Joint Channel Estimation and Signal Detection Based on MAP Criterion in MIMO-OFDM System with Phase Noise

Heng Du, Jiang Xue, *Senior Member, IEEE*, Qihong Duan, Symeon Chatzinotas, *Fellow, IEEE*

***Abstract*—The channel estimation and signal detection are key issues in the Multi-Input Multi-Output and Orthogonal Frequency Division Multiplexing (MIMO-OFDM) system. However, there exist severe impacts of phase noise (PN) on the estimations with the application of higher frequency in 5th Generation New Radio (5G-NR). In this paper, the joint channel estimation and signal detection (JCESD) method based on the maximum a posteriori (MAP) criterion under the assumption of Wiener process for PN is proposed in MIMO-OFDM system, which is called as the JCESD-PN-MAP method. Firstly, the MAP criterion is derived based on Bayesian theories and the structures of matrices in MAP criterion are analyzed to simplify the optimizations. Secondly, the optimizations for PN of receiving and transmitting antennas are translated into solving a tridiagonal linear equation and a sparse linear equation, respectively, which are optimized by the Gaussian elimination (GE) method with low computation complexity. To further reduce the computational complexity, the latter is solved by the alternating direction method of multipliers (ADMM) method. Thirdly, the optimizations for channel responses and transmitted signals are translated into solving two block diagonal linear equations, which are solved by calculating the inverse matrix with low computation complexity. The numerical results and complexity analysis confirm the effectiveness of our proposed method in terms of the accuracy and computation complexity.**

***Index Terms*—Phase noise, channel estimation, signal detection, MIMO-OFDM, maximum a posteriori.**

I. Introduction

THE orthogonal frequency division multiplexing (OFDM) [1] is the main multi-carrier modulation technology in 5th Generation New Radio (5G-NR) [2]. Meanwhile, the Multi-Input Multi-Output (MIMO) [3] is a widely applied diversity and multiplexing technology in 5G-NR. For MIMO-OFDM system, the channel estimation and signal detection are key issues [4], [5]. However, the oscillators with limited capabilities of both transmitting and receiving antennas will produce impure output signals with additional cumulative errors in their phases, which is known as the phase noise (PN) [6]. The performance of channel estimation and signal detection in MIMO-OFDM system will be affected by the PN significantly [6]–[9].

Heng Du, Jiang Xue and Qihong Duan are with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: duheng0102@stu.xjtu.edu.cn; x.jiang@xjtu.edu.cn; khtuan@126.com). Corresponding author: Jiang Xue.

Symeon Chatzinotas is with the SnT, University of Luxembourg, Luxembourg L-2721, Luxembourg (e-mail: symeon.chatzinotas@uni.lu).

A. Impacts of PN

The performance of OFDM system will be limited by the orthogonality degradation and nonlinear distortions caused by the PN, which include the common phase error (CPE) and inter-carrier interference (ICI) [6]. The CPE and ICI lead to the rotation and divergence of constellation [7], respectively. For MIMO-OFDM system, the PN of multiple transmitting and receiving antennas needs to be considered and the impacts of PN are more complex and intractable, which include the CPE and ICI within and among different antennas [8], [9]. With higher frequencies in 5G-NR, the power of PN is greater and its impacts on channel estimation and signal detection are more severe [6]–[8]. Therefore, the channel estimation and signal detection in MIMO-OFDM system with PN need to be investigated and studied.

B. Related Works

The channel estimation and signal detection methods were widely studied and summarized in [4], [5]. A number of effective channel estimation methods were proposed, such as the least squares (LS) [10], minimum mean square error (MMSE) [11], sparse Bayesian learning [12], tensor based [13] and deep learning [14] methods. Meanwhile, a lot of effective signal detection methods were also proposed, such as the decision feedback [15], QR decomposition [16], maximum likelihood (ML) [17] and deep learning [18] methods. However, the above methods considered the channel estimation and signal detection separately and better performance can be obtained by considering them jointly [19].

The joint channel estimation and signal detection (JCESD) methods in MIMO-OFDM system improved the performance of the system and were summarized in [19]. The weighted least squares (WLS) method [20] was proposed to improve the performance of LS method. In addition, the MMSE criterion [21] was proposed and achieved better performance. Moreover, the tensor based model was derived and optimized by the alternating LS method [22] with low computational complexity. Additionally, the Bayesian inference [23] and sparse Bayesian learning [24] methods were proposed to utilize the sparsity of channel matrix. Furthermore, the deep learning method [25] was proposed based on the long short term memory (LSTM) network and achieved great performance regardless of the black-box model. However, the above methods did not consider the impacts of PN and their performance was limited

by the PN [6]. Therefore, the JCESD methods in MIMO-OFDM system with PN need to be investigated and studied.

The JCESD methods in MIMO-OFDM system with PN were summarized in [26] and divided into two categories according to the estimation of PN. Firstly, the CPE estimation methods focused on the estimation of CPE while the ICI was considered as the Gaussian noise [26], [27]. The JCESD methods with CPE estimation were proposed based on the LS [28], [29] method with low computational complexity. In addition, the tensor based model was derived and optimized by the bilinear alternating LS method [30]. Secondly, the PN estimation methods focused on the estimations of both CPE and ICI [26], [31]. The extended Kalman filter (EKF) and WLS methods were combined to track the PN and estimate the channel state information (CSI) and transmitted signals [32]. In addition, the variational expectation maximization (VEM) method [33] was proposed and the PN was parameterized by the basis vectors related to the covariance matrix. Moreover, the orthogonal matching pursuit (OMP) and pattern search methods [34] were combined and dealt with the truncated PN in frequency domain. Furthermore, the maximum a posteriori (MAP) [35], [36] criterion was proposed to combine the statistical prior and likelihood information and dealt with the truncated PN [35] and complete PN [36], respectively. Generally, compared to the CPE estimation methods, the PN estimation methods achieved better performance but suffered from high computational complexity and increased the load of the system significantly [26]. Therefore, the JCESD methods with PN estimation in MIMO-OFDM system need to be studied and improved to achieve great performance with low computational complexity.

C. Motivations

In this paper, the JCESD method based on the MAP criterion under the assumption of Wiener process for PN is proposed in MIMO-OFDM system, which is called the JCESD-PN-MAP method, to achieve great performance with low computational complexity and the major motivations are described as follows.

To achieve great performance, the MAP criterion is applied to combine the statistical prior and likelihood information [36], [37]. The joint signal detection and PN estimation method based on the MAP criterion [37] achieved great performance in MIMO-OFDM system and confirmed the effectiveness of MAP criterion but it required perfect CSI, which was not realistic. In addition, The joint signal detection and PN estimation method based on the MAP criterion [36] was proposed with a separate channel estimation before the iterations, which was not sufficient and degraded the performance of MIMO-OFDM system. Therefore, the MAP criterion needs to be further studied to consider the channel estimation sufficiently and jointly.

In addition, the PN estimation methods suffered from high computational complexity and increase the load of the system significantly [26]. Especially, the computational complexity of [36], [37] increased three-order with the numbers of subcarriers and transmitting/receiving antennas. To reduce the computational complexity, the structures of matrices in MAP criterion are analyzed and the optimization processes can be simplified significantly. There are some unitary matrices due to the PN and discrete Fourier transform (DFT) and sparse or block diagonal matrices due to the channel matrix. Furthermore, many statistical methods and numerical simulations considered the PN as the Wiener process [26], [36]–[39], which can be applied as the prior information for PN, and the tridiagonal structure related to its covariance matrix is proved in this paper and can be utilized to reduce the computational complexity significantly.

D. Contributions and Novelties

In this paper, the JCESD-PN-MAP method is proposed and the major contributions and novelties of this paper are summarized as follows:

- The MAP criterion for JCESD and PN estimation is derived based on Bayesian theories in MIMO-OFDM system to achieve great performance. In addition, the PN is described by the Wiener process [26], [36]–[39] and the tridiagonal structure related to its covariance matrix is derived, which simplifies the optimization processes and reduces the computation complexity significantly.
- The optimization of MAP criterion for PN of receiving antennas is translated into solving a tridiagonal linear equation and optimized based on the Gaussian elimination (GE) method [39], [40] to reduce the computation complexity[1] from $\mathcal{O}(N_r^3N^3)$ [36], [37] to $8N_rN$, where $\mathcal{O}$ is the order of computational complexit.
- The optimization of MAP criterion for PN of transmitting antennas is translated into solving a sparse linear equation and optimized based on the GE method to reduce the computational complexity from $\mathcal{O}(N_t^3N^3)$ [36], [37] to $\mathcal{O}(N_t^3NL^2)$. When N_t and L are not very small, the computational complexity is further reduced to $\mathcal{O}(N_t^3N+N_rN\log N+K\cdot 2N_tN\log N+K\cdot 8N_tN)$ by applying the alternating direction method of multipliers (ADMM) method [41] to decompose it into two simpler subproblems.
- The optimizations of MAP criterion for channel responses and transmitted signals are translated into solving two block diagonal linear equations and solved by calculating the inverse matrix with low computation complexity, and the computational complexity for channel responses is reduced from $\mathcal{O}(N_r^3N_t^3L^3)$ [36], [37] to $\mathcal{O}(N_t^3L^3)$.

E. Organization and Notations

The rest of this paper is organized as follows. The model of MIMO-OFDM system with PN is introduced and the structures of matrices are analyzed in Section II. In addition, the models of PN are described in Section III. Moreover, the JCESD-PN-MAP method is proposed for the JCESD and PN estimation in Section IV. Furthermore, the numerical results and complexity analysis demonstrate the effectiveness of our proposed method in Section V. Lastly, the paper is concluded in Section VI.

[1] N, N_r, N_t, L and K are the numbers of subcarriers, receiving antennas, transmitting antennas, paths of channel responses and iterations of ADMM method. Generally, L is much smaller than N.

Throughout this paper, the bold lower-case and upper-case letters denote the vectors and matrices, respectively, and the unbold letters refer the scalars. In addition, $(\cdot)^{\mathrm{H}}$, $(\cdot)^{\mathrm{T}}$, $(\cdot)^{-1}$, $e^{j(\cdot)}$, $||\cdot||_{\mathrm{F}}$, $\Re\{\cdot\}$ and $\Im\{\cdot\}$ represent the Hermite transpose, transpose, inverse, complex exponent, Frobenius norm, real part and imaginary part of a matrix or vector, respectively. Moreover, $[(\cdot)^{\mathrm{T}}, (\cdot)^{\mathrm{T}}, \cdots]^{\mathrm{T}}$ is the vector composed of multiple vectors. Additionally, $\mathbb{C}^N$, $\mathbb{R}^N$, $\mathbb{C}^{N\times M}$ and $\mathbb{R}^{N\times M}$ refer the sets of all complex and real vectors and matrices, respectively. Furthermore, $\mathcal{CN}$ and $\mathcal{N}$ denote the complex and real Gaussian distributions, respectively. Finally, $\mathbf{0}_K$, $\mathbf{1}_K$ and $\boldsymbol{I}_K$ denote the K-dimensional all-zero vector, all-one vector and identity matrix, respectively, and $\boldsymbol{\Omega}_{I,J,\cdots}$ denotes the set of integers $\{(i,j,\cdots)|0\le i\le I-1, 0\le j\le J-1,\cdots\}$.

II. System Model

In this section, the MIMO-OFDM system with PN is introduced and the structures of matrices are analyzed. In additon, the equivalent channel responses and PN are mentioned.

A. MIMO-OFDM System with PN

Considering N_t transmitting antennas, N_r receiving antennas and N subcarriers, the MIMO-OFDM system with PN [36] can be represented as[2]

$$\boldsymbol{y}_q = e^{j\boldsymbol{\theta}_q} \odot \sum_{p=0}^{N_t-1} \left(\boldsymbol{h}_{p,q} \otimes (e^{j\boldsymbol{\phi}_p} \odot \boldsymbol{x}_p)\right) + \boldsymbol{w}_q \tag{1}$$

for $p,q \in \boldsymbol{\Omega}_{N_t,N_r}$, where $\bar{\boldsymbol{x}}_p \in \mathbb{C}^N$, $\boldsymbol{x}_p = \boldsymbol{F}^{\mathrm{H}}\bar{\boldsymbol{x}}_p \in \mathbb{C}^N$, $\bar{\boldsymbol{y}}_q = \boldsymbol{F}\boldsymbol{y}_q \in \mathbb{C}^N$ and $\boldsymbol{y}_q \in \mathbb{C}^N$ are the transmitted signals of p-th transmitting antenna and received signals of q-th receiving antenna in frequency and time domain, respectively, in which $\boldsymbol{F} \in \mathbb{C}^{N\times N}$ is the N-point DFT matrix. In addition, $\bar{\boldsymbol{h}}_{p,q} = \boldsymbol{F}\hat{\boldsymbol{h}}_{p,q} \in \mathbb{C}^N$ and $\boldsymbol{h}_{p,q} \in \mathbb{C}^L$ are the channel responses between p-th transmitting antenna and q-th receiving antenna in frequency and time domain, respectively, in which L is the number of paths and $\hat{\boldsymbol{h}}_{p,q} \in \mathbb{C}^N$ is obtained by padding $\boldsymbol{h}_{p,q}$ with $N-L$ zeros. Moreover, $\bar{\boldsymbol{w}}_q = \boldsymbol{F}\boldsymbol{w}_q \in \mathbb{C}^N$ and $\boldsymbol{w}_q \in \mathbb{C}^N$ are the white Gaussian noise of q-th receiving antenna in frequency and time domain, respectively. Additionally, $\boldsymbol{\phi}_p \in \mathbb{R}^N$ and $\boldsymbol{\theta}_q \in \mathbb{R}^N$ are the PN of p-th transmitting antenna and PN of q-th receiving antenna in time domain[3], respectively, which are introduced in Section III. Furthermore, $\otimes$ and $\odot$ are the cyclic convolution and Hadamard product operators, and j is the imaginary unit. Finally, it should be noted that parts of transmitted signals are known as the pilots while others are unknown and need to be estimated as described in Section IV.

It can be obtained from Eq. (1) [36] that

$$\begin{aligned}\bar{\boldsymbol{y}} &= \boldsymbol{T}(\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{h},\bar{\boldsymbol{x}}) + \bar{\boldsymbol{w}} &(2)\\ &= \boldsymbol{F}_{N_r}\boldsymbol{D}_{\boldsymbol{\theta}}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{D}_{\boldsymbol{\phi}}\boldsymbol{F}_{N_t}^{\mathrm{H}}\bar{\boldsymbol{x}} + \bar{\boldsymbol{w}}\\ &= \boldsymbol{F}_{N_r}\boldsymbol{D}_{\boldsymbol{\theta}}\boldsymbol{F}_{N_r}^{\mathrm{H}}\boldsymbol{D}_{\bar{\boldsymbol{h}}}\boldsymbol{F}_{N_t}\boldsymbol{D}_{\boldsymbol{\phi}}\boldsymbol{F}_{N_t}^{\mathrm{H}}\bar{\boldsymbol{x}} + \bar{\boldsymbol{w}}, &(3)\end{aligned}$$

where $\boldsymbol{T}(\cdot)$ represents the process described in Eq. (1). In addition, $\bar{\boldsymbol{x}} = [\bar{\boldsymbol{x}}_0^{\mathrm{T}}, \bar{\boldsymbol{x}}_1^{\mathrm{T}}, \cdots, \bar{\boldsymbol{x}}_{N_t-1}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{N_tN}$, $\boldsymbol{h}_q = [\boldsymbol{h}_{0,q}^{\mathrm{T}}, \boldsymbol{h}_{1,q}^{\mathrm{T}}, \cdots, \boldsymbol{h}_{N_t-1,q}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{N_tL}$ and $\boldsymbol{h} = [\boldsymbol{h}_0^{\mathrm{T}}, \boldsymbol{h}_1^{\mathrm{T}}, \cdots, \boldsymbol{h}_{N_r-1}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{N_rN_tL}$. Moreover, $\bar{\boldsymbol{y}} \in \mathbb{C}^{N_rN}$, $\boldsymbol{\theta} \in \mathbb{R}^{N_rN}$, $\boldsymbol{\phi} \in \mathbb{R}^{N_tN}$, $\bar{\boldsymbol{w}} \in \mathbb{C}^{N_rN}$, $\boldsymbol{y} \in \mathbb{C}^{N_rN}$, $\boldsymbol{x} \in \mathbb{C}^{N_tN}$ and $\boldsymbol{w} \in \mathbb{C}^{N_rN}$ are defined similar to $\bar{\boldsymbol{x}}$ while $\bar{\boldsymbol{h}}_q \in \mathbb{C}^{N_tN}$ and $\bar{\boldsymbol{h}} \in \mathbb{C}^{N_rN_tN}$ are defined similar to $\boldsymbol{h}_q$ and $\boldsymbol{h}$. Additionally, $\boldsymbol{D}_{\boldsymbol{\theta}} \in \mathbb{C}^{N_rN\times N_rN}$ and $\boldsymbol{D}_{\boldsymbol{\phi}} \in \mathbb{C}^{N_tN\times N_tN}$ are corresponding diagonal matrices of $e^{j\boldsymbol{\theta}}$ and $e^{j\boldsymbol{\phi}}$, respectively. Furthermore, $\boldsymbol{F}_{N_r} \in \mathbb{C}^{N_rN\times N_rN}$ and $\boldsymbol{F}_{N_t} \in \mathbb{C}^{N_tN\times N_tN}$ are the corresponding block diagonal matrices of $\boldsymbol{F}$ with different dimensions. Finally, $\boldsymbol{C}_{\boldsymbol{h}} \in \mathbb{C}^{N_rN\times N_tN}$ and $\boldsymbol{D}_{\bar{\boldsymbol{h}}} \in \mathbb{C}^{N_rN\times N_tN}$ are defined as

$$\boldsymbol{C}_{\boldsymbol{h}} = \begin{bmatrix} \boldsymbol{C}_{\boldsymbol{h}_{0,0}} & \cdots & \boldsymbol{C}_{\boldsymbol{h}_{N_t-1,0}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{C}_{\boldsymbol{h}_{0,N_r-1}} & \cdots & \boldsymbol{C}_{\boldsymbol{h}_{N_t-1,N_r-1}} \end{bmatrix} \tag{4}$$

and

$$\boldsymbol{D}_{\bar{\boldsymbol{h}}} = \begin{bmatrix} \boldsymbol{D}_{\bar{\boldsymbol{h}}_{0,0}} & \cdots & \boldsymbol{D}_{\bar{\boldsymbol{h}}_{N_t-1,0}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{D}_{\bar{\boldsymbol{h}}_{0,N_r-1}} & \cdots & \boldsymbol{D}_{\bar{\boldsymbol{h}}_{N_t-1,N_r-1}} \end{bmatrix}, \tag{5}$$

respectively, in which $\boldsymbol{C}_{\boldsymbol{h}_{p,q}} \in \mathbb{C}^{N\times N}$ and $\boldsymbol{D}_{\bar{\boldsymbol{h}}_{p,q}} \in \mathbb{C}^{N\times N}$ are the corresponding circular matrix of $\hat{\boldsymbol{h}}_{p,q}$ and corresponding diagonal matrix of $\bar{\boldsymbol{h}}_{p,q}$, respectively, and it should be noted that $\boldsymbol{C}_{\boldsymbol{h}} = \boldsymbol{F}_{N_r}^{\mathrm{H}}\boldsymbol{D}_{\bar{\boldsymbol{h}}}\boldsymbol{F}_{N_t}$ according to the property of DFT matrix.

One form of $\boldsymbol{T}(\cdot)$ is given in Eq. (3), which is used for the estimation of transmitted signals $\bar{\boldsymbol{x}}$, and other equivalent forms for the estimations of PN of receiving antennas $\boldsymbol{\theta}$, PN of transmitting antennas $\boldsymbol{\phi}$ and channel responses $\boldsymbol{h}$ are given in Section IV.

B. Structure of Matrices

It should be noted that $\boldsymbol{D}_{\boldsymbol{\theta}}$ and $\boldsymbol{D}_{\boldsymbol{\phi}}$ are diagonal and unitary matrices while $\boldsymbol{F}_{N_r}$ and $\boldsymbol{F}_{N_t}$ are block diagonal and unitary matrices. In addition, $\boldsymbol{C}_{\boldsymbol{h}_{p,q}}$ is a sparse and circular matrix with L non-zero elements per row and $\boldsymbol{D}_{\bar{\boldsymbol{h}}_{p,q}}$ is a diagonal matrix. Therefore, it can be obtained that $\boldsymbol{C}_{\boldsymbol{h}}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{h}} =$

$$\begin{bmatrix} \sum_{q=0}^{N_r-1}\boldsymbol{C}_{\boldsymbol{h}_{0,q}}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{h}_{0,q}} & \cdots & \sum_{q=0}^{N_r-1}\boldsymbol{C}_{\boldsymbol{h}_{0,q}}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{h}_{N_t-1,q}} \\ \vdots & \ddots & \vdots \\ \sum_{q=0}^{N_r-1}\boldsymbol{C}_{\boldsymbol{h}_{N_t-1,q}}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{h}_{0,q}} & \cdots & \sum_{q=0}^{N_r-1}\boldsymbol{C}_{\boldsymbol{h}_{N_t-1,q}}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{h}_{N_t-1,q}} \end{bmatrix} \tag{6}$$

$\in \mathbb{C}^{N_tN\times N_tN}$ is a sparse matrix with $N_t(2L-1)$ non-zero elements per row and $\boldsymbol{D}_{\bar{\boldsymbol{h}}}^{\mathrm{H}}\boldsymbol{D}_{\bar{\boldsymbol{h}}} =$

$$\begin{bmatrix} \sum_{q=0}^{N_r-1}\boldsymbol{D}_{\bar{\boldsymbol{h}}_{0,q}}^{\mathrm{H}}\boldsymbol{D}_{\bar{\boldsymbol{h}}_{0,q}} & \cdots & \sum_{q=0}^{N_r-1}\boldsymbol{D}_{\bar{\boldsymbol{h}}_{0,q}}^{\mathrm{H}}\boldsymbol{D}_{\bar{\boldsymbol{h}}_{N_t-1,q}} \\ \vdots & \ddots & \vdots \\ \sum_{q=0}^{N_r-1}\boldsymbol{D}_{\bar{\boldsymbol{h}}_{N_t-1,q}}^{\mathrm{H}}\boldsymbol{D}_{\bar{\boldsymbol{h}}_{0,q}} & \cdots & \sum_{q=0}^{N_r-1}\boldsymbol{D}_{\bar{\boldsymbol{h}}_{N_t-1,q}}^{\mathrm{H}}\boldsymbol{D}_{\bar{\boldsymbol{h}}_{N_t-1,q}} \end{bmatrix} \tag{7}$$

$\in \mathbb{C}^{N_tN\times N_tN}$ is a block diagonal matrix (after rearranging its rows and columns according to order $[0, N_t, \cdots, N_t(N_t-1), 1, 1+N_t, \cdots, 1+N_t(N_t-1), 2, \cdots, N_t^2-1)]^{\mathrm{T}} \in \mathbb{R}^{N_tN}$), whose inverse matrix can be calculated with the computation complexity of $\mathcal{O}(N_t^3N)$. It should be noted that $\boldsymbol{C}_{\boldsymbol{h}} = \boldsymbol{F}_{N_r}^{\mathrm{H}}\boldsymbol{D}_{\bar{\boldsymbol{h}}}\boldsymbol{F}_{N_t}$ according to the property of DFT matrix.

[2]In this paper, $\bar{\boldsymbol{a}}$ denotes the representation in frequency domain corresponding to a variable $\boldsymbol{a}$.

[3]Independent oscillators of different antennas are considered for single user [36], [37] and multiple users [33], and therefore independent PN is considered for different antennas in this paper. The shared oscillators are easier to be processed with the same mathematical principles, which is not the focus of this paper.

All the above properties and others related to PN given in Section III can simplify the optimization processes described in Section IV significantly.

C. Equivalent Channel Responses and PN

In fact, the equivalent channel responses $\widetilde{\boldsymbol{h}}_{p,q} = e^{j\hat{\phi}_p} e^{j\hat{\theta}_q} \boldsymbol{h}_{p,q}$, PN of receiving antennas $\widetilde{\boldsymbol{\theta}}_q = \boldsymbol{\theta}_q - \hat{\theta}_q \mathbf{1}_N$ and PN of transmitting antennas $\widetilde{\boldsymbol{\phi}}_p = \boldsymbol{\phi}_p - \hat{\phi}_p \mathbf{1}_N$ are estimated, which also satisfy Eq. (1), where $\hat{\theta}_q = \frac{1}{N}\sum_{n=0}^{N-1} \theta_q(n)$ and $\hat{\phi}_p = \frac{1}{N}\sum_{n=0}^{N-1} \phi_p(n)$ represent the CPE [9]. The equivalent channel responses $\widetilde{\boldsymbol{h}}$ are defined according to $\widetilde{\boldsymbol{h}}_{p,q}$ with similar relationship between $\boldsymbol{h}$ and $\boldsymbol{h}_{p,q}$ in Eq. (3). In addition, the equivalent PN $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\phi}}$ are defined according to $\widetilde{\boldsymbol{\theta}}_q$ and $\widetilde{\boldsymbol{\phi}}_p$ with similar relationships among $\boldsymbol{\theta}$, $\boldsymbol{\theta}_q$, $\boldsymbol{\phi}$ and $\boldsymbol{\phi}_p$ in Eq. (3), and their expectations are both all-zero vectors. Therefore, the initializations of PN in Section IV can be configured as all-zero vectors. For convenience, the equivalent channel responses and PN of transmitting and receiving antennas are still called as $\boldsymbol{h}$, $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$.

III. PN Models

The power spectrum density (PSD) model is an empirical model and applied to the generation of PN in the simulations, which is more practical but difficult to handle. In addition, the Wiener process model is a statistical model and applied to design of our proposed JCESD-PN-MAP method in this paper, which is easy to handle and also accurate.

A. PSD Model

The PSD model can be used to describe PN [42], [43] and a typical PSD curve[4] from 3GPP TR 38.803 is presented as

$$S_{\text{Total}}(f) = \begin{cases} S_{\text{Ref}}(f) + S_{\text{PLL}}(f), & f < f_{\text{LoopBW}} \\ S_{\text{VCO\_v2}}(f) + S_{\text{VCO\_v3}}(f), & f > f_{\text{LoopBW}} \end{cases}, \tag{8}$$

where

$$S_{\text{Ref/PLL/VCO\_v2/VCO\_v3}}(f) = P_0 \cdot \left(\frac{1 + (\frac{f}{f_z})^k}{1 + f^k} \right) \text{ dB}$$

and

$$P_0 = P_{\text{FOM}} + 20\log f_c - 10\log(\frac{P}{1 \text{ mW}}) \text{ dB},$$

in which f_{LoopBW} is the loop bandwidth, P_{FOM} is the figure of merit, f_c is the carrier frequency, P is the consumed power, f_z and k are the control parameters. Moreover, {Ref, PLL, VCO_v2, VCO_v3} represent different models of PN and the parameter settings of these models for user equipment (UE) and base station (BS) can be seen in **TABLE I**, which will be employed for the generation of PN in the simulations[5].

[4]The PSD curve is obtained from 3GPP TR 38.803, Study on new radio access technology: Radio Frequency (RF) and co-existence aspects, (Release 14), 2022-04 (https://www.3gpp.org/ftp/Specs/archive/38_series/38.803/).

[5]The method for the generation of PN is cited as: Alex Bar-Guy (2022). Phase Noise (https://www.mathworks.com/matlabcentral/fileexchange/8844-phase-noise), MATLAB Central File Exchange.

TABLE I: The parameter settings of PSD model for UE/BS

	Model: UE / BS			
	Ref	PLL	VCO_v2	VCO_v3
f_c [GHz]	$10 \sim 60$ / $10 \sim 60$			
f_{LoopBW} [kHz]	187 / 112			
P_{FOM} [dB]	-215 / -240	-240 / -245	-175 / -187	-175 / -130
f_z [Hz]	Inf / Inf	1e4 / 1e4	50.3e6 / 8e6	Inf / Inf
P [mW]	10 / 10	20 / 20	20 / 50	20 / 50
k	2 / 2	1 / 1	2 / 2	3 / 3

B. Wiener Process Model

The PN is also widely described by the Wiener process [26], [36]–[39] model as

$$\phi_p(n) = \sum_{m=0}^{n} v_p(m), \ \theta_q(n) = \sum_{m=0}^{n} u_q(m) \tag{9}$$

for $p, q, n \in \boldsymbol{\Omega}_{N_t,N_r,N}$, where $\{v_p(m)\}_{m=0}^{N-1}$ and $\{u_q(m)\}_{m=0}^{N-1}$ follow the independent identically distributed (i.i.d) real Gaussian distribution in time domain and the average powers are σ_v^2 and σ_u^2, respectively.

Consequently, the PN of both transmitting and receiving antennas follows the multivariate real Gaussian distribution as

$$P\{\boldsymbol{\phi}_p\} = \mathcal{N}(\boldsymbol{\phi}_p | \mathbf{0}_N, \sigma_v^2 \boldsymbol{\Sigma}), \ P\{\boldsymbol{\theta}_q\} = \mathcal{N}(\boldsymbol{\theta}_q | \mathbf{0}_N, \sigma_u^2 \boldsymbol{\Sigma}) \tag{10}$$

for $p, q \in \boldsymbol{\Omega}_{N_t,N_r}$, where the normalized covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{N \times N}$ and its inverse matrix $\boldsymbol{\Sigma}^{-1} \in \mathbb{R}^{N \times N}$ are

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 2 & 2 & \cdots & 2 \\ 1 & 2 & 3 & 3 & \cdots & 3 \\ 1 & 2 & 3 & 4 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & N-1 \\ 1 & 2 & 3 & \cdots & N-1 & N \end{bmatrix} \tag{11}$$

and

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} 2 & -1 & 0 & \cdots & \cdots & 0 \\ -1 & 2 & -1 & \ddots & & \vdots \\ 0 & -1 & 2 & -1 & \ddots & \vdots \\ \vdots & \ddots & -1 & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 2 & -1 \\ 0 & \cdots & \cdots & 0 & -1 & 1 \end{bmatrix}, \tag{12}$$

respectively. In fact, the (m, n)-th element of the normalized covariance matrix can be calculated by

$$\begin{aligned} \Sigma_{m,n} &= \frac{1}{\sigma_v^2} E[\phi_p(m)\phi_p(n)] = \frac{1}{\sigma_v^2} \sum_{k=0}^{m} \sum_{l=0}^{n} E[v_p(k)v_p(l)] \\ &= \frac{1}{\sigma_v^2} \sum_{k=0}^{\min(m,n)} E[v_p^2(k)] = \min(m, n) + 1 \end{aligned} \tag{13}$$

for $p, m, n \in \boldsymbol{\Omega}_{N_t,N,N}$, because $E[v_p(k)v_p(l)] = 0$ for $k \neq l \in \boldsymbol{\Omega}_{N,N}$ and $E[v_p(k)v_p(l)] = \sigma_v^2$ for $k = l \in \boldsymbol{\Omega}_{N,N}$. In addition, the same conclusion can be obtained by calculating $\Sigma_{m,n} = \frac{1}{\sigma_u^2} E[\theta_q(m)\theta_q(n)]$ for $q, m, n \in \boldsymbol{\Omega}_{N_r,N,N}$.

It can be obtained from Eq. (9) and (10) that

$$P\{\boldsymbol{\phi}\} = \mathcal{N}(\boldsymbol{\phi}|\mathbf{0}_{N_tN}, \sigma_v^2\boldsymbol{\Sigma}_{N_t}),$$
$$P\{\boldsymbol{\theta}\} = \mathcal{N}(\boldsymbol{\theta}|\mathbf{0}_{N_rN}, \sigma_u^2\boldsymbol{\Sigma}_{N_r}), \tag{14}$$

where $\boldsymbol{\Sigma}_{N_t} \in \mathbb{C}^{N_tN\times N_tN}$ and $\boldsymbol{\Sigma}_{N_r} \in \mathbb{C}^{N_rN\times N_rN}$ are the corresponding block diagonal matrices of the normalized covariance matrix $\boldsymbol{\Sigma}$ with different dimensions.

The above properties of block diagonal matrices of $\boldsymbol{\Sigma}_{N_t}^{-1}$ and $\boldsymbol{\Sigma}_{N_r}^{-1}$ as well as the tridiagonal matrix of $\boldsymbol{\Sigma}^{-1}$ can simplify the optimization process described in Section IV significantly.

IV. JCESD-PN-MAP Method

In this section, the details of our proposed JCESD-PN-MAP method will be presented. In Section IV(A), the MAP criterion for the JCESD and PN estimation is derived based on Bayesian theories. In Section IV(B) - (E), the optimizations for PN of receiving antennas, PN of transmitting antennas, channel responses and transmitted signals[6] are described, respectively. In Section IV(F), whole process of our proposed JCESD-PN-MAP method and initialization scheme are described. In Section IV(G), the computation complexity of our proposed JCESD-PN-MAP and compared methods is given.

A. MAP Criterion

According to Bayesian theories, the posteriori probability related to the process in Eq. (2) can be represented as

$$P\{\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{h},\bar{\boldsymbol{x}}|\bar{\boldsymbol{y}}\} = \frac{P\{\bar{\boldsymbol{y}}|\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{h},\bar{\boldsymbol{x}}\}P\{\boldsymbol{\theta}\}P\{\boldsymbol{\phi}\}P\{\boldsymbol{h}\}P\{\bar{\boldsymbol{x}}\}}{P\{\bar{\boldsymbol{y}}\}}, \tag{15}$$

and the MAP criterion [36], [37], [39] is to maximize the posteriori probability or minimize its negative logarithmic function as

$$\min_{\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{h},\bar{\boldsymbol{x}}} \Big\{ G(\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{h},\bar{\boldsymbol{x}}) = -\ln P\{\bar{\boldsymbol{y}}|\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{h},\bar{\boldsymbol{x}}\} - \ln P\{\boldsymbol{\theta}\} - \ln P\{\boldsymbol{\phi}\} - \ln P\{\boldsymbol{h}\} - \ln P\{\bar{\boldsymbol{x}}\}\Big\}, \tag{16}$$

where the constant $\ln P\{\bar{\boldsymbol{y}}\}$ is ignored. In addition, $P\{\boldsymbol{\phi}\}$ and $P\{\boldsymbol{\theta}\}$ are given in Eq. (14). Moreover, it is assumed that

$$P\{\bar{\boldsymbol{y}}|\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{h},\bar{\boldsymbol{x}}\} = \mathcal{CN}(\bar{\boldsymbol{y}} - \boldsymbol{T}(\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{h},\bar{\boldsymbol{x}})|\mathbf{0}_{N_rN}, \sigma_w^2\boldsymbol{I}_{N_rN}),$$
$$P\{\bar{\boldsymbol{x}}\} = \mathcal{CN}(\bar{\boldsymbol{x}}|\mathbf{0}_{N_tN}, \sigma_x^2\boldsymbol{I}_{N_tN}),$$
$$P\{\boldsymbol{h}\} = \mathcal{CN}(\boldsymbol{h}|\mathbf{0}_{N_rN_tL}, \sigma_h^2\boldsymbol{I}_{N_rN_tL}), \tag{17}$$

in which σ_x^2, σ_h^2 and σ_w^2 are the average powers of $\bar{\boldsymbol{x}}$, $\boldsymbol{h}$ and $\bar{\boldsymbol{w}}$, respectively [36].

The details of negative logarithmic function $G(\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{h},\bar{\boldsymbol{x}})$ in Eq. (16) can be seen in **Appendix**. The PN of receiving antennas, PN of transmitting antennas, channel responses and transmitted signals are estimated alternately according to Eq. (16) after the initializations of them by applying the pilots, and the negative logarithmic function decreases through each step in the above iterations. The whole process of our proposed JCESD-PN-MAP method as well as the initialization scheme is summarized in subsection (F) according to the derivations given in subsection (B) - (E).

[6]In fact, the transmitted signals that need to be estimated are the unknown ones and the pilots are known.

B. Estimation for PN of Receiving Antennas

Eq. (2) can be rewritten as

$$\bar{\boldsymbol{y}} = \boldsymbol{T}(\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{h},\bar{\boldsymbol{x}}) + \bar{\boldsymbol{w}} = \boldsymbol{F}_{N_r}\boldsymbol{D}_{\boldsymbol{s}}e^{j\boldsymbol{\theta}} + \bar{\boldsymbol{w}} \tag{18}$$

for the estimation of PN of receiving antennas $\boldsymbol{\theta}$, where $\boldsymbol{D}_{\boldsymbol{s}} \in \mathbb{C}^{N_rN\times N_rN}$ is the corresponding diagonal matrix of the vector $\boldsymbol{s} = \boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{r} = \boldsymbol{F}_{N_r}^{\mathrm{H}}\boldsymbol{D}_{\bar{\boldsymbol{h}}}\boldsymbol{F}_{N_t}\boldsymbol{r} \in \mathbb{C}^{N_rN}$. In addition, $\boldsymbol{r} = \boldsymbol{D}_{\boldsymbol{\phi}}\boldsymbol{x} \in \mathbb{C}^{N_tN}$ and $\boldsymbol{x} = \boldsymbol{F}_{N_t}^{\mathrm{H}}\bar{\boldsymbol{x}} \in \mathbb{C}^{N_tN}$.

1) Objective Function: According to Eq. (16) and (18), the MAP criterion for estimation of PN of receiving antennas $\boldsymbol{\theta}$ is a minimization problem as

$$\min_{\boldsymbol{\theta}} G_0(\boldsymbol{\theta}) = \min_{\boldsymbol{\theta}}\Big\{\frac{||\bar{\boldsymbol{y}} - \boldsymbol{F}_{N_r}\boldsymbol{D}_{\boldsymbol{s}}e^{j\boldsymbol{\theta}}||_{\mathrm{F}}^2}{\sigma_w^2} + \frac{\boldsymbol{\theta}^T\boldsymbol{\Sigma}_{N_r}^{-1}\boldsymbol{\theta}}{2\sigma_u^2}\Big\} \tag{19}$$

by giving the estimated PN of transmitting antennas $\boldsymbol{\phi}$, estimated channel responses $\boldsymbol{h}$ and estimated transmitted signals $\bar{\boldsymbol{x}}$ and ignoring other constant terms relative to $\boldsymbol{\theta}$.

Considering the linear approximation $e^{j\boldsymbol{\theta}} \approx \mathbf{1} + j\boldsymbol{\theta}$ [36]–[38] and the unitary matrix $\boldsymbol{F}_{N_r}$, the minimization problem (19) can be rewritten as

$$\min_{\boldsymbol{\theta}} G_0'(\boldsymbol{\theta}) = \min_{\boldsymbol{\theta}}\Big\{\frac{||\boldsymbol{z} - j\boldsymbol{D}_{\boldsymbol{s}}\boldsymbol{\theta}||_{\mathrm{F}}^2}{\sigma_w^2} + \frac{\boldsymbol{\theta}^T\boldsymbol{\Sigma}_{N_r}^{-1}\boldsymbol{\theta}}{2\sigma_u^2}\Big\}, \tag{20}$$

where $\boldsymbol{z} = \boldsymbol{y} - \boldsymbol{s} \in \mathbb{C}^{N_rN}$ and $\boldsymbol{y} = \boldsymbol{F}_{N_r}^{\mathrm{H}}\bar{\boldsymbol{y}} \in \mathbb{C}^{N_rN}$.

2) Optimization Method: The optimization problem (20) is convex and equivalent to solving $\frac{\partial G_0'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \mathbf{0}_{N_rN}$, which is

$$\Big(\boldsymbol{D}_{\boldsymbol{s}}^{\mathrm{H}}\boldsymbol{D}_{\boldsymbol{s}} + \frac{\sigma_w^2}{2\sigma_u^2}\boldsymbol{\Sigma}_{N_r}^{-1}\Big)\boldsymbol{\theta} = \Im\{\boldsymbol{D}_{\boldsymbol{s}}^{\mathrm{H}}\boldsymbol{z}\}, \tag{21}$$

and can be solved by the GE method with the computational complexity of $8N_rN$ [39], [40] rather than $\mathcal{O}(N_r^3N^3)$ given in [36], [37] because of the tridiagonal structure of $\boldsymbol{D}_{\boldsymbol{s}}^{\mathrm{H}}\boldsymbol{D}_{\boldsymbol{s}} + \frac{\sigma_w^2}{2\sigma_u^2}\boldsymbol{\Sigma}_{N_r}^{-1}$ and no calculation of its inverse matrix.

C. Estimation for PN of Transmitting Antennas

Eq. (2) can be rewritten as

$$\bar{\boldsymbol{y}} = \boldsymbol{T}(\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{h},\bar{\boldsymbol{x}}) + \bar{\boldsymbol{w}} = \boldsymbol{F}_{N_r}\boldsymbol{D}_{\boldsymbol{\theta}}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{D}_{\boldsymbol{x}}e^{j\boldsymbol{\phi}} + \bar{\boldsymbol{w}} \tag{22}$$

for the estimation of PN of transmitting antennas $\boldsymbol{\phi}$, where $\boldsymbol{D}_{\boldsymbol{x}} \in \mathbb{C}^{N_tN\times N_tN}$ is the corresponding diagonal matrix of $\boldsymbol{x} = \boldsymbol{F}_{N_t}^{\mathrm{H}}\bar{\boldsymbol{x}} \in \mathbb{C}^{N_tN}$.

1) Objective Function: According to Eq. (16) and (22), the MAP criterion for estimation of PN of transmitting antennas $\boldsymbol{\phi}$ is a minimization problem as

$$\min_{\boldsymbol{\phi}} G_1(\boldsymbol{\phi}) = \min_{\boldsymbol{\phi}}\Big\{\frac{||\bar{\boldsymbol{y}} - \boldsymbol{F}_{N_r}\boldsymbol{D}_{\boldsymbol{\theta}}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{D}_{\boldsymbol{x}}e^{j\boldsymbol{\phi}}||_{\mathrm{F}}^2}{\sigma_w^2} + \frac{\boldsymbol{\phi}^T\boldsymbol{\Sigma}_{N_t}^{-1}\boldsymbol{\phi}}{2\sigma_v^2}\Big\} \tag{23}$$

by giving the estimated PN of receiving antennas $\boldsymbol{\theta}$, estimated channel responses $\boldsymbol{h}$ and estimated transmitted signals $\bar{\boldsymbol{x}}$ and ignoring other constant terms relative to $\boldsymbol{\phi}$.

Considering the linear approximation $e^{j\boldsymbol{\phi}} \approx \mathbf{1} + j\boldsymbol{\phi}$ [36]–[38] as well as the unitary matrices $\boldsymbol{F}_{N_r}$ and $\boldsymbol{D}_{\boldsymbol{\theta}}$, the minimization problem (23) can be rewritten as

$$\min_{\boldsymbol{\phi}} G_1'(\boldsymbol{\phi}) = \min_{\boldsymbol{\phi}}\Big\{\frac{||\hat{\boldsymbol{z}} - j\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{D}_{\boldsymbol{x}}\boldsymbol{\phi}||_{\mathrm{F}}^2}{\sigma_w^2} + \frac{\boldsymbol{\phi}^T\boldsymbol{\Sigma}_{N_t}^{-1}\boldsymbol{\phi}}{2\sigma_v^2}\Big\}, \tag{24}$$

where $\hat{\boldsymbol{z}} = \hat{\boldsymbol{y}} - \hat{\boldsymbol{s}} \in \mathbb{C}^{N_r N}$, $\hat{\boldsymbol{y}} = \boldsymbol{D}_{\boldsymbol{\theta}}^{\mathrm{H}} \boldsymbol{y} \in \mathbb{C}^{N_r N}$ and $\hat{\boldsymbol{s}} = \boldsymbol{C}_{\boldsymbol{h}} \boldsymbol{x} = \boldsymbol{F}_{N_r}^{\mathrm{H}} \boldsymbol{D}_{\bar{\boldsymbol{h}}} \boldsymbol{F}_{N_t} \boldsymbol{x} \in \mathbb{C}^{N_r N}$. In addition, $\boldsymbol{x} = \boldsymbol{F}_{N_t}^{\mathrm{H}} \bar{\boldsymbol{x}} \in \mathbb{C}^{N_t N}$ and $\boldsymbol{y} = \boldsymbol{F}_{N_r}^{\mathrm{H}} \bar{\boldsymbol{y}} \in \mathbb{C}^{N_r N}$.

2) First Optimization Method of Eq. (24): The optimization problem (24) is convex and equivalent to solving $\frac{\partial G_1'(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}} = \boldsymbol{0}_{N_t N}$, which is

$$\left(\Re\{\boldsymbol{D}_{\boldsymbol{x}}^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{h}}^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{h}} \boldsymbol{D}_{\boldsymbol{x}}\} + \frac{\sigma_w^2}{2\sigma_v^2} \boldsymbol{\Sigma}_{N_t}^{-1}\right) \boldsymbol{\phi} = \Im\{\boldsymbol{D}_{\boldsymbol{x}}^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{h}}^{\mathrm{H}} \hat{\boldsymbol{z}}\}, \quad (25)$$

and can be solved exactly by the GE method with the computational complexity of $\mathcal{O}(N_t^3 N L^2)$ [40] rather than $\mathcal{O}(N_t^3 N^3)$ given in [36], [37] because of the same sparse structure of $\Re\{\boldsymbol{D}_{\boldsymbol{x}}^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{h}}^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{h}} \boldsymbol{D}_{\boldsymbol{x}}\} + \frac{\sigma_w^2}{2\sigma_v^2} \boldsymbol{\Sigma}_{N_t}^{-1}$ as $\boldsymbol{C}_{\boldsymbol{h}}^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{h}}$ and no calculation of its inverse matrix. Generally, L is much smaller than N.

3) Second Optimization Method of Eq. (24): The first optimization method reduces the computational complexity from $\mathcal{O}(N_t^3 N^3)$ to $\mathcal{O}(N_t^3 N L^2)$. when N_t and L are not very small, the minimization problem (24) can be optimized by the ADMM method [41] and transformed into two simpler minimization subproblems to further reduce the computational complexity, which will be simply described.

The minimization problem (24) is equivalent to a conditional minimization problem, which is

$$\begin{aligned} \min_{\hat{\boldsymbol{x}}, \boldsymbol{\phi}} G_1''(\hat{\boldsymbol{x}}, \boldsymbol{\phi}) = \min_{\hat{\boldsymbol{x}}, \boldsymbol{\phi}} \Big\{ \frac{||\hat{\boldsymbol{z}} - j\boldsymbol{C}_{\boldsymbol{h}} \hat{\boldsymbol{x}}||_{\mathrm{F}}^2}{\sigma_w^2} + \frac{\boldsymbol{\phi}^T \boldsymbol{\Sigma}_{N_t}^{-1} \boldsymbol{\phi}}{2\sigma_v^2} \Big\}, \\ s.t.\ \hat{\boldsymbol{x}} = \boldsymbol{D}_{\boldsymbol{x}} \boldsymbol{\phi}, \end{aligned} \quad (26)$$

where $\hat{\boldsymbol{x}}$ is a intermediate variable. In addition, its augmented Lagrange function is

$$\begin{aligned} L(\hat{\boldsymbol{x}}, \boldsymbol{\phi}, \boldsymbol{\lambda}) = & \frac{||\hat{\boldsymbol{z}} - j\boldsymbol{C}_{\boldsymbol{h}} \hat{\boldsymbol{x}}||_{\mathrm{F}}^2}{\sigma_w^2} + \frac{\boldsymbol{\phi}^T \boldsymbol{\Sigma}_{N_t}^{-1} \boldsymbol{\phi}}{2\sigma_v^2} + \frac{\rho}{2} ||\hat{\boldsymbol{x}} - \boldsymbol{D}_{\boldsymbol{x}} \boldsymbol{\phi}||_{\mathrm{F}}^2 \\ & + \Re\{\boldsymbol{\lambda}^{\mathrm{H}} (\hat{\boldsymbol{x}} - \boldsymbol{D}_{\boldsymbol{x}} \boldsymbol{\phi})\}, \end{aligned} \quad (27)$$

where $\rho \in R$ is the penalty factor and $\boldsymbol{\lambda} \in \mathbb{C}^{N_t N}$ is the Lagrange multiplier, and the optimization process can be represented as

$$\hat{\boldsymbol{x}}_{(k+1)} = \min_{\hat{\boldsymbol{x}}} L(\hat{\boldsymbol{x}}, \boldsymbol{\phi}_{(k)}, \boldsymbol{\lambda}_{(k)}), \quad (28)$$

$$\boldsymbol{\phi}_{(k+1)} = \min_{\boldsymbol{\phi}} L(\hat{\boldsymbol{x}}_{(k+1)}, \boldsymbol{\phi}, \boldsymbol{\lambda}_{(k)}), \quad (29)$$

$$\boldsymbol{\lambda}_{(k+1)} = \boldsymbol{\lambda}^{(k)} + \rho \frac{\partial L(\hat{\boldsymbol{x}}_{(k+1)}, \boldsymbol{\phi}_{(k+1)}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} \Big|_{\boldsymbol{\lambda} = \boldsymbol{\lambda}_{(k)}}, \quad (30)$$

where k is the index of iterations of ADMM method and $\frac{\partial L(\hat{\boldsymbol{x}}_{(k+1)}, \boldsymbol{\phi}_{(k+1)}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} \Big|_{\boldsymbol{\lambda} = \boldsymbol{\lambda}_{(k)}} = (\hat{\boldsymbol{x}}_{(k+1)} - \boldsymbol{D}_{\boldsymbol{x}} \boldsymbol{\phi}_{(k+1)})^{\mathrm{H}}$.

In addition, the optimization problem (28) is convex and equivalent to solving $\frac{\partial L(\hat{\boldsymbol{x}}, \boldsymbol{\phi}_{(k)}, \boldsymbol{\lambda}_{(k)})}{\partial \hat{\boldsymbol{x}}} \Big|_{\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}_{(k+1)}} = \boldsymbol{0}_{N_t N}$, which is

$$\begin{aligned} & \left(2\boldsymbol{D}_{\bar{\boldsymbol{h}}}^{\mathrm{H}} \boldsymbol{D}_{\bar{\boldsymbol{h}}} + \rho\sigma_w^2 \boldsymbol{I}_{N_t N}\right) \boldsymbol{F}_{N_t} \hat{\boldsymbol{x}}_{(k+1)} = \\ & \quad - 2j \boldsymbol{D}_{\bar{\boldsymbol{h}}}^{\mathrm{H}} \boldsymbol{F}_{N_r} \hat{\boldsymbol{z}} + \boldsymbol{F}_{N_t} \left(\rho\sigma_w^2 \boldsymbol{D}_{\boldsymbol{x}} \boldsymbol{\phi}_{(k)} - \sigma_w^2 \boldsymbol{\lambda}_{(k)}^{\mathrm{H}}\right) \end{aligned} \quad (31)$$

by considering that $\boldsymbol{C}_{\boldsymbol{h}} = \boldsymbol{F}_{N_r}^{\mathrm{H}} \boldsymbol{D}_{\bar{\boldsymbol{h}}} \boldsymbol{F}_{N_t}$.

The inverse matrix of $2\boldsymbol{D}_{\bar{\boldsymbol{h}}}^{\mathrm{H}} \boldsymbol{D}_{\bar{\boldsymbol{h}}} + \rho\sigma_w^2 \boldsymbol{I}_{N_t N} \in \mathbb{C}^{N_t N \times N_t N}$ can be calculated with the computation complexity of $\mathcal{O}(N_t^3 N)$ because of the same block diagonal structure of $2\boldsymbol{D}_{\bar{\boldsymbol{h}}}^{\mathrm{H}} \boldsymbol{D}_{\bar{\boldsymbol{h}}} + \rho\sigma_w^2 \boldsymbol{I}_{N_t N}$ as $\boldsymbol{D}_{\bar{\boldsymbol{h}}}^{\mathrm{H}} \boldsymbol{D}_{\bar{\boldsymbol{h}}}$. In addition, it should be noted that the matrices $\left(2\boldsymbol{D}_{\bar{\boldsymbol{h}}}^{\mathrm{H}} \boldsymbol{D}_{\bar{\boldsymbol{h}}} + \rho\sigma_w^2 \boldsymbol{I}_{N_t N}\right)^{-1}$ and $\boldsymbol{D}_{\bar{\boldsymbol{h}}}^{\mathrm{H}} \boldsymbol{F}_{N_r} \hat{\boldsymbol{z}}$ only need to be calculated once for iterations. Therefore, the computation complexity for the optimization of Eq. (31) is $\mathcal{O}(N_t^3 N + N_r N \log N + K \cdot 2N_t N \log N)$, where K is the number of iterations of ADMM method.

Furthermore, the optimization problem (29) is convex and equivalent to solving $\frac{\partial L(\hat{\boldsymbol{x}}_{(k+1)}, \boldsymbol{\phi}, \boldsymbol{\lambda}_{(k)})}{\partial \boldsymbol{\phi}} \Big|_{\boldsymbol{\phi} = \boldsymbol{\phi}_{(k+1)}} = \boldsymbol{0}_{N_t N}$, which is

$$\begin{aligned} & \left(\boldsymbol{\Sigma}_{N_t}^{-1} + \rho\sigma_v^2 \boldsymbol{D}_{\boldsymbol{x}}^{\mathrm{H}} \boldsymbol{D}_{\boldsymbol{x}}\right) \boldsymbol{\phi}_{(k+1)} = \\ & \qquad \Re\{\rho\sigma_v^2 \boldsymbol{D}_{\boldsymbol{x}}^{\mathrm{H}} \hat{\boldsymbol{x}}_{(k+1)} + \sigma_v^2 \boldsymbol{D}_{\boldsymbol{x}}^{\mathrm{H}} \boldsymbol{\lambda}_{(k)}^{\mathrm{H}}\}, \end{aligned} \quad (32)$$

and can be solved with the computational complexity of $K \cdot 8N_t N$ [39], [40] because of the tridiagonal structure of $\boldsymbol{\Sigma}_{N_t}^{-1} + \rho\sigma_v^2 \boldsymbol{D}_{\boldsymbol{x}}^{\mathrm{H}} \boldsymbol{D}_{\boldsymbol{x}} \in \mathbb{R}^{N_t N \times N_t N}$.

Moreover, the penalty factor $\rho = \frac{1}{\sigma_w^2}$ and the number of iterations $K = 3$ are selected in the simulation, and lower computational complexity can be obtained through the second optimization method when N_t and L are not very small, which can been seen in Section IV(G).

D. Estimation for Channel Responses

Eq. (2) can be rewritten as

$$\bar{\boldsymbol{y}} = \boldsymbol{T}(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{h}, \bar{\boldsymbol{x}}) + \bar{\boldsymbol{w}} = \boldsymbol{F}_{N_r} \boldsymbol{D}_{\boldsymbol{\theta}} \boldsymbol{D}_{\boldsymbol{r}} \boldsymbol{h} + \bar{\boldsymbol{w}} \quad (33)$$

for the estimation of channel responses $\boldsymbol{h}$, where

$$\boldsymbol{D}_{\boldsymbol{r}} = \begin{bmatrix} \boldsymbol{C}_{\boldsymbol{r}} & & & \\ & \boldsymbol{C}_{\boldsymbol{r}} & & \\ & & \ddots & \\ & & & \boldsymbol{C}_{\boldsymbol{r}} \end{bmatrix} \in \mathbb{C}^{N_r N \times N_r N_t L},$$

$\boldsymbol{C}_{\boldsymbol{r}} = [\boldsymbol{C}_{\boldsymbol{r}_1}, \cdots, \boldsymbol{C}_{\boldsymbol{r}_{N_t - 1}}] \in \mathbb{C}^{N \times N_t L}$ and $\boldsymbol{C}_{\boldsymbol{r}_q} \in \mathbb{C}^{N \times L}$ is the first L columns of corresponding circular matrix of $\boldsymbol{r}_q$ and $\boldsymbol{r}_q \in \mathbb{C}^N$ is composed of the qN-th to $(q+1)N$-th elements of $\boldsymbol{r}$. In addition, $\boldsymbol{r} = \boldsymbol{D}_{\boldsymbol{\phi}} \boldsymbol{x} \in \mathbb{C}^{N_t N}$ and $\boldsymbol{x} = \boldsymbol{F}_{N_t}^{\mathrm{H}} \bar{\boldsymbol{x}} \in \mathbb{C}^{N_t N}$.

1) Objective Function: According to Eq. (16) and (33), the MAP criterion for the estimation of channel responses $\boldsymbol{h}$ is a minimization problem as

$$\min_{\boldsymbol{h}} G_2(\boldsymbol{h}) = \min_{\boldsymbol{h}} \Big\{ \frac{||\bar{\boldsymbol{y}} - \boldsymbol{F}_{N_r} \boldsymbol{D}_{\boldsymbol{\theta}} \boldsymbol{D}_{\boldsymbol{r}} \boldsymbol{h}||_{\mathrm{F}}^2}{\sigma_w^2} + \frac{||\boldsymbol{h}||_{\mathrm{F}}^2}{\sigma_h^2} \Big\} \quad (34)$$

by giving the estimated PN of receiving antennas $\boldsymbol{\theta}$, estimated PN of transmitting antennas $\boldsymbol{\phi}$ and estimated transmitted signals $\bar{\boldsymbol{x}}$ and ignoring other constant terms relative to $\boldsymbol{h}$.

2) Optimization Method: The optimization problem (34) is convex and equivalent to solving $\frac{\partial G_2(\boldsymbol{h})}{\partial \boldsymbol{h}} = \boldsymbol{0}_{N_r N_t L}$, which is

$$\left(\boldsymbol{D}_{\boldsymbol{r}}^{\mathrm{H}} \boldsymbol{D}_{\boldsymbol{r}} + \frac{\sigma_w^2}{\sigma_h^2} \boldsymbol{I}_{N_r N_t L}\right) \boldsymbol{h} = \boldsymbol{D}_{\boldsymbol{r}}^{\mathrm{H}} \hat{\boldsymbol{y}}, \quad (35)$$

where $\hat{\boldsymbol{y}} = \boldsymbol{D}_{\boldsymbol{\theta}}^{\mathrm{H}} \boldsymbol{y} \in \mathbb{C}^{N_r N}$ and $\boldsymbol{y} = \boldsymbol{F}_{N_r}^{\mathrm{H}} \bar{\boldsymbol{y}} \in \mathbb{C}^{N_r N}$.

The inverse matrix of $\boldsymbol{D}_{\boldsymbol{r}}^{\mathrm{H}} \boldsymbol{D}_{\boldsymbol{r}} + \frac{\sigma_w^2}{\sigma_h^2} \boldsymbol{I}_{N_r N_t L} \in \mathbb{C}^{N_r N_t L \times N_r N_t L}$ can be calculated with the computational complexity of $\mathcal{O}(N_t^3 L^3)$ rather than $\mathcal{O}(N_r^3 N_t^3 L^3)$ given in [36] because the matrix $\boldsymbol{D}_{\boldsymbol{r}}^{\mathrm{H}} \boldsymbol{D}_{\boldsymbol{r}} + \frac{\sigma_w^2}{\sigma_h^2} \boldsymbol{I}_{N_r N_t L}$ is a low dimensional block diagonal matrix with N_r identical blocks $\boldsymbol{C}_{\boldsymbol{r}}^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{r}} + \frac{\sigma_w^2}{\sigma_h^2} \boldsymbol{I}_{N_t L} \in \mathbb{C}^{N_t L \times N_t L}$.

E. Estimation for Transmitted Signals

1) Objective Function: According to Eq. (16) and (3), the MAP criterion for the estimation of transmitted signals $\bar{\boldsymbol{x}}$ is a minimization problem as

$$\min_{\bar{\boldsymbol{x}}} G_3(\bar{\boldsymbol{x}}) = \min_{\bar{\boldsymbol{x}}} \left\{ \frac{||\bar{\boldsymbol{y}} - \boldsymbol{F}_{N_r} \boldsymbol{D}_{\boldsymbol{\theta}} \boldsymbol{C}_{\boldsymbol{h}} \boldsymbol{D}_{\boldsymbol{\phi}} \boldsymbol{F}_{N_t}^{\mathrm{H}} \bar{\boldsymbol{x}}||_{\mathrm{F}}^2}{\sigma_w^2} + \frac{||\bar{\boldsymbol{x}}||_{\mathrm{F}}^2}{\sigma_x^2} \right\} \tag{36}$$

by giving the estimated PN of receiving antennas $\boldsymbol{\theta}$, estimated PN of transmitting antennas $\boldsymbol{\phi}$ and estimated channel responses $\boldsymbol{h}$ and ignoring other constant terms relative to $\bar{\boldsymbol{x}}$.

2) Optimization Method: The optimization problem (36) is convex and equivalent to solving $\frac{\partial G_3(\bar{\boldsymbol{x}})}{\partial \bar{\boldsymbol{x}}} = \boldsymbol{0}_{N_t N}$, which is

$$\left(\boldsymbol{D}_{\bar{\boldsymbol{h}}}^{\mathrm{H}} \boldsymbol{D}_{\bar{\boldsymbol{h}}} + \frac{\sigma_w^2}{\sigma_x^2} \boldsymbol{I}_{N_t N} \right) \boldsymbol{F}_{N_t} \boldsymbol{D}_{\boldsymbol{\phi}} \boldsymbol{F}_{N_t}^{\mathrm{H}} \bar{\boldsymbol{x}} = \boldsymbol{D}_{\bar{\boldsymbol{h}}}^{\mathrm{H}} \boldsymbol{F}_{N_r} \hat{\boldsymbol{y}} \tag{37}$$

by considering the unitary matrices $\boldsymbol{D}_{\boldsymbol{\theta}}$, $\boldsymbol{D}_{\boldsymbol{\phi}}$, $\boldsymbol{F}_{N_r}$ and $\boldsymbol{F}_{N_t}$ as well as that $\boldsymbol{C}_{\boldsymbol{h}} = \boldsymbol{F}_{N_r}^{\mathrm{H}} \boldsymbol{D}_{\bar{\boldsymbol{h}}} \boldsymbol{F}_{N_t}$ described in Eq. (5). In addition, the hard decision is performed and the pilots are replaced by their true values.

The inverse matrix of $\boldsymbol{D}_{\bar{\boldsymbol{h}}}^{\mathrm{H}} \boldsymbol{D}_{\bar{\boldsymbol{h}}} + \frac{\sigma_w^2}{\sigma_x^2} \boldsymbol{I}_{N_t N} \in \mathbb{C}^{N_t N \times N_t N}$ can be calculated with the computation complexity of $\mathcal{O}(N_t^3 N)$ because the matrix $\boldsymbol{D}_{\bar{\boldsymbol{h}}}^{\mathrm{H}} \boldsymbol{D}_{\bar{\boldsymbol{h}}} + \frac{\sigma_w^2}{\sigma_x^2} \boldsymbol{I}_{N_t N}$ has the same block diagonal structure as $\boldsymbol{D}_{\bar{\boldsymbol{h}}}^{\mathrm{H}} \boldsymbol{D}_{\bar{\boldsymbol{h}}}$. Therefore, the computation complexity for the optimization of Eq. (37) is $\mathcal{O}(N_t^3 N + 2N_t N \log N + N_r N \log N)$.

F. Summary of JCESD-PN-MAP Method

According to the derivations in subsection (B) - (E), the whole process of JCESD-PN-MAP method is summarized in **Algorithm 1** and Fig. 1, in which the convergence condition is that the estimated transmitted signals do not change and M is the maximum number of iterations.

Algorithm 1 JCESD-PN-MAP Method

Require: the received signals $\bar{\boldsymbol{y}}$ and pilots $\boldsymbol{P}\bar{\boldsymbol{x}}$;
Ensure: the PN of receiving antennas $\boldsymbol{\theta}$, PN of transmitting antennas $\boldsymbol{\phi}$, channel responses $\boldsymbol{h}$, transmitted signals $\bar{\boldsymbol{x}}$;
1: initialization: the PN of receiving antennas $\boldsymbol{\theta}$ and PN of transmitting antennas $\boldsymbol{\phi}$ are both all-zero vectors and channel responses $\boldsymbol{h}$ are given by Eq. (40);
2: **while** not converged **do**
3: update the transmitted signals $\bar{\boldsymbol{x}}$ according to Eq. (37) and hard decision;
4: update the PN of receiving antennas $\boldsymbol{\theta}$ according to Eq. (21);
5: update the PN of transmitting antennas $\boldsymbol{\phi}$ according to Eq. (25) or Eq. (28) - (30);
6: update the channel responses $\boldsymbol{h}$ according to Eq. (35);
7: **end while**

The initialization scheme will be simply described. As described in Section II(C), the initializations of both PN of receiving antennas $\boldsymbol{\theta}$ and PN of transmitting antennas $\boldsymbol{\phi}$ can be configured as all-zero vectors. In this case, Eq. (3) can be rewritten as

$$\bar{\boldsymbol{y}} = \boldsymbol{F}_{N_r} \boldsymbol{C}_{\boldsymbol{h}} \boldsymbol{F}_{N_t}^{\mathrm{H}} \bar{\boldsymbol{x}} + \bar{\boldsymbol{w}}. \tag{38}$$

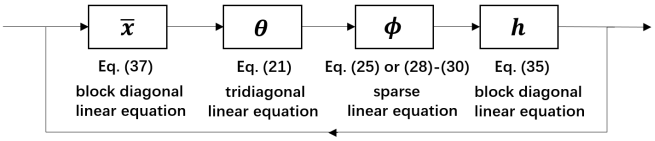


Fig. 1. One iteration of JCESD-PN-MAP method.

By considering the pilots, it can be obtained that

$$\bar{\boldsymbol{y}} = \boldsymbol{F}_{N_r} \boldsymbol{C}_{\boldsymbol{h}} \boldsymbol{F}_{N_t}^{\mathrm{H}} \boldsymbol{P} \bar{\boldsymbol{x}} + \hat{\boldsymbol{w}} = \boldsymbol{F}_{N_r} \boldsymbol{D}_{\hat{\boldsymbol{r}}} \boldsymbol{h} + \hat{\boldsymbol{w}}, \tag{39}$$

where $\hat{\boldsymbol{w}} = \boldsymbol{F}_{N_r} \boldsymbol{C}_{\boldsymbol{h}} \boldsymbol{F}_{N_t}^{\mathrm{H}} (\boldsymbol{I}_{N_t N} - \boldsymbol{P}) \bar{\boldsymbol{x}} + \bar{\boldsymbol{w}} \in \mathbb{C}^{N_r N}$ and $\boldsymbol{P} \in \mathbb{R}^{N_t N \times N_t N}$ is a diagonal matrix with the diagonal elements of 1 or 0, which represents the position of pilots. In addition, $\boldsymbol{D}_{\hat{\boldsymbol{r}}} \in \mathbb{C}^{N_r N \times N_r N_t L}$ and $\boldsymbol{C}_{\hat{\boldsymbol{r}}} \in \mathbb{C}^{N \times N_t L}$ are defined similar to $\boldsymbol{D}_{\boldsymbol{r}}$ and $\boldsymbol{C}_{\boldsymbol{r}}$ in Eq. (33) and $\hat{\boldsymbol{r}} = \boldsymbol{F}_{N_t}^{\mathrm{H}} \boldsymbol{P} \bar{\boldsymbol{x}} \in \mathbb{C}^{N_t N}$. Therefore, the initializations of channel responses $\boldsymbol{h}$ can be obtained by the LS estimation of Eq. (39), which is

$$\boldsymbol{h} = \left[\boldsymbol{D}_{\hat{\boldsymbol{r}}}^{\mathrm{H}} \boldsymbol{D}_{\hat{\boldsymbol{r}}} \right]^{-1} \boldsymbol{D}_{\hat{\boldsymbol{r}}}^{\mathrm{H}} \boldsymbol{y}. \tag{40}$$

G. Computation Complexity

The computation complexity of our proposed JCESD-PN-MAP for each iteration are analyzed as follows and summarized in **TABLE II**.

1) Estimation for PN of Receiving Antennas: The main part for the construction of Eq. (21) is the calculation of $\boldsymbol{s}$, which requires the computation complexity of $\mathcal{O}(2N_t N \log N + N_r N \log N)$. In addition, the optimization of Eq. (21) requires the computation complexity of $8N_r N$.

2) Estimation for PN of Transmitting Antennas: For the first optimization method, the main parts for the construction of Eq. (25) are the calculations of $\hat{\boldsymbol{z}}$ and $\boldsymbol{C}_{\boldsymbol{h}}^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{h}}$, which require the computation complexity of $\mathcal{O}(N_t N \log N + N_r N \log N)$ and $\mathcal{O}(N_r N_t^2 L^2)$, respectively. In addition, the optimization of Eq. (25) requires the computation complexity of $\mathcal{O}(N_t^3 N L^2)$.

For the second optimization method, the main part for the construction of Eq. (27) is the calculation of $\hat{\boldsymbol{z}}$, which requires the computation complexity of $\mathcal{O}(N_t N \log N + N_r N \log N)$. In addition, the optimization of (28)/(31) and (29)/(32) require the computation complexity of $\mathcal{O}(N_t^3 N + N_r N \log N + K \cdot 2N_t N \log N)$ and $K \cdot 8N_t N$, respectively.

3) Estimation for Channel Responses: The main parts for the construction of Eq. (35) are the calculations of $\boldsymbol{C}_{\boldsymbol{r}}^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{r}}$ and $\boldsymbol{D}_{\boldsymbol{r}}^{\mathrm{H}} \hat{\boldsymbol{y}}$, which require the computation complexity of $\mathcal{O}(N_t^2 N L^2)$ and $\mathcal{O}(N_t N L)$. In addition, the optimization of Eq. (35) requires the computation complexity of $\mathcal{O}(N_t^3 L^3)$.

4) Estimation for Transmitted Signals: The main part for the construction of Eq. (37) is the calculation of $\bar{\boldsymbol{h}}$, which requires the computation complexity of $\mathcal{O}(N_r N_t N \log N)$. In addition, the optimization of Eq. (37) requires the computation complexity of $\mathcal{O}(N_t^3 N + 2N_t N \log N + N_r N \log N)$.

The computation complexity of compared methods in [37], [36], [34] for each iteration are summarized in **TABLE II**. As described above, the computation complexity of our proposed JCESD-PN-MAP method increases at most linearly

TABLE II: The computational complexity of our proposed JCESD-PN-MAP and compared methods for each iteration

Method	Steps	Construction	Optimization
JCESD-PN-MAP Method	Eq. (21)	$\mathcal{O}(2N_tN\log N + N_rN_tNL)$	$8N_rN$
	Eq. (25) / (28) - (30)	$\mathcal{O}(N_tN\log N + N_rN\log N)$	$\mathcal{O}(N_t^3NL^2)$ or $\mathcal{O}(N_t^3N + N_rN\log N + K\cdot 2N_tN\log N + K\cdot 8N_tN)$
	Eq. (35)	$\mathcal{O}(N_t^2NL^2)$, $\mathcal{O}(N_tNL)$	$\mathcal{O}(N_t^3L^3)$
	Eq. (37)	$\mathcal{O}(N_rN_tN\log N)$	$\mathcal{O}(N_t^3N + 2N_tN\log N + N_rN\log N)$
[37]	-	$\mathcal{O}(N_r^3N^3) + \mathcal{O}(N_t^3N^3) + \mathcal{O}(N_t^3N + 2N_tN\log N + N_rN\log N)$	
[36]	-	$\mathcal{O}(N_r^3N^3) + \mathcal{O}(N_t^3N^3) + \mathcal{O}(N_r^3N_t^3L^3) + \mathcal{O}(N_t^3N + 2N_tN\log N + N_rN\log N)$	
[34]	-	$\mathcal{O}(n_gN_r^2N_tN^2) + \mathcal{O}(n_gN_r^2N_tN^2) + \mathcal{O}(N_r^3N_t^3L^3) + \mathcal{O}(N_t^3N + 2N_tN\log N + N_rN\log N)$	

with $N\log N$, N_r, K and two-order with N_t, L regardless of the computational complexity of $\mathcal{O}(N_t^3NL^2)$ and small computational complexity of N_t^3N and $\mathcal{O}\Big((N_tL)^3\Big)$. However, [37] and [36] requires the calculation of inverse matrices without the consideration of their structures and the computation complexity is more than $\mathcal{O}(N_t^3N^3 + N_r^3N^3)$. In addition, [34] requires the computation complexity of more than $\mathcal{O}(n_gN_r^2N_tN^2)$, where n_g is the number of components of truncated PN.

Therefore, under the same number of iterations, the computation complexity of our proposed JCESD-PN-MAP method is lower than methods in [37], [36] and [34], which reduces the load of the system significantly[7].

V. Numerical Results

In this section, the effectiveness of our proposed JCESD-PN-MAP method is verified through numerical results under different settings compared to existing methods.

A. Basic Settings of Simulation

In the simulation, the number of subcarriers is $N = 512$, the subcarrier interval is 60 KHz and the modulation scheme is 16-QAM or 64QAM. In addition, the numbers of transmitting antennas N_t and receiving antennas N_r are from 4 to 8. Moreover, the pilots are generated with the same modulation scheme as the transmitted signals for convenience, the proportion of pilots ρ is from 2% to 16% of all data and the position of pilots is uniformly distributed. Additionally, the channel is modeled as the multipath channel, in which each component of $\boldsymbol{h}_{p,q} \in \mathbb{C}^L$ follows the complex Gaussian distribution $\mathcal{CN}(0, \sigma_h^2)$, the number of paths is $L = 5$ and its average power is $\sigma_h^2 = 1$. Furthermore, the PN of multiple transmitting and receiving antennas is generated by the PSD model described in Section III(A) and the carrier frequency f_c is from 10 to 60 GHz. Furthermore, the signal-to-noise ratio (SNR) defined as $\frac{\sigma_h^2\cdot\sigma_x^2}{\sigma_w^2}$ is from 16 to 30 dB, in which $\sigma_x^2 = 1$ and σ_w^2 are the average powers of transmitted signals and noise, respectively. The basic settings of simulation can be seen in **TABLE III**. Finally, the performance is measured by the curves of symbol error rate (SER) and normalized mean squared error (NMSE) defined as $\frac{E||\boldsymbol{h}^e-\boldsymbol{h}||_{\mathrm{F}}^2}{E||\boldsymbol{h}||_{\mathrm{F}}^2}$, where $E(\cdot)$ and $\boldsymbol{h}^e$ are the expectations of a random variables and estimated channel responses, respectively.

TABLE III: The basic settings of simulation

	Setting
MIMO-OFDM	N_t, $N_r = 4 \sim 8$, $N = 512$, $\sigma_x^2 = 1$
Modulation	16-QAM or 64QAM
Pilot Structure	Position: Uniformly Distributed, Proportion: $\rho = 2\% \sim 16\%$
Generation of Channel	Multipath Channel, $L = 5$, $\sigma_h^2 = 1$
Generation of PN	3GPP PSD Model in Section III(A), $f_c = 10 \sim 60$ GHz
SNR	$16 \sim 30$ dB
Number of Iterations	$M = 1 \sim 5$

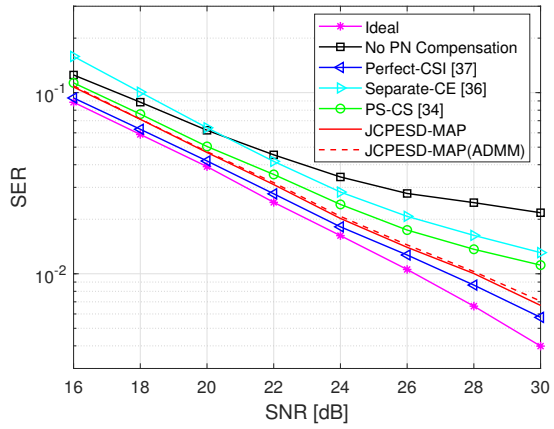
Our proposed JCESD-PN-MAP method is compared with three existing methods: (1) the joint signal detection and PN estimation method based on MAP criterion with perfect CSI [37], which is called the Perfect-CSI case, (2) the joint signal detection and PN estimation method based on MAP criterion with a separate channel estimation before the iterations [36], which is called the Separate-CE case, and (3) the pattern search and compressed sensing based joint channel and PN estimation and signal detection method [34], which is called the PS-CS method. The number of iterations of these methods is also $M = 1 \sim 5$.

In addition, the signal detection method based on MAP criterion with perfect CSI and PN is called the Ideal case. Moreover, the JCESD method based on MAP criterion without PN estimation is called the No PN Estimation case.

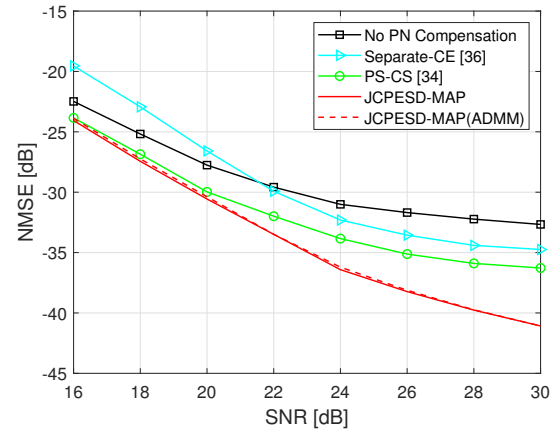
B. Simulation of Performance Comparison

Firstly, when the numbers of transmitting and receiving antennas are $N_t = N_r = 4$, the modulation scheme is 16-QAM, the proportion of pilots is $\rho = 10\%$ and the carrier frequency is $f_c = 30$ GHz, the SER curves of different methods under different SNRs are shown in Fig. 2(a). It can be seen that the SER performance of our proposed JCESD-PN-MAP method is better than the Separate-CE case significantly. A separate channel estimation is considered before the iterations in the Separate-CE case and its performance is limited by insufficient channel estimation. For our proposed JCESD-PN-MAP method, the channel estimation is considered sufficiently in the iterations to achieve better performance. In addition, the SER performance of our proposed JCESD-PN-MAP method is also better than the PS-CS method because the

[7]For example, it can be obtained that $N_t^3N^3 + N_r^3N^3 \approx 2.15\times10^9$ and $n_gN_r^2N_tN^2 \approx 2.10\times10^7$ when the numbers of subcarriers, transmitting antennas, receiving antennas, paths of channel responses, components of truncated PN and iterations of ADMM method are $N = 256$, $N_t = 4$, $N_r = 4$, $L = 5$, $n_g = 5$ and $K = 3$, respectively. In addition, it can be obtained that $N_t^3NL^2 \approx 4.10\times10^5$ and $8N_rN \approx 8.20\times10^3$ under the same settings. Moreover, it can be obtained that $N_t^3N + N_rN\log N + K\cdot 2N_tN\log N + K\cdot 8N_tN \approx 9.8\times10^4$ under the same settings.

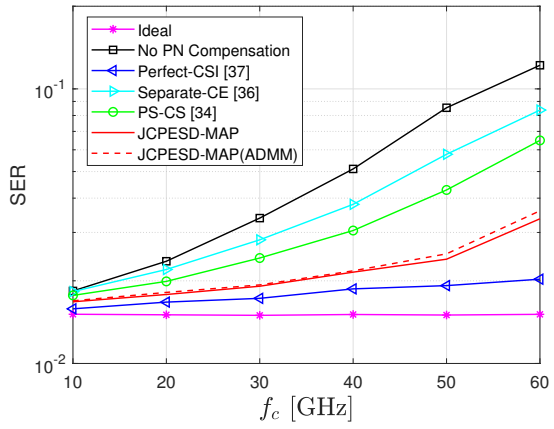


(a) SER

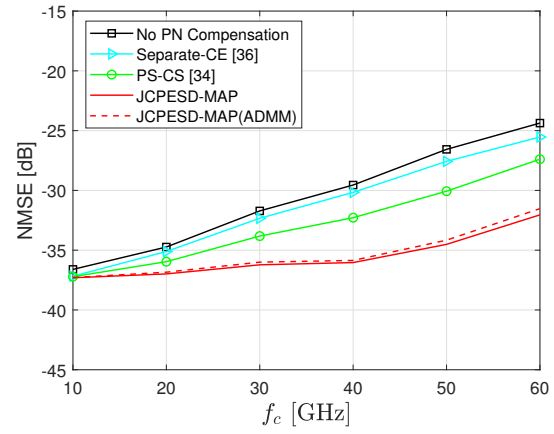


(b) NMSE

Fig. 2. The SER and NMSE curves under different SNRs: $N_t = N_r = 2$, 16-QAM, $\rho = 10\%$, $f_c = 30$ GHz.



(a) SER



(b) NMSE

Fig. 3. The SER and NMSE curves under different frequencies: $N_t = N_r = 4$, 16-QAM, $\rho = 10\%$, SNR = 25 dB.

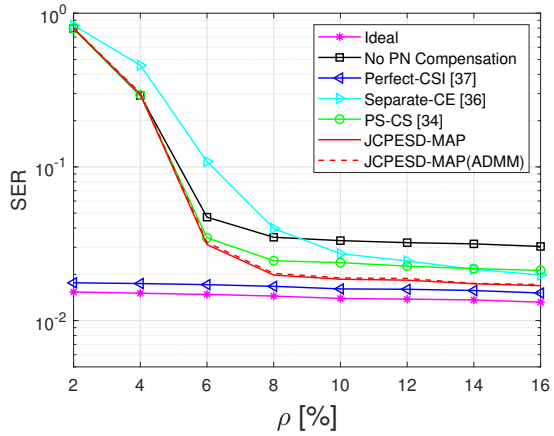
complete PN is estimated in our proposed JCESD-PN-MAP method rather than the truncated PN in the PS-CS method. Moreover, perfect CSI is required in the Perfect-CSI case and the SER performance of our proposed JCESD-PN-MAP method is close to the Perfect-CSI case, which means that our proposed JCESD-PN-MAP method can achieve great channel estimation implicitly.

Under the same settings of Fig. 2(a), the NMSE curves of different methods under different SNRs are shown in Fig. 2(b). The NMSE performance of our proposed JCESD-PN-MAP method is better than the Separate-CE case and PS-CS method, which means that our proposed JCESD-PN-MAP method can achieve better channel estimation than the Separate-CE case and PS-CS method explicitly.

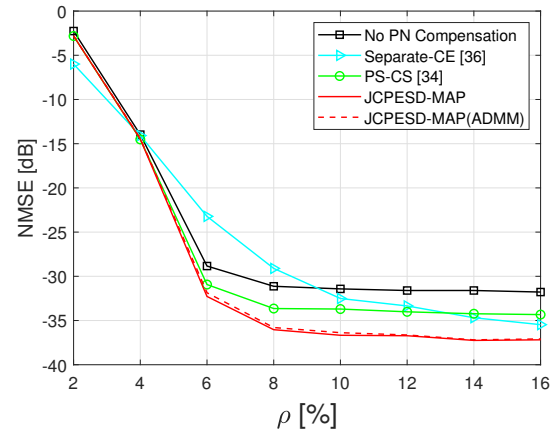
Secondly, different carrier frequencies are considered. When the numbers of transmitting and receiving antennas are $N_t = N_r = 4$, the modulation scheme is 16-QAM, the proportion of pilots is $\rho = 10\%$ and the SNR is 25 dB, the SER and NMSE curves of different methods under different carrier frequencies are shown in Fig. 3. Higher carrier frequency means higher average power of PN and it can be seen that the SER and NMSE performance of our proposed JCESD-PN-MAP method is better than the Separate-CE case and PS-CS method and a little worse than Perfect-CSI case, which confirmes the effectiveness of our proposed JCESD-PN-MAP method in signal detection and channel estimation under different carrier frequencies or average powers of PN.

Thirdly, different proportions of pilots are considered. When the numbers of transmitting and receiving antennas are $N_t = N_r = 4$, the modulation scheme is 16-QAM, the SNR is 25 dB and the carrier frequency is $f_c = 30$ GHz, the SER and NMSE curves of different methods under different proportions of pilots are shown in Fig. 4. As seen in Fig. 4(a), to approach the SER performance of the Perfect-CSI case, the proportion of pilots of 8% is enough for our proposed JCESD-PN-MAP method while the proportion of pilots of about 16% is needed for Separate-CE case. In addition, the SER performance of PS-CS method is limited even if enough proportion of pilots is given because it deals with the truncated PN. In Fig. 4(b), our proposed JCESD-PN-MAP method also outperforms the Separate-CE case and PS-CS method in the NMSE performance under different proportions of pilots.

Fourthly, different numbers of iterations are considered. When the numbers of transmitting and receiving antennas

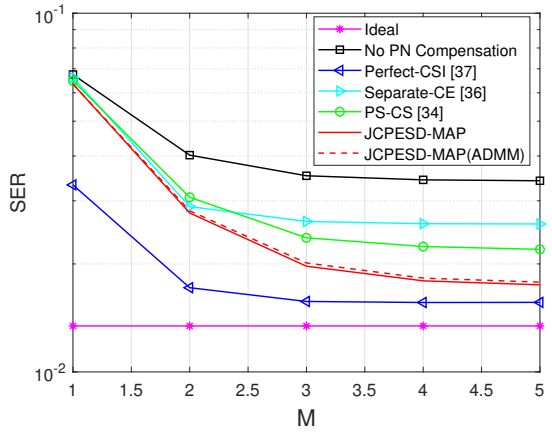


(a) SER

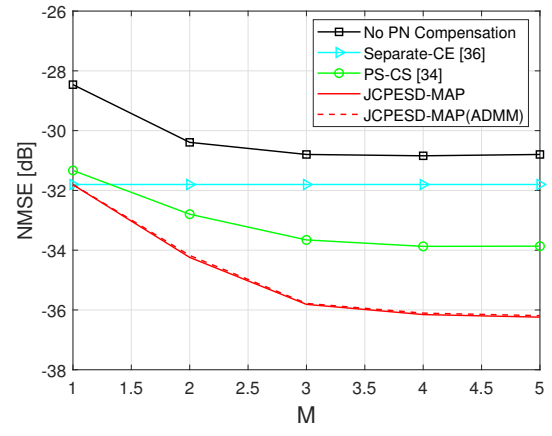


(b) NMSE

Fig. 4. The SER and NMSE curves under different proportions: $N_t = N_r = 4$, 16-QAM, $f_c = 30$ GHz, SNR = 25 dB.



(a) SER



(b) NMSE

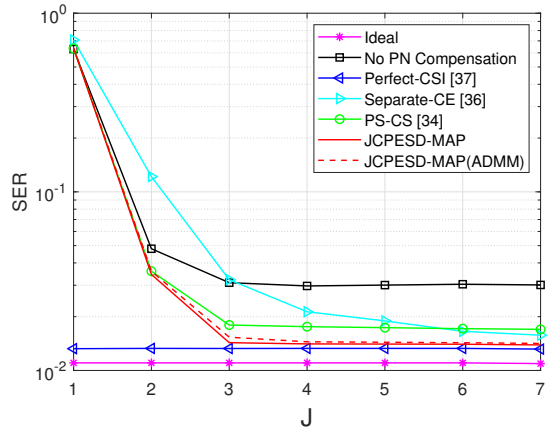
Fig. 5. The SER and NMSE curves under different numbers of iterations: $N_t = N_r = 4$, 16-QAM, $\rho = 10\%$, $f_c = 30$ GHz, SNR = 25 dB.

are $N_t = N_r = 4$, the modulation scheme is 16-QAM, the proportion of pilots is $\rho = 10\%$, the carrier frequency is $f_c = 30$ GHz and the SNR is 25 dB, the SER and NMSE curves of different methods under different numbers of iterations are shown in Fig. 5. It can also be seen that the performance of Separate-CE case and PS-CS methods is limited even if having enough number of iterations, because the separate channel estimation or truncated PN is considered. In addition, the performance of our proposed JCESD-PN-MAP method is close to that of Perfect-CSI case with enough number of iterations.

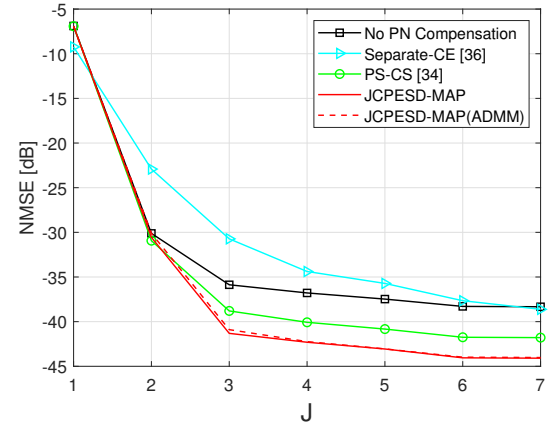
Additionally, different numbers of OFDM symbols are considered[8]. When the numbers of transmitting and receiving antennas are $N_t = N_r = 4$, the modulation scheme is 16-QAM, the proportion of pilots is $\rho = 3\%$, the carrier frequency is $f_c = 30$ GHz and the SNR is 25 dB, the SER and NMSE curves of different methods under different numbers of OFDM symbols ($J = 1 \sim 7$) are shown in Fig. 6. As seen in Fig. 6(a), to approach the SER performance of Perfect-CSI case, three OFDM symbols is enough for our proposed JCESD-PN-MAP method while more than seven OFDM symbols is needed for the Separate-CE case. In addition, the SER performance of PS-CS method is limited even if enough OFDM symbols are given because it deals with the truncated PN. In Fig. 6(b), the NMSE performance of our proposed JCESD-PN-MAP method is also better than the Separate-CE case and PS-CS method under different numbers of OFDM symbols.

Finally, different numbers of receiving antennas are considered. In addition, more receiving antennas improve the performance significantly and therefore 64-QAM is considered. When the number of transmitting antennas is $N_t = 4$, the modulation scheme is 64-QAM, the proportion of pilots is $\rho = 10\%$, the carrier frequency is $f_c = 30$ GHz and the SNR is 20 dB, the SER and NMSE curves of different methods under different numbers of receiving antennas are shown in Fig. 7. It can be seen that our proposed JCESD-PN-MAP method outperforms the Separate-CE case and PS-CS method and performs close to the Perfect-CSI case under different numbers of receiving antennas.

[8]Our proposed JCESD-PN-MAP method is extended to the case of multiple OFDM symbols in **Appendix**.

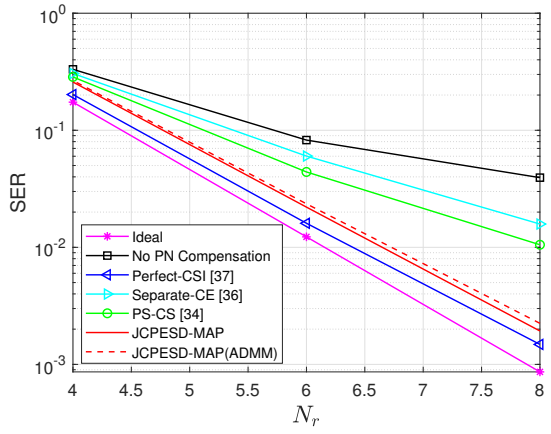


(a) SER

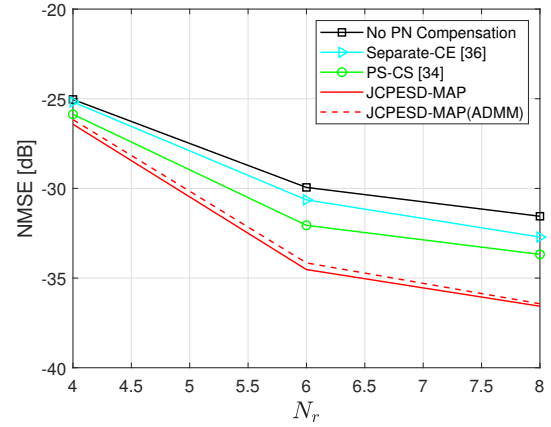


(b) NMSE

Fig. 6. The SER and NMSE curves under different numbers of OFDM symbols: $N_t = N_r = 4$, 16-QAM, $\rho = 3\%$, $f_c = 30$ GHz, SNR = 25 dB.



(a) SER



(b) NMSE

Fig. 7. The SER and NMSE curves under different numbers of receiving antennas: $N_t = 4$, 64-QAM, $\rho = 10\%$, $f_c = 30$ GHz, SNR = 20 dB.

VI. Conclusion

For MIMO-OFDM system, the channel estimation and signal detection are the key issues and thier performance is limited by the PN induced from the imperfect oscillators. In this paper, the JCESD-PN-MAP method is proposed for the JCESD and PN estimation based on the MAP criterion under the assumption of Wiener process for PN in MIMO-OFDM system. The MAP estimation criterion is derived to combine the statistical prior and likelihood information and achieve great performance. In addition, the structures of matrices in MAP criterion are analyzed to simplify the optimization processes and reduce the computation complexity. According to the numerical results, the SER/NMSE performance of our proposed JCESD-PN-MAP method is better than that of the compared methods under different settings of simulations. Furthermore, according to the complexity analysis, the computation complexity of our proposed JCESD-PN-MAP method is significantly lower than that of the compared methods. Therefore, the effectiveness of our proposed JCESD-PN-MAP method is verified in terms of the accuracy and computation complexity.

Appendix

A. Negative Logarithmic Function

By substituting Eq. (14) and (17) into Eq. (16), the negative logarithmic function can be presented as

$$\begin{aligned} G(\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{h},\bar{\boldsymbol{x}}) &= N_r N \ln 2\pi\sigma_w^2 + \frac{||\bar{\boldsymbol{y}} - \boldsymbol{T}(\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{h},\bar{\boldsymbol{x}})||_{\mathrm{F}}^2}{\sigma_w^2} \\ &+ \frac{N_r N}{2}\ln 2\pi\sigma_u^2 + \frac{1}{2}\ln|\boldsymbol{\Sigma}_{N_r}| + \frac{\boldsymbol{\theta}^T\boldsymbol{\Sigma}_{N_r}^{-1}\boldsymbol{\theta}}{2\sigma_u^2} + \frac{N_t N}{2}\ln 2\pi\sigma_v^2 \\ &+ \frac{1}{2}\ln|\boldsymbol{\Sigma}_{N_t}| + \frac{\boldsymbol{\phi}^T\boldsymbol{\Sigma}_{N_t}^{-1}\boldsymbol{\phi}}{2\sigma_v^2} + N_r N_t L\ln 2\pi\sigma_h^2 + \frac{||\boldsymbol{h}||_{\mathrm{F}}^2}{\sigma_h^2} \\ &+ N_t N\ln 2\pi\sigma_x^2 + \frac{||\bar{\boldsymbol{x}}||_{\mathrm{F}}^2}{\sigma_x^2}, \end{aligned} \tag{41}$$

where four equivalent forms of $\boldsymbol{T}(\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{h},\bar{\boldsymbol{x}})$ for the estimations of the PN of receiving antennas, PN of transmitting antennas, channel responses and transmitted signals are given in Eq. (18), (22), (33) and (3), respectively.

B. Extension to Multiple OFDM Symbols

The MIMO-OFDM system with J OFDM symbols is considered. The channel responses and pilot structure are assumed to be invariant for J OFDM symbols. Considering the cyclic prefixes, the correlation of PN among multiple OFDM symbols is small and can be ignored.

According to the similar derivations in subsection (B) - (E), (35) can be rewritten as

$$\begin{bmatrix} \boldsymbol{D}^{\mathrm{H}}_{\boldsymbol{r}^{(0)}}\boldsymbol{D}_{\boldsymbol{r}^{(0)}} + \frac{\sigma_w^2}{\sigma_h^2}\boldsymbol{I}_{N_rN_tL} \\ \vdots \\ \boldsymbol{D}^{\mathrm{H}}_{\boldsymbol{r}^{(J-1)}}\boldsymbol{D}_{\boldsymbol{r}^{(J-1)}} + \frac{\sigma_w^2}{\sigma_h^2}\boldsymbol{I}_{N_rN_tL} \end{bmatrix} \boldsymbol{h} = \begin{bmatrix} \boldsymbol{D}^{\mathrm{H}}_{\boldsymbol{r}^{(0)}}\hat{\boldsymbol{y}}^{(0)} \\ \vdots \\ \boldsymbol{D}^{\mathrm{H}}_{\boldsymbol{r}^{(J-1)}}\hat{\boldsymbol{y}}^{(J-1)} \end{bmatrix} \tag{42}$$

for multiple OFDM symbols and Eq. (37), (21) and (25) or (28) - (30) are still optimized for each OFDM symbol independently, where the added superscripts represent the indexs of OFDM symbols.

REFERENCES

[1] M. H. Mahmud, M. M. Hossain, A. A. Khan, S. Ahmed, M. A. Mahmud, and M. H. Islam, "Performance analysis of OFDM, W-OFDM and F-OFDM under Rayleigh fading channel for 5G wireless communication," in *3rd International Conference on Intelligent Sustainable Systems (ICISS)*, 2020, pp. 1172–1177.

[2] F. Hasegawa, A. Taira, G. Noh, B. Hui, H. Nishimoto, A. Okazaki, A. Okamura, J. Lee, and I. Kim, "High-speed train communications standardization in 3GPP 5G NR," *IEEE Communications Standards Magazine*, vol. 2, no. 1, pp. 44–52, 2018.

[3] M. J. Riaz, A. Sultan, M. Zahid, A. Javed, Y. Amin, and J. Loo, "MIMO antennas for future 5G communications," in *IEEE 23rd International Multitopic Conference (INMIC)*, 2020, pp. 1–4.

[4] B. Manasa and V. P., "A systematic literature review on channel estimation in MIMO-OFDM system: performance analysis and future direction," *Journal of Optical Communications*, 2022. [Online]. Available: https://doi.org/10.1515/joc-2022-0033

[5] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive MIMO detection techniques: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3109–3132, 2019.

[6] J. P. Santacruz, S. Rommel, U. Johannsen, A. Jurado-Navas, and I. T. Monroy, "Analysis and compensation of phase noise in mm-wave OFDM ARoF systems for beyond 5G," *Journal of Lightwave Technology*, vol. 39, no. 6, pp. 1602–1610, 2021.

[7] B. Sen, "On the BER analysis of OFDM receivers against phase noise imperfections," in *IEEE Radio and Wireless Symposium (RWS)*, 2021, pp. 94–97.

[8] X. Chen, H. Wang, W. Fan, Y. Zou, and J. Luo, "Phase noise effect on MIMO-OFDM systems with common and independent oscillators," *Wireless Communications and Mobile Computing*, vol. 2017, no. 7, 2017.

[9] M. H. Jeong, T. J. Lee, M. J. Kim, and Y. C. Ko, "Effect of phase noise in IEEE 802.11ad MIMO-OFDM systems with common/independent local oscillators," in *International Conference on Information and Communication Technology Convergence (ICTC)*, 2015, pp. 1332–1335.

[10] A. Hasan, S. M. A. Motakabber, F. Anwar, M. H. Habaebi, and M. I. Ibrahimy, "A computationally efficient least squares channel estimation method for MIMO-OFDM systems," in *8th International Conference on Computer and Communication Engineering (ICCCE)*, 2021, pp. 331–334.

[11] A. S. Ahmed, M. M. Hamdi, M. S. Abood, A. M. Khaleel, M. Fathy, and S. H. Khaleefah, "Channel estimation using LS and MMSE channel estimation techniques for MIMO-OFDM systems," in *International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2022, pp. 1–6.

[12] S. Wang, J. He, J. Cao, Y. Guan, M. Han, and W. Yu, "A sparse Bayesian learning algorithm for channel estimation in wideband mmwave MIMO-OFDM systems," in *International Conference on Culture-Oriented Science and Technology (CoST)*, 2022, pp. 298–302.

[13] D. C. Araújo, A. L. F. de Almeida, J. P. C. L. Da Costa, and R. T. de Sousa, "Tensor-based channel estimation for massive MIMO-OFDM systems," *IEEE Access*, vol. 7, pp. 42 133–42 147, 2019.

[14] P. Jiang, C. K. Wen, S. Jin, and G. Y. Li, "Dual CNN-based channel estimation for MIMO-OFDM systems," *IEEE Transactions on Communications*, vol. 69, no. 9, pp. 5859–5872, 2021.

[15] R. L. Chung, C. W. Chang, and J. K. Hwang, "Bidirectional decision feedback equalization for mobile MIMO-OFDM systems," in *International Symposium On Information Theory & Its Applications*, 2010, pp. 673–677.

[16] J. H. Ro, S. J. Yu, Y. H. You, S. K. Hong, and H. K. Song, "An adaptive QR-based energy efficient signal detection scheme in MIMO-OFDM systems," *Computer Communications*, vol. 149, pp. 225–231, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140366419303159

[17] S. Mondal, M. R. Laskar, and A. K. Dutta, "ML criterion based signal detection of a MIMO-OFDM system using quantum and semi-quantum assisted modified DHA/BBHT search algorithm," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 2, pp. 1688–1698, 2021.

[18] X. Zhou, J. Zhang, C. K. Wen, J. Zhang, and S. Jin, "Model-driven deep learning-based signal detector for CP-free MIMO-OFDM systems," in *IEEE International Conference on Communications Workshops (ICC Workshops)*, 2021, pp. 1–6.

[19] Juntti, Markku, Suikkanen, and Essi, "Study of adaptive detection and channel estimation for MIMO-OFDM systems," *Wireless Personal Communications An Internaional Journal*, 2017.

[20] R. N. Yang, W. T. Zhang, and S. T. Lou, "Joint adaptive blind channel estimation and data detection for MIMO-OFDM systems," *Wireless Communications and Mobile Computing*, vol. 2020, pp. 1–9, 2020.

[21] Y. Chen, L. You, A. A. Lu, X. Gao, and X. G. Xia, "Channel estimation and robust detection for IQ imbalanced uplink massive MIMO-OFDM with adjustable phase shift pilots," *IEEE Access*, vol. 9, pp. 35 864–35 878, 2021.

[22] J. Du, M. Han, Y. Chen, L. Jin, and F. Gao, "Tensor-based joint channel estimation and symbol detection for time-varying mmwave massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 69, pp. 6251–6266, 2021.

[23] W. Jiang, M. Yue, X. Yuan, and Y. Zuo, "Massive connectivity over MIMO-OFDM: Joint activity detection and channel estimation with frequency selectivity compensation," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 6920–6934, 2022.

[24] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint channel estimation and data detection in MIMO-OFDM systems: A sparse Bayesian learning approach," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5369–5382, 2015.

[25] A. K. Nair and V. Menon, "Joint channel estimation and symbol detection in MIMO-OFDM systems: A deep learning approach using Bi-LSTM," in *14th International Conference on COMmunication Systems & NETworkS (COMSNETS)*, 2022, pp. 406–411.

[26] A. Mohammadian and C. Tellambura, "RF impairments in wireless transceivers: Phase noise, CFO, and IQ imbalance – a survey," *IEEE Access*, vol. 9, pp. 111 718–111 791, 2021.

[27] J. Ma, Z. Li, Y. Xu, Q. Zhang, and M. Wang, "Projection histogram assisted common phase estimation algorithm in coherent optical OFDM system," in *Opto-Electronics and Communications Conference (OECC) and Photonics Global Conference (PGC)*, 2017, pp. 1–3.

[28] H. Huang, W. G. J. Wang, and J. He, "Phase noise and frequency offset compensation in high frequency MIMO-OFDM system," in *IEEE International Conference on Communications (ICC)*, 2015, pp. 1280–1285.

[29] I. M. Ngebani, I. Zibani, E. Matlotse, K. Tsamaase, and J. M. Chuma, "Joint channel and phase noise estimation in MIMO-OFDM systems," in *IEEE Radio and Antenna Days of the Indian Ocean (RADIO)*, 2016, pp. 1–2.

[30] B. Sokal, P. R. B. Gomes, A. L. F. d. Almeida, and M. Haardt, "Tensor-based receiver for joint channel, data, and phase-noise estimation in MIMO-OFDM systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 803–815, 2021.

[31] P. Neshaastegaran and M. Jian, "A time-domain phase noise mitigation algorithm for OFDM systems in the wireless backhaul links," in *IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, 2020, pp. 1–6.

[32] 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 Transactions on Signal Processing*, vol. 60, no. 9, pp. 4790–4807, 2012.

[33] X. Cheng, K. Xu, and S. Li, "Compensation of phase noise in uplink massive MIMO OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 18, no. 3, pp. 1764–1778, 2019.
[34] Y. Qiao, A. Hu, and X. Chen, "Pattern search and compressed sensing based phase noise and channel estimation for mmwave massive MIMO-OFDM systems," *Signal Processing*, vol. 205, p. 108870, 2023.
[35] K. Shrivastav, R. Yadav, and K. Jain, "Joint MAP channel estimation and data detection for OFDM in presence of phase noise from free running and phase locked loop oscillator," *Digital Communications and Networks*, vol. 7, no. 1, pp. 55–61, 2021.
[36] T. J. Lee and Y. C. Ko, "Channel estimation and data detection in the presence of phase noise in MIMO-OFDM systems with independent oscillators," *IEEE Access*, pp. 9647–9662, 2017.
[37] T. J. Lee, M. H. Jeong, and Y. C. Ko, "Phase noise mitigation in MIMO-OFDM systems with independent oscillators," in *IEEE Conference on Standards for Communications and Networking (CSCN)*, 2016, pp. 1–6.
[38] D. D. Lin and T. J. Lim, "The variational inference approach to joint data detection and phase noise estimation in OFDM," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1862–1874, 2007.
[39] H. Du, J. Xue, and Q. Duan, "Phase noise estimation in time domain for OFDM system," in *14th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2022, pp. 961–966.
[40] G. Strang, *Introduction to Linear Algebra*. Wellesley, MA, USA: Wellesley-Cambridge Press, 2003.
[41] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*. Hanover, MA, USA: Now Publishers Inc, 2011.
[42] X. Huang, W. Fu, P. Chen, and H. Huang, "Physical model of phase noise in feedback oscillator," in *Joint European Frequency and Time Forum & International Frequency Control Symposium (EFTF/IFC)*, 2013, pp. 952–955.
[43] D. A. Howe, C. Champagne, and N. Schlossberger, "A total imputation algorithm that fills gaps in time series measurements for ADEV and phase noise characterizations of power-law noise models," in *Joint Conference of the European Frequency and Time Forum and IEEE International Frequency Control Symposium (EFTF/IFCS)*, 2022, pp. 1–2.