M-ary Phase Position Shift Keying Demodulation Using Stacked Denoising Sparse Autoencoders

Conghui Lu 1,2,*, Peng Chen 2, Hua Zhong 1 and Mengyuan Wang 1

1 School of Communication Engineering, Hangzhou Dianzi University, Xiasha Higher Education Zone, Hangzhou 310018, China; hzhong@hdu.edu.cn (H.Z.); wmy182080133@hdu.edu.cn (M.W.)
2 State Key Laboratory of Millimeter Waves, Southeast University, Nanjing 210096, China; chenpengseu@seu.edu.cn
* Correspondence: luch@hdu.edu.cn

Abstract: A deep-learning based detector for M-ary phase position shift keying (MPPSK) systems is proposed in this paper. The major components of this detector include a special impact filter, a stacked denoising sparse autoencoder (DSAE), which was trained in unsupervised learning to extract features from the modulation signals, and a softmax classifier. The features learned by the stacked DSAE were then used to train the softmax classifier to demodulate the received signals into M classes. The architecture presented herein was trained and tested on a simple dataset extended by adding Gaussian noise only. The results from the theoretical analysis and simulation show that the detection performance of the proposed scheme is superior to that of existing detectors.

Keywords: improved M-ary phase position shift keying (MPPSK); deep learning; autoencoders; demodulation

1. Introduction

As wireless applications continue to grow, spectrally efficient modulation schemes will continue to be highly valuable [1–3]. The classical binary modulation schemes such as amplitude shift keying (ASK), frequency shift keying (FSK) and phase-shift keying (PSK, or BPSK) are spectrally inefficient. In the past two decades, a number of bandwidth efficient modulation techniques have been proposed. Wu et al. proposed an extended binary phase shift keying (EBPSK) scheme [4] which is much more flexible and has a more compact spectral structure than conventional BPSK modulation. It is shown that, by tuning the modulation parameters, EBPSK could achieve the same bit error rate (BER) performance as the conventional BPSK, but has higher spectrum utilization efficiency than the latter for wireless applications. EBPSK is also extended to M-ary phase position shift keying (MPPSK) [5] to achieve higher transmission rates. An improved MPPSK modulation with a simpler demodulator and a better performance than the original MPPSK scheme is presented [6].

Existing demodulators of MPPSK systems use back propagation (BP) [7], support vector machine (SVM) [8], joint amplitude and waveforms detection (JAWD) [9], and the phase locked loop (PLL) techniques [5]. However, the symbol error rate (SER) performance of MPPSK signaling could be further improved by using more advanced detection techniques. The demodulator for MPPSK signaling described in [5] uses a special impact filter (SIF) which could have a strong impact on the signal amplitude at the phase transition instants to significantly improve the output signal-to-noise ratio (SNR).

Deep learning has been successfully applied to a variety of problems [10–12]. The success of deep learning algorithms generally depends on how data are represented. For communication signals such as MPPSK, since the SIF affects the signal amplitude more appreciably at the phase transition points, the impacted amplitudes of MPPSK waveforms
contain abundant features. The features can be learned and used as classifiers to significantly improve detection performance. To date and to our knowledge, there is little research on unsupervised learning for detection in communication systems. Deep learning models include supervised models and unsupervised models. Compared with labeled data in supervised models, unlabeled data in the unsupervised models are easier to obtain when large-scale data are needed. Unsupervised feature learning consists of several basic modules, such as restricted Boltzmann machine (RBM) [13], autoencoder (AE) [10] and sparse coding [11]. It has been employed in the fields of computer visual perception, speech recognition, natural language processing, etc. Feature learning of unsupervised models has a great potential to improve the detection performance in communication systems [13–15]. For MPPSK signals, the AE algorithm could be employed to find the correlation characteristics among the input signals.

In this paper, we developed a detector for MPPSK systems based on the AE learning algorithm, improving the detection performance by learning the useful representations or features. Each possible transmission scenario is considered as a sub-model of deep learning, and a large number of cases are included in its database. We also commonly use training data from the database that contain samples of the modulation signals. The similarity between the training and test data is built by a feature representation to exploit. AE learning networks is categorized into sparse autoencoders (SAEs) and denoising autoencoders (DAEs). The SAE and DAE networks [16–18] are combined to form the stacked denoising sparse autoencoders (DSAEs) network, further improving the detection performance of MPPSK signals.

The remainder of this paper is organized as follows. Section 2 describes the improved MPPSK modulation and the novel demodulator based on the stacked DSAE network. Section 3 describes the principle of the proposed stacked DSAE network. The SER and computational complexity of the proposed MPPSK systems are obtained in Section 4. The database and the experimental analysis and results are proposed in Section 5. Conclusions are given in Section 6.

2. The Improved MPPSK Signaling

The demodulator of MPPSK systems uses a narrowband SIF [5] which can produce a strong impact at the phase jumping point, resulting in significant improvement in the output SNR. The SIF is followed by a threshold decision device or a geometric characteristic-based nonlinear decision device [19,20]. There are many variations of the demodulator [21] and they all require a specific threshold to distinguish zero and non-zero symbols. The position information is utilized to decode the non-zero symbols. The threshold is critical for these demodulators to function properly, and it will cause a performance degradation in practice. An issue with the MPPSK signal is that there is no phase jump for an incoming symbol “0”. Consequently, the SIF response to the symbol “0” cannot produce a similar change on the amplitude as the non-zero symbols. Therefore, the detection between the zero and non-zero symbols must exploit the amplitude information.

The improved MPPSK signaling in [6] allows demodulation by only using the position information of the received signal passed through the SIF. The basic idea is to eliminate the symbol “0”. That is, the alphabet is changed from [0, 1, 2, …, M−1] to [1, 2, …, M]. In this improved MPPSK signaling scheme, over the time period [0, T], the modulation signal is expressed as

\[
g_k(t) = \begin{cases} 
\sin 2\pi f_c t & 0 \leq t < (k - 1)(K + r_g) T, \\
\sin 2\pi f_c t + \theta, (k - 1)(K + r_g) T \leq t \leq [(k - 1)(K + r_g) + K] T, \\
\sin 2\pi f_c t & [(k - 1)(K + r_g) + K] T \leq t < NT,
\end{cases}
\] (1)

where \( T \) is the period of the carrier signal, \( 1 \leq k \leq M \), \( f_c \) is the carrier frequency, \( T_s = NT = N/f_c \) is the symbol period, \( N \) is the ratio of the symbol period to the carrier period, \( K \) is the number of carrier periods with a phase jump, \( \theta \) is the modulation angle,
0 ≤ θ ≤ π, and \( r_g \) is the control factor for the symbol protection interval. For simplicity, set \( \theta = \pi \) in this paper.

The improved MPPSK modulation given in Equation (1) is not strictly bandwidth-limited signals, since information is embedded into the position of a full carrier signal cycle. Furthermore, each symbol period consists of \( N \) cycles of the carrier signal. Many BER bounds are obtained from information theory, assuming strictly bandwidth-limited signals designed according to the Nyquist criterion for zero inter-symbol interference (ISI). It might not be directly applicable. For the improved MPPSK signaling, the relationship between SNR and \( E_b/N_0 \) is

\[
\text{SNR} = \frac{E_b}{N_0} \times \left( \frac{R_b}{W} \right) = \frac{E_b}{N_0} \times \left[ \frac{(2f_c \log_2 M)}{f_s N} \right],
\]

where \( R_b \) is a symbol rate, \( f_s \) is sampling rate, and \( W = f_s/2 \) is the effective noise bandwidth, although the signal is not strictly bandlimited. \( E_b \) is the received energy per bit, \( N_0 \) is the single-sided noise power spectral density.

The existing MPPSK demodulators include the decision threshold-based approach [5], the adaptive threshold integral discriminator [20], the coherent processing technique [21], the phase locked loop (PLL) scheme [22] and the matched filtering scheme [23, 24]. As shown in Figure 1, we develop a new demodulation scheme for the improved MPPSK modulation aiming to significantly improve detection performance. In the proposed system, there are the forward error correction (FEC) encoding before improved MPPSK modulation, and the FEC decoding after stacked DSAE demodulator. The stacked DSAE network is used for automatic feature extraction and detection. The architecture is first trained using unlabeled improved MPPSK signals. There are two detection schemes. One is the SIF directly followed by the stacked DSAE. Another is SIF, a coherent processing block and a low pass filter (LPF) followed by the stacked DSAE in the receiver.

![Figure 1. Block diagram of the stacked DSAE demodulation system.](image)

3. Stacked DSAE Network

The stacked autoencoder network is composed of multiple layers of sparse autoencoders. Furthermore, the output of the previous layer of autoencoders is used as the input of the latter layer of autoencoders. The weights of the existing neural networks are random initialization. However, in the stacked autoencoder network, the weights are learned by the autoencoder instead of that in the traditional neural network as shown in Figure 2. Furthermore, the noise is added to the training database to improve the generalization ability and robustness. Before training, the noised modulation signal is down-sampled and converted without a tag, which forms a part of the DSAE network database. After training, the weight matrices, which are learned by the network, are used to transform the data. The transformed data are used as features to train the softmax classifier. In the testing phase, the raw test data are transformed by the stacked DSAE network, and the output is fed into the softmax classifier for detection.
Figure 2. The activations of previous AEs are used to train the latter AE.

The input is the noised improved MPPSK vector \( X \) and the neurons are arranged in a column vector \( h^{(2)} = [h_1^{(2)}, h_2^{(2)}, \cdots, h_e^{(2)}]^T \) where the superscript \([\cdot]^T\) denotes transposition. It can be written in vector-matrix form as

\[
h^{(2)} = W^{(1)}X + b^{(1)},
\]

where the matrix \( W \) is weights and the vector \( b^{(1)} \) is the bias in the input layer. The network has two nonlinear functions. One is in the second layer and the other one is in the output layer. This allows the network to capture higher order statistics from the input data, leading to a higher kurtosis in the coded representation.

In the proposed stacked DSAE network, the \( W^{(1,1)} \) learned by the first autoencoder is used to link the input layer and the first hidden layer. Furthermore, the \( W^{(1,2)} \) learned by the second autoencoder is used to link the first hidden layer and the second hidden layer. Furthermore, the weights are stacked on the top of the first autoencoder, as shown in Figure 3. The encoding process of the stacked DSAE network is used to execute the coding steps of each layer of autoencoders from front to back. It can be expressed as

\[
a^{(l)} = f(z^{(l)})
\]

\[
z^{(l+1)} = W^{(l,2)}a^{(l)} + b^{(l,2)}
\]

Similarly, the decoding process is used to perform the decoding steps of each layer of autoencoders from back to front. It can be represented as

\[
a^{(n+l)} = f(z^{(n+l)})
\]

\[
z^{(n+l+1)} = W^{(n-l,2)}a^{(n+l)} + b^{(n-1,2)}
\]

where \( a^{(n)} \) is the activation value of the deepest hidden unit. It contains the information of our interest, and it is also a higher-order representation of the input value.
The proposed network architecture consists of 3 stacked DSAEs and a softmax layer. The cost function of the stacked DSAE network is given by

$$J(W, b) = \frac{1}{2m} \sum_{i=1}^{m} \left\| X^{(i)} - X^{(j)} \right\|^2 + \frac{\lambda}{2} \| W \|^2 + \beta \sum_{j=1}^{L} KL\left( \rho \parallel \hat{\rho}_j \right),$$

where $m$ is the numbers of input vectors, $\rho$ is a sparsity parameter as well as a hyper-parameter and $\lambda$ is a parameter that controls the amount of regularization. $\beta$ is a weight to control the sparsity penalty factor. The Kullback–Leibler (KL) divergence penalty is

$$KL\left( \rho \parallel \hat{\rho}_j \right) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j}. \quad (7)$$

$\hat{\rho}_j$ is the average activation of the $j$th hidden unit over the training set. The weights and biases in the input layer are updated by gradient descent.

The input features from the stacked DSAE can be used for detecting the improved MPPSK modulation signals. It is fed by the output $a^{(l)} = f \left( h^{(l)} \right)$ to a softmax classifier, and $f(\cdot)$ is a sigmoid function. Thus, the originally generated layers can be used for feature extraction. When the dataset is sent into the stacked DSAE network, every hidden layer transforms the feature data until the last hidden layer. Then, the last set of the extracted feature is fed into the softmax layer for detection. After the construction of the whole network is complete, the cost function is finally optimized by the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm.
4. SER and Computational Complexity

(1) The improved MPPSK signal can be expressed as

\[
s_1 = \sqrt{E_b}(-1, 1, 1, \ldots, 1)^T \in \mathbb{R}^{M \times 1}
\]

\[
s_2 = \sqrt{E_b}(1, -1, 1, \ldots, 1)^T \in \mathbb{R}^{M \times 1}
\]

\[
\vdots
\]

\[
s_M = \sqrt{E_b}(1, 1, 1, \ldots, -1)^T \in \mathbb{R}^{M \times 1}
\]

(8)

The probability that symbol \( p \) is erroneously detected as symbol \( q \) can be given as

\[
P_{p,q} = Q\left(\sqrt{\frac{2Nf_sSNR}{f_c}}\right),
\]

where \( Q(\cdot) \) is the Marcum Q-function. Thus, the SER of the improved MPPSK modulation is

\[
P_e = 1 - (1 - P_{p,q})^{M-1} \simeq (M - 1)Q\left(\sqrt{\frac{2Nf_sSNR}{f_c}}\right).
\]

(10)

The SER of the improved MPPSK modulation is lower than that of the existing MPPSK in [20].

(2) The computational complexity of the stacked DSAE algorithm is given by

\[
O(KX_mZ_{m3}C_{in}C_{out}),
\]

(11)

where \( C_{in} \) and \( C_{out} \) represent the input size and output size of the stacked DSAE network, respectively. \( X_m Z_{m3} \) denotes the parameter dimension in the input layer. In the training stage, it is more efficient with a much shorter training time compared to the conventional decision threshold-based approach [5], the adaptive threshold integral discriminator [20], coherent processing technique [21], the PLL scheme [22], etc., especially in the unsupervised training stage.

5. Simulation and Analysis

5.1. Database

The database used in this paper contains the training set and the test set. When the database is used to detect the 2PPSK modulation signal, the training set contains 2 classes of 60,000 sampled waveforms, and the test set contains two classes of 10,000 sampled noised waveforms at several SNR values. For example, as shown in Figure 4, there are symbol “1”, symbol “2”, and \(-3\)dB noised symbol “2”, separately. The main parameters of the improved MPPSK system are \( M = 2, K = 30, N = 105, f_c = 21 \times 10^6 \) Hz, and a sampling rate of \( f_s = 210 \times 10^6 \) Hz. When the database was used to detect the improved 4PPSK modulation signal, the training set contained four classes of 60,000 sampled waveforms, and the test set contained four classes of 100,000 sampled noised waveforms at several SNR values.
5.2. Architecture

In order to obtain a high detection performance, the number of nodes in the hidden layers may be more important than the learning algorithm chosen or the depth of the model. Considering this, we adopted different architectures corresponding to the different MPPSK parameters and then determine the appropriate numbers of the corresponding hidden units. To simplify the problem and focus on the effect of the generated nodes, the stacked architecture is used without pooling and convolution.

The stacked DSAE model proposed for the improved 2PPSK signaling is expressed in forms such as 1050-800-400-200-2, where ‘1050’ indicates 1050 nodes in the input layer; ‘800’ indicates 800 nodes in the first hidden layer; ‘400’ indicates 400 nodes in the second hidden layer; ‘200’ indicates 200 nodes in the third hidden layer; and ‘2’ indicates 2 nodes in the output layer. A 1050-800-400-200-2 network structure means that we train an SAE network of 1050-800-1050 first, then feed forward the input data to the 1050-800 network to obtain the first set of hidden features, and finally train another 800-400-800 network, using 1050-800-400 to obtain the second set of hidden features. After this, we train the 400-200-400 network and use 1050-800-400-200 to obtain the third set of hidden features. After the first SAE network is trained, the activation value of all nodes in the hidden layer is calculated and sent to the second autoencoder as its input. The activation value of each hidden layer is regarded as the extracted features of the original modulation signals. Then, the hierarchical representations of original signals are obtained. Such representations sent to a softmax classifier are effective for detection. Thus, we use softmax to train the 200-2 network, and finally, we cascade all the networks to form the set as 1050-800-400-200-2.

5.3. DSAE Training Optimization

During the pre-training process described above, the parameters of other layers remain fixed when the parameters of one layer are determined. Furthermore, we repeatedly train all unlabeled sample data with randomly initialized weights to minimize the cost function value. Then, the L-BFGS algorithm is employed to solve the non-convex optimization problem of the cost function. For better results, the whole network is fine-tuned using back-propagation and the DAE algorithm and the parameters of the whole network are adjusted.

In order to reduce the training time, we divided all the training samples into many batches, each having 100 samples, and trained with each batch. It was applied to take advantage of the efficient computational techniques available in software and hardware such as matrix multiplication in Matlab.

Figure 4. The improved 2PPSK illustrations in the database.
5.4. Results

5.4.1. Parameters of the Stacked DSAE Network

Using the proposed DSAE model with noisy and noiseless training data, we obtained the detection performance for different numbers of nodes in the hidden layers and different ratios of training and testing sets. The most appropriate ratios of training and testing sets were selected, as well as the stacked network set of the hidden layers.

For 2PPSK, \( M = 2 \), the ratio of training and testing set is 60,000:10,000 and the corresponding number of nodes in the input layer is 1050 with 2 of these nodes in the output layer. For 4PPSK, \( M = 4 \); the ratio of training and testing set is set to 60,000:100,000, and the corresponding number of nodes in the input layer is 100 with 4 of these nodes in the output layer.

The detection performances are compared, assuming different numbers of nodes in the hidden layers. The training datasets are noisy and noiseless, as shown in Tables 1 and 2. \( H_1 \) is the first hidden layer; \( H_2 \) is the second hidden layer; \( H_3 \) is the second hidden layer; ‘−19’ means the channel SNR = −19 dB; ‘(−18)−19’ means the training data SNR = −18 dB with channel SNR = −19 dB; and ‘(−3)−19’ means training data SNR = −3 dB with channel SNR = −19 dB. The main parameters of the improved MPPSK system are \( M = 2; K = 30; N = 105; f_c = 21 \times 10^6 \text{ Hz} \); sampling rate \( f_s = 210 \times 10^6 \text{ Hz} \); and the ratio of training and testing set is 60,000:10,000. The receiver consists of the SIF followed by the stacked DSAE network.

Table 1. SER comparison based on two hidden layers in the stacked DSAE network.

| H1-H2 | SNR (dB) | −19 | (−18)−19 | −3−19 |
|-------|----------|-----|----------|-------|
| 200-50 | 0.0050   | 0.0062 | 0.0051 |
| 200-100 | 0.0045   | 0.0052 | 0.0047 |
| 300-100 | 0.0043   | 0.0062 | 0.0047 |
| 300-200 | 0.0052   | 0.0038 | 0.0045 |
| 400-200 | 0.0039   | 0.0062 | 0.0044 |

Table 2. SER comparison based on three hidden layers in the stacked DSAE network.

| H1-H2-H3 | SNR (dB) | −19 | (−18)−19 | −3−19 |
|----------|----------|-----|----------|-------|
| 400-200-100 | 0.0045   | 0.4992 | 0.0037 |
| 500-200-100 | 0.0044   | 0.0041 | 0.0038 |
| 600-300-150 | 0.0037   | 0.5008 | 0.0034 |
| 800-400-200 | 0.0046   | 0.0042 | 0.0039 |

As shown in Tables 1 and 2, the SER values based on training data with \( \text{SNR} = \left( \frac{E_b}{N_0} \right) \times \left( \frac{R_b}{W} \right) = −3 \text{ dB} \) are better than that based on the training dataset of −18 dB, the case shown as ‘−3’, ‘−18’. Given this, all experiments below will use the −3dB noised training data. It is also observed that the detection performance with three hidden layers is better than that with two layers. Therefore, we adopt the combination of hidden layers 600-300-150. This means 600 nodes in the first hidden layer, 300 in the second hidden layer and 150 nodes in the third hidden layer.

5.4.2. SER Performance of the Improved 2PPSK

The main system parameters are \( M = 2; K = 30; N = 105; f_c = 21 \times 10^6 \text{ Hz} \); and a sample rate \( f_s = 210 \times 10^6 \text{ Hz} \). Corresponding to the set of system parameters, the ratio of the training and testing set is 60,000:10,000, and the structure of stacked DSAE network is 1050-600-300-150-2. There are five layers in the network, with 1050 nodes in the input layer; 600 nodes in the first hidden layer; 300 nodes in the second hidden layer; 150 nodes in the third hidden layer; and 2 nodes in the output layer.
The SER performance of the improved 2PPSK signaling is shown in Figure 5 where the receiver consists of one SIF followed by the proposed stacked DSAE network. The simulation curves are close to the theoretical curve. The SER simulation values are slightly better than the theoretical values between $-17$ dB and $-18$ dB SNR. This is because the approximate processing is introduced in the calculation of theoretical values in Equation (10).

This seems to indicate that the result violates the Shannon limit on the SNR per bit, assuming an unbounded bandwidth. Furthermore, $-1.6$ dB SNR with an arbitrarily low error rate is achievable. The result is close to the Shannon limit, with the system parameters are included in the SNR calculation defined in Equation (2). The transmission signal based on the $-3$ dB noised training data could extract signal features that are closer to reality than the case based on noiseless training data. Thus, the SER performance of the proposed scheme with the case of $-3$ dB noised training data is slightly better than that with the case of noiseless training data.

Figure 5. SER of the stacked DSAE network for the improved 2PPSK system.

5.4.3. SER Performance of Improved 4PPSK

Compared with BP [7], SVM [8] and JAWD [9], the detection performance of the improved 4PPSK system with the SIF followed by the proposed stacked DSAE network is shown in Figure 6. The system parameters are $M = 4$; $K = 2$; $N = 10$; $f_c = 10 \times 10^6$ Hz; and a sample rate $f_s = 100 \times 10^6$ Hz, and the training database is noise-free. Corresponding to the set of system parameters, the ratio of the training and testing sets is 60,000:100,000, and the structure of stacked DSAE network is 100-60-20-4. There are four layers in the network, with 100 nodes in the input layer; 60 nodes in the first hidden layer; 20 nodes in the second hidden layer; and 4 nodes in the output layer.
As shown in Figure 6, the SER of the proposed scheme based on the improved MPPSK modulation in Equation (1) is better than that based on the existing MPPSK [5,6]. It is consistent with the expected results. Furthermore, the proposed scheme performs approximately 3 dB better than the SVM at the SER of $10^{-4}$.

We also obtained the performance of the structure that employs the SIF, a coherent processing block, and LPF followed by the proposed stacked DSAE network in the receiver. The major system parameters are the same as above, except that the training database contains both noiseless and $-3$ dB noised datasets. The performances of the BP, SVM, JAWD detectors, and this stacked DSAE network are shown in Figure 6, separately.

The stacked DSAE network, based on both the noisy training database and noiseless training database, performs the best in Figure 7. In the proposed scheme, the performance with $-3$ dB noised training database is approximately 5 dB better than that with the noiseless training database at the SER of $10^{-4}$. This is consistent with the results in Tables 1 and 2 and Figure 5. This is because, in the case of the $-3$ dB, noised training data can be extracted with more signal features than with the noiseless training data.

Compared with the SER performance of the 4PPSK system with only SIF in Figure 6, the required SNR is reduced by approximately 2 dB at the SER of $10^{-4}$ in the structure with an SIF, a coherent processing block and an LPF in Figure 7.
Figuer 7. SER Comparisons of the improved 4PPSK system with various detection schemes.

6. Conclusions

The improved MPPSK modulation with a not strictly limited bandwidth is proposed in this paper. The stacked DSAE network demodulator was developed for the MPPSK signals. Given by the theoretical analysis and simulation, its proper structure and parameters were obtained. It was demonstrated that deep learning models can be utilized for data detection in communication systems. The proposed stacked DSAE network extracts features from the modulation signals and uses a softmax layer for detection. The results about the structure and system parameters are useful to generalize the features to much broader application scenarios, whose features cannot be directly derived from the training datasets. Furthermore, the technique developed in this paper results in a much better detection performance than existing detectors. We hope that this research will open up new frontiers in applying deep learning techniques for communication such as in detecting strictly band-limited signaling, which was designed by the Nyquist criterion to ensure ISI-free reception in Gaussian channels.

Author Contributions: C.L. conceived and wrote this paper; C.L. and P.C. designed the simulations and experiments; P.C. analyzed the data; H.Z. and M.W. contributed figures plotting. All authors have read and agreed to the published version of the manuscript.

Funding: This work was jointly supported by the Open Grant of State Key Laboratory of Millimeter Waves (No. K202111) and the Scientific Research Project of Zhejiang Provincial Department of Education, China (No. Y201737971).

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Yang, Y.; Dang, S.; Wen, M.; Guizani, M. Millimeter Wave MIMO-OFDM With Index Modulation: A Pareto Paradigm on Spectral-Energy Efficiency Trade-Off. *IEEE Trans. Wirel. Commun.* 2021, 20, 6371–6386. [CrossRef]
2. Salih, A.; Audah, L.; Abdullah, Q.; Abdullah, N.; Shah, N.S.M.; Saif, A. Trade-off Energy and Spectral Efficiency with Multi-Objective Optimization Problem in 5G Massive MIMO System. In Proceedings of the 2021 1st International Conference on Emerging Smart Technologies and Applications (eSmarTA), Sana’a, Yemen, 10–12 August 2021; pp.1–6.
3. Al-Obiedollah, H.M.; Cumanan, K.; Thiyagalingam, J.; Tang, J.; Burr, A.G.; Ding, Z.G.; Dobre, O.A. Spectral-Energy Efficiency Trade-Off-Based Beamforming Design for MISO Non-Orthogonal Multiple Access Systems. IEEE Trans. Wirel. Commun. 2020, 19, 6593–6606. [CrossRef]

4. Feng, M.; Wu, L.; Ding, J.; Qi, C. BER analysis and verification of EBPSK system in AWGN channel. IEICE Trans. Commun. 2011, 94, 806–809. [CrossRef]

5. Ying, P.; Wu, L. New scheme of MPPSK modem. J. Southeast Univ. (Nat. Sci. Ed.) 2012, 42, 204–208.

6. Lu, C.; Wu, L.; Chen, P.; Wang, J.; Liu, H. M-ary phase position shift keying with orthogonal signalling. IET Commun. 2015, 9, 1627–1634. [CrossRef]

7. Chen, P.; Cao, Z.; Chen, Z.; Wang, X. Off-Grid DOA Estimation Using Sparse Bayesian Learning in MIMO Radar with Unknown Mutual Coupling. IEEE Trans. Signal Process. 2019, 67, 208–220. [CrossRef]

8. Xu, H.; Wu, L. Application of LDPC code in signal detection of MPPSK based on similar-dichotomy SVM. J. Southeast Univ. (Nat. Sci. Ed.) 2014, 44, 468–471.

9. Shen, J.; Chen, X. Research on Detection and Channel Coding in MPPSK Communication System. Wirel. Commun. 2020, 10, 49–57.

10. Bengio, Y.; Aaron, C.; Pascal, V. Representation learning: A review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. 2013, 35, 1798–1828. [CrossRef]

11. Deng, L.; Xiao, L. Machine learning paradigms for speech recognition: An overview. IEEE Trans. Audio Speech Lang. Process. 2013, 21, 1060–1089. [CrossRef]

12. Miao, P.; Wu, L.; Chen, Z. An anti-noise modem for visible light communication systems using the improved M-ary position phase shift keying. AEU-Int. J. Electron. Commun. 2018, 85, 126–133. [CrossRef]

13. Hu, Z.; Liu, H. A low-complexity LDPC decoding algorithm for hierarchical broadcasting: design and implementation. IEEE Trans. Veh. Tech. 2013, 62, 1843–1849. [CrossRef]

14. Miao, P.; Chen, P.; Chen, Z. Low-Complexity PAPR Reduction Scheme Combining Multi-Band Hadamard Precoding and Clipping in OFDM-Based Optical Communications. Electronics 2018, 7, 11. [CrossRef]

15. Chen, P.; Chen, Z.; Cao, Z.; Wang, X. A New Atomic Norm for DOA Estimation With Gain-Phase Errors. IEEE Trans. Signal Process. 2020, 68, 4293–4306. [CrossRef]

16. Yuan, M.; Tang, H.; Li, H. Real-Time key point recognition using restricted boltzmann machine. IEEE Trans. Neural Netw. Learn. Syst. 2014, 25, 2119–2126. [CrossRef]

17. Liu, J.; Xu, L.; Xie, Y.; Ma, T.; Wang, J.; Tang, Z.; Gui, W.; Yin, H.; Jahanshahi, H. Toward Robust Fault Identification of Complex Industrial Processes Using Stacked Sparse-Denoising Autoencoder with Softmax Classifier. IEEE Trans. Cybern. 2021, 1–15. [CrossRef]

18. Lv, Y.; Duan, Y.; Kang, W.; Li, Z.; Wang, F.Y. Traffic flow prediction with big data: A deep learning approach. IEEE Trans. Intell. Transp. Syst. 2015, 16, 865–873. [CrossRef]

19. Zhao, C.; Li, C.; Feng, S.; Su, N. Spectral-Spatial Stacked Autoencoders Based on the Bilateral Filter for Hyperspectral Anomaly Detection. In Proceedings of the 2020 IEEE International Geoscience and Remote Sensing Symposium(IGARSS 2020),Virtual Symposium, Waikoloa, HI, USA, 26 September–2 October 2020. [CrossRef]

20. Feng, M.; Wu, L. Special non-linear filter and extension to Shannon’s channel capacity. Digit. Signal Process. 2009, 19, 861–873. [CrossRef]

21. Zhang, P.; Wu, L. A realization method for multi-carrier MPPSK modulations. Appl. Mech. Mater. 2014, 462, 615–618. [CrossRef]

22. Qi, C.; Wu, L. PLL demodulation technique for M-ray position phase shift keying. J. Electron. (China) 2009, 26, 289–295. [CrossRef]

23. Lu, C.; Wu, L.; Chen, P. New scheme of MPPSK modem. In Proceedings of the 2014 9th International Symposium on Communication Systems, Networks & Digital Sign (CSNDSP), Manchester, UK, 23–25 July 2014; pp. 564–568.

24. Chen, P.; Yang, Z.; Chen, Z.; Guo, Z. Reconfigurable Intelligent Surface Aided Sparse DOA Estimation Method with Non-ULA. IEEE Signal Process. Lett. 2021, 28, 2023–2027. [CrossRef]