Joint Channel, CFO, and Data Estimation via Bayesian Inference for Multi-User MIMO-OFDM Systems

Kenta Ito, *Graduate Student Member, IEEE*, Takumi Takahashi, *Member, IEEE*, Koji Ishibashi, *Senior Member, IEEE*, Koji Igarashi, *Member, IEEE*, and Shinsuke Ibi, *Senior Member, IEEE*

***Abstract*— In this paper, we propose a novel low-complexity Bayesian receiver design to jointly perform channel, carrier frequency offset (CFO), and data estimation from observations subject to different CFOs among users in multi-user multiple-input multiple-output orthogonal frequency-division multiplexing (MU-MIMO-OFDM) systems. Inter-sub-carrier interference (ICI) due to CFO significantly reduces channel estimation accuracy under frequency-selective fading environments, making reliable communications difficult. To tackle this difficulty, a joint channel and CFO estimation (JCCE) algorithm is designed based on belief propagation (BP). Our method uses a Bernoulli-Gaussian (BG) distribution as the prior distribution of the channel coefficient to capture its delay-domain sparsity, and a Gaussian-mixture (GM) distribution as the prior distribution of the phase shift due to CFO to perform parallel search for the allowable range of CFO defined in the 3GPP standard by the number of mixture components. The proposed algorithm can further improve the accuracy of channel, CFO, and data estimation by treating the tentatively detected data symbols as extra pilots. The efficacy of the proposed method is confirmed by numerical studies, which show that the proposed method not only significantly outperforms the state-of-the-art (SotA) methods with much lower computational cost but also approaches the performance of an idealized Genie-aided scheme.**

***Index Terms*— MU-MIMO-OFDM systems, carrier frequency offset, Bayesian inference, frequency-selective fading.**

Received 18 March 2024; revised 29 July 2024; accepted 4 December 2024. Date of publication 17 December 2024; date of current version 12 March 2025. This work was supported in part by Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP23K13335 and Grant JP23K22754; in part by the Japan science and technology agency (JST), CRONOS, Japan, under Grant JPMJCS24N1; and in part by the ministry of internal affairs and communications (MIC)/FORWARD under Grant JPMI240710001. The associate editor coordinating the review of this article and approving it for publication was V. Raghavan. *(Corresponding author: Takumi Takahashi.)*

Kenta Ito and Takumi Takahashi are with the Graduate School of Engineering, Osaka University, Suita 565-0871, Japan (e-mail: k-ito@wcs.comm.eng.osaka-u.ac.jp; takahashi@comm.eng.osaka-u.ac.jp).

Koji Ishibashi is with the Advanced Wireless and Communication Research Center (AWCC), The University of Electro-Communications, Chofu, Tokyo 182-8585, Japan (e-mail: koji@ieee.org).

Koji Igarashi is with the Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan (e-mail: iga.koji.es@osaka-u.ac.jp).

Shinsuke Ibi is with the Faculty of Science and Engineering, Doshisha University, Kyotanabe 610-0394, Japan (e-mail: sibi@mail.doshisha.ac.jp).

Digital Object Identifier 10.1109/TWC.2024.3514210

I. Introduction

LARGE multi-user multiple-input multiple-output orthogonal frequency-division multiplexing (MU-MIMO-OFDM) is one of the essential technologies in fifth generation (5G)-advanced and sixth generation (6G) networks, which serves a massive amount of wireless links simultaneously, enabling to accommodate a large number of uplink user equipment (UE) devices [1], [2], [3], [4]. In order to maximize its potential system performance, receiver design that takes full advantage of the spatial degrees of freedom (DoF) provided has been considered, *i.e.*, solving inherent problems such as channel estimation (CE), multi-user detection (MUD), and among others.

Most of these studies assume that frequency is perfectly synchronized; however, in practice, it is difficult to precisely synchronize the carrier frequencies among UEs equipped with individual local oscillators, and the resulting carrier frequency offsets (CFOs) among devices degrade system performance [5], [6]. In OFDM transmissions, CFOs cause severe inter-sub-carrier interference (ICI), making the frequency-domain equalization using orthogonality between subcarriers infeasible [7], [8]. Therefore, in order to realize a highly accurate joint channel and CFO estimation (JCCE) algorithm in practical MU-MIMO-OFDM systems, it is necessary to estimate CFO and channel impulse response (CIR) in the time domain.

The classical CE method in the presence of CFO is a two-stage scheme in which CFO is estimated with the assistance of pilot symbols/tones followed by CE, for which several *semi-blind* CFO estimation methods without channel state information (CSI) have been proposed [9], [10], [11], [12], [13], [14], [15]. In particular, the covariance-based approaches using maximum likelihood (ML) and maximum *a-posteriori* (MAP) criteria allow for highly accurate CFO estimation, and the asymptotic performance analysis based on Cramér–Rao bound (CRLB) has been investigated; however, most of these approaches have been discussed for single-user scenarios with no inter-user interference. Since the estimation accuracy depends on pilot structure, *i.e.*, its length and orthogonality, a significant increase in pilot overhead is inevitable when

semi-blind CFO estimation is directly applied to multi-user scenarios [14], [15].[1]

To alleviate the pilot overhead problem in multi-user scenarios, several JCCE algorithms have been studied that allow CFO compensation while suppressing inter-user interference using the estimated channel [19], [20], [21], [22], [23], [24], [25], [26], [27]. The most common algorithm of this approach is space-alternating generalized expectation-maximization (SAGE) [20], which efficiently performs ML estimation with computational complexity of linear orders of magnitude with respect to the system size. Specifically, a low-complexity JCCE algorithm can be derived by approximating the joint maximization based on the ML criterion for multiple CFOs to individual maximization in the signal space after interference cancellation by previous estimates and linearizing the phase variation due to CFOs by Taylor expansion. However, the SAGE algorithm based on ML criterion is a method that does not take into account estimation errors, *i.e.*, it updates the estimated solution based only on the given observations and tentative estimates; hence, when the CE accuracy is low in the initial iteration step, the iterative convergence behavior becomes unstable due to error propagation, resulting in severe performance degradation [12], [21], [28]. In other words, it becomes difficult for the SAGE algorithm to perform accurate JCCE when using short non-orthogonal pilots to reduce pilot overhead, and/or when large phase fluctuations that break the orthogonality of the pilots are expected.

The JCCE algorithms on the basis of belief propagation (BP) solve this issue [26], [27]. In the BP-based schemes, the estimation reliability can be gradually improved by iteratively exchanging the *beliefs* (likelihood information reflecting estimation reliability) on the factor graph (FG), which consists of the factor nodes (FNs) and variable nodes (VNs) corresponding to observations and unknown parameters, respectively. In addition, designing the algorithm by integrating a prior distribution that can reflect the statistical properties of the unknown parameters allows iterative estimation based on the MAP criterion, and it has been numerically shown in [26] and [27] that the Bayesian JCCE algorithms achieve more accurate estimation than the state-of-the-art (SotA) alternatives. In these algorithms, various innovations are made to treat CFOs, which are observed non-linearly as phase shifts at the receiver, within the framework of Bayesian linear inference.

For instance, in [27], the CFO estimation problem is converted into a tractable discrete search problem by setting multiple CFO candidates within the limited (allowable) range of CFO defined by the 3rd generation partnership project (3GPP) standard. More precisely, by preparing two-dimensional (2D) candidate grid with dimensions corresponding to CFOs and timing offsets, and designing the JCCE algorithm based on the received signal expanded according to the number of grid points, the time and frequency offsets can be compensated simultaneously. The resultant structured generalized approximate message passing (S-GAMP), which consists of GAMP-based CE [29] and 2D grid search via primitive BP based on a graphical model, provides highly accurate estimation with a limited pilot overhead. However, its estimation accuracy depends on the resolution of the candidate grid, and the computational complexity increases as the signal dimension increases along with the number of CFO candidates, making it not practical when estimating severe phase variations.

Another approach without such grid-based signal space expansion has also been considered. In [30], noting that the estimation of the amount of phase shift caused by CFO can be seen as a sparse signal reconstruction in the angular domain, a JCCE algorithm based on parametric bilinear inference using the Bernoulli-Gaussian (BG) distribution as a prior distribution was proposed. However, this method highly relies on the structure of single-user MIMO systems and cannot be applied directly to MU-MIMO systems since multiple independent CFOs corresponding to UEs must be estimated in MU-MIMO, leading to the completely different problem structure. In addition, sufficient angular resolution must be obtained through long-term observation of phase shifts to achieve high estimation accuracy, so the performance degrades remarkably with shorter pilot length to limit the overhead.

Based on the above previous works in this area, the major contribution of this study is that we design a novel algorithm to achieve low-complexity and high-accuracy large-scale MU-MIMO-OFDM signal demodulation with low pilot overhead in a system without perfect frequency synchronization among the base station (BS) receiver and the UEs, *i.e.*, in the presence of CFOs. In the most of previous studies on JCCE in MU-MIMO-OFDM systems, frequency-flat fading is assumed [27], and even when frequency-selective fading is assumed, the use of pilots that are longer than the product of the number of UEs and the number of delay taps is assumed [31], [32]. When using short non-orthogonal pilots, these conventional JCCE algorithms cannot perform accurate estimations. In order to reduce pilot overhead in the presence of CFOs, it is necessary to develop a JCCE algorithm that overcomes the instability of the convergence characteristics caused by the pilot non-orthogonality and achieves stable convergence behavior even under severe *underdetermined* conditions. To achieve this, it is necessary to design appropriate prior distributions that reflect the statistical properties of the phase shifts due to CFOs and channel coefficients, respectively, while making the most of the available prior knowledge, and to design an algorithm that adjusts the model (distribution) parameters.

To this end, this paper proposes a novel Bayesian JCCE algorithm that employs a BG distribution as a prior distribution of channel coefficients and a Gaussian-mixture (GM) distribution as a prior distribution of phase shifts due to CFOs. There have been several studies that use a BG prior distribution to capture the sparsity of the CIR in the delay domain, but to the best of our knowledge, this is the first paper to use the GM distribution as the prior distribution of the phase shift. By using the mean value of each Gaussian distribution comprising the GM distribution as a candidate value of CFO, it is no longer necessary to introduce the extended equivalent

[1]Blind CFO estimation schemes that estimate CFO directly from only the knowledge of the received signal without even using pilot symbols/tones have also been considered [16], [17], [18]; however, they cannot be directly applied to MU-MIMO-OFDM systems, where different CFOs of every UEs must be estimated separately.

signal model as in [27]; hence, the proposed method can search for the CFOs of UEs in parallel without increasing the computational complexity. The candidate CFO values are initialized to divide the given range of CFO occurrences into equally spaced segments and are then updated as needed within the iterative process, thus avoiding performance degradation due to insufficient angular resolution as in [30]. More specifically, the CFO estimation mechanism is incorporated into the Bayesian receiver by updating the candidate CFOs, *i.e.*, the distribution parameters of GM distribution, using the expectation-maximization (EM) algorithm that operates based on the feedback obtained at each iteration step. Only a difference from the nearest CFO candidate needs to be estimated when updating parameters, thus suppressing polynomial approximation errors due to Taylor expansion and enabling highly accurate CFO estimation. In addition, to compensate for the lack of sufficient pilot length required for time-domain CIR estimation, the proposed JCCE algorithm is extended to a joint channel, CFO, and data estimation (JCCDE) algorithm, in which the estimated data symbols are used as equivalent soft pilot symbols to improve estimation accuracy.

Our contributions, summarized below, are fundamentally algorithmic, such that our results are presented in terms of the normalized mean square errors (NMSEs) and root mean squared errors (RMSEs) of estimated quantities and bit error rates (BERs) computed from actual constellation points:

- A novel JCCE algorithm is presented, wherein time-domain CIR and CFO estimation are iterated alternately based on the time-domain received signal. The overall message passing rule is constructed using the bilinear Gaussian belief propagation (BiGaBP) framework [33], [34], [35], [36], [37], featuring that the phase shift due to CFO is searched in parallel with low-complexity and high-accuracy via updating the distribution parameters of the GM distribution using the EM algorithm and its variant. The proposed CFO estimation mechanism enables robust estimation by adaptively switching between two strategies: candidate CFO search and candidate CFO update, depending on the reliability of the phase shift estimated by BiGaBP. In addition, the time-domain CE under frequency-selective fading environments is formulated as multiple measurement vector (MMV) inference [38], exploiting sparsity in the delay domain, and the time-domain CIR estimation is achieved by designing a message passing rule using a vector-wise BG distribution [36] as a prior distribution. The proposed JCCE algorithm possesses linear complexity with respect to the dimensions of the receive antennas, number of UEs, and pilot length, respectively.
- The proposed JCCE algorithm is extended to a novel JCCDE algorithm, in which cross-domain estimation is performed by alternately executing time-domain JCCE and frequency-domain data detection. Consequently, it becomes possible to use the estimated data symbols as *effective* soft pilots, which further improves the estimation accuracy by allowing additional computational complexity even when non-orthogonal pilots are used. This is possible because the proposed JCCE algorithm is designed to make appropriate use of the estimated data symbols and their mean square errors (MSEs), and therefore the contribution of this paper from an algorithmic perspective is in the design of the JCCE algorithm.
- To confirm the efficacy of the proposed JCCE and JCCDE algorithms in large MU-MIMO-OFDM systems, the proposed methods are evaluated for various system parameters. The simulation results show that the proposed JCCE algorithm outperforms the SotA alternatives such as SAGE and S-GAMP in both frequency-flat and frequency-selective fading channels. Remarkably, the JCCDE algorithm, which uses the estimated data symbols as soft pilot symbols, approaches the performance of an idealized Genie-aided scheme in which CSI is perfectly known (without CFO) even when using short non-orthogonal pilots, which verifies the correctness of the proposed algorithm. In addition, it is shown to be robust against increases in the amount of phase shift due to an increase in CFOs, and the effectiveness of the CFO estimation mechanism based on the GM prior distribution as described above is demonstrated.

To the best of our knowledge, this paper is the first one to tackle the challenging problem of designing a low-complexity JCCE algorithm for MU-MIMO-OFDM systems under frequency-selective fading channels, and to demonstrate numerically that it is possible to achieve highly accurate JCCE (or JCCDE) using short non-orthogonal pilots.

Notation: A set of complex numbers are denoted by $\mathbb{C}$. Vectors and matrices are denoted in lower- and upper-case bold-face fonts, respectively. The conjugate, transpose, and conjugate transpose operators are denoted by $(\cdot)^*$, $(\cdot)^{\mathrm{T}}$, and $(\cdot)^{\mathrm{H}}$, respectively. Random variables and their outcomes are denoted in sans serif and italic fonts, respectively, as in a and a, such that the conditional probability density function (PDF) and the conditional expectation of the outcome a of a, given the occurrence b of b and a set $\mathcal{A}$ are respectively denoted by $p_{\mathsf{a}|\mathsf{b}}(a|b;\mathcal{A})$ and $\mathbb{E}_{\mathsf{a}}\{a|b;\mathcal{A}\}$. The complex Gaussian distribution with mean a and variance b that a variable x follows is denoted by $\mathcal{CN}(x;a,b)$, and in formula the probability density at $x=\dot{x}$ is denoted by $\mathcal{CN}(\dot{x};a,b)$. The $a\times a$ square identity matrix is denoted by $\boldsymbol{I}_a$. The (i,j)-th element of a matrix $\boldsymbol{A}$ is denoted by $[\boldsymbol{A}]_{i,j}$. The diagonal matrix constructed by placing the elements of a vector $\boldsymbol{a}$ on its main diagonal is denoted by $\mathrm{diag}[\boldsymbol{a}]$. The vector generated by cyclically shifting a vector $\boldsymbol{a}$ to the down by p is denoted by $(\boldsymbol{a})_p$. The Frobenius norm is denoted by $|\cdot|_{\mathrm{F}}$. The Dirac delta function is denoted by $\delta(\cdot)$. Finally, we use the simplified notation $\int_a f(a) \triangleq \int_{-\infty}^{\infty} f(a)da$ and $\sum_{i\neq j}^{I} a_i \triangleq \sum_{i=1}^{I} a_i - a_j$, for brevity.

II. System Model

Consider a large MU-MIMO-OFDM system composed of a BS having N receive (RX) antennas and serving U ($\leq N$) UE devices. Let $m \in \mathcal{M}$ denote a transmit (TX) antenna index with $\mathcal{M} \triangleq \{1,2,\cdots,M\}$ denoting the set of TX antenna indices. The set of indices of the TX antennas mounted on the u-th UE is denoted by $\mathcal{M}_u$, where $\mathcal{M} = \bigcup_{u=1}^{U} \mathcal{M}_u$. The TX symbol matrix in the time–frequency domain from the

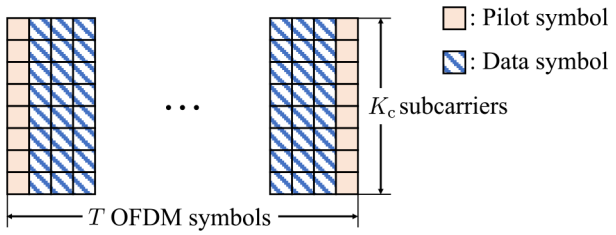


Fig. 1. Schematic of the time–frequency resource grid, where the red and blue squares represent the assigned pilot and data symbols, respectively.

m-th TX antenna is denoted by

$$\boldsymbol{X}_m = [\boldsymbol{x}_{m,1}, \ldots, \boldsymbol{x}_{m,t}, \ldots, \boldsymbol{x}_{m,T}] \in \mathbb{C}^{K_c \times T}, \tag{1a}$$

$$\boldsymbol{x}_{m,t} = [x_{m,1t}, \ldots, x_{m,k_c t}, \ldots, x_{m,K_c t}]^{\mathrm{T}} \in \mathbb{C}^{K_c \times 1}, \tag{1b}$$

where K_c and T are the numbers of subcarriers and OFDM symbols, respectively. As shown in Fig. 1, each symbol represents a (known) pilot symbol or data symbol. The data symbol represents one out of J constellation points $\mathcal{X} = \{\chi_1, \cdots, \chi_j, \cdots, \chi_J\}$, where the average power density of $\mathcal{X}$ is denoted by E_s. Denoting a K_c-point discrete Fourier transform (DFT) matrix by $\boldsymbol{F}$, the time-domain OFDM symbol can be expressed as

$$\boldsymbol{s}_{m,t} = \boldsymbol{F}^{\mathrm{H}} \boldsymbol{x}_{m,t} \in \mathbb{C}^{K_c \times 1}. \tag{2}$$

A cyclic prefix (CP) is then inserted into the time-domain OFDM symbols and the resultant signal is transmitted from each UE via individual local oscillators.

At the receiver, the RX signal observed at each RX antenna is converted to a baseband signal via a single local oscillator, and then the CP is removed. Assuming frequency-selective fading channels, the time-domain RX matrix for the t-th OFDM symbol $\boldsymbol{Y}_t \in \mathbb{C}^{N \times K_c}$ can be expressed as[2]

$$\boldsymbol{Y}_t = \sum_{u=1}^{U} \sum_{m \in \mathcal{M}_u} \sum_{p=1}^{P} \boldsymbol{h}_{p,m} \left[(\boldsymbol{s}_{m,t})_{p-1} \right]^{\mathrm{T}} \mathrm{diag}\left[\boldsymbol{d}_{u,t}\right] + \boldsymbol{W}_t, \tag{3}$$

where $p \in \{1, 2, \cdots, P\}$ is a path delay index. $\boldsymbol{h}_{p,m} \triangleq [h_{p,1m}, \ldots, h_{p,nm}, \ldots, h_{p,Nm}]^{\mathrm{T}} \in \mathbb{C}^{N \times 1}$ is a CIR vector for the m-th TX antenna at the p-th delay tap, each element of which obeys $\mathcal{CN}(h_{p,nm}; 0, \phi^h_{p,u})$ when p corresponds to a delay tap where a path exists and, zero otherwise, *i.e.*, $\phi^h_{p,u} = 0$. The total path gain for each UE, *i.e.*, $\phi^h_u = \sum_{p=1}^{P} \phi^h_{p,u}, \forall u$, is assumed to be known based on the long-term observations at the BS side. $\boldsymbol{W}_t \triangleq [\boldsymbol{w}_{t,1}, \ldots, \boldsymbol{w}_{t,k_c} \ldots, \boldsymbol{w}_{t,K_c}] \in \mathbb{C}^{N \times K_c}$ is an additive white Gaussian noise (AWGN) matrix, where $\boldsymbol{w}_{t,k_c}$ obeys $\mathcal{CN}(\boldsymbol{w}_{t,k_c}; \boldsymbol{0}, N_0 \boldsymbol{I}_N)$, with the noise spectral density N_0, *i.e.*, $\mathbb{E}_{\boldsymbol{w}_{t,k_c}}\{\boldsymbol{w}_{t,k_c}(\boldsymbol{w}_{t,k_c})^{\mathrm{H}}\} = N_0 \boldsymbol{I}_N$. $\boldsymbol{d}_{u,t} \in \mathbb{C}^{K_c \times 1}$ is a phase shift caused by the normalized CFO, *i.e.*,

$$\boldsymbol{d}_{u,t} \triangleq [\exp[\mathrm{j}\alpha_{t1}\varepsilon_u], \ldots, \exp[\mathrm{j}\alpha_{tk_c}\varepsilon_u], \ldots, \exp[\mathrm{j}\alpha_{tK_c}\varepsilon_u]]^{\mathrm{T}}, \tag{4}$$

[2]As various methods for suppressing phase noise (PN) have been proposed for both software and hardware [6], this paper does not include PN in the signal model, assuming that it will be used in conjunction with these.

with

$$\alpha_{tk_c} = \frac{2\pi}{K_c} \left\{ (t-1)(C + K_c) + C + (k_c - 1) \right\}, \tag{5}$$

where C denotes the length of the CP. The 3GPP standard requires that the synchronization error of the carrier frequency be kept within ± 0.1 [ppm] [39]; thus, defining the CFO when the synchronization error of the carrier frequency matches 0.1 [ppm] as Δf_c and when a phase rotation of 2π per sampling interval T_s occurs as $f_{\max}$, respectively, the normalized CFO is modeled as a random variable obeying a uniform distribution within $[-\varepsilon_{\max}, \varepsilon_{\max}]$ for the maximum normalized CFO $\varepsilon_{\max} \triangleq \Delta f_c / f_{\max}$ [27].

From (3), the spatio-temporal RX matrices $\boldsymbol{Y} \triangleq [\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T] \in \mathbb{C}^{N \times K}$ can be expressed as

$$\boldsymbol{Y} = \sum_{u=1}^{U} \sum_{m \in \mathcal{M}_u} \sum_{p=1}^{P} \boldsymbol{h}_{p,m} \left[\boldsymbol{s}_{p,m}\right]^{\mathrm{T}} \mathrm{diag}\left[\boldsymbol{d}_u\right] + \boldsymbol{W}, \tag{6}$$

with

$$\boldsymbol{s}_{p,m} \triangleq [\boldsymbol{s}^{\mathrm{T}}_{p,m,1}, \ldots, \boldsymbol{s}^{\mathrm{T}}_{p,m,t}, \ldots, \boldsymbol{s}^{\mathrm{T}}_{p,m,T}]^{\mathrm{T}} \in \mathbb{C}^{K \times 1}, \tag{7a}$$

$$\boldsymbol{W} \triangleq [\boldsymbol{W}_1, \ldots, \boldsymbol{W}_t, \ldots, \boldsymbol{W}_T] \in \mathbb{C}^{N \times K}, \tag{7b}$$

$$\boldsymbol{d}_u \triangleq \left[\boldsymbol{d}^{\mathrm{T}}_{u,1}, \ldots, \boldsymbol{d}^{\mathrm{T}}_{u,t}, \ldots, \boldsymbol{d}^{\mathrm{T}}_{u,T}\right]^{\mathrm{T}} \in \mathbb{C}^{K \times 1}, \tag{7c}$$

where $\boldsymbol{s}_{p,m,t} \triangleq (\boldsymbol{s}_{m,t})_{p-1}$ and $K \triangleq K_c T$. The (n,k) element of the RX matrix $\boldsymbol{Y}$ can be expressed as

$$y_{nk} = \sum_{u=1}^{U} \sum_{m \in \mathcal{M}_u} \sum_{p=1}^{P} h_{p,nm} s_{p,mk} d_{uk} + w_{nk}, \tag{8}$$

where $s_{p,mk} \triangleq [\boldsymbol{s}_{p,m}]_{k,1}$, $d_{uk} \triangleq [\boldsymbol{d}_u]_{k,1}$, and $w_{nk} \triangleq [\boldsymbol{W}]_{n,k}$, with the index $k = ((t-1) \cdot K_c + k_c) \in \mathcal{K} \triangleq \{1, \cdots, K\}$. The phase shift is defined as $d_{uk} \triangleq \exp[\mathrm{j}\alpha_k \varepsilon_u]$, $(\alpha_k = \alpha_{tk_c})$.

For later convenience, let $\mathcal{T}_{\mathrm{p}}$ denote the set of time indices to which pilot symbols are assigned and $\mathcal{T}_{\mathrm{d}}$ denote the set of time indices to which data symbols are assigned, and further define the corresponding symbol index sets by $\mathcal{K}_{\mathrm{p}}$ and $\mathcal{K}_{\mathrm{d}}$, respectively. Denoting the set of symbol indices at the t-th time index by $\mathcal{K}_t = \{(t-1) \cdot K_c + 1, \cdots, t \cdot K_c\}$, then $\mathcal{K}_{\mathrm{p}} = \bigcup_{t \in \mathcal{T}_{\mathrm{p}}} \mathcal{K}_t$ and $\mathcal{K}_{\mathrm{d}} = \bigcup_{t \in \mathcal{T}_{\mathrm{d}}} \mathcal{K}_t$, where we define $K_{\mathrm{p}} \triangleq |\mathcal{K}_{\mathrm{p}}|$ and $K_{\mathrm{d}} \triangleq |\mathcal{K}_{\mathrm{d}}|$, *i.e.*, $K = K_{\mathrm{p}} + K_{\mathrm{d}}$.

Introducing an index $q = ((m-1) \cdot P + p) \in \{1, \cdots, Q\}$, $(Q \triangleq MP)$ consisting of the path index p and TX antenna index m, (8) can be reformulated as a parametric bilinear inference (PBI) [40], [41], [42] that includes two linear inference problems (LIPs) with respect to the channel coefficient $h_{nq} (= h_{p,nm})$ and the phase shift d_{uk}, respectively, as

$$y_{nk} = \sum_{q=1}^{Q} h_{nq} \rho_{qk} + w_{nk} = \sum_{u=1}^{U} d_{uk} \sigma_{n,uk} + w_{nk}, \tag{9}$$

with

$$\rho_{qk} \triangleq s_{qk} d_{qk}, \quad d_{qk} = d_{uk}, \forall (q \in \mathcal{Q}_m, m \in \mathcal{M}_u), \tag{10a}$$

$$\sigma_{n,uk} \triangleq \sum_{m \in \mathcal{M}_u} \sum_{q \in \mathcal{Q}_m} h_{nq} s_{qk}, \quad s_{qk} = s_{p,mk}, \tag{10b}$$

where $\mathcal{Q}_m \triangleq \{q \mid (m-1) \cdot P + 1, \cdots, m \cdot P\}$.

Algorithm 1 - Proposed Bayesian Joint Channel, CFO, and Data Estimation Algorithm

Input: $\boldsymbol{Y}$, $\{x_{mk}, \forall(m, k \in \mathcal{K}_\mathrm{p})\}$, $\hat{\mathcal{P}}_q^{h,(1)}$, $\hat{\mathcal{P}}_{l,u}^{d,(1)}$

/* Initialization for Pilot Part */

$\forall(k \in \mathcal{K}_\mathrm{p}, n, q, \tau) : \hat{s}_{n,qk}^{(\tau)} = s_{qk},\ \psi_{n,qk}^{s,(\tau)} = 0$

$\forall(k \in \mathcal{K}_\mathrm{p}, n, q) : \hat{h}_{k,nq}^{(1)} = 0,\ \psi_{k,nq}^{h,(1)} = \phi_u^h / P$

$\forall(k \in \mathcal{K}_\mathrm{p}, n, u) : \hat{d}_{n,uk}^{(1)} = 1,\ \psi_{n,uk}^{d,(1)} = \alpha_k^2 \varepsilon_{\max}^2 / 3$

/* Phase I: Channel and CFO Estimation */

for $\tau = 1$ to $\tau_{1,\max}$ **do**

$\forall k \in \mathcal{K}_\mathrm{p}$: Executing Algorithm 2

end for

/* Initialization for Data Part */

$\forall(k \in \mathcal{K}_\mathrm{d}, n, q) : \hat{s}_{n,qk}^{(1)} = 0,\ \psi_{n,qk}^{s,(1)} = E_\mathrm{s}$

$\forall(k \in \mathcal{K}_\mathrm{d}, n, q) : \hat{h}_{k,nq}^{(1)} = \check{h}_{nq},\ \psi_{k,nq}^{h,(1)} = \psi_{nq}^{h}$

Compute initial replicas and their MSEs for data part, $\hat{d}_{n,uk}^{(1)}$ and $\psi_{n,uk}^{d,(1)}$, as described in Subsection III-D.

/* Phase II: Channel, CFO, and Data Estimation */

for $\tau = 1$ to $\tau_{2,\max}$ **do**

$\forall k \in \mathcal{K}_\mathrm{d}$: Executing Algorithm 3

$\forall k \in \mathcal{K}$: Executing Algorithm 2

end for

// Termination and then Hard decision

III. Joint Channel, CFO, and Data Estimation

In this section, a novel message passing algorithm via the BiGaBP [33] to jointly estimate channel, CFO, and data based on (9) is described. The BiGaBP framework relies on the scalar Gaussian approximation (SGA) in conformity with central limit theorem (CLT), whose underlying assumptions are much softer than the large-system limit assumption on which the SotA alternatives [41], [42], [43] rely, allowing for flexible algorithm design depending on the given system model [34], [35], [36].

The pseudo-code of the proposed algorithm is given in Algorithms 1–3, and a work flowchart of describing the relationship between the modules is shown in Fig. 2. Algorithm 1 is the overall algorithm of the proposed method, where Algorithms 2 and 3 show the JCCE mechanism and the data detection mechanism, respectively. As can be seen in Algorithm 1, the proposed algorithm has two phases. In Phase I, the channel coefficients and CFOs are jointly estimated in the time domain based on the BiGaBP approach [33], using the knowledge of the RX signals $y_{nk}, \forall(n, k \in \mathcal{K}_\mathrm{p})$ and the pilot symbols $s_{qk}, \forall(q, k \in \mathcal{K}_\mathrm{p})$. In Phase II, estimates of data symbols are computed via frequency-domain equalization based on the Gaussian belief propagation (GaBP) approach [44], [45], using the knowledge of $y_{nk}, \forall(n, k \in \mathcal{K}_\mathrm{d})$ and the estimates of channels and CFOs obtained in Phase I. Taking advantage of the tentative estimates of data symbols obtained from Algorithm 3 as the equivalent *soft* pilot symbols in the JCCE step, the estimation accuracy can be further improved via Algorithm 2. As shown in the work flowchart of Fig. 2, soft replicas (*i.e.*, tentative estimates) of h_{nq}, d_{uk}, and x_{mk} and their MSEs are iteratively exchanged between the three modules.

Algorithm 2 - JCCE Mechanism

Input: $\boldsymbol{Y}$, $\{\hat{h}_{k,nq}^{(\tau)}, \psi_{k,nq}^{h,(\tau)}\}$, $\{\hat{d}_{n,uk}^{(\tau)}, \psi_{n,uk}^{d,(\tau)}\}$, $\hat{\mathcal{P}}_q^{h,(\tau)}$, $\hat{\mathcal{P}}_{l,u}^{d,(\tau)}$, $\{\hat{s}_{n,qk}^{(\tau+1)}, \psi_{n,qk}^{s,(\tau+1)}\}$

Output: $\{\hat{h}_{k,nq}^{(\tau+1)}, \psi_{k,nq}^{h,(\tau+1)}\}$, $\{\hat{d}_{n,uk}^{(\tau+1)}, \psi_{n,uk}^{d,(\tau+1)}\}$, $\hat{\mathcal{P}}_q^{h,(\tau+1)}$, $\hat{\mathcal{P}}_{l,u}^{d,(\tau+1)}$, $\{\check{h}_{nq}^{(\tau+1)}, \psi_{nq}^{h,(\tau+1)}\}$

$\forall(q \in \mathcal{Q}_m, m \in \mathcal{M}_u, n) : \hat{d}_{n,qk}^{(\tau)} = \hat{d}_{n,uk}^{(\tau)}, \psi_{n,qk}^{d,(\tau)} = \psi_{n,uk}^{d,(\tau)}$

/* Channel Estimation: $\forall(n, q)$ */

Obtain $\tilde{y}_{q,nk}^{h,(\tau)}$ and $\xi_{q,nk}^{h,(\tau)}$ via (20) and (21).

Obtain $\hat{g}_{k,nq}^{h,(\tau)}$ and $\gamma_{k,nq}^{h,(\tau)}$ via (24).

Obtain $\hat{\mu}_{nq}^{(\tau)}$ and $\psi_{nq}^{\mu,(\tau)}$ via (23).

Obtain $\check{h}_{nq}^{(\tau+1)}$ and $\check{\psi}_{nq}^{h,(\tau+1)}$ via (28).

Obtain $\hat{h}_{k,nq}^{(\tau+1)}$ and $\psi_{k,nq}^{h,(\tau+1)}$ via (30).

$\forall q$: Update $\hat{\lambda}_q^{(\tau+1)}$ and $\hat{\phi}_q^{h,(\tau+1)}$ via (45).

/* CFO Estimation: $\forall(u, n)$ */

Obtain $\tilde{y}_{u,nk}^{d,(\tau)}$ and $\xi_{u,nk}^{d,(\tau)}$ via (31) and (33).

Obtain $\hat{g}_{n,uk}^{d,(\tau)}$ and $\gamma_{n,uk}^{d,(\tau)}$ via (35).

Obtain $\hat{r}_{uk}^{(\tau)}$ and $\psi_{uk}^{r,(\tau)}$ via (34).

Obtain $\check{d}_{uk}^{(\tau)}$ and $\psi_{uk}^{d,(\tau)}$ via (39).

Obtain $\hat{d}'^{(\tau)}_{n,uk}$ and $\psi'^{d,(\tau)}_{n,uk}$ via (40).

Obtain $\hat{d}_{n,uk}^{(\tau+1)}$ and $\psi_{n,uk}^{d,(\tau+1)}$ via (41) and (43).

$\forall(l, u)$: Update $\hat{\varepsilon}_{l,u}^{(\tau+1)}$ via (48).

$\forall(l, u) : \hat{\varepsilon}_{l,u}^{(\tau+1)} = \min(\max(\hat{\varepsilon}_{l,u}^{(\tau+1)}, -\varepsilon_{\max}), \varepsilon_{\max})$

$\forall(l, u)$: Update $\hat{\phi}_{l,uk}^{d,(\tau+1)}$ via (50).

Algorithm 3 - Data Detection Mechanism

Input: $\boldsymbol{Y}_\mathrm{d}$, $\{\hat{s}_{n,qk}^{(\tau)}, \psi_{n,qk}^{s,(\tau)}\}$, $\{\hat{h}_{k,nq}^{(\tau)}, \psi_{k,nq}^{h,(\tau)}\}$, $\{\check{h}_{nq}^{(\tau)}, \psi_{nq}^{h,(\tau)}\}$, $\{\hat{d}_{n,mk}^{(\tau)}, \psi_{n,mk}^{d,(\tau)}\}$

Output: $\{\hat{s}_{n,qk}^{(\tau+1)}, \psi_{n,qk}^{s,(\tau+1)}\}$

/* Data Detection: $\forall(m, n)$ */

Obtain $\tilde{y}_{m,nk}^{x,(\tau)}$ and $\xi_{m,nk}^{x,(\tau)}$ via (59) and (62).

$\gamma_{n,mk}^{x,(\tau)} = \eta \cdot \frac{\left|\check{a}_{k,nm}^{(\tau)}\right|^2}{\xi_{m,nk}^{x,(\tau)}} + (1-\eta) \cdot \gamma_{n,mk}^{x,(\tau-1)}$

$\hat{g}_{n,mk}^{x,(\tau)} = \eta \cdot \frac{\left(\check{a}_{k,nm}^{(\tau)}\right)^* \tilde{y}_{m,nk}^{x,(\tau)}}{\xi_{m,nk}^{x,(\tau)}} + (1-\eta) \cdot \hat{g}_{n,mk}^{x,(\tau-1)}$

$\psi_{n,mk}^{z,(\tau)} = \left(\sum_{i \neq n}^{N} \gamma_{i,mk}^{x,(\tau)}\right)^{-1}$

$\hat{z}_{n,mk}^{(\tau)} = \psi_{n,mk}^{z,(\tau)} \cdot \sum_{i \neq n}^{N} \hat{g}_{i,mk}^{x,(\tau)}$

$\beta(\tau) = b \cdot \frac{\tau}{\tau_{2,\max}}$

$\omega_{n,mk}(\chi_j) = \frac{\beta(\tau)}{c_\mathrm{x}^2} \left(2\Re\left\{\hat{z}_{n,mk}^{(\tau)} \chi_j^*\right\} - |\chi_j|^2\right)$

$\hat{x}_{n,mk}^{(\tau+1)} = \sum_{\chi_j \in \mathcal{X}} \frac{\chi_j \exp[\omega_{n,mk}(\chi_j)]}{\sum_{\chi'_j \in \mathcal{X}} \exp[\omega_{n,mk}(\chi'_j)]}$

$\psi_{n,mk}^{x,(\tau+1)} = \sum_{\chi_j \in \mathcal{X}} \frac{|\chi_j|^2 \exp[\omega_{n,mk}(\chi_j)]}{\sum_{\chi'_j \in \mathcal{X}} \exp[\omega_{n,mk}(\chi'_j)]} - \left|\hat{x}_{n,mk}^{(\tau+1)}\right|^2$

$\forall(n, (q,k) \in \mathcal{I}_{mk}^s) : \hat{s}_{n,qk}^{(\tau+1)} = \sum_{i \in \mathcal{K}_t} f_{\kappa(i),\kappa(k)}^* \hat{x}_{n,mi}^{(\tau+1)}$

$\forall(n, (q,k) \in \mathcal{I}_{mk}^s) : \psi_{n,qk}^{s,(\tau+1)} = \frac{1}{K_\mathrm{c}} \sum_{i \in \mathcal{K}_t} \psi_{n,mi}^{x,(\tau+1)}$

In what follows, we provide detailed descriptions of Algorithms 2 and 3 mentioned above. To that end, let us define the soft replicas of h_{nq} and d_{uk} as $\hat{h}_{k,nq}, \forall k$ and $\hat{d}_{n,uk}, \forall n$, respec-

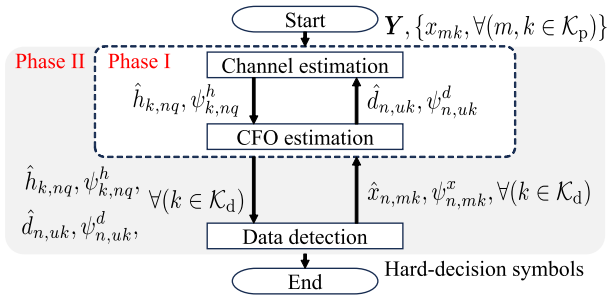


Fig. 2. Work flowchart of the proposed JCCDE receiver.

tively, such that their MSEs can be respectively expressed as

$$\psi^h_{k,nq} \triangleq \mathbb{E}_{\tilde{\mathsf{h}}_{k,nq}} \left\{ \left| \tilde{h}_{k,nq} \right|^2 \right\}, \ \psi^d_{n,uk} \triangleq \mathbb{E}_{\tilde{\mathsf{d}}_{n,uk}} \left\{ \left| \tilde{d}_{n,uk} \right|^2 \right\}, \tag{11}$$

where $\tilde{h}_{k,nq} \triangleq h_{nq} - \hat{h}_{k,nq}$ and $\tilde{d}_{n,uk} \triangleq d_{uk} - \hat{d}_{n,uk}$ denote the estimation errors, respectively. Similarly, the soft replica of the frequency-domain symbol x_{mk} and its MSE are defined as $\hat{x}_{n,mk}, \forall n$ and $\psi^x_{n,mk}, \forall n$, respectively. Notice that the time-domain symbol s_{qk} contains P overlapping s_{mk} due to the cyclic shift operation, we define the set of P 2D indices (q,k) corresponding to s_{mk} as $\mathcal{I}^s_{mk}$ ($|\mathcal{I}^s_{mk}| = P$).

Accordingly, the soft replica of $s_{qk}, \forall (q,k) \in \mathcal{I}^s_{mk}$ and its MSE $\psi^s_{n,qk}, \forall (q,k) \in \mathcal{I}^s_{mk}$ are respectively given by

$$\hat{s}_{n,qk} = \sum_{i \in \mathcal{K}_t} f^*_{\kappa(i),\kappa(k)} \hat{x}_{n,mi}, \ \psi^s_{n,qk} = \frac{1}{K_\mathrm{c}} \sum_{i \in \mathcal{K}_t} \psi^x_{n,mi}, \tag{12}$$

where $\kappa(i) \triangleq (i \mod K_\mathrm{c})$, $f_{i,j} \triangleq [\boldsymbol{F}]_{i,j}$, and $|f_{i,j}|^2 = 1/K_\mathrm{c}$.

To describe the main blocks of Algorithm 1, in lines 1 to 3, the soft replicas and their MSEs are initialized; lines 4 to 6 correspond to Phase I; in lines 7 to 9 to initialization for data detection, and lines 10 to 13 to Phase II. The predetermined number of iterations for Phase I and Phase II are denoted by $\tau_{1,\max}$ and $\tau_{2,\max}$, respectively.

In Subsections III-A through III-C, we derive the JCCE algorithm (Algorithm 2), followed by the data detection algorithm (Algorithm 3) in Subsections III-D and III-E.

A. Design of Prior Distributions in Algorithm 2

First, as a preliminary to derive the message passing algorithm, in this subsection, we design prior distributions and define their distribution parameters for time-domain channel coefficients and phase shifts due to CFOs, respectively.

1) Prior Distribution for Channel Coefficients: In the RX signal from a given UE, the delay tap where the path exists does not change for each RX antenna; hence, from the definition of $\phi^h_{p,u}$ explained after (3), the CIR has vector-wise sparsity in the delay domain. To capture its stochastic properties, $\boldsymbol{h}_q \triangleq [h_{1q}, \ldots, h_{nq}, \ldots, h_{Nq}]^\mathrm{T} \in \mathbb{C}^{N \times 1}$ can be modeled by the following *vector-wise* BG distribution as [35]

$$p_{\boldsymbol{h}_q} \left(\boldsymbol{h}_q; \hat{\mathcal{P}}^{h,(\tau)}_q \right) = \prod_{n=1}^{N} p_{\mathsf{h}_{nq}} \left(h_{nq}; \hat{\mathcal{P}}^{h,(\tau)}_q \right), \tag{13a}$$

with

$$p_{\mathsf{h}_{nq}} \left(h_{nq}; \hat{\mathcal{P}}^{h,(\tau)}_q \right) = \left(1 - \hat{\lambda}^{(\tau)}_q \right) \delta \left(h_{nq} \right) + \hat{\lambda}^{(\tau)}_q \mathcal{CN} \left(h_{nq}; 0, \hat{\phi}^{h,(\tau)}_q \right), \tag{13b}$$

where $\hat{\mathcal{P}}^{h,(\tau)}_q \triangleq \left\{ \hat{\lambda}^{(\tau)}_q, \hat{\phi}^{h,(\tau)}_q \right\}$ denotes a set of the distribution parameters (*i.e.* sparsity rate and variance) at the τ-th iteration step, and is assigned one for each channel vector $\boldsymbol{h}_q$. The sparsity rate and variance are initialized as $\hat{\lambda}^{(1)}_q = 1/P$ and $\hat{\phi}^{h,(1)}_q = \phi^h_u, \forall (q \in \mathcal{Q}_m, m \in \mathcal{M}_u)$, respectively. The vector-wise BG prior given in (13) is designed to capture the statistical properties of the CIR vector in (3) with the necessary and sufficient distribution parameters while taking into account the sparsity in the delay domain and the power deviation between paths based on the delay profile.

2) Prior Distribution for Phase Shifts: By exploiting the limited (allowable) range in which CFOs occur as prior information, a prior distribution is designed to efficiently search for the corresponding phase shifts $\boldsymbol{d}_u \triangleq [d_{u1}, \ldots, d_{uk}, \ldots, d_{uK}]^\mathrm{T}$. First, we define $\hat{\boldsymbol{\varepsilon}}^{(\tau)}_u = [\hat{\varepsilon}^{(\tau)}_{1,u}, \ldots, \hat{\varepsilon}^{(\tau)}_{l,u}, \ldots, \hat{\varepsilon}^{(\tau)}_{L,u}]^\mathrm{T} \in \mathbb{R}^{L \times 1}$ as the vector of L candidate CFOs to be used in the τ-th iteration step to search for the CFO with respect to the u-th UE, and $\hat{\theta}^{d,(\tau)}_{l,uk} \triangleq \exp[\mathrm{j}\alpha_k \hat{\varepsilon}^{(\tau)}_{l,u}]$ as the candidate phase shift experienced by the k-th symbol corresponding to $\hat{\varepsilon}^{(\tau)}_{l,u}$. Next, interpreting the candidate phase shift vectors $\hat{\boldsymbol{\theta}}^{d,(\tau)}_{l,u} \triangleq [\hat{\theta}^{d,(\tau)}_{l,u1}, \ldots, \hat{\theta}^{d,(\tau)}_{l,uk}, \ldots, \hat{\theta}^{d,(\tau)}_{l,uK}]^\mathrm{T}, \forall l$, as prior knowledge about $\boldsymbol{d}_u$, the prior distribution of $\boldsymbol{d}_u$ can be expressed using Bayes' theorem as follows:

$$p_{\boldsymbol{d}_u} (\boldsymbol{d}_u) = \sum_{l=1}^{L} p_{\boldsymbol{d}_u | \hat{\boldsymbol{\theta}}_{l,u}} \left(\boldsymbol{d}_u | \hat{\boldsymbol{\theta}}^{d,(\tau)}_{l,u} \right) p_{\hat{\boldsymbol{\theta}}_{l,u}} \left(\hat{\boldsymbol{\theta}}^{d,(\tau)}_{l,u} \right). \tag{14}$$

Finally, assuming that the candidate phase shift vectors are selected with equal probability, *i.e.*, $p_{\hat{\boldsymbol{\theta}}_{l,u}} \left(\hat{\boldsymbol{\theta}}^{d,(\tau)}_{l,u} \right) = 1/L, \forall l$, and approximating that the error between the candidate values and the true phase shift amount is Gaussian distributed, the prior distribution of $\boldsymbol{d}_u$ can be modeled by the following *vector-wise* GM distribution as

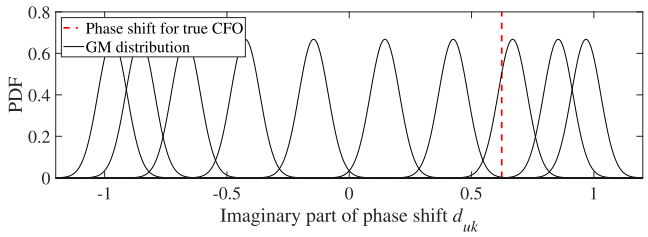
$$p_{\boldsymbol{d}_u} \left(\boldsymbol{d}_u; \hat{\mathcal{P}}^{d,(\tau)}_u \right) = \frac{1}{L} \sum_{l=1}^{L} p_{\boldsymbol{d}_u} \left(\boldsymbol{d}_u; \hat{\mathcal{P}}^{d,(\tau)}_{l,u} \right), \tag{15a}$$

with

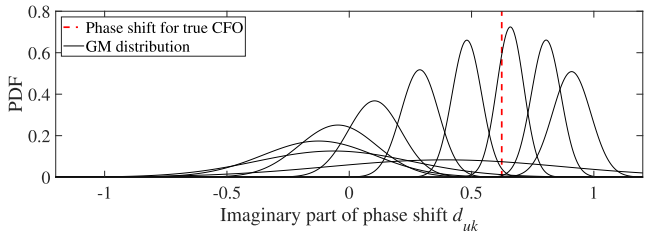
$$p_{\boldsymbol{d}_u} \left(\boldsymbol{d}_u; \hat{\mathcal{P}}^{d,(\tau)}_{l,u} \right) = \prod_{k=1}^{K} \mathcal{CN} \left(d_{uk}; \hat{\theta}^{d,(\tau)}_{l,uk}, \hat{\phi}^{d,(\tau)}_{l,uk} \right), \tag{15b}$$

where $\hat{\mathcal{P}}^{d,(\tau)}_{l,u} \triangleq \left\{ \hat{\theta}^{d,(\tau)}_{l,uk}, \hat{\phi}^{d,(\tau)}_{l,uk} \right\}_{k=1}^{K}$ denotes a set of the distribution parameters (*i.e.*, means and variances) corresponding to the l-th Gaussian components at the τ-th iteration step, and $\hat{\mathcal{P}}^{d,(\tau)}_u \triangleq \left\{ \hat{\mathcal{P}}^{d,(\tau)}_{l,u} \right\}_{l=1}^{L}$.

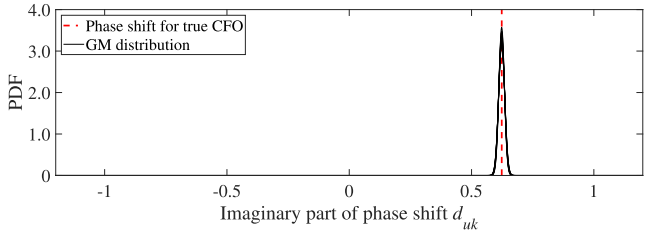
Fig. 3 is an illustration of how the true phase shift is searched (estimated) with this prior distribution. Using the allowable range of CFOs defined in the 3GPP standard [39] as prior information, the candidate CFOs are initialized to divide their existence range equally, *i.e.*, $\hat{\varepsilon}^{(1)}_{l,u} = \{ \varepsilon_\mathrm{max} \cdot (2l - 1 - L)/L \mid l = 1, \ldots, L \}$, and the corresponding GM prior is



(a) $\tau = 1$.



(b) $\tau = 4$.



(c) $\tau = 16$.

Fig. 3. Illustration of how the true phase shift is searched (estimated) using the GM distribution ($L = 10$).

set as in Fig. 3(a) for each phase shift. The means (candidate phase shifts) and their variances of this GM distribution are updated at each iteration by the EM algorithm described later in Subsection III-C. The L component Gaussian distributions are initially used to search the entire range of existence, but gradually move near the true phase shift value over the iterations, as shown in Fig. 3(b). Finally, the probability masses are aggregated into a single component Gaussian distribution, as shown in Fig. 3(c), and the true phase shift is estimated with *pinpoint* accuracy. By bridging discrete value search and continuous value estimation with the GM distribution, it becomes possible to robustly and efficiently search for phase shifts that occur over a wide range. This can be explained by the first-order approximation accuracy of the phase shift using the Taylor expansion of d_{uk} used in the parameter estimation,

$$d_{uk} = \exp\left[\mathrm{j}\alpha_k\left(\hat{\varepsilon}_{l,u}^{(\tau)} + \tilde{\varepsilon}_{l,u}^{(\tau)}\right)\right] \approx \hat{\theta}_{l,uk}^{d,(\tau)}\left(1 + \mathrm{j}\alpha_k\tilde{\varepsilon}_{l,u}^{(\tau)}\right), \tag{16}$$

where owing to the knowledge of the candidate CFOs, the residual CFO $\tilde{\varepsilon}_{l,u}^{(\tau)} \triangleq \varepsilon_u - \hat{\varepsilon}_{l,u}^{(\tau)}$ becomes much smaller than ε_u, and the accuracy of the approximation in (16) is improved because the update width of the phase shift becomes small. This is the primary reason why the proposed JCCE algorithm is robust to severe phase fluctuations. From (16), the variances are initialized as

$$\hat{\phi}_{l,uk}^{d,(1)} \approx \alpha_k^2\mathbb{E}_{\tilde{\epsilon}_{l,u}}\left\{\left(\tilde{\varepsilon}_{l,u}^{(1)}\right)^2\right\} = \alpha_k^2\varepsilon_{\max}^2/(3L^2). \tag{17}$$

If the purpose is to introduce a grid search using the candidate CFOs, a discrete uniform distribution is also a candidate for the prior distribution. However, in this case, the error between the candidate values and the true phase shift cannot be reflected; therefore, it does not exhibit the slow iterative convergence behavior shown in Fig. 3, but rather rapid convergence behavior. Consequently, the estimation performance depends deeply on the CE accuracy in the initial iteration step, and robust estimation, as in the case of the GM prior, is impossible. Only the GM prior distribution, which allows for distribution estimation that takes into account errors, enables adaptive grid-based search, as shown in Fig. 3.

B. JCCE Mechanism in Algorithm 2

Next, we design the JCCE mechanism based on the BiGaBP framework [33] using the distribution parameters defined in Subsection III-A. The complete set of update rules is derived here, but is omitted in detail due to limitation of space. For further details, we refer the reader to, *e.g.*, [33], [46], [47], [48], and [49]. In this subsection, we omit the iteration index $\cdot^{(\tau)}$ for each variable for notational simplicity.

The pseudo-code of the proposed JCCE algorithm described above is given in Algorithm 2. A well-known belief damping is introduced in lines 3 and 9, as described in [50] and summarized as follows. Let a quantity z be calculated by a function f_z. Then the τ-th damped value of z, here denoted $z^{(\tau)}$, is computed as the weighted average of $z^{(\tau-1)}$ and f_z, with weights set by the damping factor $\eta \in [0, 1]$, *i.e.*,

$$z = f_z \quad \xrightarrow{\text{damping}} \quad z^{(\tau)} = \eta f_z + (1-\eta)\, z^{(\tau-1)}. \tag{18}$$

1) FN Process for CE: Focusing on y_{nk} in (9), in the estimation of h_{nq}, the RX symbol after soft interference cancellation (Soft IC) using soft replicas can be written as

$$\begin{aligned}
&\tilde{y}_{q,nk}^{h} \\
&= y_{nk} - \underbrace{\sum_{i\neq q}^{Q}\hat{h}_{k,ni}\hat{\rho}_{n,ik}}_{\text{Inter-UE interference cancellation}} \\
&= h_{nq}\hat{\rho}_{n,qk} + \underbrace{\sum_{i\neq q}^{Q}\left(h_{ni}\rho_{ik} - \hat{h}_{k,ni}\hat{\rho}_{n,ik}\right) + h_{nq}\tilde{\rho}_{n,qk} + w_{nk}}_{\triangleq \nu_{q,nk}^{h}\text{: Residual interference plus noise}},
\end{aligned} \tag{19}$$

where the soft replicas $\hat{\rho}_{n,qk} = \hat{s}_{n,qk}\hat{d}_{n,qk}$ and $\tilde{\rho}_{n,qk} \triangleq \rho_{qk} - \hat{\rho}_{n,qk}$ are generated in VNs at the previous iteration. Assuming that $\nu_{q,nk}^{h}$ can be approximated as a complex Gaussian random variable in conformity to the CLT, the conditional PDF of $\tilde{y}_{q,nk}^{h}$, given h_{nq}, can be expressed as

$$p_{\tilde{\mathrm{y}}_{q,nk}^{h}|\mathrm{h}_{nq}}\left(\tilde{y}_{q,nk}^{h}|h_{nq}\right) \propto \exp\left[-\frac{\left|\tilde{y}_{q,nk}^{h} - h_{nq}\hat{\rho}_{n,qk}\right|^2}{\xi_{q,nk}^{h}}\right], \tag{20}$$

with

$$\begin{aligned}\xi^h_{q,nk} &= \mathbb{E}_{\{\mathsf{h}_{ni},\forall i\neq q\},\{\mathsf{s}_{qk},\forall q\},\{\mathsf{d}_{qk},\forall q\},\mathsf{w}_{nk}}\left\{\left|\nu^h_{q,nk}\right|^2|h_{nq}\right\}\\ &= \sum_{i\neq q}^{Q}\Big\{\psi^h_{k,ni}\left|\hat{s}_{n,ik}\right|^2 \\ &\quad + \left(\left|\hat{h}_{k,ni}\right|^2+\psi^h_{k,ni}\right)\left(\left|\hat{s}_{n,ik}\right|^2\psi^d_{n,ik}+\psi^s_{n,ik}\right)\Big\}\\ &\quad + \hat{\lambda}_q\hat{\phi}^h_q\left(\left|\hat{s}_{n,qk}\right|^2\psi^d_{n,qk}+\psi^s_{n,qk}\right)+N_0. \end{aligned} \tag{21}$$

2) VN Process for CE: Assuming that the effective Gaussian noise in $\tilde{y}^h_{q,nk},\forall k$, is not correlated with each other, the PDF of an *extrinsic* belief for h_{nq} can be obtained from [44]

$$\begin{aligned}p_{\hat{\mu}_{nq}|\mathsf{h}_{nq}}\left(\hat{\mu}_{nq}|h_{nq}\right) &= \prod_{k=1}^{K}p_{\tilde{\mathsf{y}}^h_{q,nk}|\mathsf{h}_{nq}}\left(\tilde{y}^h_{q,nk}|h_{nq}\right)\\ &\propto \exp\left[-\frac{\left|h_{nq}-\hat{\mu}_{nq}\right|^2}{\psi^\mu_{nq}}\right], \end{aligned} \tag{22}$$

with

$$\hat{\mu}_{nq}=\psi^\mu_{nq}\sum_{k=1}^{K}\hat{g}^h_{k,nq},\ \ \psi^\mu_{nq}=\left(\sum_{k=1}^{K}\gamma^h_{k,nq}\right)^{-1}, \tag{23}$$

where

$$\hat{g}^h_{k,nq}=\frac{\hat{\rho}^*_{n,qk}\tilde{y}^h_{q,nk}}{\xi^h_{q,nk}},\ \ \gamma^h_{k,nq}=\frac{\left|\hat{\rho}_{n,qk}\right|^2}{\xi^h_{q,nk}}. \tag{24}$$

Similarly, assuming that the effective Gaussian noise in $\hat{\mu}_{nq},\forall(n,q)$, is not correlated with each other, the conditional PDF of h_{nq}, given $\hat{\mu}_{nq}$ and $\hat{\mathcal{P}}^h_q$ in (13), can be expressed as

$$\begin{aligned}&p_{\mathsf{h}_{nq}|\hat{\mu}_{nq}}\left(h_{nq}|\hat{\mu}_{nq};\hat{\mathcal{P}}^h_q\right)\\ &\quad\triangleq\frac{p_{\hat{\mu}_{nq}|\mathsf{h}_{nq}}\left(\hat{\mu}_{nq}|h_{nq}\right)p_{\mathsf{h}_{nq}}\left(h_{nq};\hat{\mathcal{P}}^h_q\right)}{\int_{h'_{nq}}p_{\hat{\mu}_{nq}|\mathsf{h}_{nq}}\left(\hat{\mu}_{nq}|h'_{nq}\right)p_{\mathsf{h}_{nq}}\left(h'_{nq};\hat{\mathcal{P}}^h_q\right)}. \end{aligned} \tag{25}$$

When h_{nq} obeys the BG distribution, from the Gaussian-PDF multiplication rule, (25) can be rewritten as [46]

$$\begin{aligned}&p_{\mathsf{h}_{nq}|\hat{\mu}_{nq}}\left(h_{nq}|\hat{\mu}_{nq};\hat{\mathcal{P}}^h_q\right)\\ &\quad=(1-\zeta_{nq})\cdot\delta\left(h_{nq}\right)+\zeta_{nq}\cdot\mathcal{CN}\left(h_{nq};\overline{h}_{nq},\overline{\psi}^h_{nq}\right), \end{aligned} \tag{26}$$

with

$$\zeta_{nq}=\left[\frac{\left(1-\hat{\lambda}_q\right)\cdot\mathcal{CN}\left(0;\hat{\mu}_{nq},\psi^\mu_{nq}\right)}{\hat{\lambda}_q\cdot\mathcal{CN}\left(0;\hat{\mu}_{nq},\psi^\mu_{nq}+\hat{\phi}^h_q\right)}+1\right]^{-1}, \tag{27a}$$

$$\overline{h}_{nq}=\frac{\hat{\phi}^h_q\hat{\mu}_{nq}}{\psi^\mu_{nq}+\hat{\phi}^h_q},\ \ \overline{\psi}^h_{nq}=\frac{\hat{\phi}^h_q\psi^\mu_{nq}}{\psi^\mu_{nq}+\hat{\phi}^h_q}. \tag{27b}$$

From (25) and (26), the soft replica $\check{h}_{nq}$ and its MSE $\check{\psi}^h_{nq}$ can be obtained from the conditional expectation as

$$\check{h}_{nq}=\zeta_{nq}\cdot\overline{h}_{nq}, \tag{28a}$$

$$\check{\psi}^h_{nq}=(1-\zeta_{nq})\cdot\zeta_{nq}\cdot\left|\overline{h}_{nq}\right|^2+\zeta_{nq}\cdot\overline{\psi}^h_{nq}. \tag{28b}$$

Finally, following [47], [48], and [49], we approximate the posterior distribution of h_{nq} by a Gaussian distribution with (28) based on *moment matching*, and then compute the PDF of an extrinsic soft replica for h_{nq} as

$$\mathcal{CN}\left(h_{nq};\hat{h}_{k,nq},\psi^h_{k,nq}\right)\propto\frac{\mathcal{CN}\left(h_{nq};\check{h}_{nq},\check{\psi}^h_{nq}\right)}{p_{\tilde{\mathsf{y}}^h_{q,nk}|\mathsf{h}_{nq}}\left(\tilde{y}^h_{q,nk}|h_{nq}\right)}, \tag{29}$$

where we can derive

$$\hat{h}_{k,nq}=\psi^h_{k,nq}\left(\frac{\check{h}_{nq}}{\check{\psi}^h_{nq}}-\hat{g}^h_{k,nq}\right), \tag{30a}$$

$$\frac{1}{\psi^h_{k,nq}}=\frac{1}{\check{\psi}^h_{nq}}-\gamma^h_{k,nq}. \tag{30b}$$

3) FN Process for CFO Estimation: The above discussion can be applied to the estimation of an arbitrary d_{uk}. In a similar manner to (19) and (21), the cancellation process for d_{uk} and its variance can be expressed as

$$\tilde{y}^d_{u,nk}=y_{nk}-\sum_{i\neq u}^{U}\hat{d}_{n,ik}\hat{\sigma}_{n,ik}, \tag{31}$$

$$\begin{aligned}\xi^d_{u,nk} &= \sum_{i\neq u}^{U}\psi^d_{n,ik}\left|\hat{\sigma}_{n,ik}\right|^2\\ &\quad+\sum_{q=1}^{Q}\left\{\left|\hat{h}_{k,nq}\right|^2\psi^s_{n,qk}+\psi^h_{k,nq}\left(\left|\hat{s}_{n,qk}\right|^2+\psi^s_{n,qk}\right)\right\}+N_0, \end{aligned} \tag{32}$$

where $\hat{\sigma}_{n,uk}=\sum_{m\in\mathcal{M}_u}\sum_{q\in\mathcal{Q}_m}\hat{h}_{k,nq}\hat{s}_{n,qk}$.

4) VN Process for CFO Estimation: The PDF of a *posterior* belief for d_{uk} can be obtained from

$$\begin{aligned}p_{\hat{\mathsf{r}}_{uk}|\mathsf{d}_{uk}}\left(\hat{r}_{uk}|d_{uk}\right) &= \prod_{n=1}^{N}p_{\tilde{\mathsf{y}}^d_{u,nk}|\mathsf{d}_{uk}}\left(\tilde{y}^d_{u,nk}|d_{uk}\right)\\ &\propto\exp\left[-\frac{\left|d_{uk}-\hat{r}_{uk}\right|^2}{\psi^r_{uk}}\right], \end{aligned} \tag{33}$$

where

$$\hat{r}_{uk}=\psi^r_{uk}\sum_{n=1}^{N}\hat{g}^d_{n,uk},\ \ \psi^r_{uk}=\left(\sum_{n=1}^{N}\gamma^d_{n,uk}\right)^{-1}, \tag{34}$$

with

$$\hat{g}^d_{n,uk}=\frac{\hat{\sigma}^*_{n,uk}\tilde{y}^d_{u,nk}}{\xi^d_{u,nk}},\ \ \gamma^d_{n,uk}=\frac{\left|\hat{\sigma}_{n,uk}\right|^2}{\xi^d_{u,nk}}. \tag{35}$$

Assuming that the effective Gaussian noise in $\hat{r}_{uk},\forall(u,k)$, is not correlated with each other, the conditional PDF of $\boldsymbol{d}_u$, given $\hat{\boldsymbol{r}}_u\triangleq[\hat{r}_{u1},\ldots,\hat{r}_{uK}]^{\mathrm{T}}$, can be expressed as

$$\begin{aligned}&p_{\boldsymbol{d}_u|\hat{\boldsymbol{r}}_u}\left(\boldsymbol{d}_u|\hat{\boldsymbol{r}}_u;\hat{\mathcal{P}}^d_u\right)\\ &\quad\triangleq\frac{\sum_{l=1}^{L}\prod_{k=1}^{K}p_{\hat{\mathsf{r}}_{uk}|\mathsf{d}_{uk}}\left(\hat{r}_{uk}|d_{uk}\right)p_{\mathsf{d}_{uk}}\left(d_{uk};\hat{\mathcal{P}}^d_{l,u}\right)}{\int_{d'_{uk}}\sum_{l=1}^{L}\prod_{k=1}^{K}p_{\hat{\mathsf{r}}_{uk}|\mathsf{d}_{uk}}\left(\hat{r}_{uk}|d'_{uk}\right)p_{\mathsf{d}_{uk}}\left(d'_{uk};\hat{\mathcal{P}}^d_{l,u}\right)}, \end{aligned} \tag{36}$$

which can be rewritten as [46]

$$p_{\boldsymbol{d}_u|\hat{\boldsymbol{r}}_u}\left(\boldsymbol{d}_u|\hat{\boldsymbol{r}}_u;\hat{\mathcal{P}}_u^d\right) = \sum_{l=1}^{L}\pi_{l,u}\prod_{k=1}^{K}\mathcal{CN}\left(d_{uk};\overline{d}_{l,uk},\overline{\psi}_{l,uk}^d\right), \tag{37}$$

with

$$\pi_{l,u} = \frac{\prod_{k=1}^{K}\mathcal{CN}\left(0;\hat{r}_{uk}-\hat{\theta}_{l,uk}^d,\psi_{uk}^r+\hat{\phi}_{l,uk}^d\right)}{\sum_{l=1}^{L}\prod_{k=1}^{K}\mathcal{CN}\left(0;\hat{r}_{uk}-\hat{\theta}_{l,uk}^d,\psi_{uk}^r+\hat{\phi}_{l,uk}^d\right)}, \tag{38a}$$

$$\overline{d}_{l,uk} = \frac{\hat{\phi}_{l,uk}^d\hat{r}_{uk}+\psi_{uk}^r\hat{\theta}_{l,uk}^d}{\psi_{uk}^r+\hat{\phi}_{l,uk}^d},\ \overline{\psi}_{l,uk}^d = \frac{\hat{\phi}_{l,uk}^d\psi_{uk}^r}{\psi_{uk}^r+\hat{\phi}_{l,uk}^d}, \tag{38b}$$

where $\pi_{l,u}$ is the estimated mixture weight.

From (37), assuming that $\overline{d}_{l,uk},\forall k$ is not correlated, the posterior soft replica $\check{d}_{uk}$ and its MSE $\check{\psi}_{uk}^d$ can be obtained from the coefficient-wise conditional expectation as

$$\check{d}_{uk} = \sum_{l=1}^{L}\pi_{l,u}\cdot\overline{d}_{l,uk}, \tag{39a}$$

$$\check{\psi}_{uk}^d = \sum_{l=1}^{L}\pi_{l,u}\cdot\left(\left|\overline{d}_{l,uk}\right|^2+\overline{\psi}_{l,uk}^d\right)-\left|\check{d}_{uk}\right|^2. \tag{39b}$$

In a similar manner to (29), where we can obtain

$$\hat{d}'_{n,uk} = \psi'^d_{n,uk}\left(\frac{\check{d}_{uk}}{\check{\psi}_{uk}^d}-\hat{g}_{n,uk}^d\right), \tag{40a}$$

$$\frac{1}{\psi'^d_{n,uk}} = \frac{1}{\check{\psi}_{uk}^d}-\gamma_{n,uk}^d. \tag{40b}$$

Finally, since $|d_{uk}|=1$, the soft replica $\hat{d}_{n,uk}$ is given by normalizing the gain of the soft replica to 1 as follows,

$$\hat{d}_{n,uk} = \frac{\hat{d}'_{n,uk}}{\left|\hat{d}'_{n,uk}\right|}. \tag{41}$$

Besides, from (29), d_{uk} can be modeled as the output when $\hat{d}_{n,uk}$ is input to the AWGN channel, *i.e.*,

$$d_{uk} = \hat{d}'_{n,uk}+w'_{n,uk} = \hat{d}_{n,uk}+(\hat{d}'_{n,uk}-\hat{d}_{n,uk})+w'_{n,uk}, \tag{42}$$

where $w'_{n,uk}$ obeys $\mathcal{CN}\left(w'_{n,uk};0,\psi'^d_{n,uk}\right)$. From (11) and (42), the MSE for $\hat{d}_{n,uk}$ can be expressed as

$$\psi_{n,uk}^d = \left|\hat{d}_{n,uk}-\hat{d}'_{n,uk}\right|^2+\psi'^d_{n,uk}. \tag{43}$$

C. Update of Distribution Parameters in Algorithm 2

Finally, in this subsection, we explain how to update the distribution parameters of the vector-wise BG prior and vector-wise GM prior in lines 7 and 14–16 of Algorithm 2.

1) Update of Parameters for BG Distribution: The EM algorithm allows updating $\hat{\mathcal{P}}_q^{h,(\tau)}$ for every iteration step so as to maximize likelihood function $p_{\mathbf{Y}|\hat{\mathbf{P}}_q^h}\left(\boldsymbol{Y}|\hat{\mathcal{P}}_q^{h,(\tau)}\right)$, which is achieved by alternating between an *E-step*, which updates the posterior information for h_{nq} to minimize the Kullback-Leibler (KL) divergence between the likelihood function and its lower bound, and an *M-step*, which estimates the distribution parameters that maximize the lower bound [51]. Since the posterior estimates can be computed from the BiGaBP outputs, only the *M-step* solving the following maximization problem needs to be considered: [46]

$$\hat{\mathcal{P}}_q^{h,(\tau+1)} = \underset{\mathcal{P}_q^h}{\operatorname{argmax}}\sum_{n=1}^{N}\mathbb{E}_{\mathsf{h}_{nq}}\left\{\ln p_{\mathsf{h}_{nq}}\left(h_{nq};\mathcal{P}_q^h\right)\left|\hat{\mu}_{nq}^{(\tau)};\hat{\mathcal{P}}_q^{h,(\tau)}\right.\right\}, \tag{44}$$

where $\mathcal{P}_q^h$ is the set of true distribution parameters. Since it is difficult to maximize the cost function of (44) simultaneously for all parameters in $\mathcal{P}_q^h$, it is solved for each parameter separately and the update rules are derived as follows.

$$\hat{\lambda}_q^{(\tau+1)} = \sum_{n=1}^{N}\frac{\zeta_{nq}^{(\tau)}}{N},\ \hat{\phi}_q^{(\tau+1)} = \frac{\sum_{n=1}^{N}\zeta_{nq}^{(\tau)}\left(\left|\overline{h}_{nq}^{(\tau)}\right|^2+\overline{\psi}_{nq}^{h,(\tau)}\right)}{\sum_{n=1}^{N}\zeta_{nq}^{(\tau)}}, \tag{45}$$

where $\zeta_{nq}^{(\tau)}$, $\overline{h}_{nq}^{(\tau)}$, and $\overline{\psi}_{nq}^{h,(\tau)}$ are the posterior information and can be obtained from (27).

2) Update of Parameters for GM Distribution: Next, we describe the update of the GM distribution parameter set $\hat{\mathcal{P}}_u^{d,(\tau)}$. The problem arising here is that if the update rule for $\hat{\theta}_{l,uk}^{d,(\tau)}$ is simply derived using the EM algorithm, it is not possible to achieve maximal ratio combining (MRC) that takes into account the inhomogeneity of the amount of phase shift that varies for each index k according to α_k. To solve this problem, we consider instead the following update rule based on the minimum mean square error (MMSE) criterion using the posterior information of the l-th GM component as follows.

$$\begin{aligned}\hat{\varepsilon}_{l,u}^{(\tau+1)} &= \underset{\varepsilon_u}{\operatorname{argmin}}\sum_{k=1}^{K}\mathbb{E}_{\mathsf{d}_{uk}}\left\{\left|d_{uk}-\theta_{uk}^d\right|^2\left|\hat{r}_{uk}^{(\tau)};\hat{\mathcal{P}}_{l,u}^{d,(\tau)}\right.\right\}\\ &= \underset{\varepsilon_u}{\operatorname{argmax}}\underbrace{\sum_{k=1}^{K}\Re\left\{\left(\overline{d}_{l,uk}^{(\tau)}\right)^*\theta_{uk}^d\right\}}_{\triangleq J_1(\varepsilon_u)},\end{aligned} \tag{46}$$

where $\theta_{uk}^d \triangleq \exp\left[\mathrm{j}\alpha_k\varepsilon_u\right]$, the second-order Taylor approximation of which at the candidate CFO estimated in the previous iteration, *i.e.*, $\hat{\varepsilon}_{l,u}^{(\tau)}$, can be expressed as [19] and [20]

$$\begin{aligned}\theta_{uk}^d \approx{}& \hat{\theta}_{l,uk}^{d,(\tau)}+\mathrm{j}\alpha_k\left(\varepsilon_u-\hat{\varepsilon}_{l,u}^{(\tau)}\right)\hat{\theta}_{l,uk}^{d,(\tau)}\\ &-\frac{1}{2}\alpha_k^2\left(\varepsilon_u-\hat{\varepsilon}_{l,u}^{(\tau)}\right)^2\hat{\theta}_{l,uk}^{d,(\tau)}.\end{aligned} \tag{47}$$

Solving for $\frac{\partial}{\partial\varepsilon_u}J_1(\varepsilon_u)=0$ using (47), the MRC-based update formula can be obtained by

$$\hat{\varepsilon}_{l,u}^{(\tau+1)} = \hat{\varepsilon}_{l,u}^{(\tau)}-\frac{\sum_{k=1}^{K}\alpha_k\Im\left\{\left(\overline{d}_{l,uk}^{(\tau)}\right)^*\hat{\theta}_{l,uk}^{d,(\tau)}\right\}}{\sum_{k=1}^{K}\alpha_k^2\Re\left\{\left(\overline{d}_{l,uk}^{(\tau)}\right)^*\hat{\theta}_{l,uk}^{d,(\tau)}\right\}}, \tag{48}$$

where the estimation accuracy of (48) depends on the approximation accuracy of (47), *i.e.*, the magnitude of the residual CFO $\tilde{\varepsilon}_{l,u}^{(\tau)} \triangleq \varepsilon_u-\hat{\varepsilon}_{l,u}^{(\tau)}$ as described at the end of

Subsection III-A. Therefore, the proposed method, which can search for phase shifts in parallel while updating candidate CFOs, can keep $\hat{\varepsilon}_{l,u}^{(\tau)}$ small and maintain high estimation accuracy even in scenarios with large phase variations.

On the other hand, $\hat{\phi}_{l,uk}^{d,(\tau)}$ can be updated based on simple sampling average derived via the EM algorithm, in similarity to (45). To stabilize the convergence behavior, one averaged variance is estimated per OFDM symbol index t. In a similar manner to (44), the update rule for $\hat{\phi}_{l,uk}^{d,(\tau)}$ is obtained by solving the following maximization problem as [46]

$$\hat{\phi}_{l,uk}^{d,(\tau+1)} = \underset{\phi_{l,u}}{\operatorname{argmax}} \underbrace{\sum_{k\in\mathcal{K}_t} \mathbb{E}_{\mathsf{d}_{uk}} \left\{ \ln p_{\mathsf{d}_{uk}} \left(d_{uk}; \hat{\theta}_{l,uk}^{d,(\tau+1)}, \phi_{l,u} \right) \middle| \hat{r}_{uk}^{(\tau)}; \hat{\mathcal{P}}_{l,u}^{d,(\tau)} \right\}}_{\triangleq J_2(\phi_{l,u})}, \tag{49}$$

and the update rule can be obtained by solving for $\frac{\partial}{\partial \phi_{l,u}} J_2(\phi_{l,u}) = 0$, as

$$\begin{aligned} \hat{\phi}_{l,uk}^{d,(\tau+1)} &= \frac{1}{|\mathcal{K}_t|} \sum_{k\in\mathcal{K}_t} \mathbb{E}_{\mathsf{d}_{uk}} \left\{ \left| d_{uk} - \hat{\theta}_{l,uk}^{d,(\tau+1)} \right|^2 \middle| \hat{r}_{uk}^{(\tau)}; \hat{\mathcal{P}}_{l,u}^{d,(\tau)} \right\} \\ &= \frac{1}{|\mathcal{K}_t|} \sum_{k\in\mathcal{K}_t} \left(\left| \overline{d}_{l,uk}^{(\tau)} - \hat{\theta}_{l,uk}^{d,(\tau+1)} \right|^2 + \overline{\psi}_{l,uk}^{d,(\tau)} \right), \quad k \in \mathcal{K}_t. \end{aligned} \tag{50}$$

D. Initialization for Data Part in Algorithm 1

As a preliminary to describe the JCCDE algorithm, in this subsection, we describe how to initialize the soft replicas of the phase shifts and their MSEs corresponding to the data part, *i.e.*, $\hat{d}_{n,uk}^{(1)}, \psi_{n,uk}^{d,(1)}, \forall k \in \mathcal{K}_{\mathrm{d}}$, in line 9 of Algorithm 1, needed for data detection in the first iteration of Phase II. In this subsection, we omit the iteration index $\cdot^{(\tau)}$ for each variable derived in Phase I.

First, focusing on the computation of $\hat{d}_{n,uk}^{(1)}, \forall k \in \mathcal{K}_{\mathrm{d}}$, it shall be assume $\overline{d}_{l,uk} = \hat{\theta}_{l,uk}^{d}$ in (38b) because, of course, no message updates are made for the data part. Substituting this into (39a) and applying the first-order Taylor approximation, we obtain the soft replicas, given $\mathcal{P}_u^d$, as

$$\check{d}_{uk} = \sum_{l=1}^{L} \pi_{l,u} \hat{\theta}_{l,uk}^{d} \approx 1 + \mathrm{j}\alpha_k \sum_{l=1}^{L} \pi_{l,u} \hat{\varepsilon}_{l,u}, \ \forall k \in \mathcal{K}_{\mathrm{d}}. \tag{51}$$

Comparing (51) with $d_{uk} \triangleq \exp\left[\mathrm{j}\alpha_k \varepsilon_u\right] \approx 1 + \mathrm{j}\alpha_k \varepsilon_u$, we find the estimated CFO for the u-th UE can be expressed as

$$\hat{\varepsilon}_u = \sum_{l=1}^{L} \pi_{l,u} \hat{\varepsilon}_{l,u}, \tag{52}$$

and therefore we have

$$\hat{d}_{n,uk}^{(1)} = \exp\left[\mathrm{j}\alpha_k \hat{\varepsilon}_u\right], \quad \forall k \in \mathcal{K}_{\mathrm{d}}. \tag{53}$$

Next, focusing on the computation of $\psi_{n,uk}^{d,(1)}, \forall k \in \mathcal{K}_{\mathrm{d}}$, from (38b), it shall be assume $\overline{\psi}_{l,uk}^{d} = \hat{\phi}_{l,uk}^{d}$. Substituting $\overline{\psi}_{l,uk}^{d} = \hat{\phi}_{l,uk}^{d}$, $\left|\overline{d}_{l,uk}\right|^2 = \left|\hat{\theta}_{l,uk}^{d}\right|^2 = 1$, and $\left|\check{d}_{uk}\right|^2 = 1$ into (39b), we obtain the MSEs for $\hat{d}_{n,uk}^{(1)}, \forall k \in \mathcal{K}_{\mathrm{d}}$, as

$$\psi_{n,uk}^{d,(1)} = \sum_{l=1}^{L} \pi_{l,u} \hat{\phi}_{l,uk}^{d}, \quad \forall k \in \mathcal{K}_{\mathrm{d}}, \tag{54}$$

where in a similar manner to (17), we approximately have

$$\hat{\phi}_{l,uk}^{d} \approx \alpha_k^2 \mathbb{E}_{\tilde{\varepsilon}_{l,u}} \left\{ \left(\tilde{\varepsilon}_{l,u}\right)^2 \right\} = \alpha_k^2 \hat{\phi}_{l,u}^{\epsilon}. \tag{55}$$

Finally, the pilot part is used to calculate the residual variance $\hat{\phi}_{l,u}^{\epsilon} \triangleq \mathbb{E}_{\tilde{\varepsilon}_{l,u}} \left\{ \left(\tilde{\varepsilon}_{l,u}\right)^2 \right\}$. Summing both sides of (55) over all pilot symbols yields the following approximate relationship,

$$\sum_{k\in\mathcal{K}_{\mathrm{p}}} \hat{\phi}_{l,uk}^{d} \approx \hat{\phi}_{l,u}^{\epsilon} \sum_{k\in\mathcal{K}_{\mathrm{p}}} \alpha_k^2, \tag{56}$$

and therefore using (50) we have

$$\hat{\phi}_{l,u}^{\epsilon} \approx \frac{\sum_{k\in\mathcal{K}_{\mathrm{p}}} \hat{\phi}_{l,uk}^{d}}{\sum_{k\in\mathcal{K}_{\mathrm{p}}} \alpha_k^2}. \tag{57}$$

E. Data Detection in Algorithm 3

In this subsection, we describe the data detection algorithm based on the channel coefficients and phase shifts estimated in Subsections III-A through III-D. The proposed method is a *cross-domain* estimation scheme that iteratively performs time-domain inter-UE interference suppression based on the estimates obtained by the JCCE algorithm and frequency-domain equalization based on the periodicity of the time-domain channel provided by the CP as in the typical OFDM demodulation.

In the detection of an arbitrary TX symbol x_{mk}, the Soft IC is expressed as

$$\begin{aligned} \tilde{y}_{m,nk}^{s} &= y_{nk} - \sum_{i\neq m}^{M} \hat{d}_{n,ik} \hat{\sigma}_{n,ik} = \hat{d}_{n,mk} \sum_{q\in\mathcal{Q}_m} \check{h}_{nq} s_{qk} \\ &\quad + \underbrace{\sum_{q\in\mathcal{Q}_m} \tilde{c}_{k,nq} s_{qk} + \sum_{i\neq m}^{M} \left(d_{ik}\sigma_{ik} - \hat{d}_{n,ik}\hat{\sigma}_{n,ik} \right) + w_{nk}}_{\triangleq \nu_{m,nk}^{s}\text{: Time-domain residual interference plus noise}}, \end{aligned} \tag{58}$$

where $\hat{\sigma}_{n,mk} = \sum_{q\in\mathcal{Q}_m} \hat{h}_{k,nq} \hat{s}_{n,qk}$, $\tilde{c}_{k,nq} \triangleq h_{nq} d_{qk} - \check{h}_{nq} \hat{d}_{n,qk}$, and $\hat{d}_{n,mk} = \hat{d}_{n,uk}, \forall m \in \mathcal{M}_u$. Approximating $\nu_{m,nk}^{s}$ to a complex Gaussian random as in (20), (58) can be interpreted as the AWGN channel output of the time-domain TX symbol for the m-th TX antenna. Hence, by converting (58) here to the frequency domain using the cyclic nature of the time-domain channel, the frequency-domain representation can be expressed as

$$\begin{aligned} \tilde{y}_{m,nk}^{x} &= \sum_{i\in\mathcal{K}_t} f_{\kappa(k),\kappa(i)} \hat{d}_{n,mi}^{*} \tilde{y}_{m,ni}^{s} \\ &= \check{a}_{k,nm} x_{mk} + \underbrace{\sum_{i\in\mathcal{K}_t} f_{\kappa(k),\kappa(i)} \hat{d}_{n,mi}^{*} \nu_{m,ni}^{s}}_{\triangleq \nu_{m,nk}^{x}\text{: Frequency-domain residual interference plus noise}}, \end{aligned} \tag{59}$$

where $\check{a}_{k,nm}$ is the frequency-domain channel coefficient estimated from $\check{h}_{nq}, \forall q \in \mathcal{Q}_m$, as

$$\check{a}_{k,nm} = \sqrt{K_\mathrm{c}} \sum_{q \in \mathcal{Q}_m} f_{\kappa(k),\upsilon(q)} \check{h}_{nq}, \tag{60}$$

with $\upsilon(i) \triangleq (i \mod P)$. In a manner similar to (20), the conditional PDF of $\tilde{y}^x_{m,nk}$, given x_{mk}, can be expressed as

$$p_{\tilde{\mathrm{y}}^x_{m,nk}|\mathrm{x}_{mk}}\left(\tilde{y}^x_{m,nk}|x_{mk}\right) \propto \exp\left[-\frac{\left|\tilde{y}^x_{m,nk} - \check{a}_{k,nm}x_{mk}\right|^2}{\xi^x_{m,nk}}\right], \tag{61}$$

with

$$\xi^x_{m,nk} = \frac{1}{K_\mathrm{c}} \sum_{i \in \mathcal{K}_t} \xi^s_{m,ni}, \tag{62a}$$

$$\begin{aligned}\xi^s_{m,nk} = &\sum_{i \neq m}^{M} \sum_{q \in \mathcal{Q}_i} \left\{ \psi^s_{n,qk} \left|\hat{h}_{k,qk}\right|^2 \right. \\ &\left. + \left(\left|\hat{s}_{n,qk}\right|^2 + \psi^s_{n,qk}\right)\left(\left|\hat{h}_{k,qk}\right|^2 \psi^d_{n,qk} + \psi^h_{k,qk}\right)\right\} \\ &+ E_\mathrm{s} \sum_{q \in \mathcal{Q}_m} \left(\left|\check{h}_{qk}\right|^2 \psi^d_{n,qk} + \check{\psi}^h_{qk}\right) + N_0. \end{aligned} \tag{62b}$$

Based on (61), the subsequent estimation processes of x_{mk} can be performed in a manner similar to the JCDE described in [33], such that further details are omitted and offered only in a summarized form in the pseudo-code given in Algorithm 3. For further details, please see, *e.g.*, [33], [45], and [52].

The belief damping with η is introduced in lines 2 and 3 of Algorithm 3, and belief scaling [33], [45], [52], which controls convergence speed and stabilizes iterative behavior, is introduced in line 7, where $\beta(\tau)$ is the scaling parameter designed to be a monotonic increase linear function of the number of iterations as shown in line 6. All the computations of the proposed algorithm are scalar-by-scalar, the number of multiplication, division, subtraction, and addition operations; hence, the computational complexity of our JCCDE algorithm is of order $\mathcal{O}\big(\big(P + \log_2(K_\mathrm{c}) + \sqrt{J}\big)NMK_\mathrm{d} + PNMK + LMK\big)$ per iteration. Note that the proposed JCCDE algorithm requires additional computational complexity for domain switching to perform frequency-domain data detection.

IV. Performance Assessment

Numerical studies were conducted to demonstrate the performance of the proposed JCCDE receiver for uplink MUD in large MU-MIMO-OFDM systems.

Simulation parameters are summarized in Tab. I, where the maximum normalized CFO is $\varepsilon_{\max} = 0.0133$ [27]. The average RX power from each TX antenna was assumed to be identical on the basis of slow TX power control ($\phi^h_u = 1, \forall u$). The maximum number of delay taps was set to $P = 8$, and all delay paths shall fall within the CP length. The Gray-coded 4- and 16-quadrature amplitude modulation (QAM) were employed for symbol mapping, and the channel code was not used. The random pilots, *i.e.*, $x_{mk} = \sqrt{E_\mathrm{s}} \exp\left[\mathrm{j}\pi\upsilon_{mk}\right], \upsilon_{mk} \sim \mathcal{U}(-1, 1), \forall k \in \mathcal{K}_\mathrm{p}$, were used.

TABLE I
Simulation Parameters

MIMO configuration (N, M)	$(32, 16)$
OFDM configuration (K_c, T)	$(32, 14)$
Pilot resource index $(\mathcal{T}_\mathrm{p})$	$\{1, 14\}$
CP length (C)	8
Center frequency (f_c)	2 [GHz]
Subcarrier spacing (f_max)	15 [kHz]
Delay spread (T_DS)	Flat-fading: 0 [s], selective-fading: 1 [μs]
Channel delay model	Tapped delay line (TDL)-A

The number of GM components L was set to 10, the damping factor η was set to 1.0 at $\tau = 1$ and 0.5 otherwise, the scaling parameter b was set to 3 and 2.4 for 4QAM and 16QAM, respectively, and the maximum number of iterations was set constant to $(\tau_{1,\max}, \tau_{2,\max}) = (16, 16)$.

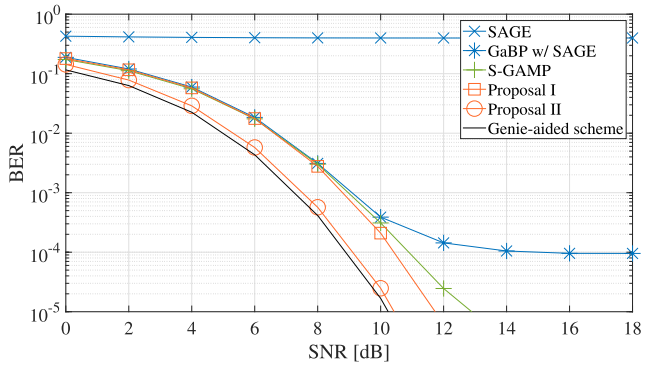
A. BER Performance in Frequency-Flat Channels

Our first set of results is given in Fig. 4, where the BER performances as a function of the signal-to-noise ratio (SNR), of the following MU-MIMO-OFDM systems with single antenna UEs, $U = 16$, $|\mathcal{M}_u| = 1$, under frequency-flat fading environments, *i.e.*, $T_\mathrm{DS} = 0$ [s], are compared:

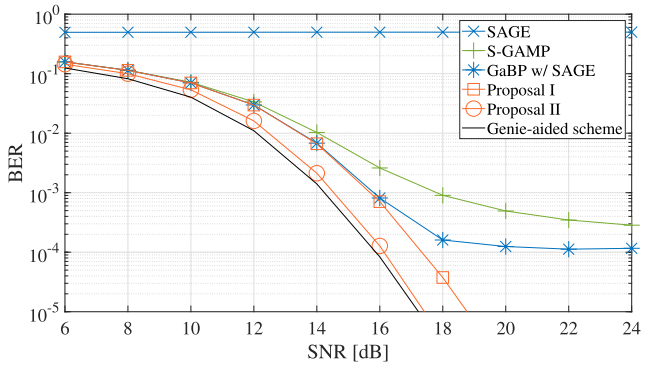
- *SAGE:* SotA Bayesian receiver based on the SAGE algorithm [20], where the channel coefficients and CFOs are iteratively estimated via the EM algorithm. This method cannot exploit the channel sparsity in the delay domain.
- *GaBP w/ SAGE:* SotA Bayesian receiver consisting of the GaBP-based CE part in Algorithm 2 and the SAGE-based CFO estimation, in order to verify the gain achieved by the proposed CFO estimation using the GM prior.
- *S-GAMP:* SotA Bayesian receiver based on the S-GAMP algorithm [27], where $L = 10$ and the belief damping is introduced as in [53]. Note that S-GAMP is only available in flat-fading channels due to its algorithmic structure.
- *Proposal I:* Proposed JCCE and data detection receiver that only performs the data detection by Algorithm 3, instead of Phase II, after Phase I, in order to verify the gain achieved by the JCCDE mechanism exploiting detected data symbols as equivalent soft pilots.
- *Proposal II:* Proposed JCCDE receiver presented in Algorithms 1–3.
- *Genie-aided scheme:* Idealized scheme in which the perfect CSI is known at the receiver, *i.e.*, without CFOs. Provides a reference lower bound which the proposed method can achieve.

Note that the channel coefficients and CFOs are estimated using only the pilot part in "SAGE," "GaBP w/ SAGE," "S-GAMP," and "Proposal I," and then the data detection is conducted by Algorithm 3.

In "SAGE," JCCE is performed based on the EM algorithm derived based on ML criterion, so it is impossible to take into account the sparsity in the delay domain when estimating the channel [54]. Therefore, "SAGE" cannot achieve highly accurate CE under severe rank-deficient conditions of $K_\mathrm{p}/Q = 0.5$, and as can be seen from Fig. 4, it fails to detect MIMO signals reliably (BER $> 10^{-1}$). In contrast, "GaBP w/ SAGE," which can exploit the sparsity of the delay-domain channel,

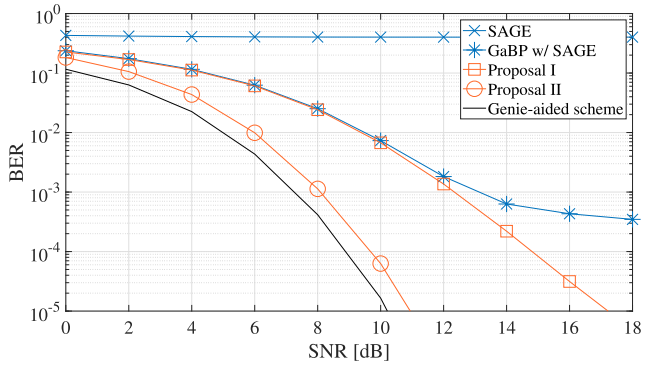


(a) Gray-coded 4QAM.

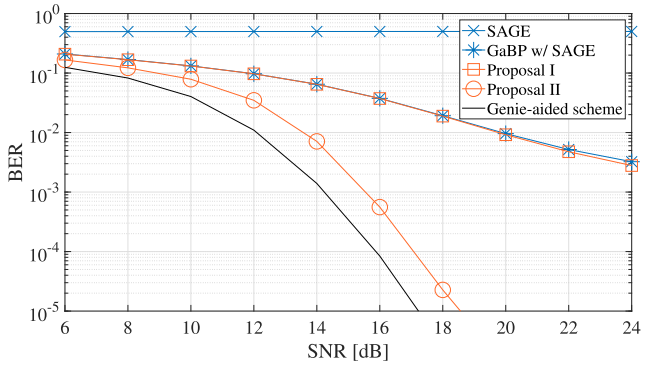


(b) Gray-coded 16QAM.

Fig. 4. BER performances of MU-MIMO-OFDM systems, where $(N, M, P, K_{\mathrm{p}}, K_{\mathrm{d}}) = (32, 16, 8, 64, 384)$ in frequency-flat fading channels.



(a) Gray-coded 4QAM.



(b) Gray-coded 16QAM.

Fig. 5. BER performances of MU-MIMO-OFDM systems, where $(N, M, P, K_{\mathrm{p}}, K_{\mathrm{d}}) = (32, 16, 8, 64, 384)$ in frequency-selective fading channels.

has significantly improved performance. However, it still suffers from high-level error floors even when using 4QAM, which is due to the fact that channel estimation error in the early iterations caused by non-orthogonal pilots degrades the Taylor approximation accuracy. Although "S-GAMP" using the grid-based CFO search performs well for 4QAM, serious performance degradation is inevitable for 16QAM due to lack of resolution of candidate CFO. In contrast, "Proposal I" and "Proposal II" can achieve stable detection performance without suffering from error floors even for 16QAM, and the latter "Proposal II" is asymptotically approaching the Genie-aided performance for both configurations. Remarkably, the degradation at BER $= 10^{-4}$ is less than 2.0 dB and 1.0 dB in "Proposal I" and "Proposal II," respectively.

B. BER Performance in Frequency-Selective Channels

Next, we evaluate the BER performance under frequency-selective fading environments in Fig. 5, where the other system parameters are the same as in Fig. 4.

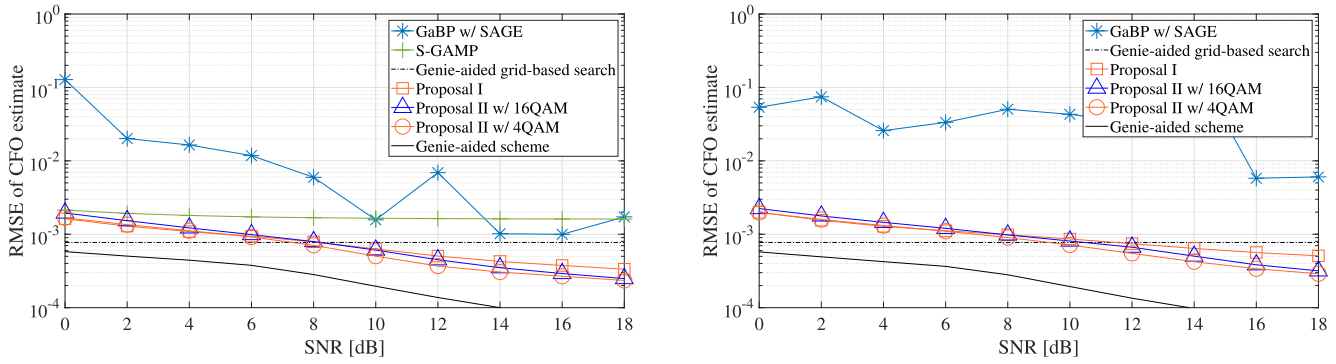
As expected from the results in Fig. 5, "SAGE" cannot detect MIMO signals. In frequency-selective channels where the channel coefficients to be estimated increase, the performances of the JCCE algorithms which use only the pilot part, "GaBP w/ SAGE" and "Proposal I," are also significantly degraded due to poor channel estimation accuracy. In contrast, "Proposal II" significantly outperforms the SotA using only the pilot part and approaches the Genie-aided scheme without suffering from error floors, thanks to the additional pilots brought in by detected symbols. Specifically, the degradation at BER $= 10^{-4}$ is less than 2.0 dB even when using 16QAM.

C. RMSE Performance of Normalized CFO Estimates

Let us shift our focus to the CFO estimation performance, in terms of the RMSE of the estimated normalized CFOs, defined as $\sqrt{\mathbb{E}_{\varepsilon_u}\left[|\varepsilon_u - \hat{\varepsilon}_u|^2\right]}$. In addition to "GaBP w/ SAGE," "S-GAMP," "Proposal I," and "Proposal II," the following performances are compared:

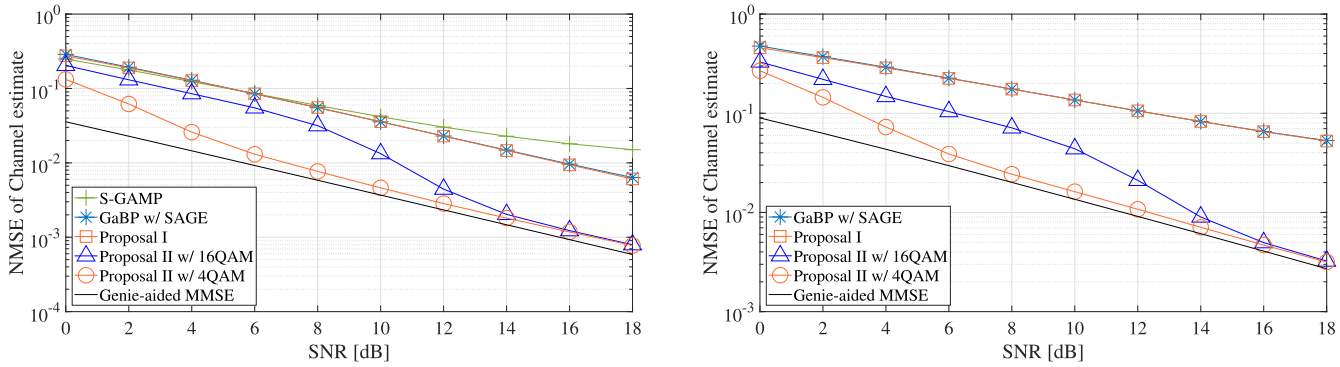
- *Genie-aided grid-based search:* Idealized scheme in which the candidate CFO closest to the true CFO is selected. It provides a reference lower bound which the grid-based search scheme such as S-GAMP can achieve.
- *Genie-aided scheme:* Genie-aided scheme in which the perfect knowledge of the channel coefficients and data symbols are provided as prior information for every iteration in "Proposal II." This performance can only be achieved by the proposed method (Proposal II) when using the perfect knowledge of instantaneous channel realizations and data symbols, and therefore serves as an absolute lower bound for Bayesian JCCDE methods in terms of the RMSE performance [34], [55].

Fig. 6 shows the RMSE performance of normalized CFO, where the system parameters are the same as in Figs. 4 and 5. The iterative convergence behavior of "GaBP w/ SAGE" is



(a) Flat-fading channel. (b) Frequency-selective channel.

Fig. 6. RMSE performances of MU-MIMO-OFDM systems, where $(N, M, P, K_{\mathrm{p}}, K_{\mathrm{d}}) = (32, 16, 8, 64, 384)$.



(a) Flat-fading channel. (b) Frequency-selective channel.

Fig. 7. NMSE performances of MU-MIMO systems as a function of SNR, where $(N, M, P, K_{\mathrm{p}}, K_{\mathrm{d}}) = (32, 16, 8, 64, 384)$.

unstable due to the mismatch in the Taylor approximation, and we can confirm that the CFO estimates have diverged. This mismatch is caused by a decrease in the accuracy of the interference cancellation that is the operating principle of SAGE, and is unavoidable when using non-orthogonal pilots. The RMSE performance of "GaBP w/ SAGE" changes drastically and irregularly because the Gram matrix of the non-orthogonal pilot matrix is not a unit matrix; therefore, the strength of the inter-UE interference changes probabilistically depending on the instantaneous channel realization. When considered together with the fact that the floor levels of "GaBP w/ SAGE" do not change regardless of the modulation order in Fig. 4, it can be inferred that the cause of the error floors is a burst error caused by this divergence behavior of the CFO estimates. In addition, the grid-based search allows 'S-GAMP" to provide stable CFO estimation; however, when using higher-order modulation such as 16QAM, the degraded CFO estimation accuracy in the high operating SNR region causes an error floor in the BER performance.

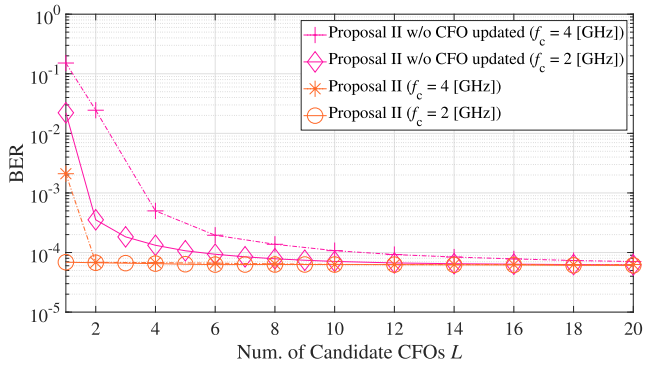
In contrast, the proposed method updates the CFO candidates based on the posterior probability of d_{uk} and achieves accurate CFO estimation at any operating SNR by adaptively switching between Taylor approximation-based scheme and candidate search-based scheme according to the estimation reliability provided by the BiGaBP algorithm. Consequently, in the low SNR region, where estimation accuracy of d_{uk} is poor, the proposed method approaches "Genie-aided grid-based search." It is not surprising that the convergence speed is different between the "Proposal II," in which the three variables are jointly estimated, and the "Genie-aided scheme," in which only CFO is estimated, and it has been mentioned in [56] and [57] that the difference in the number of estimated variables affects the convergence speed. However, as can be seen from Fig. 4, the effect on the BER performance is small.

D. NMSE Performance of Channel Estimates

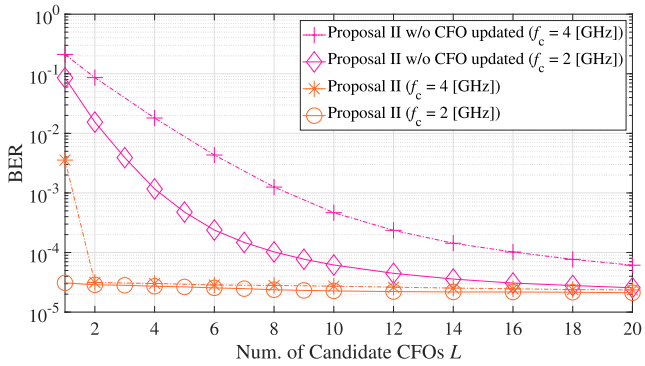
Next, we evaluate the CE performance, in terms of the NMSE of the channel coefficients, defined as $\mathbb{E}_{\boldsymbol{H}}\left[\left|\boldsymbol{H}-\check{\boldsymbol{H}}\right|_{\mathrm{F}}^2/\left|\boldsymbol{H}\right|_{\mathrm{F}}^2\right]$, where $[\boldsymbol{H}]_{n,q} \triangleq h_{nq}$ and $\left[\check{\boldsymbol{H}}\right]_{n,q} \triangleq \check{h}_{nq}$. In addition to "GaBP w/ SAGE," "S-GAMP," "Proposal I," and "Proposal II,", the following performance is compared:

- *Genie-aided MMSE:* Idealized scheme in which the channel coefficients estimated by linear MMSE filtering with the perfect knowledge of delay profile, CFO, and data symbols are provided. Provides an absolute lower bound in terms of the NMSE performance [34], [55].

Fig. 7 shows the NMSE performance, where the system parameters are the same as in Figs. 4 and 5. The performance of "S-GAMP" is deteriorated in the high SNR region due to the adverse effects of residual CFOs caused by insufficient grid resolution. In contrast, "Proposal I" achieves highly accurate channel estimation by updating and selecting the CFO candidates according to the variances of the GM distribution.

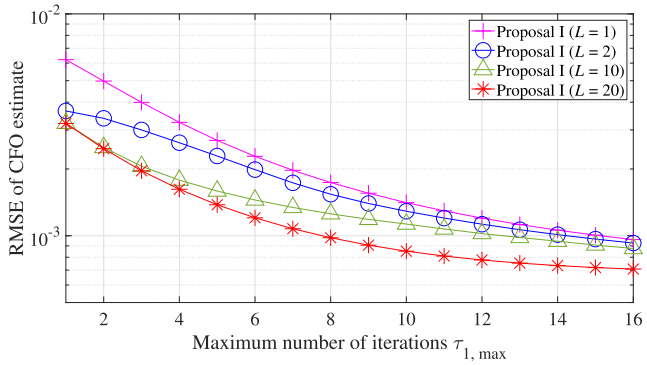


(a) Gray-coded 4QAM at SNR = 10 dB.

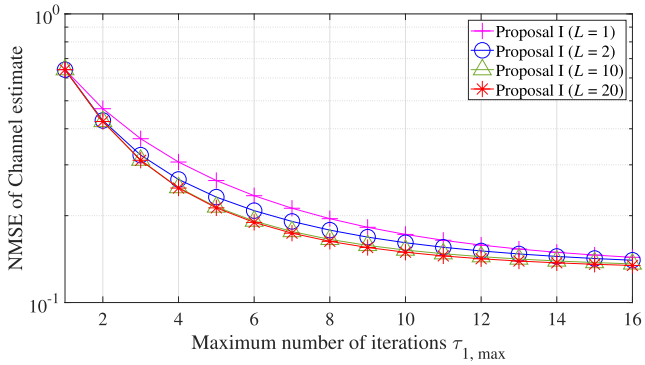


(b) Gray-coded 16QAM at SNR = 18 dB.

Fig. 8. BER performances of MU-MIMO-OFDM systems, where $(N, M, P, K_{\rm p}, K_{\rm d}) = (32, 16, 8, 64, 384)$ in frequency-selective fading channels.



(a) RMSE performance at SNR = 10 dB.



(b) NMSE performance at SNR = 10 dB..

Fig. 9. RMSE and NMSE performances of MU-MIMO-OFDM systems, where $(N, M, P, K_{\rm p}, K_{\rm d}) = (32, 16, 8, 64, 384)$ in frequency-selective fading channels.

Note that the performance degradation in "GaBP w/ SAGE" is due to the divergence of CFO estimates at a certain realization value, so no difference can be seen in terms of the average evaluation metric such as NMSE. In both fading environments, we see that "Proposal II" asymptotically approaches the Genie-aided performance in the high SNR region where the estimated data symbols serve as effective pilot symbols.

E. Assessment of Changes in the Number of CFO Candidates

Next, we focus on the BER performance of the proposed method to changes in the number of CFO candidates L. To confirm the efficacy of the candidate CFO update in "Proposal II," the following performance is compared:

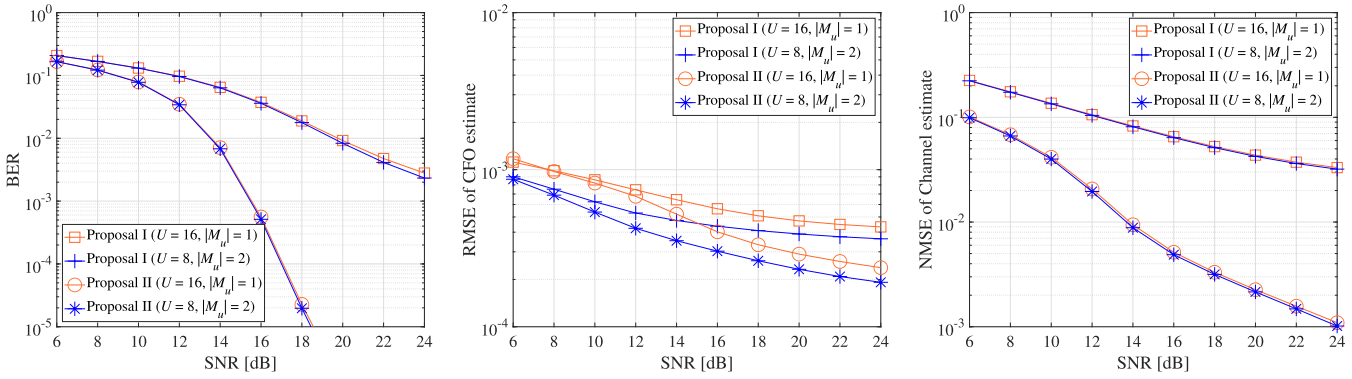
- *Proposal II w/o CFO update:* Proposed JCCDE receiver without updating candidate CFOs in (48).

Fig. 8 shows the BER performance as a function of L, where the center frequency $f_{\rm c}$ was set to 2 [GHz] and 4 [GHz], respectively, and the other system parameters are the same as in Fig. 5. The SNR is fixed at 10 dB and 18 dB in Fig. 8(a) and (b), respectively. As expected from the BER performance of "S-GAMP" in Fig. 4, "Proposal II w/o CFO update," which estimates CFO based on grid search, provides robust estimation performance, but its estimation accuracy depends deeply on the number of CFO candidates L. In contrast, "Proposal II" can significantly reduce the required value of L while maintaining the detection accuracy, by switching between the candidate CFO search and candidate CFO update strategies adaptively. This is supported by the fact that, when $L = 1$, the detection accuracy is significantly degraded at $f_{\rm c} = 4$ [GHz], while the accuracy improves rapidly for $L \geq 2$, since the candidate CFOs stabilizes the convergence performance in the early iteration steps.

F. Convergence Behavior of Proposed JCCE Algorithm

Next, we focus on the convergence behavior of the proposed JCCE algorithm in terms of the RMSE performance of the estimated CFOs and the NMSE performance of the estimated channel coefficients with different numbers of CFO candidates L. Since the convergence behavior of "Proposal II" is almost determined by the performance of "Proposal I," we will focus on the convergence behavior of "Proposal I."

Fig. 9(a) shows the RMSE performance as a function of the number of iterations $\tau_{1,\max}$, where the other system parameters are the same as in Fig. 5. First, focusing on the first iteration step ($\tau_{1,\max} = 1$), we can see that increasing L from 1 to 2 significantly improves the RMSE performance. This is because the difference between the candidate CFOs and the true CFO decreases. However, even if the number of candidate CFOs L is set to a larger value (*i.e.*, $L = 10$ and 20), the candidate CFOs are evenly distributed within their range of existence, so the estimation accuracy of d_{mk} does not change significantly, and the RMSE at $\tau_{1,\max} = 1$ do not improve significantly. Next, focusing on the iterative behavior



(a) BER performance. (b) RMSE performance. (c) NMSE performance.

Fig. 10. BER, RMSE, and NMSE performances of MU-MIMO-OFDM systems with Gray-coded 16QAM, where $(N, M, P, K_{\mathrm{p}}, K_{\mathrm{d}}) = (32, 16, 8, 64, 384)$ in frequency-selective fading channels.

in the early iteration steps, we can see that the convergence rate (speed) is slower for $L = 2$ than for $L = 1$. This is because in the candidate CFO update mechanism of the proposed method, the update step size is adjusted according to the balance between the reliability of the estimated phase shift and that of the candidate CFOs. Specifically, when $L = 1$, the reliability of the candidate CFOs is low, so the update step size based on the estimated phase shift becomes large, while when $L = 2$, the reliability of the estimated phase shift is lower than that of the candidate CFOs, so the update step size becomes small, resulting in the different convergence rate. Finally, focusing on the iterative behavior in the later iteration steps, we can see that the RMSE performance converges after approximately the same number of iterations, regardless of the value of L. The final convergence value varies depending on L, but let us see how much this difference affects the actual estimation performance in the following evaluation of CE accuracy.

Fig. 9(b) shows the NMSE performance as a function of $\tau_{1,\max}$, where the other system parameters are the same as in Fig. 9(a). When $L = 1$, the RMSE performance in the early iteration steps is poor, so we can see that the CE accuracy deteriorates compared to when multiple candidate CFOs are used (*i.e.*, $L > 2$). As L increases, the RMSE performance in the early iteration steps gradually improves, and at $L = 10$ the improvement is almost saturated. In addition, since the final convergence performance is also approximately saturated, it can be seen that improvements in the region where RMSE is less than 10^{-3} do not significantly contribute to actual estimation accuracy. The above results suggest that sufficient estimation accuracy can be achieved by setting $L = 2$, and sufficient convergence rate can be achieved by setting $L = 10$.

G. Multi-Antenna UEs

Next, we study the effect of increasing the number of TX antennas mounted on each UE. The total number of TX antennas is fixed at 16, and the performance of single-antenna UEs ($U = 16$, $|M_u| = 1, \forall u$) is compared with that of multi-antenna UEs ($U = 8$, $|M_u| = 2, \forall u$).

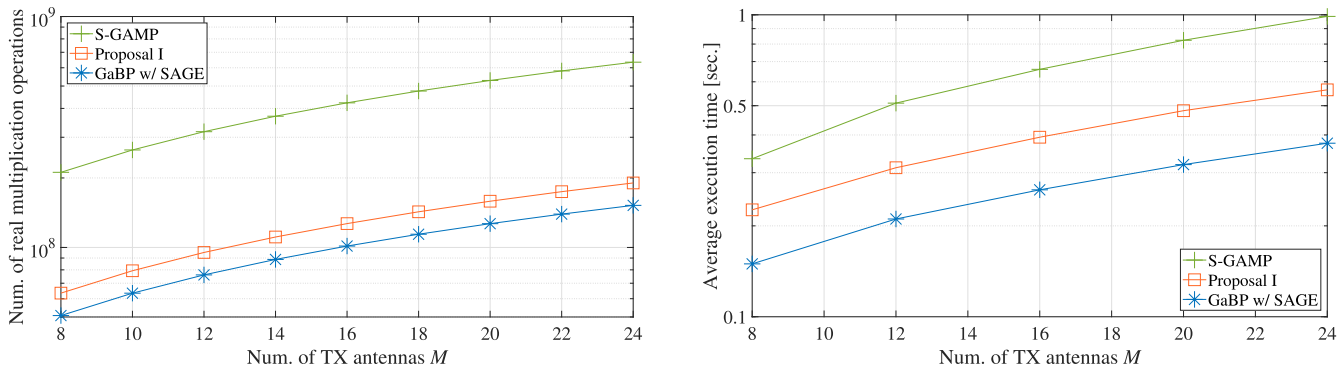
Figs. 10(a)–(c) show the BER, RMSE of the estimated CFOs, and NMSE of the estimated channel coefficients as a function of the SNR, where the other system parameters are the same as in Fig. 5(b). The BER performance in Fig. 10(a) and the NMSE performance of the estimated channel coefficients in Fig. 10(c) show that the performances of the multi-antenna UEs are slightly better than those of the single-antenna UEs. This slight (but definite) improvement is due to the improvement in the RMSE performance of the estimated CFOs in Fig. 10(b). Since each UE is equipped with a single local oscillator, in the case of multi-antenna UEs, the local oscillator is shared by $|M_u|$ TX antennas mounted on the u-th UE. Consequently, the number of CFOs estimated at the receiver is reduced from the total number of TX antennas M to the number of UEs U, improving the accuracy of CFO estimation compared to the case of single-antenna UEs.

H. Complexity Analysis

First, the computational complexity of each JCCE algorithm was evaluated in terms of the number of real multiplication operations required to estimate channel coefficients and CFOs. To evaluate the approximate number of real multiplication operations, we adopt the basic assumptions presented in [58].

The computational complexity per iteration of the proposed JCCE algorithm is $\mathcal{O}\left(PNMK_{\mathrm{p}} + LMK_{\mathrm{p}}\right)$, whereas that of the S-GAMP algorithm, which performs JCCE based on an the extended equivalent signal model [27], is $\mathcal{O}\left(LPNMK_{\mathrm{p}}\right)$. Therefore, the proposed method can operate at a low computational cost, particularly when the phase shift is large and L must be set to a large value. In CFO estimation based on SAGE [20], since the CFO for every UE is estimated individually using interference cancellation, the computational complexity per iteration of "GaBP w/ SAGE" is approximately the same as the proposed method, $\mathcal{O}\left(PNMK_{\mathrm{p}}\right)$.

Fig. 11(a) shows the number of real multiplication operations as a function of the number of TX antennas M. Similar to the results of the complexity analysis described above, it can be seen that the computational cost of "S-GAMP" is the largest, and that the proposed method and "GaBP w/ SAGE" have similar computational costs. More specifically, for $M = 16$, "Proposal I" can operate at about 0.3 times the computational cost of "S-GAMP" and about 1.2 times that of "GaBP w/ SAGE."



(a) Num. of multiplications. (b) Algorithm execution time.

Fig. 11. Complexity analysis of JCCE algorithm as a function of M with $N = 32$, $U = M$, $K_{\mathrm{p}} = 64$, $P = 8$, $\tau_{1,\max} = 16$, and $L = 10$.

For a more practical evaluation, Fig. 11(b) shows the average execution time[3] for each JCCE algorithm to estimate channel coefficients and CFOs. When the programs were actually executed and their execution times were compared, as shown in Fig. 11(b), the results show that the relative relationship between the proposed and the SotA alternatives is a similar trend to that shown in Fig. 11(a) in terms of the number of multiplication operations. However, when comparing with Fig. 11(a), the difference between "Proposal I" and "GaBP w/ SAGE" has slightly increased. This is thought to be due to the delay caused by the increased number of processing steps as a result of the proposed method taking into account the phase shift and the variance of the candidate CFOs. Looking at the Fig. 11(a) more specifically, in the case of $M = 16$, the execution time of "Proposal I" is approximately 0.6 times the execution time of "S-GAMP", and approximately 1.5 times the execution time of "GaBP w/ SAGE."

As can be seen from the results of the performance comparison shown above, considering that the proposed method can significantly improve performance and achieve performance close to the lower-bound reference in many cases, it can be said that the proposed method achieves an excellent trade-off between estimation capability and computational cost.

V. Conclusion

In this paper, we designed a novel algorithm to achieve low-complexity and high-accuracy large-scale MU-MIMO-OFDM signal demodulation with low pilot overhead in the presence of CFOs. The proposed Bayesian JCCE algorithm can significantly reduce the pilot overhead in frequency-selective fading channels, by employing the BG distribution and the GM distribution as the prior distributions of the channel coefficients and phase shifts, respectively. The CE using the BG prior can capture the statistical properties of frequency selective-fading channels, *i.e.*, the sparsity in the delay domain and the power deviation between paths. The CFO estimation using the GM prior enables robust estimation by adaptively switching between two strategies, candidate CFO search and candidate CFO update, depending on the reliability of the phase shift estimated by BiGaBP. In addition, using the estimated data symbols as effective soft pilots can further improve estimation accuracy. The numerical results show that our proposed method outperforms the SotA schemes and approaches the performance of the idealized scheme for a variety of system parameters. Furthermore, a comparative study with the SotA alternatives provided insight into the design of robust JCCE mechanisms in MU-MIMO-OFDM systems with short non-orthogonal pilots.

[3]To measure the execution time of the algorithms, we used the profiler of MATLAB R2023a. The measurements were performed on Windows 10 with Intel (R) Core (TM) i7-4770 CPU at 3.40 GHz and 32.0 GB RAM. The nonlinear function table is assumed to be read in memories in advance; hence, the processing time for computing $\exp[\cdot]$ can be ignored.

References

[1] P. Yang, Y. Xiao, M. Xiao, and S. Li, "6G wireless communications: Vision and potential techniques," *IEEE Netw.*, vol. 33, no. 4, pp. 70–75, Jul. 2019.

[2] I. F. Akyildiz, A. Kak, and S. Nie, "6G and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133995–134030, 2020.

[3] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 615–637, Mar. 2021.

[4] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 334–366, 2021.

[5] J.-J. van de Beek et al., "A time and frequency synchronization scheme for multiuser OFDM," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 11, pp. 1900–1914, Nov. 1999.

[6] A. Mohammadian and C. Tellambura, "RF impairments in wireless transceivers: Phase noise, CFO, and IQ imbalance—A survey," *IEEE Access*, vol. 9, pp. 111718–111791, 2021.

[7] K. Raghunath and A. Chockalingam, "SIR analysis and interference cancellation in uplink OFDMA with large carrier frequency/timing offsets," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2202–2208, May 2009.

[8] C. Jayawardena and K. Nikitopoulos, "Joint frequency offset compensation and detection for multi-user MIMO-OFDM systems with frequency asynchronous user access," in *Proc. IEEE Int. Conf. Commun.*, May 2023, pp. 529–534.

[9] O. Besson and P. Stoica, "On parameter estimation of MIMO flat-fading channels with frequency offsets," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 602–613, Mar. 2003.

[10] H. Minn and S. Xing, "An optimal training signal structure for frequency-offset estimation," *IEEE Trans. Commun.*, vol. 53, no. 2, pp. 343–355, Feb. 2005.

[11] J. Chen, Y.-C. Wu, S. Ma, and T.-S. Ng, "Ml joint CFO and channel estimation in OFDM systems with timing ambiguity," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2436–2440, Jul. 2008.

[12] M. Zhou, Z. Feng, X. Huang, and Y. Liu, "Maximum a posteriori probability (MAP) joint fine frequency offset and channel estimation for MIMO systems with channels of arbitrary correlation," *IEEE Trans. Signal Process.*, vol. 69, pp. 4357–4370, 2021.

[13] S. Jeong, A. Farhang, N. S. Stefan, and M. F. Flanagan, "Low-complexity joint CFO and channel estimation for RIS-aided OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 11, no. 1, pp. 203–207, Jan. 2022.

[14] J. Chen, Y.-C. Wu, S. Ma, and T.-S. Ng, "Joint CFO and channel estimation for multiuser MIMO-OFDM systems with optimal training sequences," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 4008–4019, Aug. 2008.

[15] S. Jeong, A. Farhang, N. S. Perovic, and M. F. Flanagan, "Joint CFO and channel estimation for RIS-aided multi-user massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 72, no. 9, pp. 11800–11813, Sep. 2023.

[16] Y. Yao and G. B. Giannakis, "Blind carrier frequency offset estimation in SISO, MIMO, and multiuser OFDM systems," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 173–183, Jan. 2005.

[17] W. Zhang, Q. Yin, and W. Wang, "ONE-shot blind CFO estimation for OFDM with multi-antenna receiver," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 6514–6517.

[18] Y. Meng, W. Zhang, G. L. Stüber, and W. Wang, "Blind fast CFO estimation and performance analysis for OFDM," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 11501–11514, Oct. 2020.

[19] M.-O. Pun, M. Morelli, and C.-C. Kuo, "Iterative detection and frequency synchronization for OFDMA uplink transmissions," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 629–639, Feb. 2007.

[20] T.-H. Pham, A. Nallanathan, and Y.-C. Liang, "Joint channel and frequency offset estimation in distributed MIMO flat-fading channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 648–656, Feb. 2008.

[21] X. Zeng and A. Ghrayeb, "Joint CFO and channel estimation for OFDMA uplink: An application of the variable projection method," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2306–2311, May 2009.

[22] L. He, S. Ma, Y.-C. Wu, and T.-S. Ng, "Semiblind iterative data detection for OFDM systems with CFO and doubly selective channels," *IEEE Trans. Commun.*, vol. 58, no. 12, pp. 3491–3499, Dec. 2010.

[23] S. Chakraborty and D. Sen, "Joint estimation of time, frequency offsets, and channel gains with ICIs in EF multi-relay DMIMO-OFDM system," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 5822–5838, Jul. 2017.

[24] P. Priya and D. Sen, "Data detection with CFO uncertainty and non-linearity for mmWave MIMO-OFDM systems," *IEEE Syst. J.*, vol. 16, no. 3, pp. 3734–3745, Sep. 2022.

[25] K. Ueda, T. Hara, and K. Ishibashi, "Iterative activity detection and carrier frequency offset estimation for grant-free NOMA," in *Proc. IEEE 33rd Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2022, pp. 1–6.

[26] C. Wei, H. Jiang, J. Dang, L. Wu, and H. Zhang, "Accurate channel estimation for mmWave massive MIMO with partially coherent phase offsets," *IEEE Commun. Lett.*, vol. 26, no. 9, pp. 2170–2174, Sep. 2022.

[27] G. Sun et al., "Massive grant-free OFDMA with timing and frequency offsets," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3365–3380, May 2022.

[28] L. Weng et al., "Effect of carrier frequency offset on channel estimation for SISO/MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1854–1863, May 2007.

[29] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2011, pp. 2168–2172.

[30] N. J. Myers and R. W. Heath, "Message passing-based joint CFO and channel estimation in mmWave systems with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3064–3077, Jun. 2019.

[31] A. K. Dutta, K. V. S. Hari, and L. Hanzo, "Minimum-error-probability CFO estimation for multiuser MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 64, no. 7, pp. 2804–2818, Jul. 2015.

[32] Y. Feng, H. Shen, W. Lu, N. Zhao, and A. Nallanathan, "Iterative joint frequency synchronization and channel estimation for uplink massive MIMO," *IEEE Internet Things J.*, vol. 11, no. 17, pp. 28891–28905, Sep. 2024.

[33] K. Ito, T. Takahashi, S. Ibi, and S. Sampei, "Bilinear Gaussian belief propagation for massive MIMO detection with non-orthogonal pilots," *IEEE Trans. Commun.*, vol. 72, no. 2, pp. 1045–1061, Feb. 2024.

[34] T. Takahashi, H. Iimori, K. Ando, K. Ishibashi, S. Ibi, and G. T. F. de Abreu, "Bayesian receiver design via bilinear inference for cell-free massive MIMO with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 22, no. 7, pp. 4756–4772, Jul. 2023.

[35] H. Iimori, T. Takahashi, K. Ishibashi, G. T. F. de Abreu, D. González, and O. Gonsa, "Joint activity and channel estimation for extra-large MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7253–7270, Sep. 2022.

[36] H. Iimori, T. Takahashi, K. Ishibashi, G. T. F. de Abreu, and W. Yu, "Grant-free access via bilinear inference for cell-free MIMO with low-coherence pilots," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7694–7710, Nov. 2021.

[37] K. Ito, T. Takahashi, S. Ibi, and S. Sampei, "Bilinear Gaussian belief propagation for large MIMO channel and data estimation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2020, pp. 1–6.

[38] J. Ziniel and P. Schniter, "Efficient high-dimensional inference in the multiple measurement vector problem," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 340–354, Jan. 2013.

[39] *Technical Specification Group Radio Access Network; NR; User Equipment (UE) Radio Transmission and Reception; Part 1: Range 1 Standalone (release 16)*, Standard 3GPP TS 38.101-1 V16.5.0, 3rd Generation Partnership Project, Sep. 2020.

[40] K. Ito, T. Takahashi, K. Furuta, S. Ibi, and G. Thadeu Freitas de Abreu, "Joint channel and data estimation via parametric bilinear inference for OTFS demodulation," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 6090–6105, 2024.

[41] J. T. Parker and P. Schniter, "Parametric bilinear generalized approximate message passing," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 795–808, Jun. 2016.

[42] S. Sarkar, A. K. Fletcher, S. Rangan, and P. Schniter, "Bilinear recovery using adaptive vector-AMP," *IEEE Trans. Signal Process.*, vol. 67, no. 13, pp. 3383–3396, Jul. 2019.

[43] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing—Part I: Derivation," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5839–5853, Nov. 2014.

[44] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *J. Phys. A, Math. Gen.*, vol. 36, no. 43, pp. 11111–11121, Oct. 2003.

[45] T. Takahashi, S. Ibi, and S. Sampei, "Design of adaptively scaled belief in multi-dimensional signal detection for higher-order modulation," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1986–2001, Mar. 2019.

[46] J. P. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.

[47] X. Meng, S. Wu, L. Kuang, and J. Lu, "An expectation propagation perspective on approximate message passing," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1194–1197, Aug. 2015.

[48] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, 2017.

[49] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 368–386, Jan. 2020.

[50] A. Chockalingam and B. S. Rajan, *Large MIMO Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[51] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer-Verlag, 2006.

[52] T. Takahashi, A. Tölli, S. Ibi, and S. Sampei, "Low-complexity large MIMO detection via layered belief propagation in beam domain," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 234–249, Jan. 2022.

[53] P. Schniter and S. Rangan, "Compressive phase retrieval via generalized approximate message passing," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 1043–1055, Feb. 2015.

[54] Y. Xu, J. Zhong, Y. Cai, M. Zhao, and B. Chen, "Joint CFO and sparse channel estimation for MIMO-OFDM systems via the sage algorithm," in *Proc. Int. Conf. Wireless Commun. Signal Process.*, 2013, pp. 1–5.

[55] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.

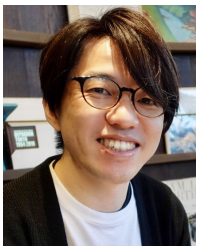
[56] X. Meng and J. Zhu, "Bilinear adaptive generalized vector approximate message passing," *IEEE Access*, vol. 7, pp. 4807–4815, 2019.

[57] M. Akrout, A. Housseini, F. Bellili, and A. Mezghani, "Bilinear generalized vector approximate message passing," 2020, *arXiv:2009.06854*.

[58] S. Yoon and C.-B. Chae, "Low-complexity MIMO detection based on belief propagation over pairwise graphs," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2363–2377, Jun. 2014.

Kenta Ito (Graduate Student Member, IEEE) received the B.E. degree in electrical engineering from Toyama University, Japan, in 2019, and the M.E. degree in communication engineering from Osaka University, Japan, in 2021, where he is currently pursuing the Ph.D. degree with the Graduate School of Engineering. His research interests include belief propagation, compressed sensing, signal processing, and wireless communications.

Takumi Takahashi (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in communication engineering from Osaka University, Osaka, Japan, in 2016, 2017, and 2019, respectively. From 2018 to 2019, he was a Visiting Researcher at the Centre for Wireless Communications, University of Oulu, Finland. In 2019, he joined the Graduate School of Engineering, Osaka University, where he is currently an Associate Professor. His current research interests include Bayesian inference, belief propagation, signal processing, and wireless communications. He received the 80th Best Paper Award from IEICE and the 2019 and 2023 Best Paper Awards from the IEICE Communication Society. He was certified as an exemplary reviewer of IEEE WIRELESS COMMUNICATIONS LETTERS in 2023.



Koji Ishibashi (Senior Member, IEEE) received the B.E. and M.E. degrees in engineering from The University of Electro-Communications, Tokyo, Japan, in 2002 and 2004, respectively, and the Ph.D. degree in engineering from Yokohama National University, Yokohama, Japan, in 2007. From 2007 to 2012, he was an Assistant Professor at the Department of Electrical and Electronic Engineering, Shizuoka University, Hamamatsu, Japan. From 2010 to 2012, he was a Visiting Scholar at the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. Since April 2012, he has been with the Advanced Wireless and Communication Research Center (AWCC), The University of Electro-Communications, Tokyo, where he is currently a Professor. His current research interests include robust beamforming, grant-free access, energy-harvesting, compressed sensing, coding, cell-free networks, and MIMO technologies. He is a Senior Member of IEICE. He was a recipient of the Takayanagi Research Encouragement Award in 2009 and the KDDI Foundation Award in 2023. He was certified as an exemplary reviewer of IEEE COMMUNICATIONS WIRELESS LETTERS in 2015 and awarded by the Telecommunication Technology Committee (TTC) for his devotion to standardization activities in 2020. He served as an Associate Editor for *IEICE Transactions on Communications* and IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and a Guest Editor for IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY.

Koji Igarashi (Member, IEEE) received the B.E. degree in electrical and computer engineering from Yokohama National University, Yokohama, Japan, in 1997, and the M.E. and Ph.D. degrees in electronic engineering from The University of Tokyo, Tokyo, Japan, in 1999 and 2002, respectively. From 2002 to 2004, he was with Furukawa Electric Corporation Ltd. Since 2004, he has been with The University of Tokyo. From 2007 to 2011, he was an Assistant Professor with the Department of Frontier Informatics and the Department of Electrical Engineering and Information Systems. From 2012 to 2013, he was with KDDI Research and Development Laboratories Inc. From 2013 to 2023, he was an Associate Professor with the Department of Electrical, Electronic and Infocommunications Engineering, Osaka University, Osaka, Japan. Since 2024, he has been a Professor with the Graduate School of Engineering Science, Osaka University. His current research interests include high-capacity long-haul optical fiber transmission systems, signal processing for coherent optical communication systems, and optical measurement techniques.

Shinsuke Ibi (Senior Member, IEEE) received the B.E. degree in advanced engineering from the Suzuka College of Technology, Japan, in 2002, and the M.E. and Ph.D. degrees in communication engineering from Osaka University, Japan, in 2004 and 2006, respectively. From 2005 to 2006, he was a Visiting Researcher at the Centre for Wireless Communications, University of Oulu, Finland. In 2006, he joined with the Graduate School of Engineering, Osaka University. From 2010 to 2011, he was a Visiting Researcher at the University of Southampton, U.K. In 2019, he joined with Doshisha University, where he is currently a Professor with the Faculty of Science and Engineering. His research interests include EXIT-based coding theory, iterative detection, digital signal processing, cognitive radio, and communication theory. He received the 64th, 71st, and 80th Best Paper Awards from IEICE, the 2017, 2018, 2019, and 2023 Best Paper Awards from IEICE Communication Society, and the 24th Telecom System Technology Award from the Telecommunication Advancement Foundation.