CONTENTS

ZTE COMMUNICATIONS
September 2023 Vol. 21 No. 3 (Issue 84)

Special Topic ▶ Reinforcement Learning and Intelligent Decision

Research Papers ▶

Serial parameters: CN 34–1294/TN*2003*q*16*124*en*P*¥30.00*2200*15*2023-09

Guest Editorial >>>

Special Topic on Reinforcement Learning and Intelligent Decision

Guest Editor

GAO Yang

Over the past few years, deep reinforcement learning (RL) has made remarkable progress in a range of applications, including Go games, vision-based control, and generative dialogue systems. Via error-and-trial mechanisms, deep RL enables data-driven optimization and sequential decision-making in uncertain environments. Compared to traditional programming or heuristic optimization methods, deep RL can elegantly balance exploration and exploitation and handle environmental uncertainties. As a result, this learning paradigm has attracted increasing attention from both academia and industry and is paving a new path for large-scale complex decision-making applications.

However, when scaling deep RL to more practical scenarios, several challenges inevitably arise that require further consideration and urgent attention in the field. Firstly, the dominant model-free deep RL is sample-demanding, as the learning process requires massive interactions with the real environment. This makes it unrealistic to implement deep RL algorithms in some sampling-expensive applications, such as robotics. Secondly, the learned policy is sensitive to changes in the environment and can easily encounter catastrophic failures when deployed in a new or unseen environment. In some time-sensitive applications, such as autonomous driving, the ability to quickly adapt to new scenes is crucial. Thirdly, in complicated scenarios, the real state of the Markov decision process may be unavailable to access, and multiple objectives may exist in scheduling. In this case, previous deep RL algorithms for a single agent with fully observable states cannot achieve the desired goal. These concerns weaken the scalability of deep RL, and scientific investigations are necessary to meet realistic requirements.

This special issue aims at overcoming previously mentioned challenges and contributes to applying deep RL to more realistic scenarios.

The call for papers of this special issue has inspired wide interest and attracted multiple submissions with high quality. After two-round peer reviews, we selected four papers that try to tackle our interested deep RL problems for publication in this special issue. The topics of these published articles include offline reinforcement learning, meta reinforcement learning, and multi-agent reinforcement learning. In terms of applications, these papers range from electroencephalogram (EEG) brain-machine interface, the grid power management to automatic radar detection with the help of deep RL.

The first paper, titled "Double Deep Q-Network Decoder Based on EEG Brain-Computer Interface," focuses on EEG signal processing. In detail, this work adopts a deep double-Q network for the decoding of EEG signals. In comparison to previous pattern recognition or signal process methods, deep RL has more adaptability of signal decoding modules in brain-machine interface to changing environments. The experimental results show the effectiveness of more precise EEG signals' decoding with deep RL.

The second paper, titled "Multi-Agent Hierarchical Graph Attention Reinforcement Learning for Grid-Aware Energy Management," considers the grid power management problem. Technically, uncertainty of environments is involved in decision-making, and multi-objectives guide the optimization process. As a result, multi-agent reinforcement learning is introduced to improve the search efficiency in the large state space, exploit the topology structure of tasks and enhance cooperation between agents at multi-levels. The proposed multi-

DOI: 10.12142/ZTECOM.202303001
Citation (Format 1): GAO Yang. Special topic on reinforcement learning and intelligent decision [J]. ZTE Communications, 2023, 21(3): 1–2. DOI: 10.12142/ZTECOM.202303001
Citation (Format 2): Y. Gao, "Special topic on reinforcement learning and intelligent decision," *ZTE Communications*, vol. 21, no. 3, pp. 1–2, Sept. 2023. doi: 10.12142/ZTECOM.202303001.

agent hierarchical graph attention reinforcement learning can well manage the grid energies and significantly reduce voltage violation numbers.

The third paper, titled "A Practical Reinforcement Learning Framework for Automatic Radar Detection," studies radar detection in a data driven way. Noting that manual adjustment in radar detection is time and money expensive, the authors in this work propose automatically achieving radar detection with the combination of offline RL and meta RL. The proposed method can reduce real-world interaction complexity and enable fast adaptation to new environments. Empirical results indicate the high efficiency of RL-based radar detection.

The fourth paper, titled "Boundary Data Augmentation for Offline Reinforcement Learning," investigates the fundamental issue in offline RL. Theoretically, offline RL can boost data efficiency. However, the existence of distribution shift results in unreliable value estimation, and this makes it difficult to construct offline RL algorithms in risk sensitive scenarios. To address these concerns, this work proposes the use of generative adversarial nets to augment the dataset and calibrate the confidence in value estimation. The experimental results show the great potential of generative modeling for improving offline RL performance.

In summary, we hope this special issue will accelerate the scientific investigation of applicable RL in more general decision-making or optimization scenarios. The articles in this special issue are not only innovative but also provide precious experimental evidence and practical experience in the field. These contributed works bring more insights into algorithm design, bottleneck circumventing, and real-world deployment of deep RL and will facilitate the development of deep RL. Last but not least, we sincerely express our gratitude to all authors, reviewers, and the editorial board, who have made efforts to the success of this special issue.

Biography

GAO Yang received his PhD degree from the Department of Computer Science and Technology, Nanjing University, China in 2000. He is currently a professor with the Department of Computer Science and Technology, Nanjing University, and is the Executive Vice President of the National Institute of Healthcare Big Data at Nanjing University. He leads the artificial intelligence reasoning and learning team of Nanjing University which was selected as the Outstanding Science and Technology Innovation Team of Jiangsu Province's Higher Education Institutions in 2019. He is the fellow of CAAI. His research interests include artificial intelligence and machine learning, especially multi-agent systems and reinforcement learning. He has published more than 200 papers in top conferences and journals in and outside China. He have more than 20 authorized patents. He translated two books: One is *Statistical Reinforcement Learning*; the other is *Algorithm Perspective of Machine Learning*, and edited one book entitled *Distributed Artificial Intelligence*. He received the second prize of the Jiangsu Provincial Science and Technology Award once, the second prize of the Wu Wenjun Natural Science Award of the Chinese Artificial Intelligence Society once, and the Research Achievement Award of the Multi Agent System Professional Group of the Chinese Computer Society once.

Double Deep Q-Network Decoder Based on EEG Brain-Computer Interface

REN Min, XU Renyu, ZHU Ting
(Southwest Jiaotong University, Chengdu 611756, China)

DOI: 10.12142/ZTECOM.202303002
https://link.cnki.net/urlid/34.1294.TN.20230907.1041.002, published online September 7, 2023
Manuscript received: 2023-06-08

Abstract: Brain-computer interfaces (BCI) use neural activity as a control signal to enable direct communication between the human brain and external devices. The electrical signals generated by the brain are captured through electroencephalogram (EEG) and translated into neural intentions reflecting the user's behavior. Correct decoding of the neural intentions then facilitates the control of external devices. Reinforcement learning-based BCIs enhance decoders to complete tasks based only on feedback signals (rewards) from the environment, building a general framework for dynamic mapping from neural intentions to actions that adapt to changing environments. However, using traditional reinforcement learning methods can have challenges such as the curse of dimensionality and poor generalization. Therefore, in this paper, we use deep reinforcement learning to construct decoders for the correct decoding of EEG signals, demonstrate its feasibility through experiments, and demonstrate its stronger generalization on motion imaging (MI) EEG data signals with high dynamic characteristics.

Keywords: brain-computer interface (BCI); electroencephalogram (EEG); deep reinforcement learning (Deep RL); motion imaging (MI) generalizability

Citation (Format 1): REN M, XU R Y, ZHU T. Double deep Q-network decoder based on EEG brain-computer interface [J]. ZTE Communications, 2023, 21(3): 3 - 10. DOI: 10.12142/ZTECOM.202303002
Citation (Format 2): M. Ren, R. Y. Xu, and T. Zhu, "Double deep Q-network decoder based on EEG brain-computer interface," *ZTE Communications*, vol. 21, no. 3, pp. 3 - 10, Sept. 2023. doi: 10.12142/ZTECOM.202303002.

1 Introduction

Brain-computer interface (BCI) offers the possibility of direct communication between the human brain and external devices that perform operational tasks[1]. Researchers capture the electrical signals generated by the brain through the electroencephalogram (EEG) and convert them into neural intentions, which will be correctly decoded and used to control external devices, such as wheelchairs, robotic arms, and automatic vehicles[2 - 4]. Among other things, the correct decoding of the neural intention is a crucial step toward this goal, and the correct interpretation of brain activity can provide the external device with the required commands so that it can perform expected tasks. Within the different EEG systems, the motor imagery (MI) BCI[5 - 6] is a very flexible EEG paradigm, which can be used to distinguish between different intracerebral instructions and to control external devices to execute commands by "what is in mind".

Many methods based on traditional machine learning have been used for MI decoding and feature extraction. Among them, filter bank common spatial patterns (FBCSP)[5, 7] based on the characteristics of common spatial patterns (CSP) have achieved good performance. And some researchers have investigated an improved feature extraction method based on CSP to further improve the performance of BCI system[8 - 9]. In addition, linear discriminant analysis (LDA), support vector machines (SVM), etc., are used to find a projection or hyperplane to separate different categories by analyzing feature distribution[10 - 11]. Due to the limited spatial resolution, low signal-to-noise ratio (SNR), and high dynamic characteristics of MI, as well as the existence of a large amount of noise in EEG signals, the extraction of robust features from EEG data is a crucial step for the successful implementation of BCI. In recent years, the success of deep learning methods has alleviated the need for manual feature extraction to a large extent. As a result, many scholars have explored the application of deep learning in EEG signals. For example, the multi-layer perceptron (MLP) was used to correctly classify EEG signals[12]. Since convolutional neural networks (CNNs) can perceive multiple small domain features with the convolution process proceeding layer by layer and can automatically extract rich features to obtain a depth representation, many studies have tried to use CNNs into BCI to build end-to-end EEG decoding models and achieved good performance[13 - 14].

Supervised learning is a popular paradigm to implement BCI, but it requires an explicit supervised signal to learn.

Even so, frequent calibration (retraining) is necessary due to the plasticity of the brain. Therefore, some scholars have focused on developing an adaptive BCI architecture that allows interaction with a dynamic environment[15-17], where BCI users learn by trial-and-error to adjust their brain activity to the decoder by observing how the external device performs the task (using feedback information). Among them, reinforcement learning (RL)[18] is the general framework that makes the system adapt to the new environment. It is an interactive learning paradigm that can improve policies through constant interaction with the environment, aiming to learn the best mapping relationship from the environmental state to the action. Thus, an RL-based BCI framework is explored, which provides a general framework for constructing dynamic mappings from neural intentions to actions adapted to changing environments, requiring only a scalar signal (reward) feedback from the environment to strengthen the decoder to complete the task, rather than a specific permanently available supervisory signal[19]. At the same time, an RL-based BCI architecture is a more reasonable learning solution to those patients unable to produce precise limb movements. In this case, they only need to understand which action will yield the greatest return when reaching the goals in their environment.

Multiple studies have shown that RL can be used in the rat EEG signal[20-21] and neuronal activity to control the basic BCI system[19, 22]. DIGIOVANNA et al.[19] first proposed a Q(λ)-learning algorithm with the temporal difference (TD) error in an RL-based BCI paradigm, which experimentally trained rats to control prostheses in a two-target selection task. Furthermore, SANCHEZ et al.[23] applied Q(λ)-learning to predict one-step actions, extending the RL-based BCI framework to primates performing center-out tasks. In addition, the BCI paradigm using RL has been successfully applied to closed-loop experiments of intracortical signals in monkeys[24-25]. BAE et al.[26] combined the kernel temporal differences (KTD) (λ) algorithm with the Q-learning algorithm to obtain a reinforcement learning-based neural decoding algorithm (Q-KTD), and the feasibility of this method for BCI decoding was demonstrated in a center-out extension task of intracortical signals in monkeys. THAPA et al.[27] further investigated the applicability and feasibility of Q-KTD in an EEG-based BCI system, demonstrating that the Q-KTD algorithm can correctly learn the mapping between neural intentions in EEG signals and external device control commands. However, there are still some challenges in EEG-based RL interface using Q-KTD: 1) The number of kernel units increases with the number of samples; 2) the curse of dimensionality limits the decoding capability of the Q-KTD algorithm; 3) the Q-KTD decoding technique based on Q-Learning has a generalization problem and requires a long training time.

To overcome the above problems, this paper proposes to use double deep Q-network (DDQN), a deep reinforcement learning algorithm, to decode EEG. DDQN[28], as a variant of deep Q-networks (DQN)[29], uses a neural network to approximate the value function and takes into account the generalization while dealing with high-dimensional inputs. In addition, the dual-value network architecture of DDQN can effectively suppress the influence of overestimation of action values on the decision-making process and is robust to EEG signals that may have random interference.

In section 2, this paper introduces the DDQN algorithm and the basic paradigm based on reinforcement learning brain-computer interface. In section 3, the EEG decoder based on DDQN is described and the network structure diagram is given. In section 4, the feasibility and advantages of DDQN for EEG signal decoding are verified by comparative experiments. In section 5, this paper is summarized, and the prospect of future research is discussed.

2 Preliminary

This paper mainly adopts DDQN to perform the end-to-end decoding operation of EEG, and the related concepts and basic knowledge are introduced as follows.

2.1 Reinforcement Learning

RL is a learning framework for dealing with sequential decision problems, which can usually be modeled as a Markov Decision Process (MDP) that can be represented by a five-tuple (S, A, P, R, γ), where:

1) S denotes the state space and $s_t \in S$ represents the state of the agent at the moment t;

2) A denotes the action space and $a_t \in A$ represents the action executed by the agent at time t;

3) $P: S \times A \times S \rightarrow [0, 1]$ denotes the state transition probability, and $P(s_{t+1} | s_t, a_t)$ denotes the probability that the agent executes the action a_t in state s_t to the next state s_{t+1};

4) $R: S \times A \rightarrow R$ denotes the reward function and $R(s_t, a_t)$ represents the immediate reward obtained by the agent by executing the action a_t in the state s_t;

5) $\gamma \in [0, 1]$ is the discount factor used to balance immediate and delayed rewards.

The action selection of an agent in reinforcement learning obeys the policy π, which is expressed as the mapping relationship $\pi: S \rightarrow A$ between the state and the executable action of the agent. RL algorithms can be classified into two categories, policy-based and value-based methods. In the value-based RL method, the policy will not be updated explicitly, but a value table or value function is maintained, and new policies are derived from this value table or value function. The state-action value function is $Q^{\pi}: S \times A \rightarrow R$. $Q^{\pi}(s_t, a_t)$ represents the expected cumulative reward obtained by the agent executing action a_t in state s_t and following the current policy π until the end of the episode, which can be expressed as:

$$Q^{\pi}(s_t, a_t) = E_{\pi}\left\{\sum_t \gamma^t R(s_t, \pi(s_t)) \mid s_t = s,\ a_t = a\right\}. \quad (1)$$

The ultimate goal of the agent is to learn an optimal policy π^*, and the value function obtained under the optimal policy satisfies the Bellman optimality equation: $V^*(s_t) = \max_{a \in A} Q^*(s,a)$, i.e., the optimal value of the state is equal to the expected cumulative reward obtained by taking the optimal action in that state, and the optimal policy can be obtained from $\pi^* \in \operatorname{argmax}_a Q^*(s,a)$.

Q-learning is an RL algorithm based on value iteration, which directly estimates the optimal state-action value function Q^*. It updates the Q-function by the following rule:

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha\Big(R(s_t,a_t) + \gamma \max_a Q(s_{t+1},a) - Q(s_t,a_t)\Big), \tag{2}$$

where s_{t+1} represents the next state reached by the agent executing action a_t in state s_t, and $\alpha \in [0,1]$ represents the learning rate. A common approach to deriving a policy based on the Q-function is the ε-greedy policy, which selects an action greedily with the probability of $1-\varepsilon$ based on the Q-function and performs any action randomly with the probability of ε. This facilitates the exploration of the agent in the environment and avoids falling into a local optimum.

2.2 Double Deep Q-Networks

When the state space is large or continuous, it is impractical to directly use the tabular Q-function for storing the values of all state-action pairs. A common solution is to approximate the Q-function using a function approximator, e. g., $Q(s_t,a_t) \approx Q(s_t,a_t,\theta)$, where $Q(s_t,a_t,\theta)$ represents the parametrized approximation of the Q-function. Specifically, DQN is a method that approximates the state action value function by the Q-learning algorithm through a neural network. In DQN, deep learning and reinforcement learning are combined through a convolutional neural network to approximate the state action value function, and high-dimensional states can be input, which solves the dimensional disaster problem faced by traditional Q-learning.

In the traditional Q-learning algorithm and DQN algorithm, directly selecting the action with the maximum Q value may cause the Q-value overestimation problem, which leads to over-optimistic estimation. DDQN[28] separates action selection and action value evaluation to avoid the overestimation problem. Like DQN, DDQN has two important ideas: the target network and experience replay mechanism. At each time step t in DDQN, the agent executes action a_t in current state s_t based on the current policy, receives the reward $R(s_t,a_t)$, and transforms to the next state s_{t+1}. The transition $(s_t,a_t,R(s_t,a_t),s_{t+1})$ is added to the experience pool D. The neural network parameters are continuously updated by a gradient descent minimization loss function. The neural network parameters are continuously updated by minimizing the loss function through gradient descent, where the loss function is expressed as the mean square error between the target value and the evaluated value, which is defined as:

$$L(\theta) = E_{(s_t,a_t,R(s_t,a_t),s_{t+1})}[\left(y^{\mathrm{DQN}} - Q(s_t,a_t;\theta)\right)^2], \tag{3}$$

where the target value y^{DQN} is defined as:

$$y^{\mathrm{DQN}} = R(s_t,a_t) + \gamma \max_{a'} Q(s_{t+1},a';\theta^-). \tag{4}$$

In Eqs. (3) and (4), θ denotes the online network parameters, θ^- denotes the target network parameters. $Q(s_t,a_t;\theta)$ denotes the online network output, and $Q(s_t,a_t;\theta^-)$ denotes the target network output, which is used to calculate the target value, where the target network has the same structure as the online network, except that its parameter values are replicated from the online network without τ steps, and the parameter representations of the target network remain unchanged during τ time steps.

The idea of DDQN is to decouple the action of selecting the maximum value in the target value and evaluating the value of the action, thus avoiding the problem of overestimation. DDQN uses the target network in DQN as the network for evaluation, without having to introduce an additional network. Therefore, in DDQN, the action is selected using the current Q-network, and then its value is evaluated using the target network. Its target value y^{DDQN} is:

$$y^{\mathrm{DDQN}} = R(s_t,a_t) + \gamma Q\Big(s_{t+1}, \arg\max_a Q(s_{t+1},a;\theta);\theta^-\Big). \tag{5}$$

The difference between DDQN and DQN is that the selection of the optimal action in DDQN is based on the online network Q with parameter θ, whereas the selection of the optimal action in DQN is based on the target network with parameter θ^-. VAN HASSELT et al.[28] have experimentally demonstrated that compared with DQN, DDQN can effectively reduce overestimation and obtain more stable learning.

2.3 Reinforcement Learning Brain Computer Interfaces

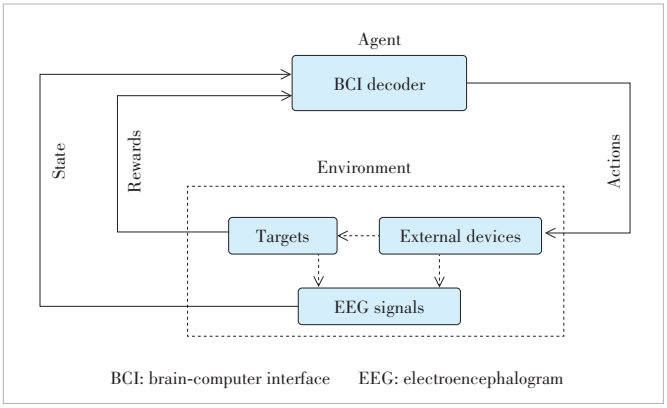
In recent years, RL has become a significant research interest in artificial intelligence. Through trial and error, the RL agent must discover which actions yield the maximum expected reward. Thus, the RL-based BCI attempts to allow BCI control algorithms to learn to complete tasks from interactions with the environment rather than explicit training signals. In fact, for many patients using BCI, the only signals available are their internal brain intention to complete the motor task and external feedback after completing the task, as opposed to specific supervised signals. The RL-based BCI attempts to learn a control policy by which, at any time t, the neural decoder observes a neural state $s_t \in S$, and the neural decoder outputs an action $a_t \in A$ based on the current policy, which

generates a control signal to the external device. After the external device completes the action, the neural decoder receives a feedback signal R_t. In future tasks, the neural decoder uses this feedback to continuously adjust the policy, which learns the optimal function mapping of the neural state to the action directly. The decoding structure is shown in Fig. 1.

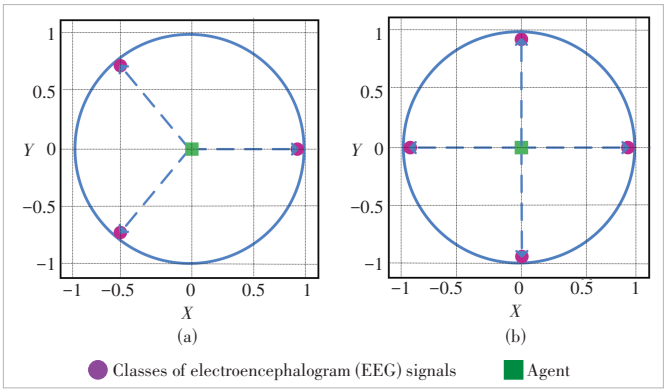
3 DDQN-Based EEG Decoder

An EEG signal decoder based on the Q-KTD RL algorithm provides the possibility of continuous learning of BCI, but its generalization is not negligible for a continuously useful decoder. Therefore, we try to use DDQN to decode the EEG signal correctly in this paper.

For the decoding task of EEG signals, a BCI decoder is considered a reinforcement learning agent, and the decoding of EEG signals is modeled as a common center-out task for BCI, associating the class of MI data with a specific direction, modeled as a single-step reinforcement learning problem, as shown in Fig. 2. A reinforcement learning environment located at the center of the origin (0, 0) with a radius of 1 is set up. In the center-out task, the reinforcement learning agent (the green square in Fig. 2) is located at the center of the origin (0, 0) at the beginning of each trial. By decoding each trial's MI data, the BCI decoder generates a specific action (one of up, down, left, and right), and the agent (located at the origin position (0, 0)) moves a distance of length 1 in the corresponding direction to a corresponding location (one of the purple circles in Fig. 2), and then receives an immediate reward based on the location reached by the agent. This paper uses a double deep Q-networks algorithm to train the agent to obtain a BCI decoder to decode EEG signals correctly.



▲Figure 1. Reinforcement learning (RL)-based BCI decoding structure



▲Figure 2. (a) Classical dataset and (b) BCI Competition IV-2a (BCI-2a) dataset are set to a center-out task. The center is located at the origin (0, 0), represented by a green square, and each class target is a purple circle

The state vector of DDQN is the EEG signal, and the agent takes action based on the current state. The optional action of the agent is the same as the label set of the EEG signal. According to the label information of the EEG signal, the feedback from the environment can be received, and the reward of the environment feedback contains two values of −1 and 1. If the current action performed by the agent is consistent with the label of the EEG signal, the environment feeds a positive reward value. Otherwise, the environment provides a negative value to the agent. The pseudocode of the algorithm is given by Algorithm 1. Moreover, we give the network architecture of the DDQN-based EEG signal decoder in Fig. 3.

Algorithm 1. DDQN-based EEG decoding

Input: the empty replay buffer D, initial network parameters θ, copy of $\theta\ \theta^{-}$, EEG signal sequences X, the training batch size N_b, explore probabilistic decay frequency N_ε, and target network replacement frequency N^{-}.
Output: action a_t
For episode=1 to M **do**
  Randomly initialize EEG signal sequences X
  If episode mod $N_\varepsilon = 0$
    $\varepsilon = \varepsilon \times$ RLepsilonDecayRate
  End if
  For t=0 to T **do**
    Set state $s_t \leftarrow X$ and select action a_t based on the ε-greedy policy
    Execute action a_t and observe reward r_t
    Store (s_t, a_t, r_t, s_{t+1}) in D
    Sample a minibatch of N_b tuples $(s, a, r, s') \sim \mathrm{Unif}(D)$
    Construct target values, one for each of the N_b tuples:

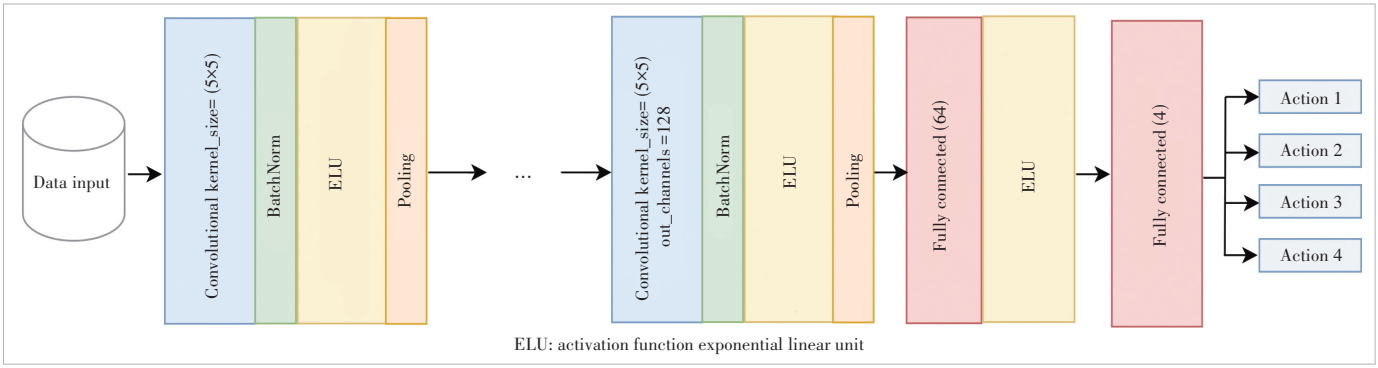
$$y_j^{\mathrm{DDQN}} = \begin{cases} r_j, & \text{if } s_{j+1} \text{ is terminal} \\ r_j + \gamma Q\left(s_{t+1}, \arg\max_a Q(s_{t+1}, a; \theta); \theta^{-}\right), & \text{otherwise} \end{cases}$$

    Do a gradient descent step with loss $\| y_j^{\mathrm{DDQN}} - Q(s_j, a_j; \theta)\|^2$
    Replace target parameters $\theta^{-} \leftarrow \theta$ every N^{-} steps
  End
End

4 Experimental Analysis

4.1 Experimental Data

We conducted experiments on two publicly available data sets: Nature's Scientific Data[30] and BCI Competition IV-2a



▲**Figure 3. Network structure of electroencephalogram (EEG) signal decoder based on double deep Q-network (DDQN)**

(BCI-2a)[31].

1) Experiment on Nature's Scientific Data: 21 electrodes were used to record EEG data from 13 healthy subjects, including 8 males and 5 females. This data set provided five different BCI paradigm data sets to imagine the movement of different body parts, such as the left hand, the right hand, or different finger movements. Among these available EEG data sets, we considered the first classical motor imagination dataset (Classical, CLA). The CLA dataset consisted of three types of motor imagination data using EEG signals, corresponding to left-hand movement, right-hand movement, and maintaining neutrality respectively. That is, the participants did not imagine anything. Six subjects in the CLA data were considered, specific to A to F subjects, since the "CLASubjectF1509163 StLRHand" data file contained only two category labels and was therefore not considered. The sampling frequency of the EEG signal was 200 Hz, and each collected data file contained 15 min sessions. Each 15-minute session included 300 trials. The total time of each trial was 3 s, which started with a one-second motion MI cue, and each trial lasted 1.5 – 2.5 s.

2) Experiment on BCI Competition IV-2a: This dataset is a publicly available dataset for BCI Competition IV, which is described in detail by TANGERMANN et al.[31] for the data characteristics of the competition. The BCI-2a dataset, which recorded EEG data from nine subjects using 22 electrodes, provided four different types of motor imagination data: left-hand movement, right-hand movement, exercise of both feet, and tongue movement imagination. The EEG signal was sampled at a frequency of 250 Hz, and the data for each subject consisted of two files, each consisting of six EEG recording blocks containing 48 trials, for a total of 576 trials.

4.2 Experimental Methods

The collection of the CLA data set is to extract the brain imagination data of 21 channels continuously for 0.85 s at a sampling frequency of 200 Hz, starting from the action stimulus for each subject[30]. For the BCI-2a dataset, some researchers have tried to use a relatively large window (about 3 s to 4 s) for their studies[32 – 33], but a relatively small window is more realistic for online BMI[27]. Therefore, the proposed method uses a 0.85-second EEG window to decode subjects' motor images. We first cropped the 0.85-second EEG signal at [0, 0.85] after the beginning of the trial for CLA and at [2, 2.85] after the beginning of the visual cue for BCI-2a. The CLA data consists of 21 channels, where the number of samples in each channel is 170 (the same as sampling frequency 200 Hz × 0.85 s), and the BCI-2a data set has a total of 22 channels, in which the number of samples per channel is 213 (about sampling frequency 250 Hz × 0.85 s).

For the CLA dataset, we evaluated the performance of the EEG signal within 0.85 s by dividing it into a training set and a test set, respectively, and as in Ref. [27], we executed 10 Monte Carlo trials at 100 episodes, and in each trial, the sequences of the trials were randomized. The performance was observed on the training set based on the success rate of the trials, which was calculated as the ratio of the number of successful trials at each step to reach the specified goal to the total number of trials considered. For the BCI-2a dataset, the performance was evaluated directly based on the existing training and test sets of each subject. In DDQN, we adopted the ε-greedy method for the exploration strategy, where the exploration probability of the intelligence was set to $\varepsilon = 0.1$ at the beginning of the trial and decays every 20 episodes, with each exploration probability decaying to half of the original one, so that the agent would be more inclined to be exploited as the trial progressed.

4.3 Feature Extraction

In order to illustrate the advantages of DDQN for EEG decoding, this paper compares the DDQN algorithm with classical supervised learning algorithms (the SVM, decision tree, and random forest) and the Q-KTD algorithm based on traditional reinforcement learning. In order to make the experiment more convincing, we follow the feature extraction approach introduced in Ref. [27] to construct Features 1 and 2, which is constructed as follows :

Feature 1: In order to obtain the complete information held in the EEG data, Feature 1 was extracted by first cropping the EEG signal data for 0.85 s, and then concatenating each channel of the cropped data as one motor imagery state vector for

each experiment. For example, for the BCI-2a dataset which has 22 channel numbers, each channel has 213 samples, so the size of the motion imagery state vector in each experiment is 4 686 (equal to 213 samples × 22 channels). Alternatively, for the CLA data, which has 21 channels, each channel contains 170 samples, so the size of the state vector in each experiment is 3 570 (equal to 170 samples × 21 channels).

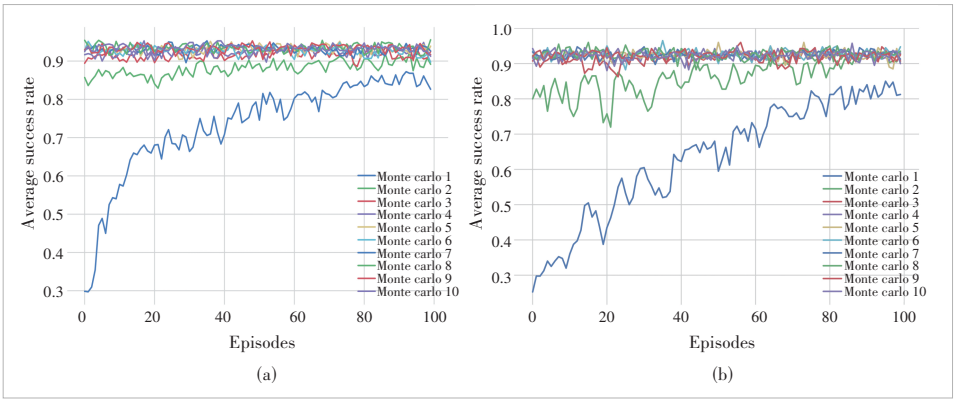
Feature 2: It has been proved that EEG signals contain frequency information[34], therefore, this paper adopts the same way as Ref. [27] to extract the frequency information as Feature 2. For the dataset BCI-2a and CLA dataset, firstly, the fast Fourier transform is performed on the EEG data with a cropped duration of 0.85 s, and then the selected complex frequency components correspond to the real and imaginary values, respectively. For the BCI-2a data, the real and imaginary values of the transformed 0 – 15 Hz were used for frequency classification, yielding a 550 dimensional feature state vector for each experiment. For the CLA dataset, using the frequency components between 0 – 5 Hz, a complex frequency feature with 5 dimensions of real values and 4 dimensions of imaginary values for each channel is obtained, and the components for each channel are concatenated to obtain a feature state vector with a dimension of 189 (equal to 9 × 21 channels).

4.4 Experimental Results and Analysis

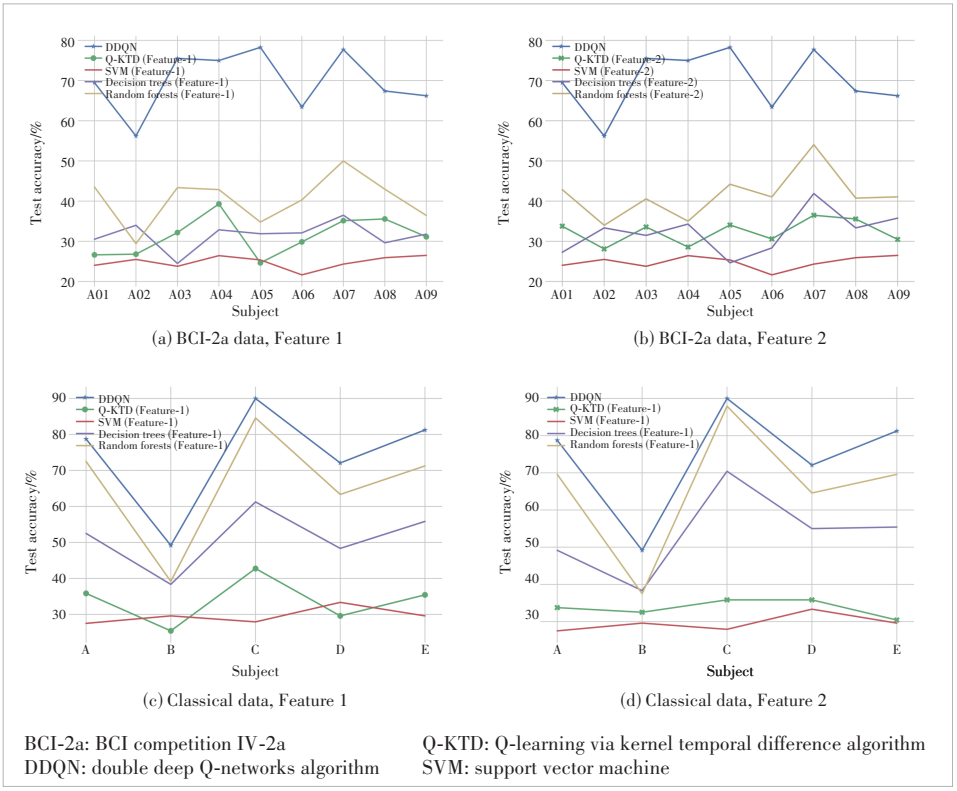
Firstly, the reinforcement learning agent is trained on the CLA and BCI-2a training sets, respectively, and its learning curve is observed. Fig. 4 shows the learning curves of the first subject on each of the two datasets. The learning curves show that the DDQN algorithm can correctly learn the correct mapping of EEG signals to actions directly in the MI center-out task with full learning by the agent as the experiment progresses.

Secondly, the generalization effect of the DDQN algorithm on the test set was compared with that of the traditional supervised learning algorithms (the SVM, decision tree, and random forest) and the Q-KTD algorithm. The result was shown in Fig. 5. For the same EEG data set, under different feature extraction, the generalization effect of DDQN algorithm on EEG decoding was significantly better than Q-KTD algorithm. Moreover, due to the small sample size and high dynamic characteristics of EEG signals, the classification performance of traditional supervised learning algorithms is poor, compared with DDQN-based EEG decoding.

Thirdly, the running time of DDQN and Q-KTD algorithms



▲ **Figure 4. Learning curves of double deep Q-network (DDQN) on (a) Classical (CLA) dataset and (b) BCI competition IV-2a EEG data (BCI-2a) dataset, where each color represents the average success rate of 10 Monte Carlo trials**



BCI-2a: BCI competition IV-2a
DDQN: double deep Q-networks algorithm
Q-KTD: Q-learning via kernel temporal difference algorithm
SVM: support vector machine

▲ **Figure 5. The generalizability of DDQN is compared with other classical algorithms for decoding based on Feature 1 (left) and Feature 2 (right) on the (a)(b)BCI-2a dataset and (c)(d) Classical (CLA) dataset, respectively**

is compared when training the agent, and the learning time of the agent under the two datasets is shown in Tables 1 and 2. According to the results, it is clear that the agent using DDQN learns much faster compared with the Q-KTD algorithm using Feature 1, and the learning time is not much different from the Q-KTD algorithm using Feature 2. However, the generalization of DDQN is much better than the Q-KTD algorithm.

Finally, in order to reflect the stability of decoding based on the deep reinforcement learning algorithm, we conducted 10 repeated experiments under different random seeds, and described their mean and standard deviation. The experimental results are shown in Fig. 6.

5 Conclusions

This paper investigates the applicability and feasibility of the deep double Q reinforcement learning algorithm in the brain-computer interface. We use two different EEG signal datasets and evaluate the performance of DDQN on both the datasets. The experimental results show that DDQN performs well in the correct decoding of EEG signals and has better generalization. This indicates that deep reinforcement learning can learn the correct decoding of EEG signals through feedback signals and has better generalization than the Q-KTD reinforcement learning algorithm. In the future, we will investigate further applications of deep reinforcement learning in EEG signals.

▼ Table 1. Comparison of learning time of DDQN and Q-KTD algorithm agent on classical (CLA) dataset

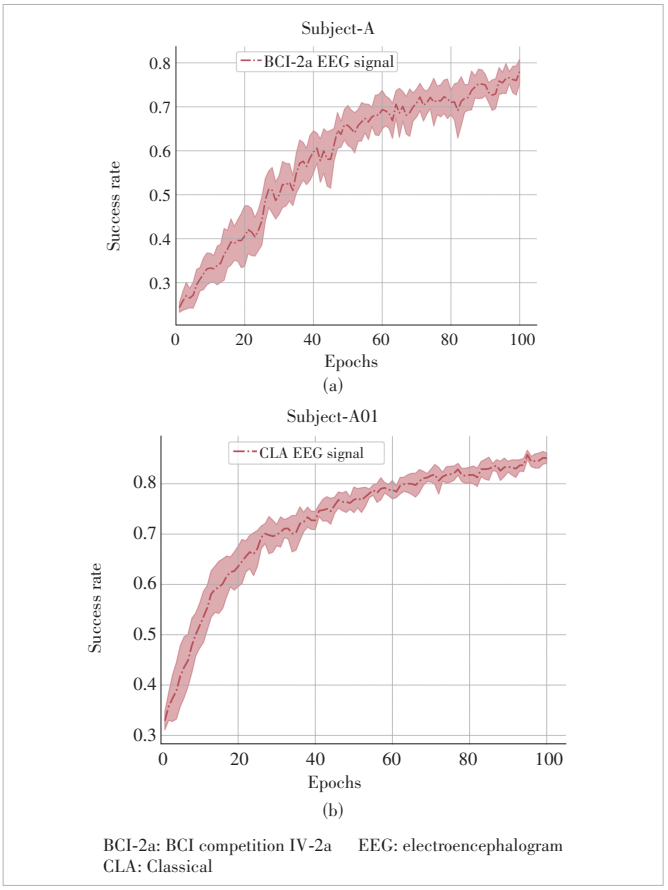
Subject	Algorithm		
	DDQN	Q-KTD_Feature 1/min	Q-KTD_Feature 2/min
A01	18.69	103.14	10.72
A02	9.3	117.8	10.94
A03	19.26	103.1	11.05
A04	16.55	79.73	8.82
A05	19.95	112.97	10.67
A06	14.96	54.86	6.43
A07	17.16	108.93	11
A08	18.11	103.73	10.53
A09	16.71	87.46	8.44

DDQN: double deep Q-networks algorithm
Q-KTD: Q-learning via kernel temporal difference algorithm

▼ Table 2. Comparison of learning time of DDQN and Q-KTD algorithm agent on BCI-2a dataset

Subject	Algorithm		
	DDQN	Q-KTD_Feature 1/min	Q-KTD_Feature 2/min
A	25.94	229.93	11.64
B	25.31	232.93	11.69
C	22.59	235.17	12.62
D	25.6	234.75	13.12
E	25.35	233.66	12.62

BCI-2a: BCI competition IV-2a
DDQN: double deep Q-networks algorithm
Q-KTD: Q-learning via kernel temporal difference algorithm



BCI-2a: BCI competition IV-2a EEG: electroencephalogram
CLA: Classical

▲ Figure 6. Results of experiments on (a) BCI-2a and (b) CLA under 10 different random seeds

References

[1] WILLETT F R, AVANSINO D T, HOCHBERG L R, et al. High-performance brain-to-text communication via handwriting [J]. Nature, 2021, 593(7858): 249 – 254. DOI: 10.1038/s41586-021-03506-2
[2] CRUZ A, PIRES G, LOPES A, et al. A self-paced BCI with a collaborative controller for highly reliable wheelchair driving: experimental tests with physically disabled individuals [J]. IEEE transactions on human-machine systems, 2021, 51(2): 109 – 119. DOI: 10.1109/THMS.2020.3047597
[3] SCHWARZ A, HÖLLER M K, PEREIRA J, et al. Decoding hand movements from human EEG to control a robotic arm in a simulation environment [J]. Journal of neural engineering, 2020, 17(3): 036010. DOI: 10.1088/1741-2552/ab882e
[4] SONG Y H, WU W F, LIN C Q, et al. Assistive mobile robot with shared control of brain-machine interface and computer vision [C]//4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). IEEE, 2020: 405 – 409. DOI: 10.1109/ITNEC48623.2020.9085096
[5] ANG K K, CHIN Z Y, WANG C C, et al. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b [J]. Frontiers in neuroscience, 2012, 6: 39. DOI: 10.3389/fnins.2012.00039
[6] TONIN L, CARLSON T, LEEB R, et al. Brain-controlled telepresence robot by motor-disabled people [C]//Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2011: 4227 – 4230. DOI: 10.1109/IEMBS.2011.6091049
[7] GROSSE-WENTRUP M, BUSS M. Multiclass common spatial patterns and information theoretic feature extraction [J]. IEEE transactions on biomedical engineering, 2008, 55(8): 1991 – 2000. DOI: 10.1109/TBME.2008.921154
[8] JIN J, XIAO R, DALY I, et al. Internal feature selection method of CSP based on L1-norm and Dempster-Shafer theory [J]. IEEE transactions on neural networks and learning systems, 2020, 32(11): 4814 – 4825. DOI: 10.1109/TNNLS.2020.3015505

[9] JIN J, MIAO Y, DALY I, et al. Correlation-based channel selection and regularized feature optimization for MI-based BCI [J]. Neural networks, 2019, 118: 262 – 270. DOI: 10.1016/j.neunet.2019.07.008
[10] FU R R, TIAN Y S, BAO T T, et al. Improvement motor imagery EEG classification based on regularized linear discriminant analysis [J]. Journal of medical systems, 2019, 43(6): 169. DOI: 10.1007/s10916-019-1270-0
[11] LIU Y X, ZHOU W D, YUAN Q, et al. Automatic seizure detection using wavelet transform and SVM in long-term intracranial EEG [J]. IEEE transactions on neural systems and rehabilitation engineering, 2012, 20(6): 749 – 755. DOI: 10.1109/TNSRE.2012.2206054
[12] SAMUEL O W, GENG Y J, LI X X, et al. Towards efficient decoding of multiple classes of motor imagery limb movements based on EEG spectral and time domain descriptors [J]. Journal of medical systems, 2017, 41(12): 194. DOI: 10.1007/s10916-017-0843-z
[13] LIN C T, CHUANG C H, HUNG Y C, et al. A driving performance forecasting system based on brain dynamic state analysis using 4-D convolutional neural networks [J]. IEEE transactions on cybernetics, 2021, 51(10): 4959 – 4967. DOI: 10.1109/TCYB.2020.3010805
[14] AMIN S U, ALSULAIMAN M, MUHAMMAD G, et al. Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion [J]. Future generation computer systems, 2019, 101: 542 – 554. DOI: 10.1016/j.future.2019.06.027
[15] TAYLOR D M, TILLERY S I H, SCHWARTZ A B. Direct cortical control of 3D neuroprosthetic devices [J]. Science, 2002, 296(5574): 1829 – 1832. DOI: 10.1126/science.1070291
[16] GAGE G J, LUDWIG K A, OTTO K J, et al. Naïve coadaptive cortical control [J]. Journal of neural engineering, 2005, 2(2): 52 – 63. DOI: 10.1088/1741-2560/2/2/006
[17] VELLISTE M, PEREL S, SPALDING M C, et al. Cortical control of a prosthetic arm for self-feeding [J]. Nature, 2008, 453(7198): 1098 – 1101. DOI: 10.1038/nature06996
[18] SUTTON R S, BARTO A G. Reinforcement learning: an introduction [M]. Cambridge, USA: MIT press, 2018
[19] DIGIOVANNA J, MAHMOUDI B, FORTES J, et al. Coadaptive brain: machine interface via reinforcement learning [J]. IEEE transactions on biomedical engineering, 2009, 56(1): 54 – 64. DOI: 10.1109/TBME.2008.926699
[20] ITURRATE I, MONTESANO L, MINGUEZ J. Robot reinforcement learning using EEG-based reward signals [C]//IEEE International Conference on Robotics and Automation. IEEE, 2010: 4822 – 4829. DOI: 10.1109/ROBOT.2010.5509734
[21] MATSUZAKI S, SHIINA Y, WADA Y. Adaptive classification for brain-machine interface with reinforcement learning [C]//18th International Conference on Neural Information Processing. ICONIP, 2011: 360 – 369. DOI: 10.1007/978-3-642-24955-6_44
[22] MAHMOUDI B, SANCHEZ J C. A symbiotic brain-machine interface through value-based decision making [J]. PLoS one, 2011, 6(3): e14760. DOI: 10.1371/journal.pone.0014760
[23] SANCHEZ J C, TARIGOPPULA A, CHOI J S, et al. Control of a center-out reaching task using a reinforcement learning brain-machine interface [C]//5th International IEEE/EMBS Conference on Neural Engineering. IEEE, 2011: 525 – 528. DOI: 10.1109/NER.2011.5910601
[24] POHLMEYER E A, MAHMOUDI B, GENG S J, et al. Using reinforcement learning to provide stable brain-machine interface control despite neural input reorganization [J]. PLoS one, 2014, 9(1): e87253. DOI: 10.1371/journal.pone.0087253
[25] MARSH B T, TARIGOPPULA V S A, CHEN C, et al. Toward an autonomous brain machine interface: integrating sensorimotor reward modulation and reinforcement learning [J]. Journal of neuroscience, 2015, 35(19): 7374 – 7387. DOI: 10.1523/jneurosci.1802-14.2015
[26] BAE J, CHHATBAR P, FRANCIS J T, et al. Reinforcement learning via kernel temporal difference [C]//Annual International Conference of IEEE Engineering in Medicine and Biology Society. IEEE, 2011: 5662 – 5665. DOI: 10.1109/IEMBS.2011.6091370
[27] THAPA B R, TANGARIFE D R, BAE J. Kernel temporal differences for EEG-based reinforcement learning brain machine interfaces [C]//44th Annual International Conference of IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2022: 3327 – 3333. DOI: 10.1109/EMBC48229.2022.9871862
[28] VAN HASSELT H, GUEZ A, SILVER D. Deep reinforcement learning with double Q-learning [C]//AAAI Conference on Artificial Intelligence. ACM, 2016: 2094 – 2100. DOI: 10.5555/3016100.3016191
[29] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari with deep reinforcement learning [C]//NIPS Deep Learning Workshop. NIPS, 2013. DOI: 10.48550/arXiv.1312.5602
[30] KAYA M, BINLI M K, OZBAY E, et al. A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces [J]. Scientific data, 2018, 5(1): 1 – 16. DOI: 10.1038/sdata.2018.211
[31] TANGERMANN M, MÜLLER K R, AERTSEN A, et al. Review of the BCI competition IV [J]. Frontiers in neuroscience, 2012, 6: 55. DOI: 10.3389/fnins.2012.00055
[32] QI F, WANG W, XIE X, et al. Single-trial eeg classification via orthogonal wavelet decomposition-based feature extraction [J]. Frontiers in Neuroscience, 2021, 15: 715855. DOI: 10.3389/fnins.2021.715855
[33] MUSALLAM Y K, ALFASSAM N I, MUHAMMAD G, et al. Electroencephalography-based motor imagery classification using temporal convolutional network fusion [J]. Biomedical signal processing and control, 2021, 69: 102826. DOI: 10.1016/j.bspc.2021.102826
[34] KEERTHI KRISHNAN K, SOMAN K P. CNN based classification of motor imaginary using variational mode decomposed EEG-spectrum image [J]. Biomedical engineering letters, 2021, 11(3): 235 – 247. DOI: 10.1007/s13534-021-00190-z

Biographies

REN Min received her BS degree from the School of Mathematics and Information, China West Normal University in 2021. She is currently working toward an MS degree at Southwest Jiaotong University, China. Her research interests include reinforcement learning and its application and data mining.

XU Renyu (ryxu@swjtu.edu.cn) received her PhD degree in mathematics and information science from Paris Saclay University, France in 2017. She is currently a lecturer in the school of mathematics, Southwest Jiaotong University, China. Her current research interests include reasoning with graph theory and machine learning.

ZHU Ting received her BS degree from the School of Mathematics and Software Science, Sichuan Normal University, China in 2021. She is now working toward an MS degree at Southwest Jiaotong University, China. Her research interests include reinforcement learning and deep reinforcement learning.

Multi-Agent Hierarchical Graph Attention Reinforcement Learning for Grid-Aware Energy Management

FENG Bingyi, FENG Mingxiao, WANG Minrui,
ZHOU Wengang, LI Houqiang
(University of Science and Technology of China, Hefei 230026, China)

DOI: 10.12142/ZTECOM.202303003
https://kns.cnki.net/kcms/detail/34.1294.TN.20230829.1650.002.html, published online August 30, 2023
Manuscript received: 2023-06-10

Abstract: The increasing adoption of renewable energy has posed challenges for voltage regulation in power distribution networks. Grid-aware energy management, which includes the control of smart inverters and energy management systems, is a trending way to mitigate this problem. However, existing multi-agent reinforcement learning methods for grid-aware energy management have not sufficiently considered the importance of agent cooperation and the unique characteristics of the grid, which leads to limited performance. In this study, we propose a new approach named multi-agent hierarchical graph attention reinforcement learning framework (MAHGA) to stabilize the voltage. Specifically, under the paradigm of centralized training and decentralized execution, we model the power distribution network as a novel hierarchical graph containing the agent-level topology and the bus-level topology. Then a hierarchical graph attention model is devised to capture the complex correlation between agents. Moreover, we incorporate graph contrastive learning as an auxiliary task in the reinforcement learning process to improve representation learning from graphs. Experiments on several real-world scenarios reveal that our approach achieves the best performance and can reduce the number of voltage violations remarkably.

Keywords: demand-side management; graph neural networks; multi-agent reinforcement learning; voltage regulation

Citation (Format 1): FENG B Y, FENG M X, WANG M R, et al. Multi-agent hierarchical graph attention reinforcement learning for grid-aware energy management [J]. ZTE Communications, 2023, 21(3): 11 - 21. DOI: 10.12142/ZTECOM.202303003
Citation (Format 2): B. Y. Feng, M. X. Feng, M. R. Wang, et al., "Multi-agent hierarchical graph attention reinforcement learning for grid-aware energy management," *ZTE Communications*, vol. 21, no. 3, pp. 11 - 21, Sept. 2023. doi: 10.12142/ZTECOM.202303003.

1 Introduction

The increasing shortage of fossil fuels and growing awareness of the need for environmental protection have made the adoption of solar photovoltaic (PV) power generation an important trend in the development of renewable energy. In recent years, more and more PV systems have been integrated into power distribution networks, owing to their low-carbon, clean, and economical benefits. However, the growing popularity of PV systems poses significant challenges to the stability of the power grid voltage. Thus, the need to make optimal use of the existing controllable resources in the power grid to ensure safe and reliable operation, reduce energy waste, and improve the acceptance of renewable energy has gained widespread attention. Prior research has suggested that using an inverter to control PV power conversion can alleviate this issue[1 - 2]. In addition, various energy storage and energy demand responses are also recommended as a means of voltage regulation[3 - 4]. Therefore, a comprehensive scheme is required to coordinate the control among these resources to ensure the stable operation of the entire power system with high PV penetration, which is referred to as grid-aware energy management[5].

Meanwhile, multi-agent reinforcement learning (MARL) has demonstrated impressive efficacy not only in games[6 - 8] but also in real-world applications[9 - 10]. Recently, MARL has also been employed to tackle issues in the power grid[11]. Under a data-driven and model-free setting, MARL does not need precise environment modeling and can be applied in situations with high PV penetration compared with traditional methods[11]. Moreover, using MARL in the power grid also potentially reduces costs and is regarded to have plug-and-play capability[12].

For grid-aware energy management, buildings established on a specific node in a power distribution network are considered as agents, which need to control the charge/discharge rate or electric energy conversion rate of multiple components, such as PV and battery. As the electric energy consumed or

This work is supported by National Key R&D Program of China under Grant No. 2022ZD0119802 and National Natural Science Foundation of China under Grant No. 61836011.
ZHOU Wengang and LI Houqiang are the corresponding authors.

generated will pass through the distribution network and cause voltage fluctuations, the goal of grid-aware energy management is to control these components inside buildings to keep the voltage within the safe range while satisfying building users' energy demands. And grid-aware energy management can be formulated as a cooperative task since all agents share one common objective which is to stabilize voltages at every node in the whole distribution network. Ref. [5] applied deep reinforcement learning to grid-aware energy management as they used independent proximal policy optimization and rule-based control to stabilize voltage.

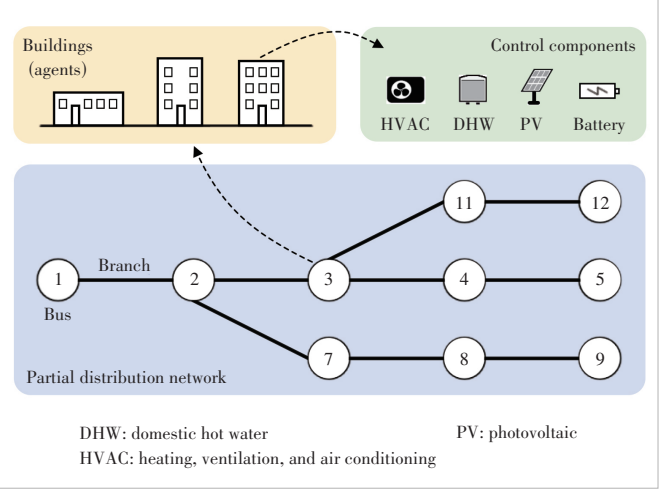
However, it is a non-trivial task to directly apply reinforcement learning algorithms to grid-aware energy management, because of the following challenges. 1) Cooperation among agents. The distribution network is a complex and nonlinear system, which results in a ripple effect to the voltage of all nodes within the distribution network if one agent takes action. Agents in previous work have been limited in their ability to learn cooperation by only utilizing their own observations during both the training and execution phases. This has resulted in difficulty in stabilizing voltage across all nodes. 2) Large state space from the large-scale agent system. There are hundreds of households on the power distribution network in reality. Directly learning a centralized agent system in a training process requires handling large state space and high-dimensional environments, which will cause serious scalability and efficiency problems[13]. 3) Topology of the distribution network. In the distribution network, each node is connected with some other nodes, forming a tree graph structure. The voltage of each node is affected by all other nodes, but the impact declines as the distance increases. Therefore, introducing the topology of the distribution network to algorithms can assist the agents in learning better correlations with each other. 4) The importance of different agents. Each building is regarded as an agent, but the building types are various and different types of buildings have different energy demands. For instance, typically, restaurants have more energy demand at noon for people to have meals, which indicates restaurants must pay more attention than offices when making decisions at noon.

To address the above challenges, we propose a multi-agent hierarchical graph attention reinforcement learning framework (MAHGA) to better stabilize voltage in a power distribution network. Our major contributions are summarized as follows: 1) We approach this task with the paradigm of centralized training and decentralized execution, enabling agents to learn better cooperation. 2) We model the whole distribution network as agent-level topology and bus-level topology. Based on these topologies, we construct an elaborate hierarchical graph attention architecture to extract correlations from agents and power grids. And it can facilitate the MARL-based methods deployed to the realistic power system. 3) Graph contrastive learning with two graph augmentations considered RL characteristics is designed as an auxiliary task in the reinforcement learning (RL) process to improve representation learning from graphs. 4) To the best of our knowledge, this is the first work to consider the topology characteristics to tackle voltage and energy tasks with large-scale agents. Experiments on several real-world datasets reveal that our approach achieves the best performance and can significantly mitigate voltage violations. The paper is organized as follows: In Section 2, we give the background of grid-aware energy management with the MARL formulations and introduce centralized training and decentralized execution. We describe the details of our method in Section 3. In Section 4, we demonstrate the results of the experiments, and in Section 5, we give a literature review of the related work. We conclude our work in Section 6.

2 Problem Formulation

2.1 Grid-Aware Energy Management

In the power system field, power distribution networks are modeled as a tree graph structure, where the node and edge represent a bus and a branch, respectively[14]. More specifically, a bus refers to a node in a power distribution network where power lines, buildings, and other electrical devices join together, and electrical power will be generated, distributed, or consumed here. The distribution network example is shown at the bottom of Fig. 1. For instance, the third bus in this figure is connected with the second bus, the fourth bus, and the eleventh bus. Hundreds of various buildings are distributed on these buses, and each building contains multiple controllable components: 1) HVAC: heating, ventilation, and air conditioning system, which consumes electricity primarily to control the temperature, humidity, and purity of the air inside a building affiliated with the storage to save cooling or thermal energy; 2) DHW: domestic hot water system, which can generate hot water by consuming electricity, affiliated with a tank to store hot water; 3) Battery: used for electricity storage or electricity supply to other equipment; 4) PV: photovoltaics, which is a micro-



▲Figure 1. An illustrative example of grid-aware energy management

generation device comprising solar cells.

The example of buildings and controllable components are shown at the top of Fig. 1. For instance, there are some buildings located on the third bus, and each building has the above four components to control to stabilize voltage after satisfying users' energy demands. Energy demand, including the use of HVAC and DHW, and other electric equipment/appliances (non-shiftable loads), as these components may constantly consume electricity from the power grid.

In terms of constructing the power models, grid-aware energy management environment GridLearn[5], grid models and AC power flows, etc., are modeled using Pandapower. The Pandapower library models the loads of the buildings with real and apparent power specifications; the PV arrays (and corresponding inverters) are modeled as PQ-controlled generators, which are defined to hold the active power P and reactive power Q constant while the voltage is allowed to vary over the limited range. It also calculates real and reactive power at each bus, load, and generator along with voltages at each bus. These values can be adapted to the state space or reward function. And they apply the preconfigured IEEE network model in it.

A large number of PV inside buildings will continuously inject power into the power grid and the power grid also needs to supply power frequently to meet the various users' energy demands, which may lead to frequent undervoltage or overvoltage problems in the power grid. Specifically, the voltage of each bus will be varied if it is injected with active power and reactive power. The exact numerical change of voltage is calculated with these two types of power through certain power flow formulas in the power flow model[5]. The formulas with physical quantities in the distribution network are complicated and non-linear in order to satisfy power system dynamics regulations[2].

The traditional control techniques for large-scale, complex, and non-linear systems are inadequate for real-time decision-making, particularly in systems with high penetration of renewable energy sources[11]. As a result, the employment of deep reinforcement learning algorithms has emerged as a potential and effective method in the literature to mitigate these difficulties.

2.2 MARL Formulations

The cooperative control process of grid-aware energy management can be modeled as a decentralized partially observable Markov decision process (DEC-POMDP)[15]. A DEC-POMDP is an extension of an MDP in decentralized multi-agent settings with partial observability. It can be defined by $\langle S, A, O, R, P, N, \gamma \rangle$, where S is the state space, A_i is the action space for agent i, $o_i = O(s; i)$ is the local observation for agent i at global state s, $P(s'|s, A)$ denotes the transition probability from S to S' given the joint action $A = (a_1, \cdots, a_n)$ for all N agents, $R(s, A)$ is the shared reward function and can also be called a global reward function, and $\gamma \in [0,1)$ is the discount factor. In a DEC-POMDP, each agent takes observation from the environment and executes an action generated by its policy to the environment. In turn, the environment provides one global feedback reward to all agents. During the interaction with the environment, the agents constantly adjust their policies to achieve the best decisions according to the rewards. Considering the grid-aware energy management problem, we describe specific elements in the DEC-POMDP in detail as follows, similar to Ref. [5].

Agent: As shown in Fig. 1, each building is regarded as an agent and will make control decisions on four components to maintain the voltage of all buses within a safe range.

Observation: The agent's observation incorporates 18 state spaces such as outdoor temperature, indoor temperature, voltage magnitude at the located bus, electricity generated by photovoltaic current, electricity consumed by base loads, current energy demand, time of day and the charging states of an HVAC storage device, a DHW storage device, and a battery.

Action: Each building controls four components, namely HVAC energy storage, DHW energy storage, battery storage, and inverters. The action made on each component is continuous and is all set in range $[-1,1]$. For the three energy storage components, the action denotes the increase (action>0) or decrease (action<0) of the energy's rate stored in the corresponding storage device. For the inverter, the action made on the inverter is used to scale the active power and reactive power supplied by PV and the battery.

Reward function: The reward function is mainly based on the voltage deviation from 1 p.u. for each bus. The term p.u. referred to "per unit" is used to express the voltage level in terms of a percentage of the nominal voltage. To alleviate the overvoltage and undervoltage problem across all buses, the reward function is calculated through the voltages on all buses. Specifically, let B denote the set of all buses in the distribution network, v_i denote the voltage on i's bus, and $\boldsymbol{\delta}_i$ a weighting factor to approximately normalize the reward function. The global reward function is calculated as follows:

$$R = -\sum_{i \in B} \left(\delta_i \left(v_i - 1 \right) \right)^2. \tag{1}$$

Note that this function limits the reward to 0 or negative and is devised to penalize the voltage rise deviation and the voltage drop deviation from 1 p.u. followed by Ref. [5]. Voltage deviations are typically measured from 1 p.u. (or 100% of the nominal voltage). For instance, if a 4% voltage deviation is allowed, the voltage safe range is from 0.96 p.u. to 1.04 p.u.

2.3 Centralized Training Decentralized Execution

Centralized training and decentralized execution (CTDE) is one of the paradigms in MARL which assumes that global in-

formation is available during training and that each agent can only use local information during execution to achieve decentralized execution[6 - 7,16]. In this paper, grid-aware energy management is formulated as a cooperative task because all agents share one common objective, which is to stabilize voltages at every bus in the whole distribution network. If each agent only observes local information on its located bus in the training phase, it is usually difficult to learn to control voltage within the safety range and guarantee service quality[2, 11]. One reason is that the environment is non-stationary if only considering the local observation where one agent's action can actually affect the whole distribution network[17].

As a result, we approach grid-aware energy management with the paradigm of CTDE. In CTDE, agents' information is shared in the training phase. In the execution phase and the time we evaluate the algorithm performance, agents are only allowed to make decisions based on their local observation. Specifically, in this paper, we improve and introduce our algorithms all based on the actor-critic class. After combining the structure of CTDE with actor-critic RL algorithms, the critic mainly assists the actor in learning during training, and the input of the critic is global information; while the input of the actor is local information, and the actor needs to make decisions independently in the execution phase. The advantages of CTDE for grid-aware energy management are twofold. On one hand, the centralized training process can motivate multiple agents to learn cooperation by perceiving a more comprehensive landscape. On the other hand, the execution process is fully decentralized without requiring complete information in the training phase, which guarantees efficiency and flexibility in online management. By applying CTDE, the learned strategies can be deployed to the power grid and achieve cooperative control without any communication device. Note that the paradigm of centralized training and centralized execution does not apply to this task due to commercial settings and users' privacy provision[18].

3 Method

In order to address the aforementioned challenges, we approach this grid-aware energy management task with the CTDE paradigm and propose a novel MAHGA approach. In the following, we first introduce the construction of the graph topology. After that, we discuss our hierarchical graph attention architecture for the critic to better extract agents' correlations. Finally, graph contrastive learning is devised as an auxiliary task in the training process to improve representation learning from graphs. The overview of MAHGA is shown in Fig. 2, where the agent takes action depending on its own observation by using the policy. In the training phase, the critic predicts global value based on all agents' observations and is updated by RL loss and graph contrastive loss. The policy is updated by corresponding RL loss with the predicted value from the critic. When in the execution phase, only the policy is used and it makes decisions by solely using agents' local observation.

3.1 Graph Topology Modeling

To capture the correlation between agents, we consider the unique characteristics of distribution networks and construct two graph structures, agent-level graph topology $G^1\left(V^1,D^1\right)$ and bus-level graph topology $G^2\left(V^2,D^2\right)$, respectively. Note that G represents graph topology, V represents the set of all nodes in the graph, and D represents the adjacency matrix which indicates how nodes are connected. For instance, if node i is connected to node j, D_{ij} equals 1; otherwise, D_{ij} equals 0. As for the agent-level graph topology, every agent is modeled as a node. The node set V_1 consists of all agents in the environment. We devise two types of operations to connect edges. The first is the operation of nodes on the same bus where all nodes on the same bus are connected with each other. Nodes on the same bus form a complete graph. The first operation for connecting edges is defined as follows:

$$D_{ij}^{1'}=\begin{cases}1, & b(i)=b(j)\\ 0, & \text{otherwise}\end{cases}, \tag{2}$$

where $b(i)$ denotes the bus, on which node i is located.

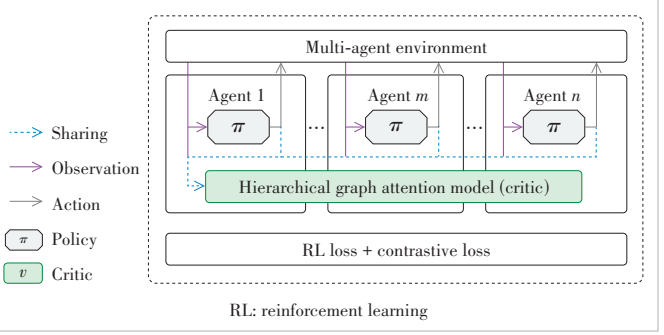
The second operation is to connect nodes from the adjacent buses. In detail, all nodes on bus i will be connected to the nodes on bus j, if bus i and bus j are connected in the distribution network. Different from the first operation, this operation makes all nodes on two adjacent buses form a complete bipartite graph. The second operation for connecting edges is defined as follows:

$$D_{ij}^{1''}=\begin{cases}1, & \text{if } b(i) \text{ and } b(j) \text{ are adjacent}\\ 0, & \text{otherwise}\end{cases}. \tag{3}$$

Then the adjacency matrix obtained from these two operations is taken as the union to form the final adjacency matrix for agent-level graph topology:

$$D_{ij}^{1}=D_{ij}^{1'}\cup D_{ij}^{1''}. \tag{4}$$

To sum up, the first operation is to model the relationship of all the buildings on the same bus, and the second is to model



▲ **Figure 2. Overview of multi-agent hierarchical graph attention (MAHGA), where each agent has one policy and shares the same critic**

the relationship of different buildings on adjacent buses.

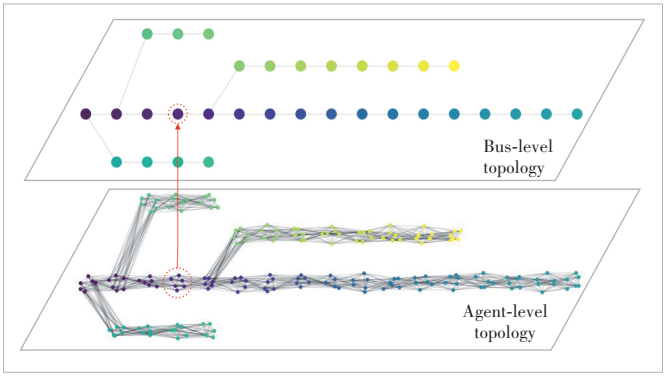
As for the bus-level graph topology, the agents from the same bus are treated as a cluster and thus every bus is modeled as a node. The node set V_2 consists of all the buses in a power distribution network. If two buses are connected in the distribution network, the corresponding node is set to be connected in G_2. The operations for connecting edges are defined as follows:

$$D_{ij}^2 = \begin{cases} 1, & \text{if } i \text{ and } j \text{ are adjacent} \\ 0, & \text{otherwise} \end{cases}. \tag{5}$$

For better illustration, we visualize one example of graph topology in Fig. 3, where nodes enclosed in the red dotted circle are one of the buses and all the agents located on it.

3.2 Hierarchical Graph Attention Architecture

We now present the architecture that exploits the graph topology to handle various observations. The pipeline of the architecture is shown in Fig. 4. The architecture consists of four main components: 1) the agent-level attention module that extracts agent-level representations from the agents' observations based on the agent-level graph topology $G^1\left(V^1,D^1\right)$; 2) the aggregation layer that clusters the agent-level nodes together and aggregates the representations to the embedding from the buses' point of view; 3) the bus-level attention module that extracts bus-level representations with the bus-level graph topology $G^2\left(V^2,D^2\right)$; 4) the readout layer and concatenation that scales down the size of representations and aggregates the representations from the above two attention layers to distill the final representations. The hierarchical characteristics of our architecture are mainly reflected in the different graph attention modules and readouts with corresponding pooling operations.



▲ **Figure 3. Visualized example of agent-level topology and bus-level topology in one case**

3.2.1 Agent-Level Attention Module

We first extract representations from agents' observations through an agent-level attention module using the agent-level graph topology mentioned above. Similar to Ref. [19], in graph attention networks, the importance of node j's feature to node i is calculated as:

$$e_{ij}^k = \vec{c}^T\left(W^k\vec{o}_i \| W^k\vec{o}_j\right), \tag{6}$$

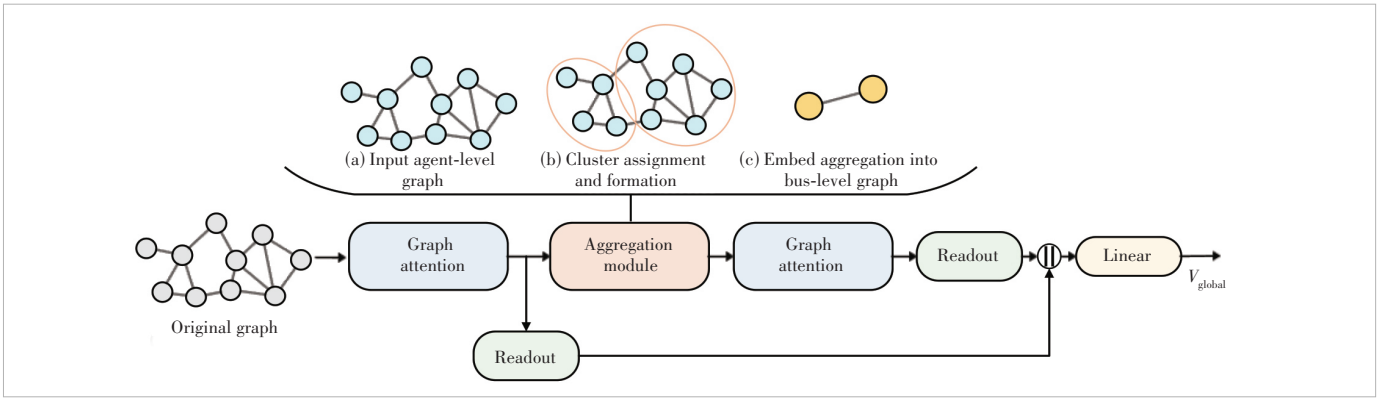
where $\vec{c}^T$ and W^k are learnable parameters, k is the k-th head among K multi-attention heads, $\cdot^T$ represents transposition and $\|$ is the concatenation operation. Then, the coefficients computed by the attention mechanism is defined as:

$$\alpha_{ij}^k = \frac{\exp\left(\text{LeakyReLU}\left(e_{ij}^k\right)\right)}{\sum_{v\in\mathcal{N}_i^1}\exp\left(\text{LeakyReLU}\left(e_{iv}^k\right)\right)}, \tag{7}$$

where $\mathcal{N}_i^1$ represents the set of node i's one-hop neighbor nodes in the graph topology G^1 and the LeakyReLU nonlinearity is applied.

Note that the mask graph attention is adopted and only the neighbor node is allowed to participate in the node i's attention coefficient calculations. The final output of node i in the attention network is formulated as:

$$\vec{h}_i^1 = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j\in\mathcal{N}_j^1}\alpha_{ij}^k W^k\vec{o}_j\right), \tag{8}$$



▲ **Figure 4. An overview of hierarchical graph attention architecture**

where σ represents the softmax nonlinearity. By applying the above steps to every node in G^1, we can get the agent-level representations h^1 for all nodes.

3.2.2 Aggregation Module and Bus-Level Attention Module

The implementation of a graph attention network is intrinsically flat, as it only propagates information across the edges of a graph. The purpose of this architecture is to define a strategy in a power distribution network that allows one to use two or more graph attention networks hierarchically to extract representations from the graph structure. Formally, given the input embedding, which is the output of the upper network, we seek to define a strategy to output a new coarsened graph embedding.

The new graph embedding contains fewer nodes and node connectivity and can then be used as input to another graph attention module.

As a result, the aggregation module is designed primarily to cluster the agent-level nodes into the classes of buses which means that the agent-level embedding $\vec{h}^1$ will be transformed to the bus-level embedding $\vec{h}^{1'}$ via this module. The aggregation process is demonstrated in the upper part of Fig. 4. Specifically, bus j' s embedding $\vec{h}_j^{1'}$ is the calculation of the embedding of all the agents situated on bus j.

$$\vec{h}_j^{1'} = \phi\left(\left\|_{v,b(v)=j}^{V^1}\left(\vec{\mathrm{h}}_v^1\right)\right.\right), \tag{9}$$

where ϕ denotes the projection function in the aggregation module and $b(v)$ indicates the bus where node v is located.

Then, the bus embedding $\vec{h}^{1'}$ is transformed into the bus representation $\vec{h}^2$ through the bus-level attention module with the graph topology G_2. The structure of the bus-level graph attention module is similar to that of the agent-level attention module which utilizes Eqs. (6) – (8) to calculate the representation but the topology inserted is G_2.

3.2.3 Readout Layer and Concatenation

Inspired by JK-net architecture[20], which has proposed a readout layer that aggregates node features to make a fixed-size representation, we apply a permutation-invariant readout layer to an extracted and integrated representation of agents. The summarized output feature of the readout layer after the agent-level attention module is as follows:

$$\vec{f}^1 = \frac{1}{\left|V^1\right|}\sum_{i=1}^{|V^1|}\vec{h}_i^1 \Big\| \max_{i=1}^{|V^1|}\vec{h}_i^1, \tag{10}$$

where $\|$ is the concatenation operation. The pooling operation in the readout layer is mainly to distill essential information of the state into latent representation while dropping redundant information.

Similarly, we can obtain the integrated representation f^2 after the bus-level graph attention module. As shown in Fig. 4, we apply a readout layer after each attention module. Then the concatenation of each readout layer is applied to aggregate the features.

$$\vec{h}^3 = \left[\vec{f}^1 \middle\| \vec{f}^2\right]. \tag{11}$$

The final representation $\vec{h}^3$ is then fed into the linear function to predict the global state value.

3.3 Graph Contrastive Learning

According to the large-scale energy management environment where hundreds of agents lead to a high dimension of model input, it can be difficult to learn representations through RL objectives as it only depends on reward from the environment. Graph contrastive learning, which has proven its effectiveness on graph prediction tasks[21], has not yet been explored in reinforcement learning, mainly due to the different nature of the problem. Inspired by graph contrastive learning already used in graph prediction tasks and image contrastive learning used in a pixel-based environments[22], we devise a graph contrast learning objective as an auxiliary task in our reinforcement learning task. The objective is devised mainly to stimulate the MAHGA to learn better representation from high-dimensional and various observation inputs.

To apply graph contrastive learning to MARL, we first introduce augmentations that should be made to the graph. The graph augmentation methods include: 1) Observation masking. We randomly select agents and mask certain ratios of agents' observations. Observation masking drives models to recover masked agent observation using their unmasked information. The underlying assumption is that missing partial node attributes does not influence the model performance much. 2) Edge dropping. It is devised to remove the connectivity in G_1 by randomly dropping a certain ratio of edges. It indicates that the semantic meaning of G_1 has certain robustness to the edge connectivity pattern variances. We also follow an independent and identically distributed (i.i.d.) uniform distribution to drop each edge.

Specifically, given the graph data Z_q composed of graph topology G_1 and observations of all nodes from the training batch of the size N, Z_q will undergo graph data augmentations mentioned above to obtain two correlated graphs Z_q^i as a positive pair. The other N–1 graphs in the batch are also augmented to generate N–1 augmented graphs. Then, we utilize the normalized temperature-scaled cross-entropy loss (NT-Xent) for graph Z_q as:

$$l_n = -\log\frac{\exp\left(Z_i^q W_s Z_j^q/\tau\right)}{\sum_{k=1,k\neq q}^{N}\exp\left(Z_i^q W_s Z_j^k/\tau\right)}, \tag{12}$$

where we employ a bilinear product to evaluate the similarity

of pairwise instances. In the formula, τ denotes the temperature parameter, N denotes the size of the training batch and W_s are learnable parameters. Objective l_n will be taken as an auxiliary task to be jointly optimized with the RL objective. Here we select multi-agent proximal policy optimization (MAPPO)[8] as the base algorithm to describe the RL objective. Specifically, following the settings in PPO's clipped surrogate objective[23], we let $r_t(\theta)$ be the probability ratio calculated by the agent's policy and $\hat{A}_t$ be an estimator of the advantage function at timestep t calculated by the global state value. ε is a hyperparameter that implicitly restricts Kullback-Leibler (KL) divergence[23]. The RL objective l_r is defined as:

$$l_r = \min\left(r_t(\theta)\hat{A}_t, \text{clip}\left(r_t(\theta), 1-\varepsilon, 1+\varepsilon\right)\hat{A}_t\right). \tag{13}$$

4 Experiments

In this section, we first introduce the experiment setup. Then we demonstrate and analyze experiment results about overall performance and ablation study. All the experiments have been conducted based on the GridLearn open-source platform[5] with the IEEE 33-bus system[24].

4.1 Experiment Setup

4.1.1 Data Description

We conduct experiments on four real-world scenarios with different climate zones respectively. Each scenario includes 192 buildings distributed on buses, and the corresponding data for the whole year in the specific climate zone[12](climate zone 2A: hot-humid; climate zone 3A: warm-humid; climate zone 4A: mixed-humid; climate zone 5A: cold-humid).

For each scenario, we select four months from four different seasons for training, as different seasons have quite different temperatures, humidity, solar radiation, and users' energy demands, which probably leads to different control strategies. The four training months are mixed and used to train algorithms until convergence. The rest eight months are used for testing. Each month containing 2 880 timesteps is regarded as an episode. The training phase lasts sixteen episodes and each experiment is conducted using 5 random seeds. After the training phase, we evaluate the learned strategy on the test dataset.

4.1.2 Comparison Algorithms

The methods that we evaluated include rule-based control (RBC), independent advantage actor-critic (IA2C), independent proximal policy optimization (IPPO), multi-agent advantage actor-critic (MAA2C), and MAPPO. Specifically, RBC devised by the used environment[5, 12] makes decisions mainly based on the time of day. For example, at 6 a.m., the battery charges and the charge value is 0.138 3. Most of the devices will choose to discharge in the daytime and early evening, and charge at night. The IA2C and IPPO are actor-critic algorithms that directly apply single-agent reinforcement learning algorithms A2C[25] and PPO[23] to MARL. All agents are completely independent. The critic network approximates the expected return only depending on agent-specific observation. MAA2C and MAPPO[8], as an extension of A2C and PPO, are actor-critic algorithms but are in the CTDE paradigm. As extensions of independent algorithms, their critic learns a joint state value function where this centralized critic conditions on all agents' observations rather than the individual observation. And their actor can only use local observation to generate actions same as IA2C and IPPO. In contrast to MAA2C, MAPPO's main advantage is its combination of on-policy optimization with its surrogate objective function.

4.1.3 Evaluation Metrics

Following the evaluation settings in GridLearn[5], we use four metrics to evaluate the performance of algorithms. For better demonstration, we name these four metrics as follows.

1) The number of soft voltage violations (NSVV). It calculates the number of all buses' voltage that is not under control within the soft safe range. Note that the soft safe range of voltage is between 0.96 p.u. and 1.04 p.u.

2) Soft reduction rate (SRR). It calculates the proportion of the algorithm to reduce the number of voltages compared with the rule-based control strategy with the soft safe range. Specifically, for the learned algorithm C, SRR is defined as:

$$\text{SRR}_\text{C} = \frac{\text{NSVV}_\text{RBC} - \text{NSVV}_\text{C}}{\text{NSVV}_\text{RBC}}, \tag{13}$$

where NSVV_RBC and NSVV_C represent the number of soft voltage violations using the rule-based control and the learned algorithm C.

3) The number of hard voltage violations (NHVV). It calculates the number of all buses' voltage that is not under control within the hard safe range. Note that the hard safe range of voltage is between 0.97 p.u. and 1.03 p.u.

4) Hard reduction rate (HRR). It calculates the proportion of the algorithm to reduce the number of voltages compared with the rule-based strategy with the hard safe range. Similar to SRR, for the learned algorithm C, HRR is defined as:

$$\text{HRR}_\text{C} = \frac{\text{NHVV}_\text{RBC} - \text{NHVV}_\text{C}}{\text{NHVV}_\text{RBC}}, \tag{14}$$

where NHVV_RBC and NHVV_C represent the number of hard voltage violations using rule-based control and the learned algorithm C.

Note that NSVV and NHVV evaluate how the algorithm can do to prevent the voltage of all buses from getting out of the safe range, and the lower number represents the better. SRR and HRR describe how much performance the algorithm can enhance compared with the rule-based control method and the higher represents the better. There are two safe voltage ranges:

the soft one and the hard one. The hard range is a more challenging one to evaluate algorithms' performance. Exceeding the safe range frequently will cause lots of problems such as equipment damage and regional power outages.

4.2 Overall Performance

Table 1 reports the median NSVV, SRR, NHVV, and HRR of all algorithms. HMAA2C and HMAPPO refer to MAA2C and MAPPO applied with MAHGA. As shown in the table, our MAHGA framework improves MAA2C and MAPPO and is superior to all other baseline algorithms on four different scenarios concerning four metrics. Owing to the CTDE paradigm and the better learned representations correlative to the grid-aware energy management task, HMAPPO and HMAA2C achieve the best performance among all other algorithms. These two algorithms reduce the number of voltage violations significantly and increase the reduction rate considering both the soft safe range and the hard safe range.

CTDE algorithms, like MAPPO and MAA2C, all perform better compared with independent learning algorithms like PPO and A2C. This proves that centralized critic integrating all agents' observations to have a global perspective can assist agents to implicitly learn better cooperation. Furthermore, HMAPPO and HMAA2C consistently perform better than MAPPO and MAA2C, which validates that MAHGA can make further improvements and motivate the agent to learn a better policy in multiple ways. More analysis of MAHGA will be discussed in an ablation study. As RBC is a well-crafted strategy, independent learning algorithms only show slightly better performance, especially in climate zone 4A.

4.3 Ablation Study

In this section, we conduct an ablation study on MAHGA to further verify the significance of each component. As MAPPO performs better in most scenarios than MAA2C, we choose MAPPO as a representative algorithm to conduct ablation experiments. And the experimental result that MAPPO outperforms PPO according to Table 1 shows that cooperation is implicitly learned and plays an important role in decision making. The following variants of HMAPPO are evaluated on all scenarios: 1) HMAPPOS removes the hierarchical graph attention architecture but uses a single graph attention network with the corresponding readout layer so as to only extract the agent-level representations from the graph attention network; 2) HMAPPOC removes the auxiliary task of graph contrastive learning. As can be seen in Fig. 5, removing any component will cause performance degradation. If we do not consider extracting representations from the bus-level topology, the performance will be significantly degraded. If the bus level and agent level are neither considered, where the algorithm is the original MAPPO, the algorithm will suffer from a large state space where all agents' observations are concatenated and are unaware of the two topologies, which finally leads to low performance. Moreover, introducing attention mechanisms into graphs can implicitly let agents learn how to make decisions with the surrounding agents of different types. These demonstrate that the application of a hierarchical graph attention framework in grid-aware energy management tasks is significantly effective. Besides, we can observe that removing auxiliary tasks of graph contrastive learning will also lead to performance degradation, which indicates that graph contrastive learning can assist the framework to learn representations better.

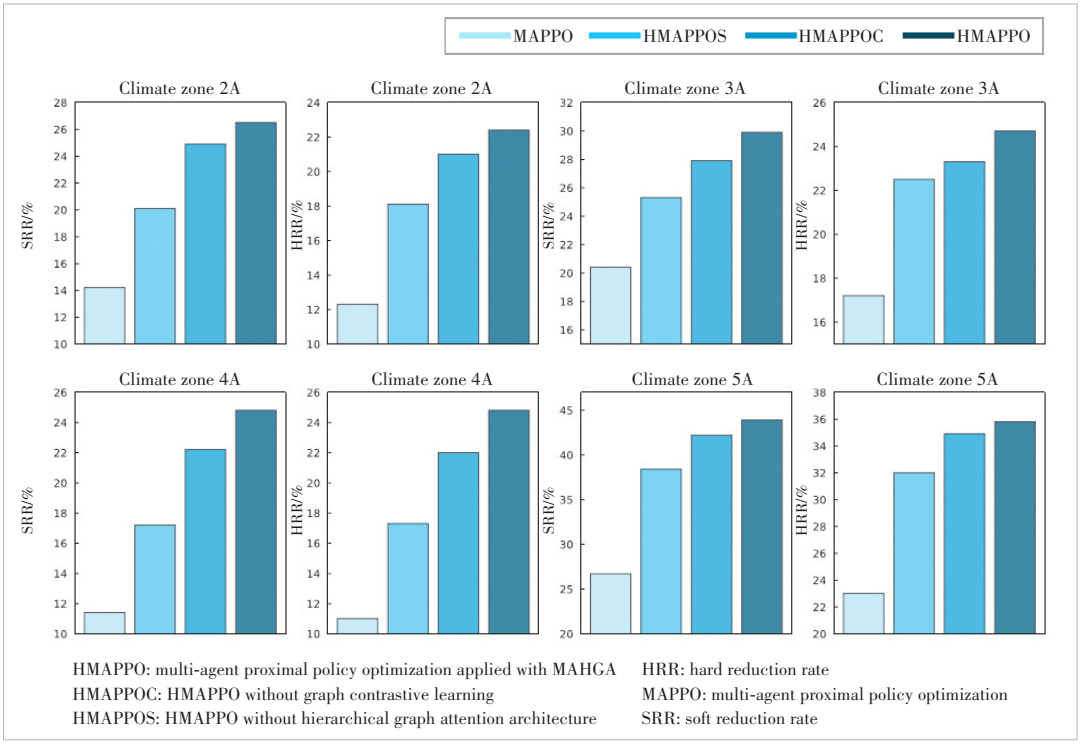
5 Related Work

1) Multi-agent reinforcement learning in power systems. Recently, efforts have been made to apply reinforcement learning to power systems for voltage regulation and energy management due to the progress of machine learning. Ref. [2, 26–27] introduce reinforcement learning in the active voltage control tasks. These works have considered managing a small number of agents and optimizing only reactive power components. In Refs. [2, 26], the load is inflexible and only the PV

▼**Table 1. Overall performance on four scenarios, where HMAA2C and HMAPPO refers to MAA2C and MAPPO applied with multi-agent hierarchical graph attention (MAHGA) (↓ denotes the lower the better, and ↑ denotes the higher the better)**

	Climate Zone 2A				Climate Zone 3A				Climate Zone 4A				Climate Zone 5A			
	NSVV ↓	SRR ↑	NHVV ↓	HRR ↑	NSVV ↓	SRR ↑	NHVV ↓	HRR ↑	NSVV ↓	SRR ↑	NHVV ↓	HRR ↑	NSVV ↓	SRR ↑	NHVV ↓	HRR ↑
RBC	86 181	0.0%	158 736	0.0%	110 902	0.0%	193 751	0.0%	83 648	0.0%	162 076	0.0%	106 823	0.0%	195 277	0.0%
A2C	79 905	7.3%	154 662	2.6%	101 102	8.8%	185 201	4.4%	81 648	2.4%	158 902	2.0%	93 365	12.6%	174 671	10.6%
PPO	79 601	7.6%	153 849	3.1%	100 954	9.0%	184 365	4.8%	81 224	2.9%	155 645	4.0%	92 920	13.0%	173 997	10.9%
MAA2C	73 264	15.0%	139 654	12.0%	89 423	19.4%	162 249	16.3%	74 569	10.9%	144 274	11.0%	79 369	25.7%	154 786	20.7%
MAPPO	73 919	14.2%	139 210	12.3%	88 236	20.4%	160 345	17.2%	74 126	11.4%	144 316	11.0%	78 314	26.7%	150 322	23.0%
HMAA2C	**64 516**	**25.1%**	**125 497**	**20.9%**	**78 158**	**29.5%**	**146 392**	**24.4%**	**63 105**	**24.6%**	**122 568**	**24.4%**	**60 766**	**43.1%**	**127 494**	**34.7%**
HMAPPO	**63 320**	**26.5%**	**123 116**	**22.4%**	**77 724**	**29.9%**	**145 946**	**24.7%**	**62 865**	**24.8%**	**121 829**	**24.8%**	**59 887**	**43.9%**	**125 386**	**35.8%**

A2C: advantage actor critic
HMAA2C: multi-agent advantage actor critic applied with MAHGA
HMAPPO: multi-agent proximal policy optimization applied with MAHGA
HRR: hard reduction rate
MAA2C: multi-agent advantage actor critic
MAPPO: multi-agent proximal policy optimization
NHVV: number of hard voltage violations
NSVV: number of soft voltage violations
PPO: proximal policy optimization
RBC: rule-based control
SRR: soft reduction rate



▲**Figure 5. Ablation studies on four scenarios**

inverter can be controlled. In Ref. [27], demand is regarded to be constant. CityLearn[12] is a platform that satisfies residential energy demands by controlling various shiftable components inside buildings. This environment has been widely used mainly to experiment with various reinforcement learning algorithms and their improvements, in comparison with reinforcement learning baselines and rule-based control. Besides, GridLearn[5], as an extension of CityLearn, considers both the grid-level and building-side objectives and introduces a more realistic environment where hundreds of agents are involved, and agents should control multiple components with corresponding resources to stabilize voltage in a power distribution network after satisfying the residential energy demand. IA2C[25] and IPPO[23] are used, which shows some effectiveness compared with the rule-based control method. In our paper, all experiments are conducted based on GridLearn. Apart from the power system field, in game-like environments, works have been proposed to better motivate the cooperation of agents, such as MAA2C and MAPPO[8], and they have shown good performance in some multi-agent game-like environments. Another approach to achieving agents' cooperation is to learn communication among multiple agents[28 – 30]. However, such approaches always lead to high communication overhead because of the large amount of information transfer.

2) Graph neural networks (GNNs) have been applied successfully to solve prediction and classification tasks in many real-world applications, including recommender systems, chemistry and bioinformatics[21, 31 – 32]. However, GNN has not yet been fully explored in MARL, mostly due to the different nature of the problem. Former works in Refs. [33 – 35] attempt to apply GNN to extract better representations from agents in game-like environments and Ref. [35] considers the unique nature of competitive games to extract information hierarchically. Their methods show encouraging performance with a few agents in games.

However, it is not yet clear whether MARL with GNN can still achieve competitive performance if applied to fully cooperative tasks with large-scale agents in real-world applications such as power systems. In the smart grid field, there are few works related to GNN. In Refs. [36 – 37], mask mechanisms and graph convolutional networks are applied to regulate voltage in single-agent reinforcement learning tasks.

6 Conclusions and Future Work

In this paper, we propose MAHGA, a novel multi-agent reinforcement learning framework for grid-aware energy management. Specifically, we first resolve the problem with the CTDE paradigm aiming to stimulate agents to learn cooperation strategy. Then, depending on modeling the distribution network to two different kinds of topology, we propose a hierarchical attention architecture to better extract agents' correlations from high-dimensional environment and capture the characteristics of grid. In addition, graph contrastive learning is designed to learn a more effective representation in the reinforcement learning training phase. Extensive experiments on four large-scale real-world scenarios have demonstrated the effectiveness of MAHGA where voltage violations can be significantly reduced compared with other baselines.

In our future work, we will explore the generalization over different climates and grid topologies, as well as the possibility of adding more energy control components that buildings can control, like electric vehicles.

References

[1] LIU H T, WU W C. Online multi-agent reinforcement learning for decentralized

inverter-based volt-VAR control [J]. IEEE transactions on smart grid, 2021, 12 (4): 2980 – 2990. DOI: 10.1109/TSG.2021.3060027
[2] WANG J H, XU W K, GU Y J, et al. Multi-agent reinforcement learning for active voltage control on power distribution networks [EB/OL]. (2021-10-27) [2023-06-10]. https://arxiv.org/abs/2110.14300
[3] WANG D X, MENG K, GAO X D, et al. Coordinated dispatch of virtual energy storage systems in LV grids for voltage regulation [J]. IEEE transactions on industrial informatics, 2018, 14(6): 2452 – 2462. DOI: 10.1109/TII.2017.2769452
[4] YI J, WANG P, TAYLOR P C, et al. Distribution network voltage control using energy storage and demand side response [C]//The 3rd IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe). IEEE, 2013: 1 – 8. DOI: 10.1109/ISGTEurope.2012.6465666
[5] PIGOTT A, CROZIER C, BAKER K, et al. GridLearn: multiagent reinforcement learning for grid-aware building energy management [J]. Electric power systems research, 2022, 213: 108521. DOI: 10.1016/j.epsr.2022.108521
[6] LOWE R, WU Y, TAMAR A, et al. Multi-agent actor-critic for mixed cooperative-competitive environments [C]//The 31st International Conference on Neural Information Processing Systems. ACM, 2017: 6382 – 6393. DOI: 10.5555/3295222.3295385
[7] RASHID T, DE WITT C S, FARQUHAR G, et al. QMIX: monotonic value function factorisation for deep multi-agent reinforcement learning [C]//35th International Conference on Machine Learning. ICML, 2018: 6846 – 6859. DOI: 10.48550/arXiv.1803.11485
[8] YU C, VELU A, VINITSKY E, et al. The surprising effectiveness of PPO in cooperative multi-agent games [J]. Advances in neural information processing systems, 2022, 35: 24611 – 24624. DOI: 10.48550/arXiv.2103.01955
[9] MA Y, HAO X, HAO J, et al. A hierarchical reinforcement learning based optimization framework for large-scale dynamic pickup and delivery problems [J]. Advances in neural information processing systems, 2021, 34: 23609 – 23620
[10] ZHANG H C, FENG S Y, LIU C, et al. CityFlow: a multi-agent reinforcement learning environment for large scale city traffic scenario [C]//The World Wide Web Conference. ACM, 2019: 3620 – 3624. DOI: 10.1145/3308558.3314139
[11] CHEN X, QU G N, TANG Y J, et al. Reinforcement learning for selective key applications in power systems: recent advances and future challenges [J]. IEEE transactions on smart grid, 2022, 13(4): 2935 – 2958. DOI: 10.1109/TSG.2022.3154718
[12] VAZQUEZ-CANTELI J R, DEY S, HENZE G, et al. Citylearn: standardizing research in multi-agent reinforcement learning for demand response and urban energy management [EB/OL]. (2020-12-18) [2023-06-10]. https://arXiv. org/abs/2012.10504
[13] PAPOUDAKIS G, CHRISTIANOS F, RAHMAN A, et al. Dealing with non-stationarity in multi-agent deep reinforcement learning [EB/OL]. (2019-6-11) [2023-06-10]. https://arXiv.org/abs/1906.04737
[14] BAKER K. Learning warm-start points for Ac optimal power flow [C]//The 29th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, 2019: 1 – 6. DOI: 10.1109/MLSP.2019.8918690
[15] OLIEHOEK F A, AMATO C. Finite-horizon dec-POMDPs [M]. A concise introduction to decentralized POMDPS. Cham: Springer, 2016: 33 – 40. DOI: 10.1007/978-3-319-28929-8_3
[16] FOERSTER J, FARQUHAR G, AFOURAS T, et al. Counterfactual multi-agent policy gradients [C]//The AAAI Conference on Artificial Intelligence. ACM, 2018: 2974 – 2982. DOI: 10.1609/aaai.v32i1.11794
[17] CHEN T Y, BU S R, LIU X, et al. Peer-to-peer energy trading and energy conversion in interconnected multi-energy microgrids using multi-agent deep reinforcement learning [J]. IEEE transactions on smart grid, 2022, 13(1): 715 – 727. DOI: 10.1109/TSG.2021.3124465
[18] GLATT R, DA SILVA F L, SOPER B, et al. Collaborative energy demand response with decentralized actor and centralized critic [C]//The 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation. ACM, 2021: 333 – 337. DOI: 10.1145/3486611.3488732
[19] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks [EB/OL]. (2017-10-30) [2023-06-10]. https://arxiv.org/abs/1710.10903
[20] XU K, LI C, TIAN Y, et al. Representation learning on graphs with jumping knowledge networks [C]//International conference on machine learning. PMLR, 2018: 5453 – 5462. DOI: 10.48550/arXiv.1806.03536
[21] YOU Y N, CHEN T L, SUI Y D, et al. Graph contrastive learning with augmentations [C]//The 34th International Conference on Neural Information Processing Systems. ACM, 2020: 5812 – 5823. DOI: 10.5555/3495724.3496212
[22] SRINIVAS A, LASKIN M, ABBEEL P. CURL: contrastive unsupervised representations for reinforcement learning [EB/OL]. (2020-04-08) [2023-06-10]. https://arxiv.org/abs/2004.04136
[23] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms [EB/OL]. (2017-08-28) [2023-06-10]. https://arXiv. org/abs/1707.06347
[24] BARAN M E, WU F F. Network reconfiguration in distribution systems for loss reduction and load balancing [J]. IEEE power engineering review, 1989, 9 (4): 101 – 102. DOI: 10.1109/MPER.1989.4310642
[25] MNIH V, BADIA A P, MIRZA M, et al. Asynchronous methods for deep reinforcement learning [C]//The 33rd International Conference on International Conference on Machine Learning. ACM, 2016: 1928 – 1937. DOI: 10.48550/arXiv.1602.01783
[26] CAO D, HU W H, ZHAO J B, et al. A multi-agent deep reinforcement learning based voltage regulation using coordinated PV inverters [J]. IEEE transactions on power systems, 2020, 35(5): 4120 – 4123. DOI: 10.1109/TPWRS.2020.3000652
[27] ZHANG Y, WANG X N, WANG J H, et al. Deep reinforcement learning based volt-VAR optimization in smart distribution systems [J]. IEEE transactions on smart grid, 2021, 12(1): 361 – 371. DOI: 10.1109/TSG.2020.3010130
[28] SUKHBAATAR S, SZLAM A, FERGUS R. Learning multiagent communication with backpropagation [C]//The 30th International Conference on Neural Information Processing Systems. ACM, 2016: 2252 – 2260. DOI: 10.5555/3157096.3157348
[29] FOERSTER J N, ASSAEL Y M, DE FREITAS N, et al. Learning to communicate with Deep multi-agent reinforcement learning [C]//The 30th International Conference on Neural Information Processing Systems. ACM, 2016: 2145 – 2153. DOI: 10.5555/3157096.3157336
[30] JIANG J C, LU Z Q. Learning attentional communication for multi-agent cooperation [C]//The 32nd International Conference on Neural Information Processing Systems. ACM, 2018: 7265 – 7275. DOI: 10.5555/3327757.3327828
[31] GILMER J, SCHOENHOLZ S S, RILEY P F, et al. Neural message passing for Quantum chemistry [C]//The 34th International Conference on Machine Learning. ACM, 2017: 1263 – 1272. DOI: 10.5555/3305381.3305512
[32] LEE J, LEE I, KANG J. Self-attention graph pooling [C]//International conference on machine learning. PMLR, 2019: 3734-3743. DOI: 10.48550/arXiv.1904.08082
[33] IQBAL S, SHA F. Actor-attention-critic for multi-agent reinforcement learning [C]//International Conference on Machine Learning. PMLR, 2019: 2961 – 2970. DOI: 10.48550/arXiv.1810.02912
[34] JIANG J C, DUN C, HUANG T J, et al. Graph convolutional reinforcement learning [EB/OL]. (2018-10-22) [2023-06-10]. https://arxiv. org/abs/1810.09202
[35] RYU H, SHIN H, PARK J. Multi-agent actor-critic with hierarchical graph attention network [C]//The AAAI Conference on Artificial Intelligence. AAAI, 2020: 7236 – 7243. DOI: 10.1609/aaai.v34i05.6214
[36] YOON D, HONG S, LEE B J, et al. Winning the L2RPN challenge: power grid management via semi-markov afterstate actor-critic [EB/OL]. (2021-01-13) [2023-06-10]. https://openreview.net/pdf?id=LmUJqB1Cz8
[37] HOSSAIN R R, HUANG Q H, HUANG R K. Graph convolutional network-based topology embedded deep reinforcement learning for voltage stability control [J]. IEEE transactions on power systems, 2021, 36(5): 4848 – 4851. DOI: 10.1109/TPWRS.2021.3084469

Biographies

FENG Bingyi received his BE degree in computer science from Anhui University, China in 2021. He is working towards his MS degree at University of Science and Technology of China. His research interest focuses on deep reinforcement learning, multi-agent reinforcement learning, and machine learning systems.

FENG Mingxiao received his BE degree in computer science from University of Science and Technology of China in 2017. Now he is working towards his PhD degree with the School of Information Science and Technology, University of Science and Technology of China. His research interests mainly include deep reinforcement learning, multi-agent reinforcement learning, and large language model.

WANG Minrui received his BE degree in computer science from Anhui University, China in 2020, and his MS degree from the University of Science and Technology of China, in 2023. His research interests mainly include deep reinforcement learning, multi-agent reinforcement learning, and machine learning for recommendation systems.

ZHOU Wengang received his BE degree in electronic information engineering from Wuhan University, China in 2006, and PhD degree in electronic engineering and information science from the University of Science and Technology of China (USTC) in 2011. From September 2011 to September 2013, he worked as a postdoc researcher in Computer Science Department at the University of Texas, USA. He is currently a professor at the EEIS Department, USTC. His research interests include reinforcement learning, multimedia information retrieval, and computer vision. In those fields, he has published over 100 papers in IEEE/ACM Transactions and CCF Tier-A International Conferences.

LI Houqiang (lihq@ustc.edu.cn) received his BS, ME, and PhD degrees in electronic engineering from University of Science and Technology of China (USTC) in 1992, 1997, and 2000, respectively. He is a professor and the Vice Dean of the School of Information Science and Technology, USTC, and the Director of MOE-Microsoft Key Laboratory of Multimedia Computing and Communication. He is a fellow of IEEE. His research interests include deep learning, reinforcement learning, image/video coding, image/video analysis, and computer vision, etc. He has authored and co-authored over 300 papers in journals and conferences, and holds over 60 granted patents.

A Practical Reinforcement Learning Framework for Automatic Radar Detection

YU Junpeng[1], CHEN Yiyu[2]

(1. Nanjing Research Institute of Electronics Technology, Nanjing 210039, China;
2. Nanjing University, Nanjing 210023, China)

DOI: 10.12142/ZTECOM.202303004

https://kns.cnki.net/kcms/detail/34.1294.TN.20230802.1625.002.html, published online August 3, 2023

Manuscript received: 2023-06-16

Abstract: At present, the parameters of radar detection rely heavily on manual adjustment and empirical knowledge, resulting in low automation. Traditional manual adjustment methods cannot meet the requirements of modern radars for high efficiency, high precision, and high automation. Therefore, it is necessary to explore a new intelligent radar control learning framework and technology to improve the capability and automation of radar detection. Reinforcement learning is popular in decision task learning, but the shortage of samples in radar control tasks makes it difficult to meet the requirements of reinforcement learning. To address the above issues, we propose a practical radar operation reinforcement learning framework, and integrate offline reinforcement learning and meta-reinforcement learning methods to alleviate the sample requirements of reinforcement learning. Experimental results show that our method can automatically perform as humans in radar detection with real-world settings, thereby promoting the practical application of reinforcement learning in radar operation.

Keywords: meta-reinforcement learning; radar detection; reinforcement learning; offline reinforcement learning

Citation (Format 1): YU J P, CHEN Y Y. A practical reinforcement learning framework for automatic radar detection [J]. ZTE Communications, 2023, 21(3): 22 - 28. DOI: 10.12142/ZTECOM.202303004

Citation (Format 2): J. P. Yu and Y. Y. Chen, "A practical reinforcement learning framework for automatic radar detection," *ZTE Communications*, vol. 21, no. 3, pp. 22 - 28, Sept. 2023. doi: 10.12142/ZTECOM.202303004.

1 Introduction

The advent of modern radar systems has brought forth a demand for higher efficiency, precision, and automation[1]. However, the current radar detection parameters heavily depend on manual adjustment and empirical knowledge, which significantly hampers automation[2]. Traditional manual adjustment methods are increasingly inadequate to meet these growing demands. This inadequacy necessitates the exploration of a new intelligent radar control learning framework and technology that can enhance the capability and automation of radar detection.

One promising learning approach is reinforcement learning, which has gained popularity in decision-task learning. Reinforcement learning is a major paradigm within the machine learning field, distinct from perceptual learning typified by image processing. Perceptual learning primarily involves supervised learning, while reinforcement learning seeks to address sequential decision-making problems through rewards. The reinforcement learning algorithm, based on the Bellman equation, continually learns and improves through trial and error within an environment, thereby accumulating experience and developing superior strategies for given tasks[3]. In recent years, deep reinforcement learning (DRL), with its powerful feature representation and function-fitting capabilities, has shown remarkable proficiency in various areas such as gaming and robotics. Notable accomplishments include AlphaGo's consecutive victories over human world champions in Go[4], AlphaStar's top master rank in StarCraft II[5], Suphx's rise to the top ten sections of the professional Japanese Mahjong platform "Tianfeng" developed by Microsoft Research Asia[6], and the flexible and universal tokamak magnetic controller architecture developed by the DeepMind team for nuclear fusion projects[7]. Furthermore, deep reinforcement learning has been progressively implemented across various industries.

However, the application of reinforcement learning in radar control tasks is hindered by the shortage of samples. The effectiveness of deep reinforcement learning is currently heavily reliant on the availability of extensive learning data and substantial computing resources. For instance, the chess benchmark algorithm, MuZero, requires approximately 10^6 steps of data[8] to achieve initial results in training. This process takes roughly 11 days at a sampling rate of 60 steps per second. Furthermore, DeepMind utilized 384 tensor processing units (TPUs) running

This work is supported by Science and Technology Innovation 2030 New Generation Artificial Intelligence Major Project under Grant No. 2021ZD0113303, the National Natural Science Foundation of China under Grant Nos. 62192783 and 62276128, and in part by the Collaborative Innovation Center of Novel Software Technology and Industrialization.
The authors contribute equally to this paper.

in parallel over a span of about 44 days to complete the reinforcement learning training for AlphaStar, the StarCraft II algorithm[5]. The high training cost associated with deep reinforcement learning significantly restricts its range of applications.

This paper aims to address these challenges by proposing a practical radar operation reinforcement learning framework that integrates offline reinforcement learning and meta-reinforcement learning methods. The framework consists of the environment modeling of radar detection, the integrated learning structure and the learning objectives. Our experimental results of the MATLAB radar detection simulator indicate that the ability of our method in automatic radar detection has basically reached the level of humans, thus promoting the practical application of reinforcement learning in radar detection. This paper is structured as follows. First, in Section 2, we introduce the related works. Then, in Section 3, we demonstrate our reinforcement learning framework for automatic radar detection. In Section 4, the experimental settings and results are introduced. Finally, in Section 5, we draw a conclusion.

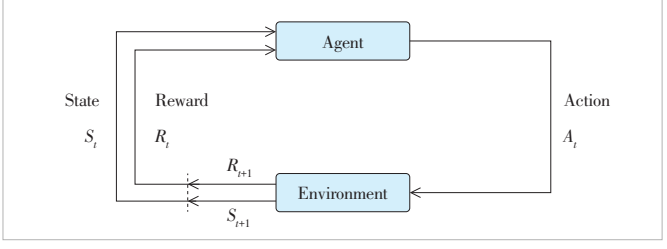
2 Related Works

In this section, we introduce the background and related works about our proposed framework, including reinforcement learning and its correlational research with radar control.

2.1 Reinforcement Learning

Reinforcement learning is one of the popular paradigms of machine learning. The framework of reinforcement learning is shown in Fig. 1, which mainly includes two parts: agent and environment. The operation of reinforcement learning is a process of continuous interaction between agents and the environment, where the environment provides agents with the current state and numerical rewards, while agents output actions to the environment according to existing information (usually the current state). The environment gives the state and rewards after the action is executed, and so forth until the environment terminates (done). In this process, agents often choose actions and learn strategies to maximize expected cumulative rewards.

The environment model of reinforcement learning is generally based on the Markov decision process (MDP). MDP is defined by a quaternion $\langle S, A, R, T \rangle$, where S is the set of environmental states, A is the set of optional actions, the state transition function $T: S \times A \times S \to [0, 1]$ gives the probability of transition from state s and action a to state s', and the reward function R: $S \times A \times S \to \mathbb{R}$ provides the reward value for each step.



▲Figure 1. Framework of reinforcement learning

The agent algorithm of reinforcement learning aims to learn a policy π, and the policy determines the execution of action a (deterministic policy) or the execution probability (non-deterministic policy) in each state s. The classical reinforcement learning algorithm considers that the MDP model of the environment is given in advance, and the optimization goal of the policy π is to maximize the expected cumulative discount reward. The parameters of the parameterization policy π_θ are θ, and the formula for calculating the optimal parameters θ^* is:

$$\theta^* = \arg\max_\theta \mathbb{E}_{\pi_\theta}\left[\sum_{t=0}^{T}\gamma^t r_t\right], \tag{1}$$

where T refers to the number of time steps that the environment runs, and the discount factor $\gamma \in [0,1]$ is used to balance long-term rewards and short-term rewards. γ significantly stabilizes the reinforcement learning algorithm in an environment with excessive T.

Reinforcement learning algorithms can be divided into two categories: value function-based and policy gradient-based. The reinforcement learning algorithm based on the value function makes decisions according to the state action value function $Q^\pi(s,a)$. In the DRL algorithms based on the value function, $Q^\pi(s,a)$ is constructed by a neural network, supplemented by some designs to enhance the stability of the algorithm[9]. Note that deep networks enable policies to adapt to tasks with a much wider range. This kind of algorithm performs better in the discrete action environment, but it is difficult to expand to the continuous action environment. Common algorithms include the deep Q-network (DQN) [9], dueling double deep Q-network (D3QN)[10], deep recurrent Q-network (DRQN)[11], etc. Reinforcement learning algorithms based on policy gradients directly calculate the policy function $\pi_\theta(a|s)$ modeling and optimization. Commonly used algorithms based on policy gradients are actor-critic architectures, which perform better in continuous action environments, including deep deterministic policy gradient (DDPG) [12], proximal policy optimization (PPO) [13], soft actor-critic (SAC)[14], twin-delayed deep deterministic policy gradient (TD3)[15], etc.

While reinforcement learning algorithms have demonstrated effective performance in simulated environments, two primary challenges exist in copying this performance to real-world scenarios: 1) The inconsistency between the simulator and the actual environment, which is often referred to as the Sim2Real gap, tends to result in catastrophic failure of deploying simulator-trained policies in the real world; 2) the high cost of real-world sampling and the complexity of the real-world tasks result in a significant difference between the collected data and the actual situation. Especially in intelligent radar detection tasks, due to the large scale of the actual environment, it is difficult to collect sufficient training data that are needed. The actual environment can change greatly at any time with factors such as weather, ter-

rain, and goals, which brings learning difficulties. Therefore, we introduce two research directions that help to solve this problem: offline reinforcement learning and meta-reinforcement learning.

2.1.1 Offline Reinforcement Learning

Offline reinforcement learning is a data-driven subset of the broader reinforcement learning field. Its primary objective is to optimize the same objective as reinforcement learning. However, in this context, intelligent agents cannot use behavioral strategies to interact with the environment or gather additional data. Instead, a learning algorithm provides a static transition dataset, denoted as $D = (s, a, r, a')$, which is used to learn the most effective strategies. This approach is more akin to the standard supervised learning problem, with D serving as the policy training set. Essentially, offline reinforcement learning requires the learning algorithm to fully understand the dynamic system that underlies the Markov decision process, using a fixed dataset to formulate a policy. When this policy is applied to interact with the Markov decision process, it aims to yield the maximum cumulative return.

Several existing model-free offline reinforcement learning methods regularize the learned policy to align closely with the behavior policy. This is achieved through techniques such as distributional matching[16], support matching[17], importance sampling[18 - 19], and learning the lower bounds of true Q-values[20]. On the other hand, model-based algorithms learn policies by leveraging a dynamic model derived from the offline dataset. Ref. [21] directly restricts the learned policy to the behavior policy, similar to model-free algorithms. To penalize the policy for visiting states where the learned model may be incorrect, MOPO[22] and MoREL[23] adjust the learned dynamics, which ensures that the value estimates are conservative when the model uncertainty exceeds a certain threshold. To eliminate the need for uncertainty quantification, COMBO[24] combines model-based policy optimization[25] and conservative policy evaluation[20]. In this paper, we employ a distributional matching method, specifically the straightforward and effective behavior cloning (BC) method, as it simplifies the learning process of meta-reinforcement learning methods.

2.1.2 Meta-Reinforcement Learning

Meta-reinforcement learning methods learn meta-policy on multiple meta-training tasks, aiming to quickly adapt to previously unseen meta-testing tasks, and thus improving the effectiveness and generalizability of reinforcement learning methods. The process of meta-reinforcement learning mirrors that of meta-learning, which consists of two stages: the meta-training stage and the meta-testing stage. During the meta-training stage, the algorithm learns from the meta-training task and prepares the model for the next stage. In the meta-testing phase, the trained model is adaptively applied to the meta-testing task to achieve testing results. Each task corresponds to a reinforcement learning environment model, typically an MDP. The meta-training task is presented in the form of task distribution $p(T)$. At the beginning of meta-training, a certain number of meta training tasks $\{T_{\text{train}}\}$ are sampled from the task distribution $p(T)$, that is, $T_{\text{Train}} \sim p(T)$. The set of meta-training tasks may be fixed by one sampling, or may be generated repeatedly by samplings in multiple rounds of meta-training.

Existing works in this field can be broadly categorized into three types: the model-agnostic-meta-learning-based (MAML-based), recurrent-based, and context-based. Some research focuses on improving and extending the meta-learning framework MAML[26]. For instance, FINN et al. proposed a simplified algorithm FO-MAML that only uses first-order derivatives in their MAML work[26]; NICHOL et al. proposed a more versatile first-order derivative algorithm Reptile[27]; The ES-MAML algorithm proposed by SONG et al. uses an evolutionary algorithm instead of derivation in outer optimization[28]; ANTONIO et al. conducted extensive experiments and concluded on the training problem of MAML[29].

Some other research reduces the uncertainty of inferring the state from observation by memorizing the history of tasks, thus improving the performance of strategies on unknown tasks. For example, the RL^2 algorithm builds a policy model based on the recurrent neural network with memory and trains between multiple tasks[30]; MISHRA et al. combined time series convolution and soft attention mechanisms to form a new depth architecture[31]; PARISOTTO uses the transformer model as a cross episodic memory module[32].

Recent popular research extracts the task context to guide policy across various tasks. SÆMUNDSSON et al. used the Gaussian process and variational inference to model the hidden variables of tasks, combined with the model-based reinforcement learning algorithm to achieve a fast meta-training algorithm ML-GP[33]; ZINTGRAF et al.[34] and LAN et al.[35] combined the MAML algorithm with a task context encoder to improve performance; HUMPLIK et al. utilized long short-term memory (LSTM) to construct a task feature inference module and implemented algorithms similar to PEARL[36]; FAKOOR et al. used gated recurrent units as the history encoder to train their reinforcement learning algorithm meta-Q-learning (MQL) based on the multi-task objective[37]. The PD-VF algorithm proposed by RAILEANU et al. used the prediction environment cumulative reward to supervise the training task hidden variable module[38]; ZINTGRAF et al. used a variational autoencoder to train the task feature inference module and proposed the VariBAD algorithm[39]. Some studies improve the generalization ability of context-based methods through comparative learning. FU et al. constructed the algorithm named contrastive learning augmented context-based meta-RL (CCM) based on MoCo[40] and CURL[41]. WANG et al. proposed a method similar to CCM, TCL, where positive and negative samples are divided according to sampling trajectories rather than task types[42].

In this paper, we utilize the context-based VariBAD algorithm[39] to consider radar detection task characteristics and requirements.

2.2 Radar Control with Reinforcement Learning

Reinforcement learning methods enable automatic learning of complex behaviors, and several studies have focused on introducing deep reinforcement learning into radar control. AZIZ et al. provided a survey of literature proposing the application of reinforcement learning to radar to overcome jamming[2]. WANG et al. suggested a cognitive frequency design method for a compressed-sensing-based frequency agile radar using reinforcement learning[43]. PATTANAYAK et al. introduced an inverse reinforcement learning approach to meta-cognitive radars in an adversarial setting[44]. ZHAI et al. proposed a reinforcement learning-based approach for multi-input multi-output (MIMO) cognitive radar[45]. OTT et al. proposed an uncertainty-based meta-reinforcement learning approach with out-of-distribution environment detection[46]. In the context of multi-agent systems, SNOW et al. proposed a multi-objective inverse reinforcement learning approach for tracking targets with a cognitive radar network[47]. MENG et al. examined the issue of target assignment when a phased-array radar network detects hypersonic-glide vehicles in near space and proposed a method for target assignment based on deep reinforcement learning[48].

The aforementioned studies illustrate that deep reinforcement learning has extensive potential applications in various aspects of radar systems. However, these related works are conducted in simple simulated scenarios, and thus it remains challenging to implement reinforcement learning methods in real-world situations. In this paper, we concentrate on the framework for extensive single radar parameter control, and we introduce realistic sample-limited settings and corresponding reinforcement learning methods to tackle this problem.

3 Reinforcement Learning Framework for Automatic Radar Detection

3.1 Environment Modeling

Environment modeling is the foundation of reinforcement learning. Existing modules of traditional radar control are: a) analog signal → plot processing; b) plot → track processing; c) track → radar parameter control module. The intelligent radar control system mainly requires intelligent automatic control of the radar while observing the processed radar data (e.g., plots, tracks, etc.), and its framework is shown in Fig. 2. In order to enhance the universality and generalization performance of our reinforcement learning algorithm, our agent focuses on processing the input data composed of original analog signals, processed plots, and mixed tracks as states, and outputs controllable radar parameters.

The radar point and track processing algorithms typically operate in cycles. After each radar scan is completed and before the next one begins, our agent makes its decisions. In this context, the input state $s = \left(s^1, s^2, s^3\right)$ includes:

a) A 3-dimensional raw echo analog signal, denoted as $s^1 \in [H, W, V]$. Here, H, W, and V represent the distance, deviation angle, and amplitude of the signal, respectively. This analog signal is the radar's echo signal in each direction. The data for each cycle is a position peak matrix.

b) Dots denoted as $s^2 = \left\{\left(x_1, y_1, v_1\right), \left(x_2, y_2, v_2\right), \cdots, \left(x_n, y_n, v_n\right)\right\}$. These are a series of points identified as target points in the analog signal. Each point has features, such as position and signal-to-noise ratio, extracted by algorithms. The number of points in the plot data for each cycle is uncertain. Each point has one row of features. Although there are much more clutter points in dots compared with tracks, it may cover more potential targets.

c) Tracks denoted as $s^3 = \left\{\left(x'_1, y'_1, v'_1, l'_1\right), \left(x'_2, y'_2, v'_2, l'_2\right), \cdots, \left(x'_{n'}, y'_{n'}, v'_{n'}, l'_{n'}\right)\right\}$. A track is a series of points in a historical track where the target point is recognized as a real target. Each point has features, such as position and velocity, extracted by algorithms. Similar to the format of dots, there are multiple dots in each cycle of dot data, each with a single line of features. However, each target is additionally marked with a unique batch number. The trajectory is the main basis for decision-making, but it often lacks some difficult-to-detect target information and performs with a certain lag.

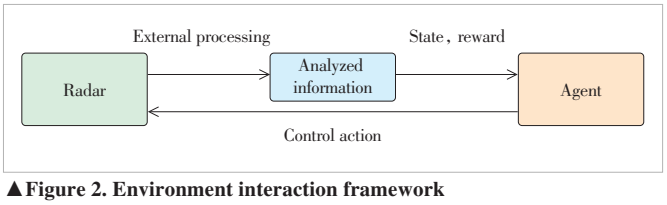
The system's output is the radar's parameter control information, which includes frequency point, speed, pitch angle, and timing transmission. Note that these are generally discrete variables.

3.2 Learning Framework

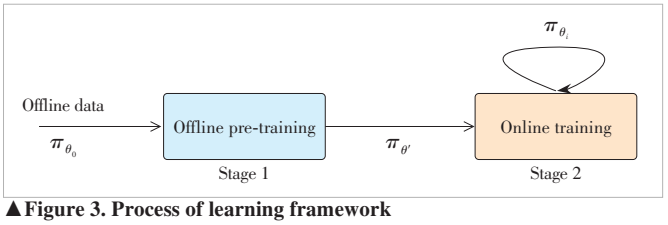
As Fig. 3 illustrates, the framework's primary process is divided into two stages:

The first stage involves data collection and offline pre-training. Although offline reinforcement learning algorithms do not impose strict requirements on offline training data, existing research indicates that the diversity of offline training data significantly impacts the learning outcome[49]. Hence, we aim to conduct offline reinforcement learning pre-training for the decision model, denoted as π_θ (where θ represents model parameters), based on as many diverse and abundant offline training data as possible. This stage initiates with model parameters θ_0 and results in pre-training parameters θ'.

The second stage involves running the pre-trained decision



▲Figure 2. Environment interaction framework



▲Figure 3. Process of learning framework

model $\pi_{\theta'}$ in an actual radar scenario and performing reinforcement learning iterations online. To facilitate a smooth transition from offline data to online scenarios for the decision model, we incorporate a meta-reinforcement learning algorithm to enhance the model's generalization. Given that radar detection requires precise environmental cognition, the introduced inference-based meta-reinforcement learning algorithm includes a task feature inference module and an auxiliary training target. This new meta-reinforcement learning network structure is also utilized in the offline pre-training phase, which indicates the combination of offline reinforcement learning and meta-reinforcement learning.

3.3 Reinforcement Learning Method

Network design: Our design integrates a fundamental RL algorithm and a variational autoencoder (VAE)[50] to encode different task scenarios automatically. These encoded features are subsequently inputted into the intelligent agent[39]. The VAE model consists of an encoder $q_\phi(s_t,a_{t-1},r_t,h_{t-1}) \rightarrow z_t$ and two decoders $p_\psi^R(z_t,s_i,a_i,s_{i+1}) \rightarrow r_{i+1}, p_\psi^T(z_t,s_i,a_i) \rightarrow s_{i+1}$. This model leverages the reconstruction constraints of rewards and states, along with constrained dimension to compress the original inputs $o_t = \{s_t,a_{t-1},r_t\}$ into low-dimensional representation z_t efficiently. The posterior distribution of a task can be interpreted as a representation of specific task characteristics, such as meteorological features, ship-type tendencies and radar models. The decoder takes a posterior distribution of tasks and some prior knowledge as input to predict the subsequent state. The context encoder q_ϕ is required to encode an indefinite length historical sequence. Therefore, a recurrent neural network (RNN) model is employed for approximation, and other models can be approximated using a multi-layer perceptron (MLP). The VAE model outputs a low-dimensional representation of tasks z_t to policy model $\pi_\theta(s_t,z_t) \rightarrow a_t$. The complete network structure is shown in Fig. 4.

Offline reinforcement learning: Although a variety of offline reinforcement learning algorithms are available, we have opted for the behavioral cloning (BC) objective due to its ease of use and scalability. In practice, our policy model takes all historical trajectories as input. Let's denote the previously collected dataset as $D = \{(o_i,a_i)\}$, where the equivalent new state is the historical trajectory $o_i = \{s_i,a_{i-1},r_i\}$, $s_i = (s_i^1,s_i^2,s_i^3)$. Thus, the offline training objective is:

$$J_1 = -E_{(s,a)}\mathrm{Dist}\left(\pi_\theta(o_i),a_i\right), \tag{2}$$

where $E_{(s,a)}$ is the expectation, $\mathrm{Dist}(\cdot)$ is the distance calculation function, which can be set as a 0 - 1 function for discrete variables. Behavioral cloning targets enable the policy model's decision actions, $\pi_\theta(o_i)$, to be closer to expert actions, thereby allowing the parameters θ to learn a certain level of strategic knowledge.

Meta-reinforcement learning: We employ reconstruction and the information bottleneck objective to constrain feature information. The forms of reward and state reconstruction objective functions are:

$$J_2 = -E_{r_i}\mathrm{Dist}\left(p_\psi^R\left(q_\phi(s_i,a_{i-1},r_i,h_{i-1}),s_i,a_i,s_{i+1}\right),r_i\right),$$

$$J_3 = -E_{s_i}\mathrm{Dist}\left(p_\psi^T\left(q_\phi(s_i,a_{i-1},r_i,h_{i-1}),s_i,a_i\right),s_i\right). \tag{3}$$

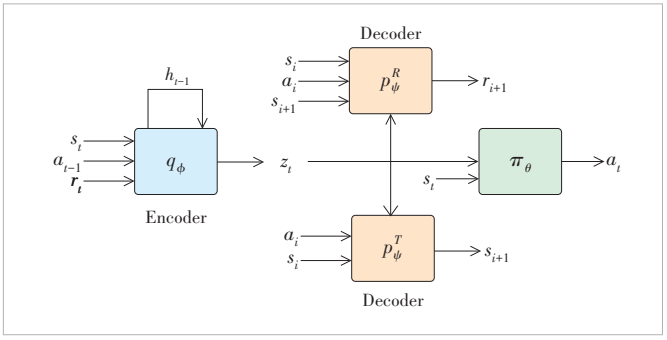
$\mathrm{Dist}(\cdot)$ can be set as the L2 distance function for continuous variables such as the state and reward. The desired feature of the target, $z_i = q_\phi(s_i,a_{i-1},r_i,h_{i-1})$, retains the original input information.

Moreover, the reconstruction objective involves the simultaneous optimization of two models, which may result in gradient descent optimization not achieving the expected results. To expedite training, we additionally introduce the information bottleneck method optimization objective[36], which is in the form of:

$$J_4 = -E_i D_{\mathrm{KL}}\left(q_\phi(s_i,a_{i-1},r_i,h_{i-1}),r(z)\right), \tag{4}$$

where $r(z)$ is the normal distribution.

During the offline pre-training stage, the overall optimization objective is $J_1 + J_2 + J_3 + J_4$ as the VAE is also trained. In the online training phase, the offline reinforcement learning objective is replaced by the traditional reinforcement learning objective J_{RL}, and the overall optimization objective is $J_{\mathrm{RL}} + J_2 + J_3 + J_4$. Note that due to the existing initialization parameters, we need to reduce the learning rate by an order of magnitude during online tuning.



▲Figure 4. Network structure of the agent

4 Experiment

Experimental tasks: Despite the universality of our constructed method, we require explicit task scenarios and objectives for the experiment. Our primary experimental scenario involves enhancing the average signal-to-noise ratio (SNR) of radar detection targets by manipulating the radar's frequency points. Given the radar's relatively short rotation time (either

1 s or 10 s), we assume that the environment will remain largely unchanged even if the radar scans every alternate turn. As for the reward setting issue, we alternate between a circle set as a frequency point generated by the algorithm and the next circle as a fixed frequency point. This approach provides a relatively standardized reward for the reinforcement learning algorithm, $r_t(a) = \mathrm{SNR}(s_t, a) - \mathrm{SNR}(s_t, a_{\mathrm{fix}})$.

Evaluation setting: For offline data, actions are expert strategies, and we compare the overlap between algorithm output actions and offline data actions. For online learning, considering the practical application, we involve relevant radar operation experts to compare the algorithm's parameter control rewards with the expert's parameter control rewards. We ask both the algorithm and experts to test each other on the same task, compare the cumulative rewards of the algorithm with the cumulative rewards of the experts, and take the average of three experiments. To benchmark real-world sample-limited applications, in the online training stage, the sample number is limited to 10 rounds, i.e. 10 000 steps.

Implementation: We employ proximal policy optimization (PPO) as the reinforcement learning algorithm during the online training phase. We use MATLAB as it supports the Radar Toolbox to construct a simulation environment. We achieve code communication between Python and MATLAB via the user datagram protocol (UDP). The environment and algorithms ran on a 2.5 GHz CPU and a single NVIDIA GeForce RTX 3080 graphics card. For offline training data, we have experts control the selection of radar parameters in the task, but we also aim to cover as many action intervals as possible, thereby obtaining data with a total of 100 000 steps with a wide distribution of action.

Experimental results: After achieving convergence in the offline training phase, the action similarity between the decision model and offline data is 99%. In online tests, the average cumulative reward of the proposed method reaches 91% of the experts' method, and the performance of a random policy is unstable and obviously weaker. Detailed online testing results with average cumulative reward are shown in Table 1. Additionally, we observe that the decision model can control different parameters for different targets, and it tends to favor some commonly used radar parameters. The experimental results indicate that the decision model has learned the preliminary radar control policy, but the potential of deep learning may not be fully exploited due to the limitation of the training sample size. According to our experience and expert judgment, our method can act as humans in basic radar automatic detection, and therefore has the potential to be applied in practical radar operation tasks. In future research, we will focus on further enhancing the effectiveness of reinforcement learning and aim to apply it to actual radar.

▼Table 1. Online testing results

Test	Random Policy	Proposed	Experts
Trial 1	−9.24	24.32	26.14
Trial 2	12.78	28.77	29.33
Trial 3	6.34	25.38	30.85
Average	3.29	26.16	28.77

5 Conclusions

In this paper, we have presented a novel practical approach to radar operation that leverages the power of reinforcement learning. By integrating offline reinforcement learning and meta-reinforcement learning methods, we have developed a practical radar operation reinforcement learning framework that can quickly adapt to unseen real-world tasks. Our experimental results have demonstrated the ability to act as humans in basic radar automatic detection with real-world settings, thereby validating our approach. Our work not only addresses the current challenges in radar operation but also paves the way for the practical application of reinforcement learning in radar operation. The proposed method has the potential to revolutionize radar detection by enhancing its efficiency, precision, and automation. Future work will focus on further refining our framework and exploring its application in real-world radar systems.

References

[1] GENG Z, YAN H, ZHANG J, et al. Deep-learning for radar: a survey [J]. IEEE access, 2021, 9: 141800-141818. DOI:10.1109/ACCESS.2021.3119561
[2] AZIZ M M, MAUD A R M, HABIB A. Reinforcement learning based techniques for radar anti-jamming [C]//International Bhurban Conference on Applied Sciences and Technologies (IBCAST). IEEE, 2021: 1021 – 1025. DOI: 10.1109/IBCAST51254.2021.9393209
[3] SUTTON R S, BARTO A G. Reinforcement learning: an introduction (2nd ed) [M]. Cambridge, USA: MIT press, 2018
[4] SILVER D, HUANG A, MADDISON C J, et al. Mastering the game of Go with deep neural networks and tree search [J]. Nature, 2016, 529(7587): 484 – 489. DOI: 10.1038/nature16961
[5] VINYALS O, BABUSCHKIN I, CZARNECKI W M, et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning [J]. Nature, 2019, 575(7782): 350 – 354. DOI: 10.1038/s41586-019-1724-z
[6] LI J J, KOYAMADA S, YE Q W, et al. Suphx: mastering mahjong with deep reinforcement learning [EB/OL]. [2023-04-16]. https://arxiv.org/abs/2003.13590
[7] DEGRAVE J, FELICI F, BUCHLI J, et al. Magnetic control of tokamak plasmas through deep reinforcement learning [J]. Nature, 2022, 602(7897): 414 – 419. DOI: 10.1038/s41586-021-04301-9
[8] SCHRITTWIESER J, ANTONOGLOU I, HUBERT T, et al. Mastering Atari, Go, chess and shogi by planning with a learned model [J]. Nature, 2020, 588 (7839): 604 – 609. DOI: 10.1038/s41586-020-03051-4
[9] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning [J]. Nature, 2015, 518(7540): 529 – 533. DOI: 10.1038/nature14236
[10] WANG Z Y, SCHAUL T, HESSEL M, et al. Dueling network architectures for deep reinforcement learning [C]//The 33rd International Conference on International Conference on Machine Learning. ACM, 2016: 1995 – 2003. DOI: 10.5555/3045390.3045601
[11] HAUSKNECHT M, STONE P. Deep recurrent Q-learning for partially observable MDPs [J]. AAAI fall symposium, 2015: 29 – 37
[12] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning [EB/OL]. [2023-04-16]. https://arxiv. org/pdf/1509.02971. pdf. DOI: 10.1016/S1098-3015(10)67722-4
[13] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms [EB/OL]. (2017-08-28) [2023-04-16]. https://arxiv.org/abs/1707.06347

[14] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor [EB/OL]. (2018-08-08) [2023-04-16]. https://arxiv.org/abs/1801.01290
[15] FUJIMOTO S, VAN HOOF H, MEGER D. Addressing function approximation error in actor-critic methods [C]//International Conference on Machine Learning. PMLR, 2018: 1587 – 1596
[16] FUJIMOTO S, MEGER D, PRECUP D. Off-policy deep reinforcement learning without exploration [EB/OL]. (2019-08-10) [2023-04-16]. https://arxiv. org/abs/1812.02900
[17] KUMAR A, FU J, TUCKER G, et al. Stabilizing off-policy Q-learning via bootstrapping error reduction [EB/OL]. (2019-11-25) [2023-04-16]. https://arxiv. org/abs/1906.00949
[18] NACHUM O, DAI B, KOSTRIKOV I, et al. AlgaeDICE: policy gradient from arbitrary experience [EB/OL]. (2019-12-04) [2023-04-16]. https://arxiv. org/abs/1912.02074
[19] LIU Y, SWAMINATHAN A, AGARWAL A, et al. Off-policy policy gradient with state distribution correction [EB/OL]. (2019-07-16) [2023-04-16]. https://arxiv.org/abs/1904.08473
[20] KUMAR A, ZHOU A, TUCKER G, et al. Conservative Q-learning for offline reinforcement learning [EB/OL]. (2020-08-19) [2023-04-16]. https://arxiv. org/abs/2006.04779
[21] MATSUSHIMA T, FURUTA H, MATSUO Y, et al. Deployment-efficient reinforcement learning via model-based offline optimization [EB/OL]. (2020-06-05) [2023-04-16]. https://arxiv.org/abs/2006.03647
[22] YU T H, THOMAS G, YU L, et al. MOPO: model-based offline policy optimization [J]. Advances in neural information processing systems, 2020, 33: 14129-14142.
[23] KIDAMBI R, RAJESWARAN A, NETRAPALLI P, et al. MOReL: Model-based offline reinforcement learning [J]. Advances in neural information processing systems, 2020, 33: 21810 – 21823
[24] YU T H, KUMAR A, RAFAILOV R, et al. COMBO: conservative offline model-based policy optimization [EB/OL]. (2022-01-27) [2023-04-16]. https://arxiv. org/abs/2102.08363
[25] JANNER M, FU J, ZHANG M, et al. When to trust your model: Model-based policy optimization [C]//The 33rd International Conference on Neural Information Processing Systems. ACM, 2019: 12519 – 12530
[26] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks [C]//The 34th International Conference on Machine Learning. ACM, 2017: 1126 – 1135. DOI: 10.5555/3305381.3305498
[27] NICHOL A, ACHIAM J, SCHULMAN J. On first-order meta-learning algorithms [EB/OL]. (2018-10-22) [2023-04-16]. https://arxiv.org/abs/1803.02999
[28] SONG X Y, GAO W B, YANG Y X, et al. ES-MAML: simple hessian-free meta learning [EB/OL]. (2020-07-07) [2023-04-16]. https://arxiv.org/abs/1910.01215
[29] ANTONIO A, STORKEY A, EDWARDS H. How to train your MAML [EB/OL]. (2019-03-15) [2023-04-16]. https://arxiv.org/abs/1810.09502
[30] DUAN Y, SCHULMAN J, CHEN X, et al. RL^2: fast reinforcement learning via slow reinforcement learning [EB/OL]. (2016-11-10) [2023-04-16]. https://arxiv.org/abs/1611.02779
[31] MISHRA N, ROHANINEJAD M, CHEN X, et al. A simple neural attentive meta-learner [EB/OL]. (2018-02-15) [2023-04-16]. http://arxiv.org/abs/1707.03141.
[32] PARISOTTO E. Meta Reinforcement Learning through Memory [D]//Pittsburgh: Carnegie Mellon University, 2021
[33] SÆMUNDSSON S, HOFMANN K, DEISENROTH M P. Meta reinforcement learning with latent variable Gaussian processes [EB/OL]. (2018-05-20) [2023-06-16]. https://arxiv.org/abs/1803.07551
[34] ZINTGRAF L, SHIARLIS K, KURIN V, et al. Fast context adaptation via meta-learning [C]//The 36th International Conference on Machine Learning. ICML, 2019: 13262 – 13276
[35] LAN L, LI Z, GUAN X, et al. Meta reinforcement learning with task embedding and shared policy [C]//The 28th International Joint Conference on Artificial Intelligence. ACM, 2019: 2794 – 2800. DOI:10.24963/ijcai.2019/387
[36] HUMPLIK J, GALASHOV A, HASENCLEVER L, et al. Meta reinforcement learning as task inference [EB/OL]. (2019-05-15) [2023-06-16]. https://arxiv.org/abs/1905.06424
[37] FAKOOR R, CHAUDHARI P, SOATTO S, et al. Meta-Q-Learning [EB/OL]. (2019-09-30) [2023-06-16]. http://arxiv.org/abs/1910.00125
[38] RAILEANU R, GOLDSTEIN M, SZLAM A D, et al. Fast adaptation to new environments via policy-dynamics value functions [C]//International Conference on Machine Learning. ICML, 2020: 7920 – 7931
[39] ZINTGRAF L, SCHULZE S, LGL M, et al. VariBAD: variational Bayes-adaptive deep RL via meta-learning [J]. The journal of machine learning research, 2021, 22 (1): 13198 – 13236
[40] HE K M, FAN H Q, WU Y X, et al. Momentum contrast for unsupervised visual representation learning [C]//Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 9726 – 9735. DOI: 10.1109/CVPR42600.2020.00975
[41] LASKIN M, SRINIVAS A, ABBEEL P C. Contrastive unsupervised representations for reinforcement learning [C]//The 37th International Conference on Machine Learning. ICML, 2020: 5595 – 5606
[42] WANG B, XU S, KEUTZER K, et al. Improving context-based meta-reinforcement learning with self-supervised trajectory contrastive learning [EB/OL]. (2021-05-10) [2023-04-16]. https://arxiv.org/abs/2103.06386
[43] WANG S S, LIU Z, XIE R, et al. Reinforcement learning for compressed-sensing based frequency agile radar in the presence of active interference [J]. Remote sensing, 2022, 14(4): 968. DOI: 10.3390/rs14040968
[44] PATTANAYAK K, KRISHNAMURTHY V, BERRY C. Meta-cognition. an inverse-inverse reinforcement learning approach for cognitive radars [C]//The 25th International Conference on Information Fusion (FUSION). IEEE, 2022: 1 – 8
[45] ZHAI W T, WANG X R, GRECO M S, et al. Weak target detection in massive MIMO radar via an improved reinforcement learning approach [C]//IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022: 4993 – 4997. DOI: 10.1109/ICASSP43922.2022.9746472
[46] OTT J, SERVADEI L, MAURO G, et al. Uncertainty-based meta-reinforcement learning for robust radar tracking [C]//The 21st IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2023: 1476 – 1483. DOI: 10.1109/ICMLA55696.2022.00232
[47] SNOW L, KRISHNAMURTHY V, SADLER B M. Identifying coordination in a cognitive radar network-a multi-objective inverse reinforcement learning approach [C]//IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1 – 5. DOI: 10.1109/ICASSP49357.2023.10096376
[48] MENG F Q, TIAN K S, WU C F. Deep reinforcement learning-based radar network target assignment [J]. IEEE sensors journal, 2021, 21(14): 16315 – 16327. DOI: 10.1109/JSEN.2021.3074826
[49] FU J, KUMAR A, NACHUM O, et al. D4RL: Datasets for deep data-driven reinforcement learning [EB/OL]. (2020-04-15) [2023-04-16]. https://arxiv. org/abs/2004.07219
[50] KINGMA D P, WELLING M. Auto-encoding variational Bayes [EB/OL]. (2013-12-10) [2023-04-16]. https://arxiv.org/abs/1312.6114

Biographies

YU Junpeng received his master's degree in communication and information systems. He is a senior engineer with the Nanjing Research Institute of Electronics Technology, the deputy secretary-general of Intelligent Perception Special Committee of Jiangsu Association of Artificial Intelligence. His research interests include radar systems and intelligent processing technologies based on artificial intelligence. He has participated in many key artificial intelligence projects sponsored by the Ministry of Science and Technology of the People's Republic of China.

CHEN Yiyu (yiyuiii@foxmail.com) is currently a PhD student in the Department of Computer Science and Technology, Nanjing University, China. His research interest includes meta-reinforcement learning and robot control.

Boundary Data Augmentation for Offline Reinforcement Learning

SHEN Jiahao[1,2], JIANG Ke[1,2], TAN Xiaoyang[1,2]
(1. College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;
2. MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, Nanjing 211106, China)

DOI: 10.12142/ZTECOM.202303005
https://link.cnki.net/urlid/34.1294.TN.20230814.1519.002.html, published online August 14, 2023
Manuscript received: 2023-07-05

Abstract: Offline reinforcement learning (ORL) aims to learn a rational agent purely from behavior data without any online interaction. One of the major challenges encountered in ORL is the problem of distribution shift, i.e., the mismatch between the knowledge of the learned policy and the reality of the underlying environment. Recent works usually handle this in a too pessimistic manner to avoid out-of-distribution (OOD) queries as much as possible, but this can influence the robustness of the agents at unseen states. In this paper, we propose a simple but effective method to address this issue. The key idea of our method is to enhance the robustness of the new policy learned offline by weakening its confidence in highly uncertain regions, and we propose to find those regions by simulating them with modified Generative Adversarial Nets (GAN) such that the generated data not only follow the same distribution with the old experience but are very difficult to deal with by themselves, with regard to the behavior policy or some other reference policy. We then use this information to regularize the ORL algorithm to penalize the overconfidence behavior in these regions. Extensive experiments on several publicly available offline RL benchmarks demonstrate the feasibility and effectiveness of the proposed method.

Keywords: offline reinforcement learning; out-of-distribution state; robustness; uncertainty

Citation (Format 1): SHEN J H, JIANG K, TAN X Y. Boundary data augmentation for offline reinforcement learning [J]. ZTE Communications, 2023, 21(3): 29 - 36. DOI: 10.12142/ZTECOM.202303005
Citation (Format 2): J. H. Shen, K. Jiang, and X. Y. Tan, "Boundary data augmentation for offline reinforcement learning," *ZTE Communications*, vol. 21, no. 3, pp. 29 - 36, Sept. 2023. doi: 10.12142/ZTECOM.202303005.

1 Introduction

Reinforcement learning (RL) is one of the major branches of machine learning that has been successfully applied in various fields in recent years, such as power grid control[1], recommendation systems[2], and robotics[3]. RL training usually involves a large number of try-and-error interactions with underlying systems. However, such "interaction hungry" behavior could have a serious negative impact on many real-world applications, especially when the online data are either costly or dangerous to collect, e.g., in healthcare[4], autonomous driving[5], and so on. To address this problem, the idea of offline RL (ORL) is to learn a new policy only from data based on offline dataset without online interaction. Unfortunately, the direct employment of the common off-policy strategy often fails to achieve the same level of performance as in the online setting[6 - 7].

The extrapolation error from out-of-distribution (OOD) actions is generally thought of as the main reason responsible for the aforementioned performance degradation[6]. The OOD actions here mean the actions taken by a model or system that is outside the range of the examples it was trained on. Since the dataset used for ORL training is generated by a behavior policy different from the new policy, possibly working in a different environment as well, it is not always possible for the agent to generalize the knowledge learned from the data to unseen real online situations. This is not uncommon in practice and usually manifests as the overestimated Q-value of OOD actions and such error compounds, leading to potentially dangerous consequences. The problem is usually referred to as distribution shift. To address this, many recent works, such as Conservative Q-Learning (CQL) [8] and Implicit Q-Learning (IQL)[9], take a conservative strategy by trying to prevent the agent from taking overestimated OOD actions, with the idea that if most of the states visited are familiar to us, the chance of error and the subsequent error compounding phenomenon could be greatly reduced. However, taking OOD actions online is almost inevitable, and avoiding the queries on them may significantly reduce the robustness of the agent. Actually, not all OOD data is non-generalizable[10].

This work is partially supported by the National Key R&D program of China under Grant No. 2021ZD0113203 and National Science Foundation of China under Grant No. 61976115.

This observation motivates some works[11-12] to formulate the ORL problem as an uncertainty-penalizing policy optimization problem, where some self-supervised methods are usually adopted. These methods utilize the model prediction to provide the supervised signal for OOD data. However, such a prediction could be unreliable, highlighting the necessity of carefully balancing the tradeoff between the so-called radicalism and conservatism when dealing with OOD data.

In this paper, we propose a novel method, named Boundary Conservative Q Learning (Boundary-CQL; BCQL), to improve the robustness of the learned policy while keeping conservative when dealing with the OOD data. Our key idea is to weaken the agent's confidence in highly uncertain regions while doing offline learning. To find those regions, we propose simulating them with a modified Generative Adversarial Net (GAN) such that the generated data follow the same distribution as the old experience but are very difficult to be dealt with by themselves, with regard to some reference policies. In practice, the reference policy could be an empirical behavior policy or a pre-trained high-capacity policy, such as a CQL policy. As the found uncertain region generated data is visually located at the boundary of the distribution of the original data, we call them boundary OOD data. Finally, we learn the new policy via the Bellman operator while simultaneously maximizing its entropy at the generated boundary OOD data, hence decreasing the confidence in the estimation of the optimal actions. In this way, we effectively maintain a balance between the minimization of Bellman error and policy conservatism. Extensive experiments on several publicly available offline RL benchmarks demonstrate the feasibility and effectiveness of the proposed method.

It is worth noting that our method can also be thought of as a self-supervised offline RL method but is unsupervised in nature, as we only use the reference policy to identify the most difficult regions within the distribution defined by the training experience, without showing the algorithm the exact actions to be taken there. This is similar to those methods in machine learning that improve their generalization capability by negative sample mining[13-15] but has never been used in the case of offline RL, to the best of our knowledge.

In what follows, a brief review of related work is given in Section 2, and a concise introduction to the background of offline RL is provided in Section 3. The description of the boundary OOD data and BCQL method is presented in detail in Section 4. Experimental results are presented in Section 5 to evaluate the effectiveness and properties of the proposed method from multiple aspects. Finally, the paper concludes with a summary of the findings and contributions.

2 Related Work

In this section, we briefly review some works most relevant to our method in literature, which mainly involve offline RL methods and OOD simulation methods.

2.1 Offline Reinforcement Learning

Offline reinforcement learning is one of the hottest research directions in RL in recent years, where, as mentioned before, the major challenge is how to handle the distribution shift problem. This can be roughly divided into three categories: the policy constraint, uncertainty-based, and regularization of the value function. Regarding the policy constraint method, the BCQ algorithm[16] addresses the exploration error caused by the distribution shift through batch-constrained restriction. The bootstrapping error accumulation reduction (BEAR) algorithm[17] solves the problem of mismatch between the learning strategy, optimal strategy, and sampling strategy through a support set matching method and can achieve better results even when the sampling strategy is poor. The behavior regularized actor critic (BRAC) algorithm[18] tries to combine the advantages of both BCQ and BEAR algorithms for better learning efficiency. On the uncertainty-based approach, a typical representative method is the random ensemble mixture (REM)[19] method, which uses multiple parameterized Q functions to estimate Q values while enforcing Behrman consistency during learning.

CQL[8] is one of the most representative methods of the third category. It uses a regularization term to the traditional Q-value network so as to learn a relatively conservative Q function. However, since CQL sets a strict restriction on OOD action evaluation, it could be overly conservative. There are some works like IQL[9] and Mildly Conservative Q-learning (MCQ)[20] trying to fix this issue, where instead of directly learning actions out of data, known state action experience is used to learn how action values vary and future outcomes are averaged with random dynamics.

2.2 OOD Solution Based on Generative Model

OOD data in machine learning usually refer to the data that are significantly different from the training data on which a model is trained. These data are harmful to a machine learning model as they could mislead the model to make incorrect predictions. Hence, they have drawn the attention of many researchers in recent years[21]. Generative networks are useful tools for high-dimensional sampling and have been widely used to generate OOD data. The most popular generative models include adversarial generation networks[22] and the autoencoder/variational autoencoder[23].

The work based on auto-encoder/variational auto-encoder (PIDHORSKYI et al.)[24] used the variational auto-encoder to calculate reconstruction errors to identify OOD. First, the parametric manifold structure implied by the normal distribution is linearized to calculate OOD probability. Next, the method of probability decomposition is given, and then the local coordinates of the tangent space of the manifold are used to calculate the reconstruction error. Competitive Reconstruction Autoencoder (CoRA)[25] trained two autoencoders on the in-distribution and abnormal data, respectively, and used the

reconstruction error of the two autoencoders as the suspicious identification signal.

On the other hand, based on an adversarial generation network, Anomaly Detection with Generative Adversarial Networks (ADGAN)[26] generates OOD samples by checking whether the sampled data are satisfactory in a hidden space, while PNET[27] uses the generation network to generate and identify OOD samples based on their reconstruction errors.

2.3 Generative Model for Offline Reinforcement Learning

Generative models have been widely used in offline reinforcement learning for different usages. In BCQ[16] and BEAR[17] algorithms, conditional variational autoencoders generate data that satisfies the constraint. Action-conditioned Q-learning (AQL)[28] replaces the conditional variational autoencoder in BEAR with a residual generative model to improve fitting performance. MCQ[20] uses a generative model such as conditional GAN to estimate the sampling policy. Diffusion Q-learning[29] uses a diffusion model to constrain target policy and add a loss of maximum action value to the original diffusion loss. Selecting from Behavior Candidates (SfBC)[30] also uses a diffusion model combined with an in-sample planning technique to further avoid selecting out-of-sample actions and increase computational efficiency. Unlike prior work, we use a generative model to generate data not only following the same distribution as the experience but also in uncertain regions for agents.

3 Preliminaries

A Markov decision process (MDP) can be specified by a tuple $\mathcal{S}, \mathcal{A}, r, \mathcal{T}, \gamma$, where $\mathcal{S}$ regards the state space, $\mathcal{A}$ is the action space, $r: \mathcal{S} \times \mathcal{A} \to \mathcal{R}$ is the reward function, which is used to evaluate the action under state s, $\mathcal{T}: \mathcal{S} \times \mathcal{A} \to \mathcal{S}$ is the transition, and γ represents the discount factor. Reinforcement learning aims to find a policy to maximize the expected cumulative rewards. Q function $Q^{\pi}(s,a) = \mathbb{E}_{\pi}[\sum_{t=0}^{\infty} \gamma^t r_t | s, a]$ measures the discounted long-term reward given the state-action pair (s,a) and the policy π. Q-learning is a classic method that trains the Q-value function by minimizing the Bellman error over Q[31]. In the setting of continuous action space, Q-learning methods use an exact or approximate maximization scheme, such as the cross entropy method (CEM)[32], to recover the greedy policy as follows:

$$\begin{aligned} Q &\leftarrow \arg\min_{Q} \mathbb{E}[\hat{\mathcal{B}}^{\pi} Q(s,a) - Q(s,a)]^2, \\ \pi &\leftarrow \arg\max_{\pi} \mathbb{E}_s \mathbb{E}_{a \sim \pi(\cdot|s)} Q(s,a), \end{aligned} \tag{1}$$

where $\hat{\mathcal{B}}^{\pi} Q(s,a)$ represents the empirical Bellman target, defined as $\hat{\mathcal{B}}^{\pi} Q(s,a) = r(s,a) + \gamma \mathbb{E}_{a' \sim \pi(\cdot|s')} Q(s',a')$.

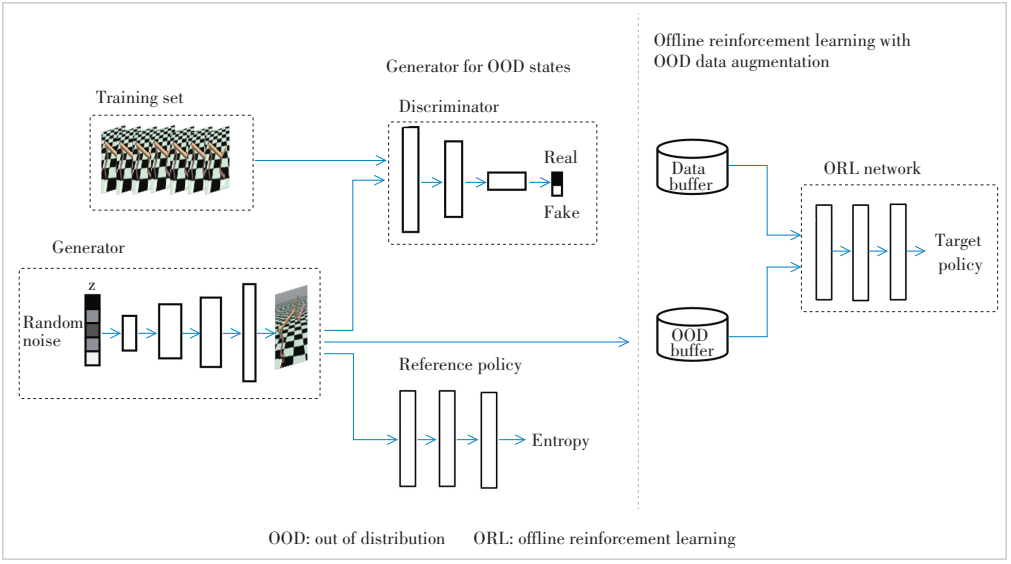
In the setting of offline reinforcement learning, Eq. (1) would be performed on a dataset $\mathcal{D}$, and the result is collected via a behavior policy π_{β}. Due to the aforementioned distribution shift issue, OOD queries usually yield incorrectly estimated Bellman targets. CQL, as a representative OOD-constraint offline RL algorithm, tries to underestimate the Q-values for OOD state-action pairs to prevent the agent from the extrapolation error[6] as follows:

$$\begin{aligned} Q \leftarrow \arg\min_{Q} \alpha \cdot \Big(& \mathbb{E}_{s \sim \mathcal{D}, a \sim \pi(a|s)} \big[Q(s,a) \big] - \\ & \mathbb{E}_{s,a \sim \mathcal{D}} \big[Q(s,a) \big] \Big) + \frac{1}{2} \mathbb{E}_{s,a,s' \sim \mathcal{D}} \left[\left(Q(s,a) - \mathcal{B}^{\pi} Q(s,a) \right)^2 \right], \end{aligned} \tag{2}$$

where π is the new policy to be learned and α is the balance-coefficient. CQL tries to prevent the Q-value of OOD actions from being overestimated, but avoiding OOD queries would degrade the robustness when the agent faces an unseen state.

4 Method

In this section, we first introduce the boundary OOD data in our work. Then we illustrate the primary approach that is utilized to generate the boundary OOD data. Finally, we give the detailed design of the proposed method BCQL. The framework of our proposed method is shown in Fig. 1, where the



▲ **Figure 1. Overall architecture of the proposed Boundary Conservative Q Learning (Boundary-CQL, BCQL) method, where the left column illustrates the pipeline to generate boundary OOD data based on an adversarial generative model, while the right column performs Offline RL with the generated data**

generator is optimized via the discriminator and the pre-trained reference policy simultaneously, and then the generated data is provided for the offline RL algorithm to enhance its robustness.

4.1 Motivation

As mentioned before, to enhance the robustness of the new policy, it is necessary to generate more OOD data. Considering that not all the OOD data are generalizable in the offline setting, we generate certain OOD states that are not very far away from the training dataset and require that the pre-trained reference policy has high entropy in making decisions at these states. In other words, our generated data should have two features: 1) OOD, which means that the distribution should be different from the training dataset, such that the reference agent would be confused about what to do at these states, and 2) boundary, which means that the generated data should not be totally unrelated to the offline dataset. Hence we name this kind of OOD data as boundary OOD data.

4.2 Generating Boundary OOD Data

In this section, we describe how to generate boundary OOD data within the adversarial generative framework.

First, recall that the loss function of GAN is defined as follows:

$$\min_G \max_D \mathbb{E}_{P_{\text{in}}(x)}\left[\log D(x)\right] + \mathbb{E}_{P_{\text{pri}(z)}}\left[\log\left(1 - D\left(G(z)\right)\right)\right], \quad (3)$$

where G denotes the generator, D the discriminator, P_{in} refers to the distribution of in-distribution (ID) data while $P_{\text{pri}(z)}$ refers to some prior distributions such as a Gaussian distribution. The loss function minimizes the error rate of the discriminator for the real data while maximizing the error rate of the discriminator for the generated data until a Nash equilibrium is achieved.

To generate the desired boundary OOD data described in the previous section, we adopt the following loss function:

$$\min_G \max_D \left[\mathbb{E}_{P_{\text{in}}(s)}\left[\log D(s)\right] + \mathbb{E}_{P_{G_B}(s')}\left[\log\left(1 - D(s')\right)\right]\right] + \beta_G \mathbb{E}_{P_{G_B}(s')} H\left[\pi_{pre}(\cdot \mid s')\right], \quad (4)$$

where s' refers to the generated OOD state and π_{pre} refers to the pre-trained reference policy. The first term of Eq. (4) requires that the generated distribution $P_{G_B}(\cdot)$ should still follow the distribution $P_{\text{in}}(\cdot)$ of the offline dataset, but under the constraint defined by the second term of Eq. (4), which forces the generated data to satisfy the requirement that the reference policy has high entropy $H\left[\pi_{\text{pre}}(\cdot \mid s')\right]$ over them, where the entropy of the reference policy π_{pre} is defined as $H\left[\pi_{\text{pre}}(\cdot \mid s')\right] = \sum_a \pi_{\text{pre}}(a \mid s') \log \pi_{\text{pre}}(a \mid s')$. The tradeoff between the above two terms is controlled by another parameter β_G, which should be set based on the specific applications.

In implementation, we use Conditional GAN[33] to generate s' conditioned on the origin state s and use the pre-trained CQL as the reference policy π_{pre}.

4.3 Boundary Conservative Q-Learning

To enhance the robustness of the new policy learned offline, we aim to weaken its confidence at the generated boundary OOD states mentioned before. A direct way to realize this is to add regularization $\mathcal{L}_{\text{BCQL}}$ to maximize the entropy of the new policy when making decisions at these states:

$$L_{\text{BCQL}} = H\left[\pi\left(a \mid \hat{s}\right)\right] = -\sum_a \pi\left(a \mid \hat{s}\right) \log \pi\left(a \mid \hat{s}\right), \quad (5)$$

where $\hat{s}$ is the boundary OOD state generated and π is the new policy. Then the whole loss function of the policy network is as follows:

$$\mathcal{L}_\pi = -\mathbb{E}_{s \sim \mathcal{D}, a \sim \pi(a|s)} Q(s, a) - \lambda \mathcal{L}_{\text{BCQL}}, \quad (6)$$

where λ is the balance-coefficient of the BCQL term, π is the new policy, and $\hat{Q}$ is the target Q network.

As Eq. (6) shows, we can keep the new policy conservative as commonly done in normal offline RL training, but it has to be uncertain as the reference policy does when facing the generated boundary OOD data. Then the loss function of Q networks can be formulated as follows:

$$\mathcal{L}_Q = \alpha \mathbb{E}_{s \sim \mathcal{D}}\left[\mathbb{E}_{a \sim \pi(a|s)}\left[Q(s, a)\right] - \mathbb{E}_{a \sim \pi_\beta(a|s)}\left[Q(s, a)\right]\right] + \frac{1}{2}\mathbb{E}_{s, a, s' \sim \mathcal{D}}\left[\left(Q - \mathcal{B}^\pi Q\right)^2\right], \quad (7)$$

where α is the weight of the CQL term and π_β denotes the empirical behavior policy.

Algorithm 1. Boundary-CQL

Input: Q networks Q, pre-trained reference policy π_{pre}, prior distribution $p_{\text{pri}}(z)$, generator G_B, offline dataset $\mathcal{D}$, generative coefficient β_G, and OOD punishment coefficient λ

1: Train G_B via Eq. (4) using π_{pre} and β_G
2: **for** each iteration **do**
3: Sample a mini-batch $B = (s, a, r, s')$ from $\mathcal{D}$
4: Sample z from $p_{\text{pri}}(z)$ and generate $\hat{s} = G_B(z)$
5: Update the new policy π with $\hat{s}$ and B via Eq. (6)
6: Update the Q networks Q with B via Eq. (7)
7: **end for**
8: Output the new policy π

Algorithm 1 summarizes the main pipeline of the proposed

method. To be specific, we first train the generator G_B via Eq. (4), and then we enter the offline RL loop, where we update the policy network π via Eq. (6) and the Q networks Q via Eq. (7) alternately. Finally, we output the policy network π for the testing stage.

5 Experiments

This section starts with an introduction to the datasets used in our research. Subsequently, the efficacy of our proposed framework is demonstrated through its assessment on Datasets for Deep Data-Driven Reinforcement Learning (D4RL) benchmarks. Further, an examination of the behavior of our generator and its impact on the new policy is conducted. Finally, a sensitive analysis is performed to elucidate the contribution of each parameter.

5.1 Datasets

The experiments presented in this study are carried out on the OpenAIGYM subset of the D4RL[34] tasks. For performance evaluation, we utilize datasets that are a combination of multiple policies, namely medium, medium-replay, medium-expert, and expert. To ensure the robustness of our findings, we conduct all experiments at four distinct random seeds.

5.2 Comparative Study

To demonstrate the superiority of our proposed framework, we conduct a comparative analysis with several state-of-the-art algorithms, including behavior cloning (BC), BEAR[17], Soft Actor-Critic (SAC)[35], twin-delayed deep deterministic policy gradient (TD3)+BC[36], and CQL[8]. In particular, we obtain the results for CQL and BEAR using our implementation, while the effects of BC and SAC are taken from Clean Offline Reinforcement Learning (CORL)[37] and MCQ[20], respectively. Additionally, we obtain the results for TD3+BC from its original publication. Table 1 presents the results, with the highest mean value being denoted in bold.

As is evident from the results presented in Table 1, our proposed algorithm outperforms the other state-of-the-art approaches in the medium, medium-expert, and medium-replay datasets, which exhibit diverse characteristics. In comparison with the basic version of CQL, our approach demonstrates superior performance in estimating OOD data. However, in an expert setting, although we employ a generative network to simulate OOD data, with the constraint on the OOD data, our approach performs better than the original CQL but worse than TD3+BC when employing behavior cloning in the half cheetah task. Overall, these findings highlight the excellent efficiency of our proposed framework in both complex tasks and expert environments.

Fig. 2 displays the performance of BCQL and CQL during the training process. As illustrated by the curves, our proposed algorithm outperforms both BC and basic CQL in the medium, medium-replay, and medium-expert environments, owing to the utilization of augmented data. In contrast, BC employs data obtained through behavior cloning. In the expert environment, the regularization imposed on the generated low-confidence data leads to a minimal impact on the performance of high-confidence data.

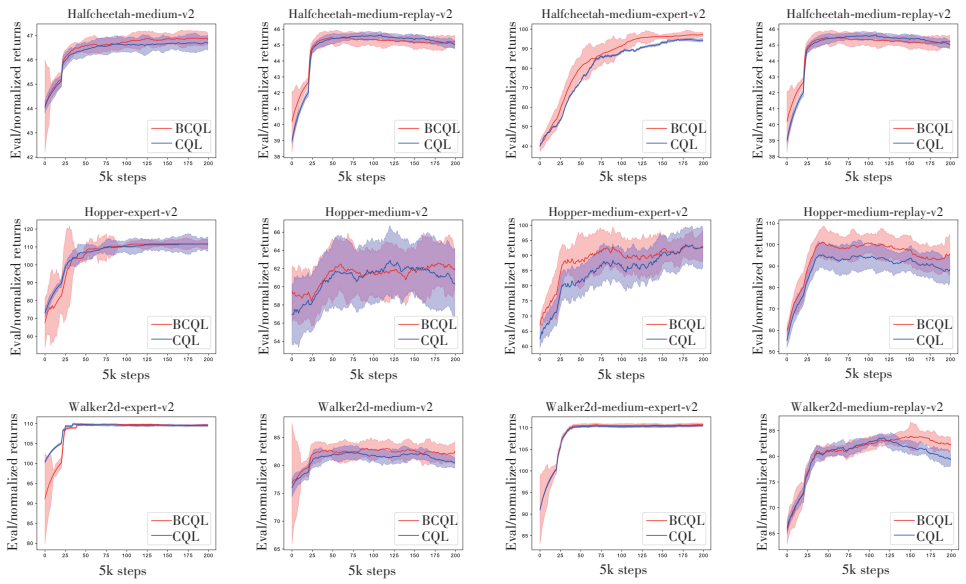
5.3 Behaviors of Generator

The present study employs a GAN-based generator to generate data in various environments, and the real and generated states are visualized, as shown in Fig. 3. The generated data is observed to be irregular yet maintained its validity in comparison to the original data. To assess the effectiveness

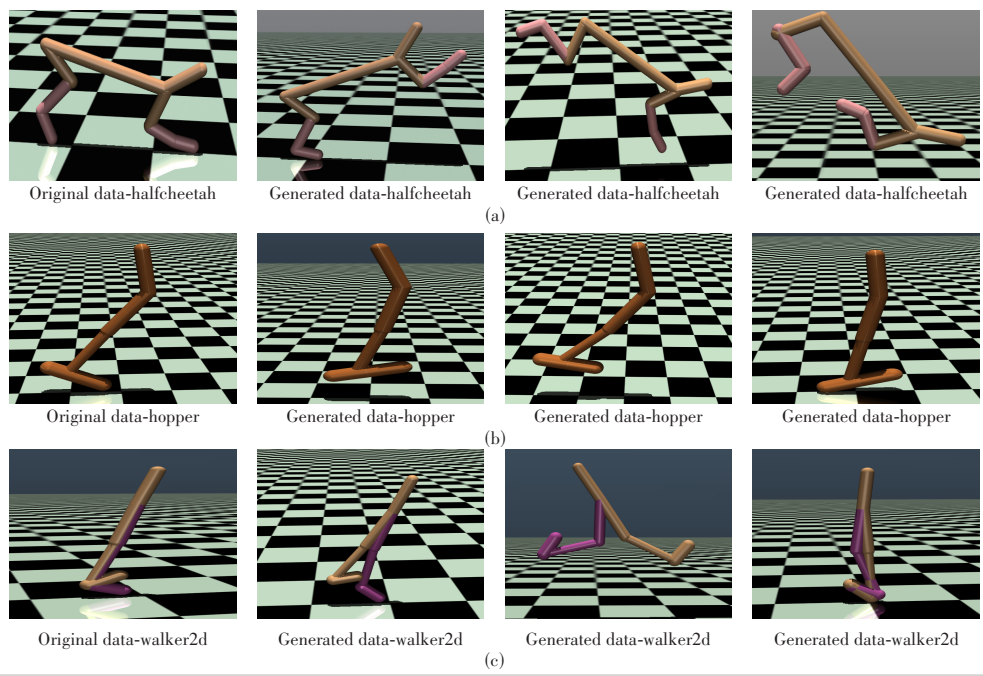
▼Table 1. Performance of BCQL and prior methods on MuJoCo tasks from D4RL, on the normalized return metric (the highest means are bolded)

Task Name	BC	BEAR	SAC	TD3+BC	CQL	BCQL
Halfcheetah-medium-v2	42.4±0.2	37.1±2.3	**55.2±27.8**	48.3±0.3	47.1±0.2	47.1±0.7
Hopper-medium-v2	53.5±2.0	30.8±0.9	0.8±0.0	59.3±4.2	64.9±4.1	**66.1±5.2**
Walker2d-medium-v2	63.2±18.8	56±8.5	-0.3±0.2	83.7±2.1	80.4±3.5	**84.6±2.5**
Halfcheetah-medium-replay-v2	35.7±2.7	36.2±5.6	0.8±1.0	44.6±0.5	45.2±0.6	**46.1±1.5**
Hopper-medium-replay-v2	29.8±2.4	31.1±7.2	7.4±0.5	60.9±18.8	87.7±14.4	**93.9±11.8**
Walker2d-medium-replay-v2	21.8±11.7	13.6±2.1	-0.4±0.3	81.8±5.5	79.3±4.9	**82.5±7.3**
Halfcheetah-medium-expert-v2	56.0±8.5	44.2±13.8	28.4±19.4	90.7±4.3	96±0.8	**97.5±3.2**
Hopper-medium-expert-v2	52.3±4.6	67.3±32.5	0.7±0.0	**98.0±9.4**	93.9±14.3	95.9±13.3
Walker2d-medium-expert-v2	99.0±18.5	43.8±6.0	1.9±3.9	110.1±0.5	109.7±0.5	**110.2±1.0**
Halfcheetah-expert-v2	91.8±1.5	**100.2±1.8**	-0.8±1.8	96.7±1.1	96.3±1.3	98.4±3.2
Hopper-expert-v2	107.7±0.7	108.3±3.5	0.7±0.0	107.8±7	109.5±14.3	**111.7±8.3**
Walker2d-expert-v2	106.7±0.2	106.1±6.0	0.7±0.3	**110.2±0.3**	108.5±0.5	109.7±1.0
Total average	63.3	56.2	7.9	82.7	83.8	87.0

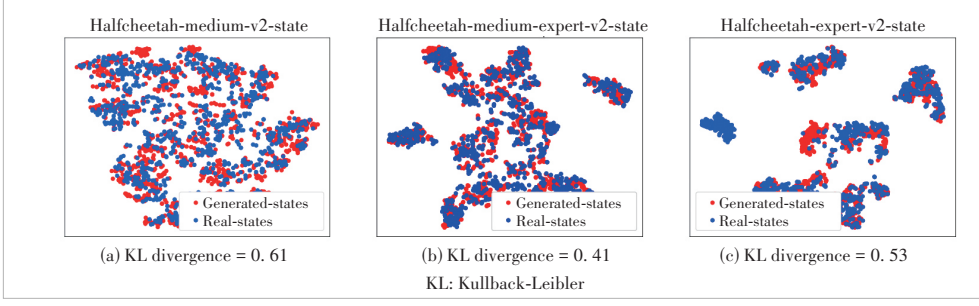
BC: behavior cloning
BCQL: Boundary Conservative Q Learning
BEAR: bootstrapping error accumulation reduction
CQL: Conservative Q-Learning
D4RL: Datasets for Deep Data-Driven Reinforcement Learning
SAC: Soft Actor-Critic
TD3: twin-delayed deep deterministic policy gradient



▲Figure 2. Policy performance during training in different environments



▲Figure 3. Visualization of data generated in different environments



▲Figure 4. Distribution of generated states and real states in different environments

of the generator, experiments are conducted in a halfcheetah environment. Fig. 4 demonstrates that the generated data closely approximates the original data, although the two are not identical. This finding satisfies the boundary requirements of the experiment. Moreover, the distribution of action possibility depicted in the figure indicates that the generated data adheres to the low confidence criteria of a pre-trained RL network. Thus, the generator is capable of producing data with the features described in Section 3.

5.4 Parameter Sensitivity Analysis

We conduct several experiments to evaluate the sensitivity of the following two parameters in our algorithm: the parameter β_G in Eq. (4) and the BCQL weight λ in Eq. (6).

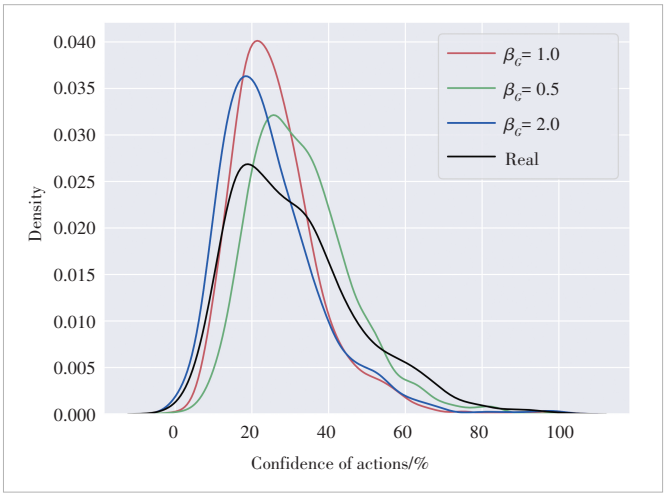
5.4.1 Study of Parameter β_G

The parameter β_G plays a crucial role in training the generator, as it affects its performance. We illustrate this by considering the task of walker2 d-medium. We set the iteration number K for the generator at a fixed value of 5 000 to avoid any confounding effects. As depicted in Fig. 5 and Table 2, the generator's performance is sensitive to changes in β_G. Specifically, increasing β_G leads to a decrease in the trained policy's confidence, while simultaneously increasing the KL divergence. The KL divergence represents the distance between the generated distribution and the original distribution. In order to strike a balance between being close to the original data and having low confidence, we use the smallest possible value for β_G that still

satisfies the low confidence requirement.

5.4.2 Study of Parameter λ

The parameter λ in Eq. (6) affects the behavior of the network when using generated data. Specifically, a higher value of λ results in a more conservative behavior, while a lower value leads to greater flexibility. To investigate the impact of λ on the performance of the network, we conduct experiments while keeping the other parameters constant, and the results are presented in Table 3, which indicates that a large value of λ can be detrimental to performance when the environment is diverse, and therefore, a milder value of λ may be more appropriate. Based on the results, a value of $\lambda = 1.0$ should be suitable in most situations.



▲**Figure 5. Confidence of actions on generated data under different β_G in a walker2d-medium environment**

▼**Table 2. KL divergence between generated states and origin states under different β_G in a walker2d-medium environment**

β_G	KL Divergence	β_G	KL Divergence
0.2	0.08	1.0	0.41
0.4	0.21	1.5	0.57
0.5	0.34	2.0	0.76

KL: Kullback-Leibler

▼**Table 3. Performance of BCQL with different λ, on the normalized return metric (the highest means are bolded)**

Task Name	λ = 0 (CQL)	λ = 0.5	λ = 1	λ = 1.5
Halfcheetah-medium-v2	**47.1±0.2**	46.8±1.3	47.1±0.7	46.1±2.2
Hopper-medium-v2	64.9±4.1	64.8±7.6	64.3±4.2	**66.1±7.2**
Walker2d-medium-v2	80.4±3.5	**84.6±2.5**	81.8±2.1	79.3±4.5
Halfcheetah-medium-replay-v2	45.2±0.6	45.0±1.0	44.9±0.5	**46.1±1.5**
Hopper-medium-replay-v2	87.7±14.4	90.2±10.5	**93.9±11.8**	83.5±14.4
Walker2d-medium-replay-v2	79.3±4.9	**82.5±7.3**	80.7±5.5	77.7±8.9
Halfcheetah-medium-expert-v2	96±0.8	95.2±0.4	**97.5±3.2**	96.9±1.1
Hopper-medium-expert-v2	93.9±14.3	94.0±8.7	**95.9±13.3**	94.8±11.3
Walker2d-medium-expert-v2	109.7±0.5	**110.2±1.0**	110.1±0.5	109.3±0.3
Halfcheetah-expert-v2	96.3±1.3	**98.4±3.2**	97.4±2.3	98.2±1.3
Hopper-expert-v2	106.5±14.3	109.7±7.9	**111.7±8.3**	111.2±10.2
Walker2d-expert-v2	108.5±0.5	108.7±0.3	**109.7±1.0**	109.5±0.5

BCQL: Boundary Conservative Q Learning CQL: Conservative Q-Learning

6 Conclusions

The proposed method BCQL improves the robustness of offline reinforcement learning algorithms while maintaining consistency with the original data distribution, based on a novel OOD simulation technique using a GAN. Extensive experiments are performed on several publicly available offline RL benchmarks, showing that the proposed BCQL method achieves state-of-the-art performance while maintaining high robustness and conservation. Our work highlights the benefits of improving the robustness of offline reinforcement learning algorithms, which is an important research direction given the increasing interest in offline RL applications.

References

[1] CHEN D, CHEN K A, LI Z J, et al. PowerNet: multi-agent deep reinforcement learning for scalable powergrid control [J]. IEEE transactions on power systems, 2022, 37(2): 1007 – 1017. DOI: 10.1109/TPWRS.2021.3100898
[2] CHEN X C, YAO L N, MCAULEY J, et al. A survey of deep reinforcement learning in recommender systems: a systematic review and future directions [EB/OL]. (2021-09-08)[2023-04-12]. https://arxiv.org/abs/2109.03540
[3] OLIFF H, LIU Y, KUMAR M, et al. Reinforcement learning for facilitating human-robot-interaction in manufacturing [J]. Journal of manufacturing systems, 2020, 56: 326 – 340. DOI: 10.1016/j.jmsy.2020.06.018
[4] YU C, LIU J M, NEMATI S, et al. Reinforcement learning in healthcare: a survey [J]. ACM computing surveys, 2023, 55(1): 1 – 36. DOI: 10.1145/3477600
[5] GRIGORESCU S, TRASNEA B, COCIAS T, et al. A survey of deep learning techniques for autonomous driving [J]. Journal of field robotics, 2020, 37(3): 362 – 386. DOI: 10.1002/rob.21918
[6] FUJIMOTO S, MEGER D, PRECUP D. Off-policy deep reinforcement learning without exploration [C]//International Conference on Machine Learning. PMLR, 2019: 2052 – 2062. DOI: 10.48550/arXiv.1812.02900
[7] LEVINE S, KUMAR A, TUCKER G, et al. Offline reinforcement learning: tutorial, review and perspectives on open problems [EB/OL]. (2020-05-04)[2023-04-12]. https://arxiv.org/abs/2005.01643
[8] KUMAR A, ZHOU A, TUCKER G, et al. Conservative Q-learning for offline reinforcement learning [EB/OL]. (2020-06-08)[2022-11-08]. https://arxiv.org/abs/2006.04779
[9] COUPRIE C, FARABET C, NAJMAN L, et al. Indoor semantic segmentation using depth information [EB/OL]. (2013-01-16)[2023-04-02]. https://arxiv.org/abs/1301.3572
[10] MCCRACKEN M W. Robust out-of-sample inference [J]. Journal of econometrics, 2000, 99(2): 195 – 223. DOI: 10.1016/S0304-4076(00)00022-1
[11] YU T H, THOMAS G, YU L T, et al. MOPO: model-based offline policy optimization [EB/OL]. (2020-05-27)[2022-10-08]. https://arxiv.org/abs/2005.13239
[12] GUO K Y, SHAO Y F, GENG Y H. Model-based offline reinforcement learning with pessimism-modulated dynamics belief [EB/OL]. (2022-10-13)[2023-04-02]. https://arxiv.org/abs/2210.06692
[13] WU C Y, MANMATHA R, SMOLA A J, et al. Sampling matters in deep embedding learning [C]//IEEE International Conference on Computer Vision (ICCV). IEEE, 2017: 2859 – 2867. DOI: 10.1109/ICCV.2017.309
[14] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training region-based object detectors with online hard example mining [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016: 761 – 769. DOI: 10.1109/CVPR.2016.89
[15] ROBINSON J, CHUANG C Y, SRA S, et al. Contrastive learning with hard negative samples [EB/OL]. (2020-10-09) [2023-03-23]. https://arxiv.org/abs/

2010.04592
[16] FUJIMOTO S, MEGER D, PRECUP D. Off-policy deep reinforcement learning without exploration [EB/OL]. (2018-12-7) [2022-10-2]. https://arxiv. org/abs/1812.02900
[17] KUMAR A, FU J, TUCKER G, et al. Stabilizing off-policy Q-learning via bootstrapping error reduction [C]//33rd International Conference on Neural Information Processing Systems. Curran Associates Inc., 2019: 11784 – 11794. DOI: 10.48550/arXiv.1906.00949
[18] WU Y F, TUCKER G, NACHUM O. Behavior regularized offline reinforcement learning [EB/OL]. (2019-11-26) [2022-09-11]. https://arxiv. org/abs/1911.11361
[19] AGARWAL R, SCHUURMANS D, NOROUZI M. An optimistic perspective on offline reinforcement learning [EB/OL]. (2019-07-10)[2022-10-11]. https://arxiv.org/abs/1907.04543
[20] LYU J F, MA X T, LI X, et al. Mildly conservative Q-learning for offline reinforcement learning [EB/OL]. (2022-07-09)[2023-04-12]. https://arxiv. org/abs/2206.04745
[21] YANG J K, ZHOU K Y, LI Y X, et al. Generalized out-of-distribution detection: a survey [EB/OL]. (2021-10-21) [2022-03-01]. https://arxiv. org/abs/2110.11334
[22] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks [J]. Communications of the ACM, 2020, 63(11): 139 – 144. DOI: 10.1145/3422622
[23] KINGMA D P and WELLING M. Auto-encoding variational bayes [EB/OL]. (2013-12-20)[2023-02-14]. https://arxiv.org/abs/1312.6114
[24] PIDHORSKYI S, ALMOHSEN R, ADJEROH D A, et al. Generative probabilistic novelty detection with adversarial autoencoders [EB/OL]. (2018-07-06) [2023-04-10]. https://arxiv.org/abs/1807.02588
[25] TIAN K, ZHOU S G, FAN J P, et al. Learning competitive and discriminative reconstructions for anomaly detection [C]//33th AAAI Conference on Artificial Intelligence. ACM, 2019: 5167 – 5174. DOI: 10.1609/aaai.v33i01.33015167
[26] DEECKE L, VANDERMEULEN R, RUFF L, et al. Image anomaly detection with generative adversarial networks [C]//European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. ECMLPKDD, 2018. DOI: 10.1007/978-3-030-10925-7_1
[27] ZHOU K, XIAO Y T, YANG J L, et al. Encoding structure-texture relation with P-net for anomaly detection in retinal images [M]//Computer Vision – ECCV 2020. Cham, Switzerland: Springer international publishing, 2020
[28] WEI H, YE D, LIU Z, et al. Boosting offline reinforcement learning with residual generative modeling [C]//Thirtieth International Joint Conference on Artificial Intelligence. IJCAI, 2021: 3574 – 3580. DOI: 10.24963/ijcai.2021/492
[29] WANG Z, HUNT J J, ZHOU M. Diffusion policies as an expressive policy class for offline reinforcement learning [EB/OL]. (2022-08-12) [2023-05-15]. https://arxiv.org/abs/2208.06193
[30] CHEN H, LU C, YING C, et al. Offline reinforcement learning via high-fidelity generative behavior modeling [EB/OL]. (2022-09-29) [2023-05-10]. https://arxiv.org/abs/2209.14548
[31] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari with deep reinforcement learning [EB/OL]. (2013-12-19)[2022-09-03]. https://arxiv.org/abs/1312.5602
[32] SZITA I, LÖRINCZ A. Learning tetris using the noisy cross-entropy method [J]. Neural computation, 2006, 18(12): 2936 – 2941. DOI: 10.1162/neco.2006.18.12.2936
[33] SOHN K, YAN X, LEE H, et al. Learning structured output representation using deep conditional generative models [C]//28th International Conference on Neural Information Processing Systems. NIPS, 2015: 3483 – 3491
[34] FU J, KUMAR A, NACHUM O, et al. D4RL: datasets for deep data-driven reinforcement learning [EB/OL]. (2020-04-15)[2022-08-15]. https://arxiv.org/abs/2004.07219
[35] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor [C]//International Conference on Machine Learning. Association for Computing Machinery, 2018. DOI: 10.48550/arXiv.1801.01290
[36] FUJIMOTO S, GU S S. A minimalist approach to offline reinforcement learning [EB/OL]. (2021-06-12)[2022-09-13]. https://arxiv.org/abs/2106.06860
[37] TARASOV D, NIKULIN A, AKIMOV D, et al. CORL: research-oriented deep offline reinforcement learning library [C]//3rd Offline Reinforcement Learning Workshop: Offline RL as a "Launchpad". NeurIPS, 2022. DOI: 10.48550/arXiv.2210.07105

Biographies

SHEN Jiahao is currently a postgraduate student in College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, China. His research interests include reinforcement learning and generative model.

JIANG Ke is currently a PhD student in the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, China. His research interest is reinforcement learning.

TAN Xiaoyang (x.tan@nuaa.edu.cn) is currently a faculty member of Department of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, China. His current research interests include machine learning and pattern recognition, computer vision, and reinforcement learning.

Differential Quasi-Yagi Antenna and Array

ZHU Zhihao[1], ZHANG Yueping[2]
(1. Key Laboratory of Ministry of Education of Design and Electromagnetic Compatibility of High-Speed Electronic Systems, Shanghai Jiao Tong University, Shanghai 200240, China;
2. Nanyang Technological University, Singapore 639798, Singapore)

DOI: 10.12142/ZTECOM.202303006
https://link.cnki.net/urlid/34.1294.TN.20230804.1848.002, published online August 7, 2023
Manuscript received: 2023-03-11

Abstract: A novel differential quasi-Yagi antenna is first presented and compared with a normal single-ended counterpart. The simulated and measured results show that the differential quasi-Yagi antenna outperforms the conventional single-ended one. The differential quasi-Yagi antenna is then used as an element for linear arrays. A study of the coupling mechanism between the two differential and the two single-ended quasi-Yagi antennas is conducted, which reveals that the TE_0 mode is the dominant mode, and the driver is the decisive part to account for the mutual coupling. Next, the effects of four decoupling structures are respectively evaluated between the two differential quasi-Yagi antennas. Finally, the arrays with simple but effective decoupling structures are fabricated and measured. The measured results demonstrate that the simple slit or air-hole decoupling structure can reduce the coupling level from −18 dB to −25 dB and meanwhile maintain the impedance matching and radiation patterns of the array over the broad bandwidth. The differential quasi-Yagi antenna should be a promising antenna candidate for many applications.

Keywords: differential quasi-Yagi antenna and array; mutual coupling; surface wave

Citation (Format 1): ZHU Z H, ZHANG Y P. Differential quasi-Yagi antenna and array [J]. ZTE Communications, 2023, 21(3): 37 – 44. DOI: 10.12142/ZTECOM.202303006
Citation (Format 2): Z. H. Zhu and Y. P. Zhang, "Differential quasi-Yagi antenna and array," *ZTE Communications*, vol. 21, no. 3, pp. 37 – 44, Sept. 2023. doi: 10.12142/ZTECOM.202303006.

1 Introduction

With the trend toward the system-on-chip (SoC) and system-in-package (SiP) solutions of radio and radar transceivers, differential antennas are getting popular for their advantages such as low cross-polarization, common-mode rejection, symmetrical radiation pattern, and seamless integration with differential circuits[1 – 4].

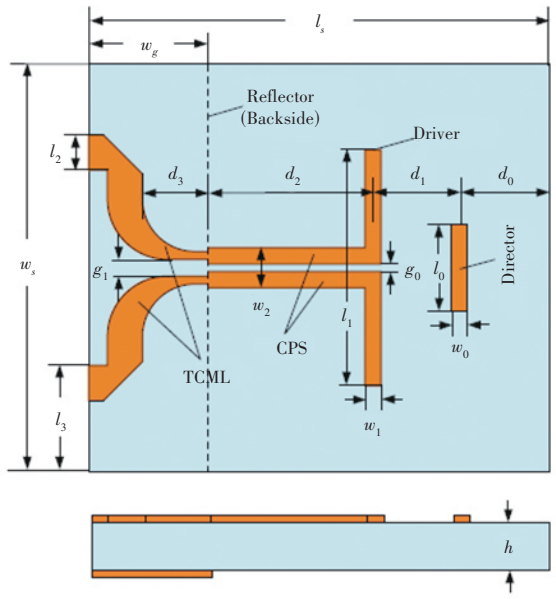
A quasi-Yagi antenna was originally proposed as a single-ended antenna[5] although the driver is differential. Now, the quasi-Yagi antenna is an important type of antennas for end-fire radiation. However, most reported quasi-Yagi antennas are single-ended antennas[5 – 8]. They need to use baluns or complicated feeding networks to convert differential drivers to single-ended inputs, which greatly increases design complexity and degrades antenna performance. In addition, the coupling between two conventional single-ended quasi-Yagi antennas was examined for the design of quasi-Yagi arrays. However, the dominant electromagnetic mechanism and the key structural part for coupling are not clear.

To our best knowledge, the first differential quasi-Yagi antenna[9] was implemented in a thin cavity-down ceramic ball grid array package in low temperature co-fired ceramic (LTCC) technology. It achieved a 10 dB impedance bandwidth of 2.3 GHz from 60.6 GHz to 62.9 GHz and a peak gain of 6 dBi at 62 GHz. The bandwidth was too narrow for 60 GHz radios, which typically require the bandwidth of 7 GHz from 57 GHz to 64 GHz. Furthermore, there has been no differential quasi-Yagi array reported up to now.

In this paper, we present the design, analysis, and measurement of a differential quasi-Yagi antenna and array on high dielectric constant substrates. We design the differential quasi-Yagi antenna and discuss the simulated and measured results in Section 2. We describe the differential quasi-Yagi linear arrays, study the coupling mechanism between the two differential quasi-Yagi antennas, evaluate the effects of decoupling structures, and discuss the simulated and measured results in Section 3. Finally, we draw the conclusion in Section 4.

2 Differential Quasi-Yagi Antenna

Fig. 1 shows the structure and dimensions of the differential quasi-Yagi antenna proposed in this paper. Note that the antenna consists of three elements, namely a driver, a reflector, and a director. The driver is fed by a differential coplanar strip (CPS) line, which is gradually transformed into two single-ended tapered coupled microstrip lines (TCML). More commonly, there are more than one director to improve the antenna gain. The driver and director are horizontally printed on the top surface of a substrate of dielectric constant ε_r, loss tangent δ, length l_s, width w_s, and thickness h. The reflector is

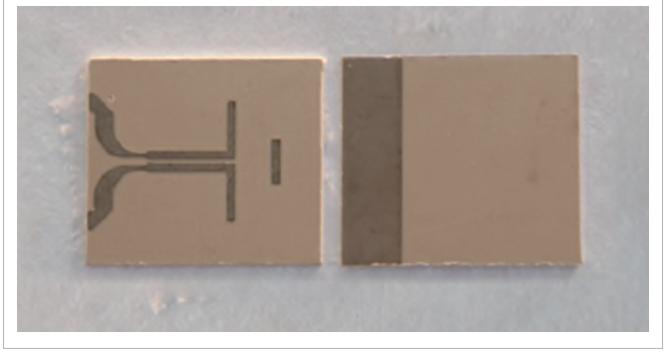


▲Figure 1. Structure and dimensions of the differential quasi-Yagi antenna

usually printed on the bottom surface of the substrate, which has another function as the ground plane.

The design procedure of the differential quasi-Yagi antenna is as follows. It starts with choosing a substrate for the maximum excitation of the surface wave of the TE_0 mode by an electric dipole on the substrate at the central frequency of the operating band. ALEXOPOULOS et al. examined how the substrate affects the excitation of surface waves[10]. They found that the surface wave of the TE_0 mode can be maximumly excited if the critical value of the substrate electrical thickness is satisfied. Using the method described in the classical paper[10], LEONG and ITOH showed that the critical values for the electrical thickness are 0.03, 0.05, and 0.08 for the substrates with ε_r =10.2, 4, and 2.2, respectively[8]. Then, the initial values are set for the length l_0 and width w_0 of the director, the distance d_1 between the director and the driver, the length l_1 and width w_1 of the driver, and the distance d_2 from the driver to the reflector. For simplicity, the same width of $0.02\lambda_0$ can be chosen for the driver and director. The length of the driver is about $0.45\lambda_g$. The length of the director should be shorter than that of the driver and can be $0.3\lambda_g$. The distances between the director and driver and between the driver and reflector are about $0.3\lambda_g$ and $0.25\lambda_g$, respectively. Next, the width w_2 and spacing g_0 of the CPS line can be estimated with the available empirical formula. Finally, the optimum values for the above design parameters can be obtained from High Frequency Structure Simulator (HFSS) simulations.

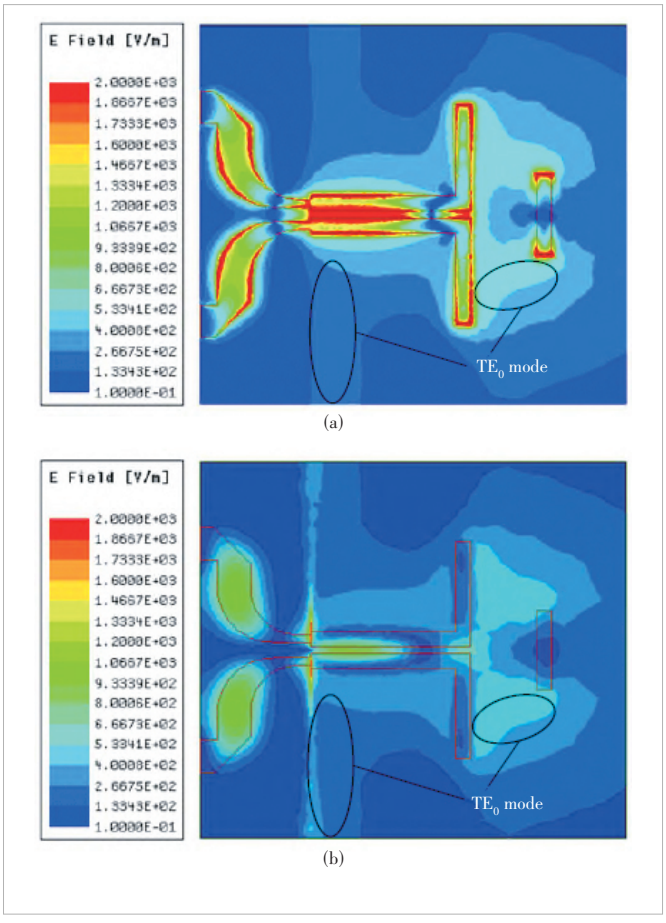
Fig. 2 shows the photo of the differential quasi-Yagi antenna designed and fabricated on a substrate of dielectric constant ε_r = 10.2 and thickness h=0.635 mm at X-band frequencies. The fabricated dimensions are w_s =15 mm, w_g=4.4 mm,



▲Figure 2. Photo of the differential quasi-Yagi antenna

w_0=w_1=0.6 mm, w_2=1.5 mm, l_s=17 mm, l_0=3.2 mm, l_1=8.7 mm, l_2 =1.3 mm, l_3 =3.9 mm, d_0 = 3.3 mm, d_1=3.2 mm, d_2=6.1 mm, g_0=0.3 mm, and g_1=0.6 mm.

Fig. 3 shows the simulated electric field distributions on the top and bottom surfaces of the differential quasi-Yagi antenna. As expected, the surface wave of the TE_0 mode is indeed strongly excited and propagated in the directions normal to the driver. On the one hand, since the polarization direction of the electric field on the driver is the same as that of the electric field on the director, there will be a strong coupling between

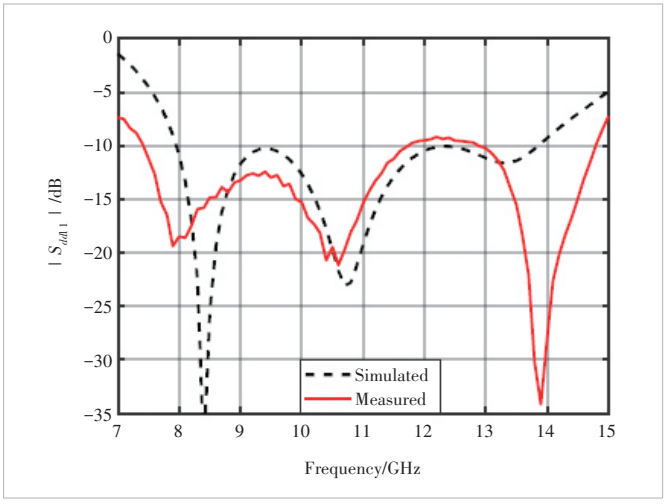


▲Figure 3. Electric field distribution on the (a) top and (b) bottom surfaces

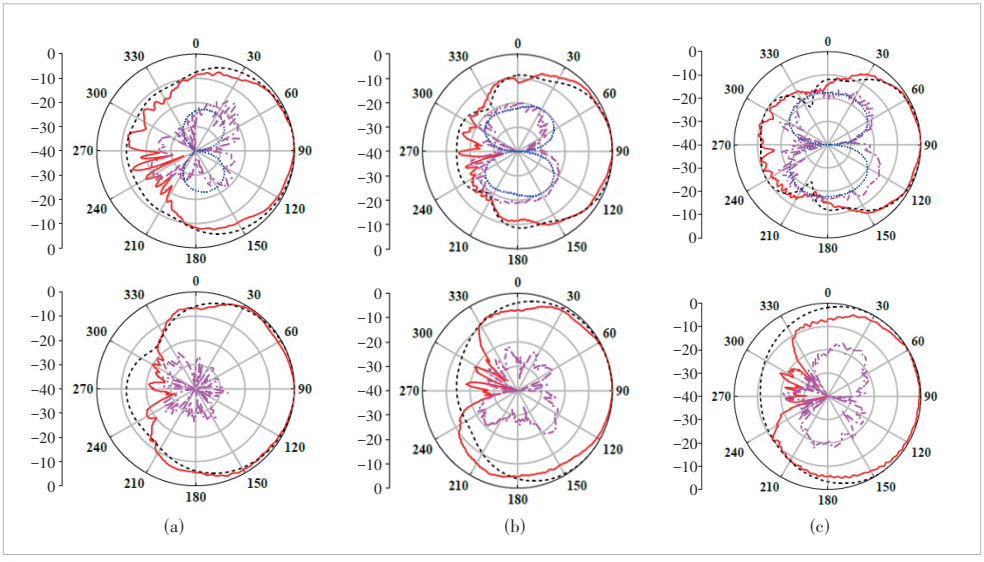
them. Thereby, the surface wave of the TE_0 mode is guided to radiate in the end-fire direction. On the other hand, due to the existence of the ground plane or the reflector, the surface wave of the TE_0 mode cannot be propagated in the grounded substrate region and will be reflected, which further strengthens the radiation in the end-fire direction.

It should be pointed out that the surface wave of the TM_0 mode is quite weakly excited, which can propagate along the axial directions of the driver in both the grounded and ungrounded substrate regions. It causes cross-polarized radiation and deteriorates antenna gain and front-to-back ratio. Therefore, in designing a quasi-Yagi antenna, the major concern is how to excite the surface waves of the TE_0 mode to the greatest extent and the surface waves of the TM_0 mode to the lowest extent.

Fig. 4 shows the simulated and measured $|S_{dd11}|$ of the differential quasi-Yagi antenna as a function of frequency from 7 GHz to 15 GHz. It is evident from the figure that although there are differences between the simulated and measured values, the antenna achieves acceptable matching from 7.5 GHz to 14.8 GHz or a fractional bandwidth of 66% for a voltage standing wave ratio ⩽ 2 at 11.15 GHz. Fig. 5 shows the simulated and measured radiation patterns at 8.2 GHz, 10.6 GHz, and 12.3 GHz. As expected, the antenna radiates an end-fire beam. Fig. 6 shows the simulated and measured gain values. The measured gain values fluctuate between 3.7 dBi and 5.4 dBi from 8 GHz to 12.3 GHz. The simulated radiation efficiency is 94% at 10 GHz. The simulated radiation efficiency at frequencies below 8 GHz and above 13.5 GHz drops quickly, which explains why the gain drops.



▲ **Figure 4. Simulated and measured $|S_{dd11}|$ of the differential quasi-Yagi antenna**

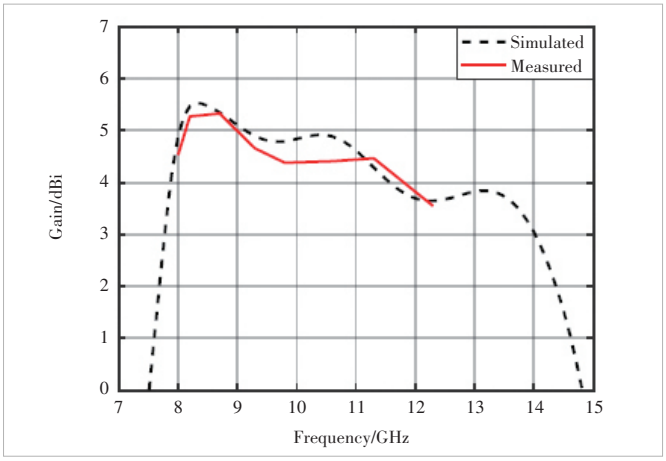


▲ **Figure 5. Simulated and measured E- and H-plane radiation patterns of differential quasi-Yagi antennas at (a) 8.2 GHz, (b) 10.6 GHz, and (c) 12.3 GHz**

Table 1 lists the bandwidth, gain, efficiency, cross-polarization (X-pol), and front-to-back ratio (FBR) for the single-ended and differential quasi-Yagi antennas at 10 GHz. It should be mentioned that these quasi-Yagi antennas have the same size and are fabricated with the same material. Note from the table that the differential quasi-Yagi antenna outperforms the single-ended counterparts.

3 Differential Quasi-Yagi Array

The application of the differential quasi-Yagi antenna as an array element is explored in this section. We limit our effort to E-plane linear arrays because we target their potential use in portable and mobile devices[11 – 12].



▲ **Figure 6. Simulated and measured gain values of the differential quasi-Yagi antenna**

▼Table 1. Single-ended and differential quasi-Yagi antennas

Reference	Bandwidth/%	Gain/dBi	Efficiency/%	X-pol/dB	FBR/dB
Ref. [7]	48	4.6	93	−12	12
This work	73	4.4	94	−21	15

FBR: front-to-back ratio X-pol: cross-polarization

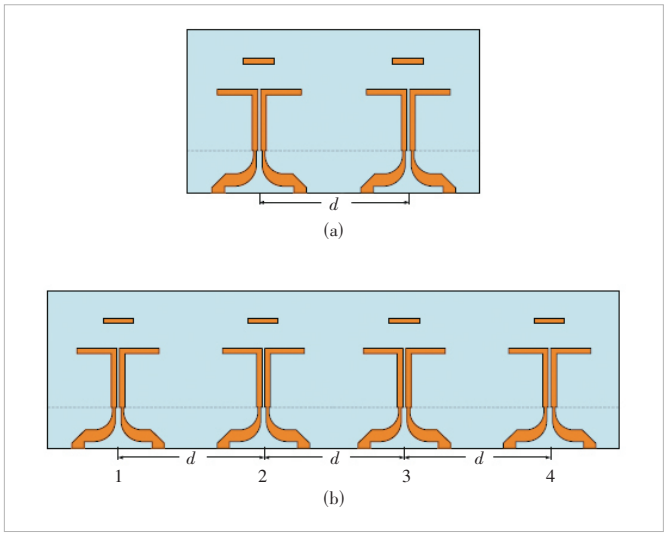
3.1 Linear Arrays

Fig. 7 shows the top views of the two- and four-element E-plane linear differential quasi-Yagi arrays. The differential quasi-Yagi element is the same as the differential quasi-Yagi antenna presented in the previous section. The distance between the two adjacent elements is d.

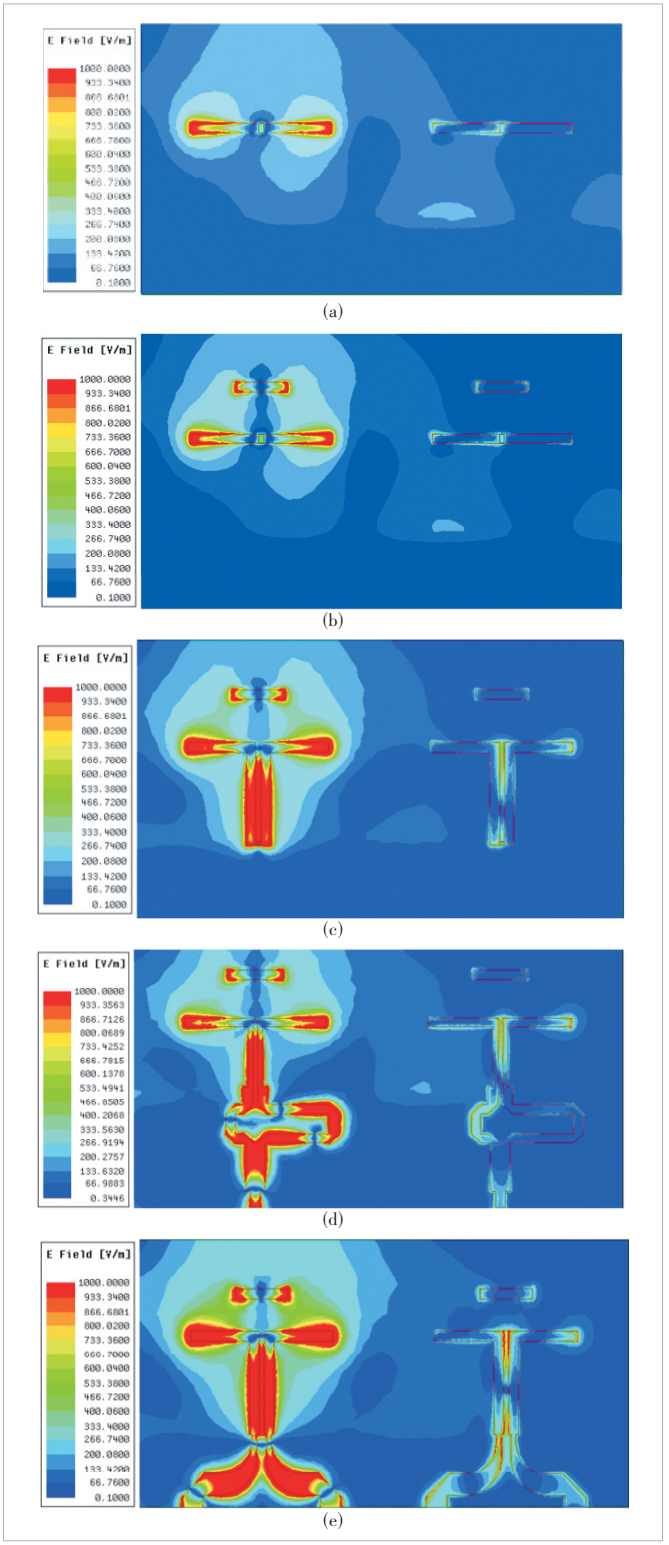
3.2 Coupling Mechanism

The mutual coupling between elements needs to be considered especially in the design of a phased array because strong mutual coupling may cause scan blindness. The mutual coupling is determined by the transmission coefficient $|S_{21}|$ or $|S_{dd21}|$ of an array. DEAL et al.[7] determined the E-plane mutual coupling between two single-ended quasi-Yagi antennas implemented on the same substrate to be below − 18 dB and, in most cases, below − 20 dB for the center-to-center spacing equal to or greater than the half wavelength at 10 GHz[7]. They also made an effort to identify the source of mutual coupling and concluded from their measurements of the testing structures that, for a 15-mm-array spacing that corresponds to the half wavelength at the central frequency of 10 GHz, mutual coupling is almost solely due to coupling through the air[7].

To get a deeper insight into the coupling mechanism, we have conducted a simulated study of mutual coupling between two single-ended and two differential quasi-Yagi antennas, respectively. First, we keep the drivers, the substrate, and the truncated ground plane but remove all the other building blocks. Fig. 8(a) shows the simulated electric field distribution on the top surface of the substrate for the case with one driver fed by a lumped port and the other driver matched to a load at



▲Figure 7. Structures of the E-plane linear differential quasi-Yagi arrays of (a) two elements and (b) four elements

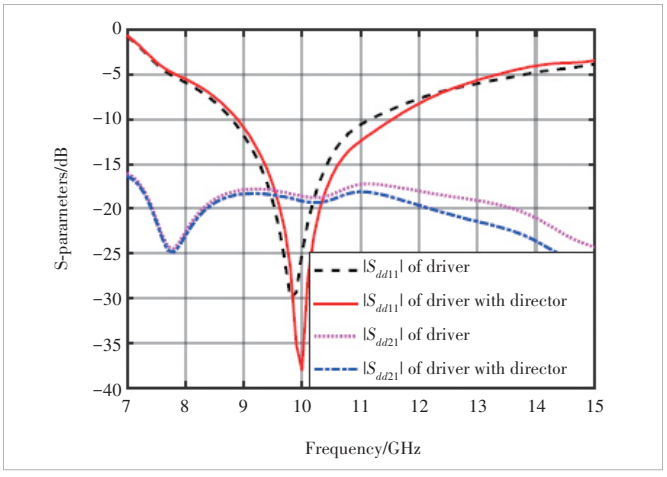


▲Figure 8. Simulated electric field distributions: (a) only driver, (b) driver and director, (c) driver, director and differential coplanar strip (CPS), (d) driver, director, CPS and balun, and (e) driver, director, CPS and tapered coupled microstrip line (TCML)

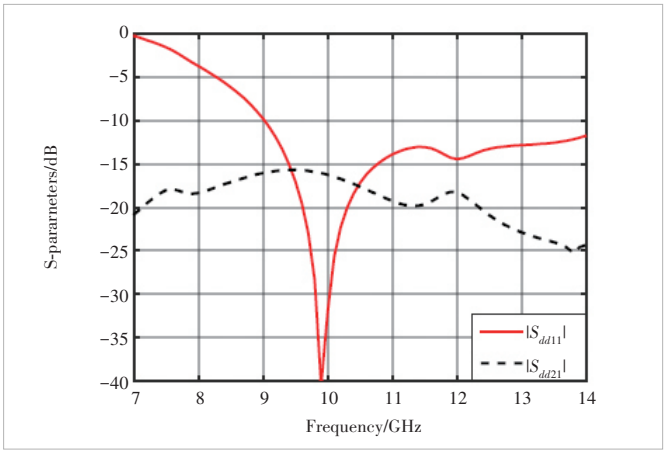
10 GHz. It is evident from the figure that the surface wave of the TE_0 mode has been strongly excited and trapped in the ungrounded substrate region. It is important to note that the grounded substrate region cuts off the surface wave of the TE_0 mode and forces it to be reflected to the ungrounded substrate region. The reflected surface wave of the TE_0 mode is an important source of mutual coupling. Then, we add two directors in the model, and the simulated electric field distribution on the top surface of the substrate at 10 GHz is shown in Fig. 8 (b). It is seen that there is a strong desirable coupling between the excited driver and its director due to the surface wave of the TE_0 mode for end-fire radiation. There seems no effect by the directors on the coupling between the two drivers. Next, we add CPS lines in the model and move the lumped port to the CPS input and the matched load to the other CPS input. Fig. 8 (c) shows the simulated electric field distribution on the top surface of the substrate at 10 GHz. Note that the CPS lines enhance the coupling between the two drivers. Finally, we add in the model baluns to realize the single-ended quasi-Yagi array and TCML to realize a differential quasi-Yagi array, respectively. Figs. 8(d) and 8(e) show the simulated electric field distributions on the top surfaces of the substrates at 10 GHz for the cases of the single-ended and differential quasi-Yagi arrays, respectively. Note that the coupling is stronger for the differential than for the single-ended quasi-Yagi array.

Fig. 9 shows the simulated $|S_{dd11}|$ and $|S_{dd21}|$ as a function of frequency for the cases in Figs. 8(a) and 8(b) with a spacing of 15 mm between the two elements. As expected, the coupling level over the acceptable matching band from 9 GHz to 11.5 GHz is almost the same between the two cases with and without the directors. Fig. 10 shows the simulated $|S_{dd11}|$ and $|S_{dd21}|$ as a function of frequency for the case in Fig. 8(c) with a spacing of 15 mm between the two elements. Note that the CPS extends the matching band to 15 GHz and reduces the coupling level over a wider bandwidth.

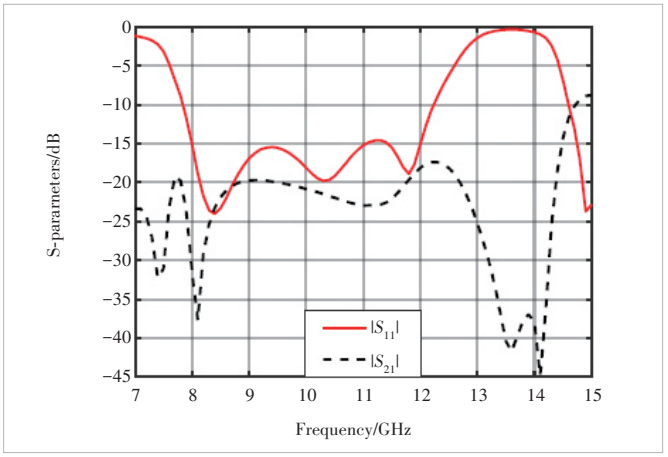
Fig. 11 shows the simulated $|S_{11}|$ and $|S_{21}|$ as a function of frequency for the single-ended quasi-Yagi array of Fig. 8(d) with a spacing of 15 mm between the two elements. Fig. 12 shows the simulated $|S_{dd11}|$ and $|S_{dd21}|$ as a function of fre-



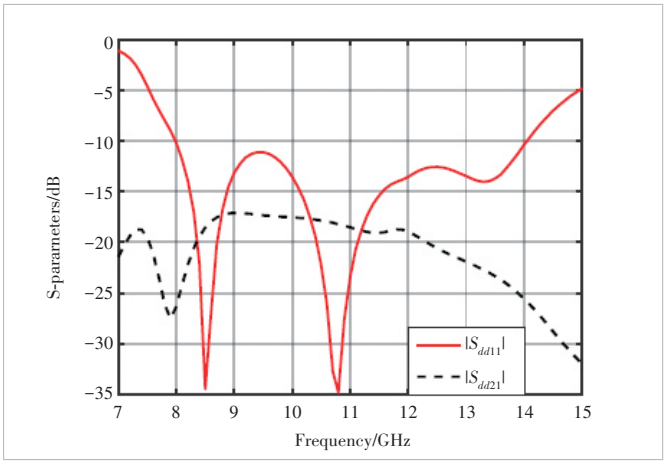
▲ **Figure 9. Simulated $|S_{dd11}|$ and $|S_{dd21}|$ as a function of frequency for the cases in Figs. 8(a) and 8(b)**



▲ **Figure 10. Simulated $|S_{dd11}|$ and $|S_{dd21}|$ as a function of frequency for the case in Fig. 8(c)**



▲ **Figure 11. Simulated $|S_{11}|$ and $|S_{21}|$ as a function of frequency for the single-ended quasi-Yagi array with the spacing of 15 mm between the two elements**



▲ **Figure 12. Simulated $|S_{dd11}|$ and $|S_{dd21}|$ as a function of frequency for the single-ended quasi-Yagi array with a spacing of 15 mm between the two elements**

quency for the differential quasi-Yagi array in Fig. 8(e) with a spacing of 15 mm between the two elements. Note that the matching band is from 7.9 GHz to 12.2 GHz and from 8.0 GHz to 14.0 GHz, respectively, for the single-ended and differential quasi-Yagi arrays. The maximum coupling level is almost the same for the two cases over the respective impedance bandwidths.

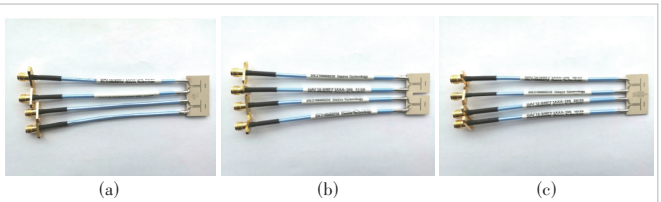
3.3 Decoupling Structures and Effects

It is seen from Fig. 12 that the simulated mutual coupling level for the center-to-center spacing of 15 mm, which is equal to the half wavelength at 10 GHz, is – 18 dB. For a multiple-input and multiple-output (MIMO) array, the mutual coupling level of – 25 dB is desirable. Hence, there is a need to further reduce the mutual coupling. A few decoupling structures such as the neutralization line[13], split-ring resonator[14], slit, and air holes have been attempted. The idea to add a meta-surface as a superstrate to reduce the mutual coupling has not been adopted to keep the low profile of the differential quasi-Yagi array[15]. The simulated decoupling effects of the above structures are summarized as follows. For the case of the neutralization line, the coupling level is greatly reduced to −25 dB but the radiation patterns are distorted, and the gain is reduced by 2 dB at 10 GHz. For the case of the split-ring resonator, it fails to reduce the coupling level but increase the gain by 0.2 – 0.7 dB over the impedance bandwidth. For the cases of slit and air holes, they are very effective to reduce the mutual coupling to −25 dB and meanwhile do not affect the impedance bandwidth and radiation patterns. Hence, the E-plane linear arrays with the slit and air holes are fabricated and measured.

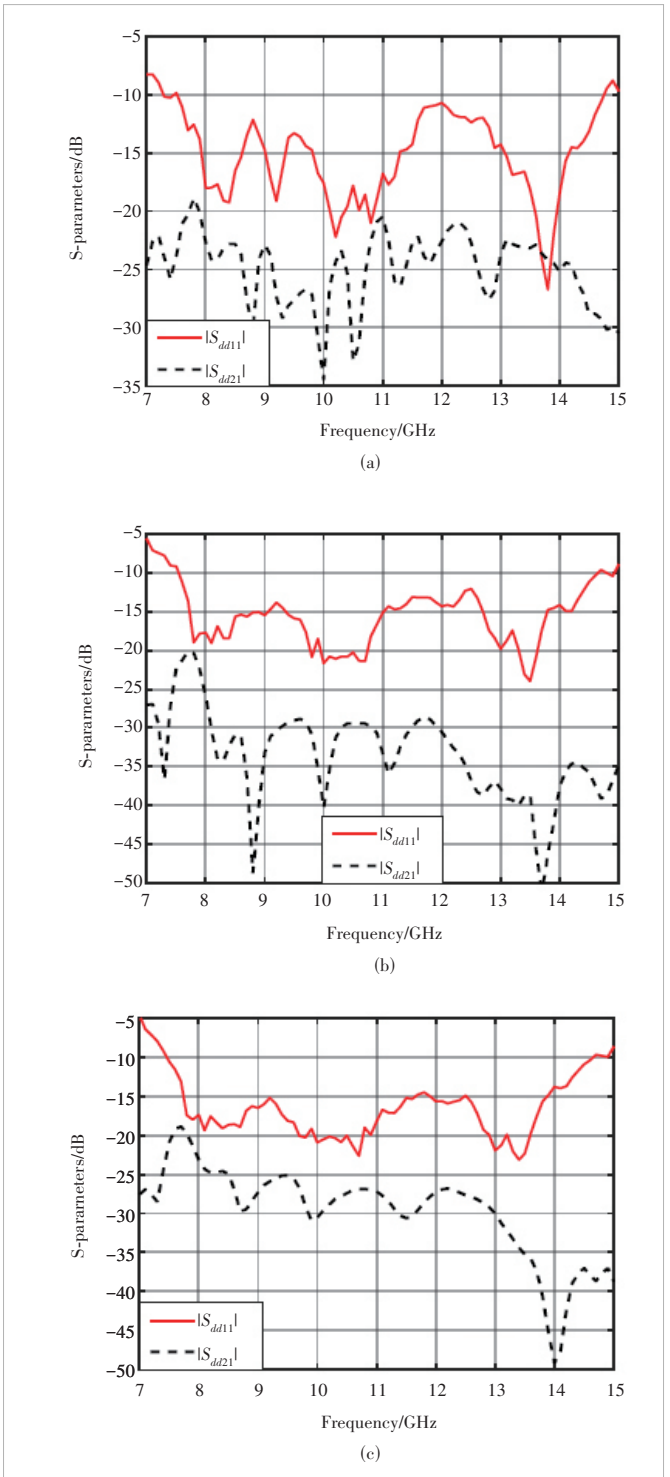
3.4 Results and Discussion

Fig. 13 shows the photos of the two-element differential quasi-Yagi arrays without any decoupling structure, with the slit, and with the air holes. The two differential quasi-Yagi antenna elements are separated by 15 mm and fed with the Subminiature version A (SMA)-connected coaxial cables. The slit is a 2 mm wide cut in between the two elements and after the truncated ground plane. The air holes that have the same diameter of 0.5 mm are punched in between the two elements and after the truncated ground plane with an optimized pattern. It should be mentioned that the details of the SMAs and coaxial cables are unknown. The irregular solder joints and curved coaxial cables are hard to be modelled exactly. Hence, only measured results are discussed.

Fig. 14 shows the measured $|S_{dd11}|$ and $|S_{dd21}|$ as a function of frequency for the two-element differential quasi-Yagi arrays without any decoupling structure, with the slit, and with the



▲ **Figure 13. Photos of the two-element differential quasi-Yagi arrays: (a) solid, (b) slit, and (c) air holes**



▲ **Figure 14. Measured $|S_{dd11}|$ and $|S_{dd21}|$ as a function of frequency for the differential quasi-Yagi antennas: (a) solid, (b) slit, and (c) air holes**

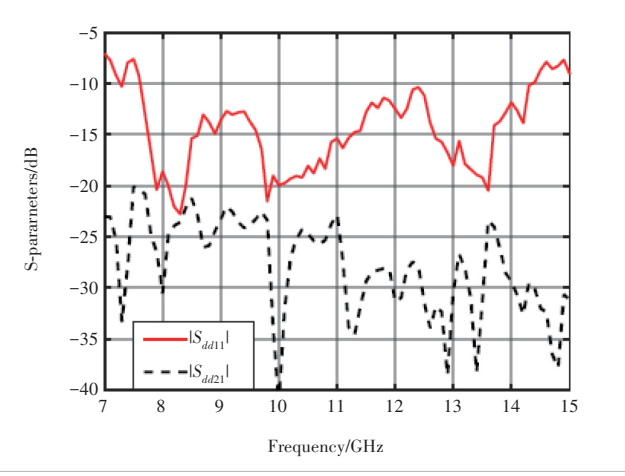
air holes, respectively. Measured results have confirmed that the simple decoupling structures are quite effective to reduce the mutual coupling level below −25 dB.

Fig. 15 shows the photo of the four-element differential quasi-Yagi array without any decoupling structure. The two adjacent differential quasi-Yagi antenna elements are separated by 0.5 free space wavelength at 10 GHz.

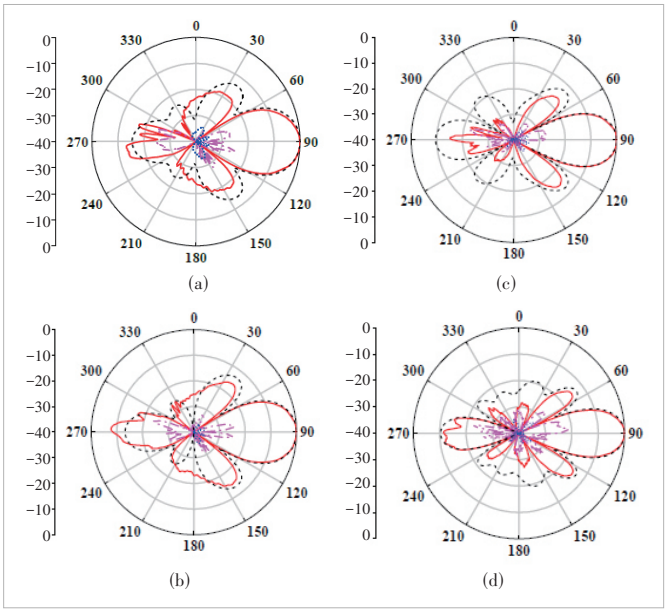
Fig. 16 shows the measured $|S_{dd11}|$ and $|S_{dd21}|$ as a function of frequency for the top two differential quasi-Yagi elements. The measured impedance bandwidth is from 7.6 GHz to 14.3 GHz. The mutual coupling level is below −20 dB over the impedance bandwidth.

Fig. 17 shows the simulated and calculated radiation patterns for the four-element differential quasi-Yagi array at 10 GHz. Due to limitations of our testing facilities, we could not measure the array patterns. We measured the element pattern and obtained the calculated patterns by considering the array factor. It is seen that the calculated and simulated patterns agree quite well for the main lobes. There are differences between the calculated and simulated side lobes.

▲Figure 15. Photo of the four-element differential quasi-Yagi array



▲Figure 16. Measured $|S_{dd11}|$ *and* $|S_{dd21}|$ as a function of frequency for the four-element differential quasi-Yagi array



▲Figure 17. Simulated and calculated E-plane radiation patterns of the four-element differential quasi-Yagi array at (a) 8.2 GHz, (b) 8.7 GHz, (c) 10.6 GHz, and (d) 12.3 GHz

4 Conclusions

In this paper, a novel differential quasi-Yagi antenna is presented and compared with a normal single-ended counterpart for the first time. It is found that the differential quasi-Yagi antenna outperforms the conventional single-ended one. The differential quasi-Yagi antenna is then used as an element for E-plane linear arrays. A study of the coupling mechanism between the two differential quasi-Yagi antennas is conducted. The driver is identified to be the decisive part for the mutual coupling. Four decoupling structures and their effects are evaluated. The arrays with simple but effective decoupling structures are fabricated and measured. The measured results demonstrate that the coupling levels of these arrays can be reduced to less than − 25 dB over the broad bandwidth and the simple slit or air-hole decoupling structure has a negligible effect on the impedance matching and radiation patterns of the arrays. It is anticipated that the differential quasi-Yagi antenna, as a promising antenna candidate, should find wide applications in wireless communication systems, power combining, phased, active, imaging, and MIMO arrays.

References

[1] ZHANG Y P, WANG J J, LI Q, et al. Antenna-in-package and transmit-receive switch for single-chip radio transceivers of differential architecture [J]. IEEE transactions on circuits and systems I: regular papers, 2008, 55(11): 3564 − 3570. DOI: 10.1109/TCSI.2008.925822

[2] ZHANG Y P, WANG J J. Theory and analysis of differentially-driven microstrip antennas [J]. IEEE transactions on antennas and propagation, 2006, 54(4): 1092 − 1099. DOI: 10.1109/TAP.2006.872597

[3] ZHANG Y P. Design and experiment on differentially-driven microstrip anten-

nas [J]. IEEE transactions on antennas and propagation, 2007, 55(10): 2701 – 2708. DOI: 10.1109/TAP.2007.905832
[4] WHITE C R, REBEIZ G M. A differential dual-polarized cavity-backed microstrip patch antenna with independent frequency tuning [J]. IEEE transactions on antennas and propagation, 2010, 58(11): 3490 – 3498. DOI: 10.1109/TAP.2010.2071364
[5] QIAN Y, DEAL W R, KANEDA N, et al. Microstrip-fed quasi-yagi antenna with broadband characteristics [J]. Electronics letters, 1998, 34(23): 2194. DOI: 10.1049/el: 19981583
[6] SOR J, QIAN Y X, ITOH T. Coplanar waveguide fed quasi-Yagi antenna [J]. Electronics letters, 2000, 36(1): 1. DOI: 10.1049/el: 20000132
[7] DEAL W R, KANEDA N, SOR J, et al. A new quasi-yagi antenna for planar active antenna arrays [J]. IEEE transactions on microwave theory and techniques, 2000, 48(6): 910 – 918. DOI: 10.1109/22.846717
[8] LEONG K M K H, ITOH T. Printed quasi-yagi antennas [M]//Printed Antennas for Wireless Communications. Chichester, UK: John Wiley & Sons, Ltd, 2007: 69 – 102. DOI: 10.1002/9780470512241.ch3
[9] SUN M, ZHANG Y P, CHUA K M, et al. Integration of yagi antenna in LTCC package for differential 60-GHz radio [J]. IEEE transactions on antennas and propagation, 2008, 56(8): 2780 – 2783. DOI: 10.1109/tap.2008.927577
[10] ALEXOPOULOS N G, KATEHI P B, RUTLEDGE D B. Substrate optimization for integrated circuit antennas [J]. IEEE transactions on microwave theory and techniques, 1983, 31(7): 550 – 557. DOI: 10.1109/TMTT.1983.1131544
[11] GU X X, LIU D X, BAKS C, et al. A multilayer organic package with four integrated 60GHz antennas enabling broadside and end-fire radiation for portable communication devices [C]//65th Electronic Components and Technology Conference (ECTC). IEEE, 2015: 1005 – 1009. DOI: 10.1109/ECTC.2015.7159718
[12] KIM H T, PARK B S, SONG S S, et al. A 28-GHz CMOS direct conversion transceiver with packaged 2 × 4 antenna array for 5G cellular system [J]. IEEE journal of solid-state circuits, 2018, 53(5): 1245 – 1259. DOI: 10.1109/JSSC.2018.2817606
[13] DIALLO A, LUXEY C, LE THUC P, et al. Study and reduction of the mutual coupling between two mobile phone PIFAs operating in the DCS1800 and UMTS bands [J]. IEEE transactions on antennas and propagation, 2006, 54 (11): 3063 – 3074. DOI: 10.1109/TAP.2006.883981
[14] BAIT-SUWAILAM M M, SIDDIQUI O F, RAMAHI O M. Mutual coupling reduction between microstrip patch antennas using slotted-complementary split-ring resonators [J]. IEEE antennas and wireless propagation letters, 2010, 9: 876 – 878. DOI: 10.1109/LAWP.2010.2074175
[15] SÁENZ E, EDERRA I, GONZALO R, et al. Coupling reduction between dipole antenna elements by using a planar meta-surface [J]. IEEE transactions on antennas and propagation, 2009, 57(2): 383 – 394. DOI: 10.1109/TAP.2008.2011249

Biographies

ZHU Zhihao received his BS degree from Hangzhou Dianzi University, China in 2016. He is currently pursuing an MS degree in electronic science and technology with Shanghai Jiao Tong University, China. His research interests include antenna theory and design, especially in shorted patch antennas, broadband printed antennas, antenna decoupling, and antenna-in-package (AiP).

ZHANG Yueping (eypzhang@ntu.edu.sg) is a full professor with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He is a Distinguished Lecturer of the IEEE Antennas and Propagation Society (IEEE AP-S) and Fellow of the IEEE. He was a member of the IEEE AP-S Field Award Committee (2015 – 2017) and an associate editor of *IEEE Transactions on Antennas and Propagation* (2010 – 2016). Prof. ZHANG has published numerous papers, including two invited papers and one regular paper in the *Proceedings of the IEEE* and one invited paper in *IEEE Transactions on Antennas and Propagation*. He is a Chinese radio scientist who has published a historical article in English learned journals such as the *IEEE Antennas and Propagation Magazine*. He received the 2012 IEEE AP-S Sergei A. Schelkunoff Prize Paper Award. Prof. ZHANG has been invited to deliver plenary speeches at the flagship conferences organized by IEEE, CIE, EurAAP, and IEICE. He received the Best Paper Award from the 2nd IEEE/IET International Symposium on Communication Systems, Networks and Digital Signal Processing, the Best Paper Prize from the 3rd IEEE International Workshop on Antenna Technology, and the Best Paper Award from the 10th IEEE Global Symposium on Millimeter-Waves. He holds seven US patents. He has made pioneering and significant contributions to the development of AiP technology. He received the 2020 IEEE AP-S John Kraus Antenna Award. His current interests are in the development of antenna-on-chip (AoC) technology for very large-scale antenna integration and characterization of chip-scale propagation channels at terahertz for wireless chip area networks (WCAN).

Massive Unsourced Random Access Under Carrier Frequency Offset

XIE Xinyu[1], WU Yongpeng[1], YUAN Zhifeng[2,3], MA Yihua[2,3]
(1. Shanghai Jiao Tong University, Shanghai 200240, China;
2. ZTE Corporation, Shenzhen 518057, China;
3. State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518055, China)

DOI: 10.12142/ZTECOM.202303007
https://link.cnki.net/urlid/34.1294.TN.20230830.1507.003, published online August 30, 2023
Manuscript received: 2022-05-22

Abstract: Unsourced random access (URA) is a new perspective of massive access which aims at supporting numerous machine-type users. With the appearance of carrier frequency offset (CFO), joint activity detection and channel estimation, which is vital for multiple-input and multiple-output URA, is a challenging task. To handle the phase corruption of channel measurements under CFO, a novel compressed sensing algorithm is proposed, leveraging the parametric bilinear generalized approximate message passing framework with a Markov chain support model that captures the block sparsity structure of the considered angular domain channel. An uncoupled transmission scheme is proposed to reduce system complexity, where slot-emitted messages are reorganized relying on clustering unique user channels. Simulation results reveal that the proposed transmission design for URA under CFO outperforms other potential methods.

Keywords: activity detection; channel estimation; frequency offset; massive machine-type communication; massive MIMO

Citation (Format 1): XIE X Y, WU Y P, YUAN Z F, et al. Massive unsourced random access under carrier frequency offset [J]. ZTE Communications, 2023, 21(3): 45 - 53. DOI: 10.12142/ZTECOM.202303007
Citation (Format 2): X. Y. Xie, Y. P. Wu, Z. F. Yuan, et al., "Massive unsourced random access under carrier frequency offset," *ZTE Communications*, vol. 21, no. 3, pp. 45 - 53, Sept. 2023. doi: 10.12142/ZTECOM.202303007.

1 Introduction

Massive machine-type communication (mMTC) [1], also known as massive access[2], is one of the three typical application scenarios in the 5G mobile communication system. MMTC aims at establishing reliable communications for a massive number of cheap devices with sporadic data stream patterns and short packet length, which is quite different from conventional human-type communications (HTC). Hence, the design of efficient massive connectivity schemes requires investigating novel theories and paradigms.

To reduce signaling overhead and latency, existing mMTC schemes generally follow a grant-free random access (RA) protocol[3] where users directly transmit data to the base station (BS) without any approval. One type of grant-free RA scheme[4 - 6] allocates unique pilots as identities to users; they are sent first as preambles for activity detection (AD) and channel estimation (CE). Data transmission happens in the next stage leveraging efficient RA techniques like sparse code multiple access (SCMA)[7] with individual user codebooks. A novel paradigm of unsourced random access (URA) was first addressed in Ref. [8]. Different from pilot-based RA schemes, URA users are forced to use the same codebook for data transmission. Since the transmitted codewords contain no user identities, the BS only acquires a list of transmitted messages without linking them to specific active users. A finite block-length (FBL) achievability bound was derived in Ref. [8], and it is found that there exists an important gap between conventional RA schemes like ALOHA and the benchmark.

It is evident that an intuitive URA scheme is closely related to a compressed sensing (CS) recovery problem, involving AD to the set of codewords each linked to a potential message sequence. However, directly applying CS techniques is impractical because the codebook size grows exponentially to the message length. To approach the FBL bound with manageable complexity, the coded compressed sensing (CCS) framework[9] is investigated which couples an outer tree code and an inner CS code. More specifically, each datum is partitioned into several sub-blocks; they are coupled by appending parity check bits generated from linear block coding, thereby forming fragments to be sent over multiple slots using a common codebook. The BS detects codeword activity via a CS recovery method, and then reconstructs the original messages by a tree-based forward error correction strategy.

Utilizing a large number of antennas at the BS, the massive

This work was supported by the ZTE Industry-University-Institute Cooperation Funds under Grant No. HC-CN-20201116001.

multiple-input and multiple-output (MIMO) technology provides high spatial resolution within the same time/frequency resource to increase spectral efficiency. Different from additive white Gaussian noise (AWGN) channels, it requires channel estimation for detection under MIMO fading channels, forming URA a joint AD and CE (JADCE) problem that is considered as a multiple measurement vector (MMV) problem and solved by CS algorithms like approximate message passing (AMP)[4]. Leveraging channel structural sparsity in the virtual angular domain, Ref. [5] presented a performance gain by performing AD at the spatial domain and CE at the angular domain. Promoting the CCS-based URA to the massive MIMO scenario, the work of Ref. [10] introduced a covariance-based paradigm and proposed a maximum likelihood decoder estimating large-scale fading coefficients (LSFCs) for AD. Ref. [11] further promoted the case of independent and identically distributed (i.i.d.) channels considered in Ref. [10] to the case of correlated channels. Leveraging the rich spatial information reserved in multiple antennas, Refs. [12 – 14] captured the strong similarity/correlation between slot-wise user channels and proposed uncoupled URA transmission schemes. In Refs. [13] and [14], the authors considered the angular domain MIMO channel and proposed an expectation-maximization-aided generalized approximate message passing algorithm with a Markov random field support structure (EM-MRF-GAMP) for JADCE. The similar statistics of slot-wise channels of each user couple the split message fragments, which eliminates the need for redundancies to improve the coding rate. Accordingly, message stitching takes on the form of a clustering decoder, recognizing slot-distributed channels of each active user based on similarity. Eliminating the need for redundancies in CCS, the uncoupled URA schemes[12 – 14] achieve higher coding rates and spectral efficiency.

Unfortunately, all the above works consider an ideal scenario where users are perfectly synchronized with the BS. However, caused by the mismatch between the carrier frequencies of the local oscillators at users and the BS, carrier frequency offset (CFO) inevitably exists. CFO corrupts the phase of channel measurements and imposes significant contamination on the JADCE results. Few works in the context of massive access address the issue of CFO estimation. In Ref. [15], the authors adopted a Lasso-based method for CFO estimation, but only focused on the AD results. The authors of Ref. [16] introduced a CS method for JADCE under CFO in the framework of orthogonal frequency division multiple access (OFDMA). The key idea is to expand the measurement matrix with a finite number of discrete frequency offsets sampled in the possible region. The JADCE problem as a generalized MMV problem with structured sparsity is then solved by the proposed structured-GAMP algorithm. However, the CFO estimation in Ref. [16] is discrete, and the measurement matrix size will increase accordingly in URA applications, resulting in greater computational complexity. To our knowledge, no existing work has discussed URA under CFO.

In this paper, confronted with the issue of CFO, we formulate URA as a JADCE problem with a bilinear signal detection structure. We consider the MIMO channel in the angular domain to promote the sparsity of the CS problem. A novel CS algorithm termed Markov-chain-aided parametric bilinear generalized approximate message passing (MC-PBiGAMP) is proposed for JADCE, which captures the clustered sparsity structure of the angular domain channel. An uncoupled transmission design for URA is then employed to reduce system complexity. With messages divided for slotted emitting, data list reconstruction is conducted in the form of a clustering decoder leveraging unique channel statistics. Simulation results show that the proposed method outperforms state-of-the-art approaches in terms of JADCE and reaches reliable URA system performance.

The rest of the paper is organized as follows. The URA system model is given in the next section. In Section 3, the MC-PBiGAMP algorithm is proposed for JADCE under CFO. Then, the uncoupled URA transmission design of low complexity is discussed in Section 4. Numerical results are presented in Section 5, followed by concluding remarks drawn in Section 6. Throughout this paper, the j-th column and i-th row of matrix $\boldsymbol{X}$ are denoted by $\boldsymbol{x}_j$ and $\boldsymbol{x}_{i,:}$, respectively, and the (i,j)-th entry of $\boldsymbol{X}$ is represented by $x_{i,j}$; $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$ stand for the conjugate, transpose, and conjugate transpose, respectively. Denote $\|\boldsymbol{x}\|$ the Euclid norm of vector $\boldsymbol{x}$, and $\|\cdot\|_2$ and $\|\cdot\|_F$ represent the the l_2-norm and the Frobenius norm, respectively. For an integer $X>0$, the shorthand notation $[X]$ stands for the set $\{1,2,\cdots,X\}$. Finally, $\mathcal{CN}(x;\hat{x},\mu^x)$ signifies the complex Gaussian distribution of a random variable x with mean $\hat{x}$ and variance μ^x.

2 System Model

In this paper, we consider an uplink transmission scenario where K_a active users communicate to a single BS equipped with a uniform linear array (ULA) of M half-wavelength spaced antennas. The channel $\tilde{\boldsymbol{h}}_k \in \mathbb{C}^M$ between the k-th user and the BS is described on a geometric basis as[5]

$$\tilde{\boldsymbol{h}}_k = \sqrt{\rho_k}\sum_{l=1}^{L} g_{k,l}\,\boldsymbol{e}\left(\theta_{k,l}\right), \tag{1}$$

where ρ_k is the LSFC which follows the standard Log-distance path loss model as $\log_{10}\rho_k = -128.1 - 37.6log_{10}\left(D_k\right)$ with distance D_k measured in km, $g_{k,l}\sim\mathcal{CN}(0,1)$ is the complex path gain of the l-th path, $\theta_{k,l}\in[-\pi/2,\pi/2]$ is the angle of arrival (AOA), and $\boldsymbol{e}(\theta_{k,l})$ is given by

$$\boldsymbol{e}\left(\theta_{k,l}\right) = \frac{1}{\sqrt{M}}\left[1, e^{-j\pi\sin\theta_{k,l}},\cdots, e^{-j\pi(M-1)\sin\theta_{k,l}}\right]^T. \tag{2}$$

The spatial domain channel can be transformed to the angu-

lar domain by

$$\boldsymbol{h}_k = \boldsymbol{U}_M^H \tilde{\boldsymbol{h}}_k, \tag{3}$$

where $\boldsymbol{U}_M^H$ is the M-dimensional discrete Fourier transform (DFT) matrix. The angular domain channel is sparse with few elements of dominant magnitude. This is attributed to the high spatial resolution of massive MIMO, leveraging large-scale antennas against finite propagation paths. Also, due to the spectral leakage of DFT, the angular domain channel reveals a clustered sparsity structure with nonzero elements spaced adjacent to dominant elements[17].

In the URA scenario, active users pick up codewords/columns from a common codebook/coding matrix to transmit the B-bit payload. We choose to generate the common codebook by the sparse regression code[18]. Each entry of the codebook $A = [a_1, \cdots, a_{2^B}] \in \mathbb{C}^{N \times 2^B}$ is generated from an i.i.d. Gaussian distribution $\mathcal{CN}(1, 1/N)$ such that $\mathbb{E}\left\{\| \mathrm{a} \|_2^2\right\} = 1$. We assume that oscillators at users are synchronized to the same frequency through calibration[19-20], i.e., CFOs are the same for all users. Considering a block fading channel where channel coefficients remain unchanged in N symbol transmissions, the received signal at the BS under CFO is represented as

$$\tilde{\boldsymbol{Y}} = \sum\nolimits_k \mathrm{diag}(\boldsymbol{\tau}(\omega)) \boldsymbol{a}_{i_k} \tilde{\boldsymbol{h}}_k^T + \tilde{\boldsymbol{W}} = \mathrm{diag}(\boldsymbol{\tau}(\omega)) \boldsymbol{A} \boldsymbol{B} \tilde{\boldsymbol{H}} + \tilde{\boldsymbol{W}}, \tag{4}$$

where $\omega \triangleq 2\pi \Delta f T$ with Δf the frequency offset between users and the BS in Hz and T the sampling period, $\boldsymbol{\tau}(\omega) = [1, e^{j\omega}, \cdots, e^{j(N-1)\omega}]^T$ is the corresponding phase rotation vector due to CFO, $\boldsymbol{B} \in \{0,1\}^{2^B \times K_a}$ is a selection matrix, $\tilde{\boldsymbol{H}} = [\tilde{\boldsymbol{h}}_1, \cdots, \tilde{\boldsymbol{h}}_{K_a}]^T \in \mathbb{C}^{K_a \times M}$, and $\tilde{\boldsymbol{W}}$ is the AWGN matrix with elements generated from i.i.d. $\mathcal{CN}(0, \sigma_w^2)$. The (i_k, k)-th entry of $\boldsymbol{B}$ is nonzero only if the k-th user transmits the information sequence $\boldsymbol{m}_k$ with decimal$(\boldsymbol{m}_k) = i_k$, where decimal$(\boldsymbol{m}_k)$ is the the radix ten equivalent of $\boldsymbol{m}_k$. The received signal in the angular domain can be calculated as

$$\boldsymbol{Y} = \mathrm{diag}(\boldsymbol{\tau}(\omega)) \boldsymbol{A} \boldsymbol{B} \tilde{\boldsymbol{H}} \boldsymbol{U}_M^* + \tilde{\boldsymbol{W}} \boldsymbol{U}_M^* = \mathrm{diag}(\boldsymbol{\tau}(\omega)) \boldsymbol{A} \boldsymbol{B} \boldsymbol{H} + \boldsymbol{W}, \tag{5}$$

where $\boldsymbol{H} = [\boldsymbol{h}_1, \cdots, \boldsymbol{h}_{K_a}]^T$, and $\boldsymbol{W} \triangleq \tilde{\boldsymbol{W}} \boldsymbol{U}_M^*$ is the equivalent noise matrix.

3 Proposed Algorithm for JADCE under CFO

3.1 Problem Formulation and Probability Model

Data detection in URA is to determine which column of $\boldsymbol{A}$ is transmitted. For better illustration, we rewrite Eq. (5) as

$$\boldsymbol{Y} = \mathrm{diag}\left(\boldsymbol{U}_N^* \mathbf{c}\right) \boldsymbol{A} \boldsymbol{X} + \boldsymbol{W} = \boldsymbol{Z} + \boldsymbol{W}, \tag{6}$$

where $\boldsymbol{U}_N$ is the N-dimensional DFT matrix, $\boldsymbol{c} = \boldsymbol{U}_N^T \boldsymbol{\tau}(\omega)$ is sparse due to the Vandermonde structure of $\boldsymbol{\tau}(\omega)$, $\boldsymbol{X} \triangleq \boldsymbol{B}\boldsymbol{H}$, and $\boldsymbol{Z} \triangleq \mathrm{diag}\left(\boldsymbol{U}_N^* \boldsymbol{c}\right) \boldsymbol{A} \boldsymbol{X}$. Since $K_a \ll 2^B$, $\boldsymbol{X}$ is row sparse. Furthermore, the sparse rate of $\boldsymbol{X}$ is promoted due to the sparsity of angular domain channels. Our purpose is to recover sparse $\boldsymbol{c}$ and $\boldsymbol{X}$ from the noisy observation $\boldsymbol{Y}$ with known $\boldsymbol{U}_N$ and $\boldsymbol{A}$, recognized as a CS recovery problem. Note that the URA receiver does not attempt to link the message to its source device, therefore, we do no need to reconstruct $\boldsymbol{B}$.

To solve the structured-matrix estimation problem, we start with reformulating the random variable dependency corresponding to Eq. (6), taking on the form

$$z_{n,m} = \sum\nolimits_{i=1}^{N} \sum\nolimits_{j=1}^{2^B} c_i z_{n,m}^{(i,j)} x_{j,m}, \tag{7}$$

where $z_{n,m}^{(i,j)} \triangleq u_{n,i} a_{n,j}$ is an element of a third-order tensor $\{\boldsymbol{z}_n^{(i,j)} = [z_{n,1}^{(i,j)}, \cdots, z_{n,M}^{(i,j)}]^T\}_{\forall i,j}$. For the sparse vector $\boldsymbol{c}$, we assign a Bernoulli-Gaussian prior to each independent component c_i, i.e.,

$$p(c_i) = \lambda_c \delta(c_n) + (1 - \lambda_c) \mathcal{CN}(0, \sigma_c^2), \tag{8}$$

where λ_c is the sparsity fraction of $\boldsymbol{c}$ and $\delta(\cdot)$ represents the Dirac-delta function. Similarly, a Bernoulli-Gaussian prior is used to model the conditional probability

$$p(x_{j,m} | s_{j,m}) = \delta(x_{j,m}) \delta(s_{j,m} + 1) + \mathcal{CN}(x_{j,m}; 0, \sigma_x^2) \delta(s_{j,m} - 1), \tag{9}$$

where $s_{j,m} \in \{-1, 1\}$ is the binary state of $x_{j,m}$ with $s_{j,m} = \pm 1$ signifying that $x_{j,m}$ is nonzero/zero. To capture the clustered sparsity of the angular domain channel support $\boldsymbol{s}_{j,:} = [s_{j,1}, \cdots, s_{j,M}]$, we employ the Markov chain (MC) model[21] as follows

$$p(\boldsymbol{s}_{j,:}) = p(s_{j,1}) \prod\nolimits_{m=2}^{M} p(s_{j,m} | | s_{j,m-1}), \tag{10}$$

where the transition probability $p(s_{j,m} | s_{j,m-1})$ is given by

$$p(s_{j,m} | s_{j,m-1}) = \begin{cases} (1 - p_{01})^{1 - s_{j,m}} p_{01}^{s_{j,m}}, s_{j,m-1} = 0 \\ (1 - p_{10})^{s_{j,m}} p_{01}^{1 - s_{j,m}}, s_{j,m-1} = 1 \end{cases}, \tag{11}$$

and $p(s_{j,1})$ is initialized as a steady-state distribution with $p(s_{j,1} = 0) = p_{10}/(p_{01} + p_{10})$ and $p(s_{j,1} = 1) = p_{10}/(p_{01} + p_{10})$. The MC model depicts the average cluster size and the average gap between two clusters by parameters p_{10} and p_{01}, respectively. A smaller value of p_{10} indicates a larger cluster length and a smaller value of p_{01} leads to a larger average gap between two clusters. And in general, the sparsity of $\boldsymbol{x}$ is measured as $\lambda_x \triangleq p_{10}/(p_{01} + p_{10})$.

Following the Bayesian theory, we derive the posterior probability density of $\boldsymbol{c}$ and $\boldsymbol{X}$ given $\boldsymbol{Y}$ as

$$p(\boldsymbol{c},\boldsymbol{X}|\boldsymbol{Y}) \propto \prod_{n,m} p\left(y_{n,m}|z_{n,m}\right)\prod_i p(c_i)\prod_j p\left(\boldsymbol{x}_{j,:}|\boldsymbol{s}_{j,:}\right)p\left(\boldsymbol{s}_{j,:}\right). \quad (12)$$

The variable dependencies are shown in a factor graph in Fig. 1. Performing the minimum mean square error (MMSE) or maximum a posteriori (MAP) estimation of Eq. (12) is impractical since it comes down to marginalizing a joint distribution with high dimensions. Therefore, we use an approximate message passing approach that is detailed in Section 3.2.

3.2 Approximate Message Passing Algorithm for Signal Reconstruction

PBiGAMP[22] employs loopy belief propagation (BP) over the factor graph to make approximate inferences of the marginal. According to the sum-product rule and the central limit theorem (CLT), messages passed between the edges of the factor graph possess Gaussian approximations under the large system limit assumption (i.e., $2^B \to \infty$). Message-passing component updates are given in Algorithm 1.

Algorithm 1. MC-PBiGAMP for JADCE under CFO

Input: $\boldsymbol{Y}$, $\boldsymbol{A}$, $\boldsymbol{U}_N$, $T_{\max}$, T_{mc}, τ

Initialize: $\forall n,m$: $\hat{s}_{n,m}(0)=0$, $\forall i,j,m$: choose $\hat{x}_{j,m}(1)$, $\mu^x_{j,m}(1)$, $\hat{c}_i(1)$, $\mu^c_i(1)$

for $t=1,\cdots,T_{\max}$ **do**

$$\forall n,m,i:\ \hat{z}^{(i,*)}_{n,m}(t)=\sum_{j=1}^{2^B} z^{(i,j)}_{n,m}\hat{x}_{j,m}(t)$$

$$\forall n,m,j:\ \hat{z}^{(*,j)}_{n,m}(t)=\sum_{i=1}^{N}\hat{c}_i(t)\,z^{(i,j)}_{n,m}$$

$$\forall n,m:\ \hat{z}^{(*,*)}_{n,m}(t)=\sum_{i=1}^{N}\hat{c}_i(t)\,\hat{z}^{(i,*)}_{n,m}(t)=\sum_{j=1}^{2^B}\hat{z}^{(*,j)}_{n,m}(t)\,\hat{x}_{j,m}(t)$$

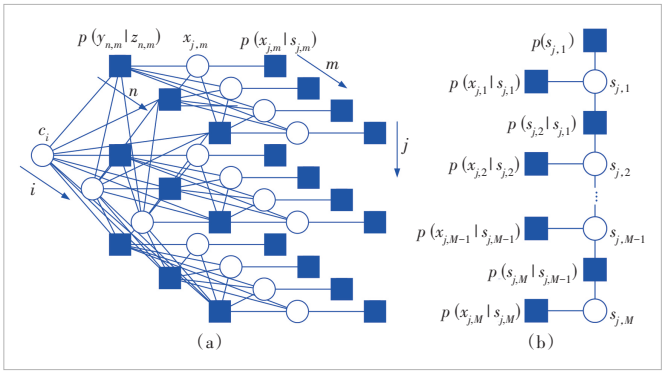
$$\bar{\mu}^p_{n,m}(t)=\sum_{i=1}^{N}\mu^c_i(t)\left|\hat{z}^{(i,*)}_{n,m}|(t)\right|^2+\sum_{j=1}^{2^B}\mu^x_{j,m}(t)\left|\hat{z}^{(*,j)}_{n,m}|(t)\right|^2$$

$$\mu^p_{n,m}(t)=\bar{\mu}^p_{n,m}(t)+\sum_{i=1}^{N}\mu^c_i(t)\sum_{j=1}^{2^B}\mu^x_{j,m}(t)\left|z^{(i,j)}_{n,m}\right|^2$$

$$\hat{p}_{n,m}(t)=\hat{z}^{(*,*)}_{n,m}(t)-\hat{s}_{n,m}(t-1)\bar{\mu}^p_{n,m}(t)$$

$$\mu^z_{n,m}(t)=\mathrm{Var}\left\{z_{n,m}|\boldsymbol{Y};\hat{p}_{n,m}(t),\mu^p_{n,m}(t),\sigma_w\right\}$$

$$\hat{z}_{n,m}(t)=\mathbb{E}\left\{z_{n,m}|\boldsymbol{Y};\hat{p}_{n,m}(t),\mu^p_{n,m}(t),\sigma_w\right\}$$



▲**Figure 1. Factor graph for Bayesian inference**

$$\mu^s_{n,m}(t)=\left(1-\mu^z_{n,m}(t)/\mu^p_{n,m}(t)\right)/\mu^p_{n,m}(t)$$

$$\hat{s}_{n,m}(t)=\left(\hat{z}_{n,m}(t)-\hat{p}_{n,m}(t)\right)/\mu^p_{n,m}(t)$$

$$\forall j,m:\ \mu^r_{j,m}(t)=\left(\sum_{n=1}^{N}\mu^s_{n,m}(t)\left|\hat{z}^{(*,j)}_{n,m}(t)\right|^2\right)^{-1}-$$

$$\mu^r_{j,m}(t)\hat{x}_{j,m}(t)\sum_{n=1}^{N}\mu^s_{n,m}(t)\sum_{i=1}^{N}\mu^c_i(t)\left|z^{(i,j)}_{n,m}\right|^2$$

$$\forall i:\ \mu^q_i(t)=\left(\sum_{n=1}^{N}\mu^s_{n,m}(t)\left|\hat{z}^{(i,*)}_{n,m}|(t)\right|^2\right)^{-1}-$$

$$\mu^q_i(t)\hat{c}_i(t)\sum_{n=1}^{N}\sum_{m=1}^{M}\mu^s_{n,m}(t)\sum_{j=1}^{2^B}\mu^c_i(t)\left|z^{(i,j)}_{n,m}\right|^2$$

$$\forall j,m:\ \mu^x_{j,m}(t+1)=\mathrm{Var}\left\{x_{j,m}|\boldsymbol{Y};\hat{r}_{j,m}(t),\mu^r_{j,m}(t),\tilde{\rho}_{j,m}(t),\sigma_x\right\}$$

$$\hat{x}_{j,m}(t+1)=\mathbb{E}\left\{x_{j,m}|\boldsymbol{Y};\hat{r}_{j,m}(t),\mu^r_{j,m}(t),\tilde{\rho}_{j,m}(t),\sigma_x\right\}$$

$$\forall i:\ \mu^c_i(t+1)=\mathrm{Var}\left\{c_i|\boldsymbol{Y};\hat{q}_i(t),\mu^q_i(t),\lambda_c,\sigma_c\right\}$$

$$\hat{c}_i(t+1)=\mathbb{E}\left\{c_i|\boldsymbol{Y};\hat{q}_i(t),\mu^q_i(t),\lambda_c,\sigma_c\right\}$$

if $\left\|\hat{\boldsymbol{Z}}^{(*,*)}(t)-\hat{\boldsymbol{Z}}^{(*,*)}(t-1)\right\|_F^2\leqslant\tau\left\|\hat{\boldsymbol{Z}}^{(*,*)}(t)\right\|_F^2$, **stop**

end for

Output: $\hat{\boldsymbol{c}}=\hat{\boldsymbol{c}}(t)$, $\hat{\boldsymbol{X}}=\hat{\boldsymbol{X}}(t)$

The reader can refer to Ref. [22] for detailed derivations. Briefly, the calculation of messages from all variable nodes $\{c_i\}_{\forall i}$ and $\{x_{j,m}\}_{\forall j,m}$ to factor node $p(y_{n,m}|z_{n,m})$ takes on the form of a complex Gaussian distribution $\mathcal{CN}(z_{n,m};\hat{p}_{n,m},\mu^p_{n,m})$. With $p(y_{n,m}|z_{n,m})=\exp(-\|y_{n,m}-z_{n,m}\|^2/\sigma_w^2)/\pi\sigma^2$ under AWGN, the marginal posterior $p(z_{n,m}|\boldsymbol{Y})$ is inferred as

$$p\left(z_{n,m}|\boldsymbol{Y};\hat{p}_{n,m},\mu^p_{n,m},\sigma_w\right)=\mathcal{CN}\left(z_{n,m};\hat{z}_{n,m},\mu^z_{n,m}\right), \quad (13)$$

with

$$\hat{z}_{n,m}=\frac{\mu^p_{n,m}y_{n,m}+\sigma_w^2\hat{p}_{n,m}}{\mu^p_{n,m}+\sigma_w^2}, \quad (14)$$

$$\mu^z_{n,m}=\frac{\mu^p_{n,m}\sigma_w^2}{\mu^p_{n,m}+\sigma_w^2}, \quad (15)$$

where $\hat{s}_{n,m}$ and $\mu^s_{n,m}$ are the scaled residual and the residual variance, respectively. Again, the message from factor node $p(y_{n,m}|z_{n,m})$ to variable node c_i is approximately Gaussian ($\mathcal{CN}(c_i;\hat{q}_i,\mu^q_i)$). We reach the estimation of the posterior mean and variance of $p(c_i|\boldsymbol{Y})$ as

$$\hat{c}_i=\mathrm{E}\left\{c_i|\boldsymbol{Y};\hat{q}_i,\mu^q_i,\lambda_c,\sigma_c\right\}=\lambda_c\frac{\hat{q}_i\sigma_c^2}{\mu^q_i+\sigma_c^2}, \quad (16)$$

$$\mu^c_i=\mathrm{Var}\left\{c_i|\boldsymbol{Y};\hat{q}_i,\mu^q_i,\lambda_c,\sigma_c\right\}=\lambda_c\frac{\mu^q_i\sigma_c^2}{\mu^q_i+\sigma_c^2}+\lambda_c(1-\lambda_c)\left|\hat{c}_i\right|^2. \quad (17)$$

As for messages passed within the MC model, the input, i.e., the message from variable node $x_{j,m}$ to factor node $p(x_{j,m}|s_{j,m})$ is a complex Gaussian distribution with mean $\hat{r}_{j,m}$ and variance $\mu^r_{j,m}$. Then, we have the message from factor node $p(x_{j,m}|s_{j,m})$ to variable node $s_{j,m}$ computed as

$$\vec{v}_{j,m} = \vec{\rho}_{j,m}\delta\left(s_{j,m} - 1\right) + \left(1 - \vec{\rho}_{j,m}\right)\delta\left(s_{j,m} + 1\right), \tag{18}$$

where

$$\vec{\rho}_{j,m} = \left(1 + \frac{\mathcal{CN}\left(0; \hat{r}_{j,m}, \mu^r_{j,m}\right)}{\int_{x_{j,m}} \mathcal{CN}\left(x_{j,m}; 0, \sigma_x^2\right)\mathcal{CN}\left(x_{j,m}; \hat{r}_{j,m}, \mu^r_{j,m}\right)}\right)^{-1}. \tag{19}$$

The forward message passing over MC $\boldsymbol{s}_{j,:}$ is conducted in a recursive way as follows[21]

$$\vec{\varphi}_{j,1} = \lambda_x = \frac{p_{01}}{p_{01} + p_{10}}, \tag{20}$$

$$\vec{\varphi}_{j,m} = \frac{p_{01}\left(1 - \vec{\rho}_{j,m-1}\right)\left(1 - \vec{\varphi}_{j,m-1}\right) + p_{11}\vec{\rho}_{j,m-1}\vec{\varphi}_{j,m-1}}{\left(1 - \vec{\rho}_{j,m-1}\right)\left(1 - \vec{\varphi}_{j,m-1}\right) + \vec{\rho}_{j,m-1}\vec{\varphi}_{j,m-1}}. \tag{21}$$

The backward message passing is performed in a similar way[21]:

$$\overleftarrow{\varphi}_{j,M} = \frac{1}{2}, \tag{22}$$

$$\overleftarrow{\varphi}_{j,m} = \frac{p_{10}\left(1 - \overleftarrow{\rho}_{j,m+1}\right)\left(1 - \overleftarrow{\varphi}_{j,m+1}\right) + \left(1 - p_{10}\right)\overleftarrow{\rho}_{j,m+1}\overleftarrow{\varphi}_{j,m+1}}{\left(p_{00} + p_{10}\right)\left(1 - \overleftarrow{\rho}_{j,m+1}\right)\left(1 - \overleftarrow{\varphi}_{j,m+1}\right) + \left(p_{11} + p_{01}\right)\overleftarrow{\rho}_{j,m+1}\overleftarrow{\varphi}_{j,m+1}}. \tag{23}$$

Subsequently, the message from variable node $s_{j,m}$ to factor node $p(x_{j,m}|s_{j,m})$ is represented as

$$\overleftarrow{v}_{j,m} = \overleftarrow{\rho}_{j,m}\delta\left(s_{j,m} - 1\right) + \left(1 - \overleftarrow{\rho}_{j,m}\right)\delta\left(s_{j,m} + 1\right), \tag{24}$$

where

$$\overleftarrow{\rho}_{j,m} = \frac{\vec{\varphi}_{j,m}\overleftarrow{\varphi}_{j,m}}{\left(1 - \vec{\varphi}_{j,m}\right)\left(1 - \overleftarrow{\varphi}_{j,m}\right) + \vec{\varphi}_{j,m}\overleftarrow{\varphi}_{j,m}}. \tag{25}$$

After that, the output of the MC support estimation module, i.e., the message from factor node $p(x_{j,m}|s_{j,m})$ to variable node $x_{j,m}$ is expressed as a Bernoulli-Gaussian distribution $\overleftarrow{\rho}_{j,m}\mathcal{CN}(x_{j,m}; 0, \sigma_x^2) + (1 - \overleftarrow{\rho}_{j,m})\delta(x_{j,m})$. Finally, we calculate the posterior mean and variance of $p(x_{j,m}|\boldsymbol{Y})$ as[13]

$$\hat{x}_{j,m} = \mathbb{E}\left\{x_{j,m}|\boldsymbol{Y}; \hat{r}_{j,m}, \mu^r_{j,m}, \vec{\rho}_{j,m}, \sigma_x\right\} = \vec{\rho}_{j,m}\frac{\hat{r}_{j,m}\sigma_x^2}{\mu^r_{j,m} + \sigma_x^2}, \tag{26}$$

$$\mu^x_{j,m} = \vec{\rho}_{j,m}\frac{\mu^r_{j,m}\sigma_x^2}{\mu^r_{j,m} + \sigma_x^2} + \vec{\rho}_{j,m}\left(1 - \vec{\rho}_{j,m}\right)\left|\hat{x}_{j,m}\right|^2. \tag{27}$$

The above message components are updated iteratively until a certain stopping criterion is satisfied. The worst-case complexity order of the proposed algorithm per iteration is $\mathcal{O}(2^B N^2 M)$. With $\hat{\boldsymbol{X}}$ (the estimation of $\boldsymbol{X}$), the active codewords are determined as

$$X = \left\{j: \left\|\hat{\boldsymbol{x}}_j\right\|^2 > \phi,\ j \in \left[2^B\right]\right\}, \tag{28}$$

where $\phi > 0$ is the threshold.

4 Uncoupled URA Transmission Scheme for Complexity Reduction

Since the size of common codebook A grows exponentially in B, it is computationally infeasible for any CS algorithm to perform AD among the codebook accommodating even small-sized messages (e.g., B = 100 bit). Many URA works take the divide-and-conquer strategy and utilize a concatenated coding scheme termed CCS. Specifically, each long message bit sequence is split into several fragments of amenable length for slotted transmission. These fragments are coupled by appending parity check bits generated by pseudo-random linear combinations of message bits from previous fragments. The decoder first determines transmitted fragments of each slot and then combines slot-wise fragments into the entire message based on a tree decoding process. With the help of the high spatial resolution provided by massive MIMO, it is suggested in Refs. [13] and [14] that the massive MIMO channels in the angular domain already offer adequate information to combine fragments scattered among different transmission slots. The sparsity and magnitude of each entry of angular domain channel vectors indicate angular propagation patterns unique to each active user. Assuming these channel statistics to be almost unchanged within the URA uplink transmission period, the message stitching process can be conducted in the form of clustering recovered channels into groups of each active user.

Following the uncoupled URA scheme in Refs. [13] and [14], we divide each B-bit message into S fragments of length $J = B/S$, which are encoded for transmission in the coherent block of length $N = N_S S$. The designed low-complexity URA receiver under CFO in this paper reconstructs the emitted message list by taking the following two steps: 1) after each transmission slot, performing JADCE to the received signal via the proposed MC-PBiGAMP algorithm with downsized common coding matrix $\tilde{\boldsymbol{A}} \in \mathbb{C}^{N_S \times 2^J}$; 2) reconstructing the en-

tire message sequence by combining slot-wise codewords according to the similarity of their corresponding channels. In the rest of the section, we discuss the slot-balanced K-means algorithm designed for message stitching.

4.1 Slot-Balanced K-Means for Clustering-Based Message Stitching

We consider an ideal case where no users transmit identical data in the same transmission slot. Therefore, K_a codewords are exactly judged to be active in every slot. The purpose of the decoder is to classify active codewords into K_a groups based on the similarity of their corresponding channel vectors. During the clustering process, two obvious constraints must be satisfied: 1) Channels from the same slot cannot be assigned to the same group; 2) each group must be composed of S channels at the end of the clustering.

Slot-balanced K-means algorithm[14] is tailored for the application scenario, performing assignment steps on a per-slot basis and obtaining groups with identical numbers of components. To meet Constraint 2 in K-means, the assignment step is performed slot by slot, i.e., all K_a active channels recognized in the same slot are simultaneously allocated to K_a groups. Then, the assignment problem with Constraint 1 is equivalent to the minimum bipartite matching problem, which can be well solved by the Hungarian algorithm[23]. We define $\boldsymbol{g}_k^s \in \mathbb{C}^M$ the k-th active channel vector at the s-th slot, $\boldsymbol{c}_{k'} \in \mathbb{C}^M$ the k'-th group center, and $\boldsymbol{C} = \left[\boldsymbol{c}_{1'}, \cdots, \boldsymbol{c}_{K_a}\right]$. The input $\boldsymbol{D} \in \mathbb{C}^{K_a \times K_a}$ of the Hungarian algorithm is calculated as each element $d_{k,k'} = \left\| \boldsymbol{g}_k - \boldsymbol{c}_{k'} \right\|_2$. The output of the algorithm is a binary matrix $\boldsymbol{\Gamma} \in \{0,1\}^{K_a \times K_a}$ with the (k,k')-th element $\gamma_{k,k'} = 1$ suggesting that the k-th channel belongs to the k'-th group. There is exactly one nonzero element within each row and each column of $\boldsymbol{\Gamma}$. Then, with the assignment result, each group center is updated as the mean of its constituent channel vectors. The algorithm iteration stops when the maximum number of iterations is reached or there are no further changes in center locations. The algorithm complexity is dominated by the Hungarian algorithm $\left(\mathcal{O}(K_a^3)\right)$ and yields the order of $\mathcal{O}\left(SK_a^3\right)$.

4.2 Codeword Collision Resolution

Codeword collision happens when more than one user chooses to send the same codeword at the same time. The corresponding channel to the reused codeword is the sum of all competitive user channels. The key idea to resolve codeword collision is to identify contaminated channels: they usually have a longer sum distance from all group center points. With a carefully designed data split profile, the probability that more than two users choose to send the same codeword tends to be zero. Therefore, if $K_s < K_a$ channels are determined active, we recognize $K_a - K_s$ channels having the largest aggregated distance to all center points as contaminated channels, and the corresponding distance row vectors are duplicately added to the original distance matrix to complete a square matrix as algorithm input. However, after data partitioning, the center points of groups involving contaminated channels cannot be updated directly. Luckily, the original center points containing representative user channel statistics can be used to eliminate interfering channels. Specifically, a unique angular domain channel support pattern is extracted from the corresponding center by selecting entries $\{c_m, m \in \mathcal{M}\}$ that concentrate most (e.g., 95%) of the energy, i.e., the elements in $\mathcal{M}$ are chosen as $\sum_{\mathcal{M}} |c_m|^2 > 0.95 \|\boldsymbol{c}\|_2^2$. Then, only elements within the set $\{g_{k,m}, m \in \mathcal{M}\}$ are taken for center location update. The complete algorithmic iteration steps are summarized in Algorithm 2.

Algorithm 2. Slot-balanced K-means for message stitching

Input: $\left\{\boldsymbol{g}_k^s : k \in \left[\left|\mathcal{X}_s\right|\right], s \in [S]\right\}, T_c$

Initialize: $\boldsymbol{C}(0,S) = \left[\boldsymbol{c}_1(0,S), \cdots, \boldsymbol{c}_{K_a}(0,S)\right]$

for $t = 1, \cdots, T_c$ **do**

Set $\boldsymbol{C}(t,0) = \boldsymbol{C}(t-1,S)$

for $s = 1, \cdots, S$ **do**

Compute $\boldsymbol{D}$ with elements $d_{k,k'} = \left\| \boldsymbol{g}_k^s - \boldsymbol{c}_{k'} \right\|$

if $K_s < K_a$ **then**

Add $K_a - K_s$ rows with the largest sum of elements

end if

Execute Hungarian algorithm with input $\boldsymbol{D}$ and output $\boldsymbol{\Gamma}$

if $K_s < K_a$ **then**

$$\forall k' : \boldsymbol{c}_{k'}(t,s) = \frac{1}{s}\left[(s-1)\boldsymbol{c}_{k'}(t,s-1) + \sum_{k=1}^{K_a} \gamma_{k,k'} \boldsymbol{g}_k^s\right]$$

else

$$\forall k' : \boldsymbol{c}_{k'}(t,s) = \frac{1}{s}\left[(s-1)\boldsymbol{c}_{k'}(t,s-1) + \sum_{k=1}^{K_a} \gamma_{k,k'} \mathrm{diag}\left(v_1, \cdots, v_M\right) \boldsymbol{g}_k^s\right],$$

where $v_m \in \{0.1\}$ is nonzero when $m \in \mathcal{M}(t,s-1)$

end if

end for

if $\boldsymbol{C}(t,S) = \boldsymbol{C}(t-1,S)$, **stop**

end for

Output: Partitioning of the data set

5 Simulation Results

In simulations, we consider a URA system with $K_a = 50$ active users randomly and uniformly located in a cell with the radius of 1 km. User channels are generated based on Eq. (1), where L=4. Each active user sends a 100 bit message which is divided into $S = 10$ fragments of length $J = 10$. Accordingly, the common codebook $\tilde{\boldsymbol{A}}$ for data transmission is an i.i.d. Gaussian matrix of size 100×2^{10}. The CFO parameter ω is chosen uniformly from the range $(-0.0133, 0.0133)$[16].

We first examine the proposed MC-PBiGAMP for JADCE under CFO. As the stopping criteria for the iterative algorithm,

we set $T_{\max} = 100$, $T_{\mathrm{mc}} = 20$, and the precision tolerance $\tau = 10^{-4}$. Since we focus on whether active codewords are properly determined, the misdetection rate is used as the AD performance metric, i.e.,

$$P_{\mathrm{m}} = \frac{|\mathcal{A} \setminus \mathcal{X}|}{|\mathcal{A}|}, \tag{29}$$

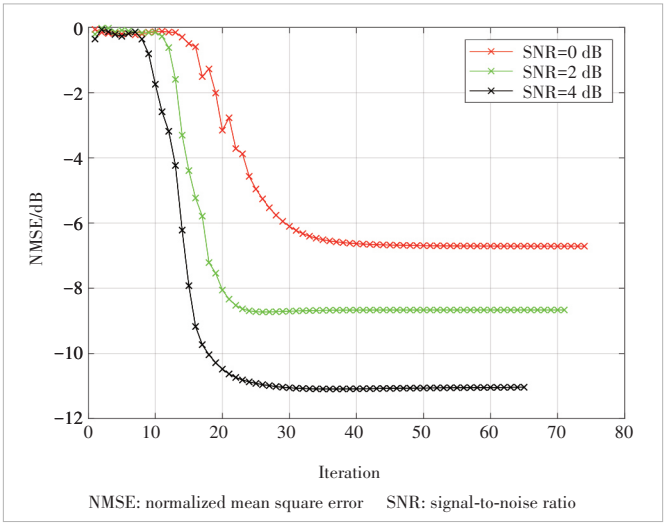
where $\mathcal{A}$ is the set of indexes of active codewords. The algorithm in the aspect of CE is assessed by the normalized mean square error (NMSE) of the recovered active channels, i.e., $\mathrm{NMSE} = \left\| \bar{\boldsymbol{X}} - \boldsymbol{X}_a \right\|_F^2 \Big/ \left\| \bar{\boldsymbol{X}} \right\|_F^2$, where $\bar{\boldsymbol{X}}$ is the original channel matrix arranged according to the indexes in Eq. (28), and $\boldsymbol{X}_a$ is the estimated active channel matrix. The NMSE performance of MC-PBiGAMP in each iteration is exhibited in Fig. 2. We observe that the proposed algorithm converges around 65–74 iteration rounds under different signal-to-noise ratios (SNRs).

For comparison, we appeal to the method in Ref. [16]. To deal with the influence of CFO, the discrete CFO parameters $[\omega_1, \cdots \omega_R]$ with $R = 9$ are uniformly sampled from the possible range. JADCE is then performed using the MMV-GAMP algorithm[14] with the measurement matrix constructed by expanding the original one with these sampled frequency offsets. We also investigate the PBiGAMP algorithm[22] with no support structure for JADCE. The JADCE performance of the aforementioned methods versus the SNR or the number of measurements N_s is depicted in Fig. 3. The proposed MC-PBiGAMP algorithm outperforms the original PBiGAMP algorithm as we take into account the correlation between adjacent angular domain channel elements by the MC support model. Although MC-PBiGAMP with respect to the AD performance is about the same as MMV-GAMP in Ref. [16], the advantage lies in that it only needs to handle a CFO estimation problem of scale $2^J \times N_s \times M$ compared with the latter of scale $D \times 2^J \times N_s \times M$. It is indicated in Fig. 3(c) that the misdetection rates of both algorithms decrease significantly when $N_s > K_a$. This coincides with the scaling law of AMP that requires measurements to reliably identify a subset of K_a active codewords among a set of size 2^J scales as $N_s = \mathcal{O}\left(K_a log \frac{2^J}{K_a}\right)$, i.e., N_s is almost linearly with K_a.

The error rate of URA transmission is defined as the average per user probability of error (PUPE)[8], i.e.,

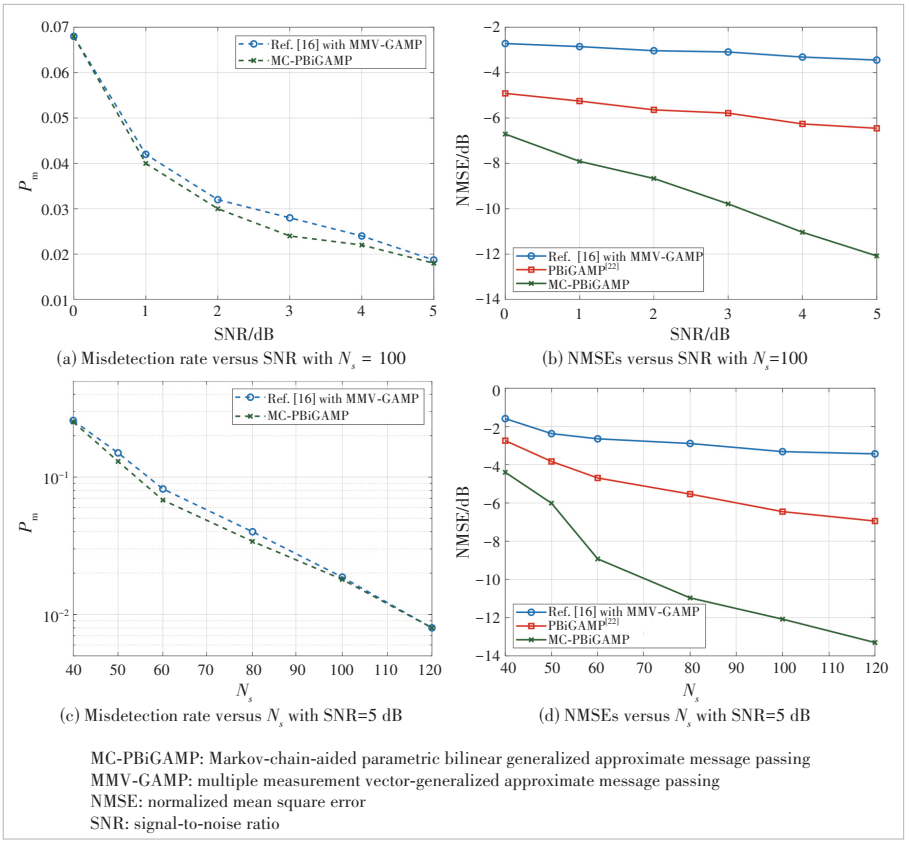
$$\mathrm{PUPE} = \mathbb{E}\left\{\frac{\left|\mathcal{L} \setminus \{m(k), k \in \mathcal{K}_{\mathrm{a}}\}\right|}{|\mathcal{L}|}\right\}, \tag{30}$$

where $m(k)$ is the message of the k-th active user in the set $\mathcal{K}_{\mathrm{a}}$, and $\mathcal{L}$ is the recovered message list. Fig. 4 presents the error rates of uncoupled compressed sensing (UCS) schemes le-



NMSE: normalized mean square error SNR: signal-to-noise ratio

▲ **Figure 2. NMSEs versus the iteration number under different SNRs with M=32, N_s=100, and K_a=50**



(a) Misdetection rate versus SNR with N_s = 100
(b) NMSEs versus SNR with N_s=100
(c) Misdetection rate versus N_s with SNR=5 dB
(d) NMSEs versus N_s with SNR=5 dB

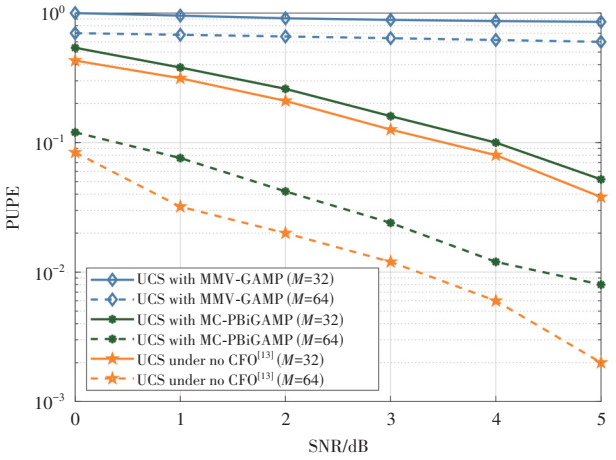
MC-PBiGAMP: Markov-chain-aided parametric bilinear generalized approximate message passing
MMV-GAMP: multiple measurement vector-generalized approximate message passing
NMSE: normalized mean square error
SNR: signal-to-noise ratio

▲ **Figure 3. Joint AD and CE (JADCE) performance of various algorithms with M=32, K_a=50**

veraging different algorithms for JADCE under CFO. As revealed in Fig. 4, the proposed UCS scheme with PBiGAMP reaches the best system performance with spectral efficiency $\Psi = BK_a/SN_s = 5$ bit/s per channel use. Fig. 4 also indicates that PUPE improves significantly when the number of receiving antennas is increased. It is due to the higher resolution offered by massive antennas, which not only makes the CS paradigm sparser but also provides more dimensional information for measuring channel similarity. We also draw in Fig. 4 the error rate of URA under no CFO using the method of Ref. [13] as the baseline. It is shown that the proposed approach under CFO pays 0.4 dB (for $M = 32$) to 1.2 dB (for $M = 64$) in terms of SNR to achieve the target PUPE = 0.05 compared with the benchmark.

6 Conclusions

This paper investigates MIMO URA countering the influence of CFO. We formulate URA under CFO as a CS recovery problem with a bilinear graphic structure. The MC-PBiGAMP algorithm is first employed for JADCE with an MC support structure to model the clustered sparsity of the considered angular domain channel. Then, an uncoupled transmission protocol is adopted to reduce the computational burden, where messages are split into several fragments for slotted transmission and stitched together upon clustering user channels. We show by simulations that the proposed scheme is capable of conducting reliable URA transmission under CFO.



CFO: carrier frequency offset
MC-PBiGAMP: Markov-chain-aided parametric bilinear generalized approximate message passing
MMV-GAMP: multiple measurement vector-generalized approximate message passing
PUPE: per user probability of error
SNR: signal-to-noise ratio
UCS: uncoupled compressed sensing

▲ **Figure 4. Error probabilities of various unsourced random access (URA) schemes under CFO as a function of SNR with N_s=100, K_a=50**

References

[1] DAWY Z, SAAD W, GHOSH A, et al. Toward massive machine type cellular communications [J]. IEEE wireless communications, 2017, 24(1): 120 – 128. DOI: 10.1109/MWC.2016.1500284WC
[2] WU Y P, GAO X Q, ZHOU S D, et al. Massive access for future wireless communication systems [J]. IEEE wireless communications, 2020, 27(4): 148 – 156. DOI: 10.1109/MWC.001.1900494
[3] SENEL K, LARSSON E G. Grant-free massive MTC-enabled massive MIMO: a compressive sensing approach [J]. IEEE transactions on communications, 2018, 66(12): 6164 – 6175. DOI: 10.1109/TCOMM.2018.2866559
[4] LIU L, YU W. Massive connectivity with massive MIMO—part I: device activity detection and channel estimation [J]. IEEE transactions on signal processing, 2018, 66(11): 2933 – 2946. DOI: 10.1109/TSP.2018.2818082
[5] KE M L, GAO Z, WU Y P, et al. Compressive sensing-based adaptive active user detection and channel estimation: massive access meets massive MIMO [J]. IEEE transactions on signal processing, 2020, 68: 764 – 779. DOI: 10.1109/TSP.2020.2967175
[6] KE M L, GAO Z, WU Y P, et al. Massive access in cell-free massive MIMO-based Internet of Things: Cloud computing and edge computing paradigms [J]. IEEE journal on selected areas in communications, 2021, 39(3): 756 – 772. DOI: 10.1109/JSAC.2020.3018807
[7] WEI F, CHEN W, WU Y P, et al. Message-passing receiver design for joint channel estimation and data decoding in uplink grant-free SCMA systems [J]. IEEE transactions on wireless communications, 2018, 18(1): 167 – 181. DOI: 10.1109/TWC.2018.2878571
[8] POLYANSKIY Y. A perspective on massive random-access [C]//International Symposium on Information Theory (ISIT). IEEE, 2017: 2523 – 2527. DOI: 10.1109/ISIT.2017.8006984
[9] AMALLADINNE V K, CHAMBERLAND J F, NARAYANAN K R. A coded compressed sensing scheme for unsourced multiple access [J]. IEEE transactions on information theory, 2020, 66(10): 6509 – 6533. DOI: 10.1109/TIT.2020.3012948
[10] FENGLER A, HAGHIGHATSHOAR S, JUNG P, et al. Non-Bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive MIMO receiver [J]. IEEE transactions on information theory, 2021, 67(5): 2925 – 2951. DOI: 10.1109/TIT.2021.3065291
[11] XIE X Y, WU Y P, GAO J Y, et al. Massive unsourced random access for massive MIMO correlated channels [C]//IEEE Global Communications Conference. IEEE, 2021: 1 – 6. DOI: 10.1109/GLOBECOM42002.2020.9347959
[12] SHYIANOV V, BELLILI F, MEZGHANI A, et al. Massive unsourced random access based on uncoupled compressive sensing: another blessing of massive MIMO [J]. IEEE journal on selected areas in communications, 2021, 39(3): 820 – 834. DOI: 10.1109/JSAC.2020.3019722
[13] XIE X Y, WU Y P. Unsourced random access with a massive MIMO receiver: Exploiting angular domain sparsity [C]//IEEE/CIC International Conference on Communications in China (ICCC). IEEE, 2021: 741 – 745. DOI: 10.1109/ICCC52777.2021.9580441
[14] XIE X Y, WU Y P, AN J P, et al. Massive unsourced random access: exploiting angular domain sparsity [J]. IEEE transactions on communications, 2022, 70(4): 2480 – 2498. DOI: 10.1109/TCOMM.2022.3153957
[15] LI Y, XIA M H, WU Y C. Activity detection for massive connectivity under frequency offsets via first-order algorithms [J]. IEEE transactions on wireless communications, 2019, 18(3): 1988 – 2002. DOI: 10.1109/TWC.2019.2901482
[16] SUN G L, LI Y N, YI X P, et al. Massive grant-free OFDMA with timing and frequency offsets [J]. IEEE transactions on wireless communications, 2022, 21(5): 3365 – 3380. DOI: 10.1109/TWC.2021.3121066
[17] BELLILI F, SOHRABI F, YU W. Generalized approximate message passing for massive MIMO mmWave channel estimation with Laplacian prior [J]. IEEE transactions on communications, 2019, 67(5): 3205 – 3219. DOI: 10.1109/TCOMM.2019.2892719
[18] FENGLER A, JUNG P, CAIRE G. SPARCs for unsourced random access [J]. IEEE transactions on information theory, 2021, 67(10): 6894 – 6915. DOI: 10.1109/TIT.2021.3081189
[19] DENG W, SIRIBURANON T, MUSA A, et al. A sub-harmonic injection-locked quadrature frequency synthesizer with frequency calibration scheme for millimeter-wave TDD transceivers [J]. IEEE journal of solid-state circuits, 2013, 48(7): 1710 – 1720. DOI: 10.1109/JSSC.2013.2253396

[20] MYERS N J, HEATH R W. Message passing-based joint CFO and channel estimation in mmWave systems with one-bit ADCs [J]. IEEE transactions on wireless communications, 2019, 18(6): 3064 – 3077. DOI: 10.1109/TWC.2019.2909865
[21] CHEN L, LIU A, YUAN X J. Structured turbo compressed sensing for massive MIMO channel estimation using a Markov prior [J]. IEEE transactions on vehicular technology, 2018, 67(5): 4635 – 4639. DOI: 10.1109/TVT.2017.2787708
[22] PARKER J T, SCHNITER P. Parametric bilinear generalized approximate message passing [J]. IEEE journal of selected topics in signal processing, 2016, 10(4): 795 – 808. DOI: 10.1109/JSTSP.2016.2539123
[23] KUHN H W. The Hungarian method for the assignment problem [J]. Naval research logistics, 2005, 52(1): 7 – 21. DOI: 10.1002/nav.20053

Biographies

XIE Xinyu received his BS degree in telecommunication engineering from Xidian University, China in 2019. He is currently working toward a PhD degree in electronic engineering at Shanghai Jiao Tong University, China. His research interests include massive random access and compressed sensing.

WU Yongpeng (yongpeng.wu@sjtu.edu.cn) received his BS degree in telecommunication engineering from Wuhan University, China in July 2007, and PhD degree in communication and signal processing from the National Mobile Communications Research Laboratory, Southeast University, China in 2013. He is currently a tenure-track associate professor with the Department of Electronic Engineering, Shanghai Jiao Tong University, China. His research interests include massive MIMO/MIMO systems, massive machine type communication, physical layer security, and signal processing for wireless communications.

YUAN Zhifeng received his MS degree in signal and information processing from Nanjing University of Post and Telecommunications, China in 2005. From 2004 to 2006, he was mainly engaged in FPGA/SOC ASIC design. He has been a member of the Wireless Technology Advance Research Department, ZTE Corporation since 2006 and has been responsible for the research of the new multiple access group since 2012. His research interests include wireless communication, MIMO systems, information theory, multiple access, error control coding, adaptive algorithm, and high-speed VLSI design.

MA Yihua received his BE degree from Southeast University, China in 2015 and MS degree from Peking University, China in 2018. He is now a pre-research expert engineer in the Department of Wireless Algorithm, ZTE Corporation. His main research interests include mMTC, grant-free transmissions, NOMA, and joint communication and sensing. In these areas, he has published and applied more than 10 papers and 20 patents.

Learning-Based Admission Control for Low-Earth-Orbit Satellite Communication Networks

CHENG Lei, QIN Shuang, FENG Gang

(University of Electronic Science and Technology of China, Chengdu 611731, China)

DOI: 10.12142/ZTECOM.202303008

https://link.cnki.net/urlid/34.1294.TN.20230830.1858.005, published online August 31, 2023

Manuscript received: 2022-07-22

Abstract: Satellite communications has been regarded as an indispensable technology for future mobile networks to provide extremely high data rates, ultra-reliability, and ubiquitous coverage. However, the high dynamics caused by the fast movement of low-earth-orbit (LEO) satellites bring huge challenges in designing and optimizing satellite communication systems. Especially, admission control, deciding which users with diversified service requirements are allowed to access the network with limited resources, is of paramount importance to improve network resource utilization and meet the service quality requirements of users. In this paper, we propose a dynamic channel reservation strategy based on the Actor-Critic algorithm (AC-DCRS) to perform intelligent admission control in satellite networks. By carefully designing the long-term reward function and dynamically adjusting the reserved channel threshold, AC-DCRS reaches a long-run optimal access policy for both new calls and handover calls with different service priorities. Numerical results show that our proposed AC-DCRS outperforms traditional channel reservation strategies in terms of overall access failure probability, the average call success rate, and channel utilization under various dynamic traffic conditions.

Keywords: satellite communications; admission control; dynamic channel reservation; actor-critic

Citation (Format 1): CHENG L, QIN S, FENG G. Learning-based admission control for low-earth-orbit satellite communication networks [J]. ZTE Communications, 2023, 21(3): 54 - 62. DOI: 10.12142/ZTECOM.202303008

Citation (Format 2): L. Cheng, S. Qin, and G. Feng, "Learning-based admission control for low-earth-orbit satellite communication networks," *ZTE Communications*, vol. 21, no. 3, pp. 54 - 62, Sept. 2023. doi: 10.12142/ZTECOM.202303008.

1 Introduction

With the increasing number of users and service types in terrestrial wireless communication networks, it is impractical to provide wireless communication services anytime and anywhere alone[1]. The satellite communication network has the prominent advantages of long-distance communications, ubiquitous coverage, large capacity and high reliability, and can be a complement to terrestrial networks. It is not restricted by complex geographic conditions and harsh environments and can provide broadband multimedia services to user terminals (UT) in any area, even where terrestrial network resources are insufficient[2].

Due to the advantages of shorter propagation delay and lower operational expenditure, low-earth-orbit (LEO) satellite communication systems have been commonly used to provide user terminals with full coverage and real-time wireless communication services[3]. Usually, LEO communication systems can exploit multi-beam technology to irradiate lots of blocks of cellular networks in their coverage area, which are called beam cells. When a UT establishes a communication connection with an LEO satellite, one challenge faced is the frequent handover from one beam cell to another, due to the fast movement of LEO satellites. If there are insufficient channel resources in the targeted beam cell, the connection would be interrupted. Frequent handover failure and new call blocking would severely degrade the network performance and/or quality of service (QoS) of users. Moreover, with the rapid development of multimedia applications, diversified service requirements pose a great challenge to the network[4 - 6]. On the other hand, the satellite channel resources are limited, which usually cannot satisfy the requirements of all services. Considering the diversified service requirements with multi-priority services, admission control is particularly critical, as it decides which services are allowed to be admitted.

A typical solution for admission control of multi-priority ser-

This work was supported by the ZTE Industry-University-Institute Cooperation Funds.

vices in satellite systems is to allocate channel resources of beam cells by a priority-based channel reservation strategy, which has been intensively investigated in satellite communication systems[4 – 10]. The basic idea is to reserve a certain number of channels for handover calls and new calls with different service priorities, to guarantee the priority of handover calls and delay-sensitive services to ensure the continuity of calls for moving UTs.

Existing channel reservation strategies are mainly classified into two categories, fixed channel reservation (FCR) and dynamic channel reservation (DCR). A guaranteed handover FCR strategy was proposed in Ref. [7], which reserves a portion of channel resources dedicated to handover calls. Some improvements have been made later, such as the channel status-based reservation strategy (CSRS) [8] and time-based channel reservation algorithm (TCRA)[9], which set the number of reserved channels based on the information including the status of the cell and/or the remaining time. However, fixed reserved channels cannot adapt to the dynamic environment and multi-service requests, causing a high blocking rate for new calls. In Refs. [10 – 15], the authors proposed adaptive DCR strategies based on different prior information to dynamically change the number of channels reserved. The authors of Ref. [10] proposed to adjust the number of reserved channels, according to the current number of ongoing calls (voice or video traffic) and the localization of users. In Ref. [11], a grey model was used to decide whether the calls need to handover and then dynamically adjust the channel reservation number based on the counter. The authors of Ref. [12] leveraged the number of mobile stations in neighbor locations and the average handover call arrival rate to reserve channels. Ref. [13] considered the varying characteristics of the wireless channel to allocate resources, aiming at maximizing spectral efficiency. The authors of Ref. [15] proposed an adaptive probability-based reservation strategy (APRS) based on mobile users' location information and the handover probability, to improve the utilization of reserved channels in reservation time. Due to the imbalance between the new call blocking rate and handover call failure rate, the system performance is not satisfactory. Some researchers have used heuristic algorithms to adjust the thresholds, and the authors of Ref. [4] proposed a probability-based channel reservation strategy for improving the quality of service. The authors of Ref. [5] proposed a threshold-based DCR scheme to set optimal thresholds for different-priority services by the genetic algorithm.

However, all aforementioned schemes cannot respond quickly to dynamic changes and uneven distribution of service requirements, since they only consider finding the optimum in the current state, while a long-term optimization is needed to improve the system performance. With this regard, some researchers resort to exploiting machine learning algorithms for designing intelligent channel reservation schemes to achieve long-term performance improvement for complex satellite networks. The authors of Ref. [15] proposed a dynamic channel allocation algorithm based on deep reinforcement learning (DRL), which uses convolutional neural networks to extract useful features to make accurate admission decisions. It can effectively reduce the blocking rate and improve system throughput. But this work focused on processing the connection relationship of the UT in the beam and considered only a single service type. The authors of Ref. [6] proposed a multi-service DCR strategy based on the deep Q network to improve the overall service quality of the system, by examining the impact of current channel reservation results on the future environment. They mainly considered how to reserve channels for new calls, while ignoring the impact of handover calls. Unfortunately, all the aforementioned works lack consideration of multi-priority services and frequent handovers in highly dynamic LEO satellite networks. Therefore, it is imperative to develop an intelligent admission control scheme to maximize long-term system performance by performing appropriate channel resource allocation for LEO satellite networks.

In this paper, we propose an intelligent DCR strategy based on the Actor-Critic algorithm (AC-DCRS), which dynamically adjusts the reserved channel thresholds for multi-service calls. While traditional solutions only obtain the optimal solution of the current state in a memoryless system, our proposed AC-DCRS based on reinforcement learning can consistently approach the long-term optimal solution by considering the Markov property of the channel reservation problem. Specifically, the Actor-Critic algorithm is leveraged to deal with continuous state space and high-dimensional action space. Through interactions with the network environment, AC-DRCS can well balance the admission of handover calls and new calls of multiple priorities by setting corresponding thresholds under the current traffic state.

The rest of the paper is structured as follows. Section 2 presents the system model and problem formulation. Section 3 elaborates on the proposed AC-based DCR strategy. We evaluate the performance of the proposed strategy in Section 4 and finally conclude the paper in Section 5.

2 System Model and Problem Formulation

We consider a typical LEO communication network shown in Fig. 1. The coverage area of a single moving LEO satellite consists of multiple adjacent beam cells. A UT in the beam cell establishes a connection to the LEO satellite directly or through the base station in the beam cell. Handover call requests will arrive during the movement of the UT and satellites. At the same time, new call requests may also arrive requiring channel resources from the connected beam cell.

2.1 LEO Mobility Model

Without loss of generality, we consider a one-dimensional square continuous beam cell model, which can be readily extended to high-dimensional models. Since the moving speed of

the UT is much slower than that of the LEO satellites, it can be regarded as relatively static. Hypothetically, a LEO satellite moving horizontally to the left at a speed v_{sat} relative to UT is equivalent to UT moving to the right at the same speed relative to beam cells. We further adopt a cyclic mechanism to simplify the periodicity of satellite movement and the simplified satellite movement model is shown in Fig. 2. When the UT leaves the rightmost cell N as it moves, it will enter the leftmost cell 1. Handover call arrivals, new call arrivals, and call termination may constantly occur during the movement.

2.2 Channel Reservation Model for Multi-Priority Service

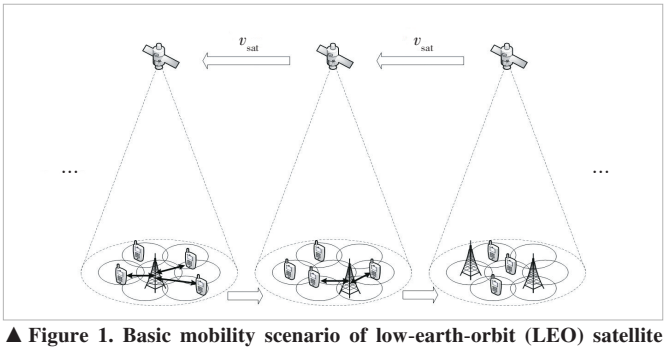
We assume that there are s types of services, and the number of new calls of the i-th service follows a Poisson distribution with an average arrival rate λ_n^i. Then the total arrival rate of the i-th service is given as the sum of the arrival rates of new calls and handover calls as follows:

$$\lambda^i = \lambda_n^i + \lambda_h^i , \tag{1}$$

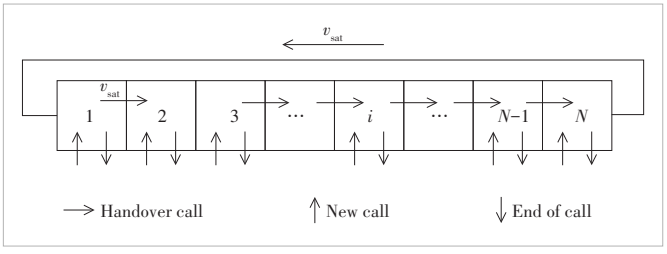
where λ_h^i is the arrival rate of the i-th service for handover calls. Obviously, the handover calls of the i-th service come from the previous beam cell and its arrival rate λ_h^i can be derived as follows[5]:

$$\lambda_h^i = \lambda_n^i \frac{(1 - P_{af}^i)P_{h1}}{1 - (1 - P_{hf}^i)P_{h2}} , \tag{2}$$

where P_{h1}, P_{h2}, P_{af} and P_{hf} are the handover success probability of the source cell, that of the target cell, the new call blocking rate of the i-th service, and the handover call failure rate of the i-th service, respectively.



▲ **Figure 1. Basic mobility scenario of low-earth-orbit (LEO) satellite communication system**



▲ **Figure 2. Simplified mobility model of low-earth-orbit (LEO) satellite communication system**

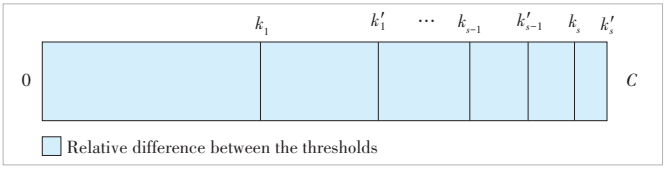
Besides, we denote the proportion of new calls of i-th service as p_i, and then the arrival rate of the i-th service can be expressed as:

$$\lambda_n^i = \lambda_n \times p_i , \tag{3}$$

where λ_n is the average arrival rate of total new calls and p_i satisfies $p_1 + p_2 + p_3 + \cdots + p_s = 1$. We further assume that the duration of all calls obeys an exponential distribution of a parameter u, so the average duration of the call is $1/u$ s.

We adopt a threshold-based channel reservation strategy to realize the admission control of satellite beam cells. We consider that the available bandwidth in a cell is equally allocated to all the channels, and each channel can be assigned to a call. Then, each admitted call will be assigned a channel with enough power to guarantee the quality of service. In this work, we focus on developing an intelligent admission control mechanism for LEO satellite communications. Thus, for simplifying the analysis, we just assume that there is enough power in a cell to guarantee the service quality of each admitted call. For a new call or a handover call, if the number of occupied channels in the current beam cell is less than the corresponding threshold, the call will be admitted successfully and assigned a channel with power and bandwidth resources, and the number of occupied channels is updated. Otherwise, the call will be blocked. After each decision period, the threshold will be updated according to our adjustment algorithm. We set a threshold k_i' for handover calls of the i-th service, while a threshold k_i is set for new calls of the i-th service. As handover calls are prior to new calls[16], we have $k_i \leqslant k_i'$. We also consider that the s types of services have certain priorities, in the way that the type with larger index numbers will be reserved for more channels than those with smaller index numbers. Therefore, the relationship between thresholds of all services satisfies $0 \leqslant k_1 \leqslant k_1' \leqslant \cdots \leqslant k_s \leqslant k_s' \leqslant C$, where C is the total number of beam cell channels. To maximize utilization of the channels, $k_s' = C$ is assumed in our model. Fig. 3 illustrates the proposed multi-priority service threshold-based channel reservation strategy.

The set of all thresholds is denoted by $K = \{k_1, k_1', k_2, k_2', \cdots, k_s, k_s'\}$. Intuitively, dynamically adjusting the thresholds K to control call admission can effectively improve the overall system performance. On the one hand, if the low-priority service threshold is set too low, low-priority service calls will be hard to get admission, even if there are no high-



▲ **Figure 3. Illustration of multi-priority service threshold-based channel reservation strategy**

priority service calls. Some channel resources may be wasted, resulting in low system channel utilization. On the other hand, if the low-priority service threshold is set too high, excessive low-priority service calls may be admitted to occupy too many channels. This will decrease the admission success rate of high-priority services in the future. Therefore, in this paper, we focus on designing a channel reservation strategy to dynamically adjust the thresholds of multi-priority calls, to realize intelligent admission control.

2.3 Problem Formulation

The overall access failure probability at time t, denoted by $O(t)$, as a system performance metric, is defined as:

$$O(t) = \alpha_0 P_{af}(t) + \alpha_1 P_{hf}(t) = \alpha_0 \sum_{i=1}^{s} \beta_i P_{af}^i(t) + \alpha_1 \sum_{i=1}^{s} \beta_i P_{hf}^i(t), \tag{4}$$

where α_0 and α_1 are the balance factors of new calls and handover calls respectively, which are used to measure the different impacts of new call blockage and handover call failure. And β_i is the balance factor of the i-th priority service, which is used to judge the significance of multi-priority services. Meanwhile, we modify $P_{af}^i(t)$ and $P_{hf}^i(t)$ as long-term metrics of the i-th service as follows:

$$P_{af}^i(t) = N_{af}^i / N_a^i, \tag{5}$$

$$P_{hf}^i(t) = N_{hf}^i / N_h^i, \tag{6}$$

where N_{af}^i, N_a^i, N_{hf}^i and N_h^i represent the number of new calls blocked for the i-th service, the total number of new calls for the i-th service, the number of handover calls failed for the i-th service, and the total number of handover call for the i-th service, respectively.

To improve the overall system performance, we minimize $O(t)$, i.e., minimize $P_{af}^i(t)$ and $P_{hf}^i(t)$ in the long term. As mentioned above, the setting of K directly affects the failure rate of new calls and handover calls. When the state space of each beam cell is modeled as a continuous-time M/M/C/C Markov chain, the closed-form relationship of the new call blocking rate, handover call failure rate and K can also be proved[5]. Therefore, we formulate our optimization problem as follows:

$$\begin{aligned} &\max \quad -O(t), \\ &\text{s.t.} \quad 0 < k_1 < k_1' < \cdots < k_s < k_s' \leqslant C \qquad (7.1), \\ &\qquad k_i, k_i' \in Z, \quad i = 1, 2, \cdots, s \qquad (7.2), \end{aligned} \tag{7}$$

where each threshold is limited to an integer for the convenience of adjustment in the dynamic channel reservation strategy. As the number of system channels C and that of services s can be large in real environments, using the brute force method to calculate the optimal thresholds in the current state will cause an exponential increase in time and space complexity. In addition, due to the rapid changes in the environment, optimal thresholds should be derived in real time. Thus, using static optimization to solve Problem (7) is infeasible and thus we resort to a learning-based solution.

3 Intelligent Admission Control Based on Dynamic Channel Reservation Strategy

In this section, we control service call admission by adjusting the reserved channel thresholds and model the problem of dynamically adjusting reserved channel thresholds as a Markov decision process (MDP). First, we slot the time as decision periods with the slot length T_Δ. In each time slot, multiple calls arrive according to the Poisson distribution. The system will adjust the reservation thresholds at the end of each decision period. The maximum number of calls in a decision period is set to N, and even if the decision period is not over, the decision will be made immediately.

3.1 MDP Model

An MDP model consists of a five-tuple $< S, A, P, R, \pi >$, where S, A, P, R and π represent state space, action space, transition probability between states, reward function, and policy for selecting actions based on the state, respectively, which are defined as follows:

1) State(S): we assume that the channel resources of the beam cell remain unchanged. The state is defined as:

$$s(t) \in \{ c\ ;\ \lambda\ ;\ K \} \qquad t = nT_\Delta,\ n = 0,1,2,\cdots, \tag{8}$$

where c is the normalized number of channels that have been occupied in the considered beam cell and satisfies $c \leqslant C$; $\lambda = \{ \lambda_n^1, \lambda_h^1, \cdots, \lambda_n^s, \lambda_h^s \}$ is the set of the call arrival rates of new calls and handover calls that satisfies Eq. (2); $K = \{ k_1, k_1', \cdots, k_s, k_s' \}$ is the set of the normalized reserved channel thresholds of new calls and handover calls that satisfies Eq. (7.1).

2) Action(A): we define the action as current normalized reserved channel thresholds, which can be expressed as:

$$a(t) = K^T = \{ k_1, k_1', \cdots, k_s, k_s' \}^T. \tag{9}$$

In each decision period, action will be taken based on the current state, which will control the admission by setting reserved channel thresholds.

3) Transition Probability(P): generally, the state transition function of the Markov decision process is a certain function $P: S \times A \times S \to [0, 1]$, which represents the probability of the transition to the state s' given state s after taking action a. Since state transition depends on not only the last action but also the traffic changes caused by the movement of satellites and UTs and call termination in the channel, it cannot be explicitly expressed in our problem.

4) Reward(R): for a single service call within a decision pe-

riod, the reward function is defined as:

$$r=\begin{cases}0 & \text{access successfully}\\ -\alpha_0\beta_i \text{ or } -\alpha_1\beta_i & \text{access failed}\\ -L & (7.1)\text{ not meet}\end{cases}. \tag{10}$$

As is defined in Eq. (10), when a new call or a handover call is blocked, a negative value $-\alpha_0\beta_i$ / $-\alpha_1\beta_i$ will be given as a punishment. L is a very large constant as a penalty for dissatisfying the constraint (7.1). Thus, the reward function of a decision period can be defined as:

$$r_\Delta=\sum_{\alpha,\beta,i} r/N^i_{\alpha,\beta}=-\alpha_0\sum_{i=1}^{s}\frac{\beta_i N^i_{af}}{N^i_a}-\alpha_1\sum_{i=1}^{s}\frac{\beta_i N^i_{hf}}{N^i_h}. \tag{11}$$

Substituting Eqs. (4) – (6) can be further expressed as:

$$r_\Delta=-\left(\alpha_0\sum_{i=1}^{s}\beta_i P^\Delta_{af_i}+\alpha_1\sum_{i=1}^{s}\beta_i P^\Delta_{hf_i}\right)=-O_\Delta(t). \tag{12}$$

Then our optimization problem of maximizing $-O(t)$ can be approximately solved by maximizing the accumulated reward $\sum_T r_\Delta$ in the long run.

5) Policy(π): we use a random policy $\pi(a|s)\rightarrow[0,1]$ to represent the probability of selecting the action a given the current state s.

In our MDP model, we use the state-value function to evaluate the value of state s, which can be expressed as:

$$V_\pi(s)=E^\pi\left[\sum_{k=0}^{\infty}\gamma^k r^{t+k}(s^{t+k},a^{t+k})|s_t\right], \tag{13}$$

where γ is the discount factor representing the discount contribution of the future states to the current state. Besides, the action-value function is used to evaluate the selected action a in the current state s, and can be expressed as:

$$Q^\pi(s,a)=E^\pi\left[\sum_{k=0}^{\infty}\gamma^k r^{t+k}(s^{t+k},a^{t+k})|s_t,a_t\right]. \tag{14}$$

Assuming that the MDP starts from the state $s^t\in S$, it experiences a trajectory as:

$$\kappa\sim\{s^t,a^t,s^{t+1},a^{t+1},\cdots,s^{t+T},a^{t+T}\}. \tag{15}$$

Since the policy is stochastic, the trajectory κ is uncertain. Denote the probability of trajectory κ as $\pi_\xi(\kappa)$, and the cumulative reward of trajectory κ is $R(\kappa)=\sum_{k=0}^{T}\gamma^k r^{t+k}$. As a result, the objective function can be rewritten as:

$$\max\ -O(t)\approx U(\pi_\xi)=E_{\kappa\sim\pi_\xi(\kappa)}[R(\kappa)]=\int_{\kappa\sim\pi_\xi(\kappa)}R(\kappa)d\kappa. \tag{16}$$

3.2 Actor-Critic-Based Dynamic Channel Reservation Strategy

This MDP problem can be solved by using the reinforcement learning (RL) algorithm. Specifically, we use the Actor-Critic framework[17] to model high-dimensional discrete action space, which is a combination of the Actor and the Critic. The Critic uses a neural network to approximate the state-value function and to judge the actions by temporal difference (TD) errors. The Actor uses another neural network to approximate the optimal policy and then selects the action while interacting with the environment.

1) Actor: The Actor will constantly improve the policy by TD errors. In our MDP problem, the policy π_ξ is modeled as a conditional probability distribution parameterized by ξ. Thus the process of modifying the policy is equivalent to the process of updating the parameter ξ. Through the back-propagation algorithm, ξ is updated as follows:

$$\xi_{\text{new}}=\xi_{\text{old}}+\alpha_{\text{actor}}\nabla_\xi U(\pi_\xi), \tag{17}$$

where α_{actor} is the learning rate of the Actor, and the gradient $\nabla_\xi U(\pi_\xi)$ is as follows:

$$\nabla_\xi U(\pi_\xi)=\nabla_\xi\log\pi_\xi(a|s)\times A^\pi(s,a), \tag{18}$$

where $A^\pi(s,a)$ is the advantage function.

In this problem, we use Gaussian probability distribution to formulate the policy, which can be expressed as:

$$\pi_\xi(a|s)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{(a-\mu(s))^2}{2\sigma^2}\right), \tag{19}$$

where $\mu(s)$ is the expectation and σ is the standard deviation of the selected action. Meanwhile, $\mu(s)$ is the action with the highest probability at the state s and σ represents the extent of exploration over all actions. Exploration and exploitation can be well balanced by exploiting the Gaussian distribution. Thus the policy can be modified through the process of updating $\mu(s)$.

To update $\mu(s)$, we extract a feature vector $\phi(s)$ from the current state as the input of the Actor neural network, which is expressed as:

$$\phi(s)=(c\,;\lambda\,;K)^T. \tag{20}$$

The neural network will then output the normalized average of reserved channel thresholds, which is denoted by $\mu(s)=(u_1,u_1',\cdots,u_s,u_s')^T$. Thus the policy can be further derived as a $2s$-dimensional Gaussian probability distribution:

$$\pi_\xi(a|s)=\frac{1}{(\sqrt{2\pi})^s(|Cov|)^{\frac{1}{2}}}e^{\left(-\frac{1}{2}(a-u(s))^T Cov^{-1}(a-u(s))\right)}, \tag{21}$$

where Cov is the covariance matrix with diagonal σ^2. Based on the policy, an action vector will be generated and the system will transit to a new state after a decision period T_Δ.

2) Critic: The Critic is used to approximate the state-value function $V_\pi(s)$. Traditional RL uses Q-value tables to record state values, which will face the problem of dimensional explosion under the scenario of large state space. To cope with this problem, a neural network parameterized by θ is utilized to approximate the state-value function $V_\theta(s)$. In the critic process, $V_\theta(s)$ will be updated by updating parameters θ.

To evaluate the gap between the actual value and the approximated value of the state-value function in the state s, the definition of TD-error is given as follows:

$$\delta_t = V_\pi(s^{t+1}) - V_\theta(s^t), \tag{22}$$

where $V_\pi(s^{t+1}) = r^{t+1} + \gamma V_\theta(s^{t+1})$ according to the bootstrapping method in the RL framework. To guide the updating of parameters and improve the performance, the objective of the critic process is designed to minimize the TD-error δ_t and can be re-expressed as:

$$\min\ 1/2\ (\delta_t)^2 . \tag{23}$$

θ is updated by the gradient descent method as follows:

$$\theta_{\text{new}} = \theta_{\text{old}} - \alpha_{\text{critic}} |\delta_t| \nabla_{\theta_{\text{old}}} V_{\theta_{\text{old}}}(s^t), \tag{24}$$

where α_{critic} is the learning rate of the Critic.

3) Actor-Critic: The Actor updates the policy based on the state-value estimated by the Critic, while the Critic updates the state-value function according to the actions selected by the Actor and the state transitions generated by interactions with the environment. Besides, the performance of the Actor can be improved by replacing δ_t with an advantage function. Then the parameter update process in Actor can be rewritten as:

$$\nabla_\xi U(\pi_\xi) = \nabla_\xi \log \pi_\xi(a^t|s^t)\delta_t, \tag{25}$$

$$\xi_{\text{new}} = \xi_{\text{old}} + \alpha_{\text{actor}} \nabla_{\xi_{\text{old}}} \log \pi_{\xi_{\text{old}}}(a^t|s^t)\delta_t. \tag{26}$$

In summary, the proposed AC-DCRS is summarized as follows:

Algorithm 1. AC-DCRS

Input: $N, M, T, T_\Delta, \sigma, \alpha_{\text{actor}}, \alpha_{\text{critic}}, \lambda, \gamma$.
Output: Optimal dynamic adjustment policy π_ξ.
Initialize: $t = 0, n = 1, \xi = \xi_0, \theta = \theta_0, a = a_0, s = s_0, \Phi(s_0)$;
Repeat:

1. Action selection:
   Input $\Phi(s^t)$ into the Actor network to get $\mu(s^t)$ and select action a, i.e., adjust the threshold once.
2. According to the threshold adjusted by action, control incoming calls, while $(t \leqslant nT_\Delta)$:
   1) if a service call arrives, judge its service type and priority
   2) determine whether the call access is successful by the corresponding threshold:
      a) if c is less than the corresponding threshold, the call is admitted successfully and can be allocated a channel resource $c \leftarrow c + 1$;
      b) else the call is blocked.
   3) record the result of this call
   4) $t \leftarrow t + 1$.
3. State transition and reward feedback:
   1) obtain the access result in this T_Δ
   2) transition into the new state s^{t+1}, get a reward r^{t+1}
   3) update state feature vector $\Phi(s^{t+1})$
   4) calculate the state-value function $V_\theta(s^{t+1})$, $V_\theta(s^t)$.
4. Update policy:
   1) Critic network calculates and outputs TD-error $\delta_t = r^{t+1} + \gamma V_\theta(s^{t+1}) - V_\theta(s^t)$
   2) update Critic network parameters $\theta \leftarrow \theta - \alpha_{\text{critic}} |\delta_t| \nabla_\theta V_\theta(s^t)$
   3) update Actor network parameters $\xi \leftarrow \xi + \alpha_{actor} \nabla_\xi \log \pi_\xi(a^t|s^t)\delta_t$.
5. $n \leftarrow n + 1$, $s^t \leftarrow s^{t+1}$.

Until: $t \geqslant T$.

3.3 Complexity Analysis

In this subsection, we analyze the computing and space complexity of AC-DCRS and compare it with three baseline algorithms, i.e., FCR, handover priority fixed channel reservation strategy (HPFCR), and DCR.

In FCR and HPFCR, the thresholds are fixed. In HPFCR, the handover calls are given higher priority, and the thresholds for new calls are set to the same value. After the initial setting, the thresholds of DCR will dynamically change according to the proportion of the number of calls of various services after the decision period T_Δ has passed. Its normalized threshold can be expressed as:

$$K_{\text{DCR}} = c' + \frac{1 - c'}{n} [n_1, n_1 + n_1', \cdots, n_1 + \cdots + n_s'], \tag{27}$$

where c' represents the normalized number of shared channels, which can be used by all calls, n represents the total number of calls arrived, and n_i and n_i' represent the number of new calls and handover calls of the i-th service respectively. As both FCR and HPFCR use a fixed threshold, the computing complexity is $O(1)$and the space complexity is $O(s)$. On the other hand, DCR will dynamically change the threshold according to Eq. (27), and thus the computing complexity and space complexity are both equal to $O(s)$.

In our AC-DCRS, the neural networks are introduced to fit the policy function and the threshold is obtained according to the output feature vector. Specifically, we use a fully-

connected neural network including two dense hidden layers. Suppose the number of neurons of the two layers is N_{L_1} and N_{L_2}. The dimension of the input feature vector is $4s + 1$ and the dimension of the output feature vector is $2s$. Therefore, the computing complexity is $O(s^2 \times N_{L_1} \times N_{L_2})$. We need to store the weights and bias of the middle layer and the values of thresholds, which results in a space complexity of $O(N_{L_1} + N_{L_2} + s)$. By using additional space and computing resources, our AC-DCRS can intelligently adjust the threshold and achieve a better performance.

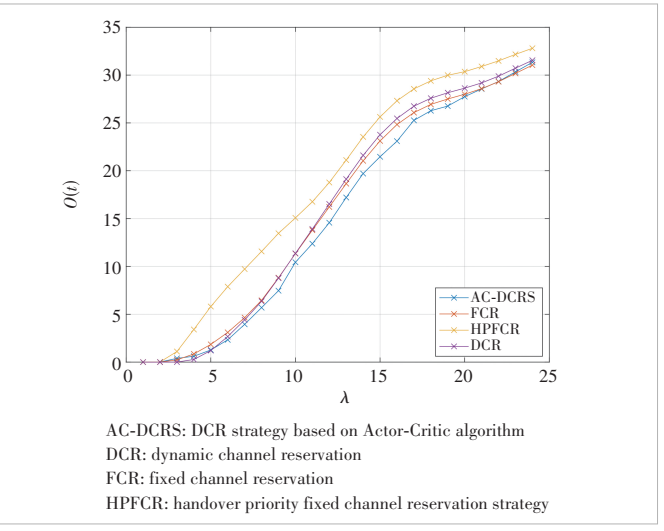
4 Performance Evaluation

4.1 Simulation Setting

We use a typical satellite system mobility model and basic assumptions, the parameters for the satellite communication network and the AC algorithm are shown in Tables 1 and 2 respectively.

4.2 Numerical Results

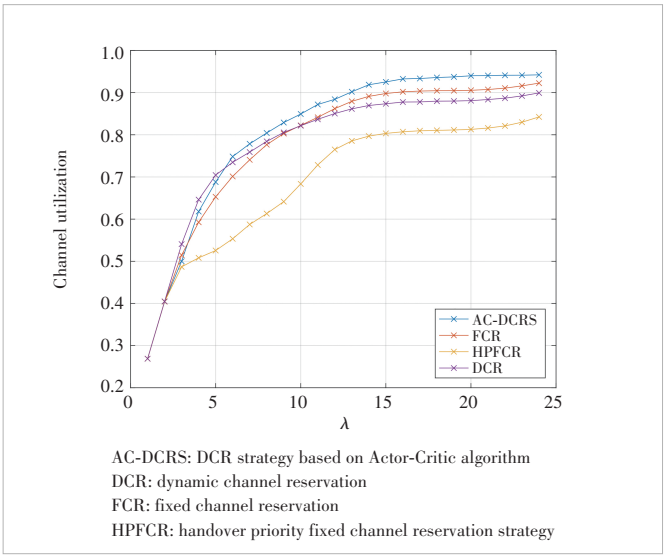
First, we examine the relationship between overall access failure probability $O(t)$, which is approximately equal to the negative long-term accumulated reward, with the varying total call arrival rate λ. As shown in Fig. 4, since our optimization objective is to minimize $O(t)$, the greater $O(t)$, the worse overall system performance. We can see that the proposed AC-DCRS algorithm can learn better admission control strategies and achieve better overall system performance, compared with FCR, HPFCR, and DCR in most traffic scenarios. However, there is no obvious advantage in the scenarios of very low and high traffic loads. This is because AC-DCRS needs some trial-and-error interactions with the environment. Reducing the thresholds of low-priority services causes some call failures in low-traffic scenarios, and multiple next-highest priority services get admission to quickly filling up the channel, which affects the overall access failure probability in high-traffic load scenarios.



AC-DCRS: DCR strategy based on Actor-Critic algorithm
DCR: dynamic channel reservation
FCR: fixed channel reservation
HPFCR: handover priority fixed channel reservation strategy

▲Figure 4. $O(t)$ with varying λ

Next, we explore the relationship between the channel utilization and the average call success rate with the varying total call arrival rate λ. From Figs. 5 and 6, we can find that the channel utilization increases with the traffic load, and the average call success rate decreases with the traffic load. We can



AC-DCRS: DCR strategy based on Actor-Critic algorithm
DCR: dynamic channel reservation
FCR: fixed channel reservation
HPFCR: handover priority fixed channel reservation strategy

▲Figure 5. Channel utilization with varying λ

▼Table 1. Simulation parameters

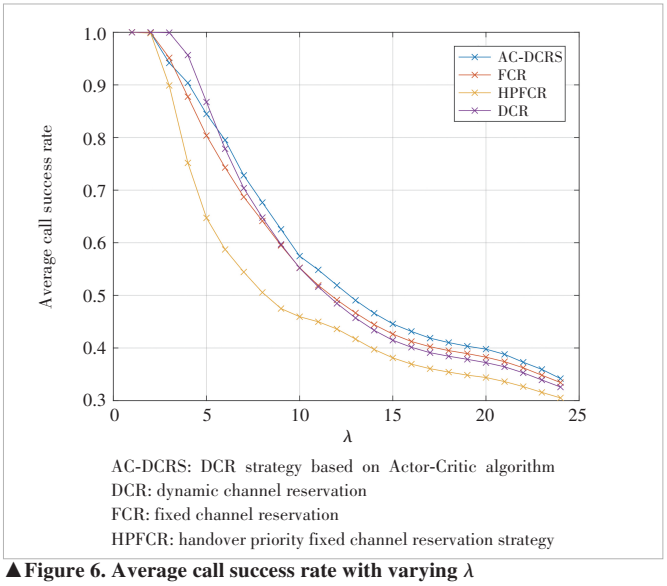
Parameter	Value
Number of services s	3
Number of beam cell channels	100
Average call duration parameter u	1/30 s
Call arrival rate ratio p_1: p_2: p_3	0.2:0.3:0.5
Decision period T_Δ	5 s
Maximum number of calls N	50
Balance factor α_0, α_1	0.4, 0.6
Balance factor β_1, β_2, β_3	0.2, 0.3, 0.5
FCR normalized fixed threshold setting	[0.73, 0.75, 0.82, 0.85, 0.86, 1]
HPFCR normalized fixed threshold setting	[0.73, 0.75, 0.73, 0.85, 0.73, 1]
DCR normalized initial threshold setting	[0.73, 0.75, 0.82, 0.85, 0.86, 1]
Number of periods played	20 000T_Δ
Total call arrival rate λ	2 – 25 calls/s

DCR: dynamic channel reservation
FCR: fixed channel reservation
HPFCR: handover priority fixed channel reservation

▼Table 2. AC algorithm parameters

Parameter	Value
Discount factor γ	0.99
Learning rate of policy α_{actor}	0.002
Learning rate of value function α_{critic}	0.005
Action selection variance σ	0.05

AC: Actor-Critic

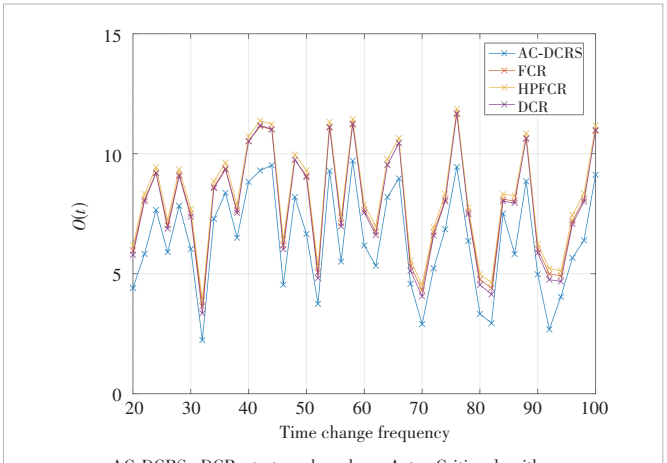


AC-DCRS: DCR strategy based on Actor-Critic algorithm
DCR: dynamic channel reservation
FCR: fixed channel reservation
HPFCR: handover priority fixed channel reservation strategy

▲**Figure 6. Average call success rate with varying λ**

also find that AC-DCRS outperforms FCR, HPFCR, and DCR in these two aspects. This is because AC-DCRS can well balance the call admission of all services from the level of the entire system. Ensuring the admission of high-priority service calls makes as many calls of multiple services as possible get admission.

We assume that the total call arrival rate changes dynamically at a certain time frequency. The initial total call arrival rate is 8 calls/s, and the range of change is $[\lambda_t - 2, \lambda_t + 2]$, where λ_t is the current total call arrival rate. As shown in Fig. 7, the AC-DCRS can achieve better system performance in different dynamic scenarios, compared with comparison algorithms. This is because our AC-DCRS can learn the optimal admission control strategy under the current traffic and can adjust the threshold in real time.



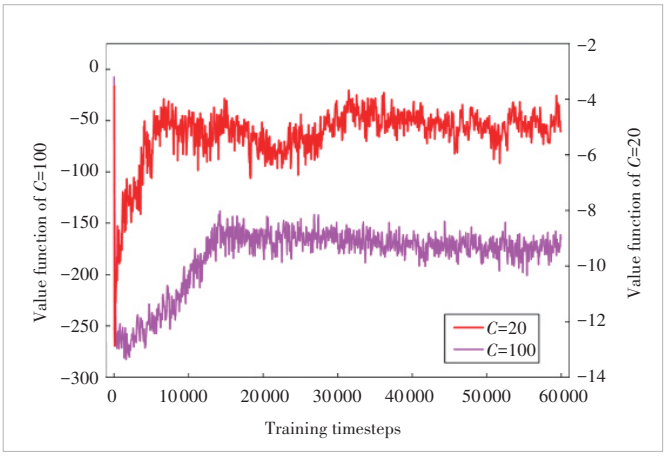
AC-DCRS: DCR strategy based on Actor-Critic algorithm
DCR: dynamic channel reservation
FCR: fixed channel reservation
HPFCR: handover priority fixed channel reservation strategy

▲**Figure 7. $O(t)$ with change frequency varying**

Finally, we show the convergence of the value function in the AC-DCRS algorithm. We separately consider the convergence in the small state space (C = 20) and big state space (C= 100) cases. We observe the dynamic change of the state-value function at a certain state as the Critic evolves. As shown in Fig. 8, we can find that after certain training steps, the value function converges. The convergence speed varies with the sizes of state space, for the reason that it requires more iterations to traverse a larger state space to reach optimal strategy. In addition, the obtained strategy through training can be applied to similar scenarios in different satellite beam cells. The training data of different satellite beam cells in similar scenarios can be shared for migration training, which will accelerate the convergence to the optimal strategy.

5 Conclusions

In this paper, we have proposed a dynamic channel reservation strategy AC-DCRS based on the Actor-Critic algorithm to realize intelligent admission control in a satellite network. AC-DCRS can learn an optimal admission policy for both new calls and handover calls with different service priorities, which will improve the performance of both the user side and the network. Numerical results show that our proposed AC-DCRS algorithm achieves better long-term overall system performance, average call success rate, and channel utilization under different traffic conditions and dynamic scenarios compared with traditional channel reservation strategies.



▲**Figure 8. Value function with time varying**

References

[1] LIU J J, SHI Y P, FADLULLAH Z M, et al. Space-air-ground integrated network: a survey [J]. IEEE communications surveys & tutorials, 2018, 20(4): 2714 – 2741. DOI: 10.1109/COMST.2018.2841996

[2] DING R, CHEN T T, LIU L, et al. 5G integrated satellite communication systems: architectures, air interface, and standardization [C]//International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, 2020: 702 – 707. DOI: 10.1109/WCSP49889.2020.9299757

[3] DENG R Q, DI B Y, ZHANG H L, et al. Ultra-dense LEO satellite constellation design for global coverage in terrestrial-satellite networks [C]// Global Communications Conference. IEEE, 2021: 1 – 6. DOI: 10.1109/

GLOBECOM42002.2020.9322362
[4] DUAN C F, DUAN R Q, FENG J, et al. A novel channel allocation strategy in low earth orbit satellite networks [C]//6th International Conference on Computer and Communications (ICCC). IEEE, 2021: 8 – 13. DOI: 10.1109/ICCC51575.2020.9345173
[5] ZHOU J, YE X G, PAN Y, et al. Dynamic channel reservation scheme based on priorities in LEO satellite systems [J]. Journal of systems engineering and electronics, 2015, 26(1): 1 – 9. DOI: 10.1109/JSEE.2015.00001
[6] LI Z W, XIE Z C, LIANG X W. Dynamic channel reservation strategy based on DQN algorithm for multi-service LEO satellite communication system [J]. IEEE wireless communications letters, 2021, 10(4): 770 – 774. DOI: 10.1109/LWC.2020.3043073
[7] MARAL G, RESTREPO J, DEL RE E, et al. Performance analysis for a guaranteed handover service in an LEO constellation with a "satellite-fixed cell" system [J]. IEEE transactions on vehicular technology, 1998, 47(4): 1200 – 1214. DOI: 10.1109/25.728509
[8] WANG X L, WANG X X. The research of channel reservation strategy in LEO satellite network [C]//11th International Conference on Dependable, Autonomic and Secure Computing. IEEE, 2014: 590 – 594. DOI: 10.1109/DASC.2013.131
[9] BOUKHATEM L, GAITI D, PUJOLLE G. A channel reservation algorithm for handover issues in LEO satellite systems based on a satellite-fixed cell coverage [C]//IEEE VTS 53rd Vehicular Technology Conference. IEEE, 2001: 2975 – 2979. DOI: 10.1109/VETECS.2001.944147
[10] BEYLOT A L, BOUMERDASSI S. Adaptive channel reservation schemes in multitraffic LEO satellite systems [C]//IEEE Global Telecommunications Conference. IEEE, 2002: 2740 – 2743. DOI: 10.1109/GLOCOM.2001.966272
[11] ZOU Q Y, ZHU L D. Dynamic channel allocation strategy of satellite communication systems based on grey prediction [C]//International Symposium on Networks, Computers and Communications (ISNCC). IEEE, 2019: 1 – 5. DOI: 10.1109/ISNCC.2019.8909122
[12] CHATTERJEE S, SAHA J, BANERJEE S, et al. Neighbour Location Based Channel Reservation scheme for LEO Satellite communication [C]//International Conference on Communications, Devices and Intelligent Systems (CODIS). IEEE, 2012: 73 – 76. DOI: 10.1109/CODIS.2012.6422139
[13] RAHMAN M, WALINGO T, TAKAWIRA F. Adaptive handover scheme for LEO satellite communication system [C]//Proceedings of AFRICON. IEEE, 2015: 1 – 5. DOI: 10.1109/AFRCON.2015.7332051
[14] CHEN L M, GUO Q, WANG H Y. A handover management scheme based on adaptive probabilistic resource reservation for multimedia LEO satellite networks [C]//WASE International Conference on Information Engineering. IEEE, 2010: 255 – 259. DOI: 10.1109/ICIE.2010.67
[15] LIU S J, HU X, WANG W D. Deep reinforcement learning based dynamic channel allocation algorithm in multibeam satellite systems [J]. IEEE access, 2018, 6: 15733 – 15742. DOI: 10.1109/ACCESS.2018.2809581
[16] CHOWDHURY P K, ATIQUZZAMAN M, IVANCIC W. Handover schemes in satellite networks: state-of-the-art and future research directions [J]. IEEE communications surveys & tutorials, 2006, 8(4): 2 – 14. DOI: 10.1109/COMST.2006.283818
[17] BHATNAGAR S, SUTTON R S, GHAVAMZADEH M, et al. Natural actor-critic algorithms [J]. Automatica, 2009, 45(11): 2471 – 2482. DOI: 10.1016/j.automatica.2009.07.008

Biographies

CHENG Lei received her BS degree in communication engineering from University of Electronic Science and Technology of China (UESTC) in 2019. She now is pursuing her PhD degree at the National Key Laboratory of Wireless Communications, UESTC. Her research interests include resource management and network control in space-air-ground/satellite-terrestrial integrated networks by using optimization theory and machine learning techniques.

QIN Shuang (blueqs@uestc.edu.cn) received his BS degree in electronic information science and technology and PhD degree in communication and information system from University of Electronic Science and Technology of China (UESTC) in 2006 and 2012, respectively. He is currently a professor with the National Key Laboratory of Wireless Communications, UESTC. His research interests include wireless and mobile networks.

FENG Gang received his BE and ME degrees in electronic engineering from University of Electronic Science and Technology of China (UESTC) in 1986 and 1989, respectively, and PhD degree in information engineering from The Chinese University of Hong Kong, China in 1998. He joined the School of Electric and Electronic Engineering, Nanyang Technological University, Singapore in December 2000 as an assistant professor and became an associate professor in October 2005. He is currently a professor with the National Laboratory of Wireless Communications, UESTC. He has extensive research experience and has published widely on wireless networking. His research interests include next generation mobile networks, mobile cloud computing, and AI-enabled wireless networking. He has received the IEEE ComSoc TAOS Best Paper Award and the ICC Best Paper Award in 2019. A number of his papers have been highly cited.

A 220 GHz Frequency-Division Multiplexing Wireless Link with High Data Rate

ZHANG Bo[1], WANG Yihui[1], FENG Yinian[1], YANG Yonghui[2,3], PENG Lin[2,3]

(1. University of Electronic Science and Technology of China, Chengdu 610054, China;
2. ZTE Corporation, Shenzhen 518057, China;
3. State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518055, China)

DOI: 10.12142/ZTECOM.202303009

https://kns.cnki.net/kcms/detail/34.1294.TN.20230810.1128.002.html, published online August 10, 2023

Manuscript received: 2023-03-15

Abstract: With the development of wireless communication, the 6G mobile communication technology has received wide attention. As one of the key technologies of 6G, terahertz (THz) communication technology has the characteristics of ultra-high bandwidth, high security and low environmental noise. In this paper, a THz duplexer with a half-wavelength coupling structure and a sub-harmonic mixer operating at 216 GHz and 204 GHz are designed and measured. Based on these key devices, a 220 GHz frequency-division multiplexing communication system is proposed, with a real-time data rate of 10.4 Gbit/s for one channel and a transmission distance of 15 m. The measured constellation diagram of two receivers is clearly visible, the signal-to-noise ratio (SNR) is higher than 22 dB, and the bit error ratio (BER) is less than 10^{-8}. Furthermore, the high definition (HD) 4K video can also be transmitted in real time without stutter.

Keywords: frequency-division multiplexing; sub-harmonic mixer; terahertz communication; terahertz duplexer

Citation (Format 1): ZHANG B, WANG Y H, FENG Y N, et al. A 220 GHz frequency-division multiplexing wireless link with high data rate [J]. ZTE Communications, 2023, 21(3): 63 - 69. DOI: 10.12142/ZTECOM.202303009
Citation (Format 2): B. Zhang, Y. H. Wang, Y. N. Feng, et al., "A 220 GHz frequency-division multiplexing wireless link with high data rate," *ZTE Communications*, vol. 21, no. 3, pp. 63 - 69, Sept. 2023. doi: 10.12142/ZTECOM.202303009.

1 Introduction

With the development of wireless communication technology, the global mobile data traffic is growing exponentially, and the demand for high-speed and low-latency data transmission is gradually growing[1]. However, the currently used millimeter wave band cannot achieve these visions, so researchers turn their attention to the terahertz (THz) band. THz waves range from 0.1 to 100 THz, and the THz wireless communication has strong directionality, excellent security, and wide bandwidth, which has a communication rate of more than 100 Gbit/s.

The NTT Laboratories in Japan first started research on THz wireless communication in 2004, the carrier frequency of which is 120 GHz under amplitude shift keying (ASK) modulation for a data rate of 10 Gbit/s[2]. In recent years, this team has developed 300 GHz power amplifiers, low-noise amplifiers, and base-wave mixers based on 80 nm InP-HEMT technology in Ref. [3], and the proposed THz wireless link has achieved a wireless data rate of up to 120 Gbit/s based on 16 quadrature amplitude modulation (QAM) modulation over a transmission distance of 9.8 m. In Ref. [4], the communication system is developed by Karlsruhe Institute of Technology with a carrier frequency of 237.5 GHz under 16QAM modulation. The proposed system achieved 100 Gbit/s over a distance of 20 m. With the support of the DOTSEVEN project, the team of the Institute for High Frequency and Communication Technology has developed a 240 GHz full integrated transceiver chip based on a 130 nm SiGe Bi-COMS technology to achieve 10 Gbit/s over a transmission distance of 15 cm and a BER of less than 10^{-9}[5]. Recently, this team has also achieved Quadrature Phase Shift Keying (QPSK) modulated signal transmission at a transmission distance of 1 m and a transmission rate of 110 Gbit/s with a measured error vector magnitude (EVM) of 31.9% based on the 130 nm SiGe heterojunction bipolar transistor (HBT) technology in the 220 - 225 GHz band[6]. The 220 GHz communication system[7] built by University of Electronic Science and Technology of China is based on the THz solid-state receiving front-end, which can realize 3.52 Gbit/s wireless communication over a distance of 200 m using QPSK modulation. In Ref. [8], the Chinese Academy of Engineering Physics developed a 140 GHz wireless link for offshore communication, which has

This work is supported by the National Natural Science Foundation of China under Grant Nos. 62022022 and 62101107, the National Key R&D Program of China under Grant No. 2018YFB1801502, China Postdoctoral Science Foundation under Grant No. 2021TQ0057, and ZTE Industry-University-Institute Cooperation Funds.

a transmit power of 33 dBm and finally achieved a transmission distance of 27 km with 500 Mbit/s. According to these works, the solid-state THz communication technology has fully demonstrated its great application potential and research value in high-rate wireless communication. However, the exploration of THz communication is still limited to point-to-point communication. Therefore, the research on the THz frequency-division multiplexing wireless link is very important to THz communication.

In this paper, we propose a 220 GHz frequency-division multiplexing communication system, which adopts the classical super outlier architecture, focusing on the design including a 220 GHz sub-harmonic mixer, duplexer and THz frequency-division multiplexing communication system, which finally realizes the transmission data rate of 10 Gbit/s for up/down link over a distance of 15 m.

2 System Architecture

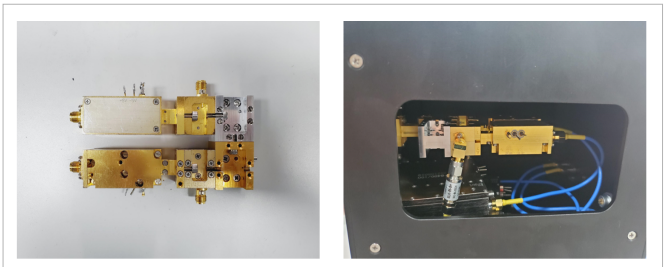
Full duplex communication systems often use a time division multiplexing (TDM) mode, a frequency-division multiplexing (FDM) mode and an isolation mode based on waveguide-guided orthogonal couplers[10]. During these modes, FDM is to divide the total bandwidth used for the transmission channel into two sub-carriers with different frequencies for receiving and transmitting. Although the two channels occupy different frequencies and large spectrum resources, the THz band has huge spectrum resources. Therefore, the FDM mode is extremely suitable for realizing THz full duplex communication.

In this paper, a 220 GHz full-duplex wireless communication system is built in a frequency-division multiplexing mode, and the system block diagram is shown in Fig. 1. The baseband signal is upconverted by a C-band frequency conversion module, which is filtered out of the upper sideband and fed into a THz mixer to upconvert to the THz band, and then filtered by a THz duplexer to remove the mirror frequency and get the radio frequency (RF) signal. The operating frequency of the uplink channel is 197.3 – 199.9 GHz, and that of the downlink channel is 207 – 213 GHz. The RF link of the communication system is shown in Fig. 2. To effectively reduce the space occupied by the link, the RF link is connected to the duplexer with a curved waveguide.

3 Critical Components of System

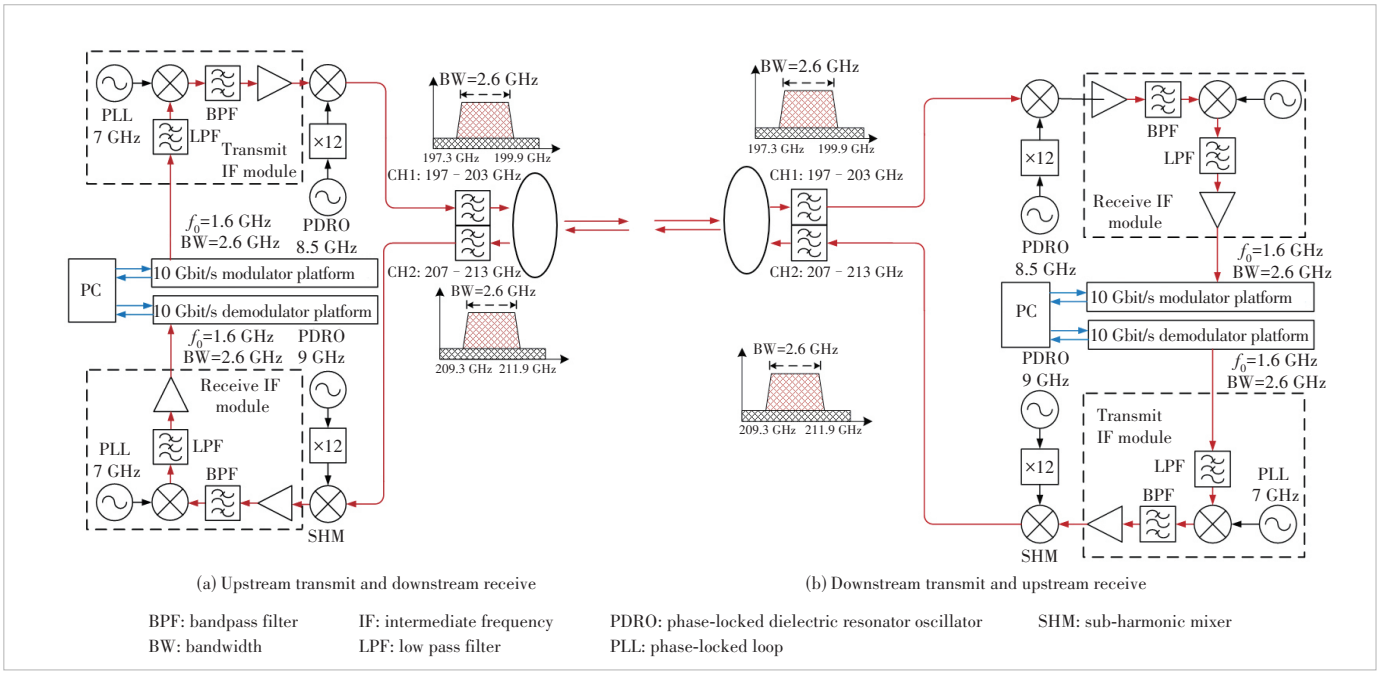
3.1 220 GHz Sub-Harmonic Mixer Study

A mixer is a crucial device of a THz communication system that uses a nonlinear device of solid-state devices to generate an output signal containing multiple frequency components[9]. The mixer consists of an RF input structure, a local oscillator (LO) input structure, multiple filters, inverted parallel diode



(a) RF link (b) Chassis diagram

▲Figure 2. Front-end link of the communication system

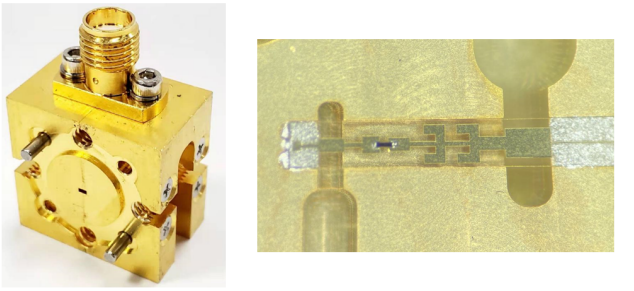


▲Figure 1. Communication system block diagram

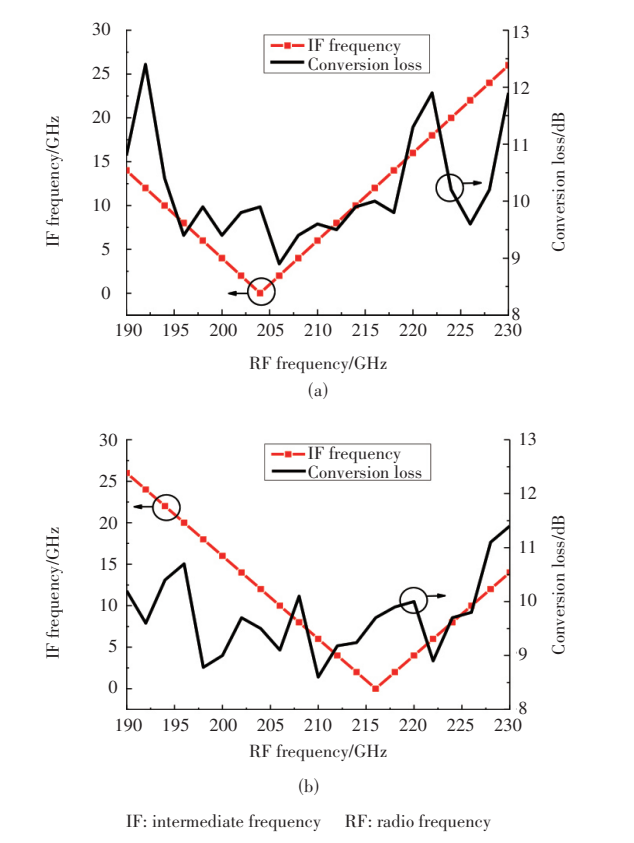
pairs, and a matching network. In this paper, we design a 220 GHz sub-harmonic mixer using Schottky diodes, and the circuit substrate is based on a 50 μm quartz substrate.

The mixer and the internal circuit are shown in Fig. 3. The cavity is made of brass, which is gold-plated on the surface. During the test, the output power of the LO link is controlled within 3 – 8 mW, and the input RF signal power is uniformly adjusted to −10 dBm.

The measured result shows that when the LO and RF signal input 204 GHz and 195 – 220 GHz, respectively, the frequency conversion loss is less than 10 dB, which is 8.8 dB at minimum, as shown in Fig. 4(a). After changing the LO frequency to 216 GHz, the frequency conversion loss is less than 10.3 dB, which is 8.9 dB at minimum, as shown in Fig. 4(b).



▲Figure 3. 220 GHz sub-harmonic mixer

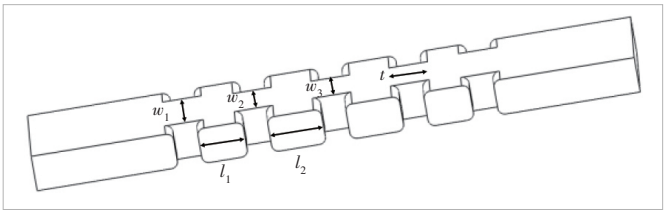


▲Figure 4. Frequency conversion loss test results of mixers with different LO frequencies: (a) 204 GHz and (b) 216 GHz

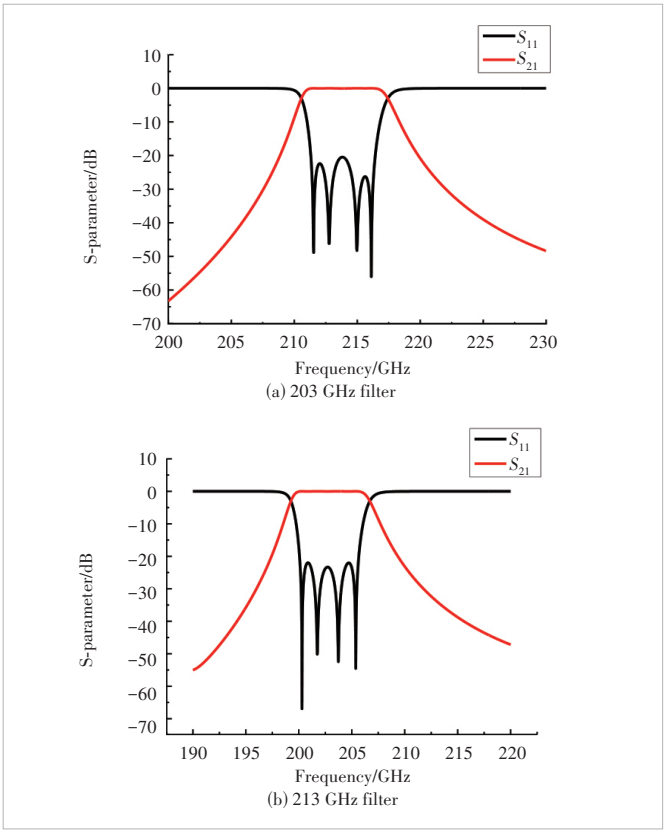
3.2 220 GHz Duplexer Study

In this paper, the designed duplexer adopts the reactance-coupled bandpass filter structure, as shown in Fig. 5. The mode of each resonant cavity is TE_{101} and the resonant cavity length is approximated as a half wavelength. The parameters affecting the filter performance are mainly the length of each resonant cavity l_i and the width of each coupling window w_i. Changing l_i has a significant effect on the center frequency, and the length of the resonant cavity near the center l_2 has a higher effect on the center frequency than l_1. Adjusting w_i has a certain effect on the in-band characteristics and bandwidth.

Two different band filters with center frequencies of 203 GHz and 214 GHz are designed. It can be seen from Fig. 6 that the simulation results show the bandwidths are 6 GHz and the in-band insertion loss (IL) is lower than 0.1 dB.

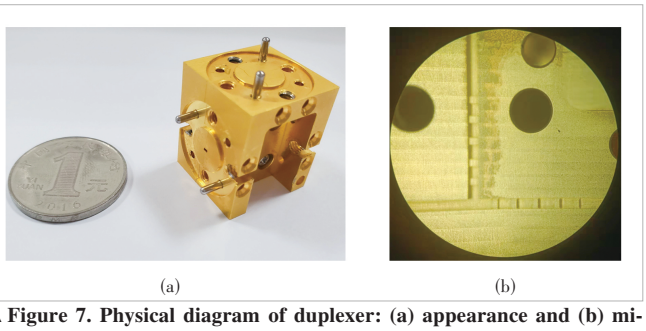


▲Figure 5. Filter structure diagram

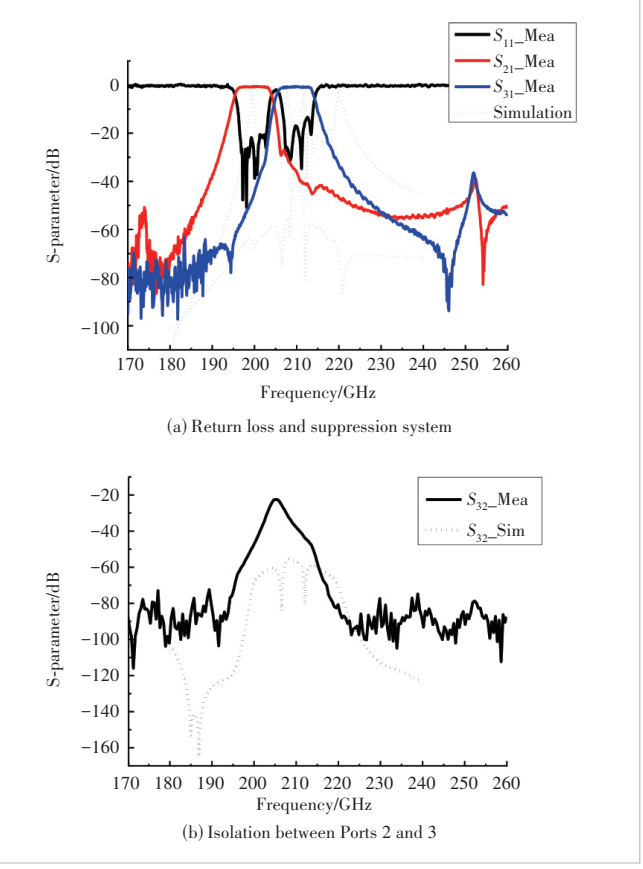


▲Figure 6. Filter simulation results

After filters are designed, every individual filter is directly connected through T-junctions[11] to form a multiplexer. The machining physical drawing and microscope photo of the internal structure are shown in Fig. 7. The proposed duplexer is fabricated by computerized numerical control (CNC) milling technology. The cavity is made of brass with a gold-plated surface. The simulation and measured results show that the duplexer passband is 197 – 203 GHz and 207 – 213 GHz, the isolation degree is 22.51 dB at 205 GHz, the average in-band return loss is less than 15 dB, and the average in-band insertion loss (IL) is less than 0.8 dB, which is 0.58 dB at minimum, as shown in Fig. 8. Compared with the simulation results, the overall frequency shift is 3 GHz and the isolation degree is 8 dB worse. However, the single-channel bandwidth is not different from the simulation results. The reason may be that the length of the resonant cavity in the center of the two channels is larger than the simulated value, and the width of the coupling window is similar to the simulated value, which leads to the same bandwidth and downward frequency shift.



▲ **Figure 7. Physical diagram of duplexer: (a) appearance and (b) microscope photo of internal structure**



▲ **Figure 8. Duplexer simulation and test data**

4 Measurement Results

For terahertz wireless communication, the link-budget of the system mainly includes transmit power, antenna gain, free space loss, and atmospheric attenuation[12]. The received power can be calculated as follows:

$$P_R = P_T + G_T + G_R + 20\lg\left(\frac{c}{4\pi R_f}\right) - L_0 - L_{ex}, \quad (1)$$

where P_T is the transmitter output power in dBm, G_T and G_R are antennas gains, R is the transmission distance, L_0 is atmospheric attenuation which is less than 0.01 dB, and L_{ex} is an excess loss set to 3 dB.

The transmitting antenna and receiving antenna both use the WR-4 band lens antenna. The baseband power input to the 220 GHz sub-harmonic mixer is less than −10 dBm, and combined with the test results in Section 3, we can get the antenna transmit power P_T less than −20 dBm. Therefore, the maximum received power of the receiver is −35.78 dBm.

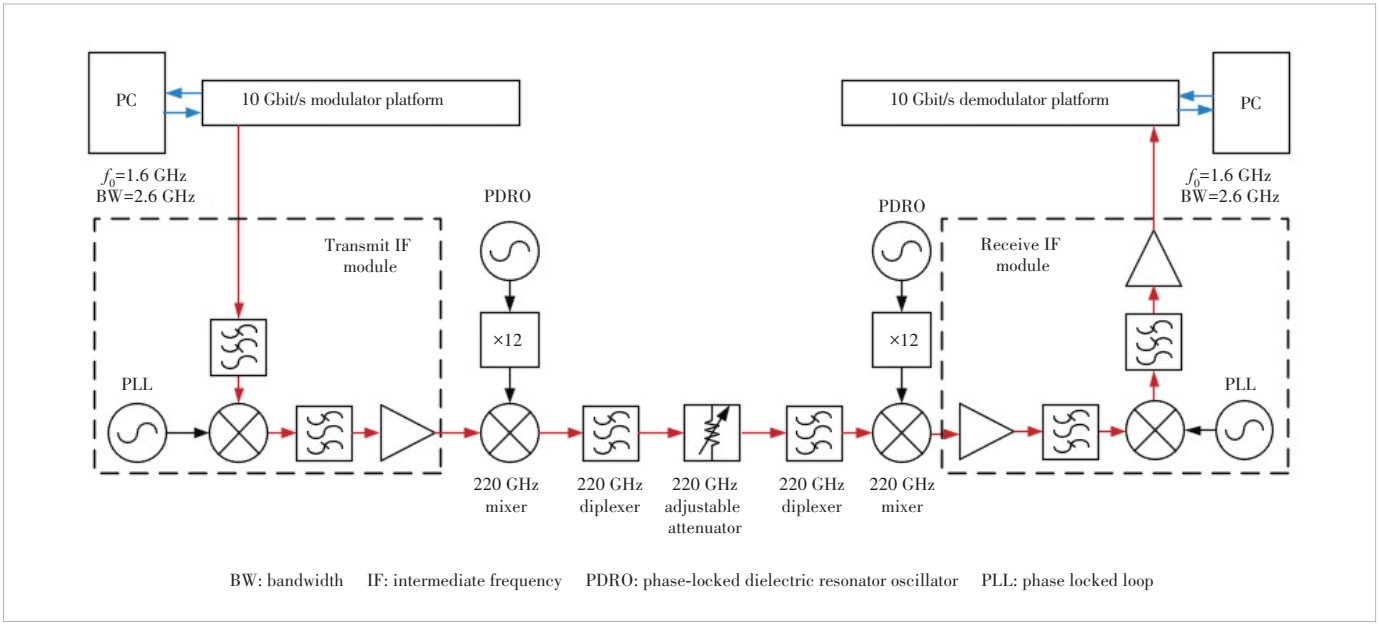
Receiver sensitivity $S_{i,\min}$, which is the minimum received power of the receiver and can be calculated as follows:

$$S_{i,\min} = -174 + 10\lg B + \mathrm{NF} + \mathrm{SNR}, \quad (2)$$

where B is signal bandwidth as 2.6 GHz, NF is the receiver noise factor that can be calculated from the system cascade noise factor calculation formula as 13 dB, and SNR is the minimum signal-to-noise ratio required for baseband demodulation which requires at least 19 dB using 16QAM modulation. Therefore, the receiver sensitivity can be calculated as −46.02 dBm. The difference between the received power and the receiver sensitivity is 10.24 dB, which means the system has a link-budget of 10.24 dB.

The block diagram of the single-channel experiment is shown in Fig. 9. The LO signals of the transmitter and receiver are both provided by the 110 GHz twelve-octave frequency multiplier. The influence of phase white noise at the far end of the local oscillator on the communication quality is great[13], and the deterioration of phase white noise at the far end of the local oscillator with the frequency multiplier is 20 $\log_{10}(n)$, so the input of the local oscillator signal should reduce its far-end white noise as much as possible. Compared with a phase-locked loop (PLL), the phase-locked dielectric resonator oscillator (PDRO) has a better phase noise. Therefore, the input of this oscillator selects PDRO at 8.5 GHz and 9 GHz, and its phase noise at 1 MHz can reach −131 dBc/Hz.

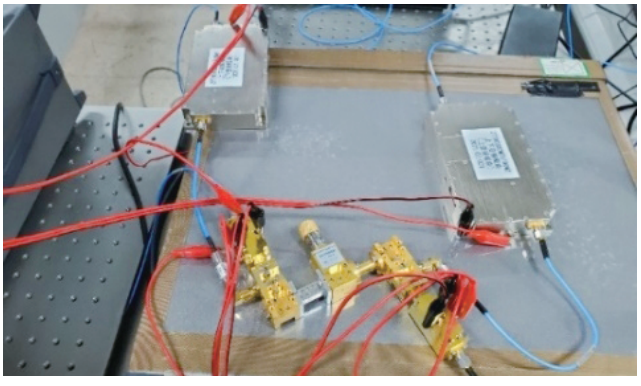
The experimental scenario is shown in Fig. 10. In the trans-



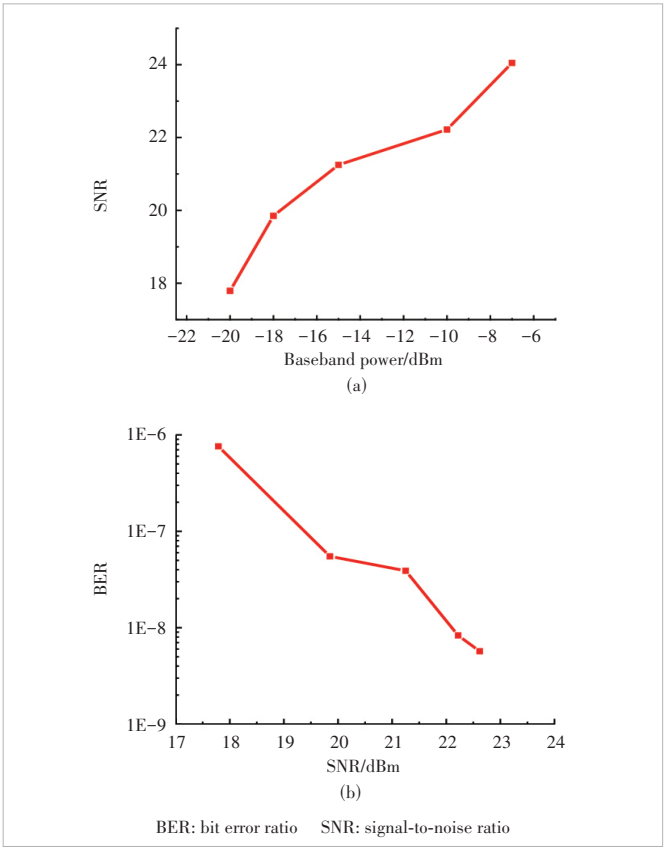
▲Figure 9. Single-channel experimental block diagram

mit link, the baseband signal is first upconverted by the C-band frequency conversion module, filtered out of the upper sideband and then fed into the THz mixer to be upconverted to the THz band, and then filtered out of the mirror frequency by the THz duplexer. In the receive link, the RF signal enters the THz mixer after the duplexer and is down-converted to the intermediate frequency (IF) signal. It is then down-converted to the baseband signal by the C-band frequency conversion module and demodulated after amplification and filtering. The transmit and receive links are connected with a THz adjustable attenuator for simulating the propagation loss of THz signals in free space. The relationship between different baseband power and the SNR is demonstrated in Fig. 11(a), which shows that there is no significant change in the demodulation SNR with the baseband power above −15 dBm and when the baseband power is below −20 dBm, the SNR deteriorates further. Fig. 11(b) shows the relationship between the demodulation SNR and BER.

The constellation diagrams corresponding to different SNRs

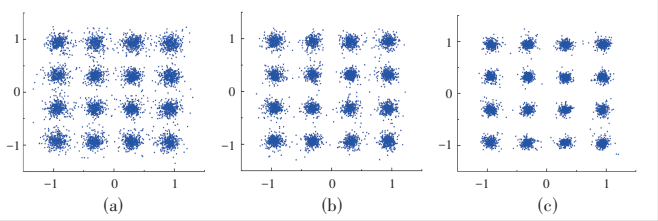


▲Figure 10. Single-channel experimental scene diagram



▲Figure 11. Single-channel test data: (a) relationship between transmit power and SNR and (b) relationship between SNR and BER

are demonstrated in Fig. 12, which shows that the communication results are basically the same when the LO frequency is 204 GHz and 216 GHz. Keeping the above test platform unchanged, the baseband modulation mode is changed to

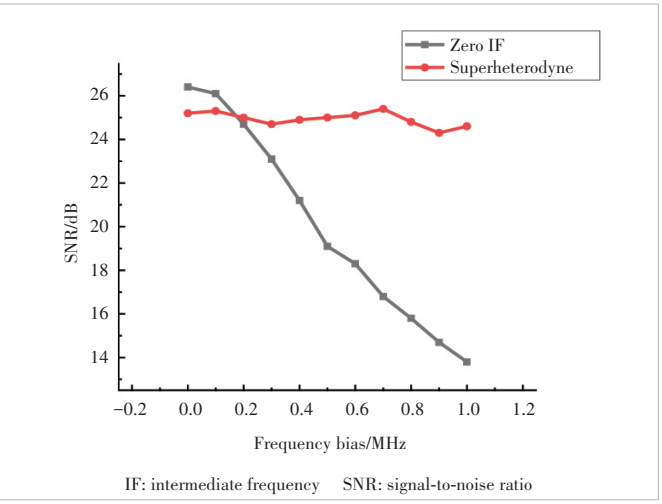


▲ **Figure 12. Constellation diagrams with different signal-to-noise ratio SNR: (a) 17.79 dB, (b) 19.85 dB, and (c) 22.62 dB**

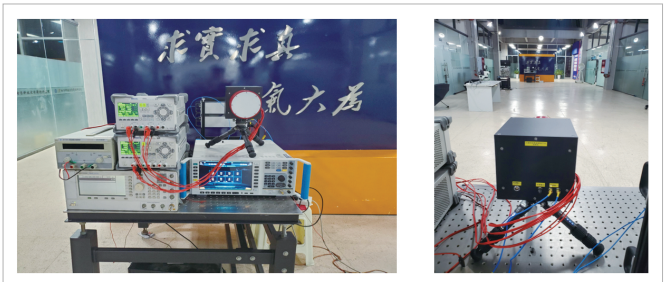
64QAM and the attenuator is adjusted to make the best performance. Then the performance at the LO frequency of 204 GHz and 216 GHz are tested respectively.

The THz attenuation is removed so that the transmitter and the receiver are directly connected back-to-back. The PDRO at the receiver side is replaced as the signal source is input to achieve the effect of transmitting and receiving LO frequency offset by changing the signal source input. The range of frequency offset measurement is 0 – 1 MHz with a step of 0.1 MHz. The variation of SNR with the amount of frequency offset is shown in Fig. 13. The baseband signal in the receiving section is a single sideband. The LO frequency bias corresponds to the overall offset of the baseband spectrum and there is no signal crossover. Therefore, as long as the frequency bias is less than the threshold, the SNR is basically unchanged. When the frequency offset exceeds the threshold value, the SNR deteriorates sharply and the BER is 1.

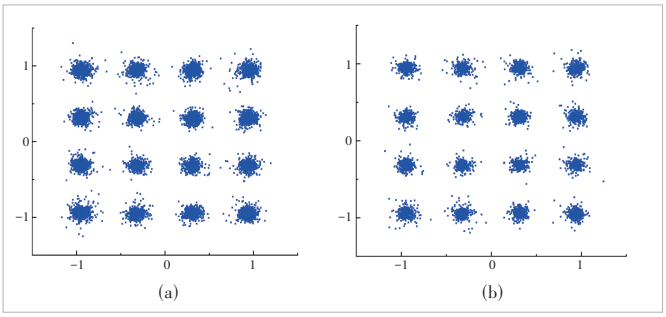
After measurement, the frequency bias threshold of this communication system is 2.5 MHz. The photograph of the communication experiment is shown in Fig. 14. The corresponding constellation diagram is shown in Fig. 15, where the demodulation SNR of the uplink channel is 22.91 dB when the LO frequency is 204 GHz and the demodulation SNR of the downlink channel is 23.19 dB when it is 216 GHz. The BER is less than 10^{-8}. The final transmission distance of this communication system is 15 m, which can reach 20 m by calculation because of spare capacity in the link. The comparison of the previously published THz wireless links with ours is shown in Table 1.



▲ **Figure 13. Variation of SNR with different frequency bias**



▲ **Figure 14. (a) Uplink transmit and downlink receive; (b) downlink transmit and uplink receive**



▲ **Figure 15. Constellation diagram: (a) 204 GHz uplink; (b) 216 GHz downlink**

5 Conclusions

In this paper, the front-end key components including a 220 GHz sub-harmonic mixer and a 220 GHz duplexer are designed for the application requirements of the THz frequency-division multiplexing communication system. The sub-harmonic mixer is designed based on the domestic Schottky diode, which has a fixed fundamental frequency of 204 GHz and 216 GHz. The frequency conversion loss is less than 10.5 dB at 20 GHz. The 220 GHz duplexer is designed based on the principle of the cavity filter. The range of the duplexer passband is 197 – 203 GHz and 207 – 213 GHz, and the return loss in the common port is less than 15 dB, which can effectively divide the channel of the communication system and suppress the mirror frequency. Based on the above circuit de-

▼ **Table 1. Experimental prototype performance comparison of terahertz (THz) wireless communication systems**

Reference	Frequency/GHz	Transmission Rate/(Gbit/s)	Modulation Format	Transmission Distance/m	Real-Time Transmission
Ref. [2]	120	10	ASK	1	No
Ref. [3]	290	120	16QAM	9.8	No
Ref. [5]	220 – 225	110	QPSK	1	No
Our proposed work	**220**	**10.4**	**16QAM**	**15**	**Yes**
	204/216	**10.4/10.4**	**16QAM**	**15**	

ASK: amplitude shift keying QPSK: Quadrature Phase Shift Keying
QAM: quadrature amplitude modulation

sign, the 220 GHz frequency-division full duplexer communication system is proposed, which realizes real-time high-speed communication with a transmission distance greater than 15 m, and uplink and downlink transmission rates of 10.4 Gbit/s, respectively. The measured BER is lower than 10^{-7}. Furthermore, the HD 4K video can be transmitted in real time. This work realizes a high-rate and long-range THz communication system that has broad application prospects in wireless communication. This paper builds a 220 GHz frequency-division multiplexing communication system, which adopts non-coherent demodulation to achieve a transmission distance of 15 m, an uplink and downlink transmission rate of 10.4 Gbit/s each, and a BER lower than 10^{-7}.

References

[1] Cisco Company. Global mobile data traffic forecast update: white paper [R]. Cisco visual networking index: forecast and trends, 2017
[2] KUKUTSU N, HIRATA A, KOSUGI T, et al. 10 Gbit/s wireless transmission systems using 120 GHz band photodiode and MMIC technologies [C]//Compound semiconductor integrated circuit symposium. IEEE, 2009: 1 – 4
[3] HAMADA H, TSUTSUMI T, MATSUZAKI H, et al. 300 GHz band 120 Gbit/s wireless front-end based on InP-HEMT PAs and mixers [J]. IEEE journal of solid-state circuits, 2020: 2316 – 2335
[4] KOENIG S, LOPEZ-DIAZ D, ANTES J, et al. Wireless sub-THz communication system with high data rate [J]. Nature photonics, 2013, 7(12):977 – 981
[5] SARMAH N, GRZYB J, STATNIKOV K, et al. A fully integrated 240-GHz direct-conversion quadrature transmitter and receiver chipset in SiGe technology [J]. IEEE transactions on microwave theory and techniques, 2016, 64(2): 562 – 574
[6] RODRÍGUEZ-VÁZQUEZ P, GRZYB J, HEINEMANN B, et al. A QPSK 110-Gb/s polarization-diversity MIMO wireless link with a 220 – 255 GHz tunable LO in a SiGe HBT technology[J]. IEEE transactions on microwave theory and techniques, 2020, 68(9): 3834 – 3851
[7] CHEN Z. Solid-state terahertz high-speed wireless communication technology [D]. University of Electronic Science and Technology, 2017
[8] LIU J, WU Q Y, WANG Y, et al. A 27 km over sea surface, 500Mbps, real time wireless communication system at 0.14 THz [C]//The 46th International Conference on Infrared, Millimeter and Terahertz Waves (IRMMW-THz), 2021:1 – 2
[9] YU M X, LI G P. Microwave solid-state circuits [M]. Chengdu, China: University of Electronic Science and Technology Press, 2008
[10] NIU Z Q, ZHANG B, ZHOU Z, et al. Terahertz full duplex high-speed communication system [J]. Radio communication technology, 2019, 45(6):643 – 647
[11] ISMAIL M A, WANG Y, YU M. Advanced design and optimization of large scale microwave devices [C]//Microwave Symposium Digest (MTT). IEEE, 2012
[12] FENG Y N, ZHANG B, et al. A 20.8 Gbps dual-carrier wireless communication link in 220 GHz band [J]. China communications, 2021, 18(05):210 – 220
[13] CHEN J, KUYLENSTIERNA D, GUNNARSSON S E, et al. Influence of white lo noise on wideband communication [J]. IEEE transactions on microwave theory and techniques, 2018, 66(7):1 – 11

Biographies

ZHANG Bo (bozhang@uestc.edu.cn) received his BE, MS, and PhD degrees in electromagnetic field and microwave technology from the University of Electronic Science and Technology of China (UESTC) in 2004, 2007, and 2011, respectively. He has been a senior member of IEEE in 2015. He is currently a professor with School of Electronic Science and Engineering, UESTC and Yangtze Delta Region Institute (Huzhou), UESTC. His research interests are terahertz solid state technology and system.

WANG Yihui received her BE degree in electronic engineering from University of Electronic Science and Technology of China (UESTC) in 2017, and is currently working toward her MS degree in terahertz solid-state circuit technology in UESTC, focusing on microwave, terahertz solid-state circuits, and terahertz communication.

FENG Yinian received his BS degree in electronic engineering from University of Electronic Science and Technology of China (UESTC) in 2016, where he is currently pursuing his PhD degree in terahertz solid-state circuit technology. His research interests include microwave, terahertz solid-state circuits, and terahertz communication.

YANG Yonghui received his BS degree in information countermeasure techniques and MS degree in electronic information engineering from Xidian University, China in 2006 and 2010, respectively. Since 2010, he has been with ZTE Corporation, where he focuses on wireless communications. His current research interests include 5G and 6G technology, millimeter-wave and terahertz communication.

PENG Lin received his BS degree in information engineering and MS degree in electromagnetic field and microwave techniques from Nanjing University of Science and Technology, China in 2004 and 2006, respectively. Since 2006, he has been with ZTE Corporation, where he focuses on the research of wireless communications. His current research interests include beyond-5G and 6G technology, millimeter-wave and terahertz communications and intelligent reflecting surface for wireless applications.

Log Anomaly Detection Through GPT-2 for Large Scale Systems

JI Yuhe[1], HAN Jing[2], ZHAO Yongxin[1], ZHANG Shenglin[1], GONG Zican[2]
(1. Nankai University, Tianjin 300071, China;
2. ZTE Corporation, Shenzhen 518057, China)

DOI: 10.12142/ZTECOM.202303010
https://kns.cnki.net/kcms/detail/34.1294.TN.20230802.1731.004.html, published online August 3, 2023
Manuscript received: 2022-12-08

Abstract: As the scale of software systems expands, maintaining their stable operation has become an extraordinary challenge. System logs are semi-structured text generated by the recording function in the source code and have important research significance in software service anomaly detection. Existing log anomaly detection methods mainly focus on the statistical characteristics of logs, making it difficult to distinguish the semantic differences between normal and abnormal logs, and performing poorly on real-world industrial log data. In this paper, we propose an unsupervised framework for log anomaly detection based on generative pre-training-2 (GPT-2). We apply our approach to two industrial systems. The experimental results on two datasets show that our approach outperforms state-of-the-art approaches for log anomaly detection.

Keywords: hybrid beamforming; hybrid architecture; weighted mean square error; manifold optimization; dynamic subarrays

Citation (Format 1): JI Y H, HAN J, ZHAO Y X, et al. Log anomaly detection through GPT-2 for large scale systems [J]. ZTE Communications, 2023, 21(3): 70 - 76. DOI: 10.12142/ZTECOM.202303010
Citation (Format 2): Y. H. Ji, J. Han, Y. X. Zhao, et al., "Log anomaly detection through GPT-2 for large scale systems," *ZTE Communications*, vol. 21, no. 3, pp. 70 - 76, Sept. 2023. doi: 10.12142/ZTECOM.202303010.

1 Introduction

As the scale of software systems expands, maintaining their stable operation has become an extraordinary challenge. System logs are the text generated in the process of software running, which records the status information of the program[1-2]. Traditional log anomaly detection relies on manual analysis by operation engineers. The increase in software systems leads to a surge in a log volume, manual analysis of logs and the design of regular expressions can consume considerable time, and the traditional log analysis is no longer feasible. In recent years, researchers have proposed a variety of automated log anomaly detection methods[3-8]. Early research on log exceptions focused on automatic exception rule generation and simple statistical methods[9]. Then, researchers divided log analyses into three directions: template extraction, model training, and anomaly detection. In recent years, many intelligent log analysis methods have been proposed with the rapid development of machine learning, especially deep neural networks. For example, models based on convolutional neural networks (CNN) and recurrent neural networks (RNN) can effectively capture the sequence characteristics and frequency characteristics of log text, and apply them to the detection of new logs[3-5]. However, the application of the deep learning model in practical scenarios faces the following challenges:

1) The normal pattern of logs is difficult to model. Most existing methods try to learn the normal pattern of log sequence and frequency. The team's O&M engineers pointed out that the semantics of logs in real-world scenarios are of considerable analytical importance. Since the production environment for generating and recording logs is not ideal, relying solely on sequence as well as frequency to encode logs can lead to a large number of false positives.

2) Complex log data contains massive noise. Due to the high concurrency of the software system and the uncertainty of network response, the log generated can be disordered and contain lots of noise.

To tackle the limitations of existing methods, in this paper, we propose an efficient and robust framework for log anomaly detection based on generative pre-training-2 (GPT-2)[10]. Inspired by the excellent performance of GPT-2 in serialization text generation and multiple types of downstream task processing, we leverage this model to capture patterns of normal log sequences.

The main contributions of this paper can be summarized as follows:

1) To tackle the first challenge, we generate sentence vectors for each log template to represent their semantic information. Specifically, we first apply Siamese Bidirectional Encoder Representations from Transformers (SBERT) networks to generate sentence vectors for all log templates and then in-

put the vectors into GPT-2 as the representation of templates. In this way, the models can study both sequential and semantic features of input logs.

2) To address the second challenge, we design an alarm strategy layer for this framework. This step will analyze the statistical characteristics through the existing log data to effectively reduce the impact of noise on model judgment and reduce the false positive rate.

The remainder of this paper is organized as follows. Section 2 introduces the background of log anomaly detection. Section 3 describes our approach. In Section 4, we present our experimental design and results. Section 5 surveys related works and Section 6 concludes this paper.

2 Background

2.1 Logs Description

Logs, which are produced by the running program and contain information generated from the logging module, are a type of semi-structured text. They reflect the running flow and real-time status of the program, and can be used to detect and localize anomalies by operators. Logs consist of a structured part (constant) and an unstructured part (variable), as shown in Fig. 1. A structured part is fixed by the designer at the beginning according to certain designed rules, and can reflect the event of the log, such as the level of logs (e.g., warning and error), the logger name (e.g., root) and so on. This part would not change into the same module. The unstructured part contains specific information on logs and varies according to the input of the program and running status. An unstructured part would reflect the real-time status of the system, and it is essential to use this part to analyze the system and detect anomalies in the system. To make full use of the important features in logs, the semi-structured texts need to be parsed into the structured text with a parsing algorithm. The useless messages in the raw log would be filtered out, and the valuable information would be extracted out from the left information to train the model and detect the anomaly.

Generally speaking, the logs generated by normally operating hardware and software systems have a good regularity. Therefore, some logs that do not match the pattern of previous characteristics are considered anomalous. In the actual dataset studied in this paper, log exceptions can be broadly classified into two categories: 1) Business exceptions caused by network blocking, resource usage, etc. When such exceptions are generated, the program will actively retry the process, so some logs will be repeatedly generated several times in a short period of time. 2) Exceptions triggered by service deployment or termination failure. This type of exception usually generates only a few error logs, which can be evidenced by the fact that the sequence of logs is abnormal, and the semantics of the error logs differ significantly from the normal logs.

```
RAS KERNEL INFO CE sym 0, at 0x0b8580c0, mask 0x10
RAS KERNEL INFO CE sym 25, at 0x04bfb9e0, mask 0x20
RAS KERNEL INFO CE sym 2, at 0x1b85fe80, mask 0x02

RAS MMCS INFO mmcs db server has been started
RAS MMCS INFO mmcs db server has been closed
```

▲ **Figure 1. Log instances, where black words are the structured part and red words are the unstructured part**

2.2 Challenges and Analysis

1) Challenge 1 is modeling the normal patterns of logs. Traditional anomaly detection algorithms always work on learning the normal patterns of logs and finding out logs different from normal patterns. Most studies mainly focus on building the normal pattern according to the sequence and frequency of logs, however, these two features are not comprehensive enough to evaluate the overall state of the system, hence the normal patterns built on these two features are not accurate[3-4]. Intuitively, each log has its semantic characteristics through the log message, and these would describe the log meaning, such as inserting in a dataset, deleting from a dataset, and failure reporting. Besides, logs are generated by triggering corresponding events, and an event always triggers a series of logs. So, there is a sequential relation between logs, which would become different from the usual and could represent the status of logs and systems. As above, if we could combine the semantic information and sequential message with other statistical information, it could perform better in detecting anomalies in logs.

2) Challenge 2 is massive noises in the log. Logs are produced by a software system, which is highly concurrent and greatly influenced by the network. A working software system can run a large number of programs at the same time. These programs can produce a series of logs as well as useless messages, such as test output, and these messages have no effect on evaluating the log status. Some programs need to interact with other programs in the network, and thus the status of the network would affect the log sequence and delays can disrupt the order of logs. The disordered log and useless information are collectively called noise. Dealing with noise properly is necessary to improve anomaly detection in logs, and the simplest way is to remove the noise, which, however, roughly brings lots of missing areas in logs and could bring new problems to the model. Based on the above observation, intuitively, if we can filter out the false positive alarms instead of roughly removing these noises, log anomaly detection can achieve higher accuracy and lower the false alarm rate.

2.3 Preliminaries

1) The log parsing algorithm. To parse the semi-structured log into a structured text, the log parsing algorithm focuses on fetching the unstructured part from the structured part and extracting the features of logs. Specifically, the log parsing algorithm would replace meaningless information (e.g., IP address,

machine ID, etc.) with markers and extract the template of the log to distinguish between logs. In this paper, we adopt Drain[11], a heuristic log parsing algorithm, to parse logs. Drain parses the template of logs by maintaining a tree, which keeps the leaves as log groups with different heuristic rules in internal nodes. New logs would be distributed into log groups in leaves according to the corresponding tree path.

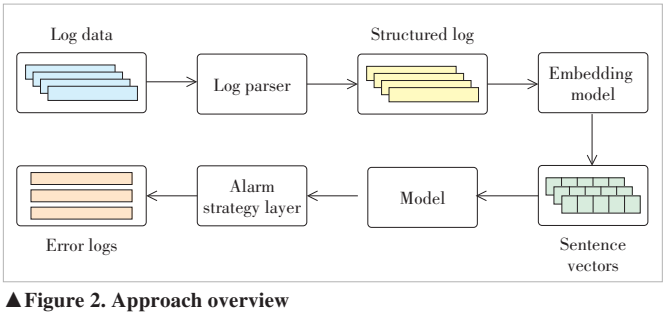
2) GPT-2. The GPT algorithm works based on the decoder of the transformer, and it pre-trains unsupervisedly on massive corpus data and then finetunes the model according to the specific tasks. GPT-2 removes the finetune step and increases the size of the training dataset and model, which makes the model perform well in multi-tasking.

3 Approach

In this section, we will introduce our log anomaly detection framework. Inspired by GPT-2, we employ transformer decoders to encode the normal pattern of system logs. Our framework is divided into four main parts: log parsing, sentence vector generation, model training and detection, and the alerting strategy layer. For Challenge 1, we embed each log template into a vector representing the semantic information. We use vectors instead of tokens as input to GPT-2, so the model can capture both semantic and sequence information to encode normal patterns. For Challenge 2, we design an alarm strategy layer for the framework, which can filter out false positives by the statistical characteristics of log data. The structure of the framework of our work is shown in Fig. 2. The raw logs collected are first transformed into structured logs by the log parser. Then a sentence vector generation model will be used to generate sentence vectors for each extracted log template.

3.1 Log Parser

System logs obtained from the database are semi-structured text, which is difficult to be used for model training. Therefore, before processing logs, we need to use a parser to convert logs into structured text. In this paper, we use Drain as our log parser, which can divide logs into templates and dynamic variables. Drain is an online log template miner, employing a parse tree with a fixed depth, and it can extract templates and variables from a stream of log messages. Drain first sorts logs into different buckets by length and then matches similar portions of the log from front to back in each bucket. Eventually, logs belonging to the same template will end up on the same leaf node. The structure of the depth-fixed tree is shown in Fig. 3. In the figure, sys is an abbreviation for system, and HEX and NUM are variables matched during the parsing process, representing hexadecimal and decimal numbers, respectively.



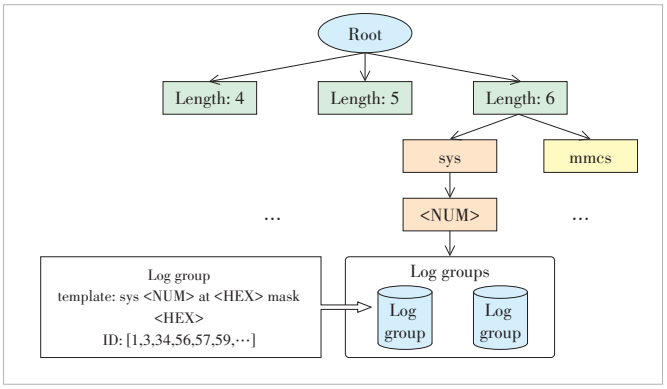
▲Figure 2. Approach overview

3.2 Sentence Vector Generation

To make GPT-2 better at encoding normal patterns, we generate semantically relevant sentence vectors for each log template. We choose SBERT[12] as our embedding model, which has been widely used in text similarity calculation and sentence classification problems and has achieved excellent results. SBERT modifies the pre-trained BERT model, and it implements Siamese or triplet net frameworks to generate semantically meaningful sentence embedding. Sentence vectors generated by templates with similar meanings have smaller cosine distances or Euclidean distances. Table 1 shows the Euclidean distance of sentence vectors in five log templates.

3.3 Detection Model

1) Model framework. We choose GPT-2 as our log anomaly detection model in this work. GPT-2 is an unsupervised Natural Language Processing (NLP) model stacked from the transformer's decoders. The structure of GPT-2 is shown in Fig. 4. Each decoder has a masked self-attention layer, a feed-forward neural network, and two normal layers. The structure of each decoder is the same, but each module maintains separate parameters. Different from the ordinary self-attention layer, the masked self-attention layer does not allow a node to

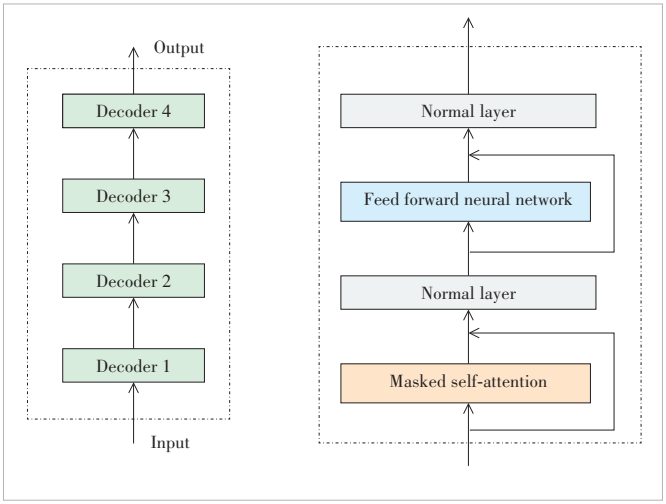


▲Figure 3. Structure of the depth-fixed tree

▼Table 1. Euclidean distance of sentence vectors of similar semantic templates

Templates	Euclidean Distance
httprequest except <*> permission denied	-
httprequest except <*> <*> permission denied	0.147 629 340 284 133 4
httprequest except <*> no such file or directory	0.595 852 332 701 891 4
httprequest except <*>	0.621 201 472 867 456 3
httprequest except EoF occurred in violation of protocol	0.838 852 193 154 771 3
httprequest except <*> connection reset by peer	0.880 359 580 380 884 6

EoF: end-of-file



▲Figure 4. Structure of generative pre-training-2 (GPT-2)

get information from subsequent nodes. This feature makes the model perform well in sequence prediction tasks.

2) Input representation. Like most NLP models, GPT-2 looks up the embedding vector corresponding to the word from the embedding matrices, and the embedding matrices are also part of the model training results. GPT-2 maintains two embedding matrices: a token embedding matrix and a position embedding matrix. Each row of the token embedding matrix is a vector that represents a token, and these vectors are randomly initialized and continuously adjusted during training. Each row of the position embedding matrix represents the position information of the token. Tokens in the same position in different sentences have the same position embedding vector. Each template embedding inputted into GPT-2 is the sum of position embedding and token embedding. In our framework, we initialize the token embedding matrix with sentence vectors generated by SBERT, which unlike the original random initialization, can make the model catch the initial semantic information of log templates more accurately. Besides, this approach gives operations more control over the model. We can adjust the dimension or generation method of sentence vectors according to the characteristics of logs and actual business requirements.

3) Model training. The core concept of GPT-2 is language modeling. Language modeling refers to distribution estimation from a group of unsupervised samples (x_1, x_2, x_3, ⋯). Each sample consists of symbol sequences of variable length (s_1, s_2, s_3, ⋯). Because there are explicit sequential relationships between phrases in natural languages, language modeling typically decomposes the joint probability of symbols as a product of conditional probabilities.

$$p_{(x)} = \prod_{i=1}^{n} p\left(s_i | s_1, s_2, \cdots, s_{n-1}\right). \tag{1}$$

Transforming the above equation into a logarithmic form, the goal of the language model is to maximize the probability of the following equation.

$$p(x) = \sum_{i=1}^{n} \log p\left(s_i | s_1, s_2, \cdots, s_{n-1}; \theta\right), \tag{2}$$

where n is the length of the language sequence, and the conditional probability p is modeled by the neural network with parameter θ. These parameters are trained by the stochastic gradient descent.

The input of the decoder at the first layer consists of the token embedding vector and position embedding vector of log templates. The output of each decoder is processed from the previous layer's output. The output probability obtained by the model can be expressed as follows:

$$\begin{aligned} h_0 &= \boldsymbol{X}\boldsymbol{W}_e + \boldsymbol{W}_p, \\ h_l &= \text{Decoder}\left(h_{l-1}\right) \ \forall l \in [1, n], \\ p(x) &= \text{softmax}(h_n \boldsymbol{W}_e^T), \end{aligned} \tag{3}$$

where h_l represents the output of the l-th layer decoder, $\boldsymbol{X}$ is the matrix composed of the unique thermal encoding of the input log sequence, $\boldsymbol{W}_e$ is the token embedding matrix, and $\boldsymbol{W}_P$ is the position embedding matrix. To maximize the prediction probability, the model adjusts the decoder parameters of each layer during the learning process.

3.4 Alarm Strategy

Analyzing the actual data, we find that in the production environment, the logs printed by the machine are not always sequential. The main differences between these noises and severe systems are as follows: logs corresponding to noise appear frequently and usually have a certain seasonality, while severe anomalies occur infrequently and are difficult to predict. Based on the above observation, we use frequency and periodicity as criteria to determine whether model error reporting is noisy or a true anomaly.

For the abnormal log templates detected by the model, we first calculate the time interval of their occurrence in the training data and then use the auto-correlation coefficient to analyze whether the time interval of template occurrence has a specific pattern. The auto-correlation coefficient is a common parameter for finding repetitive patterns (e.g., periodic signals masked by noise) and is often used in signal processing problems. The sequence consisting of the auto-correlation coefficients is known as the auto-correlation function. For the obtained template interval sequence, its autocorrelation function is calculated, and the peak of the function is the possible period of the corresponding template. If the value of the function at a certain time is higher than a threshold value set based on expert experience, we consider the template to be periodic and its basic impossibility to be an error log template.

Then, for non-periodic templates, we go through the statis-

tics of their daily frequency of occurrence. Based on the observation of the data and the communication with the operation and maintenance staff, we have learned that the probability of daily serious anomalies in a smoothly running system is extremely low. An exception log with a high frequency is often caused by minor errors such as network blocking and data locking. The system can often recover from such errors quickly, so such alarms are not necessary. Therefore, for exception log templates that occur more frequently than a certain threshold, the alert policy layer will filter them out as less serious exception logs. The selection of this frequency threshold is strongly correlated with the type of machine logs and relies on the involvement of business experts.

In this layer, error logs detected by GPT-2 will be analyzed, and if their characteristics are more like noise, the exception will not be reported. This step can greatly reduce the model false positives caused by noise, improve the accuracy of the log anomaly detection framework, and reduce the disturbance to operation engineers.

4 Experiment and Evaluation

4.1 Experimental Setup

1) Research questions. In this section, we evaluate the performance of our framework with the following research questions (RQs):

a) RQ1: How effective is our framework in log-based anomaly detection?

b) RQ2: How effective is the sentence vector generation and alerting strategy layer in improving the effectiveness of the model.

2) Datasets. We evaluate our approach on two large-scale systems called Ada and Bob. Ada is a framework for microservice deployment applications. Bob is a hardware network consisting of a large number of switch systems. The statistics of the datasets are shown in Table 2.

3) Baselines. We compare our framework with two baselines, LogAnomaly[4] and NeuralLog[8]. LogAnomaly is a log anomaly detection method based on a long short-term memory (LSTM) network. LogAnomaly first uses a Frequent Term Tree (FT-Tree) to analyze the semi-structured log text. Then, Template2Vec is implemented to generate vectors for each log template. Finally, the vectors representing log semantics and frequency are input into the LSTM to enable the model to learn the normal pattern of logs. NeuralLog is a novel log-based anomaly detection framework. Different from the traditional process, the algorithm does not require template parsing. NeuralLog generates semantic vectors for row logs. These representation vectors are then used to detect anomalies through a transformer-based classification model.

4) Evaluation metrics. We use Precision, Recall, and F1-Score (F1S) as our evaluation metrics, which are defined as follows. True Positive (TP) is the number of abnormal logs that are correctly detected by the model, False Positive (FP) is the number of normal logs that are wrongly identified as anomalies, and False Negative (FN) is the number of abnormal logs that are not detected by the model.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} . \tag{4}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} . \tag{5}$$

$$\text{F1S} = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision} + \text{Recall}} . \tag{6}$$

4.2 Experimental Results

In this section, we will give response to the RQs mentioned above.

1) RQ1: How effective is our framework in log-based anomaly detection?

In this RQ, we evaluate whether our framework can work effectively on logs generated in the production environment. We compare our framework with two baselines: LogAnomaly[4] and NeuralLog[8].

Table 3 shows the results of our method as well as two baselines on Ada and Bob. Both LogAnomaly and NeuralLog show poor Precision and Recall performance on Ada. LogAnomaly has very limited learning capability due to the limitation of model size, which makes it difficult to obtain good results on log datasets with a large number of templates and complex processes. Also, in the production environment logs, log templates that are not present in the training data often appear in the test set, making LogAnomly generate a large number of false positives often[18]. NeuralLog tends to consider every log unlikely to be anomalous on large unsupervised datasets due to the problem of data dilution. Our framework analyzes and learns the actual semantics of the logs, and designs an alert policy layer that incorporates the actual business characteristics of the machine. These measures make our model more capable of capturing the normal patterns of real logs in a produc-

▼Table 2. Statistics of evaluation datasets

Dataset	Training Data	Number of Templates	Test Dataset	
			Normal	Anomalous
Ada	6 626 865	599	7 911 944	2 648
Bob	7 021 577	84	1 067 850	904

▼Table 3. Evaluation results of our method vs the other two methods

Approach	Ada			Bob		
	Precision	Recall	F1S	Precision	Recall	F1S
LogAnomaly	0.394	0.190	0.256	0.353	0.332	0.342
NeuralLog	0.297	0.354	0.323	0.638	0.872	0.736
Our method	0.738	1.00	0.850	0.857	1.00	0.923

tion environment.

The complexity of Bob is much less than that of Ada, mainly because of less templates and a relatively fixed log sequence. However, due to problems such as network latency and data washout, the logs generated by the switch have a large number of errors and retry messages. In most cases, these error messages are not of concern because the program can be restored to normal after several retries and will not have a significant impact on the execution of the business. Also, these alarms of variable duration can disrupt the log sequence, making it more difficult to learn the normal pattern of the logs purely from frequency or sequence. The sequence confusion greatly affects the learning ability of LogAnomaly on this dataset, and NeuralLog has a lower Precision due to the report of the unimportant exceptions. Our framework captures the normal characteristics of logs from multiple perspectives and designs alerting policies based on the frequency and periodicity of logs, thus achieving good performance on the Bob dataset.

The experimental results show that our framework has good performance for complex log datasets in practice. The analysis of log semantics makes the model show good robustness on poor stability sequences, and the addition of the alert policy layer reduces the model's disturbance to engineers and screens out some of the exceptions that can be fixed automatically.

2) RQ2: How effective is the sentence vector generation and alerting strategy layer in improving the effectiveness of the model?

As mentioned in the previous sections, we add the sentence vector matrix as well as the alert alarm strategy layer to GPT-2. In this RQ, we will verify the effect of the main parts of the framework on its effectiveness, by removing one or two components, namely the sentence vector (SV) and the alarm strategy (AS).

OM w/o SV & AS: We remove the sentence vector generation and alert policy layers from our framework. That is, we use only the GPT-2 model for sequence prediction to diagnose anomaly logs.

OM w/o AS: We remove the alarm strategy generation from our framework.

OM w/o SV: We remove the sentence vector layer from our framework.

OM: Anomaly detection work on logs using the completed framework proposed in this paper.

The experimental results in Table 4 show that the sentence vector generation part of the framework can improve the accuracy of the model to some extent and greatly enhance Recall. Because GPT-2 achieves better results in capturing the semantic information of normal and abnormal templates after receiving the sentence vectors of the templates as prior knowledge. This is demonstrated by the fact that log templates containing the same abnormal keywords, the vector representations of which have a closer distance, are easily detected together in the anomaly detection stage. At the same time, log templates that symbolize normal patterns are more difficult for the model to detect as false positives because they often contain positive-meaning words. Since the alarm strategy layer is built based on expert experience and has accurate filtering rules, it can filter out a large number of false abnormal logs and effectively improve the Recall of the model.

With the inclusion of both components, the effectiveness of our framework has been significantly improved. For complex logging environments, more accurate exception identification, very low FPs and a fairly high Recall can be achieved.

5 Related Work

As a kind of operational data, logs are widely used for system anomaly detection in practice. To take advantage of the logs, previous work mainly focuses on detecting abnormal logs with artificial rules, which is not appropriate in scenarios with a large number of logs[5]. And deep learning is widely used in automatic anomaly detection in logs. DeepLog[3] uses LSTM to learn the normal pattern of the system and predict the next log template by log sequence. LogAnomaly[5] uses a word embedding model to mine semantic information of log templates and learn the sequential patterns and quantitative relationships for logs with LSTM. LogRobust[13] represents the semantic information of the log by word vector and takes advantage of the bi-directional LSTM to learn the normal pattern of the log. OneLog[14] merges components (such as parsers and classifiers) into a deep neural network to detect log anomalies. LogMerge[15] learns the semantic similarity of multi-syntax logs to realize the transfer of log exception patterns across log types, which greatly reduces the overhead of exception annotation. Transformers could also be used to represent the semantics of the log and model the log sequence, and anomalies would be detected with learned information[10, 15–17]. GPT-2[10] is proposed for unsupervised learning of text information based on transformers and performs well on text generation, text classification, semantic judgment, etc.

6 Conclusions and Future Work

In practical scenarios, large-scale systems produce logs that are different from the vast majority of laboratory open-source datasets. The log sequence is more complex, and normal patterns are harder to capture. In this work, we introduced a way

▼Table 4. Experimental results

Approach	Ada			Bob		
	Precision	Recall	F1S	Precision	Recall	F1S
OM w/o SV & AS	0.128	0.835	0.222	0.510	0.940	0.661
OM w/o AS	0.427	1.00	0.598	0.718	1	0.836
OM w/o SV	0.627	0.807	0.705	0.833	0.940	0.883
OM	0.738	1.00	0.850	0.857	1.00	0.923

AS: alarm strategy SV: sentence vector
OM: our method

to input semantic analysis data into GPT-2 and designed an alarm strategy layer. Through these ways, we improved the complex log sequence learning ability of the model and reduced the noise effect on the model prediction. Experimental results on two industrial datasets have shown that the false alarm rate of the model is significantly reduced, and our framework shows good performance in the actual operation scenario.

In the future, we will continue to improve the performance of the model on multiple datasets and reduce the dependence of the alarm strategy layer on expert experience.

References

[1] ZHANG S L, LIU Y, PEI D, et al. Rapid and robust impact assessment of software changes in large Internet-based services [C]//The 11th ACM Conference on Emerging Networking Experiments and Technologies. ACM, 2015: 1 – 13. DOI: 10.1145/2716281.2836087
[2] ZHU J M, HE S L, LIU J Y, et al. Tools and benchmarks for automated log parsing [C]//IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP). IEEE, 2019: 121 – 130. DOI: 10.1109/ICSE-SEIP.2019.00021
[3] DU M, LI F F, ZHENG G N, et al. DeepLog: anomaly detection and diagnosis from system logs through deep learning [C]//ACM SIGSAC Conference on Computer and Communications Security. ACM, 2017: 1285 – 1298. DOI: 10.1145/3133956.3134015
[4] MENG W B, LIU Y, ZHU Y C, et al. LogAnomaly: unsupervised detection of sequential and quantitative anomalies in unstructured logs[C]//The 28th International Joint Conference on Artificial Intelligence. ACM, 2019, 19(7): 4739 – 4745
[5] ZHANG X, XU Y, LIN Q W, et al. Robust log-based anomaly detection on unstable log data [C]//The 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. ACM, 2019: 807 – 817. DOI: 10.1145/3338906.3338931
[6] EKELHART A, EKAPUTRA F J, KIESLING E. The SLOGERT framework for automated log knowledge graph construction [C]//European Semantic Web Conference. ESWC, 2021: 631 – 646. DOI: 10.1007/978-3-030-77385-4_38
[7] GUO H X, YUAN S H, WU X T. LogBERT: log anomaly detection via BERT [C]//Proceedings of 2021 International Joint Conference on Neural Networks (IJCNN). IEEE, 2021: 1 – 8. DOI: 10.1109/IJCNN52387.2021.9534113
[8] LE V H, ZHANG H Y. Log-based anomaly detection without log parsing [C]//The 36th IEEE/ACM International Conference on Automated Software Engineering. ACM, 2021: 492 – 504. DOI: 10.1109/ASE51524.2021.9678773
[9] HE S L, HE P J, CHEN Z B, et al. A survey on automated log analysis for reliability engineering [J]. ACM computing surveys, 54(6): 1 – 37. DOI: 10.1145/3460345
[10] RADFORD A, WU J, CHILD R, et al. Language models are unsupervised multitask learners [EB/OL]. [2023-03-10]. https://www.semanticscholar.org/paper/Language-Models-are-Unsupervised-Multitask-Learners-Radford-Wu/9405cc0d6169988371b2755e573cc28650d14dfe
[11] HE P J, ZHU J M, ZHENG Z B, et al. Drain: an online log parsing approach with fixed depth tree [C]//IEEE International Conference on Web Services (ICWS). IEEE, 2017: 33 – 40. DOI: 10.1109/ICWS.2017.13
[12] REIMERS N, GUREVYCH I. Sentence-BERT: sentence embeddings using Siamese BERT-networks [C]//Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Association for Computational Linguistics, 2019: 3982 – 3992. DOI: 10.18653/v1/d19-1410
[13] HASHEMI S, MÄNTYLÄ M. OneLog: towards end-to-end training in software log anomaly detection [EB/OL]. [2022-12-12]. https://arxiv.org/abs/2104.07324
[14] CHEN R, ZHANG S L, LI D W, et al. LogTransfer: cross-system log anomaly detection for software systems with transfer learning [C]//IEEE 31st International Symposium on Software Reliability Engineering. IEEE, 2020: 37 – 47. DOI: 10.1109/ISSRE5003.2020.00013
[15] HUANG S H, LIU Y, FUNG C, et al. HitAnomaly: Hierarchical transformers for anomaly detection in system log [J]. IEEE transactions on network and service management, 2020, 17(4): 2064 – 2076. DOI: 10.1109/TNSM.2020.3034647
[16] YANG H T, ZHAO X, SUN D G, et al. Sprelog: log-based anomaly detection with self-matching networks and pre-trained models [C]//International Conference on Service-Oriented Computing. 2021: 736 – 743
[17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. ACM, 2017: 6000 – 6010. DOI: 10.5555/3295222.3295349
[18] LE V H, ZHANG H Y. Log-based anomaly detection with deep learning: how far are we? [C]//IEEE/ACM 44th International Conference on Software Engineering (ICSE). IEEE, 2022: 1356 – 1367

Biographies

JI Yuhe received his bachelor's degree in software engineering from the College of Software, Nankai University, China in 2022. He is now pursuing his master's degree at the School of Software, Nankai University. His research interests include anomaly detection and natural language processing.

HAN Jing (han.jing28@zte.com.cn) received her master's degree from Nanjing University of Aeronautics and Astronautics, China. She has been with ZTE Corporation since 2000. She had been engaged in 3G/4G key technologies, from 2000 to 2016, and has become a technical director responsible for intelligent operation of cloud platforms and wireless networks since 2016. Her research interests include machine learning, data mining, and signal processing.

ZHAO Yongxin received her bachelor's degree in software engineering from Nankai University, China in 2021. She is currently pursuing her master's degree at the School of Software, Nankai University. Her research interests include anomaly detection and failure diagnosis.

ZHANG Shenglin received his BS degree in network engineering from the School of Computer Science and Technology, Xidian University, China in 2012 and PhD degree in computer science from Tsinghua University, China in 2017. He is currently an associate professor with the College of Software, Nankai University, China. His current research interests include failure detection, diagnosis and prediction for service management. He is an IEEE Member.

GONG Zican received his master's degree in professional computing and artificial intelligence from the Australian National University in 2019. He has been a machine learning engineer in ZTE Corporation since 2020. His research interests include machine learning, professional computing and system architecture.

Robust Beamforming Under Channel Prediction Errors for Time-Varying MIMO System

ZHU Yuting[1], LI Zeng[2,3], ZHANG Hongtao[1]

(1. Key Lab of Universal Wireless Communications, Ministry of Education of China, Beijing University of Posts and Telecommunications, Beijing 100876, China;
2. ZTE Corporation, Shenzhen 518057, China;
3. State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518055, China)

DOI: 10.12142/ZTECOM.202303011

https://link.cnki.net/urlid/34.1294.TN.20230830.1912.007, published online August 31, 2023

Manuscript received: 2022-12-23

Abstract: The accuracy of acquired channel state information (CSI) for beamforming design is essential for achievable performance in multiple-input multiple-output (MIMO) systems. However, in a high-speed moving scene with time-division duplex (TDD) mode, the acquired CSI depending on the channel reciprocity is inevitably outdated, leading to outdated beamforming design and then performance degradation. In this paper, a robust beamforming design under channel prediction errors is proposed for a time-varying MIMO system to combat the degradation further, based on the channel prediction technique. Specifically, the statistical characteristics of historical channel prediction errors are exploited and modeled. Moreover, to deal with random error terms, deterministic equivalents are adopted to further explore potential beamforming gain through the statistical information and ultimately derive the robust design aiming at maximizing weighted sum-rate performance. Simulation results show that the proposed beamforming design can maintain outperformance during the downlink transmission time even when channels vary fast, compared with the traditional beamforming design.

Keywords: time-varying channels; time-division duplex; robust beamforming; channel prediction errors; weighted sum-rate maximization

Citation (Format 1): ZHU Y T, LI Z, ZHANG H T. Robust beamforming under channel prediction errors for time-varying MIMO system [J]. ZTE Communications, 2023, 21(3): 77 - 85. DOI: 10.12142/ZTECOM.202303011
Citation (Format 2): Y. T. Zhu, Z. Li, and H. T. Zhang, "Robust beamforming under channel prediction errors for time-varying MIMO system," *ZTE Communications*, vol. 21, no. 3, pp. 77 - 85, Sept. 2023. doi: 10.12142/ZTECOM.202303011.

1 Introduction

The rapid growth of intelligent devices and applications generates large demands for high data rate transmission. Massive multiple-input multiple-output (MIMO) technique, as a key technology in the fifth generation of wireless networks, provides huge potential capacity gain by employing multiple antennas and exploiting the extra degree of freedom without extending the extra bandwidth[1]. The achievable performance heavily relies on the exploitation of spatial multiplexing gain, as well as enough channel knowledge of the base station[2]. The former is tightly related to multi-user MIMO (MU-MIMO) communications and appropriate beamforming design to enhance the intended signal and suppress unintended interference. As to the latter, in a time-division duplex (TDD) mode, the strategy used to obtain channel state information (CSI) is reciprocity, where the transmitter uses uplink (UL) channel information to speculate the downlink (DL) channel state in the next transmission interval[3]. However, this strategy cannot face the channel aging problem, especially in a high-mobility environment where the channel characteristic is varying fast during the transmission period, which leads to a serious impact on the performance of beamforming in the MIMO system.

Traditional beamforming algorithms, such as classic zero-forcing (ZF) and weighted minimum mean-squared-error (WMMSE), actually work well only when accurate and instantaneous CSI is obtained. Note that the WMMSE precoder[4 - 5] is designed according to the sum-rate maximization criterion so it performs better than ZF. In the time-varying TDD system, the channel estimation for multiple users is based on their sounding reference signals (SRS), which means we can only obtain the accurate CSI of the current SRS transmission time slot while cannot measure the channel state of the intermediate DL time slots in the case of a high-speed moving scene.

This work was supported by the ZTE Industry-University-Institute Cooperation Funds under Grant No. 2021ZTE01-03.

Thus, the acquired CSI is inevitably outdated, leading to outdated beamforming design and then performance degradation. Considering the CSI used for DL transmission is not always perfect in practice, several studies of robust beamforming have been carried out[6 - 7]. Ref. [6] introduced an iterative algorithm to maximize the linear assignment weighted sum rate with statistical CSI under channel fading conditions. In Ref. [7], a robust coordinated beamforming method is proposed to maintain the weighted sum-rate performance with inaccurate CSI. The channel aging effect still weakens the robust benefit dramatically[8], so more accurate CSI is needed from the viewpoint of CSI, which can be achieved by channel prediction technology.

Channel prediction technology is an effective way to combat the detrimental effects of channel aging and keep the track of the time-varying channel[9], which predicts the future channels by exploiting the potential temporal correlation from those in the past[10]. There are various prediction methods from autoregressive (AR) model-based methods to deep learning[11]. The AR model is commonly used, which treats the time-varying channel as a wide-sense stationary stochastic process[9, 12], and deep learning requires a mass of samples for training[13]. Besides, by utilizing the time and spatial properties of channels, the authors in Ref. [14] proposed an angle-domain channel tracking scheme for high-speed railway communications, and spatial-temporal sparse structures are further exhibited to propose a novel estimation and prediction scheme for time-varying massive MIMO channels[15]. Furthermore, to solve the problem of tracking the nonlinear channels, the kernel recursive least squares (KRLS)[16 - 17] algorithm is proposed, mapping the channel sample space to the high-dimensional space, and it shows great adaptiveness.

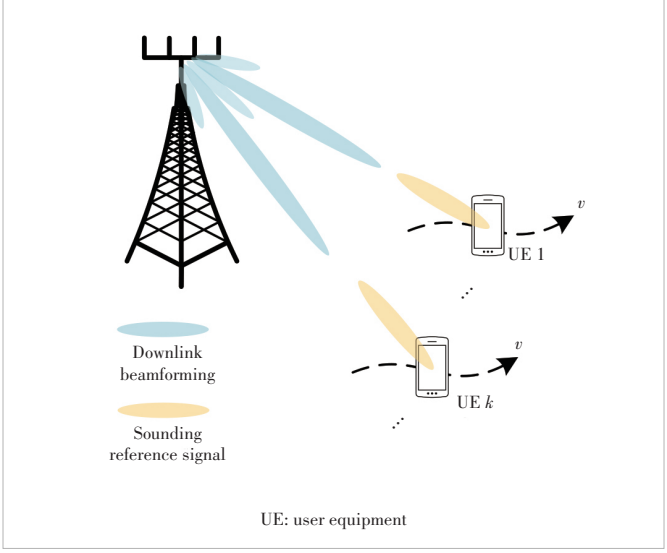
Based on the predicted CSI, linear or nonlinear interpolation can be employed for predicting the DL channel between the current channel and the predicted one[18]. Ref. [19] evaluated the benefit of channel prediction in adaptive beamforming systems and showed the tolerable Doppler spread increase for a given outage performance by using prediction. However, there still exist errors in the channel prediction, which becomes more obvious as the mobility increases[20]. The existence of errors between the acquired CSI and the real-time CSI will result in the performance error cost since the designed beamforming does not match the real channel. However, most of the existing works focus on the impact that the channel prediction error brings[21], while no robust beamforming for improving the performance under channel prediction errors is considered. Moreover, the channel prediction errors are often constructed as complex Gaussian random variables with independent and identically distributed (i.i.d.), zero mean and unit variance entries, which does not consider the statistical characteristics.

Motivated by the above, we propose a robust beamforming design under channel prediction errors based on channel prediction for the time-varying MIMO system in this paper. The main contributions of this work are summarized as follows. 1) A framework of TDD system for beamforming based on channel prediction is introduced, with a novel prediction error model that combines channel statistical information and a Gaussian random matrix, including the joint correlation properties of realistic massive MIMO channels. 2) To maximize the sum-rate performance, we propose a robust beamforming design based on channel prediction, utilizing the deterministic equivalents method to explore further potential beamforming gain brought by the static statistical information, breaking through the performance limitation brought by the channel prediction errors. 3) The beamforming gains compared with traditional benchmarks under different mobile scenes are validated by simulation, and we investigate the performance during the half-frame period to reflect the loss caused by the time-varying effect.

The rest of this paper is organized as follows. Section 2 presents the proposed framework of the system including the system model and the formulation of the discussed problem. Section 3 introduces the proposed robust beamforming design. Section 4 provides the simulation results and performance discussion. Finally, we conclude this paper in Section 5.

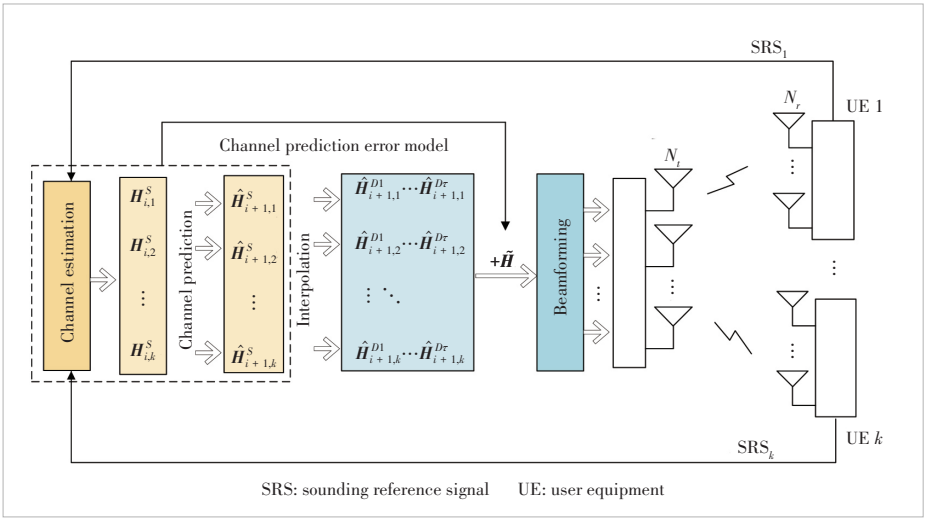
2 System Model and Problem Formulation

In this paper, we focus on the TDD system, and the CSI used for DL transmission is obtained depending on the channel reciprocity. As shown in Fig. 1, we consider a MU-MIMO system where one base station (BS) equipped with N_t transmit antennas serves K pieces of mobile user equipment (UE), and each is equipped with N_r receive antennas. Denote $\boldsymbol{H}_{i,k}^{S} \in \mathbb{C}^{N_r \times N_t}$ as the CSI matrix spanning from the BS to the k-th piece of UE



▲ **Figure 1. Illustration of a time-division duplex (TDD) multi-user multiple-input multiple-output (MU-MIMO) system where v denotes the speed of mobile UE**

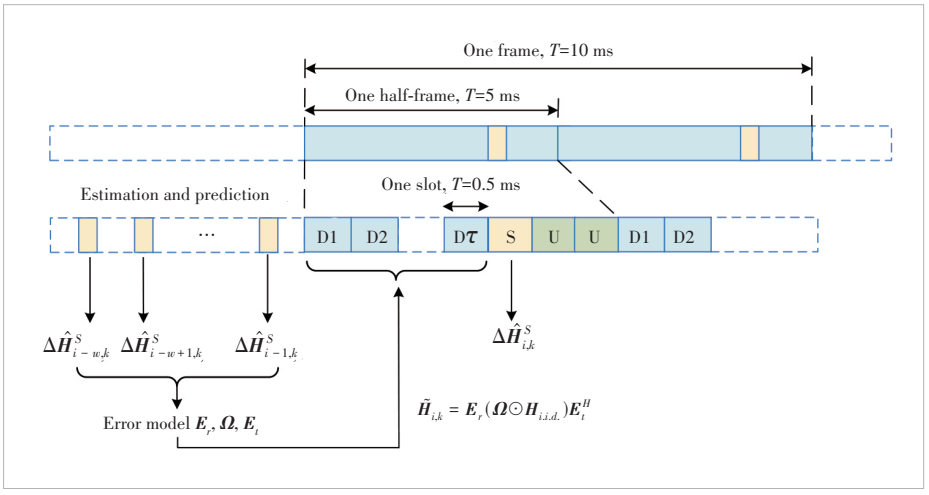
obtained through SRS in the i-th half-frame and $\boldsymbol{H}_{i,k}^{Dt} \in \mathbb{C}^{N_r \times N_t}$ as the CSI matrix for DL transmission in the t-th DL time slot of the i-th half-frame period. As illustrated in Fig. 2, the whole procedure of channel acquisition can be divided into three steps, which are channel estimation, channel prediction and interpolation. Based on the channel prediction algorithm and interpolation for the DL channel between the currently estimated channel and the predicted one, $\boldsymbol{H}_{i,k}^{Dt}$ used for DL beamforming depends on the predicted DL channel matrix $\hat{\boldsymbol{H}}_{i,k}^{Dt}$, and the channel prediction error is considered further.



▲Figure 2. System model for proposed beamforming scheme

2.1 Channel Prediction

In the TDD system, firstly, by exploiting the reciprocity between UL and DL, we can obtain the uplink channel information by channel estimation such as minimum mean square error (MMSE) and least square (LS), and use them to calculate parameters including beamforming for DL transmission. The channel estimation is to diminish the effect of noise and extract the channel model from received data. In this paper, we assume the channel estimation is perfect and we can obtain the estimated CSI for UE $\boldsymbol{H}_{i,1}^{S}, \boldsymbol{H}_{i,2}^{S}, \cdots, \boldsymbol{H}_{i,K}^{S}$ through the processing of SRS.



▲Figure 3. Channel prediction error model and time slot structure

Then, based on a certain number of past CSI samples, we can predict the future state of channels for UE $\hat{\boldsymbol{H}}_{i+1,1}^{S}, \hat{\boldsymbol{H}}_{i+1,2}^{S}, \cdots, \hat{\boldsymbol{H}}_{i+1,K}^{S}$ by exploiting the correlation characteristics of the channels. In this paper, to track nonlinear channels and predict the time-varying channel, we adopt a sparse sliding-window KRLS algorithm, where the sample set is updated dynamically based on the correlation analysis among samples. And it is to be mentioned that the proposed beamforming scheme in the following is not limited to a certain prediction method.

During the special time slot for SRS, with the currently estimated channel and predicted one $\hat{\boldsymbol{H}}_{i+1,k}^{S}$, the next step is to interpolate the DL channel between them and here we take the linear interpolation method. Define τ and N_{slot} as the number of DL time slots and that of the total slots in one half-frame, respectively. Assuming that besides the DL time slots there is one SRS slot and two UL slots, with the detail of the time slot structure shown in Fig. 3, the deployed linear interpolation is expressed as

$$\hat{\boldsymbol{H}}_{i+1,k}^{Dt} = \boldsymbol{H}_{i,k}^{S} + (t+2) \cdot \frac{\hat{\boldsymbol{H}}_{i+1,k}^{S} - \boldsymbol{H}_{i,k}^{S}}{N_{\text{slot}}}, \quad 1 \leqslant t \leqslant \tau . \tag{1}$$

2.2 Channel Prediction Error Model

Considering that the predicted channels obtained by channel prediction techniques still exist errors inevitably, compared with the real channels, we take channel prediction errors into consideration and introduce an error model. With $\boldsymbol{H}_{i,k}^{S}$ denoting the perfect estimated CSI for the k-th piece of UE in the i-th SRS period, the former channel prediction errors collected can be described as

$$\Delta\boldsymbol{H}_{i,k}^{S} = \boldsymbol{H}_{i,k}^{S} - \hat{\boldsymbol{H}}_{i,k}^{S}, \tag{2}$$

and a sliding-window size w is set for collection, as shown in Fig. 3. By exploiting the statistical characteristics of the historical channel prediction error sample set $\left[\Delta\boldsymbol{H}_{i-w,k}^{S}, \Delta\boldsymbol{H}_{i-w+1,k}^{S}, \cdots, \Delta\boldsymbol{H}_{i-1,k}^{S}\right]$, we use the jointly corre-

lated channel model to describe the potential correlations of each channel prediction error in the past slots and express the error model for the k-th piece of UE in the DL time slots of the i-th half-frame period,

$$\tilde{\boldsymbol{H}}_{i,k} = \boldsymbol{E}_r(\boldsymbol{\Omega} \odot \boldsymbol{H}_{\text{i.i.d.}})\boldsymbol{E}_t^H . \tag{3}$$

The statistical information $\boldsymbol{E}_r$, $\boldsymbol{E}_t$ and $\boldsymbol{\Omega}$ are gathered from practical error samples for the UE, where $\boldsymbol{E}_r$ and $\boldsymbol{E}_t$ are deterministic unitary matrices that reveal the correlation of the transmit and receive antennas respectively, while each element of $\boldsymbol{\Omega}$ is the square root of the element of the eigenmode channel coupling matrix. $\boldsymbol{H}_{\text{i.i.d.}}$ is a random matrix with zero mean and unit variance entries, following i.i.d. For illustration purposes, we depict the error model with the time slot structure in Fig. 3.

For simplicity, we use $\boldsymbol{H}_k$ as a substitute for $\boldsymbol{H}_{i,k}^{Dl}$ in the following part to denote the CSI matrix for the k-th UE considering the beamforming design for only one typical DL transmission time slot. Based on the analysis mentioned above, the channel matrix for UE can be divided into two parts: the determined prediction channel matrix $\hat{\boldsymbol{H}}_k$ and the random error matrix $\tilde{\boldsymbol{H}}_k$.

2.3 Problem Formulation

Denote the linear precoding matrix for the k-th piece of UE as $\boldsymbol{V}_k \in \mathbb{C}^{N_t \times 1}$. The receive signal of the k-th piece of UE $\boldsymbol{y}_k$ at this DL time slot is

$$\boldsymbol{y}_k = \boldsymbol{H}_k\boldsymbol{V}_k\boldsymbol{s}_k + \sum_{m=1, m\neq k}^{K} \boldsymbol{H}_k\boldsymbol{V}_m\boldsymbol{s}_m + \boldsymbol{n}_k , \tag{4}$$

where the transmit signal to the k-th piece of UE $\boldsymbol{s}_k$ is independent random variables with zero mean and unit variance; $\boldsymbol{n}_k$ represents the additive white Gaussian noise with the distribution $\text{CN}(0, \sigma^2\boldsymbol{I}_{N_r})$. Then, the achievable data rate for UE at a DL time slot can be expressed as

$$R_k = \mathbb{E}\left\{\log\det\left(\boldsymbol{I} + \boldsymbol{H}_k\boldsymbol{V}_k\boldsymbol{V}_k^H\boldsymbol{H}_k^H\left(\sigma^2\boldsymbol{I} + \sum_{m=1, m\neq k}^{K} \boldsymbol{H}_k\boldsymbol{V}_m\boldsymbol{V}_m^H\boldsymbol{H}_k^H\right)^{-1}\right)\right\} . \tag{5}$$

To simplify the objective, we treat the aggregate interference-plus-noise as the Gaussian noise for UE and denote its covariance matrix as

$$\boldsymbol{Q}_k = \sigma^2\boldsymbol{I} + \mathbb{E}\left\{\sum_{m=1, m\neq k}^{K} \boldsymbol{H}_k\boldsymbol{V}_m\boldsymbol{V}_m^H\boldsymbol{H}_k^H\right\} . \tag{6}$$

Then the achievable rate for UE at a DL time slot can be rewritten as

$$R_k = \mathbb{E}\left\{\log\det\left(\boldsymbol{I} + \boldsymbol{H}_k\boldsymbol{V}_k\boldsymbol{V}_k^H\boldsymbol{H}_k^H\boldsymbol{Q}_k^{-1}\right)\right\} . \tag{7}$$

Furthermore, the achievable weighted sum rate of the system can be expressed as

$$R = \frac{\tau}{N_{\text{slot}}}\sum_{k=1}^{K}\alpha_k R_k . \tag{8}$$

Aiming at optimizing the beamforming matrix to maximize the expected weighted sum rate at a typical DL slot, we can express the optimization problem as

$$\begin{aligned} &\max_{\{\boldsymbol{V}_k\}} \sum_{k=1}^{K}\alpha_k R_k , \\ &\text{s.t.} \sum_{k=1}^{K}\text{tr}\left(\boldsymbol{V}_k\boldsymbol{V}_k^H\right) \leqslant P , \end{aligned} \tag{9}$$

where $\{\alpha_k\}$ is the weighting coefficient to ensure fairness among the UE and P denotes the power budget of the BS. Based on the channel model where prediction errors are taken into consideration, we investigate the optimal robust beamforming design for the proposed problem in the next section.

3 Proposed Robust Beamforming Design

Due to the randomness introduced by the channel prediction error, it is difficult to obtain the close-form expressions of the objective. To deal with the randomness, a robust beamforming design is proposed by considering deterministic equivalents. The NP-hard Problem (9) can be resolved by exploiting the relationship between the rate and the mean-square error (MSE) matrix, inspired by Ref. [5]. Denote the estimated signal at the receiver as $\hat{\boldsymbol{s}}_k = \boldsymbol{U}_k^H y_k$, where $\boldsymbol{U}_k$ denotes the receive beamformer of the k-th piece of UE. Assuming that signal $\boldsymbol{s}_k$ and noise $\boldsymbol{n}_k$ are independent, the MSE matrix can be written as

$$\begin{aligned} &\boldsymbol{MSE}_k = \mathbb{E}_{\boldsymbol{s},\boldsymbol{n}}\{(\hat{\boldsymbol{s}}_k - \boldsymbol{s}_k)(\hat{\boldsymbol{s}}_k - \boldsymbol{s}_k)^H\} = \\ &(\boldsymbol{I} - \boldsymbol{U}_k^H\boldsymbol{H}_k\boldsymbol{V}_k)(\mathbf{I} - \boldsymbol{U}_k^H\boldsymbol{H}_k\boldsymbol{V}_k)^H + \sum_{m\neq k}^{K}\boldsymbol{U}_k^H\boldsymbol{H}_k\boldsymbol{V}_m\boldsymbol{V}_m^H\boldsymbol{H}_k^H\boldsymbol{U}_k + \sigma^2\boldsymbol{U}_k^H\boldsymbol{U}_k . \end{aligned} \tag{10}$$

Fixing all of the transmitting beamforming matrices and adopting the well-known MMSE receiver, we can rewrite the MSE matrix as

$$\begin{aligned} &\boldsymbol{E}_k = \boldsymbol{MSE}_k(\boldsymbol{U}_k^{mmse}) = \\ &\boldsymbol{I} - \boldsymbol{V}_k^H\boldsymbol{H}_k^H(\boldsymbol{Q}_k + \boldsymbol{H}_k\boldsymbol{V}_k\boldsymbol{V}_k^H\boldsymbol{H}_k^H)^{-1}\boldsymbol{H}_k\boldsymbol{V}_k . \end{aligned} \tag{11}$$

Note that $R_k = \mathbb{E}\{\log\det(\boldsymbol{E}_k^{-1})\}$ can be expressed as a convex function of $\boldsymbol{E}_k$. Using the first-order condition of the convex function, we can obtain

$$\mathbb{E}\left\{\log\det(\boldsymbol{E}_k)^{-1}\right\} \geqslant \mathbb{E}\left\{\log\det(\boldsymbol{E}_k^{(d)})^{-1}\right\} - \mathbb{E}\left\{\mathrm{tr}(\boldsymbol{E}_k^{(d)})^{-1}\left(\boldsymbol{E}_k - \boldsymbol{E}_k^{(d)}\right)\right\} \geqslant \mathbb{E}\left\{\log\det(\boldsymbol{E}_k^{(d)})^{-1}\right\} + \mathrm{tr}(\boldsymbol{I}) - \mathbb{E}\left\{\mathrm{tr}(\boldsymbol{E}_k^{(d)})^{-1}\boldsymbol{MSE}_k\right\}, \quad (12)$$

where $\boldsymbol{E}_k^{(d)}$ represents $\boldsymbol{E}_k$ with a fixed set of beamforming matrix $\{\boldsymbol{V}_k^{(d)}\}$ at the d-th iteration, $k = 1,\cdots,K$. By ignoring the first two constant terms of the right side of Inequality (12), the last part can be expressed as a function of $\boldsymbol{V}_k$. For brevity, define functions $\eta_k^{pri}[f] = \mathbb{E}\{\boldsymbol{H}_k f \boldsymbol{H}_k{}^H\}$, $\tilde{\eta}_k^{pri}[f] = \mathbb{E}\{\boldsymbol{H}_k{}^H f \boldsymbol{H}_k\}$, $\eta_k[f] = \mathbb{E}\{\tilde{\boldsymbol{H}}_k f \tilde{\boldsymbol{H}}_k{}^H\}$, and $\tilde{\eta}_k[f] = \mathbb{E}\{\tilde{\boldsymbol{H}}_k{}^H f \tilde{\boldsymbol{H}}_k\}$. Denote the whole right side of Inequality (12) as $g(\boldsymbol{V}_k|\{\boldsymbol{V}_k^{(d)}\})$, which is described in detail as

$$g(\boldsymbol{V}_k|\{\boldsymbol{V}_k^{(d)}\}) = c_k^{(d)} + \mathrm{tr}((\boldsymbol{A}_k^{(d)})^H\boldsymbol{V}_k) + \mathrm{tr}(\boldsymbol{A}_k^{(d)}\boldsymbol{V}_k^H) - \mathrm{tr}\left(\boldsymbol{B}_k^{(d)}\boldsymbol{V}_k\boldsymbol{V}_k^H\right) - \mathrm{tr}\left(\boldsymbol{C}_k^{(d)}\sum_{m\neq k}\boldsymbol{V}_m\boldsymbol{V}_m{}^H\right), \quad (13)$$

where

$$c_k^{(d)} = \mathbb{E}\left\{\log\det(\boldsymbol{E}_k^{(d)})^{-1}\right\} + \mathrm{tr}(\boldsymbol{I}) - \mathbb{E}\left\{\mathrm{tr}\left((\boldsymbol{E}_k^{(d)})^{-1}\right)\right\} - \sigma^2\mathbb{E}\left\{\mathrm{tr}\left((\boldsymbol{E}_k^{(d)})^{-1}\left(\boldsymbol{U}_k^{(d)}\right)^H\boldsymbol{U}_k^{(d)}\right)\right\}, \quad (14)$$

$$\boldsymbol{A}_k^{(d)} = \hat{\boldsymbol{H}}_k^H\left(\boldsymbol{Q}_k^{(d)}\right)^{-1}\hat{\boldsymbol{H}}_k\boldsymbol{V}_k^{(d)} + \tilde{\eta}_k[\left(\boldsymbol{Q}_k^{(d)}\right)^{-1}]\boldsymbol{V}_k^{(d)}, \quad (15)$$

$$\boldsymbol{B}_k^{(d)} = \hat{\boldsymbol{H}}_k^H\left(\boldsymbol{Q}_k^{(d)}\right)^{-1}\hat{\boldsymbol{H}}_k + \tilde{\eta}_k\left[\left(\boldsymbol{Q}_k^{(d)}\right)^{-1}\right] - \mathbb{E}\left\{\boldsymbol{H}_k{}^H(\boldsymbol{Q}_k^{(d)} + \boldsymbol{H}_k\boldsymbol{V}_k^{(d)}\left(\boldsymbol{V}_k^{(d)}\right)^H\boldsymbol{H}_k{}^H)^{-1}\boldsymbol{H}_k\right\}, \quad (16)$$

$$\boldsymbol{C}_k^{(d)} = \mathbb{E}\left\{\boldsymbol{H}_k^H\left(\left(\boldsymbol{Q}_k^{(d)}\right)^{-1} - \mathbb{E}\left\{(\boldsymbol{Q}_k^{(d)} + \boldsymbol{H}_k\boldsymbol{V}_k^{(d)}\left(\boldsymbol{V}_k^{(d)}\right)^H\boldsymbol{H}_k^H)^{-1}\right\}\right)\boldsymbol{H}_k\right\}. \quad (17)$$

Denote $f(\{\boldsymbol{V}_k^{(d)}\})$ as the left side of Inequality (12) and $g(\boldsymbol{V}_k|\{\boldsymbol{V}_k^{(d)}\})$ can be regarded as the lower-bounding function which minorizes the objective function $f(\{\boldsymbol{V}_k^{(d)}\})$, since it is a convex function of $\boldsymbol{V}_k$ that satisfies $g(\boldsymbol{V}_k|\{\boldsymbol{V}_k^{(d)}\}) \leqslant f(\{\boldsymbol{V}_k\})$ and $g(\boldsymbol{V}_k^{(d)}|\{\boldsymbol{V}_k^{(d)}\}) = f(\{\boldsymbol{V}_k^{(d)}\})$. It can be iteratively optimized and converge to a stationary point, which has been proved in Ref. [22].

The update of $\boldsymbol{V}_k$ for all users can be decoupled across transmitters. Then, we update beamforming matrices as $\boldsymbol{V}_k^{(d+1)} = \arg\max_{\boldsymbol{V}_k} g\left(\boldsymbol{V}_k\middle|\{\boldsymbol{V}_k^{(d)}\}\right)$, and the optimal solution to the equation is obtained by the Lagrange multipliers method with Lagrange multiplier μ. According to the first-order optimal conditions, the iterative equation of the beamformer can be obtained as

$$\boldsymbol{V}_k^{(d+1)} = (\boldsymbol{D}_k^{(d)} + \mu\boldsymbol{I})^{-1}(\alpha_k\boldsymbol{A}_k^{(d)}), \quad (18)$$

where

$$\boldsymbol{D}_k^{(d)} = \alpha_k\boldsymbol{B}_k^{(d)} + \sum_{m\neq k}\alpha_m\boldsymbol{C}_m^{(d)}, \quad (19)$$

and μ can be obtained by using the bisection method.

For the last part of $\boldsymbol{B}_k^{(d)}$ and $\boldsymbol{C}_m^{(d)}$, which have not been transformed into the deterministic matrix, there are complicated calculations of the random variables, leading to difficulty in obtaining closed-form expressions. Drawing on the essence of Ref. [23], we will utilize deterministic equivalents to obtain the approximate closed-form expression of MIMO capacity, and the derivation process is presented as follow.

Note that $R_k = \mathbb{E}\left\{\log\det(\boldsymbol{I} + \boldsymbol{H}_k\boldsymbol{V}_k\boldsymbol{V}_k{}^H\boldsymbol{H}_k{}^H\boldsymbol{Q}_k^{-1})\right\}$ can be rewritten as $\mathbb{E}\left\{\log\det(\boldsymbol{I} + \boldsymbol{Q}_k^{-\frac{1}{2}}\boldsymbol{H}_k\boldsymbol{V}_k\boldsymbol{V}_k{}^H\boldsymbol{H}_k{}^H\boldsymbol{Q}_k^{-\frac{1}{2}})\right\}$. Take $\boldsymbol{Q}_k^{-\frac{1}{2}}\boldsymbol{H}_k\boldsymbol{V}_k$ as a whole part $\boldsymbol{\mathcal{H}}_k$ and denote the term $\boldsymbol{Q}_k^{-\frac{1}{2}}\boldsymbol{H}_k\boldsymbol{V}_k\boldsymbol{V}_k{}^H\boldsymbol{H}_k{}^H\boldsymbol{Q}_k^{-\frac{1}{2}}$ as a Hermitian matrix $\boldsymbol{Z}_{N_r,k}$. Let $F_{\boldsymbol{Z}_{N_r,k}}(\lambda)$ denote the expected cumulative distribution of the eigenvalues $\lambda_1,\cdots,\lambda_{N_r}$ of $\boldsymbol{Z}_{N_r,k}$, and the Shannon transform $\mathcal{V}_{\boldsymbol{Z}_{N_r,k}}$ is described as $\mathcal{V}_{\boldsymbol{Z}_{N_r,k}}(x) = \int_0^\infty\log\left(1+\frac{1}{x}\lambda\right)dF_{\boldsymbol{Z}_{N_r,k}}(\lambda)$. Then R_k can be simplified as $\sum\mathbb{E}_\lambda\left[\log(1+\lambda_i)\right]$ and translated into $N_r\int_0^\infty\log(1+\lambda)\,dF_{\boldsymbol{Z}_{N_r,k}}(\lambda)$, finally equivalent to $N_r\mathcal{V}_{\boldsymbol{Z}_{N_r,k}}(1)$. The Stieltjes transform for $F_{\boldsymbol{Z}_{N_r,k}}(\lambda)$ is

$$\mathcal{S}_{\boldsymbol{Z}_{N_r,k}}(y) = \int\frac{1}{y-\lambda}\mathrm{d}F_{\boldsymbol{Z}_{N_r,k}}(\lambda) = \frac{1}{N_r}\mathbb{E}\{\mathrm{tr}[(y\boldsymbol{I}_{N_r} - \boldsymbol{Z}_{N_r,k})^{-1}]\}. \quad (20)$$

Then the relation between the Stieltjes transform $\mathcal{S}_{\boldsymbol{Z}_{N_r,k}}(y)$ and the Shannon transform $\mathcal{V}_{\boldsymbol{Z}_{N_r,k}}$ can be established as

$$\mathcal{V}_{\boldsymbol{Z}_{N_r,k}}(x) = \int_x^\infty\left(\frac{1}{y} + \mathcal{S}_{\boldsymbol{Z}_{N_r,k}}(-y)\right)\mathrm{d}y. \quad (21)$$

Thus, we can obtain the closed-form expression of the ergodic user rate R_k by establishing the closed-form expression of the Stieltjes transform and Shannon transform. The approximate closed-form Stieltjes transform expressions can be calculated by applying the free probability theory[24], denoted as Eqs. (22) and (23) with Notations (24) – (27).

$$\mathcal{S}_{\boldsymbol{Z}_{N_r,k}}(y\boldsymbol{I}_{N_r}) = (y\tilde{\boldsymbol{\Phi}}_k - \left(\boldsymbol{Q}_k^{(d)}\right)^{-\frac{1}{2}}\hat{\boldsymbol{H}}_k\boldsymbol{V}_k^{(d)}\boldsymbol{\Phi}_k^{-1}\boldsymbol{V}_k^{(d)H}\hat{\boldsymbol{H}}_k^H\left(\boldsymbol{Q}_k^{(d)}\right)^{-\frac{1}{2}})^{-1}, \tag{22}$$

$$\mathcal{S}_{\boldsymbol{\mathcal{H}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k}(y\boldsymbol{I}_{N_t}) = (y\boldsymbol{\Phi}_k - \left(\boldsymbol{V}_k^{(d)}\right)^H\hat{\boldsymbol{H}}_k^H\left(\boldsymbol{Q}_k^{(d)}\right)^{-\frac{1}{2}}\tilde{\boldsymbol{\Phi}}_k^{-1}\left(\boldsymbol{Q}_k^{(d)}\right)^{-\frac{1}{2}}\hat{\boldsymbol{H}}_k\boldsymbol{V}_k^{(d)})^{-1}, \tag{23}$$

$$\tilde{\boldsymbol{\Phi}}_k = \boldsymbol{I}_{N_r} - \left(\boldsymbol{Q}_k^{(d)}\right)^{-\frac{1}{2}}\eta_k\left[\left(\boldsymbol{V}_k^{(d)}\right)^H\mathcal{S}_{\boldsymbol{\mathcal{H}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k}(y\boldsymbol{I}_{N_t})\boldsymbol{V}_k^{(d)}\right]\left(\boldsymbol{Q}_k^{(d)}\right)^{-\frac{1}{2}}, \tag{24}$$

$$\boldsymbol{\Phi}_k = \boldsymbol{I}_{d_k} - \left(\boldsymbol{V}_k^{(d)}\right)^H\tilde{\eta}_k\left[\left(\boldsymbol{Q}_k^{(d)}\right)^{-\frac{1}{2}}\mathcal{S}_{\boldsymbol{Z}_{N_r,k}}(y\boldsymbol{I}_{N_r})\left(\boldsymbol{Q}_k^{(d)}\right)^{-\frac{1}{2}}\right]\boldsymbol{V}_k^{(d)}, \tag{25}$$

$$\boldsymbol{\Gamma}_k = -\tilde{\eta}_k\left[\left(\boldsymbol{Q}_k^{(d)}\right)^{-\frac{1}{2}}\mathcal{S}_{\boldsymbol{Z}_{N_r,k}}(-\boldsymbol{I}_{N_r})\left(\boldsymbol{Q}_k^{(d)}\right)^{-\frac{1}{2}}\right] + \hat{\boldsymbol{H}}_k^H\left(\boldsymbol{Q}_k^{(d)}\right)^{-\frac{1}{2}}\tilde{\boldsymbol{\Phi}}_k^{-1}\left(\boldsymbol{Q}_k^{(d)}\right)^{-\frac{1}{2}}\hat{\boldsymbol{H}}_k, \tag{26}$$

$$\tilde{\boldsymbol{\Gamma}}_k = -\eta_k\left[\boldsymbol{V}_k^{(d)}\mathcal{S}_{\boldsymbol{\mathcal{H}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k}(-\boldsymbol{I}_{N_t})\left(\boldsymbol{V}_k^{(d)}\right)^H\right] + \hat{\boldsymbol{H}}_k\boldsymbol{V}_k^{(d)}\boldsymbol{\Phi}_k^{-1}\left(\boldsymbol{V}_k^{(d)}\right)^H\hat{\boldsymbol{H}}_k^H. \tag{27}$$

Then the deterministic equivalent form of rate $\bar{R}_k$ can be expressed as:

$$\bar{R}_k = \log\det(\boldsymbol{I}_{N_r} + \tilde{\boldsymbol{\Gamma}}_k\boldsymbol{Q}_k^{-1}) + \log\det(\boldsymbol{\Phi}_k) - \mathrm{tr}\left(\eta_k[\boldsymbol{V}_k\mathcal{S}_{\boldsymbol{\mathcal{H}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k}(-\boldsymbol{I}_{N_t})\boldsymbol{V}_k^H]\boldsymbol{Q}_k^{-\frac{1}{2}}\mathcal{S}_{\boldsymbol{Z}_{N_r,k}}(-\boldsymbol{I}_{N_r})\boldsymbol{Q}_k^{-\frac{1}{2}}\right), \tag{28}$$

$$\bar{R}_k = \log\det(\boldsymbol{I}_{N_t} + \boldsymbol{\Gamma}_k\boldsymbol{V}_k\boldsymbol{V}_k^H) + \log\det(\tilde{\boldsymbol{\Phi}}_k) - \mathrm{tr}\left(\tilde{\eta}_k[\boldsymbol{Q}_k^{-\frac{1}{2}}\mathcal{S}_{\boldsymbol{Z}_{N_r,k}}(-\boldsymbol{I}_{N_r})\boldsymbol{Q}_k^{-\frac{1}{2}}]\boldsymbol{V}_k\mathcal{S}_{\boldsymbol{\mathcal{H}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k}(-\boldsymbol{I}_{N_t})\boldsymbol{V}_k^H\right). \tag{29}$$

Based on the aforementioned analysis, the approximate closed-form expressions of the second part of $\boldsymbol{B}_k^{(d)}$ can be derived as Eq. (30) by exploiting the relationship between it and the original expression R_k. The closed-form expressions of $\boldsymbol{C}_k^{(d)}$ can be derived similarly as Eq. (31).

$$\mathbb{E}\left\{\boldsymbol{H}_k^H(\boldsymbol{Q}_k^{(d)} + \boldsymbol{H}_k\boldsymbol{V}_k^{(d)}\left(\boldsymbol{V}_k^{(d)}\right)^H\boldsymbol{H}_k^H)^{-1}\boldsymbol{H}_k\right\} = \frac{\partial R_k}{\partial\left(\boldsymbol{V}_k\boldsymbol{V}_k^H\right)} = \frac{\partial\bar{R}_k}{\partial\left(\boldsymbol{V}_k\boldsymbol{V}_k^H\right)} = (\boldsymbol{I}_{N_t} + \boldsymbol{\Gamma}_k\boldsymbol{V}_k\boldsymbol{V}_k^H)^{-1}\boldsymbol{\Gamma}_k, \tag{30}$$

$$\mathbb{E}\left\{\boldsymbol{H}_k^H\left(\left(\boldsymbol{Q}_k^{(d)}\right)^{-1} - \mathbb{E}\left\{(\boldsymbol{Q}_k^{(d)} + \boldsymbol{H}_k\boldsymbol{V}_k^{(d)}\left(\boldsymbol{V}_k^{(d)}\right)^H\boldsymbol{H}_k^H)^{-1}\right\}\right)\boldsymbol{H}_k\right\} = \frac{\partial R_m}{\partial\left(\boldsymbol{V}_m\boldsymbol{V}_m^H\right)} = \frac{\partial\bar{R}_k}{\partial\left(\boldsymbol{V}_m\boldsymbol{V}_m^H\right)} = \tilde{\eta}_k^{pri}\left(\boldsymbol{Q}_k^{-1}\right) - \tilde{\eta}_k^{pri}\left((\boldsymbol{Q}_k + \tilde{\boldsymbol{\Gamma}}_k)^{-1}\right). \tag{31}$$

We then obtain

$$\boldsymbol{B}_k^{(d)} = \hat{\boldsymbol{H}}_k^H\left(\boldsymbol{Q}_k^{(d)}\right)^{-1}\hat{\boldsymbol{H}}_k + \tilde{\eta}_k[\left(\boldsymbol{Q}_k^{(d)}\right)^{-1}] - (\boldsymbol{I}_{N_t} + \boldsymbol{\Gamma}_k\boldsymbol{V}_k\boldsymbol{V}_k^H)^{-1}\boldsymbol{\Gamma}_k, \tag{32}$$

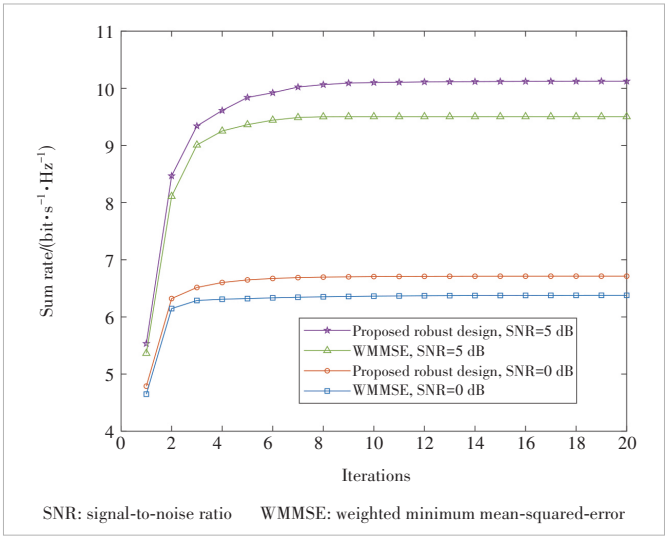
$$\boldsymbol{C}_k^{(d)} = \tilde{\eta}_k^{\mathrm{pri}}\left(\boldsymbol{Q}_k^{-1}\right) - \tilde{\eta}_k^{\mathrm{pri}}\left((\boldsymbol{Q}_k + \tilde{\boldsymbol{\Gamma}}_k)^{-1}\right). \tag{33}$$

By substituting Eqs. (15), (19), (32) and (33) into Eq. (18), we can derive the close-form of $\boldsymbol{V}_k^{(d+1)}$ and the sum-rate maximization problem (9) is finally able to be solved by alternately optimizing $\boldsymbol{V}_k$ until the convergence or the maximum iteration times are reached.

4 Simulation Results

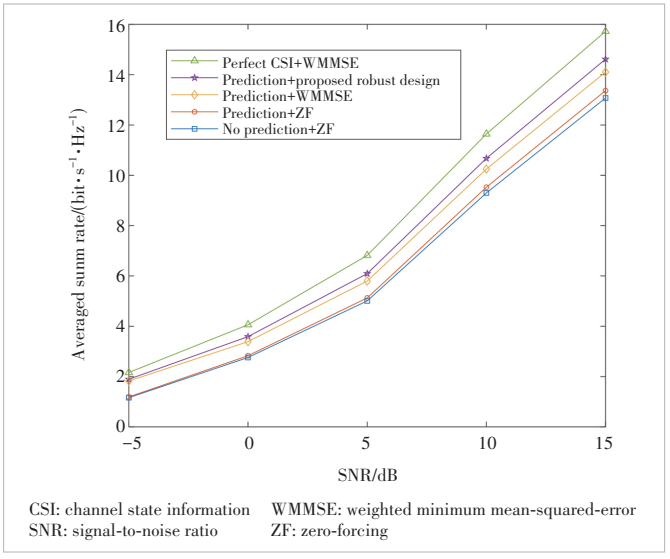
In this section, our simulation results show the system performance of the proposed beamforming design under channel prediction errors. We consider a MU-MIMO system with the number of UE K set to 4 and each piece of UE is equipped with $N_r = 4$ receive antennas. The transmit antenna number $N_t = 32$ and the transmit power budget P is set to 1. The weight α_k for each piece of UE is set equally. In addition, $d_k = N_r$, $\tau = 7$, $N_{\mathrm{slot}} = 10$, and $w = 10$. The cluster delay line (CDL) channel model is used in the simulation to generate a time-varying channel for each piece of UE, where the implementations exactly follow the 3GPP 5G new radio standard protocol TR 38.901[25]. Specifically, we adopt an urban macro scenario and consider a CDL-A delay profile where the delay spread is set as 100 ns. The Doppler shift is computed by $f_d = vf_c/c$, which is a combination of user speed v, carrier frequency f_c and the speed of light c, with $f_c = 2$ GHz. Different UE velocities are set to indicate different time-varying channels for performance comparisons. Several combinations of acquired CSI and beamforming methods are discussed here. The simplest case with no prediction and ZF algorithm is considered, where the CSI used for DL beamforming is the estimated CSI obtained in the previous SRS time slot. With the predicted CSI for DL time slots, ZF and WMMSE are considered for comparisons. Moreover, the case with perfect CSI known in the transmitter and WMMSE beamforming is also discussed as an ideal case.

Fig. 4 illustrates the convergence behavior of the proposed beamforming design and WMMSE for the cases of SNR = 0 dB and SNR = 5 dB. Since the extra prediction errors are taken into consideration in the proposed design and then the



SNR: signal-to-noise ratio WMMSE: weighted minimum mean-squared-error

▲Figure 4. Convergence behavior of the proposed beamforming design and WMMSE



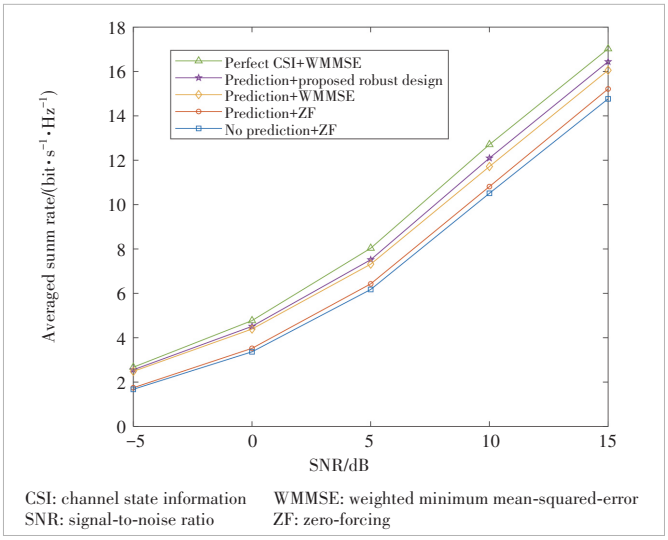
CSI: channel state information WMMSE: weighted minimum mean-squared-error
SNR: signal-to-noise ratio ZF: zero-forcing

▲Figure 6. Averaged sum rate versus SNR under different beamforming methods with v=120 km/h

randomness of the objective is to be solved, it is supposed to be more complex to conduct. As shown in this figure, the proposed one converges a little slower than WMMSE but both behave almost the same. And the proposed converges within 10 iterations, which confirms the practicality of our beamforming design. Moreover, it is observed that both algorithms take more iterations to converge as the SNR increases.

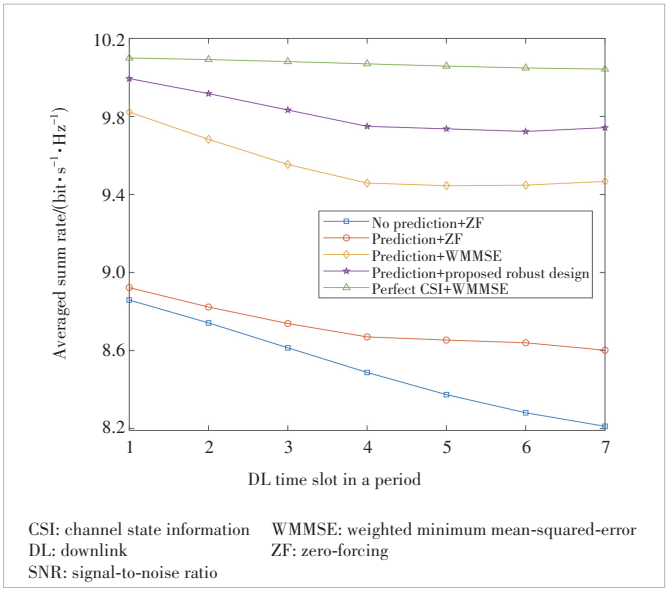
From Figs. 5 and 6, we can observe the averaged sum-rate performance of the system under different SNRs for different beamforming cases. Firstly, it is easy to notice the performance gain achieved by the channel prediction, which grows as the SNR increases. With the predicted CSI, WMMSE cannot attain equivalent gain compared with our proposed method that takes the error into consideration, due to the imperfection of CSI brought by the prediction errors which cause the performance limitation. The proposed beamforming design outperforms WMMSE more as the SNR goes higher. Comparing these two figures, we can find that the overall performance goes down as the UE velocity increases, with the performance gaps from the ideal perfect case to others becoming larger, but our robust method can maintain certain gains under such a severe time-varying effect. The proposed design can even achieve more gain against WMMSE with a higher velocity.

Fig. 7 shows the sum-rate performance comparisons during the DL time slot within the half-frame period. The statistics of each DL time slot plotted in the figure are obtained by averaging 100 periods. As time goes by, the outdating of acquired



CSI: channel state information WMMSE: weighted minimum mean-squared-error
SNR: signal-to-noise ratio ZF: zero-forcing

▲Figure 5. Averaged sum rate versus SNR under different beamforming methods with v=60 km/h



CSI: channel state information WMMSE: weighted minimum mean-squared-error
DL: downlink ZF: zero-forcing
SNR: signal-to-noise ratio

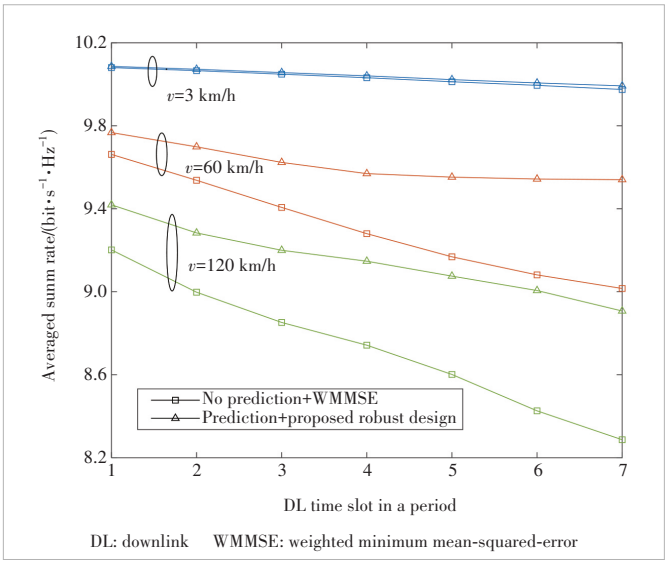
▲Figure 7. Sum rate versus DL time slots under different beamforming methods with v=60 km/h

CSI is more serious, so the averaged sum rates decrease during the DL transmission period. However, the channel prediction can combat this fading, with more obvious gain in the last few time slots as shown in this figure. By considering the channel prediction errors and utilizing the statistical characteristics, the proposed beamforming can even achieve better performance which is close to the ideal case.

Fig. 8 presents the fall of the performance during the DL time slots within the half-frame period under different UE velocities. Considering the advantage of joint channel prediction and proposed robust beamforming design, we take the WMMSE method with no prediction as the benchmark. The averaged sum rates decrease more rapidly with the increasing of velocity. This is because the velocity is a sign of channel variation. Though the achievable gain of channel prediction gets smaller as the velocity increase, which can be observed by comparing Figs. 5 and 6, the joint robust design can implement more substantial performance gain as the time-varying effects become severer.

5 Conclusions

In this paper, we investigate the beamforming design under channel prediction errors in a time-varying MIMO system. By collecting the prediction errors and exploiting the statistical characteristics of the error samples, we propose a robust beamforming design to further combat the detriment caused by channel aging based on the channel prediction system. Due to the uncertainty brought by the error part, we exploit deterministic equivalents to obtain the closed-form expression to derive the robust beamforming matrix. Our simulation results show the effectiveness of the proposed beamforming design and reveal that the achievable gain even grows as UE velocities increase. The proposed design can maintain the outperformance during the DL transmission time while the channels vary fast.



▲**Figure 8. Sum rate versus DL time slots under different beamforming methods with v=60 km/h**

References

[1] NADEEM Q U A, KAMMOUN A, ALOUINI M S. Elevation beamforming with full dimension MIMO architectures in 5G systems: a tutorial [J]. IEEE communications surveys & tutorials, 2019, 21(4): 3238 – 3273. DOI: 10.1109/COMST.2019.2930621
[2] LARSSON E G, EDFORS O, TUFVESSON F, et al. Massive MIMO for next generation wireless systems [J]. IEEE communications magazine, 2014, 52(2): 186 – 195. DOI: 10.1109/MCOM.2014.6736761
[3] CASTAÑEDA E, SILVA A, GAMEIRO A, et al. An overview on resource allocation techniques for multi-user MIMO systems [J]. IEEE communications surveys & tutorials, 2017, 19(1): 239 – 284. DOI: 10.1109/COMST.2016.2618870
[4] CHRISTENSEN S S, AGARWAL R, DE CARVALHO E, et al. Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design [J]. IEEE transactions on wireless communications, 2008, 7(12): 4792 – 4799. DOI: 10.1109/T-WC.2008.070851
[5] SHI Q J, RAZAVIYAYN M, LUO Z Q, et al. An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel [J]. IEEE transactions on signal processing, 2011, 59(9): 4331 – 4340. DOI: 10.1109/TSP.2011.2147784
[6] WU Y P, JIN S, GAO X Q, et al. Transmit designs for the MIMO broadcast channel with statistical CSI [J]. IEEE transactions on signal processing, 2014, 62(17): 4451 – 4466. DOI: 10.1109/TSP.2014.2336637
[7] WU D Y, ZHANG H T. Tractable modelling and robust coordinated beamforming design with partially accurate CSI [J]. IEEE wireless communications letters, 2021, 10(11): 2384 – 2387. DOI: 10.1109/LWC.2021.3101037
[8] ZHENG J K, ZHANG J Y, BJÖRNSON E, et al. Impact of channel aging on cell-free massive MIMO over spatially correlated channels [J]. IEEE transactions on wireless communications, 2021, 20(10): 6451 – 6466. DOI: 10.1109/TWC.2021.3074421
[9] LIU L H, FENG H, YANG T, et al. MIMO-OFDM wireless channel prediction by exploiting spatial-temporal correlation [J]. IEEE transactions on wireless communications, 2014, 13(1): 310 – 319. DOI: 10.1109/TWC.2013.112613.130455
[10] YIN H F, WANG H Q, LIU Y Z, et al. Addressing the curse of mobility in massive MIMO with prony-based angular-delay domain channel predictions [J]. IEEE journal on selected areas in communications, 2020, 38(12): 2903 – 2917. DOI: 10.1109/JSAC.2020.3005473
[11] WU C, YI X P, ZHU Y M, et al. Channel prediction in high-mobility massive MIMO: from spatio-temporal autoregression to deep learning [J]. IEEE journal on selected areas in communications, 2021, 39(7): 1915 – 1930. DOI: 10.1109/JSAC.2021.3078503
[12] THEODORIDIS S. Machine learning: a Bayesian and optimization perspective [M]. Orlando, USA: Academic press, 2015
[13] YUAN J D, NGO H Q, MATTHAIOU M. Machine learning-based channel prediction in massive MIMO with channel aging [J]. IEEE transactions on wireless communications, 2020, 19(5): 2960 – 2973. DOI: 10.1109/TWC.2020.2969627
[14] XU K, SHEN Z X, WANG Y R, et al. Location-aided mMIMO channel tracking and hybrid beamforming for high-speed railway communications: an angle-domain approach [J]. IEEE systems journal, 2020, 14(1): 93 – 104. DOI: 10.1109/JSYST.2019.2911296
[15] XIA X C, XU K, ZHAO S B, et al. Learning the time-varying massive MIMO channels: robust estimation and data-aided prediction [J]. IEEE transactions on vehicular technology, 2020, 69(8): 8080 – 8096. DOI: 10.1109/TVT.2020.2968637
[16] ENGEL Y, MANNOR S, MEIR R. The kernel recursive least-squares algorithm [J]. IEEE transactions on signal processing, 2004, 52(8): 2275 – 2285. DOI: 10.1109/TSP.2004.830985

[17] LIU W F, PARK I, PRINCIPE J C. An information theoretic approach of designing sparse kernel adaptive filters [J]. IEEE transactions on neural networks, 2009, 20(12): 1950 – 1961. DOI: 10.1109/TNN.2009.2033676
[18] MIN C, CHANG N, CHA J, et al. MIMO-OFDM downlink channel prediction for IEEE802.16e systems using Kalman filter [C]//IEEE Wireless Communications and Networking Conference. IEEE, 2007: 942 – 946. DOI: 10.1109/WCNC.2007.179
[19] RAMYA T R, BHASHYAM S. On using channel prediction in adaptive beamforming systems [C]//2nd International Conference on Communication Systems Software and Middleware. IEEE, 2007: 1 – 6. DOI: 10.1109/COMSWA.2007.382451
[20] MARTOS-NAYA E, PARIS J F, FERNANDEZ-PLAZAOLA U, et al. Exact BER analysis for M-QAM modulation with transmit beamforming under channel prediction errors [J]. IEEE transactions on wireless communications, 2008, 7(10): 3674 – 3678. DOI: 10.1109/T-WC.2008.070192
[21] PARIS J F, MARTOS-NAYA E, FERNANDEZ-PLAZAOLA U, et al. Analysis of adaptive MIMO transmit beamforming under channel prediction errors based on incomplete lipschitz – hankel integrals [J]. IEEE transactions on vehicular technology, 2009, 58(6): 2815 – 2824. DOI: 10.1109/TVT.2008.2011990
[22] SUN Y, BABU P, PALOMAR D P. Majorization-minimization algorithms in signal processing, communications, and machine learning [J]. IEEE transactions on signal processing, 2017, 65(3): 794 – 816. DOI: 10.1109/TSP.2016.2601299
[23] LU A N, GAO X Q, XIAO C S. Free deterministic equivalents for the analysis of MIMO multiple access channel [J]. IEEE transactions on information theory, 2016, 62(8): 4604 – 4629. DOI: 10.1109/TIT.2016.2573309
[24] TULINO A M, VERDÚ S. Random matrix theory and wireless communications [M]. Norwell, USA: Now Publishers Inc, 2004. DOI: 10.1561/9781933019505
[25] 3GPP. Study on channel model for frequencies from 0.5 to 100 GHz: TR 38.901 [S]. 2019

Biographies

ZHU Yuting received her bachelor's degree in communication engineering from Beijing University of Posts and Telecommunications (BUPT), China in 2022. She is currently working toward a master's degree in communication and information engineering at the School of Artificial Intelligence, BUPT. Her research interests include the emerging technologies of 5G wireless communication networks.

LI Zeng (li.zeng@zte.com.cn) received his master's degree in information and communication engineering from Harbin Institute of Technology, China in 2016. He is currently working at the Algorithm Department, Wireless Product R&D Institute, Wireless Product Operation Division, ZTE Corporation. His research interests include 5G wireless communications and signal processing.

ZHANG Hongtao received his PhD degree in communication and information systems from Beijing University of Posts and Telecommunications (BUPT), China in 2008. He is currently a full professor with BUPT. He has authored or co-authored more than 100 articles on international journals and conferences, and has filed more than 50 patents. He is also the author of ten technical books. His research interests include 5G wireless communications and signal processing.

Design of Raptor-Like LDPC Codes and High Throughput Decoder Towards 100 Gbit/s Throughput

LI Hanwen, BI Ningjing, SHA Jin
(School of Electronic Science & Engineering, Nanjing University, Nanjing 210000, China)

DOI: 10.12142/ZTECOM.202303012
https://kns.cnki.net/kcms/detail/34.1294.TN.20230718.1034.002.html, published online July 19, 2023
Manuscript received: 2023-01-29

Abstract: This paper proposes a raptor-like low-density parity-check (RL-LDPC) code design together with the corresponding decoder hardware architecture aiming at next-generation mobile communication. A new kind of protograph different from the 5G new radio (NR) LDPC basic matrix is presented, and a code construction algorithm is proposed to improve the error-correcting performance. A multi-core layered decoder architecture that supports up to 100 Gbit/s throughput is designed based on the special protograph structure.

Keywords: RL-LDPC; error floor; high throughput; protograph; 5G NR

Citation (Format 1): LI H W, BI N J, SHA J. Design of raptor-like LDPC codes and high throughput decoder towards 100 Gbit/s throughput [J]. ZTE Communications, 2023, 21(3): 86 - 92. DOI: 10.12142/ZTECOM.202303012
Citation (Format 2): H. W. Li, N. J. Bi, and J. Sha, "Design of raptor-like LDPC codes and high throughput decoder towards 100 Gbit/s throughput," *ZTE Communications*, vol. 21, no. 3, pp. 86 - 92, Sept. 2023. doi: 10.12142/ZTECOM.202303012.

1 Introduction

5G is emerging as a promising technology on a global scale. However, with the demand for virtual reality (VR), Metaverse, and other high throughput applications, the transmission speed needs to be improved. As one of the most important technologies in wireless communication, channel coding is of great significance to the improvement of system reliability[1].

Low-density parity-check (LDPC) codes have been applied in many communication systems due to their top performance, which is close to the Shannon limit and suitable for parallel structures. LDPC codes were first proposed by GALLAGER in 1963[2], and in 1996, MACKAY and NEAL proposed the positional degree propagation iterative decoding algorithm for LDPC codes[3], which significantly improved performance.

Code construction is a critical part of LDPC code design because it directly affects error correction performance. The random construction method cannot guarantee the absence of short cycles. Based on this, the progressive edge growth (PEG) algorithm was proposed[4], which tries to avoid short cycles. Some improved PEG algorithms were proposed in Refs. [5 - 6], which failed to guarantee the good performance of the code. The approximate cycle extrinsic (ACE) message degree algorithm was suggested in Ref. [7], which could eliminate the cycles selectively in the LDPC codes. The codes constructed by the ACE algorithm have lower error floors but perform worse in the waterfall region than the PEG algorithm.

Typically, the LDPC code can be obtained by lifting the protograph, which is a structure that governs the macroscopic statistics of the code. The protograph is an important optimization object in LDPC design because it determines the error-correcting performance of LDPC codes. Some wireless communication LDPC codes are designed with this approach, like the 5G NR LDPC code.

The belief propagation (BP) algorithm is the classic decoding algorithm for LDPC codes. The BP algorithm is computationally intensive and not conducive to hardware implementation, so researchers have proposed the min-sum decoding algorithm, which still uses the idea of iterative decoding with probabilistic computation but reduces the hardware resource greatly. It occupies an important position in the implementation of LDPC decoders, and most of the decoders adopt this method as a guiding idea.

The code construction algorithm and the decoding algorithm tend to affect the performance of the code, while the

This work is supported in part by ZTE Industry-University-Institute Cooperation funds under Grant No. 2020ZTE01-03.

hardware implementation of the decoder has more impact on the throughput. Many researchers have proposed different decoder architectures in the hardware implementation aspect.

In Ref. [8], researchers used a fully parallel decoding scheme to partition the global interconnection and weaken the routing pressure by grouping the variable nodes. The architecture achieves a throughput of 92.8 Gbit/s. A partial parallel decoder architecture with multi-frame pipeline decoding is proposed in Ref. [9] to increase the parallelism. It can achieve higher throughput while having high decoding flexibility. In Ref. [10], researchers proposed an iterative unfolding architecture, which fully expanded the hardware architecture used for one iteration so that one iteration could be performed per cycle. This approach lacks flexibility while significantly increasing throughput. The multi-core decoder was implemented in Ref. [11], which could satisfy high throughput while being relatively flexible.

In this paper, we design a new kind of RL-LDPC code aiming at next-generation mobile communication and give the construction results of the core matrix. A code construction algorithm is proposed to optimize the performance of the code by eliminating cycles of lengths 6 and 4. In addition, a hardware architecture is proposed to match the above codes using a multi-core row-layered decoding approach to achieve a throughput of 100 Gbit/s. The multi-core decoder mainly includes an input buffer, a controller, decoder cores and their memory blocks, an output buffer, and an output selection module.

The remainder of this paper is organized as follows. Section 2 introduces the code construction method we proposed. Then we describe the hardware architecture in Section 3. In Section 4, the performance and hardware resource consumption of the code are presented. Finally, the conclusions are drawn in Section 5.

2 Codes Construction

2.1 RL-LDPC Codes Construction

LDPC code is determined by an $m \times n$ parity check matrix $\boldsymbol{H}$, where one element in the $\boldsymbol{H}$ matrix is much less than zero, m represents the number of rows, and n represents the number of columns. Each column corresponds to a variable node (VN), and each row corresponds to a check node (CN). The information block length is $k = m - n$. The elements in the parity check matrix $\boldsymbol{H}$ indicate whether there is information exchange between the VNs and the CNs. The tanner diagram can be used to represent the information transmission network that connects the CNs and VNs. As shown in Fig. 1, the circles represent the CNs and the squares represent the VNs. The structure of starting from one node and making a circle along the connection to return to the node is called a "cycle". Short cycles can have a bad effect on performance.

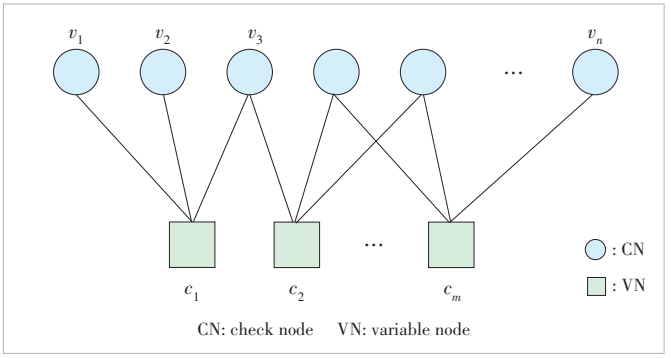
The protograph structure of RL-LDPC is shown in Fig. 2. $\boldsymbol{H}_{\mathrm{HRC}}$ is the core matrix framed in green in Fig. 2, which corresponds to the highest code rate of the LDPC code, and $\boldsymbol{H}_{\mathrm{IRC}}$ is the extension matrix framed in blue in Fig. 2. The sizes are marked in the figure. The size of the extension matrix can be adjusted according to different code rates. $\boldsymbol{H}_{\mathrm{HRC}}$ and $\boldsymbol{H}_{\mathrm{IRC}}$ are constructed and optimized separately[12]. Matrix $\boldsymbol{B}$ is a square matrix with a double diagonal structure and its size is $P \times P$. Matrix $\boldsymbol{I}$ is an identity matrix, which corresponds to the variable node part of RL-LDPC with a degree of 1.

$\boldsymbol{H}_{\mathrm{HRC}}$ consists of a number of elements shown as follows:

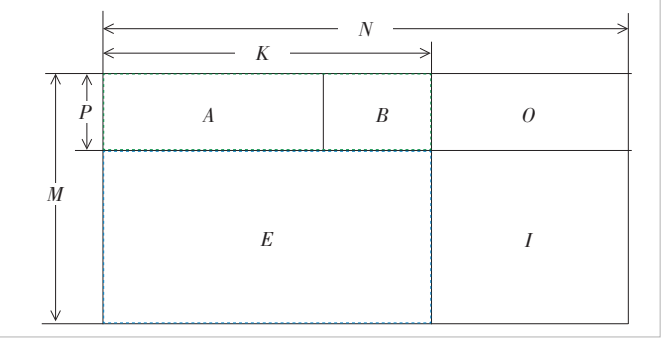
$$\boldsymbol{H}_{\mathrm{HRC}} = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,K+P} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,K+P} \\ \vdots & \vdots & \vdots & \vdots \\ h_{P,1} & h_{P,2} & \cdots & h_{P,K+P} \end{bmatrix}. \quad (1)$$

The parity check matrix is obtained by expanding the base matrix above by a lifting size Z. Therefore, each element of $\boldsymbol{H}_{\mathrm{HRC}}$ is replaced by a square matrix of $Z \times Z$. Elements "−1" are replaced by the all-zero matrix, while $h_{i,j}>0$ elements are replaced by a circulant matrix, obtained by right shifting the identity matrix by $h_{i,j}$ positions. The range of cyclic lifting number $h_{i,j}$ is 0 to $Z - 1$.

The method of code construction based on protograph is used in this paper. A protograph with a low iterative decoding threshold needs to be constructed first, and then the lifting numbers are generated to obtain our LDPC code. Following



▲Figure 1. Tanner graph



▲ Figure 2. Structure of raptor-like low-density parity-check (RL-LDPC) code

the below principles when constructing a protograph can ensure the basic performance of the code.

Keeping the variable node degree ⩾ 3 can ensure linear minimum distance growth with the blocklength in the process of constructing $\boldsymbol{H}_{\mathrm{HRC}}$. The linear minimum distance growth property ensures that the performance of the code does not degrade as the block length increases. Adding some variable nodes with the degree −2 or even degree −1 in the code is beneficial to the iterative decoding threshold. A small number of such nodes will not destroy the linear minimum distance growth property, but they need to be added prudently. Adding punctured variable nodes in the $\boldsymbol{H}_{\mathrm{HRC}}$ can improve the threshold of the code[13]. During the protograph construction, almost every check node in $\boldsymbol{H}_{\mathrm{IRC}}$ should be connected to the punctured variable node.

2.2 Protograph Construction Constraint

We present the constraints that need to be followed when constructing a protograph. "1" in Fig. 3 represents a non-zero submatrix, and "0" represents a zero submatrix. The specific constraints can be described as follows. Each protograph row is obtained by shifting the above row cyclically. The first row is shifted by one column to generate the second row, the second row is shifted by one column to generate the third row, and so on. The number of shifts can be flexibly adjusted as needed. Fig. 3 shows the protograph structure with a size of 4×16. The positions of "1" elements in rows are obtained by shifting the above row by one column. Such a construction structure can simplify the design of the reading and writing networks in traditional architecture, which reduces the consumption of hardware resources.

In practical applications, this constraint will not be as regular as shown in the figure above. In our decoder design, the constraint is added only in the first K columns of the protograph. The left columns are directly solidified into a double diagonal matrix to adapt to the invertibility of the matrix and encoding requirements. The corresponding reading and writing networks are consistent with the traditional structure.

Our construction method can directly construct the protograph of the required size randomly: determining the protograph of the first row first, and then shifting to obtain the protograph of other rows. The check part of the protograph is fixed as a double diagonal structure and does not participate in shifting.

The specific steps are as follows:

1	1	0	1	1	0	1	0	1	1	1	0	1	1	1	1
1	1	1	0	1	1	0	1	0	1	1	1	0	1	1	1
1	1	1	1	0	1	1	0	1	0	1	1	1	0	1	1
1	1	1	1	1	0	1	1	0	1	0	1	1	1	0	1

▲**Figure 3. Construction constraint structure**

1) Generate the first row of the protograph randomly;

2) Generate additional rows of the protograph by shifting the first row;

3) Randomly reduce or increase the number of "1" elements in the protograph to improve the threshold.

2.3 Codes Construction Method

In addition to searching for protograph and optimizing the iterative decoding threshold, a code construction algorithm is designed, which outperforms the conventional PEG algorithm. The codes constructed by using PEG algorithms contain a small number of 6 cycles, so we suggest a new algorithm for code construction to guarantee the absence of 6 cycles in the construction.

Starting from the code we randomly construct, our algorithm searches for all 6 cycles and 4 cycles in the code and modifies the submatrix's lifting number, which connects the maximum number of cycles, so that the number of cycles passing through this submatrix is reduced. This process is repeated until the number of cycles cannot be reduced through all submatrices.

Set the parity check matrix of the constructed LDPC code as $\boldsymbol{H}$. P is the corresponding protograph, $h_{i,j}$ is each element of $\boldsymbol{H}$, $p_{i,j}$ is each element of P, and n is the iteration number. Set $g4_{i,j,n}$, $g6_{i,j,n}$ as the number of 4 cycles and 6 cycles associated with $h_{i,j}$. The element with the largest number of associated 4 cycles is $g4_{\max\_i,\max\_j}$. The element with the largest number of associated 6 cycles is $g6_{\max\_i,\max\_j}$. Set the lifting size as Z.

The specific algorithm is shown below:

Algorithm 1: Cycle elimination construction algorithm

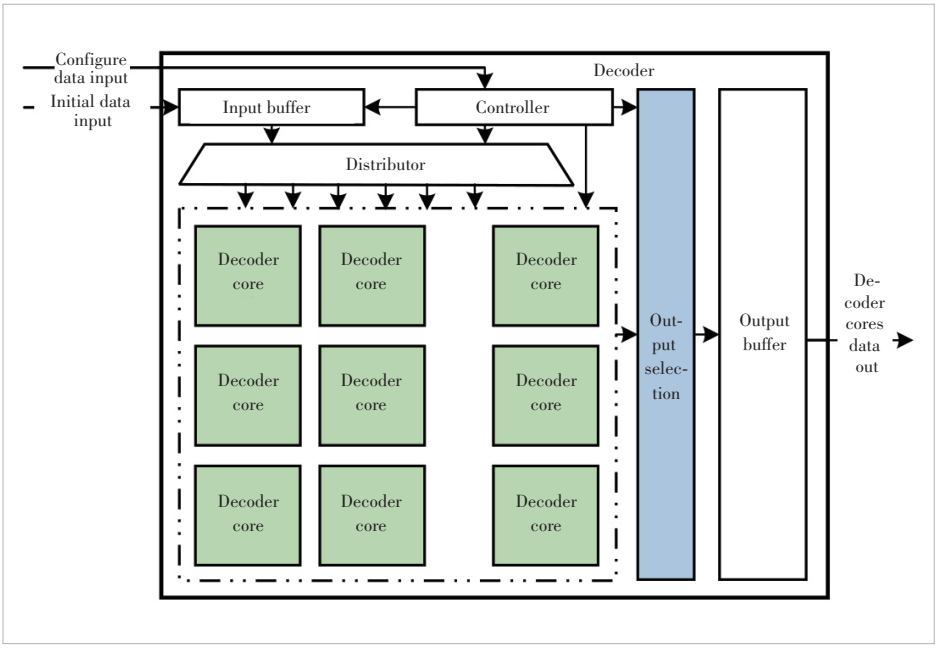
Input: the protograph P, the size of protograph M,N
Output: the parity check matrix $\boldsymbol{H}$
for (i = 0 to M) do
  for (j = 0 to N) do
    if $p_{i,j}$≠0 **then**
      $h_{i,j}$= random(0, Z). // Initializations
      else
      $h_{i,j}$=−1.
    end if
  end for
end for
n = 0.
while (n< Maximum number of iterations)
for (i = 0 to M) do
  for (j =0 to N) do
    if $p_{i,j}$≠0 **then**
      $g4_{i,j,n}$ = cycle4_calculate(H,i,j).
      $g6_{i,j,n}$ = cycle6_calculate(H,i,j). // Cycle calculation
      else
  $g4_{i,j,n}$=0.
      $g6_{i,j,n}$=0.

end if
end for
end for
$g4_{max\_i,max\_j} = MAX(g4_{i,j,n})$.
$h_{max\_i,max\_j} = random(0,Z)$.
$g6_{max\_i,max\_j} = MAX(g6_{i,j,n})$.
$h_{max\_i,max\_j} = random(0,Z)$.// Elemental position calculation.
$n = n + 1$.
if $MAX(g4_{i,j,n}) == MAX(g4_{i,j,n-1})$ and $MAX(g6_{i,j,n}) == MAX(g6_{i,j,n-1})$ **then**
Record current H. // Exit judgment
break.
end if
end while

3 Hardware Design

We design a multi-core decoder architecture to achieve a throughput of 100 Gbit/s. The architecture is shown in Fig. 4.

The multi-core decoder includes an input buffer, controller, decoder core, output buffer, and output selection module. The function of the output selection module is to select the decoded data from multiple decoder cores according to the controller configuration. The input buffer module converts the data fragments into a complete codeword. The input buffer module operates at a different higher frequency clock domain to accommodate the data stream with the internal blocks. The controller receives external configuration information, distributes the complete codeword to the free decoder core, and controls the output of the data frame at the end of the decoding.



▲**Figure 4. Multi-core decoder architecture**

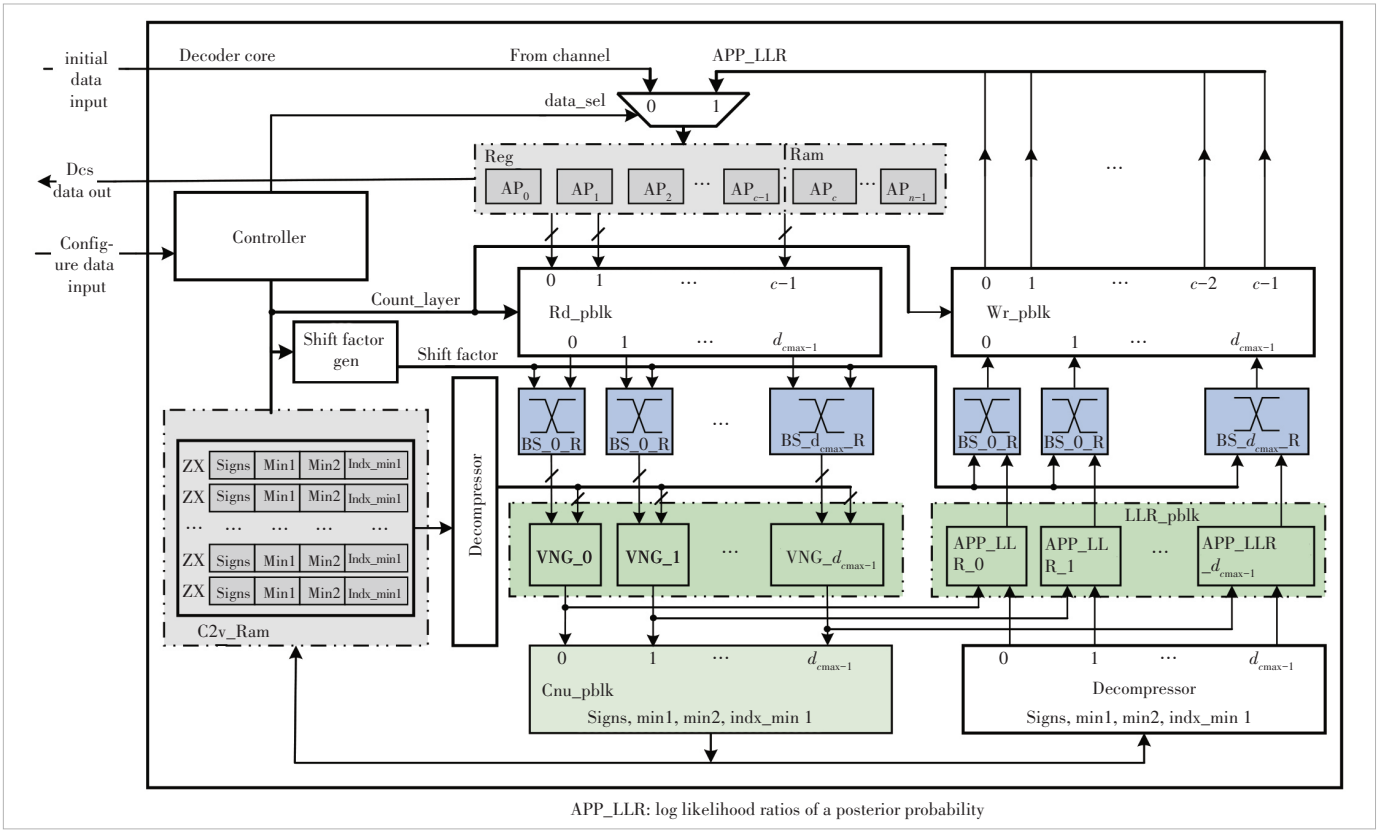
The decoder core, shown in Fig. 5, is the core computing module in the multi-core decoder architecture. It includes the log likelihood ratios of a posterior probability (APP_LLR) information storage module, check-node to variable-node (C2V) message storage module, and other main computing units.

The decoder core[14] receives data and configuration information from the controller. According to the configuration information, iterative decoding is carried out layer by layer. The initial log likelihood ratios (LLR) information or the updated LLR information of the current layer is read from the LLR storage module during the decoding of each layer. The LLR storage module contains two memory components: shift registers and static random-access memory (SRAM). The initial LLR information of each codeword to be decoded or the LLR information that is continuously updated during an iteration is stored in these two pieces of memory. The shift registers receive the initial information and shift according to the current layer index to accommodate the matrix's shift constraints during decoding.

The LLR information to be updated by up to d_{cmax} sets is selected and fed into barrel shifters by the reading network (Rd_pblk). Note that Rd_pblk is implemented by multiplexers in Ref. [15]. But based on our proposed special protograph structure, the Rd_pblk selects the same sets of LLR information from the shift registers without using multiplexers.

The selected information is aligned to the check node through the barrel shifters. In the same clock cycle, the compressed stored check-node to variable-node (C2V) messages are read from the C2V_Ram and distributed to the variable node processing array (VNU_pblk) through the decompressor. The LLR and C2V messages are input to the VNU_pblk to calculate the new variable-node to check-node (V2C) messages. The new V2C messages are input to the check node processing array (CNU_pblk) to calculate C2V messages. At last, the C2V messages are stored in the C2V_Ram.

New V2C messages and new C2V messages are added in the LLR_pblk to obtain new LLR information, which is then realigned to the VNs via barrel shifters. The aligned new LLR information is written back to the registers and RAM by writing network (Wr_pblk) according to the current layer index. Based on the special protograph structure, Wr_pblk only updates LLR information at fixed locations. The decoding of one layer is completed after the LLR information is written back. When the decoding of all layers is complete, the current it-



▲Figure 5. Decoder core architecture

eration is finished. The early termination function in the decoding process is implemented by LLR_pblk, which can calculate the parity check result of the current layer. When all the parity checks are satisfied, the decoding will stop successfully and enter the waiting state for the decision information output.

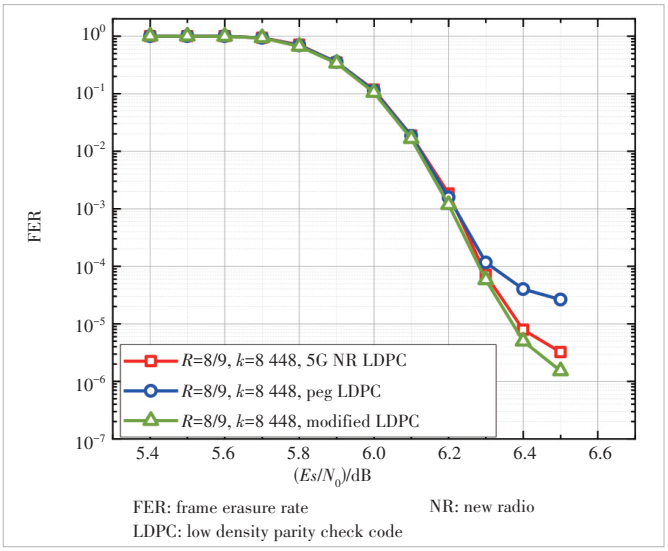
4 Implementation Results

4.1 Results of Improved Error Floor

We build a compute unified device architecture (CUDA)-based code performance testing platform, including a noise generator, a decoder, and a decoding result statistics module. Noise is added directly to the all-zero codeword to improve throughput during testing. And we test the code performance on a single GeForce GTX 1080 Ti GPU.

We evaluate the performance of the proposed LDPC codes in the additive white Gaussian noise (AWGN) channel using quadrature phase shift keying (QPSK) modulation. The belief propagation (BP) flooding decoding schedule is used and the maximum number of iterations is set to 50.

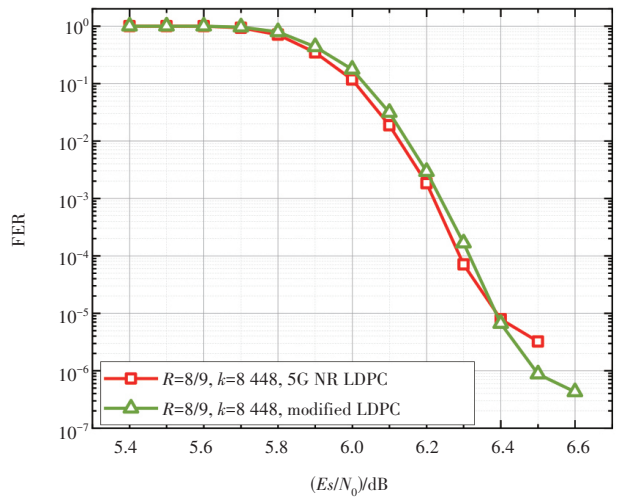
Compared with the 5G NR LDPC code and the PEG code, the code performance constructed by our algorithm has a certain optimization effect, as shown in Fig. 6. As we reduce the number of 6 cycles, the error floor is optimized.



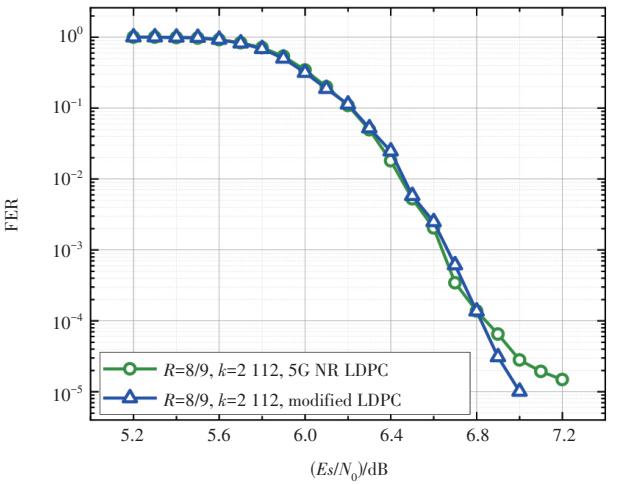
▲Figure 6. Performance comparison after optimizing the number of 6-cycle

4.2 Optimized LDPC Code Performance

After applying the code constraints, the code is constructed using the construction algorithm described above, and performance curves with an information blocklength of k=8 448 bits are shown in Fig. 7. Although the code we constructed still has a gap of less than 0.05 dB compared with the 5G NR



(a) Information blocklength is 8 848 bits



(b) Information blocklength is 2 112 bits

FER: frame erasure rate NR: new radio
LDPC: low density parity check code

▲**Figure 7. Performance comparison under rate of 8/9**

LDPC in the waterfall region, it has an improvement in the error floor region. The size of the parity check matrix we used during the test is 5×29, and the lifting size is 352.

The performance of the matrix with a shorter information blocklength of k=2 112 bits is shown in Fig. 7. The lifting size is 88. The code we constructed also has the advantage in the error floor region.

4.3 Hardware Implementation Results

For LDPC codes with a lifting size of 352, we design the corresponding decoder prototype to evaluate and compare the hardware resources accurately before and after adding the protograph construction constraints. The decoder consists of buffers, a controller, several decoder cores, etc. We use a quantization length of 6 bits for the LLR information and 4 bits for the internal messages.

The implementation results, as well as the corresponding comparisons, are reported in this section. The formula for calculating throughput is as follows:

$$\text{Throughput} = R \times \frac{N_{\text{cores}} \times n}{N_{\text{iter}} \times M} \times f_{\text{clk}}, \quad (2)$$

where R is the code rate, N_{cores} is the number of decoder cores, n is the length of the codeword, M is the number of basic matrix rows used for decoding, and N_{iter} is the average number of iterations. To achieve the desired throughput with the code, the information blocklength of which is 8 448 bits, 3 cores are required.

In order to evaluate the effectiveness of our optimization methods, the synthesis results of decoders with and without applying the optimizations proposed in Sections 2 and 3 are listed. We use the architecture from Ref. [15] as the baseline. Table 1 summarizes the hardware resources of the decoder core on the Xilinx Virtex UltraScale XCVU440 field programmable gate array (FPGA). The decoder core occupies about 34.0% of the lookup tables and 1.2% of the flip-flops available on the device. Compared with the architecture used in Ref. [15], the proposed decoder reduces the resources of the lookup table by about 15%.

▼**Table 1. Implementation results**

Resources	Implementation in Ref. [15]	Proposed
LUTs	1 015 935	861 574
FFs	62 460	62 460
Brams	66	66

Brams: block-RAMs FFs: flip-flops LUTs: lookup tables

5 Conclusions

This paper proposes an RL-LDPC code aiming at next-generation mobile communication, with a corresponding hardware architecture with a throughout up to 100 Gbit/s. In this code, the designed structure of the protograph is different from the conventional 5G NR LDPC basic matrix, and an improved code construction algorithm is investigated for better performance. The decoder implemented in our work can achieve a throughput higher than 100 Gbit/s, and compared with the traditional architecture, the resources of the lookup table are reduced by about 15% on Xilinx Virtex UltraScale XCVU440 FPGA.

References

[1] VENUGOPAL T, RADHIKA S. A survey on channel coding in wireless networks [C]//Proceedings of 2020 International Conference on Communication and Signal Processing (ICCSP). IEEE, 2020: 784 – 789. DOI: 10.1109/ICCSP48568.2020.9182213
[2] GALLAGER R. Low-density parity-check codes [J]. IRE transactions on infor-

mation theory, 1962, 8(1): 21 – 28. DOI: 10.1109/TIT.1962.1057683
[3] MACKAY D J C, NEAL R M. Near Shannon limit performance of low density parity check codes [J]. Electronics letters, 1996, 32(18): 1645. DOI: 10.1049/el:19961141
[4] HU X Y, ELEFTHERIOU E, ARNOLD D M. Regular and irregular progressive edge-growth tanner graphs [J]. IEEE transactions on information theory, 2005, 51(1): 386 – 398. DOI: 10.1109/TIT.2004.839541
[5] DIOUF M, DECLERCQ D, FOSSORIER M, et al. Improved PEG construction of large girth QC-LDPC codes [C]//The 9th International Symposium on Turbo Codes and Iterative Information Processing (ISTC). IEEE, 2016: 146 – 150. DOI: 10.1109/ISTC.2016.7593094
[6] LIU Y H, ZHANG M L, NIU X L. Quasi-cyclic low-density parity-check codes based on progressive cycle growth algorithm [C]//The sixth International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC). IEEE, 2016: 854 – 857. DOI: 10.1109/IMCCC.2016.167
[7] TIAN T, JONES C R, VILLASENOR J D, et al. Selective avoidance of cycles in irregular LDPC code construction [J]. IEEE transactions on communications, 2004, 52(8): 1242 – 1247. DOI: 10.1109/TCOMM.2004.833048
[8] CHENG C C, YANG J D, LEE H C, et al. A fully parallel LDPC decoder architecture using probabilistic Min-sum algorithm for high-throughput applications [J]. IEEE transactions on circuits and systems I: regular papers, 2014, 61(9): 2738 – 2746. DOI: 10.1109/TCSI.2014.2312479
[9] LI M, DERUDDER V, DESSET C, et al. A 100 gbps LDPC decoder for the IEEE 802.11ay standard [C]//The 10th International Symposium on Turbo Codes & Iterative Information Processing (ISTC). IEEE, 2019: 1 – 5. DOI: 10.1109/ISTC.2018.8625286
[10] BONCALO O, AMARICAI A. Ultra high throughput unrolled layered architecture for QC-LDPC decoders [C]//IEEE Computer Society Annual Symposium on VLSI (ISVLSI). IEEE, 2017: 225 – 230. DOI: 10.1109/ISVLSI.2017.47
[11] LI M, DERUDDER V, BERTRAND K, et al. High-speed LDPC decoders towards 1 Tb/S [J]. IEEE transactions on circuits and systems I: regular papers, 2021, 68(5): 2224 – 2233. DOI: 10.1109/TCSI.2021.3060880
[12] FANG Y, BI G A, GUAN Y L, et al. A survey on protograph LDPC codes and their applications [J]. IEEE communications surveys & tutorials, 2015, 17(4): 1989 – 2016. DOI: 10.1109/COMST.2015.2436705
[13] CHEN T Y, DIVSALAR D, WESEL R D. Protograph-based Raptor-like LDPC codes with low thresholds [C]//IEEE International Conference on Communications (ICC). IEEE, 2012: 2161 – 2165. DOI: 10.1109/ICC.2012.6363996
[14] NGUYEN-LY T T, GUPTA T, PEZZIN M, et al. Flexible, cost-efficient, high-throughput architecture for layered LDPC decoders with fully-parallel processing units [C]//Euromicro Conference on Digital System Design (DSD). IEEE, 2016: 230 – 237. DOI: 10.1109/DSD.2016.33
[15] CUI H X, GHAFFARI F, LE K, et al. Design of high-performance and area-efficient decoder for 5G LDPC codes [J]. IEEE transactions on circuits and systems I: regular papers, 2021, 68(2): 879 – 891. DOI: 10.1109/TCSI.2020.3038887

Biographies

LI Hanwen received his BS degree from Fuzhou University, China in 2020, and MS degree from Nanjing University, China in 2023. His research interests include very-large scale integration design for digital signal processing and error control coding.

BI Ningjing received her BS degree and MS degree in electrical engineering from Nanjing University, China, in 2020 and 2023, respectively. Her research interest focuses on digital very-large scale integration design.

SHA Jin (shajin@nju.edu.cn) received his BS degree in physics and PhD degree in electrical engineering from Nanjing University, China in 2002 and 2007, respectively. From 2012 to 2013, he was a visiting professor with Stanford University, USA. He is currently an associate professor with School of Electronic Science and Engineering, Nanjing University. His current research interests include very-large scale integration design for digital signal processing, error control coding, flash memory error control, and heterogeneous computing systems.

Hybrid Architecture and Beamforming Optimization for Millimeter Wave Systems

TANG Yuanqi[1], ZHANG Huimin[1], ZHENG Zheng[2], LI Ping[2], ZHU Yu[1]
(1. Department of Communication Science and Engineering, Fudan University, Shanghai 200433, China;
2. ZTE Corporation, Shenzhen 518057, China)

DOI: 10.12142/ZTECOM.202303013
https://kns.cnki.net/kcms/detail/34.1294.TN.20230726.1546.002.html, published online July 27, 2023
Manuscript received: 2023-02-21

Abstract: Hybrid beamforming (HBF) has become an attractive and important technology in massive multiple-input multiple-output (MIMO) millimeter-wave (mmWave) systems. There are different hybrid architectures in HBF depending on different connection strategies of the phase shifter network between antennas and radio frequency chains. This paper investigates HBF optimization with different hybrid architectures in broadband point-to-point mmWave MIMO systems. The joint hybrid architecture and beamforming optimization problem is divided into two sub-problems. First, we transform the spectral efficiency maximization problem into an equivalent weighted mean squared error minimization problem, and propose an algorithm based on the manifold optimization method for the hybrid beamformer with a fixed hybrid architecture. The overlapped subarray architecture which balances well between hardware costs and system performance is investigated. We further propose an algorithm to dynamically partition antenna subarrays and combine it with the HBF optimization algorithm. Simulation results are presented to demonstrate the performance improvement of our proposed algorithms.

Keywords: hybrid beamforming; hybrid architecture; weighted mean square error; manifold optimization; dynamic subarrays

Citation (Format 1): TANG Y Q, ZHANG H M, ZHENG Z, et al. Hybrid architecture and beamforming optimization for millimeter wave systems [J]. ZTE Communications, 2023, 21(3): 93 - 104. DOI: 10.12142/ZTECOM.202303013
Citation (Format 2): Y. Q. Tang, H. M. Zhang, Z. Zheng, et al., "Hybrid architecture and beamforming optimization for millimeter wave systems," *ZTE Communications*, vol. 21, no. 3, pp. 93 - 104, Sept. 2023. doi: 10.12142/ZTECOM.202303013.

1 Introduction

Millimeter-wave (mmWave) communication has emerged as a key technology for fifth-generation (5G) wireless networks to cope with the dilemma between scarce sub-6 GHz spectrum resources and people's rapidly growing demand for higher data transmission[1-4]. MmWave's short wavelength makes it convenient to arrange large-scale antenna arrays at the transceiver end to compensate for the high propagation loss, and thus massive multiple-input and multiple-output (MIMO) becomes attractive in mmWave systems. However, conventional fully digital beamforming (FDBF) requires a separate radio frequency (RF) chain for each antenna and will result in huge hardware costs and power consumption in massive MIMO systems[5]. Therefore, hybrid beamforming (HBF) that requires very few RF chains has become a research hotspot recently[6-8].

This work was supported by ZTE Industry-University-Institute Cooperation Funds, the Natural Science Foundation of Shanghai under Grant No. 23ZR1407300, and the National Natural Science Foundation of China under Grant No. 61771147.

HBF has different hybrid architectures depending on different connection strategies between antennas and RF chains. The fully-connected (FC) architecture and the partially-connected (PC) architecture are two conventional hybrid architectures. Most previous works considered the FC architecture. In Ref. [9], the authors applied the orthogonal matching pursuit algorithm to design the column vectors of the analog precoding matrix based on codebooks. The authors in Ref. [10] proposed an HBF algorithm based on the coordinate update iteration method in the narrowband point-to-point MIMO system. The authors in Ref. [11] proposed an alternating minimization algorithm based on manifold optimization (MO) by minimizing the Frobenius norm between the HBF matrix and the FDBF matrix. The authors in Ref. [12] took the mean square error minimization as the optimization goal and the designed HBF algorithms based on MO and generalized eigenvalue decomposition (EVD).

On the other hand, the PC architecture, where each antenna is only connected to only one RF chain instead of all RF chains, can reduce the power consumption and hardware costs compared with the FC architecture at the cost of certain system performance loss. A low-complexity HBF optimization al-

gorithm based on the positive semi-definite relaxation for the PC architecture has been proposed in Ref. [11]. The HBF optimization algorithms based on element iteration and MO for the PC architecture in the broadband system have also been proposed in Ref. [13].

To achieve a good compromise between hardware costs and system performance, other fixed hybrid architectures have also attracted research attention recently. The authors in Ref. [14] proposed a partially-fully connected architecture that combines the FC and PC architectures and designed algorithms based on continuous interference cancellation and matrix factorization. The authors in Ref. [15] designed an alternative minimization algorithm for this architecture. An overlapped (OL) subarray architecture and a heuristic unified low-rank sparse recovery algorithm were proposed in Ref. [16]. The authors in Ref. [17] proposed a generalized subarray-connected architecture, and developed a successive interference cancellation-based HBF algorithm along with an exhaustive search algorithm to maximize the system energy efficiency.

Since the hardware costs and power consumption of switches in mmWave massive MIMO systems are relatively small[18-19], the dynamic hybrid architecture becomes a promising approach to achieving a better balance between hardware costs and system performance. The authors in Ref. [20] proposed a greedy algorithm with low complexity to partition the antennas over RF chains. A low complexity algorithm to design the optimal partition using statistical channel state information was proposed in Ref. [21]. The authors in Ref. [22] considered the scenario of ultra-wideband mmWave and terahertz frequency band and decomposed the precoding problem into multiple subproblems under the FC architecture.

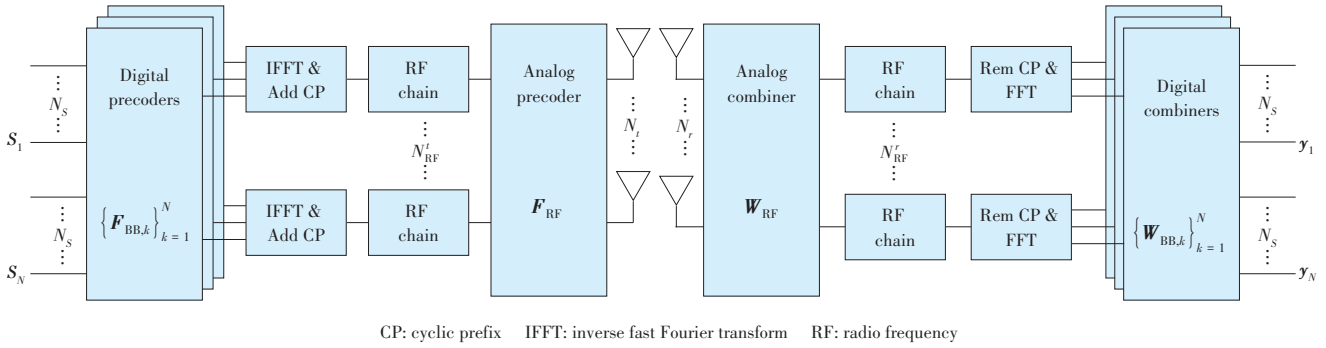
In this paper, we investigate the HBF algorithms with different hybrid architectures for broadband mmWave massive MIMO systems, aiming at maximizing the spectral efficiency. Based on the equivalence between the spectral efficiency maximization (SEM) problem and the weighted minimum mean square error minimization (WMMSE) problem, we design the beamforming optimization algorithm to directly tackle the original SEM optimization problem instead of the conventional indirect design approach of approximating the FDBF matrix with the HBF matrix. We adopt the alternating minimization method to decompose the joint transmitting and receiving HBF optimization problem into two sub-problems. It shows that both the digital precoding and combining optimization sub-problems have closed-form optimal solutions. To further optimize the analog precoder and combiner, we apply the MO method to deal with the constant modulus constraint. In contrast to Ref. [11], where the MO method was applied to solve the matrix approximation problem with the objective of minimizing the Frobenius norm between the FDBF matrix and the HBF matrix of the FC architecture, in our work, the MO method is applied to solve the HBF problem with the WMMSE objective and for arbitrary hybrid architectures by introducing the Hadamard product of the analog precoder and a connection matrix. Apart from the conventional FC and PC architectures, we consider the OL architecture and the PC architecture with dynamic subarrays (PC-dynamic architecture). In particular, we simulate three specific types of fixed OL architectures with a uniform planar array (UPA) and find that our proposed HBF optimization algorithm could achieve a compromise between hardware costs and system performance compared with conventional fixed architectures. Besides, for the PC-dynamic architecture, we derive a lower bound of the original WMMSE objective, based on which, and with some approximations we formulate an eigenvalue maximization problem. Then, we propose a greedy partition algorithm to optimize the dynamic partition of subarrays. Simulation results show that the PC-dynamic architecture with the proposed dynamic partition algorithm can achieve significant performance improvement over the fixed PC architecture.

We denote matrices and vectors by boldface capitals and lower-case letters respectively. $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and the complex conjugate transpose of a matrix or vector, respectively. $\mathrm{tr}(\cdot)$ and $\|\cdot\|_F$ represent the trace and the Frobenius norm of a matrix, respectively. $\mathbb{E}[\cdot]$ is the statistical expectation, $\odot$ is the Hadamard product of two matrices, $\boldsymbol{I}_N$ denotes the $N \times N$ identity matrix, and $CN(0, \boldsymbol{K})$ represents the circularly symmetric complex Gaussian distribution with zero mean and covariance matrix $\boldsymbol{K}$.

2 System Model and Problem Formulation

2.1 System Model

In this paper, we consider the downlink of a broadband mmWave MIMO-orthogonal frequency division multiplexing (OFDM) system with HBF, as shown in Fig. 1. The transmitter first precodes N_s data streams, denoted by the vector $\boldsymbol{s}_k \in \mathbb{C}^{N_s \times 1}$, and at the k-th subcarrier uses a digital precoder $\boldsymbol{F}_{\mathrm{BB},k} \in \mathbb{C}^{N_{\mathrm{RF}}^t \times N_s}$, for $k = 0, \cdots, N-1$ with N denoting the number of subcarriers. Then, N_{RF}^t output streams are transformed into the time domain by the N-point inverse fast Fourier transform. After adding cyclic prefixes (CPs), the signals are further precoded by an analog precoder $\boldsymbol{F}_{\mathrm{RF}} \in \mathbb{C}^{N_t \times N_{\mathrm{RF}}^t}$ composed of a number of phase shifters. It is worth noting that in the HBF design for broadband systems, the digital beamformers can be optimized for different subcarriers, in contrast, the analog one is invariant for the whole frequency band and thus $\boldsymbol{F}_{\mathrm{RF}}$ is not related to the subcarrier index. It is also worth noting that $\boldsymbol{F}_{\mathrm{RF}}$ can represent different hybrid architectures. In particular, we define a connection matrix $\boldsymbol{U}_{\mathrm{p}} \in \mathbb{C}^{N_t \times N_{\mathrm{RF}}^t}$, $\left[\boldsymbol{U}_{\mathrm{p}}\right]_{ij} \in \{0,1\}$ to represent the connection strategy with any specific hybrid architecture, where $\left[\boldsymbol{U}_{\mathrm{p}}\right]_{i,j} = 1$ indicates that the j-th RF chain is connected to the i-th antenna.



▲Figure 1. Downlink single-user mmWave multiple-input multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) system with hybrid beamforming (HBF)

The analog beamformer with any arbitrary fixed hybrid architecture can be represented by:

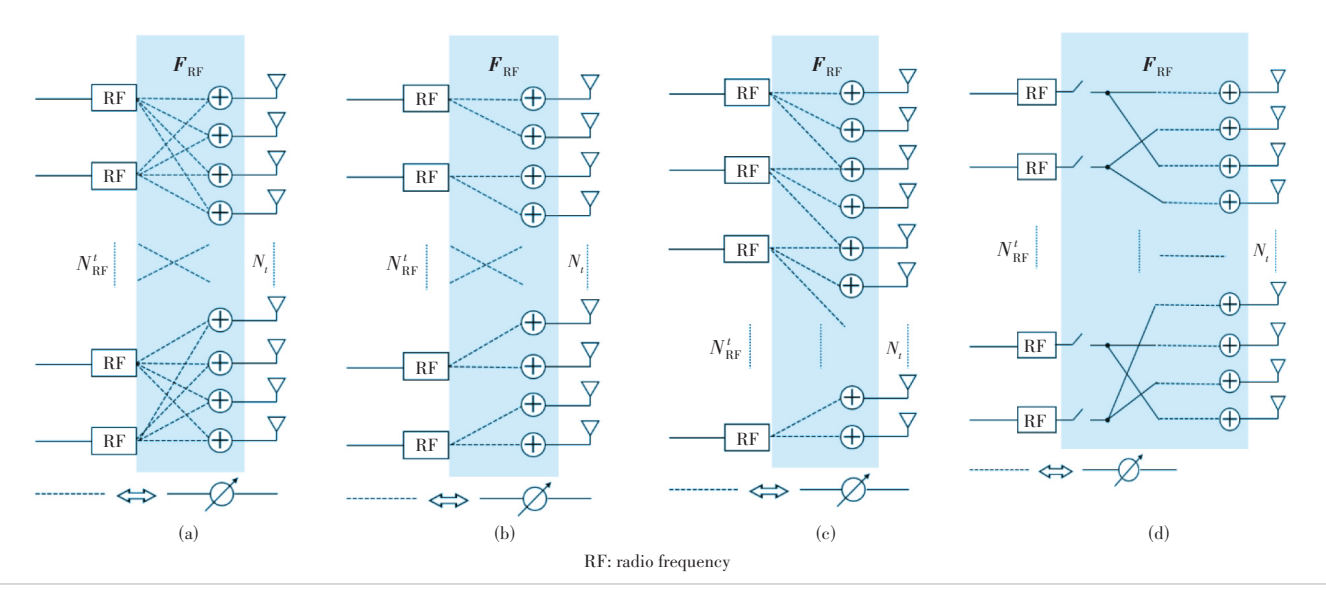
$$\boldsymbol{F}_{\mathrm{RF}} = \boldsymbol{F}_{\mathrm{RF}}^{\mathrm{FC}} \odot \boldsymbol{U}_{\mathrm{p}}, \tag{1}$$

$$\boldsymbol{W}_{\mathrm{RF}} = \boldsymbol{W}_{\mathrm{RF}}^{\mathrm{FC}} \odot \boldsymbol{U}_{\mathrm{c}}, \tag{2}$$

where $\boldsymbol{F}_{\mathrm{RF}}^{\mathrm{FC}}$ and $\boldsymbol{W}_{\mathrm{RF}}^{\mathrm{FC}}$ represent the analog precoder and combiner with the FC architecture, respectively.

We consider four hybrid architectures for analog beamforming as depicted in Fig. 2. In the FC architecture shown in Fig. 2(a), each RF chain is connected to all antenna elements so that a total of $N_t N_{\mathrm{RF}}^t$ phase shifters are required. In the PC architecture shown in Fig. 2(b), each RF chain is connected to an antenna subarray while each antenna is connected to only one RF chain, so that a total number of N_t phase shifters are required. In the OL architecture shown in Fig. 2(c), the antenna subarrays connected to each RF chain can overlap, where the overlapped antennas are connected to multiple RF chains at the same time. The number of phase shifters required lies between $\left[N_t, N_t N_{\mathrm{RF}}^t\right]$. In the PC-dynamic architecture based on a switch network in Fig. 2(d), the partition of the antenna subarrays can be dynamically adjusted by turning on or off the switches according to the system state, and a total number of N_t phase shifters and $N_t N_{\mathrm{RF}}^t$ switches are required.

The transmitted signal at the k-th subcarrier via N_t antennas is represented by $\boldsymbol{x}_k = \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB},k} \boldsymbol{s}_k$, where $\boldsymbol{F}_{\mathrm{BB},k}$ and $\boldsymbol{F}_{\mathrm{RF}}$ satisfy the power constraint $\left\| \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB},k} \right\|_F^2 \leqslant 1$. After passing through the channel matrix at the k-th subcarrier $\boldsymbol{H}_k \in \mathbb{C}^{N_r \times N_t}$, the signals reach the receiver which is equipped with N_r antennas. The received signals are first processed by an analog combiner $\boldsymbol{W}_{\mathrm{RF}} \in \mathbb{C}^{N_r \times N_{\mathrm{RF}}^t}$, which is also shared by all subcarriers. Then, after removing CPs and performing the fast Fourier transform, a digital combiner $\boldsymbol{W}_{\mathrm{BB},k} \in \mathbb{C}^{N_{\mathrm{RF}}^t \times N_s}$ is deployed at



▲Figure 2. Diagram of four hybrid architectures

each subcarrier. Finally, the processed signal at the k-th subcarrier can be expressed as:

$$\boldsymbol{y}_k = \boldsymbol{W}_{\mathrm{BB},k}^H \boldsymbol{W}_{\mathrm{RF}}^H \boldsymbol{H}_k \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB},k} \boldsymbol{s}_k + \boldsymbol{W}_{\mathrm{BB},k}^H \boldsymbol{W}_{\mathrm{RF}}^H \boldsymbol{n}_k, \quad \text{for } k = 0,\cdots,N-1, \tag{3}$$

where $\boldsymbol{n}_k \sim CN(0, {\sigma_k}^2 \boldsymbol{I}_{N_r})$ denotes the additive white Gaussian noise at the k-th subcarrier.

2.2 Channel Model

We consider the clustered delay line (CDL) model developed by 3GPP, which takes the mmWave propagation characteristics into account, and can better characterize spatial correlations for 3D channels. Besides the normalized delay and the power, the azimuth angle of departure (AOD), the azimuth angle of arrival (AOA), the zenith angle of departure (ZOD), and the zenith angle of arrival (ZOA) are also defined in the CDL model. Three types of the CDL model, i.e., CDL-A, CDL-B and CDL-C, are constructed to represent different channel profiles for the non-line of sight (NLOS) scenarios, while two types, i.e., CDL-D and CDL-E, are constructed for the line-of-sight scenarios[23]. The NLOS channel coefficient of the n-th cluster with M rays between the transmit and receive antennas (u and s respectively) at time instant t and delay τ is given by:

$$H_{u,s,n}^{\mathrm{NLOS}}(t) = \sqrt{\frac{P_n}{M}} \sum_{m=1}^{M} \begin{bmatrix} F_{\mathrm{rx},u,\theta}\left(\theta_{n,m,\mathrm{ZOA}}, \phi_{n,m,\mathrm{AOA}}\right) \\ F_{\mathrm{rx},u,\phi}\left(\theta_{n,m,\mathrm{ZOA}}, \phi_{n,m,AOA}\right) \end{bmatrix}^T \times$$
$$\begin{bmatrix} \exp\left(j\Phi_{n,m}^{\theta\theta}\right) & \sqrt{K_{n,m}^{-1}} \exp\left(j\Phi_{n,m}^{\theta\phi}\right) \\ \sqrt{K_{n,m}^{-1}} \exp\left(j\Phi_{n,m}^{\phi\theta}\right) & \exp\left(j\Phi_{n,m}^{\phi\phi}\right) \end{bmatrix} \begin{bmatrix} F_{\mathrm{tx},u,\theta}\left(\theta_{n,m,\mathrm{ZOD}}, \phi_{n,m,\mathrm{AOD}}\right) \\ F_{\mathrm{tx},u,\phi}\left(\theta_{n,m,\mathrm{ZOD}}, \phi_{n,m,\mathrm{AOD}}\right) \end{bmatrix} \times$$
$$\exp\left(j2\pi\lambda_0^{-1}\left(r_{\mathrm{rx},n,m}^T \dot{d}_{\mathrm{rx},s}\right)\right) \exp\left(j2\pi\lambda_0^{-1}\left(r_{\mathrm{tx},n,m}^T \cdot \dot{d}_{\mathrm{tx},s}\right)\right) \exp\left(j2\pi v_{n,m} t\right),$$

where the definitions of parameters are given in Table 1.

2.3 Problem Formulation

We jointly optimize the hybrid architecture and the hybrid beamformers to maximize the spectral efficiency over N subcarriers subject to the constant modulus constraint of the analog beamformers and the power constraint of the transmitter. The problem can be formulated as follows:

▼ Table 1. Definitions of some parameters in the clustered delay line (CDL) channel model

Parameter	Definition
P_n	Power of the n-th cluster
$F_{\mathrm{rx},u}, F_{\mathrm{tx},s}$	Radiation patterns of the receiving and the transmitting antennas
$\Phi_{n,m}$	Random initial phases of different polarization combinations
$K_{n,m}$	Cross polarization power ratio for the m-th ray in the n-th cluster
λ_0	Carrier wavelength
$r_{\mathrm{rx},n,m}, r_{\mathrm{tx},n,m}$	Spherical unit vectors of the receiving and the transmitting antennas
$\boldsymbol{v}_{n,m}$	Velocity vector

$$\begin{aligned} \underset{\boldsymbol{F}_{\mathrm{BB},k}, \boldsymbol{F}_{\mathrm{RF}}^{\mathrm{FC}}, \boldsymbol{W}_{\mathrm{RF}}^{\mathrm{FC}}, \boldsymbol{W}_{\mathrm{BB},k}, \boldsymbol{U}_p, \boldsymbol{U}_c}{\text{maximize}} \quad & \frac{1}{N} \sum_{k=1}^{N} R_k \\ \text{subject to} \quad & \left\| \boldsymbol{F}_k \right\|_{\mathrm{F}}^2 \leqslant 1, \forall k; \\ & \left| \left[\boldsymbol{F}_{\mathrm{RF}}^{\mathrm{FC}} \right]_{ij} \right| = 1, \left| \left[\boldsymbol{W}_{\mathrm{RF}}^{\mathrm{FC}} \right]_{ij} \right| = 1, \forall i,j; \\ & \left[\boldsymbol{U}_{\mathrm{p}} \right]_{ij}, \left[\boldsymbol{U}_{\mathrm{c}} \right]_{ij} \in 0,1, \forall i,j; \\ & \left\| \boldsymbol{U}_p \right\|_1 = N_{p1}, \left(\boldsymbol{U}_p \right)_2 = N_{p2}, \end{aligned} \tag{4}$$

where $R_k = \log \left| \boldsymbol{I}_{N_s} + \sigma_k^2 \left(\boldsymbol{W}_k{}^H \boldsymbol{W}_k \right)^{-1} \boldsymbol{W}_k{}^H \boldsymbol{H}_k \boldsymbol{F}_k \boldsymbol{F}_k^H \boldsymbol{H}_k^H \boldsymbol{W}_k \right|$ is the achievable spectral efficiency at each subcarrier, and $\boldsymbol{F}_k = \left(\boldsymbol{F}_{\mathrm{RF}}^{\mathrm{FC}} \odot \boldsymbol{U}_{\mathrm{p}} \right) \boldsymbol{F}_{\mathrm{BB},k}$, $\boldsymbol{W}_k = \left(\boldsymbol{W}_{\mathrm{RF}}^{\mathrm{FC}} \odot \boldsymbol{U}_{\mathrm{c}} \right) \boldsymbol{W}_{\mathrm{BB},k}$. N_{p1} and N_{p2} are the predetermined numbers of phase shifters used at the transmitter and the receiver, respectively.

It has been proved in Ref. [13] that the SEM problem can be transformed into an equivalent WMMSE problem, which is more tractable. The modified mean square error (MSE) is defined as:

$$\boldsymbol{E} \triangleq \mathbb{E}\left[\left(\beta^{-1} \boldsymbol{y} - \boldsymbol{s} \right) \left(\beta^{-1} \boldsymbol{y} - \boldsymbol{s} \right)^H \right], \tag{5}$$

where β is a scaling factor to be jointly optimized with the hybrid beamformers. The WMMSE problem in the broadband scenario can be formulated as:

$$\begin{aligned} \underset{\boldsymbol{F}_{\mathrm{U},k}, \boldsymbol{F}_{\mathrm{RF}}^{\mathrm{FC}}, \boldsymbol{W}_{\mathrm{RF}}^{\mathrm{FC}}, \boldsymbol{W}_{\mathrm{BB},k}, \boldsymbol{U}_{\mathrm{p}}, \boldsymbol{U}_{\mathrm{c}}, \beta_k, \boldsymbol{T}_k}{\text{minimize}} \quad & \frac{1}{N} \sum_{k=1}^{N} \left(\boldsymbol{T}_k \boldsymbol{E}_k \right) - \log \left| \boldsymbol{T}_k \right| \\ \text{subject to} \quad & \left(\boldsymbol{F}_k \right)_{\mathrm{F}}^2 \leqslant 1, \forall k; \\ & \left| \left[\boldsymbol{F}_{\mathrm{RF}}^{\mathrm{FC}} \right]_{ij} \right| = 1, \left| \left[\boldsymbol{W}_{\mathrm{RF}}^{\mathrm{FC}} \right]_{ij} \right| = 1, \forall i,j; \\ & \left[\boldsymbol{U}_{\mathrm{p}} \right]_{ij}, \left[\boldsymbol{U}_{\mathrm{c}} \right]_{ij} \in \{0,1\}, \forall i,j; \\ & \left(\boldsymbol{U}_{\mathrm{p}} \right)_1 = N_{\mathrm{p1}}, \left(\boldsymbol{U}_{\mathrm{c}} \right)_2 = N_{\mathrm{p2}}, \end{aligned} \tag{6}$$

where $\boldsymbol{T}_k$ and $\boldsymbol{E}_k = \boldsymbol{I}_{N_s} - \beta_k^{-1} \boldsymbol{F}_k^H \boldsymbol{H}_k^H \boldsymbol{W}_k - \beta_k^{-1} \boldsymbol{W}_k^H \boldsymbol{H}_k \boldsymbol{F}_k + \beta_k^{-2} \sigma_k^2 \boldsymbol{W}_k^H \boldsymbol{W}_k + \beta_k^{-2} \boldsymbol{W}_k^H \boldsymbol{H}_k \boldsymbol{F}_k \boldsymbol{F}_k^H \boldsymbol{H}_k^H \boldsymbol{W}_k$ are respectively the weight matrix and the MSE matrix for the k-th subcarrier, and $\boldsymbol{F}_{\mathrm{U},k} = \beta_k^{-1} \boldsymbol{F}_{\mathrm{BB},k}$. Since the joint optimization of the hybrid beamforming and architecture is hard to solve, we decompose the problem into two subproblems: the HBF optimization problem with a fixed architecture and the architecture optimization problem.

3 HBF Optimization with Fixed Architecture

In this section, we apply the alternating minimization method and the MO method to optimize the hybrid beamformers with a fixed hybrid architecture. The WMMSE problem is formulated as:

$$\begin{aligned}\underset{\boldsymbol{F}_{\mathrm{U},k},\boldsymbol{F}_{\mathrm{RF}},\boldsymbol{W}_{\mathrm{RF}},\boldsymbol{W}_{\mathrm{BB},k}\beta_k,\boldsymbol{T}_k}{\text{minimize}} \quad & \frac{1}{N}\sum_{k=1}^{N}\left(\boldsymbol{T}_k\boldsymbol{E}_k\right)-\log\left|\boldsymbol{T}_k\right| \\ \text{subject to} \quad & \left(\boldsymbol{F}_{\mathrm{U},k}\boldsymbol{F}_{\mathrm{RF}}\right)_{\mathrm{F}}^{2}\leqslant\beta_k^{-2},\forall k; \\ & \left|\left[\boldsymbol{F}_{\mathrm{RF}}\right]_{ij}\right|=\begin{cases}1,i,j\in\left\{i,j\middle|\left|\left[\boldsymbol{U}_{\mathrm{p}}\right]_{ij}\right|=1\right\}\\ 0,\text{ else}\end{cases}; \\ & \left|\left[\boldsymbol{W}_{\mathrm{RF}}\right]_{ij}\right|=\begin{cases}1,i,j\in\left\{i,j\middle|\left|\left[\boldsymbol{U}_{\mathrm{c}}\right]_{ij}\right|=1\right\}\\ 0,\text{ else}\end{cases},\forall i,j\end{aligned} \tag{7}$$

Since the joint optimization problem of the five variables in Eq. (7) is hard to solve, we adopt the alternating minimization method to decouple the optimization of the transmitter and receiver and solve the two subproblems separately.

3.1 Transmitter Design

In this subsection, we fix the hybrid combiner $\boldsymbol{W}_k$ and optimize the hybrid precoder. Firstly, the closed form solution of $\boldsymbol{T}_k$ can be obtained as follows by differentiating the objective function with respect to $\boldsymbol{T}_k$

$$\boldsymbol{T}_k=\boldsymbol{E}_k^{-1}. \tag{8}$$

Secondly, the optimal digital precoder $\boldsymbol{F}_{\mathrm{BB},k}$ and the scaling factor β_k at each subcarrier can be derived with fixed $\boldsymbol{F}_{\mathrm{RF}}$. Considering the power constraint, it can be proved that the optimal β_k can only be achieved with the maximum transmit power, and the optimal β_k is given by:

$$\beta_k=1/\sqrt{\left\|\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{U},k}\boldsymbol{F}_{\mathrm{U},k}^{H}\boldsymbol{F}_{\mathrm{RF}}^{H}\right\|_F^2}. \tag{9}$$

According to the Karush-Kuhn-Tucker (KKT) conditions, $\boldsymbol{F}_{U,k}$ has a closed-form solution as follows:

$$\boldsymbol{F}_{U,k}=\left(\boldsymbol{F}_{\mathrm{RF}}^{H}\boldsymbol{G}_k\boldsymbol{G}_k^{H}\boldsymbol{F}_{\mathrm{RF}}+\xi_k\boldsymbol{F}_{\mathrm{RF}}^{H}\boldsymbol{F}_{\mathrm{RF}}\right)^{-1}\boldsymbol{F}_{\mathrm{RF}}^{H}\boldsymbol{G}_k, \tag{10}$$

where $\xi_k=(\sigma_k^2\mathrm{tr}(\boldsymbol{T}_k\boldsymbol{W}_k^{H}\boldsymbol{W}_k))^{-1}$ and $\boldsymbol{G}_k=\boldsymbol{H}_k^{H}\boldsymbol{W}_k$.

Thirdly, by substituting $\boldsymbol{T}_k$,β_k and $\boldsymbol{F}_{\mathrm{BB},k}$ back into the original objective function, the optimization problem of $\boldsymbol{F}_{\mathrm{RF}}$ can be obtained as follows:

$$\begin{aligned}\underset{\boldsymbol{F}_{\mathrm{RF}}}{\text{minimize}} \quad & f\left(\boldsymbol{F}_{\mathrm{RF}}\right) \\ \text{subject to} \quad & \left|\left[\boldsymbol{F}_{\mathrm{RF}}\right]_{ij}\right|=\begin{cases}1,i,j\in\left\{i,j\middle|\left|\left[\boldsymbol{U}_{p}\right]_{ij}\right|=1\right\}\\ 0,\text{ else}\end{cases},\end{aligned} \tag{11}$$

where $f(\boldsymbol{F}_{\mathrm{RF}})=\frac{1}{N}\sum_{k=1}^{N}\mathrm{tr}\left(\left(\boldsymbol{T}_k^{-1}+\xi_k\boldsymbol{G}_k^{H}\boldsymbol{F}_{\mathrm{RF}}(\boldsymbol{F}_{\mathrm{RF}}^{H}\boldsymbol{F}_{\mathrm{RF}})^{-1}\boldsymbol{F}_{\mathrm{RF}}^{H}\boldsymbol{G}_k\right)^{-1}\right)$.

Next, we use the MO method to design $\boldsymbol{F}_{\mathrm{RF}}$. The basic idea is to define a Riemannian manifold considering the constant modulus constraint, and iteratively update $\boldsymbol{F}_{\mathrm{RF}}$ along the direction of the Riemann gradient in a way similar to the conventional Euclidean gradient descent algorithm[12]. The key is to derive the Euclidean conjugate gradient of $f\left(\boldsymbol{F}_{\mathrm{RF}}\right)$ with the FC architecture, which is given by:

$$\begin{aligned}\nabla_{\boldsymbol{F}_{\mathrm{RF}}^{\mathrm{FC}}}f\left(\boldsymbol{F}_{\mathrm{RF}}\right)=&\frac{1}{N}\sum_{k=1}^{N}\xi_k\left(\boldsymbol{F}_{\mathrm{RF}}\left(\boldsymbol{F}_{\mathrm{RF}}^{H}\boldsymbol{F}_{\mathrm{RF}}\right)^{-1}\boldsymbol{F}_{\mathrm{RF}}^{H}-\boldsymbol{I}_{N_t}\right)\\ &\boldsymbol{G}_k\boldsymbol{\Omega}_k^{-2}\boldsymbol{G}_k^{H}\boldsymbol{F}_{\mathrm{RF}}\left(\boldsymbol{F}_{\mathrm{RF}}^{H}\boldsymbol{F}_{\mathrm{RF}}\right)^{-1},\end{aligned} \tag{12}$$

where $\boldsymbol{\Omega}_k\triangleq\boldsymbol{T}_k^{-1}+\xi_k\boldsymbol{G}_k^{H}\boldsymbol{F}_{\mathrm{RF}}\left(\boldsymbol{F}_{\mathrm{RF}}^{H}\boldsymbol{F}_{\mathrm{RF}}\right)^{-1}\boldsymbol{F}_{\mathrm{RF}}^{H}\boldsymbol{G}_k$. Since $f\left(\boldsymbol{F}_{\mathrm{RF}}\right)$ is only related to the antennas that are connected to each RF chain with any specified hybrid architecture and calculating the gradient involves the derivative with respect to each entry of $\boldsymbol{F}_{\mathrm{RF}}$[13], with $\boldsymbol{F}_{\mathrm{RF}}=\boldsymbol{F}_{\mathrm{RF}}^{\mathrm{FC}}\odot\boldsymbol{U}_{\mathrm{p}}$, it can be shown that:

$$\nabla f\left(\boldsymbol{F}_{\mathrm{RF}}\right)=\nabla_{\boldsymbol{F}_{\mathrm{RF}}^{\mathrm{FC}}}f\left(\boldsymbol{F}_{\mathrm{RF}}\right)\odot\boldsymbol{U}_{\mathrm{p}}. \tag{13}$$

Then, we can obtain the Riemannian gradient by projecting the Euclidean gradient $\nabla f\left(\boldsymbol{F}_{\mathrm{RF}}\right)$ onto the tangent space, and update $\boldsymbol{F}_{\mathrm{RF}}$ with a proper step size determined by the well-known Armijo backtracking algorithm. Finally, the retraction operation is applied to make the result satisfy the constant modulus constraint[11] as follows:

$$\mu\boldsymbol{d}\mapsto\mathrm{Retr}_{\boldsymbol{x}}\left(\mu\boldsymbol{d}\right)=\mathrm{vec}\left[\frac{\left(\boldsymbol{x}+\mu\boldsymbol{d}\right)_i}{\left|\left(\boldsymbol{x}+\mu\boldsymbol{d}\right)_i\right|}\right]. \tag{14}$$

It is worth noting that with Eqs. (12) and (13), the above algorithm based on the MO method can be adopted in the HBF design with arbitrary hybrid architectures as there is no specific requirement to the connection matrix $\boldsymbol{U}_{\mathrm{p}}$. Finally, the precoder design with arbitrary fixed hybrid architectures is summarized in Algorithm 1.

Algorithm 1: Hybrid precoder design based on the MO method

Input: $\xi_k,\boldsymbol{G}_k,\boldsymbol{T}_k,\boldsymbol{U}_{\mathrm{p}}$
1: Initialize $\boldsymbol{F}_{\mathrm{RF},0}$ with random phases, $i=0$
2: **repeat**
3: Select the step size μ
4: Update $\mathrm{vec}\left(\boldsymbol{F}_{\mathrm{RF},i+1}\right)$ according to Eq. (14)
5: Update the Riemannian gradient $\boldsymbol{g}_i=\nabla f\left(\boldsymbol{F}_{\mathrm{RF},i+1}\right)$ according to Eqs. (12) and (13)
6: Calculate $\boldsymbol{g}_i^{+},\boldsymbol{d}_i^{+}$ from $\boldsymbol{x}_i$ to $\boldsymbol{x}_{i+1}$
7: Select Polak-Ribiere parameter η_{i+1}
8: Calculate the conjugate direction $\boldsymbol{d}_{i+1}=-\boldsymbol{g}_{i+1}+\eta_{i+1}\boldsymbol{d}_i^{+}$
9: Update $i\leftarrow i+1$
10: **until** a stopping condition is satisfied
Output: $\boldsymbol{F}_{\mathrm{RF}}$

3.2 Receiver Design

With the fixed hybrid precoder, we can get the optimization problem for the hybrid combiner. By differentiating Eq. (7) with respect to $\boldsymbol{W}_{\mathrm{BB},k}$, the closed-form solution of $\boldsymbol{W}_{\mathrm{BB},k}$ is given by:

$$\boldsymbol{W}_{\mathrm{BB},k} = \left(\boldsymbol{W}_{\mathrm{RF}}^H \tilde{\boldsymbol{G}}_k \tilde{\boldsymbol{G}}_k^H \boldsymbol{W}_{\mathrm{RF}} + \tilde{\xi}_k \boldsymbol{W}_{\mathrm{RF}}^H \boldsymbol{W}_{\mathrm{RF}}\right)^{-1} \boldsymbol{W}_{\mathrm{RF}}^H \tilde{\boldsymbol{G}}_k, \tag{15}$$

where $\tilde{\boldsymbol{G}}_k = \beta_k^{-1} \boldsymbol{H}_k \boldsymbol{F}_k$, $\tilde{\xi}_k = \sigma_k^2 \beta_k^{-2}$. By substituting Eq. (15) back into Eq. (7), we can get the optimization problem of $\boldsymbol{W}_{\mathrm{RF}}$ as follows:

$$\begin{aligned} & g\left(\boldsymbol{W}_{\mathrm{RF}}\right) = \\ \underset{\boldsymbol{W}_{\mathrm{RF}}}{\text{minimize}} \quad & \frac{1}{N}\sum_{k=1}^{N} \mathrm{tr}\left(\boldsymbol{T}_k\left(\boldsymbol{I}_{N_s} + \tilde{\xi}_k^{-1} \tilde{\boldsymbol{G}}_k^H \boldsymbol{W}_{\mathrm{RF}}\left(\boldsymbol{W}_{\mathrm{RF}}^H \boldsymbol{W}_{\mathrm{RF}}\right)^{-1} \boldsymbol{W}_{\mathrm{RF}}^H \tilde{\boldsymbol{G}}_k\right)^{-1}\right) \\ \text{subject to} \quad & \left|\left[\boldsymbol{W}_{\mathrm{RF}}\right]_{ij}\right| = \begin{cases} 1, i, j \in \left\{i, j \middle| \left|\left[\mathbf{U}_c\right]_{ij}\right| = 1\right\} \\ 0, \text{else} \end{cases}. \end{aligned} \tag{16}$$

This problem is difficult to tackle due to the non-convex constraint. However, since it has a similar form to the design problem of the analog precoder in Eq. (11), we can also adopt the MO method to optimize the analog combiner in the same way as we optimize the analog precoder. Firstly, the key step is to derive the Euclidean gradient of $g\left(\boldsymbol{W}_{\mathrm{RF}}\right)$ with the FC architecture, which is given by:

$$\begin{aligned} g\left(\boldsymbol{W}_{\mathrm{RF}}\right) = & \frac{1}{N}\sum_{k=1}^{N} \tilde{\xi}_k \left(\boldsymbol{W}_{\mathrm{RF}}\left(\boldsymbol{W}_{\mathrm{RF}}^H \boldsymbol{W}_{\mathrm{RF}}\right)^{-1} \boldsymbol{W}_{\mathrm{RF}}^H - \right. \\ & \left. \boldsymbol{I}_{N_t}\right) \tilde{\boldsymbol{G}}_k \boldsymbol{T}_k \boldsymbol{J}_k^{-2} \tilde{\boldsymbol{G}}_k^H \boldsymbol{W}_{\mathrm{RF}}\left(\boldsymbol{W}_{\mathrm{RF}}^H \boldsymbol{W}_{\mathrm{RF}}\right)^{-1}, \end{aligned} \tag{17}$$

where $\boldsymbol{J}_k = \boldsymbol{T}_k\left(\boldsymbol{I}_{N_s} + \tilde{\xi}_k \tilde{\boldsymbol{G}}_k^H \boldsymbol{W}_{\mathrm{RF}}\left(\boldsymbol{W}_{\mathrm{RF}}^H \boldsymbol{W}_{\mathrm{RF}}\right) \boldsymbol{W}_{\mathrm{RF}}^H \tilde{\boldsymbol{G}}_k\right)$. Secondly, we can use the formula $\nabla g\left(\boldsymbol{W}_{\mathrm{RF}}\right) = \nabla_{\boldsymbol{W}_{\mathrm{RF}}^{\mathrm{FC}}} g\left(\boldsymbol{W}_{\mathrm{RF}}\right) \odot \mathbf{U}_c$ to obtain the Euclidean gradient of $g\left(\boldsymbol{W}_{\mathrm{RF}}\right)$ with any specified architecture. Finally, with the derived gradient, we can optimize $\boldsymbol{W}_{\mathrm{RF}}$ by applying a procedure similar to that in Algorithm 1.

3.3 Alternating Optimization

We develop a joint hybrid precoding and combining optimization algorithm based on the WMMSE criterion by iteratively and alternatively using Algorithm 1. During each iteration, with the fixed hybrid combiner $\boldsymbol{W}_k$ and the weight matrix $\boldsymbol{T}_k$, we first optimize $\boldsymbol{F}_{\mathrm{RF}}$,$\boldsymbol{F}_{\mathrm{BB},k}$ according to Algorithm 1. Then, with the fixed hybrid precoder, we optimize $\boldsymbol{W}_{\mathrm{RF}}$,$\boldsymbol{W}_{\mathrm{BB},k}$. Finally, we update $\boldsymbol{T}_k$ according to Eq. (8). These steps are repeated until the stopping condition is satisfied. The stopping condition could be set as a maximum number of iterations or it depends on whether the relative difference between the objective function values of two consecutive iterations is smaller than a specific value. The HBF algorithm with a fixed hybrid architecture is summarized in Algorithm 2, which is referred to as the hybrid beamforming algorithm using manifold optimization under the WMMSE criterion (HBF-WMO algorithm).

Algorithm 2: The HBF-WMO algorithm

Input: $\sigma_k, \boldsymbol{H}_k, \forall k \in \{1,\cdots,N\}, \boldsymbol{U}_{\mathrm{p}}, \boldsymbol{U}_{\mathrm{c}}$
1: Initialize $\boldsymbol{W}_{\mathrm{RF},0}, \boldsymbol{F}_{\mathrm{RF},0}, \boldsymbol{T}_{k,0}, \boldsymbol{W}_{\mathrm{BB},k,0}, i = 0$
2: **repeat**:
3: Compute $\boldsymbol{F}_{\mathrm{RF},i}$ according to Algorithm 1
4: Compute $\beta_{k,i}, \boldsymbol{F}_{\mathrm{U},k,i}$ according to Eqs. (9) and (10)
5: Compute $\boldsymbol{W}_{\mathrm{RF},i}$ according to Algorithm 1
6: Compute $\boldsymbol{W}_{\mathrm{BB},k,i}$ according to Eq. (15)
7: Compute $\boldsymbol{T}_{k,i} = \boldsymbol{E}_{k,i}^{-1}$
8: $i \leftarrow i + 1$
9: **until** a stopping condition is satisfied
10: $\boldsymbol{F}_{\mathrm{BB},k} = \beta_k \boldsymbol{F}_{\mathrm{U},k}$
Output: $\boldsymbol{F}_{\mathrm{RF}}, \boldsymbol{F}_{\mathrm{BB},k}, \boldsymbol{W}_{\mathrm{RF}}, \boldsymbol{W}_{\mathrm{BB},k}$

4 Subarray Partition Optimization with the PC-Dynamic Architecture

In this section, we propose an algorithm to dynamically allocate antennas to each RF chain through a switch network in accordance with the channel state variation, and under the constraint that each antenna element is allocated only once. Since the size of the receive antenna array is relatively small compared with that of the transmitter, we only consider the architecture optimization of the transmitter here. Based on the derived objective function in Eq. (11), the WMMSE problem for the optimization of the subarray partition with the PC-dynamic architecture is given by

$$\begin{aligned} \underset{\boldsymbol{F}_{\mathrm{RF}}^{\mathrm{FC}}, \boldsymbol{U}_{\mathrm{p}}}{\text{minimize}} \quad & f\left(\boldsymbol{F}_{\mathrm{RF}}^{\mathrm{FC}} \odot \boldsymbol{U}_{\mathrm{p}}\right) \\ \text{subject to} \quad & \left|\left[\boldsymbol{F}_{\mathrm{RF}}^{\mathrm{FC}}\right]_{ij}\right| = 1; \\ & \left[\boldsymbol{U}_{\mathrm{p}}\right]_{ij} \in \{0,1\}; \\ & \left(\boldsymbol{U}_{\mathrm{p}i\cdot}\right)_1 = 1, \forall i,j, \end{aligned} \tag{18}$$

where $\boldsymbol{U}_{\mathrm{p}i\cdot}$ dedicates the i-th row of $\boldsymbol{U}_{\mathrm{p}}$. The original problem is difficult to solve directly, so we transform it into a more tractable one. First, let S_r denote the antenna subset connected to the r-th RF chain. We partition N_t antennas into N_{RF}^t subsets as:

$$\bigcup_{r=1}^{N_{\mathrm{RF}}^t} S_r = S_{\mathrm{ant}}, \left|S_r\right| > 0, S_i \cap S_j = \varnothing, \forall i,j \in \left\{1,\cdots,N_{\mathrm{RF}}^t\right\}, i \neq j, \tag{19}$$

where $\left|S_r\right|$ dedicates the number of elements in S_r and $S_{\mathrm{ant}} = \left\{1,\cdots,N_{\mathrm{t}}\right\}$. The optimization problem can be formulated as:

$$\begin{aligned}&\underset{\{S_r\}_{r=1}^{N_{\mathrm{RF}}^t}}{\text{minimize}} \quad \frac{1}{N}\sum_{k=1}^{N}\mathrm{tr}\left(\left(\boldsymbol{T}_k^{-1}+\xi_k\boldsymbol{G}_k^H\boldsymbol{F}_{\mathrm{RF}}\left(\boldsymbol{F}_{\mathrm{RF}}^H\boldsymbol{F}_{\mathrm{RF}}\right)^{-1}\boldsymbol{F}_{\mathrm{RF}}^H\boldsymbol{G}_k\right)^{-1}\right)\\&\text{subject to} \quad \left|\left[\boldsymbol{F}_{\mathrm{RF}}\right]_{ij}\right|=\begin{cases}1, i\in S_j\\0, \text{else}\end{cases};\\&\bigcup_{r=1}^{N_{\mathrm{RF}}^t}S_r=S_{\mathrm{ant}}, |S_r|>0, S_i\cap S_j=\varnothing, i\neq j.\end{aligned} \tag{20}$$

It can be shown that the analog precoder with the conventional fixed PC architecture satisfies $\boldsymbol{F}_{\mathrm{RF}}^H\boldsymbol{F}_{\mathrm{RF}}=\frac{N_t}{N_{\mathrm{RF}}^t}\boldsymbol{I}_{N_{\mathrm{RF}}^t}$, since the number of antennas connected to each RF chain is assumed to be the same. The equality does not hold in general with the PC-dynamic architecture since the number of antennas connected to each RF chain is not the same. However, as all the RF chains are treated equally and the number of RF chains is assumed to be much less than the number of antennas, it is very likely that the number of antennas connected to different RF chains tends to be close to each other, i. e., $\boldsymbol{F}_{\mathrm{RF}}^H\boldsymbol{F}_{\mathrm{RF}}\approx\frac{N_t}{N_{\mathrm{RF}}^t}\boldsymbol{I}_{N_{\mathrm{RF}}^t}$. According to the simulation results, the number of antennas in each subarray varies little with the optimized partition. Thus, by using this approximation, the objective function in Eq. (20) can be written as:

$$J(\boldsymbol{F}_{\mathrm{RF}})=\frac{1}{N}\sum_{k=1}^{N}\mathrm{tr}\left(\left(\frac{\xi_k}{N_{\mathrm{t}}}\boldsymbol{G}_k^H\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{RF}}^H\boldsymbol{G}_k\right)^{-1}\right). \tag{21}$$

Inspired by Ref. [12], where the authors derived a lower bound of the original MMSE problem and proposed the EVD algorithm with the FC architecture, we derive the lower bound of $J(\boldsymbol{F}_{\mathrm{RF}})$ as:

$$\begin{aligned}&\sum_{k=1}^{N}t\left(\boldsymbol{M}_k^{-1}\right)=\sum_{k=1}^{N}\sum_{s=1}^{N_s}\lambda_s^{-1}\left(\boldsymbol{M}_k\right)\geqslant\\&\sum_{k=1}^{N}N_s^2\left(\sum_{s=1}^{N_s}\lambda_s\left(\boldsymbol{M}_k\right)\right)^{-1}\geqslant N^2N_s^2\left(\sum_{k=1}^{N}\mathrm{tr}\left(\boldsymbol{M}_k\right)\right)^{-1},\end{aligned} \tag{22}$$

where $\boldsymbol{M}_k\triangleq\boldsymbol{T}_k^{-1}+\frac{\xi_k}{N_{\mathrm{t}}}\boldsymbol{G}_k^H\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{RF}}^H\boldsymbol{G}_k$, and $\lambda_s(\cdot)$ dedicates the eigenvalues of a matrix. The equality holds only if the values of $\mathrm{tr}\left(\boldsymbol{M}_k\right)$ at all subcarriers are the same. By taking the lower bound as the objective function and omitting the constants, the problem becomes:

$$\begin{aligned}&\underset{\{S_r\}_{r=1}^{N_{\mathrm{RF}}^t}}{\text{maximize}} \quad \mathrm{tr}\left(\frac{1}{N}\sum_{k=1}^{N}\xi_k\boldsymbol{G}_k^H\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{RF}}^H\boldsymbol{G}_k\right)\\&\text{subject to} \quad \left|\left[\boldsymbol{F}_{\mathrm{RF}}\right]_{ij}\right|=\begin{cases}1, i\in S_j\\0, \text{else}\end{cases};\\&\bigcup_{r=1}^{N_{\mathrm{RF}}^t}S_r=S_{\mathrm{ant}}, |S_r|>0, S_i\cap S_j=\varnothing, i\neq j.\end{aligned} \tag{23}$$

We can write $\boldsymbol{G}_k$ as a combination of N_{RF}^t block matrixes:

$$\boldsymbol{G}_k^H=\left[\boldsymbol{G}_{k,S_1}^H\ \boldsymbol{G}_{k,S_2}^H\ \cdots\ \boldsymbol{G}_{k,S_{N_{\mathrm{RF}}^t}}^H\right], \tag{24}$$

where $\boldsymbol{G}_{k,S_r}^H=\boldsymbol{G}_k^H(:,S_r)$, so the objective function in Eq. (23) can be written as:

$$\begin{aligned}&\mathrm{tr}\left(\frac{1}{N}\sum_{k=1}^{N}\xi_k\boldsymbol{G}_k^H\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{RF}}^H\boldsymbol{G}_k\right)=\\&\frac{1}{N}\sum_{k=1}^{N}\xi_k\left(\boldsymbol{G}_k^H\boldsymbol{F}_{\mathrm{RF}}\right)_F^2=\\&\frac{1}{N}\sum_{k=1}^{N}\xi_k\left(\left[\boldsymbol{G}_{k,S_1}^H\boldsymbol{f}_{\mathrm{RF},S_1}\cdots\boldsymbol{G}_{k,S_{N_{\mathrm{RF}}^t}}^H\boldsymbol{f}_{\mathrm{RF},S_{N_{\mathrm{RF}}^t}}\right]\right)_F^2=\\&\sum_{r=1}^{N_{\mathrm{RF}}^t}\left(\boldsymbol{f}_{\mathrm{RF},S_r}^H\boldsymbol{R}_{S_r}\boldsymbol{f}_{\mathrm{RF},S_r}\right),\end{aligned} \tag{25}$$

where $\boldsymbol{R}_{S_r}=\frac{1}{N}\sum_{k=1}^{N}\xi_k\boldsymbol{G}_{k,S_r}^H\boldsymbol{G}_{k,S_r}\boldsymbol{f}_{\mathrm{RF},S_r}\in\mathbb{C}^{N_t\times 1}$, $\left[\left|\boldsymbol{f}_{\mathrm{RF},S_r}\right|\right]_i=\begin{cases}1, i\in S_j\\0, i\notin S_j\end{cases}$. Without consideration of the constant modulus constraint, the maximum value of Eq. (25) is given by $\sum_{r=1}^{N_{\mathrm{RF}}^t}\lambda_1\left(\boldsymbol{R}_{S_r}\right)$, where $\lambda_1(\cdot)$ is the maximum eigenvalue of a matrix. Therefore, we can solve Eq. (23) by maximizing the sum of the maximum eigenvalues of $\boldsymbol{R}_{S_r}$ corresponding to each subarray. The optimization problem of the subarray partition can be formulated as:

$$\begin{aligned}&\underset{\{S_r\}_{r=1}^{N_{\mathrm{RF}}^t}}{\text{maximize}} \quad \sum_{r=1}^{N_{\mathrm{RF}}^t}\lambda_1\left(\boldsymbol{R}_{S_r}\right)\\&\text{subject to} \quad \bigcup_{r=1}^{N_{\mathrm{RF}}^t}S_r=S_{ant}, |S_r|>0, S_i\cap S_j=\varnothing, i\neq j.\end{aligned} \tag{26}$$

The optimal solution of Eq. (26) is a complex combinatorial optimization problem and calculating the eigenvalues leads to relatively high computational cost. Therefore, two operations are adopted here to further simplify the optimization problem. Firstly, according to Ref. [20], the maximum eigenvalue can be approximated with the l_1 norm of a matrix as follows:

$$\hat{\lambda}_1\left(\boldsymbol{R}_{S_r}\right)\triangleq\frac{1}{|S_r|}\sum_{i,j\in S_r}\left|\left[\boldsymbol{R}\right]_{i,j}\right|, \tag{27}$$

where $\boldsymbol{R}=\frac{1}{N}\sum_{k=1}^{N}\xi_k\boldsymbol{G}_{k,S_r}^H\boldsymbol{G}_{k,S_r}$ is the average channel covariance matrix of all subcarriers. Then, the problem in Eq. (26) becomes[1]:

$$\begin{aligned}&\underset{\{S_r\}_{r=1}^{N_{\mathrm{RF}}^t}}{\text{maximize}} \quad \frac{1}{|S_r|}\sum_{i,j\in S_r}\left|\left[\mathbf{R}\right]_{i,j}\right|\\&\text{subject to} \quad \bigcup_{r=1}^{N_{\mathrm{RF}}^t}S_r=S_{\mathrm{ant}}, |S_r|>0, S_i\cap S_j=\varnothing, i\neq j.\end{aligned} \tag{28}$$

Secondly, a greedy algorithm is proposed here to solve the

1 It is worth noting that compared with the original optimization objective function in Eq. (20), the omission of the constant modulus constraint and the use of several approximations above will bring in some performance loss, which is of interest for further investigation in future work.

problem. Since $\boldsymbol{R}$ is a Hermitian matrix, we only need to consider the upper triangular part when calculating the objective function in Eq. (28). The main idea of the greedy algorithm is that the objective function does not decrease after adding $\left|[\boldsymbol{R}]_{ij}\right|, i<j$ into or removing it out of the subset S_r. The algorithm consists of two steps. The first one is to define an initial full antenna set $S_0=\{1,\cdots,N_t\}$ and sort all the elements $\left|[\boldsymbol{R}]_{ij}\right|, i<j$ in descending order. To ensure that the constraint $|S_r|>0$ is satisfied, two antennas i and j corresponding to $[\boldsymbol{R}]_{ij}$ are partitioned into S_r in order. The second step is to consider different cases of partitioning antenna i,j into different subsets and choose the one that maximizes the objective function. In this process, if two antennas i and j belong to two different subsets, only four cases related to two subsets are considered. Otherwise, N_{RF}^t cases are considered if both antennas belong to S_0. For simplicity of description, we define a function $f_{\boldsymbol{R}_{S_r}}$ as follows:

$$f_{\boldsymbol{R}_{S_r}}(S_r,n_{sel},r)=\begin{cases}0, \ |S_r|=0 \text{ or } \{n_{\mathrm{sel}}=N_{\mathrm{RF}}^t, \text{and } r=0\}\\ \frac{1}{|S_r|}\left(\sum_{i,j\in S_r}\left|[\boldsymbol{R}_{\mathrm{up}}]_{i,j}\right|+\frac{1}{2}\sum_{i\in S_r}\left|[\boldsymbol{R}]_{i,i}\right|\right), \text{otherwise,}\end{cases} \tag{29}$$

where n_{sel} denotes the number of subsets that already have elements, and $\boldsymbol{R}_{\mathrm{up}}$ is the upper triangular part of $\boldsymbol{R}$. The partition optimization algorithm is summarized in Algorithm 3. With $\{S_r\}_{r=1}^{N_{\mathrm{RF}}^t}$, we can easily get $\boldsymbol{U}_{\mathrm{p}}$, and then use the HBF-WMO algorithm proposed in Section 3 to optimize $\boldsymbol{F}_{\mathrm{RF}}$ with the PC-dynamic architecture.

Algorithm 3: Dynamic subarrays partition optimization

Input: $\boldsymbol{R}, N_{\mathrm{RF}}^t, S_0, n_{\mathrm{sel}}=0$

1: Sort $\boldsymbol{R}_{\mathrm{up}}$ in descending order:
$\left|[\boldsymbol{R}]_{i,j}\right|\geqslant\left|[\boldsymbol{R}]_{i_kj_k}\right|\geqslant\cdots\geqslant\left|[\boldsymbol{R}]_{i_Kj_K}\right|, 1\leqslant i_k<j_k\leqslant N_t, K=N_t(N_t-1)/2$

2: For $k=1{:}K$ **repeat**:

3: If $i_k,j_k\in S_0$:

4: If $n_{\mathrm{sel}}<N_{\mathrm{RF}}^t$:
$S_{n_{\mathrm{sel}}}\leftarrow\{i_k,j_k\}, S_0\backslash\{i_k,j_k\}, n_{\mathrm{sel}}\leftarrow n_{\mathrm{sel}}+1$

5: Else: $r_{\max}=\underset{r}{\mathrm{argmax}}\ (f_{\boldsymbol{R}_{S_r}}(S_r\cup\{i_k,j_k\},n_{\mathrm{sel}},r))$
$S_{r_{\max}}\leftarrow\{i_k,j_k\}, S_0\backslash\{i_k,j_k\}$

6: Else if $i_k\in S_m, j_k\in S_l, \forall m,l\in\{0,1,\cdots,n_{sel}\}, m\neq l$:
$u_{\mathrm{crt}}=f_{\boldsymbol{R}_{S_r}}(S_m,n_{sel},m)+f_{\boldsymbol{R}_{S_r}}(S_l,n_{\mathrm{sel}},l)$
$u_{\mathrm{newj}}=f_{\boldsymbol{R}_{S_r}}(S_m\cup\{j_k\},n_{\mathrm{sel}},m)+f_{\boldsymbol{R}_{S_r}}(S_l\backslash\{j_k\},n_{\mathrm{sel}},l)$
$u_{\mathrm{newi}}=f_{\boldsymbol{R}_{S_r}}(S_m\backslash\{i_k\},n_{sel},m)+f_{\boldsymbol{R}_{S_r}}(S_l\cup\{i_k\},n_{sel},l)$
$u_{\mathrm{newij}}=f_{\boldsymbol{R}_{S_r}}((S_m\backslash\{i_k\})\cup\{j_k\},n_{\mathrm{sel}},m)+f_{\boldsymbol{R}_{S_r}}((S_l\backslash\{j_k\})\cup\{i_k\},n_{\mathrm{sel}},l)$
$u=[u_{\mathrm{crt}},u_{\mathrm{newj}},u_{\mathrm{newi}},u_{\mathrm{newij}}]$
$\max(u)=u_{\mathrm{newj}}$ and $m\neq 0$, $S_m\leftarrow\{j_k\}, S_l\backslash\{j_k\}$
$\max(u)=u_{\mathrm{newi}}$ and $l\neq 0$, $S_m\backslash\{i_k\}, S_l\leftarrow\{i_k\}$
$\max(u)=u_{\mathrm{newij}}$ and $m,l\neq 0$,
$(S_m\backslash\{i_k\})\leftarrow\{j_k\}, (S_l\backslash\{j_k\})\leftarrow\{i_k\}$

Output: $\{S_r\}_{r=1}^{N_{\mathrm{RF}}^t}$

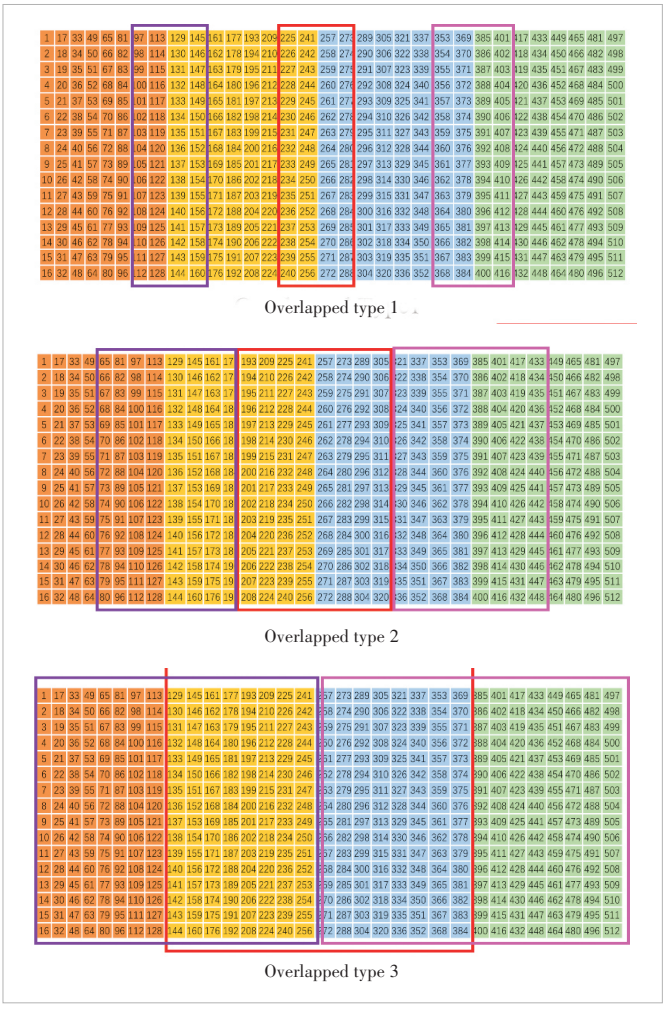
5 Simulation Results

In this section, we first provide some simulation results to show the spectral efficiency performance of the proposed HBF-WMO algorithm in Section 3 with fixed hybrid architectures, in comparison with the optimal fully-digital one. Then, we compare the spectral efficiency performance of fixed and dynamic partitions of antenna subarrays with the PC architecture.

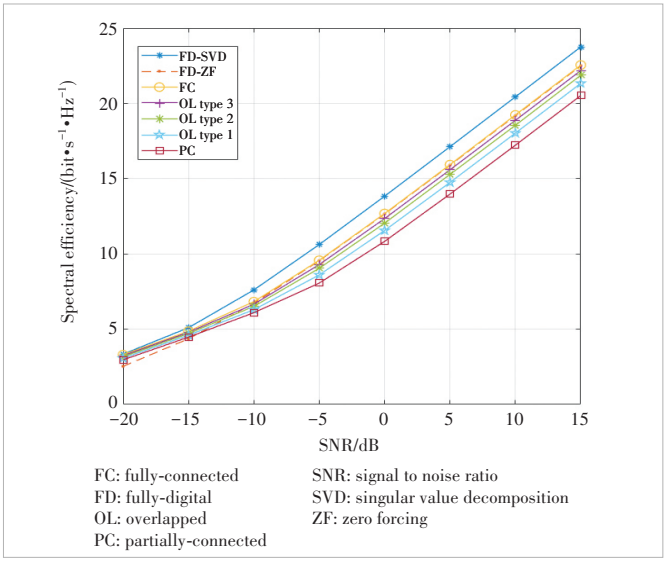
Considering an mmWave MIMO-OFDM system as that in Fig. 1, we assume that the transmitter takes a half-wavelength spaced UPA with $N_t=512$ antennas for the transmission of $N_s=2$ data streams. Five fixed hybrid architectures at the transmitter are evaluated, including the FC and PC architectures, and three types of the OL architecture. Considering the issue of practical implementation of the HBF architecture, we propose three specified OL architectures when four RF chains are employed with a 16×32 UPA at the transmitter, which are shown in Fig. 3. The numbers indicate the antenna indexes, and the units in the same color mean that the corresponding antenna elements are connected to the same RF chain. The antennas within framed squares are overlapped and connected to multiple RF chains. The receiver takes a UPA with $N_r=8$ antennas and $N_{\mathrm{RF}}^r=2$ RF chains with a fixed PC architecture. The number of subcarriers is set to $N=64$, the bandwidth is 100 MHz and the center frequency is 28 GHz. The signal-to-noise ratio (SNR) is defined as $\frac{1}{\sigma^2}$. We take CDL-A as the channel model to evaluate the system performance in a more practical NLOS scenario. In the simulation, the stopping condition is set as the relative difference between the objective function values of two consecutive iterations becomes smaller than $\delta=10^{-3}$.

5.1 Performance with Different Fixed Hybrid Architectures

Fig. 4 shows the performance of spectral efficiency as a function of SNR for the proposed HBF-WMO algorithm with different fixed architectures in CDL-A when four RF chains are equipped at the transmitter. The performance curves of the FC, PC, and OL architectures are labeled as "FC", "PC", and "OL", respectively. For comparison, two FDBF algorithms, namely the FDBF with singular value decomposition (SVD) on



▲ **Figure 3. Diagram of three types of the overlapped subarray-connected architecture at the transmitter with N_t=512, N_{RF}^t = 4**

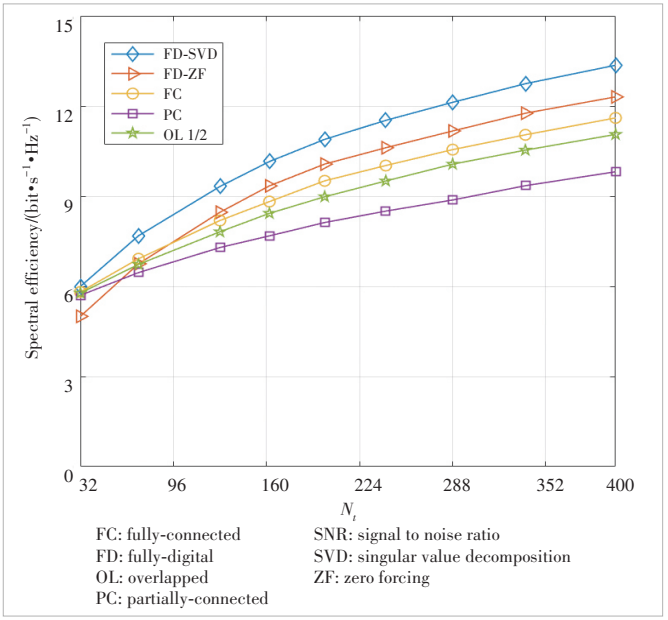


▲ **Figure 4. Spectral efficiency vs SNR for the hybrid beamforming (HBF-WMO) algorithm with different fixed hybrid architectures for a massive multiple-input multiple-output-orthogonal frequency division multiplexing (MIMO-OFDM) system with N=64, N_t=512, N_r=8, N_{RF}^t = 4, N_{RF}^r = 2, N_s=2**

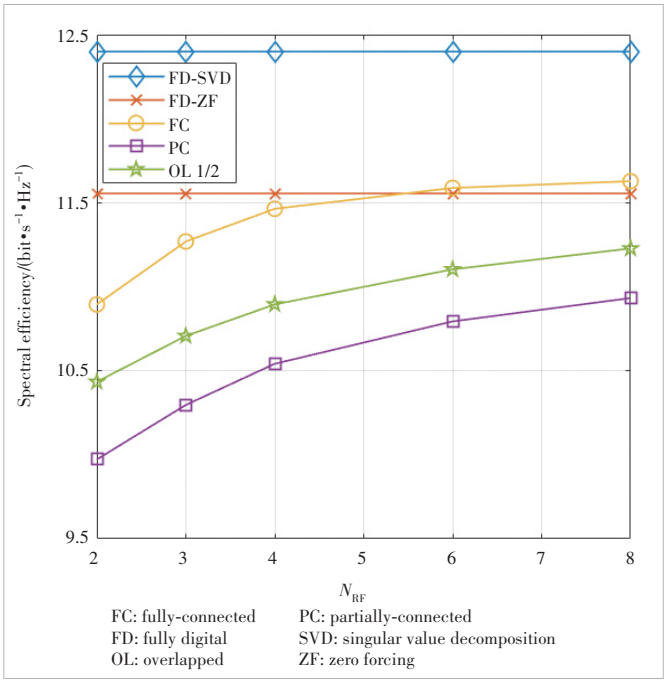
each subcarrier and the FDBF with zero-forcing (ZF), are adopted, and their performance curves are labeled as "FD-SVD" and "FD-ZF", respectively.

It can be seen from this figure that the proposed HBF-WMO algorithm with the FC architecture performs well with far fewer RF chains than antenna elements, and its performance gain over the PC architecture is about 4 dB in CDL-A. The reason is that there are much fewer entries that can be optimized in the HBF matrix of the PC architecture than the FC architecture, since the PC architecture employs far fewer phase shifters. Results also show that the OL architecture achieves the compromise between the system performance and hardware costs, when compared with the FC and PC architectures. In particular, it achieves higher spectral efficiency than the PC architecture at the cost of higher power consumption and implementation complexity introduced by more phase shifters. For example, with 192 more phased shifters employed, the first type of OL architecture has a performance gain of about 1 dB over the PC architecture. According to Ref. [24], based on the state-of-the-art technique, the power consumed by each phase shifter is about 10 mW. With the size of overlapped antennas among different subarrays growing, the performance improvement of the OL architecture over the PC architecture gets bigger. For example, with the third type of OL architecture, the performance gain is about 3 dB over the PC one.

To verify the generality of the proposed HBF-WMO algorithm, we provide in Figs. 5 and 6 the spectral efficiency performance with different numbers of transmit antennas and RF chains under fixed hybrid architectures. The label "OL 1/2"



▲ **Figure 5. Spectral efficiency vs number of transmit antennas for the HBF-WMO algorithm with different fixed hybrid architectures for a massive multiple-input multiple-output-orthogonal frequency division multiplexing (MIMO-OFDM) system with SNR=0 dB, N=64, N_r=8, N_{RF}^t = N_{RF}^r = 2, N_s=2**



▲ **Figure 6. Spectral efficiency vs number of transmit RF chains for the HBF-WMO algorithm with different fixed hybrid architectures for a massive multiple-input multiple-output-orthogonal frequency division multiplexing (MIMO-OFDM) system with SNR=0 dB, N=64, N_t=512, N_r=8, N_{RF}^t = 2, N_s=2**

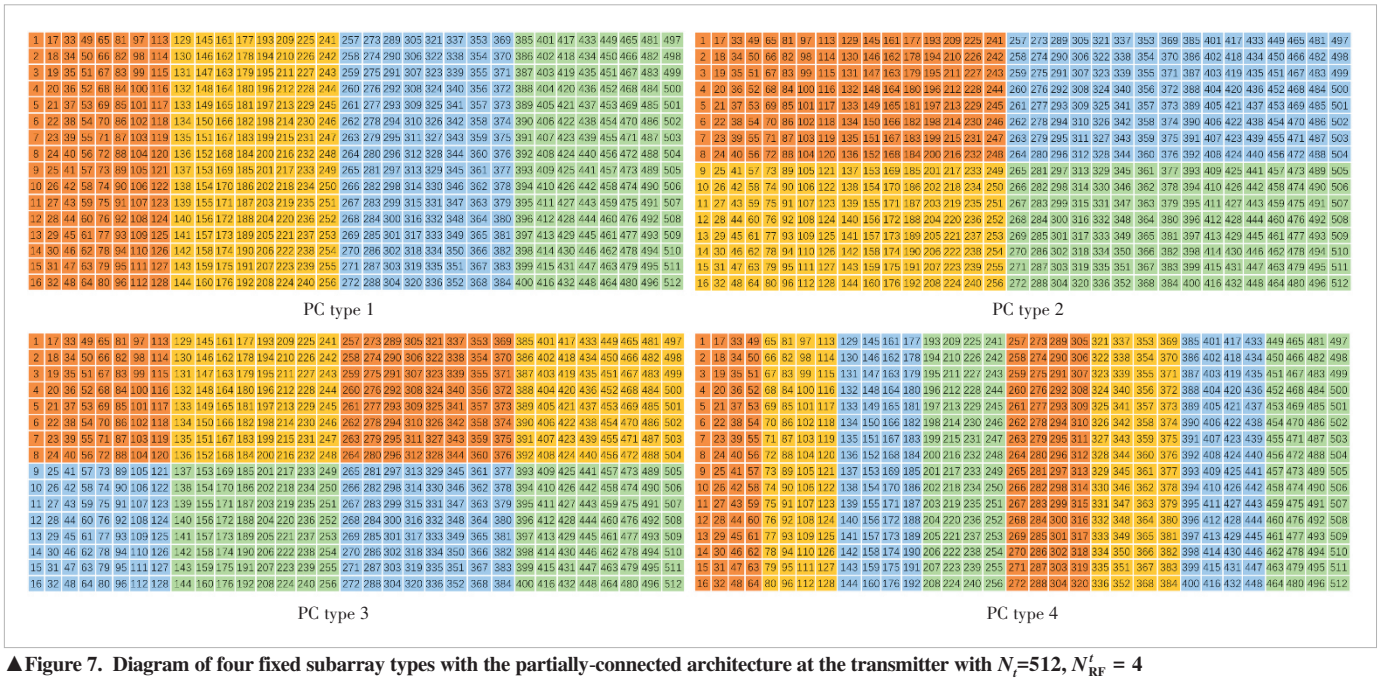
refers to the case where the number of overlapped antennas equals half the number of transmit antennas in the OL architecture. As shown in Fig. 5, the performance of the HBF-WMO algorithm improves with more transmit antennas. Fig. 6 also shows that the gap between the performance of the HBF-WMO algorithm and the optimal FDBF algorithm narrows with more RF chains equipped at the transmitter.

5.2 Performance with Different Partitions of Antenna Subarrays

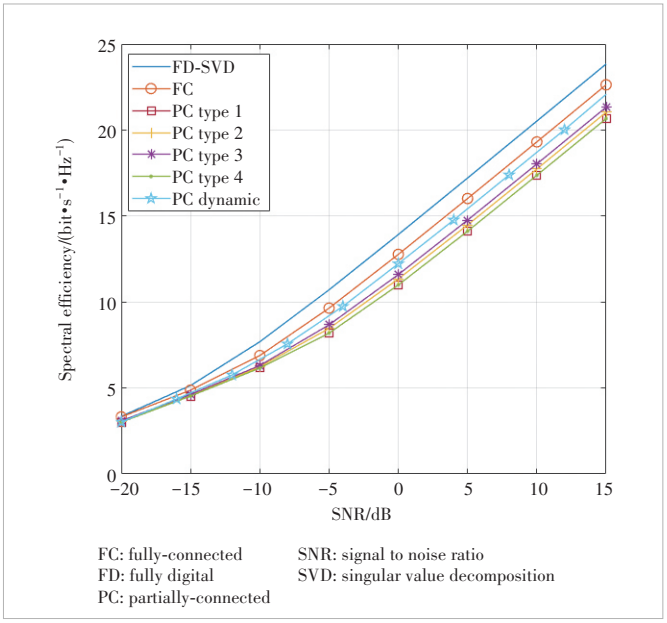
Next, we compare the performance of fixed and dynamic partitions of antenna subarrays with the PC architecture. Four types of partitions under the fixed PC architecture are considered at the transmitter, as shown in Fig. 7. For the HBF system with the PC-dynamic architecture, the antenna subarrays are first dynamically partitioned based on the algorithm proposed in Section 4 and then the HBF matrices are optimized based on the HBF-WMO algorithm in Section 3. Fig. 8 shows the average spectral efficiency performance as a function of SNR with fixed and dynamic antenna subarrays. It can be seen that of the four fixed partition types, the third one achieves the best performance. This is mainly due to the balanced horizontal and vertical angle resolution of the subarray in the third type and the original distribution of angles in CDL-A. It is also shown that the dynamic partition outperforms the third fixed partition type with about 1 dB at the cost of more power consumption and the associated complex circuit brought by N_t = 512 more switches employed in the switch network. According to Ref. [24], the power consumed by each switch is about 5 mW.

6 Conclusions

With relatively small hardware costs and performance loss compared with FDBF, HBF for mmWave communication systems has attracted much attention. Meanwhile, the design of hybrid architecture has also become a research hotspot considering the practical implementation complexity. We have inves-



▲ **Figure 7. Diagram of four fixed subarray types with the partially-connected architecture at the transmitter with N_t=512, N_{RF}^t = 4**



▲Figure 8. Spectral efficiency vs SNR for the HBF-WMO algorithm with different partitions of subarrays for a massive multiple-input multiple-output-orthogonal frequency division multiplexing (MIMO-OFDM) system with N=64, N_t=512, N_r=8, N_{RF}^t = 4, N_{RF}^r = 2, N_s=2

tigated HBF with different hybrid architectures in this paper. After transforming the original SEM problem to a more tractable equivalent WMMSE problem, we propose the HBF-WMO algorithm with different fixed architectures. Simulation results have shown that the OL architecture achieves a compromise between the hardware costs and system performance compared with the conventional fixed architectures. We have also proposed a low-complexity subarray partition optimization algorithm based on the maximum eigenvalue approximation with the PC-dynamic architecture and combined it with the HBF-WMO algorithm. Simulation results show that the PC-Dynamic architecture achieves some performance gain over the fixed PC architecture.

In this paper, certain problems in the practical implementation of the proposed architecture and HBF-WMO algorithm under massive MIMO-OFDM systems have not been investigated. On one hand, the dynamic architecture and the OL architecture would lead to more power consumption and insertion power loss with more required phase shifters, splitters, combiners and switches than the conventional PC architecture, and thus the energy efficiency could be considered for HBF optimization. On the other hand, some studies have made efforts to alleviate the effect of beam squint while developing hybrid precoding schemes for massive MIMO-OFDM systems, such as carrying out a phase compensation operation at each subcarrier[25] or projecting all frequencies to the central frequency[26]. In future work, we will make efforts to study the energy efficiency performance of different HBF architectures and extend our investigation to the scenario when the system is subject to the beam squint effect.

Reference

[1] PI Z Y, KHAN F. An introduction to millimeter-wave mobile broadband systems [J]. IEEE communications magazine, 2011, 49(6): 101 – 107. DOI: 10.1109/MCOM.2011.5783993
[2] ROH W, SEOL J Y, PARK J, et al. Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results [J]. IEEE communications magazine, 2014, 52(2): 106 – 113. DOI: 10.1109/MCOM.2014.6736750
[3] RANGAN S, RAPPAPORT T S, ERKIP E. Millimeter-wave cellular wireless networks: potentials and challenges [J]. Proceedings of the IEEE, 2014, 102(3): 366 – 385. DOI: 10.1109/JPROC.2014.2299397
[4] PAULRAJ A J, GORE D A, NABAR R U, et al. An overview of MIMO communications: a key to gigabit wireless [J]. Proceedings of the IEEE, 2004, 92(2): 198 – 218. DOI: 10.1109/JPROC.2003.821915
[5] AKDENIZ M R, LIU Y P, SAMIMI M K, et al. Millimeter wave channel modeling and cellular capacity evaluation [J]. IEEE journal on selected areas in communications, 2014, 32(6): 1164 – 1179. DOI: 10.1109/JSAC.2014.2328154
[6] ZHANG J, YU X H, LETAIEF K B. Hybrid beamforming for 5G and beyond millimeter-wave systems: a holistic view [J]. IEEE open journal of the communications society, 2019, 1: 77 – 91. DOI: 10.1109/OJCOMS.2019.2959595
[7] LARSSON E G, EDFORS O, TUFVESSON F, et al. Massive MIMO for next generation wireless systems [J]. IEEE communications magazine, 2014, 52(2): 186 – 195. DOI: 10.1109/MCOM.2014.6736761
[8] MOLISCH A F, RATNAM V V, HAN S Q, et al. Hybrid beamforming for massive MIMO: a survey [J]. IEEE communications magazine, 2017, 55(9): 134 – 141. DOI: 10.1109/MCOM.2017.1600400
[9] AYACH O E, RAJAGOPAL S, ABU-SURRA S, et al. Spatially sparse precoding in millimeter wave MIMO systems [J]. IEEE transactions on wireless communications, 2014, 13(3): 1499 – 1513. DOI: 10.1109/TWC.2014.011714.130846
[10] SOHRABI F, YU W. Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays [J]. IEEE journal on selected areas in communications, 2017, 35(7): 1432 – 1443. DOI: 10.1109/JSAC.2017.2698958
[11] YU X H, SHEN J C, ZHANG J, et al. Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems [J]. IEEE journal of selected topics in signal processing, 2016, 10(3): 485 – 500. DOI: 10.1109/JSTSP.2016.2523903
[12] LIN T, CONG J Q, ZHU Y, et al. Hybrid beamforming for millimeter wave systems using the MMSE criterion [J]. IEEE transactions on communications, 2019, 67(5): 3693 – 3708. DOI: 10.1109/TCOMM.2019.2893632
[13] ZHAO X Y, LIN T, ZHU Y, et al. Partially-connected hybrid beamforming for spectral efficiency maximization via a weighted MMSE equivalence [J]. IEEE transactions on wireless communications, 2021, 20(12): 8218 – 8232. DOI: 10.1109/TWC.2021.3091524
[14] ZHANG D D, WANG Y F, LI X H, et al. Hybridly connected structure for hybrid beamforming in mmWave massive MIMO systems [J]. IEEE transactions on communications, 2018, 66(2): 662 – 674. DOI: 10.1109/TCOMM.2017.2756882
[15] GAUTAM P R, ZHANG L. Hybrid precoding for partial-full mixed connection mmWave MIMO [C]//The IEEE Statistical Signal Processing Workshop (SSP). IEEE, 2021: 271 – 275. DOI: 10.1109/SSP49050.2021.9513847
[16] SONG N, YANG T, SUN H. Overlapped subarray based hybrid beamforming for millimeter wave multiuser massive MIMO [J]. IEEE signal processing letters, 2017, 24(5): 550 – 554. DOI: 10.1109/LSP.2017.2681689
[17] CHEN Y, CHEN D, JIANG T, et al. Millimeter-wave massive MIMO systems relying on generalized sub-array-connected hybrid precoding [J]. IEEE transactions on vehicular technology, 2019, 68(9): 8940 – 8950. DOI: 10.1109/TVT.2019.2930639
[18] MÉNDEZ-RIAL R, RUSU C, GONZÁLEZ-PRELCIC N, et al. Hybrid MIMO architectures for millimeter wave communications: phase shifters or switches? [J]. IEEE access, 2016, 4: 247 – 267. DOI: 10.1109/ACCESS.2015.2514261
[19] ALKHATEEB A, NAM Y H, ZHANG J Z, et al. Massive MIMO combining with switches [J]. IEEE wireless communications letters, 2016, 5(3): 232 – 235. DOI: 10.1109/LWC.2016.2522963
[20] PARK S, ALKHATEEB A, HEATH R W. Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems [J]. IEEE transactions on wire-

less communications, 2017, 16(5): 2907 – 2920. DOI: 10.1109/TWC.2017.2671869
[21] JIN J N, XIAO C S, CHEN W, et al. Channel-statistics-based hybrid precoding for millimeter-wave MIMO systems with dynamic subarrays [J]. IEEE transactions on communications, 2019, 67(6): 3991 – 4003. DOI: 10.1109/TCOMM.2019.2899628
[22] YAN L F, HAN C, YUAN J H. A dynamic array-of-subarrays architecture and hybrid precoding algorithms for terahertz wireless communications [J]. IEEE journal on selected areas in communications, 2020, 38(9): 2041 – 2056. DOI: 10.1109/JSAC.2020.3000876
[23] 3GPP. Study on channel model for frequencies from 0.5 to 100 GHz: TR 38.901 V16.2.0 [S]. 2020
[24] LI H Y, LI M, LIU Q. Hybrid beamforming with dynamic subarrays and low-resolution PSs for mmWave MU-MISO systems [J]. IEEE transactions on communications, 2020, 68(1): 602 – 614. DOI: 10.1109/TCOMM.2019.2950905
[25] CHEN Y, CHEN D, JIANG T, et al. Channel-covariance and angle-of-departure aided hybrid precoding for wideband multiuser millimeter wave MIMO systems [J]. IEEE transactions on communications, 2019, 67(12): 8315 – 8328. DOI: 10.1109/TCOMM.2019.2942307
[26] CHEN Y, XIONG Y F, CHEN D, et al. Hybrid precoding for wideband millimeter wave MIMO systems in the face of beam squint [J]. IEEE transactions on wireless communications, 2021, 20(3): 1847 – 1860. DOI: 10.1109/TWC.2020.3036945

Biographies

TANG Yuanqi received her BS degree in communication science and engineering from Fudan University, China in 2022, where she is currently pursuing her MS degree. Her research interests include hybrid beamforming for massive MIMO systems, millimeter wave signal processing and reconfigurable intelligent surface.

ZHANG Huimin received her BS degree in communication science and engineering from Fudan University, China in 2021, where she is currently pursuing her MS degree. Her current research interests include hybrid beamforming for massive MIMO systems and energy efficiency in intelligent reflecting surface-aided systems.

ZHENG Zheng received his BS and PhD degrees in information science and electronic engineering from Zhejiang University, China in 2013 and 2019, respectively. He is currently a senior algorithm engineer working on physical layer algorithms in ZTE Corporation. His research interests include wireless communications, array signal processing and artificial intelligence algorithms.

LI Ping received her MS degree in communication and information engineering from Xi'an Jiaotong University, China in 2004. She is currently a senior algorithm system engineer at ZTE Corporation, responsible for national key projects. Her research interests include digital signal processing, multiple antenna, system performance optimization, reconfigurable intelligent surface, networking technology, network planning, integrated sensing and communications (ISAC), and key technologies in 5G-A. She has applied for nearly 100 patents and published over 10 papers in various journals and conferences.

ZHU Yu (zhuyu@fudan.edu.cn) received his BE degree (Hons.) in electronics engineering and ME degree (Hons.) in communication and information engineering from the University of Science and Technology of China in 1999 and 2002, respectively, and got his PhD degree in electrical and electronic engineering from the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, China in 2007. Since 2008, he has been with Fudan University, China, where he is currently a professor with the School of Information Science and Technology. His current research interests include broadband wireless communication systems and networks and signal processing for communications. He has served as an editor for the *IEEE Wireless Communications Letters*, and as an editor of the *Journal of Communications and Information Networks*.

Simulation and Modeling of Common Mode EMI Noise in Planar Transformers

LI Wei[1], JI Jingkang[1], LIU Yuanlong[1], SUN Jiawei[2], LIN Subin[2]
(1. ZTE Corporation, Shenzhen 518057, China;
2. Fuzhou University, Fuzhou 350108, China)

DOI: 10.12142/ZTECOM.202303014
https://kns.cnki.net/kcms/detail/34.1294.TN.20230810.1409.004.html, published online August 10, 2023
Manuscript received: 2022-11-21

Abstract: The transformer is the key circuit component of the common-mode noise current when an isolated converter is working. The high-frequency characteristics of the transformer have an important influence on the common-mode noise of the converter. Traditionally, the measurement method is used for transformer modeling, and a single lumped device is used to establish the transformer model, which cannot be predicted in the transformer design stage. Based on the transformer common-mode noise transmission mechanism, this paper derives the transformer common-mode equivalent capacitance under ideal conditions. According to the principle of experimental measurement of the network analyzer, the electromagnetic field finite element simulation software three-dimensional (3D) modeling and simulation method is used to obtain the two-port parameters of the transformer, extract the high-frequency parameters of the transformer, and establish its electromagnetic compatibility equivalent circuit model. Finally, an experimental prototype is used to verify the correctness of the model by comparing the experimental measurement results with the simulation prediction results.

Keywords: planar transformer; electromagnetic compatibility; common-mode noise; simulation; modeling

Citation (Format 1): LI W, JI J K, LIU Y L, et al. Simulation and modeling of common mode EMI noise in planar transformers [J]. ZTE Communications, 2023, 21(3): 105 - 116. DOI: 10.12142/ZTECOM.202303014
Citation (Format 2): W. Li, J. K. Ji, Y. L. Liu, et al., "Simulation and modeling of common mode EMI noise in planar transformers," *ZTE Communications*, vol. 21, no. 3, pp. 105 - 116, Jun. 2023. doi: 10.12142/ZTECOM.202303014.

1 Introduction

Isolated power converters are widely used in applications that require isolation between the input side and the output side. However, due to the high voltage and current rate of change when switching devices in the power converter are turned on or off, serious electromagnetic interference (EMI) noise is generated[1]. For converters including isolation transformers, the critical path of common-mode noise includes the distributed capacitance between the primary and secondary windings of the transformer in addition to the direct-to-ground coupling capacitance of the voltage trip point[2-4]. The transmission mechanism of common mode noise and the establishment of an electromagnetic compatibility model are of great significance for the analysis and suppression of common mode noise of isolated power converters.

Magnetic components in power converters often occupy a large volume. As power converters develop toward high power density, traditional wound transformers are gradually replaced by planar transformers due to their disadvantages of large size and weight. Planar transformers generally use multi-layer printed circuit board (PCB) traces as windings. The primary and secondary windings are tightly coupled, with low leakage inductance. At the same time, the core height is greatly reduced, so that planar transformers have a smaller size, which is widely used in isolated converters with high power density[5-7]. However, the copper thickness of the printed circuit board is limited by the process. In order to improve the current capacity, it is necessary to increase the width of the planar transformer windings and adopt the method of multi-layer parallel wiring, which will greatly increase the facing area between the primary and secondary windings of the planar transformer, increase the distributed capacitance between the windings, and deteriorate the high-frequency characteristics of the planar transformer[8]. For the integrally installed planar transformer, its printed circuit board windings and the main circuit share the same PCB, and its discrete devices cannot be obtained to evaluate its electromagnetic compatibility characteristics. Therefore, the relevant parameters of the electromagnetic compatibility equivalent circuit model of the planar transformer can be extracted through electromagnetic field simulation software. Compared with the traditional wound transformer, the structural parameters of the planar transformer are stable and consistent, and it is easy to model in the electromagnetic field finite element simulation software.

In Ref. [9], considering the winding potential gradient distribution of planar transformers, a two-capacitance model based on common-mode current equivalence is derived based on the winding structure theory. However, the results obtained from the theoretical calculation consider the common-mode equivalent capacitance to be constant, ignoring the case that the common-mode equivalent capacitance is no longer a pure capacitance at high frequencies. Refs. [10 – 11] propose a method for evaluating the common-mode EMI characteristics of transformers using vector network analyzer experimental measurements. In this method, the transformer is treated as a common-mode noise filter, and the insertion loss parameter S_{21} [12 – 13] used to evaluate the EMI filter performance is referenced into the transformer. The experiment proves that it is feasible and effective to use the insertion loss measurement principle to evaluate the common mode noise rejection capability of the transformer. However, Refs. [10 – 11] only use S_{21} as the evaluation basis without further analyzing the specific meaning of the complete S-parameters obtained by the vector network analyzer to measure the transformer.

Section 2 analyzes the common mode noise transmission mechanism of the planar transformer, and deduces the expression of the induced charge Q between the primary and secondary windings according to the potential distribution of the windings when the planar transformer is working. From the expression, two capacitors are sufficient to represent the common-mode noise of the transformer. Section 3 establishes the electromagnetic compatibility equivalent circuit model of planar transformers according to the common mode noise transmission mechanism and the energy transmission characteristics of the transformer and introduces a method to establish the electromagnetic compatibility model of the full-bridge transformer through the experimental measurement of the network analyzer. Section 4 is based on the principle of experimental measurement and modeling of the network analyzer, and the two-port parameters of the planar transformer are extracted through the electromagnetic field finite element simulation software without physical objects, and are equivalent to a circuit model. Finally, the accuracy of the simulation modeling is verified by a full-bridge circuit prototype experiment in Section 5. Section 6 concludes this paper.

2 Transmission Mechanism of Common Mode Noise in Planar Transformers

2.1 Structural Capacitance Calculation

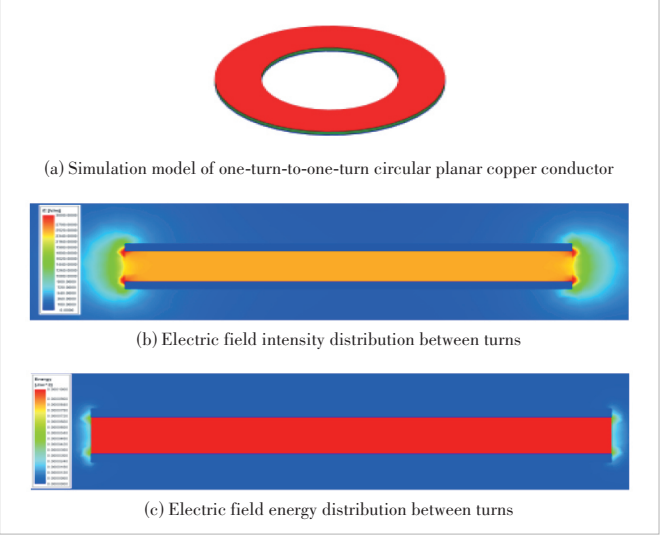
The common-mode noise transmission of the planar transformer is mainly through the distributed capacitance between the primary winding and the secondary winding. It is worth noting that the distributed capacitance of the transformer is not equal to the equivalent capacitance of the common mode noise, because the voltage on the planar transformer winding is not a constant value. However, the derivation of the common-mode noise equivalent capacitance is based on the distributed capacitance of the planar transformer. The simulation results of a one-turn-to-one-turn circular plane copper wire are shown in Fig. 1. Fig. 1(a) shows a simulation model of a one-turn-to-one-turn circular planar copper conductor. In Fig. 1(c), the planar transformer winding uses PCB wiring, and the distance d between the turns of the primary and secondary windings is the distance between the layers of the PCB. For a planar transformer with a single-layer single-turn structure, considering the current flow factor, the wiring width w of the winding is generally much larger than d, and the distributed capacitance between the primary winding and the secondary winding can be approximately regarded as a parallel plate capacitor, ignoring the edge effect. Then the structural capacitance C_0 between turns in Fig. 1(c) can be expressed as:

$$C_0 = \frac{\varepsilon_r \varepsilon_0 S}{d}, \tag{1}$$

where ε_r is the dielectric constant of the filling material between the PCB layers, ε_0 is the dielectric constant in vacuum, S is the facing area between turns, and d is the distance between turns.

Since the method of approximating parallel plate capacitors is used in the calculation of the capacitance of the planar transformer structure, in order to test the rationality and accuracy of the approximation, the electromagnetic field finite element simulation software is used to verify the winding structural capacitance.

Taking a turn-to-turn ring-shaped flat copper wire as an example, we list the model parameters as follows: the copper thickness is 1 mm, the inner diameter is 12.3 mm, the outer



(a) Simulation model of one-turn-to-one-turn circular planar copper conductor

(b) Electric field intensity distribution between turns

(c) Electric field energy distribution between turns

▲ **Figure 1. Simulation results of one-turn-to-one-turn toroidal planar copper conductors**

diameter is 20.6 mm, the distance between turns is 0.4 mm, and the relative permittivity of the filler material is 4.4.

The simulation result of the finite element simulation software is 21.769 pF, and the calculation result of Eq. (1) is 20.89 pF, with an error of 4.04%. The main source of error is the edge effect of the parallel plate capacitor. In Fig. 1(b), the electric field strength between conductors is basically uniform, and the electric field distribution at the edge of the conductor is obviously uneven. However, Fig. 1(c) shows that most of the electric field energy is stored between conductors, so it is reasonable to use an approach that approximates a parallel plate capacitor.

2.2 Calculation of Equivalent Capacitance for Common Mode Noise of Planar Transformers

For a planar transformer, due to the close magnetic coupling of each layer of windings, the magnetic flux or induced electromotive force linked by each turn of the winding is basically the same, and the alternating current (AC) resistance of each turn of the winding is much smaller than the excitation inductance. If the leakage inductance is ignored, it can be approximated that the potential of the winding is linearly distributed along the number of turns and the length of the winding.

Assuming that the potential on the transformer winding in Fig. 1(a) is linearly distributed along the turn length, and the distributed capacitance is evenly distributed along the winding, the integral of the primary winding and the secondary winding along the turn length l can be expressed as:

$$\int_0^l v_{\mathrm{p}}(x)\mathrm{d}x = \frac{v_{\mathrm{ph}} + v_{\mathrm{pl}}}{2} \cdot l, \tag{2}$$

$$\int_0^l v_{\mathrm{s}}(x)\mathrm{d}x = \frac{v_{\mathrm{sh}} + v_{\mathrm{sl}}}{2} \cdot l. \tag{3}$$

The charge Q stored between the turns of the primary and secondary windings in Fig. 1(a) can be expressed as:

$$Q = \int_0^l \frac{C_0}{l} \cdot \left[v_{\mathrm{p}}(x) - v_{\mathrm{s}}(x)\right]\mathrm{d}x = \frac{C_0}{l}\left[\int_0^l v_{\mathrm{p}}(x)\mathrm{d}x - \int_0^l v_{\mathrm{s}}(x)\mathrm{d}x\right]. \tag{4}$$

Substituting Eqs. (2) and (3) into Eq. (4), we can get:

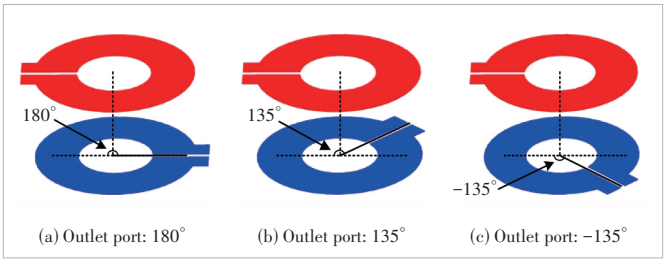
$$Q = C_0\left(\frac{v_{\mathrm{ph}} + v_{\mathrm{pl}}}{2} - \frac{v_{\mathrm{sh}} + v_{\mathrm{sl}}}{2}\right). \tag{5}$$

Considering that when the windings between different turns of the planar transformer are connected, the outlet ports of the windings on the same side need to be staggered by a certain angle. At this time, the voltage difference between the relative position of the primary winding and the secondary winding will change with the staggered angle. However, it can be obtained from Eq. (5) that the induced charge Q between turns is determined by the structural capacitance C_0 between the primary and secondary windings and the midpoint potential of the primary and secondary windings. As long as the assumption of uniform distribution of potential on the windings is established, the charge Q is not affected by the voltage difference distribution of the primary and secondary windings. As long as the assumption of uniform distribution of potential on the windings holds, the charge Q is not affected by the voltage difference distribution of the primary and secondary windings. As shown in Fig. 2, the angles between the primary winding outlet port and the secondary winding outlet port are 180°, 135° and −135°. The potential difference between the primary winding and the secondary winding changes. However, ignoring the influence of port voids, after the integration operation of Eq. (4), the final induced charge Q between turns and the turns of three different outlet port angles is the same.

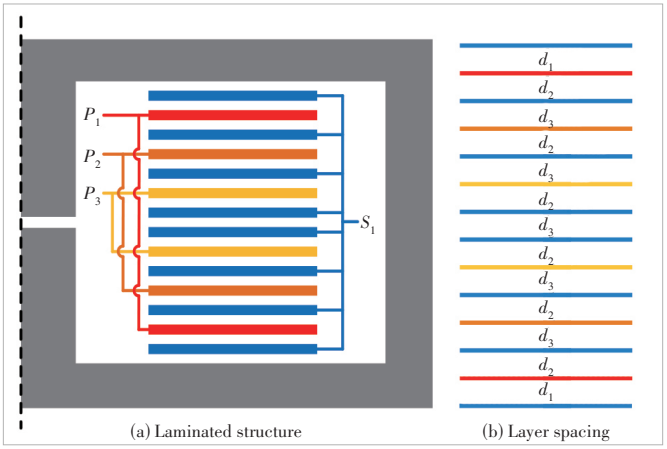
The primary and secondary windings of a planar transformer are multi-turn, so it is only necessary to follow the calculation method of Eq. (5). The inductive charge between each turn of the primary and secondary windings is added up to the charge Q_{all} stored between the primary and secondary windings of the entire planar transformer. Q_{all} characterizes the transfer capability of transformer common mode noise.

Taking the single-layer single-turn planar transformer in Fig. 3 as an example to calculate Q_{all}, we find that the turn ratio of the primary and secondary windings is 3:1, the primary winding has 3 turns, which are P_1, P_2, and P_3, respectively, and the secondary winding has 1 turn, which is S_1. The transformer adopts a 2-layer parallel winding on the primary side, an 8-layer parallel winding on the secondary side, and a staggered winding on the primary and secondary sides. The specific stacked structure is shown in Fig. 3. At the same time, the laminated structure parameters of the planar transformer are given in Table 1. Under the symmetrical structure, the facing areas between P_1, P_2, P_3 and S_1 are basically the same, and S is uniformly taken as the facing area between the turns of the primary and secondary windings.

The charge $Q_{P_1S_1}$ stored between the first turn P_1 of the primary winding and the secondary winding S_1 can be expressed as:



▲**Figure 2. Schematic diagrams of different angles of the outlet ports of the primary and secondary windings**



▲ **Figure 3. Schematic diagram of the laminated structure of a planar transformer**

▼**Table 1. Planar transformer laminated structure parameters**

d_1/mil	d_2/mil	d_3/mil	S/mm^2
4.5	4.3	5	214.5

$$Q_{p_1s_1} = C_{p_1s_1}(v_{p_1m} - v_{s_1m}), \tag{6}$$

where $C_{p_1s_1}$ is the total structural capacitance between P_1 and S_1, v_{p_1m} is the midpoint voltage of P_1, and v_{s_1m} is the midpoint voltage of S_1.

In the same way, the charge $Q_{P_2S_1}$ stored between the second turn P_2 of the primary winding and the secondary winding S_1 is:

$$Q_{p_2s_1} = C_{p_2s_1}(v_{p_2m} - v_{s_1m}), \tag{7}$$

where $C_{p_2s_1}$ is the total structural capacitance between P_2 and S_1, and v_{p_2m} is the midpoint voltage of P_2.

The charge $Q_{P_3S_1}$ stored between the third turn P_3 of the primary winding and the secondary winding S_1 is:

$$Q_{p_3s_1} = C_{p_3s_1}(v_{p_3m} - v_{s_1m}), \tag{8}$$

where $C_{p_3s_1}$ is the total structural capacitance between P_3 and S_1, and v_{p_3m} is the midpoint voltage of P_3.

Then the charge Q_{all} stored between the primary and secondary windings of the entire planar transformer can be expressed as:

$$Q_{\text{all}} = Q_{p_1s_1} + Q_{p_2s_2} + Q_{p_3s_3}, \tag{9}$$

From the perspective of the entire planar transformer, the planar transformer in Fig. 4 has four endpoints: A, B, C, and D, where A and C are ends with the same name. According to the relationship between the voltage ratio of the primary and secondary sides of the transformer and the turn ratio, it can be known that the voltage drop Δv_p per turn of the primary winding of the transformer is the same as the voltage drop Δv_s per turn of the secondary winding. If the voltage drop per turn is Δv, then we have:

$$\Delta v = \frac{v_p}{N_p} = \frac{v_s}{N_S}, \tag{10}$$

where v_p is the port voltage of the entire primary winding of the planar transformer, v_s is the port voltage of the entire secondary winding of the planar transformer, N_p is the number of turns of the primary winding of the planar transformer, and N_s is the number of turns of the secondary winding of the planar transformer. According to the stacked structure of the planar transformer, N_p takes 3 and N_s takes 1.

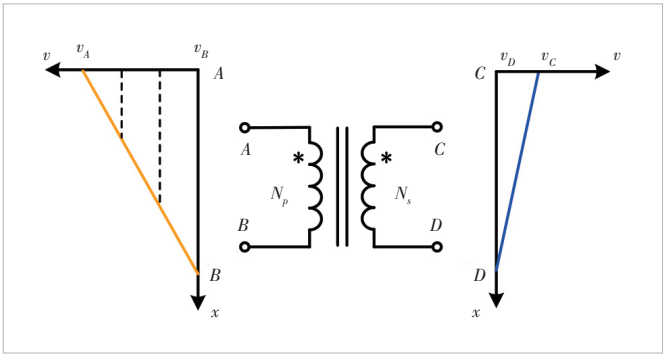
Considering the voltage distribution of the planar transformer winding in Fig. 4, it is only necessary to select the voltage of a certain point on the primary winding of the transformer as the voltage reference point, and the potential of any other point can be expressed by the voltage reference point and the voltage drop Δv per turn; the same is true for the secondary winding.

The point inside the transformer winding is used as the voltage reference point. When modeling, the model of the transformer needs to be split, which complicates the modeling. Therefore, in order to simplify the modeling, the four terminals A, B, C, and D of the transformer are used as alternative voltage reference points. By selecting point B of the primary winding and point D of the secondary winding as the potential reference points, the midpoint potentials v_{p_1m}, v_{p_2m}, v_{p_3m} and v_{S_1m} of each turn winding in the planar transformer can be expressed as:

$$v_{p_1m} = v_B + 2.5\Delta v, \tag{11}$$

$$v_{p_2m} = v_B + 1.5\Delta v, \tag{12}$$

$$v_{p_3m} = v_B + 0.5\Delta v, \tag{13}$$



▲ **Figure 4. Flat transformer winding voltage distribution**

$$v_{s_1 m} = v_D + 0.5\Delta v\,. \tag{14}$$

Substituting Eqs. (6) – (8) and Eqs. (11) – (14) into Eq. (9), the total induced charge Q_{all} can be expressed as:

$$Q_{\text{all}} = \left(C_{p_1 s_1} + C_{p_2 s_1} + C_{p_3 s_1}\right) v_{\text{BD}} + \left(2C_{p_1 s_1} + C_{p_2 s_1}\right)\Delta v, \tag{15}$$

where $C_{p_1 s_1}$, $C_{p_2 s_1}$, and $C_{p_3 s_1}$ are quantitative after the planar transformer structure and material are determined, and the induced charge Q_{all} can be calculated with only two voltage variables v_{BD} and Δv. Q_{all} reflects the induced charge between the primary and secondary windings, and replaces the voltage variable Δv with the voltage variables v_{AD} and v_{BD} between the primary and secondary windings, as shown in Eq. (16).

$$\Delta v = \frac{v_A - v_B}{3} = \frac{v_{AD} - v_{BD}}{3}. \tag{16}$$

Substituting Eq. (16) into Eq. (15) and eliminating Δv, we can get:

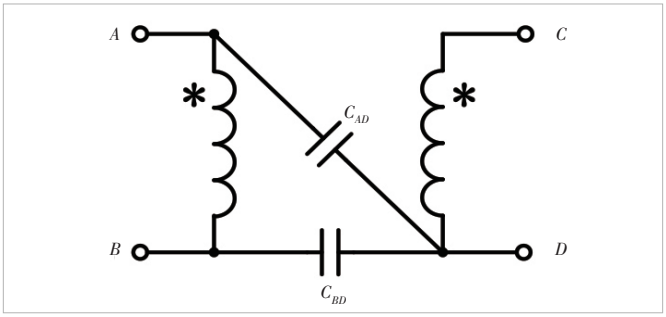
$$Q_{\text{all}} = \left(\frac{1}{3}C_{p_1 s_1} + \frac{2}{3}C_{p_2 s_1} + C_{p_3 s_1}\right) v_{BD} + \left(\frac{2}{3}C_{p_1 s_1} + \frac{1}{3}C_{p_2 s_1}\right) v_{AD}. \tag{17}$$

Eq. (17) shows that the induced charge Q_{all} is divided into two parts, and the induced charge Q_{all} can be reduced to the transformer port BD and port AD. Then the expressions of capacitance C_{BD} and C_{AD} are:

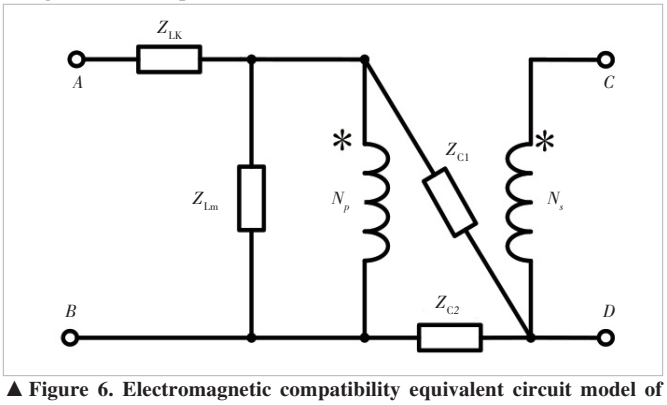
$$C_{BD} = \frac{1}{3}C_{p_1 s_1} + \frac{2}{3}C_{p_2 s_1} + C_{p_3 s_1}, \tag{18}$$

$$C_{AD} = \frac{2}{3}C_{p_1 s_1} + \frac{1}{3}C_{p_2 s_1}\,. \tag{19}$$

The capacitance C_{BD} between B and D and the capacitance C_{AD} between A and D are the equivalent common-mode capacitance of the planar transformer. The sum of the induced charges at both ends of the capacitance C_{BD} and C_{AD} is Q_{all}. The corresponding two-capacitor models of the transformer are shown in Fig. 5.



▲Figure 5. Two-capacitor model



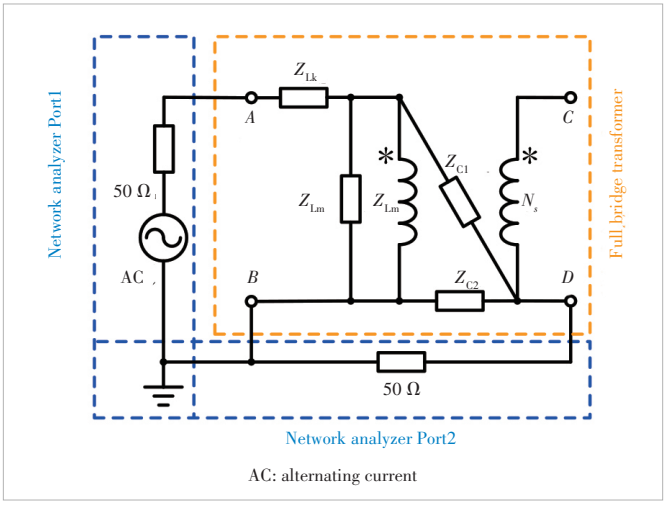
▲Figure 6. Electromagnetic compatibility equivalent circuit model of planar transformer

3 Electromagnetic Compatibility Equivalent Circuit Model of Planar Transformer

The electromagnetic compatibility equivalent circuit model of the full-bridge transformer is shown in Fig. 6. This model can not only realize the energy transmission characteristics of the transformer but also reflect its electromagnetic compatibility characteristics.

N_p represents the number of turns of the primary winding, N_s the number of turns of the secondary winding, Z_{Lk} the leakage inductance impedance, Z_{Lm} the magnetizing inductance impedance, and Z_{C1} and Z_{C2} the common mode equivalent capacitance impedance. When the frequency is low, Z_{Lk} and Z_{Lm} can be represented by pure inductance, and Z_{C1} and Z_{C2} can be represented by pure capacitance. However, as the frequency increases, the electromagnetic environment inside the transformer becomes more and more complex, and the electromagnetic compatibility characteristics inside the transformer can no longer be represented by the pure inductance or pure capacitance model of these four impedance parameters. At the same time, in order to take into account the basic circuit functions of transmitting energy, the lumped concept is adopted, and the positions of Z_{Lk}, Z_{Lm}, Z_{C1}, and Z_{C2} are kept unchanged. According to the specific impedance curve of the four parameters, a high-order model composed of inductance, resistance and capacitance is used to represent the electromagnetic compatibility equivalent circuit model of the planar transformer. At the operating frequency of the switch tube of the isolated converter, the model can realize the energy transfer from the primary side to the secondary side, and at the frequency of conducted electromagnetic interference, the model can reflect the electromagnetic compatibility characteristics of the transformer.

To extract Z_{C1} and Z_{C2}, it is necessary to simulate the actual working conditions of the planar transformer in the full-bridge circuit, and to form the same potential distribution on



▲**Figure 7. Schematic diagram of the simulation wiring diagram of the actual working condition of the full-bridge transformer**

the primary and secondary windings of the transformer as the actual working condition. Fig. 7 shows the schematic diagram of the simulation wiring diagram of the actual working condition of the full-bridge transformer. The network analyzer Port1 applies voltage excitation. The excitation output terminal of Port1 is connected to the primary winding point *A* of the full-bridge transformer, and the other excitation output reference terminal is connected to the primary winding point *B* of the full-bridge transformer. The network analyzer Port2 receives the common mode noise current generated by the planar transformer, the signal-receiving terminal of Port2 is connected to the secondary winding point *D*, and the other signal-receiving reference terminal is connected to the primary winding point *B*.

The measurement result of the network analyzer is the scattering matrix, that is, the *S*-parameter matrix, which describes the relationship between the reflected wave of the port voltage and the incident wave. According to the relationship between the port voltage and the current, the *S*-parameter matrix can be converted into a *Z*-parameter matrix, as shown in Eq. (20).

$$\begin{cases} Z_{11} = Z_0 \dfrac{(1+S_{11})(1-S_{22})+S_{12}S_{21}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}} \\ Z_{12} = Z_0 \dfrac{2S_{12}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}} \\ Z_{21} = Z_0 \dfrac{2S_{21}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}} \\ Z_{22} = Z_0 \dfrac{(1-S_{11})(1+S_{22})+S_{12}S_{21}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}} \end{cases}, \quad (20)$$

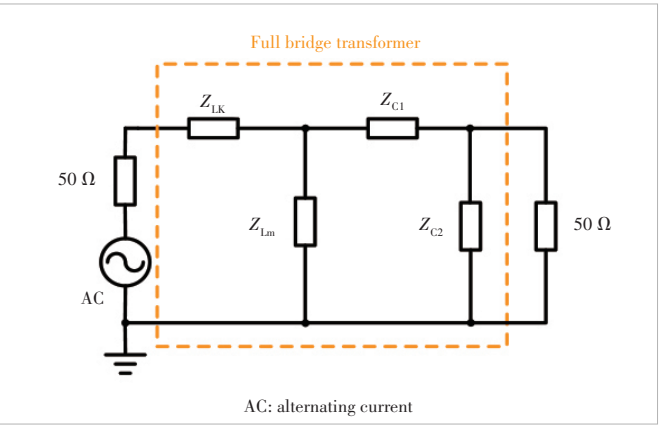
where S_{21} and S_{12} are the transmission coefficients in the scattering matrix, S_{11} and S_{22} are the reflection coefficients in the scattering matrix, and Z_0 is the characteristic impedance.

According to the test principle, the magnetic field of the magnetic core of the planar transformer is excited during the test, the potential distribution of the primary and secondary windings of the transformer is formed, and the leakage magnetic field between the primary and secondary windings is formed. Therefore, the network analyzer measures a two-port network containing transformer leakage inductance Lk, excitation inductor Lm, and equivalent common-mode capacitors C1 and C2. We simplify the wiring schematic of Fig. 7 and obtain an equivalent circuit diagram for measurement as shown in Fig. 8.

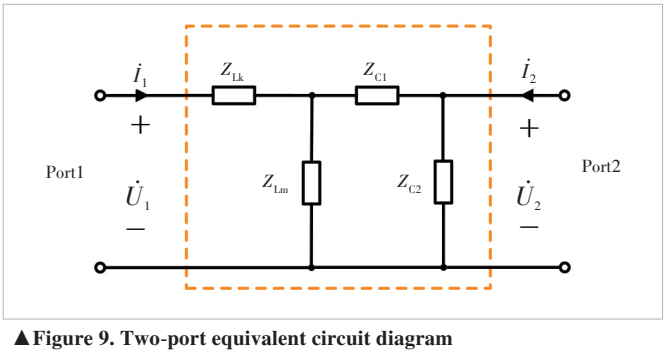
According to the circuit theory, the two-port *Z*-parameter can be expressed as:

$$\begin{cases} \dot{U}_1 = Z_{11}\dot{I}_1 + Z_{12}\dot{I}_2 \\ \dot{U}_2 = Z_{21}\dot{I}_1 + Z_{22}\dot{I}_2 \end{cases}. \quad (21)$$

The test result of the network analyzer is a two-port matrix, which makes Fig. 8 further equivalent to the two-port network of Fig. 9, and the *Z*-parameter expression of the two ports is shown in Eq. (21). For the reciprocal two-port network with three variables Z_{11}, $Z_{12}(=Z_{21})$, and Z_{22}, three equations can be listed, and three unknowns can be solved theoretically.



▲**Figure 8. Equivalent circuit diagram of simulated wiring of full-bridge transformer in actual working conditions**



▲**Figure 9. Two-port equivalent circuit diagram**

Therefore, for the above two-port equivalent circuit, first, the transformer secondary port *CD* in Fig. 6 is short-circuited, and the leakage inductance impedance Z_{Lk} is measured using a network analyzer or impedance analyzer. In Fig. 9, given Z_{Lk}, opening the port Port2, we can get Eqs. (22) and (23):

$$Z_{11} = \frac{\dot{U}_1}{\dot{I}_1}\bigg|_{\dot{I}_2 = 0} = Z_{Lk} + Z_{Lm} // \left(Z_{C1} + Z_{C2}\right), \tag{22}$$

$$Z_{12} = Z_{21} = \frac{\dot{U}_2}{\dot{I}_1}\bigg|_{\dot{I}_2 = 0} = \frac{\left[Z_{Lm} // \left(Z_{C1} + Z_{C2}\right)\right] Z_{C2}}{Z_{C1} + Z_{C2}}. \tag{23}$$

By opening Port1, Z_{22} can be represented as:

$$Z_{22} = \frac{\dot{U}_2}{\dot{I}_2}\bigg|_{\dot{I}_1 = 0} = Z_{C2} // \left(Z_{C1} + Z_{Lm}\right). \tag{24}$$

Combining Eqs. (22) – (24), we can get the expressions of Z_{Lm}, Z_{C1} and Z_{C2} with Z_{Lk}, and Z_{11}, Z_{12} and Z_{22} as variables:

$$Z_{Lm} = \frac{Z_{11}Z_{22} - Z_{21}^{\ 2} - Z_{Lk}Z_{22}}{Z_{22} - Z_{21}}, \tag{25}$$

$$Z_{C1} = \frac{Z_{11}Z_{22} - Z_{21}^{\ 2} - Z_{Lk}Z_{22}}{Z_{21}}, \tag{26}$$

$$Z_{C2} = \frac{Z_{11}Z_{22} - Z_{21}^{\ 2} - Z_{Lk}Z_{22}}{Z_{11} - Z_{Lk} - Z_{21}}. \tag{27}$$

Z_{Lm}, Z_{C1} and Z_{C2} can be solved according to Eqs. (25) – (27). After obtaining the impedance curves of the parameters Z_{Lm}, Z_{C1}, Z_{C2} and Z_{Lk} required for modeling, the circuit parameters of the electromagnetic compatibility equivalent circuit model of the planar transformer are determined by the method of circuit model fitting.

4 Planar Transformer Simulation

4.1 Extraction of Material Parameters of Planar Transformer Core

To obtain the impedance curves of magnetic core materials such as inductors and transformers with the help of simulation tools, it is first necessary to extract the complex permeability of the magnetic core material. The complex permeability of the magnetic core material has an important influence on the electromagnetic compatibility characteristics of magnetic components, and is of great significance for the product design selection of magnetic components and the research on electromagnetic compatibility characteristics. Although the data sheets provided by some magnetic core material manufacturers will include the complex permeability of the magnetic core material, the complex permeability frequency range provided in the data sheet is usually in the range of several kHz to several MHz, which cannot completely cover the frequency band range of conducted electromagnetic interference research (150 kHz – 30 MHz). Therefore, it is necessary to extract the complex permeability of the magnetic core material through experimental measurement.

Because the distribution parameters of magnetic components are very rich, there are distributed capacitances between the magnetic core and the windings, and between the windings. To reduce the effect of distributed capacitance on complex permeability measurements, the complex permeability of a magnetic core is measured using a single-turn centering method, as shown in Fig. 10. The single-turn winding uses copper wires with a short length and a major diameter to reduce the influence of the winding loss and lead inductance of the single-turn winding on the measurement of the complex magnetic permeability of the magnetic core.

The impedance characteristic of the single-turn coil is measured with an impedance analyzer. The equivalent model of the measurement is:

$$Z = R_s + j\omega L_s, \tag{28}$$

where R_s and L_s are the equivalent series resistance and equivalent series inductance of the magnetic component under test, respectively. It can be approximated that the R_s in the test result is only generated by the core loss, and L_s is only generated by the magnetic permeability of the magnetic core.

Thus according to the definition of complex permeability:

$$\mu = \mu' - j\mu'', \tag{29}$$

▲ **Figure 10. Experimental diagram of the single-turn through-center measurement method**

where μ' is the real permeability and μ'' is the imaginary permeability. The equivalent series model of the core impedance can be expressed as:

$$Z = j\omega \frac{\mu N^2 A_e}{l_e}, \tag{30}$$

where N is the number of turns of the winding, l_e is the equivalent magnetic circuit length, A_e is the cross-sectional area of the magnetic core, and ω is the measured angular frequency.

Substituting Eq. (29) into Eq. (30), we can get:

$$Z = j\omega \frac{(\mu' - j\mu'')N^2 A_e}{l_e}. \tag{31}$$

Comparing Eqs. (28) and (31), we can get the real and imaginary parts of the complex permeability of the magnetic core as follows:

$$\mu' = \frac{L_s l_e}{N^2 A_e}, \tag{32}$$

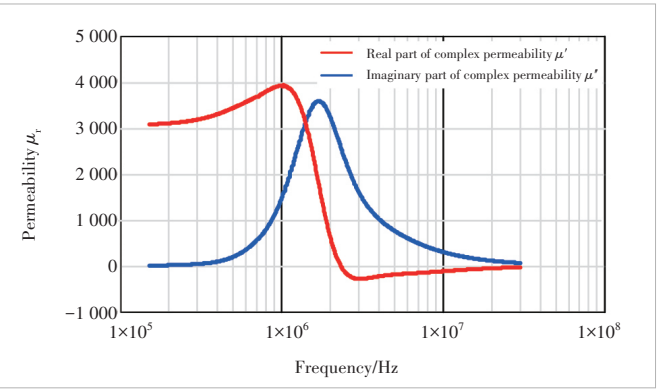
$$\mu'' = \frac{R_s l_e}{\omega N^2 A_e}. \tag{33}$$

The core loss tangent is:

$$\tan(\theta) = \frac{\mu''}{\mu'}. \tag{34}$$

In order to verify that the complex permeability obtained by the experimental measurement can truly represent the core parameters of the actual sample in the finite element simulation software, an impedance analyzer is used to extract the complex permeability of the magnetic ring in Fig. 11 according to the method in Fig. 10. The impedance analyzer

▲Figure 11. Actual core samples



▲Figure 12. Real and imaginary parts of core complex permeability extracted by the impedance analyzer

model is WK 6500B, the core material of the magnetic ring is DMR96, the measured outer diameter is 25 mm, the inner diameter is 15 mm, and the height is 7 mm.

The real and imaginary parts of the complex magnetic permeability of the magnetic ring obtained by the impedance analyzer are shown in Fig. 12.

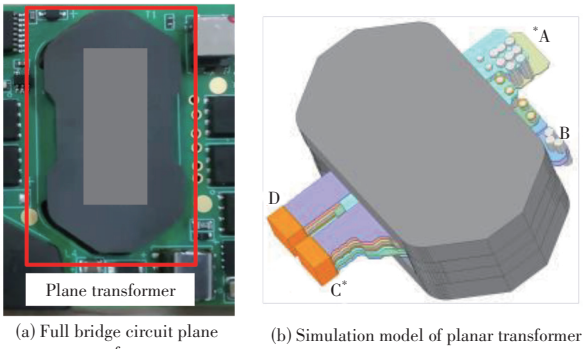
4.2 Planar Transformer Simulation Modeling

The measurement of the two-port parameters of the planar transformer by the network analyzer is introduced previously. According to the measurement results, the impedance curves of the leakage inductance, magnetizing inductance and equivalent common mode capacitance of the transformer are obtained, and the electromagnetic compatibility equivalent circuit model of the transformer is established. However, the method based on the actual sample measurement of the transformer cannot realize the prediction of conducted EMI noise in the design stage of planar transformer, and the method of deriving the equivalent capacitance of common-mode noise based on the theory of transformer winding stacked structure can only reflect the transmission characteristics of transformer common-mode noise at low frequency. The electromagnetic field finite element simulation software HFSS of Ansys can simulate the function of the network analyzer, obtain the two-port Z-parameters of the transformer, and realize the extraction of the wide-band parameters of the planar transformer without physical objects. At the same time, due to the use of a planar transformer, the accuracy and consistency of the winding structure parameters are higher than those of the traditional manual winding transformer, which is conducive to improving the consistency between the simulation model and the real thing.

Take the planar transformer of the full-bridge circuit as an example. The magnetizing inductance is designed to be 110 μH, the turn ratio of the primary winding and the secondary winding is 3:1, the number of turns on the primary side is 3, and the number of turns on the secondary side is 1. The core material is DMR96, the relative permittivity of

plate FR4 is set to 4.4, and the physical diagram and simulation model of the planar transformer are shown in Fig. 13. The physical diagram of Fig. 13(a) shows that the planar transformer of the full-bridge circuit is integrally installed. We import the PCB of the full-bridge circuit into the simulation software and separate the PCB layout of the winding part of the planar transformer to establish the simulation model of the planar transformer.

The actual planar transformer magnetic core is composed of two magnetic cores. Generally, in order to control the inductance of the magnetizing inductance during the production process, an air gap will be opened on the central column of the magnetic core. This air gap is very small and difficult to measure accurately, but the size of the air gap can be determined by comparing the magnetizing inductance. In the simulation, the transformer's secondary winding is left open and the primary winding is excited, and the input impedance curve can be obtained, which includes the magnetizing inductance, the leakage inductance and the inter-turn capacitance of the transformer. Since the planar transformer adopts the interleaved winding method, the leakage inductance is far smaller than the magnetizing inductance. The input impedance curve at a low frequency can be approximately regarded as the impedance curve of the magnetizing inductance. After determining the size of the air gap, we short-circuit the secondary winding to obtain the leakage inductance Z_{Lk} impedance curve of the transformer. After that, we add excitation according to the measurement method in Fig. 7 to obtain the Z-parameter impedance curve of the transformer at two ports, and derive Z_{Lm}, Z_{C1}, and Z_{C2} in the model according to Eqs. (25) – (27). The simulation results are shown in Fig. 14.



(a) Full bridge circuit plane transformer

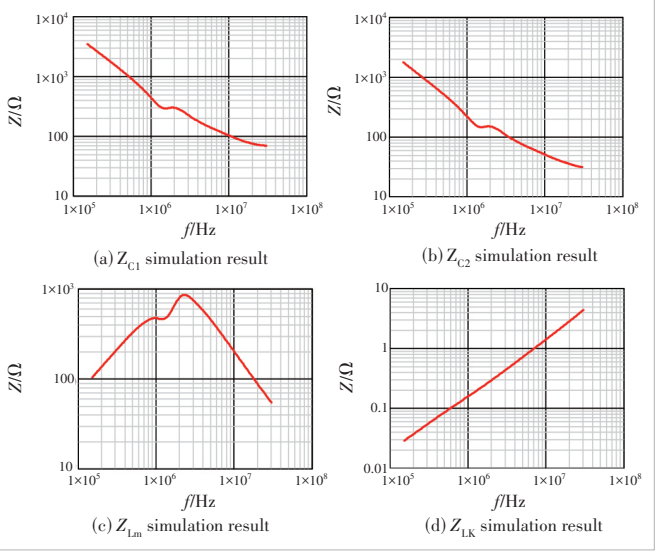
(b) Simulation model of planar transformer

▲Figure 13. Full-bridge circuit planar transformer simulation model

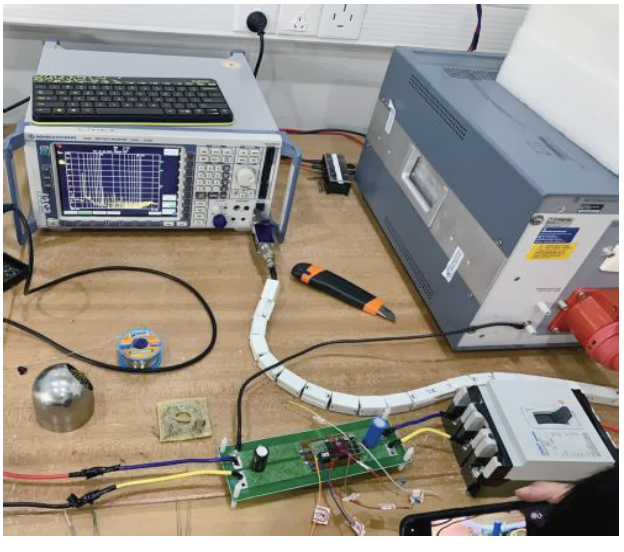
5 Experimental Verification

We use the full-bridge circuit module in the wireless base station power supply as the experimental prototype. The rated input voltage of the full-bridge circuit module is 48 V, the rated output voltage is 12 V, the rated output current is 35.8 A, and the operating frequency of the circuit is 170 kHz. The experimental prototype is shown in Fig. 15. The planar transformer used in the experimental prototype is the transformer mentioned above. The conduction test of the full bridge circuit is shown in Fig. 15.

According to the generation and conduction mechanism of common mode noise which is mainly caused by the instantaneous potential change when the switching device is turned on or off, and the distributed capacitance that exists between the power supply and ground is affected by the rising and falling edges of the switching device voltage, resulting in noise currents flowing through the line impedance stabilization network (LISN), ground, and P and N lines with the same amplitude and direction. The common-mode noise cir-



▲Figure 14. Planar transformer simulation results



▲Figure 15. Full bridge circuit conduction test

cuit of the full-bridge circuit is shown in Fig. 16. There are 5 main potential jump points, of which the four potential jump points are the 4 terminals of the transformer. Part of the common-mode noise current at the two potential trip points on the primary side of the transformer conducts the common-mode noise current to ground through the switching device's coupling capacitor to the ground, and the other part of the common-mode noise current is transmitted from the primary side to the secondary side through the interturn capacitance of the transformer through the potential change of the transformer primary side and the secondary side winding, and then flows into the earth, LISN, and back to the primary side. For the two potential jump points on the secondary side of the transformer, since the secondary side is connected to the ground, the common-mode noise current flowing through the coupling capacitor between the switching device and ground flows directly back on the secondary side without flowing through the LISN.

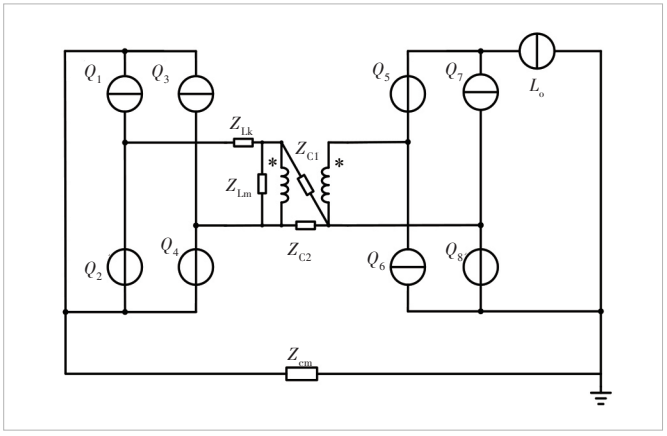
According to the substitution theorem, the switches Q_1, Q_3, Q_6, and Q_7 and the output inductor L_o are equivalent to current sources, and the switches Q_2, Q_4, Q_5, and Q_8 are equivalent to voltage sources. The input capacitor C_{in} and the output capacitor C_o can be regarded as a short circuit due to their small impedance, and the common mode impedance Z_{cm} of the LISN can be regarded as 25 Ω. The equivalent circuit diagram after the application of the substitution theorem is shown in Fig. 17.

When the current sources work together, all voltage sources are short-circuited, and an equivalent circuit diagram can be obtained. The primary current source is short-circuited, the voltage at the transformer port is 0, and the secondary side current source is connected in parallel across the secondary winding of the transformer, which is also short-circuited. All current sources are short-circuited when the current source is active, and no common mode noise current is formed in Zcm. The equivalent circuit diagram considering only the current source is shown in Fig. 18.

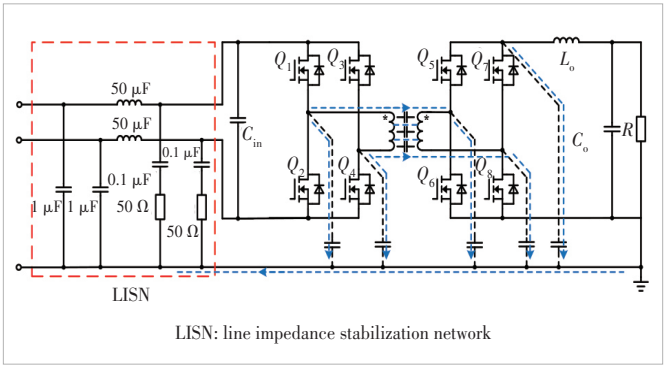
When the voltage sources work together, open all current sources to obtain the equivalent circuit diagram. The equivalent circuit diagram considering only the voltage sources is shown in Fig. 19. Q_2, Q_4 and Q_8 will generate common mode noise current on Z_{cm}. Using the superposition theorem for the voltage sources Q_2, Q_4, and Q_8, the common-mode noise voltage $\dot{V}_{cm}$ generated on Z_{cm} can be found:

$$\dot{V}_{cm}=\frac{-1}{\left(\frac{1}{Z_{cm}}+\frac{1}{Z_{AD}}+\frac{1}{Z_{BD}}\right)-\frac{1}{Z_{AD}}\cdot\frac{\frac{1}{Z_{AD}}}{\frac{1}{Z_{Lk}}+\frac{1}{Z_{Lm}}+\frac{1}{Z_{AD}}}}\cdot\left[\frac{1}{Z_{Lm}}\cdot\dot{V}_{Q_8}+\frac{1}{Z_{BD}}\cdot\dot{V}_{Q_4}+\frac{1}{Z_{AD}}\cdot\left(\frac{\frac{1}{Z_{Lk}}\cdot\dot{V}_{Q_2}+\frac{1}{Z_{Lm}}\dot{V}_{Q_4}}{\frac{1}{Z_{Lk}}+\frac{1}{Z_{Lm}}+\frac{1}{Z_{AD}}}\right)\right]+\dot{V}_{Q_8} \quad (35)$$

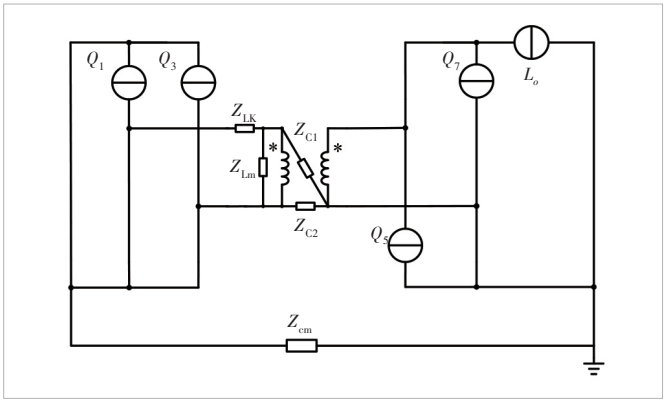
We use an oscilloscope to measure the voltage waveforms $\dot{V}_{Q_2}$, $\dot{V}_{Q_4}$, and $\dot{V}_{Q_8}$ across the switching tubes Q_2, Q_4, and Q_8, perform Fourier decomposition on them, substitute the results into Eq. (35) to obtain the predicted noise spectrum, and compare it with the measured results, which is shown in



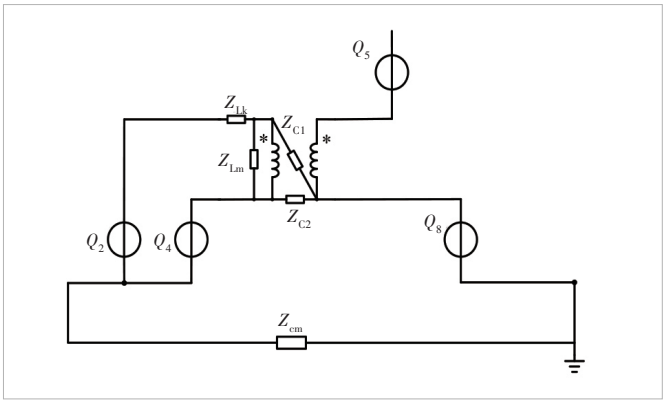
▲Figure 17. Equivalent circuit diagram after applying the substitution theorem



▲Figure 16. Full-bridge circuit noise path



▲Figure 18. Equivalent circuit diagram considering only current source



▲Figure 19. Equivalent circuit diagram considering only the voltage source

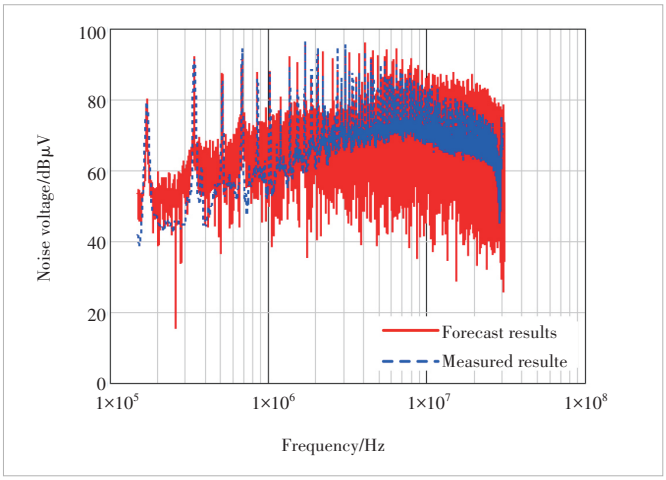
Fig. 20. The noise prediction results are in good agreement with the measured results, which shows the accuracy of the extraction and modeling of the common mode noise equivalent circuit of the planar transformer.

6 Conclusions

In this paper, the mechanism of common-mode noise conduction of planar transformers is analyzed and a method for extracting planar transformer models through simulation is introduced. The following conclusions are obtained:

1) The common-mode noise transmission capability of planar transformers is reflected in the induced charge Q between the primary and secondary windings of planar transformers. At the same time, according to the theoretical derivation of the potential distribution on the planar transformer winding, the induced charge Q can be represented by two capacitors.

2) When using a network analyzer to measure a planar transformer, the measurement result is a two-port network including leakage inductance, magnetizing inductance, and inter-turn capacitance.



▲Figure 20. Predicted noise spectrum and measured noise spectrum

3) With the help of the simulation software, the Z-parameters of the planar transformer can be obtained, and the EMI model of the planar transformer can be established through the Z-parameters. Finally, the accuracy of the simulation and modeling is verified through a full-bridge circuit prototype experiment.

References

[1] MOHAMMADI M, ADIB E, YAZDANI M R. Family of soft-switching single-switch PWM converters with lossless passive snubber [J]. IEEE transactions on industrial electronics, 2015, 62(6): 3473 - 3481. DOI: 10.1109/TIE.2014.2371436
[2] LI Y M, ZHANG H, WANG S, et al. Investigating switching transformers for common mode EMI reduction to remove common mode EMI filters and Y-capacitors in flyback converters [J]. IEEE journal of emerging and selected topics in power electronics, 2018, 6(4): 2287 - 2301. DOI: 10.1109/JESTPE.2018.2827041
[3] ZHANG H, WANG S, LI Y M, et al. Two-capacitor transformer winding capacitance models for common-mode EMI noise analysis in isolated DC - DC converters [J]. IEEE transactions on power electronics, 2017, 32(11): 8458 - 8469. DOI: 10.1109/TPEL.2017.2650952
[4] CHU Y B, WANG S. A generalized common-mode current cancelation approach for power converters [J]. IEEE transactions on industrial electronics, 2015, 62(7): 4130 - 4140. DOI: 10.1109/TIE.2014.2387335
[5] TRIA L A R, ZHANG D M, FLETCHER J E. Planar PCB transformer model for circuit simulation [J]. IEEE transactions on magnetics, 2016, 52(7): 1 - 4. DOI: 10.1109/tmag.2016.2516995
[6] ALI SAKET M, ORDONEZ M, SHAFIEI N. Planar transformers with near-zero common-mode noise for flyback and forward converters [J]. IEEE transactions on power electronics, 2018, 33(2): 1554 - 1571. DOI: 10.1109/TPEL.2017.2679717
[7] MASWOOD A I, SONG L K. Design aspects of planar and conventional SMPS transformer: A cost benefit analysis [J]. IEEE transactions on industrial electronics, 2003, 50(3): 571 - 577. DOI: 10.1109/TIE.2003.812469
[8] ALI SAKET M, SHAFIEI N, ORDONEZ M. Planar transformer winding technique for reduced capacitance in LLC power converters [C]//Proceedings of 2016 IEEE Energy Conversion Congress and Exposition (ECCE). IEEE, 2017: 1 - 6. DOI: 10.1109/ECCE.2016.7855352
[9] ZHANG Z L, HE B H, HU D D, et al. Common-mode noise modeling and reduction for 1-MHz eGaN multioutput DC - DC converters [J]. IEEE transactions on power electronics, 2019, 34(4): 3239 - 3254. DOI: 10.1109/TPEL.2018.2850351
[10] CHEN Q B, CHEN W. An evaluation method of transformer behaviors on the suppression of common-mode conduction noise in switch mode power supply [J]. Proceedings of the CSEE, 2012, 32(18): 73-79+180
[11] FU K N, CHEN W. Evaluation method of flyback converter behaviors on common-mode noise [J]. IEEE access, 2019, 7: 28019 - 28030. DOI: 10.1109/ACCESS.2019.2902462
[12] LIU S, ZHANG Y T, YU D Y. Research and design of EMI digital filters using scattering parameters [C]//International Conference on Wireless Communications & Signal Processing. IEEE, 2009: 1 - 5. DOI: 10.1109/WCSP.2009.5371392
[13] WANG S, LEE F C, ODENDAAL W G. Characterization and parasitic extraction of EMI filters using scattering parameters [J]. IEEE transactions on power electronics, 2005, 20(2): 502 - 510. DOI: 10.1109/TPEL.2004.842949
[14] FRICKEY D A. Conversions between S, Z, Y, H, ABCD, and T parameters which are valid for complex source and load impedances [J]. IEEE transactions on microwave theory and techniques, 1994, 42(2): 205 - 211. DOI: 10.1109/22.275248

Biographies

LI Wei (li.wei27@zte.com.cn) received his MS degree in mechatronic engineering from Southeast University, China in 2017. Currently he is working as an EMC engineer in ZTE Corporation. His research interests include switching power supply of EMC and lightning protection.

JI Jingkang received his MS degree in mechatronic engineering from Southeast University, China in 2020. He currently works as an EMC engineer in ZTE Corporation. His research interests include switching power supply of EMC and reverberation chambers.

LIU Yuanlong received his MS degree from Harbin Institute of Technology, China. Currently he is working as an EMC engineer in ZTE Corporation. His research interests include lightning protection, switching power supply of EMC, and engineering safety.

SUN Jiawei received his BE degree in electrical engineering and automation and ME degree in power electronics and power transmission both from Fuzhou University, China in 2018 and 2022, respectively. His research interest focuses on power electronics high frequency magnetic technology.

LIN Subin received his doctor's degree from Fuzhou University, China. He currently works as an associate professor with the School of Electrical Automation, Fuzhou University. He has been engaged in the theoretical research and technical development of power electronic magnetic components for a long time. His main research direction is power electronic electromagnetic component technology, electromagnetic compatibility analysis and diagnosis.

Statistical Model of Path Loss for Railway 5G Marshalling Yard Scenario

DING Jianwen[1], LIU Yao[1], LIAO Hongjian[1], SUN Bin[1], WANG Wei[2]

(1. Beijing Jiaotong University, Beijing 100044, China;
2. ZTE Corporation, Shenzhen 518057, China)

DOI: 10.12142/ZTECOM.202303015

https://link.cnki.net/urlid/34.1294.TN.20230905.2004.002, published online September 6, 2023

Manuscript received: 2022-10-25

Abstract: The railway mobile communication system is undergoing a smooth transition from the Global System for Mobile Communications-Railway (GSM-R) to the Railway 5G. In this paper, an empirical path loss model based on a large amount of measured data is established to predict the path loss in the Railway 5G marshalling yard scenario. According to the different characteristics of base station directional antennas, the antenna gain is verified. Then we propose the position of the breakpoint in the antenna propagation area, and based on the breakpoint segmentation, a large-scale statistical model for marshalling yards is established.

Keywords: 5G-R; marshalling yard; path loss prediction; statistical modeling

Citation (Format 1): DING J W, LIU Y, LIAO H J, et al. Statistical model of path loss for railway 5G marshalling yard scenario [J]. ZTE Communications, 2023, 21(3): 117 - 122. DOI: 10.12142/ZTECOM.202303015
Citation (Format 2): J. W. Ding, Y. Liu, H. J. Liao, et al., "Statistical model of path loss for railway 5G marshalling yard scenario," *ZTE Communications*, vol. 21, no. 3, pp. 117 - 122, Sept. 2023. doi: 10.12142/ZTECOM.202303015.

1 Introduction

With the deployment and advancement of the 5G "new infrastructure" strategy, a series of technological innovations for the new generation of railway mobile communication are actively being carried out in China. Taking 5G as an opportunity, 5G for Railway (5G-R) has been widely regarded as a solution to meeting the diverse requirements of railway wireless communications[1-2]. Railway 5G is composed of 5G-R and the public 5G network dedicated to the railway. Among them, 5G-R refers to a private 5G network that provides services for the railway system, which is completely independent and highly reliable. However, the current 5G-R bandwidth is limited and cannot carry a large number of high-bandwidth services in railways[3]. It is necessary to use the dedicated frequency of the public 5G network to carry out some services uncorrelated to driving safety because the 5G-R frequency has not been approved. At present, the development of the Railway 5G is in the preliminary stage, without enough technical standards for the construction of a railway wireless network. For various scenarios with higher frequencies, it is necessary to conduct a large number of field measurements and tests to collect test data. The current technical standards can be summarized and improved more accurately based on these data. Compared with the previous, the current railway system has changed a lot. Therefore, we need to determine the field strength prediction model under typical railway scenarios and provide corresponding technical standards for the construction of the Railway 5G through the evaluation of wireless network coverage. At the same time, high-speed railway is also one of the important application scenarios of the 6G mobile communication technology in the future. A 6G network can provide more comprehensive performance indicators, such as ultra-low delay jitter, ultra-high security, stereo coverage, and ultra-high positioning accuracy. With the help of 6G, more high-speed railway business and application requirements can be realized, and the development of railway digitalization can be promoted[4]. Therefore, we not only need to build a 5G-R network in an all-round way, but also make theoretical and technical preparations for 6G.

The channel parameters of 5G including path loss exponent and shadow fading were measured in the scene of campus in Ref. [5]. A new path loss prediction model for the viaduct area was deduced through statistical analysis of the measurement results in Ref. [6]. To model the path loss of viaducts and plains, Ref. [7] proposed a modified free space model with good results. The authors in Ref. [8] divided viaducts into four areas: suburban, open, mountainous and urban. They calculated respective path loss exponent and standard deviations of shadow fading. Ref. [9] measured and analyzed the path loss

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant No. 2022JBXT001, in part by NSFC under Grant No. 62171021, in part by the Project of China State Railway Group under Grant No. P2021G012, and in part by ZTE Industry-University-Institute Cooperation Funds under Grant No. I21L00220.

for the 915 MHz railway yard environment. For the scenarios of high-speed railway noise barriers and forests, the channel characteristics including delay spread and Doppler spread were analyzed based on ray tracing in Ref. [10]. On the basis of Ref. [10], the authors utilized multiple antennas for modeling and supplemented the discussion of received power in Ref. [11]. Curved and straight tunnels were measured at 900 MHz and 2 100 MHz, and the path loss modeling under train-to-train (T2T) communication was fitted[12]; Based on ray tracing simulation technology, the channel parameters such as path loss, delay spread, and Doppler spread under the 5G system in the urban rail viaduct scenario were analyzed[13].

The main contributions made in this paper are summarized as follows.

Considering the influence of the main lobe and side lobe of the directional antenna pattern of a base station, two methods of dividing the propagation area are proposed: one is based on the side lobe and main lobe coverage area division, and the other is based on the center of the main lobe of the antenna pattern. And the above two methods were compared with the Fresnel band gap division method.

Based on the field strength data measured in the actual marshalling yard scene, the above-mentioned conjecture of the division of the propagation area is verified, and a large-scale fading empirical model is established. The radio wave propagation area is divided into Area A and Area B for segmental modeling.

It is concluded that the path loss model of Area A is similar to the two-path model, and that of Area B is consistent with the traditional empirical model of large-scale fading.

Combined with the Hata model, the relationship between the correction factor and the antenna height of the base station in the marshalling station scenario is fitted. Then we propose a correction factor for Area B and establish an empirical model about this area.

2 Data Preprocessing

2.1 Antenna Gain Check

The effects of the main lobe and side lobes of the antenna pattern should be taken into account to build an accurate path loss model. By verifying the antenna gain of the original data samples, the application scope of the model can be expanded effectively. Firstly, we verify the original measurement data to obtain the received power of the reference signal at the test point. The path loss of the test point is given by

$$PL(d) = P_t - P_r(d) + G(d,\theta_1) - L_{\text{loss}}, \quad (1)$$

where P_t defines the reference signal transmitting power, $P_r(d)$ is the test point reference signal received power, L_{loss} denotes the feeder and its connection loss in the test system, and $G(d,\theta_1)$ expresses the vertical gain sum of the transmitting and receiving antennas of the base station.

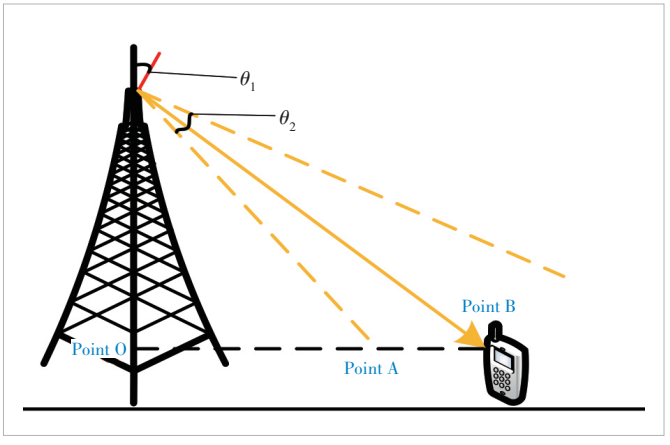
The horizontal beam of the transmitting antenna we adopted is wide, and the receiving antenna is a horizontal omnidirectional antenna. The antennas are linear coverage in the scenario of a marshalling yard. It can be approximately considered that the gain of the transmitting and receiving antennas is invariable on the horizontal plane. Therefore, only the vertical antenna pattern is considered and the horizontal is irrespective in this paper.

Result $PL(d)$ obtained after antenna gain verification is the sum of path loss and shadow fading. Relative to the transmitting antenna, the receiving antenna moves slowly and the shadow fading has less effect on the overall results. Therefore, a large-scale path loss model is built based on the results obtained through antenna gain verification in the following work of this paper.

2.2 Radio Wave Propagation Area Division

The base station antennas generally used in the construction of rail transit wireless communication systems are high-gain directional antennas. Directional antennas are characterized by a small coverage angle, strong directivity, and a large signal gap between the main lobe and the side lobe. Fig. 1 is a 2D schematic diagram of a directional antenna system, where θ_1 denotes the downtilt angle of the base station antenna, taking into account both mechanical downtilt and electrical downtilt, and θ_2 denotes the 3 dB lobe width of the base station antenna in the vertical direction.

In Fig. 1, Point O represents the location of the base station tower, Point A represents the position of the vertical side lobe edge of antenna, and Point B represents the center position of the vertical main lobe of antenna. In other words, the direction of the base station antenna pointing to Point B is the main radiation direction of the antenna, which indicates the direction with the largest gain in the antenna pattern. Generally, the gain of the directional antenna pattern is large and uniform in the main lobe, while the gain in the side lobes is small and fluctuates greatly. The main lobe and side lobes will have a great impact on the received power. The antenna pattern of the transceiver antenna should be considered as a whole to ensure accurate results.



▲Figure 1. 2D schematic diagram of directional antenna system

The antenna pattern in the near-field area of the base station has low receiving intensity and large fluctuation, which is different from the variation trend of the received power and distance of the reference signal outside the near-field area. From the antenna pattern of the base station antenna, it can be seen that the side lobe gain of the base station antenna fluctuates by 10 – 20 dB. The received power of the reference signal will be affected by both the gain of the transmit antenna and that of the receive antenna. It is indicated that the path loss model in this part of the region no longer follows the logarithmic fading model[14].

The path loss model divides the antenna coverage area into two parts based on the side lobe and main lobe coverage areas, which are referred to here as Area A and Area B. Segmentation is performed with Fresnel zone gaps in Ref. [15]. There are three ways to divide the propagation area:

1) The antenna coverage area is divided based on the side lobe and main lobe coverage areas, which is the distance of the line segment *OA* in Fig. 1. The distance of *OA* is given by:

$$d_1 = \frac{\Delta h}{\tan\left(\theta_1 + \frac{\theta_2}{2}\right)}. \tag{2}$$

2) The propagation area can be divided based on the center position of the main lobe of the antenna pattern, which is the distance of the line segment *OB* in Fig. 1. The distance to *OB* is given by:

$$d_1 = \frac{\Delta h}{\tan(\theta_1)}. \tag{3}$$

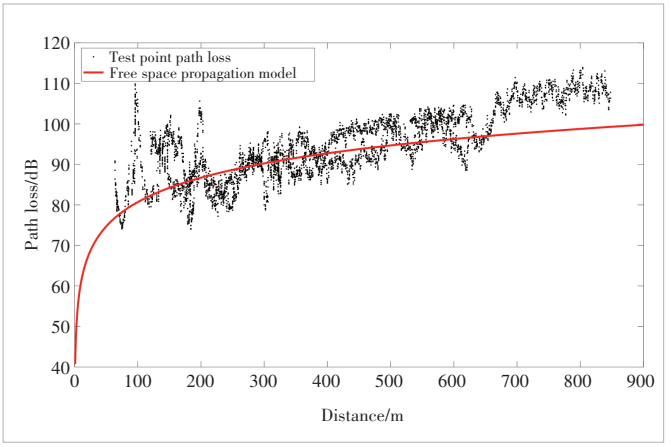
3) The propagation area can also be divided according to the Fresnel zone gap, and the distance is given by:

$$d_1 = \frac{4h_t h_r}{\lambda}. \tag{4}$$

3 Path Loss Statistical Modeling for Area A

The marshalling yard scenario tested is somewhat similar to the cutting scenario in rail transportation. The railway yard is located in the middle, with slopes and mountains covered with trees on one side and several workshops on the other side. According to the analysis, the received signal may be affected by direct waves, reflected waves, and scattered waves. The test area includes three base stations and nine cells. In the measurement, the transmitter uses a 5G AAU base station with an antenna gain of 24.5 dBi, the receiver uses a PCTEL horizontal omnidirectional antenna, and SPARK software is used for data storage and visual analysis. We select one of the cells for specific analysis.

An example of a single measurement result of the path loss after the antenna gain verification process of 150 cells in the marshalling station scenario is shown in Fig. 2. It can be seen from the figure that the path loss in the near-field area of the base station fluctuates greatly, which is different from the variation trend with a distance of the far-field area. When the conditions of 150 cells is substituted, it is calculated that *OA* = 154.64 m, *OB* = 214.66 m, and the Fresnel zone gap = 1212.17 m. From the actual test results in Fig. 2, it can be concluded that the boundary points in Definition (2) can better represent the boundary points of Area A and Area B, and the conclusions of other cells are similar. To sum up, Point B is divided as the breakpoint of the propagation area, and the models built in subsequent segmentation are based on this.



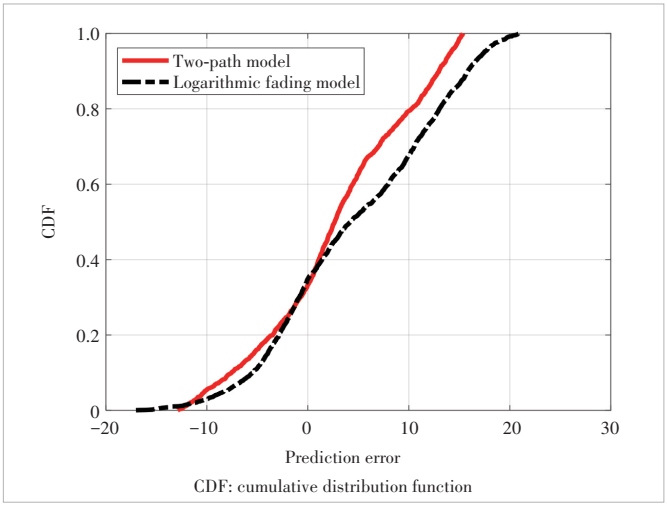
▲Figure 2. Path loss of 150 cells of the marshalling station

Due to the presence of a large number of metal products in the marshalling yard scenario, the influence of the reflection path in the near-field area is difficult to ignore. By contrasting the accuracy of the two-path model and the logarithmic fading model in Area A, we make statistics based on the path loss prediction error $e(i)$ of the sample points in each cell. The calculation formula of $e(i)$ is given by:

$$e(i) = PL_{\text{measure}}(i) - PL_{\text{predict}}(i), \tag{5}$$

where i denotes the cell sample point number, $PL_{\text{measure}}(i)$ is the i-th measured data sample, and $PL_{\text{predict}}(i)$ defines the i-th predicted data sample.

Under the same conditions, probability statistics are performed based on $e(i)$ of the path loss results predicted by the logarithmic fading model and the two-path model in the test cell of Area A. The corresponding cumulative distribution function (CDF) results are shown in Fig. 3. The average prediction error of the two-path model is 2.80 dB, and the standard deviation is 7.24 dB. The average prediction error of the logarithmic fading model is 4.75 dB, and the standard deviation is 8.26 dB. At the same time, it can be seen from Fig. 3 that in Area A, compared with the logarithmic fading model, the average prediction error of the two-path model is smaller and the change is more stable, which is more consistent with the actual measured data. Therefore, according to the calculation results, it can be concluded that in the marshalling station



▲Figure 3. Prediction error statistics of marshalling yard Area A

scenario, the change of path loss in Area A obeys the two-path model. The received power of the two-path model is as:

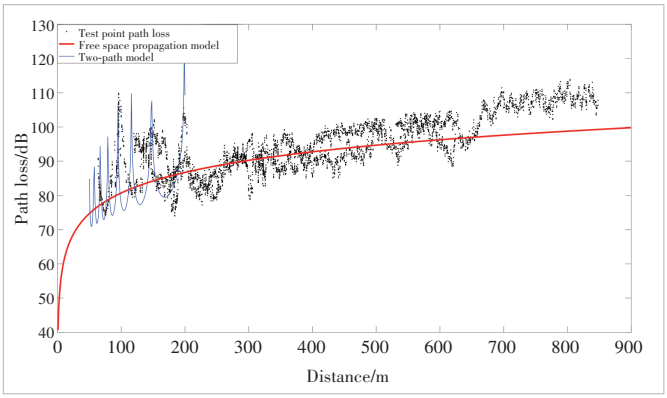
$$P_r = P_t\left(\frac{\lambda}{4\pi}\right)^2\left|\frac{\sqrt{G_l}}{d_d} + \xi\frac{\sqrt{G_r}\,e^{-jk_\lambda(d_r-d_d)}}{d_r}\right|^2, \tag{6}$$

where d_d is the length of the direct path of the transmitter and receiver, d_r is the total length of the reflection path of the transmitter and receiver, k_λ is the wave number, and ζ is the reflection coefficient of the ground. In general, $\zeta=-1$, $G_l = G_aG_b$, and $\sqrt{G_l}$ is the total gain of the transmitting and receiving antennas on the direct path from the transmitter to the receiver; $G_r = G_cG_d$, and $\sqrt{G_r}$ is the total gain of the transmitting and receiving antennas on the reflective path from the transmitter to the receiver.

The fitting results of the two-path model are shown in Fig. 4.

4 Path Loss Statistical Modeling for Area B

In this section, an empirical model of large-scale fading of radio waves is established based on a large number of test samples obtained in the marshalling station Area B. Among them, the carrier frequency of the base station is fixed at 2.6 GHz, and the height of the receiver antenna is fixed at 1 m. We build a statistical model of Area B utilizing the path loss PL, the distance d of the transceiver antenna, and the height of the base station antenna h_t in the test data.



▲Figure 4. Fitting two-path model for path loss in marshalling station Area A

The path loss formula of the Hata model is as follows:

$$PL_{\text{Hata}} = 69.55 + 26.16\lg f_c - 13.82\lg h_t - \alpha(h_r) + (44.9 - 6.55\lg h_t)\lg d, \tag{7}$$

$$\alpha(h_r) = \begin{cases}(1.1\lg f_c - 0.7)h_r - (1.56\lg f_c - 0.8), & \text{medium and small cities}\\ 8.29\lg^2(1.54h_r) - 1.1, & \text{large city, } f_c \leqslant 200\text{ MHz}\\ 3.2\lg^2(11.75h_r) - 4.97, & \text{large city, } f_c \geqslant 400\text{ MHz}\\ 0, & h_r = 1.5\text{ m}\end{cases}, \tag{8}$$

where f_c is the frequency (the unit is MHz), h_t and h_r define the effective heights of the transmitter antenna and the receiver antenna (the unit is m), and $\alpha(h_r)$ varies with the city size (medium and small cities; large cities) and frequency (≤ 300 MHz or >300 MHz).

In different scenarios, the Hata model changes the path loss model by adding a correction factor, and the expression after adding the correction factor is shown in Eq. (9).

$$PL_{\text{Hata}} = \Delta_1 + 74.52 + 26.16\lg f_c - 13.82\lg h_t - 3.2(\lg(11.75h_r))^2 + (\Delta_2 + 44.9 - 6.55\lg h_t)\lg d. \tag{9}$$

One part of the correction factors Δ_1 and Δ_2 in the Hata model is linear with the logarithm of the height of the transmitting antenna, and the rest are the constants in Table 1. The correction factors are given by:

$$\begin{cases}\Delta_1 = p_1\log_{10}(h_t) + q_1\\ \Delta_2 = p_2\log_{10}(h_t) + q_2,\end{cases} \tag{10}$$

where p_1, p_2, q_1 and q_2 are the undetermined coefficients.

The correction factors Δ_1 and Δ_2 are fitted to the height of the base station antenna based on the calculation results. Taking the results of multiple tests as statistical data, we perform a least-square (LS) fitting model based on the path loss test data of nine cells. The fitting results are shown in Table 2.

▼Table 1. Correction factors of Hata model in different scenarios

Scenario	Correction Factor Δ_1	Correction Factor Δ_2
Urban area	−20.47	−1.82
Suburbs	$5.74\lg h_t - 30.42$	−6.72
Rural	$6.43\lg h_t - 30.44$	−6.71
Viaduct	−21.42	−9.62
Cutting	−18.78	$51.34\lg h_t - 78.99$
Station	$34.29\lg h_t - 70.75$	−8.86

According to the results of the nine cells in the marshalling yard scenario, we fit the optimal regression model under each cell through linear regression, including the corresponding constant term and path loss exponent n, and then calculate the corresponding correction factor for each cell. By conducting joint analysis of the height information of base stations in the corresponding cell, it is determined whether there is a significant linear relationship between the correction factors (Δ_1 and Δ_2) and the height of the base station antenna. The fitting results are shown in Figs. 5 and 6. The correction factors are given by:

$$\Delta_1 = -141.73, \tag{11}$$

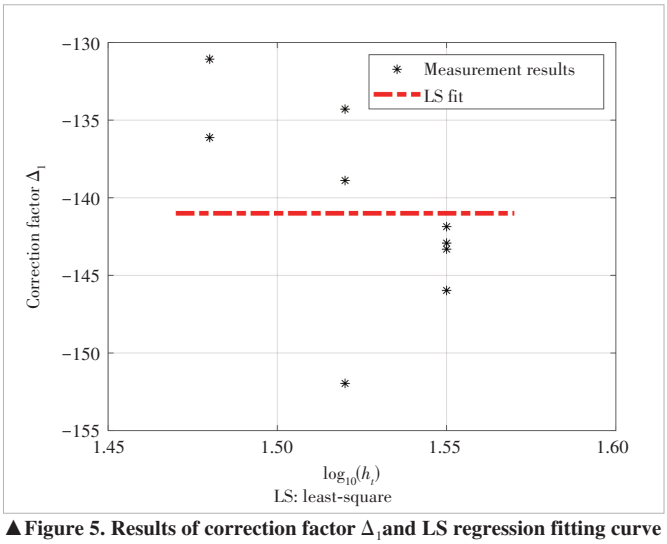
$$\Delta_2 = 17.64\lg h_t - 17.22. \tag{12}$$

5 Validation and Evaluation

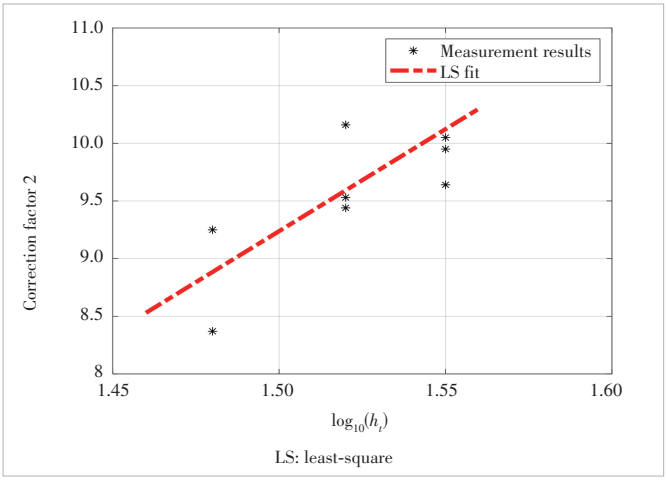
The empirical model proposed in this paper is compared with several traditional empirical models that are improved Hata models as well. The traditional empirical models including the Stanford University temporary model (SUI) and the Hata-Okumura extended model are selected as the reference model for comparison and verification. The validation data consist of three cells, and the validation data have the same measurement system as the statistical modeling data. The validation data are not used in statistical modeling, so they can be used to verify the accuracy and generalizability of the model.

▼Table 2. Fitting measurement results

Cell Number	Path Loss PL_0	Path-Loss Exponent n	Correction Factor Δ_1	Correction Factor Δ_2
62	4.565 0	4.513 81	−134.29	10.16
63	7.949 0	4.347 35	−131.07	9.25
64	3.651 2	4.358 74	−136.12	8.37
143	−3.014 8	4.542 59	−141.86	10.64
150	−7.115 8	4.472 55	−145.97	9.95
151	−4.066 2	4.482 93	−142.92	10.05
156	−4.460 3	4.442 59	−143.31	9.64
184	−13.105 2	4.438 65	−151.96	9.44
185	0.043 9	4.447 31	−138.89	9.53

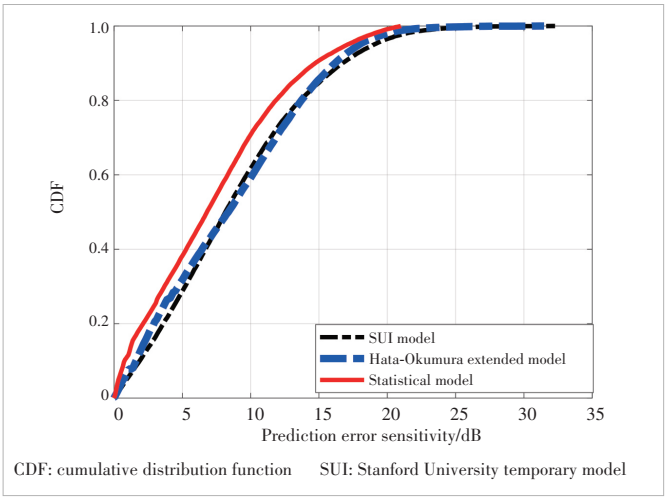


▲Figure 5. Results of correction factor Δ_1 and LS regression fitting curve



▲Figure 6. Results of correction factor Δ_2 and LS regression fitting curve

The SUI model, Hata-Okumura extended model and statistical model are represented by Model 1, Model 2 and Model 3, respectively. Fig. 7 shows a comparison of the accuracy of different models on the validation data at different prediction error sensitivities. When the maximum prediction error sensitivity is required to be 10 dB, the maximum prediction accuracy of the SUI model is 61.66%, that of the Hata-Okumura model is 59.55%, and that of the statistical model is 59.55% with an accuracy of 71.06%. In this scenario, compared with the existing SUI and Hata-Okumura models, the accuracy of the self-built statistical model is improved by about 11.06%. However, the prediction accuracies of the SUI model, the Hata-Okumura extended model and the self-built model in the marshalling yard scenario are not much different, indicating that the SUI model and the Hata-Okumura extended model have



▲Figure 7. Accuracy comparison of different models based on validation data

certain applicable values in the marshalling yard scenario.

6 Challenges

In the statistical model, the path loss will increase with the growing of the distance between transceiver antennas as a whole. There is no need to consider detailed environmental information when the large-scale fading empirical model is applied to the scenario. In specific environments, such as railway marshalling yards, the transmission of radio waves may encounter fading and other conditions, resulting in changes in the final received power to reach the user, making it difficult to determine the specific value of the calculated path loss. This is also a deficiency of large-scale fading empirical models. In the future, it is hoped to re-validate predictions with the help of deterministic modeling, machine learning and other methods to achieve higher prediction accuracy.

7 Conclusions

In this paper, the calculation method of the propagation demarcation point for marshalling yard scenarios is proposed and verified to improve the accuracy of the subsequent empirical model. Based on numerous measured data, a large-scale path loss statistical empirical model for marshalling yard scenarios is established, and the propagation boundary point is regarded as the model segmentation point. The path loss in Area A conforms to the two-path model, and that in Area B is close to the logarithmic model. According to the measurement data in different cells, the correction factor in the marshalling yard scenario is fitted with the Hata model as the benchmark. Finally, the shortcomings and improvements that need to be made to the statistical model are discussed.

References

[1] AI B, MOLISCH A F, RUPP M, et al. 5G key technologies for smart railways [J]. Proceedings of the IEEE, 2020, 108(6): 856 – 893. DOI: 10.1109/JPROC.2020.2988595
[2] ZHONG Z D, GUAN K, CHEN W. Challenges and perspective of new generation of railway mobile communications [J]. ZTE Technology Journal, 2021, 27(4): 44–50. DOI:10.12142/ZTETJ.202104009
[3] AI B, GUAN K, RUPP M, et al. Future railway services-oriented mobile communications network [J]. IEEE communications magazine, 2015, 53(10): 78 – 85. DOI: 10.1109/MCOM.2015.7295467
[4] ZHAO Y J, ZHANG J Y, AI B. Applications of reconfigurable intelligent surface in smart High speed train communications [J]. ZTE Technology Journal, 2021, 27(4): 36-43. DOI: 10.12142/ZTETJ.202104008
[5] YANG M, HE R S, AI B, et al. Measurement-based channel characterization for 5G wireless communications on campus scenario [J]. ZTE communications, 2017, 15(1): 8 – 13. DOI: 10.3969/j.issn.1673-5188.2017.01.002
[6] LU J H, ZHU G, AI B. Radio propagation measurements and modeling in railway viaduct area [C]//6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM). IEEE, 2010: 1 – 5. DOI: 10.1109/WICOM.2010.5600926
[7] WEI H, ZHONG Z D, GUAN K, et al. Path loss models in viaduct and plain scenarios of the High-speed Railway [C]//5th International ICST Conference on Communications and Networking in China. IEEE, 2010: 1 – 5. DOI: 10.4108/iwoncmm.2010.3
[8] HE R S, ZHONG Z D, AI B. Path loss measurements and analysis for high-speed railway viaduct scene [C]//6th International Wireless Communications and Mobile Computing Conference. New York: ACM, 2010: 266 – 270. DOI: 10.1145/1815396.1815458
[9] NEWHALL W G, SALDANHA K J, RAPPAPORT T S. Propagation time delay spread measurements at 915 MHz in a large train yard [C]//IEEE Vehicular Technology Conference. IEEE, 2002: 864 – 868. DOI: 10.1109/VETEC.1996.501434
[10] KNORZER S, BALDAUF M A, FUGEN T, et al. Channel modelling for an OFDM train communications system including different antenna types [C]//64th Vehicular Technology Conference. IEEE, 2006: 1 – 5. DOI: 10.1109/VTCF.2006.55
[11] KNORZER S, BALDAUF M A, FUGEN T, et al. Channel analysis for an OFDM-MISO train communications system using different antennas [C]//66th Vehicular Technology Conference. IEEE, 2007: 809 – 813. DOI: 10.1109/VETECF.2007.178
[12] BRISO-RODRÍGUEZ C, FRATILESCU P, XU Y Y. Path loss modeling for train-to-train communications in subway tunnels at 900/2400 MHz [J]. IEEE antennas and wireless propagation letters, 2019, 18(6): 1164 – 1168. DOI: 10.1109/LAWP.2019.2911406
[13] TANG P. Channel characteristics for 5G systems in urban rail viaduct based on ray-tracing [C]//4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI). IEEE, 2022: 24 – 28. DOI: 10.1109/ISRITI54043.2021.9702771
[14] MOLISCH A F. Wireless communications [M]. Hoboken, USA: John Wiley & Sons, 2005
[15] WEN Y H, MA Y S, ZHANG X Y, et al. Channel fading statistics in high-speed mobile environment [C]//IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications (APWC). IEEE, 2012: 1209 – 1212. DOI: 10.1109/APWC.2012.6324963

Biographies

DING Jianwen received his MS and PhD degrees from Beijing Jiaotong University, China in 2005 and 2019, respectively. He is currently a professor of National Research Center of Railway Safety Assessment, Beijing Jiaotong University. He received the first prize of progress in science and technology of the Chinese Railway Society. His research interests are broadband mobile communications, dedicated mobile communication system for railway, and safety communication technology for train control system.

LIU YAO (21120095@bjtu.edu.cn) received her BE degree in communication engineering from Lanzhou Jiaotong University, 2021, China. She is currently working toward a master's degree with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, China. Her research interests include wireless communications and 5G-R.

LIAO Hongjian received his BE degree in communication engineering from Beijing Jiaotong University, China in 2020. He is currently working toward a master's degree with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, China. His research interests include wireless communications and 5G-R.

SUN Bin received his BS degree in electronic engineering and MS degree in electronic engineering from Beijing Jiaotong University, China in 2004 and 2007, respectively. From 2007 to 2015, he was a R&D manager with BeiJing LiuJie Technology Co., Ltd. He is currently an assistant researcher with National Research Center of Railway Safety Assessment, Beijing Jiaotong University. His main research interest is the interconnection and interworking of core network for dedicated railway mobile communication system.

WANG Wei is the LTE-R technical director and a railway wireless communication system expert of ZTE Corporation, with rich experience of the GSM-R system design. He has a deep understanding of GSM-R and LTE-R and has undertaken several major railway-related projects on wireless communication systems.