Activity Detection in Distributed MIMO: Distributed AMP via Likelihood Ratio Fusion

Jianan Bai and Erik G Larsson

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-189476

N.B.: When citing this work, cite the original publication.
Bai, J., Larsson, E. G, (2022), Activity Detection in Distributed MIMO: Distributed AMP via Likelihood Ratio Fusion, IEEE Wireless Communications Letters, 11(10), 2200-2204.
https://doi.org/10.1109/LWC.2022.3197053

Original publication available at:
https://doi.org/10.1109/LWC.2022.3197053
©2022 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

http://www.ieee.org/index.html
Activity Detection in Distributed MIMO: Distributed AMP via Likelihood Ratio Fusion

Jianan Bai and Erik G. Larsson

Abstract—We develop a new algorithm for activity detection for grant-free multiple access in distributed multiple-input multiple-output (MIMO). The algorithm is a distributed version of the approximate message passing (AMP) based on a soft combination of likelihood ratios computed independently at multiple access points. The underpinning theoretical basis of our algorithm is a new observation that we made about the state evolution in the AMP. Specifically, with a minimum mean-square error denoiser, the state maintains a block-diagonal structure whenever the covariance matrices of the signals have such a structure. We show by numerical examples that the algorithm outperforms competing schemes from the literature.

Index Terms—Distributed AMP, activity detection, distributed MIMO.

I. INTRODUCTION

LIMITLESS connectivity is envisioned to be one of the key features in next-generation wireless networks. Distributed multiple-input multiple-output (MIMO), also known as cell-free massive MIMO, is a promising technology to achieve this goal [1]. To support massive connectivity and low latency, grant-free multiple access (GFMA) has been proposed to reduce signaling compared to grant-based access.

In GFMA, an access point (AP) needs to identify the active users and estimate their channels based on received pilots. Due to the massive number of devices and the limited coherence block size, assigning mutually orthogonal pilot sequences to all devices becomes impractical. The resulting non-orthogonality of the pilots makes the problem of joint activity detection and channel estimation (JADCE) challenging. In typical application scenarios, devices only sporadically access the network. This results in the system being denoted by $\sum_{n \in \mathcal{N}} a_n \phi_n^T h_{kn}^T + W_k$, at the $k$-th AP can be expressed as

$$Y_k = \sum_{n \in \mathcal{N}} \sqrt{L_p} a_n \phi_n^T h_{kn}^T + W_k,$$  \hspace{1cm} (1)
where \( p_n \in [0, p_{\text{max}}] \) is the transmit power of device \( n \). The channel between AP \( k \) and device \( n \) is modeled by \( h_{kn} \sim \mathcal{C}\mathcal{N}(0, R_{kn}) \), where \( R_{kn} \in \mathbb{C}^{M \times M} \) is the spatial correlation matrix, and \( \beta_{kn} = \text{tr}(R_{kn})/M \) can be interpreted as the large-scale fading coefficient (LSFC). The channel is assumed to be uncorrelated between different APs and devices. The noise matrix \( W_k \in \mathbb{C}^{L \times M} \) has i.i.d. entries with i.e.d. \( \mathcal{C}\mathcal{N}(0, \sigma^2) \) elements, where \( \sigma^2 \) is the noise variance.

For brevity of notation, we define the effective channel \( h_{kn} \triangleq \sqrt{L p_n} h_{kn} \), which has the distribution \( \mathcal{C}\mathcal{N}(0, R_{kn}) \), where \( R_{kn} = L p_n R_{kn} \), and \( p_n = L p_n \beta_{kn} \) can be interpreted as the received signal strength of device \( n \) at AP \( k \).

Denoting the pilot matrix by \( \Phi = [\phi_1, \cdots, \phi_N] \), the effective channel matrix by \( H_k = [h_{k1}, \cdots, h_{kN}]^T \), and the vector of device activities by \( a = [a_1, \cdots, a_N]^T \), the received signal model in (1) can be written as

\[
Y_k = \Phi D_a H_k + W_k. \tag{2}
\]

By combining the received signal at all APs, we obtain

\[
Y = \Phi D_a \left[ H_1, \cdots, H_K \right] + \left[ W_1, \cdots, W_K \right], \tag{3}
\]

where \( H = [h_1, \cdots, h_N]^T \) and \( h_n = [h_{1n}, \cdots, h_{Kn}]^T \) is the channel from device \( n \) to all APs. Note that by assuming uncorrelated fading across different APs, \( h_n \) has the distribution \( \mathcal{C}\mathcal{N}(0, R_n) \), where \( R_n \) is block-diagonal: \( R_n = \text{blockdiag}(R_{1n}, \cdots, R_{Kn}) \).

### A. Power Allocation

In distributed MIMO, since the APs are spread out, the channel gains from a device to different APs vary significantly. The signal strength from a device is generally larger at APs that are physically close to the device than at other APs.

We propose a user-centric power allocation scheme that comes in a few different variations. The details are as follows:

1) Each device \( n \) is associated with the subset of APs, say \( K_n^p \), for which the LSFCs exceed a threshold \( \beta^p_n \),

\[
K_n^p = \{ k \in K : \beta_{kn} > \beta^p_n \}. \tag{4}
\]

If no AP satisfies this requirement, we associate the device to the AP with the largest LSFC, i.e.,

\[
\tilde{K}_n^p = K_n^p \cup \{ \arg \max_{k \in K} \beta_{kn} \}. \tag{5}
\]

2) For each device, a coefficient \( s_n \) is calculated. We consider the three different choices:

\[
s_n = \begin{cases} 
1, & \text{FullPower} \\
\frac{1}{|K_n^p|} \sum_{k \in K_n^p} \beta_{kn}, & \text{MasterAP} \\
\max_{k \in \tilde{K}_n^p} \beta_{kn}, & \text{AvgAP} \end{cases} \tag{6}
\]

3) For each device, the transmit power is set to

\[
p_n = \min \left\{ s_{\text{min}}/s_n, 1 \right\} p_{\text{max}}, \tag{7}
\]

where \( s_{\text{min}} \) is the minimum coefficient among all devices for which at least one AP satisfies the LSFC requirement, i.e., \( \beta_{kn} > \beta^p_n \).

### III. Activity Detection in Distributed MIMO

The system model in (3) is an instance of the linear measurement model \( Y = \Phi X + W \), where the unknown signal matrix \( X \) is row sparse, and each row \( x_n^T = a_n h_n^T \) has a Bernoulli-Gaussian distribution. Therefore, the activity detection becomes a support recovery problem in CS, which can be solved using the AMP algorithm.

#### A. AMP with MMSE Denoiser and Likelihood-Ratio Test

By initializing \( Z^0 = Y \) and \( X^0 = 0_{N \times M_k} \), the AMP iteration \( t \in \{0, 1, \cdots \} \) for complex-valued signals is [2],

\[
\tilde{x}^{t+1}_n = \text{g}_t\left( (Z^t)^T \Phi_n^+ \tilde{x}^t_n \right), \quad \forall n \in N, \tag{8}
\]

\[
Z^{t+1} = Y - \Phi \tilde{X}^{t+1} + \frac{1}{L} Z^t \sum_{n \in N} g_t(\xi_n^t), \tag{9}
\]

where \( \tilde{X}^t = [\tilde{x}_1^t, \cdots, \tilde{x}_N^t] \). Here, \( g_t(\cdot) : \mathbb{C}^{M_k} \rightarrow \mathbb{C}^{M_k} \) is the denoiser and \( g_t(\xi) \) represents its Jacobian at \( \xi \).

As demonstrated in the state evolution analysis [3], under some mild conditions and in the large-system limit, \( \xi_n^t \) behaves like a Gaussian-noise corrupted version of \( x_n \), i.e.,

\[
\xi_n^t \sim x_n + \mathcal{C}\mathcal{N}(0, \Sigma^t). \tag{10}
\]

In (10), \( \Sigma^t \) is referred to as the state; this state evolves by

\[
\Sigma^{t+1} = \sigma^2 I + \frac{1}{L} \sum_{n \in N} \mathbb{E} \left[ (g_t(x_n+v^t) - x_n) (g_t(x_n+v^t) - x_n)^H \right], \tag{11}
\]

where \( v^t \) has distribution \( \mathcal{C}\mathcal{N}(0, \Sigma^t) \) and is independent of \( x_n \), and the expectation is taken over the joint distribution of \( x_n \) and \( v^t \). The initial state is given by

\[
\Sigma^0 = \sigma^2 I + \frac{1}{L} \sum_{n \in N} R_n. \tag{12}
\]

The minimum mean-square error (MMSE) denoiser is given by the MMSE estimate of \( x_n \) given \( \xi_n^t \),

\[
g_t(\xi_n^t) = \mathbb{E}[x_n|\xi_n^t] = \theta_n^t(\xi_n^t) \cdot \Psi_n^t \xi_n^t, \tag{13}
\]

where

\[
\theta_n^t(\xi) = \left( 1 + \frac{1-\epsilon_n}{\epsilon_n} \frac{|R_n+\Sigma^t|}{|\Sigma^t|} \exp \left( -\xi^H \Omega_n^t \xi \right) \right)^{-1}, \tag{14}
\]

\[
\Psi_n^t = R_n (R_n + \Sigma^t)^{-1}, \tag{15}
\]

\[
\Omega_n^t = (\Sigma^t)^{-1} - (R_n + \Sigma^t)^{-1}. \tag{16}
\]

The support recovery problem is equivalent to the detection of the non-zero entries in the binary vector \( a \). To determine the value of \( a_n \), we consider the binary hypothesis test

\[
\mathcal{H}_0 : a_n = 0 \quad \text{and} \quad \mathcal{H}_1 : a_n = 1. \tag{17}
\]

The likelihood-ratio test (LRT) is given by

\[
\ell_n = \frac{p(\xi_n|a_n = 0)}{p(\xi_n|a_n = 1)} \frac{\alpha_t}{\gamma}, \tag{18}
\]

\( ^1 \)For brevity, we henceforth omit the iteration index \( t \) in the superscripts.
where $\gamma > 0$ is the decision threshold. According to (10), the likelihood-ratio can be written as
\[
\ell_n = \frac{CN(\xi_n, 0, \Sigma)}{CN(\xi_n, 0, R_n + \Sigma)} = \frac{|R_n + \Sigma|}{|\Sigma|} \exp \left( -\xi_n^H \Omega_n \xi_n \right). \tag{19}
\]

Notice that (14) can be rewritten as $\theta_n^k = (1 + \frac{1}{\gamma} \ell_n^k)^{-1}$.

With a large number of antennas, $M_{\text{tot}}$, a naive implementation of the AMP algorithm has two major drawbacks: 1) calculating the determinants and inverting the $M_{\text{tot}} \times M_{\text{tot}}$ matrices in (14), (15) and (16) can be computationally demanding; 2) sending the $L \times M_{\text{tot}}$-dimensional matrix $Y$ requires high fronthaul capacity.

B. Covariance Structure in the AMP State Evolution

The received signal model in distributed MIMO, see (1), has a special property: the covariance matrices $\{R_n\}$ are block-diagonal. In the following theorem, we show that during the state evolution in AMP, the states maintain the same block-diagonal structure during all iterations.

**Theorem 1.** Assume that $\{R_n\}$ have a block-diagonal structure: $R_n = \text{bdiag}(R_{n1}, \cdots, R_{nK})$. By using the MMSE denoiser in (13), the state $\Sigma^l$ in the state evolution (11) stays as a block-diagonal matrix with the same structure for each block, i.e., $\Sigma^l = \text{bdiag}(\Sigma_{1}^l, \cdots, \Sigma_{K}^l)$, for all $t$.

**Proof.** See Appendix A. $\square$

According to Theorem 1, the inversion of the $M_{\text{tot}} \times M_{\text{tot}}$ matrices in (15) and (16) can be performed by inverting their diagonal blocks, which are of dimension $M \times M$.

When the channel vector from device $n$ to AP $k$ is modeled by i.i.d. Rayleigh fading, the channel covariance matrix becomes $R_{nk} = \beta_{nk} I_M$. Correspondingly, the effective channel $h_n$ from device $n$ to all APs has the distribution $CN(0, R_n)$ with $R_n = \text{bdiag}(\rho_{n1}I_M, \cdots, \rho_{nK}I_M)$. The following corollary can be viewed as a generalization of [2, Theorem 1] to the scenario of distributed MIMO.

**Corollary 1.** Assume that $\{R_n\}$ have the diagonal structure $R_n = \text{bdiag}(\rho_{n1}I_M, \cdots, \rho_{nK}I_M)$. By using the MMSE denoiser in (13), the state $\Sigma^l$ stays as a scaled identity matrix for each diagonal block, i.e., $\Sigma^l = \text{bdiag}(\tau_1^2 I, \cdots, \tau_K^2 I)$, for all $t$.

**Proof.** By setting the size of the diagonal blocks in Theorem 1 to one, we conclude that the state $\Sigma^l$ stays as a diagonal matrix. Then, by using the symmetry, we conclude that the elements corresponding to the same AP are equal. $\square$

In the i.i.d. Rayleigh case, the calculations of all matrix inversions and determinants simplify to scalar operations.

C. Distributed Activity Detection

Since by Theorem 1, $\Sigma$ and $\Omega_n$ are block-diagonal, we can rewrite the likelihood-ratio in (19) as
\[
\ell_n = \prod_{k \in K} \left[ \frac{|R_{nk} + \Sigma_k|}{|\Sigma_k|} \exp \left( -\xi_n^H \Omega_{nk} \xi_{nk} \right) \right] \pm \ell_n^k. \tag{20}
\]

Algorithm 1 distributed AMP (dAMP)

**Input:** $\Phi, \{Y_k\}, \{R_{kn}\}$

**Initialize:** $Z_k^0 = Y_k, Z_{nk}^0 = 0, \text{ and } \Sigma_k^0 = \frac{1}{T} Y_k^H Y_k, \forall k, \forall n$

1: for each $k \in K$, independently do
2: for $t = 0, 1, \cdots$ do
3: for each $n \in N_k$ do
4: $\xi_{nk}^t = (Z_{nk}^t)^H \phi_n + x_{nk}^t$
5: $\Psi_{nk}^t = R_{nk}(R_{nk} + \Sigma_k)^{-1}$
6: $\Omega_{nk}^t = (\Sigma_k)^{-1} - (R_{nk} + \Sigma_k)^{-1}$
7: $\ell_{nk}^t = \frac{|R_{nk} + \Sigma_k|}{|\Sigma_k|} \exp \left( -\xi_{nk}^t \Omega_{nk} \xi_{nk} \right)$
8: $\theta_{nk}^t = (1 + \frac{1}{\gamma} \ell_{nk}^t)^{-1}$
9: $x_{nk}^{t+1} = \theta_{nk}^t \Psi_{nk}^t x_{nk}^t$
10: end for
11: $U_k^t = \frac{1}{N} \sum_{n \in N_k} \theta_{nk}^t \Psi_{nk}^t (I + (1 - \theta_{nk}^t) \xi_{nk}^t \xi_{nk}^t)^H \Omega_{nk}^t$
12: $Z_k^{t+1} = Y_k - \sum_{n \in N_k} \phi_n (x_{nk}^{t+1})^T + \frac{1}{T} Z_k^t U_k^t$
13: $\Sigma_k^{t+1} = \frac{1}{T} (Z_k^{t+1})^H (Z_k^{t+1})$
14: end for
15: end for

Equivalently, the LLR is
\[
\log \ell_n = \sum_{k \in K} \log \ell_{kn}, \tag{21}
\]

where
\[
\log \ell_{kn} = \log \frac{|R_{nk} + \Sigma_k|}{|\Sigma_k|} - \xi_{nk}^H \Omega_{kn} \xi_{nk}. \tag{22}
\]

Here, $\Sigma_k$ and $\Omega_{kn}$ are the $k$-th diagonal blocks of $\Sigma$ and $\Omega_n$, respectively, and $\xi_{nk}$ is the corresponding subvector of $\xi_n$. This means that the LLR $\log \ell_{kn}$ can be written as the sum of $\{\log \ell_{kn}\}$, which can be interpreted as the local LLRs after coherently processing the received signals at each AP. In the special case of i.i.d. Rayleigh fading, the LLR can be further simplified into
\[
\log \ell_{kn} = M \log \left( 1 + \frac{\rho_{kn}}{\tau_k} \right) - \frac{\rho_{kn} \xi_{nk}^2}{\tau_k (\rho_{kn} + \tau_k)}, \tag{23}
\]

where the quantity $\rho_{kn}/\tau_k$ can be interpreted as the signal-to-noise ratio (SNR).

Inspired by the factorization in (20), we propose a distributed approach to activity detection in distributed MIMO. The procedure is as follows: each AP runs the AMP algorithm locally by using only the received signal $Y_k$ and sends the local LLR $\log \ell_{kn}$ to the aggregator. Then, the aggregator computes the LLR $\log \ell_n = \sum_{k \in K} \log \ell_{kn}$ for activity detection.

D. Dynamic Cooperation Clustering

We assumed that each device was served by all APs. This configuration is not scalable in complexity and resource requirements as $N \to \infty$. Meanwhile, the AMP algorithm, or more generally, CS techniques, are known to work in the regime where the measurement size (pilot length) is larger than or equal to the support size (number of active devices).

To address these issues, we consider a dynamic cooperation (DCC) framework, such that a device is served only by the APs with indices in the set $K_n^d \subset K$. Conversely, an
AP only serves a subset of devices $\Lambda^d_{\ell} = \{n \in \mathcal{N} : k \in K^d_{\ell} \}$. There are two advantages of using the DCC framework: 1) the computational complexity is reduced; 2) the effective number of active devices served by an AP decreases.

Finally, by exploiting the DCC framework and our new findings about the MMSE denoiser, we propose a distributed AMP (dAMP) which is detailed in Algorithm 1. A centralized AMP (cAMP) is also developed in a similar way, while the step-wise details are omitted owing to space constraints. The key distinction in cAMP is that the denoiser for device $n$ is designed using $\theta_n^k = \left(1 + \frac{1}{\tau_n^k} \prod_{k \in K^d_{\ell}} \theta_{kn}^k \right)^{-1}$ by combining the local LLRs from all its serving APs in each iteration. These algorithms can be modified for other network structures. For example, multiple neighboring APs can coherently process the received signals. In this respect, cAMP (fully coherent) and dAMP (noncoherent) represent two extreme cases.

### E. Complexity Analysis

The computational complexity of dAMP with correlated fading is dominated by the calculation of matrix inversions and determinants in steps 5-7 of Algorithm 1 with complexity $O(M^3)$. Therefore, the overall complexity is $O(KTNM^3)$. For the i.i.d. Rayleigh case, the complexity of matrix-vector multiplications in steps 4, 7, and 9 is $O(M^2)$, and the matrix multiplications in steps 12 and 13 have complexity $O(LM^2)$. Since we are interested in the regime where $L \ll N$, the overall complexity becomes $O(KTNM^2)$. Notice that dAMP can be distributed, and the processing per AP has complexity $O(TNM^2)$. Furthermore, by using the DCC framework, we can replace $N$ by $\max_k |\Lambda^d_{\ell}|$.

For comparison, the covariance-based method in [7] has overall complexity $O(TN(K^3_{dom} + KL^2))$, where $K_{dom}$ is the number of dominant APs; for the typical case $K_{dom} < L$, the complexity becomes $O(KTNL^2)$. Note, however, that method of [7] is developed for the i.i.d. Rayleigh case and while extensions are possible, they are likely to incur higher complexity. Since the number of antennas is typically small on an AP, we have $M < L$, and our algorithms have lower complexity than that of [7].

### IV. Simulations

We consider a distributed MIMO system with $K = 20$ APs with $M = 3$ antennas each. A total of $N = 400$ devices are randomly dropped in a $2 \times 2$ km area with activity probability $\epsilon_n = 0.1$, $\forall n$. By using a wrap-around technique, we approximate an infinitely large network with 15 antennas and 10 active devices per square km. The pilots are random Gaussian sequences normalized to unit energy. The maximum transmit power is 23 dBm. The bandwidth is 1 MHz. The noise power spectral density is $-169$ dBm/Hz. The LSFC is generated by $-140.6 - 36.7 \log_{10}(d_n) + \Upsilon_i$ in dB, where $d_n$ is the distance from device $n$ to the AP in km, and $\Upsilon_i$ is the shadow fading effect distributed as $\mathcal{N}(0, \sigma_d^2)$, with standard deviation $\sigma_d = 4$ dB. The small-scale fading is modeled by i.i.d. Rayleigh for each pair of AP and device. The LSFC threshold for power allocation is set to satisfy $p_{\max, \ell n} = 6$ dB, $\forall n$. For the DCC framework, we connect each device to the 10 APs with the largest LSFC.

The performances of cAMP and dAMP are examined in Fig. 1 with or without the DCC framework and with different power allocation schemes.\(^3\) The covariance-based approach in [7] (with 3 dominant APs) and the hard-decision-and-fusion based AMP method\(^4\) in [8] are used as baselines for the simulations.

\(^3\)Code available at https://github.com/jiananbai/distributed-AMP.

\(^4\)Since [8] provided neither theoretical results nor algorithm details for the multi-antenna AP case, we use the expressions of probabilities of missed detection and false alarm in [2] to perform the decision fusion. Notice that this method uses a minimum-probability-of-error criterion and does not produce receiver operating characteristic (ROC) curves.
comparison. A runtime comparison is provided in Table I.

The results for pilot length \( L = 40 \) are shown in Fig. 1a. The following observations can be made: (i) When the pilot length is larger than or equal to the average number of active devices, AMP outperforms the covariance-based approach in almost all configurations since our AMP algorithms can coherently process received signals from more APs. (ii) AMP works better with full power. We hypothesize that this is because of the macro-diversity in distributed MIMO: for each device there are almost always some APs to which the path gain is better with full power. We showed that for activity detection in distributed MIMO: for each device there are almost always some APs to which the path gain is larger than or equal to the average number of active devices, \( \frac{\eta}{2} \theta(\mathbf{x}+\mathbf{v})^2(\mathbf{x}+\mathbf{v})^H \) (25)

\[ \hat{Q} \]

where \( \theta(\cdot) \) is defined in (14). By denoting as \( p_x(\mathbf{x}) \) and \( p_v(\mathbf{v}) \) the density functions of \( \mathbf{x} \) and \( \mathbf{v} \), respectively, the \( (i,j) \)-th element of \( Q \) is given by

\[ Q_{i,j} = \int_{\mathbf{x},\mathbf{v}} (x_i+u_j)(x_j+u_j)^\theta(x+v)^2p_x(x)p_v(v) \]  

\[ \hat{Q}_{i,j}(\mathbf{x},\mathbf{y}) \]

Denote by \( M_k \) the row (column) indices corresponding to the \( k \)-th diagonal block. Then \( f_{i,j}(\mathbf{x},\mathbf{y}) \) is partially odd in \( M_k \) if \( i \in M_k \) and \( j \notin M_k \), or \( i \notin M_k \) and \( j \in M_k \), and partially even in \( M_k \) otherwise. That is, \( Q_{i,j}=0 \) if the indices \( i \) and \( j \) are not in the same diagonal block. This means that \( Q \) is also block-diagonal with the same structure as \( \{R_n\} \). Then, the first term in (24), which equals to \( \Psi Q \Psi^H \), keeps the same block-diagonal structure.

By using similar arguments, one can show that the remaining terms have the same structure. We omit the details due to the limited space.

V. CONCLUSION

We showed that for activity detection in distributed MIMO: 1) the AMP algorithm can be implemented in a distributed manner with an acceptable performance loss; 2) the AMP algorithm outperforms the covariance-based approach when the pilot length is larger than or equal to the average number of active devices, although this is not the case in co-located MIMO; 3) the problem of pilot correlation can be alleviated by using the DCC framework, when the pilot length is less than the average number of active devices.

APPENDIX A

We prove Theorem 1 by induction. First, when the covariance matrices \( \{R_n\} \) share a block-diagonal structure, the initial state \( \Sigma^0 \) in (12) has the same block-diagonal structure. Then, assuming that \( \Sigma^t \) stays in this structure, we show that \( \Sigma^{t+1} \) has the same structure.

Definition 1. (Partially Odd or Even Function) A function \( f : \mathbb{R}^M \rightarrow \mathbb{R} \) is partially odd or even in indices \( \mathcal{I} \subset \mathcal{M} = \{1, \ldots, M\} \) if \( f(\eta_\mathcal{I}(x)) = -f(x) \) or \( f(\eta_\mathcal{I}(x)) = f(x) \), respectively. Here, \( \eta_\mathcal{I}(\cdot) \) is an element-wise operator with \( \eta_\mathcal{I}(x)_i \) equals to \(-x_i \) for \( i \in \mathcal{I} \), and \( x_i \) otherwise.

\[ \text{The simulations were performed on an Intel Xeon Gold 6130 Processor.} \]