IEICE TRANS. COMMUN., VOL.E107–B, NO.4 APRIL 2024

PAPER

A Lightweight Graph Neural Networks Based Enhanced Separated Detection Scheme for Downlink MIMO-SCMA Systems*

Zikang CHEN[†,††], *Student Member*, **Wenping GE**[†,††a)], **Henghai FEI**[†,††], **Haipeng ZHAO**[†,††], *and* **Bowen LI**[†,††], *Nonmembers*

SUMMARY The combination of multiple-input multiple-output (MIMO) technology and sparse code multiple access (SCMA) can significantly enhance the spectral efficiency of future wireless communication networks. However, the receiver design for downlink MIMO-SCMA systems faces challenges in developing multi-user detection (MUD) schemes that achieve both low latency and low bit error rate (BER). The separated detection scheme in the MIMO-SCMA system involves performing MIMO detection first to obtain estimated signals, followed by SCMA decoding. We propose an enhanced separated detection scheme based on lightweight graph neural networks (GNNs). In this scheme, we raise the concept of coordinate point relay and full-category training, which allow for the substitution of the conventional message passing algorithm (MPA) in SCMA decoding with image classification techniques based on deep learning (DL). The features of the images used for training encompass crucial information such as the amplitude and phase of estimated signals, as well as channel characteristics they have encountered. Furthermore, various types of images demonstrate distinct directional trends, contributing additional features that enhance the precision of classification by GNNs. Simulation results demonstrate that the enhanced separated detection scheme outperforms existing separated and joint detection schemes in terms of computational complexity, while having a better BER performance than the joint detection schemes at high E_b/N_0 (energy per bit to noise power spectral density ratio) values.

key words: *MIMO-SCMA, multi-user detection (MUD), bit error rate (BER), deep learning (DL)*

1. Introduction

1.1 Background

With the rapid development of the internet of things (IoT) [1], the demands placed on next generation wireless communication networks have become increasingly rigorous, requiring higher spectrum efficiency, reduced latency, and improved communication quality. While orthogonal multiple access (OMA) techniques have been successful in previous communication eras by mitigating inter-user interference through the allocation of orthogonal resource elements (REs) [2], the scarcity of spectrum resources driven by the pursuit of high throughput makes it challenging to rely solely on OMA techniques for resolution. The advent of non-orthogonal multiple access (NOMA) technology has revitalized the field of multiple access techniques, allowing for the transmission of signals from different users on the same RE, thereby increasing the overloading factor of REs to users and effectively alleviating the strain on limited spectrum resources [3].

Sparse code multiple access (SCMA) technology, as one of the various NOMA techniques, employs combinations of sparse code vectors, enabling simultaneous reception and decoding of multi-user signals [4]. This reduces the complexity of NOMA based systems while providing excellent anti-interference performance due to the high mutual information between different user signals. Multiple-input multiple-output (MIMO) technology, which utilizes spatial multiplexing, is another crucial technique for enhancing spectrum efficiency in next generation wireless communication networks [5].

In this context, MIMO-SCMA holds great promise in further improving spectrum efficiency, which is a primary reason for the sustained interest of the academic community in this field [6].

1.2 Related Work and Motivation

MIMO-SCMA is a technology that utilizes codebooks to map user data into multidimensional sparse codewords for transmission via multiple antennas. In order to improve the performance of MIMO-SCMA systems, a large-scale codebook optimization algorithm was proposed by [7]. Additionally, the design of the receiver plays a critical role in determining the performance of MIMO-SCMA systems. Separated detection algorithms, which combine MIMO detection algorithms [8] and SCMA detection scheme (message passing algorithm (MPA) [9]), suffer from inferior bit error rate (BER) performance and have a high computational complexity. To enhance the decoding performance, a joint sparse graph-detector that integrates the single graph of MIMO channels and SCMA codewords was proposed in [10]. However, while this technique effectively reduces the BER, it does not exhibit a significant decrease in computational complexity. Building upon the ideas presented in [10], [11] introduced two innovative low-complexity detectors based on an extended MIMO-SCMA factor graph for downlink MIMO-SCMA systems. The experimental results

Manuscript received August 23, 2023.
Manuscript revised October 23, 2023.
Manuscript publicized January 30, 2024.

[†]The authors are with College of Information Science and Engineering, Xinjiang University, 830046, China.
[††]The authors are also with the Signal Detection and Processing Key Laboratory, 830046, China.

*This work was partially supported by the Natural Science Foundation of Xinjiang Uygur Autonomous Region under Grant 2022D01C426 and supported by the Graduate Research Innovation Project of Xinjiang Uygur Autonomous Region under Grant XJ2023G100.

a) E-mail: wenpingge@xju.edu.cn

DOI: 10.23919/transcom.2023EBP3144

indicated that, while there was a noticeable reduction in computational complexity, there was a slight decline in BER performance. Therefore, existing separated and joint detection schemes do not achieve a balanced trade-off between computational complexity and BER performance.

Deep learning (DL) has become a widely utilized technology across various domains. In the SCMA systems, researchers in [12] applied DL to the receiver and developed a decoder with lower computational complexity than MPA while achieving a comparable BER performance. Building upon this work, [13] proposed an automatic encoder-decoder based on deep neural networks (DNNs) specifically designed for SCMA systems, demonstrating even lower computational complexity and superior BER performance. However, it is worth noting that these DL based SCMA decoders did not take into consideration the integration of SCMA with MIMO technology. [14] presented a DL based network model for application in the MIMO detection, thereby advancing the development of DL based MIMO-SCMA systems.

In this paper, we directly employ DL techniques for downlink MIMO-SCMA systems, and propose an enhanced separated detection algorithm based on lightweight graph neural networks (GNNS). We propose the concept of coordinate point relay, which maps the amplitude and phase of the estimated signal obtained by MIMO detection and the channel characteristics it have experienced into a K-point polyline graph with trend features. The differences in types of K-point polyline graphs can be ultimately attributed to the differences in the corresponding transmission symbol combinations. The GNNs all adopt the same modified MobileNet architecture [15]. Furthermore, we propose the concept of full-category training, where the utilized image dataset for the training process encompasses all categories of K-point polyline graphs, in contrast to the random training approach in [12]. This ensures a more scientific and rigorous training process. Our proposed scheme surpasses existing separated and joint detection schemes in terms of computational complexity, while achieving a better BER performance than the joint detection schemes across the high E_b/N_0 (energy per bit to noise power spectral density ratio) values.

1.3 Contributions

- We propose the concept of coordinate point relay, which allows us to generate K-point polyline graphs with trend features for training purposes. The eigenvalues of the K-point polyline graphs contain crucial information, including the amplitude and phase of estimated signals, as well as channel characteristics. Additionally, different types of K-point polyline graphs exhibit diverse trend directions, providing extra features that aid in accurate classification by the GNNs.
- We propose the concept of full-category training, whereby the employed image dataset for training comprises all distinct categories of K-point polyline graphs. This approach guarantees a more methodical and rigorous training process, lending greater scientific validity to our study.
- Our proposed algorithm presents a novel research perspective by combining traditional communication and computer vision techniques. This approach offers a fresh insight into multi-user detection (MUD) in the downlink MIMO-SCMA system, by replacing the role of the MPA at the SCMA receiver with image classification techniques based on lightweight GNNs. Additionally, our algorithm balance both computational complexity and BER performance. In comprehensive evaluations, it demonstrates superior performance compared to existing separated and joint detection schemes.

1.4 Organization

The remainder of this article is organized as follows. Section 2 introduces the downlink MIMO-SCMA system model. Section 3 introduces the conventional separated detection scheme, and describes our lightweight GNNs based enhanced separated detection scheme (LG-ESDS) in detail. Section 4 presents and evaluates the simulation results. Finally, Sect. 5 presents the conclusions.

2. Downlink MIMO-SCMA System Model

Figure 1 illustrates a downlink MIMO-SCMA system with J independent users multiplexed over K orthogonal REs, achieving an overloading factor of $\lambda = J/K$. In this system, the base station is equipped with N_t transmit antennas, while each user is equipped with N_r receive antennas. For the n_t-th antenna of user u, where $n_t = 1, 2, \ldots, N_t$ and $u = 1, 2, \ldots, J$, the input $\log_2(M)$ binary bits $\mathbf{b}_u^{n_t}$ are mapped into a K-dimensional complex codeword $\mathbf{x}_u^{n_t} = \left[x_{u,1}^{n_t}, x_{u,2}^{n_t}, \ldots, x_{u,K}^{n_t}\right]^T$, which is selected from the known corresponding SCMA codebook $\mathbf{C}_u^{n_t} \in \mathbb{C}^{K \times M}$ with size M. Based on the size of the codebook, each user can be considered to have M possible transmission symbols (e.g., $0, 1, \ldots, M-1$). Therefore, at the n_t-th antenna, the transmitted overlapping codeword corresponding to the transmission symbols combination (TSC) of J users can be represented

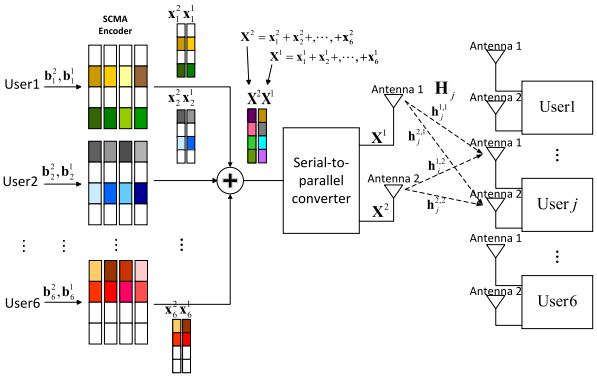


Fig. 1 Architecture of downlink MIMO-SCMA system with $J = 6$, $K = 4$, $M = 4$, $N_t = 2$, and $N_r = 2$.

as

$$\mathbf{x}^{n_t} = \sum_{u=1}^{J} \mathbf{x}_u^{n_t}. \tag{1}$$

The received signal at the n_r-th antenna of user j, where $j = 1, 2, \ldots, J$ and $n_r = 1, 2, \ldots, N_r$, can be expressed as

$$\mathbf{y}_j^{n_r} = \sum_{n_t=1}^{N_t} \mathrm{diag}\left\{\mathbf{h}_j^{n_r,n_t}\right\} \mathbf{x}^{n_t} + \mathbf{n}_j^{n_r}, \tag{2}$$

where $\mathbf{h}_j^{n_r,n_t} = \left[h_{j,1}^{n_r,n_t}, h_{j,2}^{n_r,n_t}, \ldots, h_{j,K}^{n_r,n_t}\right]^T$ represents the channel gain vector between the n_t-th antenna of base station and the n_r-th antenna of j-th user, and $\mathbf{n}_j^{n_r} = \left[n_{j,1}^{n_r}, n_{j,2}^{n_r}, \ldots, n_{j,K}^{n_r}\right]^T$ is the additive white Gaussian noise (AWGN) with zero mean and variance $\sigma_{n_r}^2$. By stacking the signals at all N_r receive antennas together, we can obain the received signal $\mathbf{y}_j$ of user j, which can be expressed as

$$\mathbf{y}_j = \sum_{n_t=1}^{N_t} \sum_{u=1}^{J} \mathrm{diag}\left\{\mathbf{h}_j^{n_t}\right\} \tilde{\mathbf{x}}_u^{n_t} + \mathbf{n}_j, \tag{3}$$

where

$$\begin{aligned}
\mathbf{y}_j &= \left[\left(\mathbf{y}_j^1\right)^T, \left(\mathbf{y}_j^2\right)^T, \ldots, \left(\mathbf{y}_j^{N_r}\right)^T\right]^T, \\
\mathbf{h}_j^{n_t} &= \left[\left(\mathbf{h}_j^{1,n_t}\right)^T, \left(\mathbf{h}_j^{2,n_t}\right)^T, \ldots, \left(\mathbf{h}_j^{N_r,n_t}\right)^T\right]^T, \\
\tilde{\mathbf{x}}_u^{n_t} &= \left[\left(\mathbf{x}_u^{n_t}\right)^T, \left(\mathbf{x}_u^{n_t}\right)^T, \ldots, \left(\mathbf{x}_u^{n_t}\right)^T\right]^T, \\
\mathbf{n}_j &= \left[\left(\mathbf{n}_j^1\right)^T, \left(\mathbf{n}_j^2\right)^T, \ldots, \left(\mathbf{n}_j^{N_r}\right)^T\right]^T.
\end{aligned} \tag{4}$$

3. Separated Detection Scheme for Downlink MIMO-SCMA System

In Sect. 3.1, we commence by introducing the concept of conventional separated detection algorithm. Following that, we present a comprehensive and detailed exposition of our novel LG-ESDS in Sect. 3.2.

3.1 Conventional Separated Detection Algorithm

Figure 2 illustrates the architecture of the conventional separated detection algorithm, depicting four distinct categories of nodes. These categories include RA nodes representing the receive antennas, TA nodes denoting the transmit antennas, R nodes embodying the REs, and U nodes signifying the users. The conventional separated detection algorithm refers to a two-step process, involving MIMO detection followed by MPA decoding of the estimated signals obtained from MIMO detection. During this process, MIMO detection and MPA decoding are performed independently. However, MPA necessitates numerous iterative loops, which hinders meeting the low latency requirements of modern communication systems. Moreover, MPA relies on continuous message exchange between R nodes and U nodes, resulting in the transmission and reception of both relevant and irrelevant information. This makes it difficult to directly exploit the effective information of the estimated signals. In light of this, we propose the LG-ESDS for MIMO-SCMA systems.

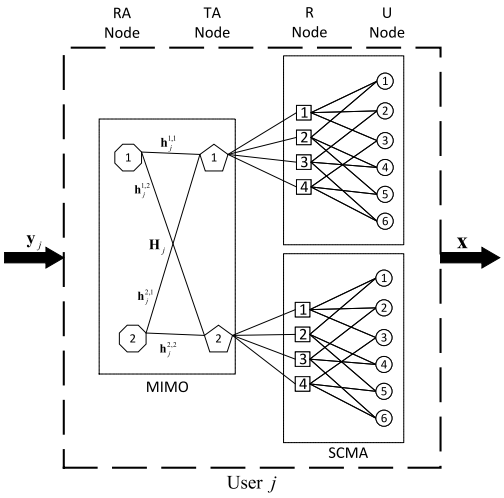


Fig. 2 Architecture of conventional separated detection algorithm with $J = 6$, $K = 4$, $M = 4$, $N_t = 2$, and $N_r = 2$.

3.2 LG-ESDS

In Sect. 3.2.1, we provide an introduction to the MIMO detection algorithm: minimum mean square error (MMSE) detectionn algorithm [8], which is utilized in our proposed LG-ESDS. In Sect. 3.2.2, we provide a detailed description of the concept of coordinate point relay, which enables the replacement of MPA in SCMA decoding with GNNs based image classification technology. Moving to Sect. 3.2.3, we propose the concept of full-category training to enhance the scientific and rigorous nature of the training process. Then, we elucidate the logical derivation for optimizing GNNs parameters. Section 3.2.4 focuses on the experimental setup and parameter configuration employed in our simulation experiment. We present the model parameters of the MIMO-SCMA system used and describe the specific structure of the GNNs. Moreover, we outline the parameter settings of our proposed LG-ESDS. While the detection scheme we have proposed can be applied to larger MIMO-SCMA systems, it is important to note that the design of codebooks for such systems falls outside the scope of our study. Therefore, we consider a downlink MIMO-SCMA system where different transmit antennas employ the same mapping codebook. Based on this, we only need to clarify how our proposed LG-ESDS helps user j retrieve his own information transmitted by the antenna n_t.

3.2.1 MMSE Detection Algorithm

Based on Eqs. (3), a more simplified representation of $\mathbf{y}_j$ can

be achieved as

$$\mathbf{y}_j = \mathbf{H}_j\mathbf{x} + \mathbf{n}_j, \tag{5}$$

where $\mathbf{H}_j = \left[\mathbf{h}_j^1, \mathbf{h}_j^2, \ldots, \mathbf{h}_j^{N_t}\right]$ is the MIMO channel matrix, $\mathbf{x} = \left[\left(\mathbf{x}^1\right)^T, \left(\mathbf{x}^2\right)^T, \ldots, \left(\mathbf{x}^{N_t}\right)^T\right]^T$, and $\mathbf{n}_j$ is the complex Gaussian noise vector at j-th user with zero mean and variance σ^2.

The fundamental principle of the MMSE algorithm is to minimize the expected value of the mean square error between the estimated signal and the actual transmitted signal [8]. Mathematically, this can be expressed as follows:

$$F_{MMSE} = \arg\min_F E\left\|F\mathbf{y}_j - \mathbf{x}\right\|^2, \tag{6}$$

where F_{MMSE} is defined as the objective function of the MMSE algorithm. According to the principle of orthogonality, we can derive

$$E\left\{\left(F_{MMSE}\mathbf{y}_j - \mathbf{x}\right)\mathbf{y}^H\right\} = 0, \tag{7}$$

where $\mathbf{y}^H$ is the conjugate transpose of $\mathbf{y}_j$. By combining Eqs. (5) and (7) and employing the principle of matrix inversion, we can obtain the simplified expression

$$F_{MMSE} = \left(\mathbf{H}_j^H\mathbf{H}_j + \sigma^2\mathbf{I}_K\right)^{-1}\mathbf{H}_j^H, \tag{8}$$

where $\mathbf{H}_j^H$ is the conjugate transpose of $\mathbf{H}_j$, and $\mathbf{I}_K$ is a K-dimensional identity matrix. Therefore, the estimated signal $\mathbf{e}$ obtained through the MMSE algorithm can be represented as

$$\mathbf{e} = F_{MMSE}\mathbf{y}_j = \left(\mathbf{H}_j^H\mathbf{H}_j + \sigma^2\mathbf{I}_K\right)^{-1}\mathbf{H}_j^H\mathbf{y}_j, \tag{9}$$

where $\mathbf{e} = \left[\left(\mathbf{e}^1\right)^T, \left(\mathbf{e}^2\right)^T, \ldots, \left(\mathbf{e}^{N_t}\right)^T\right]^T$, and $\mathbf{e}^{n_t} = \left[e_1^{n_t}, e_2^{n_t}, \ldots, e_K^{n_t}\right]^T$ represents the estimated value of the signal transmitted by the antenna n_t.

3.2.2 Coordinate Point Relay

The realization of the coordinate point relay is accomplished in the PyTorch environment, utilizing the matplotlib module [16]. We map the first component, $e_1^{n_t}$, of $\mathbf{e}^{n_t}$, including both the real and imaginary parts (i.e., $\mathrm{Re}\left(e_1^{n_t}\right)$ and $\mathrm{Im}\left(e_1^{n_t}\right)$), onto a Cartesian coordinate system, resulting in the coordinate point c_1, which can be expressed as

$$(X_1, Y_1) = \left(\mathrm{Re}\left(e_1^{n_t}\right), \mathrm{Im}\left(e_1^{n_t}\right)\right), \tag{10}$$

where (X_1, Y_1) represents the coordinate of c_1. Based on this, we can obtain the corresponding point c_2 in the Cartesian coordinate system for the second component, $e_2^{n_t}$, of $\mathbf{e}^{n_t}$, that can be expressed as

$$(X_2, Y_2) = (X_1, Y_1) + \left(\mathrm{Re}\left(e_2^{n_t}\right) + \mathrm{RF}, \mathrm{Im}\left(e_2^{n_t}\right)\right), \quad \mathrm{Re}\left(e_2^{n_t}\right) + \mathrm{RF} > 0, \tag{11}$$

where (X_2, Y_2) represents the coordinate of c_2, $\mathrm{Re}\left(e_2^{n_t}\right)$ and $\mathrm{Im}\left(e_2^{n_t}\right)$ represent the real and imaginary parts of $e_2^{n_t}$, and RF is the rightwalk factor (RF), which is a novel concept proposed by us, ensuring that c_2 lies to the right of c_1. Similarly, we can obtain the corresponding points in the Cartesian coordinate system for the remaining $K-2$ components of $\mathbf{e}^{n_t}$. This can be expressed as

$$(X_{z+1}, Y_{z+1}) = (X_z, Y_z) + \left(\mathrm{Re}\left(e_{z+1}^{n_t}\right) + \mathrm{RF}, \mathrm{Im}\left(e_{z+1}^{n_t}\right)\right), \tag{12}$$

where (X_{z+1}, Y_{z+1}) represents the coordinate of c_{z+1} ($z = 2, 3, \ldots, K-1$), (X_z, Y_z) represents the coordinate of c_z, $\mathrm{Re}\left(e_{z+1}^{n_t}\right)$ and $\mathrm{Im}\left(e_{z+1}^{n_t}\right)$ represent the real and imaginary parts of $e_{z+1}^{n_t}$, and RF satisfies

$$\mathrm{Re}\left(e_{z+1}^{n_t}\right) + \mathrm{RF} > 0, \tag{13}$$

where RF guarantees that c_{z+1} is positioned to the right of c_z. For a receiver with perfect channel state information (CSI), the channel characteristic $\mathbf{h}_j^{n_r,n_t}$ is calculated instantaneously in real-time to evaluate the characteristics of the channel. The module matplotlib.colors and the colormap class in matplotlib allow for mapping floating-point numbers in the range of 0 to 1 to color values. Leveraging this capability, we can define the color value for the c_k ($k = 1, 2, \ldots, K$). The floating-point number corresponding to the color value of the c_k is set as follows:

$$\sum_{n_r=1}^{Nr} \frac{sig\left(\left|h_{j,k}^{n_r,n_t}\right|\right)}{N_r}, \tag{14}$$

where $\left|h_{j,k}^{n_r,n_t}\right|$ represents the magnitude of $h_{j,k}^{n_r,n_t}$, and $sig(\cdot)$ refers to the sigmoid activation function. By not displaying the coordinate system and sequentially connecting $c_1, c_2, \ldots, c_K$, with the connecting lines during this process set to black, we obtain a K-point polyline graph with trend features, as shown in Fig. 3. It is worth noting that, during the process of handling the output of K-point polyline graph, we have made efforts to retain only the relevant portion of K-point polyline graph, thereby removing any excess blank areas surrounding K-point polyline graph. A K-point polyline graph with excessively big size can result in increased computational complexity for LG-ESDS, whereas a K-point polyline graph with excessively small size may distort the image. Therefore, the output format of the K-point polyline graph is appropriately set as 3 * 32 * 32 by the figsize function [16], where "3" refers to the number of color channels (i.e., red, green, and blue), and "32 * 32" specifies the size of the image in terms of width and height in pixels.

The above represents the mapping process of a certain K-point polyline graph. Considering J users, each with M possible transmission symbols, there are a total of M^J distinct types of K-point polyline graphs, corresponding to the M^J types of transmission symbol combinations (TSCs).

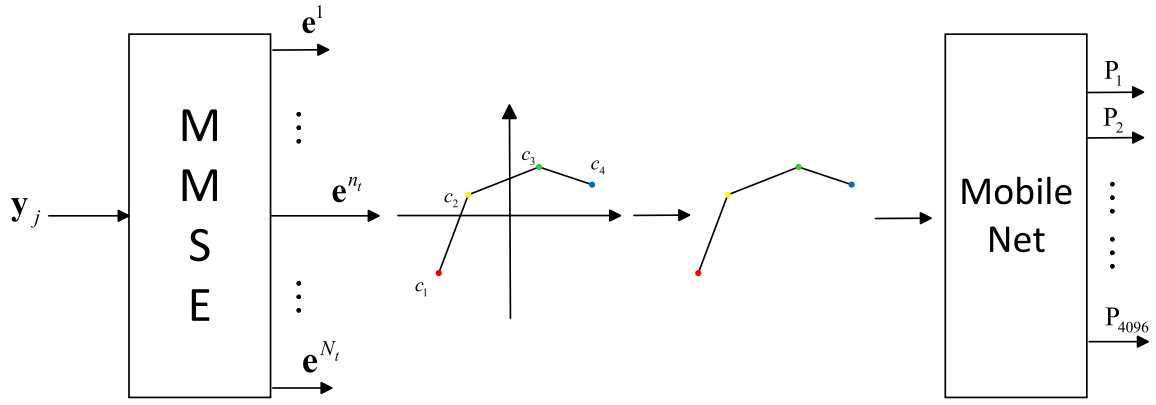


Fig. 3 Architecture of our proposed LG-ESDS with $J = 6$, $K = 4$, and $M = 4$.

The coordinate point relay can generate various types of K-point polyline graphs, whose feature values include crucial information required for decoding, such as the amplitude and phase of the estimated signals, as well as the channel characteristics they have encountered. Additionally, since the coordinate of c_{k+1} in the K-point polyline graph are derived from c_k, each type of K-point polyline graph exhibits distinct trend features. RF ensures that the Euclidean distance between the mapped coordinate points in the K-point polyline graph is sufficiently large. This property is beneficial for GNNs to differentiate between different types of K-point polyline graphs.

3.2.3 Full-Category Training

In the MIMO-SCMA system, the overlapping codeword corresponding to the TSC transmitted by antenna n_t is mapped as a K-point polyline graph after MMSE detection and coordinate point relay. Based on the one-to-one correspondence between the types of TSCs and K-point polyline graphs, we can obtain the corresponding type of K-point polyline graph, by controlling the type of TSC transmitted by antenna n_t. The TSC corresponding to the K-point polyline graph is treated as a J-bit M-ary number, which equals a decimal value. And the decimal value is defined as the category of the K-point polyline graph. This can be expressed as follow

$$l = \sum_{j=1}^{J} m_j M^{J-j}, \tag{15}$$

where l is the category of the K-point polyline graph, m_j $(m = 0, 1, \ldots, M-1)$ is the user j's transmission symbol.

Unlike [12] which generates simulated data of TSCs with random categories for training DNNs, our proposed LG-ESDS adopts a full-category training approach. Specifically, an equal proportion of each type of TSCs' simulated data is generated at the transmitter of the MIMO-SCMA system, allowing GNNs to learn the features of all types of K-point polyline graphs in a systematic manner. During each communication process of generating the K-point polyline graph, we consider the dynamic changes of $\mathbf{h}_j^{n_r,n_t}$ and perform real-time calculations accordingly to determine the color value of c_k.

In order to find the optimal E_b/N_0 value for training, we test the following scenarios in this paper.

- S: train the model using a E_b/N_0 value of 6 dB
- M: train the model using a E_b/N_0 value of 8 dB
- B: train the model using a E_b/N_0 value of 10 dB

To classify a K-point polyline graph and predict the $\log_2(M)$ data bits for user j, we train the GNNs' parameters by minimizing the following loss function:

$$L(\mathbf{p}, \mathbf{b}) = -\sum_{i=1}^{M^J} b_i \log(p_i), \tag{16}$$

where function $L(\cdot)$ is the well-known cross-entropy loss, $\mathbf{p} = [p_1, \ldots, p_{M^J}]^T$ is the output of GNNs' softmax layer, and $\mathbf{b}$ represents the corresponding one-hot label of the index allocated by the class_to_idx function [16].

After successfully categorizing the K-point polyline graph, user j may determine the associated TSC since there is a one-to-one correlation between the K-point polyline graph and the TSC in categories. User j is able to recreate his own original $\log_2(M)$ binary bits $\mathbf{b}_j^{n_t}$ after masking the transmission symbols of other users in the known TSC.

3.2.4 Model Configuration

In this letter, we consider a basic downlink MIMO-SCMA system model with $J = 6$, $K = 4$, and $M = 4$. Regarding the number of antennas, we also consider both cases of $n_t = n_r = 2$ and $n_t = n_r = 4$ simultaneously. The components of the channel gain vector $\mathbf{h}_j^{n_r,n_t}$ are modeled as independently and identically distributed (i.i.d.) complex Gaussian random variables with zero mean and unit variance. Each method in this letter is using the same codebook provided by [17]. The whole K-point polyline graph set has 2,048,000 samples. The batch size is set to 64, and the number of iterations is set to 400. In order to minimize the loss function in Eqs. (16), we adopt stochastic gradient descent (SGD) optimizer [18], in which the learning rate is set as 0.002 and the momentum

Table 1 Structure of modified version of the MobileNet model.

Type / Stride	Filter Shape	Input Size
Conv / s2	3×3×3×8	32×32×3
Conv dw / s1	3×3×8 dw	16×16×8
Conv / s1	1×1×8×8	16×16×8
Conv dw / s2	3×3×8 dw	16×16×8
Conv / s1	1×1×8×8	8×8×8
Conv dw / s1	3×3×8 dw	8×8×8
Conv / s1	1×1×8×16	8×8×8
Conv dw / s1	3×3×16 dw	8×8×16
Conv / s1	1×1×16×16	8×8×16
Conv dw / s1	3×3×16 dw	8×8×16
Conv / s1	1×1×16×16	8×8×16
Conv dw / s1	3×3×16 dw	8×8×16
Conv / s1	1×1×16×32	8×8×16
Conv dw / s2	3×3×32 dw	8×8×32
Conv / s1	1×1×32×32	4×4×32
Avg Pool / s1	Pool 4×4	4×4×32
FC / s1	32×4096	1×1×32
Softmax / s1	Classifier	1×1×4096

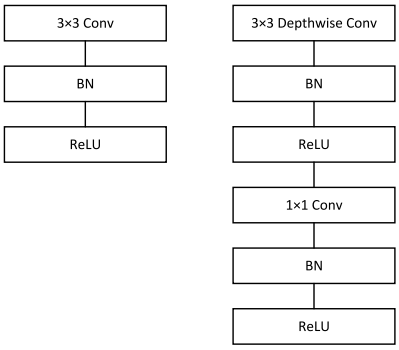


Fig. 4 Left: standard convolutional layer with batchnorm and ReLU. Right: depthwise Separable convolutions with depthwise and pointwise layers followed by batchnorm and ReLU.

is set as 0.9. The colormap utilized in the experiment is jet [16], and the value of RF is determined in Sect. 4.2.

A modified MobileNet [15] model has been adopted as the GNN, and its architecture is shown in Table 1. In the structure of the GNN, apart from the first layer of convolutional layers which is a full convolutional, all other convolutional layers are depthwise separable convolutional layers. The distinction between depthwise separable convolutional layers and standard convolutional layers is illustrated in Fig. 4.

4. Analysis of Simulation Results

In Sect. 4.1, we identify the optimal E_b/N_0 value for LG-ESDS training. In Sect. 4.2, we determine the optimum value of RF for generating the K-point polyline graphs. In Sect. 4.3, we compare the performance of our LG-ESDS with the conventional separated detection algorithm and the joint detection scheme on BER over different MIMO channel configurations. In Sect. 4.4, We evaluate the computational complexity of our LG-ESDS, along with the conventional separated detection algorithm and the joint detection schemes over the 2×2 MIMO channel.

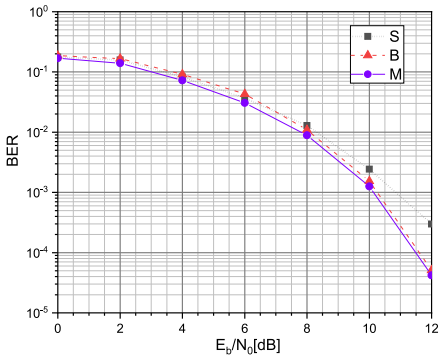


Fig. 5 Find the optimal E_b/N_0 value for the training of LG-ESDS over the 2×2 MIMO channel.

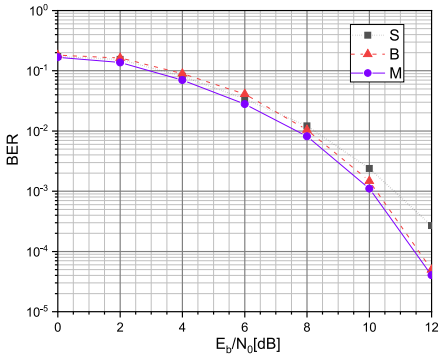


Fig. 6 Find the optimal E_b/N_0 value for the training of LG-ESDS over the 4×4 MIMO channel.

4.1 Choice of the Optimal E_b/N_0 Value

Figure 5 and Fig. 6 show the BER performance of the LG-ESDS over the 2×2 and 4×4 MIMO channels respectively, after it has been trained using each of the aforementioned scenarios. During the experiment, the value of RF is temporarily set as 5. The simulation results demonstrate that, in comparison to the alternative scenarios, M emerges as the optimal training strategy. Therefore the E_b/N_0 value for training is set as 8 dB in the rest of this work.

4.2 Determination of the Optimum Value of RF

We run simulations for each of the following three scenarios to determine the optimum value of RF, over the 2×2 and 4×4 MIMO channels respectively.

- RFS: set the value of RF as 4
- RFM: set the value of RF as 5
- RFB: set the value of RF as 6

As shown in Fig. 7 and Fig. 8, the simulation achieves the greatest outcomes in scenario RFM. Therefore, in LG-ESDS, the value of RF is set as 5.

4.3 BER Comparison

Figure 9 and Fig. 10 compare the BER performance of our

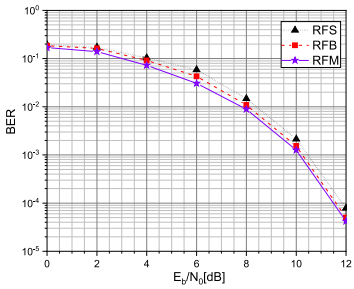


Fig. 7 Determine the optimum value of RF over the 2×2 MIMO channel.

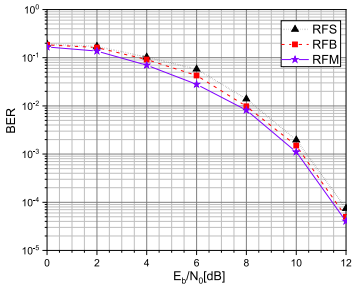


Fig. 8 Determine the optimum value of RF over the 4×4 MIMO channel.

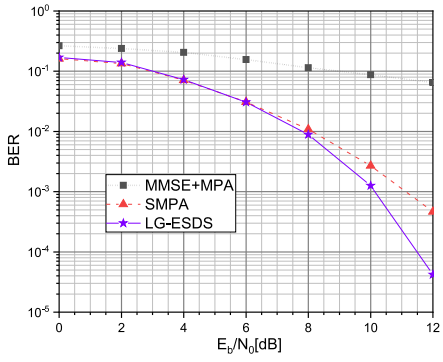


Fig. 9 BER comparison of MMSE+MPA, SMPA and LG-ESDS over the 2×2 MIMO channel.

LG-ESDS with the conventional separated detection algorithm (MMSE+MPA) and the joint detection scheme (Serial Schedule strategy based MPA (SMPA) [10]) over the 2×2 and 4×4 MIMO channels respectively. Our LG-ESDS consistently outperforms MMSE+MPA (8 iterations) across different E_b/N_0 values, and also achieves lower BER than SMPA (5 iterations) at high E_b/N_0 values. It is noteworthy that when the value of E_b/N_0 exceeds 6 dB, as E_b/N_0 increases, the BER performance of LG-ESDS compared to SMPA becomes more significant. This can be explained that compared to other MIMO-SCMA decoding strategies, our LG-ESDS does not require continuous message exchange between R nodes and U nodes, it can directly utilize the effective information of the estimated signals. Furthermore, our LG-ESDS exploits more features (the trend features of the K-point polyline graphs), which is beneficial for image classification.

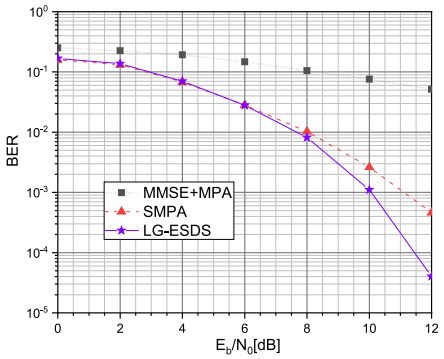


Fig. 10 BER comparison of MMSE+MPA, SMPA and LG-ESDS over the 4×4 MIMO channel.

4.4 Complexity Analysis

The computational cost of conventional convolution can be expressed as follows:

$$K_s \cdot K_s \cdot N_{in} \cdot N_{out} \cdot D \cdot D, \tag{17}$$

where $K_s \times K_s$ is the kernel size, N_{in} denotes the number of input channels, N_{out} represents the number of output channels, and $D \times D$ is the output feature map size. The computational cost of depthwise separable convolution can be expressed as follows:

$$K_s \cdot K_s \cdot N_{in} \cdot D \times D + N_{in} \cdot N_{out} \cdot D \times D \tag{18}$$

And the computational cost of fully connected layer can be expressed as follows:

$$N_n \cdot N_c, \tag{19}$$

where N_n is the number of neurons in the fully connected layer, and N_c denotes the number of neurons in the output layer.

The computational complexity of the MMSE detection and the generation of the K-point polyline graph can be considered negligible compared to the computational complexity of the convolutional computation. Based on Eqs. (17), Eqs. (18), and Eqs. (19), we calculate the computational complexity of our proposed LG-ESDS and compare it with MMSE+MPA (8 iterations), SMPA (5 iterations), and improved maximum distance MPA (IMDMPA (3 iterations)) [11], as illustrated in Fig. 11. Our LG-ESDS exhibits lower computational complexity compared to the other three decoding strategies. It should be mentioned that in order to reach performance convergence, different decoding strategies may require varying numbers of iterations. In order to ensure fairness in our comparison of computational complexity, we fix the number of iterations for each strategy to be the bare minimum needed to achieve performance convergence. Although our adopted GNNs in comparison to the original version of MobileNet has undergone significant simplifications, the extreme similarity in the distribution of the same type of K-point polyline graph features ensures that

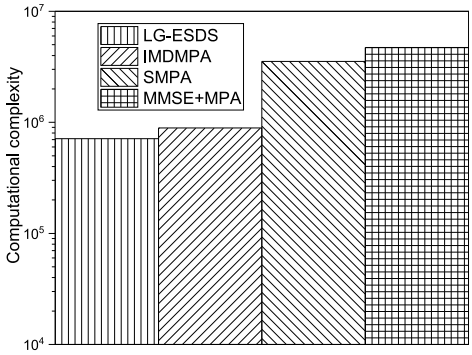


Fig. 11 Computational complexity comparison of MMSE+MPA, SMPA, IMDMPA and LG-ESDS over the 2×2 MIMO channel.

our LG-ESDS can achieve the decoding performance illustrated in Sect. 4.3. It is worth mentioning that as the number of antennas increases in the MIMO-SCMA system, the computational complexity of various decoding algorithms also significantly increases. In such cases, our LG-ESDS exhibits even more pronounced advantages over other decoding algorithms in terms of computational complexity.

5. Conclusion

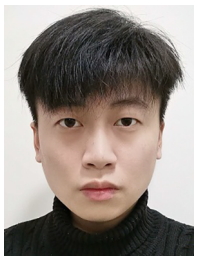
We have proposed a lightweight GNNs based enhanced separated detection scheme to accomplish the multi-user detection tasks at the receiver of the downlink MIMO-SCMA systems. We have raised the concepts of coordinate point relay and full-category training, which replacing the MPA in SCMA decoding with GNNs based image classification technology and offering a more methodical and rigorous experimental approach. Our LG-ESDS achieves a balance between computational complexity and BER performance, outperforming other decoding algorithms.

Acknowledgments

This work was partially supported by the Natural Science Foundation of Xinjiang Uygur Autonomous Region under Grant 2022D01C426 and supported by the Graduate Research Innovation Project of Xinjiang Uygur Autonomous Region under Grant XJ2023G100.

References

[1] Y.-P.E. Wang, X. Lin, A. Adhikary, A. Grovlen, Y. Sui, Y. Blankenship, J. Bergman, and H.S. Razaghi, "A primer on 3GPP narrowband Internet of Things," IEEE Commun. Mag., vol.55, no.3, pp.117–123, March 2017.
[2] D.W.K. Ng and R. Schober, "Resource allocation and scheduling in-multicell OFDMA systems with decode-and-forward relaying," IEEE Trans. Wireless Commun., vol.10, no.7, pp.2246–2258, July 2011.
[3] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-lin, and Z. Wang, "Nonorthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," IEEE Commun. Mag., vol.53, no.9, pp.74–81, Sept. 2015.
[4] H. Nikopour and H. Baligh, "Sparse code multiple access," Proc. IEEE 24th Annu. Int. Symp. Pers. Indoor Mobile Radio Commun., London, U.K., pp.332–336, 2013.
[5] F. Rusek, D. Persson, B.K. Lau, E.G. Larsson, T.L. Marzetta, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," IEEE Signal Process. Mag., vol.30, no.1, pp.40–60, Jan. 2013.
[6] M. Moltafet, S. Parsaeefard, M.R. Javan, and N. Mokari, "Robust radio resource allocation in MISO-SCMA assisted C-RAN in 5G networks," IEEE Trans. Veh. Technol., vol.68, no.6, pp.5758–5768, June 2019.
[7] S. Qian, W. Ge, Y. Zhang, and P. Zhang, "A large-scale SCMA codebook optimization and codeword allocation method," IEICE Trans. Commun., vol.E105-B, no.7, pp.788–796, July 2022.
[8] S.S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," IEEE Commun. Surveys Tuts., vol.17, no.4, pp.1941–1988, 4th Quart., 2015.
[9] L. Yang, Y. Liu, and Y. Siu, "Low complexity message passing algorithm for SCMA system," IEEE Commun. Lett., vol.20, no.12, pp.2466–2469, 2016.
[10] Y. Du, B. Dong, Z. Chen, P. Gao, and J. Fang, "Joint sparse graph-detector design for downlink MIMO-SCMA systems," IEEE Wireless Commun. Lett., vol.6, no.1, pp.14–17, Feb. 2017.
[11] H. Cheng, C. Zhang, Y. Huang, and L. Yang, "Efficient message passing receivers for downlink MIMO-SCMA systems," IEEE Trans. Veh. Technol., vol.71, no.5, pp.5073–5086, May 2022.
[12] M. Kim, N.-I. Kim, W. Lee, and D.-H. Cho, "Deep learning-aided SCMA," IEEE Commun. Lett., vol.22, no.4, pp.720–723, 2018.
[13] Q. Yuan, Z. Wang, D. Li, Q. Guo, and W. Xiang, "Sparse code multiple access scheme based on variational learning," IEICE Trans. Commun.,, vol.70, no.12, pp.7989–8002, Dec. 2022.
[14] S. Dörner, S. Cammerer, J. Hoydis, and S. ten Brink, "Deep learning based communication over the air," IEEE J. Sel. Topics Signal Process., vol.12, no.1, pp.132–143, Feb. 2017.
[15] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W.J. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, [online] Available: https://arxiv.org/abs/1704.04861
[16] A. Paszke, et al., "PyTorch: An imperative style, high-performance deep learning library," Proc. Adv. Neural Inf. Process. Syst., 2019, pp.8026–8037.
[17] Altera Innovate Asia FPGA Design Contest. (2015) 5G Algorithm Innovation Competition.
[18] S. Ruder, "An overview of gradient descent optimization algorithms," arXiv:1609.04747, 2016.



Zikang Chen received the B.S. degree in Communication Engineering at China University of Geosciences (Wuhan) in 2020. At present, he majors in Information and Communication Engineering at Xinjiang University for M.S. degree.

Wenping Ge received the Ph.D. degree in Electromagnetic and Microwave Technology at Shanghai Jiaotong University in 2003. Her research interests include wireless communication, optical crystal fiber, and optoelectronic materials. She is full professor in College of Information Science and Technology, Xinjiang University. Her current research interests include sparse code multiple access and codebook design in wireless communications.

Henghai Fei received the B.S. degree in Communication Engineering at Jilin University in 2020. At present, he majors in Information and Communication Engineering at Xinjiang University for M.S. degree.

Haipeng Zhao received the B.S. degree in Communication Engineering at Jinan University in 2020. At present, he majors in Electronic and Communication Engineering at Xinjiang University for M.E. degree.

Bowen Li received the B.S. degree in Communication Engineering at Wuhan Textile University in 2021. At present, he majors in Electronic and Communication Engineering at Xinjiang University for M.E. degree.