Bilinear Gaussian Belief Propagation for Massive MIMO Detection With Non-Orthogonal Pilots

Kenta Ito , *Graduate Student Member, IEEE*, Takumi Takahashi , *Member, IEEE*,
Shinsuke Ibi , *Member, IEEE*, and Seiichi Sampei, *Life Fellow, IEEE*

Abstract— We propose a novel joint channel and data estimation (JCDE) algorithm via bilinear Gaussian belief propagation (BiGaBP) for massive multi-user MIMO (MU-MIMO) systems with non-orthogonal pilot sequences. The contribution aims to reduce significantly the communication overhead required for channel acquisition by enabling the use of short non-orthogonal pilots, while maintaining multi-user detection (MUD) capability. Bilinear generalized approximate message passing (BiGAMP), which is systematically derived by extending approximate message passing (AMP) to the bilinear inference problem (BIP), provides computationally efficient approximate implementations of large-scale JCDE via sum-product algorithm (SPA); however, as the pilot length decreases, the estimation accuracy is severely degraded. To tackle this issue, the proposed BiGaBP algorithm generalizes BiGAMP by relaxing its dependence on the large-system limit approximation and leveraging the belief propagation (BP) concept. In addition, a novel belief scaling method complying with the data detection accuracy for each iteration step is designed to avoid the divergence behavior of iterative estimation in the early iterations due to the use of non-orthogonal pilots, especially in insufficient large-system conditions. Simulation results show that the proposed method outperforms the state-of-the-art schemes and approaches the performance of idealized (genie-aided) scheme in terms of mean square error (MSE) and bit error rate (BER) performances.

Index Terms— Massive multi-user MIMO, Bayesian bilinear inference, belief propagation, BiGAMP, belief scaling.

I. INTRODUCTION

MASSIVE multiple-input multiple-output (MIMO) systems have been considered as one of the key technologies for the fifth generation (5G) advanced and sixth generation (6G) networks, which promise significant performance improvements in many aspects, such as spectral efficiency, detection reliability, and energy efficiency [1], [2], [3]. In particular, massive MU-MIMO technology, where a base station (BS) is equipped with massive number of antenna arrays, can simultaneously serve a massive amount of wireless links, and brings enormous connectivity in the

uplink access [4], [5]. Under such a scenario, low-complexity and large-scale MUD consisting of high-dimensional channel estimation and subsequent data detection play a key role in the operation of massive MU-MIMO [6], [7].

Conventional MUD is composed of the following two steps, channel estimation based on a pilot (training) sequence and subsequent data detection based on estimated channel state information (CSI). To obtain highly accurate CSI knowledge with this scheme, orthogonal pilot sequences must be used under the assumption that the pilot length K_{p} is, at least, larger than the maximum number of uplink users M communicating simultaneously [8]. This implies that the communication overhead required for CSI acquisition increases rapidly with the number of simultaneous connections, which means that it might be infeasible to obtain accurate CSI in the case of fast-fading environments due to its short channel coherence time [9], [10]. In addition, the facile use of short non-orthogonal pilots ($K_{\mathrm{p}} < M$) for channel estimation results in rank-deficient conditions in the estimated channel, leading to severe performance deterioration in subsequent data detection.

One promising solution to shorten the pilot length without sacrificing the estimation accuracy is JCDE [11], [12], [13], [14]. By exploiting the statistical quasi-orthogonality of data sequences, JCDE takes advantage of estimated data symbols as equivalent soft pilot symbols, providing significant improvements in system performance. The classical JCDE scheme assumes the exchange of log-likelihood ratios (LLRs) between the signal detector and channel decoder based on the turbo-principle [11], [12], where the reliability of tentative data symbols is enhanced at every iteration through an error correction process and then used as additional pilot symbols. However, iterative decoding increases power consumption and causes severe processing delays, which have been obstacles to the practical use of these typical JCDE algorithms. To address this issue, the JCDE scheme based on Bayesian message passing (MP) without requiring channel decoding at every iteration has been investigated [15], [16].

The most common algorithm of this approach is BiGAMP [15], which is derived by approximating the SPA designed for the BIP according to the generalized approximate message passing (GAMP) framework [17], [18] in the large-system limit.[1] However, the original GAMP has been proposed

---

[1] The system limit condition, where input and output dimensions, M and N, and symbol lengths K, respectively, are infinity for given compression rates $\rho_1 \triangleq N/M$ and $\rho_2 \triangleq K/M$.

in the context of linear inference problems (LIPs), which is appropriately designed on the premise of a bipartite factor graph (FG) consisting of factor nodes (FNs) and variable nodes (VNs) corresponding to observations and unknown parameters, respectively. This makes it difficult to straightforwardly extend GAMP to the BIP which is represented by a tripartite FG consisting of FNs and two VNs. Therefore, the BiGAMP algorithm in [15], which is derived by extending the GAMP algorithm to the BIP forcibly, cannot properly decouple the self-feedback between the two VNs across iterations using the *Onsager* correction term [16]. Another problem is that in aligning the resultant message update rules to correspond with the GAMP rules, the approximation accuracy assumed in the derivation of each message is not consistent throughout the algorithm. Because of the above inconveniences, the JCDE algorithm via BiGAMP in [15] can achieve high estimation accuracy only when a sufficient system size can be assumed and prior information, *e.g.*, sufficiently long pilots, is available [19].

On the one hand, while accepting the drawback of BiGAMP described above, some methodologies have been proposed to achieve robust signal recovery by incorporating the knowledge of problem structure into the MP rules. In [20], [21], and [22], for instance, the sparse structure of massive MIMO channel in the beam domain brought about by a limited range of angle of arrival (AoA) is exploited as additional prior information to mitigate the under-determined condition of MUD.

On the other hand, in an attempt to solve the drawback of BiGAMP described above, the authors in [16] focus on the FG structure of BIP and propose to modify the MP rule of BiGAMP itself. Specifically, by interpreting a tripartite graph corresponding to BIP as consisting of two bipartite graphs, and by designing the MP rule while considering both the self-feedback that occurs in each bipartite graph and the self-feedback that occurs in information exchange between the graphs, this enables us to derive appropriate Onsager correction terms. The resulting MP rule is a reasonable extension of the GAMP framework to BIP, eliminating the mathematical inconsistencies found in the conventional BiGAMP algorithm. Through numerical simulations, it has been shown in [16] that the modified algorithm achieves more robust performance against changes in system size than the conventional alternatives. However, poor convergence behavior due to mismatch with the asymptotic conditions in the large-system limit is still unavoidable. The mitigation of this behavior requires the introduction of successive MP mechanism that results in an increase in the processing delay [16].

Inspired by these works, we proposed the JCDE algorithm via BiGaBP in [23][2] to achieve more robust signal recovery in more realistic system sizes, with the same order of complexity as BiGAMP. The core idea of the proposed method is to relax approximations in the algorithm derivation process. Gaussian belief propagation (GaBP) [24], [25] relies only on scalar Gaussian approximation (SGA) based on mild central limit

theorem (CLT), in contrast to the much harder asymptotic conditions required by the GAMP algorithm. Indeed, the GAMP algorithm in the context of LIPs is proven to be systematically derived from a rigorous approximation of the GaBP algorithm in the large-system limit condition [26]. The proposed BiGaBP algorithm, which can be systematically derived by extending the GaBP framework to BIPs, also relies only on SGA, which provides more stable iterative convergence behavior in insufficient system sizes even when using non-orthogonal pilots [23]. The BiGaBP framework was recently employed in the receiver design of cell-free massive MIMO (CF-mMIMO) systems adopting low-resolution analog-to-digital converters (ADCs) [27], the joint activity and channel estimation (JACE) scheme of extra-large MIMO (XL-MIMO) systems [28], and the receiver design of grant-free (GF) access [29], and in all of them the BiGaBP-based approaches were shown to outperform earlier state-of-the-art schemes.

However, there are no literature that mention in detail the differences and relationships among the algorithmic structures of BiGAMP [15], modified BiGAMP [16], and BiGaBP. The performance comparison via computer simulations presented in the literature [23] was done only in a limited problem setting in a specific system configuration; hence, its systematic position and effective scope have not yet been fully reported. Therefore, to validate the efficacy of the BiGaBP framework to design the JCDE algorithm, this article extends our conference paper [23] by presenting a detailed process flow of the JCDE algorithm via BiGaBP with a novel belief scaling method, clarifying the relationship between BiGAMP, and then providing a detailed simulation analysis of estimation performance.

The contributions of the article are summarized as follows[3]:

- A novel JCDE algorithm via BiGaBP is presented. The derivation relies only on the SGA in conformity with mild CLT, whose underlying assumptions are much softer than the large-system limit assumption on which the BiGAMP algorithm heavily relies, is presented. In addition, to suppress unstable iterative convergence behavior due to using non-orthogonal pilots, we also propose a new belief scaling method that extends adaptively scaled belief (ASB), which was proposed in [25], [26], and [33] so as to improve detection capability of the MP algorithm in LIPs, to BIPs. Specifically, the proposed scaling method combines the beliefs of the pilot and data parts propagated on the FG with different weights depending on the number of iterations, thereby suppressing iterative divergence behavior in the first stage of iteration, which is a problem when using non-orthogonal pilots. Note that this extension is only valid for BiGaBP which can suppress the self-feedback via generating extrinsic values based on the BP regime rather than Onsager term.
- To clarify that BiGaBP is truly a relaxed approximation version of BiGAMP, we prove that the BiGAMP

---

[2]This conference paper is an earlier version of this article, which was presented at the IEEE GLOBECOM 2020.

[3]Please note that higher-complexity alternatives based on bilinear adaptive vector AMP (BAd-VAMP), which is an extension of vector AMP (VAMP) [30] to BIP [31], [32], will not be considered in this article, as they require matrix inversion operation in each iteration process, and therefore do not offer the same scalability potential of the technique proposed here.

algorithm in [16] can be derived by approximating the BiGaBP algorithm in asymptotic conditions of the large-system limit. Furthermore, we compare the BiGAMP algorithms proposed in [15] and [16], respectively, to identify differences in algorithm structure resulting from the different interpretations of FGs described above.

- To confirm the efficacy of the JCDE algorithm via BiGaBP in massive MU-MIMO systems, we compare the performance of conventional and proposed methods on various system parameters. Our numerical results are presented in terms of normalized mean square error (NMSE) of estimated quantities and BER. In addition to evaluating performance as a function of conventional signal-to-noise ratio (SNR), we also evaluate the estimation performance for different pilot length and data length settings. The simulation results are shown to outperform the current state-of-the-art for any system parameter and approach the performance of an idealized scheme in which channel coefficients are perfectly known.

Finally, this section concludes by discussing deep learning (DL)-based JCDE schemes that have been attracting attention in recent years and then clarifying the position of the proposed method relative to these methods. Starting with the proposal of the direct demodulation technique without prior explicit channel estimation was proposed for orthogonal frequency-division multiplexing (OFDM) systems in [34], the JCDE methods using deep neural networks (DNNs)-based channel equalization, have been actively investigated [35], [36], [37]. More recently, the DL-based JCDE algorithms incorporate *model-driven* algorithm design, such as the deep unfolding (DU) technique [38], have been proposed for various systems in [39], [40], and [41]. These methods have been reported to achieve more efficient learning by leveraging *domain* knowledge and can achieve better performance at lower learning cost than the traditional *black-box* counterparts, especially in massive MIMO scenarios. As a template for such DU-aided methods, this article contributes to provide a novel framework to design a more robust JCDE algorithm.

Notation: Throughout this paper, vectors, and matrices are denoted by lower- and upper-case bold-face letters, respectively. The conjugate, transpose, and conjugate transpose operators are denoted by $\cdot^*$, $\cdot^\mathrm{T}$, and $\cdot^\mathrm{H}$, respectively. $P_{\mathsf{a|b}}[a|b]$ and $p_{\mathsf{a|b}}(a|b)$ respectively represent the conditional probability mass function (PMF) and the probability density function (PDF) of a realization a of random variable a given the occurrence of a realization b of random variable b. $\mathbb{E}_{\mathsf{a}}\{\cdot\}$ is the expected value of random variable a. $\mathbb{E}_{\mathsf{a|b}=b}\{\cdot\}$ denotes the conditional expectation of random variable a given the occurrence of a realization b of random variable b. $\mathbb{C}^{a\times b}$ denotes a complex field of size $a\times b$. $\mathcal{CN}(\dot{x};a,b)$ indicates that $\dot{x}$ obeys a complex-valued Gaussian process with mean a and variance b. $\sum_{i\neq b}^{B} a_i$ means the operation of adding the elements of a_i corresponding to $i\in\{1,\cdots,b-1,b+1,\cdots,B\}$. $\boldsymbol{I}_a$ represents an $a\times a$ square identity matrix. $\mathrm{tr}\,[\cdot]$ denotes the trace of the matrix and $|\cdot|_\mathrm{F}$ denotes the Frobenius norm. $[\boldsymbol{A}]_{ab}$ is the (a,b) element in the matrix $\boldsymbol{A}$. $\int_a f(a)$ is an indefinite integral of $f(a)$.

II. PRELIMINARIES

A. Signal Model

Consider a single-cell massive MU-MIMO system, where a BS has N receive (RX) antennas and M ($\leq N$) user equipment (UE) devices are equipped with a single transmit (TX) antenna. At the k-th discrete time instance, the m-th UE transmits a TX symbol x_{mk}, which represents one among Q constellation points $\mathcal{X} \triangleq \{\chi_1,\cdots,\chi_q,\cdots,\chi_Q\}$. The average energy density of constellations in $\mathcal{X}$ is denoted by E_s. Denoting the TX vector at the k-th discrete time instance by $\boldsymbol{x}_k \triangleq [x_{1k},\ldots,x_{mk},\ldots,x_{Mk}]^\mathrm{T} \in \mathcal{X}^{M\times 1}$, the RX symbol received at the n-th RX antenna under the assumption of frequency-flat and slow-fading is given by

$$\boldsymbol{y}_k \triangleq [y_{1k},\ldots,y_{nk},\ldots,y_{Nk}]^\mathrm{T} = \boldsymbol{H}\boldsymbol{x}_k + \boldsymbol{w}_k, \quad (1)$$

where $\boldsymbol{H} \in \mathbb{C}^{N\times M}$ denotes an $N \times M$ MIMO channel matrix, where the (n,m) element, h_{nm}, obeys $\mathcal{CN}(h_{nm};0,\phi)$, with $\phi = 1/N$, owing to slow TX power control. The complex additive white Gaussian noise (AWGN) vector is denoted by $\boldsymbol{w}_k \triangleq [w_{1k},\ldots,w_{nk},\ldots,w_{Nk}]^\mathrm{T} \in \mathbb{C}^{N\times 1}$ obey $\mathcal{CN}(w_{nk};0,N_0)$, where N_0 is the noise spectral density, and thus the covariance matrix of $\boldsymbol{w}_k$ is given by $\mathbb{E}_{\mathsf{w}}\{\boldsymbol{w}_k\boldsymbol{w}_k^\mathrm{H}\} = N_0\boldsymbol{I}_N$.

Then, the spatial-temporal matrices can be expressed as $\boldsymbol{X} \triangleq [\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k,\ldots,\boldsymbol{x}_K] \in \mathbb{C}^{M\times K}$, $\boldsymbol{Y} \triangleq [\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k,\ldots,\boldsymbol{y}_K] \in \mathbb{C}^{N\times K}$, and $\boldsymbol{W} \triangleq [\boldsymbol{w}_1,\ldots,\boldsymbol{w}_k,\ldots,\boldsymbol{w}_K] \in \mathbb{C}^{N\times K}$. Assuming that the channel matrix $\boldsymbol{H}$ is constant during K successive transmissions, concatenating K successive RX vectors yields the following compact spatial-temporal RX signal representation as

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{W}. \quad (2)$$

In the TX symbol matrix $\boldsymbol{X}$, each UE device forms a frame with a length of K symbols, which includes K_p pilot symbols with the index $k_\mathrm{p} \in \mathcal{K}_\mathrm{p} \triangleq \{1,\cdots,K_\mathrm{p}\}$ and K_d data symbols with the index $k_\mathrm{d} \in \mathcal{K}_\mathrm{d} \triangleq \{K_\mathrm{p}+1,\cdots,K\}$, *i.e.*, $K = K_\mathrm{p} + K_\mathrm{d}$. Consequently, the spatial-temporal matrices can be sub-divided into pilot and data parts as follows: $\boldsymbol{X} \triangleq [\boldsymbol{X}_\mathrm{p},\boldsymbol{X}_\mathrm{d}],(\boldsymbol{X}_\mathrm{p} \in \mathbb{C}^{M\times K_\mathrm{p}}, \boldsymbol{X}_\mathrm{d} \in \mathbb{C}^{M\times K_\mathrm{d}})$, $\boldsymbol{Y} \triangleq [\boldsymbol{Y}_\mathrm{p},\boldsymbol{Y}_\mathrm{d}]$, $(\boldsymbol{Y}_\mathrm{p} \in \mathbb{C}^{N\times K_\mathrm{p}}, \boldsymbol{Y}_\mathrm{d} \in \mathbb{C}^{N\times K_\mathrm{d}})$, and $\boldsymbol{W} \triangleq [\boldsymbol{W}_\mathrm{p},\boldsymbol{W}_\mathrm{d}]$, $(\boldsymbol{W}_\mathrm{p} \in \mathbb{C}^{N\times K_\mathrm{p}}, \boldsymbol{W}_\mathrm{d} \in \mathbb{C}^{N\times K_\mathrm{d}})$. The goal of the JCDE algorithm is to detect the intended TX symbol matrix $\boldsymbol{X}_\mathrm{d}$ and accurately estimate the channel matrix $\boldsymbol{H}$, out of the spatial-temporal RX matrix $\boldsymbol{Y}$ and pilot matrix $\boldsymbol{X}_\mathrm{p}$.

B. Channel Estimation Using Spatial Filtering

In this subsection, we briefly review basic channel estimation techniques based on spatial filtering using $\boldsymbol{X}_\mathrm{p}$ to obtain the initial estimates input to the JCDE algorithm.

1) Least square (LS): When using orthogonal pilots ($K_\mathrm{p} \geq M$), the channel estimator based on the LS criterion is given by

$$\hat{\boldsymbol{H}}_\mathrm{LS} = \arg\min_{\dot{\boldsymbol{H}}} \left|\boldsymbol{Y}_\mathrm{p} - \dot{\boldsymbol{H}}\boldsymbol{X}_\mathrm{p}\right|_\mathrm{F}^2 = \boldsymbol{Y}_\mathrm{p}\boldsymbol{X}_\mathrm{p}^\dagger, \quad (3)$$

where $\boldsymbol{X}_\mathrm{p}^\dagger \triangleq \boldsymbol{X}_\mathrm{p}^\mathrm{H}\left[\boldsymbol{X}_\mathrm{p}\boldsymbol{X}_\mathrm{p}^\mathrm{H}\right]^{-1}$ is Moore-Penrose pseudo-inverse of $\boldsymbol{X}_\mathrm{p}$. The linear filtering error is uniformly

superimposed on the estimates; hence, the MSE of each element in $\boldsymbol{H}_{\mathrm{LS}}$ can be expressed as

$$\hat{\psi}_{\mathrm{LS}}^{h} = \frac{1}{MN} \cdot \mathbb{E}_{\boldsymbol{W}_{\mathrm{p}}} \left\{ \left| \hat{\boldsymbol{H}}_{\mathrm{LS}} - \boldsymbol{H} \right|_{\mathrm{F}}^{2} \right\} = \frac{N_0}{E_{\mathrm{s}} K_{\mathrm{p}}}. \quad (4)$$

2) Minimum norm solution (MNS): When using short non-orthogonal pilots ($K_{\mathrm{p}} < M$), we can no longer compute $\boldsymbol{X}_{\mathrm{p}}^{\dagger}$ due to the *underdetermined* condition. Instead, one can use the MNS to find a unique solution, which is given by the Lagrange multiplier method as

$$\hat{\boldsymbol{H}}_{\mathrm{MNS}} = \arg\min_{\dot{\boldsymbol{H}}} \left| \dot{\boldsymbol{H}} \right|_{\mathrm{F}}^{2} \ \text{s.t.} \ \boldsymbol{Y}_{\mathrm{p}} = \dot{\boldsymbol{H}} \boldsymbol{X}_{\mathrm{p}} + \boldsymbol{W}_{\mathrm{p}}$$
$$= \boldsymbol{Y}_{\mathrm{p}} \boldsymbol{X}_{\mathrm{p}}^{\ddagger}, \quad (5)$$

where $\boldsymbol{X}_{\mathrm{p}}^{\ddagger} \triangleq \left[\boldsymbol{X}_{\mathrm{p}}^{\mathrm{H}} \boldsymbol{X}_{\mathrm{p}} \right]^{-1} \boldsymbol{X}_{\mathrm{p}}^{\mathrm{H}}$. Similarly, the MSE of each element is given by

$$\hat{\psi}_{\mathrm{MNS}}^{h}$$
$$= \frac{1}{MN} \cdot \mathbb{E}_{\boldsymbol{W}_{\mathrm{p}}, \boldsymbol{H}} \left\{ \left| \hat{\boldsymbol{H}}_{\mathrm{MNS}} - \boldsymbol{H} \right|_{\mathrm{F}}^{2} \right\}$$
$$= \frac{\phi}{M} \mathrm{tr} \left[(\boldsymbol{G}_{\mathrm{p}} - \boldsymbol{I}_M)^{\mathrm{H}} (\boldsymbol{G}_{\mathrm{p}} - \boldsymbol{I}_M) \right] + \frac{N_0}{M} \mathrm{tr} \left[\left[\boldsymbol{X}_{\mathrm{p}}^{\mathrm{H}} \boldsymbol{X}_{\mathrm{p}} \right]^{-1} \right], \quad (6)$$

where $\boldsymbol{G}_{\mathrm{p}} \triangleq \boldsymbol{X}_{\mathrm{p}} \left[\boldsymbol{X}_{\mathrm{p}}^{\mathrm{H}} \boldsymbol{X}_{\mathrm{p}} \right]^{-1} \boldsymbol{X}_{\mathrm{p}}^{\mathrm{H}}$ is a resolution matrix. From (6), severe estimation error is inevitable in channel estimation based on the MNS owing to the off-diagonal elements in $\boldsymbol{G}_{\mathrm{p}}$, which is caused by the non-orthogonality of $\boldsymbol{X}_{\mathrm{p}}$.
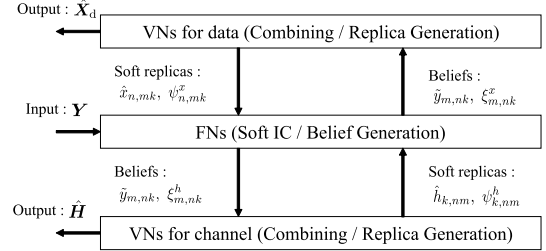
III. BILINEAR GAUSSIAN BELIEF PROPAGATION

This section describes a MUD process based on a BiGaBP mechanism to design JCDE receivers for massive MU-MIMO systems. Fig. 1a shows the tripartite FG consisting of FNs and two VNs, which correspond to the channel coefficients and TX symbols, respectively. The edges between nodes indicate dependencies between corresponding variables, and the information is propagated and exchanged across these edges to perform bilinear inference. In addition, Fig. 1b shows the entities of the propagation information and the direction in which they propagate between each node. As can be seen from these figures, the estimation procedure of the proposed method is performed by exchanging *beliefs* (*i.e.*, likelihood information reflecting estimation reliability) and *soft replicas* (*i.e.*, tentative estimates) on the tripartite FG.

To elaborate, in the FNs, soft interference cancellation (Soft IC) is performed on each RX symbol using the soft replicas of $\boldsymbol{H}$ and $\boldsymbol{X}$ estimated in the previous iteration. Then, the beliefs are computed based on the Soft IC outputs, which are propagated to the VNs. In the VNs, the beliefs from the FNs are combined following the typical BP regime [24] to generate *extrinsic* likelihood information. Then, the soft replicas of $\boldsymbol{H}$ and $\boldsymbol{X}$ are computed on the basis of the conditional expectation, given the beliefs, which are used in the FNs at the next iteration step.
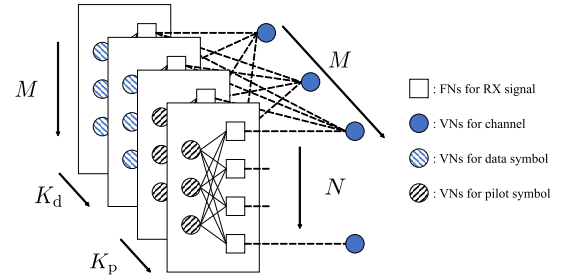
(a) Tripartite FG for BIP: $(N, M, K_{\mathrm{p}}, K_{\mathrm{d}}) = (4, 3, 2, 2)$.

(b) Beliefs exchanged between FNs and VNs.

Fig. 1. Schematic of the belief propagation process in the proposed BiGaBP algorithm.

We emphasize that all belief distributions propagated from FNs to VNs are approximated by a scalar Gaussian distribution based on CLT, so that only information on the mean and variance of the estimated belief distribution is actually required [25]. That is, please note that the entities of the beliefs are their mean and variance values, as shown in Fig. 1b.

For latter convenience, let us define the soft replicas of x_{mk} and h_{nm} as $\{\hat{x}_{n,mk}, \forall n\}$ and $\left\{\hat{h}_{k,nm}, \forall k\right\}$, respectively, such that their MSEs can be respectively expressed as

$$\psi_{n,mk}^{x} \triangleq \mathbb{E}_{\mathsf{x}_{mk}} \left\{ |x_{mk} - \hat{x}_{n,mk}|^{2} \right\}$$
$$= \mathbb{E}_{\tilde{\mathsf{x}}_{n,mk}} \left\{ |\tilde{x}_{n,mk}|^{2} \right\}, \quad \forall n, \quad (7a)$$
$$\psi_{k,nm}^{h} \triangleq \mathbb{E}_{\mathsf{h}_{nm}} \left\{ \left| h_{nm} - \hat{h}_{k,nm} \right|^{2} \right\}$$
$$= \mathbb{E}_{\tilde{\mathsf{h}}_{k,nm}} \left\{ \left| \tilde{h}_{k,nm} \right|^{2} \right\}, \quad \forall k, \quad (7b)$$

where $\tilde{x}_{n,mk} \triangleq x_{mk} - \hat{x}_{n,mk}$ and $\tilde{h}_{k,nm} \triangleq h_{nm} - \hat{h}_{k,nm}$ denote the estimation errors, respectively.

Herein, the JCDE algorithm for estimating x_{mk} and h_{nm} is focused upon. Please note that in practice the following process is performed on all the RX indices in parallel to estimate all channel coefficients and TX symbols simultaneously.

A. FNs: Soft IC and Belief Generation

Let us start with the Soft IC for the (n, k) symbol in $\boldsymbol{Y}$, y_{nk}, with the aid of the soft replicas $\left\{\hat{x}_{n,mk}, \hat{h}_{k,nm}, \forall m\right\}$ generated in the previous iteration step. At the first iteration ($t = 1$), the soft replicas are appropriately initialized. In the detection of an arbitrary TX symbol x_{mk}, the cancellation

process is expressed as

$$
\begin{aligned}
\tilde{y}_{m,nk} &= y_{nk} - \underbrace{\sum_{i \neq m}^{M} \hat{h}_{k,ni} \hat{x}_{n,ik}}_{\text{Inter-UE interference cancellation}} \\
&= \hat{h}_{k,nm} x_{mk} \\
&\quad + \underbrace{\sum_{i \neq m}^{M} \left(h_{ni} x_{ik} - \hat{h}_{k,ni} \hat{x}_{n,ik} \right) + \tilde{h}_{k,nm} x_{mk} + w_{nk}}_{\triangleq \nu_{m,nk}^{x}: \text{ Residual interference plus noise}}.
\end{aligned}
$$
(8)

Under large-system conditions, the residual interference-plus-noise component in (8) can be approximated as a complex Gaussian variable in conformity with CLT; this approximation is referred to as SGA [1]. Accordingly, under an SGA of the effective noise $\nu_{m,nk}^{x}$, the conditional PDF of $\tilde{y}_{m,nk}$, given x_{mk}, can be expressed as

$$
p_{\tilde{y}_{m,nk}|x_{mk}} \left(\tilde{y}_{m,nk}|x_{mk} \right) \propto \exp \left[-\frac{\left| \tilde{y}_{m,nk} - \hat{h}_{k,nm} x_{mk} \right|^2}{\xi_{m,nk}^{x}} \right],
$$
(9)

with

$$
\begin{aligned}
\xi_{m,nk}^{x} &= \mathbb{E}_{\{h_{nm}, \forall m\}, \{x_{ik}, \forall i \neq m\}, w_{nk}|x_{mk}=x_{mk}} \left\{ \left| \nu_{m,nk}^{x} \right|^2 \right\} \\
&= \sum_{i \neq m}^{M} \left\{ \left| \hat{h}_{k,ni} \right|^2 \psi_{n,ik}^{x} + \left(|\hat{x}_{n,ik}|^2 + \psi_{n,ik}^{x} \right) \psi_{k,ni}^{h} \right\} \\
&\quad + |x_{mk}|^2 \psi_{k,nm}^{h} + N_0,
\end{aligned}
$$
(10)

where $|x_{mk}|^2 = E_{\mathrm{s}}$ when the modulation scheme is phase-shift keying (PSK). Even when using quadrature amplitude modulation (QAM) signaling, we may approximately use the true variance, i.e., $|x_{mk}|^2 \approx \mathbb{E}_{x_{mk}} \left\{ |x_{mk}|^2 \right\} = E_{\mathrm{s}}$.

In a similar manner, the estimation of an arbitrary channel coefficient h_{nm} can be obtained. When using QAM, under SGA conditions in (8), the conditional PDF of $\tilde{y}_{m,nk}$, given h_{nm}, can be expressed as

$$
p_{\tilde{y}_{m,nk}|h_{nm}} \left(\tilde{y}_{m,nk}|h_{nm} \right) \propto \exp \left[-\frac{\left| \tilde{y}_{m,nk} - h_{nm} \hat{x}_{n,mk} \right|^2}{\xi_{m,nk}^{h}} \right],
$$
(11)

with

$$
\begin{aligned}
\xi_{m,nk}^{h} &= \sum_{i \neq m}^{M} \left\{ \left| \hat{h}_{k,ni} \right|^2 \psi_{n,ik}^{x} + \left(|\hat{x}_{n,ik}|^2 + \psi_{n,ik}^{x} \right) \psi_{k,ni}^{h} \right\} \\
&\quad + |h_{nm}|^2 \psi_{k,nm}^{x} + N_0,
\end{aligned}
$$
(12)

where the instantaneous channel gain $|h_{nm}|^2$ is not available; hence, we here approximately use the true variance, i.e., $|h_{nm}|^2 \approx \mathbb{E}_{h_{nm}} \left\{ |h_{nm}|^2 \right\} = \phi$.

B. VNs: Belief Combining and Replica Generation

Assuming a high-precision SGA of the effective noise components in $\{\tilde{y}_{m,nk}, \forall n\}$, the beliefs corresponding to x_{mk} are combined over all the RX indices except for the n-th RX index, which results in the extrinsic belief $p_{\hat{r}_{n,mk}|x_{mk}} \left(\hat{r}_{n,mk}|x_{mk} \right)$ for x_{mk}. This is expressed as

$$
\begin{aligned}
p_{\hat{r}_{n,mk}|x_{mk}} \left(\hat{r}_{n,mk}|x_{mk} \right) &= \prod_{i \neq n}^{N} p_{\tilde{y}_{m,ik}|x_{mk}} \left(\tilde{y}_{m,ik}|x_{mk} \right) \\
&\propto \exp \left[-\sum_{i \neq n}^{N} \frac{\left| \tilde{y}_{m,ik} - \hat{h}_{k,im} x_{mk} \right|^2}{\xi_{m,ik}^{x}} \right] \\
&\propto \exp \left[-\frac{|x_{mk} - \hat{r}_{n,mk}|^2}{\psi_{n,mk}^{r}} \right],
\end{aligned}
$$
(13)

where

$$
\hat{r}_{n,mk} = \psi_{n,mk}^{r} \sum_{i \neq n}^{N} \frac{\hat{h}_{k,im}^{*} \tilde{y}_{m,ik}}{\xi_{m,ik}^{x}},
$$
(14a)

$$
\psi_{n,mk}^{r} = \left(\sum_{i \neq n}^{N} \frac{\left| \hat{h}_{k,im} \right|^2}{\xi_{m,ik}^{x}} \right)^{-1}.
$$
(14b)

The extrinsic combining operation in (13) enables significantly suppress the correlation between y_{nk} and $\hat{r}_{n,mk}$ by removing the belief propagated from the n-th FN, $\tilde{y}_{m,nk}$, from the extrinsic belief and decoupling the self-noise regression due to w_{nk}. Consequently, the trapping of the iterative process in a poor local solution can be avoided.

In a similar manner, the extrinsic belief $p_{\hat{q}_{k,nm}|h_{nm}} \left(\hat{q}_{k,nm}|h_{nm} \right)$ for h_{nm} is expressed as

$$
\begin{aligned}
p_{\hat{q}_{k,nm}|h_{nm}} \left(\hat{q}_{k,nm}|h_{nm} \right) &= \prod_{i \neq k}^{K} p_{\tilde{y}_{m,ni}|h_{nm}} \left(\tilde{y}_{m,ni}|h_{nm} \right) \\
&\propto \exp \left[-\frac{|h_{nm} - \hat{q}_{k,nm}|^2}{\psi_{k,nm}^{q}} \right],
\end{aligned}
$$
(15)

where

$$
\hat{q}_{k,nm} = \psi_{k,nm}^{q} \sum_{i \neq k}^{K} \frac{\hat{x}_{n,mi}^{*} \tilde{y}_{m,ni}}{\xi_{m,ni}^{h}},
$$
(16a)

$$
\psi_{k,nm}^{q} = \left(\sum_{i \neq k}^{K} \frac{|\hat{x}_{n,mi}|^2}{\xi_{m,ni}^{h}} \right)^{-1}.
$$
(16b)

Assuming that the effective noise components in $\{\hat{r}_{n,mk}, \forall m, k\}$ are not correlated to each other under SGA conditions, using Bayes' rule, the soft replica of x_{mk} and its MSE can be in general obtained from the symbol-wise conditional expectation, given $\hat{r}_{n,mk}$, as

$$
\begin{aligned}
\hat{x}_{n,mk} &\triangleq \mathbb{E}_{x_{mk}|\hat{r}_{n,mk}=\hat{r}_{n,mk}} \left\{ x_{mk} \right\} \\
&= \sum_{\chi_q \in \mathcal{X}} \frac{\chi_q \cdot p_{\hat{r}_{n,mk}|x_{mk}} \left(\hat{r}_{n,mk}|\chi_q \right) p_{x_{mk}} \left(\chi_q \right)}{\sum_{\chi_q' \in \mathcal{X}} p_{\hat{r}_{n,mk}|x_{mk}} \left(\hat{r}_{n,mk}|\chi_q' \right) p_{x_{mk}} \left(\chi_q' \right)},
\end{aligned}
$$
(17a)

$$
\begin{aligned}
\psi_{n,mk}^{x} \\
&\triangleq \mathbb{E}_{\mathsf{x}_{mk}|\hat{\mathsf{r}}_{n,mk}=\hat{r}_{n,mk}}\left\{|\tilde{x}_{mk}|^{2}\right\} \\
&= \sum_{\chi_q \in \mathcal{X}} \frac{|\chi_q|^{2} \cdot p_{\hat{\mathsf{r}}_{n,mk}|\mathsf{x}_{mk}}\left(\hat{r}_{n,mk}|\chi_q\right) p_{\mathsf{x}_{mk}}\left(\chi_q\right)}{\sum_{\chi'_q \in \mathcal{X}} p_{\hat{\mathsf{r}}_{n,mk}|\mathsf{x}_{mk}}\left(\hat{r}_{n,mk}|\chi'_q\right) p_{\mathsf{x}_{mk}}\left(\chi'_q\right)} - |\hat{x}_{n,mk}|^{2},
\end{aligned}
$$

$$(17b)$$

where the denominators in the summation is introduced for normalization purpose.

In Gray-coded quadrature PSK (QPSK) signaling, *i.e.*, $\mathcal{X} = \{\pm c_x \pm \mathrm{j} c_x\}$, $c_x = \sqrt{E_s/2}$, with $p_{\mathsf{x}_{mk}}(\chi_q) = 1/Q$, $\forall \chi_q \in \mathcal{X}$, (17) can be readily obtained by the following *denoiser* as [42]

$$
\begin{aligned}
\hat{x}_{n,mk} \\
&\triangleq f_{\mathrm{x}}\left(\hat{r}_{n,mk}, \psi_{n,mk}^{r}\right) \\
&= c_x \cdot \left(\tanh\left[2c_x \frac{\Re\{\hat{r}_{n,mk}\}}{\psi_{n,mk}^{r}}\right] + \mathrm{j}\tanh\left[2c_x \frac{\Im\{\hat{r}_{n,mk}\}}{\psi_{n,mk}^{r}}\right]\right),
\end{aligned}
$$

$$(18a)$$

$$
\begin{aligned}
\psi_{n,mk}^{x} \\
&\triangleq V_{\mathrm{x}}\left(\hat{r}_{n,mk}, \psi_{n,mk}^{r}\right) = E_s - |\hat{x}_{n,mk}|^{2}.
\end{aligned}
$$

$$(18b)$$

In a similar manner, the soft replica of h_{nm} can be obtained from the coefficient-wise conditional expectation, given $\hat{q}_{k,nm}$, as

$$
\begin{aligned}
\hat{h}_{k,nm} \\
&\triangleq \mathbb{E}_{\mathsf{h}_{nm}|\hat{\mathsf{q}}_{k,nm}=\hat{q}_{k,nm}}\left\{h_{nm}\right\} \\
&= \int_{h_{nm}} h_{nm} \frac{p_{\hat{\mathsf{q}}_{k,nm}|\mathsf{h}_{nm}}\left(\hat{q}_{k,nm}|h_{nm}\right) p_{\mathsf{h}_{nm}}\left(h_{nm}\right)}{\int_{h'_{nm}} p_{\hat{\mathsf{q}}_{k,nm}|\mathsf{h}_{nm}}\left(\hat{q}_{k,nm}|h'_{nm}\right) p_{\mathsf{h}_{nm}}\left(h'_{nm}\right)}.
\end{aligned}
$$

$$(19)$$

When h_{nm} obeys $\mathcal{CN}(h_{nm}; 0, \phi)$, using the Gaussian-PDF multiplication rule [15], yields

$$
\begin{aligned}
\hat{h}_{k,nm} \\
&= C \int_{h_{nm}} h_{nm} \cdot \mathcal{CN}\left(h_{nm}; \hat{q}_{k,nm}, \psi_{k,nm}^{q}\right) \cdot \mathcal{CN}\left(h_{nm}; 0, \phi\right) \\
&= C' \int_{h_{nm}} h_{nm} \cdot \mathcal{CN}\left(h_{nm}; \frac{\phi \hat{q}_{k,nm}}{\psi_{k,nm}^{q} + \phi}, \frac{\phi \psi_{k,nm}^{q}}{\psi_{k,nm}^{q} + \phi}\right),
\end{aligned}
$$

$$(20)$$

where C and C' are Pearl's normalization constants. From (20), the soft replica of $\hat{h}_{nm}$ and its MSE are expressed as

$$
\hat{h}_{k,nm} \triangleq f_{\mathrm{h}}(\hat{q}_{k,nm}, \psi_{k,nm}^{q}) = \frac{\phi \hat{q}_{k,nm}}{\psi_{k,nm}^{q} + \phi}, \quad (21a)
$$

$$
\psi_{k,nm}^{h} \triangleq V_{\mathrm{h}}(\hat{q}_{k,nm}, \psi_{k,nm}^{q}) = \frac{\phi \psi_{k,nm}^{q}}{\psi_{k,nm}^{q} + \phi}. \quad (21b)
$$

Thus, here the entire processing in the JCDE algorithm via GaBP is completed.

C. Design of ASB for JCDE via BiGaBP

Although the large-system approximation assumption is relaxed in BiGaBP schemes compared to the conventional BiGAMP, the operating principle of BiGaBP still depends on an accurate SGA of residual interference components in (8). When this approximation accuracy is not sufficient due to, *e.g.*, physical limitations of the receiver, mismatches between the SGA and the stochastic behavior of actual effective noise may result in belief *outlier* [25]. This makes it hard to generate accurate soft replicas using the denoiser functions of (18) and (21), causing performance deterioration due to error propagation [43], especially during earlier iterations of the JCDE algorithms and/or in systems with short non-orthogonal pilots.

As a simple and highly effective solution to mitigate such potential issues, belief damping [19], [44], which prevents the algorithm from converging to local minima and belief scaling [25], [26], [45], which controls convergence speed, have been proposed. In this subsection, we propose an extension of ASB [25], [26], a belief scaling method designed for the Bayesian linear inference via low-complexity MP algorithms, to BIPs to further improve the convergence property of the JCDE algorithm via BiGaBP.

1) ASB for Data Detection: The main cause of significant performance deterioration in data detection via the MP algorithms under insufficient large-system conditions is the error propagation of erroneous hard-decision symbols as the soft replicas due to the input of the aforementioned beliefs outliers to the denoiser function of (18a). This is due to the fact that the variance $\psi_{n,mk}^{r}$ in (18a) gives the shape of the optimal denoiser function only when the algorithm is operating with ideal behavior and thus cannot properly handle belief outliers. To address this issue, the ASB introduces a parameter instead of $\psi_{n,mk}^{r}$ to adjust the denoiser function according to the iteration number, while taking into consideration the instantaneous CSI. In QPSK signaling, when using ASB, the denoiser functions of (18) can be replaced with [25]

$$
\begin{aligned}
\hat{x}_{n,mk} \\
&\triangleq f_{\mathrm{ASB}}\left(\hat{r}_{n,mk}, \gamma\right) \\
&= c_x \cdot \left(\tanh\left[\frac{\gamma}{c_x}\Re\{\hat{r}_{n,mk}\}\right] + \mathrm{j}\tanh\left[\frac{\gamma}{c_x}\Im\{\hat{r}_{n,mk}\}\right]\right),
\end{aligned}
$$

$$(22a)$$

$$
\begin{aligned}
\psi_{n,mk}^{x} \\
&\triangleq V_{\mathrm{ASB}}\left(\hat{r}_{n,mk}, \gamma\right) = E_s - |\hat{x}_{n,mk}|^{2},
\end{aligned}
$$

$$(22b)$$

where γ is the scaling parameter. According to [25], the scaling parameter is designed to be a monotonic increasing function of the number of iterations, *i.e.*, $\gamma(t) = \tau_0 + \tau_1 \cdot \left(\frac{t}{T}\right)$, where (t) indicates the variable at the t-th iteration step, T is the maximum number of iterations, and (τ_0, τ_1) are the predetermined parameters.

Fig. 2 shows the dynamics of $f_{\mathrm{ASB}}\left(\hat{r}_{n,mk}, \gamma\right)$ as a function of $\hat{r}_{n,mk}$ with different scaling parameters, which indicates that the scaling parameter changes the *slope* of the denoiser function, instead of $\psi_{n,mk}^{r}$ in (18a), enabling adjustment of the iterative convergence speed of the algorithm. In the early
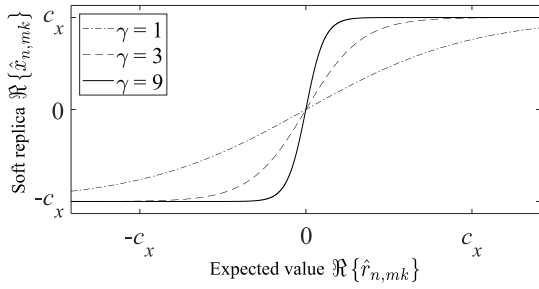
Fig. 2. The dynamics of $f_{\mathrm{ASB}}(\cdot)$ with $\gamma = 1, 3$, and 9, as a function of $\hat{r}_{n,mk}$, in Gray-coded QPSK signaling.

iterations, γ is set to lower values to prevent convergence to local minima due to the erroneous hard-decision symbols, and in the later iterations, γ is set to higher values to facilitate the convergence of the algorithm.

2) ASB for Channel Estimation: Next, we consider how ASB for channel estimation should be designed. From (22a), it is found that the scaling parameter in ASB is multiplied by the log-likelihood, *i.e.*, $\hat{r}_{n,mk}$, with respect to the detected symbol. In addition, belief combining for channel estimation is performed for discrete-time indices $k \in \mathcal{K}$ as shown in (15); however, it is not reasonable to multiply the beliefs propagated from the discrete-time indices corresponding to the pilot symbols and to those corresponding to the data symbols, by the same scaling parameter. This is because they should have different statistical properties throughout the iterative process, as described later.

Based on the above, a straight extension of ASB for data detection to channel estimation would replace the extrinsic belief in (15) with

$$p_{\check{\mathsf{q}}_{k,nm}|\mathsf{h}_{nm}}\left(\check{q}_{k,nm}|h_{nm}\right)$$

$$= \left(\prod_{i \in \mathcal{K}_{\mathrm{p}} \backslash k} \left[p_{\check{\mathsf{y}}_{m,ni}|\mathsf{h}_{nm}}\left(\tilde{y}_{m,ni}|h_{nm}\right) \right] \right)^{\alpha}$$

$$\times \left(\prod_{i \in \mathcal{K}_{\mathrm{d}} \backslash k} \left[p_{\check{\mathsf{y}}_{m,ni}|\mathsf{h}_{nm}}\left(\tilde{y}_{m,ni}|h_{nm}\right) \right] \right)^{\beta}$$

$$\propto \exp\left[-\frac{|h_{nm} - \check{q}_{k,nm}|^2}{\check{\psi}^q_{k,nm}} \right], \tag{23}$$

with

$$\check{q}_{k,nm}$$

$$= \check{\psi}^q_{k,nm} \cdot \left(\alpha \sum_{i \in \mathcal{K}_{\mathrm{p}} \backslash k} \frac{x^*_{mi} \tilde{y}_{m,ni}}{\xi^h_{m,ni}} + \beta \sum_{i \in \mathcal{K}_{\mathrm{d}} \backslash k} \frac{\hat{x}^*_{n,mi} \tilde{y}_{m,ni}}{\xi^h_{m,ni}} \right),$$

$$\tag{24a}$$

$$\check{\psi}^q_{k,nm}$$

$$= \left(\alpha \sum_{i \in \mathcal{K}_{\mathrm{p}} \backslash k} \frac{|x_{mi}|^2}{\xi^h_{m,ni}} + \beta \sum_{i \in \mathcal{K}_{\mathrm{d}} \backslash k} \frac{|\hat{x}_{n,mi}|^2}{\xi^h_{m,ni}} \right)^{-1}, \tag{24b}$$

where α and β are scaling parameters corresponding to the pilot and data parts, respectively. From (24), α and β are responsible for adjusting the ratio of the beliefs propagated

from the pilot and data parts, and (23) coincides with (15) if and only if $\alpha = \beta = 1$. It is also worth mentioning that when $\alpha = \beta$, $\check{q}_{k,nm}$ in (24a) coincides with $\hat{q}_{k,nm}$ in (16), and only $\psi^q_{k,nm}$ is multiplied by the scaling parameter, *i.e.*, $\check{\psi}^q_{k,nm} = \alpha \psi^q_{k,nm}$, which is equivalent to the operation to change the *slope* in the denoiser function of $f_{\mathrm{h}}(\hat{q}_{k,nm}, \check{\psi}^q_{k,nm})$ as well as the ASB mechanism for data detection adjusting the equivalent variance of extrinsic beliefs, as can be inferred from (22). Since $f_{\mathrm{h}}(\hat{q}_{k,nm}, \check{\psi}^q_{k,nm})$ is a simple linear function of $\hat{q}_{k,nm}$, it is not suitable for adjusting the iterative convergence speed by the scaling parameters taking advantage of the non-linearity of denoiser function. Instead, we can utilize the scaling parameters, α and β, as weight parameters to stabilize the convergence behavior of iterative estimation in the early iteration steps.

From (24), the beliefs propagated from the VNs corresponding to the pilot symbols contain only the estimated channel coefficient uncertainty, while the beliefs propagated from the VNs corresponding to the data symbols contain both the estimated channel coefficient and estimated data symbol uncertainty. In other words, the reliability of the former beliefs is much higher than that of the latter beliefs, especially in the early iteration steps when the reliability of detected symbols, *i.e.*, soft replicas of $\{x_{mk}, \forall(m, k)\}$, is low.

In light of the above, it may be possible to suppress the divergence behavior of iterative estimation by setting α and β to large and small values, respectively, in the early iteration steps. As the iterative process increases the reliability of the soft replicas, gradually decreasing the value of α and gradually increasing the value of β can facilitate the iterative convergence of the algorithm; thus, it is reasonable for the dynamics of α and β to be monotonically decreasing and monotonically increasing functions, respectively, with respect to the number of iterations. In addition, the dynamics of α and β should satisfy the following two requirements: a) $\alpha = \beta = 1$ at the final iteration step so as to converge to the fixed point of the original BiGaBP algorithm, and b) $K_{\mathrm{p}} + K_{\mathrm{d}} = \alpha K_{\mathrm{p}} + \beta K_{\mathrm{d}}$ so as to change the weights between the pilot and data parts without varying the scale of the extrinsic belief.

Based on the above, if a simple linear function is chosen as the monotonic function, the dynamics of scaling parameters is given by a function of the number of iterations as

$$\alpha(t) = 1 + (1 - \beta(t)) \cdot \left(\frac{K_{\mathrm{d}}}{K_{\mathrm{p}}} \right), \tag{25a}$$

$$\beta(t) = \tau_2 + (1 - \tau_2) \cdot \left(\frac{t}{T} \right), \tag{25b}$$

where $0 \le \tau_2 \le 1$ is the predetermined parameter to determine the initial value of β, *i.e.*, $\beta(0)$. The operation of ASB for channel estimation can be interpreted as a *soft* scheduling strategy for belief update, where the reliability of the beliefs is sequentially enhanced from the nodes near the VNs corresponding to the pilot symbols to the whole of the FG via iterative processes, leading to stabilization of the convergence behavior of BiGaBP.

D. Algorithmic Description

The pseudo-code of the JCDE algorithm via BiGaBP with belief damping and scaling is given in Alg. 1. Besides the RX matrix $\boldsymbol{Y}$ and the pilot matrix $\boldsymbol{X}_{\mathrm{p}}$, the algorithm requires the initial estimates of the channel coefficients $\hat{\boldsymbol{H}}$ and their MSEs $\hat{\psi}^h$, which are given by the spatial filtering based on MNS or LS criteria according to the pilot length, as described in Section II-B, outputting the hard-decision estimates of the TX matrix and the estimates of the channel coefficients. The belief damping [15], [19], [44] is also introduced in lines 11, 12, 15, and 16 to prevent the convergence to local minima by averaging the current beliefs based on the past information, where $\eta \in [0, 1]$ is a damping factor. Unlike [15], the introduction is changed to apply the damping process only once for each belief, to be consistent with the MP rule.

In addition, as the most vital algorithmic structural difference between BiGaBP and conventional BiGAMP, lines 9, 10, 13, and 14 corresponding to the self-feedback removal mechanism via extrinsic belief generation are highlighted with gray boxes in Alg. 1.

IV. DERIVATION OF BIGAMP FROM BIGABP

In this section, we prove that the BiGaBP algorithm presented in Section III can be rigorously approximated under the large-system limit assumption, *i.e.*, $N, M, K \to \infty$ for a given compression ratio $\rho_1 \triangleq N/M, \rho_2 \triangleq K/M$, to derive the BiGAMP algorithm proposed in [16]. We then compare these algorithms to the original BiGAMP algorithm proposed in [15] and elaborate on the differences in the algorithmic structure. It is worth noting here that in both references [15] and [16], the BiGAMP algorithm is derived by approximating the linear inference via SPA in the large-system limit; however, the derivation process from the BiGaBP algorithm, *i.e.*, the bilinear inference algorithm designed based on the GaBP framework [24], has not yet been reported explicitly, to the best of our knowledge.

Since the process of data detection and channel estimation is the same except for the soft replica generation, the approximation process is described mainly using the data detection part of the BiGaBP algorithm. To make the process across iterations easier to understand, (t) is used in this section. The derivation process of BiGAMP from BiGaBP consists of three processes: a) approximation of the second-order moments, b) approximation of the first-order moments, and c) closing the loop of the algorithm.

A. Approximation of Second-Order Moments

Let us consider the process at the t-th iteration step. First, the variance $\xi^x_{m,nk}(t)$ in (10) can be rewritten as (26), shown at

the bottom of the page, where the third term in (26) converges almost surely to zero as $N \to \infty$ in the asymptotic conditions. Similar to (26), the variance $\psi^r_{n,mk}(t)$ in (14) can be rewritten as

$$\frac{1}{\psi^r_{n,mk}(t)} = \underbrace{\sum_{i=1}^{N} \frac{\left|\hat{h}_{k,im}(t)\right|^2}{\xi^x_{ik}(t)}}_{\triangleq 1/\psi^r_{mk}(t)} + \mathcal{O}(N^{-1}). \quad (27)$$

From the above, the variance, *i.e.* the second-order moment, can be approximated with the accuracy of $\mathcal{O}(1)$, removing the dependence of the index n or m, respectively.

B. Approximation of First-Order Moments

Next, using (26) and (27), the first-order moments is approximated with the accuracy of $\mathcal{O}(1)$. The Soft IC output $\tilde{y}_{m,nk}(t)$ in (8) can be rewritten as

$$\tilde{y}_{m,nk}(t) = y_{nk} - \sum_{i=1}^{M} \hat{h}_{k,ni}(t)\hat{x}_{n,ik}(t) + \hat{h}_{k,ni}(t)\hat{x}_{n,ik}(t)$$

$$= \underbrace{y_{nk}(t) - \hat{p}_{nk}(t)}_{\triangleq \tilde{y}_{nk}(t)} + \hat{h}_{k,ni}(t)\hat{x}_{n,ik}(t), \quad (28)$$

with the soft replica of interference components,

$$\hat{p}_{nk}(t) \triangleq \sum_{i=1}^{M} \hat{h}_{k,ni}(t)\hat{x}_{n,ik}(t). \quad (29)$$

Substituting (28) into (14), the expected value $\hat{r}_{n,mk}(t)$ can be rewritten as

$$\hat{r}_{n,mk}(t) = \psi^r_{mk}(t) \sum_{i \neq n}^{N} \frac{\hat{h}^*_{k,im}(t)\tilde{y}_{m,ik}(t)}{\xi^x_{ik}(t)}$$

$$= \hat{r}_{mk}(t) - \psi^r_{mk}(t) \frac{\hat{h}^*_{k,nm}(t)\tilde{y}_{nk}(t)}{\xi^x_{nk}(t)} + \mathcal{O}(N^{-1}), \quad (30)$$

with

$$\hat{r}_{mk}(t) \triangleq \psi^r_{mk}(t) \sum_{i=1}^{N} \frac{|\hat{h}^*_{k,im}(t)|^2 \hat{x}_{i,mk}(t)}{\xi^x_{ik}(t)} + \psi^r_{mk}(t) \sum_{i=1}^{N} \frac{\hat{h}^*_{k,im}(t)\tilde{y}_{ik}(t)}{\xi^x_{ik}(t)}, \quad (31)$$

where the second term in (30) is the component that is removed in the extrinsic belief combining in the BiGaBP algorithm.

Let us see how this self-feedback component is handled in the BiGAMP algorithm. Consider a Taylor series expansion of

$$\xi^x_{m,nk}(t) = \underbrace{\sum_{i=1}^{M}\left\{\left|\hat{h}_{k,ni}(t)\right|^2 \psi^x_{n,ik}(t) + \left(|\hat{x}_{n,ik}(t)|^2 + \psi^x_{n,ik}(t)\right)\psi^h_{k,ni}(t)\right\} + N_0}_{\triangleq \xi^x_{nk}(t)} + \mathcal{O}(N^{-1}), \quad (26)$$

Algorithm 1 JCDE Algorithm via BiGaBP With Belief Damping and Scaling

---

Input: $\boldsymbol{Y}$, $\boldsymbol{X}_{\mathrm{p}}$, $\hat{\boldsymbol{H}}$, $\hat{\psi}^h$, T
Output: $\{\hat{x}_{mk}, \forall(k \in \mathcal{K}_{\mathrm{d}}, m)\}$, $\{\hat{h}_{nm}, \forall(m,n)\}$
1: $\forall(k \in \mathcal{K}_{\mathrm{p}}, m, n, t) : \hat{x}_{n,mk}(t) = [\boldsymbol{X}_{\mathrm{p}}]_{mk}$, $\psi^x_{n,mk}(t) = 0$
2: $\forall(k \in \mathcal{K}_{\mathrm{d}}, m, n) : \hat{x}_{n,mk}(1) = 0$, $\psi^x_{n,mk}(1) = E_{\mathrm{s}}$
3: $\forall(k, m, n) : \hat{h}_{k,nm}(1) = \left[\hat{\boldsymbol{H}}\right]_{nm}$, $\psi^h_{k,nm}(1) = \hat{\psi}^h$
4: **for** $t = 1$ to T **do**
   /* FNs */
5: $\forall(k, m, n) : \tilde{y}_{m,nk}(t) = y_{nk} - \sum_{i \neq m}^M \hat{h}_{k,ni}(t)\hat{x}_{n,ik}(t)$
6: $\forall(k, m, n) : \xi^y_{m,nk}(t) = \sum_{i \neq m}^M \left\{ |\hat{h}_{k,ni}(t)|^2 \psi^x_{n,ik}(t) \right.$
      $\left. + \psi^h_{k,ni}(t) \left(|\hat{x}_{n,ik}(t)|^2 + \psi^x_{n,ik}(t) \right) \right\} + N_0$
7: $\forall(k \in \mathcal{K}_{\mathrm{d}}, m, n) : \xi^x_{m,nk}(t) = \xi^y_{m,nk}(t) + E_{\mathrm{s}} \psi^h_{k,nm}(t)$
8: $\forall(k, m, n) : \xi^h_{m,nk}(t) = \xi^y_{m,nk}(t) + \phi \psi^x_{n,mk}(t)$
   /* VNs */
9: $\forall(k \in \mathcal{K}_{\mathrm{d}}, m, n) : \psi^r_{n,mk}(t) = \left(\sum_{i \neq n}^N \frac{|\hat{h}_{k,im}(t)|^2}{\xi^x_{m,ik}(t)} \right)^{-1}$

10: $\forall(k \in \mathcal{K}_{\mathrm{d}}, m, n) : \hat{r}_{n,mk}(t)$
      $= \psi^r_{n,mk}(t) \sum_{i \neq n}^N \frac{\hat{h}^*_{k,im}(t)\tilde{y}_{m,ik}(t)}{\xi^x_{m,ik}(t)}$

11: $\forall(k \in \mathcal{K}_{\mathrm{d}}, m, n) : \hat{x}_{n,mk}(t+1)$
      $= \eta \cdot f_{\mathrm{ASB}}(\hat{r}_{n,mk}(t), \gamma(t)) + (1-\eta) \cdot \hat{x}_{n,mk}(t)$
12: $\forall(k \in \mathcal{K}_{\mathrm{d}}, m, n) : \psi^x_{n,mk}(t+1)$
      $= \eta \cdot V_{\mathrm{ASB}}(\hat{r}_{n,mk}(t), \gamma(t)) + (1-\eta) \cdot \psi^x_{n,mk}(t)$

13: $\forall(k, m, n) : \check{\psi}^q_{k,nm}(t)$
      $= \left(\alpha(t) \sum_{i \in \mathcal{K}_{\mathrm{p}} \backslash k} \frac{|x_{mi}(t)|^2}{\xi^h_{m,ni}(t)} + \beta(t) \sum_{i \in \mathcal{K}_{\mathrm{d}} \backslash k} \frac{|\hat{x}_{n,mi}(t)|^2}{\xi^h_{m,ni}(t)} \right)^{-1}$

14: $\forall(k, m, n) : \check{q}_{k,nm}(t)$
      $= \check{\psi}^q_{k,nm}(t) \cdot \left(\alpha(t) \sum_{i \in \mathcal{K}_{\mathrm{p}} \backslash k} \frac{x^*_{mi}(t)\tilde{y}_{m,ni}(t)}{\xi^h_{m,ni}(t)} \right.$
      $\left. + \beta(t) \sum_{i \in \mathcal{K}_{\mathrm{d}} \backslash k} \frac{\hat{x}^*_{n,mi}(t)\tilde{y}_{m,ni}(t)}{\xi^h_{m,ni}(t)} \right)$

15: $\forall(k, m, n) : \hat{h}_{k,nm}(t+1)$
      $= \eta \cdot f_{\mathrm{h}}(\check{q}_{k,nm}(t), \check{\psi}^q_{k,nm}(t)) + (1-\eta) \cdot \hat{h}_{k,nm}(t)$
16: $\forall(k, m, n) : \psi^h_{k,nm}(t+1)$
      $= \eta \cdot V_{\mathrm{h}}(\check{q}_{k,nm}(t), \check{\psi}^q_{k,nm}(t)) + (1-\eta) \cdot \psi^h_{k,nm}(t)$
17: **if** $\sum_{k,n,m} |\hat{h}_{k,nm}|^2 / (K \cdot N \cdot M) \geqq \tau_{\mathrm{BiGaBP}} \cdot \phi$ **then** $T \leftarrow t$
      and **break** ▷ Stop criterion
18: **end if**
19: **end for**
   /* Decision */
20: $\forall(k \in \mathcal{K}_{\mathrm{d}}, m) : \psi^r_{mk} = \left(\sum_{i=1}^N \frac{|\hat{h}_{k,im}(T)|^2}{\xi^x_{m,ik}(T)} \right)^{-1}$
21: $\forall(k \in \mathcal{K}_{\mathrm{d}}, m) : \hat{x}_{mk}$
      $= \arg\max_{\chi_q \in \mathcal{X}} \left| \chi_q - \psi^r_{mk} \sum_{i=1}^N \frac{\hat{h}^*_{k,im}(T)\tilde{y}_{m,ik}(T)}{\xi^x_{m,ik}(T)} \right|$
22: $\forall(m, n) : \check{\psi}^q_{nm} = \left(\sum_{i=1}^K \frac{|\hat{x}_{mi}(T)|^2}{\xi^h_{m,ni}(T)} \right)^{-1}$
23: $\forall(m, n) : \check{q}_{nm} = \check{\psi}^q_{nm} \cdot \left(\sum_{i=1}^K \frac{\hat{x}^*_{mi}(T)\tilde{y}_{m,ni}(T)}{\xi^h_{m,ni}(T)} \right)$
24: $\forall(m, n) : \hat{h}_{nm} = f_{\mathrm{h}}(\check{q}_{nm}, \check{\psi}^q_{nm})$

---

the denoiser function in (18a) about the point $\hat{r}_{mk}(t)$, the soft replica used in the next iteration $\hat{x}_{n,mk}(t+1)$ can be expressed as

$$
\begin{aligned}
&\hat{x}_{n,mk}(t+1) \\
&= f_{\mathrm{x}}\left(\hat{r}_{n,mk}(t), \psi^r_{n,mk}(t)\right) \\
&= f_{\mathrm{x}}\left(\hat{r}_{mk}(t), \psi^r_{mk}(t)\right) + \mathcal{O}(N^{-1}) \\
&\quad - \psi^r_{mk}(t) \frac{\partial f_{\mathrm{x}}\left(\hat{r}_{mk}(t), \psi^r_{mk}(t)\right)}{\partial \hat{r}_{mk}(t)} \frac{\hat{h}^*_{k,nm}(t)\tilde{y}_{nk}(t)}{\xi^x_{nk}(t)} \\
&\quad - \psi^r_{mk}(t) \frac{\partial f_{\mathrm{x}}\left(\hat{r}_{mk}(t), \psi^r_{mk}(t)\right)}{\partial \hat{r}^*_{mk}(t)} \frac{\hat{h}_{k,nm}(t)\tilde{y}^*_{nk}(t)}{\xi^x_{nk}(t)}.
\end{aligned} \tag{32}
$$

When using the Bayes-optimal denoiser function as $f_{\mathrm{x}}(\cdot)$, *i.e.*, the denoiser function in (17a), (32) can be rewritten as

$$
\begin{aligned}
&\hat{x}_{n,mk}(t+1) \\
&= \hat{x}_{mk}(t+1) - \psi^x_{mk}(t+1) \frac{\hat{h}^*_{k,nm}(t)\tilde{y}_{nk}(t)}{\xi^x_{nk}(t)} + \mathcal{O}(N^{-1}),
\end{aligned} \tag{33}
$$

where $\hat{x}_{mk}(t+1) \triangleq f_{\mathrm{x}}(\hat{r}_{mk}(t), \psi^r_{mk}(t))$ and $\psi^x_{mk}(t+1) \triangleq V_{\mathrm{x}}(\hat{r}_{mk}(t), \psi^r_{mk}(t))$. The second term in (33) denotes the self-feedback component propagating to the next iteration step through the denoiser function. Note that (32) can be rewritten in a simpler form as in (33) if and only if the denoiser function is designed according to the Bayes optimal criterion so that the identity between the expectation and variance holds [16]; otherwise, the Wirtinger derivatives in equation (32) must be calculated, as in [46]. In a similar manner, $\hat{h}_{k,nm}$ can be rewritten as

$$
\begin{aligned}
&\hat{h}_{k,nm}(t+1) \\
&= \hat{h}_{nm}(t+1) - \psi^h_{nm}(t+1) \frac{\hat{x}^*_{n,mk}(t)\tilde{y}_{nk}(t)}{\xi^h_{nk}(t)} + \mathcal{O}(N^{-1}),
\end{aligned} \tag{34}
$$

where $h_{nm}(t+1) \triangleq f_{\mathrm{h}}(\hat{q}_{nm}(t), \psi^q_{nm}(t))$ and $\psi^h_{nm}(t+1) \triangleq V_{\mathrm{h}}(\hat{q}_{nm}(t), \psi^q_{nm}(t))$.

C. Closing Loop

Finally, the loop is closed by replacing $\hat{x}_{n,mk}$ and $\hat{h}_{k,nm}$ with $\hat{x}_{mk}$ and $\hat{h}_{nm}$, respectively, throughout the MP rule. From (33) and (34),

$$
\begin{aligned}
&\hat{x}_{n,mk}(t+1) \\
&= \hat{x}_{mk}(t+1) - \psi^x_{mk}(t+1) \frac{\hat{h}^*_{nm}(t)\tilde{y}_{nk}(t)}{\xi^x_{nk}(t)} + \mathcal{O}(N^{-1}),
\end{aligned} \tag{35}
$$

$$
\begin{aligned}
&\hat{h}_{k,nm}(t+1) \\
&= \hat{h}_{nm}(t+1) - \psi^h_{nm}(t+1) \frac{\hat{x}^*_{mk}(t)\tilde{y}_{nk}(t)}{\xi^h_{nk}(t)} + \mathcal{O}(N^{-1}),
\end{aligned} \tag{36}
$$

and $\psi^x_{n,mk}(t) = \psi^x_{mk}(t) + \mathcal{O}(N^{-1})$ and $\psi^h_{k,nm}(t) = \psi^h_{nm}(t) + \mathcal{O}(N^{-1})$.

Substituting (35) and (36) into (26), $\xi^x_{nk}(t)$ can be rewritten as

$$
\xi^x_{nk}(t)
$$

$$= \sum_{i=1}^{M} \left\{ \left| \hat{h}_{ni}(t) \right|^2 \psi_{ik}^x(t) + \left(|\hat{x}_{ik}(t)|^2 + \psi_{ik}^x(t) \right) \psi_{ni}^h(t) \right\}$$
$$+ N_0 + \mathcal{O}(N^{-\frac{1}{2}}), \tag{37}$$

Similarly, $\psi_{mk}^r(t)$ in (27) can be rewritten as

$$\frac{1}{\psi_{mk}^r(t)} = \sum_{i=1}^{N} \frac{\left| \hat{h}_{im}(t) \right|^2}{\xi_{ik}^x(t)} + \mathcal{O}(N^{-\frac{1}{2}}). \tag{38}$$

Substituting (35) and (36) into (29), $\hat{p}_{nk}(t)$ can be rewritten as (39), shown at the bottom of the next page, where the third term in (39) is the *Onsager* correction term, which is found to be responsible for predicting and canceling the self-feedback component in the large-system limit assumption, from the derivation process so far.

Similarly, substituting (35) and (36) into (31), $\hat{r}_{mk}(t)$ can be rewritten as

$$\hat{r}_{mk}(t) = \hat{x}_{mk}(t) + \psi_{mk}^r(t) \sum_{i=1}^{N} \frac{\hat{h}_{k,im}^*(t) \tilde{y}_{ik}(t)}{\xi_{ik}^x(t)} + \mathcal{O}(N^{-1/2})$$
$$= \hat{x}_{mk}(t) + \psi_{mk}^r(t) \sum_{i=1}^{N} \frac{\hat{h}_{im}^*(t) \tilde{y}_{ik}(t)}{\xi_{ik}^x(t)} + \mathcal{O}(N^{-1/2})$$
$$\underbrace{- \psi_{mk}^r(t) \hat{x}_{mk}(t-1) \sum_{i=1}^{N} \frac{\psi_{im}^h(t) \tilde{y}_{ik}(t) \tilde{y}_{ik}^*(t-1)}{\xi_{ik}^x(t) \xi_{ik}^h(t-1)}}_{\text{Onsager correction term}},$$
$$\tag{40}$$

where the fourth term in (40) is the *Onsager* correction term for the VNs. The above discussion is equally applicable for the channel estimation part.

Under the large-system limit assumption, the derivation of the JCDE algorithm via BiGAMP proposed in [16] is completed by asymptotically setting the remainder terms to zero. The pseudo-code is given in Alg. 2, where belief damping is introduced in lines 14–15 and 19–20 as in [16].

D. Differences in Algorithmic Structure From BiGAMP [15]

Finally, we clarify the differences between the JCDE algorithm via the original BiGAMP proposed in [15] and JCDE algorithm via BiGAMP [16] derived above, in terms of algorithmic structure. Due to space limitations, the derivation of the original BiGAMP algorithm [15] is omitted and offered only in a summarized form in the pseudo-code of Alg. 3. Focusing on the points that differ from Alg. 2, first, line 8 in Alg. 3 is the Onsager correction term, which is obtained by approximating (39) as $|\hat{x}_{mk}(t)|^2 \approx \hat{x}_{mk}(t) \hat{x}_{mk}^*(t-1)$ and $|\hat{h}_{nm}(t)|^2 \approx \hat{h}_{nm}(t) \hat{h}_{nm}^*(t-1)$. Next, in lines 14 and 19, the output of the ramp function is used to calculate the equivalent gains of the soft replicas in the VNs, which is derived as a result of approximating the expectation and variance of the beliefs with different precision. Therefore, if the large-system limit assumptions are not tightly satisfied, these mathematical inconsistencies cause stochastic misalignments between the belief expectation and variance, leading to unstable convergence behavior of iterative estimation. Finally,

---

Algorithm 2 JCDE Algorithm via BiGAMP [16]

Input: $\mathbf{Y}, \mathbf{X}_\mathrm{p}, \hat{\mathbf{H}}, \hat{\psi}^h, T$
Output: $\{\hat{x}_{mk}, \forall(k \in \mathcal{K}_\mathrm{d}, m)\}, \{\hat{h}_{nm}, \forall(m, n)\}$
1: $\forall(k \in \mathcal{K}_\mathrm{p}, m, t): \hat{x}_{mk}(t) = [\mathbf{X}_\mathrm{p}]_{mk}, \ \psi_{mk}^x(t) = 0$
2: $\forall(k \in \mathcal{K}_\mathrm{d}, m): \hat{x}_{mk}(1) = 0, \ \psi_{mk}^x(1) = E_\mathrm{s}$
3: $\forall(m, n): \hat{h}_{nm}(1) = \left[\hat{\mathbf{H}} \right]_{nm}, \ \psi_{nm}^h(1) = \hat{\psi}^h$
4: $\forall(k, n): \hat{s}_{nk}(0) = 0$
5: **for** $t = 1$ to T **do**
   /* FNs */
6: $\forall(k, n): \overline{\xi}_{nk}^p(t) = \sum_{i=1}^{M} \Big\{ \psi_{ni}^h(t) \hat{x}_{ik}(t) \hat{x}_{ik}^*(t-1)$
                           $+ \hat{h}_{ni}(t) \hat{h}_{ni}^*(t-1) \psi_{ik}^x(t) \Big\}$
7: $\forall(k, n): \xi_{nk}^p(t) = \sum_{i=1}^{M} \Big\{ |\hat{h}_{ni}(t)|^2 \psi_{ik}^x(t)$
                           $+ \psi_{ni}^h(t) \left(|\hat{x}_{ik}(t)|^2 + \psi_{ik}^x(t) \right) \Big\}$
8: $\forall(k, n): \hat{p}_{nk}(t) = \sum_{i=1}^{M} \hat{h}_{ni}(t) \hat{x}_{ik}(t) - \overline{\xi}_{nk}^p(t) \hat{s}_{nk}(t-1)$
9: $\forall(k, n): \xi_{nk}^s(t) = \frac{1}{\xi_{nk}^p(t) + N_0}$
10: $\forall(k, n): \hat{s}_{nk}(t) = \frac{y_{nk} - \hat{p}_{nk}(t)}{\xi_{nk}^p(t) + N_0}$
   /* VNs */
11: $\forall(k \in \mathcal{K}_\mathrm{d}, m): \psi_{mk}^r(t) = \left(\sum_{i=1}^{N} |\hat{h}_{im}(t)|^2 \xi_{ik}^s(t) \right)^{-1}$
12: $\forall(k \in \mathcal{K}_\mathrm{d}, m): \overline{r}_{mk}(t)$
                  $= \hat{x}_{mk}(t) + \psi_{mk}^r(t) \sum_{i=1}^{N} \hat{h}_{im}^*(t) \hat{s}_{ik}(t)$
13: $\forall(k \in \mathcal{K}_\mathrm{d}, m): \hat{r}_{mk}(t)$
                  $= \overline{r}_{mk}(t) - \psi_{mk}^r(t) \hat{x}_{mk}(t-1) \sum_{i=1}^{N} \psi_{im}^h(t) \hat{s}_{ik}(t) \hat{s}_{ik}^*(t-1)$
14: $\forall(k \in \mathcal{K}_\mathrm{d}, m): \hat{x}_{mk}(t+1)$
                  $= \eta \cdot f_\mathrm{x}(\hat{r}_{mk}(t), \psi_{mk}^r(t)) + (1-\eta) \cdot \hat{x}_{mk}(t)$
15: $\forall(k \in \mathcal{K}_\mathrm{d}, m): \psi_{mk}^x(t+1)$
                  $= \eta \cdot V_\mathrm{x}(\hat{r}_{mk}(t), \psi_{mk}^r(t)) + (1-\eta) \cdot \psi_{mk}^x(t)$
16: $\forall(m, n): \psi_{nm}^q(t) = \left(\sum_{i=1}^{K} |\hat{x}_{mi}(t)|^2 \xi_{ni}^s(t) \right)^{-1}$
17: $\forall(m, n): \overline{q}_{nm}(t) = \hat{h}_{nm}(t) + \psi_{nm}^q(t) \sum_{i=1}^{K} \hat{x}_{mi}^*(t) \hat{s}_{ni}(t)$
18: $\forall(m, n): \hat{q}_{nm}(t)$
                  $= \overline{q}_{nm}(t) - \psi_{nm}^q(t) \hat{h}_{nm}(t-1) \sum_{i=1}^{K} \psi_{mi}^x(t) \hat{s}_{ni}(t) \hat{s}_{ni}^*(t-1)$
19: $\forall(m, n): \hat{h}_{nm}(t+1)$
                  $= \eta \cdot f_\mathrm{h}(\hat{q}_{nm}(t), \psi_{nm}^q(t)) + (1-\eta) \cdot \hat{h}_{nm}(t)$
20: $\forall(m, n): \psi_{nm}^h(t+1)$
                  $= \eta \cdot V_\mathrm{h}(\hat{q}_{nm}(t), \psi_{nm}^q(t)) + (1-\eta) \cdot \psi_{nm}^h(t)$
21: **if** $\sum_{n,m} |\hat{h}_{nm}|^2/(N \cdot M) \geqq \tau_{\mathrm{BiGAMP}} \cdot \phi$ **then** $T \leftarrow t$ and
   break   ▷ Stop criterion
22: **end if**
23: **end for**
24: $\forall(k \in \mathcal{K}_\mathrm{d}, m): \hat{x}_{mk} = \arg \min_{\chi_q \in \mathcal{X}} |\chi_q - \hat{r}_{mk}(T)|$
25: $\forall(m, n): \hat{h}_{nm} = f_\mathrm{h}(\hat{q}_{nm}(T), \hat{\psi}_{nm}^q(T))$

---

the biggest difference in algorithmic structure is the absence of the Onsager correction term in the VNs in Alg. 3 to make the algorithm structure in the VNs closer to the linear inference algorithm via GAMP [18]. However, as described above, the Onsager correction term is intended to cancel the self-feedback, and the original BiGAMP algorithm without this mechanism cannot completely decouple the self-feedback of beliefs across iterations between the VNs. This is the main reason why the BiGAMP algorithm proposed in [16]

Algorithm 3 JCDE Algorithm via BiGAMP [15]

Input: $\boldsymbol{Y}, \boldsymbol{X}_{\mathrm{p}}, \hat{\boldsymbol{H}}, \hat{\psi}^h, T$
Output: $\{\hat{x}_{mk}, \forall(k \in \mathcal{K}_{\mathrm{d}}, m)\}, \{\hat{h}_{nm}, \forall(m, n)\}$

1: $\forall(k \in \mathcal{K}_{\mathrm{p}}, m, t): \hat{x}_{mk}(t) = [\boldsymbol{X}_{\mathrm{p}}]_{mk}, \ \psi^x_{mk}(t) = 0$
2: $\forall(k \in \mathcal{K}_{\mathrm{d}}, m): \hat{x}_{mk}(1) = 0, \ \psi^x_{mk}(1) = E_{\mathrm{s}}$
3: $\forall(m, n): \hat{h}_{nm}(1) = \left[\hat{\boldsymbol{H}}\right]_{nm}, \ \psi^h_{nm}(1) = \hat{\psi}^h$
4: $\forall(k, n): \hat{s}_{nk}(0) = 0$
5: **for** $t = 1$ to T **do**
   /* FNs */
6: $\forall(k, n): \overline{\xi}^p_{nk}(t)$
   $= \eta \sum_{i=1}^M \left\{|\hat{h}_{ni}(t)|^2 \psi^x_{ik}(t) + \psi^h_{ni}(t)|\hat{x}_{ik}(t)|^2\right\}$
   $+ (1 - \eta) \cdot \overline{\xi}^p_{nk}(t-1)$
7: $\forall(k, n): \xi^p_{nk}(t) = \eta \left(\overline{\xi}^p_{nk}(t) + \sum_{i=1}^M \psi^h_{ni}(t)\psi^x_{ik}(t)\right)$
   $+ (1 - \eta) \cdot \xi^p_{nk}(t-1)$
8: $\forall(k, n): \hat{p}_{nk}(t) = \sum_{i=1}^M \hat{h}_{ni}(t)\hat{x}_{ik}(t) - \overline{\xi}^p_{nk}(t)\hat{s}_{nk}(t-1)$
9: $\forall(k, n): \xi^s_{nk}(t) = \eta \frac{1}{\xi^p_{nk}(t) + N_0} + (1 - \eta) \cdot \xi^s_{nk}(t-1)$
10: $\forall(k, n): \hat{s}_{nk}(t) = \eta \frac{y_{nk} - \hat{p}_{nk}(t)}{\xi^p_{nk}(t) + N_0} + (1 - \eta) \cdot \hat{s}_{nk}(t-1)$
    /* VNs */
11: $\forall(k \in \mathcal{K}_{\mathrm{d}}, m): \overline{x}_{mk}(t) = \eta \cdot \hat{x}_{mk}(t) + (1 - \eta) \cdot \overline{x}_{mk}(t-1)$
12: $\forall(m, n): \overline{h}_{nm}(t) = \eta \cdot \hat{h}_{nm}(t) + (1 - \eta) \cdot \overline{h}_{nm}(t-1)$
13: $\forall(k \in \mathcal{K}_{\mathrm{d}}, m): \psi^r_{mk}(t) = \left(\sum_{i=1}^N |\overline{h}_{im}(t)|^2 \xi^s_{ik}(t)\right)^{-1}$
14: $\forall(k \in \mathcal{K}_{\mathrm{d}}, m): g^x_{mk}(t)$
    $= \max\left(1 - \psi^r_{mk}(t)\sum_{i=1}^N \psi^h_{im}(t)\xi^s_{ik}(t), 0\right)$
15: $\forall(k \in \mathcal{K}_{\mathrm{d}}, m): \hat{r}_{mk}(t)$
    $= g^x_{mk}(t)\overline{x}_{mk}(t) + \psi^r_{mk}(t)\sum_{i=1}^N \overline{h}^*_{im}(t)\hat{s}_{ik}(t)$
16: $\forall(k \in \mathcal{K}_{\mathrm{d}}, m): \hat{x}_{mk}(t+1) = f_{\mathrm{x}}(\hat{r}_{mk}(t), \psi^r_{mk}(t))$
17: $\forall(k \in \mathcal{K}_{\mathrm{d}}, m): \psi^x_{mk}(t+1) = V_{\mathrm{x}}(\hat{r}_{mk}(t), \psi^r_{mk}(t))$
18: $\forall(m, n): \psi^q_{nm}(t) = \left(\sum_{i=1}^K |\overline{x}_{mi}(t)|^2 \xi^s_{ni}(t)\right)^{-1}$
19: $\forall(m, n): g^h_{nm}(t)$
    $= \max\left(1 - \psi^q_{nm}(t)\sum_{i=1}^K \psi^x_{mi}(t)\xi^s_{ni}(t), 0\right)$
20: $\forall(m, n): \hat{q}_{nm}(t)$
    $= g^h_{nm}(t)\overline{h}_{nm}(t) + \psi^q_{nm}(t)\sum_{i=1}^K \overline{x}^*_{mi}(t)\hat{s}_{ni}(t)$
21: $\forall(m, n): \hat{h}_{nm}(t+1) = f_{\mathrm{h}}(\hat{q}_{nm}(t), \psi^q_{nm}(t))$
22: $\forall(m, n): \psi^h_{nm}(t+1) = V_{\mathrm{h}}(\hat{q}_{nm}(t), \psi^q_{nm}(t))$
23: **if** $\sum_{n,m} |\hat{h}_{nm}|^2 / (N \cdot M) \geqq \tau_{\mathrm{BiGAMP}} \cdot \phi$ **then** $T \leftarrow t$ and
    break   ▷ Stop criterion
24: **end if**
25: **end for**
26: $\forall(k \in \mathcal{K}_{\mathrm{d}}, m): \hat{x}_{mk} = \underset{\chi_q \in \mathcal{X}}{\arg\min} |\chi_q - \hat{r}_{mk}(T)|$
27: $\forall(m, n): \hat{h}_{nm} = f_{\mathrm{h}}(\hat{q}_{nm}(T), \hat{\psi}^q_{nm}(T))$

---

can achieve more robust signal recovery than the original scheme [15].

V. PERFORMANCE ASSESSMENT

Computer simulations were conducted to demonstrate the performances of the proposed JCDE algorithm via BiGaBP for uplink MUD in massive MU-MIMO systems. In all subsequent simulations, the average RX power from each TX antenna was assumed to be identical on the basis of slow TX power control as mentioned in Section II-A, and the time and frequency synchronization was assumed to be perfect. The pilot sequences, *i.e.*, rows of $\boldsymbol{X}_{\mathrm{p}}$, are given by Zadoff-Chu sequences [47], [48]. In cases when $K_{\mathrm{p}} \geq M$, orthogonal pilot sequences are used, while in cases when $K_{\mathrm{p}} < M$, the codebook of pilot sequences is given by the rows of a matrix constructed from an $M \times K_{\mathrm{p}}$ sub-matrix of an $M \times M$ orthogonal pilot matrix consisting of Zadoff-Chu sequences. Gray-coded QPSK-modulated signals were used for $\boldsymbol{X}_{\mathrm{d}}$, and the channel code is not used. As for algorithmic parameters, the damping factor η was set to 0.5 in all the JCDE algorithms, the predetermined parameters in ASB[4] were set to $(\tau_0, \tau_1) = (1, 4)$ in $\gamma(t)$ and $\tau_2 = 0.4$ in $\beta(t)$, respectively, and the maximum number of iterations was set constant to $T = 32$, although other criteria based on convergence could also be employed. A stop criterion was introduced for each JCDE algorithm to avoid divergence of estimates caused by unstable convergence behavior, and the predetermined parameter for the stop criterion was set to $\tau_{\mathrm{BiGaBP}} = \tau_{\mathrm{BiGAMP}} = 5$ in all the JCDE algorithms. In addition, we define the compression ratio as the ratio between the pilot length K_{p} and the number of UE devices M; thus, $\kappa \triangleq K_{\mathrm{p}}/M$.

A. BER Performance

Our first set of results is given in Fig. 3, where the performances in terms of BER as a function of the SNR, of the massive MU-MIMO systems with short non-orthogonal pilots in uncorrelated Rayleigh fading channels, *i.e.*, $h_{n,m} \sim \mathcal{CN}(h_{n,m}; 0, 1/N), \forall(n, m)$, are compared:

- *MMSE:* Baseline two-stage receiver consisting of spatial filtering-based channel estimation presented in Section II-B and linear minimum mean square error (MMSE)-based data detection.
- *GaBP:* Two-stage receiver consisting of spatial filtering-based channel estimation presented in Section II-B and GaBP-based data detection [24], [25].
- *BiGAMP (Alg.3):* JCDE receiver based on BiGAMP [15] presented in Alg. 3.

---

[4]Several simulations were conducted to find the sub-optimal parameters for minimizing the BER at SNR = 10 dB in the setting of Fig. 3 shown on the next page. Learning optimization of these scaling parameters as trainable parameters using DU techniques [38] remains as future work. For further details, we refer the readers to [49].

$$\hat{p}_{nk}(t) = \sum_{i=1}^M \hat{h}_{ni}(t)\hat{x}_{ik}(t) - \underbrace{\left(\frac{\tilde{y}_{nk}(t)}{\xi^x_{nk}(t)}\sum_{i=1}^M \hat{h}_{ni}(t)\hat{h}^*_{ni}(t-1)\psi^x_{ik}(t) + \frac{\tilde{y}_{nk}(t)}{\xi^h_{nk}(t)}\sum_{i=1}^M \hat{x}_{ik}(t)\hat{x}^*_{ik}(t-1)\psi^h_{ni}(t)\right)}_{\text{Onsager correction term}} + \mathcal{O}(N^{-\frac{1}{2}}), \quad (39)$$
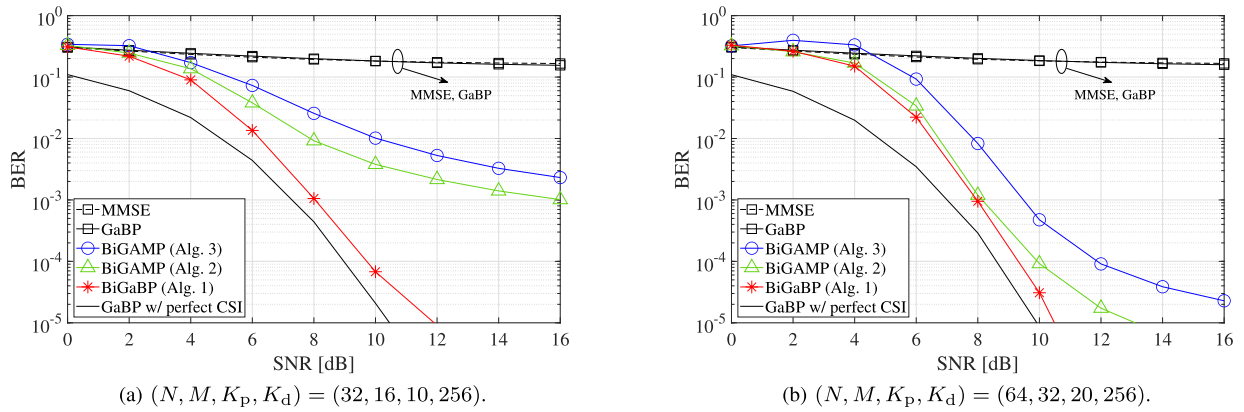
(a) $(N, M, K_{\mathrm{p}}, K_{\mathrm{d}}) = (32, 16, 10, 256)$.

(b) $(N, M, K_{\mathrm{p}}, K_{\mathrm{d}}) = (64, 32, 20, 256)$.

Fig. 3.   BER performances of MU-MIMO systems: ratio of pilot length κ fixed at 0.625.

- *BiGAMP (Alg.2):* JCDE receiver based on modified BiGAMP [16] presented in Alg. 2.
- *BiGaBP (Alg.1):* Proposed JCDE receiver based on BiGaBP presented in Alg. 1.
- *GaBP w/ perfect CSI:* Genie-aided scheme in which the perfect CSI is known at the receiver.

The results in Fig. 3 show the BER performance in massive MU-MIMO systems with $\rho_1 \triangleq N/M = 2$, $\kappa \triangleq K_{\mathrm{p}}/M = 0.625$, and $K_{\mathrm{d}} = 256$. The two-stage MUD schemes, "MMSE" and "GaBP," fail to detect MIMO signals reliably (BER $> 10^{-1}$) due to fatal errors in channel estimation using MNS under severe underdetermined conditions of $\kappa = 0.625$. On the one hand, "BiGAMP (Alg. 3)" can improve estimation accuracy significantly by taking advantage of the quasi-orthogonality of data structure. However, its performance deviates significantly from "GaBP w/ perfect CSI" because the mismatches between the finite MIMO system and large-system limit assumptions break the mathematical consistency in the MP rule. On the other hand, "BiGAMP (Alg. 2)" can further improve the performance owing to the MP rule that enables suppression of the harmful effect of the self-feedback with higher precision. However, there remains a non-negligible degradation from the Genie-aided performance, and the high-level error floor is inevitable due to poor initial estimation accuracy caused by the non-orthogonal pilots, especially in Fig. 3a, where the system size is relatively small. In contrast, our proposed "BiGaBP (Alg. 1)" significantly outperforms both "BiGAMP (Alg. 2)" and "BiGAMP (Alg. 3)" at the operating SNR with the same computational complexity and approaches the Genie-aided performance without suffering from error floors, owing to the ASB mechanism. Remarkably, the degradation at BER $= 10^{-4}$ is less than 1.0 dB for both configurations.

As shown in Section IV, BiGAMP is derived as a rigorous approximation of BiGaBP in the large-system limit; therefore, the performance difference between these two methods decreases as the system size increases. In fact, as can be seen from the comparison in Figs. 3a and 3b, the gain of the proposed method compared to "BiGAMP (Alg. 2)" becomes smaller, especially in the low SNR region. However, this apparently small difference is due to burst errors in symbol blocks, which have a significant impact on transmission

performance in the actual systems. In addition, as will be discussed later in Section V-D, when spatial correlation exists among the fading coefficients, the improvement by our proposal is significant, even for massive configurations.

B. NMSE Performance

Next, we evaluate the channel estimation performance in terms of the NMSE of the channel estimates, defined as $\mathrm{NMSE}(\hat{\boldsymbol{H}}) \triangleq \mathbb{E}_{\boldsymbol{H}}\left[\frac{|\boldsymbol{H} - \hat{\boldsymbol{H}}|_{\mathrm{F}}^2}{|\boldsymbol{H}|_{\mathrm{F}}^2} \right]$. In addition to "BiGAMP (Alg. 2)," "BiGAMP (Alg. 3)," and "BiGaBP (Alg. 1)," the following performances are compared:

- *MNS:* Baseline MNS-based channel estimator.
- *MMSE limit:* Genie-aided scheme in which the perfect knowledge of $\boldsymbol{X}_{\mathrm{d}}$ is provided as prior information at each iteration of the JCDE algorithm [27], [50]. Provides an absolute lower bound in terms of the NMSE performance.

Fig. 4 shows the NMSE performance, where the system parameters are the same as in Fig. 3. As expected from the results in Fig. 3, serious estimation errors occur in "MSE," indicating that increasing the SNR does not improve the NMSE performance significantly due to prohibitive inter-pilot interference with $\kappa = 0.625$. Although "BiGAMP (Alg. 3)" can improve the performance by relaxing the underdetermined conditions by the JCDE mechanism, its operation becomes unstable when the system size is small, and its operation at low SNR induces divergence behavior of iterative estimation, making it difficult to achieve highly accurate channel estimation. Note that the concave shape of the MSE curve in the low SNR region in Fig. 4b is due to the stopping criterion for iterative divergence behavior. Next, "BiGAMP (Alg. 2)" can alleviate this inconvenience, but still deviates from the lower bound. In contrast, "BiGaBP (Alg. 1)" operates robustly even in the low SNR region and approaches the Genie-aided performance, "MMSE-limit," in the high SNR region for both configurations.

C. Robustness to Changes in Symbol Length

Let us shift our focus to the robustness of the proposed method to changes in frame configurations. Fig. 5 shows the BER performance as a function of κ ($\triangleq K_{\mathrm{p}}/M$), where the
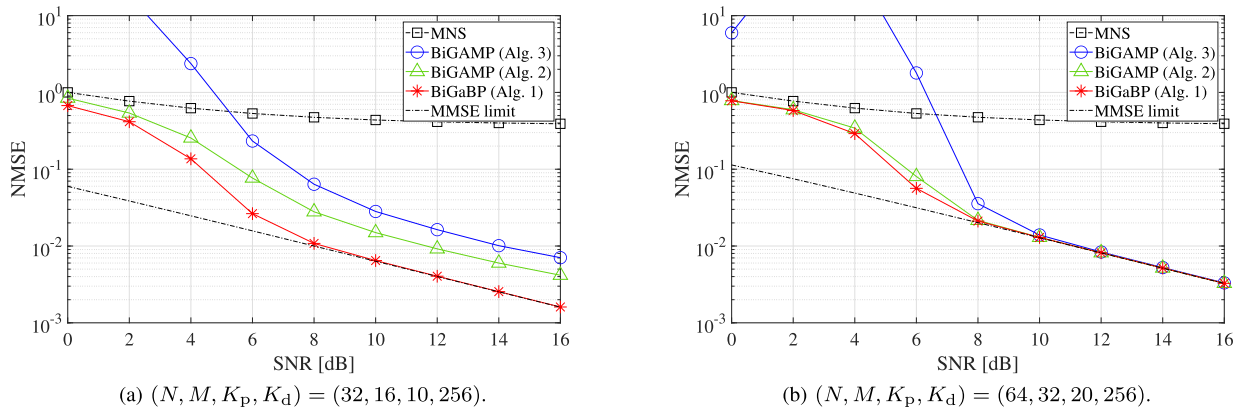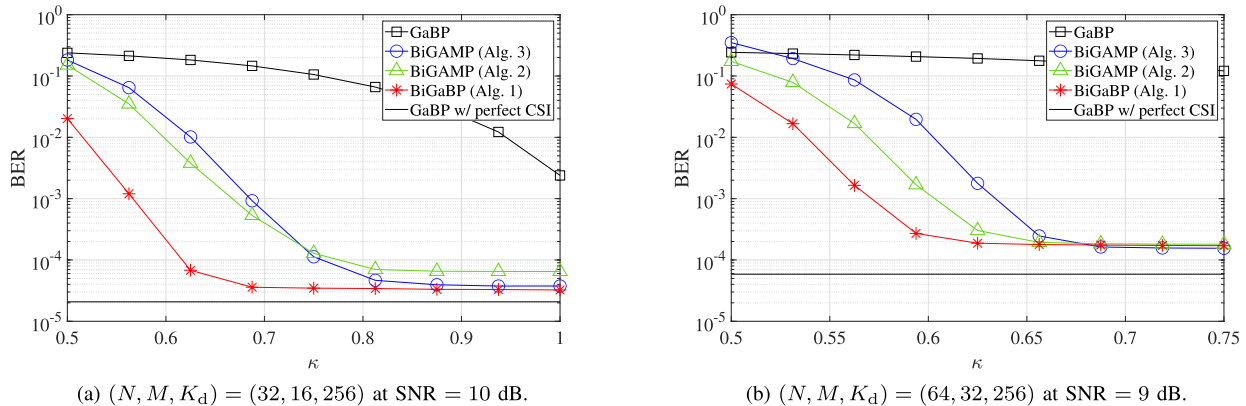
(a) $(N, M, K_{\mathrm{p}}, K_{\mathrm{d}}) = (32, 16, 10, 256)$.

(b) $(N, M, K_{\mathrm{p}}, K_{\mathrm{d}}) = (64, 32, 20, 256)$.

Fig. 4.   NMSE performances of MU-MIMO systems: ratio of pilot length κ fixed at 0.625.

(a) $(N, M, K_{\mathrm{d}}) = (32, 16, 256)$ at SNR $= 10$ dB.

(b) $(N, M, K_{\mathrm{d}}) = (64, 32, 256)$ at SNR $= 9$ dB.

Fig. 5.   BER performances of MU-MIMO systems as a function of the pilot length.

other system parameters are the same as in Fig. 3. The SNR is fixed at 10 dB and 9 dB in Figs. 5a and 5b, respectively. For both configurations, it is clearly confirmed that the proposed method can work at lower κ than the BiGAMP-based counterparts. More specifically, "BiGaBP (Alg. 1)" approaches the Genie-aided performance, "GaBP w/ perfect CSI," up to about $\kappa = 0.7$ in Fig. 5a and $\kappa = 0.6$ in Fig. 5b, indicating that the improvement from the conventional methods becomes significant, especially when the system size is small.

Fig. 6 shows the BER performance as a function of K_{d}. Intuitively, it would seem that the longer K_{d}, the better the orthogonality of the pilot-plus-data sequence and the better the estimation performance. In fact, when using orthogonal pilots, the channel estimation performance improves due to the improved diversity gain over the discrete-time dimension as K_{d} increases. However, since we use non-orthogonal pilots in Fig. 6, the initial channel estimation accuracy is poor; hence, the convergence behavior of iterative estimation becomes unstable due to error propagation if K_{d} is too large. As a result, it is not necessarily advantageous to increase K_{d} for the BiGAMP-based methods that average beliefs in the large system limit, and the performance degrades when K_{d} is set very long, *e.g.*, 512. "BiGaBP (Alg. 1)" can alleviate the above problem by the MP rule that does not rely on belief averaging using the large-system limit approximation and the ASB mechanism for stabilizing convergence. Consequently,

as K_{d} increases, the performance improves and asymptotically approaches the Genie-aided reference.

D. Robustness to Correlated Massive MIMO Channels

In practice, employing a large number of antennas at a BS leads to spatial correlation among fading coefficients; hence, it is vital to confirm that our method works well even in correlated channels. In this article, we use two channel models: the geometric one-ring model [27], [51], [52] and the finite path model [53], [54], [55], which are commonly used to represent spatial correlation among fading coefficients in massive MIMO scenarios.

1) Geometric One-Ring Model: In practice, the wireless channels between the BS and the UE exhibit a small angular spread from the perspective of the BS, as a result of local scatterers around the UE and the high placement of the BS antennas [33], [51], [52], [56]. Assuming diffuse $2 \times$ D field of isotropic scatters around UEs, the (i, j) element of the RX spatial correlation matrix for the m-th UE, $\boldsymbol{\Theta}_m \triangleq \mathbb{E}_{\mathbf{h}_m} \left\{ \boldsymbol{h}_m \boldsymbol{h}_m^{\mathrm{H}} \right\} \in \mathbb{C}^{N \times N}$ is given by

$$[\boldsymbol{\Theta}_m]_{i,j} = \frac{1}{\Delta \theta_m} \int_{\theta_m^{\min}}^{\theta_m^{\max}} \exp \left[\mathrm{j} \pi (i - j) \cos(\theta) \right] d\theta, \quad (41)$$

which denotes the correlation coefficient between the i-th and j-th RX antenna elements. Here, waves arrive from the m-th UE with an angular spread $\Delta \theta_m \triangleq \theta_m^{\max} - \theta_m^{\min}$, and
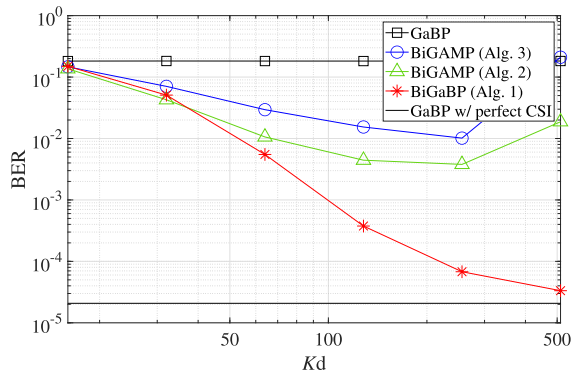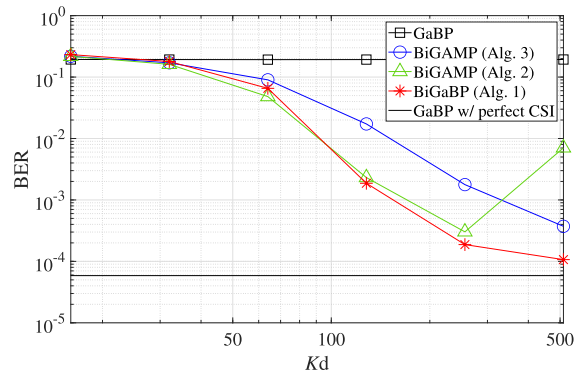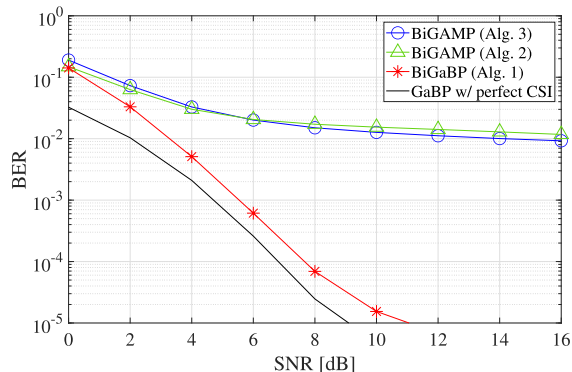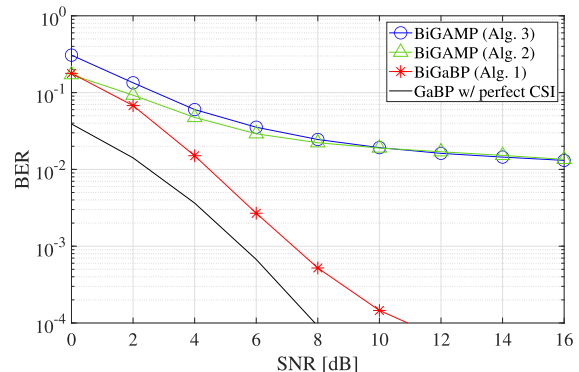
(a) $(N, M, K_{\mathrm{p}}) = (32, 16, 10)$ at SNR = 10 dB.

(b) $(N, M, K_{\mathrm{p}}) = (64, 32, 20)$ at SNR = 9 dB.

Fig. 6.   BER performances of MU-MIMO systems as a function of the data length.

(a) One-ring model.

(b) Finite path model.

Fig. 7.   BER performance of MU-MIMO systems in correlated channels with $(N, M, K_{\mathrm{p}}, K_{\mathrm{d}}) = (64, 16, 10, 256)$.

the antenna element spacing is fixed to half the wavelength. The m-th column vector of $\boldsymbol{H}$ is computed by $\boldsymbol{h}_m = \boldsymbol{\Theta}_m^{1/2} \boldsymbol{\nu}_m, \boldsymbol{\nu}_m \sim \mathcal{CN}(\boldsymbol{\nu}_m; \boldsymbol{0}, \phi \boldsymbol{I}_N)$.

Fig. 7a shows the BER performance of massive MU-MIMO systems in correlated channels following the one-ring model, where $(N, M) = (64, 16)$ and the other system parameters were the same as in Fig. 3. A sector antenna of 120 degrees opening was considered. The angular spread for each UE was set to 30 degrees. The UEs were naturally partitioned into 8 segments with $M/8$ UEs randomly dropped in each segment.

Although the ratio of the number of spatially multiplexed streams to the number of receive antennas is $\rho_1 \triangleq N/M = 4$, a configuration that is relatively easy to obtain diversity gain, the BiGAMP-based JCDE algorithms are not able to provide highly accurate estimation due to the high level of error floor. This is because the correlation among the fading coefficients greatly degrades the accuracy of the large-system limit approximation, making it difficult to decouple the self-feedback across iterations due to the Onsager correction term. In contrast, the BiGaBP-based JCDE algorithm, which relies only on SGA based on mild CLT, is relatively robust against correlation among fading coefficients, and significant performance improvement can be achieved by adjusting the convergence speed with the assistance of the proposed ASB. Specifically, our method can achieve BER = 10^{-4}, and the degradation from the lower-bound reference is suppressed to within 1.0 dB at BER = 10^{-4}.

2) Finite Path Model: Since the geometric one-ring model assumes infinite scatterers around every UE as found in (41),

the Gaussianity is relatively high; however, such a rich scattering environment cannot always occur in practice. For example, in millimeter-wave (mmWave) wireless communications, where diffraction and scattering rarely occur, the number of paths arriving at the receiver is limited [57], [58]. To represent such wireless channels, the finite path model is often utilized. Assuming that L propagation paths arrive at the BS from each UE, the m-th column vector of $\boldsymbol{H}$ can be expressed as [53], [54], and [55]

$$\boldsymbol{h}_m = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} g_{l,m} \cdot \boldsymbol{s}\left(\Omega_{l,m}\right), \tag{42}$$

where $g_{l,m} \sim \mathcal{CN}(g_{l,m}; 0, \phi)$ is the channel gain along the l-th path of the m-th UE. This is obtained from the steering vector $\boldsymbol{s}\left(\Omega_{l,m}\right) \triangleq \left[1, \exp\left[\mathrm{j}\pi\Omega_{l,m}\right], \ldots, \exp\left[\mathrm{j}\pi\left(N-1\right)\Omega_{l,m}\right]\right]^{\mathrm{T}}$. where $\Omega_{l,m} \triangleq \cos\theta_{l,m}$ with $\theta_{l,m}$ denoting the azimuth angle of the l-th propagation path of the m-th UE. The antenna element space is fixed to half the wavelength.

Fig. 7b shows the BER performance of massive MU-MIMO systems in correlated channels following the finite path model. A sector antenna with a 120 degrees opening was considered, and the UE devices were randomly dropped in the above angular region around the BS. The number of paths was set to $L = 8$.

As can be inferred from the results of Fig. 7a, the BiGAMP-based JCDE algorithms fail to provide highly accurate estimation and suffer from a high level of error floor with BER > 10^{-2}. The lower the Gaussianity of the massive

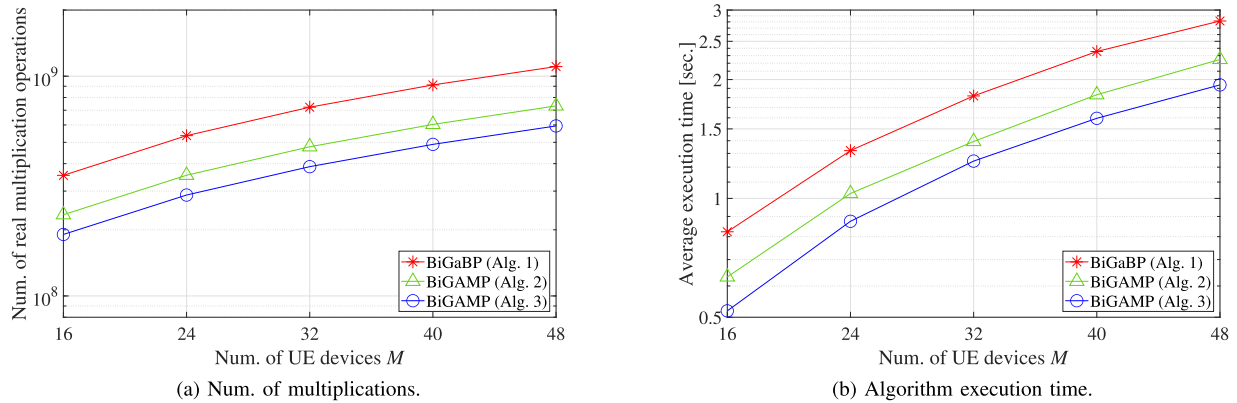(a) Num. of multiplications.

(b) Algorithm execution time.

Fig. 8. Complexity analysis of JCDE algorithm as a function of M with $\kappa = 0.625$, $N = 64$, $K = 256$ and $T = 32$.

MIMO channel, the greater the discrepancy between the Onsager correction term and the actual self-feedback value, which causes error propagation and significantly degrades the iterative convergence property. In contrast, the proposed method can significantly reduce the error floor level, achieving BER $= 10^{-4}$. The deviation from the lower-bound reference is also suppressed within 2.0 dB at BER $= 10^{-3}$, confirming that the self-feedback cancellation by exchanging extrinsic values and adjusting the iterative convergence speed by ASB are effective for highly accurate estimation in such deterministic wireless channels.

E. Complexity Analysis

First, the computational complexity of each JCDE algorithm was evaluated in terms of the number of real multiplication operations required to detect data symbols and estimate channel coefficients. To evaluate the approximate number of real multiplication operations, we adopt the following basic assumptions presented in [59].

Fig. 8a shows the number of real multiplication operations as a function of the number of UE devices M, and the compression ratio is fixed to $\kappa \triangleq K_{\mathrm{p}}/M = 0.625$. The dominant factors determining the computational complexity in the conventional BiGAMP-based methods are the process to compute $\hat{p}_{nk}(t)$, $\hat{q}_{nm}(t)$, and $\hat{r}_{mk}(t)$ in Alg. 2 (and Alg. 3), whose complexity is of order $\mathcal{O}(MNK)$ per iteration, which is similar to that of the proposed BiGaBP-based method. This fact is confirmed by the results in Fig. 8a, which shows that the proposed method can significantly improve the estimation accuracy without largely increasing the amount of computation. More specifically, at $M = 32$, "BiGaBP (Alg. 1)" can operate at about twice the computational cost of "BiGAMP (Alg. 3)" and about 1.5 times that of "BiGAMP (Alg. 2)," respectively.

For a more practical evaluation, Fig. 8b shows the average execution time[5] for each JCDE algorithm to detect data symbols and estimate channel coefficients. When the programs were actually executed and their execution times were compared, as shown in Fig. 8b, the results show that the relative

relationship between the proposed and conventional methods is a similar trend to that shown in Fig. 8a in terms of the number of multiplication operations.

The above results show that BiGAMP and BiGaBP have a similar complexity order, but in terms of more practical processing cost, BiGAMP, which can reduce the number of beliefs propagated by the large-system limit approximation, can achieve lower computational cost. However, considering that BiGAMP are extremely vulnerable to insufficient system size and channel correlations, and that BiGaBP can significantly improve performance and achieve performance close to the lower-bound reference in many cases, the proposed method can be seen to achieve an excellent trade-off between estimation capability and computational cost.

VI. Conclusion

In this paper, we proposed a novel JCDE scheme via BiGaBP for uplink massive MU-MIMO systems with short non-orthogonal pilots. The proposed BiGaBP framework operates based on the SGA under CLT, which is a milder assumption than the large-system limit condition the BiGAMP algorithm relies on. In addition, the ASB mechanism is extended to BIPs, realizing stable convergent behavior of JCDE when using non-orthogonal pilots. It is also shown that BiGAMP is derived by approximating BiGaBP under the large-system limit assumption, and the relationship between the original BiGAMP framework and the proposed scheme is clarified. The numerical results show that our proposed method outperforms the state-of-the-art scheme and approaches the performance of the idealized scheme for a variety of system parameters.

---

[5]To measure the execution time of the algorithms, we used the MATLAB profiler of MATLAB R2023a. The measurements were run on Windows 10 with Intel (R) Core (TM) i7-11700 CPU at 2.50GHz and 16.0GB RAM. The non-linear function table is assumed to be read in memories in advance; hence, the processing time for computing $\tanh(\cdot)$ can be ignored.

References

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

[2] 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.

[3] 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.

[4] X. Lin et al., "5G new radio: Unveiling the essentials of the next generation wireless access technology," *IEEE Commun. Standards Mag.*, vol. 3, no. 3, pp. 30–37, Sep. 2019.

[5] 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.

[6] T. Wild, V. Braun, and H. Viswanathan, "Joint design of communication and sensing for beyond 5G and 6G systems," *IEEE Access*, vol. 9, pp. 30845–30857, 2021.

[7] 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.

[8] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[9] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Proc. 40th Asilomar Conf. Signals, Syst. Comput.*, Oct. 2006, pp. 359–363.

[10] X. Yuan, C. Fan, and Y. J. Zhang, "Fundamental limits of training-based uplink multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7544–7558, Nov. 2018.

[11] M. C. Valenti and B. D. Woerner, "Iterative channel estimation and decoding of pilot symbol assisted turbo codes over flat-fading channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 9, pp. 1697–1705, Sep. 2001.

[12] K. Takeuchi, R. R. Müller, M. Vehkaperä, and T. Tanaka, "On an achievable rate of large Rayleigh block-fading MIMO channels with no CSI," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6517–6541, Oct. 2013.

[13] J. Ma and L. Ping, "Data-aided channel estimation in large antenna systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3111–3124, Jun. 2014.

[14] Y. Akiba, T. Ishihara, and S. Sugiura, "Variable-Block-Length joint channel estimation and data detection for spatial modulation over time-varying channels," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13964–13969, Nov. 2020.

[15] 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.

[16] T. Fujitsuka and K. Takeuchi, "Pilot contamination in massive MIMO uplink via approximate message-passing," *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. E103.A, no. 12, pp. 1356–1366, 2020.

[17] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.

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

[19] J. Vila, P. Schniter, S. Rangan, F. Krzakala, and L. Zdeborová, "Adaptive damping and mean removal for the generalized approximate message passing algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 2021–2025.

[20] J. Zhang, X. Yuan, and Y. A. Zhang, "Blind signal detection in massive MIMO: Exploiting the channel sparsity," *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 700–712, Feb. 2018.

[21] W. Yan and X. Yuan, "Semi-blind channel-and-signal estimation for uplink massive MIMO with channel sparsity," *IEEE Access*, vol. 7, pp. 95008–95020, 2019.

[22] L. Chen and X. Yuan, "Blind multiuser detection in massive MIMO channels with clustered sparsity," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1052–1055, Aug. 2019.

[23] 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.

[24] 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.

[25] 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.

[26] R. Tamaki, K. Ito, T. Takahashi, S. Ibi, and S. Sampei, "Suppression of self-noise feedback in GAMP for highly correlated large MIMO detection," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2022, pp. 1300–1305.

[27] 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.

[28] 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.

[29] 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.

[30] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6664–6684, Oct. 2019.

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

[32] 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.

[33] 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.

[34] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018.

[35] X. Yi and C. Zhong, "Deep learning for joint channel estimation and signal detection in OFDM systems," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2780–2784, Dec. 2020.

[36] Z. Zhou, L. Liu, and H.-H. Chang, "Learning for detection: MIMO-OFDM symbol detection through downlink pilots," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 3712–3726, Jun. 2020.

[37] L. Sun, Y. Wang, A. L. Swindlehurst, and X. Tang, "Generative-adversarial-network enabled signal detection for communication systems with unknown channel models," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 47–60, Jan. 2021.

[38] J. R. Hershey, J. L. Roux, and F. Weninger, "Deep unfolding: Model-based inspiration of novel deep architectures," 2014, *arXiv:1409.2574*.

[39] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Model-driven deep learning for MIMO detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 1702–1715, 2020.

[40] X. Wang, H. Hua, and Y. Xu, "Pilot-assisted channel estimation and signal detection in uplink multi-user MIMO systems with deep learning," *IEEE Access*, vol. 8, pp. 44936–44946, 2020.

[41] Y. Zhang, J. Sun, J. Xue, G. Y. Li, and Z. Xu, "Deep expectation-maximization for joint MIMO channel estimation and signal detection," *IEEE Trans. Signal Process.*, vol. 70, pp. 4483–4497, 2022.

[42] L. Hanzo, T. H. Liew, B. L. Yeap, R. Y. S. Tee, and S. X. Ng, *Turbo Coding, Turbo Equalisation and Space-Time Coding: EXIT-Chart-Aided Near-Capacity Designs for Wireless Channels*. Hoboken, NJ, USA: Wiley, Mar. 2011.

[43] T. Takahashi, S. Ibi, and S. Sampei, "Design of criterion for adaptively scaled belief in iterative large MIMO detection," *IEICE Trans. Commun.*, vol. E102.B, no. 2, pp. 285–297, 2019.

[44] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing—Part II: Applications," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5854–5867, Nov. 2014.

[45] M. Taskaldiran, R. C. S. Morling, and I. Kale, "A comparative study on the modified max-log-MAP turbo decoding by extrinsic information scaling," in *Proc. Wireless Telecommun. Symp.*, Apr. 2007, pp. 1–5.

[46] A. Maleki, L. Anitori, Z. Yang, and R. G. Baraniuk, "Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP)," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4290–4308, Jul. 2013.

[47] D. Chu, "Polyphase codes with good periodic correlation properties (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 4, pp. 531–532, Jul. 1972.

[48] T. Kawamura, Y. Kishiyama, K. Higuchi, and M. Sawahashi, "Orthogonal pilot channel using combination of FDMA and CDMA in single-carrier FDMA-based evolved UTRA uplink," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2007, pp. 2403–2408.

[49] D. Shirase, T. Takahashi, S. Ibi, K. Muraoka, N. Ishii, and S. Sampei, "Deep unfolding-aided Gaussian belief propagation for correlated large MIMO detection," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2020, pp. 1–6.

[50] 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.

[51] W. C. Jakes, *Microwave Mobile Communications*, 1st ed. Hoboken, NJ, USA: Wiley, 1994.

[52] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[53] P. V. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212–2223, Jun. 2015.

[54] D. Fan et al., "Angle domain channel estimation in hybrid millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8165–8179, Dec. 2018.

[55] L. Xu, C. Qian, F. Gao, W. Zhang, and S. Ma, "Angular domain channel estimation for mmWave massive MIMO with one-bit ADCs/DACs," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 969–982, Feb. 2021.

[56] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York, NY, USA: Cambridge Univ. Press, 2005.

[57] R. W. Heath Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[58] M. Xiao et al., "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.

[59] 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.

Shinsuke Ibi (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 with the Centre for Wireless Communications, University of Oulu, Finland. In 2006, he joined the Graduate School of Engineering, Osaka University. From 2010 to 2011, he was a Visiting Researcher with the University of Southampton, U.K. He moved to Doshisha University in 2019, 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 and 71st Best Paper Awards from IEICE and the 24th Telecom System Technology Award from the Telecommunication Advancement Foundation.

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 with the Centre for Wireless Communications, University of Oulu, Finland. In 2019, he joined the Graduate School of Engineering, Osaka University, as an Assistant Professor. His current research interests include belief propagation, compressed sensing, signal processing, and wireless communications.

Seiichi Sampei (Life Fellow, IEEE) received the B.E., M.E., and Ph.D. degrees in electrical engineering from the Tokyo Institute of Technology, Japan, in 1980, 1982, and 1991, respectively. From 1982 to 1993, he was with the Communications Research Laboratory, Ministry of Posts and Telecommunications. From 1991 to 1992, he was a Visiting Researcher with the University of California at Davis, Davis, CA, USA. In 1993, he joined the Faculty of Engineering, Osaka University. He is currently a Professor with the Department of Information and Communications Technology, Osaka University. He has been involved in the development of adaptive modulation, intelligent radio transmission/access, cognitive wireless networking, and wireless distributed network techniques. He is a member of the Institute of Image Information and Television Engineers (ITE). He is a fellow of the Institute of Electronics, Information and Communication Engineers (IEICE). He received the Shinohara Young Engineering Award, the Achievements Award, the Communications Society Best Paper Award, the Best Paper Award from IEICE, the Telecom System Technology Award from the Telecommunication Advancement Foundation, the DoCoMo Mobile Science Award from the Mobile Communication Fund, the Ericsson Telecommunications Award, and the Shida Rinzaburo Award from the President of Council for Info-Communications Promotion Month.