

Graph Neural Network Aided Detection for the Multi-User Multi-Dimensional Index Modulated Uplink

Xinyu Feng, *Member, IEEE*, Mohammed EL-Hajjar, *Senior Member, IEEE*, Chao Xu, *Senior Member, IEEE*, and Lajos Hanzo, *Life Fellow, IEEE*

Abstract—The concept of Compressed Sensing-aided Space-Frequency Index Modulation (CS-SFIM) is conceived for the Large-Scale Multi-User Multiple-Input Multiple-Output Uplink (LS-MU-MIMO-UL) of Next-Generation (NG) networks. Explicitly, in CS-SFIM, the information bits are mapped to both spatial- and frequency-domain indices, where we treat the activation patterns of the transmit antennas and of the subcarriers separately. Serving a large number of users in an MU-MIMO-UL system leads to substantial Multi-User Interference (MUI). Hence, we design the Space-Frequency (SF) domain matrix as a joint factor graph, where the Approximate Message Passing (AMP) and Expectation Propagation (EP) based MU detectors can be utilized. In the LS-MU-MIMO-UL scenario considered, the proposed system uses optimal Maximum Likelihood (ML) and Minimum Mean Square Error (MMSE) detectors as benchmarks for comparison with the proposed MP-based detectors. These MP-based detectors significantly reduce the detection complexity compared to ML detection, making the design eminently suitable for LS-MU scenarios. To further reduce the detection complexity and improve the detection performance, we propose a pair of Graph Neural Network (GNN) based detectors, which rely on the orthogonal AMP (OAMP) and on the EP algorithm, which we refer to as the GNN-AMP and GEPNet detectors, respectively. The GEPNet detector maximizes the detection performance, while the GNN-AMP detector strikes a performance versus complexity trade-off. The GNN is trained for a single system configuration and yet it can be used for any number of users in the system. The simulation results show that the GNN-based detector approaches the ML performance in various configurations.

Index Terms—Index Modulation (IM), Multi-User, Graph Factor, Message Passing (MP), Machine Learning, Graph neural network (GNN).

A. ABBREVIATIONS AND ACRONYMS

NOMENCLATURE

Acronym

AMP	Approximate Message Passing
AWGN	Additive white Gaussian Noise
BER	Bit Error Ratio
BS	Base Station
CP	Cyclic Prefix
CS	Compressed Sensing
CS-SFIM	CS-aided Space-Frequency Index Modulation
CSI	Channel State Information
DL	Downlink
DNN	Deep Neural Network
EE	Energy Efficiency
EP	Expectation Propagation

FD	Frequency Domain
FFT	Fast Fourier Transform
FN	Factor Node
GFIM	Generalized OFDM-IM
GNN	Graph Neural Network
GSM	Generalized SM
HD	Hard-Decision
IFFT	Inverse Fast Fourier Transform
IM	Index Modulation
KL	Kullback-Leibler
LMMSE	Linear MMSE
LS-MU-MIMO	Large Scale MU-MIMO
MAP	Maximum A Posteriori
MIM	Multi-dimensional IM
MIMO	Multiple-Input Multiple-Output
ML	Maximum Likelihood
MLP	Multiple Layer Perceptron
MMSE	Minimum-Mean-Square-Error
MP	Message Passing
MRF	Markov Random Field
MU	Multi-User
MUD	Multi-User Detector
MUI	Multi-User Interference
NG	Next-Generation
NN	Neural Network
OAMP	Orthogonal AMP
OFDM	Orthogonal Frequency Division Multiplexing
OFDM-IM	OFDM associated with Index Modulation
PAPR	Peak to Average Power Ratio
PE	Power Efficiency
PMF	Probability Mass Function
RA	Receive Antennas
RC	Reduced-Complexity
ReLU	Rectified Linear Unit
RF	Radio Frequency
SD	Spatial-Domain
SE	Spectral Efficiency
SF	Space-Frequency
SIM	Subcarrier-IM
SM	Spatial Modulation
ST	Space-Time
TA	Transmit Antennas
TD	Time-Domain
UL	Uplink
VD	Virtual Domain

VN Variable Node

Symbols

U	number of users
G	number of subcarrier group over entire OFDM symbol
$b_{u,g}^1$	bits for FIM of u -th user at g -th group subcarriers of CS-SFIM
$b_{u,g}^2$	bits for SM of u -th user at g -th group subcarriers of CS-SFIM
$b_{u,g}^3$	bits for APM of u -th user at g -th group subcarriers of CS-SFIM
K	number of activated subcarrier for each subcarrier group
N_v	number of subcarriers in VD
N_f	number of subcarrier in FD
N_c	number of subcarrier of entire OFDM symbol
N_t	number of TAs
N_r	number of RAs
N_a	number of activated TAs
$S_{u,g}$	SF matrix of u -th user spread over the g -th group subcarriers of CS-SFIM
$\mathcal{L}$	number of constellation realizations of APM
$\mathbf{s}_{u,g}$	FD signal transmitted from the N_t TAs associated with g -th subcarrier block of the u -th user
$\mathbf{A}$	CS measurement matrix for each user and each subcarrier group
$\mathbf{H}$	channel model in FD
$\mathbf{w}$	Additive White Gaussian noise
$\mathbf{y}$	Received signal
$\mathbf{A}$	CS equivalent measurement matrix for each subcarrier group of all users
I_{AC}^g	selection pattern of active TAs for g -th subcarrier group of all users
I_{SI}^g	selection pattern of active subcarrier for g -th subcarrier group of all users
I^g	selection pattern of equivalent SF matrix for g -th subcarrier group of all users
$\mathbf{X}_{\mathcal{L}}^g$	realization selection of APM for g -th subcarrier group of all users
$\hat{\alpha}_g$	estimated realization of APM symbols for g -th subcarrier group of all users
$\hat{\beta}_g$	estimated index pattern of active subcarrier for g -th subcarrier group of all users
$\hat{\gamma}_g$	estimated index pattern of active TAs for g -th subcarrier group of all users
$\hat{\mathbf{s}}$	search space for the g -th subcarrier group of U users
$\mathbf{H}_g$	equivalent channel matrix in the g -th subcarrier group of all users
$\mathbf{x}_g$	equivalent signal transmitted by all users over the g -th subcarrier group and modulated by TA index, subcarrier index and APMs
λ	average SNR per symbol of g -th subcarrier group
Ω	size of the transmit signal vector x for each subcarrier group of all users
Φ	size of the received signal vector y for each subcarrier group of all users
L	number of active index elements in the SF matrix for each subcarrier group of all users
$\mathcal{R}_i$	non-zero element set for $\mathbf{H}$ of i -th transmit signal x_i

$\mu_{x_i \rightarrow y_j}^t(\mathbf{x}_i)$	message sent from VN x_i to FN y_j
$\mu_{y_j \rightarrow x_i}^t(\mathbf{x}_i)$	message feed back from FN y_j to FN x_i
$\phi_{\phi_i \rightarrow x_i}(\mathbf{x}_i)$	a priori probability of the transmitted symbol
$\mathbf{r}$	vector from x_i
Σ	mean value of the approximate Gaussian variable x_i
$\hat{r}$	variance value of the approximate Gaussian variable x_i
$\hat{x}$	updated residual after message exchanging
$\hat{v}$	estimated x_i mean value after message exchanging
$\mathcal{Q}$	estimated x_i variance after message exchanging
V_i	all pattern of $\mathcal{L}$ -APM set
η_i	mean value of an approximate unnormalized Gaussian distribution variable x_i
μ_{ep}	variance value of an approximate unnormalized Gaussian distribution variable x_i
Σ_{ep}	mean value of a modified approximate distribution variable in 42
$m_{o,i}^{(t)}$	variance value of a modified approximate distribution variable in 43
$v_{o,i}^{(t)}$	mean value of the EP approximates tractable distribution $q(\mathbf{x})$ in t -th iteration in 44
$m_i^{(t)}$	variance value of the EP approximates tractable distribution $q(\mathbf{x})$ in t -th iteration in 44
$v_i^{(t)}$	mean value of the cavity distribution of x_i in t -th iteration in 45
$\mathbf{a}_i$	variance value of the cavity distribution of x_i in t -th iteration in 45
$e_{i,j}$	attributes of node in GNN network layer
$\mathbf{f}_{ij}$	edge between node i and j
$\mathbf{U}_i$	attributes of edge $e_{i,j}$
$\mathbf{m}_{ji}$	hidden vector update in next round of the GNN node i
$\phi(x_i)$	message from node j to node i to update the feature vector
$\psi(x_i, x_j)$	characterized self potential of i -th node
	characterized pair potential of edge $e_{i,j}$

I. INTRODUCTION

MULTIPLE -Input Multiple-Output (MIMO) schemes have been widely used in wireless systems for improving performance and Spectral Efficiency (SE). They have also been combined with Orthogonal Frequency Division Multiplexing (OFDM) for communication over frequency-selective channels [1]. Multi-User (MU) MIMO systems, which benefit from further improved SE [1] and Power Efficiency (PE) [2], are widely considered as a key technique for Next-Generation (NG) wireless systems [3]. However, a persistent challenge in MU-MIMO systems is the presence of Multi-User Interference (MUI) [4].

Index Modulation (IM) [5] constitutes a potential candidate for next-generation wireless systems as a benefit of its flexible resource activation [6]. The concept of IM has been derived from that of Spatial Modulation (SM) [7] [8], which is capable of striking a flexible performance versus complexity trade-off using a single Radio Frequency (RF) chain [9], [10] along with low-complexity detection. To achieve improved flexibility, generalized SM (GSM) has been proposed, where multiple

Transmit Antennas (TA) can be activated simultaneously [11] and utilized for MU scenarios [12].

Then, the concept of SM has been extended to the frequency domain, where the resultant philosophy of IM has been proposed in [13]. OFDM associated with index modulation (OFDM-IM), which has also been referred to as Subcarrier-IM (SIM) [14] was proposed for increasing the Energy Efficiency (EE) and performance compared to traditional OFDM, where only a fraction of the subcarriers are activated in the Frequency Domain (FD) with the index of the active subcarriers implicitly conveying extra information [15]. In [16], Fan *et al.* generalized the OFDM-IM principle for increasing the spectral efficiency. Additionally, both the Uplink (UL) and Downlink (DL) of MU OFDM-IM have been investigated in [17], where the Peak-to-Average Power Ratio (PAPR) has been reduced compared to that of classical MU OFDM systems. However, subcarrier-index modulated OFDM suffers from a potential throughput reduction compared to the classic OFDM due to the deactivation of a number of subcarriers. Hence, Zhang *et al.* [18] proposed an improved SIM concept relying on Compressed Sensing (CS) [19], which benefits from the sparsity of symbols in the FD by compressing the sparse transmit vector.

Recently, the concept of IM has also been applied to multiple dimensions, which we refer to as Multi-dimensional IM (MIM) [20]–[24]. CS techniques can also be harnessed in the MIM scheme. For instance, in [25] a CS-aided MIM scheme was conceived, which applied IM both in the frequency and spatial-domain. Then, in [26], a CS-aided joint MIM system has been proposed, which jointly designed multiple dimensions and harnessed the CS technique. Furthermore, applying the MIM in Large Scale MU-MIMO (LS-MU-MIMO) scenarios is promising. Hence, CS-aided Space-Frequency Index Modulation (CS-SFIM) was proposed in [27] for LS-MU-MIMO-UL scenarios. *However, a significant open challenge for LS-MU-MIMO systems is the design of a reliable, yet low-complexity Multi-User Detector (MUD) for employment at the Base Station (BS).*

Conventional linear detectors, such as the MMSE detector, are effective in MU-OFDM scenarios, but their performance is far from that of the complex exhaustive-search-based Maximum Likelihood (ML) detector [17]. However, ML detectors may only be harnessed for systems supporting a low number of users due to their potentially excessive computational complexity. As a remedy, non-linear detectors such as Message Passing (MP) schemes [28] [29] achieve a good performance at a reasonable complexity for MIMO systems. MP-based detectors rely on Gaussian distributions for approximating the posterior probability distribution of the transmitted symbols conditioned on the received signal. Hence, the Approximate Message Passing (AMP) [28] detector has been considered both in SM [30] and OFDM-IM [31] schemes. Then, Chakrapani *et al.* [32] proposed a MP-based low-complexity detection method for reducing the complexity of SFIM detection. If we extend the single user scenario to massive MIMO and MU-MIMO schemes, a high MUI is introduced, where the AMP [28] detector performs poorly in the presence of imperfect Channel State Information (CSI). These

problems have been partially resolved by the Orthogonal AMP (OAMP) detector [33] that integrates the AMP with the linear MMSE filtering. As a further design alternative, Expectation Propagation (EP) [34] [35] based detectors rely on a posterior distribution approximation based on the independent Gaussian approximation and the Gaussian cavity distribution in the detection process. Hence, they can significantly outperform the OAMP detector, as a benefit of harnessing regularization parameters into the MMSE filter, which are iteratively adjusted based on both the CSI and on the level of MUI. However, it has been shown that the use of this posterior distribution approximation leads to the partial loss of MUI information in these detectors and a severe performance degradation in a high-MUI scenario. *Therefore, despite the aforementioned improvements in MP algorithms, there remains a significant performance gap compared to the ML detector.*

Neural Network (NN)-based detectors have recently been proposed in [36]–[39] for striking an improved performance versus complexity trade-off in the case of imperfect CSI and MUI. To be more specific, this may be achieved by embedding the iteration process of the MP detectors into NN layers and optimizing the NN parameters during the training phase. For example, the OAMPNet detector of [36] combines the OAMP and NN to deal with potential ill-conditioned channel matrices, which results in a significant performance improvement compared to the conventional OAMP detector. Additionally, the Graph Neural Network (GNN) aided detector [39] [40] is based on the belief propagation algorithm and utilizes a pairwise Markov Random Field (MRF) model, which significantly reduces the number of NN parameters. However, a significant performance gap remains between the practical state-of-the-art detectors and the ML detector in the presence of ill-conditioned channel matrices and high MUI. Then, a AMP-based GNN model was proposed in [41] for MU-OFDM scenarios, which significantly reduces the detection complexity, while attaining near-ML performance. Furthermore, the EP-based GNN model (GEPNet) was proposed in [42], [43] for achieving near-optimal performance, while slightly increasing the computational complexity compared to the GNN-AMP.

Against the above backdrop, in this paper, we propose a CS-SFIM scheme for the LS-MU-MIMO-UL, which combines the benefits of LS-MU-MIMO, SM, GSM, OFDM-IM and CS for UL transmission over frequency-selective fading channels. In our proposed solution, we design the index of the TAs and the subcarriers separately for attaining higher implementation flexibility than the joint space-frequency indexing of [20], while the Space-Frequency (SF) domain indexing is considered as a joint SF matrix for factor graph design. Then, we design an MP-based reduced-complexity MUD for the proposed system, which incorporates AMP and EP philosophies in order to strike an attractive Bit Error Rate (BER) versus complexity trade-off. Guided by this analysis, we propose a GNN based framework for improving the detection performance by fine-

TABLE I: Contrasting our contributions to the literature

Contribution	proposed*	[44]	[45]	[46]	[40]	[43]	[41]	[42]	[27]	[17]	[47]
MU scenario	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Large-scale MIMO	✓	✓	✓						✓	✓	✓
CS at the transmitter	✓								✓		
Separate MIM	✓								✓		✓
MMSE-based GNN detector	✓					✓					
AMP-GNN detector	✓						✓				
GEPnet detector	✓							✓			

tuning the cavity distribution¹. Compared to the conventional Gaussian cavity function parameterized only by its mean and variance, the GNN based framework introduces additional parameters into the cavity function to capture more accurate MUI information [39]. The choice of the AMP-based GNN detector is motivated by its low computational complexity, in contrast to the EP detector since it does not have to perform any matrix inversion operation. We refer to the newly developed GEPNet detector, which can iteratively achieve an improved detection performance compared to the conventional EP [34] [35]. We design both GEPNet and GNN-AMP detectors exhibiting robustness to the fluctuating number of users, which allows the proposed detectors to be trained only once even for systems supporting a time-variant number of users. Our simulation results show that the GEPNet and GNN-AMP detectors are capable of significantly outperforming the EP [34], the AMP [33] and the GNN [40] based detectors.

In summary, Table I boldly contrasts the novelty of this paper to the literature. More explicitly, the contributions of this paper are elaborated on as follows:

- 1) We propose a CS-SFIM for the LS-MIMO-MU-UL scenario to strike a compelling complexity versus performance trade-off. For each user, the information is conveyed over the FD-IM, the Spatial-Domain (SD) IM and the classic Amplitude Phase Modulation (APM) symbols. Additionally, to attain an improved diversity gain, in contrast to the MIM scheme of [25], different TA activation patterns are proposed for the different transmitted APM symbols instead of constraining the active TAs for the entire subcarrier group of OFDM symbol. The received signal model can be formulated based on a SF domain matrix, which is a sparse matrix that can facilitate CS-SFIM for LS-MIMO-MU-UL.
- 2) We design MP detectors for the CS-SFIM-MU system, where we evaluate the accuracy of their joint posterior distribution approximations and analyze the relationship between this accuracy and the cavity distributions used for calculating the symbol estimates. Then, we compare the computational complexity of the MP-based detector to that of the ML and MMSE detectors. We demonstrate that the AMP and EP detectors are capable of striking a better BER performance versus complexity trade-off

than the conventional MU-MIMO-OFDM system using the ML and MMSE-based detector.

- 3) As a further advance, we develop a GNN-based framework to improve the posterior distribution approximation by fine-tuning the cavity distributions in the MP detectors. The proposed GNN-based framework is designed to allow processing of the *a priori* knowledge generated by the MP detectors. We propose a pair of novel GNN based detectors that use the GNN-based framework developed to improve the classic EP and AMP. The first detector, namely the AMP-GNN scheme is capable of achieving a reasonable detection performance at low complexity, while the GEPNet detector is proposed for attaining an improved detection performance by introducing iterations.
- 4) We demonstrate that training the proposed GNN-based detector only once is sufficient, even if we have a variable number of users supported by the proposed system. This conclusion is verified by mathematical proof that the proposed detectors share GNN's permutation equivariance property. Our simulation results further demonstrate that the proposed GNNs are robust to changes in the number of users. By contrast, conventional Deep Neural Network (DNN) based detectors require individual training for each specific number of users.

The rest of the paper is organized as follows. In Section II, we present the proposed system model. In Section III, we present our linear and non-linear MU detectors. Then, in Section III-C we develop our proposed GNN-based MP detector for the CS-SFIM system considered. In Section IV, we present our simulation results, while our conclusions are offered in Section V.

II. SYSTEM MODEL

In this section, we introduce the transceiver model of the CS-SFIM aided LS-MIMO-MU-UL supporting U UL users, where each user is equipped with N_t TAs communicating with a BS having N_r Receive Antennas (RA), where N_r is on the order of tens to hundreds. The modulation of the proposed CS-SFIM scheme at the transmitter is illustrated in Fig. 1, where both Generalized OFDM-IM (GFIM) and SM techniques are harnessed for each user to improve the UL throughput and diversity gain. The specific procedures of both subcarrier and TA index modulation designs in CS-SFIM systems are detailed in the following subsections.

¹In EP based detection, EP approximate the distribution with a tractable distribution, to achieve more accurate estimation compared with Gaussian distribution. Then, to further increase the accuracy of the replaced distribution, a cavity distribution is introduced to update the approximate factor of aimed distribution.

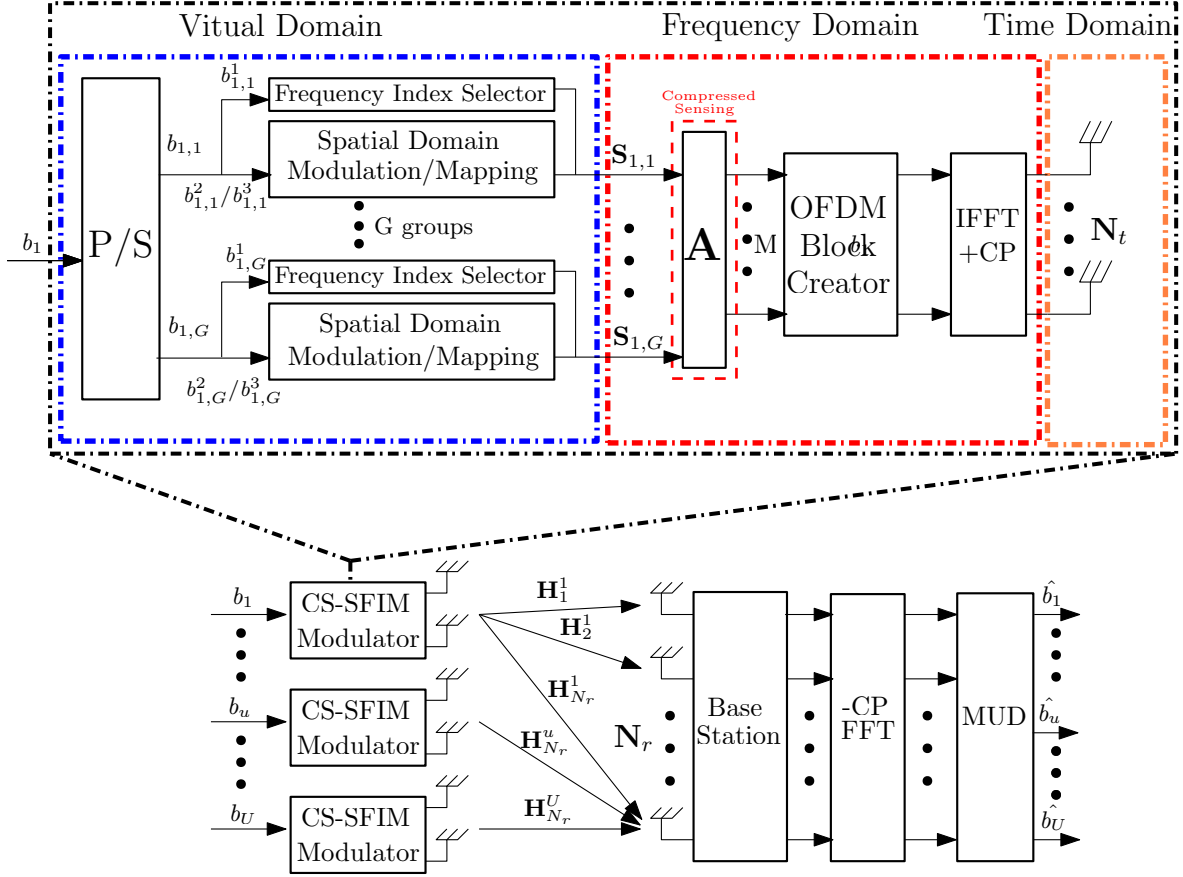


Fig. 1: MU CS-SFIM UL transceiver architecture, where the BS has N_r RAs to serve the U UL users and each user has N_t TAs.

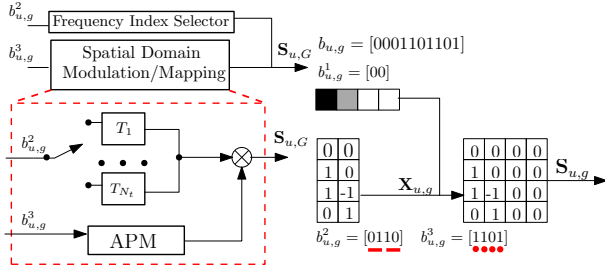


Fig. 2: structure and Example of CS-SFIM modulation at g -th group of u -user with $k = 2$ and bit sequence of [0001101101].

A. UL Transmission

As shown in Fig. 1, in the block diagram of the LS-MU CS-SFIM system, there are U UL users communicating with a BS, where IM is employed by each user. As shown in Fig. 1, the bit sequence b_u of the u -th UL user is split into G subcarrier groups for the CS-SFIM schemes, where all the G subcarrier groups have the same modulation process. For the CS-SFIM scheme, the information sequence $b_{u,g}$ is split into $b_{u,g}^1$ and $b_{u,g}^2$ and $b_{u,g}^3$. Explicitly, $b_{u,g}^3$ represents the APM constellation, $b_{u,g}^1$ is conveyed by the Virtual-Domain (VD) Frequency Index Selector block and $b_{u,g}^2$ is conveyed by the SM modulation.

1) *Virtual-Domain Frequency Index Selection*: With the aid of CS techniques, we can readily extend the set of available subcarriers by introducing the VD², where the bit sequence $b_{u,g}^1$ of length $\lfloor \log_2(C_{N_v}^K) \rfloor$ is entered into the Frequency index selector of Fig. 1. For each virtual subcarrier block, we consider N_v available subcarriers for the GFIM modulation. Specifically, the bit sequence $b_{u,g}^1$ controls the subcarrier selection. A simple example is presented in Table II for $K = 2$ and $N_v = 4$. If $b_{u,g}^1 = [0 \ 0]$, the first and the second subcarriers are activated for carrying information, while the other subcarriers remain inactive.

2) *Spatial-Domain Modulation/Mapping*: As shown in Fig. 1, for each subcarrier group, the information sequence $b_{u,g}^2$ is conveyed by the SD mapping block. To increase the SE in the SD, generalized SM is utilized in the system of Fig. 1, where N_a out of N_t TAs associated with $N_a \in (1, 2, \dots, N_t)$ are activated by the antenna selector. The information sequence $b_{u,g}^2$ in the g -th subcarrier group of the u -th user has a length of $K \lfloor \log_2 C(N_t, N_a) \rfloor$, where $\lfloor \log_2 C(N_t, N_a) \rfloor$ bits are needed for selecting the activated N_a TAs, which are applied to K activated subcarriers.

3) *Space-Frequency Index Modulation*: As shown in Fig. 1, the active subcarrier of the VD frequency index selector and

²VD is the domain where the FIM is deployed. This concept was first introduced by Zhang *et al.* in [48] in OFDM-IM systems for improving the spectral efficiency.

the active antenna of the SD antenna selector can form a SF matrix $\mathcal{S}_{u,g} \in \mathbb{C}^{N_t \times N_v}$. More specifically, the SF matrix is obtained by assigning the column index to the SD active antenna and the row index to the active VD subcarriers, which results in $N_a \cdot K$ non-zero elements in the sparse SF matrix $\mathcal{S}_{u,g}$.

TABLE II: A Look-up table of subcarrier selection for CS-SFIM systems having $K = 2, N_v = 4$.

b_2	Indices	Allocation
[0 0]	(1,2)	$[K_1 \ K_2 \ 0 \ 0]$
[0 1]	(2,3)	$[0 \ K_1 \ K_2 \ 0]$
[1 0]	(3,4)	$[0 \ 0 \ K_1 \ K_2]$
[1 1]	(1,4)	$[K_1 \ 0 \ 0 \ K_2]$

4) *APM*: The bit sequence $\mathbf{b}_{u,g}^3$ of length $N_a K \log_2 \mathcal{L}$ is mapped to the $N_a K$ positions of the non-zero elements in $\mathcal{S}_{u,g}$ to create the $(N_a \cdot K)$ $\mathcal{L}$ -ary APM symbols of Fig. 2. Assuming that we have $K = 2, N_a = 2, N_t = 4, N_v = 4$ and $\mathcal{L} = 2$, the information sequence $\mathbf{b}_{u,g} = \mathbf{b}_{u,g}^1 + \mathbf{b}_{u,g}^2 + \mathbf{b}_{u,g}^3 = \lfloor \log_2(C_{N_v}^K) \rfloor + K \lfloor \log_2 C_{N_a}^{N_t} \rfloor + N_a K \log_2 \mathcal{L}$, has a length of $(\log_2 4 + 2 \times \log_2 4 + 2 \times 2 \log_2 2) = 10$ bits. As shown in the example in Fig. 2, we have an information bit sequence of $\mathbf{b}_{u,g} = [0001101101]$. Then $\mathbf{b}_{u,g}^1 = [00]$ represents that the first and second subcarriers in a subcarrier group are activated as shown in Table II, while $\mathbf{b}_{u,g}^2 = [0110]$ indicates that the second and third TAs are activated for the first subcarrier and that the third and fourth antennas are activated for the second subcarrier, while the total number of activated subcarriers in each VD subcarrier group is $K = 2$. Then a SF matrix $\mathcal{S}_{u,g} \in \mathbb{C}^{4 \times 4}$ is constructed with the aid of $\mathbf{b}_{u,g}^1$ and $\mathbf{b}_{u,g}^2$, which has four non-zero elements, as shown in Fig. 2. Lastly, $\mathbf{b}_{u,g}^3 = [1101]$ corresponds to allocating BPSK symbols of (1), (1), (-1) and (1) to the corresponding position of non-zero elements in the SF matrix.

5) *Compressed Sensing and Block Assembly*: For G subcarrier groups per OFDM symbol, each subcarrier group symbol is modulated by the FIM-SM block in parallel, G SF matrices $\mathcal{S}_{u,g} (g = 1, 2, \dots, G)$ are established for the u -th user. Then, for each subcarrier group of the τ -th TA of the u -th user, the CS measurement matrix $\mathbf{A} \in \mathbb{C}^{N_f \times N_v}$ is harnessed for compressing the N_v -dimensional vector $\mathcal{S}_{u,g}^\tau (\tau = 1, 2, \dots, N_t)$ gleaned in the VD into a N_f -dimensional vector $\mathbf{s}_{u,g}^\tau$ from the τ -th TA in the FD, which can be expressed as:

$$\mathbf{s}_{u,g}^\tau = \mathbf{A} \mathcal{S}_{u,g}^\tau. \quad (1)$$

Then, the FD signals $\mathbf{s}_{u,g}$ transmitted from the N_t TAs associated with g -th subcarrier block of the u -th user, where $\mathbf{s}_{u,g} \in \mathbb{C}^{N_t \times N_f}$. Afterwards, the Block Assembler of Fig. 1 gathers G FD signals $\mathbf{s}_{u,g}$ to form a long SF frame $\mathbf{s}_u \in \mathbb{C}^{N_t \times N_f G}$, which contains N_t OFDM symbols for transmission from N_t TAs and can be formulated as $\mathbf{s}_u = [\mathbf{s}_{u,1}, \mathbf{s}_{u,2}, \dots, \mathbf{s}_{u,G}]$. Additionally, the entire OFDM has N_c subcarrier number, which meets the condition that $N_c = N_f \times G$. Then, Inverse Fast Fourier transform (IFFT) is applied to the OFDM symbol at each TA to obtain the Time-Domain (TD) symbol, which is then extended by the Cyclic Prefix (CP). Afterwards, the TD signals of the u -th user are transmitted from N_t TAs

over the wireless channel. Hence, for the LS-MU-MIMO-UL scenario, the maximum achievable rate of the proposed CS-SFIM system in bits/s/Hz can be expressed as [27]:

$$R_t = U \times \frac{G(\log_2(C_{N_v}^K) + K \log_2 C_{N_a}^{N_t}) + N_a K \log_2 \mathcal{L}}{N_c + N_{cp}}. \quad (2)$$

To achieve the theoretical maximum transmission rate, a powerful detector is required in this scenario at an affordable complexity.

III. MULTI-USER DETECTION

In Fig. 1, the MU-UL signals are superimposed at the BS and received through distinct MIMO channels. Despite simultaneous transmission, user separability is achieved due to the unique SF signatures imposed by each user's channel [49] [50]. The MIMO channel model of U UL users can be expressed as $\mathbf{H}^u \in \mathbb{C}^{N_r \times N_t} (u = 1, 2, \dots, U)$, where $\mathbf{H}_i^u[n] \in \mathbb{C}^{1 \times N_t} (i = 1, 2, \dots, N_r, u = 1, 2, \dots, U, n = 1, 2, \dots, N_c)$ denotes the complex channel gain of the n -th subcarrier received from the N_t TAs of the u -th user at the i -th RA. Assuming perfect synchronization³, then the FD signal $\mathbf{y}_i^u[n]$ received from the u -th user at the i -th RA after CP removal and Fast Fourier Transform (FFT)⁴ is given by

$$\mathbf{y}_i^u[n] = \mathbf{H}_i^u[n] \mathbf{s}_u[n] + \mathbf{w}_i^u[n], \quad (3)$$

where $\mathbf{s}_u \in \mathbb{C}^{N_t \times 1}$ denotes the FD symbols transmitted on the n -th subcarrier of the N_t TAs and $\mathbf{w}_i^u[n]$ is the Additive White Gaussian Noise (AWGN) with zero mean and variance of σ^2 . Then we can express the signal $\mathbf{y}_i^u \in \mathbb{C}^{N_c \times 1}$ received by the i -th RA from the u -th user as:

$$\mathbf{y}_i^u = \mathbf{H}_i^u \mathbf{s}_u + \mathbf{w}_i^u, \quad (4)$$

where $\mathbf{H}_i^u = \text{diag}\{\mathbf{H}_i^u[n]\}, (n = 1, 2, \dots, N_c)$ $\mathbf{H}_i^u \in \mathbb{C}^{N_c \times (N_t N_c)}$, denotes the diagonal structure of the complex channel matrix from the N_t TA to the i -th RA. Furthermore, $\mathbf{s}_u \in \mathbb{C}^{(N_t N_c) \times 1} = [(\mathbf{S}_u[1])^T (\mathbf{S}_u[2])^T \dots (\mathbf{S}_u[N_c])^T]^T$ represents the FD symbols vector of length N_c for the u -th user transmitted from the N_t TAs, which is sparse with $(N_a \cdot K \cdot G)$ non-zero elements. Moreover, $\mathbf{w}_i^u \in \mathbb{C}^{N_t \times 1}$ denotes the AWGN vector, with $\mathcal{CN}(0, \sigma_n^2)$. Then, the FD symbols $\mathbf{y}_i \in \mathbb{C}^{N_c \times 1}$ received at the i -th RA are expressed as:

$$\mathbf{y}_i = \sum_{u=1}^U \mathbf{y}_i^u = \mathbf{H}_i \mathbf{s} + \mathbf{w}_i, \quad (5)$$

where we have $\mathbf{H}_i = [\mathbf{H}_i^1, \mathbf{H}_i^2, \dots, \mathbf{H}_i^U]$, $\mathbf{s} = [(\mathbf{s}_1})^T, (\mathbf{s}_2})^T, \dots, (\mathbf{s}_U})^T] \in \mathbb{C}^{(N_t N_c U) \times 1}$, and $\mathbf{w}_i = \sum_{u=1}^U \mathbf{w}_i^u$. For the overall MIMO model, the

³While the synchronization among multiple users (MUs) is an important practical consideration, it is beyond the scope of this paper. Several techniques addressing MU synchronization have been proposed in the literature, including [51] [52].

⁴All users simultaneously transmit their signal in the UL from N_t TAs, which impinge on multiple RAs at the BS. The cyclic prefix (CP) removal and FFT based demodulation operations are then applied jointly to these stacked signals, rather than to each user's signal individually.

signal $\mathbf{y} \in \mathbb{C}^{(N_r N_c) \times 1}$ received over N_r RAs at the BS can be modeled by

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{w}, \quad (6)$$

where we have $\mathbf{y} = [\mathbf{y}_1^T \mathbf{y}_2^T \cdots \mathbf{y}_{N_r}^T]^T$, $\mathbf{H} = [\mathbf{H}_1^T \mathbf{H}_2^T \cdots \mathbf{H}_{N_r}^T]^T \in \mathbb{C}^{(N_r N_c) \times (N_t N_c U)}$ and $\mathbf{w} = [\mathbf{w}_1^T \mathbf{w}_2^T \cdots \mathbf{w}_{N_r}^T]^T \in \mathbb{C}^{(N_r N_c) \times 1}$. As shown in Fig. 1, the FD signal y can be split into G groups. Hence, the sub-group signal $\mathbf{y}_g \in \mathbb{C}^{(N_r N_f) \times 1}$ of the g -th subcarrier group of the U users received from the N_t TAs over the N_r RAs at the BS can be expressed as:

$$\mathbf{y}_g = \mathbf{H}_g \mathbf{s}_g + \mathbf{w}_g. \quad (7)$$

In (7) the g -th subcarrier group $\mathbf{s}_g \in \mathbb{C}^{(N_t N_f U) \times 1}$ denotes the g -th subcarrier group FD signals transmitted from the N_t TAs of the U users. Additionally, $\mathbf{w}_g \in \mathbb{C}^{(N_r N_f U) \times 1}$ is the Gaussian noise vector. The equivalent channel matrix $\mathbf{H}_g \in \mathbb{C}^{(N_r N_f) \times (N_t N_f U)}$ corresponding to the g -th subcarrier group of the U users transmitted signals from the N_t TAs over the N_r RAs can be expressed as:

$$\mathbf{H}_g = \begin{bmatrix} \mathbf{H}_{1,g}^1 & \mathbf{H}_{1,g}^2 & \cdots & \mathbf{H}_{1,g}^U \\ \mathbf{H}_{2,g}^1 & \mathbf{H}_{2,g}^2 & \cdots & \mathbf{H}_{2,g}^U \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{N_r,g}^1 & \mathbf{H}_{N_r,g}^2 & \cdots & \mathbf{H}_{N_r,g}^U \end{bmatrix}, \quad (8)$$

where $\mathbf{H}_{i,g}^u \in \mathbb{C}^{N_f \times (N_t N_f)}$, represents the complex-valued channel from N_t TAs to the i -th RA of the u -th user received. Then, $\mathbf{H}_{i,g}$ can be written as:

$$\mathbf{H}_{i,g}^u = \begin{bmatrix} \mathbf{H}_{i,g}^u[1] & 0 & \cdots & 0 \\ 0 & \mathbf{H}_{i,g}^u[2] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{H}_{i,g}^u[N_f] \end{bmatrix}, \quad (9)$$

where $\mathbf{H}_{i,g}^u[n] \in \mathbb{C}^{1 \times N_t}$, $n = 1, 2, \dots, N_f$ represents the channel of the n -th FD subcarrier in the g -th subcarrier group of the u -th user transmitted from the N_t TAs to the i -th RA. Then, the signals $\mathbf{s}_g$ in (7) transmitted to the g -th group can be expressed as:

$$\mathbf{s}_g = \bar{\mathbf{A}} \mathbf{I}_{SI}^g \mathbf{I}_{AC}^g \mathbf{X}_L^g + \mathbf{w}_g, \quad (10)$$

where $\bar{\mathbf{A}} \in \mathbb{C}^{(N_t N_f U) \times (N_t N_v U)}$ denotes the equivalent⁵ CS measurement matrix of $\mathbf{A}$ in (4), $\mathbf{I}_{SI}^g \in \mathbb{C}^{(N_t N_v U) \times (N_t K U)}$ represents the subcarrier index activation pattern, while $\mathbf{I}_{AC}^g \in \mathbb{C}^{(N_t K U) \times (K N_a U)}$ represents the equivalent TA index activation pattern for U users in the g -th subcarrier group.

Furthermore, $\mathbf{X}_L^g \in \mathbb{C}^{(K N_a U) \times 1}$ in (18) represents the vector, which has $K N_a U$ APM symbols in the g -th subcarrier group for U users. Therefore, we can extend (7) as:

$$\mathbf{y}_g = \mathbf{H}_g \bar{\mathbf{A}} \mathbf{I}_{SI}^g \mathbf{I}_{AC}^g \mathbf{X}_L^g + \mathbf{w}_g = \mathbf{H}_g \bar{\mathbf{A}} \mathbf{I}_g \mathbf{X}_L^g + \mathbf{w}_g, \quad (11)$$

where $\mathbf{I}_g \in \mathbb{C}^{(N_t N_v U) \times (K N_a U)} = \mathbf{I}_{SI}^g \mathbf{I}_{AC}^g$ is the equivalent SF matrix containing the information of the active TAs and the active subcarriers of U users in the g -th subcarrier group.

⁵In this case, $\bar{\mathbf{A}}$ is a larger matrix than $\mathbf{A}$, which is a grouped measurement matrix for compressing U users' signal transmitted from N_t TAs.

Then, we consider the case that not all users are active at a time. We introduce a activity indicator $\rho_u \in \{1, 0\}$ and modify (5) as

$$\mathbf{y}_i = \sum_{u=1}^U \rho_u \mathbf{y}_i^u = \mathbf{H}_i \mathbf{s} + \mathbf{w}_i, \quad (12)$$

and (6) can be modified as

$$\mathbf{y} = \mathbf{H} \mathbf{I}_\rho \mathbf{s} + \mathbf{w}, \quad (13)$$

and for each subcarrier group

$$\mathbf{y}_g = \mathbf{H}_g \mathbf{I}_\rho \bar{\mathbf{A}} \mathbf{I}_g \mathbf{X}_L^g + \mathbf{w}_g, \quad (14)$$

where $\mathbf{I}_\rho$ indicates the active users, which is decided by K_a .

In the following, we will detail both conventional detectors and learning-based detectors.

A. Conventional Detectors

Conventional exhaustive-search based ML detection can be applied at the receiver, albeit this may lead to excessive complexity [13]. In the following section, we present the conventional ML-based Hard-Decision (HD) detector, followed by the MMSE detector. Then, the MP-based AMP and EP detectors are discussed, followed by the proposed GNN-based methods, where the GNN is applied to improve the MP and AMP detector performance iteratively.

1) *Maximum Likelihood Detection:* As shown in Fig. 1 and Fig. 1, the information sequence $\mathbf{b}_u$ ($u = 1, 2, \dots, U$) is processed group by group by the CS-SFIM modulator. Then the signal received at the BS is detected by the MU detector of Fig. 1. To simplify the analysis, we only introduce the ML-aided detector for the g -th subcarrier group, which can be generalized to the entire received signal. Then, we assume that $\langle \alpha_g, \beta_g, \gamma_g \rangle$ represents the information in the g -th subcarrier group for U users of the CS-SFIM system shown in Fig. 1. Here, $(\mathbf{X}_L^g)_{\alpha_g}$ denotes the α_g -th realization of the $K N_a U$ APM symbols for $\alpha_g = 1, 2, \dots, L^{K N_a U}$. Furthermore, $(\mathbf{I}_{SI}^g)_{\beta_g}$ represents the β_g -th indexing pattern of the active subcarrier of the U users, where $\beta_g = 1, 2, \dots, N_{SI}^U$. $(\mathbf{I}_{AC}^g)_{\gamma_g}$ denotes the γ_g -th active TA selection pattern for the $K U$ active subcarriers in the g -th subcarrier group of U users, where we have $\gamma_g = 1, 2, \dots, (N_{AC}^K)^U$. Correspondingly, we use $\langle \hat{\alpha}_g, \hat{\beta}_g, \hat{\gamma}_g \rangle$ to represent the estimation of the input information. Hence, we have the optimal HD ML detector based on (7) and (11) as:

$$\begin{aligned} \langle \hat{\alpha}, \hat{\beta}, \hat{\gamma} \rangle &= \arg \min_{\alpha, \beta, \gamma} \|\mathbf{Y}_g - \mathbf{H}_g \bar{\mathbf{A}} (\mathbf{I}_{SI}^g)_\alpha (\mathbf{I}_{AC}^g)_\beta (\mathbf{X}_L^g)_\gamma\|^2, \\ &= \arg \min_{\alpha, \beta, \gamma} \|\mathbf{Y}_g - \mathbf{H}_g \bar{\mathbf{A}} \hat{\mathbf{s}}\|^2, \end{aligned} \quad (15)$$

where $\hat{\mathbf{s}} = (\mathbf{I}_{SI}^g)_\alpha (\mathbf{I}_{AC}^g)_\beta (\mathbf{X}_L^g)_\gamma$ denotes the search space of the g -th subcarrier group. The ML detector achieves the optimal detection performance with excessive complexity that increases exponentially with throughput.

2) *MMSE*: Based on the signal $\mathbf{y}_g$ of g -th subcarrier group in (16) and (11) received from all users, the signal of the g -th subcarrier group can be reformulated as:

$$\mathbf{y}_g = \bar{\mathbf{H}}_g \mathbf{x}_g + \mathbf{w}_g, \quad (16)$$

where $\bar{\mathbf{H}}_g \in \mathbb{C}^{(N_r N_f) \times (N_t N_v U)} = \mathbf{H}_g \bar{\mathbf{A}}$ is the equivalent channel matrix of all users in the g -th subcarrier group and $\mathbf{x}_g \in \mathbb{C}^{N_t N_v U \times 1}$ is the equivalent information signal transmitted by all users over the g -th subcarrier group and modulated by TA index, subcarrier index and APM.

Then, the Linear MMSE (LMMSE) detector is applied to the received signal $\mathbf{y}_g$ of (16) for mitigating the MU interference and acquiring the estimator $\hat{\mathbf{x}}_g \in \mathbb{C}^{(N_t N_v U) \times 1}$ of the transmitted signal $\mathbf{x}_g$ of (16), which can be expressed as:

$$\hat{\mathbf{x}}_g = (\bar{\mathbf{H}}_g^H \bar{\mathbf{H}}_g + \frac{1}{\lambda} \mathbf{I}_{(N_t N_f U)})^{-1} \bar{\mathbf{H}}_g^H \mathbf{y}_g, \quad (17)$$

where $\lambda = E\|\bar{\mathbf{H}}_g^H \mathbf{x}\|_2^2 / E\|\mathbf{w}_g\|_2^2$ represents the average SNR per symbol and is expressed as $\lambda = \frac{1}{\sigma^2}$. Afterwards, the estimated signal $\hat{\mathbf{s}}_g$ is split into U vectors in parallel, and transformed into matrix form as $[\hat{\mathbf{x}}_g^1 \hat{\mathbf{x}}_g^2 \cdots \hat{\mathbf{x}}_g^U]^T \in \mathbb{C}^{U \times N_t N_f}$, where $\hat{\mathbf{x}}_g^u \in \mathbb{C}^{(N_t N_f) \times 1}$ ($u = 1, 2, \dots, U$) is the MMSE estimate of the u -th users' signal. Based on (11) and (17), the signal $\hat{\mathbf{x}}_g^u$ estimated for the u -th user can be expressed as:

$$\hat{\mathbf{x}}_g^u = \mathbf{I}_g^u \mathbf{X}_L^{g,u} + \hat{\mathbf{w}}_g^u, \quad (18)$$

where $\mathbf{I}_g^u \in \mathbb{C}^{N_t N_v \times K N_a} = \mathbf{I}_{SI}^{g,u} \mathbf{I}_{AC}^{g,u}$ denotes the joint equivalent index pattern of the SF matrix. It contains the activation patterns $\mathbf{I}_{AC}^{g,u} \in \mathbb{C}^{N_t N_v \times N_t K}$ of TAs and subcarriers $\mathbf{I}_{SI}^{g,u} \in \mathbb{C}^{N_t K \times K N_a}$ arriving from the u -th user in the g -th subcarrier group. Furthermore, $\mathbf{X}_L^{g,u} \in \mathbb{C}^{K N_a \times 1}$ represents the APM symbol vector of the u -th user and $\hat{\mathbf{w}}_g^u \in \mathbb{C}^{N_t N_v \times 1}$ is the AWGN vector. Finally, we can demodulate $\hat{\mathbf{x}}_g^u$ without encountering MUI, which is suppressed in (17).

B. Maximum a Posteriori Detector and Factor Graph

Given the received signal $\mathbf{y}$ and fading matrix $\bar{\mathbf{H}}$ of (16), the *a posteriori* probability $p(\mathbf{x}|\mathbf{y})$ is used for demodulating $\mathbf{x}$ based on the received signal $\mathbf{y}$ as:

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x} \in \mathcal{S}^\Omega} p(\mathbf{x}|\mathbf{y}), \quad (19)$$

where $\Omega = N_t N_v U$ is the size of the transmit signal vector $\mathbf{x}$ of a specific subcarrier group. To simplify the calculation, we assume that $\mathbf{y}$ corresponds to a single subcarrier group. The i -th element x_i , $1 \leq i \leq \Omega$, of $\mathbf{x}$ can be estimated by the *a posteriori* Probability Mass Function (PMF). Given (19), we have

$$\hat{x}_i = \arg \max_{\mathbf{x} \in \mathcal{S}^\Omega} \sum p(\mathbf{x}|\mathbf{y}). \quad (20)$$

According to Bayes' theorem, we can have

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}), \quad (21)$$

where $p(\mathbf{x}) = \prod_{i=1}^\Omega p(x_i)$ is the joint *a priori* probability function of all the symbols in the SF matrix domain and

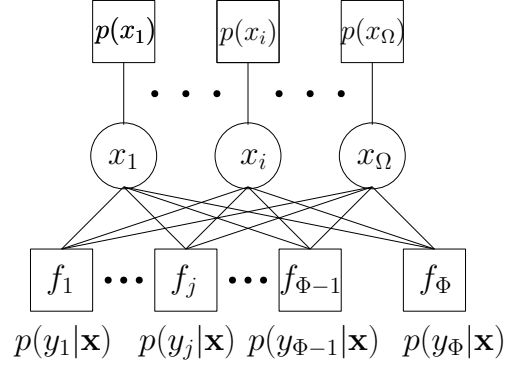


Fig. 3: Factor graph for CS-SFIM

$p(\mathbf{y}) = \sum_{\mathbf{x} \in \mathcal{S}^\Omega} p(\mathbf{y}|\mathbf{x})$ is the PDF of the received signal $\mathbf{y}$ represented as:

$$p(\mathbf{y}|\mathbf{x}) = \prod_{j=1}^\Phi p(y_j|\mathbf{x}), \quad (22)$$

where $\Phi = N_r N_f$ is the size of the received signal vector $\mathbf{y}$ and $p(y_j|\mathbf{x})$, $1 \leq j \leq \Phi$, is the conditional probability of the j -th observation y_j , given the transmitted symbol $\mathbf{x}$. Bearing in mind the sparsity of the index distribution in the SF domain matrix, each symbol is interfered by $(L - 1)$ symbols of L active index elements in the SF matrix. By combining (21) and (22) into (23), we obtain

$$\hat{x}_i = \arg \max_{x_i \in \mathcal{S}} \sum_{\mathbf{x} \in \mathcal{S}^{\Omega(\Omega-1), x_i}} p(\mathbf{x}) \prod_{j \in \mathcal{R}_i} p(y_j|\mathbf{x}), \quad (23)$$

where $\mathcal{R}_i$ represents the non-zero element set for $\bar{\mathbf{H}}$ and $p(y_j|\mathbf{x})$ is formulated as

$$p(y_j|\mathbf{x}) = \frac{1}{\pi \sigma^2} \exp\left(-\frac{|y_j - \bar{\mathbf{H}}\mathbf{x}|^2}{\sigma^2}\right). \quad (24)$$

The factor graph of the CS-SFIM system is illustrated as shown in Fig. 3, where the SF signal $\mathbf{x}$ is represented by a vector and the non-zero elements x_i are mapped to the i -th Variable Node (VN) associated with the circles. Then the received signal y_j is treated as the j -th Factor Node (FN) with solid represented by the squares in Fig. 3. Hence, we can describe the factorization of the joint distribution as:

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) = \prod_n^{N_r N_f} f_j(y_j|\mathbf{x}) \prod_i^\Omega p(\mathbf{x}_i), \quad (25)$$

where the likelihood function $f_j(y_j|\mathbf{x})$ associated with the observation vector $\mathbf{y}$ is formulated as:

$$f_j(y_j|\mathbf{x}) \propto p(y_j|\mathbf{x}) \propto \exp\left(-\frac{|y_j - \bar{\mathbf{H}}\mathbf{x}|^2}{\sigma^2}\right). \quad (26)$$

Based on the factor graph, for the i -th element, let $\mu_{x_i \rightarrow y_j}^t(\mathbf{x}_i)$ and $\mu_{y_j \rightarrow x_i}^t(\mathbf{x}_i)$ denote the message sent from VN x_i to FN y_j and in the opposite direction at the t -th iteration, respectively. Then, following the principle of the sum-product algorithm [53], the message passing rule can be formulated as:

$$\mu_{x_i \rightarrow f_j}^t(\mathbf{x}_i) = \phi_{\phi_i \rightarrow x_i}(\mathbf{x}_i) \prod_{j' \neq j} \mu_{f_{j'} \rightarrow x_i}^{t-1}(\mathbf{x}_i), \quad (27)$$

$$\mu_{f_j \rightarrow x_i}^t(\mathbf{x}_i) = \sum_{\mathbf{x} \setminus \mathbf{x}_i} f_j(y_j | \mathbf{x}) \prod_{i' \neq i} \mu_{x_{i'} \rightarrow f_j}^t(\mathbf{x}_i), \quad (28)$$

where $\mathbf{x} \setminus \mathbf{x}_i$ denotes all the enumerations of $\mathbf{x}$ except for x_i and $\phi_{\phi_i \rightarrow x_i}(\mathbf{x}_i) = p(\mathbf{x}_i)$ represents the *a priori* probability of the transmitted symbol vector.

1) *AMP-Aided Detection*: It has been shown in [28] that the AMP algorithm benefits from low complexity and rapid convergence. Inspired by the application of AMP in OFDM-IM [54] and GSM [55], we harness the AMP detector over the CS-SFIM scheme to reduce the detection complexity compared to the full search ML and to acquire an improved performance compared to that of the MMSE detector. The AMP-based detector aims for iteratively decoupling the *a posteriori* probability $p(\mathbf{x} | \mathbf{y})$ into a series $p(x_i | \mathbf{y})$.

To simplify the calculations, we approximate the distribution of information $\mu_{x_i \rightarrow y_j}^t(\mathbf{x}_i)$ by a Gaussian distribution and assume the PDF follows the $\mathcal{N}(\mathbf{x} : \mathbf{r}, \Sigma)$ [56], which can be expressed as:

$$\mu_{x_i \rightarrow f_j}^t(\mathbf{x}_i) \propto p(x_i) \mathcal{N}(x_i; r_{x_i \rightarrow f_j}^{t-1}, \Sigma_{x_i \rightarrow f_j}^{t-1}), \quad (29)$$

where $r_{x_i \rightarrow f_j}^{t-1}$ and $\Sigma_{x_i \rightarrow f_j}^{t-1}$ are the mean and variance of the Gaussian variable x_i .

Furthermore, the information flowing from FN to VN $\mu_{f_j \rightarrow x_i}^t(\mathbf{x}_i)$ can also be approximated by a Gaussian PDF as:

$$\mu_{f_j \rightarrow x_i}^t(\mathbf{x}_i) \propto p(x_i; \frac{y_j - \mathcal{Z}_{f_j \rightarrow x_i}^{t-1}}{h_{ji}}, \frac{\sigma^2 + \mathcal{V}_{f_j \rightarrow x_i}^{t-1}}{|h_{ji}|^2}), \quad (30)$$

where the mean is $\hat{x}_{f_j \rightarrow x_i}^{t-1} = \frac{y_j - \mathcal{Z}_{f_j \rightarrow x_i}^{t-1}}{h_{ji}}$ and the variance is $\hat{v}_{f_j \rightarrow x_i}^{t-1} = \frac{\sigma^2 + \mathcal{V}_{f_j \rightarrow x_i}^{t-1}}{|h_{ji}|^2}$. With the consideration of both forward and backward oriented information, the AMP algorithm can be further simplified as described in [57], which is presented in **Algorithm 1**.

Finally, (35) and (36) can be further expressed with the aid of the $\mathcal{L}$ -APM set $\mathcal{Q} = \{q_1, q_2, \dots, q_k, \dots, q_L\}$, as follows:

$$\hat{x}_i^{(t+1)} = \frac{\sum_{q_k \in \mathcal{Q}} q_k \mathcal{N}(q_k; r_n^{(t)}, \Sigma_n^{(t)}) p(q_k)}{\sum_{q_k \in \mathcal{Q}} \mathcal{N}(q_k; r_n^{(t)}, \Sigma_n^{(t)}) p(q_k)}, \quad (37)$$

$$\hat{v}_i^{(t+1)} = \frac{\sum_{q_k \in \mathcal{Q}} |q_k|^2 \mathcal{N}(q_k; r_n^{(t)}, \Sigma_n^{(t)}) p(q_k)}{\sum_{q_k \in \mathcal{Q}} \mathcal{N}(q_k; r_n^{(t)}, \Sigma_n^{(t)}) p(q_k)} - |\hat{x}_i^{(t+1)}|^2. \quad (38)$$

2) *EP-Aided Detection*: As a message passing-based algorithm, the EP [35] detector also has two modules: *observation module* and *estimation module*.

The posterior probability distribution of the transmitted signal conditioned on the received signal in (16), based on (21) can be reformulated as:

$$p(\mathbf{x} | \mathbf{y}) = \frac{p(\mathbf{y} | \mathbf{x}) p(\mathbf{x})}{p(\mathbf{y})} \propto \underbrace{\mathcal{N}(\mathbf{y} : \bar{\mathbf{H}} \mathbf{x}, \sigma^2 \mathbf{I})}_{p(\mathbf{y} | \mathbf{x})} \underbrace{\prod_{i=1}^{\Omega} p(x_i)}_{p(\mathbf{x})}, \quad (39)$$

and the *a priori* probability of the transmitted symbol vector of each subcarrier group $p(\mathbf{x})$ can be simplified using an unnormalized Gaussian distribution $t(x_i | V_i, \eta_i) = \exp(-\frac{1}{2} V_i x_i^2 +$

Algorithm 1: AMP Detector

Input: $\mathbf{y}, \bar{\mathbf{H}}, \sigma^2, p(x_i)$

Initialization: $\hat{x}_i^{t=1} = 0, \hat{v}_i^{t=1} = 1, \mathcal{Z}_j^{t=0} = y_j$

Iteration:

1 **for** $t = 1, 2, \dots, T$ **do**

2

$$\mathcal{V}_j^{(t)} = \sum |h_{ji}|^2 \hat{v}_i^t \quad (31)$$

3

$$\mathcal{Z}_j^{(t)} = \sum h_{ji} \hat{x}_i^t - \frac{\mathcal{V}_j^{(t)} (y_j - \mathcal{Z}_j^{(t-1)})}{\sigma^2 + \mathcal{V}_j^{(t-1)}} \quad (32)$$

4

$$\hat{\Sigma}_i^{(t)} = \left(\sum \frac{|h_{ji}|^2}{\sigma^2 + \mathcal{V}_j^{(t-1)}} \right)^{-1} \quad (33)$$

5

$$\hat{r}_i^{(t)} = \hat{x}_i^{(t)} + \hat{\Sigma}_i^{(t)} \sum \frac{h_{ji}^* y_j - \mathcal{Z}_j^{(t)}}{\sigma^2 + \mathcal{V}_j^{(t)}} \quad (34)$$

6

$$\hat{x}_i^{(t+1)} = \mathbb{E}(x_i; \hat{r}_i^{(t)}, \hat{\Sigma}_i^{(t)}) \quad (35)$$

7

$$\hat{v}_i^{(t+1)} = \text{Var}(x_i; \hat{r}_i^{(t)}, \hat{\Sigma}_i^{(t)}) \quad (36)$$

8 **end**

Output: $\hat{\mathbf{x}}^{(T)}$

$\eta_i x_i$) to approximate the *a priori* distribution $p(\mathbf{x})$, where V_i and η_i represent the parameters of the Gaussian PDF. Then, (39) can be rewritten as:

$$\begin{aligned} \hat{p}(\mathbf{x} | \mathbf{y}) &\propto p(\mathbf{y} | \mathbf{x}) \prod_{i=1}^{\Omega} t(x_i) \\ &\propto \mathcal{N}(\mathbf{y} : \bar{\mathbf{H}} \mathbf{x}, \sigma^2 \mathbf{I}) \times \prod_{i=1}^{\Omega} \exp\left(-\frac{1}{2} V_i x_i^2 + \eta_i x_i\right) \quad (40) \\ &\propto \mathcal{N}(\mathbf{y} : \bar{\mathbf{H}} \mathbf{x}, \sigma^2 \mathbf{I}) \times \exp\left(-\frac{1}{2} \mathbf{x}^T \mathbf{V} \mathbf{x} + \boldsymbol{\eta}^T \mathbf{x}\right), \end{aligned}$$

where we have $\boldsymbol{\eta} = [\eta_1, \eta_2, \dots, \eta_{N_t N_f U}]^T$ and $\mathbf{V} = \text{diag}\{V_1, V_2, \dots, V_{N_t N_f U}\}$. Furthermore, by multiplying (16) with the pseudo-inverse matrix of $\bar{\mathbf{H}}$ as $\bar{\mathbf{H}}^\dagger = (\bar{\mathbf{H}}^T \bar{\mathbf{H}})^{-1} \bar{\mathbf{H}}^T$, we arrive at $p(\mathbf{y} | \mathbf{x}) \propto \mathcal{N}(\mathbf{x} : \bar{\mathbf{H}}^\dagger \mathbf{y}, \sigma^2 (\bar{\mathbf{H}}^T \bar{\mathbf{H}})^{-1})$ for a given vector $\mathbf{y}$ [35]. To iteratively approximate $p(\mathbf{y} | \mathbf{x})$, we reformulate (40) as:

$$\begin{aligned} \hat{p}(\mathbf{x} | \mathbf{y}) &\propto \mathcal{N}(\mathbf{x} : \bar{\mathbf{H}}^\dagger \mathbf{y}, \sigma^2 (\bar{\mathbf{H}}^T \bar{\mathbf{H}})^{-1}) \times \mathcal{N}(\mathbf{x} : \boldsymbol{\eta}_i V_i^{-1}, V_i^{-1}) \\ &\propto \mathcal{N}(\mathbf{x} : \boldsymbol{\mu}_{ep}, \boldsymbol{\Sigma}_{ep}), \end{aligned} \quad (41)$$

where $\boldsymbol{\mu}_{ep}$ and $\boldsymbol{\Sigma}_{ep}$ denote the mean vector and covariance matrix formulated as

$$\boldsymbol{\Sigma}_{ep} = \left(\sigma^{-2} \bar{\mathbf{H}}^T \bar{\mathbf{H}} + \mathbf{V} \right)^{-1}, \quad (42)$$

and

$$\boldsymbol{\mu}_{ep} = \boldsymbol{\Sigma}_{ep} (\sigma^{-2} \bar{\mathbf{H}}^T \mathbf{y} + \boldsymbol{\eta}). \quad (43)$$

To reduce the computational complexity, we can have a tractable distribution $q(\mathbf{x})$ to approximate the intractable distribution $p(\mathbf{x}|\mathbf{y})$ by minimizing the Kullback-Leibler (KL) divergence, which is equivalent to matching the moments of mean and variance for these two probability distributions. EP then approximates $p(\mathbf{x}|\mathbf{y})$ by a marginal distribution $q(\mathbf{x})$ and iteratively updates the mean and variance as:

$$q^{(t)}(\mathbf{x}) \propto \prod_{i=1}^{\Omega} \underbrace{\mathcal{N}(x_i : \mu_i^{(t)}, \Sigma_{ii}^{(t)})}_{q^{(t)}(x_i)}, \quad (44)$$

where t denotes the iteration index. Afterwards, we calculate the cavity distribution of x_i [35] based on $q^{(t)}(x_i)$ as:

$$\begin{aligned} q^{(t)\setminus i}(x_i) &= \frac{q^{(t)}(x_i)}{t^{(t)}(x_i)} \\ &\propto \frac{\mathcal{N}(x_i : \mu_i^{(t)}, \Sigma_{ii}^{(t)})}{\mathcal{N}(x_i : \eta_i^{(t)} V_i^{-1(t)}, V_i^{-1(t)})} \\ &\propto \mathcal{N}(x_i : m_{o,i}^{(t)}, v_{o,i}^{(t)}), \end{aligned} \quad (45)$$

where we have

$$m_{o,i}^{(t)} = \frac{\Sigma_{ii}^{(t)}}{1 - \Sigma_{ii}^{(t)} V_i^{(t)}}, \quad (46)$$

and

$$v_{o,i}^{(t)} = m_{o,i}^{(t)} \left(\frac{\mu_i^{(t)}}{\Sigma_{ii}^{(t)}} - \eta_i^{(t)} \right), \quad (47)$$

where $\mathbf{m}_o^{(t)} = [m_{o,1}^{(t)}, m_{o,2}^{(t)}, \dots, m_{o,N_t N_f U}^{(t)}]^T$ and the covariance vector $\mathbf{V}_o^{(t)} = [v_{o,1}^{(t)}, v_{o,2}^{(t)}, \dots, v_{o,N_t N_f U}^{(t)}]^T$ are forwarded to the *estimation module*. We can then compute the mean and variance of the approximated PDF $q(x_i) = q^{(t)\setminus i}(x_i)p(x_i)$, where $p(x_i)$ denotes the discrete uniform distribution. Based on the moment matching method [58], we can calculate the mean and variance of (45) as:

$$\hat{m}_i^{(t)} = \sum_{a \in \Omega} a \times q^{(t)\setminus i}(x_i = a), \quad (48)$$

$$v_i^{(t)} = \sum_{a \in \Omega} (x_i - \hat{x}^{(t)})^2 \times q^{(t)\setminus i}(x_i = a). \quad (49)$$

The EP detector iteratively estimates the transmitted symbols and carries out hard decision at iteration T . The hard decision of $\hat{\mathbf{x}}_{(T)}$ is based on comparing the Euclidean distances from the elements of the constellation set $\mathcal{Q}$, which is the mean of the cavity distribution $q^{(t)\setminus i}(x_i)$, as shown in (46). Furthermore, in the case of $t < T$, the parameter pair $(\eta_i^{(t+1)}, V_i^{(t+1)})$ is updated by matching the mean and variance of

$$V_i^{(t+1)} = \frac{1}{v_i^{(t)}} - \frac{1}{v_{o,i}^{(t)}}, \quad (50)$$

and

$$\eta_i^{(t+1)} = \frac{\hat{x}_i^{(t)}}{v_i^{(t)}} - \frac{m_{o,i}^{(t)}}{v_{o,i}^{(t)}}. \quad (51)$$

The parameter update in (49) may result in a negative $V_i^{(t+1)}$, which causing no matching pair of $(\eta_i^{(t+1)}, V_i^{(t+1)})$. In this

Algorithm 2: Expectation Propagation Detector

Input: $\mathbf{y}, \bar{\mathbf{H}}, \sigma^2, \Omega, \Phi$, transmit power E_s

Initialization: $\hat{\mathbf{x}}^{t=1} = 0, \boldsymbol{\eta}^{t=1} = 0, \mathbf{V}^{t=1} = \frac{1}{E_s} \mathbf{I}, \epsilon = 0.05$

```

1 for  $t = 1, 2, \dots, T$  do
2   for  $i = 1, 2, \dots, \Omega$  do
3     Compute  $\Sigma$  and  $\mu$  based on (42) and (43);
4     Compute the mean and variance of the cavity
       distribution  $q^{(t)\setminus i}(x_i)$  based on (47) and (46);
5     Compute the mean and variance of the cavity
       distribution  $q^{(t)}(x_i)$  based on (49) and (48);
6     Match the moment based on (52) and (53)
7   end
8   ;
9   Update the approximate mean  $\mathbf{m}^{(t)}$  and variance
      $\mathbf{v}^{(t)}$ ;
10   $\hat{\mathbf{x}}^{(t)} = \hat{\mathbf{m}}^{(t)}$ 
11 end
Output:  $\hat{\mathbf{x}}^{(T)}$ 

```

case, when $V_i^{(t+1)} < 0$, we have $V_i^{(t+1)} = V_i^{(t)}$ and $\eta_i^{(t+1)} = \eta_i^{(t)}$. Finally, we can harness the following damping actions

$$V_i^{(t+1)} = (1 - \epsilon)V_i^{(t+1)} + \epsilon V_i^{(t)}, \quad (52)$$

and

$$\eta_i^{(t+1)} = (1 - \epsilon)\eta_i^{(t+1)} + \epsilon \eta_i^{(t)}. \quad (53)$$

where $\epsilon \in [0, 1]$ is a weighting coefficient. Finally, the *estimation module* sends the parameter pair to the observation module for the next iteration. The operation of the EP detectors is summarized in **Algorithm 2**.

C. GNN-based Detector

1) *GNN structure*: GNNs offer a transformative approach to MU detection by addressing the inherent limitations of traditional detectors, such as the MMSE, AMP, and EP schemes [39] [59]. Although conventional detection methods rely on predefined algorithms and assumptions, they struggle to handle high-dimensional data and hence suffer from MUI. By contrast, GNNs exploit the structured nature of MU systems in a graph framework, dynamically learn and adapt, while improving both the detection accuracy and computational efficiency. In the next section, we explore the application of GNN-based detectors to our CS-SFIM system, including the GNN-MIMO, GNN-AMP, and GEPnet schemes. Each method demonstrates how GNNs can enhance the performance by integrating graph-based processing techniques with traditional detection frameworks.

Since the posterior distribution approximation accuracy critically depends on the scale of the factor graph in the MP-based detector, for a limited number of nodes, we need advanced damping methods for estimating the information of each node. As shown in Fig. 3, we consider the TAs and RAs as the nodes and the CSI as the edges in the construction of the GNNs. Then, we can deploy the GNN framework for capturing the structured dependency of the transmitted signal by harnessing

learning method into the MP on the pair-wise MRF model as shown in Fig. 4 and Fig. 5.

- **Node:** Each node $i \in \Omega$ represents the i -th element in the SF matrix of the equivalent symbol $\mathbf{x}$.
- **Node attributes:** Each node is assigned its node attribute $\mathbf{a}_i$ that is constant, when exchanging the information between nodes.
- **Edge:** An edge $e_{i,j} \in \mathcal{E}$ connects node $i \in \Omega$ and $j \in \Omega$.
- **Edge attributes:** Each edge $e_{i,j}$ has an assigned edge attribute $\mathbf{f}_{ij}$ that is constant, when computing the message, where we use the CSI and noise as edge attribute.
- **Hidden vector:** Each node $\mathbf{U}_i$ has a hidden vector updated in each round of the GNN and used for computing the output of the GNN.
- **Message:** the messages $\mathbf{m}_{ji}$ arriving from the connected edges are used to update the node feature vector.

The i -th variable node is characterized by a self potential $\phi(x_i)$, and the (i,j) -th pair of an edge is characterized by a pair potential $\psi(x_i, x_j)$, which are given by

$$\phi(x_i) = \exp\left(\frac{1}{\sigma^2} \mathbf{y}^T \mathbf{h}_i x_i - \frac{1}{2} \mathbf{h}_i^T \mathbf{h}_i x_i^2\right) p(x_i), \quad (54)$$

and

$$\psi(x_i, x_j) = \exp\left(-\frac{1}{\sigma^2} \mathbf{h}_i^T \mathbf{h}_j x_i x_j\right), \quad (55)$$

where $i, j \in 1, \dots, \Omega$ and $i \neq j$. Finally, the joint probability $p_{out}(\mathbf{x})$ corresponding to the pair-wise MRF can be obtained as [39]

$$p_{out}(\mathbf{x}) = \frac{1}{Z} \prod_{i=1}^{\Omega} \phi(x_i) \prod_{j=1, j \neq i}^{\Omega} \psi(x_i, x_j), \quad (56)$$

where Z is a normalization constant.

GNN can be used to infer the *a posteriori* probability $p(\mathbf{x}|\mathbf{y})$ and to recover the equivalent transmitted symbols $\mathbf{x}$ in the CS-SFIM detection problem of (16). In this case, GNNs are constructed for the MIMO pair-wise MRF presented in (54) and (55). The MRF models the structured dependency of a set of random variables $\mathbf{x} = x_1, \dots, x_i, \dots, x_{\Omega}$ by an undirected graph $G = \{\Omega, \mathcal{E}\}$, where Ω and $\mathcal{E}$ are the set of nodes and edges, respectively.

Then, the input of the GNN is extracted from $\phi(x_i)$ and $\psi(x_i, x_j)$ and the information along each edge $e_{i,j}$ represents the feature vector $\mathbf{f}_{ij} = [\mathbf{h}_i^T \mathbf{h}_j, \sigma^2]$, which can be obtained from (56) and utilized for the message passing of the GNN. The hidden vector $\mathbf{u}_i^0$ of each node i is initialized with

$$\mathbf{u}_i^0 = \mathbf{W}_1 [\mathbf{y}^T \mathbf{h}_i, \mathbf{h}_i^T \mathbf{h}_i, \sigma^2]^T + \mathbf{b}_1, \quad (57)$$

where we have $\mathbf{W} \in \mathbb{R}_{N_u} \times \mathbb{R}^3$ and $\mathbf{b} \in \mathbb{R}_{N_u}$ is a learnable vector, while N_u is the size of the feature vector. A typical GNN framework consists of three main modules: propagation module, aggregation module and readout module. The first two modules operate at every iteration l , while the readout module is activated only at final iteration L to make the inference. In the following we elaborate on the GNN processing.

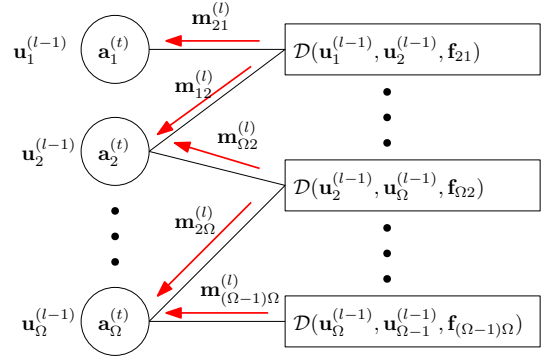


Fig. 4: Propagation process of the GNN

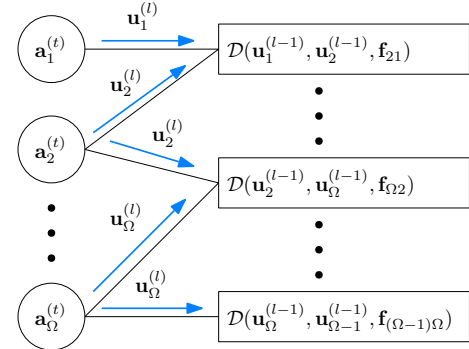


Fig. 5: Aggregation process of the GNN

a) *Propagation module:* Fig. 4 shows a typical GNN structure and its propagation process. The messages $u_k^{(l)}, k \in [\Omega], l \in [L]$ are sent to the neighboring factor nodes and the propagation module will gather the messages from pairs of variable nodes, and feed back the update message for each direct edge. For example, the $m_{21}^{(l)}$ is the update message of edge e_{21} in the l iteration and fed back to node 1. The propagation module outputs the updated message $\mathbf{m}_{ji}^{(i)}$ for each direct edge $e_{ij} \in \mathcal{E}$ as:

$$\mathbf{m}_{ji}^{(l)} = \mathcal{D}(\mathbf{u}_i^{(l-1)}, \mathbf{u}_j^{(l-1)}, \mathbf{f}_{ji}), \quad (58)$$

where $\mathbf{f}_{ji}$ is the information associated with the edge e_{ij} and $\mathcal{D}$ is a Multiple Layer Perceptron (MLP) using the classic Rectified Linear Unit (ReLU) as activation function. In the propagation module, each edge has an MLP associated with two hidden layers of size N_{h_1} and N_{h_2} and an output layer of size N_u . Finally, the outputs $\mathbf{m}_{ji}^{(l)}$ are fed back to the nodes.

b) *Aggregation module:* The aggregation module of each variable node adds all the incoming messages $\mathbf{m}_{jk}^{(l)}, j, k \in [\Omega]$ gleaned from its connected edges and concatenates the sum of $\mathbf{m}_{ji}^{(l)}$ with the node attribute $\mathbf{a}_i^{(t)}$ to obtain the gathering message $\mathbf{m}_i^{(l)}$ for the current iteration l , which is used to update the node message $\mathbf{u}_i^{(l)}$ to the propagation module for the next iteration, as shown in Fig. 5:

$$\mathbf{m}_i^{(l)} = \left[\sum_{j=1, j \neq i}^{\Omega} \mathbf{m}_{ji}^{(l)}, \mathbf{a}_i^{(t)} \right]. \quad (59)$$

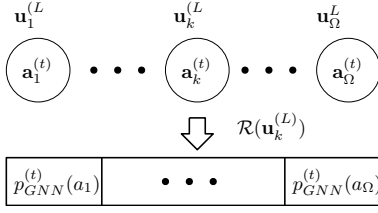


Fig. 6: Readout process of the GNN

The message $\mathbf{m}_i^{(l)}$ is used for computing the nodes' hidden vector $\mathbf{u}_i^{(l)}$ as follow:

$$\mathbf{g}_i^{(l)} = U(\mathbf{g}_i^{(l-1)}, \mathbf{m}_i^{(l)}), \quad (60)$$

$$\mathbf{u}_i^{(l)} = \mathbf{W}_2 \mathbf{g}_i^{(l)} + \mathbf{b}_2, \quad (61)$$

where the function U is constituted by a Gated Recurrent Unit (GRU) [60] based network associated with the current and the previous states. The previous hidden states specified as $\mathbf{g}_i^{(l)} \in \mathbb{R}^{N_{h_1}}$ and $\mathbf{g}_i^{(l-1)} \in \mathbb{R}^{N_{h_1}}$, respectively. Furthermore, $\mathbf{W}_2 \in \mathbb{R}^{N_u \times N_{h_1}}$ is a learnable matrix and $\mathbf{b}_2 \in \mathbb{R}^{N_u}$ is a learnable vector. The updated hidden vector in (61) is passed to the propagation module for the next iteration, as shown in Fig. 5.

c) *Readout module*: In Fig. 6, after L iterations of the MP between the propagation and aggregation modules, a readout module is harnessed for generating the estimated distribution $p_{GNN}^{(t)}(x_i = s_i | \mathbf{y})$ for the t -layer GNN given by

$$\tilde{p}_{GNN}^{(t)}(x_i = a | \mathbf{y}) = R(\mathbf{u}_i^{(t)}), a \in \Omega, \quad (62)$$

and

$$p_{GNN}^{(t)}(x_i = a | \mathbf{y}) = \frac{\exp(\tilde{p}_{GNN}^{(t)}(x_i = a | \mathbf{y}))}{\sum_{a \in \Omega} \exp(\tilde{p}_{GNN}^{(t)}(x_i = a | \mathbf{y}))}, a \in \Omega. \quad (63)$$

The readout function R consists of an MLP having two hidden layers of sizes N_{h_1} and N_{h_2} along with the classic RELU activation after each hidden layer. Hence, we can compute

$$\mathbf{g}_i^{(0)} \leftarrow \mathbf{g}_i^{(L)} \text{ and } \mathbf{u}_i^{(0)} \leftarrow \mathbf{u}_i^{(L)}, i = 1, 2, \dots, \Omega. \quad (64)$$

for the next iteration.

2) *GNN-MMSE*: In this section, to improve the prior information, we incorporate the MMSE *a posteriori* as the *a priori* $p(x)$ derived from (16) such that

$$p(x_i) = \frac{1}{\pi h_{c_{ii}}} \exp\left[-\frac{(z_i - x_i)^2}{h_{c_{ii}}}\right], \quad (65)$$

where z_i is the i -th element of the MMSE estimation vector $\mathbf{z} = (\bar{\mathbf{H}}^T \bar{\mathbf{H}} + \sigma^2 \mathbf{I}_\Omega)^{-1} \bar{\mathbf{H}}^T \mathbf{y}$ and $\sigma_{h_{c_{ii}}}$ is the (i, i) -th element of $\mathbf{C} = \sigma^2 (\bar{\mathbf{H}}^T \bar{\mathbf{H}} + \sigma^2 \mathbf{I}_\Omega)^{-1}$. The prior correlation coefficient ρ_{ij} between the variable x_i and x_j , namely $\rho_{ij} = \frac{h_{c_{ij}}^2}{h_{c_{ii}} h_{c_{jj}}}$ is added to the feature vector $\mathbf{f}_{ij}$. Afterwards, the same model of Section III-C1 is applied subject to the slight modification of the information $\mathbf{f}_{ij}$ between the edges and the initial value of the hidden vector $\mathbf{u}_i^{(0)}$. Then, the information along the edge e_{ij} is defined as $e_{ij} = [\rho_{ij}, \mathbf{h}_i^T \mathbf{h}_j, \sigma^2]$ and the initial hidden vector $\mathbf{u}_i^{(0)}$ of each node i is initialized with $\mathbf{u}_i^{(0)} = \mathbf{W}_1 [z_i, h_{c_{ii}}, \mathbf{y}^T \mathbf{h}_i, \mathbf{h}_i^T \mathbf{h}_i, \sigma^2]^T + \mathbf{b}_1$.

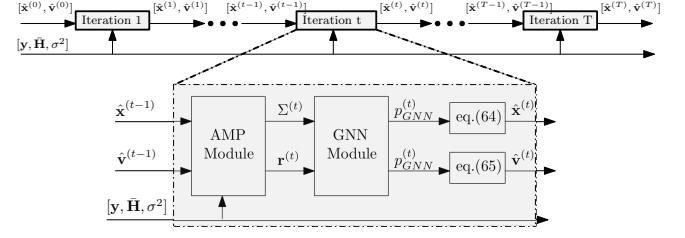


Fig. 7: The structure of the AMP-GNN network.

3) *AMP-GNN*: The GNN can also be harnessed for the MP-based method to reduce the inaccuracy of approximation. In this case, the conventional AMP approximates the distribution of information by the Gaussian distribution, but the GNN can be invoked for achieving a more accurate distribution of the corresponding information. As shown in Fig. 7, in the t -th iteration, the AMP-GNN takes its input from the linear module in the AMP algorithm and then the mean $r_i^{(t)}$ and variance $\Sigma_i^{(t)}$ from (33) and (34) are fed into the twin-tuple attribute $\mathbf{a}_n^t$ of the variable node x_i as follows:

$$\mathbf{a}_n^t = [r_i^{(t)}, \Sigma_i^{(t)}]. \quad (66)$$

After a few iterations, we can invoke $p_{GNN}^{(t)}(x_i = s_i | \mathbf{y})$ of Fig. 7 to compute the *a posteriori* mean and variance for the AMP, formulated as:

$$\hat{x}_i^{(t+1)} = \mathbb{E}(x_i; p_{GNN}^{(t)}), \quad (67)$$

and

$$\hat{v}_i^{(t+1)} = \text{Var}(x_i; p_{GNN}^{(t)}), \quad (68)$$

where the expectation and variance of the estimated signal x_i are computed by taking into account $p_{GNN}^{(t)}$. The most significant difference between conventional AMP and AMP-GNN is that $p_{GNN}^{(t)}$ is not approximated by the Gaussian distribution for the AMP-GNN detector. Then, $\hat{x}_i^{(t+1)}$ and $\hat{v}_i^{(t+1)}$ are forwarded to the next AMP-GNN iteration, where the iterations terminate at a fixed number of layers. Accordingly, we can modify **Algorithm 1** leading to the AMP-GNN procedure of **Algorithm 3**.

4) *GEPnet*: To mitigate the inaccuracy of the posterior distribution approximation in the EP detector, we harness a feature proposed in [42], which treats the EP's cavity $q^{(t)\setminus i}(x_i)$ in (45) as the prior information for the variable node i in our GEPnet detector. The cavity $q^{(t)\setminus i}(x_i)$ is given by a Gaussian distribution with a mean $m_{o,i}^{(t)}$ and variance $v_{o,i}^{(t)}$ defined in (46) and (47). Then, the variable node attribute can be represented as:

$$\mathbf{a}_n^t = [m_{o,i}^{(t)}, v_{o,i}^{(t)}]. \quad (69)$$

Hence, we can modify the GEPnet readout module of (63), where the appropriately adjusted cavity distribution $q_{GNN}^{(t)\setminus i}(x_i = a)$ can be represented as

$$q_{GNN}^{(t)\setminus i}(x_i = a) = \frac{\exp[\tilde{p}_{GNN}^{(t)}(x_i = a)]}{\sum_{a \in \Omega} \exp[\tilde{p}_{GNN}^{(t)}(x_i = a)]}, a \in \Omega. \quad (70)$$

Algorithm 3: AMP-GNN Detector

Input: $\mathbf{y}, \bar{\mathbf{H}}, \sigma^2, p(x_i)$
Initialization: $\hat{x}_i^{t=1} = 0, \hat{v}_i^{t=1} = 1, Z_j^{t=0} = y_j, \mathbf{g}^{(0)} = \mathbf{0}$

```

1 for  $t = 1, 2, \dots, T$  do
2   Compute  $\Sigma_i^{(t)}$  and  $r_i^{(t)}$  in (33) and (34),  $i \in [\Omega]$ 
3   The GNN Module
4   Compute (66)
5   if  $t = 1$  then
6     Compute  $\mathbf{u}_i^{(0)}$  in (57),  $i \in [\Omega]$ 
7   end
8   for  $l = 1, 2, \dots, L$  do
9     Compute  $\mathbf{m}_{ji}^{(l)}$  and  $\mathbf{m}_i^{(l)}$  in (58) and (59),
         $i, j \in [\Omega], i \neq j$ 
10    Compute  $\mathbf{g}_i^{(l)}$  and  $\mathbf{u}_i^{(l)}$  in (60) and (61),  $i \in [\Omega]$ 
11  end
12  Compute  $p_{GNN}^{(l)}(x_i)$  in (62) and (63)
13  Compute (64)
14  Compute  $\hat{x}_i^{(t+1)}$  and  $\hat{v}_i^{(t+1)}$  based on (67) and (68),
         $i \in [\Omega]$ 
15 end
Output:  $\hat{\mathbf{x}}^{(T)}$ 

```

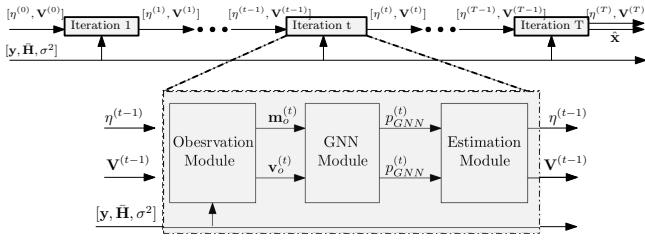


Fig. 8: The structure of the GEPnet.

Then, we can use the GNN modules, trained distribution for replacing the distribution in the *estimation module* of the EP as

$$\hat{m}_i^{(t)} = \sum_{a \in \Omega} a \times q_{GNN}^{(t)\setminus i}(x_i = a), \quad (71)$$

$$v_i^{(t)} = \sum_{a \in \Omega} (x_i - \hat{x}^{(t)})^2 \times q_{GNN}^{(t)\setminus i}(x_i = a). \quad (72)$$

The GEPnet structure is shown in Fig. 8 and the GEPnet procedure is presented in **Algorithm 4**.

D. Complexity analysis

In this section, we analyze the computational complexity of the GNN-based detector of our CS-SFIM system, then compare it to conventional MP-based MU detectors and classic DNN-based detector, which use results of ML detection as training target. The complexity of the AMP detector is on the order of $\mathcal{O}(\Omega\Phi)$ due to the associated matrix-vector multiplication and the complexity of the EP is $\mathcal{O}(\Omega^2\Phi)$. Additionally, the complexity of the DNN [61], GNN-MMSE, AMP-GNN and GEPnet is $\mathcal{O}(N_{h_1}N_{h_2})$, $\mathcal{O}(\Omega^2 + \Omega^2F^2)$, $\mathcal{O}(\Omega\Phi + \Omega^2F^2)$ and

Algorithm 4: GEPnet Detector

Input: $\mathbf{y}, \bar{\mathbf{H}}, \sigma^2, \Omega, \Phi$, transmit power E_s
Initialization: $\hat{\mathbf{x}}^{t=1} = \mathbf{0}, \boldsymbol{\eta}^{t=1} = \mathbf{0}, \mathbf{V}^{t=1} = \frac{1}{E_s}\mathbf{I}, \epsilon = 0.1, \mathbf{g}^{(0)} = \mathbf{0}$

Iteration:

```

1 for  $t = 1, 2, \dots, T$  do
2   EP Observation Module
3   Compute  $\Sigma$  and  $\boldsymbol{\mu}$  based on (42) and (43)
4   Compute  $m_{o,i}^{(t)}$  and  $v_{o,i}^{(t)}$  based on (46) and (47)
5   Compute (69)
6   The GNN Module
7   if  $t = 1$  then
8     Compute  $\mathbf{u}_i^{(0)}$  in (57),  $i \in [\Omega]$ 
9   end
10  for  $l = 1, 2, \dots, L$  do
11    Compute  $\mathbf{m}_{ji}^{(l)}$  and  $\mathbf{m}_i^{(l)}$  in (58) and (59),
         $i, j \in [\Omega], i \neq j$ 
12    Compute  $\mathbf{g}_i^{(l)}$  and  $\mathbf{u}_i^{(l)}$  in (60) and (61),  $i \in [\Omega]$ 
13  end
14  Compute  $q_{GNN}^{(l)}(x_i)$  in (70)
15  Compute (64)
16  EP Estimation Module
17  Compute  $\hat{m}_i^{(t)}$  and  $v_i^{(t)}$  based on (48) and (49),
         $i \in [\Omega]$ 
18  Compute  $\eta_i^{(t+1)}$  and  $V_i^{(t+1)}$  based on (50) and
        (51),  $i \in [\Omega]$ 
19  Optimize  $\eta_i^{(t+1)}$  and  $V_i^{(t+1)}$  based on (52) and (53)
20 end
21  $\hat{\mathbf{x}}^{(t)} = \hat{\mathbf{m}}^{(t)}$ ;
Output:  $\hat{\mathbf{x}}^{(T)}$ 

```

$\mathcal{O}(\Omega^2\Phi + \Omega^2F^2)$.⁶ Then we can characterize the complexity of each detector in different scenarios in Table IV and computational complexity based on the number of multiplications as our metric quantified in Table III.

IV. PERFORMANCE ANALYSIS

In this section, we first characterize the MU CS-SFIM system proposed in Section II. We then compare the performance of the GNN-based detectors to that of conventional detectors. More specifically, the BER is evaluated both as a function of the number of users and of the transmit-to-receive antenna ratio. We consider $U = 4$ and 16 users. For $U = 4$ users, we employ 4 TAs per user and 16 RAs. Hence, the transmit-to-receive antenna ratio is $\varrho \equiv \frac{N_t \cdot U}{N_r} = 1$. By contrast, we set $\varrho \in [\frac{1}{2}, 1]$ for the $U = 16$ users' scheme. We investigated a set of schemes, which are summarized as follows:

- 1) **Scheme 1:** MU-CS-SFIM, which supports $U = 4$ users, activating $N_a = 2$ out of 4 TAs, 16 RAs, APM constellation $\mathcal{L} = 2$ and $N_f = 2$ subcarriers per group, while considering $N_v = 4$ subcarriers per group in the

⁶ h_1 and h_2 represent the input and output of the DNN. F represents the features of each node in GNN network.

TABLE III: Multiplications number of different detectors for MU-CS-SFIM.

MU setting			MMSE	AMP	EP	DNN	GNN-MMSE	AMP-GNN	GEPnet
User	TA	RA							
4	4	16	4.31×10^6	1.12×10^6	2.61×10^7	2.43×10^6	2.13×10^7	2.75×10^7	6.72×10^7
16	4	64	1.42×10^7	4.35×10^6	5.23×10^8	9.83×10^6	9.22×10^7	1.33×10^8	8.84×10^8
16	4	128	1.42×10^7	8.71×10^6	1.12×10^9	1.05×10^8	1.15×10^8	1.54×10^8	1.25×10^9

VD and $K = 2$ subcarriers activated. Then we apply the following detectors:

- MMSE based detector.
- AMP detector.
- EP detector.
- DNN detector.
- GNN-MMSE detector
- AMP-GNN detector.
- GEPnet detector.
- ML detector.

- Scheme 2:** MU-CS-SFIM, which supports $U = 4$ users, activating $N_a = 2$ out of 4 TAs, 16 RAs, APM constellation $\mathcal{L} = 4$ and $N_f = 2$ subcarriers per group, while considering $N_v = 4$ subcarriers per group in the VD and $K = 2$ subcarriers activated. Then we apply the following detectors:

- MMSE based detector.
- AMP detector.
- EP detector.
- DNN detector.
- GNN-MMSE detector
- AMP-GNN detector.
- GEPnet detector.
- ML detector.

- Scheme 3:** MU-CS-SFIM, which supports $U = 16$ users, activated $N_a = 2$ out of 4 TAs, 64 RAs, APM constellation $\mathcal{L} = 2$ and $N_f = 2$ subcarriers per group, while considering $N_v = 4$ subcarriers per group in the VD and $K = 2$ subcarriers activated. Then we apply the following detectors:

- MMSE based detector.
- AMP detector.
- EP detector.
- GNN-MMSE detector
- AMP-GNN detector.
- GEPnet detector.

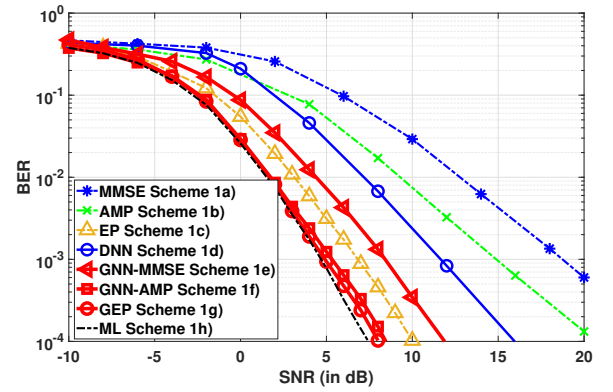
- Scheme 4:** MU-CS-SFIM, which supports $U = 16$ users, activated $N_a = 2$ out of 4 TAs, 128 RAs, APM constellation $\mathcal{L} = 2$ and $N_f = 2$ subcarriers per group, while considering $N_v = 4$ subcarriers per group in the VD and $K = 2$ subcarriers activated. Then we apply the following detectors:

- MMSE based detector.
- AMP detector.
- EP detector.
- GNN-MMSE detector
- AMP-GNN detector.
- GEPnet detector.

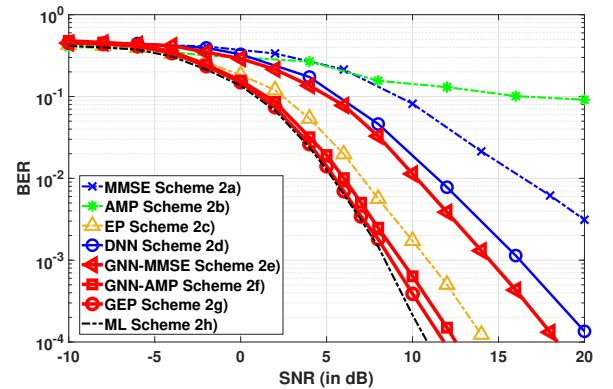
- Scheme 5:** MU-CS-SFIM, which supports $U = 4$ users, where $K_a = 3$ only 3 out of 4 users randomly selected

are activated each time, $N_a = 2$ out of 4 TAs, 64 RAs, APM constellation $\mathcal{L} = 2$ and $N_f = 2$ subcarriers per group, while considering $N_v = 4$ subcarriers per group in the VD and $K = 2$ subcarriers activated. Then we apply the following detectors:

- MMSE based detector.
- AMP detector.
- EP detector.
- DNN detector.
- GNN-MMSE detector
- AMP-GNN detector.
- GEPnet detector.
- ML detector.



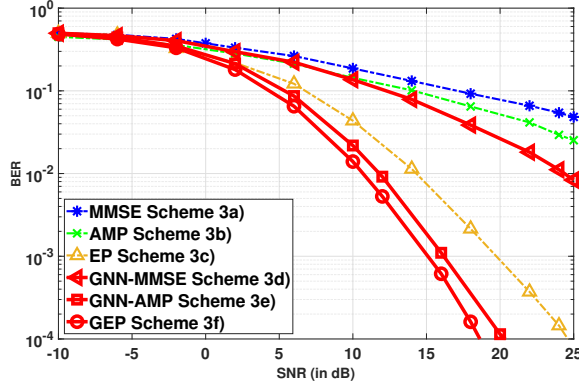
(a) Performance comparison of detectors in Scheme 1.



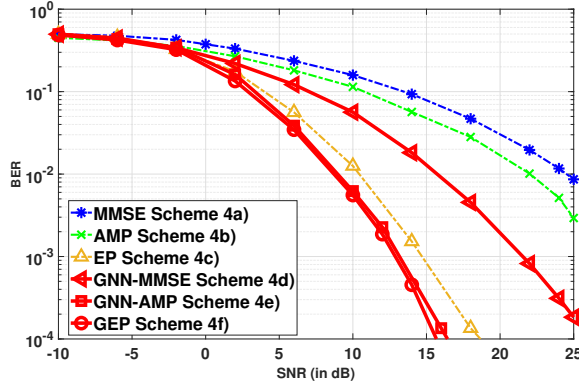
(b) Performance comparison of detectors in Scheme 2.

Fig. 9: BER performance comparison of the different MU CS-SFIM schemes under different QAM schemes for transmission over Rayleigh channel.

Graph-based modeling naturally matches the factor-graph interpretation of the joint detection problem in CS and IM, enabling direct incorporation of the sparse prior and Markov



(a) Performance comparison of detectors in Scheme 3.



(b) Performance comparison of detectors in Scheme 4.

Fig. 10: BER performance comparison of the different MU CS-SFIM schemes under different U for transmission over Rayleigh channel.

structure [39]. DNNs do not inherently exploit the sparse factor-graph structure, and might need more parameters to learn an equivalent message-passing approach.

In Fig. 9(a), we compare the performance of various detectors for our MU CS-SFIM system that supports $U = 4$ users in **scheme 1**. As discussed in Section III, for every subcarrier group of each user, we have N_{SI} possible active subcarrier index combinations, N_{AC}^K possible active TA index realizations, and $L^{K \cdot N_a}$ possible classical APM symbols combinations. The complexity order of **Scheme 1a**) is $\mathcal{O}_{MMSE}[U \cdot (N_{FI} N_{AC}^K L^{K N_a})]$ and of **Scheme 1g**) is $\mathcal{O}_{ML}[(N_{FI} N_{AC}^K L^{K N_a})^U]$, which is feasible for a small number of users. **Scheme 1g**) achieves a BER of 10^{-4} at an SNR of 7.12 dB, as shown in Fig. 9(a). **Scheme 1a**) shows significant performance degradation compared to **Scheme 1g**), since it requires 20 dB SNR to achieve a BER of 10^{-4} . The **Scheme 1b**), which had knowledge of *a priori* information of the transmitted signals' PDF and structured sparsity exhibited improved performance compared to **Scheme 1a**). Specifically, it achieves a BER of 10^{-3} at 14.85 dB of SNR, which is 3.16 dB lower than that of **Scheme 1a**), but 10.36 dB higher than the SNR of **Scheme 1g**). The EP algorithm of **Scheme 1c**) imposes a high computational complexity due to iterative

TABLE IV: The SNR required by each Scheme at $\text{BER} = 10^{-2}$.

Scheme index	SNR at BER of 10^{-2}	Computational Complexity $\mathcal{O}(\cdot)$
Scheme 1	a)	12.8
	b)	9.47
	c)	3.12
	d)	7.44
	e)	4.41
	f)	1.77
	g)	1.75
	h)	1.73
Scheme 2	a)	17.13
	b)	-
	c)	6.71
	d)	3.59
	e)	10.05
	f)	5.33
	g)	5.25
	h)	5.13
Scheme 3	a)	37.31
	b)	32.29
	c)	14.23
	d)	24.37
	e)	11.75
	f)	10.72
Scheme 4	a)	24.51
	b)	22.04
	c)	10.41
	d)	15.72
	e)	8.92
	f)	8.71

updates of the messages and requires a careful selection of approximations for the posterior distributions. However, it substantially improves the performance of the MU detection scheme. As shown in Fig. 9(a), **Scheme 1c**) achieves a BER of 10^{-4} at an SNR of 10 dB. The conventional DNN detector is trained by the data of ML detection and it requires about 8.3 dB SNR for attaining a BER of 10^{-4} .

The GNN-aided detector of **Scheme 1d**)-**Scheme 1f**) exhibits a performance gain by relying on an accurate PDF of the transmitted signal information. Firstly, **Scheme 1d**) iteratively updates the estimated signal with the aid of the MMSE *a posteriori* information and achieves a BER of 10^{-4} at 12.22 dB of SNR. Then, for the GNN-AMP detector of **Scheme 1e**), the simple Gaussian distribution approximation of the transmitted signal distribution associated with the iteratively updated GNN module also leads to substantial performance improvement compared to **Scheme 1b**), which achieves a BER of 10^{-4} at 7.37 dB. This is only 0.3 dB worse than the SNR of **Scheme 1g**), despite its significant reduction in complexity. Furthermore, **Scheme 1f**) shows near-ML performance albeit at an increased detection complexity at a BER of 10^{-4} .

As the APM order increases from $\mathcal{L} = 2$ to $\mathcal{L} = 4$, the BER performance of all detection schemes deteriorates due to the reduced minimum Euclidean distance between constellation points, making the system more susceptible to noise and interference. As shown in Fig. 9(b), **Scheme 2h**) achieves the best performance, reaching a BER of 10^{-4} at an SNR of approximately 7.02 dB. **Scheme 2a**) demonstrates significantly worse performance, requiring around 16.32 dB to achieve the BER of 10^{-2} . This is due to its limited ability to exploit signal sparsity or *a priori* information. **Scheme 2b**) fails to reach the BER of 10^{-2} within the SNR range considered,

indicating its weak performance. By contrast, **Scheme 2c)** can still achieve a BER of 10^{-4} at 10.13 dB, outperforming both the MMSE and AMP detectors, thanks to its use of local Gaussian approximations and iterative updates, albeit at a higher computational cost.

The DNN-based detector of **Scheme 2d)** is trained with the aid of ML based detection and reaches a BER of 10^{-4} at 14.78 dB, showing about 10 dB gap compared to the ML benchmark. This gap indicates a limitation in the generalization capacity of the DNN model in multi-user environments. On the other hand, the GNN-aided detectors of **Scheme 2e)** and **Scheme 2f)** significantly improve the performance due to their ability to incorporate graph-based message passing that models the inter-user interference structure.

Specifically, **Scheme 2e)** achieves a BER of 10^{-4} at an SNR of 12.21 dB, while **Scheme 2f)** reaches the same BER at 7.37 dB. This is only 0.35 dB worse than **Scheme 2h)**, making it an efficient design alternative, as a benefit of its dramatically reduced complexity. **Scheme 2g)**, which amalgamates a graph-based structure and expectation propagation principles, further improves the performance and achieves a BER of 10^{-4} at 7.09 dB, approaching the optimal ML performance with the aid of a practical detection architecture.

Fig. 10(a) illustrates the performance of **Scheme 3** for $U = 16$ users, 4 TAs per user and 64 RAs deployed. As the number of users increases, the ML complexity increases exponentially, which becomes excessive. Compared to Fig. 9(a), **Scheme 3a)** has worse performance due to the increased number of users. As shown in Fig. 10(a), **Scheme 3b)** achieves similar performance to **Scheme 3a)**, failing to achieve a BER of 10^{-3} even at 25 dB SNR. Then, **Scheme 3c)** has even worse performance in this situation, exhibiting a BER of 10^{-4} at 24.82 dB SNR.

Next, the GNN-based detector of **Scheme 3d)**-**Scheme 3f)** is characterized in Fig. 10(a). **Scheme 3d)** does not perform well in this case, resulting in similar performance to **Scheme 3a)** and **Scheme 3b)**. **Scheme 3e)** shows a good performance in $U = 16$ users achieving a BER of 10^{-4} at 20.16 dB SNR, while **Scheme 3f)** performs 1.8 dB better than the **Scheme 3e)**.

In **Scheme 4**, we double the number of RAs to attain additional diversity gain. As shown in Fig. 10(b), the performance of **Scheme 4a)** and **Scheme 4b)** slightly improves. **Scheme 4c)** can succeeds in exploiting the benefits of increased number of RAs, then it achieves a BER of 10^{-4} at 18.65 dB SNR, which is about 6 dB better the performance of **Scheme 3**. With an increased number of RAs, all GNN-aided detectors of **Scheme 4** achieve improved performance compared to **Scheme 3**. The **Scheme 4d)** detector obtains a BER of 10^{-3} at 21.8 dB SNR, while **Scheme 3d)** attains a BER of 10^{-2} at 24.37 dB SNR. Both **Scheme 4e)** and **Scheme 4f)** achieve a BER of 10^{-3} at about 15.52 dB, where **Scheme 4e)** only suffers from a 0.2 dB SNR degradation with respect to **Scheme 4f)**.

Additionally, we consider a more realistic case, where not all users are active simultaneously. Fig. 11 illustrates the detection performance of our MU CS-SFIM framework under different maximum numbers of users, $U_{\max} = 4, 8, 16$, in a (MU-UL setting). In order to compare the robustness

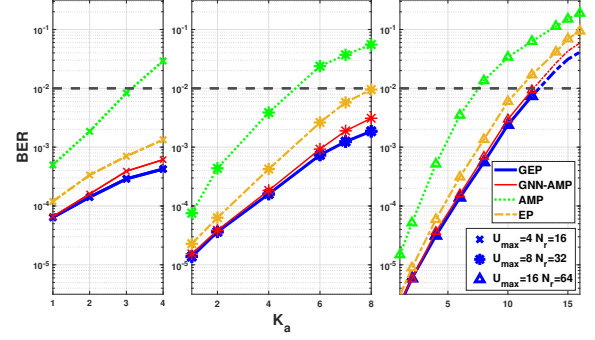


Fig. 11: BER performance versus the number of active users K_a for $N_r = 16, 32, 64$. at SNR of 5 dB for transmission over Rayleigh channel.

of both GNN-aided methods, the GEP and GNN-AMP with conventional MP methods, namely AMP and EP under a fixed SNR of 5 dB while varying the number of active users K_a . A threshold of BER at 10^{-2} is used as the benchmark for effective transmission. Each curve, distinguished by different markers, corresponds to a distinct scheme evaluated for scenarios with K_a ranging from 1 up to the maximum number of users. All GNN-aided methods are trained by a scheme that applies a mixture of a single active user to the maximum number of users to see the full range, ensuring fair comparison across different loads.

Across all scenarios, both GNN-AMP and GEP demonstrate significantly better BER performance than conventional AMP and EP, particularly as the number of active users increases. For instance, when $U_{\max} = 4$, AMP suffers from a rapid BER degradation exceeding a BER of 10^{-2} even at a low load of $K_a = 3$, while the GNN-aided methods and EP still maintain reasonable performance under 10^{-2} of BER. For a $U_{\max} = 8$ scheme, the AMP method reaches the 10^{-2} of BER at $K_a = 5$, while EP reaches the 10^{-2} of BER when all users are activated. GNN-AMP and GEP maintain viable performance under the same settings. In the scenario of a larger maximum user load of $U_{\max} = 16$ and large scale of RAs, each scheme can achieves better performance compared with less $U_{\max}$ and less RAs. However, with the load of user increasing, detection performance of each scheme degraded and exceed the effective threshold at a BER of 10^{-2} . The GNN solutions maintain a viable BER of 10^{-2} at $K_a = 12$, while the AMP method is ineffective at $K_a = 8$ and EP exceeds a BER of 10^{-2} at $K_a = 11$. The EP methods can outperform the AMP based low-complexity detector under varying of active users' but at a high computational complexity. Harnessing a GNN in both the AMP and EP can further increase the detection performance, despite varying the number of users.

Fig. 11 also shows that the GEP provides a marginally better performance than GNN-AMP, albeit at a higher complexity. GNN-AMP, on the other hand, achieves a better trade-off between complexity and performance. Furthermore, while all algorithms suffer from a degraded performance upon increasing K_a , the GNN-aided detectors exhibit a marginally better performance in MU-UL environments than conventional MP-

based detectors.

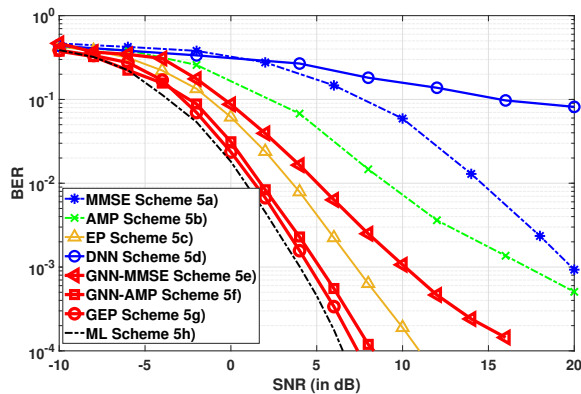


Fig. 12: Performance comparison of detectors in Scheme 5 for transmission over Rayleigh channel.

For example, the **Scheme 5** settings are used for characterizing the performance, where we consider $K_a = 3$, 3 out of 4 users being activated at a time. The DNN-based detector is trained with a fixed number of antennas and users. However, the number of users in practical MIMO systems changes dynamically. As shown in Fig. 12, we train the GNN-aided methods for $U = 4$ users and employ a 16×16 MIMO size. Then, we simulate the scenario, where 3 of the 4 users are communicating simultaneously. As shown in Fig. 12, both **Scheme 5a)** and **Scheme 5b)** result in performance degradation with respect to **Scheme 1**. As the MP methods have the ability to deal with a dynamic number of users, **Scheme 5c)** still performs well in this case, as shown in Fig. 12. Furthermore, both **Scheme 5e)** and **Scheme 5f)** exhibit robustness to a fluctuating number of users, due to the flexibility of the GNN.

V. CONCLUSION

GNN-based detectors were conceived for MU-CS-SFIM systems, namely the GNN-MMSE, AMP-GNN and GEPnet by providing the GNN module with the accurate distribution of *a priori* information and by intrinsically amalgamating the GNN module with MP algorithms. The MP-based GNN methods benefit from the low complexity of the MP algorithm and the improved efficiency brought by the GNN module. We demonstrated that the GNN based MP algorithms significantly outperform the corresponding conventional MP methods. The simulation results also confirmed that the GEPnet outperformed AMP-GNN at the cost of a higher complexity. Furthermore, we also validated that the GNN-aided detectors, particularly GNN-AMP and GEPnet, exhibit robustness to dynamic variations in the number of active users, which demonstrates that the GNN-based detectors are capable of supporting a flexible number of users. Furthermore, since the *a priori* information has already been exploited by our MP-based GNN detector, channel coding can be naturally incorporated into the proposed GNN-based framework for iteratively exchanging soft information by harnessing joint detection and decoding. This constitutes an interesting area for future research.

REFERENCES

- [1] M. Jiang and L. Hanzo, "Multiuser MIMO-OFDM for Next-Generation Wireless Systems," *Proceedings of the IEEE*, vol. 95, no. 7, pp. 1430–1469, 2007.
- [2] S. Yang and L. Hanzo, "Fifty Years of MIMO Detection: The Road to Large-Scale MIMOs," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 4, pp. 1941–1988, 2015.
- [3] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive MIMO Detection Techniques: A Survey," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 4, pp. 3109–3132, 2019.
- [4] K. K. Vaigandla and D. N. Venu, "Survey on Massive MIMO: Technology, Challenges, Opportunities and Benefits," 2021.
- [5] Y. Chau and S.-H. Yu, "Space Modulation On Wireless Fading Channels," in *IEEE 54th Vehicular Technology Conference. VTC Fall 2001. Proceedings (Cat. No. 01CH37211)*, vol. 3, 2001, pp. 1668–1671 vol.3.
- [6] N. Ishikawa, S. Sugiura, and L. Hanzo, "50 Years of Permutation, Spatial and Index Modulation: From Classic RF to Visible Light Communications and Data Storage," *IEEE Communications Surveys and Tutorials*, vol. 20, no. 3, pp. 1905–1938, 2018.
- [7] R. Y. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial Modulation," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2228–2241, 2008.
- [8] C. Xu, S. Sugiura, S. X. Ng, and L. Hanzo, "Spatial Modulation and Space-Time Shift Keying: Optimal Performance at a Reduced Detection Complexity," *IEEE Transactions on Communications*, vol. 61, no. 1, pp. 206–216, 2013.
- [9] L. He, J. Wang, and J. Song, "On Massive Spatial Modulation MIMO: Spectral Efficiency Analysis and Optimal System Design," in *2016 IEEE Global Communications Conference (GLOBECOM)*, 2016, pp. 1–6.
- [10] P. Yang, M. Di Renzo, Y. Xiao, S. Li, and L. Hanzo, "Design Guidelines for Spatial Modulation," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 1, pp. 6–26, 2015.
- [11] T. Datta and A. Chockalingam, "On Generalized Spatial Modulation," in *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, 2013, pp. 2716–2721.
- [12] T. Lakshmi Narasimhan, P. Raviteja, and A. Chockalingam, "Generalized Spatial Modulation in Large-Scale Multiuser MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 7, pp. 3764–3779, 2015.
- [13] E. Basar, "Index Modulation Techniques For 5G Wireless Networks," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 168–175, 7 2016.
- [14] R. Abu-alhiga and H. Haas, "Subcarrier-Index Modulation OFDM," *2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 177–181, 2009.
- [15] M. Wen, X. Cheng, M. Ma, B. Jiao, and H. V. Poor, "On the Achievable Rate of OFDM With Index Modulation," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 1919–1932, 2016.
- [16] R. Fan, Y. J. Yu, and Y. L. Guan, "Generalization of Orthogonal Frequency Division Multiplexing With Index Modulation," *IEEE Transactions on Wireless Communications*, vol. 14, no. 10, pp. 5350–5359, 2015.
- [17] M. Yüzgeçcioğlu and E. Jorswieck, "Uplink and Downlink Transceiver Design for OFDM with Index Modulation in Multi-user Networks," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017, pp. 1–5.
- [18] H. Zhang, L.-L. Yang, and L. Hanzo, "Compressed Sensing Improves the Performance of Subcarrier Index-Modulation-Assisted OFDM," *IEEE Access*, vol. 4, pp. 7859–7873, 2016.
- [19] D. L. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, apr 2006.
- [20] T. Datta, H. S. Eshwaraiah, A. Chockalingam, and S. Member, "Generalized Space-and-Frequency Index Modulation," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 4911–4924, 2016.
- [21] I. A. Hemadeh, M. El-Hajjar, and L. Hanzo, "Hierarchical Multi-Functional Layered Spatial Modulation," *IEEE Access*, vol. 6, pp. 9492–9533, 2018.
- [22] B. Shamasundar, S. Bhat, S. Jacob, and A. Chockalingam, "Multidimensional Index Modulation in Wireless Communications," *IEEE Access*, vol. 6, pp. 589–604, 2018.
- [23] P. Yang, Y. Xiao, Y. L. Guan, M. Di Renzo, S. Li, and L. Hanzo, "Multidomain Index Modulation for Vehicular and Railway Communications: A Survey of Novel Techniques," *IEEE Vehicular Technology Magazine*, vol. 13, no. 3, pp. 124–134, 2018.
- [24] C. Xu, Y. Xiong, N. Ishikawa, R. Rajashekar, S. Sugiura, Z. Wang, S.-X. Ng, L.-L. Yang, and L. Hanzo, "Space-, Time- and Frequency-Domain Index Modulation for Next-Generation Wireless: A Unified

- Single-/Multi-Carrier and Single-/Multi-RF MIMO Framework,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 3847–3864, 2021.
- [25] S. Lu, I. A. Hemadeh, M. El-Hajjar, and L. Hanzo, “Compressed Sensing-Aided Multi-Dimensional Index Modulation,” *IEEE Transactions on Communications*, vol. 67, no. 6, pp. 4074–4087, 2019.
- [26] X. Feng, M. El-Hajjar, C. Xu, and L. Hanzo, “Near-Instantaneously Adaptive Learning-Assisted and Compressed Sensing-Aided Joint Multi-Dimensional Index Modulation,” *IEEE Open Journal of Vehicular Technology*, vol. 4, pp. 893–912, 2023.
- [27] S. Lu, M. El-Hajjar, and L. Hanzo, “Two-Dimensional Index Modulation for the Large-Scale Multi-User MIMO Uplink,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7904–7918, 2019.
- [28] D. L. Donoho, A. Maleki, and A. Montanari, “Message Passing Algorithms for Compressed Sensing,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18914–18919, 2009.
- [29] C. Jeon, R. Ghods, A. Maleki, and C. Studer, “Optimality of Large MIMO Detection via Approximate Message Passing,” in *2015 IEEE International Symposium on Information Theory (ISIT)*, 2015, pp. 1227–1231.
- [30] L. Wei, J. Zheng, and Q. Liu, “Approximate Message Passing Detector for Index Modulation With Multiple Active Resources,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 972–976, 2019.
- [31] Z. Sui, S. Yan, H. Zhang, L.-L. Yang, and L. Hanzo, “Approximate Message Passing Algorithms for Low Complexity OFDM-IM Detection,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 9607–9612, 2021.
- [32] B. Chakrapani, T. L. Narasimhan, and A. Chockalingam, “Generalized Space-Frequency Index Modulation: Low-Complexity Encoding and Detection,” in *2015 IEEE Globecom Workshops (GC Wkshps)*, 2015, pp. 1–6.
- [33] J. Ma and L. Ping, “Orthogonal AMP,” *IEEE Access*, vol. 5, pp. 2020–2033, 2017.
- [34] J. Céspedes, P. M. Olmos, M. Sánchez-Fernández, and F. Perez-Cruz, “Expectation Propagation Detection for High-Order High-Dimensional MIMO Systems,” *IEEE Transactions on Communications*, vol. 62, no. 8, pp. 2840–2849, 2014.
- [35] H. Wang, A. Kosasih, C.-K. Wen, S. Jin, and W. Hardjawana, “Expectation Propagation Detector for Extra-Large Scale Massive MIMO,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2036–2051, 2020.
- [36] H. He, C.-K. Wen, S. Jin, and G. Y. Li, “Model-Driven Deep Learning for MIMO Detection,” *IEEE Transactions on Signal Processing*, vol. 68, pp. 1702–1715, 2020.
- [37] H. He, S. Jin, C.-K. Wen, F. Gao, G. Y. Li, and Z. Xu, “Model-Driven Deep Learning for Physical Layer Communications,” *IEEE Wireless Communications*, vol. 26, no. 5, pp. 77–83, 2019.
- [38] N. Samuel and T. Diskin, “Learning to Detect,” *IEEE Transactions on Signal Processing*, vol. 67, no. 10, pp. 2554–2564, may 2019.
- [39] A. Scotti, N. N. Moghadam, D. Liu, K. Gafvert, and J. Huang, “Graph Neural Networks for Massive MIMO Detection,” *arXiv preprint arXiv:2007.05703*, 2020.
- [40] Y. Shen, J. Zhang, S. H. Song, and K. B. Letaief, “Graph Neural Networks for Wireless Communications: From Theory to Practice,” *IEEE Transactions on Wireless Communications*, vol. 22, no. 5, pp. 3554–3569, 2023.
- [41] H. He, A. Kosasih, X. Yu, J. Zhang, S. Song, W. Hardjawana, and K. B. Letaief, “GNN-Enhanced Approximate Message Passing for Massive/Ultra-Massive MIMO Detection,” in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, 2023, pp. 1–6.
- [42] A. Kosasih, V. Onasis, W. Hardjawana, V. Miloslavskaya, V. Andrean, J.-S. Leu, and B. Vucetic, “Graph Neural Network Aided Expectation Propagation Detector for MU-MIMO Systems,” in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, 2022, pp. 1212–1217.
- [43] A. Kosasih, V. Onasis, V. Miloslavskaya, W. Hardjawana, V. Andrean, and B. Vucetic, “Graph Neural Network Aided MU-MIMO Detectors,” *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2540–2555, 2022.
- [44] J. Li, Q. Li, S. Dang, M. Wen, X.-Q. Jiang, and Y. Peng, “Low-Complexity Detection for Index Modulation Multiple Access,” *IEEE Wireless Communications Letters*, vol. 9, no. 7, pp. 943–947, 2020.
- [45] S. Narayanan, M. J. Chaudhry, A. Stavridis, M. Di Renzo, F. Graziosi, and H. Haas, “Multi-User Spatial Modulation MIMO,” in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, 2014, pp. 671–676.
- [46] K.-H. Ngo, M. Guillaud, A. Decurninge, S. Yang, S. Sarkar, and P. Schniter, “Non-Coherent Multi-User Detection Based on Expectation Propagation,” in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, 2019, pp. 2092–2096.
- [47] Y. Liu, M. Zhang, H. Wang, and X. Cheng, “Spatial Modulation Orthogonal Frequency Division Multiplexing with Subcarrier Index Modulation for V2X Communications,” in *2016 International Conference on Computing, Networking and Communications (ICNC)*, 2016, pp. 1–5.
- [48] H. Zhang, L.-L. Yang, and L. Hanzo, “Compressed Sensing Improves the Performance of Subcarrier Index-Modulation-Assisted OFDM,” *IEEE Access*, vol. 4, pp. 7859–7873, 2016.
- [49] M. Jiang and L. Hanzo, “Multiuser MIMO-OFDM for Next-Generation Wireless Systems,” *Proceedings of the IEEE*, vol. 95, no. 7, pp. 1430–1469, 2007.
- [50] S. Chen, S. X. Ng, E. F. Khalaf, A. Morfeq, and N. D. Alotaibi, “Multiuser Detection for Nonlinear MIMO Uplink,” *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 207–219, 2020.
- [51] E. Ekrem, M. Koca, and H. Deliç, “Iterative Synchronization of Multiuser Ultra-Wideband Signals,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 10, pp. 3040–3051, 2010.
- [52] W. M. Jang, B. Vojcic, and R. Pickholtz, “Joint transmitter-receiver optimization in synchronous multiuser communications over multipath channels,” *IEEE Transactions on Communications*, vol. 46, no. 2, pp. 269–278, 1998.
- [53] F. Kschischang, B. Frey, and H.-A. Loeliger, “Factor Graphs and the Sum-Product Algorithm,” *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, 2001.
- [54] Z. Sui, S. Yan, H. Zhang, L.-L. Yang, and L. Hanzo, “Approximate Message Passing Algorithms for Low Complexity OFDM-IM Detection,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 9607–9612, 2021.
- [55] L. Wei, J. Zheng, and Q. Liu, “Approximate Message Passing Detector for Index Modulation With Multiple Active Resources,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 972–976, 2019.
- [56] J. P. Vila and P. Schniter, “Expectation-Maximization Gaussian-Mixture Approximate Message Passing,” *IEEE Transactions on Signal Processing*, vol. 61, no. 19, pp. 4658–4672, 2013.
- [57] S. Rangan, “Generalized Approximate Message Passing for Estimation with Random Linear Mixing,” in *2011 IEEE International Symposium on Information Theory Proceedings*, 2011, pp. 2168–2172.
- [58] T. P. Minka, “Expectation Propagation for Approximate Bayesian Inference,” *arXiv preprint arXiv:1301.2294*, 2013.
- [59] B. Khemani, S. Patil, K. Kotecha, and S. Tanwar, “A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions,” *Journal of Big Data*, vol. 11, no. 1, p. 18, 2024.
- [60] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation,” *arXiv preprint arXiv:1406.1078*, 2014.
- [61] X. Feng, M. El-Hajjar, C. Xu, and L. Hanzo, “Near-instantaneously adaptive learning-assisted and compressed sensing-aided joint multi-dimensional index modulation,” *IEEE Open Journal of Vehicular Technology*, vol. 4, pp. 893–912, 2023.