Model-Driven IEP-GNN Framework for MIMO Detection With Bayesian Optimization

Zishen Liu, Dongxuan He, *Member, IEEE*, Nan Wu, *Member, IEEE*, Qinsiwei Yan, and Yonghui Li, *Fellow, IEEE*

***Abstract*—In this letter, a model-driven detector called IEP-GNN is proposed for massive multiple-input and multiple-output (MIMO) systems. Graph neural network (GNN) and improved moment matching (IMM) are integrated into the expectation propagation (EP) algorithm to improve the accuracy of posterior distribution approximation and leverage the self-correction ability of EP algorithm. Moreover, to acquire the training experiences and optimize initial parameters, hotbooting and Bayesian parameter optimization (BPO) are employed respectively, which can further improve the performance of the proposed IEP-GNN. Simulation results show that our proposed IEP-GNN with BPO outperforms other state-of-the-art EP-based detectors while maintaining an acceptable convergence and computational complexity.**

***Index Terms*—Model-driven, expectation propagation, graph neural network, Bayesian parameter optimization.**

I. Introduction

MASSIVE multiple-input and multiple-output (MIMO) has shown great potential in achieving remarkable spectral efficiency, making it a promising candidate in the upcoming sixth-generation (6G) [1]. However, traditional detection methods, such as maximum-likelihood (ML) sequence detectors, are computationally infeasible for massive MIMO due to their prohibitive complexity, especially when the transmitters and receivers are equipped with tens or even hundreds of antennas [2]. Therefore, it is critical to design low-complexity detection methods for massive MIMO to meet its practical implementation requirements.

To tackle this challenge, several message passing (MP) based detectors have been proposed, offering reliable detection through iterations with manageable complexity [3], [4], [5]. Particularly, expectation propagation (EP) based detectors have gained widespread adoption in massive MIMO systems due to their near-optimal performance regardless of the antenna configuration [6]. For example, by strategically selecting initial parameters of EP algorithm, Yao et al. proposed a novel EP detector with excellent performance at the cost of lower convergence [7]. Zhang et al. proposed an EP detector based on neural networks (NN), which unfolded the EP algorithm using multi-layer deep forward networks and demonstrated good performance [8]. To enhance the performance of approximate EP algorithm, Ge et al. proposed a deep learning (DL) based detector against correlated channels by adding learnable parameters into unfolded approximated EP framework [9]. However, the aforementioned EP-based detectors are often limited by the accuracy of the generated marginal distributions during approximation. This issue requires careful attention and mitigation strategies to further enhance their performance.

Benefiting from its outstanding feature extraction ability, model-driven graph neural networks (GNN)-based learning has been integrated into EP based detectors to mitigate inaccurate marginal distributions. In [10], a GNN-based EP detector was proposed in an uplink multiuser-MIMO (MU-MIMO) system, which demonstrated improved detection performance compared to its counterparts by enhancing the accuracy of approximated posterior distributions. Nevertheless, the performance of such GNN-based detector can be further improved since the self-correction ability and parameters selection of EP are ignored, especially under high signal-to-noise ratios (SNRs). However, to our best knowledge, such problem has not been well investigated in the existing GNN-based EP detectors.

In light of the above, a model-driven improved EP-GNN (IEP-GNN) detector is proposed in this letter. Our main contributions are summarized as follows: 1) By intergrating GNN and improved moment matching (IMM) into the EP based detector, an IEP-GNN detection framework is proposed, which can improve the accuracy of posterior distribution approximation and exploit the self-correction ability of the EP algorithm. 2) In addition, hotbooting and Bayesian parameter optimization (BPO) are introduced into our proposed IEP-GNN detector to acquire the training experiences and optimize initial parameters, which can further improve the detection performance. Simulation results demonstrate that our proposed model-driven IEP-GNN detection framework can achieve better performance compared to other state-of-the-art EP-based detectors while maintaining an acceptable convergence and complexity.

II. System Model

We consider a MIMO system consisting of N_t transmitting antennas and a base station (BS) equipped with N_r receiving

Manuscript received 10 October 2023; accepted 31 October 2023. Date of publication 3 November 2023; date of current version 9 February 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62101306 and Grant 61971041. The associate editor coordinating the review of this article and approving it for publication was S. Sugiura. *(Corresponding author: Dongxuan He.)*

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

Digital Object Identifier 10.1109/LWC.2023.3329876

antennas. Accordingly, the received signal at BS can be expressed as

$$\widetilde{\mathbf{y}} = \widetilde{\mathbf{H}}\widetilde{\mathbf{x}} + \widetilde{\mathbf{w}}, \tag{1}$$

where $\widetilde{\mathbf{H}} \in \mathbb{C}^{N_r \times N_t}$ is the Rayleigh fading channel matrix, $\widetilde{\mathbf{x}} \in \mathbb{C}^{N_t \times 1}$ is the $\widetilde{M}$-quadrature amplitude modulation (QAM) modulated complex transmitted symbol vector, and $\widetilde{\mathbf{w}}$ is additive white Gaussian noise (AWGN) with zero mean and variance $\widetilde{\sigma}^2$. For simplicity, the complex model shown in (1) is typically transformed into an equivalent real-valued model, given by [9]

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{2}$$

where $\mathbf{x} \in \mathbb{R}^{2N_t \times 1}$, $\mathbf{y} \in \mathbb{R}^{2N_r \times 1}$, $\mathbf{w} \in \mathbb{R}^{2N_r \times 1}$, and $\mathbf{H} \in \mathbb{R}^{2N_r \times 2N_t}$. To be noticed, the channel state information (CSI) is assumed to be perfectly known to the detector in this letter. According to Bayes' theorem, the posterior distribution of $\mathbf{x}$ can be presented as

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \propto \mathcal{N}\Big(\mathbf{y} : \mathbf{H}\mathbf{x}, \sigma^2\mathbf{I}_N\Big) \prod_{i=1}^{2N_t} p(x_i), \tag{3}$$

where $\sigma^2 = \widetilde{\sigma}^2/2$, $p(x_i) = \frac{1}{|\Theta|}\sum_{x\in\Theta}\delta(x_i - x)$ is the indicator function of $\mathbf{x} \in \Theta$, with Θ representing M-QAM constellation with size $\sqrt{\widetilde{M}}$, where the elements of Θ are the values of the real or imaginary parts of $\widetilde{M}$-QAM.

III. Improved EP-GNN Detection

A. EP Detector

To obtain the approximated posterior distribution, EP detector approximates the joint prior probability $p(\mathbf{x}) = \prod_{i=1}^{2N_t} p(x_i)$ by independent Gaussian distributions, i.e., $q^{(t-1)}(\mathbf{x}) = \prod_{i=1}^{2N_t} q^{(t-1)}(x_i)$, whose corresponding approximate posterior distribution is given by

$$\begin{aligned} q^{(t)}(\mathbf{x}|\mathbf{y}) &\propto p(\mathbf{y}|\mathbf{x})q^{(t-1)}(\mathbf{x}) \\ &= \mathcal{N}\Big(\mathbf{y}:\mathbf{H}\mathbf{x}, \sigma^2\mathbf{I}_{2N_t}\Big)\mathcal{N}\Big(\mathbf{x}:(\mathbf{\Lambda}^{(t-1)})^{-1}\boldsymbol{\gamma}^{(t-1)}, (\mathbf{\Lambda}^{(t-1)})^{-1}\Big) \\ &= \mathcal{N}\Big(\mathbf{x}:\boldsymbol{\mu}^{(t)}, \mathbf{\Sigma}^{(t)}\Big), \end{aligned} \tag{4}$$

where $\boldsymbol{\gamma}^{(t-1)} = [\gamma_1^{(t-1)}, \gamma_2^{(t-1)}, \ldots, \gamma_{2N_t}^{(t-1)}]^T$ and $\mathbf{\Lambda}^{(t-1)} = diag\{[\lambda_1^{(t-1)}, \lambda_2^{(t-1)}, \ldots, \lambda_{2N_t}^{(t-1)}]\}$ are the information vector and precision matrix of $q^{(t-1)}(\mathbf{x})$ in the $(t-1)$-th iteration, respectively. $\boldsymbol{\mu}^{(t)}$ and $\mathbf{\Sigma}^{(t)}$ can be obtained as

$$\mathbf{\Sigma}^{(t)} = \Big(\sigma^{-2}\mathbf{H}^T\mathbf{H} + \mathbf{\Lambda}^{(t-1)}\Big)^{-1}, \tag{4a}$$

$$\boldsymbol{\mu}^{(t)} = \mathbf{\Sigma}^{(t)}\Big(\sigma^{-2}\mathbf{H}^T\mathbf{y} + \boldsymbol{\gamma}^{(t-1)}\Big). \tag{4b}$$

Due to the inherent complexity in calculating the values of $\boldsymbol{\gamma}^{(t)}$ and $\mathbf{\Lambda}^{(t)}$ that satisfy the moment matching conditions, an iterative approximation method is typically adopted in EP detector [6]. More specifically, $\boldsymbol{\mu}^{(t)}$ and $\mathbf{\Sigma}^{(t)}$ are computed iteratively in (5) using $\boldsymbol{\gamma}^{(t-1)}$ and $\mathbf{\Lambda}^{(t-1)}$.

Then, to update $\boldsymbol{\gamma}^{(t)}$ and $\mathbf{\Lambda}^{(t)}$, the cavity distribution of x_i is defined, which is given as $q^{(t)\backslash i}(x_i) = \frac{q^{(t)}(x_i|\mathbf{y})}{q^{(t)}(x_i)} \approx \frac{\mathcal{N}(x_i : \mu_i^{(t)}, \Sigma_{ii}^{(t)})}{\exp(-\frac{1}{2}\lambda_i^{(t-1)}x_i^2 + \gamma_i^{(t-1)}x_i)} \propto \mathcal{N}(x_i : \mu_{cav,i}^{(t)}, \Sigma_{cav,i}^{(t)})$, where the marginal distribution $q^{(t)}(x_i|\mathbf{y})$ is approximated by $\mathcal{N}(x_i : \mu_i^{(t)}, \Sigma_{ii}^{(t)})$ to reduce the computational complexity.[1] As a result, the elements of $\boldsymbol{\mu}_{cav}^{(t)}$ and $\mathbf{\Sigma}_{cav}^{(t)}$ can be calculated in parallel as follows

$$\Sigma_{cav,i}^{(t)} = \frac{\Sigma_{ii}^{(t)}}{1 - \Sigma_{ii}^{(t)}\lambda_i^{(t-1)}}, \tag{5a}$$

$$\mu_{cav,i}^{(t)} = \Sigma_{cav,i}^{(t)}\left(\frac{\mu_i^{(t)}}{\Sigma_{ii}^{(t)}} - \gamma_i^{(t-1)}\right). \tag{5b}$$

Then, an approximated marginal posterior distribution $\hat{p}^{(t)}(x_i|\mathbf{y})$ is introduced, which can utilize the cavity distribution and true prior information to improve the local posterior information, given by $\hat{p}^{(t)}(x_i|\mathbf{y}) = q^{(t)\backslash i}(x_i)p(x_i)$. Therefore, the expectation and variance of $\hat{p}^{(t)}(x_i|\mathbf{y})$, i.e., $\mu_i^{'(t)}$ and $\Sigma_i^{'(t)}$, can be presented as

$$\mu_i^{'(t)} = E_{\hat{p}^{(t)}(x_i|\mathbf{y})}[x_i], \tag{6a}$$

$$\Sigma_i^{'(t)} = E_{\hat{p}^{(t)}(x_i|\mathbf{y})}\left[\left(x_i - \mu_i^{'(t)}\right)^2\right], \tag{6b}$$

where $E_{\hat{p}^{(t)}(x_i|\mathbf{y})}[\cdot]$ denotes the expectation operation among the distribution $\hat{p}^{(t)}(x_i|\mathbf{y})$.

To update $\boldsymbol{\gamma}^{(t)}$ and $\mathbf{\Lambda}^{(t)}$, $\hat{p}^{(t)}(x_i|\mathbf{y})$ is approximated as a Gaussian distribution by moment matching [6]. Accordingly, $\boldsymbol{\gamma}^{(t)}$ and $\mathbf{\Lambda}^{(t)}$ can be updated as follows

$$\lambda_i^{(t)} = \frac{1}{\Sigma_i^{'(t)}} - \frac{1}{\Sigma_{cav,i}^{(t)}}, \tag{7a}$$

$$\gamma_i^{(t)} = \frac{\mu_i^{'(t)}}{\Sigma_i^{'(t)}} - \frac{\mu_{cav,i}^{(t)}}{\Sigma_{cav,i}^{(t)}}. \tag{7b}$$

It is worth noting that due to the non-negativity of $\lambda_i^{(t)}$, the EP detector will keep $\lambda_i^{(t)} = \lambda_i^{(t-1)}$ and $\gamma_i^{(t)} = \gamma_i^{(t-1)}$ when $1/\Sigma_i^{'(t)} - 1/\Sigma_{cav,i}^{(t)} < 0$. Otherwise, a damping factor β will be introduced into the EP detector to smooth the parameter updating process, i.e., $\lambda_i^{(t)} = \beta\lambda_i^{(t-1)} + (1-\beta)(1/\Sigma_i^{'(t)} - 1/\Sigma_{cav,i}^{(t)})$ and $\gamma_i^{(t)} = \beta\gamma_i^{(t-1)} + (1-\beta)(\mu_i^{'(t)}/\Sigma_i^{'(t)} - \mu_{cav,i}^{(t)}/\Sigma_{cav,i}^{(t)})$. Once converged, $\hat{\boldsymbol{\mu}}_{EP} = \boldsymbol{\mu}^{(t_{max})}$ can be obtained by the posterior approximation corresponding to the minimum Kullback-Leibler divergence (KLD), where t_{max} denotes the maximum number of iteration. Subsequently, hard decision is applied to detect the received symbols, given by

$$\hat{x}_{EP,i} = \arg\min_{x_i\in\Theta} |x_i - \hat{\mu}_{EP,i}|^2. \tag{8}$$

B. Proposed Model-Driven IEP-GNN Detector

The conventional EP detector suffers from a non-negligible performance loss due to the inaccurate approximation during iterations. To improve the detection reliability, a model-driven

[1]The off-diagonal elements of $\mathbf{\Sigma}^{(t)}$ are ignored during the approximation [10].

IEP-GNN detector is proposed in this section, which integrates a GNN module and an IMM module into the EP framework to improve the approximation accuracy and exploit the self-correction ability.

1) GNN Module: By properly adjusting the cavity distribution, the correlated and non-Gaussian residual noise caused by the operation of ignoring off-diagonal elements of $\mathbf{\Sigma}^{(t)}$ can be handled. As a result, the detection performance of EP detector can be significantly improved [10]. Inspired by this, we consider to utilize GNN to obtain more accurate cavity distribution in our proposed IEP-GNN.[2]

After calculating $\boldsymbol{\mu}_{cav}^{(t)}$ and $\mathbf{\Sigma}_{cav}^{(t)}$ by (5), GNN is further introduced to compensate the approximation loss in cavity distribution. Accordingly, the node state of x_i can be initialized as

$$\mathbf{u}_i^{(0)} = \mathbf{W}_1 \left[\mathbf{y}^T \mathbf{h}_i, \mathbf{h}_i^T \mathbf{h}_i, \sigma^2\right]^T + \mathbf{b}_1, \tag{9}$$

and the edge attribution between x_i and x_j is expressed as $\boldsymbol{\epsilon}_{i,j} = [-\mathbf{h}_i^T \mathbf{h}_j, \sigma^2]$. Moreover, during the initialization of $\mathbf{u}_i^{(0)}$, both $\mathbf{W}_1$ and $\mathbf{b}_1$ can be learned by a multiple layer perception (MLP) easily.

The considered GNN module consists of three components, namely a propagation, an aggregation, and a readout component. More specifically, the propagation and aggregation can be expressed as

$$\mathbf{m}_{i\to j}^{(l)} = \mathcal{M}\left(\mathbf{u}_i^{(l-1)}, \mathbf{u}_j^{(l-1)}, \boldsymbol{\epsilon}_{i,j}, \mathbf{W}_2\right), \tag{10}$$

$$\mathbf{u}_i^{(l)} = \mathcal{U}\left(\mathbf{u}_i^{(l-1)}, \left[\sum_{j\in[2N_t]\backslash i} \mathbf{m}_{j\to i}^{(l)}, \mu_{cav,i}^{(t)}, \Sigma_{cav,i}^{(t)}\right], \mathbf{W}_3\right), \tag{11}$$

where $\mathcal{M}(\cdot)$ and $\mathcal{U}(\cdot)$ represent the MLP and the gated recurrent unit (GRU), respectively. $\mathbf{W}_2$ and $\mathbf{W}_3$ are their learnable parameter matrixes, respectively. $\mathbf{m}_{i\to j}^{(l)}$ denotes the message passing from node i to node j at the l-th iteration, and $\mathbf{u}_i^{(l)}$ is the state of node i at the l-th iteration. As for the readout component, we have

$$\{g_{i,m}\} = \mathcal{R}\left(u_i^L, \mathbf{W}_4\right), \tag{12a}$$

$$\hat{q}_{cav}^{(t)\backslash i}(x_i = s_m) = \frac{\exp\left(g_{i,m}\right)}{\sum_{m=1}^{|\Theta|}\exp\left(g_{i,m}\right)}, s_m \in \Theta, \tag{12b}$$

where $\mathcal{R}(\cdot)$ represents the MLP with learnable parameter matrix $\mathbf{W}_4$. Then, by replacing $q^{(t)\backslash i}(x_i)$ by $\hat{q}_{cav}^{(t)\backslash i}(x_i)$ generated by GNN module, the moment computation operation shown in (6) can be handled.

2) IMM Module: For the conventional EP detectors, the negative variance and its corresponding expectation, which are treated as outliers due to the non-negativity property of variance, are neglected during the update of $\boldsymbol{\gamma}$ and $\mathbf{\Lambda}$. However, such negative variance is helpful to determine the incorrect detection having minimum probability, which has been already validated as the self-correction ability of the EP algorithm [7].

[2]Since the inference methods are jointly learned with the latent variables, GNN can learn the graphical relationships among transmitted symbols accurately, thus leading to a more accurate distribution for classification [13].

Algorithm 1 The Proposed IEP-GNN Detector

Input: $\mathbf{H}$, $\mathbf{y}$, σ^2, α^*, β^*, $\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \mathbf{W}_4]$
Initialization: Set $\boldsymbol{\gamma}^{(0)} = \mathbf{0}$, $\mathbf{\Lambda}^{(0)} = \alpha^* \mathbf{I}_M$, $\beta = \beta^*$, $\mathbf{u}^{(0)} = \mathbf{0}$, $\boldsymbol{\epsilon}_{i,j} = [-\mathbf{h}_i^T \mathbf{h}_j, \sigma^2]$.
for $t = 1$ to t_{max} **do**
  Compute $\mathbf{\Sigma}^{(t)}$ and $\boldsymbol{\mu}^{(t)}$ using (4);
  Compute $\Sigma_{cav,i}^{(t)}$ and $\mu_{cav,i}^{(t)}$ using (5);
  Compute $\mathbf{u}_i^{(0)}$ using (9);
  for $l = 1$ to L **do**
    Compute $\mathbf{m}_{i\to j}^{(l)}$ and $\mathbf{u}_i^{(l)}$ using (10) and (11);
  end for
  Compute $\hat{q}_{cav}^{(t)\backslash i}(x_i)$ using (12);
  Compute $\mu_i^{\prime(t)}$ and $\Sigma_i^{\prime(t)}$ utilizing $\hat{q}_{cav}^{(t)\backslash i}(x_i)$ using (6);
  Compute $\lambda_i^{(t)}$ and $\gamma_i^{(t)}$ using (13);
end for
Output: $\hat{\mathbf{x}} = \boldsymbol{\mu}^{(t_{max})}$

Therefore, to make full use of such property, we reformulate the update process of $\lambda_i^{(t)}$ and $\gamma_i^{(t)}$ as

$$\lambda_i^{(t)} = \lambda_i^{(t-1)} + (1-\beta)\left(\frac{1}{\Sigma_i^{\prime(t)}} - \frac{1}{\Sigma_{ii}^{(t)}}\right), \tag{13a}$$

$$\gamma_i^{(t)} = \gamma_i^{(t-1)} + (1-\beta)\left(\frac{\mu_i^{\prime(t)}}{\Sigma_i^{\prime(t)}} - \frac{\mu_i^{(t)}}{\Sigma_{ii}^{(t)}}\right). \tag{13b}$$

Based on the above-mentioned analysis, the cavity distribution of our proposed IEP-GNN detector can be obtained by the output of GNN, i.e., $\hat{q}_{cav}^{(t)\backslash i}(x_i)$. Moreover, the statistical properties of $q(\mathbf{x})$ can be updated by (13), where the outliers will not be ignored, thus leading to a better detection performance. The details of our proposed IEP-GNN detector are summarized in **Algorithm** 1.

IV. Hotbooting and BPO for IEP-GNN

Noted that the performance of our proposed IEP-GNN detector largely relies on the selection of $\mathbf{W}$, which will be deteriorated when the training of GNN fails to converge. To guarantee the reliability of our proposed IEP-GNN detector, hotbooting scheme and Bayesian parameter optimization are employed in our detector in this section.

In Section III-B, the learnable parameters of GNN are randomly initialized, which will lead to a uniform distribution output of $\hat{q}_{cav}^{(t)\backslash i}(x_i)$. As a result, the incorrect uniform distribution will interact with the outliers in (13), which will result in the overload of self-correction ability in EP. This will cause the detector to be difficult to converge during the training. To solve this problem, a hotbooting scheme is proposed, which can acquire the previous training experiences in similar scenarios to find an efficient initialization without IMM. Then, the hotbooted GNN with initial parameters $\mathbf{W}_{\text{NI}}$ is further trained with IMM, where less samples and smaller learning rate are required.

Moreover, it has been verified that initial parameters (i.e., α and β) sensitive to SNRs are helpful to improve the performance of EP detectors [7]. More specifically, the

Algorithm 2 The Proposed Hotbooting and BPO Training Scheme

1: Initialize the learnable parameters of the GNN module as $\mathbf{W} = \mathbf{W}_{\mathrm{NI}}$.
2: Initialize a set of hyper-parameters $\boldsymbol{\Theta} = \{\boldsymbol{\theta}, \mathcal{L}(\boldsymbol{\theta})\}$, objective function $\mathcal{L}$, iterations N_b, top performers ζ, the prior distribution over hyper-parameters $\alpha \sim U(0, 1/E_s), \beta \sim U(0.01, 1)$.
3: **for** $i = 1$ to N_b **do**
4: Sample a set of hyper-parameters $\boldsymbol{\theta}_i = [\alpha_i, \beta_i]$ from the prior distribution;
5: Train the model using $\boldsymbol{\theta}_i$ and compute the objective function $\mathcal{L}(\boldsymbol{\theta}_i)$;
6: Store the hyper-parameters and associated objective function $(\boldsymbol{\theta}_i, \mathcal{L}(\boldsymbol{\theta}_i))$ in $\boldsymbol{\Theta}$;
7: **if** $i > \zeta$ **then**
8: According to the objective function, split $\zeta\%$ best performing hyper-parameters to $\boldsymbol{\Theta}_g$, the rest to $\boldsymbol{\Theta}_b$;
9: Update the prior distribution as a mixture $p_g(\boldsymbol{\theta})$ and $p_b(\boldsymbol{\theta})$ estimated from $\boldsymbol{\Theta}_g$ and $\boldsymbol{\Theta}_b$ by kernel functions;
10: **end if**
11: **end for**
12: **return** $\boldsymbol{\theta}^* = [\alpha^*, \beta^*]$ that yields the minimum objective function and the corresponding parameters $\mathbf{W}^*$.

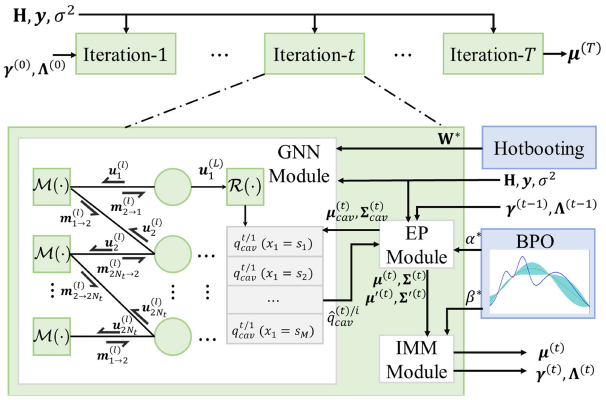

Fig. 1. Structure of the proposed IEP-GNN detector with hotbooting and BPO training.

damping factor β in (13), which can regulate the influence of outliers, should decrease as SNR increases, and this can avoid the variance to converge too fast. In addition, in conventional EP detector, setting $\alpha = 1/E_s$ is too excessive, which will lead to an overconfidence of distribution approximation. Accordingly, a derivative-free BPO method for initial parameters optimization is considered, which works during offline training.

Based on the tree-structured parzen estimator (TPE) approach in [14], α and β can be treated as continuous hyper-parameters of GNN, the distributions of which satisfy uniform distributions, i.e., $\alpha \sim U(0, 1/E_s)$ and $\beta \sim U(0.01, 1)$ [7]. The symbol error rate (SER) is used to evaluate whether the selected $[\alpha, \beta]$ is optimal. The detail of our proposed hotbooting and BPO training scheme is summarized in **Algorithm** 2.

The structure of our proposed IEP-GNN detector with hotbooting and BPO training scheme is shown as Fig. 1.

V. Simulation Results

In this section, numerical results are presented to demonstrate the superiority of our proposed IEP-GNN detector, where we compare IEP-GNN detector with MMSE, EP [6], mEP [11], GEPNet [10]. Moreover, the convergence performance is also investigated to show the effectiveness of our proposed IEP-GNN detector. Throughout this section, 16-QAM is assumed to be transmitted. For comparison, both IEP-GNN detector and GEPNet detector adopt similar parameters, where the numbers of inner GNN and outer IEP iterations are set to be $L = 2$ and $T = 10$, respectively.

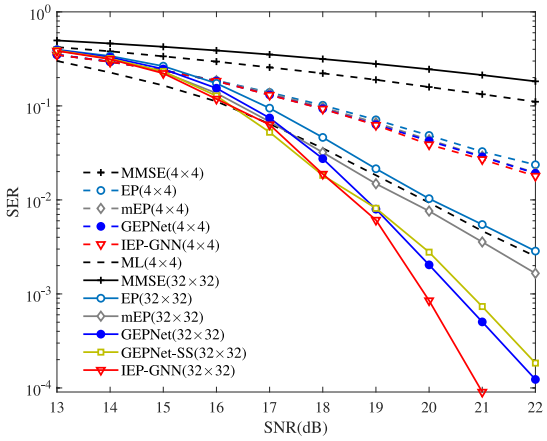

Fig. 2. SER performance comparison in the Rayleigh-fading channel with different antenna configurations.

For the GNN module, the output size of the MLP corresponding to $\mathbf{W}_1$ is N_u. The output sizes of three MLP layers corresponding to $\mathbf{W}_2$ for $\mathcal{M}(\cdot)$ are N_{h_1}, N_{h_2}, and N_u, respectively. The output sizes of the hidden MLP layers in $\mathcal{U}(\cdot)$ corresponding to $\mathbf{W}_3$ are N_{h_1}, N_u. The output sizes of MLP layers corresponding to $\mathbf{W}_4$ are N_{h_1}, N_{h_2}, and $|\Theta|$, respectively. According to [10], we set $N_u = 8$, $N_{h_1} = 64$, and $N_{h_2} = 32$. All simulations are run on a desktop with GPU NVIDIA Quadro RTX 4000 and Pytorch 2.0.0.

As for hotbooting, our proposed detector ignoring IMM is trained by Adam optimizer, where the detector corresponding to each SNR realization is trained by 20 epochs. Besides, for each epoch, 5×10^5 samples are divided as batches with size $N_{ba} = 64$. The learning rate is initially set to 10^{-3} and the decay factor is 0.91. Once the learning rate drops to 10^{-4}, the dropping stops in the sequential epochs. To train the posterior distribution $\hat{p}^{(t)}(x_i|\mathbf{y})$ accurately, cross-entropy loss function is adopted as the learning performance metric, given by

$$L = -\frac{1}{N_{ba}} \sum_{n=1}^{N_{ba}} \sum_{i=1}^{M} \sum_{s \in \Theta} \delta(x_i = s) \log\left(\hat{p}^{(t_{max})}(s)\right), \quad (14)$$

where $\hat{p}^{(t_{max})}(s) = \hat{q}_{cav}^{(t_{max})\backslash i}(s) p(s)$.

As for BPO, we have $N_b = 100$ and $\zeta = 20$. The hyper-parameters of IEP-GNN detector (i.e., α and β) are obtained by BPO detailed in **Algorithm** 2, where each set of hyper-parameters for a specific SNR realization is trained over 10 epochs with 10^5 samples.

To show the superiority of our proposed IEP-GNN detector, Fig. 2 illustrates SER performance of MMSE, EP, mEP, GEPNet and our proposed IEP-GNN detector with different MIMO system configurations. The damping factors β for EP,

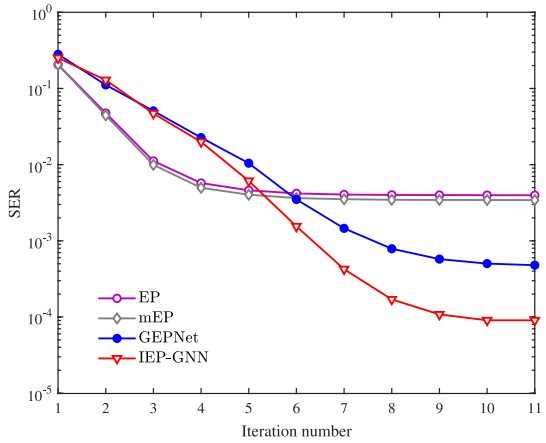

Fig. 3. The convergence comparison at the SNR = 21dB with antenna configuration 32×32.

mEP and GEPNet are set to 0.9, 0.9 and 0.7, respectively [10]. It can be seen that our proposed IEP-GNN detector has superior performance compared with other detectors regardless of the system configuration. In particular, 1dB performance improvement can be achieved for the proposed IEP-GNN detector compared with GEPNet at the SER $= 10^{-4}$ in the large-scale scenario. Furthermore, to mitigate the impact of single SNR (SS) training on learning based methodology, the SER performance of GEPNet-SS is also provided. Since the GEPNet-SS without BPO suffers a performance loss caused by unsuitable initial parameters, its performance is worse than GEPNet at high SNRs. Therefore, it is evident that our proposed IEP-GNN framework can achieve better performance than the learning based counterparts with different training schemes.[3]

To verify the convergence of our proposed IEP-GNN detector, the SER performance of four iterative detectors is shown in Fig. 3, with SNR = 21dB. All detectors are able to converge in 10 iterations. Although our proposed IEP-GNN detector needs a slightly greater iteration number than the conventional EP and mEP detectors, it can significantly enhance the performance while maintaining convergence compared with learning based GEPNet[4].

Moreover, the computational complexity of our proposed IEP-GNN detector is on the order of $\mathcal{O}(8N_t^3 t_{max} + 2N_t N_{h_1} N_{h_2} L t_{max})$, which is the same as the GEPNet.[5] Nevertheless, the SER performance of our proposed detector is superior.

[3]In addition, our proposed detector could show performance superiority under realistic correlated MIMO channels, while supplementary approaches like LDPC coding [7] and data stream reduction [15] may be needed for receivers in low-rank scenarios.

[4]To achieve convergence with fewer iterations, learning method can be utilized to add learnable parameters into the LMMSE operation in EP, which can accelerate the convergence of EP framework [16].

[5]Since the training of GNN-based detectors can be realized offline, only the complexity of the online detection is considered here.

VI. Conclusion

In this letter, an improved model-driven GNN-based MIMO detector, called IEP-GNN, was proposed. In particular, graph neural network and improved moment matching were introduced into our proposed detector, which can exploit the symbol correlation and the outlier information of prior distribution, thus guaranteeing the detection performance. To further improve the performance of our proposed detector, hotbooting and Bayesian parameter optimization were introduced into the detector to acquire the training experiences and select proper initial parameters, respectively. Simulation results validate the superiority of our proposed detector, which outperforms the state-of-the-art EP-based counterparts.

References

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

[2] 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.

[3] W. Yuan, N. Wu, Q. Guo, Y. Li, C. Xing, and J. Kuang, "Iterative receivers for downlink MIMO-SCMA: Message passing and distributed cooperative detection," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3444–3458, May 2018.

[4] S. Wu, L. Kuang, Z. Ni, J. Lu, D. Huang, and Q. Guo, "Low-complexity iterative detection for large-scale multiuser MIMO-OFDM systems using approximate message passing," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 902–915, Oct. 2014.

[5] J. Zeng, J. Lin, and Z. Wang, "Low complexity message passing detection algorithm for large-scale MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 708–711, Oct. 2018.

[6] J. Céspedes, P. M. Olmos, M. Sanchez-Fernandez, and F. Perez-Cruz, "Expectation propagation detection for high-order high-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2840–2849, Aug. 2014.

[7] G. Yao, H. Chen, and J. Hu, "An improved expectation propagation based detection scheme for MIMO systems," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2163–2175, Apr. 2021.

[8] J. Zhang, Y. He, Y.-W. Li, C.-K. Wen, and S. Jin, "Meta learning-based MIMO detectors: Design, simulation, and experimental test," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1122–1137, Feb. 2021.

[9] Y. Ge, X. Tan, Z. Ji, Z. Zhang, Y. Xiaohu, and C. Zhang, "Improving approximate expectation propagation massive MIMO detector with deep learning," *IEEE Wireless Commun. Lett.*, vol. 10, no. 10, pp. 2145–2149, Oct. 2021.

[10] 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, Jul. 2022.

[11] H. Chen, G. Yao, and J. Hu, "Algorithm parameters selection method with deep learning for EP MIMO detector," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 10146–10156, Oct. 2021.

[12] L. Xiao, D. Jiang, D. Xu, H. Zhu, Y. Zhang, and H. V. Poor, "Two-dimensional antijamming mobile communication based on reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9499–9512, Oct. 2018.

[13] 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. (ICML)*, Vienna, Austria, 2020, pp. 1–4.

[14] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, Feb. 2012.

[15] Y. Ding, Y. Wang, J.-F. Diouris, and Z. Yao, "Robust fixed-complexity sphere decoders for rank-deficient MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4297–4305, Sep. 2013.

[16] S. Li, C. Ding, L. Xiao, X. Zhang, G. Liu, and T. Jiang, "Expectation propagation aided model driven learning for OTFS signal detection," *IEEE Trans. Veh. Technol.*, vol. 12, no. 9, pp. 12407–12412, Sep. 2023.