GNN-Assisted BiG-AMP: Joint Channel Estimation and Data Detection for Massive MIMO Receiver

Zishen Liu, Nan Wu, *Senior Member, IEEE*, Dongxuan He, *Member, IEEE*, Weijie Yuan, *Senior Member, IEEE*, Yonghui Li, *Fellow, IEEE*, and Tony Q. S. Quek, *Fellow, IEEE*

***Abstract*— In this paper, we develop a graph neural network (GNN)-assisted bilinear inference approach to enhance the receiver performance of the MIMO system through message passing-based joint channel estimation and data detection (JCD). Specifically, based on the bilinear generalized approximate message passing (BiG-AMP) framework and conditional correlation of signal, we propose a GNN-assisted BiG-AMP (GNN-BiGAMP) approach, which integrates a GNN module into the data-detection-loop to compensate the inaccurate marginal likelihood approximation. By leveraging the coupling between the channel and received symbols, a bilinear GNN-assisted BiG-AMP (BiGNN-BiGAMP) JCD receiver is further proposed. This method incorporates two GNNs with similar graph representation into the bilinear posterior estimation loops, which not only compensates for approximation errors but also alleviates performance loss due to premature variance convergence, thereby enhancing the receiver performance significantly. To fully exploit the supervised information from channel estimation and data detection, we propose a multitask learning based training scheme, which coordinates GNNs with different tasks in two loops. Simulation results show that our proposed GNN-assisted JCD receivers significantly outperform other JCD counterparts in terms of both channel estimation and data detection.**

***Index Terms*— Joint channel estimation and data detection, graph neural network, bilinear generalized approximate message passing, MIMO receiver, multitask learning.**

Received 30 August 2024; revised 17 December 2024 and 11 February 2025; accepted 12 February 2025. Date of publication 27 February 2025; date of current version 12 June 2025. This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFB2900600; in part by the National Natural Science Foundation of China under Grant 61971041, Grant 62371045, Grant 62301060, and Grant 62471208; in part by Guangdong Provincial Natural Science Foundation under Grant 2024A151510098; and in part by the National Research Foundation, Singapore, and Infocomm Media Development Authority under its Future Communications Research and Development Programme. The associate editor coordinating the review of this article and approving it for publication was M. Wen. *(Corresponding authors: Nan Wu; Dongxuan He; Tony Q. S. Quek.)*

Zishen Liu, Nan Wu, and Dongxuan He are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: zishen_liu@bit.edu.cn; wunan@bit.edu.cn; dongxuan_he@bit.edu.cn).

Weijie Yuan is with the School of System Design and Intelligent Manufacturing and Shenzhen Key Laboratory of Robotics and Computer Vision, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: yuanwj@sustech.edu.cn).

Yonghui Li is with the School of Electrical and Information Engineering, The University of Sydney, Sydney, NSW 2008, Australia (e-mail: yonghui.li@sydney.edu.au).

Tony Q. S. Quek is with the Information Systems Technology and Design Pillar, Singapore University of Technology and Design, Singapore 487372, and also with the Department of Electronic Engineering, Kyung Hee University, Yongin 17104, South Korea (e-mail: tonyquek@sutd.edu.sg).

Digital Object Identifier 10.1109/TWC.2025.3543176

I. Introduction

MASSIVE multiple-input and multiple-output (MIMO) technology has demonstrated great potential in providing outstanding spectral efficiency and low latency, thus making it a crucial technology for the upcoming sixth-generation (6G) wireless communication [1], [2]. However, the performance of massive MIMO heavily depends on the estimation accuracy of channel state information (CSI), requiring the accurate estimation of a large number of channel coefficients at the receiver [3]. Typically, the CSI is acquired through additional pilot symbols, which decreases the energy efficiency and data rate, particularly in rapidly time-variant environment [4], [5]. Therefore, it is greatly challenging to develop a receiver with accurate and efficient channel estimation for massive MIMO systems.

To tackle this challenge, various works have explored a range of approaches [6], [7]. In particular, joint channel estimation and data detection (JCD) has been shown to provide improved estimation and detection performance while using fewer pilots [8], [9]. Particularly, message passing based algorithms and their derivatives, have been utilized as a unified framework for iterative JCD in MIMO receivers owing to their superior performance and tractable complexity [10], [11], [12], [13]. In [10], the framework of expectation propagation was employed to perform joint channel estimation and decoding in 3D massive MIMO systems with orthogonal frequency division multiplexing (OFDM), which can achieve matched filter bound with acceptable complexity in spatially correlated channels. In [11], a Bayes-optimal estimator based on the bilinear generalized approximate message passing (BiG-AMP), was developed, which can realize satisfactory JCD even with low-precision analog-to-digital converters. In [12], a variational Bayesian (VB) inference procedure was proposed for joint channel estimation and soft symbol decoding, which shows superior performance with fast convergence. In [13], two approximate message passing (AMP)-based JCD schemes were designed to guarantee the performance of a reflecting intelligent surface (RIS)-assisted uplink wireless communication system, where the cascaded channels and transmitted signal were accurately estimated jointly. However, in practical MIMO systems, the performance of the above JCD receivers is often limited due to the approximation inaccuracy during the iterative process [16]. Therefore, how to realize acceptable receiver performance in practical MIMO systems is still very challenging.

Recently, with its excellent ability in handling intractable problems, deep learning has been introduced into wireless communications [14], [15]. For our considered JCD receiver, deep learning shows superior performance in improving both the estimation and detection performance [17], [18], [19], [20], [21]. For example, convolutional neural networks (CNNs) were employed to construct channel estimation network and channel conditioned recovery network in [17], which were used as channel estimator and data detector, respectively. By treating the receiver as a uninterpretable system, residual networks and deep complex-valued convolutional networks have been used to design JCD receivers in [18] and [19], respectively. However, the above studies ignore the domain knowledge, resulting in significant training time and extensive data set requirements.

To address this issue, the authors in [20] and [21] employed model-driven approaches, which unfold the traditional algorithms like orthogonal approximate message passing (OAMP) and generalized expectation maximization (GEM), to reduce the number of parameters to be trained. Nevertheless, the loss function of the above approaches is based on the data detection results, which ignores the supervised information generated from channel estimation [22]. Moreover, the turbo-like architecture used by these approaches cannot perform channel estimation and data detection in parallel during iterations, suffering the severe system latency. On the other hand, the performances of these model-driven receivers rely on their corresponding unfolded traditional algorithms. Therefore, the approximation inaccuracy in traditional algorithms will adversely influence the receiver performance.

Benefiting from its excellent generalization ability and interpretability, graph neural networks (GNN) have been applied in MIMO receivers to realize accurate channel estimation [23] and data detection [24], [25], [26], [27]. For example, a GNN-based channel tracking framework was designed in [23], which estimates time-varying MIMO channels by exploiting the information of channel spatial correlation with the prior information obtained from least square (LS) estimation, showing significant adaptability in channel estimation. A GNN framework based on the graph model of Markov random field (MRF) was developed in [24], which learns a message-passing solution to realize the inference task of MIMO detection. Leveraging its ability in extracting the correlation of the graph, GNN was further employed as a learnable module to compensate the inaccuracy of cavity distribution in expectation propagation (EP) algorithms, which achieves superior performance when compared to the state-of-the-art counterparts [25], [26]. Inspired by the above works, where GNNs can be applied to enhance signal processing performance at receiver in terms of channel estimation or data detection, it shows to be potential to integrate GNNs into the traditional message-passing framework to solve JCD problems.

Motivated by the above background, a novel GNN-assisted bilinear inference approach is proposed to improve the performance of the message passing significantly in JCD problems. The main contributions of this paper are summarized as follows:

- Leveraging the conditional correlation between channel and data symbols in BiG-AMP framework, a GNN-assisted BiG-AMP (GNN-BiGAMP) JCD receiver is proposed by integrating GNN module into the data-detection-loop to compensate the inaccuracy of marginal likelihoods.
- By analyzing the coupling between channel coefficients and data symbols, a bilinear GNN-assisted BiG-AMP (BiGNN-BiGAMP) JCD receiver is further proposed by employing the bilinear GNN-assisted posterior estimation loops with two GNNs sharing the same graph representation.
- Inspired by pre-trained models and multitask learning, a step-by-step training scheme for the BiGNN-BiGAMP receiver is proposed, which utilizes the supervised information generated during both data detection and channel estimation.
- Based on the numerical results, our proposed GNN-BiGAMP and BiGNN-BiGAMP receivers have considerable performance gain compared to various JCD receivers with respect to channel estimation and data detection performance. Furthermore, our proposed GNN-assisted receivers demonstrate significant robustness to the spacial correlation of MIMO channels, in contrast to conventional bilinear inference algorithms.

The rest of this paper is organized as follows. Section II presents the signal model and statistic problem formulation. Our proposed GNN-BiGAMP receiver and its corresponding training scheme are described in Section III. In Section IV, we present the design of the proposed BiGNN-BiGAMP receiver, where the role of GNN and the proposed step-by-step training scheme with multitask learning are detailed. In Section V, the computational complexity of the proposed receivers is analyzed. The simulation results are provided in Section VI, followed by conclusions in Section VII.

Notations–Matrices and vectors are denoted by uppercase and lowercase boldface letters, e.g., $\mathbf{X}$ and $\mathbf{x}$. $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^*$ denote the transpose, conjugate transpose and conjugate operation, respectively. $\|\cdot\|$ denotes the Frobenius norm of a vector or a matrix, while $|\cdot|$ denotes the magnitude operator. $\widetilde{\mathbf{X}}$ represents that $\mathbf{X}$ is defined in the complex domain, while $\mathbb{R}^{m\times n}$ and $\mathbb{C}^{m\times n}$ denote the $m\times n$ real and complex space, respectively. $\mathbb{E}[\cdot]$ denotes the expectation operator, while $\mathbb{E}_{\xi}[\cdot]$ denotes expectation with respect to the distribution ξ. Var$[\cdot]$ denotes the variance operator. $\mathbf{I}_n$ denotes an identity matrix of size n. Furthermore, $\mathcal{CN}(x;a,\sigma^2)$ denotes a complex Gaussian distribution with mean a and variance σ^2:

$$\mathcal{CN}(x;a,\sigma^2)=\frac{1}{\sqrt{\pi\sigma^2}}\exp\left(-\frac{|x-a|^2}{\sigma^2}\right).$$

II. System Model

We consider a MIMO uplink system where a base station (BS) equipped with N_r receive antennas serves N_t single-antenna users. We consider the block-fading channel, which is assumed to be constant over T transmitted symbols intervals within a block. Accordingly, the received signal

$\widetilde{\mathbf{Y}} = [\widetilde{y}_{nt}] \in \mathbb{C}^{N_r \times T}$ can be expressed as

$$\widetilde{\mathbf{Y}} = \widetilde{\mathbf{H}}\widetilde{\mathbf{X}} + \widetilde{\mathbf{W}}, \tag{1}$$

where $\widetilde{\mathbf{H}} = [\widetilde{h}_{nm}] \in \mathbb{C}^{N_r \times N_t}$ is channel matrix with $\widetilde{h}_{nm}$ denoting the fading coefficient between user m and receive antenna n. $\widetilde{\mathbf{X}} = [\widetilde{x}_{mt}] \in \mathbb{C}^{N_t \times T}$ is the $\widetilde{M}$-quadrature amplitude modulation (QAM) modulated complex transmitted symbols of constellation $\widetilde{\Theta}$ within a block, $\widetilde{\mathbf{W}} \in \mathbb{C}^{N_r \times T}$ denotes additive white Gaussian noise (AWGN) with zero mean and element-wise variance $\widetilde{\sigma}^2$. Besides, we define $\widetilde{\mathbf{Z}} \triangleq \widetilde{\mathbf{H}}\widetilde{\mathbf{X}} \in \mathbb{C}^{N_r \times T}$. In particular, the CSI is assumed to be unknown, which needs to be estimated at the receiver.

To detect the received symbols accurately, the symbols transmitted during the first T_p time slots are employed as pilots, while the remaining $T_d = T - T_p$ time slots are used to transmit information. Accordingly, $\widetilde{\mathbf{X}}_p \in \mathbb{C}^{N_t \times T_p}$ and $\widetilde{\mathbf{X}}_d \in \mathbb{C}^{N_t \times T_d}$ are used to represent pilot symbols and information symbols, respectively. Besides, $\widetilde{\mathbf{X}}$ is assumed to be independent and identically distributed (i.i.d.) random variables $\widetilde{x}_{mt}$ with known probability distribution, given by

$$\begin{aligned} &p(\widetilde{\mathbf{X}}) \\ &= \prod_{m=1}^{N_t}\prod_{t=1}^{T} p(\widetilde{x}_{mt}) = \prod_{m=1}^{N_t}\prod_{t=1}^{T_p} p(\widetilde{x}_{p,mt}) \prod_{m=1}^{N_t}\prod_{t=1}^{T_d} p(\widetilde{x}_{d,mt}), \end{aligned} \tag{2}$$

where $\widetilde{x}_{p,mt}$ and $\widetilde{x}_{d,mt}$ are generated uniformly from the constellation distributed with $p(\widetilde{x}_{p,mt}) = \delta(\widetilde{x}_{p,mt} - \widetilde{x})|_{\widetilde{x}\in\widetilde{\Theta}}$ and $p(\widetilde{x}_{d,mt}) = \frac{1}{|\widetilde{\Theta}|}\sum_{\widetilde{x}\in\widetilde{\Theta}} \delta(\widetilde{x}_{d,mt} - \widetilde{x})$, respectively. Here, $\delta(\cdot)$ denotes the indicator function.

Accordingly, we have

$$\widetilde{\mathbf{Y}}_i = \widetilde{\mathbf{H}}\widetilde{\mathbf{X}}_i + \widetilde{\mathbf{W}}_i, \tag{3}$$

where $\widetilde{\mathbf{Y}}_i \in \mathbb{C}^{N_r \times T_i}$ with $i \in (p, d)$ representing the receive pilot and information signal, respectively.

According to [11], $\widetilde{\mathbf{H}}$ is assumed to be a realization of an independent random matrix with known separable probability density functions (PDF), given by

$$p(\widetilde{\mathbf{H}}) = \prod_{n=1}^{N_r}\prod_{m=1}^{N_t} p(\widetilde{h}_{nm}). \tag{4}$$

To guarantee satisfactory receiver performance, iterative joint channel estimation and data detection is considered. Instead of first estimating $\widetilde{\mathbf{H}}$ based on $\widetilde{\mathbf{Y}}_p$ and $\widetilde{\mathbf{X}}_p$ and then detecting $\widetilde{\mathbf{X}}_d$, JCD estimates $\widetilde{\mathbf{H}}$ and $\widetilde{\mathbf{X}}_d$ jointly from $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{X}}_p$, which can achieve equivalent performance but only requires relatively fewer pilot symbols. As such, the Bayesian inference framework is employed to achieve the minimum mean square error (MMSE) estimates of $\widetilde{\mathbf{H}}$ and $\widetilde{\mathbf{X}}$ [28]. Accordingly, the estimators of $\widetilde{h}_{nm}$ and $\widetilde{x}_{mt}$ can be expressed as

$$\hat{h}_{MMSE,nm} = \int \widetilde{h}_{nm} p(\widetilde{h}_{nm}|\widetilde{\mathbf{Y}}) d\widetilde{h}_{nm}, \tag{5a}$$

$$\hat{x}_{MMSE,mt} = \int \widetilde{x}_{mt} p(\widetilde{x}_{mt}|\widetilde{\mathbf{Y}}) d\widetilde{x}_{mt}, \tag{5b}$$

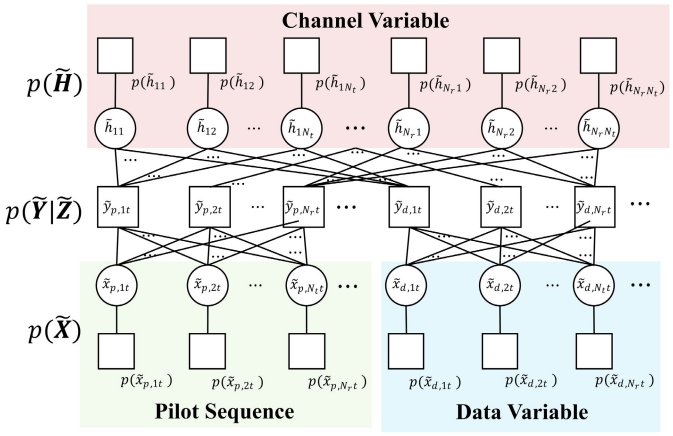


Fig. 1. Factor graph representation of the JCD problem.

where $p(\widetilde{h}_{nm}|\widetilde{\mathbf{Y}})$ and $p(\widetilde{x}_{mt}|\widetilde{\mathbf{Y}})$, i.e., marginal posterior distributions of elements in $\widetilde{\mathbf{H}}$ and $\widetilde{\mathbf{X}}$, are given by

$$p(\widetilde{h}_{nm}|\widetilde{\mathbf{Y}}) = \int_{\widetilde{\mathbf{H}}\setminus\widetilde{h}_{nm}} \int p(\widetilde{\mathbf{X}}, \widetilde{\mathbf{H}}|\widetilde{\mathbf{Y}}) d\widetilde{\mathbf{X}} d\widetilde{\mathbf{H}}, \tag{6a}$$

$$p(\widetilde{x}_{mt}|\widetilde{\mathbf{Y}}) = \int \int_{\widetilde{\mathbf{X}}\setminus\widetilde{x}_{mt}} p(\widetilde{\mathbf{X}}, \widetilde{\mathbf{H}}|\widetilde{\mathbf{Y}}) d\widetilde{\mathbf{X}} d\widetilde{\mathbf{H}}, \tag{6b}$$

respectively. $\int_{\widetilde{\mathbf{H}}\setminus\widetilde{h}_{nm}} d\widetilde{\mathbf{H}}$ denotes the integration over all elements in $\widetilde{\mathbf{H}}$ except for $\widetilde{h}_{nm}$, while $\int_{\widetilde{\mathbf{X}}\setminus\widetilde{x}_{mt}} d\widetilde{\mathbf{X}}$ denotes the integration over all elements in $\widetilde{\mathbf{X}}$ except for $\widetilde{x}_{mt}$.

To obtain these marginal posterior distributions effectively, a factorization method is employed, where the posterior distribution $p(\widetilde{\mathbf{X}}, \widetilde{\mathbf{H}}|\widetilde{\mathbf{Y}})$ can be factorized as

$$\begin{aligned} p(\widetilde{\mathbf{X}}, \widetilde{\mathbf{H}}|\widetilde{\mathbf{Y}}) &\propto p(\widetilde{\mathbf{Y}}|\widetilde{\mathbf{X}}, \widetilde{\mathbf{H}}) p(\widetilde{\mathbf{X}}) p(\widetilde{\mathbf{H}}) \\ &= p(\widetilde{\mathbf{Y}}|\widetilde{\mathbf{Z}}) p(\widetilde{\mathbf{X}}) p(\widetilde{\mathbf{H}}) \\ &= \left[\prod_{n=1}^{N_r}\prod_{t=1}^{T} p(\widetilde{y}_{nt}|\sum_{m=1}^{N_t} \widetilde{h}_{nm}\widetilde{x}_{mt})\right] \\ &\quad \times \left[\prod_{m=1}^{N_t}\prod_{t=1}^{T} p(\widetilde{x}_{mt})\right] \left[\prod_{n=1}^{N_r}\prod_{m=1}^{N_t} p(\widetilde{h}_{nm})\right]. \end{aligned} \tag{7}$$

Furthermore, (7) can be further derived as the following two transmission stages with respect to (2), given by

$$\begin{aligned} p(\widetilde{\mathbf{X}}, \widetilde{\mathbf{H}}|\widetilde{\mathbf{Y}}) \propto &\prod_{n=1}^{N_r}\prod_{t=1}^{T} p(\widetilde{y}_{nt}|\sum_{m=1}^{N_t} \widetilde{h}_{nm}\widetilde{x}_{mt}) \times \prod_{n=1}^{N_r}\prod_{m=1}^{N_t} p(\widetilde{h}_{nm}) \\ &\times \underbrace{\prod_{m=1}^{N_t}\prod_{t=1}^{T_p} p(\widetilde{x}_{p,mt})}_{\text{Pilot stage}} \times \underbrace{\prod_{m=1}^{N_t}\prod_{t=1}^{T_d} p(\widetilde{x}_{d,mt})}_{\text{Data stage}}. \end{aligned} \tag{8}$$

To observe JCD problem clearly, the factor graph of (8) is shown in Fig.1, where a circle denotes a variable node associated with $\widetilde{h}_{nm}$, $\widetilde{x}_{p,mt}$, or $\widetilde{x}_{d,mt}$, while a square represents a factor node associated with $p(\widetilde{y}_{nt}|\sum_{m=1}^{N_t} \widetilde{h}_{nm}\widetilde{x}_{mt})$, $p(\widetilde{x}_{p,mt})$, $p(\widetilde{x}_{d,mt})$, or $p(\widetilde{h}_{nm})$.

III. GNN-Asisted BiG-AMP JCD Receiver

A. BiG-AMP Based JCD Receiver

Due to the high-dimensional integration and summation involved in typical sum-product algorithms, the JCD problem shown in (8) and Fig. 1, which can be regarded as a bilinear Bayesian inference, is hard to solve directly. To tackle this problem, BiG-AMP-based receiver, which employs generalized approximate message passing (GAMP)-like approximation, can be adopted to solve the bilinear inference problem iteratively.

The BiG-AMP based JCD receiver is presented in **Algorithm** 1. According to the large-system limit (LSL) and central limit theorem (CLT), most variables within **Algorithm** 1 can be regarded as Gaussian variables, which can be completely characterized by its mean and variance [29]. In lines 2-7, an estimation of the matrix product $\widetilde{\mathbf{Z}}$ is provided. To be specific, a plug-in estimate of $\widetilde{z}_{nt}$ is obtained based on AMP approximation, characterized by its variance $\bar{V}_{nt}^{(l)}$ and mean $\bar{Z}_{nt}^{(l)}$ in lines 2 and 3 respectively. Then, an Onsager correction is adopted to improve the approximation accuracy of distribution of $\widetilde{z}_{nt}$, characterized by its variance $V_{nt}^{(l)}$ and mean $Z_{nt}^{(l)}$ as given in lines 4 and 5 respectively, where $\hat{s}_{nt}$ is the mean of the elements of Gaussian residual term $\mathbf{S}$ [30]. Since $\widetilde{\mathbf{Z}}$ and $\widetilde{\mathbf{W}}$ in (1) are independent, the Gaussian distribution of $p(\widetilde{\mathbf{Y}}|\widetilde{\mathbf{Z}})$ can be characterized by its variance $V_{Z,nt}^{(l)}$ and mean $\hat{Z}_{nt}^{(l)}$ as presented in lines 6 and 7 respectively. In lines 8 and 9, the variance $V_{s,nt}^{(l)}$ and mean $\hat{s}_{nt}^{(l)}$ of the residual s_{nt} are updated based on the approximation in [31]. Furthermore, the posterior estimations of $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{H}}$ are updated in lines 10-17. Specifically, the variance and mean of approximated marginal likelihood of $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{H}}$, i.e., $\{\Sigma_{\widetilde{x},mt}^{(l)}, R_{\widetilde{x},mt}^{(l)}\}$ and $\{\Sigma_{\widetilde{h},nm}^{(l)}, R_{\widetilde{h},nm}^{(l)}\}$, are firstly computed in lines 10-11 and 12-13, respectively. Accordingly, the marginal posterior estimation of $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{H}}$ can be updated, characterized by their mean and variance, i.e., $\{\mu_{\widetilde{x},mt}^{(l)}, v_{\widetilde{x},mt}^{(l)}\}$ and $\{\mu_{\widetilde{h},nm}^{(l)}, v_{\widetilde{h},nm}^{(l)}\}$ as presented in lines 14-17, where the marginal posterior distribution $\xi_{\widetilde{x},mt}^{(l)}$ and $\xi_{\widetilde{h},nm}^{(l)}$ can defined based on the Bayes' theorem, given by

$$\xi_{\widetilde{x},mt}^{(l)} = \frac{p\left(\widetilde{x}_{mt}^{(l)}\right)\mathcal{CN}\left(\widetilde{x}_{mt}^{(l)}; R_{\widetilde{x},mt}^{(l)}, \Sigma_{\widetilde{x},mt}^{(l)}\right)}{\int p(\widetilde{x})\mathcal{CN}\left(\widetilde{x}; R_{\widetilde{x},mt}^{(l)}, \Sigma_{\widetilde{x},mt}^{(l)}\right)d\widetilde{x}}, \tag{9a}$$

$$\xi_{\widetilde{h},nm}^{(l)} = \frac{p\left(\widetilde{h}_{nm}^{(l)}\right)\mathcal{CN}\left(\widetilde{h}_{nm}^{(l)}; R_{\widetilde{h},nm}^{(l)}, \Sigma_{\widetilde{h},nm}^{(l)}\right)}{\int p(\widetilde{h})\mathcal{CN}\left(\widetilde{h}; R_{\widetilde{h},nm}^{(l)}, \Sigma_{\widetilde{h},nm}^{(l)}\right)d\widetilde{h}}. \tag{9b}$$

Since the pilot sequences are assumed to be perfectly known at the receiver, the marginal posterior mean and variance of $\widetilde{\mathbf{X}}_p$ can be regarded as the true symbols $\widetilde{x}_{p,mt}$ and 0, respectively, which will not be updated in lines 14-17.

After L iterations, the output of the BiG-AMP receiver is given by the posterior estimation $\boldsymbol{\mu}_{\widetilde{h}}^{(L)}$ and $\boldsymbol{\mu}_{\widetilde{x}}^{(L)}$ with a hard decision.

Though the above BiG-AMP based receiver can be utilized to solve the JCD problem, such a receiver suffers from a significant performance loss due to the inaccurate approximation from LSL and CLT [29]. In addition, the data detection is conducted based on the Bayesian MMSE criterion, which leads to a significant performance loss when compared to the optimal maximum likelihood (ML) detection [32].

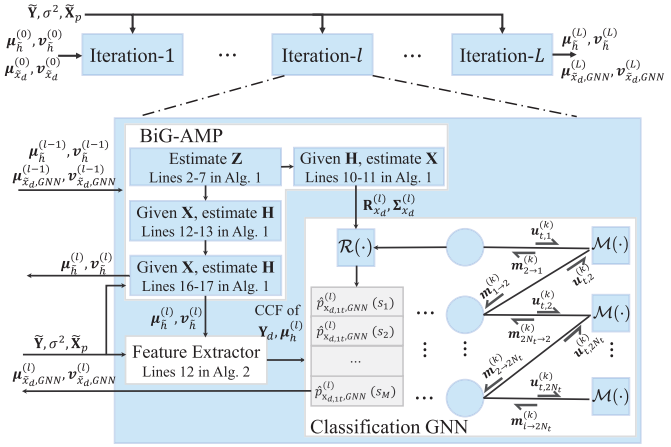


Fig. 2. The proposed GNN-BiGAMP receiver.

B. Improved BiG-AMP JCD Receiver With GNN-Assisted Detection

Against the receiver detailed in Section III-A, a GNN-assisted BiG-AMP (GNN-BiGAMP) JCD receiver is further proposed to improve both the channel estimation and data detection performance. Specifically, by integrating a GNN module into the BiG-AMP framework, our newly proposed GNN-BiGAMP receiver can compensate the inaccurate approximation and generate improved message from data-detection-loop to channel-estimation-loop, thus improving the receiver performance significantly.

As shown in Fig. 2, the proposed GNN-assisted JCD receiver includes the BiG-AMP module, feature extractor, and GNN module, which are detailed as follows.

1) GNN Module: By combining deep learning and graph structure, GNN can capture the data structure into feature vectors through iterative message-passing between nodes [24]. Specifically, GNN encodes the data information as the node state $\mathbf{u}$ and edge feature $\boldsymbol{\epsilon}$ to represent the graph structure inherent in data. Then, the value of $\mathbf{u}$ is updated by a recurrent neural network (RNN), which exploits the results obtained from previous iteration with the features from the neighbor nodes $\mathbf{u}_j : j \in ne(i)$, where $ne(\cdot)$ represents the neighboring variable set.

Typically, a GNN consists of three main components, namely a propagation, an aggregation, and a readout component. The propagation and aggregation components operate at every iteration k as the implementation of message-passing, while the readout component serves as a classification head to map the distribution learned by GNN to output a specific result required for a given task. More specifically, the propagation can be expressed as

$$\mathbf{m}_{i\to j}^{(k)} = \mathcal{M}\left(\mathbf{u}_i^{(k-1)}, \mathbf{u}_j^{(k-1)}, \boldsymbol{\epsilon}_{i,j}\right), \tag{10}$$

Algorithm 1 The BiG-AMP Based JCD Receiver

Input: $\widetilde{\mathbf{X}}_p$, $\widetilde{\mathbf{Y}}$, $\widetilde{\sigma}^2$, L

Output: $\hat{\mathbf{H}} = \boldsymbol{\mu}_{\widetilde{h}}^{(L)}$, $\hat{\mathbf{X}}_d = \arg\min_{\hat{\mathbf{X}}_d \in \Theta} \left| \hat{\mathbf{X}}_d - \boldsymbol{\mu}_{\widetilde{x}_d}^{(L)} \right|^2$.

Initialization[1]: $\boldsymbol{\mu}_{\widetilde{h}}^{(0)} = \widetilde{\mathbf{Y}}_p \widetilde{\mathbf{X}}_p^H (\widetilde{\mathbf{X}}_p \widetilde{\mathbf{X}}_p^H + \widetilde{\sigma}^2 \mathbf{I}_{N_t})^{-1}$

$\forall m,n,t: v_{\widetilde{h},nm}^{(0)} = 1$

$\mu_{\widetilde{x}_d,mt}^{(0)} = 0, v_{\widetilde{x}_d,mt}^{(0)} = 1$

$\hat{s}_{nt}^{(0)} = 0, V_{s,nt}^{(0)} = 1,$

1 **for** $l = 1$ *to* L **do**

// Plug-in Estimation of $\widetilde{\mathbf{Z}}$

2 $\forall n,t$:
$\bar{V}_{nt}^{(l)} = \sum_m \left[|\mu_{\widetilde{h},nm}^{(l-1)}|^2 v_{\widetilde{x},mt}^{(l-1)} + v_{\widetilde{h},nm}^{(l-1)} |\mu_{\widetilde{x},mt}^{(l-1)}|^2 \right]$;

3 $\forall n,t$: $\bar{Z}_{nt}^{(l)} = \sum_m \mu_{\widetilde{h},nm}^{(l-1)} \mu_{\widetilde{x},mt}^{(l-1)}$;

// Onsager Correction

4 $\forall n,t$: $V_{nt}^{(l)} = \bar{V}_{nt}^{(l)} + \sum_m v_{\widetilde{h},nm}^{(l-1)} v_{\widetilde{x},mt}^{(l-1)}$;

5 $\forall n,t$: $Z_{nt}^{(l)} = \bar{Z}_{nt}^{(l)} - \hat{s}_{nt}^{(l-1)} \bar{V}_{nt}^{(l)}$;

// Posterior Estimation of $\widetilde{\mathbf{Z}}$

6 $\forall n,t$: $V_{Z,nt}^{(l)} = \frac{V_{nt}^{(l)} \widetilde{\sigma}^2}{V_{nt}^{(l)} + \widetilde{\sigma}^2}$;

7 $\forall n,t$: $\hat{Z}_{nt}^{(l)} = V_{Z,nt}^{(l)} \left[\frac{y_{nt}}{\widetilde{\sigma}^2} + \frac{Z_{nt}^{(l)}}{V_{nt}^{(l)}} \right]$;

// Updating of Residual Terms $\mathbf{S}$

8 $\forall n,t$: $V_{s,nt}^{(l)} = \frac{V_{nt}^{(l)} - V_{Z,nt}^{(l)}}{(V_{nt}^{(l)})^2}$;

9 $\forall n,t$: $\hat{s}_{nt}^{(l)} = \frac{\hat{Z}_{nt}^{(l)} - Z_{nt}^{(l)}}{V_{nt}^{(l)}}$;

// Likelihood Updating of $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{H}}$

10 $\forall m,t$: $\Sigma_{\widetilde{x},mt}^{(l)} = \left(\sum_n |\mu_{\widetilde{h},nm}^{(l-1)}|^2 V_{s,nt}^{(l)} \right)^{-1}$;

11 $\forall m,t$: $R_{\widetilde{x},mt}^{(l)} = \mu_{\widetilde{x},mt}^{(l-1)} \left[1 - \Sigma_{\widetilde{x},mt}^{(l)} \sum_n v_{\widetilde{h},nm}^{(l-1)} V_{s,nt}^{(l)} \right] + \Sigma_{\widetilde{x},mt}^{(l)} \sum_m (\mu_{\widetilde{h},nm}^{(l-1)})^* \hat{s}_{nt}^{(l)}$;

12 $\forall m,t$: $\Sigma_{\widetilde{h},nm}^{(l)} = \left(\sum_t |\mu_{\widetilde{x},mt}^{(l-1)}|^2 V_{s,nt}^{(l)} \right)^{-1}$;

13 $\forall m,t$: $R_{\widetilde{h},nm}^{(l)} = \mu_{\widetilde{h},nm}^{(l-1)} \left[1 - \Sigma_{\widetilde{h},nm}^{(l)} \sum_t v_{\widetilde{x},mt}^{(l-1)} V_{s,nt}^{(l)} \right] + \Sigma_{\widetilde{h},nm}^{(l)} \sum_t (\mu_{\widetilde{x},mt}^{(l-1)})^* \hat{s}_{nt}^{(l)}$;

// Posterior Estimation of $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{H}}$

14 $\forall m,t$: $\mu_{\widetilde{x},mt}^{(l)} = \mathbb{E}_{\xi_{\widetilde{x},mt}^{(l)}} [\widetilde{x}_{mt}]$;

15 $\forall m,t$: $v_{\widetilde{x},mt}^{(l)} = \mathbb{E}_{\xi_{\widetilde{x},mt}^{(l)}} \left[(\widetilde{x}_{mt} - \mu_{\widetilde{x},mt}^{(l)})^2 \right]$;

16 $\forall m,t$: $\mu_{\widetilde{h},nm}^{(l)} = \mathbb{E}_{\xi_{\widetilde{h},nm}^{(l)}} \left[\widetilde{h}_{nm} \right]$;

17 $\forall m,t$: $v_{\widetilde{h},nm}^{(l)} = \mathbb{E}_{\xi_{\widetilde{h},nm}^{(l)}} \left[(\widetilde{h}_{nm} - \mu_{\widetilde{h},nm}^{(l)})^2 \right]$;

where $\mathbf{m}_{i \to j}^{(k)}$ denotes the message passing from node i to node j at the k-th iteration, $\mathbf{u}_i^{(k)}$ is the state of node i at the k-th iteration, and $\boldsymbol{\epsilon}_{i,j}$ represents the feature of edge connecting node i to node j. $\mathcal{M}(\cdot)$ represents a multiple layer perception (MLP) with active function ReLU. This propagation aims to combine current feature vectors of two nodes and the information along the direct edge between them. The aggregation component serves for updating the node feature vector by aggregating the current incoming messages $\mathbf{m}_{i \to j}^{(k)}$ at node i, with $j \in ne(i)$, as follow

$$\mathbf{u}_i^{(k)} = \mathcal{U} \left(\mathbf{u}_i^{(k-1)}, \sum_{j \in ne(i)} \mathbf{m}_{j \to i}^{(k)} \right), \tag{11}$$

and $\mathcal{U}(\cdot)$ is the gated recurrent unit (GRU) [33].

After K iterations, the readout component is employed to infer variables using the feature vectors $\mathbf{u}_i^{(K)}$, where the readout component for data detection can be constructed by a MLP with softmax function, given by

$$\{g_{i,m}\} = \mathcal{R}(\mathbf{u}_i^{(K)}), \tag{12a}$$

$$\hat{p}(x_i = s_m) = \frac{\exp(g_{i,m})}{\sum_m \exp(g_{i,m})}, \tag{12b}$$

where $\mathcal{R}(\cdot)$ represents the MLP for readout component and s_m is the m-th constellation point.

2) Feature Extraction: Given that the GNN relies on the corresponding message passing, its input must be capable of accurately representing the graph structure inherent to the Bayesian inference problem. To apply the GNN module to the BiG-AMP JCD receiver, a feature extraction scheme is proposed, which utilizes the conditional correlation between channel estimation and data detection to establish the graph structure.

Within the BiG-AMP framework, messages calculated from variable nodes of data symbols to factor nodes are mutually dependent on posterior channel estimation results, which can be verified from lines 10-11 in **Algorithm 1**. Therefore, the relationship between data and channel at the l-th iteration can be expressed as

$$\widetilde{\mathbf{Y}}_d = \boldsymbol{\mu}_{\widetilde{h}}^{(l)} \widetilde{\mathbf{X}}_d + \widetilde{\mathbf{W}}_d, \tag{13}$$

where $\boldsymbol{\mu}_{\widetilde{h}}^{(l)}$ denotes the posterior estimation of channel estimation at l-iteration, i.e., $[\mu_{\widetilde{h},nm}^{(l)}]$, which can represent the conditional correlation between different data symbols.

Relying on the above conditional correlation, a pair-wise Markov random field (MRF) representation can be adopted to obtain the column coupling features (CCF) for graph structure. In particular, the node feature vector $\boldsymbol{\alpha}_{x_d,it}^{(l)} = [\mathbf{y}_{d,t}^T \mathbf{h}_i^{(l)}, (\mathbf{h}_i^{(l)})^T \mathbf{h}_i^{(l)}, \frac{\widetilde{\sigma}^2}{2}]$ can be derived by the information from self-potential function, given by [24, Eq. (6)]

$$\phi(x_{d,it}) = \exp\left(\frac{2}{\widetilde{\sigma}^2} \mathbf{y}_{d,t}^T \mathbf{h}_i^{(l)} x_{d,it} - \frac{1}{2} (\mathbf{h}_i^{(l)})^T \mathbf{h}_i^{(l)} x_{d,it}^2\right) p(x_{d,it}), \tag{14}$$

where $x_{d,mt}$ is the element of the equivalent real-valued representation of $\widetilde{\mathbf{X}}_d$, i.e., $\mathbf{X}_d \in \mathbb{R}^{2N_t \times T_d}$. $\mathbf{y}_{d,t}$ and $\mathbf{h}_i^{(l)}$ are the t-th column of $\widetilde{\mathbf{Y}}_d$ and i-th column of real-valued $\mathbf{Y}_d \in \mathbb{R}^{2N_r \times T_d}$ and $\boldsymbol{\mu}_h^{(l)} \in \mathbb{R}^{2N_r \times 2N_t}$, respectively. Moreover, the edge feature vector $\boldsymbol{\epsilon}_{x_d,ijt}^{(l)} = [-(\mathbf{h}_i^{(l)})^T \mathbf{h}_j^{(l)}, \frac{\widetilde{\sigma}^2}{2}]$

[1]Since the initialization $\mu_{\widetilde{h},nm}^{(0)} = 0$ will lead to infinite values and indeterminate problems during the first iteration, a LMMSE estimation is employed as the initialization based on pilots [31].

is obtained through pair-potential function, given by [25, Eq. (6)]

$$\psi\left(x_{d,it}, x_{d,jt}\right) = \exp(-\frac{2}{\widetilde{\sigma}^2}(\mathbf{h}_i^{(l)})^T \mathbf{h}_j^{(l)} x_{d,it} x_{d,jt}). \tag{15}$$

However, since $\boldsymbol{\mu}_h^{(l)}$ is derived by the estimation of the BiG-AMP module, information loss cannot be ignored when uncertainty information is not included during feature extraction. To address this issue, an extended feature extraction is proposed to measure the uncertainty of JCD problem in virtue of the variance information during the BiG-AMP iterations. Specifically, the feature vectors of nodes and edges are defined as

$$\boldsymbol{\alpha}_{x_d,it}^{(l)} = \left[\mathbf{y}_{d,t}^T \mathbf{h}_i^{(l)}, F_{1,it}^{(l)}, (\mathbf{h}_i^{(l)})^T \mathbf{h}_i^{(l)}, F_{2,i}^{(l)}, \frac{\widetilde{\sigma}^2}{2}\right], \tag{16a}$$

$$\boldsymbol{\epsilon}_{x_d,ijt}^{(l)} = \left[-(\mathbf{h}_i^{(l)})^T \mathbf{h}_j^{(l)}, F_{3,ij}^{(l)}, \frac{\widetilde{\sigma}^2}{2}\right], \tag{16b}$$

where

$$\begin{cases} F_{1,it}^{(l)} = (\mathbf{y}_{d,t}^2)^T \mathbf{v}_{h,i}^{(l)} \\ F_{2,i}^{(l)} = 2\Big((\mathbf{v}_{h,i}^{(l)})^T \mathbf{v}_{h,i}^{(l)} + \Big((\mathbf{h}_i^{(l)})^2\Big)^T \mathbf{v}_{h,i}^{(l)} + (\mathbf{v}_{h,i}^{(l)})^T (\mathbf{h}_i^{(l)})^2\Big) \\ F_{3,ij}^{(l)} = (\mathbf{v}_{h,i}^{(l)})^T \mathbf{v}_{h,j}^{(l)} + \Big((\mathbf{h}_i^{(l)})^2\Big)^T \mathbf{v}_{h,j}^{(l)} + (\mathbf{v}_{h,i}^{(l)})^T (\mathbf{h}_j^{(l)})^2, \end{cases}$$

and $\mathbf{v}_{h,i}^{(l)}$ is the i-th column of the real-valued form of $\mathbf{V}_{\widetilde{h}}^{(l)} = [v_{\widetilde{h},nm}^{(l)}]$. The detailed derivation of these variance features is given in Appendix.

3) GNN-Assisted JCD Receiver: By properly adjusting the marginal likelihoods, the correlated and non-Gaussian residual noise caused by the operation of ignoring inter-symbol correlation in iterative Bayesian inference can be handled. As a result, the detection performance can be considerably improved [25]. Inspired by this, we consider utilizing the GNN module with the proposed feature extraction scheme in our proposed JCD receiver to obtain more accurate marginal likelihoods of received symbols, which can further enhance the performance of channel estimation through the bilinear structure of the BiG-AMP framework.

To be specific, after calculating the marginal posterior estimation in line 17 of **Algorithm** 1 at the l-th iteration, $\mu_{\widetilde{h},nm}^{(l)}$ and $v_{\widetilde{h},nm}^{(l)}$ are used to generate feature vectors through (16) as the input of GNN module. Besides, MLP is employed to encode the feature vectors into the node feature for message passing, given by

$$\mathbf{u}_{x_d,it}^{(0)} = \mathbf{W}_1(\boldsymbol{\alpha}_{x_d,it}^{(l)})^T + \mathbf{b}_1, \tag{17}$$

where $\mathbf{W}_1 \in \mathbb{R}^{N_u \times 5}$ and $\mathbf{b}_1 \in \mathbb{R}^{N_u}$ are both learnable matrices.

Then, the message passing including propagation in (10) and aggregation in (11) are conducted. However, since our aim is to improve the accuracy of the marginal likelihoods of data detection, $R_{\widetilde{x},it}^{(l)}$ and $\Sigma_{\widetilde{x},it}^{(l)}$ generated in lines 11-12 of **Algorithm** 1 by BiG-AMP outer iteration are added into the aggregation component as prior information, given by

$$\mathbf{u}_{x_d,it}^{(k)} = \mathcal{U}\left(\mathbf{u}_{x_d,it}^{(k-1)}, \left[\sum_{j\in ne(i)} \mathbf{m}_{t,j\to i}^{(k)}, R_{x_d,it}^{(l)}, \Sigma_{x_d,it}^{(l)}\right]\right), \tag{18}$$

where $R_{x_d,it}^{(l)}$ and $\Sigma_{x_d,it}^{(l)}$ are the element of data sequence part of the real-valued form of $\mathbf{R}_{\widetilde{x}_d}^{(l)} = \left[R_{\widetilde{x}_d,it}^{(l)}\right]$ and $\boldsymbol{\Sigma}_{\widetilde{x}_d}^{(l)} = \left[\Sigma_{\widetilde{x}_d,it}^{(l)}\right]$, i.e., $\mathbf{R}_{x_d}^{(l)} \in \mathbb{R}^{2N_t \times T_d}$ and $\boldsymbol{\Sigma}_{x_d}^{(l)} \in \mathbb{R}^{2N_t \times T_d}$, respectively.

After inner iterations, the readout component is performed to yield the improved marginal likelihoods, where the readout process in (12) can be replaced by the following steps, given by

$$[g_{x_d,it}(s_m)] = \mathcal{R}_c(\mathbf{u}_{x_d,it}^{(K)}), \tag{19a}$$

$$\hat{p}_{x_d,it,GNN}^{(l)}(s_m) = \frac{\exp\left(g_{x_d,it}(s_m)\right)}{\sum_{s_q\in\Theta}\exp\left(g_{x_d,it}(s_q)\right)}, s_m \in \Theta, \tag{19b}$$

where Θ represents M-QAM constellation with size $M = \sqrt{\widetilde{M}}$, and the elements of Θ are the real or imaginary parts of $\widetilde{M}$-QAM. Considering that the prior distribution of real-valued $x_{d,it}$ is also uniform indicator function, i.e., $p(x_{d,it}) = \frac{1}{|\Theta|}\sum_{x\in\Theta}\delta(x_{d,it} - x)$, the real-valued data detection estimation can be derived as

$$\mu_{x_{d,it},GNN}^{(l)} = \mathbb{E}_{\xi_{x_{d,it},GNN}^{(l)}}\left[x_{d,it}\right], \tag{20a}$$

$$v_{x_{d,it},GNN}^{(l)} = \mathbb{E}_{\xi_{x_{d,it},GNN}^{(l)}}\left[(x_{d,it} - \mu_{x_{d,it},GNN}^{(l)})^2\right], \tag{20b}$$

where

$$\xi_{x_{d,it},GNN}^{(l)} = \frac{p(x_{d,it}^{(l)})\hat{p}_{x_{d,it},GNN}^{(l)}}{\int p(x')\hat{p}_{x_{d,it},GNN}^{(l)}(x')dx'}. \tag{21}$$

Then, the outputs of GNN are converted to complex-valued form and substituted into the BiG-AMP module in the subsequent iterations, serving as the enhanced marginal posterior estimation of $\widetilde{x}_{d,mt}$.

The details of our proposed GNN-BiGAMP JCD receiver are shown in **Algorithm** 2.

C. Training Scheme

The remaining problem of our proposed GNN-BiGAMP receiver is to train the GNN efficiently, consisting of the design of training data structure and loss function.

Considering that the symbols over different time slots within a transmission block are independent, batch processing is more suitable to train the GNN, where the batch size is set as $N_{\text{batch}} = T$ to take both the pilot and data into consideration.

To train the network to achieve performance comparable to the optimal ML estimator, the cross-entropy is selected as the loss function, given by [34]

$$\mathcal{L}_c = -\frac{1}{T_d}\sum_{t=1}^{T_d}\sum_{i=1}^{2N_t}\sum_{s_m\in\Theta}\delta(x_{d,it} = s_m)\log(\hat{p}_{x_{d,it},GNN}^{(L)}(s_m)). \tag{22}$$

Algorithm 2 The Proposed GNN-BiGAMP Receiver

Input: $\widetilde{\mathbf{X}}_p$, $\widetilde{\mathbf{Y}}$, $\widetilde{\sigma}^2$, L, K

Output: $\hat{\mathbf{H}} = \boldsymbol{\mu}_{\widetilde{h}}^{(L)}$, $\hat{\mathbf{X}}_d = \boldsymbol{\mu}_{\widetilde{x}_d,GNN}^{(L)}$

Initialization: $\boldsymbol{\mu}_{\widetilde{h}}^{(0)} = \widetilde{\mathbf{Y}}_p\widetilde{\mathbf{X}}_p^H(\widetilde{\mathbf{X}}_p\widetilde{\mathbf{X}}_p^H + \widetilde{\sigma}^2\mathbf{I}_{N_t})^{-1}$;
$\forall m,n,t: v_{\widetilde{h},nm}^{(0)} = 1$;
$\mu_{\widetilde{x}_d,mt}^{(0)} = 0, v_{\widetilde{x}_d,mt}^{(0)} = 1$;
$\hat{s}_{nt}^{(0)} = 0, V_{s,nt}^{(0)} = 1$,;

1 **for** $l = 1$ *to* L **do**

`// The BiG-AMP Module:`

2 **if** $l = 1$ **then**

3 Refer to steps of lines 2-3 in **Algorithm** 1;

4 **else**

5 Refer to steps of lines 2-3 in **Algorithm** 1 using $\mu_{\widetilde{x}_{d,mt},GNN}^{(l-1)}$ and $v_{\widetilde{x}_{d,mt},GNN}^{(l-1)}$;

6 Refer to steps of lines 4-9 in **Algorithm** 1;

7 **if** $l = 1$ **then**

8 Refer to steps of lines 10-13 in **Algorithm** 1;

9 **else**

10 Refer to steps of lines 10-13 in **Algorithm** 1 using $\mu_{\widetilde{x}_{d,mt},GNN}^{(l-1)}$ and $v_{\widetilde{x}_{d,mt},GNN}^{(l-1)}$;

11 Refer to steps of lines 16-17 in **Algorithm** 1;

`// The Feature Extractor:`

12 $\forall t,i,j$: Compute $\boldsymbol{\alpha}_{x_d,it}^{(l)}$ and $\boldsymbol{\epsilon}_{x_d,ijt}^{(l)}$ via (16);

`// The GNN Module:`

13 $\forall t,i$: Compute $\mathbf{u}_{x_d,it}^{(0)}$ via (17);

14 **for** $k = 1$ *to* K **do**

15 $\forall t,i,j$: Compute $\mathbf{m}_{i\to j}^{(k)}$ and $\mathbf{u}_{x_d,it}^{(k)}$ via (10) and (18), respectively;

16 $\forall t,i$: Compute $\hat{p}_{x_{d,it},GNN}^{(l)}(s_m)$ via (19);

17 $\forall t,i$: Compute $\mu_{x_{d,it},GNN}^{(l)}$ and $v_{x_{d,it},GNN}^{(l)}$ via (20);

It is important to note that since the pilots are perfectly known at the receiver, the pilots would not be used to calculate the loss.

IV. Bilinear GNN-Assisted JCD Receiver Design

The BiG-AMP receiver regards the channel estimation and data detection as two independent processes during posterior estimation process among each iteration, and our proposed GNN-BiGAMP receiver only utilizes the unidirectional assistance from channel estimation to data detection. However, the coupling between channel and symbols, which is significant to estimate channel and detect data mutually, has not been utilized in the above two receivers.

Inspired by this, a new bilinear GNN-assisted BiG-AMP (BiGNN-BiGAMP) JCD receiver is developed in this section. Compared to GNN-BiGAMP, BiGNN-BiGAMP utilizes the GNN to enhance the marginal likelihoods of the two branches of the bilinear inference. Benefiting from such a design, the inaccuracy caused by distribution approximation can be improved and the premature variance convergence of BiG-AMP algorithm can be handled.

A. Bilinear GNN-Assisted JCD Receiver

For our considered system model, the coupling between channel and symbols can be regarded as invariant which follows the linear process $\widetilde{\mathbf{Z}} = \widetilde{\mathbf{H}}\widetilde{\mathbf{X}}$. More specifically, as shown in (23), shown at the bottom of the next page, the information from rows of channel matrix $\widetilde{\mathbf{H}}$ can provide a quantitative measurement of the relevance strength for different symbols within a time slot under the observation condition, vice versa.

To utilize such coupling properties efficiently, a graph representing is established to reflect conditional correlations between different channel coefficients of a row. In particular, such a graph for the block-fading channels has the same structure as data detection, i.e., a full-connect graph as presented in Fig. 3. Consequently, the core component of the GNN, which is responsible for message passing, can maintain a uniform structure and dimensionality.

Accordingly, in light of the CCF generated by real-valued observation $\mathbf{Y}_d$ and measurement matrix $\boldsymbol{\mu}_h^{(l)}$ for data detection, a shared feature extraction scheme is proposed by introducing the row coupling features (RCF) of real-valued $\mathbf{Y}_p \in \mathbb{R}^{N_r\times 2T_p}$ and $\mathbf{X}_p \in \mathbb{R}^{2N_t\times 2T_p}$ as the input features of the channel estimation branch,[2] given as

$$\boldsymbol{\alpha}_{h,ni} = \left[\mathbf{y}_{p,n}\mathbf{x}_{p,i}^T, \mathbf{x}_{p,i}\mathbf{x}_{p,i}^T, \frac{\widetilde{\sigma}^2}{2}\right], \tag{24a}$$

$$\boldsymbol{\epsilon}_{h,nij} = \left[-\mathbf{x}_{p,i}\mathbf{x}_{p,j}^T, \frac{\widetilde{\sigma}^2}{2}\right] \tag{24b}$$

where $\mathbf{y}_{p,n}$ and $\mathbf{x}_{p,i}$ are the n-th and i-th row of $\mathbf{Y}_p$ and $\mathbf{X}_p$, respectively. Since $\mathbf{X}_p$ is perfectly known with zero variance, the corresponding zero variance features are not added in RCF. Besides, these input feature vectors wound not be updated due to the fact that they only contain the pilot information.

Based on a shared feature extraction scheme derived by the coupling relationship, a BiGNN-BiGAMP receiver as shown in Fig. 4 is proposed. More specifically, two GNN modules with classification and regression head for readout are respectively employed to improve data detection and channel estimation using the CCF and RCF, which can be regarded as classification and regression problems, respectively.[3] Then, the enhanced marginal likelihood estimations from two GNNs are merged into the BiG-AMP model part by the message passing from factor nodes $p(\widetilde{\mathbf{Y}}|\widetilde{\mathbf{Z}})$ to variable nodes. Benefiting from the calculations in BiG-AMP and feature extractor, the improved channel estimation and signal detection are integrated into each other's loop based on the conditional correlation. Subsequently, two GNN modules are enabled to obtain more accurate prior information and feature inputs for marginal posterior estimations. Thus, the performance of JCD receiver can be further improved during the iterations.

[2]Due to the accumulation of variance with the length of the time sequence, only the plot information is utilized for feature generation in GNN for channel estimation, rather than using the estimation of $\widetilde{\mathbf{X}}_d$.

[3]Since the channel and data to be estimated have entirely different distributions, the backbone of GNN is difficult to be shared for these two tasks [35].

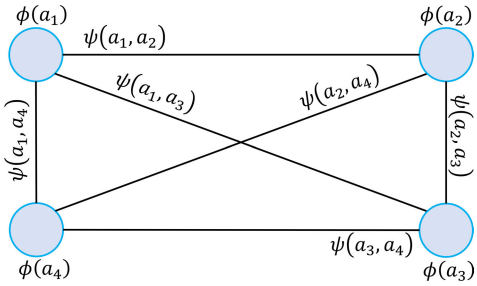


Fig. 3. An example of the full connect graph representing, $a_i \in (h_{ni}, x_{it})$.

The details of BiGNN-BiGAMP receiver are presented in **Algorithm** 3. Firstly, the initial node feature vector is obtained by a MLP with learnable parameters $\mathbf{W}_2 \in \mathbb{R}^{N_u \times 3}$ and $\mathbf{b}_2 \in \mathbb{R}^{N_u}$. Similar to (10) and (11), the propagation and aggregation components of the GNN for channel estimation are conducted in inner iterative steps of lines 9-10, where $R^{(l)}_{h,ni}$ and $\Sigma^{(l)}_{h,ni}$ are the element of real-valued form of $\mathbf{R}^{(l)}_{\widetilde{h}} = \left[R^{(l)}_{\widetilde{h},nm}\right]$ and $\mathbf{\Sigma}^{(l)}_{\widetilde{h}} = \left[\Sigma^{(l)}_{\widetilde{h},nm}\right]$, i.e., $\mathbf{R}^{(l)}_{h} \in \mathbb{R}^{N_r \times 2N_t}$ and $\mathbf{\Sigma}^{(l)}_{h} \in \mathbb{R}^{N_r \times 2N_t}$, respectively. The real-valued marginal likelihood estimation of channel estimation output by GNN is obtained in line 11, in which $\mathcal{R}_r$ is denoted as a regression head to generate the mean and variance of marginal likelihoods. To complete computations in the complex domain, the outputs of GNN are transferred to the complex-valued form in lines 14-15. Then, the enhanced marginal posterior estimation is derived in line 16, where

$$\xi^{(l)}_{\widetilde{h}_{nm},\mathrm{GNN}} = \frac{p\left(\widetilde{h}^{(l)}_{nm}\right)\mathcal{CN}\left(\widetilde{h}^{(l)}_{nm}; R^{(l)}_{\widetilde{h}_{nm},\mathrm{GNN}}, \Sigma^{(l)}_{\widetilde{h}_{nm},\mathrm{GNN}}\right)}{\int p(\widetilde{h})\mathcal{CN}\left(\widetilde{h}; R^{(l)}_{\widetilde{h}_{nm},\mathrm{GNN}}, \Sigma^{(l)}_{\widetilde{h}_{nm},\mathrm{GNN}}\right) d\widetilde{h}}. \tag{25}$$

Additionally, after the first iteration, the marginal posterior estimation outputs by two GNN branches substitute for all marginal posterior mean and variance of BiG-AMP in other steps.

Algorithm 3 The Proposed BiGNN-BiGAMP Receiver

Input: $\widetilde{\mathbf{X}}_p$, $\widetilde{\mathbf{Y}}$, $\widetilde{\sigma}^2$, L, K
Output: $\hat{\mathbf{H}} = \boldsymbol{\mu}^{(L)}_{\widetilde{h},GNN}$, $\hat{\mathbf{X}}_d = \boldsymbol{\mu}^{(L)}_{\widetilde{x}_d,GNN}$
Initialization: $\boldsymbol{\mu}^{(0)}_{\widetilde{h}} = \widetilde{\mathbf{Y}}_p\widetilde{\mathbf{X}}^H_p(\widetilde{\mathbf{X}}_p\widetilde{\mathbf{X}}^H_p + \widetilde{\sigma}^2\mathbf{I}_{N_t})^{-1}$;
$\forall m,n,t: v^{(0)}_{\widetilde{h},nm} = 1$;
$\mu^{(0)}_{\widetilde{x}_{d,mt}} = 0, v_{\widetilde{x}^{(0)}_{d,mt}} = 1$;
$\hat{s}^{(0)}_{nt} = 0, V^{(0)}_{s,nt} = 1$,;
1 $\forall n,i,j$: Compute $\boldsymbol{\alpha}_{h,ni}$ and $\boldsymbol{\epsilon}_{h,nij}$ using (24);
2 **for** $l = 1$ *to* L **do**
`// The BiG-AMP Module:`
3 Refer to the BiG-AMP module in **Algorithm** 2 but use $\mu^{(l-1)}_{\widetilde{x}_{d,mt},GNN}$, $v^{(l-1)}_{\widetilde{x}_{d,mt},GNN}$, $\mu^{(l-1)}_{\widetilde{h}_{nm},GNN}$ and $v^{(l-1)}_{\widetilde{h}_{nm},GNN}$ when $l > 1$;
`// The Feature Extractor:`
4 Refer to the feature extractor in **Algorithm** 2;
`// The GNN Module:`
5 $\forall t,i$: $\mathbf{u}^{(0)}_{x_{d,it}} = \mathbf{W}_1(\boldsymbol{\alpha}^{(l)}_{x_{d,it}})^T + \mathbf{b}_1$;
6 $\forall n,i$: $\mathbf{u}^{(0)}_{h,ni} = \mathbf{W}_2(\boldsymbol{\alpha}_{h,ni})^T + \mathbf{b}_2$,;
7 **for** $k = 1$ *to* K **do**
8 Refer to the step of line 15 in **Algorithm** 2;
9 $\forall n,i,j$: $\mathbf{m}^{(k)}_{n,i\to j} = \mathcal{M}\left(\mathbf{u}^{(k-1)}_{h,ni}, \mathbf{u}^{(k-1)}_{h,nj}, \boldsymbol{\epsilon}_{h,nij}\right)$;
10 $\forall n,i$: $\mathbf{u}^{(k)}_{h,ni} = \mathcal{U}\left(\mathbf{u}^{(k-1)}_{h,ni}, \left[\sum_{j\in \mathrm{ne}(i)} \mathbf{m}^{(k)}_{n,j\to i}, R^{(l)}_{h,ni}, \Sigma^{(l)}_{h,ni}\right]\right)$;
11 $\forall n,i$: $[R^{(l)}_{h_{ni},GNN} \quad \Sigma^{(l)}_{h_{ni},GNN}] = \mathcal{R}_r(\mathbf{u}^{(K)}_{h,ni})$;
12 $\forall n,m$: $R^{(l)}_{\widetilde{h}_{nm},GNN} = R^{(l)}_{h_{nm},GNN} + j \cdot R^{(l)}_{h_{n(N_t+m)},GNN}$;
13 $\Sigma^{(l)}_{\widetilde{h}_{nm},GNN} = \Sigma^{(l)}_{h_{nm},GNN} + \Sigma^{(l)}_{h_{n(N_t+m)},GNN}$;
14 $\forall n,i$: $\mu^{(l)}_{\widetilde{h}_{nm},GNN} = \mathbb{E}_{\xi^{(l)}_{\widetilde{h}_{nm},GNN}}[\widetilde{h}_{nm}]$;
15 $v^{(l)}_{\widetilde{h}_{nm},GNN} = \mathbb{E}_{\xi^{(l)}_{\widetilde{h}_{nm},GNN}}[(\widetilde{h}_{nm} - \mu^{(l)}_{\widetilde{h}_{nm},GNN})^2]$;
16 Refer to steps of lines 16-17 in **Algorithm** 2;

B. Impact of the GNN

1) Approximation Compensation: As mentioned in Section III-A, the effectiveness of scalar iterative BiG-AMP algorithm relies on the large system limit, i.e., infinite output and input dimensions [29]. However, for a MIMO uplink system with finite transmit and receive antennas, approximation inaccuracy of message passing is inevitable. To solve this problem, our proposed BiGNN-BiGAMP receiver applies GNNs into both branches to exploit the correlation between channel matrix and data matrix, which can be reflected from the output of the marginal likelihoods by incorporating relevant information from RCF and CCF.

2) Overconfidence Alleviation: Typically, a small variance indicates a more accurate estimation. As demonstrated in Fig. 5, the channel estimation variance of our proposed GNN-BiGAMP receiver is obviously smaller than that of BiG-AMP, which benefits from the enhanced data detection in bilinear inference. However, this phenomenon also leads to overconfidence in the model, making it challenging to

$$\begin{pmatrix} \tilde{z}_{11} & \tilde{z}_{12} & \cdots & \tilde{z}_{1T} \\ \tilde{z}_{21} & \tilde{z}_{22} & \cdots & \tilde{z}_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{z}_{N_r1} & \tilde{z}_{N_r2} & \cdots & \tilde{z}_{N_rT} \end{pmatrix} = \begin{pmatrix} \overbrace{\begin{matrix} \tilde{h}_{11} & \tilde{h}_{12} & \cdots & \tilde{h}_{1N_t} \\ \tilde{h}_{21} & \tilde{h}_{22} & \cdots & \tilde{h}_{2N_t} \end{matrix}}^{\widetilde{\mathbf{h}}_1} \\ \begin{matrix} \vdots & \vdots & \ddots & \vdots \\ \tilde{h}_{N_r1} & \tilde{h}_{N_r2} & \cdots & \tilde{h}_{N_rN_t} \end{matrix} \end{pmatrix} \overset{\text{Coupling}}{\leftrightarrow} \begin{pmatrix} \underbrace{\begin{matrix} \tilde{x}_{11} \\ \tilde{x}_{21} \\ \vdots \\ \tilde{x}_{N_t1} \end{matrix}}_{\widetilde{\mathbf{x}}_1} & \begin{matrix} \tilde{x}_{12} & \cdots & \tilde{x}_{1T} \\ \tilde{x}_{22} & \cdots & \tilde{x}_{2T} \\ \vdots & \ddots & \vdots \\ \tilde{x}_{N_t2} & \cdots & \tilde{x}_{N_tT} \end{matrix} \end{pmatrix} \tag{23}$$

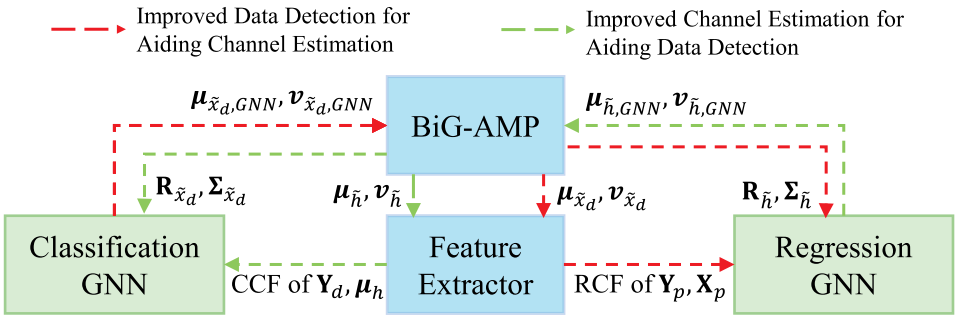


Fig. 4. The structure of the proposed BiGNN-BiGAMP receiver.

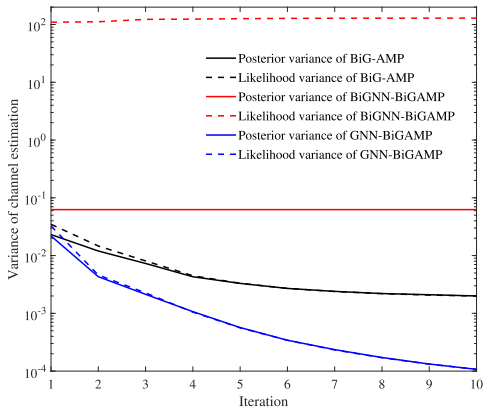


Fig. 5. Variance convergence of different JCD receivers at SNR = 18dB with antenna configuration 32×16.

update the computed messages more accurate, particularly at high signal-to-noise ratio (SNR) [37]. Although our proposed GNN-BiGAMP receiver can avoid the premature convergence generated by Gaussian products in data-detection-loop of BiG-AMP framework [38], the overconfidence problem caused by excessively small variances fails to be addressed.

To tackle this issue, our BiGNN-BiGAMP receiver learns a strategy that can reduce interpretability but significantly improve performance simultaneously. Here, the Rayleigh fading channel is demonstrated as an example, where $\tilde{h}_{nm}$ is drawn from a complex Gaussian distribution $\mathcal{CN}(\tilde{h}_{nm}; 0, \sigma_h^2)$. As illustrated in Fig. 5, the GNN outputs regression results including a high likelihood variance to maintain the posterior variance at such noise level. Accordingly, the GNN designed for regression learns to adjust the outputted marginal likelihood mean to incorporate the amplified variance, developing lines 14-15 of **Algorithm** 3

$$v^{(l)}_{\tilde{h}_{nm},GNN} = \left((\Sigma^{(l)}_{\tilde{h}_{nm},GNN})^{-1} + (\sigma_h^2)^{-1} \right)^{-1} \approx \sigma_h^2, \quad (26a)$$

$$\mu^{(l)}_{\tilde{h}_{nm},GNN} = \frac{R^{(l)}_{\tilde{h}_{nm},GNN}}{\Sigma^{(l)}_{\tilde{h}_{nm},GNN}} v^{(l)}_{\tilde{h}_{nm},GNN} \approx \frac{R^{(l)}_{\tilde{h}_{nm},GNN}}{\Sigma^{(l)}_{\tilde{h}_{nm},GNN}} \sigma_h^2, \quad (26b)$$

with $\Sigma^{(l)}_{\tilde{h}_{nm},GNN} \gg \sigma_h^2$. As a result, the $\mu^{(l)}_{\tilde{h}_{nm},GNN}$ can be updated towards better channel estimation, while $v^{(l)}_{\tilde{h}_{nm},GNN}$ is almost maintained constant to overcome the problem of overconfidence by participating in the computation during other steps.

C. Training Scheme

Due to its iterative structure, end-to-end training is challenging for our proposed BiGNN-BiGAMP receiver.[4] To solve this problem, a step-by-step training scheme is proposed, which loads the pre-trained models and utilizes multitask learning to balance the classification and regression.

As shown in Fig. 6(a), the GNN responsible for classification is first trained by the scheme in III-C with the cross-entropy loss function in Step 1. Then, the GNN responsible for regression in channel estimation branch is trained in Step 2, where RCF information at a high SNR is utilized to effectively improve decoupling ability with reduced noise interference. To be noticed, Step 2 is empowered by a mean square error (MSE) loss function, given by

$$\mathcal{L}_r = \frac{1}{2N_t N_r} \sum_{n=1}^{N_r} \sum_{i=1}^{2N_t} |h_{ni} - \mu^{(L)}_{h_{ni},GNN}|^2, \quad (27)$$

where h_{ni} is the element of real-valued true CSI.

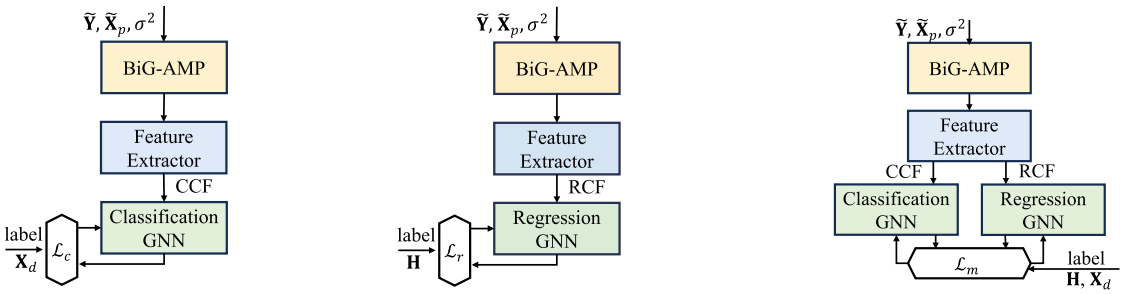
With the two pre-trained models obtained from Step 1 and Step 2, a multitask learning approach is proposed as the final training step shown in Fig. 6(c), where both CCF and RCF information are used as the input of learnable part. To coordinate the tasks of two GNN branches during this step, an uncertainty loss with multiple terms is introduced into the final training step, given by [39]

$$\mathcal{L}_m = \frac{1}{2e^{\gamma_r}} \mathcal{L}_r + \frac{1}{e^{\gamma_c}} \mathcal{L}_c + \frac{\gamma_r}{2} + \frac{\gamma_c}{2}, \quad (28)$$

where γ_r and γ_c are both learnable parameters that are dynamically adjusted during the training process.

It is important to note that the pre-trained GNN models utilized in this training scheme are universal. In other words, once the two models are trained, they can be applied under various conditions, such as different channel states and transmitted block schemes, requiring only the adaptive multitask training outlined in the third step (see Fig. 6(c)), which will be verified through simulation in Section VI.

[4] In the early stage of end-to-end training, two GNNs interact dependently in the iterative structure, leading to instability during the training process [36].



(a) Step 1: Train GNN of classification (b) Step 2: Train GNN of regression (c) Step 3: Multitask learning with pre-trained models

Fig. 6. A step-by-step training scheme for BiGNN-BiGAMP receiver.

V. Complexity Comparisons

In this section, the computational complexity of different receivers is analyzed in terms of number of real-valued multiplications (RVMs), as presented in Table I.[5]

As for GNN-BiGAMP, the whole complexity is determined by three modules, i.e., BiG-AMP, feature extraction and GNN. For BiG-AMP, the RVMs per iteration of lines 2-11 in **Algorithm** 2 is $4(6N_rN_t + 5N_rN_tT_d + 5N_rN_tT + 3N_tT_d + 9N_rT) + 2N_tT_dM$. For feature extraction, the cost per iteration in expressions (16) is $8N_rN_tT_d + 32N_rN_t^2T_d$. The rest of operations belong to the GNN components, including propagation, aggregation and readout with other operations, whose RVMs are respectively $((2N_u+3)N_{h_1}N_{h_1}N_{h_2} + N_uN_{h_2})2N_t(2N_t - 1)KT_d$, $((N_u+2)N_{h_1} + N_{h_1}^2 + N_uN_{h_1})2N_tKT_d$ and $(5N_u + N_uN_{h_1} + N_{h_1}N_{h_2} + N_{h_2}M)2N_tT_d$. As for BiGNN-BiGAMP, the RVMs is calculated by considering the sum of two GNNs' batches, i.e., $T_d + N_r$, as batch size due to the fact that two GNNs have the same structure. However, fewer features are contained in (24) compared to (16). Therefore, the RVMs of BiGNN-BiGAMP have two terms to subtract.

To numerically illustrate the complexity, RVMs of each receiver are calculated and shown in the last column of Table I. As expected, the proposed receivers have a higher complexity compared to conventional JCD algorithms like iterative LMMSE, BiG-AMP, and VB, but their complexity is significantly lower than that of turbo-like deep learning-based OAMPNet2 receiver, showing the superiority of GNN-assisted bilinear inference approach. Moreover, it can be seen that BiGNN-BiGAMP has the similar RVMs compared to GNN-BiGAMP, though it employs another GNN for regression.

VI. Simulation Results

In this section, numerical results are presented to demonstrate the performance of our proposed GNN-BiGAMP and BiGNN-BiGAMP receiver, where the symbol error rate (SER) and the normalized MSE (NMSE) are employed as the performance evaluation metrics of data detection and channel estimation, respectively. To illustrate the superiority of our proposed GNN-BiGAMP and BiGNN-BiGAMP receiver, we adopt four receivers employing the JCD framework as baselines[6]:

- **Iterative LMMSE**: A traditional solution with iterative LMMSE-based channel estimation and LMMSE-based data detection for the JCD problem given in [40].
- **BiG-AMP**: A classical message passing algorithm for the JCD problem given in [11].
- **VB**: An uplink MIMO receiver developed by variational Bayes framework proposed in [12].
- **OAMPNet2**: A model-driven deep learning MIMO receiver proposed in [20], developed by a turbo-like JCD architecture containing LMMSE channel estimator and OAMP detector with learnable parameters.

A. Implementation Details

Throughout the experiments, we assume the transmission of 16-QAM for both pilot and data sequences, setting the system parameters of a block as $T = 160$. Moreover, a Rayleigh block fading channel is considered, i.e., channel coefficients are generated from $\mathcal{CN}(\widetilde{h}_{nm}; 0, 1)$, and the SNR of the system is defined as SNR $= \|\widetilde{\mathbf{H}}\widetilde{\mathbf{X}}\|_2^2 / \|\widetilde{\mathbf{W}}\|_2^2$.

For comparison, All receivers containing the GNN module adopt similar parameters, where the numbers of inner GNN and outer BiG-AMP iterations are set to be $K = 2$ and $L = 10$, respectively. Furthermore, the GNN modules of both classification and regression tasks have layers of the same dimensions. More specifically, the output sizes of three MLP layers for $\mathcal{M}(\cdot)$ are N_{h_1}, N_{h_2}, and N_u, respectively. The output sizes of the hidden layers in $\mathcal{U}(\cdot)$ are N_{h_1}, N_u. The output sizes of first two MLP layers in $\mathcal{R}(\cdot)$ are N_{h_1}, N_{h_2}, respectively. According to [25], we set $N_u = 8$, $N_{h_1} = 64$, and $N_{h_2} = 32$. The number of iterations of iterative LMMSE, BiG-AMP, VB and OAMPNet2 is 10. All simulations are executed on a desktop with GPU NVIDIA GeForce RTX 4090 and Pytorch 2.0.0 environment.

As for the training details, we employ the Adam optimizer to train our proposed receivers. The learning rate is initially set to 10^{-3} and reduces every 50 epochs by a decay factor of 0.91. Once the learning rate drops to 10^{-4}, the dropping stops in the sequential epochs. For our proposed GNN-BiGAMP receiver,

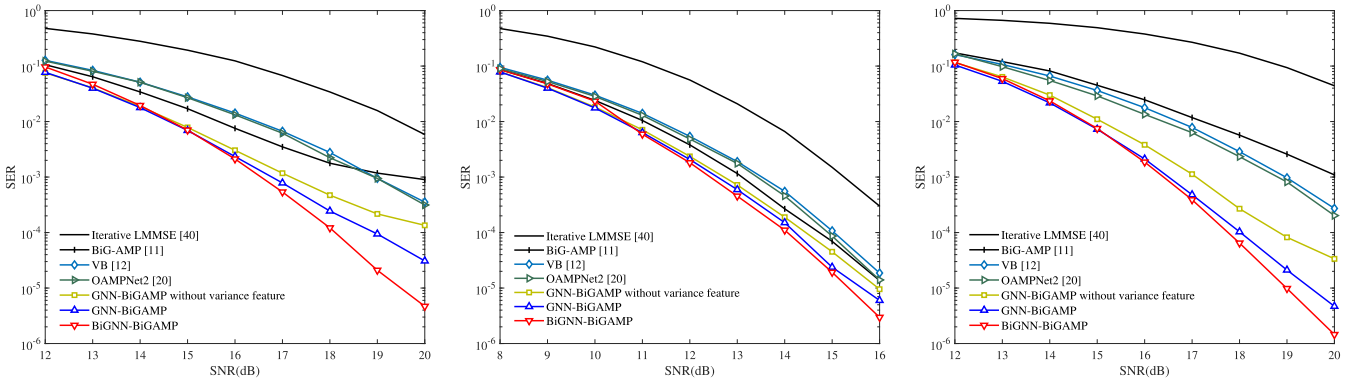
[5]Due to the fact that the training of GNN-based receivers can be realized offline, only the complexity of the online operation is considered here.

[6]Due to the fact that a further increasing iteration number only offers a slight performance gain, the same number of iterations in four baseline methods was set for fair comparison.

TABLE I
COMPUTATIONAL COMPLEXITY COMPARISON OF DIFFERENT RECEIVERS

Receiver	Number of RVMs per iteration for a transmit block	RVMs of 16-QAM $N_r = 32$ $N_t = 16^*$
Iterative LMMSE [40]	$4(N_rN_t^2 + N_rN_t(T_d+T) + \frac{10}{3}N_t^3 + N_t^2(T_d+2T) + N_tT_d\widetilde{M})$	1.27×10^6
BiG-AMP [11]	$4(6N_rN_t + 5N_rN_tT_d + 5N_rN_tT + 3N_tT_d + 9N_rT + 4N_tT_d\widetilde{M})$	3.70×10^6
VB [12]	$4(N_t^2 + 3N_rN_tT_d + 6N_tT_d\widetilde{M} + N_tT_d + 3N_rN_t + 2N_rN_tT + N_rN_t^2 + N_t^3 + N_t^2 + 2N_t + N_t^2T_d)$	2.43×10^6
OAMPNet2 [20]	$(8N_r^3N_t^2T_p + 2(8N_r^3N_t^3 + 4N_r^2N_t^2 + 4N_rN_t) + 8N_r^2N_tT_p^2 + 4N_r^2N_tT_p + 4N_r^2N_t^2 + 30N_rN_t)L_1 + 8N_r^2N_t + 8N_t^3T + 4N_t^2T + (8N_r^2T_d + 8N_r^3T_d + 20N_rN_tT_d + 12N_rN_t^2T_d + 2N_rT_d + 2T_d + 8N_t^3T_d + 18N_tT_d)L_1$	2.69×10^{10}
GNN-BiGAMP	$4(6N_rN_t + 5N_rN_tT_d + 5N_rN_tT + 3N_tT_d + 9N_rT) + 2N_tT_dM + (8N_rN_tT_d + 32N_rN_t^2T_d) + ((2N_u+3)N_{h_1} + N_{h_1}N_{h_2} + N_uN_{h_2})2N_t(2N_t-1)KT_d + ((N_u+2)N_{h_1} + N_{h_1}^2 + N_uN_{h_1})2N_tKT_d + (5N_u + N_uN_{h_1} + N_{h_1}N_{h_2} + N_{h_2}M)2N_tT_d$	3.72×10^8
BiGNN-BiGAMP	$4(6N_rN_t + 5N_rN_tT_d + 5N_rN_tT + 3N_tT_d + 9N_rT) + 2N_tT_dM + (8N_rN_tT_d + 32N_rN_t^2T_d) + (((2N_u+3)N_{h_1} + N_{h_1}N_{h_2} + N_uN_{h_2})2N_t(2N_t-1)K(T_d+N_r) - N_{h_1}2N_t(2N_t-1)KN_r) + ((N_u+2)N_{h_1} + N_{h_1}^2 + N_uN_{h_1})2N_tK(T_d+N_r) + ((5N_u + N_uN_{h_1} + N_{h_1}N_{h_2} + N_{h_2}M)2N_t(T_d+N_r) - 4N_uN_tN_r)$	4.53×10^8

* The transmitted block structure are set as $T_p = 32$, $T_d = 128$ and $T = 160$, while the parameters of GNN are set as $N_u = 8$, $N_{h_1} = 64$, $N_{h_2} = 32$ and $K = 2$. The inner iteration number of OAMPNet2 is set as $L_1 = 6$ [20].



(a) 16×8, $T_p = 32$ (b) 32×8, $T_p = 32$ (c) 32×16, $T_p = 32$

Fig. 7. The SER performance comparison in Rayleigh block fading channel.

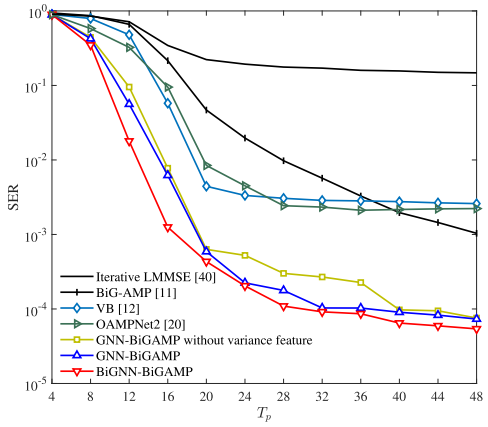


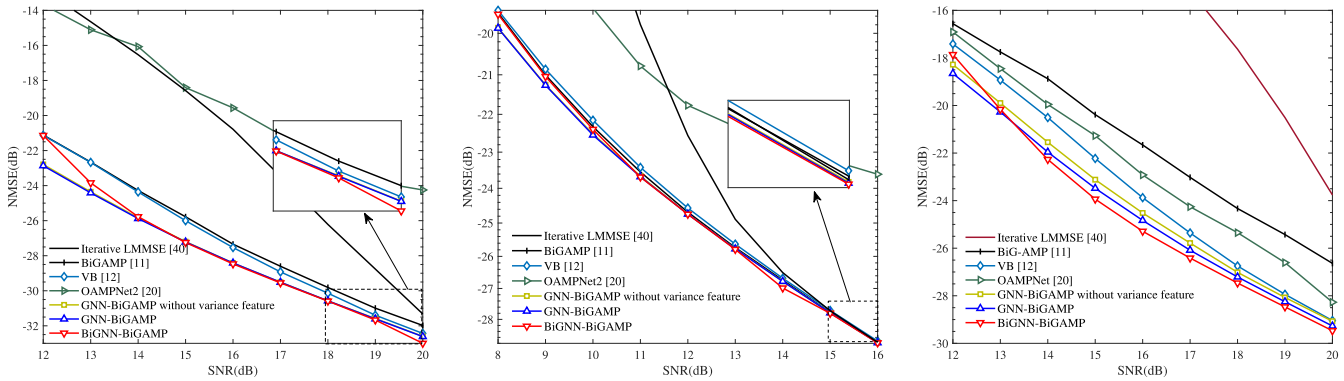
Fig. 8. The SER performance comparison with respect to T_p at SNR = 18dB with antenna configuration 32×16.

the training data is uniformly sampled in the SNR range $[5, 30]$dB. Utilizing the training scheme in Section III-C, this receiver undergoes a training process spanning 100 epochs. Each epoch processes a data set consisting of 10^5 samples, which are segmented into the transmitted blocks. For the proposed BiGNN-BiGAMP receiver, we train the pre-trained model of classification as presented in Fig. 6(a) with the details above with $T_p = 32$. Then, step 2 in Fig. 6(b) is implemented to train the pre-trained model of regression for 100 epochs specifically at SNR = 20dB with $T_p = 32$. Afterwards, the multitask learning step in Fig. 6 is engaged by loading these two pre-trained models and continuing the process for additional 50 epochs. To be noted, the pre-trained models for the BiGNN-BiGAMP in the Kronecker correlated channel in Fig. 11 and Fig. 12 are trained on the Rayleigh block fading channel to verify the adaptation of the proposed training scheme over different channel states.

B. Performance of Data Detection

To show the superiority of our proposed receivers in terms of data detection during JCD, Fig. 7 compares the SER performance achieved by different receivers under different MIMO configurations.

Obviously, our proposed GNN-assisted receivers significantly outperform the classical iterative LMMSE, BiG-AMP and VB receivers and learning-based OAMPNet2 regardless of system configuration, especially at high SNR regimes. When the ratio of the number of receive antennas to transmit antennas (i.e., N_r/N_t) remains constant, SER performance is



(a) 16×8, $T_p = 32$ (b) 32×8, $T_p = 32$ (c) 32×16, $T_p = 32$

Fig. 9. The NMSE performance comparison in Rayleigh block fading channel.

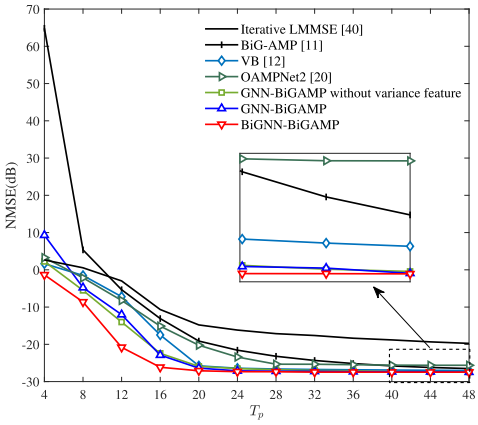


Fig. 10. The NMSE performance comparison with respect to T_p at SNR $= 18$dB with antenna configuration 32×16.

supposed to improve as the increase of system size. However, by comparing Fig. 7(a) with Fig. 7(c), it can be observed that detection performance of the conventional BiG-AMP receiver does not improve in larger systems, which is attributed to the degraded channel estimation performance with the same transmission block setting. With the integration of GNN, our proposed receivers based on the BiG-AMP framework, particularly the BiGNN-BiGAMP receiver, can mitigate the limitations associated with insufficient pilot length in large-scale systems. When N_r/N_t becomes larger, the performance gaps between five receivers become narrower.[7] Nevertheless, 1dB performance gain can still be obtained at the SER of 10^{-5} for our proposed receivers when compared to BiG-AMP, VB and OAMPNet2.

To show how the variance features influence the receiver performance, the SER curve of GNN-BiGAMP without variance features is provided. Obviously, the information uncertainty will result in similar confidence levels being assigned to CCFs with varying accuracy, thus leading to significant performance degradation for GNN-BiGAMP without variance features. As shown in Fig.7, the SER performance of the GNN-BiGAMP outperforms that of GNN-BiGAMP without variance features over all SNRs, where about 1dB performance gain can be obtained at the SER of 10^{-4} in Fig. 7(a) and Fig. 7(c).

Additionally, as depicted in Fig. 7, the SER performance of our proposed BiGNN-BiGAMP outperforms GNN-BiGAMP at high SNRs. For example, when the antenna configuration is 32×16, over 0.5dB performance improvement can be realized by BiGNN-BiGAMP when compared to GNN-BiGAMP at the SER of 10^{-5}, which validates that the introduction of GNN in channel-estimation-loop can also improve the data detection performance.

To show the impact of pilot length, Fig. 8 is presented, which illustrates the SER performance of different JCD receivers with respect to T_p when SNR $= 18$dB. It can be seen that the proposed receivers based on GNN-assisted bilinear inference outperform the iterative LMMSE, BiG-AMP and VB receivers, as well as the OAMPNet2, regardless of the pilot length. Notably, our proposed receivers, especially the BiGNN-BiGAMP model, require fewer pilots to achieve a satisfactory SER performance. For example, as shown in Fig. 8, the BiGNN-BiGAMP receiver only requires approximately 13 pilots to achieve the SER of 10^{-2}, which reduces about $1/3$, $1/3$ and $4/7$ pilots compared to the OAMPNet2, VB and conventional BiG-AMP receivers, respectively. Consequently, our proposed JCD receivers can dramatically enhance spectral efficiency while maintaining similar detection performance.

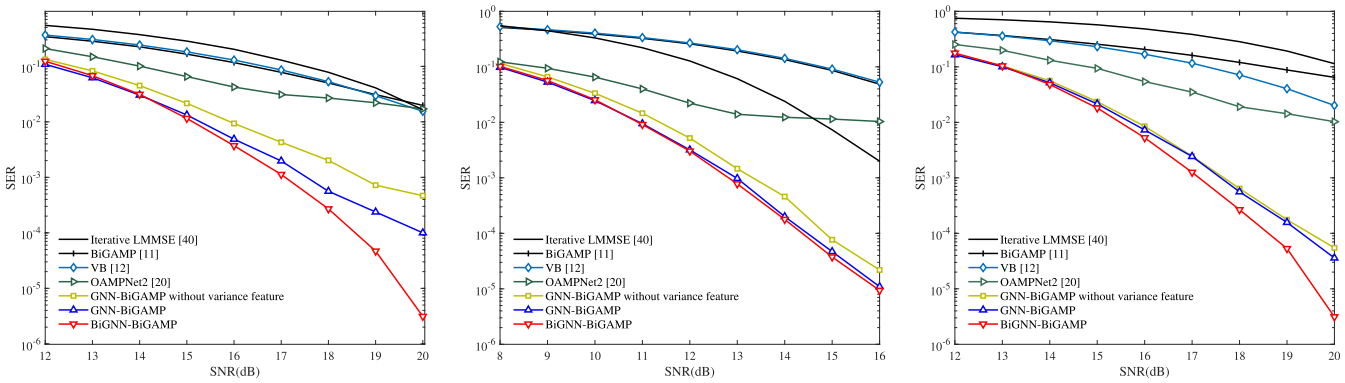
C. Performance of Channel Estimation

Figure 9 shows the channel estimation performance, where the performance of MIMO systems configured with 16×8, 32×8, and 32×16 is provided. In particular, NMSE is defined as

$$\text{NMSE} = \frac{1}{N_b} \sum_{q=1}^{N_b} \frac{\|\mathbf{H} - \hat{\mathbf{H}}\|_F^2}{\|\mathbf{H}\|_F^2}, \tag{29}$$

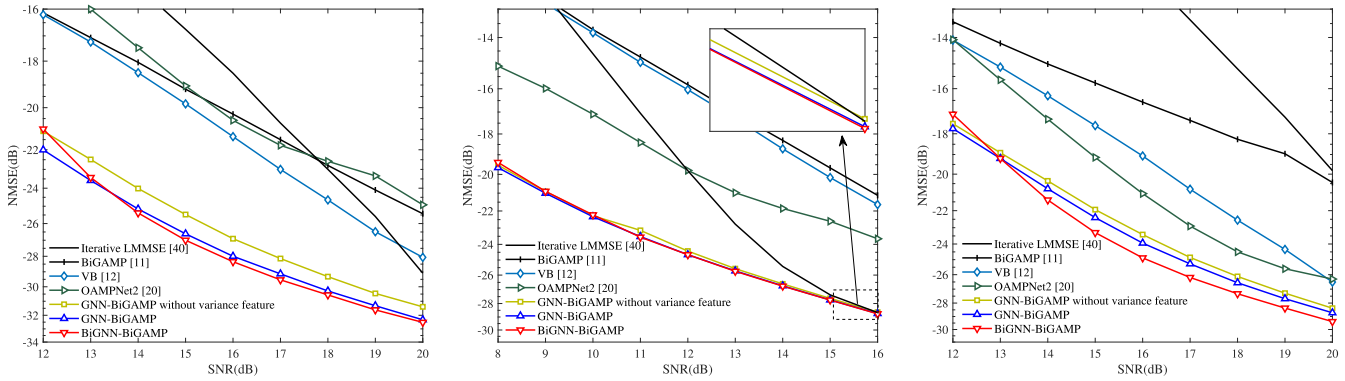
where N_b is the number of transmitted symbol blocks. It can be observed that the proposed GNN-BiGAMP and BiGNN-BiGAMP always outperform the iterative LMMSE, BiG-AMP,

[7]Since increasing the N_r/N_t ratio further approximately satisfies the LSL [29], the inaccuracy caused by the approximation from LSL in conventional JCD algorithms is mitigated, thus leading to slighter performance loss that can be compensated by the GNN.



(a) 16×8, $T_p = 32$ (b) 32×8, $T_p = 32$ (c) 32×16, $T_p = 32$

Fig. 11. The SER performance comparison in Kronecker correlated channel.



(a) 16×8, $T_p = 32$ (b) 32×8, $T_p = 32$ (c) 32×16, $T_p = 32$

Fig. 12. The NMSE performance comparison in Kronecker correlated channel.

VB and OAMPNet2 receivers,[8] where about 2dB gain can be achieved by our proposed receivers compared to BiG-AMP at the NMSE of -20dB as illustrated in Fig. 9(c). Besides, the NMSE performance of GNN-BiGAMP and BiGNN-BiGAMP turns to be the same as N_r/N_t increases. In addition, when the pilot is sufficient in the small size systems, such as Fig. 9(a), the performance gain narrows yet remains discernible, with around 1dB gain at the NMSE of -30dB. Moreover, the channel estimation performance of BiGNN-BiGAMP is poorer than GNN-BiGAMP at low SNR regimes, which is attributed to the first two steps of step-by-step training shown in Fig. 6(b) and Fig. 6(c). However, when the SNR is high, the superior decoupling ability of BiGNN-BiGAMP can be demonstrated, thus making BiGNN-BiGAMP outperform GNN-BiGAMP.

Figure 10 illustrates the channel estimation performance of different JCD receivers with respect to the number of pilots. The GNN-assisted bilinear inference notably enhances the performance of BiG-AMP based framework but requires fewer pilots. As we can see, the GNN-BiGAMP and BiGNN-BiGAMP only need $T_p = 20$ pilots to realize a satisfactory performance, while the BiG-AMP receiver requires up to $T_p = 48$ and OAMPNet2 requires up to $T_p = 28$ to reach similar performance. In comparison, both the VB and GNN-BiGAMP based receivers achieve satisfactory performance with $T_p = 20$, as evidenced in Fig 10. Meanwhile, BiGNN-BiGAMP can work well with only $T_p = 16$, as shown in Fig. 10. This highlights the significant benefits of incorporating dual GNN-assisted loops in BiGNN-BiGAMP, particularly in scenarios involving short transmission blocks with a small number of pilots.

D. Robustness to Spatial Correlation

To show the robustness of our proposed GNN-assisted receivers to spatial correlation, a Kronecker correlated channel is considered, where channel coefficients are generated from $\mathcal{CN}(\widetilde{h}_{nm}; 0, 1)$ and the correlated coefficient of Kronecker channel is $\rho = 0.3$ to make its elements non-i.i.d [41].

Figure 11 demonstrates the SER comparison among different receivers in Kronecker correlated channel. Obviously, the performance gaps between the GNN-assisted receivers and their counterparts, i.e., iterative LMMSE, BiG-AMP, VB, and OAMPNet2, are magnified regardless of MIMO system configurations compared to the Rayleigh fading channels. In contrast to the serious performance deterioration observed in classical BiG-AMP and VB receivers, our proposed GNN-assisted

[8]Since OAMPNet2 fails to balance the training losses from JCD and prematurely utilizes inaccurate data detection results during the channel estimation process in its turbo-like architecture, its channel estimation performance deteriorates severely when pilots are sufficient, as shown in Fig. 9(a) and Fig. 9(b).

receivers can still work efficiently by handling the inaccuracy caused by the separable PDF assumption of the bilinear inference framework by utilizing the adaptive compensation from GNNs. For instance, as presented in Fig. 11(c), our proposed GNN-BiGAMP and BiGNN-BiGAMP outperform iterative LMMSE, BiG-AMP, VB, and OAMPNet2 more than 5dB at SER $= 10^{-2}$, being much larger than 2dB performance gain in Fig. 7(c). Besides, the performance improvement compared to the OAMPNet2 can verify that our proposed GNN-assisted approaches can leverage the adaptability of deep learning more effectively than the deep unfolding schemes with some trainable parameters in JCD problems.

Figure 12 demonstrates the channel estimation performance in Kronecker correlated channel. It is clear that the proposed GNN-BiGAMP and BiGNN-BiGAMP still outperform the iterative LMMSE, BiG-AMP, VB, and OAMPNet2 regardless of antenna configurations, e.g., more than 4dB gain at the NMSE of -24dB in Fig. 12(a), more than 5dB gain at the NMSE of -22dB in Fig. 12(b) and more than 2dB gain at the NMSE of -22dB in Fig. 12(c). Similar to the detection performance in Fig. 11, the performance gaps between the GNN-assisted receivers and their counterparts are significantly magnified due to the fact the learnable GNN parts can handle the inaccuracy caused by separable PDF assumption of bilinear inference framework.

VII. Conclusion

In this paper, to enhance the performance of joint channel estimation and data detection, a GNN-assisted bilinear inference approach was developed. Based on the approximate message passing in the factor graph, the classical BiG-AMP framework was analyzed to iteratively solve the bilinear inference problems. Motivated by the approximation and criterion to derive the BiG-AMP framework, a GNN-BiGAMP receiver was proposed by integrating a GNN module into the data-detection-loop, which could compensate the inaccurate marginal likelihood estimation and deliver this improvement with message flowing to channel-estimation-loop. According to the coupling between channel and symbols, a BiGNN-BiGAMP receiver was further proposed by developing bilinear GNN-assisted posterior estimation loops with two GNNs sharing the same graph representation. To guarantee the convergence of the training of BiGNN-BiGAMP, a step-by-step training scheme empowered by multitask learning was further developed to balance the classification and regression problem in JCD. Simulation results demonstrated that our proposed receivers can achieve significantly better performance compared to their JCD counterparts.

Appendix
Derivation of Variance Features

To derive the variance features in (16), the uncertainty in feature vectors needs to be analyzed. Since the elements of $\mathbf{H}$ have separable PDFs, the column coupling feature between observation and channel information, i.e., $\mathbf{y}_{d,t}^T\mathbf{h}_i^{(l)} = \sum_{q=1}^{2N_r} y_{d,qt}h_{qi}^{(l)}$, has a corresponding variance to measure its uncertainty, given by

$$F_{1,it}^{(l)} = \mathrm{Var}\left[\sum_{q=1}^{2N_r} y_{d,qt}h_{qi}^{(l)}\right] = \sum_{q=1}^{2N_r}\mathrm{Var}\left[y_{d,qt}h_{qi}^{(l)}\right] = (\mathbf{y}_{d,t}^2)^T\mathbf{v}_{h,i}^{(l)}. \tag{30}$$

Similarly, for the self column coupling of channel information, i.e., $(\mathbf{h}_i^{(l)})^T\mathbf{h}_i^{(l)} = \sum_{q=1}^{2N_r}(h_{qi}^{(l)})^2$, the variance feature can be derived by

$$\begin{aligned} F_{2,i}^{(l)} &= \mathrm{Var}\left[\sum_{q=1}^{2N_r}(h_{qi}^{(l)})^2\right] = \sum_{q=1}^{2N_r}\mathrm{Var}\left[(h_{qi}^{(l)})^2\right] \\ &= \sum_{q=1}^{2N_r}\left(\mathbb{E}\left[(h_{qi}^{(l)})^4\right] - \mathbb{E}^2\left[(h_{qi}^{(l)})^2\right]\right) \\ &= 2\left((\mathbf{v}_{h,i}^{(l)})^T\mathbf{v}_{h,i}^{(l)} + \left((\mathbf{h}_i^{(l)})^2\right)^T\mathbf{v}_{h,i}^{(l)} + (\mathbf{v}_{h,i}^{(l)})^T(\mathbf{h}_i^{(l)})^2\right), \end{aligned} \tag{31}$$

where $\mathbb{E}\left[(h_{qi}^{(l)})^4\right]$ can be calculated according to the moment generating function [42].

Moreover, to measure the uncertainty of edge feature, the inter-column coupling $-(\mathbf{h}_i^{(l)})^T\mathbf{h}_j^{(l)} = -\sum_{q=1}^{2N_r} h_{qi}^{(l)}h_{qj}^{(l)}$ can obtain the corresponding variance feature, given by

$$\begin{aligned} F_{3,ij}^{(l)} &= \mathrm{Var}\left[-\sum_{q=1}^{2N_r} h_{qi}^{(l)}h_{qj}^{(l)}\right] = \sum_{q=1}^{2N_r}\mathrm{Var}\left[h_{qi}^{(l)}h_{qj}^{(l)}\right] \\ &= \sum_{q=1}^{2N_r}\left(\mathbb{E}\left[(h_{qi}^{(l)})^2\right]\mathbb{E}\left[(h_{qj}^{(l)})^2\right] - \mathbb{E}^2\left[h_{qi}^{(l)}\right]\mathbb{E}^2\left[h_{qj}^{(l)}\right]\right) \\ &= (\mathbf{v}_{h,i}^{(l)})^T\mathbf{v}_{h,j}^{(l)} + \left((\mathbf{h}_i^{(l)})^2\right)^T\mathbf{v}_{h,j}^{(l)} + (\mathbf{v}_{h,i}^{(l)})^T(\mathbf{h}_j^{(l)})^2. \end{aligned} \tag{32}$$

References

[1] G. Gui, M. Liu, F. Tang, N. Kato, and F. Adachi, "6G: Opening new horizons for integration of comfort, security, and intelligence," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 126–132, Oct. 2020.

[2] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, Aug. 2019.

[3] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive MIMO detection techniques: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3109–3132, 4th Quart., 2019.

[4] H. Wu, L. Lu, and Z. Wang, "Near-field channel estimation in dual-band XL-MIMO with side information-assisted compressed sensing," *IEEE Trans. Commun.*, early access, Aug. 16, 2024, doi: 10.1109/TCOMM.2024.3445282.

[5] Y. Chen, Y. Wang, Z. Wang, and Z. Han, "Angular-distance based channel estimation for holographic MIMO," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 6, pp. 1684–1702, Jun. 2024.

[6] L. Lu, Z. Wang, Z. Gao, S. Chen, and H. V. Poor, "Block-sparse tensor recovery," *IEEE Trans. Inf. Theory.*, vol. 70, no. 12, pp. 9293–9326, Dec. 2024.

[7] X. Guo, Y. Chen, and Y. Wang, "Compressed channel estimation for near-field XL-MIMO using triple parametric decomposition," *IEEE Trans. Veh. Technol.*, vol. 72, no. 11, pp. 15040–15045, Nov. 2023.

[8] Y. Ma, N. Wu, W. Yuan, D. W. K. Ng, and L. Hanzo, "Joint channel estimation and equalization for index-modulated spectrally efficient frequency division multiplexing systems," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6230–6244, Oct. 2020.

[9] Y. Zhang, B. Li, N. Wu, Y. Ma, W. Yuan, and L. Hanzo, "Message passing-aided joint data detection and estimation of nonlinear satellite channels," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 1763–1774, Feb. 2023.

[10] S. Wu, L. Kuang, Z. Ni, D. Huang, Q. Guo, and J. Lu, "Message-passing receiver for joint channel estimation and decoding in 3D massive MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8122–8138, Dec. 2016.

[11] C.-K. Wen, C.-J. Wang, S. Jin, K.-K. Wong, and P. Ting, "Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2541–2556, May 2016.

[12] S. S. Thoota and C. R. Murthy, "Variational Bayes' joint channel estimation and soft symbol decoding for uplink massive MIMO systems with low resolution ADCs," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3467–3481, May 2021.

[13] L. Wei et al., "Joint channel estimation and signal recovery for RIS-empowered multiuser communications," *IEEE Trans. Commun.*, vol. 70, no. 7, pp. 4640–4655, Jul. 2022.

[14] Z. Qin, H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep learning in physical layer communications," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 93–99, Apr. 2019.

[15] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2224–2287, 3rd Quart., 2019.

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

[17] 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.

[18] M. Honkala, D. Korpi, and J. M. J. Huttunen, "DeepRx: Fully convolutional deep learning receiver," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3925–3940, Jun. 2021.

[19] Z. Zhao, M. C. Vuran, F. Guo, and S. D. Scott, "Deep-waveform: A learned OFDM receiver based on deep complex-valued convolutional networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2407–2420, Aug. 2021.

[20] 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.

[21] 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.

[22] D. Bahn, J.-I. Jung, J.-I. Jang, C. Yoon, C. Park, and I. Park, "MTCNet: Multi-task complex network for concurrent channel estimation and equalization," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Madrid, Spain, Dec. 2021, pp. 1–6.

[23] Y. Yang, S. Zhang, F. Gao, J. Ma, and O. A. Dobre, "Graph neural network-based channel tracking for massive MIMO networks," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1747–1751, Aug. 2020.

[24] A. Scotti, N. N. Moghadam, D. Liu, K. Gafvert, and J. Huang, "Graph neural networks for massive MIMO detection," in *Proc. Int. Conf. Mach. Learn.*, Vienna, Austria, Jan. 2020, pp. 1–5.

[25] A. Kosasih, V. Onasis, V. Miloslavskaya, W. Hardjawana, V. Andrean, and B. Vucetic, "Graph neural network aided MU-MIMO detectors," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 9, pp. 2540–2555, Sep. 2022.

[26] Z. Liu, D. He, N. Wu, Q. Yan, and Y. Li, "Model-driven IEP-GNN framework for MIMO detection with Bayesian optimization," *IEEE Wireless Commun. Lett.*, vol. 13, no. 2, pp. 387–391, Feb. 2024.

[27] Q. Lu, S. Luan, and X.-W. Chang, "GCEPNet: Graph convolution-enhanced expectation propagation for massive MIMO detection," 2024, *arXiv:2404.14886*.

[28] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, vol. 2. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.

[29] 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.

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

[31] Q. Zou, H. Zhang, and H. Yang, "Multi-layer bilinear generalized approximate message passing," *IEEE Trans. Signal Process.*, vol. 69, pp. 4529–4543, 2021.

[32] H. Artes, D. Seethaler, and F. Hlawatsch, "Efficient detection algorithms for MIMO channels: A geometrical approach to approximate ml detection," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2808–2820, Nov. 2003.

[33] K. Yoon et al., "Inference in probabilistic graphical models by graph neural networks," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2019, pp. 868–875.

[34] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[35] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 12, pp. 5586–5609, Dec. 2022.

[36] T. Glasmachers, "Limits of end-to-end learning," in *Proc. Asian Conf. Mach. Learn. (ACML)*, Seoul, (South) Korea, 2017, pp. 17–32.

[37] G. Colavolpe, D. Fertonani, and A. Piemontese, "SISO detection over linear channels with linear complexity in the number of interferers," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1475–1485, Dec. 2011.

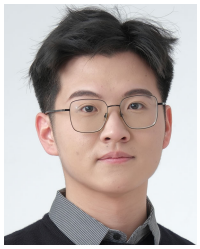
[38] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[39] R. Cipolla, Y. Gal, and A. Kendall, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, Jun. 2018, pp. 7482–7491.

[40] S. Buzzi, M. Lops, and S. Sardellitti, "Performance of iterative data detection and channel estimation for single-antenna and multiple-antennas wireless communications," *IEEE Trans. Veh. Technol.*, vol. 53, no. 4, pp. 1085–1104, Jul. 2004.

[41] K. Maki and T. Ikegami, "A from-scratch approach to deep unfolding-aided bilinear Gaussian belief propagation for correlated large MIMO," in *Proc. IEEE Global Commun. Conf.*, Kuala Lumpur, Malaysia, Dec. 2023, pp. 5671–5676.

[42] C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed., Secaucus, NJ, USA: Springer-Verlag, 2006.



Zishen Liu received the B.S. degree in electronics and information engineering from Beijing Institute of Technology (BIT), Beijing, China, in 2022, where he is currently pursuing the M.S. degree in information and communication engineering with the School of Information and Electronics. His current research interests include massive MIMO systems, Bayesian inference, and AI-empowered wireless communications. He was a recipient of the National Scholarship in 2021 and 2024. He serves as a reviewer for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and IEEE WIRELESS COMMUNICATIONS LETTERS.

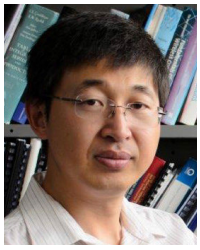
Nan Wu (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees from Beijing Institute of Technology (BIT), Beijing, China, in 2003, 2005, and 2011, respectively.

He was a Visiting Ph.D. Student with the Department of Electrical Engineering, The Pennsylvania State University, University Park, PA, USA, from 2008 to 2009. He is currently a Professor with the School of Information and Electronics, BIT. His research interests include signal processing in wireless communication networks.

He was a recipient of the National Excellent Doctoral Dissertation Award by MOE of China in 2013. He serves as an Editorial Board Member for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and IEEE WIRELESS COMMUNICATIONS LETTERS.

Dongxuan He (Member, IEEE) received the B.S. degree in automation and the Ph.D. degree in information and communication systems from Beijing Institute of Technology (BIT), China, in 2013 and 2019, respectively. From 2017 to 2018, he was a Visiting Student with Singapore University of Technology and Design (SUTD). From 2019 to 2022, he was a Post-Doctoral Researcher with the Department of Electronic Engineering, Tsinghua University. He is currently an Assistant Professor with the School of Information and Electronics, BIT. His current interests include integrated sensing and communication (ISAC), terahertz communication, and AI empowered wireless communications. He is currently serving as a Guest Editor for the IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY, *Electronics*, and *Space: Science and Technology*. He was also an Exemplary Reviewer of IEEE WIRELESS COMMUNICATIONS LETTERS.

Weijie Yuan (Senior Member, IEEE) received the Ph.D. degree from the University of Technology Sydney, Australia, in 2019. In 2016, he was a Visiting Ph.D. Student with the Institute of Telecommunications, Vienna University of Technology, Austria. He was a Research Assistant with The University of Sydney, a Visiting Associate Fellow with the University of Wollongong, and a Visiting Fellow with the University of Southampton, from 2017 to 2019. From 2019 to 2021, he was a Research Associate with the University of New South Wales. He is currently an Assistant Professor with Southern University of Science and Technology. He currently serves as an Associate Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, *IEEE Communications Standards Magazine*, IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, IEEE COMMUNICATIONS LETTERS, IEEE OPEN JOURNAL OF COMMUNICATIONS SOCIETY, and *EURASIP Journal on Advances in Signal Processing*. He is the Lead Editor for the series on ISAC in *IEEE Communications Magazine*. He was an Organizer/the Chair of several workshops and special sessions on OTFS and ISAC in flagship IEEE and ACM conferences, including IEEE ICC, IEEE/CIC ICCC, IEEE SPAWC, IEEE VTC, IEEE WCNC, IEEE ICASSP, and ACM MobiCom. He is the Founding Chair of the IEEE ComSoc Special Interest Group on OTFS (OTFS-SIG). He has been listed in the World's Top 2% Scientists by Stanford University for citation impact since 2021. He was a recipient of the Best Editor Award from IEEE COMMUNICATIONS LETTERS, the Exemplary Reviewer from IEEE TRANSACTIONS ON COMMUNICATIONS/IEEE WIRELESS COMMUNICATIONS LETTERS, and the Best Paper Award from IEEE ICC 2023, IEEE/CIC ICCC 2023, and IEEE GlobeCom 2024.



Yonghui Li (Fellow, IEEE) received the Ph.D. degree from Beijing University of Aeronautics and Astronautics, Beijing, China, in November 2002.

Since 2003, he has been with the Centre of Excellence in Telecommunications, The University of Sydney, Australia. He is currently a Professor and the Director of the Wireless Engineering Laboratory, School of Electrical and Information Engineering, The University of Sydney. His current research interests include wireless communications, with a particular focus on MIMO, millimeter wave communications, machine-to-machine communications, coding techniques, and cooperative communications. He holds a number of patents granted and pending in these fields. He was a recipient of Australian Queen Elizabeth II Fellowship in 2008 and Australian Future Fellowship in 2012. He was also received the Best Paper Award from the IEEE International Conference on Communications (ICC) in 2014, the IEEE PIRMC 2017, and the IEEE Wireless Days Conferences (WD) in 2014. He was an Editor of IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He also served as the Guest Editor for several IEEE journals, such as IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, *IEEE Communications Magazine*, IEEE INTERNET OF THINGS JOURNAL, and IEEE ACCESS.

Tony Q. S. Quek (Fellow, IEEE) received the B.E. and M.E. degrees in electrical and electronics engineering from Tokyo Institute of Technology in 1998 and 2000, respectively, and the Ph.D. degree in electrical engineering and computer science from Massachusetts Institute of Technology in 2008.

Currently, he is the Cheng Tsang Man Chair Professor with Singapore University of Technology and Design (SUTD) and a ST Engineering Distinguished Professor. He is also the Director of the Future Communications Research and Development Programme, the Head of ISTD Pillar, and the AI on RAN Working Group Chair in AI-RAN Alliance. His current research interests include wireless communications and networking, network intelligence, non-terrestrial networks, open radio access networks, and 6G. He is a fellow of WWRF and the Academy of Engineering Singapore. He was honored with the 2008 Philip Yeo Prize for Outstanding Achievement in Research, the 2012 IEEE William R. Bennett Prize, the 2015 SUTD Outstanding Education Awards–Excellence in Research, the 2016 IEEE Signal Processing Society Young Author Best Paper Award, the 2017 CTTC Early Achievement Award, the 2017 IEEE ComSoc AP Outstanding Paper Award, the 2020 IEEE Communications Society Young Author Best Paper Award, the 2020 IEEE Stephen O. Rice Prize, the 2020 Nokia Visiting Professor, the 2022 IEEE Signal Processing Society Best Paper Award, and the 2024 IIT Bombay International Award For Excellence in Research in Engineering and Technology.