Pre-Trained Hybrid-Damping Message Passing Detector for Multiple MIMO Scenarios

Yingmeng Ge*†‡, Baiping Xiong*†‡, Haibo Wang*†‡, and Zaichen Zhang*†‡
*National Mobile Communications Research Laboratory, Southeast University, Nanjing, 210096, P.R.China
†Frontiers Science Center for Mobile Information Communication and Security, Southeast University, Nanjing, 210096, P.R.China
‡Purple Mountain Laboratory, Nanjing, 211111, P.R.China
Email: {230208082, xiongbp, haibowang, zczhang}@seu.edu.cn

2023 IEEE 23rd International Conference on Communication Technology (ICCT) | 979-8-3503-2595-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCT59356.2023.10419595

***Abstract*—Among various multiple-input multiple-output (MIMO) detection algorithms, message passing algorithms (MPAs) have been widely considered to balance performance and complexity. Damping factors can greatly affect MPAs' convergence behavior and stability. Deep neural network (DNN) based detectors with learnable damping factors have shown improved convergence performance in MIMO systems but require retraining in different system configurations. In this paper, a pre-trained hybrid-damping (PHD) message passing detection (MPD) for multi-scenarios using multi-objective evolutionary algorithm (MOEA) is proposed, which only requires one pre-training step and can be adapted to multiple scenarios. Numerical results indicate that the proposed PHD scheme can provide better convergence and flexibility than unified damping (UD) and DNN-based ones. Furthermore, the proposed PHD scheme can also be applied to other MPAs.**

***Index Terms*—MIMO detector, message passing, damping factor, multi-objective evolutionary algorithm.**

I. Introduction

In the context of the ever-increasing demand for communication rate, multiple-input multiple-output (MIMO) has gained wide attention recently due to its potential for high spectral efficiency. The complexity of optimal MIMO detection algorithm, maximum-likelihood (ML) or maximum a posterior (MAP), becomes unaffordable in massive MIMO scenarios. To address this issue, message passing algorithms (MPAs), e.g., expectation propagation (EP) [1], approximate message passing (AMP) [2], and channel hardening-exploiting message passing (CHEMP) [3], have been widely considered in the literature. In [4], a CHEMP detector with nearestneighbor approximation (NNA), named message passing detection (MPD), is proposed and implemented, presenting nearoptimal performance in Rayleigh fading channels and superior hardware efficiency compared to other state-of-the-art (SOA) detectors.

In MPAs, the damping factor, which can affect the convergence behavior and stability of algorithms, controls the amount of adjustment made to the updated messages in each iteration. The choice of damping factor depends on various factors, such as the channel conditions, system parameters, and performance requirements. In practice, the damping factor, being unified at each iteration, is often tuned empirically through simulations or experiments to find the optimal value for a specific MIMO system configuration [5], which is laborious. Recently, model-driven detectors, which unfold the existing MPAs and map each iteration to the layer of deep neural network (DNN) with learnable damping factors, have attracted much attention [6], [7]. By exploiting the power of DNN, model-driven detectors realize better BER and convergence performance as compared to conventional MPAs without learnable damping factors. However, DNN-based MPAs usually require retraining in different scenarios, being inflexible in practical use.

In this paper, a pre-trained[1] hybrid-damping (PHD) MPD based on multi-objective evolutionary algorithm (MOEA) is proposed. The main contributions of this paper are listed as follows,

1) The PHD MPD for multi-scenarios is modeled as a multi-objective optimization problem (MOOP). The determination of the converging iteration number is also illustrated in detail.
2) The MOEA optimization for solving hybrid damping factors is presented, including initialization, fitness, selection, crossover, mutation, and stopping criterion.
3) Numerical results are presented to demonstrate that the proposed PHD scheme only requires one pre-training step and can greatly improve the convergence performance as compared to unified-damping (UD) one under different system configurations. Furthermore, the superiority of PHD scheme compared to the DNN-based one is presented.

The structure of this paper is organized as follows. In Section II, we review the basic theory of MIMO system model, MPD, and MOEA. In Section III, the proposed PHD MPD is demonstrated in detail. The numerical results of the proposed PHD MPD are presented in Section IV. Section V concludes the entire paper.

This work is supported by NSFC project 61960206005 and the Fundamental Research Funds for the Central Universities 2242022k60001. Zaichen Zhang is the corresponding author.

[1]Here, pre-trained means the hybrid damping factors are trained offline and fixed for online detection.

Notations: $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. $\mathbf{X}^\top$ is the transpose operation of the matrix $\mathbf{X}$. $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multi-variate *Gaussian* distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

II. Preliminary

In this section, we introduce the narrow-band MIMO system model, MPD, and MOEA.

A. MIMO System Model

We consider the narrow-band MIMO communication system with N_t transmitting (Tx) antennas and N_r receiving (Rx) antennas. The corresponding system model in real domain can be expressed as,

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

where $\mathbf{y} \in \mathbb{R}^{2N_r \times 1}$ and $\mathbf{x} \in \Omega^{2N_t \times 1}$ are the received and transmitted vectors, respectively. Q-QAM modulation is assumed here and $\Omega = \{\omega_1, ..., \omega_{\sqrt{Q}}\}$ is the set of in-phase or quadrature parts of points in the complex constellation. $\mathbf{H} \in \mathbb{R}^{2N_r \times 2N_t}$ is the channel matrix, assumed to be the independent and identically distributed (i.i.d.) Rayleigh channel with mean zero and variance $1/N_r$ in this paper. $\mathbf{n} \sim \mathcal{N}\big(\mathbf{0}, \sigma_n^2 \mathbf{I}_{2N_r}\big)$ is the additive white *Gaussian* noise (AWGN) vector. We assume the channel state information (CSI) is perfectly known at the receiver.

B. MPD

The MPD detector is summarized in Algorithm 1, where $\mathbf{z} = \mathbf{H}^\top \mathbf{y}$ denotes the received vector after matched filter and z_i is the i-th element of $\mathbf{z}$. $\mathbf{J} = \mathbf{H}^\top \mathbf{H}$ is the Gram matrix and J_{ij} is the (i, j)-th element of $\mathbf{J}$. L is the maximum iteration number of MPD. E_s denotes the mean symbol energy. $\delta^{(l)} (l = 1, ..., L)$ denotes the damping factor at l-th iteration. In conventional UD MPD, $\delta^{(l)} (l = 1, ..., L)$ maintains the same value throughout the iteration. m_1 and m_2 denote the index of the first and second nearest neighbor symbols in real constellation Ω, respectively. The probabilities $\rho_i^{(l)}(\omega_m)$ can be computed as follows,

$$\rho_i^{(l)}(\omega_m) = \begin{cases} \dfrac{1}{1 + \exp(\beta_i^{(l)})}, & m = m_1, \\ 1 - \rho_i^{(l)}(\omega_{m_1}), & m = m_2, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

and

$$\beta_i^{(l)} = - \left| \frac{s_\omega \left(z_i^{(l)} - a_\omega \right)}{\tau_i^{(l)}} \right|, \tag{3}$$

where $a_\omega = (\omega_{m_1} + \omega_{m_2})/2$ and $s_\omega = \omega_{m_2} - \omega_{m_1}$.

Algorithm 1: MPD Detector

Input : $\mathbf{z}, \mathbf{J}, \sigma_n^2, L, \delta^{(l)} (\forall l = 1, ..., L)$,
$\hat{x}_i^{(1)} = 0, \xi_i^{(1)} = E_s (\forall i = 1, ..., 2N_t)$.
Output: $\hat{x}_i^{(L+1)}$ $(\forall i = 1, ..., 2N_t)$.

for $l = 1, 2, ..., L$ **do**
  for $i = 1, 2, ..., 2N_t$ **do**
    $\mu_i^{(l)} = \sum_{j=1, j \neq i}^{2N_t} J_{ij} \hat{x}_j^{(l)}$;
    $\zeta_i^{(l)} = \sum_{j=1, j \neq i}^{2N_t} J_{ij}^2 \xi_j^{(l)} + \sigma_n^2$;
    $r_i^{(l)} = (z_i^{(l)} - \mu_i^{(l)})/J_{ii}$;
    $\tau_i^{(l)} = \zeta_i^{(l)}/J_{ii}^2$;
    find m_1 and m_2;
    calculate $\rho_i^{(l)}(\omega_m)$;
    $\hat{x}_i' = \sum_{m'=1}^{\sqrt{Q}} \omega_{m'} \rho_i^{(l+1)}(\omega_{m'})$;
    $\xi_i' = \sum_{m'=1}^{\sqrt{Q}} \omega_{m'}^2 \rho_i^{(l+1)}(\omega_{m'}) - (\hat{x}_i')^2$;
    $\hat{x}_i^{(l+1)} = \delta^{(l)} \hat{x}_i' + (1 - \delta^{(l)}) \hat{x}_i^{(l)}$;
    $\xi_i^{(l+1)} = \delta^{(l)} \xi_i' + (1 - \delta^{(l)}) \xi_i^{(l)}$;

C. MOEA

MOEA is a class of evolutionary algorithms designed to solve MOOP, providing an effective approach for problems with multiple conflicting objectives. In MOEA, each individual represents a potential solution and is evaluated based on its performance across multiple objective functions. These individuals evolve through genetic operators, such as selection, crossover, and mutation, to generate new candidate solutions, where population diversity is maintained to effectively explore the solution space of the multi-objective problem.

One of the key advantages of MOEA is its ability to generate a set of solutions that exhibit balance across multiple objectives, rather than a single optimal solution. This allows decision-makers to trade-off and select among multiple objectives, gaining a better understanding of the problem nature and various trade-off possibilities. Commonly used MOEAs include non-dominated sorting genetic algorithm II (NSGA-II), adaptive geometry estimation based MOEA (AGE-MOEA) [8], etc. In this paper, we adopt AGE-MOEA for solving the MOOP due to its superior performance and low complexity as compared to conventional MOEA, e.g., NSGA-II.

III. The Proposed PHD Scheme

In this section, we first illustrate the optimization model of hybrid damping factors. Then the determination of converging iteration number is provided. Finally, the MOEA optimization process for hybrid damping factors is introduced in detail.

A. Optimization Model

The proposed PHD scheme aims at minimizing the converging iteration number while maintaining the BER performance under different system configurations. Specifically, the problem can be modeled as follows,

$$\begin{aligned} \min_{\boldsymbol{\delta}} \quad & \tilde{L}_m, \quad m = 1, ..., M, \\ s.t. \quad & \bar{P}_m^{(\tilde{L}_m)} / P_m^{(L)} - 1 \leq \varepsilon_1, \quad m = 1, ..., M, \\ & \delta^{(l)} \in \{\Delta, 2\Delta, ..., 1\}, \quad l = 1, ..., L. \end{aligned} \tag{4}$$

Here, $\boldsymbol{\delta} = \{\delta^{(1)}, ..., \delta^{(L)}\}$ is the set of learnable damping factors. $\tilde{L}_m$ denotes the converging iteration number of PHD MPD under system configuration $\mathcal{S}_m$. $\mathcal{S}_m$ includes N_t, N_r, Q, and tested signal-to-noise ratio (SNR) $R = 10\log\big(\mathbb{E}[\|\mathbf{Hx}\|^2]/\mathbb{E}[\|\mathbf{n}\|^2]\big)$, which is abbreviated as $\mathcal{S}_m = (N_t, N_r, Q, R)$ in the following. M denotes the number of different configurations considered. $\bar{P}_m^{(\tilde{L}_m)}$ is the converging bit error rate (BER) of PHD MPD with $\boldsymbol{\delta}$ under system configuration $\mathcal{S}_m$, while $P_m^{(L)}$ is the converging BER of UD MPD. Here, L needs to be set to a value large enough for UD MPD to converge. ε_1 is the threshold of relative BER error between $\bar{P}_m^{(\tilde{L}_m)}$ and $P_m^{(L)}$, which measures the loss of BER performance of PHD MPD.

Though the values of $\delta^{(l)}$ are taken continuously within the interval $(0, 1]$, the convergence performance of MPD is not highly sensitive to $\delta^{(l)}$. Thus, $\delta^{(l)}$ can take values from a discrete set instead, that is $\delta^{(l)} \in \{\Delta, 2\Delta, ..., 1\}$ with Δ being the step size, which greatly reduces the search space of $\delta^{(l)}$. The value of the step size Δ could be adjusted flexibly to enable different accuracy of the damping factor.

B. Determination of $\tilde{L}_m$

Considering that the convergence of PHD MPD with feasible $\boldsymbol{\delta}$ is guaranteed due to the constraints in Eq. (4), we propose a method for determining the converging iteration number $\tilde{L}_m$ based on the relative BER error. Specifically, at l-th ($l > 1$) iteration, determine whether the relative error between $\bar{P}_m^{(l-1)}$ and $\bar{P}_m^{(l)}$ is less than the threshold ε_2. If so, the detector has converged; otherwise, it is not converging. Therefore, the method of obtaining $\tilde{L}_m$ can be summarized in Algorithm 2.

Algorithm 2: Computing $\tilde{L}_m$

Input : $\varepsilon_2, \bar{P}_m^{(l)}$ ($\forall l = 1, ..., L$).
Output: $\tilde{L}_m$.
$l = L$;
while $l > 1$ **do**
  if $\bar{P}_m^{(l-1)}/\bar{P}_m^{(l)} - 1 > \varepsilon_2$ **then**
    break;
  $l = l - 1$;
$\tilde{L}_m = l$;

C. MOEA Optimization

The *gene* is defined as $\delta^{(l)}(l = 1, ..., L)$, while the *chromosome* is defined as the sequence of genes (i.e., $\boldsymbol{\delta}$). *Population* is a collection of chromosomes (also called *individual*). Individuals with good fitness survive as the "optimal" solutions through evolution. The proposed MOEA-based PHD MPD is illustrated as follows,

- **Initialization**: In initialization step, each $\delta^{(l)}$ is randomly generated from set $\{\Delta, 2\Delta, ..., 1\}$ to compose a chromosome and the initial population contains N_{pop} chromosomes.
- **Fitness**: The fitness of an individual denotes its proximity to meeting the overall requirements of the intended solution. It is related to the object functions and constraints in Eq. (4). Individuals with high fitness are more likely to survive and reproduce the next generation. In AGE-MOEA, two metrics, i.e., non-dominated rank and survival score, are considered to guarantee both the diversity and proximity of solutions.
- **Selection**: Selection operators are used to select individuals from the current population for the creation of the next generation. The binary tournament selection is considered in this paper.
- **Crossover**: The crossover operator recombines the genetic information of two or more parent individuals to generate new offspring individuals, increasing the diversity of the population and facilitating the evolutionary process. Considering the problem here is real-coded, simulated binary crossover (SBX) is adopted.
- **Mutation**: The mutation operator plays an important role in maintaining genetic diversity and exploring new regions of the search space, introducing random changes to the genetic information of individuals. The polynomial mutation (PM) is considered here.
- **Stopping Criterion**: The evolutionary process stops once it exceeds the maximum generation.

Overall, the pre-training process of the hybrid-damping factor can be summarized as Algorithm 3. After the set of optimal damping factors is obtained, it is used for online detection based on Algorithm 1.

Algorithm 3: Pre-training Hybrid Damping Factors.

Initialize population with N_{pop} individuals;
repeat
  Generate offspring with selection, crossover, and mutation;
  Combine current and offspring population;
  Fast non-dominated sort;
  Compute survival scores;
  Form next population according to non-dominated rank and survival scores;
until *Meet stopping criterion*;
Obtain the set of optimal $\boldsymbol{\delta}$;

IV. Simulation Results

In this section, the implementation details are first demonstrated. Then the numerical results of the proposed PHD MPD are presented, compared with minimum mean square error (MMSE), UD MPD [3] and DNN-based MPD [7].

A. Implementation Details

We consider a massive MIMO system with $M = 3$ different configurations, which are listed in Table I. Here, we choose the value of R to guarantee the BER of MPD falls below 10^{-3}, considering that MPD is more sensitive to damping factors at low BER point. The population number N_{pop} is set to 50, and the maximum generation is set to 100. AGE-MOEA with binary tournament selection, SBX, and PM is adopted as the

MOEA for optimizing the hybrid damping factors. ε_1, ε_2, and Δ are fixed to 0.5, 0.02, and 0.05, respectively. Our simulation is performed in Pymoo 0.6.0 [9] with AMD Ryzen™ 9 5950X CPU and NVIDIA Tesla P40. The δ for UD MPD is set to 0.5 and $L = 30$, which is enough for MPD to converge in the 3 configurations. DNN-based MPD with learnable damping factors is trained separately in 3 different configurations.

TABLE I: System Configurations.

m	N_t	N_r	Q	R
1	16	128	256	20
2	32	128	256	24
3	48	128	64	20

B. Numerical Results

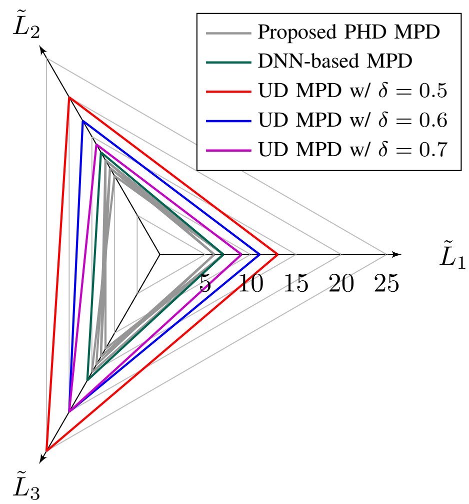


Fig. 1: The objective space comparison of PHD, UD, and DNN-based MPDs using radar plot.

Fig. 1 presents the objective space comparison of the proposed PHD MPD, UD MPD, and DNN-based MPD using radar plot. PHD scheme has 12 different non-dominated solutions, and some are overlapped because of having the same values of objective function. UD scheme with $\delta = 0.8$ is not plotted because of divergence under configuration $\mathcal{S}_3$. The results from Fig. 1 indicate that the proposed PHD MPD have smaller $\tilde{L}_1$, $\tilde{L}_2$, and $\tilde{L}_3$ than UD MPD and DNN-based MPD, demonstrating the efficiency of the proposed PHD scheme in terms of convergence performance.

Selecting a feasible $\boldsymbol{\delta}$ from the non-dominated solutions, Fig. 2 gives the BER performance of the proposed PHD MPD, UD MPD, and DNN-based MPD versus iterations under configuration $\mathcal{S}_2$. The results show that the proposed PHD scheme converges faster than UD and DNN-based ones without loss of BER performance. Fig. 3 compares the BER performance among MMSE detector, proposed PHD MPD, UD MPD, and DNN-based MPD under configurations $\mathcal{S}_2$ and $\mathcal{S}_3$. It can be seen from the figure that the proposed PHD

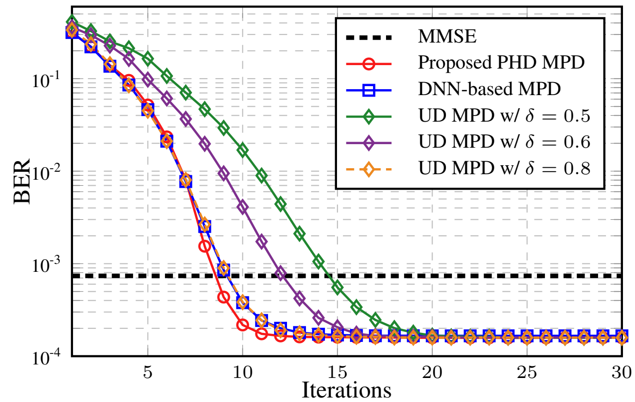


Fig. 2: BER performance of different algorithms versus iterations under configuration $\mathcal{S}_2$.

scheme achieves similar BER performance as UD and DNN-based MPDs, all superior to the MMSE detector.

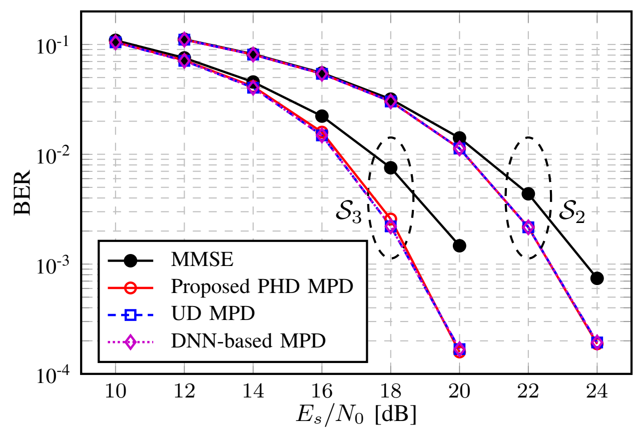
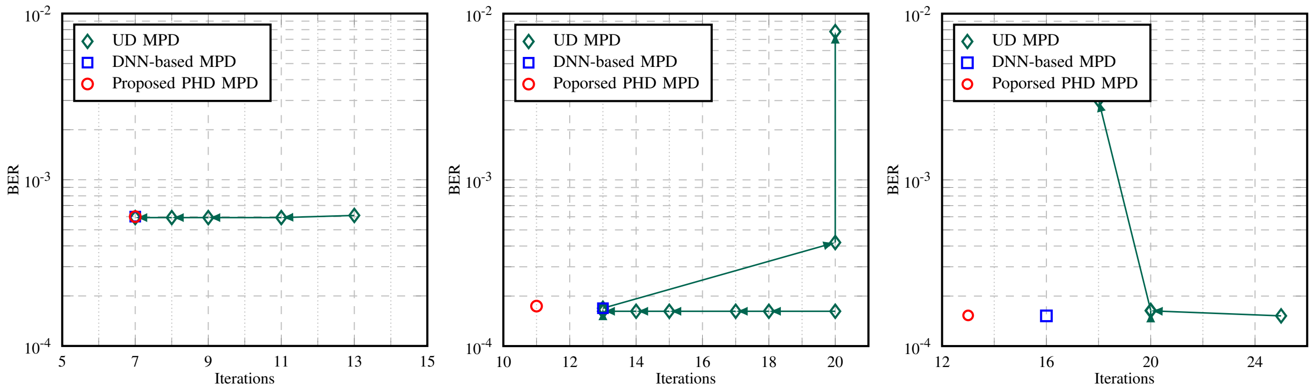


Fig. 3: BER comparison of MMSE, proposed PHD MPD ($\mathcal{S}_2 : L = 11$; $\mathcal{S}_3 : L = 13$), UD MPD ($\mathcal{S}_2 : \delta = 0.8, L = 13$; $\mathcal{S}_3 : \delta = 0.7, L = 20$), and DNN-based MPD ($\mathcal{S}_2 : L = 13$; $\mathcal{S}_3 : L = 16$).

To further illustrate the superiority of PHD scheme, the trade-off of BER and converging iteration number under different configurations is presented in Fig 4. Here, $v_{\min} : d : v_{\max}$ denotes the damping factor of UD MPD changing from $v_{\min}$ to $v_{\max}$ with interval d. For example, the direction of the arrow in Fig. 4a denotes the change in damping factor value: $0.5 \rightarrow 0.6 \rightarrow 0.7 \rightarrow 0.8 \rightarrow 0.9$. Though UD MPD can have the same iterations as PHD one when $\delta = 0.9$ under configuration $\mathcal{S}_1$, its convergence performance is worse than PHD one under $\mathcal{S}_2$ and $\mathcal{S}_3$. Similarly, the proposed PHD MPD also presents better convergence performance than the DNN-based one.

C. Complexity

The proposed PHD MPD includes two steps: offline pre-training and online detection. Though the training procedure can be done offline by powerful computation devices, PHD



(a) $\mathcal{S}_1 = (16, 128, 256, 20)$ (b) $\mathcal{S}_2 = (32, 128, 256, 24)$ (c) $\mathcal{S}_3 = (48, 128, 64, 20)$

Fig. 4: Trade-off of BER and iterations at convergence under different configurations. The change of damping factor in UD scheme: (a) $0.5 : 0.1 : 0.9$; (b) $0.5 : 0.05 : 0.9$; (c) $0.5 : 0.1 : 0.8$.

MPD can find damping factors suitable for multi-scenarios, while DNN-based MPD requires training in different scenarios to obtain targeted damping factors. For online detection step, the proposed PHD MPD has smaller converging iteration number than UD and DNN-based ones (shown in Fig. 5), requiring fewer computation resources.

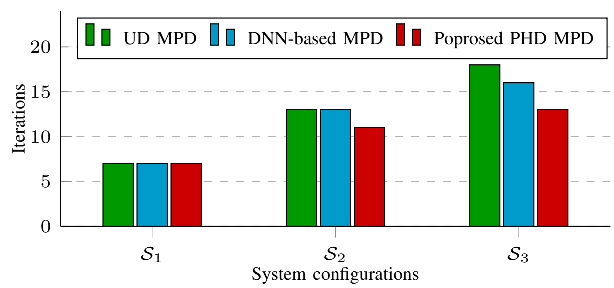


Fig. 5: Comparison of converging iteration number between UD, DNN-based, and PHD MPDs under different configurations.

In summary, the detailed comparison between the proposed PHD MPD, UD MPD, and DNN-based MPD are listed as follows:

1) *Compared with UD MPD*: The proposed scheme has faster convergence while maintaining the BER performance, resulting in lower complexity and higher throughput;
2) *Compared with DNN-based MPD*: The proposed scheme has faster convergence than the DNN-based one. The training process of the proposed scheme is unsupervised, requiring no labeled data, while DNN-based MPD needs labeled data. Only one training procedure is required in advance for the proposed scheme and the training results can be reused for different system configurations, while DNN-based MPD requires retraining in different scenarios.

V. Conclusion

In this paper, we proposed a PHD MPD applicable to multiple MIMO scenarios. By leveraging MOEA, the PHD scheme only requires one pre-training and can accelerate the convergence rate of conventional UD MPD under different system configurations. Numerical results demonstrate that the proposed PHD scheme provides better convergence performance compared to UD and DNN-based ones. Also, the flexibility of the proposed PHD scheme is demonstrated by comparing it to the DNN-based one. Note that the proposed PHD scheme can also be applied to other iterative MPAs, which will be investigated in our further works.

References

[1] J. Cespedes, P. M. Olmos, M. Sánchez-Fernández, 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.

[2] C. Jeon, R. Ghods, A. Maleki, and C. Studer, “Optimality of large MIMO detection via approximate message passing,” in *Proc. IEEE Int. Symp. Inf. Theory. (ISIT)*, Jun. 2015, pp. 1227–1231.

[3] T. L. Narasimhan and A. Chockalingam, “Channel hardening-exploiting message passing (CHEMP) receiver in large-scale MIMO systems,” *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 847–860, 2014.

[4] W. Tang, C.-H. Chen, and Z. Zhang, “A 0.58-mm^2 2.76-Gb/s 79.8-pJ/b 256-QAM message-passing detector for a 128×32 massive MIMO uplink system,” *IEEE J. Solid-State Circuits*, vol. 56, no. 6, pp. 1722–1731, 2021.

[5] J. Yang, W. Song, S. Zhang, X. You, and C. Zhang, “Low-complexity belief propagation detection for correlated large-scale MIMO systems,” *J. Signal Process. Syst.*, vol. 90, no. 4, pp. 585–599, Aug. 2018.

[6] S. Şahin, C. Poulliat, A. M. Cipriano, and M.-L. Boucheret, “Doubly iterative turbo equalization: Optimization through deep unfolding,” in *Proc. IEEE Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2019, pp. 1–6.

[7] Y. Ge, X. Tan, Z. Ji, Z. Zhang, X. You, and C. Zhang, “Improving approximate expectation propagation massive MIMO detector with deep learning,” *Wireless Commun. Lett.*, vol. 10, no. 10, pp. 2145–2149, 2021.

[8] A. Panichella, “An adaptive evolutionary algorithm based on non-Euclidean geometry for many-objective optimization,” in *Proc. Conf. Genet. Evol. Comput.*, 2019, pp. 595–603.

[9] J. Blank and K. Deb, “Pymoo: Multi-objective optimization in Python,” *IEEE Access*, vol. 8, pp. 89 497–89 509, 2020.