A New Message Passing Algorithm Based on Sphere Decoding Improvement

Hongwei Zhao(✉), Yue Yan, and Zichun Zhang

School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China
13359237963@163.com

Abstract. In a multi-satellite measurement and control system, when multiple satellites transmit data to the ground station at the same time, the ground station needs to efficiently detect the data of each satellite. Large overload access leads to strong mutual interference of satellite signals, and non-positive signals based on sparse codes are used. Cross multiple access (SCMA) technology can provide access and distribution capabilities that exceed the limit of traditional channel capacity, and is suitable for a large number of constellation measurement and control systems. However, the traditional MPA algorithm is extremely complex and has a large processing delay. It has a great test of the complexity and real-time performance that needs to be considered in the application. Therefore, it is of great significance to study the low-complexity space borne reception detection algorithm. A new multi-user detection algorithm based on partial codeword marginalization and sphere decoding, which is called PMSD-MPA, is proposed in this paper. PMSD-MPA considers both the mapping between resource nodes and user nodes and reduce the number of iteration. Proved by simulation, the convergence rate of the iteration process is accelerated, thus reducing the computational complexity of the detection algorithm.

Keywords: SCMA · MPA · Sphere decoding

1 Introduction

The vision of mobile communication is to achieve "anytime, anywhere access to everything", so people put forward many new requirements for 5G system. Compared with 4G, the spectral efficiency of 5G system needs to be increased 5–15 times, the connection density needs to be increased more than 10 times, the end-to-end delay in the system needs to be reduced to the order of milliseconds, and the reliability needs to be close. The new demand of 5G brings new challenges to research and development in many aspects [1].

SCMA is a code-based non-orthogonal multi-access technology with nearly optimal spectrum efficiency, which is designed to meet the requirements of 5G. Overall, SCMA's overall performance is better than NOMA and MUSA's, and its complexity is lower than PDMA's, making it well suited to the 5G standard. Based on the design idea of LDS system, SCMA is a flexible non-orthogonal multi-access scheme to control the complexity of multi-user detection algorithm at the receiving end within a reasonable

B. Li et al. (Eds.): IoTaaS 2020, LNICST 346, pp. 685–695, 2021.
https://doi.org/10.1007/978-3-030-67514-1_54

range through the sparse design of sending end code words [2]. Based on 5G large connection, low latency and low power consumption, as a key technology in the physical layer, the receiver complexity of multi-access technology affects device power consumption, system delay and other indicators from the perspective of underlying devices. Therefore, it is of great practical significance to study low complexity decoding algorithm.

When SCMA technology was first proposed, MPA multi-user detection algorithm was adopted. For the receiving end decoding scheme of SCMA, the message passing algorithm (MPA) can be used to decode the receiving end. MPA can obtain decoding performance close to the maximum likelihood probability criterion on the premise of ensuring reasonable complexity [3, 4]. The complexity of MPA is proportional to the exponential power of constellation size, which leads to limited application in 5G low-delay network. Based on this, scholars at home and abroad focus on reducing the complexity of decoding algorithm. Literature [5] proposed the Log-MPA algorithm, which transformed the messages to be transmitted in the MPA algorithm into the logarithmic domain, and applied the Jacobian logarithmic formula to further reduce the computational burden, eliminate all exponential operations, and convert about 90% multiplication operations into addition operations, saving at least 20% running time compared with the traditional MPA algorithm. Literature [6, 7] proposed a detection algorithm of fixed complexity based on Partial Marginalization. Literature [8] proposed two improved algorithms. One is the sign flip detection algorithm (SFA). The idea of SFA comes from the bit flip algorithm, which is often used for decoding low density parity (LDPC). The symbol flip algorithm firstly makes local hard judgment on each resource node, then reverses the hard judgment result of the resource node with the lowest reliability, and realizes the reliability update and symbol flip operation through iteration, until the condition of stopping iteration meets. The simulation results show that the computational complexity of SFA is obviously lower than that of traditional MPA algorithm. The other is the dynamic user node marginalization messaging algorithm, which USES the reliability defined in SFA to dynamically decide whether to stop updating the soft information of some user nodes during the iteration of MPA, so as to reduce the complexity. The algorithm is called DPM-MPA. Simulation analysis shows that THE performance and complexity of DPM-MPA are superior to that of PM-MPA.

However, the above literatures only improved the MPA algorithm once, but did not improve it the second time. Through careful study of literature [9], this paper finds that the complexity of the improved MPA algorithm is still high, and there is still room for further improvement, which can reduce the complexity of the MPA algorithm again. Specifically, in message update, the nodes are marginalized first, and then the trusted part of the updated message is updated by using the spherical decoding principle. Simulation results show that when the threshold value is small, the complexity of the algorithm proposed in this paper will decrease again, while BER has no loss.

The structure of the paper is as follows: in Sect. 2, the model of SCMA system and the PMSD-MPA algorithm are presented. Section 3 analyzes the algorithm performance. Section 4 summarizes the conclusions.

2 MPA and Improvement Based on Spherical Decoding

2.1 Basic Model of SCMA System

Basic Model

SCMA is gradually evolved from the combination of OFDMA-CDMA and the idea of sparse spread spectrum. At the transmitting end, the biggest difference between SCMA and LTE systems is that the traditional quadrature amplitude modulation (QAM) is changed to SCMA coding. The key technology of the coding is codebook design. SCMA codebook design can be regarded as multi-dimensional a joint optimization of modulation and sparse spread spectrum. Data of multiple users are superimposed on the air interface to form a superimposed non-orthogonal sequence for transmission. At the receiving end, the SCMA system combines the single-user receiving equalization module, the QAM modulation and the de-mapping module into one, forming a multi-user SCMA detection process. Because SCMA data is non-orthogonal superposition, users cannot be distinguished directly, so the receiving end adopts a multi-user joint detection message passing algorithm for signal detection.

The simplified model of the SCMA system uplink is shown in the Fig. 1 below.

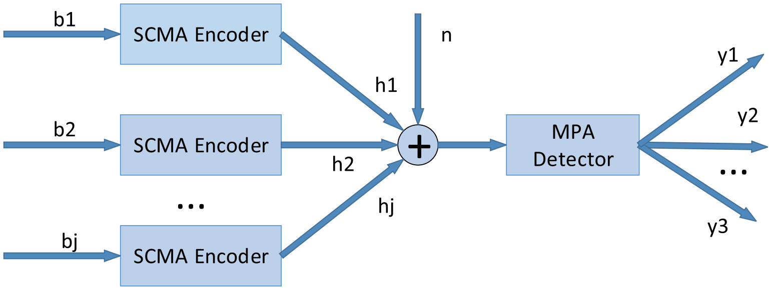

Fig. 1. Basic model of SCMA system

Assume that the SCMA system contains J users and K orthogonal time-frequency resource blocks, similar to OFDMA orthogonal subcarriers. Each user transmits data to the base station through orthogonal time-frequency resource blocks, and the overload factor can be obtained by

$$OF = \frac{J}{K} \tag{1}$$

At the transmitting end, the input bit stream is SCMA encoded and then mapped into a multi-dimensional complex number domain code-word. The code-word is mapped through physical resources and then transmitted through the channel. The code-words selected by different users are based on the same orthogonal time-frequency resources. The method of sparse spread spectrum is non-orthogonal superposition.

When transmitting data, the SCMA encoder selects code-words from the designed codebook according to the input coded bits. When the code-word is selected, it will carry user data for transmission. To store, it can be represented in the form of a tanner graph. The Fig. 2 shows the tanner graph form of a codebook. In the figure, the 6 data layers are multiplexed on 4 orthogonal resource blocks. The connection point means that the codebook has data transmission on this data layer. There are 3 data streams on each physical resource node, and each data stream is connected. To 2 physical resource nodes. Multiple users superimpose on the air interface to form a non-orthogonal sequence.

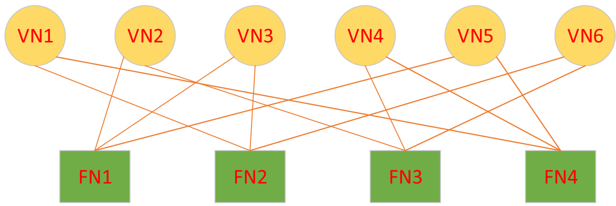

Fig. 2. Tanner graph form of a codebook

In the uplink access system, after the code word is selected, it will carry user data for transmission. Multiple user data is superimposed on the air interface to form a superimposed non-orthogonal sequence, so different user information is superimposed on the 4 resource blocks at the receiving end.

The receiver of SCMA mainly includes two modules: multi-user detector and channel decoder. The multi-user detector is mainly responsible for separating the multi-user data superimposed on the shared channel. The channel decoding module removes the check bits and restores the user's original binary bit stream, which corresponds to the channel encoder at the sending end. Due to the sparse nature of SCMA code-words, a new form of multi-user detection algorithm is given, and the channel decoder can apply the decoding algorithm of the traditional orthogonal system. Therefore, the main purpose is to discuss and study the multi-user detection algorithm.

If all users are synchronized in time and the signal received by the base station is the superimposed signal of all users, the received signal can be expressed as:

$$y = \sum_{j} diag(h_j)x_j + n \tag{2}$$

Where x_j represents the code word sent by the j user; h_j represents the channel gain vector of the j user; n is Gaussian white noise, and At the receiving end, the SCMA system combines the single-user receiving equalization module and the QAM modulation and demapping module into one, forming a multi-user SCMA detection process. Then the signal received at the resource node K is:

$$y_k = \sum_{j \in \xi_j} h_{jk}x_{jk} + n_k \tag{3}$$

MPA

The multi-user detection algorithm of SCMA system makes use of the sparsity of codebook and adopts MPA algorithm to detect superposition of multiple users on the same physical resource block. However, as the number of access users increases, the constellation map superposed on a single physical resource block becomes more and more dense, and the number of constellation points to be searched for detection will increase exponentially. At the same time, in order to ensure the accuracy of the detection, the number of iterations in the MPA detection process will gradually increase. The increase of computational complexity will lead to the increase of system delay, and a large number of intermediate variables will be generated during iteration, which will also take up more storage space. How to reduce the complexity of detection is one of the key problems in SCMA.

An example is given to illustrate the transmission of 6 users over 4 orthogonal time-frequency resources. There is a one-to-one correspondence between factor graph and mapping matrix. The mapping matrix corresponding to the figure is shown in the following formula:

$$F = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \quad (4)$$

The flow chart of MPA decoding algorithm is shown in the Fig. 3. As can be seen from the flow chart of MPA algorithm, the first step of MPA algorithm initializes the conditional probability of factor graph and the probability of each code word of each user. The second step is to update FN and VN nodes of MPA iteratively until the target iteration times or result convergence is achieved. The third step is decision output.

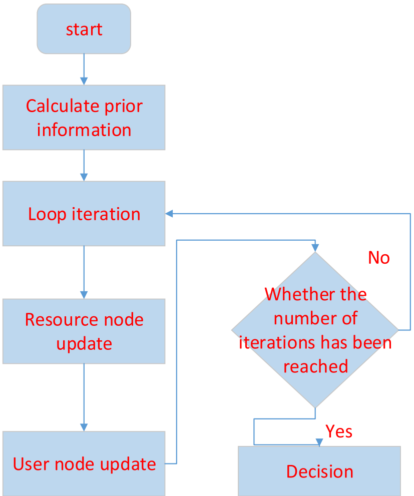

Fig. 3. MPA decoding algorithm process

Use f_k to represent FN, use v_j to represent VN. Each FN corresponds to an orthogonal time-frequency resource, which is a row in F_{K-J}, and each VN corresponds to a user, which is a column in F_{K-J}.

Use $I^t_{f_k \to v_j}(w_{kj})$ to indicate the message from f_k passed to v_j, which is also the updated message on f_k; Use $I^t_{v_j \to f_k}(w_{kj})$ to indicate the message from v_j passed to f_k, which is also the updated message on v_j; w_{kj} indicates the status of the message, $w_{kj} \in \{0, 1, 2, \ldots, M-1\}$

The maximum posterior probability criterion is mainly used for the multi-user detection of SCMA using MPA. First, the posterior probability is converted into prior probability through Bayesian formula, and the prior and other concepts are assumed. Then, the message is constantly updated through iteration until the maximum iteration times are reached, and finally the decoding output is achieved. The decoding process of the messaging algorithm can be expressed as follows:

Step 1: Initialization: Assume the prior and other concepts, as shown in the following formula

$$I^0_{f_k \to v_j}(w_{kj}) = 1/M \tag{5}$$

Step 2: Message iteration. Message update on FN, as shown in Eq. (5.3):

$$I^t_{f_k \to v_j}(w_{kj}) = \sum_{w^{exit}_{kj}} \left(I^{t-1}_{f_k \to v_j}\left(w^{exit}_{kj}\right)\right) \tag{6}$$

where w^{exit}_{kj} indicates the information status of other nodes on FN except v_j, $I^t_{f_k \to v_j}\left(w^{exit}_{kj}\right)$ is external information.

$$I^t_{f_k \to v_j}(w_{kj}) = \sum \left\{ \frac{1}{2\pi\sigma} \exp\left(-\frac{1}{2\sigma^2} \left\| y_k - \sum_{l \in \xi_k} h_{kl} x_{kl} \right\|^2 \right) \bullet \prod_{l \in \xi_k / \{j\}} I^t_{v_l \to f_k}(w_{kl}) \right\} \tag{7}$$

t indicates the number of iterations, ξ_k represents the VN collection connected to f_k, $\xi_k / \{j\}$ represents a collection of VNs except v_j connected to f_k

The message update on VN is shown in Equation

$$I^t_{v_j \to f_k}(w_{kj}) = K_j \prod_{p \in \varepsilon_j / \{k\}} I^t_{f_k \to v_j}(w_{pj}) \tag{8}$$

where, K_j is the regulatory factor, which satisfies the following equation

$$\sum_{w_{kj}=1}^{M} I^t_{v_j \to f_k}(w_{kj}) = 1 \tag{9}$$

The ε_j represents the VN collection connected to v_j

Step 3: Decoding detection:

After a finite number of iterations, VN receives the message from FN, which is the guess of the signal sent by the user, and takes it as the detection result, as shown in Eq. (10). The $t_{\max}$ represents the maximum number of iterations.

$$Q_j\left(w_{kj}\right) = \prod_{k \in \xi_j} I_{f_k \to v_j}^{t_{\max}}\left(w_{kj}\right) \tag{10}$$

In order to further reduce the computational complexity of the MPA algorithm, the PM-MPA algorithm introduces the idea of partial marginalization. Since the MPA algorithm needs to traverse the user code words used in each iteration, resulting in a lot of repeated calculations, in order to reduce redundant calculations and reduce the computational complexity, the PM-MPA algorithm selects only those confidences in each iteration. Part of the messages with higher degrees are updated, and the PM-MPA algorithm can achieve a balance between computational complexity and BER performance through reasonable settings of related parameters. The basic principle of the PM-MPA algorithm is: In order to avoid exhaustive traversal in the MPA algorithm, a threshold p needs to be set first. When <, then continue to use the message update iteration process of the previous MPA algorithm; when =, then judge the user information and complete the decoding output; when >, then only select $0 \leq \;\leq$ users' messages to participate in the iterative update in subsequent iteration.

2.2 Sphere Decoding and the PMSD-MPA

In order to further reduce the complexity of MPA, based on the theory of spherical decoding and combined with the non-orthogonality of SCMA, this paper introduces the new decision standard of channel quality to screen the user code words and reduce the code words involved in iteration so as to reduce the computational complexity of MPA.

According to the spherical decoding theory, the closer the Euclidean distance of the synthesized constellation point is to the receiving signal point, the more likely it is to be correctly decoded. Then, the user code words participating in the iteration can be reduced by setting the spherical radius, whose expression is:

$$D_k\left(\left[m_j\right]\right) = \left\| y_k - \sum_{j \in \varepsilon_k} h_{kj} x_{kj}\left(m_j\right) \right\| \leq R = \beta\sigma \tag{11}$$

Where $D_k\left(\left[m_j\right]\right)$ is the Euclidean distance between the synthetic constellation point and the received signal; β is a real number greater than 0. when $D_k\left(\left[m_j\right]\right)$ is greater than the radius R, the corresponding SCP is discarded and only retained the part of code word that $D_k\left(\left[m_j\right]\right)$ is less than half R to participate in the calculation of MPA. The spherical decoding radius R is determined by the noise power σ^2. A compromise between computational complexity and bit error rate can be achieved by adjusting the size of R. Among them, when β = 1, 2, 3, the correspondence between the confidence interval and the probability of correct decoding is shown in the table below. when

$R = 2\sigma$, the confidence interval is $(-2\sigma, 2\sigma)$, which can guarantee the probability of correct decoding on resource block K to reach 95.4% (Table 1).

Table 1. Normal distribution.

Confidence interval	Probability value/%
$(-\sigma,\ \sigma)$	68.3
$(-2\sigma,\ 2\sigma)$	95.4
$(-3\sigma,\ 3\sigma)$	99.7

If the filtered code word is substituted into Eq. 7, it can be rewritten as

$$I^t_{f_k \to v_j}(w_{kj}) = \sum_{D_k([m_j] \le R)} \left\{ f(y_k | x_k([m_j])) \bullet \prod_{l \in \xi_k/\{j\}} I^t_{v_l \to f_k}(w_{kl}) \right\} \tag{12}$$

The setting of radius R value should be reasonable. If it is too small, part of the code words that can be correctly decoded will be screened out, thus affecting the detection result.

The MPA algorithm based on the principle of sphere decoding achieves the goal of reducing complexity by reducing the number of user information superimposed on each resource. If, on the basis of reducing the number of user information superimposed on each resource, the number of iterations of some user information is reduced, the complexity of the algorithm will be more reduced, that is, the PMSD-MPA algorithm proposed in this paper.

3 Results and Discussion

In this paper, six users are multiplexed for transmission on four physical resource blocks, and the code book published by Huawei is adopted. Each user occupies two resources for data transmission. The specific simulation parameters are shown in the Table 2:

Table 2. Simulation parameters.

Parameters	Value
Orthogonal resource number K	4
User number V	6
Codebook dimension M	4
Frame size N	3000
Channel	AWGN

Figure 4 shows the comparison of bit error rate performance with SNR of PM-MPA algorithm and conventional MPA algorithm in 6 iterations. Figure 4 shows that the bit error rate performance of PM MPA algorithm is similar to that of traditional MPA algorithm.

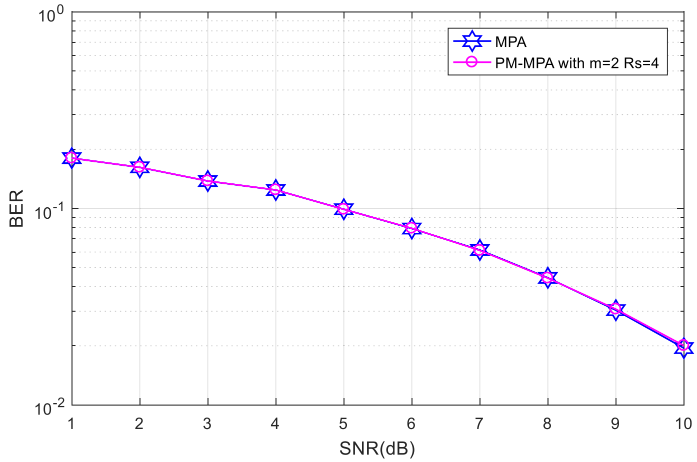

Fig. 4. Comparison of BER performance of MPA and PM-MPA

Figure 5 shows the curve of bit error rate with SNR of traditional MPA algorithm, PM-MPA algorithm, SD-MPA algorithm and PMSD-MPA algorithm proposed in this paper when $t_{\max} = 6$.

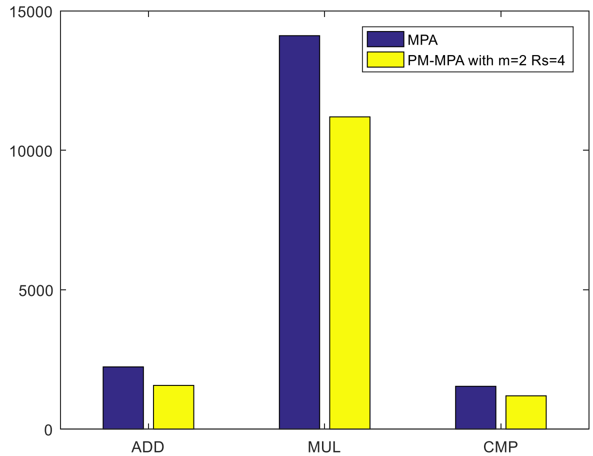

Fig. 5. Comparison of BER performance

When R = 1 and R = 2 are set, the number of iterations is set to 6, and the BER performance of SD-MPA algorithm is shown as Fig. 5: As can be seen from the Fig. 6, the smaller the decision radius is, the worse the BER performance will be, because too small R will over filter the user code word information. When $R = 2\sigma$, bit error rate performance of SD-MPA and MPA is similar, mainly because word screening based on spherical decoding theory ensures the accuracy of multi-user detection with high probability.

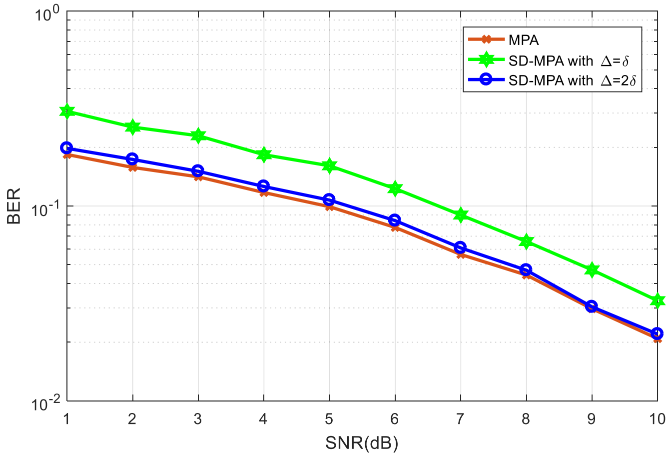

Fig. 6. Comparison of BER performance at different R of SD-MPA

Figure 7 shows the curve of bit error rate with SNR of traditional MPA algorithm, PM-MPA algorithm, SD-MPA algorithm and PMSD-MPA algorithm proposed in this paper when $t_{\max} = 6$. When the decoding radius R is set appropriate, the bit error rate performance of PMSD-MPA does not suffer too much.

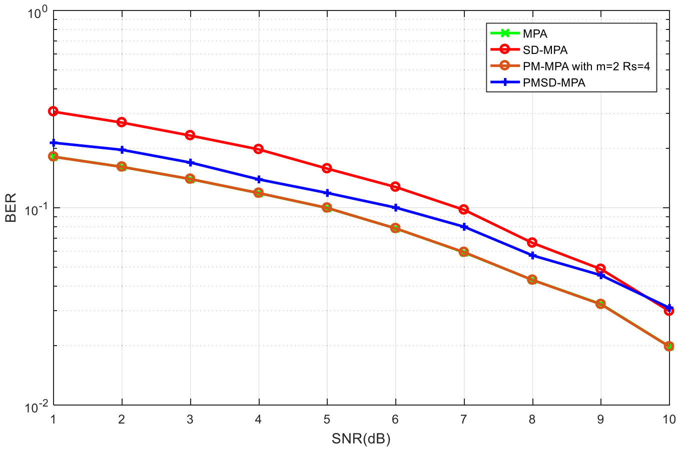

Fig. 7. Comparison of BER performance

Taken together, the algorithm proposed in this paper can achieve a good balance between bit error rate performance and complexity.

4 Conclusions

In this paper, the spherical decoding principle is adopted to implement the MPA algorithm. Based on the partial code word marginalization method, the PMSD-MPA algorithm is proposed. In other words, on the basis of setting the minimum value of the probability density function, constellation points far away from the received signal are discarded again to reduce the complexity of the original MPA again. Simulation results show that the algorithm proposed in this paper is lower in complexity than the algorithm before the improvement, and has no impact on the system BER. The algorithm proposed in this paper is feasible and has practical value.

References

1. Qi, B., Lin, L., Shan, Y.: Non-orthogonal multiple access technology for 5G. Telecommun. Sci. **31**(5), 14–21.2 (2015)
2. Jing, L.: Research on SCMA low-complexity detection algorithm based on information transfer. Beijing University of Posts and Telecommunications. Master's degree thesis (2016)
3. Ning, W.: Research on multi-user detection algorithm in SCMA system. South-Central University for Nationalities (2018)
4. Morrow, R.K., Lehnert, J.S.: Bit-to-bit error dependence in slotted DS/SSMA packet systems with random signature sequences. IEEE Trans. Commun. **37**(10), 1052–1061 (1989)
5. Lian, J., Zhou, S., Zhang, X., et al.: Low complexity decoding method for SCMA in uplink random access. In: Global Communications Conference. IEEE (2017)
6. Mu, H., Ma, Z., Alhaji, M.: A fixed low complexity message pass algorithm detector for uplink SCMA system **4**(6) (2015)
7. Bo, H.: Research on sparse code multiple access technology, Xidian University. Master's thesis (2017), vol. 8 (2017)
8. Jianchao, S., Hua, H.: The structure of LAS code and the advantages of LAS-CDMA over traditional CDMA. Commun. Technol. **40**(12) (2007)
9. Kai, S., Bei, Y., Guangyu, W.: Multiuser detection scheme for SCMA with partial extrinsic information transmission. Syst. Eng. Electron. (2017)