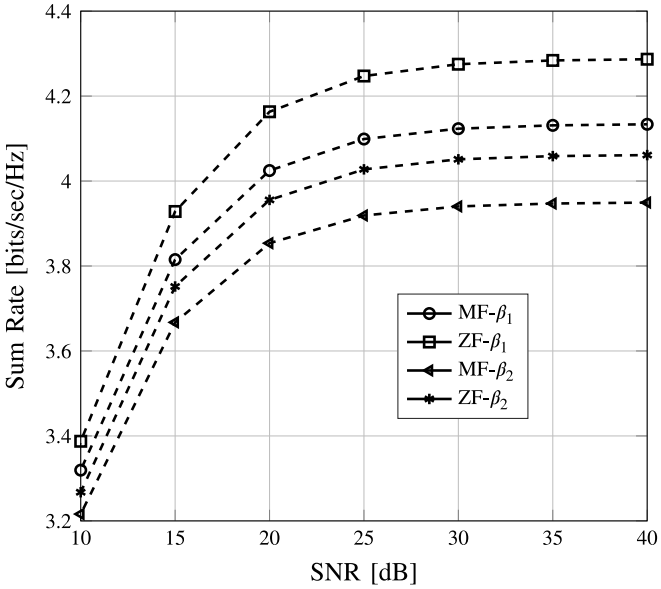


FIGURE 4. Sum-rate versus SNR for MF and ZF receivers.

In Fig. 3, we plot derived SINRs for Case 1 (46), Case 2 (48), and Case 3 (51) versus $\text{SINR}_{\text{ideal}}$ in (46). For these three cases, we consider both linear and optimal interpolation. Moreover, the phase noise bandwidth is set as 100 Hz or 1 kHz. Dashed lines show the theoretical results. The observations can be summarized as follows:

- We observe that, the simulation results fully match with the derived effective SINRs in (46), (48), (51).
- When the phase noise bandwidth increases, the effective SINRs of the last two cases decrease, e.g., without phase noise compensation, and when the phase noise increases from 100 Hz to 1 kHz at an SNR of 30 dB, the SINR decreases by about 10 dB. In other words, when phase noise increases, the power of interference terms increases, reducing the SINR.
- Promisingly, the proposed algorithm significantly improves effective SINR. Moreover, optimal interpolation achieves higher effective SINR than linear interpolation. Furthermore, a larger interpolation vector Q increases the effective SINR in both cases, e.g., optimal interpolation with Q_2 achieves an SINR more or less that of ideal case for 100-Hz phase noise.

We next compare the derived sum-rates of the MF and ZF receivers, with simulation results in Fig. 4. For a fair comparison, the interpolation size must be carefully chosen. But optimal interpolation with $Q = 32$ achieved a higher effective SINR (Fig. 3). Thus, we assume the same here. Fig. 4 shows that the derived results (dashed lines) and the simulation results match perfectly. Furthermore, ZF outperforms MF vis-à-vis the sum-rate. The reason is that the ZF receiver suppresses more interference terms than the MF receiver does. Moreover, when the phase noise increases, the sum-rate achieved by both ZF and MF receivers decreases, which is due to phase noise causing increased interference.

To examine the performance of Algorithm 1, we plot the MSE of channel and phase noise estimates in Fig. 5(a) and Fig. 5(b), respectively. Both linear and optimal interpolation are considered. Moreover, CRLB's of these estimations confirms the accuracy of Algorithm 1. Note that the number of iterations $I_{\max}$ is set at three, and phase noise bandwidth is 1 kHz.

The estimators developed thus far do not use statistics of the signals and channel gains. However, to compare the performance of Algorithm 1 with other estimators, we develop an iterative linear minimum mean-squared error (LMMSE) estimator. Unlike Algorithm 1, this estimator requires the channel correlation matrix, phase noise correlation matrix and the noise variance. In (24), at the i-th iteration, phase noise $\hat{\mathbf{c}}_{i-1}$ is assumed given. Therefore, the LMMSE channel estimate is computed by

$$\hat{\mathbf{h}}_i = \sigma_h^2 \mathbf{F}_h^H \mathbf{X}^H \mathbf{J}_i'^H \left(\sigma_w^2 \mathbf{I}_N + \sigma_h^2 \mathbf{J}'_{i-1} \mathbf{X} \mathbf{F}_h \mathbf{F}_h^H \mathbf{X}^H \mathbf{J}_{i-1}'^H \right)^{-1} \mathbf{y}, \tag{67}$$

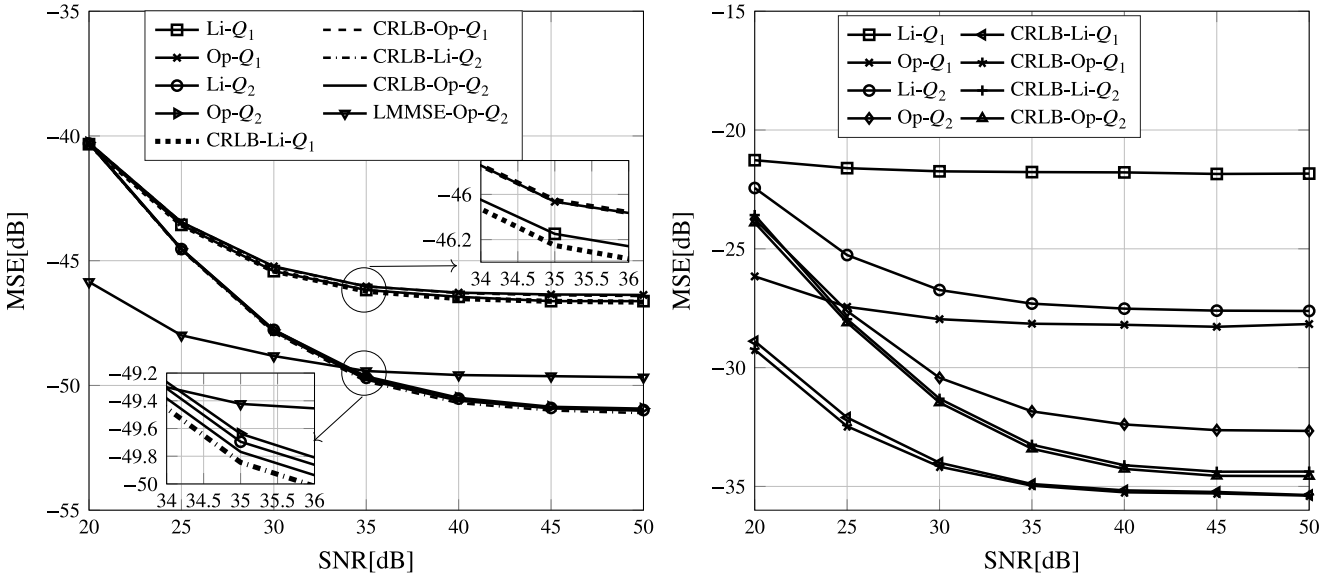
where $\mathbf{J}'_{i-1}$ is phase noise circulant matrix formed by elements of $\mathbf{F}_c \hat{\mathbf{c}}_{i-1}$. By using estimate $\hat{\mathbf{h}}_i$, LMMSE phase noise estimate is computed as

$$\hat{\mathbf{c}}_i = \mathbf{R}_{\tilde{\mathbf{p}}} \mathbf{F}_c^H \mathbf{T}_i'^H \left(\sigma_w^2 \mathbf{I}_N + \mathbf{T}'_i \mathbf{F}_c \mathbf{R}_{\tilde{\mathbf{p}}} \mathbf{F}_c^H \mathbf{T}_i'^H \right)^{-1} \mathbf{y}, \tag{68}$$

where $\mathbf{T}'_i$ is the circulant matrix formed by elements of $\hat{\mathbf{H}}'_i \mathbf{x}$, $\hat{\mathbf{H}}'_i = \text{diag}(\mathbf{F}_h \hat{\mathbf{h}}_i)$, and $\mathbf{R}_{\tilde{\mathbf{p}}}$ is phase noise correlation matrix in (23). Similar to Algorithm 1, the LMMSE estimator recursively works until the objective function stops improving or until it reaches the maximum number of iterations. In Fig. 5(a), for LMMSE, optimal interpolation with $Q = 32$ is considered and phase noise bandwidth is 1 kHz.

The observations of Fig. 5 can be summarized as follows:

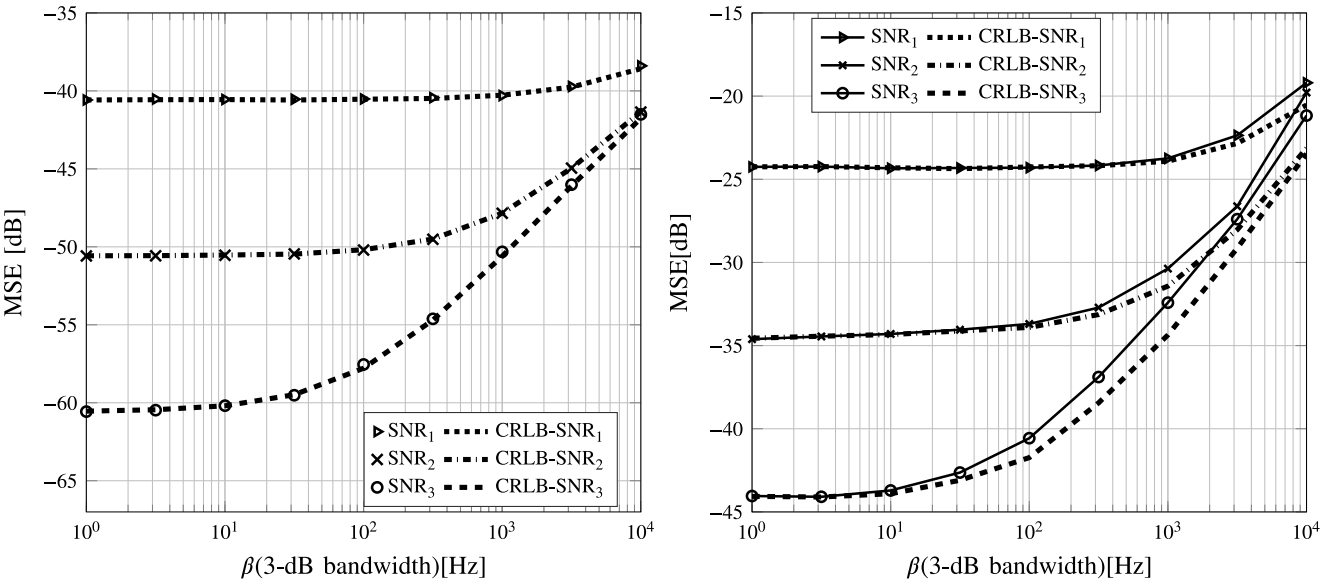
- The MSE of the channel estimates closely match the CRLB [Fig. 5(a)]. This match confirms that Algorithm 1 is an efficient unbiased estimator and its channel estimates can be reliably deployed for data detection.
- In addition, in low SNRs, the biased LMMSE estimator obtains lower MSE than the Algorithm 1. However, Algorithm 1 outperforms the LMMSE estimator in high SNRs.
- In Fig. 5(b), a gap exists between the MSE of phase noise estimates and the CRLB in all cases, which is due to the use of interpolation. However, this gap decreases when the interpolation size increases. Furthermore, the gap can be reduced by choosing optimal interpolation rather than linear interpolation.
- Promisingly, the MSE of both channel and phase noise estimates decreases for larger interpolation vectors, Q. For example, when we use optimal interpolation, the difference between the MSE of channel estimates for Q_1 and Q_2 at an SNR of 30 dB is 3 dB. However, larger Q increases the computational complexity. Therefore, complexity and MSE performance exhibit a trade-off.



(a) Estimated Channel

(b) Estimated Phase noise

FIGURE 5. MSE performance of Algorithm 1 with linear ("Li") and optimal ("Op") interpolations.



(a) Estimated Channel

(b) Estimated Phase noise

FIGURE 6. MSE of channel and phase noise estimates as a function of the phase noise bandwidth.

Moreover, to verify the impacts of phase noise on the performance of the proposed NLS estimator in Algorithm 1, MSE's of estimated channel and phase noise versus phase noise bandwidth β are illustrated in Fig. 6(a) and Fig. 6(b), respectively. Optimal interpolation ($Q = 32$) and also SNR values 20, 30 and 40 dB are considered. Furthermore, the number of iterations $I_{\max}$ is set at three. The observations from Fig. 6 are summarized as follows:

- From Fig. 6(a), close match between the MSE of channel estimates and CRLB is observed. This match shows that the proposed algorithm is able to compensates for the effect of different phase noise bandwidths.
- From Fig. 6(b), we observe that MSE of the phase noise estimate and CRLB are fairly close. However, a gap exists, which is due to the interpolation errors of phase noise. This points to the trade off between the MSE performance and computational complexity.
- Unsurprisingly, when phase noise bandwidth increases, the MSE's of both channel and phase noise estimation increase. Moreover, for the large phase noise regime, the MSE becomes more or less independent of the SNR,

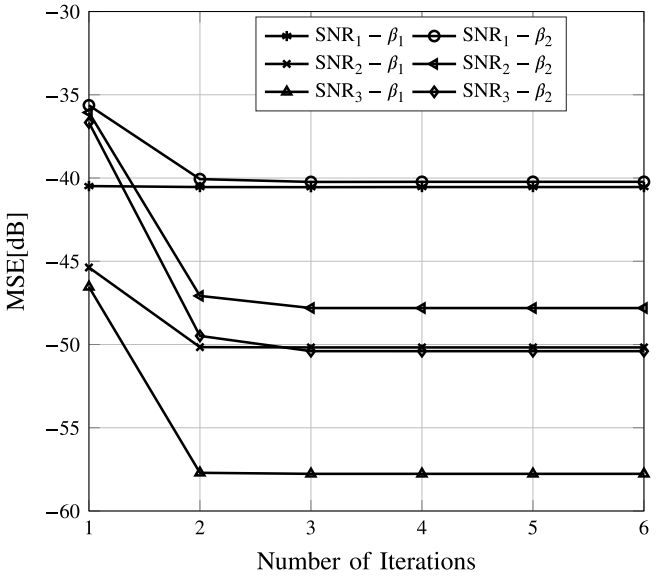


FIGURE 7. MSE of Algorithm 1 as a function of the number of iterations.

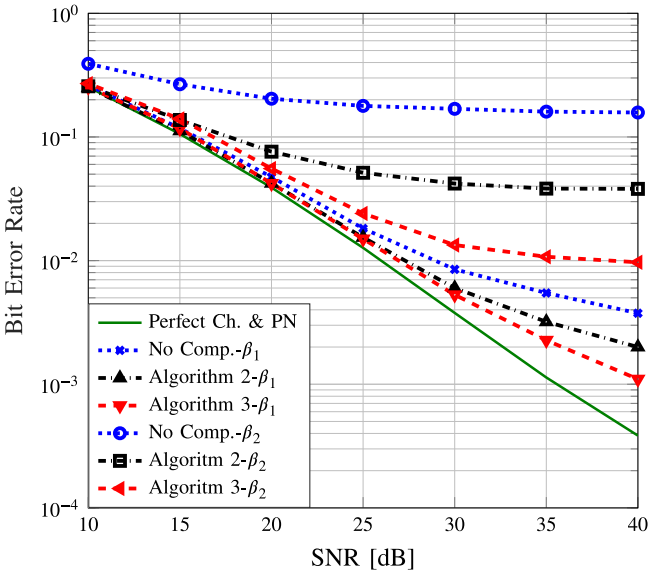


FIGURE 8. BER versus SNR for Case 1, Case 2 and Case 3 – Algorithm 2 and Case 3 – Algorithm 3.

and the power of interference terms (due to phase noise) increases and dominates the additive white noise terms.

To gauge the convergence rate of Algorithm 1, in Fig. 7, we plot the MSE of channel estimates (Algorithm 1) versus the number of iterations. The same SNR values as per Fig. 6 and 100 Hz and 1 kHz phase noise bandwidths, and optimal interpolation ($Q = 32$) are considered. We observe that for the 100-Hz phase noise, our algorithm converges after two iterations for all SNRs, while just three iterations are needed for the 1 kHz phase noise in high SNRs. These observations suggest that the proposed estimator is fast and requires just three iterations to ensure convergence.

Thus far, we have shown that the proposed joint NLS channel and phase noise estimator obtains satisfactory channel estimates, which can then be deployed for data detection. This is done in Algorithm 2. It uses comb-type GFDM symbols with pilot subcarrier spacing $\Delta_k = 4$. We consider Case 1, with perfect channel and phase noise knowledge, (legend Perfect Ch. & PN), Case 2, with perfect knowledge of channel but no information about the phase noise, (legend No comp.), Case 3 – Algorithm 2 in Section III-B1 (legend Algorithm 2), and Case 3 – closed-form data detection algorithm in Section III-B2 (legend Algorithm 3). Note that Case 1 yields the BER lower bound of BER and thus serves as a performance benchmark for the proposed algorithms. On the other hand, Case 2 yields the BER upper bound and is utilized to evaluate the improvement of the proposed algorithms.

Moreover, $I_{\max}$ in Algorithm 2 is set at three and also phase noise bandwidths values 100 Hz and 1 kHz and optimal interpolation are considered. Furthermore, since Fig. 4 illustrates that ZF receiver outperforms MF one in the presence of phase noise, we examine it exclusively in this figure. Note that the ZF receiver can cancel more interference terms than the MF, which boosts the desired signal and hence improves the BER. The observations of Fig. 8 can be summarized as follows:

- When the phase noise bandwidth increases, the BER increases in all cases. Undoubtedly, the increased phase noise degrades the MSE's of channel and phase noise estimates (Fig. 6). Fortunately, this degradation is more or less completely eliminated by both Algorithm 2 and Algorithm 3. For example, for 100 Hz phase noise, Algorithm 3 achieves a BER close to that of the perfect channel and phase noise estimation scenario.
- Promisingly, Algorithm 2 and Algorithm 3 compensate the phase noise effects and decrease the BER. In comparison with Case 1 and Case 2, they significantly improve the BER; e.g., for extreme phase noise (1 kHz) and additive noise (SNR of 30 dB), Algorithm 2 reduces the BER by 400% (compared to the no compensation case).
- Furthermore, Algorithm 3 outperforms Algorithm 2 and significantly decreases the BER, e.g., for 1 kHz phase noise and additive noise (SNR of 30 dB), Algorithm 3 reduces the BER by 300% in comparison with Algorithm 2. Besides, Algorithm 3 is closed-form; thus, its computational complexity is less than Algorithm 2, which requires more iterations for convergence.
- In Algorithm 3, the GFDM demodulation matrix $\mathbf{B}$ is a part of the data detector in (43). For the considered ZF receiver, the GFDM demodulator becomes $\mathbf{B}_{\text{ZF}} = (\mathbf{P}\text{diag}(\mathbf{W}, \ldots, \mathbf{W}))^{-1}$, which contains GFDM filter parameters. In contrast, Algorithm 2 is independent of the GFDM demodulator, since first GFDM symbols are detected and then are passed through the demodulator. Therefore, the performance gap between

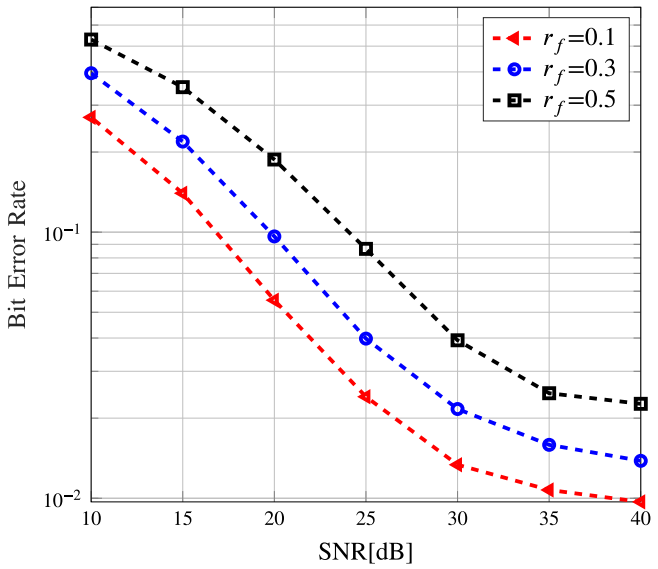


FIGURE 9. BER versus SNR for Algorithm 3 with three roll-off factors.

two algorithms is because of including GFDM demodulation in data detection in Algorithm 3 and handling the GFDM interference together with phase noise.

Finally, to show the impacts of non-orthogonality of GFDM on the system performance, in Fig. 9, we plot the BER versus SNR with three values of roll-off factor r_f for GFDM prototype filter $g[n]$. We consider Algorithm 3, phase noise bandwidth $\beta = 1$ kHz and other parameters as in Fig. 8. Clearly, the increasing roll-off factor results in the degradation of the BER. According to (43) in Algorithm 3, data detection depends on the GFDM modulator and demodulator matrices, which in turn are constructed based on the prototype filter. Thus, when the filter roll-off factor increases, the degree of non-orthogonality increases, which results in more interference terms and degrades the performance of data detection.

VI. CONCLUSION

This paper investigated phase noise compensation for GFDM systems. First, it proposed an estimator for joint channel and phase noise estimation, which utilizes the NLS approach and a block pilot GFDM symbol. Second, it also offered an iterative LS algorithm for data detection and phase noise compensation. Furthermore, to reduce computational complexity, a closed-form LS algorithm for data detection was developed. The complexity of all algorithms is reduced by deploying interpolation techniques and relating channel frequency and time responses.

Moreover, the impacts of phase noise on GFDM are quantified via the analysis of the SINR and sum-rate. The proposed algorithms improve the effective SINR. Additionally, the CRLBs for the channel and phase noise estimates were derived; and, we found that their MSE reaches the CRLB. Finally, the BER with the joint data detection and phase noise compensation algorithms was evaluated. Overall, it is possible to fully compensate for the impacts of phase noise and thus reduce GFDM receivers' sensitivity to RF impairments. For future research, this work can be extended in several directions. First, optimal pilot designs that use the MSE or a similar measure can be developed. Second, other RF impairments such as CFO and IQ imbalance can be incorporated with our proposed algorithms, and further afield, they may also be compensated for full-duplex transceivers [14], [21], [69].

APPENDIX A
DERIVATION OF Γ IN (35)

To obtain the matrix $\mathbf{\Gamma}$ derivation in (35), we first rewrite the cost function in (33) as

$$\begin{aligned}&\left\|\mathbf{z}-\mathbf{T}'_{i-1}\mathbf{F}'_c\Delta\bar{\mathbf{c}}'-\hat{\mathbf{J}}'_{i-1}\mathbf{X}\mathbf{F}_h\Delta\bar{\mathbf{h}}\right\|^2\\&=\left\|\mathbb{Re}\{\mathbf{z}\}-\mathbb{Re}\left\{\mathbf{T}'_{i-1}\mathbf{F}'_c\Delta\bar{\mathbf{c}}'\right\}-\mathbb{Re}\left\{\hat{\mathbf{J}}'_{i-1}\mathbf{X}\mathbf{F}_h\Delta\bar{\mathbf{h}}\right\}\right\|^2\\&\quad+\left\|\mathbb{Im}\{\mathbf{z}\}-\mathbb{Im}\left\{\mathbf{T}'_{i-1}\mathbf{F}'_c\Delta\bar{\mathbf{c}}'\right\}-\mathbb{Im}\left\{\hat{\mathbf{J}}'_{i-1}\mathbf{X}\mathbf{F}_h\Delta\bar{\mathbf{h}}\right\}\right\|^2.\end{aligned} \tag{69}$$

By using (34), we can have

$$\begin{bmatrix}\mathbb{Re}\{\mathbf{F}_h\Delta\bar{\mathbf{h}}\}\\ \mathbb{Im}\{\mathbf{F}_h\Delta\bar{\mathbf{h}}\}\\ \mathbb{Re}\{\mathbf{F}'_c\Delta\bar{\mathbf{c}}'\}\\ \mathbb{Im}\{\mathbf{F}'_c\Delta\bar{\mathbf{c}}'\}\end{bmatrix}=\mathbf{\Gamma}_1\mathbf{\Theta}, \tag{70}$$

where

$$\mathbf{\Gamma}_1=\begin{bmatrix}\mathbb{Re}\{\mathbf{F}_h\} & -\mathbb{Im}\{\mathbf{F}_h\} & \mathbf{0} & \mathbf{0}\\ \mathbb{Im}\{\mathbf{F}_h\} & \mathbb{Re}\{\mathbf{F}_h\} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbb{Re}\{\mathbf{F}'_c\} & -\mathbb{Im}\{\mathbf{F}'_c\}\\ \mathbf{0} & \mathbf{0} & \mathbb{Im}\{\mathbf{F}'_c\} & \mathbb{Re}\{\mathbf{F}'_c\}\end{bmatrix}. \tag{71}$$

Now, according to (69) and (70), $\mathbf{\Gamma}$ can be derived as

$$\mathbf{\Gamma}=\mathbf{\Gamma}_2\mathbf{\Gamma}_1, \tag{72}$$

where

$$\mathbf{\Gamma}_2=\begin{bmatrix}\mathbb{Re}\left\{\hat{\mathbf{J}}'_{i-1}\mathbf{X}\right\} & -\mathbb{Im}\left\{\hat{\mathbf{J}}'_{i-1}\mathbf{X}\right\} & \mathbb{Re}\{\mathbf{T}'_{i-1}\} & -\mathbb{Im}\{\mathbf{T}'_{i-1}\}\\ \mathbb{Im}\left\{\hat{\mathbf{J}}'_{i-1}\mathbf{X}\right\} & \mathbb{Re}\left\{\hat{\mathbf{J}}'_{i-1}\mathbf{X}\right\} & \mathbb{Im}\{\mathbf{T}'_{i-1}\} & \mathbb{Re}\{\mathbf{T}'_{i-1}\}\end{bmatrix}. \tag{73}$$

APPENDIX B
SINR (46) WITH IDEAL COMPENSATION

First, based on (10) and $\mathbb{E}\{\mathbf{d}\mathbf{d}^H\}=\sigma_d^2\mathbf{I}_N$, we derive $\mathbf{R}_x=\mathbb{E}\{\mathbf{x}\mathbf{x}^H\}$ as

$$\begin{aligned}\mathbf{R}_x&=\mathbf{P}\text{diag}(\mathbf{W},..,\mathbf{W})\mathbb{E}\left\{\mathbf{d}\mathbf{d}^H\right\}\text{diag}\left(\mathbf{W}^H,..,\mathbf{W}^H\right)\mathbf{P}^H\\&=\sigma_d^2\sum_{k=0}^{K-1}\mathbf{P}_k\mathbf{W}\mathbf{W}^H\mathbf{P}_k.\end{aligned} \tag{74}$$

After straightforward manipulation, $\mathbf{R}_x$ can be derived as

$$\mathbf{R}_x = \sigma_d^2 \begin{bmatrix} \boldsymbol{\Psi}' & \boldsymbol{\Psi_2}\boldsymbol{\Psi_1}^H & \cdots & \boldsymbol{\Psi_1}\boldsymbol{\Psi_2}^H \\ \boldsymbol{\Psi_1}\boldsymbol{\Psi_2}^H & \boldsymbol{\Psi}' & \cdots & \mathbf{0}_M \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Psi_2}\boldsymbol{\Psi_1}^H & \mathbf{0}_M & \cdots & \boldsymbol{\Psi}' \end{bmatrix}, \quad (75)$$

where $\boldsymbol{\Psi}' = \boldsymbol{\Psi_1}\boldsymbol{\Psi_1}^H + \boldsymbol{\Psi_2}\boldsymbol{\Psi_2}^H$, $\boldsymbol{\Psi}_1 = \text{diag}(\gamma_0, \ldots, \gamma_{M-1})$ and $\boldsymbol{\Psi}_2 = \text{diag}(\gamma_M, \ldots, \gamma_{2M-1})$.

Second, we derive $\sigma_{J,l}^2 = \mathbb{E}\{J[l]J^*[l]\}$ as

$$\sigma_{J,l}^2 = \frac{1}{N^2}\sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1} e^{-2|n_1-n_2|\pi\beta T_s} e^{-j\frac{2\pi(n_1-n_2)l}{N}}. \quad (76)$$

It can be easily shown that $\sum_{l=0}^{N-1}\sigma_{J,l}^2 = \frac{1}{N}$. Now, we have

$$\begin{aligned}\mathbb{E}\left\{\mathbf{x}^H\mathbf{H}^H\mathbf{J}^H\mathbf{J}\mathbf{H}\mathbf{x}\right\} &= \mathbb{T}\mathrm{r}\left\{\mathbb{E}\left\{\mathbf{J}\mathbf{H}\mathbf{x}\mathbf{x}^H\mathbf{H}^H\mathbf{J}^H\right\}\right\} \\ &= \mathbb{T}\mathrm{r}\left\{\mathbf{J}\left(\mathbf{R}_x \circ \mathbb{E}\left\{\mathbf{h}\mathbf{h}^H\right\}\right)\mathbf{J}^H\right\}, \quad (77)\end{aligned}$$

where $\mathbf{h} = [H[0], \ldots, H[N-1]] \in \mathbb{C}^{N\times 1}$. Since the channel coefficients $H[l]$ are circularly symetric Gaussian distributed with mean zero and variance σ_h^2, $\mathbb{E}\{\mathbf{h}\mathbf{h}^H\} = \sigma_h^2\mathbf{I}_N$. Thus, according to (75), we have

$$\mathbf{R}_x \circ \mathbb{E}\left\{\mathbf{h}\mathbf{h}^H\right\} = \sigma_h^2\sigma_d^2\,\text{diag}\underbrace{\left(\boldsymbol{\Psi}', \ldots, \boldsymbol{\Psi}'\right)}_{\text{K}}. \quad (78)$$

Now, by using (78), we can derive (77) as

$$\begin{aligned}\mathbb{T}\mathrm{r}\left\{\mathbf{J}\left(\mathbf{R}_x \circ \mathbb{E}\left\{\mathbf{h}\mathbf{h}^H\right\}\right)\mathbf{J}^H\right\} &= \sigma_h^2\sigma_d^2\mathbb{T}\mathrm{r}\left\{\text{diag}\left(\boldsymbol{\Psi}', \ldots, \boldsymbol{\Psi}'\right)\right\} \\ &\quad \times \sum_{l=0}^{N-1}\sigma_{J,l}^2 \\ &= \frac{\sigma_h^2\sigma_d^2}{N}\sum_{l_1=0}^{N-1}\left(|\gamma_{l_1'}|^2 + |\gamma_{M+l_1'}|^2\right), \quad (79)\end{aligned}$$

where $l_1' = (l_1)_M$.

APPENDIX C
DERIVATION OF CONSTANT μ

We have

$$\begin{aligned}\mu = \mathbb{R}\mathrm{e}\{\mathbb{E}\{J[0]\}\} &= \frac{1}{N}\mathbb{R}\mathrm{e}\left\{\mathbb{E}\left\{\sum_{n=0}^{N-1} e^{j\phi[n]}\right\}\right\} \\ &= \frac{1}{N}\left(\sum_{n=0}^{N-1} e^{-\frac{1}{2}\mathbb{E}\{|\phi[n]|^2\}}\right). \quad (80)\end{aligned}$$

Since $[\phi[n+1] - \phi[n]] \sim \mathcal{N}(0, 4\pi\beta T_s)$, $\mathbb{E}\{|\phi[n]|^2\} = (n+1)4\pi\beta T_s$, $n = 0, 1, \ldots, N-1$. Thus, μ is derived as

$$\begin{aligned}\mu &= \frac{1}{N}\left(\sum_{n=0}^{N-1} e^{-2\pi\beta T_s(n+1)}\right) \\ &= \begin{cases} 1 & \beta = 0 \\ \frac{1}{N}\frac{e^{-2\pi\beta T_s} - e^{-2\pi\beta T_s(N+1)}}{1-e^{-2\pi\beta T_s}} & \beta > 0. \end{cases} \quad (81)\end{aligned}$$

APPENDIX D
SINR WITH THE PROPOSED ALGORITHM

Similar to Appendix B and based on the SINR formulation (51) and using the SINR for ideal compensation in (46), we find

$$\begin{aligned}\text{SINR}_{\text{prop}} &= \frac{\mathbb{T}\mathrm{r}\left\{\mathbb{E}\left\{\hat{\mathbf{J}}\hat{\mathbf{J}}^H\right\}\right\}\text{SINR}_{\text{ideal}}}{\mathbb{T}\mathrm{r}\left\{\mathbb{E}\left\{\left(\mathbf{J}-\hat{\mathbf{J}}\right)\left(\mathbf{J}-\hat{\mathbf{J}}\right)^H\right\}\right\}\text{SINR}_{\text{ideal}} + \frac{1}{N}} \\ &= \frac{\mathbb{E}\left\{\left\|\tilde{\mathbf{j}}_{\mathbf{app}}\right\|^2\right\}\text{SINR}_{\text{ideal}}}{\mathbb{E}\left\{\left\|\tilde{\mathbf{j}} - \tilde{\mathbf{j}}_{\mathbf{app}}\right\|^2\right\}\text{SINR}_{\text{ideal}} + \frac{1}{N}}. \quad (82)\end{aligned}$$

Based on $\tilde{\mathbf{j}}_{\mathbf{app}} = \mathbf{F}_c\bar{\mathbf{p}}$ and $\mathbf{F}_c = \frac{1}{N}\mathbf{F}_N\boldsymbol{\Upsilon}$, we have

$$\begin{aligned}\mathbb{E}\left\{\left\|\tilde{\mathbf{j}}_{\mathbf{app}}\right\|^2\right\} &= \frac{1}{N^2}\mathbb{T}\mathrm{r}\left\{\mathbb{E}\left\{\boldsymbol{\Upsilon}\bar{\mathbf{p}}\bar{\mathbf{p}}^H\boldsymbol{\Upsilon}^H\right\}\right\} \\ &= \frac{1}{N^2}\mathbb{T}\mathrm{r}\left\{\boldsymbol{\Upsilon}\mathbf{R}_{\bar{\mathbf{p}}}\boldsymbol{\Upsilon}^H\right\}. \quad (83)\end{aligned}$$

Moreover, based on $\tilde{\mathbf{j}} = \frac{1}{N}\mathbf{F}_N\tilde{\mathbf{p}}$, we have

$$\begin{aligned}\mathbb{E}\left\{\left\|\tilde{\mathbf{j}} - \tilde{\mathbf{j}}_{\mathbf{app}}\right\|^2\right\} &= \frac{1}{N^2}\mathbb{E}\left\{\|\tilde{\mathbf{p}} - \boldsymbol{\Upsilon}\bar{\mathbf{p}}\|^2\right\} \\ &= \frac{1}{N^2}\Big(\mathbb{T}\mathrm{r}\left\{\mathbb{E}\left\{\tilde{\mathbf{p}}\tilde{\mathbf{p}}^H\right\}\right\} + \mathbb{T}\mathrm{r}\left\{\mathbb{E}\left\{\boldsymbol{\Upsilon}\bar{\mathbf{p}}\bar{\mathbf{p}}^H\boldsymbol{\Upsilon}^H\right\}\right\} \\ &\quad - 2\mathbb{R}\mathrm{e}\left\{\mathbb{T}\mathrm{r}\left\{\mathbb{E}\left\{\tilde{\mathbf{p}}\bar{\mathbf{p}}^H\boldsymbol{\Upsilon}^H\right\}\right\}\right\}\Big) \\ &= \frac{1}{N} + \frac{1}{N^2}\mathbb{T}\mathrm{r}\left\{\boldsymbol{\Upsilon}\mathbf{R}_{\bar{\mathbf{p}}}\boldsymbol{\Upsilon}^H\right\} \\ &\quad - 2\mathbb{R}\mathrm{e}\left\{\frac{1}{N^2}\mathbb{T}\mathrm{r}\left\{\mathbf{R}_{\tilde{\mathbf{p}},\bar{\mathbf{p}}}\boldsymbol{\Upsilon}^{\mathbf{H}}\right\}\right\}. \quad (84)\end{aligned}$$

By substituting (83) and (84) in (82), effective SINR for proposed method in (51) is derived.

APPENDIX E
POWER OF SIGNAL COMPONENTS (??)

By using signal model in (57), we derive the power of desired signal $P_{k'M+m'}^s = \mathbb{E}\{\bar{d}_{k'M+m'}\bar{d}_{k'M+m'}^H\}$, as

$$\begin{aligned}P_{k'M+m'}^s &= \mathbb{E}\Big\{\mathbf{U}_{k',m'}\mathbf{F}_N\mathbf{E}_{\mathbf{app}}\mathbf{F}_N^H\mathbf{H}\mathbf{C}_{k',m'}\mathbf{d}_{k'M+m'} \\ &\qquad \mathbf{d}_{k'M+m'}^H\mathbf{C}_{k',m'}^H\mathbf{H}^H\mathbf{F}_N\mathbf{E}_{\mathbf{app}}{}^H\mathbf{F}_N^H\mathbf{U}_{k',m'}^H\Big\} \\ &= \sigma_d^2\mathbb{E}\Big\{\mathbf{U}_{k',m'}\mathbf{F}_N\mathbf{E}_{\mathbf{app}}\mathbf{F}_N^H\Big\{\mathbf{C}_{k',m'}\mathbf{C}_{k',m'}^H \circ \mathbb{E}\{\mathbf{h}\mathbf{h}^H\}\Big\} \\ &\qquad \mathbf{F}_N\mathbf{E}_{\mathbf{app}}{}^H\mathbf{F}_N^H\mathbf{U}_{k',m'}^H\Big\} = \sigma_h^2\sigma_d^2\mathbf{U}_{k',m'}\mathbf{F}_N \\ &\quad \Big\{\Big\{\mathbf{F}_N^H\text{diag}\Big\{\mathbf{C}_{k',m'}\mathbf{C}_{k',m'}^H\Big\}\mathbf{F}_N\Big\} \circ \mathbb{E}\{\boldsymbol{\Upsilon}\bar{\mathbf{p}}\bar{\mathbf{p}}^H\boldsymbol{\Upsilon}^H\}\Big\}\mathbf{F}_N^H\mathbf{U}_{k',m'}^H \\ &= \sigma_h^2\sigma_d^2\mathbf{U}_{k',m'}\mathbf{F}_N\Big\{\Big\{\mathbf{F}_N^H\text{diag}\Big\{\mathbf{C}_{k',m'}\mathbf{C}_{k',m'}^H\Big\}\mathbf{F}_N\Big\} \circ \{\boldsymbol{\Upsilon}\mathbf{R}_{\bar{\mathbf{p}}}\boldsymbol{\Upsilon}^H\}\Big\} \\ &\quad \mathbf{F}_N^H\mathbf{U}_{k',m'}^H. \quad (85)\end{aligned}$$

Similar to (85) and after some manipulations, we can derive power of interference term caused by other symbols as $P_{k'M+m'}^I{}^d = P_{k'M+m'}^R{}^d - P_{k'M+m'}^s$, where $P_{k'M+m'}^R{}^d =$

$\mathbb{E}\{R_{d,k'M+m'}R^H_{d,k'M+m'}\}$ is equal to

$$P^{R^d}_{k'M+m'} = \sigma_h^2\sigma_d^2\mathbf{U}_{k',m'}\mathbf{F}_N \left\{\left\{\mathbf{F}_N^H\text{diag}\big(\mathbf{\Psi}',\ldots,\mathbf{\Psi}'\big)\mathbf{F}_N\right\}\circ\left\{\mathbf{\Upsilon}\mathbf{R}_{\tilde{\mathbf{p}}}\mathbf{\Upsilon}^H\right\}\right\}\mathbf{F}_N^H\mathbf{U}^H_{k',m'}. \quad (86)$$

Moreover, by deploying derivation in (84) and signal model in (57), we can daculate power of interference term due to approximation error $P^{I^\phi}_{k'M+m'} = \mathbb{E}\{I_{\phi,k'M+m'}I^H_{\phi,k'M+m'}\}$ as

$$\begin{aligned}P^{I^\phi}_{k'M+m'} = {}& \sigma_h^2\sigma_d^2\mathbf{U}_{k',m'}\mathbf{F}_N \\ &\times\left\{\left\{\mathbf{F}_N^H\text{diag}\big(\mathbf{\Psi}',\ldots,\mathbf{\Psi}'\big)\mathbf{F}_N\right\}\right. \\ &\left.\circ\left\{\mathbf{R}_{\tilde{\mathbf{p}}}+\mathbf{\Upsilon}\mathbf{R}_{\tilde{\mathbf{p}}}\mathbf{\Upsilon}^H-2\mathbb{R}\text{e}\left\{\mathbf{R}_{\tilde{\mathbf{p}},\tilde{\mathbf{p}}}\mathbf{\Upsilon}^{\mathbf{H}}\right\}\right\}\right\} \\ &\times\mathbf{F}_N^H\mathbf{U}^H_{k',m'}, \end{aligned} \quad (87)$$

where $\mathbf{R}_{\tilde{\mathbf{p}}} = \mathbb{E}\{\tilde{\mathbf{p}}\tilde{\mathbf{p}}^H\}$. Finally, we can derive the power of equivalent noise $P^w_{k'M+m'} = \mathbb{E}\{w_{k'M+m'}w^H_{k'M+m'}\}$ as

$$P^w_{k'M+m'} = \sigma_w^2\mathbf{U}_{k',m'}\mathbf{U}^H_{k',m'}. \quad (88)$$

APPENDIX F
$\partial\mathbf{S}_\Theta/\partial\Theta$ DERIVATION

Based on (14), we have $\mathbf{S}_{\mathbf{\Theta}} = \mathbf{J}'_{\mathbf{app}}\mathbf{X}\mathbf{F}_h\bar{\mathbf{h}}$. Moreover, based on (18) and (31), we have $\mathbf{S}_{\mathbf{\Theta}} = \mathbf{T}'\mathbf{F}'_c\mathbf{c}'$, where $\mathbf{T}'$ is a circulant matrix formed by elements of $\mathbf{H}'\mathbf{x}$. According to aforementioned expressions for $\mathbf{S}_{\mathbf{\Theta}}$ and (34), $\partial\mathbf{S}_{\mathbf{\Theta}}/\partial\mathbf{\Theta}$ can be derived in an element by element as

$$\begin{aligned}\frac{\partial\mathbf{S}_{\mathbf{\Theta}}}{\partial\mathbb{R}\text{e}\{h'[n]\}} &= \Big(\mathbf{J}'_{\mathbf{app}}\mathbf{X}\mathbf{F}_h\Big)a_{n+1},\ n=0,\ldots,L_{ch}-1\\ \frac{\partial\mathbf{S}_{\mathbf{\Theta}}}{\partial\mathbb{I}\text{m}\{h'[n]\}} &= j\Big(\mathbf{J}'_{\mathbf{app}}\mathbf{X}\mathbf{F}_h\Big)a_{n+1},\ n=0,\ldots,L_{ch}-1\\ \frac{\partial\mathbf{S}_{\mathbf{\Theta}}}{\partial\mathbb{R}\text{e}\{c'[n]\}} &= \big(\mathbf{T}'\mathbf{F}'_c\big)e_n,\ n=1,\ldots,Q-1\\ \frac{\partial\mathbf{S}_{\mathbf{\Theta}}}{\partial\mathbb{I}\text{m}\{c'[n]\}} &= j\big(\mathbf{T}'\mathbf{F}'_c\big)e_n,\ n=1,\ldots,Q-1,\end{aligned} \quad (89)$$

where a_{n+1} is $(n+1)$-th columns of $\mathbf{I}_{L_{ch}}$ and e_n is (n)-th columns of $\mathbf{I}_{Q-1}$.

REFERENCES

[1] Z. Zhang *et al.*, "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 28–41, Sep. 2019.

[2] T. S. Rappaport *et al.*, "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78729–78757, 2019.

[3] M. Ataeeshojai, R. C. Elliott, W. A. Krzymien, C. Tellambura, and J. Melzer, "Energy-efficient resource allocation in single-RF load-modulated massive MIMO HetNets," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1738–1764, 2020.

[4] A. Osseiran *et al.*, "Scenarios for 5G mobile and wireless communications: The vision of the METIS project," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26–35, May 2014.

[5] G. Wunder *et al.*, "5GNOW: Non-orthogonal, asynchronous waveforms for future mobile applications," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 97–105, Feb. 2014.

[6] B. Farhang-Boroujeny, "OFDM versus filter bank multicarrier," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 92–112, May 2011.

[7] F. Schaich, T. Wild, and Y. Chen, "Waveform contenders for 5G—Suitability for short packet and low latency transmissions," in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, 2014, pp. 1–5.

[8] N. Michailow *et al.*, "Generalized frequency division multiplexing for 5th generation cellular networks," *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3045–3061, Sep. 2014.

[9] S. Tiwari and S. S. Das, "Low-complexity joint-MMSE GFDM receiver," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1661–1674, Apr. 2018.

[10] P. Chen, B. Su, and Y. Huang, "Matrix characterization for GFDM: Low complexity MMSE receivers and optimal filters," *IEEE Trans. Signal Process.*, vol. 65, no. 18, pp. 4940–4955, Sep. 2017.

[11] Z. Na *et al.*, "Turbo receiver channel estimation for GFDM-based cognitive radio networks," *IEEE Access*, vol. 6, pp. 9926–9935, 2018.

[12] A. Mohammadian, M. Baghani, and C. Tellambura, "Optimal power allocation of GFDM secondary links with power amplifier nonlinearity and ACI," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 93–96, Feb. 2019.

[13] A. Mohammadian and C. Tellambura, "GFDM-modulated full-duplex cognitive radio networks in the presence of RF impairments," in *Proc. IEEE 30th Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, 2019, pp. 1–6.

[14] A. Mohammadian and C. Tellambura, "Cognitive GFDM full-duplex radios with RF impairments and ACI constraints," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 732–749, 2020.

[15] Z. Wang, L. Mei, X. Sha, and V. C. M. Leung, "Minimum BER power allocation for space-time coded generalized frequency division multiplexing systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 717–720, Jun. 2019.

[16] F. Li, K. Zheng, L. Zhao, H. Zhao, and Y. Li, "Design and performance of a novel interference-free GFDM transceiver with dual-filter," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4695–4706, May 2019.

[17] K. Liu, W. Deng, and Y. Liu, "Theoretical analysis of the peak-to-average power ratio and optimal pulse shaping filter design for GFDM systems," *IEEE Trans. Signal Process.*, vol. 67, no. 13, pp. 3455–3470, Jul. 2019.

[18] S. Ehsanfar, M. Matthe, M. Chafii, and G. P. Fettweis, "Pilot- and CP-aided channel estimation in MIMO non-orthogonal multi-carriers," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 650–664, Jan. 2019.

[19] Z. Na, J. Lv, F. Jiang, M. Xiong, and N. Zhao, "Joint subcarrier and subsymbol allocation based simultaneous wireless information and power transfer for multiuser GFDM in IoT," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 5999–6006, Aug. 2016.

[20] F. Tian *et al.*, "A novel concatenated coded modulation based on GFDM for access optical networks," *IEEE Photon. J.*, vol. 10, no. 2, Apr. 2018, Art. no. 7200808.

[21] A. Mohammadian, C. Tellambura, and M. Valkama, "Analysis of self-interference cancellation under phase noise, CFO, and IQ imbalance in GFDM full-duplex transceivers," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 700–713, Jan. 2020.

[22] Q. Zou, A. Tarighat, and A. H. Sayed, "Joint compensation of IQ imbalance and phase noise in OFDM wireless systems," *IEEE Trans. Commun.*, vol. 57, no. 2, pp. 404–414, Feb. 2009.

[23] P. Mathecken, T. Riihonen, N. N. Tchamov, S. Werner, M. Valkama, and R. Wichman, "Characterization of OFDM radio link under PLL-based oscillator phase noise and multipath fading channel," *IEEE Trans. Commun.*, vol. 60, no. 6, pp. 1479–1485, Jun. 2012.

[24] N. N. Tchamov, J. Rinne, A. Hazmi, M. Valkama, V. Syrjälä, and M. Renfors, "Enhanced algorithm for digital mitigation of ICI due to phase noise in OFDM receivers," *IEEE Wireless Commun. Lett.*, vol. 2, no. 1, pp. 6–9, Feb. 2013.

[25] Q. Zou, A. Tarighat, and A. H. Sayed, "Compensation of phase noise in OFDM wireless systems," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5407–5424, Nov. 2007.

[26] G. Liu and W. Zhu, "Compensation of phase noise in OFDM systems using an ICI reduction scheme," *IEEE Trans. Broadcast.*, vol. 50, no. 4, pp. 399–407, Dec. 2004.

[27] V. Syrjälä, M. Valkama, N. N. Tchamov, and J. Rinne, "Phase noise modelling and mitigation techniques in OFDM communications systems," in *Proc. Wireless Telecommun. Symp.*, Apr. 2009, pp. 1–7.

[28] E. Costa and S. Pupolin, "M-QAM-OFDM system performance in the presence of a nonlinear amplifier and phase noise," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 462–472, Mar. 2002.

[29] P. Mathecken, T. Riihonen, S. Werner, and R. Wichman, "Performance analysis of OFDM with Wiener phase noise and frequency selective fading channel," *IEEE Trans. Commun.*, vol. 59, no. 5, pp. 1321–1331, May 2011.
[30] M. R. Gholami, S. Nader-Esfahani, and A. A. Eftekhar, "A new method of phase noise compensation in OFDM," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 5, May 2003, pp. 3443–3446.
[31] R. A. Casas, S. L. Biracree, and A. E. Youtz, "Time domain phase noise correction for OFDM signals," *IEEE Trans. Broadcast.*, vol. 48, no. 3, pp. 230–236, Sep. 2002.
[32] S. Negusse, P. Zetterberg, and P. Händel, "Phase-noise mitigation in OFDM by best match trajectories," *IEEE Trans. Commun.*, vol. 63, no. 5, pp. 1712–1725, May 2015.
[33] P. Mathecken, T. Riihonen, S. Werner, and R. Wichman, "Phase noise estimation in OFDM: Utilizing its associated spectral geometry," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 1999–2012, Apr. 2016.
[34] P. Mathecken, T. Riihonen, S. Werner, and R. Wichman, "Constrained phase noise estimation in OFDM using scattered pilots without decision feedback," *IEEE Trans. Signal Process.*, vol. 65, no. 9, pp. 2348–2362, May 2017.
[35] J.-H. Lee, J.-S. Yang, S.-C. Kim, and Y.-W. Park, "Joint channel estimation and phase noise suppression for OFDM systems," in *Proc. IEEE 61st Veh. Technol. Conf.*, vol. 1, May 2005, pp. 467–470.
[36] F. Munier, T. Eriksson, and A. Svensson, "An ICI reduction scheme for OFDM system with phase noise over fading channels," *IEEE Trans. Commun.*, vol. 56, no. 7, pp. 1119–1126, Jul. 2008.
[37] R. Carvajal, J. C. Aguero, B. I. Godoy, and G. C. Goodwin, "EM-based maximum-likelihood channel estimation in multicarrier systems with phase distortion," *IEEE Trans. Veh. Technol.*, vol. 62, no. 1, pp. 152–160, Jan. 2013.
[38] O. H. Salim, A. A. Nasir, H. Mehrpouyan, W. Xiang, S. Durrani, and R. A. Kennedy, "Channel, phase noise, and frequency offset in OFDM systems: Joint estimation, data detection, and hybrid Cramar–Rao lower bound," *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3311–3325, Sep. 2014.
[39] W. Liu, S. Schwarz, M. Rupp, D. Chen, and T. Jiang, "Preamble-based channel estimation for OQAM/FBMC systems with delay diversity," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7169–7180, Nov. 2020.
[40] V. K. Singh, M. F. Flanagan, and B. Cardiff, "Generalized least squares based channel estimation for FBMC-OQAM," *IEEE Access*, vol. 7, pp. 129411–129420, 2019.
[41] D. Ren, J. Li, G. Zhang, G. Lu, and J. Ge, "Multi-tap channel estimation for preamble-based FBMC/OQAM systems," *IEEE Access*, vol. 8, pp. 176232–176240, 2020.
[42] J. C. Estrada-Jimenez, K. Chen-Hu, M. J. F. Garcia, and A. Garcia Armada, "Power allocation and capacity analysis for FBMC-OQAM with superimposed training," *IEEE Access*, vol. 7, pp. 46968–46976, 2019.
[43] L. D. Le and H. H. Nguyen, "Phase noise compensation for CFBMC–OQAM systems under imperfect channel estimation," *IEEE Access*, vol. 8, pp. 47247–47263, 2020.
[44] B. Lim and Y.-C. Ko, "SIR analysis of OFDM and GFDM waveforms with timing offset, CFO, and phase noise," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6979–6990, Oct. 2017.
[45] A. Mohammadian and C. Tellambura, "Full-duplex GFDM radio transceivers in the presence of phase noise, CFO and IQ imbalance," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.
[46] J. Jeong, Y. Park, S. Weon, J. Kim, S. Choi, and D. Hong, "Eigendecomposition-based GFDM for interference-free data transmission and pilot insertion for channel estimation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6931–6943, Oct. 2018.
[47] H. Shayanfar, H. Saeedi-Sourck, and A. Farhang, "CFO and channel estimation techniques for GFDM," in *Proc. IEEE Int. Microw. Workshop Series 5G Hardw. Syst. Technol. (IMWS-5G)*, 2018, pp. 1–3.
[48] H. Shayanfar, H. Saeedi-Sourck, and A. Farhang, "Low-complexity search method for CFO estimation in GFDM," *Electron. Lett.*, vol. 55, no. 6, pp. 355–357, 2019.
[49] Y. Liu, X. Zhu, E. G. Lim, Y. Jiang, and Y. Huang, "Robust semi-blind estimation of channel and CFO for GFDM systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2019, pp. 1–7.
[50] N. Tang, S. He, H. Wang, Y. Huang, and L. Yang, "Training sequence design for channel estimation and IQ imbalance compensation in GFDM systems," in *Proc. 9th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, 2017, pp. 1–6.
[51] H. Cheng, Y. Xia, Y. Huang, L. Yang, and D. P. Mandic, "A normalized complex LMS based blind I/Q imbalance compensator for GFDM receivers and its full second-order performance analysis," *IEEE Trans. Signal Process.*, vol. 66, no. 17, pp. 4701–4712, Sep. 2018.
[52] H. Cheng *et al.*, "Joint channel estimation and Tx/Rx I/Q imbalance compensation for GFDM systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1304–1317, Feb. 2019.
[53] L. D. Le and H. H. Nguyen, "Compensation of phase noise and IQ imbalance in multi-carrier systems," *IEEE Access*, vol. 8, pp. 191263–191277, 2020.
[54] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
[55] R. Bellman, *Introduction to Matrix Analysis*. Philadelphia, PA, USA: SIAM, 1997.
[56] R. M. Gray, *Toeplitz and Circulant Matrices: A Review*. New York, NY, USA: Now, 2006.
[57] S. Han, Y. Sung, and Y. H. Lee, "Filter design for generalized frequency-division multiplexing," *IEEE Trans. Signal Process.*, vol. 65, no. 7, pp. 1644–1659, Apr. 2017.
[58] V. Syrjälä, M. Valkama, L. Anttila, T. Riihonen, and D. Korpi, "Analysis of oscillator phase-noise effects on self-interference cancellation in full-duplex OFDM radio transceivers," *IEEE Trans. Wireless Commun.*, vol. 13, no. 6, pp. 2977–2990, Jun. 2014.
[59] H. Mehrpouyan, A. A. Nasir, S. D. Blostein, T. Eriksson, G. K. Karagiannidis, and T. Svensson, "Joint estimation of channel and oscillator phase noise in MIMO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4790–4807, Sep. 2012.
[60] A. Hajimiri, S. Limotyrakis, and T. H. Lee, "Jitter and phase noise in ring oscillators," *IEEE J. Solid-State Circuits*, vol. 34, no. 6, pp. 790–804, Jun. 1999.
[61] T. Cui, C. Tellambura, and Y. Wu, "Low-complexity pilot-aided channel estimation for OFDM systems over doubly-selective channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 3, 2005, pp. 1980–1984.
[62] G. Wang, F. Gao, Y. Wu, and C. Tellambura, "Joint CFO and channel estimation for OFDM-based two-way relay networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 456–465, Feb. 2011.
[63] T. Cui and C. Tellambura, "Semiblind channel estimation and data detection for OFDM systems with optimal pilot design," *IEEE Trans. Commun.*, vol. 55, no. 5, pp. 1053–1062, May 2007.
[64] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*. New York, NY, USA: Springer, 2008.
[65] T. Strutz, *Data Fitting and Uncertainty: A Practical Introduction to Weighted Least Squares and Beyond*. Heidelberg, Germany: Springer Vieweg, 2016.
[66] D. Ham and A. Hajimiri, "Virtual damping and Einstein relation in oscillators," *IEEE J. Solid-State Circuits*, vol. 38, no. 3, pp. 407–418, Mar. 2003.
[67] A. Demir, "Phase noise and timing jitter in oscillators with colored-noise sources," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 49, no. 12, pp. 1782–1791, Dec. 2002.
[68] Z. Xiao, Y. Li, L. Bai, and J. Choi, "Achievable sum rates of half- and full-duplex bidirectional OFDM communication links," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1351–1364, Feb. 2017.
[69] M. Mohammadi, H. A. Suraweera, Y. Cao, I. Krikidis, and C. Tellambura, "Full-duplex radio for uplink/downlink wireless access with spatially random nodes," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 5250–5266, Dec. 2015.

AMIRHOSSEIN MOHAMMADIAN (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the Amirkabir University of Technology, Tehran, Iran, in 2014 and 2017, respectively. He is currently pursuing the Ph.D. degree in electrical engineering with the University of Alberta, Edmonton, AB, Canada.

His current research interests include wireless communication, full-duplex transmission, cognitive radio networks, dirty RF, and MIMO structure. He was a recipient of the Recruitment Doctor of Philosophy Scholarship and the University of Alberta Graduate Fellowships in 2017 and 2019. He was also a recipient of Alberta Innovates Graduate Student Data-Enabled Innovation Scholarship in 2020.

CHINTHA TELLAMBURA (Fellow, IEEE) received the B.Sc. degree in electronics and telecommunications from the University of Moratuwa, Sri Lanka, the M.Sc. degree in electronics from the Kings College, University of London, and the Ph.D. degree in electrical engineering from the University of Victoria, Canada.

He was with Monash University, Australia, from 1997 to 2002. Since 2002, he has been with the Department of Electrical and Computer Engineering, University of Alberta, where he is currently a Full Professor. He has authored or coauthored over 560 journal and conference papers, with an H-index of 72 (Google Scholar). He has supervised or co-supervised 66 M.Sc., Ph.D., and PDF trainees. His current research interests include cognitive radio, heterogeneous cellular networks, fifth-generation wireless networks, and machine learning algorithms. He received the Best Paper Awards from the IEEE International Conference on Communications in 2012 and 2017. He is the Winner of the prestigious McCalla Professorship and the Killam Annual Professorship from the University of Alberta. He served as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS from 1999 to 2012 and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2001 to 2007. He was an Area Editor of *Wireless Communications Systems and Theory* from 2007 to 2012. He was elected as a Fellow of The Canadian Academy of Engineering in 2017.