Deep Learning-Based Long-Term Power Allocation Scheme for NOMA Downlink System in S-IoT

YUNYU SUN, YE WANG, (Member, IEEE), JIAN JIAO[ID], (Member, IEEE),
SHAOHUA WU[ID], (Member, IEEE), AND QINYU ZHANG, (Senior Member, IEEE)
Communication Engineering Research Centre, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China
Peng Cheng Laboratory, Shenzhen 518055, China

Corresponding authors: Ye Wang (wangye83@hit.edu.cn) and Jian Jiao (jiaojian@hit.edu.cn)

ABSTRACT In this paper, we formulate a long-term resource allocation problem of non-orthogonal multiple access (NOMA) downlink system for the satellite-based Internet of Things (S-IoT) to achieve the optimal decoding order and power allocation. This long-term resource allocation problem of the satellite NOMA downlink system can be decomposed into two subproblems, i.e., a rate control subproblem and a power allocation subproblem. The latter is a non-convex problem and the solution of which relies on both queue state and channel state. However, the queue state and the channel state continually change from one time slot to another, which makes it extremely strenuous to characterize the optimal decoding order of successive interference cancellation (SIC). Therefore, we explore the weight relationship between the queue state and the channel state to derive an optimal decoding order by leveraging deep learning. The proposed deep learning-based long-term power allocation (DL-PA) scheme can efficiently derive a more accurate decoding order than the conventional solution. The simulation results show that the DL-PA scheme can improve the performance of the S-IoT NOMA downlink system, in terms of long-term network utility, average arriving rate, and queuing delay.

INDEX TERMS Satellite-based Internet of Things, deep learning, non-orthogonal multiple access, successive interference cancellation.

I. INTRODUCTION

With the acceleration of internet of everything (IoE) process, the demand of anywhere and anytime broadband access capability is becoming more and more urgent [1]. At the same time, the satellite communications, machine to machine communications and sensor technology are made breakthrough [2]. Thus, the satellite-based Internet of Things (S-IoT) with next generation of high throughput satellite (HTS) which can provide broadband access in a global coverage and cost-effective manner [3], is viewed as a vital role to address issues related to human living such as energy management, climate change, transportation, healthcare, business

The associate editor coordinating the review of this manuscript and approving it for publication was Guan Gui.

logistics, building automation, and disaster recovery, etc [4]–[6]. A simple structure of S-IoT is shown in Fig. 1.

To realize massive access in S-IoT and adapt to the tight spectrum resources, non-orthogonal multiple access (NOMA) is viewed as an effective solution for enhancing spectral efficiency. Recently, NOMA has drawn significant interest from researchers because of its very promising applications in fifth generation (5G) networks [7]. In general, the existing NOMA schemes can be divided into two categories: power domain NOMA [8]–[10] and code domain NOMA [11], [12]. Power domain NOMA allows multiple UEs to share the same subcarrier simultaneously with different power levels at the transmitter, which increases the sum spectral efficiency but also introduces multi-user interference. To cope this problem, successive interference
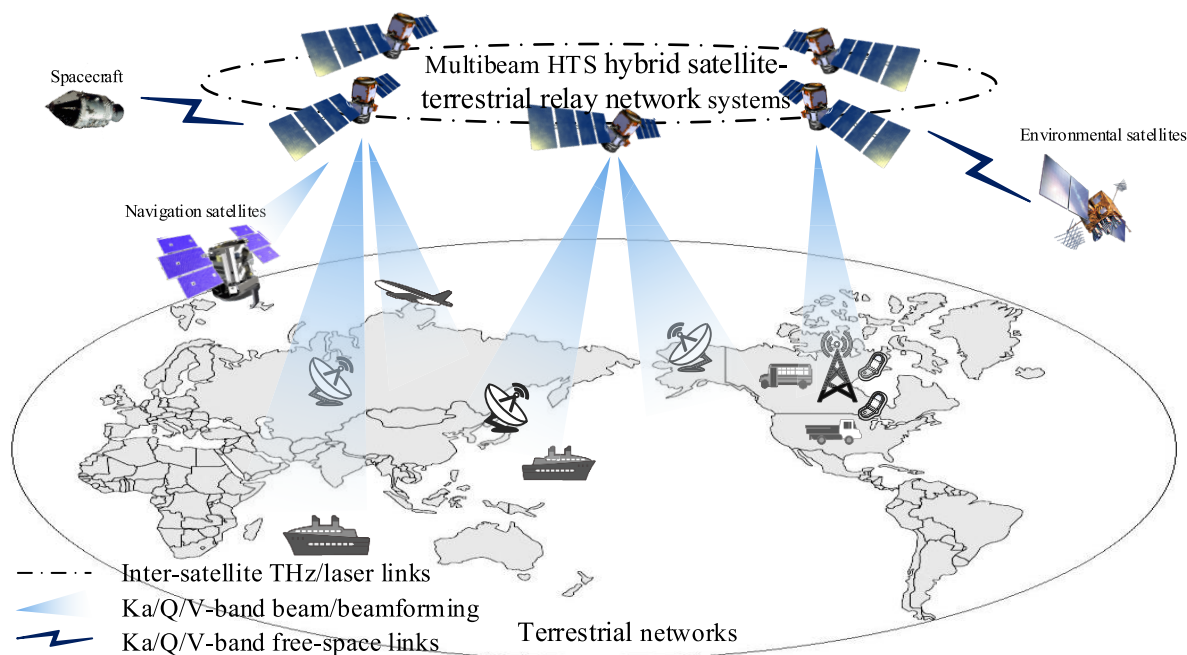
FIGURE 1. The architecture of satellite based internet of things systems.

cancelation (SIC) is applied at the receiver to separate multiplexed UEs signals at the cost of increased computational complexity [13].

The superiority of NOMA has been demonstrated in previous literature [14]. The authors in [8] studied the performance of NOMA when users were randomly deployed, and showed that NOMA can improve the ergodic sum rates of the 5G land mobile system. The user fairness was investigated in [15], where two cases of instantaneous channel state information (CSI) and average CSI were discussed.

Note that the power and storage resources on the satellites are limited, therefore, the outage events may occur due to the insufficient power to allocate, and the limited storage resource on HTS would lead to overflow as well. To address these issues, we propose a long-term joint power allocation and rate control problem, and convert this long term problem into a series of online subproblems according to Lyapunov optimization. In addition, we notice that the power allocation subproblem is non-convex and depends not only on channel state but also on queue state. However, it is extremely strenuous to approximate the optimal SIC decoding order through conventional theoretical analysis.

Thus, we turn to deep learning (DL), also known as deep structured learning or hierarchical learning, which is part of a broader family of machine learning methods based on artificial neural networks [16]–[18]. In the past few years, deep learning architectures such as deep neural networks, deep belief networks, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, speech recognition, natural language

processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, material inspection and board game programs, where they have produced results comparable to and in some cases superior to human experts [19], [20].

Moreover, there have been some works applying DL technology into physical-layer wireless communications in recent years. In [21], the authors propose an effective DL-aided NOMA scheme which can model the channel state information (CSI) by learning the environment automatically via offline learning. In addition, a DL based pilot allocation scheme is designed in [22] to improve the performance in cellular networks with severe pilot contamination by learning the relationship between pilot assignment and the users' location pattern. A novel decentralized resource allocation mechanism for vehicle-to-vehicle (V2V) communications based on deep reinforcement learning is developed in [23], which can be applied to both unicast and broadcast scenarios. Furthermore, in [24] and [25], DL is introduced to learn optimal resource allocation policies in wireless communication systems. Similar to these works, our proposed long-term resource allocation problem is also non-convex and lack of SIC decoding order model knowledge, thus, we adopt the concept of universal function approximation of deep neural networks and develop a deep learning-based approach to train the model of SIC decoding order.

In this paper, considering that both the power and storage on satellites are limited, we first establish a long-term joint power allocation and rate control scheme for the S-IoT NOMA downlink system. Then, by leveraging the

Lyapunov optimization framework, we convert the long-term optimization problem into a series of online power allocation and rate control subproblems. However, it is worth noting that our proposed long-term power allocation subproblem is non-convex, and the solution of which depends on the SIC decoding order determined by both queue state and channel state. Considering the queue state and channel state continually change from one time slot to another, which make it extremely strenuous to characterize the optimal SIC decoding order. Therefore, we approximate the SIC decoding order by leveraging deep learning to training through large amounts of data. We further simulate and evaluate our proposed DL-PA scheme, and simulation results show that our proposed DL-PA scheme can improve the performance than the rough conventional estimation scheme.

The remaining part of this paper is organized as follows. In Section II, the system model is provided in detail. The long-term optimal problem is formulated and converted into multiple online problems in Section III. In Section IV, we introduce deep learning to approximate the SIC decoding order. The simulation results and analysis are given in Section V. Some key notations and nomenclature are summarized in Table 1.

TABLE 1. Summary of notations.

Symbol	Meaning
T	Number of time slots.
K	Number of UEs.
P_{max}	The short-term peak power.
a_{max}	The maximum arriving rate.
P_{mean}	The long-term average power.
$D_i(t)$	The distance between S and UE_i at time slot t.
$g_i(t)$	The channel gain of i-th downlink between S and UE_i at time slot t.
$p_i(t)$	The transmit power allocated to UE_i at time slot t.
$Q_i(t)$	The queue backlog state of UE_i at time slot t.
$a_i(t)$	The data rate arriving at the i-th queue at time slot t.
$b_i(t)$	The transmit data rate targeted UE_i at time slot t.
$Z(t)$	The power debt state at time slot t.
n	AWGN.
η	The noise power.
W	The bandwidth.
τ	The duration of each time slot.
$SINR_i(t)$	Signal-to-interference-plus-noise ratio in i-th downlink at time slot t.
$s_i(t)$	The index of UE with the i-th largest power level at time slot t.
lr	Learning rate.
$Epoch$	Number of rounds of learning.
ε	Minimum constant number for numerical stability.

II. SYSTEM MODEL

Consider a downlink NOMA system consisting of a satellite source node and K terrestrial terminals, denoted as S and $UE_1, UE_2, \cdots, UE_K$, respectively. Note that the focus of this paper is to study the optimal resource allocation scheme within a certain NOMA group, and we regard that all the UEs covered by a same spot beam. Thus, we focus on the scenario where a NOMA group under a single spot beam. By taking advantages of power domain NOMA, all the UEs in the same spot beam coverage with different positions can share the same frequency at the same time.

We divide time into multiple time slots t, $t \in 0, 1, 2, \ldots, T - 1$. At the beginning of each time slot, the superposition coding (SC) is employed at node S, and S multicasts the following SC signals to the K UEs:

$$x(t) = \sqrt{p_1(t)}x_1(t) + \sqrt{p_2(t)}x_2(t) + \cdots + \sqrt{p_K(t)}x_K(t), \tag{1}$$

where $x_i(t)$ is the desired signal for UE_i during the time slot t with $E[|x_i(t)|^2] = 1 (i = 1, 2, \cdots, K)$, and $p_i(t)$ denotes the transmit power allocated to UE_i at time slot t.

Consistent with most existing works [26], [27], we assume that the links between S and UEs experience independent and identically distributed (i.i.d.) block shadowed-Rician fading distribution and additive white Gaussian noise (AWGN). We define $g_i(t)$ as the composite channel gain, which is consisting of antenna gain, beam gain, fading channel coefficient and free space loss at time slot t. Through the i-th downlink, the received signal at UE_i can be expressed by

$$y_i(t) = g_i(t)x(t) + n. \tag{2}$$

At UE_i, the UEs are sorted in an ascending order of power level as $p_{s_1(t)}(t) < p_{s_2(t)}(t) < \cdots < p_{s_K(t)}(t)$, which is a key step to find an optimal decoding order of SIC, and this is very challenging for us to solve by our proposed scheme. Then, SIC is used to decode the signals of the UEs with higher power levels. Specifically, UE_i need to decode the signal for each UE_j with $j > i$, then subtract them from the received signal until decode the UE_i's own signal. Furthermore, UE_i treats the signals of UE_l with $l < i$ as interference and the signal-to-interference-plus-noise ratio (SINR) of UE_i is given by

$$SINR_{s_1(t)}(t) = \frac{p_{s_1(t)}(t)g_{s_1(t)}(t)}{\eta},$$

$$SINR_{s_2(t)}(t) = \frac{p_{s_2(t)}(t)g_{s_2(t)}(t)}{\eta + p_{s_1(t)}(t)g_{s_2(t)}(t)},$$

$$\cdots$$

$$SINR_{s_K(t)}(t) = \frac{p_{s_K(t)}(t)g_{s_K(t)}(t)}{\eta + \sum_{j=1}^{K-1} p_{s_j(t)}(t)g_{s_K(t)}(t)}. \tag{3}$$

During each time slot, with a pre-determined SIC decoding order, we can derive the leaving data rate targeted UE_i $b_i(t)$ as

$$b_i(t) = W\tau \log_2 (1 + SINR_i(t)), \tag{4}$$

where W represents the frequency bandwidth, and τ is the duration of each time slot. Let $Q_i(t)$ to denote the amount of data buffered at queue Q_i during time slot t waiting to be transmitted. And we denote $a_i(t)$ as the data targeted UE_i arriving at the HTS S from the backhaul link, which are first buffered at queue Q_i and then forwarded to UE_i over the wireless channel.

III. PROBLEM FORMULATION AND CONVERSION
A. PROBLEM FORMULATION
To convert the two long-term constraints: long-term mean rate stability and long-term mean power stability, we establish the

following two virtual queue Q and Z.

$$Q_i(t+1) = \max\left(Q_i(t) - b_i(t), 0\right) + a_i(t). \quad (5)$$

We find that when a UE has large queue backlog, there are two ways to keep the queue stable and avoid overflow, one is to decrease the arriving rate $a_i(t)$ and the other is to increase the leaving rate $b_i(t)$. According to [28], a discrete time process $Q(t)$ is *network stable*, and the overflow can be avoided if the following conditions are satisfied:

$$\lim_{t\to\infty} \frac{E[|Q(t)|]}{t} = 0, \quad (6)$$

and

$$Z(t+1) = \max\left(Z(t) - P_{mean}, 0\right) + \sum_{i=1}^{K} p_i(t), \quad (7)$$

where $Z(t)$ denotes the power debt state at time slot t to guarantee the long-term mean power constraint P_{mean}, i.e., (8c). When the allocated power during the last time slot t exceeds P_{mean} to maximize the expected throughput (the long-term network utility), i.e., $Z(t) - P_{mean} > 0$, then the power can be used in current time slot $(t+1)$ is adjusted to be less or even zero, which ensures that long-term power constraint.

Considering the actual satellite communication environment, the transmission power of HTS is limited by both its long-term average power P_{mean} and short-term peak power P_{max}. Therefore, to maximize the long-term network utility U, which is denoted as a non-decreasing and strictly concave function of the long-term time average arriving data rate [28], we state the original problem which depends on the long-term time average arriving data rate $a_i(t)$ and the coefficient of power allocation $p_i(t)$ as follows:

$$\max \sum_{i=1}^{K} U_i\left(\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} E[a_i(t)]\right), \quad (8a)$$

$$s.t. \sum_{i=1}^{K} p_i(t) \le P_{max}, \quad (8b)$$

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} E\left[\sum_{i=1}^{K} p_i(t)\right] \le P_{mean}, \quad (8c)$$

$$0 \le a_i(t) \le a_{max}, \forall i \in [1, K], \quad (8d)$$

$$\lim_{t\to\infty} \frac{E[|Q(t)|]}{t} = 0, \quad (8e)$$

$$p_{s_i(t)}(t) > \eta/g_{s_i(t)}(t) + \sum_{j=1}^{i-1} p_{s_j(t)}(t), \quad (8f)$$

$$i = 1, 2, \cdots, K.$$

B. PROBLEM CONVERSION

It is obvious that the original problem is a time average optimization problem and depends on multiple time slots. According to the Lyapunov theory [28], we can decompose the original problem and get a series of single time slot power allocation sub-problems as follows:

$$\max \ W\tau\left[Q_{s_1(t)}(t)\log_2\left(1 + \frac{p_{s_1(t)}(t)g_{s_1(t)}(t)}{\eta}\right)\right.$$

$$\left. +Q_{s_2(t)}(t)\log_2\left(1 + \frac{p_{s_2(t)}(t)g_{s_2(t)}(t)}{\eta + p_{s_1(t)}(t)g_{s_2(t)}(t)}\right) + \ldots\right.$$

$$\left. +Q_{s_K(t)}(t)\log_2\left(1 + \frac{p_{s_K(t)}(t)g_{s_K(t)}(t)}{\eta + \sum_{j=s_1(t)}^{s_{K-1}(t)} p_j(t)g_{s_K(t)}(t)}\right)\right]$$

$$-Z\left(\sum_{j=1}^{K} p_j(t)\right), \quad (9a)$$

$$s.t. \sum_{i=1}^{K} p_i(t) \le P_{max}, \quad (9b)$$

$$p_{s_i(t)}(t) > \eta/g_{s_i(t)}(t) + \sum_{j=1}^{i-1} p_{s_j(t)}(t), \quad (9c)$$

$$i = 1, 2, \cdots, K. \quad (9d)$$

The derivation of the decomposition is deduced in Appendix A. We can see that every single time slot online optimization problem only depends on the information during the current time slot. It is worth noting that the optimization objective function depends on both Q and g, which play the key role in the power distribution process. Moreover, the relationship of Q and g directly determines the power level order, which is equal to the SIC decoding order.

Specifically, in the initial state, all the queue states of UEs are same, i.e., all the queue states Q equal to zero. And we allocate power only depends on the channel gain g, which is consistent with conventional power domain NOMA scheme. While after several of time slots, due to the differential power allocation and arriving rate, the queue state Q has changed and produced a difference with each other. At the same time, power allocation no longer depends only on the channel state g, but is also greatly affected by the queue state Q.

Therefore, an optimized SIC decoding order is the key to achieve the maximum utility, we dedicate to derive the optimal SIC decoding order to enhance the performance of our proposed long-term optimization problem in the next section.

However, the objective function is non-convex and it is difficult to deduce an accurate and optimal SIC decoding order. Thus, we propose a *FuS* algorithm in which a function $F = Q + ug$ is defined to roughly estimate the SIC decoding order, where u is a weight to emphasize the importance of the proportion between the queue state Q and the channel state g. In addition, we fix the weight u to a constant at the beginning of each time slot, and sort all the UEs in an ascending order of $F = Q + ug$ which we regard as a suboptimal decoding order. After reordering all UEs in the suboptimal decoding order, we then use particle swarm optimization (PSO) which iteratively tries to improve a candidate solution with regard to a given measure of quality to find the optimal power allocation. We denote this algorithm as *FuS* Algorithm (Algorithm 1).

IV. IMPROVE THE SIC DECODING ORDER VIA DEEP LEARNING METHOD

Since the result of *FuS* algorithm is suboptimal, in this section, we aim to further approximate the SIC decoding order by leveraging deep learning method, which is denoted *DLS* algorithm to enhance the performance of our proposed *DL-PA* scheme, as shown in Algorithm 2. The framework of

Algorithm 1 *FuS* Algorithm

Input: Q, Z, D, τ, P_{max}, P_{mean}, a_{max}, B, η, and T;
Output: The channel gain g, the optimal rate R and the optimal power allocation

1 Sort UEs according to $F(Q, g)$;
2 //Network stability control;
3 **for** $i = 1 : K$ **do**
4      $a_i = \frac{V_i}{Q_i}$;
5   **if** $a_i < 0$ **then**
6         $a_i = 0$;
7   **end**
8   **if** $a_i > a_{max}$ **then**
9         $a_i > a_{max}$;
10   **end**
11 **end**
12 Power allocation with PSO ;
13 Restore the original order;
14 Update Q and Z with the new allocated rate and power.

Algorithm 2 *DP-PA* Scheme

Input: lr, *batchsize*, $[Q_1(t), \cdots , Q_K(t)]$, T, $[g_1(t), \cdots , g_K(t)]$, W, τ, p_{mean} and p_{max}; //Set the learning parameters for training process and the practical constraints for satellite communications;
Output: The optimal SIC decoding order $[\hat{s}_1(t), \cdots , \hat{s}_K(t)]$ and optimal power allocation $[p_1(t), \cdots , p_K(t)]$;

1 Generate *Nbatchsize* sets of training samples $\{[s_1, \cdots , s_K], [\hat{s}_1, \cdots , \hat{s}_K]\}$;
2 Train the framework of *DLS* Algorithm;
3 Finish training process;
4 Power allocation process.
5 **for** $t = 1{:}T$ **do**
6   Use the above model trained, $[\hat{s}_1(t), \cdots , \hat{s}_K(t)]$ returns when $[Q_1(t), \cdots , Q_K(t)]$ and $[g_1(t), \cdots , g_K(t)]$ is input.
7   Resort all *UE*s according to $[\hat{s}_1(t), \cdots , \hat{s}_K(t)]$;
8   Power allocation through the *PA* Algorithm;
9   Return $[p_1(t), \cdots , p_K(t)]$;//The optimal power allocation;
10   Recover the original index of UEs using $[\hat{s}_1(t), \cdots , \hat{s}_K(t)]$.
11 **end**

our proposed DL aided long-term power allocation NOMA scheme is demonstrated in Fig. 2.

The core idea behind in our *DL-PA* scheme is to use the neural network in Fig. 3 (consisting of three hidden layers of neural network) as an approximation function that computes the SIC decoding order based on the given queue state and channel state. The output of the S-IoT network is then compared to a given optimal SIC decoding order. The aim of
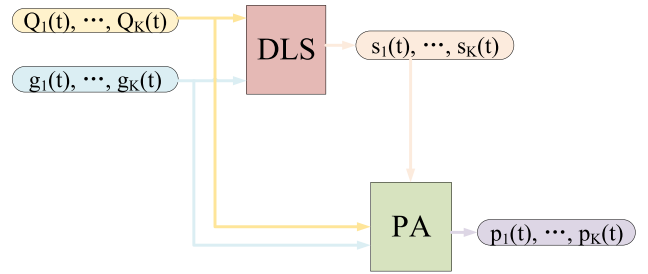

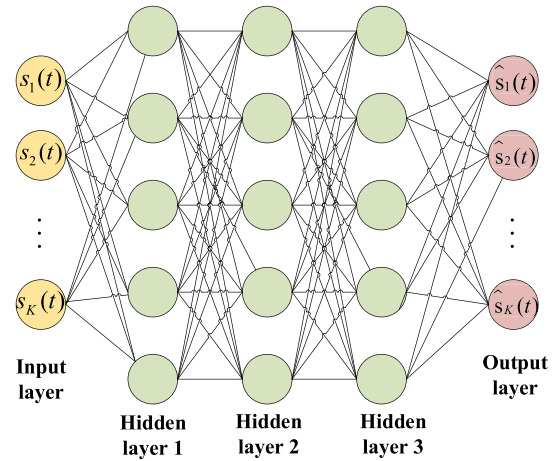**FIGURE 2.** The framework of DL-PA scheme.

FIGURE 3. The framework of DL aided long-term power allocation NOMA scheme.

our output is to be as close to the target as possible, and we define a loss function according to mean squared error (MSE) principle as follows

$$L(\hat{s}_i, s_i) = MSE(\hat{s}_i, s_i) = \frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{K} \frac{||\hat{s}_i - s_i||^2}{s_i^2}. \quad (10)$$

Moreover, in our *DLS* algorithm, we use backpropagation to alter our network's weights, then the future outputs will be closer to our desired target. In this section, we choose the adaptive moment estimation (Adam) algorithm for our *DLS* algorithm, which is a popular algorithm in the field of deep learning. Compare to other favorably stochastic optimization methods, Adam achieves good results in a fast convergence speed, which is proved to efficiently solve the practical deep learning problems [29]. Moreover, there are some attractive benefits [30] of using Adam algorithm on non-convex optimization problems as follows:

1) Straightforward to implement.
2) Computationally efficient.
3) Little memory requirements.
4) Invariant to diagonal rescale of the gradients.
5) Well suited for problems that are large in terms of data and/or parameters.
6) Appropriate for non-stationary objectives.
7) Appropriate for problems with very noisy/or sparse gradients.

8) Hyper-parameters have intuitive interpretation and typically require little tuning.

Before training, we need to prepare a lot of data for training. For each different set of queue state $[Q_1(t), \cdots, Q_K(t)]$ and channel state $[g_1(t), \cdots, g_K(t)]$ combinations, we first traverse all possible sequences, find the decoding order that maximizes the utility and record it as the optimal SIC decoding order $[\hat{s}_1(t), \cdots, \hat{s}_K(t)]$, and the detail of this *DLS* algorithm is described in Algorithm 3.

Algorithm 3 *DLS* Algorithm

Input: lr, $batchsize$, $[Q_1(t), \cdots, Q_K(t)]$,
    $[g_1(t), \cdots, g_K(t)]$; //Set the learning parameters
    for training process;
Output: The optimal SIC decoding order
    $[\hat{s}_1(t), \cdots, \hat{s}_K(t)]$;
1 Generate *Nbatchsize* sets of training samples
    $\{[s_1, \cdots, s_K], [\hat{s}_1, \cdots, \hat{s}_K]\}$;
2 Set the minimum constant number for numerical
    stability ε, exponential decay rates of the moment
    estimation ρ_1, ρ_2;
3 **for** *Epoch* $= 1 : 100$ **do**
4   **for** $n = 1{:}N$ **do**
5   Adam optimization process;
6   Initialize the first-order moment $r_1 = 0$ and the
        second-order moment $r_2 = 0$;
7   Calculate the gradient
        $\theta = \frac{1}{batchsize}\Sigma_{m=1}^{batchsize}\Sigma_{i=1}^{K}\nabla L(\hat{s}_i, s_i)$ each time
        taking *batchsize* samples from the training set;
8   **while** $L(\hat{s}_i, s_i)$ *is not converged* **do**
9   Update the first-order moment:
           $r_1 \longleftarrow \rho_1 r_1 + (1 - \rho_1)\tau$;
10   Update the second-order moment:
           $r_2 \longleftarrow \rho_1 r_2 + (1 - \rho_2)\tau$;
11   Correct the deviation of the first-order
           moment: $\hat{r}_1 \longleftarrow \frac{r_1}{1-\rho_1}$;
12   Correct the deviation of the second-order
           moment: $\hat{r}_2 \longleftarrow \frac{r_2}{1-\rho_2}$;
13   Update $s \longleftarrow s - lr\frac{\hat{r}_1}{\sqrt{\hat{r}_2}+\varepsilon}$;
14   **end**
15   **end**
16 **end**
17 Finish training process;
18 Return $[\hat{s}_1(t), \cdots, \hat{s}_K(t)]$;//The optimal SIC decoding
    order;

Then, we split our prepared data into a training and a testing subset, respectively, where the latter is used to evaluate the performance of our *DL-PA* scheme after the training procedure. In this way, we can make sure that our agent will not overfit on the training data, where overfitting is the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably [31].
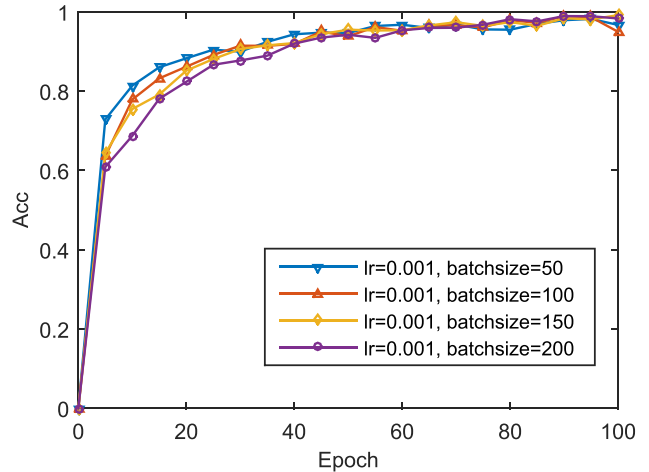

FIGURE 4. Accuracy with fixed learning rate versus Epoch for different batchsize based on the validation data set for the DL-PA scheme.

After finishing the training process, we use the trained model to approximate the optimal SIC decoding order, and then we resort all the UEs and perform the power allocation process, which is illustrated in *DP-PA* Algorithm (Algorithm 2).

V. SIMULATION AND ANALYSIS

In this section, we first provide the parameter selection process of our proposed *DLS* algorithm, and then we analyze the simulation results and compare with *FuS* scheme in the case of 5 UEs.

During the training process, we use the parameter settings suggested by the Adam optimization proposer: $r_1 = 0.9$, $r_2 = 0.999, \varepsilon = 10^{-8}$. We observe the performance in the S-IoT NOMA downlink scenario of about 500 km away from the 5 UEs to the HTS S, where the SNR is defined as $SNR = \frac{P_{mean}}{\eta}$.

We first fix the learning rate to 0.001, and change the batchsize from 50 to 200. The simulation results are shown in Fig. 4, where *Acc* denotes the accuracy, which can reach more than 96 percent after 100 iterations with different batchsize values. Thus, it is feasible to use this model to perform training since high accuracy and fast convergence can both obtained.

Then, it is important to choose the appropriate learning parameters and we adjust the parameters by observing MSE value in Fig. 5 and Fig. 6. In Fig. 5, we fix the batchsize to 50 and adjust the learning rate *lr* from 0.01 to 0.0001. It is obvious that the lowest MSE is obtained when *lr* = 0.001, and we set *lr* = 0.001 in our training process. Similarly, we observe the trend of MSE with fixed *lr* and different batchsize values. All the curves in Fig. 6 converge, while the curve of *batchsize* = 50 is smoother and more stable. Therefore, we choose *batchsize* = 50 in our training process.

After setting the feasible learning parameters, we simulate the performance of proposed *DLS* algorithm in terms of utility, data rate and queue delay. We compare the utility
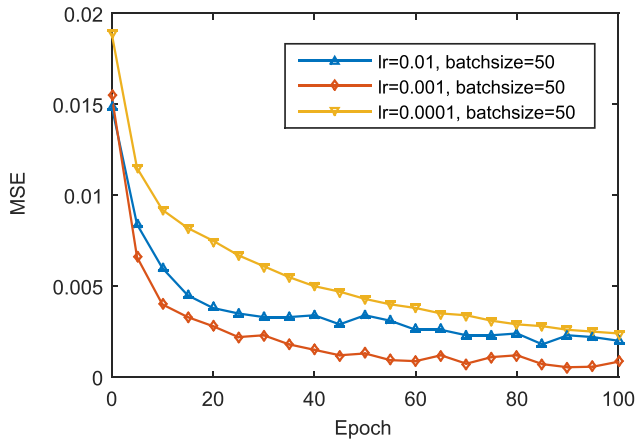
FIGURE 5. MSE with fixed batchsize versus Epoch for different learning rate based on the validation data set for DL-PA scheme.
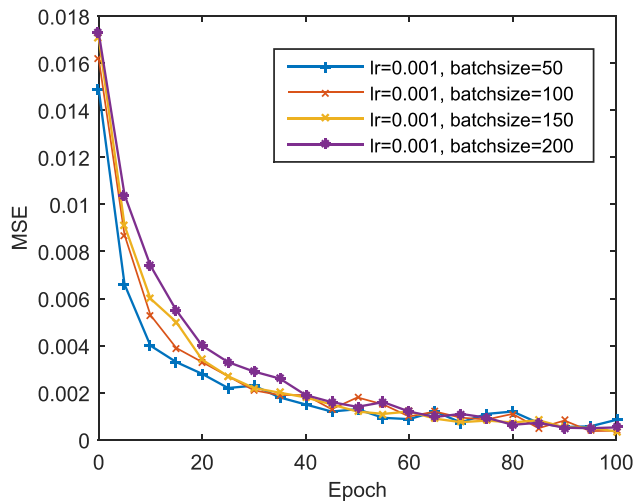

FIGURE 6. MSE with fixed learning rate versus Epoch for different batchsize based on the validation data set for DL-PA scheme.
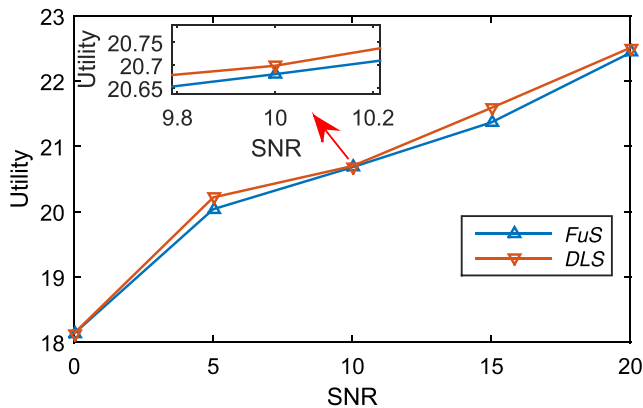

FIGURE 7. Utility under different SNR, where $K = 5, lr = 0.001, batchsize = 50, Epoch = 100, T = 5000$.

FIGURE 8. Data rate of each individual UE under different SNR, where $K = 5, lr = 0.001, batchsize = 50, Epoch = 100, T = 5000$.
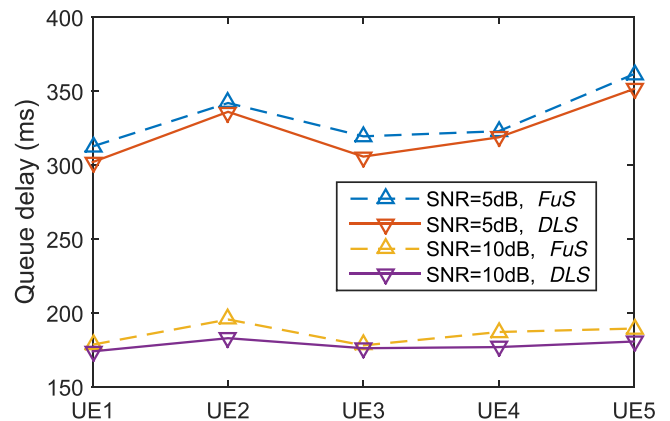
FIGURE 9. Queue delay of each individual UE under different SNR, where $K = 5, lr = 0.001, batchsize = 50, Epoch = 100, T = 5000$.

of individual performance of per UE in Fig. 8 and Fig. 9. Simulation results demonstrate that the data rate of each UE in the *DLS* algorithm outperforms that in the *FuS* algorithm, and the queue delay in the *DLS* algorithm is also lower than that in the *FuS* algorithm.

VI. CONCLUSION

In this paper, with the help of deep learning, an improved long-term power allocation *DL-PA* scheme for S-IoT NOMA downlink system is proposed, which can approximate the optimal SIC decoding order to further improve the performance of our long-term power allocation scheme. Taking the advantage of the Adam optimization algorithm, our training accuracy can achieve more than 96 percent in few iterations. In addition, simulation results demonstrate that the proposed *DLS* algorithm can efficiently allocate power due to the optimal SIC decoding order than that of *FuS* algorithm, and the performance of the long-term power allocation scheme is further improved, in terms of long-term network utility, average arriving rate and queue delay.

of *DLS* algorithm and *FuS* algorithm in Fig. 7 under different SNR. It can be seen that the utility of the total system has been improved in *DLS* algorithm across a wide range of SNRs. In addition, we can observe that the comparison
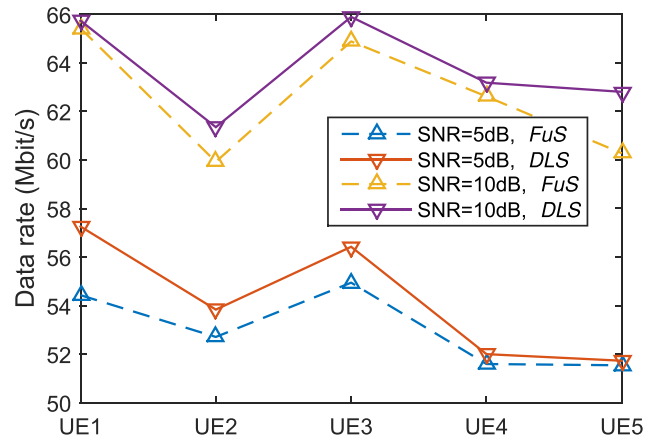
APPENDIX. A

According to Lyapunov theory, we first define a set of virtual queue state $\Phi(t) = [[Q_1(t), \cdots, Q_K(t)], Z(t)]$ and construct a Lyapunov function as follows:

$$L(\Phi(t)) \triangleq \frac{\left[\sum_{i=1}^{K} Q_i^2(t) + Z^2(t)\right]}{2}. \tag{11}$$

Then, we can derive the Lyapunov drift function as follows:

$$\triangle\left(\Phi(t)\right) \triangleq E\left[L\left(\Phi(t+1)\right) - L\left(\Phi(t)\right)\right], \tag{12}$$

where a new target generates, to minimize this Lyapunov drift function $\triangle\left(\Phi(t)\right)$. A small Lyapunov drift means that the queue is stable and the long-term average power constraint is satisfied. Recalling the previous objective function (8a), our current goal has changed to be minimizing Lyapunov drift function $\triangle\left(\Phi(t)\right)$ while maximizing the objective function (8a). In order to unite two optimization goals, we construct a drift-minus-penalty function as follows:

$$DMP(\Phi(t)) = \triangle\left(\Phi(t)\right) - VE\left[U|\Phi(t)\right], \tag{13}$$

where V is a parameter used to emphasize the importance between Lyapunov drift minimization and utility maximization, greater than 0. Bring eq. (5) and eq. (7) into eq. (13), we can further bound $DMP(\Phi(t))$ as the follows:

$$
\begin{aligned}
DMP(\Phi(t)) = E[& \frac{\left[\sum_{i=1}^{K} Q_i^2(t+1) + Z^2(t+1)\right]}{2} \\
& - \frac{\left[\sum_{i=1}^{K} Q_i^2(t) + Z^2(t)\right]}{2}] \\
\leq E[& \frac{\sum_{i=1}^{K}(Q_i(t) - b_i(t) + a_i(t))^2}{2} \\
& + \frac{(Z(t) - P_{mean} + \Sigma_{i=1}^{K} p_i(t))^2}{2}] \\
& - E[\frac{\left[\sum_{i=1}^{K} Q_i^2(t) + Z^2(t)\right]}{2}] \\
\leq B + & E\left[\sum_{i=1}^{K} Q_i(t)a_i(t)|\Phi(t)\right] \\
& - E\left[\sum_{i=1}^{K} Q_i(t)b_i(t)|\Phi(t)\right] \\
& + E\left[Z(t)\left(\sum_{i=1}^{K} p_i(t) - P_{mean}\right)|\Phi(t)\right] \\
& - VE[U|\Phi(t)], \tag{14}
\end{aligned}
$$

where B is a bounded value because $a_i(t)$, $b_i(t)$ and $p_i(t)$ are all bounded. And our objective function is converted into minimizing the right side of eq. (14). As we can see, eq. (14) can be broken down into two parts that one is only depending on arriving rate $a_i(t)$ named *rate control (RC)* problem, and the other is only relying on $p_i(t)$ named *power allocation (PA)* problem. In addition, the optimal solution can be obtained if and only if the two sub-problems are both solved optimally.

The *RC* problem is given as follows:

$$RC\ Problem : \min\ Q_i(t)a_i(t) - VU_i(a_i(t)), \tag{15a}$$

$$s.t.\ 0 \leq a_i(t) \leq a_{max}, \quad \forall i \in [1, K], \tag{15b}$$

which is a convex optimization problem, we can find the optimal solution at the extreme value easily.

The expression of *PA* problem is as follows:

$$
\begin{aligned}
PA\ Problem : \max\ & \sum_{i=1}^{K} Q_i(t)b_i\left(p_i(t), g_i(t)\right) \\
& - Z(t)\sum_{i=1}^{K} p_i(t), \tag{16a}
\end{aligned}
$$

$$s.t.\ \sum_{i=1}^{K} p_i(t) \leq P_{max}, \tag{16b}$$

$$p_{s_i(t)}(t) > \eta/g_{s_i(t)}(t) + \sum_{j=1}^{i-1} p_{s_j(t)}(t), \tag{16c}$$

$$i = 1, 2, \cdots, K.$$

Bring the complete expression of $b_i(t)$ eq. (4) into eq. (16a), we can get eq. (9).

REFERENCES

[1] H. Yao, L. Wang, X. Wang, Z. Lu, and Y. Liu, "The space-terrestrial integrated network: An overview," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 178–185, Sep. 2018.

[2] N. Zhang, S. Zhang, P. Yang, O. Alhussein, W. Zhuang, and X. Shen, "Software defined space-air-ground integrated vehicular networks: Challenges and solutions," *IEEE Commun. Mag.*, vol. 55, no. 7, pp. 101–109, Jul. 2017.

[3] S. Gu, J. Jiao, Z. Huang, S. Wu, and Q. Zhang, "ARMA-based adaptive coding transmission over millimeter-wave channel for integrated satellite-terrestrial networks," *IEEE Access*, vol. 6, pp. 21635–21645, 2018.

[4] M. Chernyshev, Z. Baig, O. Bello, and S. Zeadally, "Internet of Things (IoT): Research, simulators, and testbeds," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1637–1647, Jun. 2018.

[5] N. Zhang, P. Yang, J. Ren, D. Chen, Y. Li, and X. Shen, "Synergy of big data and 5G wireless networks: Opportunities, approaches, and challenges," *IEEE Wireless Commun.*, vol. 25, no. 1, pp. 12–18, Feb. 2018.

[6] Y. Wang, Q. Li, J. Jiao, S. Wu, and Q. Zhang, "ARM: Adaptive random-selected multi-beamforming estimation scheme for satellite-based Internet of Things," *IEEE Access*, vol. 7, pp. 63264–63276, 2019.

[7] Y. Sun, J. Jiao, S. Wu, Y. Wang, and Q. Zhang, "Joint power allocation and rate control for NOMA-based space information networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, May 2019, pp. 1-6.

[8] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[9] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE 77th Veh. Technol. Conf. (VTC Spring)*, Jun. 2013, pp. 1–5.

[10] S. M. A. Kazmi, N. H. Tran, T. M. Ho, A. Manzoor, D. Niyato, and C. S. Hong, "Coordinated device-to-device communication with non-orthogonal multiple access in future wireless cellular networks," *IEEE Access*, vol. 6, pp. 39860–39875, 2018.

[11] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[12] M. Gan, J. Jiao, L. Li, S. Wu, and Q. Zhang, "Performance analysis of uplink uncoordinated code-domain NOMA for SINs," in *Proc. 10th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2018, pp. 1–6.

[13] L. Song, Y. Li, Z. Ding, and H. V. Poor, "Resource management in non-orthogonal multiple access networks for 5G and beyond," *IEEE Netw.*, vol. 31, no. 4, pp. 8–14, Jul./Aug. 2017.

[14] Z. Chen, Z. Ding, X. Dai, and R. Zhang, "An optimization perspective of the superiority of noma compared to conventional OMA," *IEEE Trans. Signal Process.*, vol. 65, no. 19, pp. 5191–5202, Oct. 2017.

[15] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1647–1651, Oct. 2015.

[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Neural Inf. Process. Syst. (NIPS)*, Lake Tahoe, NV, USA, 2012.

[18] A. H. Marblestone, "Toward an integration of deep learning and neuroscience," *Frontiers Comput. Neurosci.*, vol. 10, p. 94, Sep. 2012.

[19] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, Sep. 2018.

[20] K. Kim, J. Lee, and J. Choi, "Deep learning based pilot allocation scheme (DL-PAS) for 5G massive MIMO system," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 828–831, Apr. 2018.

[21] H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep reinforcement learning based resource allocation for V2V communications," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3163–3173, Apr. 2019.

[22] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[23] M. Eisen, C. Zhang, L. F. O. Chamon, D. D. Lee, and A. Ribeiro, "Learning optimal resource allocations in wireless systems," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2775–2790, Apr. 2019.

[24] J. Luo, J. Tang, D. K. C. So, G. Chen, K. Cumanan, and J. A. Chambers, "A deep learning-based approach to power minimization in multi-carrier NOMA with SWIPT," *IEEE Trans. Signal Process.*, vol. 7, pp. 17450–17460, 2019.

[25] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[26] X. Liang, J. Jiao, S. Wu, and Q. Zhang, "Outage analysis of multirelay multiuser hybrid satellite-terrestrial millimeter-wave networks," *IEEE Wireless Commun. Lett.*, vol. 7, no. 6, pp. 1046–1049, Dec. 2018.

[27] J. Jiao, Y. Hu, Q. Zhang, and S. Wu, "Performance modeling of LTP-HARQ schemes over OSTBC-MIMO channels for hybrid satellite terrestrial networks," *IEEE Access*, vol. 6, pp. 5256–5268, 2018.

[28] M. J. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synth. Lect. Commun. Netw.*, vol. 3, no. 1, p. 211, 2010.

[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[30] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*. [Online]. Available: https://arxiv.org/abs/1609.04747

[31] C. Bitsakos, I. Konstantinou, and N. Koziris, "DERP: A deep reinforcement learning cloud system for elastic resource provisioning," in *Proc. IEEE Int. Conf. Cloud Comput. Technol.*, Dec. 2018, pp. 21–29.

YUNYU SUN received the B.S. degree in communication engineering from Dalian University, Liaoning, China, in 2017. She is currently a Graduate Student with the Harbin Institute of Technology (Shenzhen), Guangdong, China. Her current research interests include resource allocation and optimization techniques.

YE WANG received the M.S. and Ph.D. degrees in information and communication engineering from the Harbin Institute of Technology (HIT), Shenzhen, China, in 2009 and 2013, respectively, where he is currently an Assistant Professor. From 2013 to 2014, he was a Postdoctoral Research Fellow with the University of Ontario Institute of Technology, Canada. His research interests include satellite communications, resource allocation, and the mobile Internet.

JIAN JIAO (M'16) received the M.S. and Ph.D. degrees in communication engineering from the Harbin Institute of Technology (HIT), in 2007 and 2011, respectively. From 2011 to 2015, he was a Postdoctoral Research Fellow with the Communication Engineering Research Centre, Shenzhen Graduate School, HIT, Shenzhen, China. From 2016 to 2017, he was a China Scholarship Council Visiting Scholar with the School of Electrical and Information Engineering, The University of Sydney, Sydney, Australia. Since 2017, he has been an Assistant Professor with the Department of Electrical and Information Engineering, HIT Shenzhen. His current interests include error control codes, space information networks, random multiple access, and machine-to-machine communications.

SHAOHUA WU (S'07–M'11) received the Ph.D. degree in communication engineering from the Harbin Institute of Technology, in 2009. From 2009 to 2011, he held a postdoctoral position with the Department of Electronics and Information Engineering, Shenzhen Graduate School, Harbin Institute of Technology, where he has been an Associate Professor, since 2012. From 2014 to 2015, he was a Visiting Researcher with BBCR, University of Waterloo. His current research interests include wireless image/video transmission, deep space communication, advanced channel coding techniques, and 5G wireless transmission technologies. He has authored or coauthored over 70 papers in these fields and holds over 30 Chinese patents.

QINYU ZHANG (M'08–SM'12) received the bachelor's degree in communication engineering from the Harbin Institute of Technology (HIT), in 1994, and the Ph.D. degree in biomedical and electrical engineering from the University of Tokushima, Japan, in 2003. From 1999 to 2003, he was an Assistant Professor with the University of Tokushima. From 2003 to 2005, he was an Associate Professor with the Shenzhen Graduate School, HIT, and was the Founding Director of the Communication Engineering Research Center, School of Electronic and Information Engineering, where he has been a Full Professor, since 2005, and serves as the Dean of the EIE School. His research interests include aerospace communications and networks, wireless communications and networks, cognitive radios, signal processing, and biomedical engineering. He has been a TPC Member of INFOCOM, ICC, GLOBECOM, WCNC, and other flagship conferences in communications. He received the National Science Fund for Distinguished Young Scholars, the Young and Middle-Aged Leading Scientist of China, and the Chinese New Century Excellent Talents in University and obtained three scientific and technological awards from governments. He was the TPC Co-Chair of the IEEE/CIC ICCC'15, the Symposium Co-Chair of the IEEE VTC'16 Spring, an Associate Chair for Finance of ICMMT'12, and the Symposium Co-Chair of CHINACOM'11. He was the Founding Chair of the IEEE Communications Society Shenzhen Chapter. He is on the Editorial Board of some academic journals, such as the *Journal on Communications*, the *KSII Transactions on Internet and Information Systems*, and *Science China: Information Sciences*.

● ● ●