

Please cite the Published Version

Das, SK, Mudi, R, Rahman, MS , Rabie, KM and Li, X (2024) Federated Reinforcement Learning for Wireless Networks: Fundamentals, Challenges and Future Research Trends. IEEE Open Journal of Vehicular Technology, 5. pp. 1400-1440.

DOI: https://doi.org/10.1109/OJVT.2024.3466858

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Version: Published Version

Downloaded from: https://e-space.mmu.ac.uk/636820/

Usage rights: Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Additional Information: This is an open access article which first appeared in IEEE Open Journal of Vehicular Technology

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines)



Received 25 August 2024; accepted 14 September 2024. Date of publication 24 September 2024; date of current version 17 October 2024. The review of this article was coordinated by Editor Jianbo Du.

Digital Object Identifier 10.1109/OJVT.2024.3466858

Federated Reinforcement Learning for Wireless Networks: Fundamentals, Challenges and Future Research Trends

SREE KRISHNA DAS [©]¹ (Student Member, IEEE), RATNA MUDI [©]², MD. SIDDIKUR RAHMAN [©]³ (Student Member, IEEE), KHALED M. RABIE ^{©4,5} (Senior Member, IEEE), AND XINGWANG LI ^{©6,7} (Senior Member, IEEE)

¹Department of Electrical, Electronic and Communication Engineering, Military Institute of Science and Technology, Dhaka 1216, Bangladesh ²Department of Computer Science and Engineering, Jahangirnagar University, Savar Dhaka 1342, Bangladesh ³Department of Electrical and Electronics Engineering, Universiti Teknologi PETRONAS, Seri Iskandar, Perak 32610, Malaysia

⁴Department of Computer Engineering and Center for Communication Systems and Sensing, King Fahd University of Petroleum and Minerals, Dhahran 31261,

Saudi Arabia

⁵Department of Electrical and Electronic Engineering Science, University of Johannesburg, Johannesburg 2092, South Africa ⁶School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454099, China ⁷National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China

CORRESPONDING AUTHOR: KHALED M. RABIE (e-mail: k.rabie@kfupm.edu.sa).

The increasing popularity of Internet of Things (IoT)-based wireless services highlights the ABSTRACT urgent need to upgrade fifth-generation (5G) wireless networks and beyond to accommodate these services. Although 5G networks currently support a variety of wireless services, they might not fully meet the high computational and communication resource demands of new applications. Issues such as latency, energy consumption, network congestion, signaling overhead, and potential privacy breaches contribute to this limitation. Machine learning (ML) frequently offers solutions to these problems. As a result, sixth-generation (6G) wireless technologies are being developed to address the deficiencies of 5G networks. Traditional ML methods are generally centralized. However, the vast amount of wireless data generated, growing privacy concerns, and the increasing computational capabilities of edge devices have led to a shift towards optimizing system performance in a distributed manner. This paper provides a thorough analysis of distributed learning techniques, including federated learning (FL), multi-agent reinforcement learning (MARL), and the multiagent federated reinforcement learning (FRL) framework. It explains how these techniques can be effectively and efficiently implemented in wireless networks. These methods offer potential solutions to the challenges faced by current wireless networks, promising to create a more robust, capable, and versatile network that meets the growing demands of IoT and other emerging applications. Implementing the FRL framework can significantly improve the learning efficiency of wireless networks. To tackle the challenges posed by rapidly changing radio channels, we propose a robust FRL framework that enables local users to perform distributed power allocation, bandwidth allocation, interference mitigation, and communication mode selection. Finally, the paper outlines several future research directions aimed at effectively integrating the FRL framework into wireless networks.

INDEX TERMS Federated reinforcement learning, power allocation, bandwidth allocation, interference mitigation, communication mode selection.

I. INTRODUCTION

The growing number of connected user equipment (UE), including industrial machines, Internet of Things (IoT) devices, and smartphones and, is causing bottlenecks in the

inadequate radio resources of cellular networks. Consequently, there is a constant essential to improve the current wireless network structure to fulfill various needs. Fifthgeneration (5G) networks are seen as the cornerstone of future Internet of Everything (IoE)-based wireless services, supporting four main application categories: ultra-reliable and low-latency communications (URLLC), massive machinetype communications (mMTC), virtual reality (VR), and enhanced mobile broadband (eMBB) [1]. Since 2020, 5G networks have been partially deployed in some countries. While 5G is a significant advancement toward a fully connected society, it is acknowledged that it alone is not sufficient to achieve this alteration [2]. Significant enhancements are necessary to manage forthcoming heterogeneous networks and address new trends in user and application demands, such as increased authenticity and improved quality video streaming.

Huawei has introduced several additional wireless applications within the realm of 5.5G networks. These frameworks encompass machine vision (MV), augmented reality (AR), extended reality (XR), high-definition video uploading, real-time broadband communication (RTBC), vehicleto-everything (V2X), harmonized communication and sensing (HCS), and uplink-centric broadband communication (UCBC) [1]. Consequently, both academia and industry have initiated discussions on a new standard, termed 6G, to delineate the essential requirements, needs, and potential use cases for 6G networks [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20].

Anticipated for 6G networks are three innovative wireless applications and services: Contextually Agile eMBB Communications (CAeC), Event-Defined URLLC (EDuRLLC), and Computation-Oriented Communications (COC). Given the intricate nature and diverse dimensions of 6G systems, tackling these challenges will heavily rely on machine learning (ML) and resource optimization strategies in wireless communication. As advancements progress in radio data collection, learning models and methods, and software and hardware platforms, the adoption of federated learning (FL) algorithms will be crucial for advancing cellular technologies. These technologies will expedite the development, calibration, and deployment of 6G networks, addressing issues such as latency, energy consumption, network congestion, and privacy concerns. Moreover, FL will augment digital transformation and efficiency gains across various industries.

Specifically, FL is a distributed learning paradigm that can be integrated with the multi-agent reinforcement learning (MARL) algorithm. In this review article, this integration is referred to as the multi-agent federated reinforcement learning (FRL) framework. The integration of FRL into mobile edge computing (MEC) is expected to bring about genuine intelligence in intricate wireless environments, thereby unlocking the complete capabilities of FRL across diverse intelligent 6G wireless applications. This amalgamation aims to improve both Quality of Experience (QoE) and Quality of Service (QoS), catering to the extensive intelligence requirements of forthcoming societies. In this review article, we discuss the FRL framework for optimized model design in wireless networks, focusing on power allocation (PA), bandwidth allocation (BA), interference mitigation (IM), and communication mode selection mechanism (CMSM).

A. RELATED WORKS

To achieve the deployment of 6G technology by 2030, a multidisciplinary approach and numerous disruptive wireless technologies are necessary [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45]. These include radio resource technologies [46], electronic circuit technologies [47], massive random-access [48], finite blocklength (FBL) regimes [49], non-orthogonal multiple access (NOMA) with asynchronized transmission [50], unsourced random access [51], edge computing [2], wireless sensing [12], and AI functionalities [1].

In [48], the authors addressed the challenge of providing multiple access (MAC) to a large number of uncoordinated users. They also explore the asymptotic coding problem for a K-user Gaussian MAC, where K is proportional to blocklength, with each user having a fixed payload. They identify an interesting tradeoff between energy-per-bit and spectral efficiency in this context. In [49], the authors explored the maximum channel coding rate achievable at a given blocklength and error probability, presenting new, tighter achievability and converse bounds for a broad range of parameters. These bounds provide close approximations of the maximum achievable rate for given blocklengths. In [50], the authors investigated the achievable rate for narrowband uplink NOMA with asynchronized transmission systems, where each user experiences random link delays. By leveraging the bandlimited property of asynchronized NOMA signals, the study examines the upper and lower bounds of the achievable rates under these conditions. In [51], the authors examined the problem of user activity detection (AD) and large-scale fading coefficient (LSFC) estimation in random access wireless uplink systems with a massive MIMO base station (BS) using an iterative component-wise minimization, resulting in a scheme with complexity comparable to nonnegative least squares (NNLS) and an adapted version of multiple measurement vector-approximate message passing (MMV-AMP) algorithms. While traditional methods exist to manage these advanced communication scenarios, they often struggle with scalability, efficiency, and complexity. ML can offer improvements by predicting traffic patterns, optimizing resource allocation (RA), and managing interference, but without it, these systems tend to be less adaptable and more challenging to optimize in real-time.

ML particularly artificial neural networks (ANN), plays a crucial role in building and optimizing future cellular networks across the physical, medium access control, and application layers [1]. Deep learning (DL) plays a pivotal role in advancing beyond 5G (B5G) air interfaces by optimizing the smart radio environment (SRE), enhancing source-channel coding, improving semantic communication (SC), and holistically supporting URLLC wireless networks and services [22], [23], [24], [25]. However, traditional DL methods are generally static and require substantial computational and communication resources, especially for large-scale ML models, highlighting the need for innovative solutions to better address these challenges.

In [14], the authors critically assessed recent literature on FL, focusing on studies related to IoT applications. They evaluated network performance using key metrics such as scalability, quantization, robustness, sparsification, security, and privacy. In [32], the authors explored a novel concept within vehicular networks called a federated vehicular network (FVN), characterized as a resilient distributed vehicular network. To support transactions and deter malicious activities, they incorporated auxiliary blockchain-based systems and identified open problems and future research directions for this disruptive technology. In [33], the authors reviewed current research, technical challenges, potential solutions, and unresolved questions related to deploying FL in vehicular IoT. They outlined future research avenues for combining FL with vehicular IoT, focusing on both using FL to improve vehicular IoT and advancing vehicular IoT technologies to better support FL. In [34], the authors presented a tutorial on FL and a comprehensive survey on implementation issues. They provided detailed reviews, analyses, and comparisons of approaches for emerging challenges in FL implementation, including communication cost, RA, data privacy, and data security. Additionally, in [35], the authors conducted a comprehensive study on the efficient and effective deployment of distributed learning over wireless edge networks. They presented an overview of several emerging distributed learning paradigms, including FL, federated distillation, distributed inference, and MARL, and offered a holistic set of guidelines for deploying a broad range of distributed learning frameworks over real-world wireless communication networks.

In [36], the authors introduced a novel FL algorithm that extends the federated averaging (FedAvg) approach by incorporating a weight-based proximal term into each local loss function. This modification addresses challenges posed by non-independent and identically distributed (non-IID) data, data imbalance, and heterogeneity among UEs, significantly reducing training time and energy consumption compared to traditional FL methods, such as those involving full user participation and equal BA. In [37], the authors proposed an AI-enabled architecture for 6G networks designed to facilitate knowledge discovery, smart resource management, automatic network adjustments, and intelligent service provisioning. They also highlighted important future research directions and potential solutions for AI-enabled 6G networks, including computation efficiency, algorithm robustness, hardware development, and energy management. In [38], a comprehensive survey was provided on communication-efficient techniques in FL, covering wireless communications for FL and FL applications within wireless communication settings. They also discussed open problems for FL and provided future directions helpful for researchers working at the intersection of the two emerging paradigms, FL and next-generation wireless communications. In [39], the discussion focused on the motivations for employing FL in the operation, design, and optimization of FL-based wireless networks. They identified TABLE 1 List of the Main Acronyms and Their Definitions

AI	Artificial intelligence
ANN	Artificial neural networks
AR	Augmented reality
BA	Bandwidth allocation
B5G	Beyond 5G
CAeC	Contextually agile eMBB communications
CE	Communication efficiency
CD	Client drift
CMSM	Communication mode selection mechanism
CNN	Convolutional neural networks
COC	Computation-oriented communications
DL	Deep learning
DRL	Deep reinforcement learning
EDuRLLC	Event-defined URLLC
EE	Energy efficiency
eMBB	Enhanced mobile broadband
FedAvg	Federated averaging
FL	Federated learning
FRL	Federated reinforcement learning
FBL	Finite blocklength
GD	Gradient descent
HCS	Harmonized communication and sensing
ІоТ	Internet of things
IoE	Internet of everything
IM	Interference mitigation
MAC	Multiple access
mMTC	Massive machine-type communications
MARL	Multi-agent reinforcement learning
MEC	Mobile edge computing
ML	Machine learning
M2M	Machine-to-machine
NOMA	Non-orthogonal multiple access
NS	Network slicing
OWC	Optical wireless communication
PA	Power allocation
QoS	Quality of service
QoE	Quality of experience
RISs	Reconfigurable intelligent surfaces
RA	Resource allocation
RL	Reinforcement learning
RNN	Recurrent neural networks
RTBC	Real-time broadband communication
SC	Semantic communications
SGD	Stochastic gradient descent
SRE	Smart radio environment
SE	Spectrum efficiency
URLLC	Ultra-reliable and low-latency communications
UCBC	Uplink-centric broadband communication
VR	Virtual reality
V2X	Vehicle-to-everything
VLC	Visible light communication
XR	Extended reality
XL-MIMO	Extremely large-scale MIMO
56	Fifth generation
	Sixth generation
00	Sixui generation

techniques required to meet the challenges of using FL in practical wireless communication situations. Finally, in [40], the authors outlined the benefits of using FL in IoT environments and explored several significant applications. They highlighted key research challenges that need to be addressed to advance the development of FL in the IoT domain.

In [41], the authors introduced a novel scheduling policy and PA strategy for NOMA settings, aiming to maximize the weighted sum data rate under realistic constraints throughout the learning process. This strategy enhances FL testing accuracy in NOMA-based wireless networks, outperforming existing schemes within equivalent learning durations.

Ref.	Key contributions	FL	MARL	FRL	PA	BA	IM	CMSM
[2]	Describes the FL architecture, categories, operation, and optimization schemes for 6G communications	\checkmark	×	×	×	\checkmark	×	×
[13]	Explores the challenges of statistical analysis and communication efficiency (CE) for FL-based systems	\checkmark	×	×	×	\checkmark	×	×
[14]	Discusses the scalability, quantization, robustness, sparsification, secu- rity, and privacy challenges of FL towards enabling FL-powered IoT applications	\checkmark	×	×	×	\checkmark	\checkmark	×
[19]	Presents the application of FL-enabled visible light communication (VLC) in 6G networks, envisioned as an integrated communication and computing platform	\checkmark	×	×	×	×	×	×
[21]	Discusses the FL concept, technologies, and learning approaches for wireless communications	\checkmark	×	×	×	\checkmark	×	×
[31]	31] Presents the deep reinforcement learning (DRL)-aided CMSM to enhance coverage and improve QoS for fog radio access networks		\checkmark	×	\checkmark	\checkmark	×	\checkmark
[34]	Analyzes the communication costs, privacy, and security of FL in MEC networks	\checkmark	×	×	×	\checkmark	×	×
[35]	Presents the FL, its characteristics, challenges, and forthcoming direc- tions in wireless networks	\checkmark	×	×	\checkmark	\checkmark	×	×
[40]	Discusses some of the unprecedented challenges across systems, se- curity, and practical implementation issues in real-world deployments, which act as the main limitations to allowing FL for the IoT domain and the prospects to tackle these challenges	\checkmark	×	×	×	×	×	×
[41]	1] Investigates the fully automated PA using FL schemes in NOMA networks		×	×	\checkmark	×	×	×
[42]	2] Analyzes the FL approach-based BA for wireless communications		X	Х	×	\checkmark	×	×
[43]] Presents the over-the-air FL-aided IM for multi-cell wireless networks		X	X	X	X	\checkmark	X
[44]	Analyzes FL and its enabling real-life applications, protocols, and use- cases	$\overline{\mathbf{v}}$	×	×	×	×	×	×
This article	Provides an extensive review of the FRL framework for future wireless networks, focusing on PA, BA, IM, and CMSM	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark

TABLE 2 Comparison of This Article With Selected Recent Works on Existing FL and MARL Techniques

In [42], the authors proposed a FL-based RA strategy for wireless communication networks, enabling users to cooperatively train an RA policy in a distributed manner. This approach allows traditional DL-based RA methods to apply and adapt their policies in distributed scenarios and time-varying environments without needing a computationally intensive server. In [43], the authors proposed a cooperative multi-cell FL optimization framework to effectively manage interference in both downlink and uplink transmissions. The algorithm shows significantly improved average learning performance across multiple cells compared to non-cooperative baseline approaches. The authors in [44] introduced a method aimed at handling stochastic radio channels to optimize joint resource blocks (RBs) management and PA in real-time IoT applications. They investigated the complexities and benefits of FL and provided specific service use-cases to demonstrate how various architectures and protocols leveraging FL can be integrated to achieve desired outcomes. In [45], the authors addressed a problem aiming to minimize the combined weighted sum of system and learning costs through the joint optimization of bandwidth, computation frequency, transmission PA, and subcarrier assignment. The proposed algorithm shows superior performance compared to benchmark schemes.

B. NOTATIONS, CONTRIBUTIONS, AND ORGANIZATIONS

Table 1 presents the main acronyms and their definitions. The main contributions of these works on existing FL and

MARL techniques are summarized in Table 2 for comparison. From an implementation perspective, the wide range of computing and networking resources available on client UEs can cause significant delays in the existing FL training process, known as "stragglers" [52], [53]. This issue is worsened by the uneven distribution of client data sizes, with clients holding larger data volumes typically experiencing increased training latency. Moreover, the existing FL training process often incurs substantial communication costs due to the repeated need for model updates between client UEs and the central server. To address these issues, we present the FRL framework for wireless networks. FRL combines the robust adaptability of DRL for addressing complex challenges in uncertain environments with the collaborative and privacy-preserving characteristics of FL through model aggregation. This offers a ground-breaking approach to enhance the network performance of conventional MARL techniques. However, fine-tuning learning parameters, such as aggregation frequency, and optimizing the architecture are essential to balance network performance, communication costs, and privacy. The main contributions of this review article are as follows:

• We explore various distributed learning paradigms, including FL, MARL, and FRL, and provide a comprehensive analysis of the FRL framework for future wireless networks. This analysis covers elements related to the design of wireless communication systems, evaluation of performance, and the influence of wireless factors on FRL parameters.





- A comprehensive discussion is presented on traditional ML-aided PA, BA, IM, and communication mode selection process techniques in wireless networks. This discussion explores their strengths, weaknesses, and constraints, prompting the adoption of a FRL framework in a decentralized way.
- We present several critical research challenges and propose potential directions for the development of next-generation communication networks. In summary, we provide comprehensive guidelines for implementing FRL frameworks, addressing key issues crucial for fully realizing the potential of intelligent wireless networks.

The structure of this review article is organized as follows (see also Fig. 1). Section II explains the operating principles of FL, covering its architecture and model aggregation process. Section III discusses distributed MARL algorithms. Section IV discusses the FRL framework. Section V provides a concise outline of the design considerations in the FRL framework for wireless communications. Sections VI, VII, and VIII delve into the challenges associated with designing future intelligent wireless networks focused on FRL-based RA, IM, and communication mode selection process, respectively. Lastly, Section IX outlines prospective avenues for future research in wireless networks, followed by Section X which concludes the paper.





(a) Typical architecture of FL algorithm [54].

FIGURE 2. Classical architecture and communication process of the FL scheme.

II. FL FRAMEWORK

This section discusses the operating principles of FL, including its architecture, classification, and model aggregation process.

A. FUNDAMENTALS

Centralizing data from multiple BSs and terminals into a single fusion server for processing faces significant challenges due to extensive storage requirements, high computational complexity, and privacy considerations. FL addresses these challenges by facilitating local model training on distributed datasets without the need to transmit raw data to a central server [54], [55], [56], [57]. FL minimizes the data transmitted to the server by communicating only model updates, thereby reducing the strain on network resources.

B. WORKING PROCEDURE OF FL

In a FL-based wireless network setup, there is typically a central server alongside multiple end devices, depicted in Fig. 2(a). Each end device conducts local model training, and the global model is updated by the aggregator until convergence. As illustrated in Fig. 2(b), the learning process unfolds in three stages:

- Task initialization: Select a subset of IoT devices with updates and good channel states to minimize communication overhead.
- Upload local model: IoT devices update their local models based on the global model and transmit them to the aggregator.
- Download global model: The global model is updated by the aggregator through aggregation of local models, which are subsequently distributed to selected end devices for further learning.

In an on-device FL system, each device stores its own training data, ensuring higher user confidentiality, lower power requirements, and reduced delay. Fig. 2(a) shows a FL system where a BS with I IoT devices conducts distributed ML tasks.

During each learning round, the central server estimates the global model, selects participating devices based on criteria like user mobility and signal coverage, and manages the learning process. The *i*th client receives initial global parameters A_o and uses its dataset \mathcal{D}_i , represented by input-output pairs (x_i^k, y_i^k) , for local model training, determining the loss function gradient using stochastic gradient descent (SGD) at the tth communication round.

C. PROBLEM FORMULATION OF TYPICAL FL

If I client devices participate in model training managed by a parameter server, the standard FL training objective is:

$$\min_{m_i^k} \sum_{i=1}^{l} \frac{c_i}{D_i} \sum_{k \in \mathcal{D}_i} f\left(m_i^k, x_i^k, y_i^k\right),\tag{1}$$

where the primary aim is to estimate the ML model, denoted by $m_i^k \in \mathbb{R}^d$, where the loss function $f(\cdot)$ depends on the input vector x_i^k and output vector y_i^k . Additionally, the scaling parameter c_i helps to adjust the weight of the average loss for the *i*th client.

$$\frac{1}{D_i} \sum_{k \in \mathcal{D}_i} f\left(m_i^k, x_i^k, y_i^k\right),\tag{2}$$

on the total training loss with

$$\sum_{i=1}^{I} c_i = 1.$$
(3)

To train efficient ML models, each IoT device may collect limited data, necessitating additional inputs. FL addresses this challenge by managing extensive training data across multiple edge devices within intelligent wireless systems [58], [59]. While current distributed methods typically assume independent and identically distributed (IID) data among training agents, FL effectively handles non-IID data through advanced selection and utilization techniques. FL also addresses issues of unbalanced data sharing and introduces concepts like personalized FL for non-IID data scenarios. Specifically, Model-Agnostic Meta-Learning (MAML) and Federated Multi-Task Learning (FMTL) algorithms are tailored to tackle the complexities associated with non-IID data distributions.

 FMTL: FMTL involves UEs performing distinct yet interconnected learning tasks, where each UE handles a unique task within a non-IID data distribution. The training objective enabled by FMTL can be formulated as:

$$\min_{M,\Omega} \sum_{i=1}^{I} \sum_{k \in \mathcal{D}_i} f\left(m_i^k, x_i^k, y_i\right) + r(M, \Omega), \qquad (4)$$

where $M = [m_1^k, \ldots, m_I^k]$, function $r(\cdot)$ performs regularization, and Ω represents the correlation of various learning activities of clients. Problem (4) can be divided into sub-problems for distributed solving, as shown using quadratic approximation and dual methods [45]. Each device optimizes its model, updating Ω at the parameter server, resulting in diverse converged models and lower total training loss compared to traditional FL.

 MAML-based FL: The goal of MAML-based FL is to find a global ML model that allows each client to obtain a personalized ML model through gradient descent iterations. The training objective for MAML-enabled FL is represented as:

$$\min_{m_i^k} \sum_{i=1}^{I} \frac{c_i}{D_i} \sum_{k \in \mathcal{D}_i} f\left(m_i^k - \lambda_i^k \nabla f_i^k, x_i^k, y_i^k\right), \quad (5)$$

where, the gradient descent and learning rate for the *i*th client are denoted by ∇f_i^k and λ_i^k , respectively. Each device updates the global model through gradient descent iterations to develop a personalized ML model.

D. MODEL AGGREGATION

We present two primary model aggregation processes, namely, federated averaging (FedAvg) and gradient descent (GD), as explained below.

Federated averaging: Model aggregation refers to the process of combining models from multiple UEs to generate a new model, as elaborated in the following.

The foundational method for model aggregation is the FedAvg approach. In this method, the averaging of weights from local models occurs at the central server, specifically the BS, to update the training of the global model [60], [61]. Here, p_i represents the percentage of the number of data samples at the *i*th UE over the total number of data samples across all UEs, s_i indicates the number of data samples in dataset D_i , A denotes the learning weights, and s is the total number of data samples, computed as $s = \sum_{i=1}^{I} s_i$. The learning aim of FedAvg is provided as follows [60],

$$\min_{A} \mathcal{L}(A_G) = \sum_{i=1}^{I} p_i f_i(A_i), \tag{6}$$

where

$$p_i = \frac{s_i}{s} \text{ and } f_i(A_i) = \frac{1}{s_i} \sum_{k \in \mathcal{D}_i} l_k(A_i).$$
(7)

In (7), $l_k(A_i)$ signifies the loss of the FL model linked to the *k*th data sample, calculated based on the local model weights A_i of the *i*th UE, with A_G indicating the weight of the global model.

Two methods are employed for updating the training of the global model. The initial approach involves computing the gradient for each UE, followed by the BS aggregating these gradients from the *i*th UE to update the global model,

$$A_{G}^{t+1} = A_{G}^{t} - \eta \sum_{i=1}^{l} p_{i} g_{i}^{t}, \text{ with } g_{i}^{t} = \nabla f_{i} \left(A_{i}^{t} \right), \qquad (8)$$

wherein $\nabla f_i(A_i^t)$ represents the gradient computed at the *i*th UE and η is the learning rate. The alternative method involves updating the weights of the local model at each UE using this gradient,

$$A_i^{t+1} = A_G^t - \eta g_i^t, \tag{9}$$

wherein g_i^t denotes the gradient computed employing (8). Subsequently, the global model at the BS undergoes an update as follows,

$$A_G^{t+1} = \sum_{i=1}^{I} p_i A_i^{t+1}.$$
 (10)

In the alternate method, each UE initiates the GD process for its local model using local datasets, and the BS then averages these local models. This allows each UE to iterate the local update multiple times according to (10) before uploading the local models, thereby accelerating convergence speed.

Despite the significant success of FedAvg, one of the most renowned methods in FL, challenges due to statistical heterogeneity in the data remain. Specifically, the training data exhibit non-IID characteristics, which negatively impact convergence behavior.

GD: Conventional federated optimization techniques, such as FedAvg [60], may demonstrate suboptimal convergence performance, especially in heterogeneous wireless networks. This is primarily due to two factors: 1) client drift (CD), where local models diverge from the optimal global model, causing unstable and slow convergence; and 2) lack of adaptivity, where FedAvg may not be suitable for large datasets with heavy-tailed stochastic gradient noise distributions, a common issue in natural language processing research [62]. Heavy-tailed distributions are probability distributions with tails that are heavier than those of the exponential distribution [63].

Several innovative GD methods have been proposed to address the challenges of CD and lack of adaptivity, as described below:

1) CD: Addressing the issue of CD, in [64], the authors introduced a novel stochastic controlled averaging (SCAFFOLD) method. This method incorporates control variates for the *i*th UE (e_i) and the variate for the server $e_G = \frac{1}{I} \sum_{k=1}^{I} e_k$ in the GD process to update the training of local and global models, respectively. In contrast to FedAvg, the GD of the *i*th UE in the SCAFFOLD method is expressed as follows,

$$A_{i}^{t+1} = A_{i}^{t} - \eta \left(g_{i}^{t} + e_{G}^{t} - e_{i}^{t} \right), \qquad (11)$$

wherein $e_G^t - e_i^t$ ensures that the GD moves in the correct direction, and e_i^{t+1} is computed utilizing,

$$e_i^{t+1} = e_i^t - e_G^t + \frac{1}{N_i \eta} \left(A_G^t - A_i^t \right), \qquad (12)$$

where N_i denotes the number of iterations for updating the *i*th UE with its local information in the *t*th time slot. In (12), the SCAFFOLD method employs gradients computed in previous steps to modernize the control variate. Following this, the control variate of global model e_G is aggregated as follows,

$$e_G^{t+1} = e_G^t + \frac{1}{I} \sum_{k=1}^{I} \left(e_k^{t+1} - e_k^t \right).$$
(13)

The corrective term $(e_G^t - e_i^t)$ in (11) guarantees that the updates to the local model move in the optimal direction, effectively addressing the CD problem observed in FedAvg.

2) Adaptivity: The adaptive learning method includes adjustable training parameters, such as the learning rate, which can automatically adapt to the statistics of obtained data, available computational radio resources, or other relevant information in its operating environment. Incorporating adaptive variants helps learning algorithms improve convergence performance and training accuracy [60]. To enhance convergence performance in FedAvg, three approaches have been introduced in [60]: adaptive optimizer, fast-convergent FL, and federated proximal.

III. DISTRIBUTED MARL ALGORITHMS

This section delves into the fundamentals of reinforcement learning (RL) and the operational principles of distributed MARL algorithms, including their classification and challenges.

A. PRELIMINARY ON RL ALGORITHMS

The fundamentals of RL revolve around goal-oriented training and intelligent decision-making. Then, RL utilizes a decisionmaker, commonly known as an agent, which is trained to make beneficial decisions through interactions with its environment. Central to these interactions is a reward system, which reinforces actions that lead to positive outcomes, guiding the agent towards maximizing cumulative rewards over time. The iterative refinement of the agent's approach, referred to as a policy, is crucial in directing actions within the RL framework. The agent begins its interaction with the environment using an initial policy, observes the outcomes of its actions, and adjusts the policy based on feedback. This process of iterative training persists until the agent achieves an optimal policy. In real-world wireless applications, RL is operationalized through its mathematical model known as a Markov Decision Process (MDP), providing a structured framework for modeling and analyzing the entire process of training and decision-making in RL techniques.

Formally, a MDP is denoted as a tuple (S, A, P, R, γ) , wherein *S* and *A* denote for the sets of the agent's states and actions, respectively. The state transition probability set, defined as $P: S \times A \times S \rightarrow [0, 1]$, indicates the likelihood of transitions. On the other hand, $R: S \times A \times S \rightarrow \mathbb{R}$ indicates the set of rewards denoted as $r(s^t, a^t, s^{t+1})$, obtained through the agent from the radio environment. Typically, a finite MDP with *T* time steps is considered, and the interaction within it can be represented as an order $\{s^0, a^0, s^1, a^1, \ldots, s^T, a^T\}$, commonly referred to as an episode.

At each time step *t*, the agent makes a decision to select an action a^t based on the present state s^t . Subsequently, the radio environment transitions to a new state s^{t+1} and issues a reward $r(s^t, a^t, s^{t+1})$ to the agent. It is essential to focus that the agent's objective is to boost the discounted cumulative return $G^t = \sum_{k=0}^{T} \gamma^k r(s^{t+k}, a^{t+k}, s^{t+1+k})$ rather than the instant reward $r(s^t, a^t, s^{t+1})$. The symbol $\gamma \in [0, 1]$ provides as a discount rate, striking a balance amid instant and upcoming rewards.

As emphasized, the goal of RL is to determine the optimal strategy. The agent's strategy is denoted as $\pi(a^t | s^t)$, representing the probability of selecting action a^t given state s^t .

Classical RL approaches can be categorized into two types, such as: value-based RL and policy-based RL, as explained below.

Value-based RL methods calculate the value function, predicting the anticipated cumulative reward for each state-action pair. The objective is to determine the optimal value function, indicating the maximum anticipated cumulative reward for each state-action pair. This optimal value function then guides the optimal strategy, where the agent chooses the action that outcomes in the state having the maximum anticipated value at each stage.

Policy-based RL techniques concentrate on directly pinpointing the optimal strategy without requiring the training of a value function. In these methodologies, the strategy gets parameterized, frequently employing neural networks in intricate radio environments, and the model parameters experience iterative updates to enhance the strategy. The fundamental principle guiding strategy-based methods is known as policy gradient. In straightforward language, policy gradient methods aim to boost the anticipated return by modifying strategy parameters to optimally boost performance.





(a) Typical architecture of MARL algorithm.

(b) Communication process of MARL algorithm.

FIGURE 3. Classical architecture and communication process of the MARL scheme.

B. FUNDAMENTALS OF MARL ALGORITHMS

MARL algorithms are designed to enable multiple agents to learn and make decisions in an environment where the actions of each agent affect not only their own outcomes but also the outcomes of other agents. This field extends the principles of single-agent RL to scenarios involving multiple decision-makers, introducing unique challenges and opportunities.

MARL provides an ideal framework for making sequential decisions in dynamic settings by interactively engaging with the highly dynamic radio environment, as illustrated in Fig. 3(a). In MARL scenarios, each agent independently selects its action based on its own observation of the radio environment, as illustrated in Fig. 3(b). In forthcoming 6G wireless communications networks, an agent's state may encompass transmission PA status, radio channel information, and throughput. Actions taken by the agent can involve radio RBs and power levels. Rewards are typically associated with the learning objectives, such as enhancing the aggregate energy efficiency (EE) or spectrum efficiency (SE), and are defined as the currently attained EE or user rate.

MARL is often designed as a decision-making and training model within a discrete-time stochastic control process, such as the MDP [65]. Traditional RL methods include value-based algorithms (e.g., Q-learning), actor-critic schemes, modelenhanced algorithms, and policy-driven algorithms. In [66], the authors developed an asynchronous algorithm incorporating parallel computing to tackle non-convex problems using RL. However, in modern intelligent IoT applications, such as smart transportation systems and robotics communications, employing the MARL algorithm is essential. This approach involves multiple agents collaborating with a radio environment to achieve a common goal and maximize the shared team's reward across various local action spaces [67]. Given the vast state-action spaces, delayed rewards and feedback, high mobility, and stochastic radio environments that must be coordinated with heterogeneous agents' behaviors, an effective communication policy among multiple agents is crucial for achieving better and stable network performance with the MARL algorithm.

In the server-client enhanced MARL architecture, the proposed scheme manages the training process for all edge agents. In [68], the authors introduced a multi-agent actorcritic method featuring decentralized actors at each agent and a centralized critic for sharing parameters among the edge agents. To improve the CE of the gradient function for the distribution strategy in the MARL scheme, [69] presents a gradient function for a loosely aggregated system that reduces communication rounds by only using informative gradients from selected edge agents and reusing outdated gradients for the remaining agents. For IoT applications without central controllers, such as smart transportation systems, [70] proposes a communication connectivity graph for a decentralized MARL scheme where edge agents only allow information exchange. In [71], the authors introduced a decentralized actor-critic algorithm with function approximation, where each agent makes independent decisions based on local data observation and communication messages shared through a consensus stage over the wireless network. In [72], a decentralized gradient method for an entropy-regularized strategy is proposed, which requires data exchange only with nearby edge agents to train a single strategy for a multi-task RL algorithm involving multiple edge agents operating in different radio environments. Based on these studies, we describe some common MARL algorithms as follows.



- *Independent Q-learning (IQL):* Treats other agents as part of the environment, learning independently.
- *Deep distributed Q-networks (DDQN):* Extends deep Q-network (DQN) to multi-agent settings for handling high-dimensional state and action spaces.
- *Multi-agent deep deterministic policy gradient (MAD-DPG):* An actor-critic approach that learns continuous actions in a cooperative or competitive setting.
- *Counterfactual multi-agent (COMA) policy gradients:* Focuses on the credit assignment problem by using a counterfactual baseline to evaluate an agent's contribution to the collective outcome.

C. CHALLENGES OF MARL FRAMEWORK

Despite substantial progress made in MARL, especially in the domain of distributed learning, there are still numerous challenges that need to be addressed before it can be fully utilized in practical applications.

Non-stationary environment: In multi-agent systems, all agents are learning their strategies simultaneously, and each agent must consider not only its own actions but also those of other agents. This ongoing interaction with other agents constantly changes the environment, making it difficult for agents to identify optimal strategies. When applying MARL in a distributed manner, a popular solution is to use centralized training and distributed execution (CTDE). This technique is particularly effective for optimizing transmissions within multiuser wireless networks, as it offers extra spatial degrees of freedom for signal manipulation.

Partial observation: In real-world scenarios, individual agents often only have access to a subset of the total state data, restricting their capacity to learn the globally optimal strategy. To mitigate this issue in dynamic environments that involve machine-to-machine (M2M) communications, a consensus communication method that utilizes a graph network-based self-attention mechanism can significantly reduce the effects of partial observation on MARL.

Training method: There are various multi-agent approaches that either adopt a fully centralized or fully distributed training method. In the fully centralized scheme, a central unit takes on the responsibility of strategy learning, utilizing data from all agents. However, this method encounters considerable computational complexity. Conversely, fully distributed training methods often struggle with convergence problems due to the absence of comprehensive state data for training. Studies have demonstrated that the CTDE method outperforms both fully centralized and fully distributed training schemes. In CTDE, a centralized network uses global data for centralized training, whereas agents execute the learned strategy in a distributed fashion using their individual local data samples. This strategy effectively counters the challenges posed by non-stationary environments, guarantees convergence, and minimizes training overhead.



FIGURE 4. Typical architecture and communication process of FRL framework.

IV. FRL FRAMEWORK

This section delves into the fundamentals of the FRL framework and the operational principles of extended FRL algorithms, highlighting their advantages and classification.

A. FUNDAMENTALS

FL-enabled systems protect raw data by transmitting model parameters for aggregation, but increasing FL training efficiency and accuracy is challenging. Designing techniques to enhance communication efficiency, reduce delay, and improve accuracy is crucial. Integrating MARL with FL can improve client selection using a two-layer perceptron-based MARL agent at the aggregation server, optimizing global model accuracy and communication delay. MARL agents then learn value decomposition to maximize team rewards, making distributed FRL adaptable to various real-time systems. Unlike traditional RL, FRL ensures fast convergence for large state and action spaces, as shown in Fig. 4(a).

B. WORKFLOW OF FRL FRAMEWORK

The workflow of the FRL framework, illustrated in Fig. 4(b), involves three main components at the central server: the model storage block, the MARL block, and the statistics collection block [53]. The model storage block stores and updates the global deep neural network (DNN) model. The MARL block executes trained MARL agents for client selection. The statistics collection block gathers client information. The following summarizes the functionalities at each stage per iteration:

Stage 1: At each training round, *K* client devices are selected from the client pool.

Stage 2: Selected clients receive the global model parameters.

Stage 3: Clients conduct training and report losses to the MARL block.

Stage 4: Clients also send latency information to the statistics collection block.

Stage 5: The MARL agents receive and combine loss information with previously stored data.

Stage 6: Client selection decisions are generated.

Stage 7: Chosen clients perform local training and send updated parameters to the model storage block, which updates the global model. Latency data is updated in the statistics collection block. Clients report initial training losses to the BS,

which collects loss values and latency information. Clients share model weights in a FL manner to train the DNN model quickly via the DRL controller. In subsequent iterations, stages 6 and 7 transition into stages 1 and 3, respectively.

C. PROBLEM FORMULATION OF FRL FRAMEWORK

Nevertheless, the decentralized arrangement poses a risk to the agent's generalization performance as the diversity of data within isolated multi-user agent systems is restricted. This limitation could potentially lead the agent into a local optimum. In response to this challenge, we introduce a FRL algorithm, leveraging FL to enhance the agent's generalization during training while upholding data privacy.

In FL, two roles are distinguished: the participant and the collaborator [73]. The participant $k, k \in [1, n_n]$, is represented as a DNN model $f_{A_k}^k$. It undergoes local self-training and periodically uploads its parameters A_k to the collaborator. Here, n_n is the number of participants processed concurrently. Due to data privacy constraints, the participant $f_{A_k}^k$ only trains on the local dataset, leading to potential issues with insufficient training due to limited data capacity and diversity. FL addresses this challenge through the following steps. Initially, at training epoch $p, p \in [1, N_p]$, the model of the kth participant is defined as $f_{A_{p}^{p}}^{p}$, undergoing self-training to acquire parameters A_k^p , with N_p representing the total number of training epochs. Subsequently, each participant uploads its parameters to the collaborator, forming a parameter list $A^p =$ $[A_1^p, A_2^p, \ldots, A_{n_n}^p]$. The collaborator calculates the weighted average of A^p to estimate a global model f_G^p with parameters \bar{A}_{G}^{p} . After aggregation, the collaborator broadcasts \bar{A}_{G}^{p} to all participants, replacing their individual parameters, i.e., $\bar{A}_G^p = A_1^{p+1} = A_2^{p+1} \dots = A_{n_n}^{p+1}$. The local loss function and the learning rate of the kth participant are denoted as $F_k(.)$ and η , respectively. The aggregation mechanism of FL is mathematically formulated as follows,

$$A_k^{p+1} = \bar{A}_G^p - \eta \nabla F_k \left(A_k^p \right), \qquad (14)$$

where

$$\bar{A}_{G}^{p} = \sum_{k=1}^{n_{n}} \frac{1}{n_{n}} A_{k}^{p+1}.$$
(15)

In this comprehensive review, one can regard the participant as the agent within each multi-user system, while the collaborator assumes the role of a server responsible for aggregating and broadcasting the parameters. The objective of FRL is to address the distributed optimization model presented below,

$$\min_{\bar{A}_{G}^{p}} F\left(\bar{A}_{G}^{p}\right) = \sum_{k=1}^{n_{n}} p_{k} F_{k}\left(A_{k}^{p}\right), \qquad (16)$$

wherein p_k signifies the relative weight assigned to each multi-user agent in the global model, F(.) denotes the global loss, with $p_k > 0$ and $\sum_{k=1}^{n_n} p_k = 1$. We define $p_k = \frac{|D_k|}{\sum_{k=1}^{n_n} |D_k|}$, where D_k denotes the data size utilized for the local training of the *k*th multi-user agent. It is important to note

that the direct computation of F(.) is not feasible without the exchange of information among participants.

The architecture of extended FRL framework is depicted in Fig. 5(a). At epoch p, the global agent in three multi-user agent is initially replaced by the global agent in the $(p + 1)^{th}$ epoch. Subsequently, the three multi-user agents engage in self-training to acquire parameters, which are then sent to the server for aggregation. Following this, the global agent is constructed on the server, and the parameters are broadcasted to the multi-user agent for the $(p + 1)^{th}$ epoch. The FRL comprises two main components: one executed on the server, serving as the collaborator, and the other executed on multi-user agent, acting as the participant. The procedures carried out on the server and multi-user agent are explained below, respectively.

1) SERVER PART

In the initial stages of FRL training epochs on the server, the primary focus lies in aggregating and broadcasting agent parameters. At the commencement of the FRL training epoch, the server establishes a global agent with the parameter \bar{A}_{G}^{0} , which is subsequently disseminated to each multi-user agent for self-training. As the agents concurrently update their parameters, the server consolidates the parameter list $A^{p} = [A_{1}^{p}, A_{2}^{p}, \ldots, A_{n_{n}}^{p}]$ using (15). Moreover, the aggregated parameters \bar{A}_{G}^{p} are employed to update the parameters of global model and broadcast to the multi-user agents for the training of epoch p + 1.

2) MULTI-USER AGENT PART

The self-training process is embraced by the multi-user agent in the FRL procedure, and it collaborates with the server. Upon receiving the parameter \bar{A}_G^p from the global model at epoch *p*, the parameter of each multi-user agent is substituted by \bar{A}_G^p , denoted as $A_k^p = \bar{A}_G^p$. Subsequently, each multi-user agent undergoes N_m individual self-training epochs concurrently. Following this, the parameters of the multi-user agent at the last self-training epoch, specifically s_{N_m} and u_{N_m} , are stored and transmitted to the server.

Each multi-user agent performs self-training with a famous DRL algorithm, namely, proximal policy optimization (PPO), to obtain the optimal policy e. Here are two types of DNN, namely, actor and critic, defined by the muti-user agent. Actor e^s is parameterized by s, which aims produce the action, and the critic is denoted as v^u , which is parameterized by u.

The training process of the self-training procedure within a single episode is depicted in Fig. 5(b). Initially, the experiment tuples *T* are sampled as,

$$T = \left\{ < s^{o}, a^{o}, r^{o}, s^{1} >, < s^{1}, a^{1}, r^{1}, s^{2} >, \\ \dots, < s^{L}, a^{L}, r^{L}, s^{L+1} > \right\},$$
(17)

where L signifies the length of T. Subsequently, the loss function of the actor at the *j*th episode is computed, defined as





(a) Typical architecture of extended FRL framework.

(b) Communication process of self-training of the multi-user agent.

FIGURE 5. Typical architecture and communication process of extended FRL framework.

follows,

$$\mathcal{L}_{c} = \mathbb{E}^{s, a \sim T} \left[\left(\min \frac{e_{j}^{s}(a|s)}{e_{j-1}^{s}(a|s)} A_{e_{j}^{s}}^{s,a}, \\ \operatorname{clip} \left(\frac{e_{j}^{s}(a|s)}{e_{j-1}^{s}(a|s)}, 1 - \gamma, 1 + \gamma \right) A_{e_{j}^{s}}^{s,a} \right) \right], \quad (18)$$

wherein $\mathbb{E}^{s,a\sim T}[.]$ denotes the empirical average across the sampled experiment tuples $T. e_{j-1}$ and e_j represent the previous and new policies, respectively. γ serves as the clip parameter, and $A_{e_j}^{s,a}$ signifies the advantage, assessing the worthiness of an action by comparing the action value and state value,

$$A_{e_{j}}^{s^{t},a^{t}} = \mathbb{E}\left[L|s^{o}=s^{t}, a^{o}=a^{t}\right] - v\left(s^{t}\right) = Q\left(s^{t}, a^{t}\right) - v\left(s^{t}\right).$$
(19)

Nonetheless, obtaining $A_{e_j}^{s,a}$ directly poses challenges due to the complexity in determining $Q(s^t, a^t)$. Consequently, this study employs the generalized advantage estimation method,

$$A_{e_{j}^{s}}^{s,a} = \xi_{v}^{o} + (\pi\theta)\xi_{v}^{1} + (\pi\theta)^{2}\xi_{v}^{2}, \dots, + (\pi\theta)^{L-t+1}\xi_{v}^{L-t+1},$$
(20)

wherein $\theta \in [0, 1]$ and $\pi \in [0, 1]$ represent the a hyperparameter and discount factor adjusting the tradeoff between bias and variance in the estimation, respectively. It is worth noting that increasing θ raises the variance while decreasing bias. Following Schulman et al.'s recommendation [73], θ is set to 0.95. The calculation of ξ_v^j is expressed as follows,

$$\xi_{v}^{j} = r^{j} + \pi v_{j}^{u} \left(s^{t+1} \right) - v_{j}^{u} \left(s^{t} \right), \qquad (21)$$

wherein $v_j^u(s^{t+1})$ and $v_j^u(s^t)$ are provided through the critic, which is trained using the loss function \mathcal{L}_v ,

$$\mathcal{L}_{v} = \mathbb{E}^{s, a \sim T} \left[\left(\pi v_{j}^{u} \left(s^{t+1} \right) + r \left(s^{t}, a^{t} \right) - v_{j}^{u} \left(s^{t} \right) \right)^{2} \right].$$
(22)

The actor's parameter updates are expressed as follows,

$$s^{j+1} = s^j + \eta_c \nabla_{s_j} \mathcal{L}_c, \tag{23}$$

VOLUME 5, 2024

wherein η_c denotes the learning rate of the actor. The critic's parameter updates are expressed as follows,

$$u^{j+1} = u^j + \eta_v \nabla_{u_j} \mathcal{L}_v, \tag{24}$$

wherein η_v denotes the learning rate of the critic. As both \mathcal{L}_c and \mathcal{L}_v are optimized within each multi-user agent in the extended FRL algorithm, they serve as the local loss functions contributing to the construction of the global loss according to (16).

D. ADVANTAGES

FRL utilizes RL agents to address a problem underneath the guidance of a central processing entity while safeguarding private information. Through the aggregation of local models, FRL effectively reduces the communication burden that conventional DRL methods entail, enhancing privacy. Through tackling problems at the edge of the wireless network, it significantly decreases overall complexity, leading to lower processing delays. As a result, FRL emerges as a promising solution aligned with the expectations of advanced networks like 6G. However, the distributed manner of FRL presents various challenges. Relying on the specific network scenario, it is crucial to explore various implementations of FRL methods to strike the optimal balance between system performance, communication costs, and privacy preservation.

E. CLASSIFICATION

FRL methods can be classified based on their aggregation frequency, specifically as slot-aided aggregation and multi-slotaided aggregation methods [20]. In the slot-aided aggregation approach within the FRL framework, model aggregation takes place following each local model update for every communication round. Although this method yields comparable system performance to conventional MARL, it does incur a substantial communication overhead due to frequent information exchanges. In contrast, when considering communication rounds greater than one, multi-slot-aided aggregation methods in FRL conduct model aggregation after multiple local model



FIGURE 6. FRL framework for 6G networks.

updates have occurred. In addition, multi-slot-aided aggregation methods enhance the independence of UEs, reducing the need for coordination, communication burden, and the threat of information leakage.

F. FRL FRAMEWORK FOR 6G WIRELESS NETWORKS

As a result of the growing size of wireless networks, the increasing density of UE connections, and the presence of system heterogeneity, it has become increasingly difficult, and in certain instances, impractical to model such a dynamic cellular network using traditional ML techniques. Conventional ML-based wireless network optimization relies on the assumption that the convex loss function can be represented in easily manageable geometric shape, enabling an optimizer to evaluate solutions through simple computations [1]. Nevertheless, establishing the connection between a decision and its consequences on the physical wireless system is excessively communication costly and may not be amenable to non-convex problem solutions. Modern advancements in ML technologies, i.e., statistical learning, FL, and MARL algorithms, have the potential to effectively tackle complex network optimization problems in upcoming cellular networks. These DL methodologies have the capacity to iteratively identify asymptotically optimal solutions through the use of SGD techniques. To elaborate further, FRL techniques, which include FL, multi-armed bandit theory, and MARL algorithms, create a feedback loop within the physical system and decision-maker. This arrangement allows the decision-maker to progressively refine its actions based on the feedback it receives from the system, ultimately leading to the attainment of optimality. As demonstrated in Fig. 6, FRL techniques have been extensively applied to address a wide array of emerging challenges in the realm of communication and networking. These challenges include tasks like BA, PA, IM, CMSM, and others.

V. DESIGN ASPECTS OF FRL FRAMEWORK FOR WIRELESS NETWORKS

This section describes a FRL framework for wireless communications, performance evaluation, and the impact of wireless factors on FRL parameters. Below, we provide a brief review of the design aspects of the FRL framework for wireless networks to properly set the stage for our contributions in this paper.

- Choosing and scheduling client: The random clientchoosing technique in FRL frameworks for wireless communications can require additional time and resources, leading to system heterogeneity. Selecting and scheduling clients effectively addresses convergence time, learning accuracy, and client capabilities but increases communication overhead and decreases reliability in spectrum-limited systems [74], [75], [76]. Proper client-choosing and resource management mechanisms are crucial. Maximizing participating devices while balancing accuracy, reliability, and resource requirements is essential, particularly for straggler clients. Including switches in RF client scheduling strategies enhances global model transmission efficiency.
- *Combined learning and communication:* Within a FRL framework designed for wireless systems, local models are transmitted to the central server through RF signals during the uplink phase, while the updated global model is disseminated via radio signals during the downlink phase. Dynamic propagation and wireless impairments can impact model accuracy, which can be mitigated through error detection and correction methods such as longitudinal redundancy check, cyclic redundancy check, and parity checking. Enhancing accuracy and training speed involves adjusting training parameters according to client capabilities and employing model-based optimization schemes that consider system bandwidth, memory constraints, and computational capabilities.
- *Communication efficiency:* In extensive FRL systems, minimal data exchange per transmission can lead to low CE. Researchers have proposed several solutions to address this, including reducing the size of updated models, adjusting transmission categories, and decreasing communication frequency, as detailed in Fig. 7 [77], [78], [79], [80], [81], [82], [83].
- Predicting user mobility: Predicting user mobility is crucial for maintaining high performance in outdoor wireless systems, as learning parameters vary with device movement. Effective wireless RA, position updates, signal transmission, and handover management depend on accurate mobility prediction. Current schemes combine comprehensive data with mobility models and localization data. A large number of FL-aided clients can aid in mobility prediction, ensuring only low-mobility clients are selected for local training to minimize communication errors. Developing robust training strategies



			Communication efficiency	
_		_	¥	 *
	Model updates dimension minimization		Communication frequency	Communication category
)	Quantization: it involves reducing the volume of data transferred from edge servers to the cloud node [77-78].	-	Minimizing the volume of exchanged parameters: it involves local clients downloading and uploading only the escential parts of the model [70,80]	Synchronous: the server completes model aggregation by obtaining local model updates from all participating clients. However, it is
],	Sparsification: It is a technique used in distributed training process to reduce communication costs by sparsifying stochastic gradients, slightly increasing the required number of iterations.		Periodic aggregation: clients perform numerous local interactions before distributing their local updates with the computing server to eliminate the straggler	susceptible to stragglers [82]. Asynchronous: the computing server carries out the model aggregation process, even when alignt model undertage are only particular.
	Hybrid: it is commonly adopted in compression-based techniques, where they incorporate both quantization and sparsification methods that leverage the strengths of both methods.	,	issue in FL systems [81]. Over-the-air computation: aggregation procedure is accomplished over the air via the client in a distributed way that provides low communication delay and saves implementation costs.	received. It addresses the straggler problem by integrating client updates as soon as they are available, instead of waiting to gather updates from all clients and subsequently discarding any delayed updates [83].

FIGURE 7. Various CE increment methods in FRL framework.

that account for asynchronous cooperation and predictive techniques is essential for dynamic wireless networks.

- Trade-off between training accuracy and processing delay: The agents in the FRL framework utilize an optimal approach to strike a balance between two key objectives: achieving optimal local training accuracy and minimizing total processing delay [84], [85]. During the initial stages of FL training, a relatively smaller number of UEs are selected compared to the later stages. This selection takes place both at the beginning and at the end of the training rounds. There is a rationale behind this approach. In the early phases of training, DNNs typically acquire low-complexity functional components before progressing to further advanced features. These early components tend to be more robust in the face of noise and perturbations [84]. Consequently, it becomes possible to achieve effective training with less training data during the early stages, resulting in reduced processing delay and communication costs [85].
- Incentive mechanism: Typically, in distributed learning setups, it is assumed that all users will willingly take part in the global model aggregation process without expecting any compensation in return. Nevertheless, in practical scenarios, participants may hesitate to join this federation process as training MARL models consumes valuable resources [86], [87], [88], [89]. In [90], the authors explored the concept of incentivizing participants in FL by introducing an incentive-compatible scoring system to establish a payment framework. Fig. 8 illustrates the architecture of this incentive mechanism in FL, where users can encompass mobile UEs, edge UEs, IoT UEs in cross-UE FL, or large corporations in cross-silo FL. These users contribute various types of resources, not limited to just data, all of which significantly impact the learning performance. Following global ML model



2 Local model training

FIGURE 8. Typical architecture of incentive mechanism.

aggregation, the server compensates each user based on their individual contributions to the FL process. In [91] and [89], the authors conducted comprehensive surveys on incentive mechanisms in recent FL research. These surveys identified the challenges in designing incentive mechanisms for FL and categorized existing incentive mechanisms into various techniques, including Stackelberg games [92], auctions [93], contract theory [94], [95], Shapley values [96], RL [97], and blockchain [98]. Stackelberg games, auctions, and contract theory are primarily employed to select users and allocate payments to incentivize their participation in the FL process, while the Shapley value is utilized for an equitable evaluation of user contributions to FL. FRL and blockchain are crucial to enhance the performance and resilience of these incentive schemes.

• *Scalability:* The projection is for the number of mobile UEs to reach 10.3 billion and the count of IoT UEs to hit 30.9 billion by 2025 [99], [100], creating a notable scalability challenge. The traditional cloud infrastructure struggles to provide scalability for both data and applications because of the high likelihood of network congestion caused by the data transmission from tens of millions of end UEs. Edge computing offers a solution whereby, if one edge server becomes congested and cannot meet incoming requests, the corresponding service can be seamlessly shifted to another nearby edge server to handle the computational workload.

On the contrary, edge computing shares a similar operational concept with cloud computing but brings computational resources closer to local UEs. Instead of sending resource-intensive computational tasks to a remote cloud, end UEs turn to nearby edge servers for computational resources. Typically, there are several nearby edge servers accessible to each end UE. However, it is important to note that edge servers possess limited power and computational resources compared to the cloud server, which is typically assumed to be highly potent. This introduces additional complexity into the computation offloading problem, as it necessitates considerations regarding edge server selection and resource management [100].

In cloud computing, the central aspect of computation offloading revolves around determining whether to offload, the extent of offloading, and what should be offloaded. In the context of edge computing, in addition to these aspects, we must also address where and how to offload and the allocation of resources. Recent research has delved into the joint problem of computation offloading and resource management with the objective of minimizing energy consumption and processing delay [101], [102]. Researchers have formulated this joint problem as a combinatorial optimization challenge with non-linear constraints and have proposed computation offloading algorithms based on convex optimization [103], [104], Lyapunov optimization [105], [106], and game theory [107], [108]. Furthermore, the design of computation offloading schemes can be likened to the decisionmaking process for offloading and RA within a dynamic environment, a facet explored through the application of FRL methods in numerous research endeavors [109], [110].

• *Privacy:* A key advantage of FRL is its ability to enable UEs to train a learning model without the need to share the raw data of those UEs. However, it is important to note that some level of private information can still be revealed through the analysis of differences between the uploaded models [111], [112], [113], [114], [115], [116], [117], [118]. For instance, in [113], the authors demonstrated that with access only to the trained ML model of a specific hospital's private information, attackers can deduce whether an individual had been a

patient at that hospital. In general, within the FL system, honest but inquisitive coordinators, untrustworthy users, and potential eavesdroppers in the wireless network may exploit the system to glean information about UEs.

To tackle this issue and protect user privacy in the context of 6G networks, it is essential to employ privacypreserving techniques such as differential privacy (DP) and secure multiparty computation (SMC). Table 3 provides an overview of various approaches aimed at safeguarding user privacy within the FRL framework, along with pertinent references.

Security: One of the primary goals of FRL is to ensure privacy. To implement FRL over wireless networks while safeguarding the privacy of each user from both external and internal threats, it is essential to employ a dependable approach [119]. However, it is important to acknowledge that these methods can only offer a limited level of privacy protection. Moreover, FRL remains vulnerable to various security challenges, including poisoning attacks [120], backdoor attacks [121], and channel attacks [122]. Additionally, these privacy approaches primarily focus on safeguarding data. The nature of wireless access mediums is also susceptible to communication-based attacks, such as jamming [123] and Denial-of-Service (DoS) [124]. For a comprehensive exploration of security issues in FRL, we suggest to read in [119].

Remark: Transmitting model parameters imposes greater requirements on reliable, low-latency radio links, as the convergence performance of the ML model is intricately linked to the efficiency of wireless communication. Traditional communication technologies such as PA, IM, CMSM, and BA must be improved with a focus on prioritizing ML model convergence as a key performance metric. From the preceding discussion, it is evident that deploying the FRL framework in real-world scenarios entails considering factors such as over-the-air computation, privacy, scalability, gradient compression, and the allocation of communication and computing resources. This is done to enhance the performance of wireless networks and support distributed FRL effectively in 6G wireless networks.

A. PERFORMANCE EVALUATION OF FRL FRAMEWORK

Executing a FRL framework in a cellular network involves several key steps during each communication round. At the outset, each end UE trains a local model and subsequently uploads the locally computed parameters of the FRL model. The central server then aggregates these local models to generate and broadcast the global model. Performance of FRL frameworks is indicated by training loss, convergence time, latency, power consumption, and reliability. These factors are crucial for evaluating the efficiency and effectiveness of FRL systems in wireless environments.



TABLE 3 Methods to Address the Privacy Challenges Within the FRL Framework

Ref.	Method	Main concept	Limitation
[111]	Generative privacy	Use generative adversarial network (GAN) for the generation of synthetic data	Challenges related to convergence
		at the local level to substitute real data	and model accuracy
[112], [113]	DP	Inject Gaussian noise into the locally modernized models as a means to attain	Reduce the training time
		DP	
[114]	DP and SMC	By integrating DP and SMC, it is possible to mitigate the increase in noise	Unable to handle user dropouts
		injection while still preserving a predefined level of trust	
[115], [116]	DP	Leverage the channel noise through an analog approach to ensure DP	Reduce the training time
[117]	Reputation-based	Design a metric for reputation to selectively choose dependable and trustwor-	Challenges related to convergence
	user selection	thy UEs	



FIGURE 9. Time performance in the FRL framework, where *M* represents the total number of UEs.

- *Training loss:* The loss function measures training loss in a FRL framework, which is influenced by the clients' ML models. Poor wireless channel conditions can cause errors in transmitted ML models, leading to increased training loss. Additionally, since only selected clients participate due to computational and power constraints, the use of fewer local models in generating the global ML model further increases training loss. This reduction in client participation can compromise the overall model's accuracy and effectiveness.
- *Convergence time:* The convergence time *T* of a FRL framework for a wireless network depends on three factors (τ_i, l_i, t_i) as illustrated in Fig. 9 and can be calculated by

$$T = \max(\tau_1 l_1, \tau_2 l_2, \dots, \tau_M l_M) + \max(t_1, t_2, \dots, t_M),$$
(25)

where τ_i is the training time for updating the local model per iteration at UE *i*, l_i is the number of iterations for convergence at UE *i*, i = 1, 2, ..., M, and *M* denotes the total number of UEs in the learning process. Note that τ_i and l_i are interdependent, meaning that the required number of iterations for convergence can be reduced by increasing the SGD step amounts while updating a local ML model in each learning round. Finally, t_i is the training time for transmitting the ML model in each learning round by UE *i*.

- Latency: Latency in FRL systems can occur during local training, uplink transmission, model aggregation, and downlink transmission, as shown in Fig. 9. To minimize latency and improve performance, combined optimization of both computation and transmission processes should be implemented. This holistic approach ensures that delays in one part of the system do not disproportionately affect the overall performance, leading to a more efficient and responsive FRL framework.
- *Power consumption:* FRL clients use their limited battery power to compute locally and transmit parameters to the central server. This process, repeated multiple times, can deplete battery life and reduce system efficiency. To prolong battery life and enhance efficiency, it is crucial to minimize the frequency of these activities. Implementing strategies to reduce computational and transmission demands on clients will help maintain their operational longevity and optimize the overall performance of the FRL system.
- *Reliability:* Usually unreliability of radio channels and the finite availability of wireless resources can lead to communication errors in FRL frameworks, resulting in performance degradation. To mitigate these issues, it is essential to implement robust error correction techniques and optimize the allocation of radio resources, ensuring a more reliable and efficient communication process. This approach helps maintain the integrity of the transmitted data and enhances the overall performance of the FRL system.
- Channel conditions: In FRL, the convergence performance of the distributed ML model is significantly influenced by transmitting of model parameters. Consequently, it is of utmost importance to take into account the conditions of the radio channel when scheduling users. In [125], a thorough investigation is carried out regarding the user scheduling technique known as "proportional fair," with a particular focus on its applicability in various radio channel environments. Additionally, in [126], researchers immersed themselves in the domain of FRL within the context of wireless multi-path fading propagation channels. They developed a user scheduling method that chooses a user for transmitting signals based on the complex wireless propagation environment. Furthermore, they extended this method to establish the "best radio channel scheduling method"



FIGURE 10. The accuracy of learning performance across various learning architectures.

through choosing multiple users with the most favorable channel gains [68].

Age of update: User scheduling strategies aim to opti-٠ mize the use of limited radio resources or leverage the diversity of local datasets to maximize updates that the BS can accumulate in each round of global communication. However, these strategies often neglect the issue of update staleness. In [127], the authors introduced a new metric called age-of-update (AoU) to evaluate the staleness of local model updates in each communication round. They then formulated a user scheduling method that considers both the straggler effect and communication quality, aiming to minimize AoU. This approach ensures the freshness of all local updates while maintaining fairness among users. Furthermore, in [128], the authors used AoU as a performance metric for user fairness, and optimization of user selection policy, throughput, transmission power, spectral efficiency, mobility management, and CPU-cycle frequency can be conducted based on this metric.

In the FRL context, the BS can perform model aggregation as soon as it receives model updates from local users, without waiting for potentially delayed users. Fig. 10 shows the accuracy of learning performance across various learning architectures. Furthermore, our assessment encompasses three aggregation scenarios: FL and centralized MARL, both of which aggregate models after each communication round, and FRL, which aggregates models periodically every 10 DRL communication rounds. Given the asynchrony in local users completing their model updates, it is important to note that updates submitted within the same round may contain distinct and potentially outdated information. This is because local models are trained using global model versions received at various time points. Additionally, the time-varying radio channel conditions contribute to the asynchrony in transmitting model updates from multiple local users. Consequently, it is imperative to develop an effective and efficient FRL method tailored for wireless networks. Such an algorithm should adeptly address staleness within the system while working within the constraints of limited wireless resources.

B. EFFECTS OF WIRELESS COMMUNICATION FACTORS ON FRL FRAMEWORK

Various components of cellular networks, such as computational resources, transmission power, and bandwidth, can significantly impact the efficiency of FRL frameworks. The relationship between FRL efficiency and these elements is summarized in Table 4, wherein every tick mark identifies the specific influence on efficiency through each communication component. The subsequent section provides a detailed explanation of how these wireless elements influence the performance and efficiency of FRL systems.

- Based on the BA to each UE, the user throughput, error probability, and signal-to-interference-plus-noise ratio (SINR) is calculated. Consequently, bandwidth distribution significantly impacts latency, reliability, transmission power, training loss, and transmission time. Proper management of bandwidth ensures efficient data transfer, minimizes delays, and maintains high reliability and performance in the FRL framework. By optimizing these factors, the overall system efficiency and effectiveness are enhanced, leading to improved learning outcomes and reduced operational costs.
- As the SGD updates in each iteration are based on computational ability, the required transmission power and training time for local model training are directly impacted by this capability. Additionally, an increase in the number of SGD updates leads to higher training loss and more iterations needed for convergence. Efficiently managing computational resources can thus help minimize training loss and reduce the overall convergence time in the FRL framework.
- Wireless link quality and transmission power influence the data rate, SINR, and transmission error probability. As transmission power increases, the number of iterations, reliability, training time, and training loss all improve [35]. Effective management of transmission power is essential for enhancing the overall performance of the FRL framework by ensuring faster convergence, higher reliability, and reduced training loss. This approach helps optimize the learning process and maintains efficient communication within the wireless network.
- Increasing the number of participating clients in a FRL framework enhances reliability and training time, while reducing the number of iterations and training loss. This is because a larger number of clients contribute more diverse data and computational resources, leading to a more robust and accurate global model. Consequently, the learning process becomes more efficient, achieving convergence with fewer iterations and lower training loss.
- Increasing the size of parameters in local training of a FRL model generally leads to higher power and time requirements. Conversely, it results in fewer iterations,

TABLE 4 Influence of Wireless Communication Factors on FRL Framework

Factors	Training loss	Local training time	Transmission time	Total learning phases	Energy for local training	Energy for data transmission	Latency	Reliability
Radio resources	\checkmark			\checkmark		\checkmark		\checkmark
Computational capability	\checkmark	\checkmark		\checkmark	\checkmark		V	
Transmission power per UE	\checkmark		\checkmark	\checkmark		\checkmark	\checkmark	\checkmark
Communication channel	\checkmark		\checkmark	\checkmark		\checkmark	\checkmark	\checkmark
Participating UEs	\checkmark		\checkmark	\checkmark			\checkmark	\checkmark
Size of trained parameters	\checkmark	\checkmark		\checkmark	\checkmark		\checkmark	\checkmark
Size of transmit- ted parameters			\checkmark			\checkmark		

improved reliability, and reduced training loss. This is because larger parameter sizes can capture more detailed information, which enhances the accuracy and stability of the model but requires more computational resources and time for processing. This trade-off is crucial for optimizing the efficiency and performance of FRL frameworks.

C. INTERPLAY OF FRL FRAMEWORK AND WIRELESS NETWORKS

The learning performance in the FRL framework depends on radio environments, as well as the communication resources and energy constraints of the workers, since all communications between workers and the server occur over wireless links. Factors such as path loss, interference, and fading can influence this performance.

Increasing transmission power raises energy consumption for communication, while reducing the transmission rate requires faster local computation within a fixed time period for each communication round, potentially leading to higher energy consumption for local computation. Given the typically limited battery capacity of mobile devices, minimizing their energy consumption is crucial. This can be achieved by appropriately adjusting local computation and communication parameters, while ensuring the learning performance requirements are met.

The centralized architecture with a parameter server operates similarly to current cellular networks, Wi-Fi, and IoT networks with a central controller. These controllers can be an app, a router, or an IoT device. Even in data centers, communication is often a bottleneck for centralized ML, leading to issues such as fading, additive noise, and bandwidth limitations. These problems can cause network congestion, high energy consumption by user devices, or wireless interference.

In distributed ML, the training goal is global, meaning all participating UEs share a common objective. Therefore, it is essential to use limited wireless resources efficiently. The philosophy of "wireless for FL" focuses on task-oriented approaches, where the aim of the communication system is to derive intelligence from data. The FRL framework can be directly applied to RA, where environment factors include channel quality and interference level. The action space encompasses spectrum access, PA, and spatial resources, while the reward function can be defined in terms of latency, data rate, EE, user throughput, and other relevant metrics.

The framework can be used to obtain an optimal policy for RA to maximize the desired reward. Initially proposed to address concerns of privacy, device computation and storage, and communication bandwidth, the framework has already found numerous wireless applications, such as PA, IM, transmission mode selection, and BA, as detailed in Section VI. The primary advantages of the FRL framework include: i) agents account for the specific nature and environment of individual applications; ii) local interactions between agents can be modeled and examined; iii) challenges in modeling and computation can be tackled in distributed ways.

D. DISCUSSION AND OUTLOOK

A major challenge in autonomous wireless networks is managing the heterogeneity of wireless propagation and adapting effectively without adding complexity. In intelligent cellular networks with vast amounts of sensed information and highly dynamic environments, data or model-assisted optimization methods need to be more consistent. The FRL framework addresses this by handling the stochastic nature of wireless channels and efficiently utilizing limited resources for adaptive optimization in real-time network environments. It synergistically employs both data-driven and model-driven techniques through transfer learning, leveraging wireless propagation capabilities [129]. By learning from interactions with unreliable or dynamic environments, the FRL framework can determine the optimal policy by observing radio environments and the policies of other entities. This approach significantly reduces latency, energy consumption, and network congestion in various use cases, such as interference cancellation, PA, BA, and transmission mode selection for 6G wireless networks.

By enhancing the communication design to focus on learning performance, the efficiency and effectiveness of FRL algorithms can be improved. In [130], the authors introduced a grant-free massive random access scheme for online FL in environments with extensive connectivity, addressing dynamic device participation due to intermittent local updates by

Ref.	Strategy	Main concept	Benefits
[68]	Over the air computation	Perform model aggregation analogously without the need for converting to	Decrease processing delay
		discrete symbols	
[68]	Joint UEs scheduling and	Effectively schedule users to optimize limited wireless resources, account	Enhance FRL performance and
	RA	for the user heterogeneity, and accommodate various data samples	minimize aggregation delay
[129]	Gradient compression	Condense gradient updates prior to transmission	Decrease communication data sam-
			ple, save limited wireless resource

TABLE V. Wireless Communication-Efficient Strategies for FRL Framework

reducing the impacts of transmission outages and device activity on learning outcomes. Despite their potential, practical deployment of FRL systems faces challenges such as managing numerous active UEs, significant delays, and risks to UEs' privacy. To address these issues, [131] described the integration of unsourced random access into FL systems, which supports massive connectivity and safeguards UEs' identity privacy through its unsourced property.

It is noted in [132] that the secrecy capacity cannot be achieved using FL coding schemes for unsourced random access, as it relies on random binning coding schemes. This raises the question: is it possible to develop a coding scheme for edge servers to maximize confusion for eavesdroppers? Addressing this, [133] introduced a practical FBL coding scheme for wireless FL with physical layer security, achieving near-perfect secrecy without compromising learning performance. Conversely, traditional designs of FRL models have primarily focused on optimizing key performance metrics, such as convergence speed, latency, EE, and accuracy, for both synchronous and asynchronous transmissions. In [134], the authors explored optimal RA strategies to improve EE in multi-carrier NOMA and simultaneous wireless information and power transfer-based FL systems supporting asynchronous transmission, aiming to minimize energy use while adhering to latency constraints in FRL systems.

In the context of deploying FRL within cellular networks, the quality of radio links plays a pivotal role in determining the sharing of model parameters. This necessitates greater standards for traditional wireless communication emerging technologies, prompting the need for innovative solutions that can enhance network performance by offering greater reliability and smaller processing delay. Table 5 provides an overview of contemporary wireless technologies, encompassing gradient compression, over-the-air computation, and UEs scheduling and RA, all aimed at improving the sharing of model parameters in a distributed ML manner while enhancing CE. UEs scheduling and RA prove to be effective strategies for addressing the challenges posed by limited wireless resources and a diverse user base, with the ultimate goal of optimizing the convergence performance of distributed ML. However, it is worth noting that the complexity inherent in the UEs scheduling method itself can impact convergence time and should be minimized to ensure efficiency.

In addition to optimizing RA and scheduling strategies, over the air computation takes a unique approach by bypassing digital conversion and employing analog transmission for direct model aggregation. Additionally, decreasing the dimension of local model modernizes at the data samples before sharing proves to be an effective method. This method not only conserves wireless resource but also decreases processing delay. However, it is important to note that many of the techniques mentioned above have primarily been studied in isolation. Consequently, it is becoming increasingly evident that there is a growing need to develop an effective framework that promotes synergy among these wireless communication technologies.

VI. RESOURCE ALLOCATION

RA focuses on minimizing traffic delays and enhancing SE and EE by dynamically distributing the available time-frequency RBs to users based on evolving radio conditions. With the advent of 6G networks, there is a pressing need for significantly increased capacity and reduced latency compared to existing 5G networks.

RA in wireless networks, when based on conventional optimization approaches, faces several challenges that stem from the dynamic and complex nature of wireless environments, as well as from the limitations inherent in traditional optimization methods. These challenges can hinder the effectiveness, efficiency, and practical implementation of RA strategies. Here is a detailed exploration of the key challenges associated with using conventional optimization approaches for RA [9], [54]:

- *Dynamic network conditions:* Wireless networks are characterized by highly dynamic conditions, with varying user demands, mobility patterns, and channel conditions. Conventional optimization methods, which typically rely on static models or assumptions, may not adapt quickly enough to these changes, leading to suboptimal RA.
- *Non-convex problems:* Many RA problems are inherently non-convex, making them difficult to solve using conventional optimization techniques that are designed for convex problems. Non-convexity arises due to various reasons, including interference management and user fairness considerations, and can lead to local optima rather than global solutions.
- *Real-time constraints:* For RA to be effective, it often needs to be performed in real-time or near real-time. Traditional optimization methods may require significant computation time, especially for complex network models, which can make them impractical for real-time applications.
- User fairness: Ensuring fairness among users is a critical aspect of RA. Conventional optimization methods may focus on maximizing overall network throughput or



efficiency without adequately addressing the fairness of bandwidth distribution among users, leading to dissatisfaction and degraded service quality for dynamic users.

- Spectrum efficiency: Maximizing SE is essential for meeting the growing demand for wireless services. Traditional optimization techniques may not effectively balance the trade-offs between maximizing spectrum utilization and ensuring QoS, especially in congested or interference-limited environments.
- Integration with existing infrastructure: Implementing new RA strategies based on conventional optimization approaches may require significant changes to existing network infrastructure and protocols, posing challenges in terms of compatibility, cost, and deployment.
- *Cross-layer dependencies:* RA decisions can have implications across multiple layers of the network stack, from the physical layer up to the application layer. Conventional optimization approaches may not fully account for these cross-layer interactions, leading to solutions that are optimal from a narrow perspective but suboptimal in terms of overall network performance.

To address these issues, numerous researchers have explored ML as a tool to enhance radio channel conditions, thereby supporting RA in upcoming wireless networks. Below, we first review some recent centralized ML approaches for PA and briefly comment on their limitations. We then discuss some related works fostering the use of FRL for PA. Finally, we review recent centralized ML-based methods for BA and introduce a FRL framework to improve network performance and privacy.

A. POWER ALLOCATION

In [135], the authors introduced a Deep-Q-Fully-Connected-Network (DQFCNet) for multicell PA. The simulation results demonstrate that DQFCNet significantly improves both convergence speed and stability when compared to traditional water-filling and Q-learning methods. In [136], an algorithm based on convolutional neural network (CNN) was developed to predict and allocate power factors for each user in a Multi-Input Single-Output (MISO)-NOMA cell. The simulation results show that the proposed approach can improve the system perform compared to selected benchmark methods in terms outage probability and bit error rate (BER). In [137], the authors designed a DNN structure that utilizes the mathematical model of data rates to enhance the learning of PA policies in multicell systems. This design incorporates parameter sharing between the dimension reduction network and the update network within the proposed data-rate based DNN (DRNN), leveraging permutation equivariance (PE) properties to improve learning efficacy. Simulation results show that the sum rate achieved by the learned policy can be enhanced with a given number of training samples, or alternatively, training complexity can be significantly reduced. In [138], the focus was on a device-to-device (D2D) network where ML methods were applied to address power optimization challenges. It was demonstrated that the ML approach, specifically a feedforward neural network (FNN), achieved better performance compared to selected several benchmark methods in terms of QoS metrics across different optimization models.

The main contributions of these works, which focus on centralized ML-based PA, faces several challenges that stem from the intrinsic characteristics of centralized architectures and the dynamic nature of wireless environments. These challenges can impact the effectiveness, efficiency, and practical implementation of PA strategies. Here are some key challenges associated with using centralized ML for PA:

- *Complexity:* Centralized ML models require processing vast amounts of data from all nodes in the network, which can lead to scalability issues as the network grows. The complexity of managing and analyzing this data in real-time increases exponentially with the number of nodes and the variability of network conditions.
- *Communication overhead:* All data generated by endpoints (like sensors, user devices, etc.) must be sent to the central server. As the number of endpoints grows, the volume of data transmission increases, which can lead to congestion in the network and increased latency.
- *Power constrains:* In systems where endpoints are battery-operated or have limited power sources, such as IoT devices, this can be a significant limitation. The need to transmit all data to a central point can lead to inefficiencies, particularly if the data has to travel over busy networks or during peak times, which could result in bottlenecks. This can limit the applicability of such models in power-constrained environments.
- Latency concerns: The need to aggregate data at a central point and then disseminate decisions back to the nodes introduces latency, which in communication networks can include over-the-air delay, backhaul delay, and routing delay. In dynamic environments where conditions change rapidly, these delays can result in PA decisions becoming outdated by the time they are implemented.
- *Single point of failure:* Centralizing the ML model creates a single point of failure. If the central server experiences a fault or becomes compromised, the entire system's ability to allocate power effectively can be jeopardized, impacting network performance and reliability.
- *Privacy and security risks:* Centralizing data aggregation from all nodes raises concerns about privacy and security. This process can expose sensitive information to unauthorized access or attacks, making it challenging to secure data both in transit and when stored.
- Adaptability and flexibility: Adapting centralized ML models to changes in network topology, user behavior, or power availability can be cumbersome. Modifying the model to reflect new conditions or to incorporate new nodes requires retraining and redeployment, which may not be feasible in real-time.
- Model generalization: Centralized ML models trained on data from specific network conditions or configurations may not generalize well to different scenarios. This

can limit the model's effectiveness in environments that differ from those on which it was trained.

To address these challenges, there is growing interest in decentralized FL approaches, where ML models are trained locally on the nodes and only model updates are shared, reducing communication overhead, enhancing privacy, and improving scalability. Additionally, exploring lightweight FL models, optimizing data transmission for model training, and implementing robust security measures are crucial for overcoming the limitations of centralized ML in PA. Below, we provide a brief review of some key works on FL-aided PA to properly set the stage for our contributions in this paper.

1) FL FRAMEWORK FOR POWER ALLOCATION

The study in [139] introduced a PA method aimed at minimizing a loss function while operating within a resource budget, achieving optimal balance between local updates and global parameter aggregation. The authors in [140] addressed FL over cellular networks, focusing on two trade-offs: training time versus power consumption employing the Pareto efficiency model, and computation versus communication training time by optimizing the training accuracy constraint. In [141], a joint optimization problem was formulated for training, transmission PA, and client selection to minimize the loss function. The study in [142] explored FL in wireless networks, formulating a stochastic optimization problem for joint UEs selection and PA under finite energy constraints to maximize throughput. In [143], a cooperative computation and transmission PA and edge association problem for users was analyzed under a hierarchical FL framework to minimize global costs.

While FL can reduce energy consumption during the parameter updating process, it also incurs significant communication costs owing to numerous UEs and communication rounds. Nevertheless, achieving additional reductions in energy consumption remains challenging, particularly in scenarios involving large state and action spaces influenced by time-varying radio propagation. The inherent difficulty in acquiring precise channel state information (CSI) poses considerable challenges for PA problems. Therefore, the development of innovative PA techniques and energy conservation methods for intelligent networks is imperative.

FRL-aided PA in wireless networks offers several advantages over traditional FL approaches. It can enhance EE and spectrum utilization, dynamically adjust user scheduling and power control to improve system performance, and enable more effective and efficient PA in future B5G/6G networks. Additionally, FRL can help reduce latency and ensure sustainable operation by optimizing various network performance indicators, including energy and QoE factors, which are essential for AI-based applications in complex edge computation and wireless communication environments. Below, we provide a brief review of some key works on FRL for PA and energy consumption to properly set the stage for our contributions in this paper.

2) FRL FRAMEWORK FOR POWER ALLOCATION

Recent studies suggest employing DRL algorithms to tackle various optimization challenges in cellular communication networks [144]. PA is a critical issue, and numerous studies have applied DRL algorithms to determine the appropriate transmission power for every device. Some of these studies employ the DQN method for discrete power levels, while others employ advanced DRL methods like trust region policy optimization (TRPO) [145] and deep deterministic policy gradient (DDPG) [146] for PA in multi-cell network scenarios. Current research investigates the potential of DRL algorithms for addressing various wireless resource utilization problems, including optimal PA. However, the optimization problem remains non-convex and challenging due to interference terms in the SINR denominator.

The non-convexity of the problem adds complexity to solving the PA issue. Iterative schemes, although capable of achieving satisfactory network performance, necessitate computationally intensive procedures such as the singular value decomposition (SVD), bisection method, solving NP-hard problems, and channel matrix inversion in each communication round, complicating their execution. Moreover, these methods need knowledge of the CSI for all devices to allocate the proper transmission power for every device.

Hence, finding a nearly optimal solution becomes crucial to ensure efficient performance and high network quality despite having incomplete knowledge of the dynamic environment. In this scenario, employing MARL allows a network entity to learn a more stable policy compared to single-agent RL and DRL, which operate without leveraging information from other network entities. This transforms the problem into a MARL framework [56], where each BS acts as an independent agent determining transmission PA for its connected users in each time slot. Through FRL-based PA methods, these agents fine-tune their transmission power levels based on feedback regarding throughput and transmission powers of other users and neighboring BSs. Hence, the FRL framework emerges as an effective approach to address the PA problem.

Fig. 11 illustrates the variation of EE with respect to the transmission power of each UE for different frameworks. It can be seen that with the increase of the transmission power, the EE performance first increases and then decreases after reaching the maximum. Fig. 11 clearly shows that the proposed FRL framework consumes lower power compared to MARL and FL frameworks. This advantage stems from the FRL framework's ability to allocate the optimal transmit power for each UE in scenarios with a large number of users. This significantly reduces power consumption of each UE while maintaining user desired data rate, ultimately improving EE. The figure demonstrates that the proposed FRL framework achieves higher EE without requiring additional



FIGURE 11. Impact of varying transmission power on EE across different frameworks.

transmission power. Furthermore, it is worth noting that the proposed FRL framework can substantially improve the system performance in terms of EE.

3) FRL FRAMEWORK FOR TACKLING ENERGY LIMITATIONS IN UE

The FRL framework concerning energy consumption presents significant challenges when deployed in wireless networks, primarily due to the energy constraints imposed on UEs [147], [148], [149], [150], [151], [152], [153], [154]. These UEs encompass both mobile and IoT UEs, typically relying on battery power, and must carefully manage their energy utilization during both local model computation and model updates with the central server. This update process involves broadcasting the updated global model from the central server and transmitting the modernized local model to the central server.

In order to attain a predetermined global accuracy target for the training model, UEs are entrusted with the responsibility of carefully apportioning their limited energy resources between computational and transmission tasks. It is worth noting that this allocation directly impacts the overall training time, encompassing both transmission and local computation durations. Consequently, there exists a delicate equilibrium to be established, considering the allocation of energy across computation, transmission, and the ensuing training time. Moreover, the computational process itself faces constraints related to UE resources, including limitations like the maximum CPU cycles per second. Simultaneously, the transmission phase is governed by UE capabilities, i.e., the maximum data transmission rate, as well as the available transmission power.

In a typical FRL system, there is a synchronized computational phase wherein all UEs are required to successfully solve their respective local problems up to a predetermined level of local training accuracy within a specified timeframe. Following this, a communication phase is initiated. This particular implementation approach has the capability to utilize channel access methods such as time division multiple access (TDMA). The trade-off within training time and the overall

rounds until they attain a predetermined level of local training accuracy. In the synchronized execution approach, it has been shown, especially for loss functions that exhibit strong convexity, that there exists an upper limit on the total number of global communication rounds. This upper limit is expressed as $\frac{\mathcal{O}(\log \frac{1}{\beta})}{1-\alpha}$ [148], [149] and is dependent on both the local accuracy (α) and global accuracy (β), as explained in [149]. This implies that, when aiming for a predefined global accuracy target, there exists an inverse relationship between the number of transmissions to the central server and the local accuracy. This, in turn, determines the count of local communication rounds. By utilizing dynamic computational resources, i.e., the ability to adjust CPU cycles per second, and dynamic transmission radio resources, i.e., modernizing upload data rate, the network model can achieve reduced power consumption

power consumption of UEs in this synchronized execution

has been examined in [140]. In this approach, which deviates

from the conventional FL scheme conducting a set number

of local communication rounds, UEs continue communication

and improved training times, particularly wherein CSI is accessible [148], [149]. Moreover, by making efficient choices regarding local accuracy, we have the flexibility to fine-tune the balance between local computation and the number of transmissions, leading to reduced energy usage and enhanced training times [148], [149]. Alternatively, we can explore energy-conscious scheduling, selectively engaging UEs with sufficient computational capacity and favorable channel conditions for global model updates, even if this results in a slight reduction in the training model's accuracy [147], [152]. Table 6 presents a discussion of various methods for addressing the challenge posed by energy-constrained UEs within the FRL framework, along with relevant references.

B. BANDWIDTH ALLOCATION

There is increasing interest in utilizing advanced techniques such as ML, which can provide more adaptive, scalable, and efficient solutions for BA. Centralized ML approaches are more capable of handling the dynamic, non-convex, and real-time aspects of BA in modern wireless networks compared to traditional optimization methods [155], [156], [157], [158], [159], [160], [161], [162], [163], [164], [165], [166], [167]. Although initial efforts focused on centralized ML approaches, recent shifts have favored FRL due to its numerous benefits within wireless network contexts. Below, we provide a brief review of some key works on centralized ML-aided BA to properly set the stage for our contributions in this paper.

1) CENTRALIZED ML TECHNIQUES FOR BANDWIDTH ALLOCATION

In [155], the liquid state machine (LSM) algorithm was introduced to enhance Wi-Fi multiple access performance in Unmanned Aerial Vehicle (UAV)-based LTE networks. It formulates an optimization problem integrating user association,

Ref	Method	Main concent	Limitation
[140]	Execution of	Allowing asynchronous transmission to more enhance both transmission time and	Require the ability to adjust radio
[140]	computation in an	energy consumption	resources for users
	sevenchronous manner	chergy consumption	resources for users
[1.47]		Anninis des modulans of initials and initials and in another sets to be the former of the set	Need west time information shout
[147]	Energy-conscious	Analysis the problem of jointly optimizing radio resource scheduling and mini-	Need real-time information about
	scheduling	mizing energy consumption in FL-aided MIMO networks	the energy levels of users
[148]	Energy-conscious	Choose UEs possessing substantial computational capabilities and favorable radio	Decrease the global accuracy of
	scheduling	channel conditions for modernizing the global ML model	ML model
[149]	Dynamic energy	Dynamically balance CPU cycles per second, transmission data rate, and local	Require the ability to adjust radio
	consumption	accuracy to simultaneously reduce data transmission time and energy consumption	resources for users
[150]	Dynamic training	Employ SGD along with variable batch sizes and adaptable gradient sparsification	The limitation of gradient sparsi-
		to flexibly control computation energy expenditure and communication costs	fication is determined through the
		5 1 05 1	learning rate
[151]	RL	Use RL to identify the optimal power level without any prior knowledge of the	Absence of theoretical assurance
r1		available radio resources	
[152]	Incentive mechanism	Leverage contract theory to create an incentive mechanism that motivates users	Inability to fully leverage all the
L3		with high-quality information to engage in FL	accessible information within the
		whit high quarty information to engage in 12	network
[153]	Incentive mechanism	Apply the Stackelberg game to incentivize user participation	Inefficient utilization of radio re-
[155]	meentive meentinsm	rippi) the blackelooig game to meentivize user participation	sources
[154]	Modifying the loss	Integrate a regularization term into the loss function to penalize the disperity	Enhance the training time
[134]	function	hat was the summer least model and the slabel model. Furthermore, allow	Emilance the training time
	runction	between the current local model and the global model. Furthermore, allow	
		individual users to conduct a variable number of local communication rounds	

TABLE 6 Strategies to Mitigate the Challenges Posed by Energy Constrained UEs Within the FRL Framework

RBs allocation, and content caching to minimize transmission overhead. Compared to traditional learning algorithms like Q-learning, LSM reduces convergence time by up to 20%. Meanwhile, in [156], the DQN algorithm was applied to jointly optimize user association and beamforming in symbiotic radio networks (SRNs), aiming for spectrum, energy, and infrastructure-efficient communications in IoT-cellular networks. This DRL algorithm performs competitively with the optimal user association policy, which requires perfect real-time information.

The authors in [157] presented a model-free DRL framework to address the dynamic BA challenge within a weighted fair queueing (WFQ) system. This WFQ-DRL framework enables the system to derive a control policy that effectively minimizes both average delay and packet loss rate, despite limited bandwidth resources. Trained controllers show superior performance over traditional rule-based policies, such as the longest connected queue (LCQ), especially under real traffic conditions. The WFQ-DRL framework optimizes bandwidth use in telecommunication networks with limited bandwidth, complex traffic patterns, and finite buffering capacities, aligning well with modern routers' operational realities.

In [158], a DRL approach utilizing single-agent actor-critic was proposed for channel assignment in NOMA-based B5G networks. It demonstrated superior performance in terms of sum rate and spectral efficiency compared to traditional methods. Meanwhile, in [159], the DQN method was introduced for cooperative spectrum sensing in CR, enabling spectrum sensing for potential information transmission while avoiding interference with primary users. This approach achieves faster convergence and improved reward performance compared to traditional RL methods employing ϵ -greedy exploration. Lastly, in [160], researchers investigated single-agent DQN and DDQN based DRL methods for dynamic RBs management in cellular networks. These methods enable sensing of discrete frequency channels for potential information transmission, allowing the system to learn to avoid collisions and achieve near-optimal performance even in complex scenarios.

In [161], a deep actor-critic-based DRL model was proposed for dynamic multi-channel access in cellular networks, addressing channel selection. This framework demonstrates competitive performance across 16 channels and superior performance with 32 and 64 channels. Meanwhile, in [162], a DDQN-based DRL model was introduced for distributed spectrum access in B5G networks. It aims to maximize user throughput by prioritizing users with the highest number of packets in their queues, ensuring fair BA. Additionally, in [163], an recurrent neural network (RNN)-based DQN model was presented to enhance channel utilization and reduce packet loss rates in vehicular communications, outperforming existing algorithms. In [164], a DQN-based DRL method was proposed for RB assignment in multi-beam satellite communication systems, improving traffic capacity and spectral efficiency while reducing blocking probability compared to other allocation algorithms. Furthermore, in [165], a Q-learning-aided RL model was explored for LEO satellite communication systems. It focuses on optimizing the joint distribution of fixed channel pre-allocation and dynamic channel scheduling to enhance channel resource efficiency. Lastly, in [166], an MEC-based vehicular network utilized a DDPG-based DRL algorithm for joint spectrum, computing, and storage RA, achieving high satisfaction ratios for delay and QoS with the proposed RBs management strategies.

Finally, in the aforementioned works focused on centralized ML-based methods for BA in various promising technologies. Nevertheless, conventional DRL-based RA is typically not scalable or manageable, making it difficult for DRL methods to converge. Consequently, several new research challenges emerge in the context of the conventional DRL process, as explained below.



Challenge-1: At the start of each iteration in the conventional DRL process, the BS must first handle spectrum allocation duties, such as selecting UE and allocating RBs, which involve solving both convex and non-convex optimization problems. However, the reliability and dependability of these links can be severely compromised in wireless networks due to shadowing and multi-path fading, preventing some UEs from reliably transmitting their CSI to the BS through direct links. Consequently, the BS may struggle to efficiently utilize the available RBs in such scenarios.

Challenge-2: Considering the potential instability of wireless links and the constraints on bandwidth, the swift convergence of the DRL process might be compromised due to the reduced accuracy of the uploaded local models. Consequently, the efficacy of DRL predictions may decline since numerous local models could contain out-of-date CSI at the BS.

The FRL framework represents a revolutionary technology for addressing wireless RA and management in modern wireless networks. It facilitates a global approach to solving complex optimization problems without requiring data sharing among BSs; instead, each BS independently resolves its optimization problem and shares results with neighboring BSs. This approach is particularly beneficial for managing RBs in wireless networks, addressing intricate optimization challenges like UE selection and beamforming for extensive state and action spaces. Consequently, it markedly decreases communication overhead and traffic delays.

2) FRL FRAMEWORK FOR BANDWIDTH ALLOCATION

The above-mentioned challenges have spurred the development of an architecture in which local parameters can be trained utilizing local information, such as CSI. These local parameters can be integrated into a global model to enable devices to learn from each other, thus increasing the user data rate. A widely recognized method to achieve this involves averaging distributed networks to form a global FL model [57], [168], [169], [170], [171], [172].

Fig. 12 illustrates the FRL framework, which is specifically designed to allocate communication bandwidth for each UE. This approach enhances the user throughput and SE. Unlike supervised learning, unsupervised learning, and DRL methods that need large datasets and predefined network models, the FRL framework trains local parameters using CSI obtained from each UE's interaction with the environment. Each network entity relies solely on its local observations to create action decisions, eliminating the need for information exchange between entities. This reduction in communication overhead significantly decreases the computation offloading burden on the UEs.

In this context, we present leveraging DRL to address the combinatorial optimization problem in a decentralized manner. Next, we transform the challenge of joint channel selection and BA as a multi-agent DRL problem. Each M2M pair acts as an independent agent, autonomously executing



FIGURE 12. FRL framework for BA.



FIGURE 13. Impact of the number of RBs on EE for different frameworks.

and refining its BA strategy. We utilize a common FL algorithm, specifically FedAvg [60], for the model iteration. FedAvg organizes the training of a global model into rounds, wherein each round involves updating a local model at each multi-user agent.

In Fig. 13, we plot EE versus the number of RBs for different frameworks. We consider a total system bandwidth of 4MHz divided into 20 RBs and a total of 20 UEs in this plot. Each RB has a bandwidth of 200 kHz. Examining the EE achieved by the three frameworks (FRL, MARL, and FL), we observe that EE increases with the number of RBs. This is because a larger system bandwidth allows for higher throughput and, consequently, higher EE. The reason is that when the number of UEs increases, the proposed FRL framework has more flexibility in choosing RBs. We can therefore conclude that the proposed FRL framework is more energy-efficient than existing schemes like MARL and FL.

D.C.	N. (1)		X • • / /•
Ref.	Method	Main concept	Limitation
[81]	Sparsification	Diminish the parameter count for transmission by mapping the original vector onto	Reduce the rate of convergence in
		a sparse vector	scenarios with non-IID data sam-
			ples distribution
[110]	Model compression	Alter the user's model to a simpler one to decrease the quantity of model	Necessitate design decisions tai-
		parameters	lored to each particular model
[110]	Decreasing the	Increase the count of local communication steps for SGD to alleviate the burden	Reduce the rate of convergence in
	frequency of data		scenarios with non-IID data sam-
	transmission		ples distribution
[170]	Quantization	Implement a combination of sparsification and quantization techniques for FRL	Additional research is necessary to
		when operating over the medium access control layer in cellular networks	understand the impact of sparsifi-
			cation on the rate of convergence
[171]	Autoencoder	Decrease the volume of transmitted data samples through encoding model param-	Necessitate extra computation for
		eters and performing decoding at the central server employing an autoencoder	training the model based on au-
			toencoder
[172]	Analog transmission	Employ multiple antennas at the BS and create an optimal receive beamforming	Depend on CSI and precise syn-
		vector at the BS to alleviate fading impacts, particularly wherein CSI is accessible	chronization for aligning transmis-
		at the BS	sions, making it challenging to em-
			ploy error correcting codes

	ABLE 7 Strategies to Mitigate the Challenges Posed by Limit	ited Wireless Bandwidth Within the FRL Framewo
--	---	--

3) FRL FRAMEWORK FOR EFFICIENT UTILIZATION OF LIMITED WIRELESS BANDWIDTH

FRL necessitates periodic exchanges of model parameters, and these models can be exceedingly large. Consequently, the primary concern in real-world FRL implementation is the substantial communication burden, particularly for UEs with limited wireless bandwidth. For example, the VGGNet design boasts about 138 million parameters (equivalent to 4264 Mb) [168], while UEs often have restricted wireless resources. In ideal radio channel conditions, the maximum wireless uplink data rate is 75 Mb/s [169]. Transmitting VG-GNet in each FRL communication round over a wireless network would take approximately one minute under ideal radio channel conditions. This inherent wireless communication bottleneck in FRL hampers the computational complexity of DL models and user engagement. Moreover, disparities in network capacity among users can lead to performance bottlenecks in FRL framework. Hence, the effective implementation of FRL hinges on communication-efficient constructs capable of accommodating wireless resource variations among UEs. Numerous approaches need to be investigated for mitigating the communication burden in FRL networks. Table 7 presents a discussion of various methods for addressing the challenge posed by limited wireless bandwidth within the FRL framework, along with relevant references.

C. LESSON LEARNED

This section discussed the FRL framework for RA in 6G networks to improve performance and privacy by leveraging local resources at the UEs. Existing ML techniques, particularly RL approaches, offer a promising avenue for addressing enduring and challenging optimization problems. This is primarily due to their advanced decision-making capabilities, especially in dynamic scenarios characterized by uncertainty [173]. Motivated by the success of RL in handling NP-hard and nonconvex problems, we draw inspiration from RL approaches to reevaluate longstanding PA challenges in wireless networks. RL has proven its superiority and potential in wireless communication networks [174], but its deployment

typically assumes a centralized system with global information, which may not be practical in reality [175] and may face feasibility challenges [176]. To overcome the constraints of observing only local information, distributed MARL, has been proposed. For example, in [177], the authors considered each M2M link as an agent, creating a multi-agent communication system in a distributed manner. Nevertheless, the learning quality and stability of the FRL system encounter limitations due to the neglect of strategic interactions among UEs or link agents, especially with a growing number of agents. This issue needs thorough investigation.

Implementing a FRL framework for PA in wireless networks, while ensuring data privacy, introduces a set of challenges. These challenges arise from the need to balance efficient PA with the imperative to protect sensitive information inherent in the data used for learning. PA decisions can reveal sensitive information about network traffic patterns, user behavior, and device locations. Ensuring that the learning process does not inadvertently expose this information to other agents or external entities is crucial. To address this challenges, the implementation of the FRL framework for PA requires a careful design that incorporates advanced cryptographic techniques, robust privacy-preserving algorithms, secure aggregation protocols, and effective anomaly detection mechanisms. Additionally, ongoing research and collaboration between academia, industry, and regulatory bodies are essential to develop secure, efficient, and scalable FRL frameworks for PA.

In [178], the authors proposed a multi-agent DRL approach to allocate radio resources for both unicast and broadcast wireless applications without relying on global knowledge of the CSI. Similarly, in [179], the authors designed an energy-efficient BA protocol based on DRL to maximize the number of admissible service requests in a wireless network and conserve the communication bandwidth. However, these studies require centralized training on data collected from all participating UEs to obtain robust learning models. Since the data is typically decentralized among UEs, centralized data aggregation and training would incur significant communication overhead and raise privacy concerns. The FRL framework can induce the training instability problem induced by the cooperative multi-agent environment for optimal BA. Therefore, robust FRL algorithms for BA should be investigated to address the issue.

Addressing user incentive issues within a FRL framework for BA presents several challenges. These challenges stem from the need to motivate user participation in the learning process while ensuring fair and efficient BA across the network. Ensuring that the incentive mechanism promotes fairness in BA among users. Users contributing more valuable data or computational resources might expect higher rewards, which could lead to disparities in BA. Addressing these expectations while maintaining equitable network access is challenging. To overcome these challenges, the design and implementation of incentive mechanisms within the FRL framework for BA require a multidisciplinary approach that combines insights from network theory, and privacy protection. Moreover, continuous testing, feedback, and adaptation are essential to refine the incentive mechanisms and ensure they effectively motivate user participation while achieving fair and efficient BA.

VII. INTERFERENCE MITIGATION

IM is a critical concern in M2M networks, particularly when reusing the same RBs. To prevent degradation in successive interference cancellation (SIC) performance with increasing average data rates, it is essential to balance the traffic load for user-cell association and PA [180], [181], [182], [183], [184], [185], [186], [187], [188], [189], [190], [191], [192], [193], [194], [195], [196], [197], [198], [199], [200], [201], [202], [203], [204], [205]. Similar to emerging technologies such as millimeter Wave (mmWave), ultra-dense networks (UDNs), beamforming, reconfigurable intelligent surfaces (RISs), satellite, terahertz (THz), semantic, and NOMA communication, significant efforts are being made to develop IM methods to enhance the performance of next-generation cellular networks.

Below, we first review some recent conventional approaches for IM and briefly comment on their limitations. We then discuss some related works promoting the use of centralized ML for IM. Finally, we review recent works on FRL-based methods for IM and introduce a FRL framework that provides additional flexibility in implementing distributed learning for IM.

A. CONVENTIONAL OPTIMIZATION APPROACHES FOR INTERFERENCE MITIGATION

Current research on IM in cellular networks can be classified into four categories: i) time-domain, ii) frequency-domain, iii) transmission power optimization, and iv) spatial-domain methods. Frequency-domain approaches, such as fractional frequency reuse and soft frequency reuse, focus on achieving frequency reuse by using orthogonal polarization states for transmission in networks on the outskirts of a centralized network. In [183], [184], the authors proposed reducing co-channel interference (CCI) by addressing a combined scheduling issue based on dynamic fractional frequency reuse, involving more than two schedulers operating on different time scales. The authors in [185] analyzed a decentralized soft frequency reuse scheme that improves the average cell data rate at the cell edge without requiring data exchanges among small BSs. However, frequency-domain IM approaches are inadequate when wireless resources are limited.

In [186], the authors examined the asymptotic and finite frame length performance of a frame asynchronous coded slotted ALOHA (FA-CSA) system for uncoordinated multiple access. In this system, users join on a slot-by-slot basis according to a Poisson random process and, unlike in standard frame synchronous CSA (FS-CSA), they are not frame-synchronized. FA-CSA generally outperforms FS-CSA in both the error floor (EF) and waterfall regions. Additionally, FA-CSA demonstrates superior delay properties compared to FS-CSA. In FA-CSA, collisions occur when multiple users attempt to transmit in the same slot. The FA-CSA typically handles collisions using simple retransmission strategies, which may not be efficient in highly congested networks to mitigate interference. In [187], the authors provided an extensive literature review of state-of-the-art hybrid automatic repeat request (HARQ) techniques and discuss their integration into various wireless technologies. The review offers insights into the advantages and disadvantages of different automatic repeat request types, as well as open problems and future directions. Traditional HARQ mechanisms do not dynamically adjust transmission parameters, such as modulation and coding schemes, based on time-changing channel conditions. This can lead to inefficient use of bandwidth and power, as these parameters may not be optimally allocated to varying communication channel conditions for mitigating interference.

In [188], a game theory approach was introduced to optimize transmission PA for each device in D2D networks to mitigate interference. However, this solution necessitates additional data exchanges and must account for dynamic radio channel conditions, which diminish its efficacy. Meanwhile, authors in [189] proposed a game theory-based method for joint radio resource and PA in D2D networks, enhancing performance for both cellular UEs (CUEs) and D2D UEs (DUEs) through a mixed game approach aimed at reducing average power consumption. Additionally, in [190], researchers investigated a game theory-based PA method for NOMA networks to mitigate CCI. Despite the potential improvements in network performance offered by game theory and bisection algorithms, solving consecutive decision problems remains challenging and impractical due to prolonged convergence times.

Proper network planning, such as mounting antennas at lower altitudes or adjusting the down-tilt of the victim or aggressor to balance CCI levels, can also mitigate CCI but may decrease cell coverage and increase deployment costs. In the spatial domain, dynamic solutions like beam nulling, interference rejection combining, or beam selection can be applied by the victim. The authors in [191] proposed interference management algorithms for large-scale MIMO networks over backhaul links and joint downlink transmission using zero-forcing (ZF) beamforming to provide the same number of spatial degrees of freedom per user. However, this approach may not be feasible for real-time wireless applications due to the high cost of adding extra antennas at each cell site and the stochastic nature of the wireless channel.

Conventional optimization-based IM methods are unsuitable for mitigating interference in 6G networks. By leveraging global network insights, centralized ML can optimize RA across the network better than conventional optimization schemes. This includes dynamically adjusting bandwidth, power levels, and channel assignments to minimize interference and improve overall network quality and capacity. Below, we review some key works on centralized ML for IM to properly set the stage for our contributions in this paper.

B. CENTRALIZED ML TECHNIQUES FOR INTERFERENCE MITIGATION

In [192], an IM scheme utilizing a distributed PA method based on Q-learning for each UE was proposed. However, the practicality of this solution is questionable due to its assumption that a small BS serves only one user. Meanwhile, authors in [193] examined a RL-based intelligent handoff strategy with information dissemination to mitigate handoff overhead. They also investigated reinforcement-aided edge caching across diverse network configurations, including fixed access points, fog-enabled paradigms, cooperative schemes, and aerial and ground vehicles. The research illustrated that integrating learning with edge caching yields significant advantages, surpassing traditional optimization methods by autonomously and dynamically meeting service requirements online. Nevertheless, these methods necessitate radio signaling exchanges among small BSs or macro BSs, which are undesirable in intelligent networks. Moreover, several approaches overlook transmission power control for each UE or only consider single-cell, single-user scenarios. Hence, developing a fully distributed transmission PA scheme is crucial to minimize radio signaling exchanges among small BSs, thereby reducing interference and enabling intelligent network operations.

In [194], a PA-based time-domain CCI mitigation method employing a Q-learning approach in wireless networks was proposed to improve user throughput. In [195], a new RLbased framework utilizing an adaptive multi-thresholding policy was proposed to efficiently mitigate interference for dynamic scenarios without prior knowledge about the interference links. As the time-domain method for dynamic environments increases the complexity of the IM approach, such as typical joint RBs and PA algorithms, it needs greater processing capabilities of the BSs, thus impacting network performance.

In [196], the authors introduced an algorithm for mitigating narrow-band interference (NBI) and wide-band interference (WBI) utilizing a deep residual network (ResNet).

Subsequently, a detection model based on a conventional CNN framework was developed to ascertain the presence of interference in echo signals. The efficacy of this mitigation algorithm was validated through simulations and on synthetic aperture radar (SAR) data derived from terrain observation by progressive scans (TOPS) mode. Moreover, the performance comparison with notch filtering and eigensubspace filtering demonstrated the superiority of the proposed IM algorithm.

In [197], the authors proposed a novel real-time nonlinear self-interference cancellation strategy, denoted as DL-based self-interference cancellation (DSIC), to facilitate in-band full-duplex (IBFD) wireless communication. The study addressed three critical questions: 1) the method for collecting synchronized wireless channel data for training the DL model, 2) the approach to modeling a wireless channel using a DNN, and 3) the strategy for implementing an open-source software-defined radio (SDR) IBFD wireless framework in real-world scenarios.

In [198], the authors introduced an innovative algorithm based on RNN aimed at reducing interference in environments using frequency modulated continuous wave (FMCW) and Orthogonal Frequency Division Multiplexing (OFDM) radar systems. By integrating an attention module into the existing gated recurrent unit (GRU) model, the enhanced approach more effectively discerns the relationships within time sequences. This advanced model not only eliminates interference but also restores the original signal, outperforming the current leading methods in this domain.

In [199], the authors proposed a novel strategy for radar IM, involving the training of a CNN-based autoencoder to denoise range-Doppler (RD) images affected by interference. The proposed neural network shows significant improvement with respect to SINR compared to other state-of-the-art mitigation techniques, while better preserving phase information of the spectrum.

In [200], the authors introduced a new multi-cell clusterfree NOMA framework, where coordinated beamforming and cluster-free SIC are jointly optimized to mitigate both intra-cell and inter-cell interference. To address the complexities of this mixed integer nonlinear programming (MINLP) problem, a novel communication-efficient distributed autolearning graph neural network (AutoGNN) architecture was developed. This architecture autonomously adjusts the GNN structure, reducing computational and communication demands. Numerical results highlighted the superiority of the cluster-free NOMA approach over traditional cluster-based methods in multi-cell environments and demonstrated the computational and communicational advantages of AutoGNN over existing algorithms.

In [201], the authors proposed a distributed DRL algorithm for BSs to alleviate inter-cell interference in multi-cell networks, relying solely on limited inter-cell information sharing, power measurements at the target cell, and user coordinates. Simulations indicated that this distributed approach closely matches the effectiveness of centralized techniques, with spectral efficiency increasing as the number of cells grows.

TABLE 8 A Summary of Works on Centralized ML Methods-Based IM

Ref.	Approach	Drawbacks
[196]	CNN	 Training large datasets typically demands substantial memory resources. Elevated error vector magnitude can induce distortions in radar applications.
[197]	DNN	 The computational complexity is notably high. Varied complexities, such as differing numbers of hidden layers and parameters, can markedly degrade performance.
[198]	RNN	Vulnerable to exploding gradient problems.Slow convergence rate, making it difficult to track fast time-varying interference.
[199]	Autoencoder	 Poor performance at low SINR with insufficient data. Requires substantial processing time to validate the model.
[200]	GNN	 GNNs can be computationally expensive, particularly for large graphs. As the size of the graph increases, the amount of computation required grows significantly, making them less practical for very large-scale graphs. Reduced consideration of graph edges based on their types and relationships in the GNN model can lead to a significant decline in network performance.
[201]	DRL	 In multi-agent environments, the behavior of one agent can change over time as other agents adapt their strategies. This non-stationarity can make it challenging to learn effective policies since the environment's dynamics are not fixed. As the number of agents increases, the state space grows exponentially, while the action space usually remains independent. This exponential growth makes training multi-agent DRL models computationally expensive and challenging in large-agent scenarios.
[202]	LSTM	 LSTMs are designed to mitigate the vanishing gradient problem in standard RNNs, but they are not entirely immune to it. When dealing with very long sequences, gradients can still vanish, making it difficult for the network to capture long-term dependencies. LSTMs process sequences sequentially, which limits their potential for parallelism during training and inference. LSTMs produce complex internal representations, which can make them challenging to interpret and explain. Understanding why a model made a particular prediction can be difficult.
[203]	GAN	 GAN training can be notoriously unstable. Modal collapse in GAN occurs when the generator model produces a limited set of outputs that fail to capture the full diversity of the real data distribution. Fine-tuning hyperparameters and architectural choices can help mitigate instability, but it remains a challenge. GANs are sensitive to hyperparameter choices, including learning rates, network architectures, and optimization techniques. Finding the right set of hyperparameters can be time-consuming.
[204]	Deep unfolding	 Deep unfolding can result in neural networks with a large number of layers, which can be computationally expensive to train and deploy, especially for real-time or resource-constrained applications. Deeper networks are more prone to vanishing or exploding gradients, and finding the right initialization and training hyperparameters can be time-consuming. Deep unfolding may not scale well to highly complex optimization problems or problems with a large number of variables, as the resulting neural network can become too large and difficult to train effectively.
[205]	Meta-learning	 Some meta-learning methods can be computationally expensive, particularly when dealing with large datasets or complex base models. Meta-learning algorithms often assume that the tasks or domains in the meta-training set share some common structure or similarity. When this assumption is not met, the performance of meta-learners may degrade significantly.

In [202], the authors formulated a long short-term memory (LSTM)-based soft cooperative fusion model to capture the spatial and temporal dependencies in spectrum detection data. This model significantly enhances performance for secondary users (SUs) located near low-cost spectrum sensors (LCSS), suggesting that SUs can leverage cooperative prediction to optimize energy consumption and enhance cognitive capabilities without direct environmental sensing.

In [203], the authors employed a generative adversarial network (GAN) as an innovative interference mitigation strategy for the fast Fourier transform of fast-time samples (RFFT spectrum) in automotive radar systems. This approach significantly boosts the signal-to-interference-plus-noise ratio (SINR) and maintains robustness in complex disturbance scenarios beyond the training dataset's scope.

In [204], the authors discussed deep unfolding methods, including the Analytical Learned Iterative Shrinkage Thresholding Algorithm (ALISTA) and the Analytic Learned Fast Iterative Shrinkage Thresholding Algorithm (ALFISTA). These methods reconstruct corrupted time-domain samples by exploiting the sparsity attribute of radar targets in the range-Doppler plane, utilizing all available uncorrupted data. While ALFISTA demonstrates superior performance, ALISTA is advantageous under constraints of limited computational power and memory.

In [205], the authors addressed the adaptability challenges in current DL-based methods for wireless interference identification using meta-learning. The study simulated a practical scenario where the interference identifier must adapt to unfamiliar signals from new technologies and frequencies with minimal samples. Comparative performance evaluations between a meta-learning model and a traditional DL model in a coexistence system were conducted to assess the impact of a meta-learning-based solution for wireless interference identification.

Table 8 outlines the limitations of various centralized MLbased IM methods. Therefore, developing new distributed ML-aided IM techniques for intelligent networks that aim to maximize the sum rate is crucial. Unlike the current predominant approach that utilizes an infinite state and centralized



FIGURE 14. FRL framework for mitigating interference.

method to tackle the IM problem, the FRL framework is crucial for addressing stochastic optimization problems with extensive state and action spaces. FRL minimizes transmission power per UE, decreases CCI, confirms individual user desired throughput, and enhances overall system performance. This includes accommodating more UEs and decreasing the outage ratio in the proposed model [206], [207], [208], [209], [210], [211].

C. FRL FRAMEWORK FOR INTERFERENCE MITIGATION

Existing MARL methods depend on a combined reward from all agents to accomplish tasks in an uncertain radio propagation environment, where each agent receives a unique reward and shares it with others. However, this method faces practical challenges in real-world wireless applications because sharing observations and rewards raises privacy and security concerns. Within the FRL framework, agents share their local observations and rewards with one another, subsequently updating their strategies to maximize long-term local rewards. MARL can be described as a tuple consisting of sets of states for all agents, observation space, a group of actions, transition function, reward function, and discounting factor. The objective of the proposed network is to optimize the combined reward, calculated as the sum of each agent's reward weighted by its importance. The FRL framework is depicted in Fig. 14. The proposed model utilizes SINR as a metric for predicting channel quality, incorporating factors like additive white Gaussian noise (AWGN), transmission PA for each UE, and CCI among concurrent M2M pairs. Furthermore, the model



FIGURE 15. CDF of the interference power for different frameworks.

determines the data rate by performing actions to obtain rewards for reducing interference, based on a dynamic RB allocation strategy and UE selection [59].

Additionally, a well-trained global model, developed from local learning samples, can accurately predict optimal channels for users across various channel states. This effectiveness extends to newly joined users as long as similar channel states are encountered in the local training samples. In the FRL mechanism, each user undertakes minimal training with local samples, avoiding the need for individual RL model training. Compared to centralized RL, global aggregation significantly reduces users' computing requirements, making it more suitable for mobile users engaged in dynamic spectrum access to mitigate interference.

Fig. 15 illustrates the cumulative distribution function (CDF) curves of interference power for various frameworks. The figure indicates that the proposed framework experiences lower interference compared to the other frameworks, namely MARL and FL, which suffer from more interference in the network. This is beacuse for a large number of UEs in the proposed FRL framework, most UEs utilize the proper optimal transmission power, thereby mitigating interference.

D. INTERFERENCE MITIGATION IN DYNAMIC WIRELESS CHANNELS USING FRL

Another significant obstacle in the deployment of FRL over cellular networks is the impact of radio channel characteristics and their dynamic fluctuations on FRL performance. Unfavorable radio channel conditions can lead to inaccuracies or failures in uploading the updated model parameters. For example, inaccuracies in CSI estimation, quantization of feedback information, signal acquisition delays in multi-path fading radio channels, and similar factors can result in inaccuracies in the obtained updated model by UE or the central server [206]. In real-world implementations where wireless channel noise is an inherent component of the transmitted model, this can prolong the training process, particularly for models that are not resilient to noise, i.e., DNN [207]. This

Ref.	Method	Main concept	Limitation
[149]	Noise-conscious	Allocate resources and schedule UEs based on their superior proportional AWGN	Reduce the overall model precision
	scheduling	conditions to mitigate the influence of interference	_
[206]	Modifying the loss	Incorporate a regularization term into the ML model to estimate the main loss	The model might become trapped
	function	function, smoothing out its region in order to mitigate the influence of noise and	in a local minimum
		interference	
[209]	Dynamic SINR	Employ diversity combining or dynamic transmission power adjustment techniques	Demand real-time noise and inter-
		to balance both uplink and downlink SINR and maintain a good convergence rate	ference information
		in FRL comparable to that achieved with ideal communication conditions	
[211]	Scheduling involving	Apply Gaussian process regression to forecast precise CSI and subsequently	Vulnerable to forecasting inaccura-
	CSI estimation	incorporate this estimation into scheduling decisions using the drift-plus-penalty	cies
		approach derived from the Lyapunov optimization framework	

TABLE 9 Strategies to Mitigate the Challenges Posed by the Dynamic Fluctuations of Wireless Channel to Mitigate Interference Within the FRL Framework

results from the fact that the UE necessitates a higher transmission power, which in turn leads to increased interference among concurrent communication pairs.

Furthermore, radio channels can render the transmitted information of the modernized model undetectable and result in its loss. In cases where an upload fails, waiting for retransmission may not be efficient, especially given that the FRL method is inherently iterative and can incorporate the update in subsequent global model transmissions. Additionally, due to the characteristics of training models, it may even be advantageous to randomly dispose of the UE's modernized model parameters as a measure to preclude overfitting [208]. However, wherein the number of failure radio signal transmissions accumulates due to a decline in the SINR, the FRL algorithm forfeits crucial information about the UE's data samples and their computations. Consequently, in order to achieve a predetermined level of training accuracy, it becomes necessary to execute more global model iterations to retransmit the lost data sample information, ultimately prolonging the training process.

Addressing the challenges of failure signal transmissions is crucial. In cases where networks suffer from failure signal transmissions, one effective approach is to modify the main loss function through introducing a regularizer. This can help mitigate the impact of statistically known radio channel properties, such as quantization errors, to some extent and enhance the training efficiency [206]. Additionally, in the context of the objective loss functions is strictly convex, mitigating the effects of failure signal transmission in FRL over cellular networks can be achieved by scaling the transmission SINR. This can be accomplished through methods like diversity combining or dynamic transmission power adjustment [209]. For cellular networks prone to failure signal transmissions due to factors like AWGN and interference, it is advisable to employ channel-conscious scheduling strategies. These policies allocate communication resources preferentially to UEs with better radio channel conditions, thereby optimizing the training process. This approach is particularly beneficial because UEs with superior radio channel conditions can compensate for the interfering signal information from other UEs [210]. Table 9 presents a discussion of various methods for addressing the challenge posed by the dynamic fluctuations of wireless channel to mitigate interference within the FRL framework, along with relevant references.

E. LESSON LEARNED

This section discussed the FRL framework for IM to meet the high demand while enhancing the QoS. The authors of [212], [213] addressed the spectrum access problem using a distributed game-theoretic stochastic learning method without information interactions. The concept of adjusting the probability of each action based on individual action-reward experiences after each iteration is intriguing. However, their methods are not suitable for solving our fully distributed PA problem, given the continuous action space and the probability of a specific action being zero. Due to the stochastic nature of the radio channel and user arrival process, the adjustment of power for IM in intelligent wireless networks poses a sequential decision problem within the realm of stochastic optimization for the FRL framework [214]. Fortunately, there is a need to investigate and address the challenges of solving sequential decision problems in complex dynamic radio environments for FRL-based IM.

The robustness of optimization models for IM problems in communication networks has been an under-explored area. Most existing algorithms for solving robust optimization problems are centralized [215], making them unsuitable for IM problems that require distributed solutions. FRL with distributionally robust optimization (DRO) addresses this challenge by designing a communication strategy that leverages DRO for IM, allowing control over gradient bias. This aspect should be further investigated to mitigate interference in wireless networks.

VIII. COMMUNICATION MODE SELECTION

The communication modes in M2M networks include direct mode (i.e., M2M mode), indirect mode (i.e., traditional cellular communications), and hybrid mode. These modes are used to deliver various IoT applications, i.e., crowd sensing and video streaming. UEs requiring frequent access to the internet or computing servers with massive capacity for M2M communications utilize these modes [216], [217], [218], [219], [220], [221], [222], [223], [224], [225], [226], [227], [228], [229], [230], [231], [232], [233], [234], [235], [236]. To fully exploit the potential of underlaid M2M communications based on the communication mode selection process, it is crucial to provide the appropriate resources for each UE using the RA scheme and to design an efficient ML-based resource utilization policy that mitigates interference.

Below, we first review some recent conventional approaches for communication mode selection and briefly comment on their limitations. We then discuss some related works fostering the use of centralized ML for CMSM. Finally, we review recent FRL-based methods for the communication mode selection process that can further improve network performance and privacy.

A. CONVENTIONAL OPTIMIZATION APPROACHES FOR COMMUNICATION MODE SELECTION

A location-aware communication mode selection mechanism and RB allocation strategy for M2M networks is proposed in [226], [227]. Simulation results show that this approach significantly enhances system performance by increasing peruser throughput and reducing the traffic load on BSs. In [228], a novel power control mechanism based on location-aware communication mode selection using the water-filling algorithm was presented to reduce interference in M2M networks. Extensive simulations demonstrated the performance of this mechanism compared to selected benchmark algorithms. In [229], the authors described a security-critical message of 1.2 kilobytes requiring a maximum traffic delay of 5 ms and a dependability of 99.999%. However, conventional RA methods in radio networks struggle to meet such diverse QoS requirements, especially for URLLC.

Transmission mode selection is crucial due to its impact on wireless resource utilization and the behavior of the wireless propagation environment. Traditional optimization algorithms, such as the water-filling algorithm, heuristic algorithms, bisection algorithm, and NP-hard algorithms, have been used to address transmission mode selection issues. However, these methods often assume M2M communications occur in a dedicated resource pool, while severe interference between UEs in the shared resource pool remains a challenge. To tackle this, a ML algorithm can be employed, but the dynamic nature of wireless networks requires continuous updates to the M2M pairs' transmission modes, making highly complex algorithms less suitable. Therefore, innovative centralized ML-assisted transmission mode selection-based RA methods are crucial for ensuring latency and reliability requirements. Below, we review key works on centralized ML for transmission mode selection to set the stage for our contributions in this paper.

B. CENTRALIZED ML TECHNIQUES FOR COMMUNICATION MODE SELECTION

To reduce the computational complexity and overhead associated with acquiring complete CSI in traditional methods, adopting AI methods for efficient M2M communications is essential. In [230], the authors explored a DRL-based RA method for M2M networks, where each M2M source acts as an independent agent making decisions based on local observations of radio RBs. This DRL method enables M2M users to autonomously select available channels and power levels to maximize SE while minimizing interference, significantly improving BA and power control.

In [231], the authors described a mode selection-based joint RBs management and PA problem using a RL algorithm for various network load scenarios, including light and heavy network loads, to enhance UE QoS. In [232], the authors proposed a distributed technique for communication mode selection and RB distribution for M2M networks, wherein M2M pairs update their strategies using a RL process. This scheme allows UEs to autonomously select available channels and optimal power to maximize SE while reducing co-tier interference, with convergence reducing computational complexity compared to traditional schemes. Lastly, in [233], the authors proposed a transmission mode selection scheme based on the Q-learning algorithm for resource utilization policy, developing DQN-based and DDPG-based optimization approaches to adjust the PA of cluster heads and the scheduling and BA of UAVs during their missions to improve overall network data transmission performance. The validity and superiority of the proposed approaches were compared with other benchmark policies from different perspectives.

Despite the various sensing elements and realistic radio channel gains, the Q-learning approach may prove ineffective due to large state and action spaces. However, the DRL approach can address these challenges. In [43], the authors investigated a DRL scheme for training an optimal communication mode selection policy from high-dimensional inputs in M2M communications. This approach reduces traffic burden for time-changing radio channels and highly dynamic topology in M2M communications while ensuring QoS necessities. In the aforementioned works focused on centralized ML-aided communication mode selection processes, various aim functions were considered. However, centralized ML algorithm-based communication mode selection process for wireless networks are infeasible for practical applications due to limited radio resources. To address this issue, we explain the FRL-aided transmission mode selection for 6G networks as follows.

C. FRL FRAMEWORK FOR COMMUNICATION MODE SELECTION

The above-mentioned works based on conventional ML systems are typically trained in a centralized way. However, due to constraints like limited resource, communication delay, and privacy concerns, uploading training information is not feasible. Additionally, despite the assumptions in [234], the time-changing radio channel remains unknown at the UE due to the highly dynamic propagation environment. Imperfect training information on each UE also limits the robustness of the DRL method, and improper clustering can significantly degrade network performance. Therefore, a distributed FRL framework is necessary for enabling UEs to create intelligent decisions.

A FL-based mode selection and RA in M2M networks is required to address the challenges posed by unreliable M2M links and heterogeneous QoS requirements. Fig. 16 illustrates the typical architecture and communication mode selection





(a) Typical architecture of FRL framework-aided CMSM.

(b) Controlling mode selection processes through classification.

FIGURE 16. Typical architecture and communication mode selection process of FRL framework.

process of the FRL framework. Additionally, Fig. 16(b) illustrates the type of control that can be employed to establish M2M links. Next, we describe the control of the communication mode selection process.

- *Centralized:* The BS fully controls the UEs and acts as the central entity to control the communication path, co-ordinate interference, establish the communication path, and perform other essential functions within the cell.
- Conventional ML algorithm: The tasks of coordinating interference, creating the communication path, and performing other functions in the cell are handled by a classical ML method, which independently controls the communication path of the UEs, reducing computational load and burden. However, this classical ML method may not be appropriate for many M2M pairs due to the stochastic nature of the channel. Combining neural networks and RL can complicate managing a high number of M2M pairs. In the conventional centralized DQN approach, interactive experiences of the radio environment, such as state transitions, are stored in replay memory and used to train the DQN network. The highly dynamic topology and time-varying spectrum states prevent local observations from representing the global radio environment state, significantly decreasing the efficiency of replay memory. Although some authors have suggested using replay memory for multi-agent-based DRL methods [43], this approach lacks scalability and does not offer an optimal trade-off between communication burden and network performance. In this context, a specific distributed ML method can achieve superior network performance compared to centralized ML methods.
- *Distributed ML algorithm:* To manage the communication path, a specific instance of a distributed FRL algorithm is utilized, with all control processes

VOLUME 5, 2024

asynchronously handled by the UEs. The FRL framework is illustrated in Fig. 16(a). The BSs periodically create undirected graphs based on the proximity of M2M UEs (MUEs) and large-scale radio channel gains, facilitating a high-performance communication mode selection process. For each cluster, selected RBs for each UE are calculated to reduce network size and alleviate communication overload on the BSs. Conversely, the average local networks of M2M pairs are located within the same cluster, using FL methods where M2M pairs simultaneously select their actions and train local networks in every subframe. These are then uploaded and averaged, with the resulting global model responding to the entire set of M2M pairs. Furthermore, the global model can be broadcast to newly operative M2M pairs to reduce training time and achieve fast convergence. Given the training limitations of local DRL models, a FRL algorithm is developed to aid in obtaining robust models.

• *Hybrid:* Interference coordination, path link establishment, and other tasks within the cell are managed by the BSs, using distributed or centralized ML methods to manage the communication path. The goal is to utilize these methods to enhance network performance.

Fig. 17 illustrates the impact of varying the communication mode selection threshold distance on EE for different frameworks. We observe that EE decreases as the mode selection threshold distance increases. This is because the UE experiences more interference, which degrades EE. Additionally, the figure shows that when the mode selection threshold distance increases from 0 meters to 60 meters, EE decreases rapidly. As the mode selection threshold distance increases, the UE requires more transmit power. This is due to the necessity of higher transmission power to overcome the deteriorating



FIGURE 17. Impact of varying the communication mode selection threshold distance on EE for different frameworks.

channel conditions. Furthermore, it is evident that the mode selection threshold distance between M2M pairs significantly affects network performance. The figure demonstrates that the proposed FRL framework achieves superior system performance in the network. This is because when a large number of UEs use optimal RBs and transmission power, they can significantly mitigate interference, thereby enhancing EE as the mode selection threshold distance increases.

D. LESSON LEARNED

This section discussed the FRL framework for transmission mode selection techniques aimed at reducing data processing time and complexity for real-time operation while achieving lower learning errors and faster convergence in model training. Unstable M2M links and the high signaling overhead associated with centralized transmission mode selection methods can significantly constrain safety-critical applications. To address this, a joint optimization problem for communication mode selection and RBs allocation for M2M communications is proposed to meet diverse QoS requirements, particularly URLLC needs.

As noted in [237], UEs frequently experience highly dynamic, rapidly changing, fast-fading radio channels that are unknown to them. To enable UEs to make independent decisions, a distributed DRL method is required. However, the limited local training data available on each UE can impede the robust learning of the DRL model, and improperly federated clusters can reduce network performance. To address these issues, a FRL framework is proposed. This framework functions as a DRL agent, making adaptive decisions based on local observations such as interference levels, traffic loads, and large-scale radio channel qualities. These factors must be examined to select a transmission mode and maximize capacity.

IX. FUTURE DIRECTIONS

Despite notable progress in the configuration, construction, security, and other aspects of wireless network technology, its application and implementation are still in their early stages. This section focuses key challenges and exciting research guidelines for next-generation communication networks on promising FRL technologies that could improve 6G wireless networks.

A. XL-MIMO SYSTEM FOR OWC

Optical wireless communication (OWC) systems can utilize the widespread deployment of light-emitting diodes (LEDs) to enable a distributed MIMO configuration. Extremely largescale MIMO (XL-MIMO) is considered a promising technology for managing the growing UE data traffic by employing extremely large-scale arrays, which significantly enhance spectral efficiency and spatial resolution by increasing the number of antennas. However, a major challenge lies in developing effective CSI estimation approaches to obtain accurate CSI at the source for creating efficient precoders, which require substantial pilot signals and complicate the understanding of XL-MIMO in OWC systems. Additionally, OWC based XL-MIMO systems face significant self-interference due to high spatial correlation in the CSI.

Traditionally, designing precoders involves solving optimization problems in an iterative manner, posing challenges in attaining optimal solutions and managing computational complexity. Consequently, FRL schemes are crucial for the development of robust precoders in multi-antenna OWC networks. It is anticipated that each UE will possess its own set of training data pairs, with the channel matrix as the input and the precoder values as the output. During training, the gradient values from local training at each UE are consolidated at a central server. Once the desired accuracy is achieved, the trained global model is disseminated to the UEs, enabling them to predict the corresponding precoder. Thus, XL-MIMO systems for OWC can significantly enhance the network performance within the FRL framework by exploiting substantial spatial multiplexing gains, which facilitate the parallel execution of tasks such as inference, training, communication, and computation. These aspects warrant further investigation.

B. RISS-BASED OWC SYSTEM

RISs have recently emerged as a revolutionary technology for B5G networks, enhancing radio signal coverage, reliability, and EE. Composed of numerous reconfigurable metasurfaces with unique electromagnetic properties, RISs can manipulate incoming radio signals in various ways [9], [167], [238], [239]. These capabilities include reflection, refraction, beam focusing, wavefront shaping, frequency shifting, splitting, absorption, nonreciprocity, and polarization [9], [167]. Consequently, RISs are particularly significant in FL-based wireless networks, especially in RIS-assisted OWC systems.

In RIS-enhanced OWC systems, multiple RISs can be installed on indoor walls to facilitate various functions that support the transmission of the global model's radio signal.



TABLE 10 Research Directions

Key	Research questions	Research directions
technology	*	
NS	What potential benefits could arise from using promising NS technology in pervasive intelligence with FRL?	Introducing FRL methods for the creation of network slice instances, RBs management, and enhancing QoE in 6G networks represents a promising avenue for research. With FRL, each network slice instance can autonomously learn from its own data and collaborate with other slice instances to optimize RA. Furthermore, integrating FRL into the NS process can elevate the automation and intelligence of network management while reducing operational costs.
XL-MIMO	How can confidentiality be enhanced and traffic delay reduced in cell-free XL-MIMO networks?	FL-aided intelligent 6G networks present a promising technique for enhancing confidentiality and reducing traffic delays. By utilizing the extensive antenna arrays, several researchers have investigated integrating the FL scheme into XL-MIMO networks. However, it is crucial to address some key goals, such as CSI estimation and beamforming design, to fully leverage the benefits of FRL, including improved privacy and decreased traffic delays.
OWC	How can we develop effective meth- ods to decrease the traffic burden and enhance the security of OWC-based outdoor wireless applications?	Further investigation is needed to realize the behavior of OWC communications in practical scenarios, with the goal of improving realistic radio propagation models at the physical layer and developing receiver designs that adapt to variations in transmission power caused by weather conditions, and UE movement. Implementing the FRL framework in OWC with RIS methods could offer solutions that leverage the advantages of both technologies to enhance security and decrease transmission burdens. Therefore, exploring the potential of the FRL framework for OWC using RIS schemes represents a promising research direction for 6G networks.
RISs	How does FRL assist in CSI acquisition?	To attain radio channel estimation, various studies have employed ML-based methods to reduce CSI acquisition overhead. These methods require datasets for supervised or unsupervised training, using received pilot signals as input and predicted channel state as output. Centralized training, as proposed in [238], [239], employs a twin CNN architecture to estimate both direct and cascaded channels, but this approach burdens wireless links. To mitigate this, FRL for RIS-aided communications will improve the network performance.
SC	How can SC be executed and accom- plished efficiently?	In [240], a DNN-based method for classic Huffman coding was proposed to transmit text data. However, performance improvement with DNN was limited, as per Shannon's information theory. Ensuring accurate data restoration in SC, whether through precise information bit retrieval or direct semantic restoration, remains a significant challenge and a constructional problem in SC, as noted by Shannon and Weaver. This raises the question of whether to implement a FRL method for training within SC or to explore different design stages of communication.

Specifically, RISs can help establish line-of-sight (LoS) connections between connected UEs and the computing server, which is crucial for OWC systems as any obstacle can lead to a connection failure. Moreover, RISs can reflect radio signals for energy harvesting, enabling resource-limited UEs to reliably connect their models. By intelligently controlling the electromagnetic properties of incoming wireless signals, beam focusing can be achieved by adjusting RISs located at the source front-end, resulting in improved global model performance and enhanced radio signal coverage for more UEs. Additionally, RISs can bolster physical layer security by destructively transmitting the global model, preventing eavesdroppers from intercepting the signal.

However, to realize these promising benefits in FL-based OWC systems using RISs, it is crucial to properly satisfy and adjust the RIS constraints to achieve the desired outcomes. Notably, the optimization of FRL frameworks for OWC with RISs remains unexplored in the literature, making it a compelling area for future research.

C. EXTENDED REALITY

XR applications in wireless technology represent a modern trend that enhances interactive and immersive experiences by integrating virtual visual and auditory content with realworld dynamic radio environments. These XR-enabled UEs, equipped with GPS modules and instruments, are designed to enrich everyday experiences. However, XR-based wireless applications are often highly localized and particularly sensitive to traffic delays. Additionally, these applications generate vast amounts of data from multiple users, such as images, requiring intensive information processing and efficient use of limited radio resources. Consequently, high-quality XR-based applications demand a high data rate.

Moreover, with the growing need for multi-target virtualization, accurate identification and taxonomy are crucial for enhancing the immersive experience of UEs. To address traffic delays, improve client confidentiality, and reduce traffic burden, XR processes can be managed at both the source and receiver sides using the FRL framework. Additionally, SC has the potential to provide secure and high-speed data rates for XR applications [240], [241].

Thus, creating high-speed cellular connections between XR users and a centralized computing server can offload models and streamline traffic from the congested radio spectrum. Incorporating the FRL framework into SC and OWC is essential for enhancing the user experience of XR-based wireless applications. This area deserves further research.

D. METAVERSE

The metaverse is anticipated to symbolize the next stage of Internet development, succeeding the mobile networks era. In this innovative setup, users-represented by digital avatars-can engage with others and software applications in a 3D virtual space via head-mounted displays [242]. These services demand ultra-high bandwidth and URLLC to provide immersive, delay-free virtual experiences at scale, introducing new challenges for distributed intelligence. To address the diverse and rigorous needs of various metaverse processes, an efficient orchestrator is crucial for managing cloud and edge interactions [243], [244]. Therefore, creating a distributed system, like the FRL framework backed by synchronized endedge-cloud computing, presents a promising research avenue. This system would merge cloud and edge resources with data processing capabilities, ensuring a seamless experience for metaverse users.

Dynamic network slicing (NS) and RA, essential features of the FRL framework for next-generation metaverse applications, such as the allocation of RBs. This enhances QoS, flexibility, and battery life by optimizing network infrastructure for real-time data transmission from sensors, devices, high-resolution videos, social media, and other systems. Lastly, Table 10 describes potential research directions for FRL-aided promising technologies in future wireless networks.

X. CONCLUSION

This paper has delved into various distributed learning frameworks, including FL, MARL, and FRL. Furthermore, we conducted an in-depth analysis of the FRL framework, designed for wireless networks. This analysis covered aspects of wireless communication design, performance assessment, and the impact of wireless factors on FRL parameters. Moreover, we provided a detailed discussion of conventional ML-aided PA, BA, IM, and communication mode selection process techniques for wireless networks. We then addressed their capabilities, shortcomings, and limitations, which have paved the way for integrating the FRL framework in a distributed manner. We also highlighted several critical research challenges and proposed potential directions for advancing next-generation communication networks. In summary, we have provided a comprehensive set of guidelines for implementing FRL frameworks, addressing key issues essential to fully unlocking the potential of intelligent wireless networks.

REFERENCES

- K. B. Letaief, Y. Shi, J. Lu, and J. Lu, "Edge artificial intelligence for 6G: Vision, enabling technologies, and applications," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 5–36, Jan. 2022.
- [2] M. Al-Quraan et al., "Edge-native intelligence for 6G communications driven by federated learning: A survey of trends and challenges," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 7, no. 3, pp. 957–979, Jun. 2023.
- [3] Z. Zhang et al., "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 28–41, Sep. 2019.
- [4] D. C. Nguyen et al., "6G Internet of Things: A comprehensive survey," IEEE Internet Things J., vol. 9, no. 1, pp. 359–383, Jan. 2022.
- [5] F. Guo, F. R. Yu, H. Zhang, X. Li, H. Ji, and V. C. M. Leung, "Enabling massive IoT toward 6G: A comprehensive survey," *IEEE Internet Things J.*, vol. 8, no. 15, pp. 11891–11915, Aug. 2021.
- [6] W. Xu, Z. Yang, D. W. K. Ng, M. Levorato, Y. C. Eldar, and M. Debbah, "Edge learning for B5G networks with distributed signal processing: Semantic communication, edge computing, and wireless sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 1, pp. 9–39, Jan. 2023.
- [7] D. C. Nguyen et al., "Enabling AI in future wireless networks: A data life cycle perspective," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 553–595, Firstquarter 2021.
- [8] N. Kato, B. Mao, F. Tang, Y. Kawamoto, and J. Liu, "Ten challenges in advancing machine learning technologies toward 6G," *IEEE Wireless Commun.*, vol. 27, no. 3, pp. 96–103, Jun. 2020.

- [9] S. K. Das et al., "Comprehensive review on ML-based RIS-enhanced IoT systems: Basics, research progress and future challenges," *Comput. Netw.*, vol. 224, 2023, Art. no. 109581.
- [10] F. Tariq, M. R. A. Khandaker, K.-K. Wong, M. A. Imran, M. Bennis, and M. Debbah, "A speculative study on 6G," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 118–125, Aug. 2020.
- [11] A. Chaoub et al., "6G for bridging the digital divide: Wireless connectivity to remote areas," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 160–168, Feb. 2022.
- [12] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May/Jun. 2020.
- [13] O. A. Wahab, A. Mourad, H. Otrok, and T. Taleb, "Federated machine learning: Survey, multi-level classification, desirable criteria and future directions in communication and networking systems," *IEEE Commun. Surv. Tut.*, vol. 23, no. 2, pp. 1342–1397, Secondquarter 2021.
- [14] L. U. Khan, W. Saad, Z. Han, E. Hossain, and C. S. Hong, "Federated learning for Internet of Things: Recent advances, taxonomy, and open challenges," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1759–1799, Thirdquarter 2021.
- [15] L. Liu, J. Zhang, S. Song, and K. B. Letaief, "Client-edge-cloud hierarchical federated learning," in *Proc. IEEE Int. Conf. Commun.*, 2020, pp. 1–6.
- [16] T. Li, W. Liu, Z. Zeng, and N. N. Xiong, "DRLR: A deepreinforcement-learning-based recruitment scheme for massive data collections in 6G-based IoT networks," *IEEE Internet Things J.*, vol. 9, no. 16, pp. 14595–14609, Aug. 2022.
- [17] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, Mar. 2020.
- [18] P. S. Bouzinis, P. D. Diamantoulakis, and G. K. Karagiannidis, "Wireless federated learning (WFL) for 6G networks-Part I: Research challenges and future trends," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 3–7, Jan. 2022.
- [19] S. Naser et al., "Toward federated-learning-enabled visible light communication in 6G systems," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 48–56, Feb. 2022.
- [20] H. De Oliveira, M. Kaneko, and L. Boukhatem, "Federated multiagent deep reinforcement learning for intelligent IoT wireless communications," Nov. 2023, doi: 10.36227/techrxiv.24007503.v2.
- [21] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation, opportunities, and challenges," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 46–51, Jun. 2020.
- [22] K. M. Faisal and W. Choi, "Machine learning approaches for reconfigurable intelligent surfaces: A survey," *IEEE Access*, vol. 10, pp. 27343–27367, 2022.
- [23] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 2663–2675, 2021.
- [24] D. Wang, S. A. Al-Ahmed, and M. Z. Shakir, "Optimized link distribution schemes for ultrareliable and low-latent communications in multilayer airborne networks," *IEEE Trans. Ind. Inform.*, vol. 16, no. 9, pp. 5866–5873, Sep. 2020.
- [25] S. Roy, Y. P. Singh, U. Biswas, D. S. Gurjar, and T. Goel, "Machine learning in smart transportation systems for mode detection," in *Proc. IEEE 18th India Council Int. Conf.*, 2021, pp. 1–6.
- [26] O. Salman, I. H. Elhajj, A. Chehab, and A. Kayssi, "A machine learning based framework for IoT device identification and abnormal traffic detection," *Trans. Emerg. Telecommun. Technol.*, vol. 33, no. 3, 2022, Art. no. e3743.
- [27] M. S. Mollel et al., "A survey of machine learning applications to handover management in 5G and beyond," *IEEE Access*, vol. 9, pp. 45770–45802, 2021.
- [28] F. Liang, W. Yu, X. Liu, D. Griffith, and N. Golmie, "Toward edgebased deep learning in industrial Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4329–4341, May 2020.
- [29] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Alhussain, "Speech emotion recognition using deep learning techniques: A review," *IEEE Access*, vol. 7, pp. 117327–117345, 2019.
- [30] F. Hussain, R. Hussain, S. A. Hassan, and E. Hossain, "Machine learning in IoT security: Current solutions and future challenges," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1686–1721, Thirdquarter 2020.

- [31] Y. Sun, M. Peng, and S. Mao, "Deep reinforcement learning-based mode selection and resource management for green fog radio access networks," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1960–1971, Apr. 2019.
- [32] J. Posner, L. Tseng, M. Aloqaily, and Y. Jararweh, "Federated learning in vehicular networks: Opportunities and solutions," *IEEE Netw.*, vol. 35, no. 2, pp. 152–159, Mar./Apr. 2021.
- [33] Z. Du, C. Wu, T. Yoshinaga, K.-L. A. Yau, Y. Ji, and J. Li, "Federated learning for vehicular Internet of Things: Recent advances and open issues," *IEEE Open J. Comput. Soc.*, vol. 1, pp. 45–61, 2020.
- [34] W. Y. B. Lim et al., "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 2031–2063, Thirdquarter 2020.
- [35] M. Chen et al., "Distributed learning in wireless networks: Recent progress and future challenges," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 12, pp. 3579–3605, Dec. 2021.
- [36] V.-D. Nguyen, S. K. Sharma, T. X. Vu, S. Chatzinotas, and B. Ottersten, "Efficient federated learning algorithm for resource allocation in wireless IoT networks," *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3394–3409, Mar. 2021.
- [37] H. Yang, A. Alphones, Z. Xiong, D. Niyato, J. Zhao, and K. Wu, "Artificial-intelligence-enabled intelligent 6G networks," *IEEE Netw.*, vol. 34, no. 6, pp. 272–280, Nov./Dec. 2020.
- [38] D. Shome, O. Waqar, and W. U. Khan, "Federated learning and next generation wireless communications: A survey on bidirectional relationship," *Trans. Emerg. Telecommun. Technol.*, vol. 33, no. 7, 2022, Art. no. e4458.
- [39] Z. Yang, M. Chen, K.-K. Wong, H. V. Poor, and S. Cui, "Federated learning for 6G: Applications, challenges, and opportunities," *Engineering*, vol. 8, pp. 33–41, 2022.
- [40] T. Zhang, L. Gao, C. He, M. Zhang, B. Krishnamachari, and A. S. Avestimehr, "Federated learning for the Internet of Things: Applications, challenges, and opportunities," *IEEE Internet Things Mag.*, vol. 5, no. 1, pp. 24–29, Mar. 2022.
- [41] X. Ma, H. Sun, and R. Q. Hu, "Scheduling policy and power allocation for federated learning in NOMA based MEC," in *Proc. IEEE Glob. Commun. Conf.*, 2020, pp. 1–7.
- [42] P. Behmandpoor, P. Patrinos, and M. Moonen, "Federated learning based resource allocation for wireless communication networks," in *Proc. 30th Eur. Signal Process. Conf.*, 2022, pp. 1656–1660.
- [43] Z. Wang, Y. Zhou, Y. Shi, and W. Zhuang, "Interference management for over-the-air federated learning in multi-cell wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 8, pp. 2361–2377, 2022.
- [44] M. Aledhari, R. Razzak, R. M. Parizi, and F. Saeed, "Federated learning: A survey on enabling technologies, protocols, and applications," *IEEE Access*, vol. 8, pp. 140699–140725, 2020.
- [45] J. Ren, J. Sun, H. Tian, W. Ni, G. Nie, and Y. Wang, "Joint resource allocation for efficient federated learning in Internet of Things supported by edge computing," in *Proc. IEEE Int. Conf. Commun. Workshops*, 2021, pp. 1–6.
- [46] F. Kooshki, M. A. Rahman, M. M. Mowla, A. G. Armada, and A. Flizikowski, "Efficient radio resource management for future 6G mobile networks: A cell-less approach," *IEEE Netw. Lett.*, vol. 5, no. 2, pp. 95–99, Jun. 2023.
- [47] J. W. Lambrechts et al., "Intelligent integrated circuits and systems for 5G/6G telecommunications," *IEEE Access*, vol. 12, pp. 21402–21419, 2024.
- [48] Y. Polyanskiy, "A perspective on massive random-access," in Proc. IEEE Int. Symp. Inf. Theory, 2017, pp. 2523–2527.
- [49] A. Fengler, S. Haghighatshoar, P. Jung, and G. Caire, "Non-Bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive MIMO receiver," *IEEE Trans. Inf. Theory*, vol. 67, no. 5, pp. 2925–2951, May 2021.
- [50] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [51] S. Li, Z. Wei, W. Yuan, J. Yuan, B. Bai, and D. W. K. Ng, "On the achievable rates of uplink NOMA with asynchronized transmission," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2021, pp. 1–7.
- [52] T. Li et al., "Applications of multi-agent reinforcement learning in future internet: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 1240–1279, Secondquarter 2022.

- [53] S. Q. Zhang, J. Lin, and Q. Zhang, "A multi-agent reinforcement learning approach for efficient client selection in federated learning," in *Proc. AAAI Conf. Artif. Intell.*, Jun. 2022, vol. 36, pp. 9091–9099.
- [54] S. K. Das, B. Champagne, I. Psaromiligkos, and Y. Cai, "A survey on federated learning for reconfigurable intelligent metasurfacesaided wireless networks," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 1846–1879, 2024.
- [55] D. C. Nguyen et al., "Federated learning for industrial Internet of Things in future industries," *IEEE Wireless Commun.*, vol. 28, no. 6, pp. 192–199, Dec. 2021.
- [56] P. Tehrani, F. Restuccia, and M. Levorato, "Federated deep reinforcement learning for the distributed control of NextG wireless networks," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw.*, 2021, pp. 248–253.
- [57] Z. Ji and Z. Qin, "Federated learning for distributed energy-efficient resource allocation," in *Proc. IEEE Int. Conf. Commun.*, 2022, pp. 1–6.
- [58] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Statist.*, 2017, pp. 1273–1282.
- [59] Y. Song, H.-H. Chang, Z. Zhou, S. Jere, and L. Liu, "Federated dynamic spectrum access," 2021, arXiv:2106.14976.
- [60] X. Liu, Y. Deng, A. Nallanathan, and M. Bennis, "Federated learning and meta learning: Approaches, applications, and directions," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 1, pp. 571–618, Firstquarter 2024.
- [61] T. K. Ho, J. Hull, and S. Srihari, "Decision combination in multiple classifier systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 1, pp. 66–75, Jan. 1994.
- [62] S. Ek, F. Portet, P. Lalanda, and G. Vega, "Evaluation of federated learning aggregation algorithms: Application to human activity recognition," in Proc. Adjunct ACM Int. Joint Conf. Pervasive Ubiquitous Comput. Proc. ACM Int. Symp. Wearable Comput., 2020, pp. 638–643.
- [63] E. Gorbunov, M. Danilova, and A. Gasnikov, "Stochastic optimization with heavy-tailed noise via accelerated gradient clipping," in *Proc.* 34th Int. Conf. Neural Inf. Process. Syst., 2020, pp. 15042–15053.
- [64] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "SCAFFOLD: Stochastic controlled averaging for federated learning," in *Proc. 37th Int. Conf. Mach. Learn.*, 2020, pp. 5132–5143.
- [65] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Nov. 2017.
- [66] V. Mnih et al., "Asynchronous methods for deep reinforcement learning," in *Proc. 33rd Int. Conf. Mach. Learn.*, New York, NY, USA, 2016, pp. 1928–1937.
- [67] M. M. Amiri and D. Gündüz, "Federated learning over wireless fading channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3546–3557, May 2020.
- [68] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 6382–6393.
- [69] T. Chen, K. Zhang, G. B. Giannakis, and T. Başar, "Communicationefficient policy gradient methods for distributed reinforcement learning," *IEEE Trans. Control Netw. Syst.*, vol. 9, no. 2, pp. 917–929, Jun. 2022.
- [70] K. Zhang, Z. Yang, and T. Başar, Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms, vol. 325. Cham, Switzerland: Springer, 2021, pp. 321–384.
- [71] K. Zhang, Z. Yang, H. Liu, T. Zhang, and T. Basar, "Fully decentralized multi-agent reinforcement learning with networked agents," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, pp. 5872–5881.
- [72] S. Zeng, M. A. Anwar, T. T. Doan, A. Raychowdhury, and J. Romberg, "A decentralized policy gradient approach to multi-task reinforcement learning," in *Proc. 37th Conf. Uncertainty Artif. Intell.*, 2021, pp. 1002–1012.
- [73] Y. Li, S. He, Y. Li, Y. Shi, and Z. Zeng, "Federated multiagent deep reinforcement learning approach via physics-informed reward for multimicrogrid energy management," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 5, pp. 5902–5914, May 2024.
- [74] B. Yan et al., "FedCM: A real-time contribution measurement method for participants in federated learning," in *Proc. Int. Joint Conf. Neural Netw.*, 2021, pp. 1–8.

- [75] M. B. Mashhadi, N. Shlezinger, Y. C. Eldar, and D. Gündüz, "Fedrec: Federated learning of universal receivers over fading channels," in *Proc. IEEE Stat. Signal Process. Workshop*, 2021, pp. 576–580.
- [76] F. Sattler, A. Marban, R. Rischke, and W. Samek, "CFD: Communication-efficient federated distillation via soft-label quantization and delta coding," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 4, pp. 2025–2038, Jul./Aug. 2022.
- [77] E. T. Martinez Beltran et al., "Decentralized federated learning: Fundamentals, state of the art, frameworks, trends, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 4, pp. 2983–3013, Fourthquarter 2023.
- [78] N. Rahaman et al., "On the spectral bias of neural networks," in *Proc.* 36th Int. Conf. Mach. Learn., 2019, pp. 5301–5310.
- [79] J. Park et al., "Communication-efficient and distributed learning over wireless networks: Principles and applications," *Proc. IEEE*, vol. 109, no. 5, pp. 796–819, May 2021.
- [80] W. Wen et al., "TernGrad: Ternary gradients to reduce communication in distributed deep learning," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 1508–1518.
- [81] S. U. Stich, J.-B. Cordonnier, and M. Jaggi, "Sparsified SGD with memory," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 4452–4463.
- [82] S. Dorner, S. Cammerer, J. Hoydis, and S. t. Brink, "Deep learning based communication over the air," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 132–143, Feb. 2018.
- [83] S. Hosseinalipour, C. G. Brinton, V. Aggarwal, H. Dai, and M. Chiang, "From federated to fog learning: Distributed machine learning over heterogeneous wireless networks," *IEEE Commun. Mag.*, vol. 58, no. 12, pp. 41–47, Dec. 2020.
- [84] Z.-Q. J. Xu, "Frequency principle: Fourier analysis sheds light on deep neural networks," *Commun. Comput. Phys.*, vol. 28, no. 5, pp. 1746–1767, 2020.
- [85] M. Beitollahi and N. Lu, "Federated learning over wireless networks: Challenges and solutions," *IEEE Internet Things J.*, vol. 10, no. 16, pp. 14749–14763, Aug. 2023.
- [86] L. Qian et al., "Distributed learning for wireless communications: Methods, applications and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 3, pp. 326–342, Apr. 2022.
- [87] X. Liu et al., "Distributed intelligence in wireless networks," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 1001–1039, 2023.
- [88] R. Zeng, C. Zeng, X. Wang, B. Li, and X. Chu, "A comprehensive survey of incentive mechanism for federated learning," 2021, arXiv:2106.15406.
- [89] Y. Liu and J. Wei, "Incentives for federated learning: A hypothesis elicitation approach," 2020, arXiv:2007.10596.
- [90] Y. Zhan, J. Zhang, Z. Hong, L. Wu, P. Li, and S. Guo, "A survey of incentive mechanism design for federated learning," *IEEE Trans. Emerg. Topics Comput.*, vol. 10, no. 2, pp. 1035–1044, Apr.–Jun. 2022.
- [91] S. R. Pandey, N. H. Tran, M. Bennis, Y. K. Tun, A. Manzoor, and C. S. Hong, "A crowdsourcing framework for on-device federated learning," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3241–3256, May 2020.
- [92] Y. Jiao, P. Wang, D. Niyato, B. Lin, and D. I. Kim, "Toward an automated auction framework for wireless federated learning services market," *IEEE Trans. Mobile Comput.*, vol. 20, no. 10, pp. 3034–3048, Oct. 2021.
- [93] J. Kang, Z. Xiong, D. Niyato, S. Xie, and J. Zhang, "Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10700–10714, Dec. 2019.
- [94] W. Y. B. Lim et al., "Towards federated learning in UAV-enabled internet of vehicles: A multi-dimensional contract-matching approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 8, pp. 5140–5154, Aug. 2021.
- [95] T. Wang, J. Rausch, C. Zhang, R. Jia, and D. Song, A Principled Approach to Data Valuation for Federated Learning. Berlin, Germany: Springer, 2020, pp. 153–167.
- [96] Y. Zhan, P. Li, Z. Qu, D. Zeng, and S. Guo, "A learning-based incentive mechanism for federated learning," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6360–6368, Jul. 2020.
- [97] Y. Liu, S. Sun, Z. Ai, S. Zhang, Z. Liu, and H. Yu, "Fedcoin: A peer-topeer payment system for federated learning," in *Federated Learning*. Berlin, Germany: Springer, 2020, pp. 125–138.

- [98] P. Mach and Z. Becvar, "Mobile edge computing: A survey on architecture and computation offloading," *IEEE Commun. Surv. Tut.*, vol. 19, no. 3, pp. 1628–1656, Thirdquarter 2017.
- [99] V. Andalibi, J. Dev, D. Kim, E. Lear, and L. J. Camp, "Making access control easy in IoT," in *Proc. Hum. Aspects Inf. Secur. Assurance*, 2021, pp. 127–137.
- [100] S. Sardellitti, G. Scutari, and S. Barbarossa, "Joint optimization of radio and computational resources for multicell mobile-edge computing," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 1, no. 2, pp. 89–103, Jun. 2015.
- [101] C. You, K. Huang, H. Chae, and B.-H. Kim, "Energy-efficient resource allocation for mobile-edge computation offloading," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1397–1411, Mar. 2017.
- [102] M. Chen and Y. Hao, "Task offloading for mobile edge computing in software defined ultra-dense network," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 3, pp. 587–597, Mar. 2018.
- [103] T. X. Tran and D. Pompili, "Joint task offloading and resource allocation for multi-server mobile-edge computing networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 856–868, Jan. 2019.
- [104] Y. Mao, J. Zhang, and K. B. Letaief, "Dynamic computation offloading for mobile-edge computing with energy harvesting devices," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3590–3605, Dec. 2016.
- [105] Y. Mao, J. Zhang, S. H. Song, and K. B. Letaief, "Stochastic joint radio and computational resource management for multi-user mobile-edge computing systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5994–6009, Sep. 2017.
- [106] J. Zhang, W. Xia, F. Yan, and L. Shen, "Joint computation offloading and resource allocation optimization in heterogeneous networks with mobile edge computing," *IEEE Access*, vol. 6, pp. 19324–19337, 2018.
- [107] Q.-V. Pham, T. Leanh, N. H. Tran, B. J. Park, and C. S. Hong, "Decentralized computation offloading and resource allocation for mobile-edge computing: A matching game approach," *IEEE Access*, vol. 6, pp. 75868–75885, 2018.
- [108] J. Li, H. Gao, T. Lv, and Y. Lu, "Deep reinforcement learning based computation offloading and resource allocation for MEC," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2018, pp. 1–6.
- [109] X. Chen, H. Zhang, C. Wu, S. Mao, Y. Ji, and M. Bennis, "Optimized computation offloading performance in virtual edge computing systems via deep reinforcement learning," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4005–4018, Jun. 2019.
- [110] A. Triastcyn and B. Faltings, "Federated generative privacy," *IEEE Intell. Syst.*, vol. 35, no. 4, pp. 50–57, Jul./Aug. 2020.
- [111] C. Ma et al., "On safeguarding privacy and security in the framework of federated learning," *IEEE Netw.*, vol. 34, no. 4, pp. 242–248, Jul./Aug. 2020.
- [112] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in *Proc. IEEE Symp. Secur. Privacy*, Los Alamitos, CA, USA, 2017, pp. 3–18.
- [113] S. Truex et al., "A hybrid approach to privacy-preserving federated learning," in *Proc. 12th ACM Workshop Artif. Intell. Secur.*, 2019, pp. 1–11.
- [114] M. Seif, R. Tandon, and M. Li, "Wireless federated learning with local differential privacy," in *Proc. IEEE Int. Symp. Inf. Theory*, 2020, pp. 2604–2609.
- [115] D. Liu and O. Simeone, "Privacy for free: Wireless federated learning via uncoded transmission with adaptive power control," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 170–185, Jan. 2021.
- [116] J. Kang, Z. Xiong, D. Niyato, Y. Zou, Y. Zhang, and M. Guizani, "Reliable federated learning for mobile networks," *IEEE Wireless Commun.*, vol. 27, no. 2, pp. 72–80, Apr. 2020.
- [117] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, "A survey on security and privacy of federated learning," *Future Gener. Comput. Syst.*, vol. 115, pp. 619–640, 2021.
- [118] K. Wei et al., "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 3454–3469, 2020.
- [119] A. N. Bhagoji, S. Chakraborty, P. Mittal, and S. Calo, "Analyzing federated learning through an adversarial lens," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 634–643.
- [120] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, "How to backdoor federated learning," in *Proc. 23rd Int. Conf. Artif. Intell. Statist.*, 2020, pp. 2938–2948.

Vehicular Technology

- [121] A. Rehman, I. Razzak, and G. Xu, "Federated learning for privacy preservation of healthcare data from smartphone-based side-channel attacks," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 2, pp. 684–690, Feb. 2023.
- [122] W. Xu, K. Ma, W. Trappe, and Y. Zhang, "Jamming sensor networks: Attack and defense strategies," *IEEE Netw.*, vol. 20, no. 3, pp. 41–47, May/Jun. 2006.
- [123] F. Lau, S. Rubin, M. Smith, and L. Trajkovic, "Distributed denial of service attacks," in *Proc. IEEE Int. Conf. Syst. Man Cybern. Evolving Syst. Hum. Org. Complex Interact.*, 2000, pp. 2275–2280.
- [124] H. B. McMahan, D. Ramage, K. Talwar, and L. Zhang, "Learning differentially private recurrent language models," in *Proc. Int. Conf. Learn. Representations*, 2018.
- [125] H. H. Yang, Z. Liu, T. Q. S. Quek, and H. V. Poor, "Scheduling policies for federated learning in wireless networks," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 317–333, Jan. 2020.
- [126] M. M. Amiri, D. Gündüz, S. R. Kulkarni, and H. V. Poor, "Convergence of update aware device scheduling for federated learning at the wireless edge," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3643–3658, Jun. 2021.
- [127] H. H. Yang, A. Arafa, T. Q. S. Quek, and H. Vincent Poor, "Age-based scheduling policy for federated learning in mobile edge networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 8743–8747.
- [128] B. Yin, Z. Chen, and M. Tao, "Joint user scheduling and resource allocation for federated learning over wireless networks," in *Proc. IEEE Glob. Commun. Conf.*, 2020, pp. 1–6.
- [129] Y. Lin, S. Han, H. Mao, Y. Wang, and W. J. Dally, "Deep gradient compression: Reducing the communication bandwidth for distributed training," 2020, arXiv:1712.01887.
- [130] S. Xia, Y. Shi, Y. Zhou, Y. Wu, L. F. Yang, and K. B. Letaief, "Federated learning with massive random access," *IEEE Trans. Wireless Commun.*, early access, Jun. 03, 2024, doi: 10.1109/TWC.2024.3405451.
- [131] Y. Tian, J. Che, Z. Zhang, and Z. Yang, "Federated learning with unsourced random access," in *Proc. IEEE 97th Veh. Technol. Conf.*, 2023, pp. 1–5.
- [132] H. Zhang, C. Yang, and B. Dai, "When wireless federated learning meets physical layer security: The fundamental limits," in *Proc. IEEE Conf. Comput. Commun. Workshops*, 2022, pp. 1–6.
- [133] H. Zhang, C. Yang, and B. Dai, "A finite blocklength approach for wireless hierarchical federated learning in the presence of physical layer security," in *Proc. IEEE Conf. Comput. Commun. Workshops*, 2023, pp. 1–6.
- [134] M. C. Ho, A. T. Tran, D. Lee, J. Paek, W. Noh, and S. Cho, "A DDPGbased energy efficient federated learning algorithm with SWIPT and MC-NOMA," *ICT Exp.*, vol. 10, no. 3, pp. 600–607, 2024.
- [135] Y. Zhang, C. Kang, T. Ma, Y. Teng, and D. Guo, "Power allocation in multi-cell networks using deep reinforcement learning," in *Proc. IEEE* 88th Veh. Technol. Conf., 2018, pp. 1–6.
- [136] M. Gaballa, M. Abbod, and S. Alnasur, "Hybrid deep learning for channel estimation and power allocation for MISO-NOMA system," in *Proc. IEEE Future Netw. World Forum*, 2022, pp. 361–366.
- [137] J. Guo and C. Yang, "Learning power allocation for cellular systems with data rate-based deep neural network," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2022, pp. 1033–1038.
- [138] J. Zhou, X. Liu, and C. Huang, "Machine learning for power allocation of a D2D network," in *Proc. IEEE Glob. Commun. Conf.*, 2019, pp. 1–6.
- [139] S. Wang et al., "Adaptive federated learning in resource constrained edge computing systems," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1205–1221, Jun. 2019.
- [140] N. H. Tran, W. Bao, A. Zomaya, M. N. H. Nguyen, and C. S. Hong, "Federated learning over wireless networks: Optimization model design and analysis," in *Proc. IEEE Conf. Comput. Commun.*, 2019, pp. 1387–1395.
- [141] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "Performance optimization of federated learning over wireless networks," in *Proc. IEEE Glob. Commun. Conf.*, 2019, pp. 1–6.
- [142] J. Xu and H. Wang, "Client selection and bandwidth allocation in wireless federated learning networks: A long-term perspective," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1188–1200, Feb. 2021.

- [143] S. Luo, X. Chen, Q. Wu, Z. Zhou, and S. Yu, "HFEL: Joint edge association and resource allocation for cost-efficient hierarchical federated edge learning," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6535–6548, Oct. 2020.
- [144] B. S. Ciftler, A. Alwarafy, and M. Abdallah, "Distributed DRL-based downlink power allocation for hybrid RF/VLC networks," *IEEE Photon. J.*, vol. 14, no. 3, Jun. 2022, Art. no. 8632510.
- [145] M. G. Khoshkholgh and H. Yanikomeroglu, "Faded-experience trust region policy optimization for model-free power allocation in interference channel," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 659–663, Mar. 2021.
- [146] J. Ren and S. Xu, "DDPG based computation offloading and resource allocation for MEC systems with energy harvesting," in *Proc. IEEE* 93rd Veh. Technol. Conf., 2021, pp. 1–5.
- [147] R. Hamdi, M. Chen, A. B. Said, M. Qaraqe, and H. V. Poor, "Federated learning over energy harvesting wireless networks," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 92–103, Jan. 2022.
- [148] Z. Yang, M. Chen, W. Saad, C. S. Hong, and M. Shikh-Bahaei, "Energy efficient federated learning over wireless communication networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1935–1949, Mar. 2021.
- [149] C. Ma et al., "Distributed optimization with arbitrary local solvers," Optim. Methods Softw., vol. 32, no. 4, pp. 813–848, 2017.
- [150] Q. Zeng, Y. Du, K. Huang, and K. K. Leung, "Energy-efficient radio resource allocation for federated edge learning," in *Proc. IEEE Int. Conf. Commun. Workshops*, 2020, pp. 1–6.
- [151] L. Li, D. Shi, R. Hou, H. Li, M. Pan, and Z. Han, "To talk or to work: Flexible communication compression for energy efficient federated learning over heterogeneous mobile edge devices," in *Proc. IEEE Conf. Comput. Commun.*, 2021, pp. 1–10.
- [152] T. T. Anh, N. C. Luong, D. Niyato, D. I. Kim, and L.-C. Wang, "Efficient training management for mobile crowd-machine learning: A deep reinforcement learning approach," *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1345–1348, Oct. 2019.
- [153] J. Kang, Z. Xiong, D. Niyato, H. Yu, Y.-C. Liang, and D. I. Kim, "Incentive design for efficient federated learning in mobile networks: A contract theory approach," in *Proc. IEEE VTS Asia-Pacific Wireless Commun. Symp.*, 2019, pp. 1–5.
- [154] L. U. Khan et al., "Federated learning for edge networks: Resource optimization and incentive mechanism," *IEEE Commun. Mag.*, vol. 58, no. 10, pp. 88–93, Oct. 2020.
- [155] M. Chen, W. Saad, and C. Yin, "Liquid state machine learning for resource allocation in a network of cache-enabled LTE-U UAVs," in *Proc. IEEE Glob. Commun. Conf.*, 2017, pp. 1–6.
- [156] Q. Zhang, Y.-C. Liang, and H. V. Poor, "Intelligent user association for symbiotic radio networks using deep reinforcement learning," in *Proc. IEEE Glob. Commun. Conf.*, 2019, pp. 1–6.
- [157] J. Pan, G. Chen, H. Wu, X. Peng, and L. Xia, "Deep reinforcement learning-based dynamic bandwidth allocation in weighted fair queues of routers," in *Proc. IEEE 18th Int. Conf. Automat. Sci. Eng.*, 2022, pp. 1580–1587.
- [158] J. Zheng, X. Tang, X. Wei, H. Shen, and L. Zhao, "Channel assignment for hybrid NOMA systems with deep reinforcement learning," *IEEE Wireless Commun. Lett.*, vol. 10, no. 7, pp. 1370–1374, Jul. 2021.
- [159] Y. Zhang, P. Cai, C. Pan, and S. Zhang, "Multi-agent deep reinforcement learning-based cooperative spectrum sensing with upper confidence bound exploration," *IEEE Access*, vol. 7, pp. 118898–118906, 2019.
- [160] Y. Xu, J. Yu, W. Headley, and R. Buehrer, "Deep reinforcement learning for dynamic spectrum access in wireless networks," in *Proc. IEEE Mil. Commun. Conf.*, 2018, pp. 207–212.
- [161] C. Zhong, Z. Lu, M. C. Gursoy, and S. Velipasalar, "A deep actor-critic reinforcement learning framework for dynamic multichannel access," *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, no. 4, pp. 1125–1139, Dec. 2019.
- [162] S. B. Janiar and V. Pourahmadi, "Deep-reinforcement learning for fair distributed dynamic spectrum access in wireless networks," in *Proc. IEEE 18th Annu. Consum. Commun. Netw. Conf.*, 2021, pp. 1–4.
- [163] Y. Wang, X. Li, P. Wan, and R. Shao, "Intelligent dynamic spectrum access using deep reinforcement learning for VANETs," *IEEE Sensors J.*, vol. 21, no. 14, pp. 15554–15563, Jul. 2021.

- [164] S. Liu, X. Hu, and W. Wang, "Deep reinforcement learning based dynamic channel allocation algorithm in multibeam satellite systems," *IEEE Access*, vol. 6, pp. 15733–15742, 2018.
- [165] F. Zheng, Z. Pi, Z. Zhou, K. Wang, and C.-H. Chen, "LEO satellite channel allocation scheme based on reinforcement learning," *Mobile Inf. Syst.*, vol. 2020, 2020, Art. no. 8868888.
- [166] H. Peng and X. Shen, "Deep reinforcement learning based resource management for multi-access edge computing in vehicular networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 4, pp. 2416–2428, Oct.– Dec. 2020.
- [167] Z. T. Azim, S. A. Shaon, S. K. Das, and M. G. Mostafa, "Optimization of reconfigurable intelligent surface for M2M communications over cellular networks," in *Proc. 5th Int. Conf. Elect. Eng. Inf. Commun. Technol.*, 2021, pp. 1–7.
- [168] Z. Liu, J. Li, Z. Shen, G. Huang, S. Yan, and C. Zhang, "Learning efficient convolutional networks through network slimming," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2755–2763.
- [169] Z. Shen, A. Papasakellariou, J. Montojo, D. Gerstenberger, and F. Xu, "Overview of 3GPP LTE-advanced carrier aggregation for 4G wireless communications," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 122–130, Feb. 2012.
- [170] M. M. Amiri and D. Gündüz, "Machine learning at the wireless edge: Distributed stochastic gradient descent over-the-air," in *Proc. IEEE Int. Symp. Inf. Theory*, 2019, pp. 1432–1436.
- [171] M. Beitollahi and N. Lu, "Flac: Federated learning with autoencoder compression and convergence guarantee," in *Proc. IEEE Glob. Commun. Conf.*, 2022, pp. 4589–4594.
- [172] M. M. Amiri, T. M. Duman, D. Gunduz, S. R. Kulkarni, and H. V. Poor, "Blind federated edge learning," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 5129–5143, Aug. 2021.
- [173] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduc*tion. Cambridge, MA, USA: MIT Press, 2018.
- [174] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.
- [175] Y. He, N. Zhao, and H. Yin, "Integrated networking, caching, and computing for connected vehicles: A deep reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 67, no. 1, pp. 44–55, Jan. 2018.
- [176] L. Liang, S. Xie, G. Y. Li, Z. Ding, and X. Yu, "Graph-based resource sharing in vehicular communication," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4579–4592, Jul. 2018.
- [177] L. Liang, H. Ye, and G. Y. Li, "Spectrum sharing in vehicular networks based on multi-agent reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2282–2292, Oct. 2019.
- [178] H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep reinforcement learning based resource allocation for V2V communications," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3163–3173, Apr. 2019.
- [179] R. F. Atallah, C. M. Assi, and J. Y. Yu, "A reinforcement learning technique for optimizing downlink scheduling in an energy-limited vehicular network," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 4592–4601, Jun. 2017.
- [180] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. Machine Learn. Syst.*, I. Dhillon, D. Papailiopoulos, and V. Sze, Eds., 2020, vol. 2, pp. 429–450.
- [181] N. H. Mahmood, I. S. Ansari, G. Berardinelli, P. Mogensen, and K. A. Qaraqe, "Analysing self interference cancellation in full duplex radios," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2016, pp. 1–6.
- [182] E. Ntuli, C. Du, and F. Mekuria, "Review of interference mitigation techniques for white space networks," in *Proc. IEEE 2nd Wireless Afr. Conf.*, 2019, pp. 1–6.
- [183] M. Jia, X. Zhang, X. Gu, Q. Guo, Y. Li, and P. Lin, "Interbeam interference constrained resource allocation for shared spectrum multibeam satellite communication systems," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6052–6059, Aug. 2019.
- [184] N. Saquib, E. Hossain, and D. I. Kim, "Fractional frequency reuse for interference management in LTE-advanced hetnets," *IEEE Wireless Commun.*, vol. 20, no. 2, pp. 113–122, Apr. 2013.
- [185] F. Bernardo, R. Agustí, J. Pérez-Romero, and O. Sallent, "Intercell interference management in OFDMA networks: A decentralized approach based on reinforcement learning," *IEEE Trans. Syst., Man, Cybern., Part C (Appl. Rev.)*, vol. 41, no. 6, pp. 968–976, Nov. 2011.

- [186] E. Sandgren, A. Graell i Amat, and F. Brännström, "On frame asynchronous coded slotted aloha: Asymptotic, finite length, and delay analysis," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 691–704, Feb. 2017.
- [187] A. Ahmed, A. Al-Dweik, Y. Iraqi, H. Mukhtar, M. Naeem, and E. Hossain, "Hybrid automatic repeat request (HARQ) in wireless communications systems and standards: A contemporary survey," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 4, pp. 2711–2752, Fourthquarter 2021.
- [188] G. Zhang, J. Hu, W. Heng, X. Li, and G. Wang, "Distributed power control for D2D communications underlaying cellular network using stackelberg game," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2017, pp. 1–6.
- [189] S. Najeh and A. Bouallegue, "Distributed interference and power control based-game theory for D2D communication," in *Proc. 15th Int. Wireless Commun. Mobile Comput. Conf.*, 2019, pp. 991–996.
- [190] K. Kang, Z. Pan, J. Liu, and S. Shimamoto, "A game theory based power control algorithm for future MTC NOMA networks," in *Proc.* 14th IEEE Annu. Consum. Commun. Netw. Conf., 2017, pp. 203–208.
- [191] K. Hosseini, W. Yu, and R. S. Adve, "Large-scale MIMO versus network MIMO for multicell interference mitigation," in *Proc. IEEE* 15th Int. Workshop Signal Process. Adv. Wireless Commun., 2014, pp. 70–74.
- [192] S. Nie, Z. Fan, M. Zhao, X. Gu, and L. Zhang, "Q-learning based power control algorithm for D2D communication," in *Proc. IEEE 27th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun.*, 2016, pp. 1–6.
- [193] N. Nomikos, S. Zoupanos, T. Charalambous, and I. Krikidis, "A survey on reinforcement learning-aided caching in heterogeneous mobile edge networks," *IEEE Access*, vol. 10, pp. 4380–4413, 2022.
- [194] W. Lu, Q. Fan, Z. Li, and H. Lu, "Power control based time-domain inter-cell interference coordination scheme in DSCNs," in *Proc. IEEE Int. Conf. Commun.*, 2016, pp. 1–6.
- [195] R. Barazideh, O. Semiari, S. Niknam, and B. Natarajan, "Reinforcement learning for mitigating intermittent interference in terahertz communication networks," in *Proc. IEEE Int. Conf. Commun. Workshops*, 2020, pp. 1–6.
- [196] W. Fan, F. Zhou, P. Rong, and X. Yao, "Interference mitigation for synthetic aperture radar using deep learning," in *Proc. 6th Asia-Pacific Conf. Synthetic Aperture Radar*, 2019, pp. 1–6.
- [197] H. Guo, S. Wu, H. Wang, and M. Daneshmand, "DSIC: Deep learning based self-interference cancellation for in-band full duplex wireless," in *Proc. IEEE Glob. Commun. Conf.*, 2019, pp. 1–6.
- [198] J. Mun, S. Ha, and J. Lee, "Automotive radar signal interference mitigation using RNN with self attention," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 3802–3806.
- [199] J. Fuchs, A. Dubey, M. Lübke, R. Weigel, and F. Lurz, "Automotive radar interference mitigation using a convolutional autoencoder," in *Proc. IEEE Int. Radar Conf.*, 2020, pp. 315–320.
- [200] X. Xu, Y. Liu, Q. Chen, X. Mu, and Z. Ding, "Distributed autolearning GNN for multi-cell cluster-free NOMA communications," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 4, pp. 1243–1258, Apr. 2023.
- [201] M. Dahal and M. Vaezi, "Multi-agent deep reinforcement learning for multi-cell interference mitigation," in *Proc. 57th Annu. Conf. Inf. Sci. Syst.*, 2023, pp. 1–6.
- [202] B. S. Shawel, D. Hailemariam Woledegebre, and S. Pollin, "Deeplearning based cooperative spectrum prediction for cognitive networks," in *Proc. Int. Conf. Inf. Commun. Technol. Convergence*, 2018, pp. 133–137.
- [203] S. Chen, W. Shangguan, J. Taghia, U. Kühnau, and R. Martin, "Automotive radar interference mitigation based on a generative adversarial network," in *Proc. IEEE Asia-Pacific Microw. Conf.*, 2020, pp. 728–730.
- [204] J. Overdevest, A. Koppelaar, M. Bekooij, J. Youn, and R. v. Sloun, "Signal reconstruction for FMCW radar interference mitigation using deep unfolding," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2023, pp. 1–5.
- [205] A. Owfi, F. Afghah, and J. Ashdown, "Meta-learning for wireless interference identification," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2023, pp. 1–6.
- [206] F. Ang, L. Chen, N. Zhao, Y. Chen, W. Wang, and F. R. Yu, "Robust federated learning with noisy communication," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3452–3464, Jun. 2020.



- [207] Y. Tang and C. Eliasmith, "Deep networks for robust visual recognition," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 1055–1062.
- [208] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [209] X. Wei and C. Shen, "Federated learning over noisy channels: Convergence analysis and design examples," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 2, pp. 1253–1268, Jun. 2022.
- [210] W. Shi, S. Zhou, Z. Niu, M. Jiang, and L. Geng, "Joint device scheduling and resource allocation for latency constrained wireless federated learning," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 453–467, 2021.
- [211] M. M. Wadu, S. Samarakoon, and M. Bennis, "Federated learning under channel uncertainty: Joint client scheduling and resource allocation," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2020, pp. 1–6.
- [212] Y. Xu, J. Wang, Q. Wu, J. Zheng, L. Shen, and A. Anpalagan, "Dynamic spectrum access in time-varying environment: Distributed learning beyond expectation optimization," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5305–5318, Dec. 2017.
- [213] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, and Y.-D. Yao, "Opportunistic spectrum access in unknown dynamic environment: A game-theoretic stochastic learning solution," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1380–1391, Apr. 2012.
- [214] Y. Wang, G. Feng, Y. Sun, S. Qin, and Y.-C. Liang, "Decentralized learning based indoor interference mitigation for 5G-and-beyond systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12124–12135, Oct. 2020.
- [215] X. Shen, X. Wang, Q. Xu, W. Ge, and X. Xue, "Towards scalable and fast distributionally robust optimization for data-driven deep learning," in *Proc. IEEE Int. Conf. Data Mining*, 2022, pp. 448–457.
- [216] S. K. Das and R. Mudi, "A location-aware resource efficiency and energy efficiency optimization for M2M communications over downlink cellular systems," in *Proc. IEEE 31st Annu. Int. Symp. Pers., Indoor Mobile Radio Commun.*, 2020, pp. 1–6.
- [217] M. C. Lucas-Estañ and J. Gozalvez, "Mode selection for 5G heterogeneous and opportunistic networks," *IEEE Access*, vol. 7, pp. 113511–113524, 2019.
- [218] X. Zhang, M. Peng, S. Yan, and Y. Sun, "Deep-reinforcementlearning-based mode selection and resource allocation for cellular V2X communications," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6380–6391, Jul. 2020.
- [219] G. Wang, S. Zhou, and Z. Niu, "Mode selection in UAV-aided vehicular network: An evolutionary game approach," in *Proc. 10th Int. Conf. Wireless Commun. Signal Process.*, 2018, pp. 1–6.
- [220] G. Wang, C. Wu, T. Yoshinaga, W. Bao, and R. Yin, "Deep reinforcement learning based mode selection for coexistence of D2D-U and Wi-Fi," in *Proc. Int. Conf. Inf. Commun. Technol. Disaster Manage.*, 2021, pp. 9–14.
- [221] H. Xiang, S. Yan, and M. Peng, "A deep reinforcement learning based content caching and mode selection for slice instances in fog radio access networks," in *Proc. IEEE 90th Veh. Technol. Conf.*, 2019, pp. 1–5.
- [222] S. K. Das and R. Mudi, "Kalman filter based NLOS identification and mitigation for M2M communications over cellular networks," in *Proc. IEEE 93rd Veh. Technol. Conf.*, 2021, pp. 1–6.
- [223] L. Feng, Z. Yang, Y. Yang, X. Que, and K. Zhang, "Smart mode selection using online reinforcement learning for VR broadband broadcasting in D2D assisted 5G hetnets," *IEEE Trans. Broadcast.*, vol. 66, no. 2, pp. 600–611, Jun. 2020.
- [224] S. K. Das and R. Mudi, "A location-aware energy-efficient scheme for M2M communications by resource efficiency optimization," in *Proc. IEEE Int. Conf. Adv. Netw. Telecommun. Syst.*, 2020, pp. 1–6.
- [225] S. K. Das and M. F. Hossain, "TDOA based localization architecture for M2M communications over cellular networks," in *Proc. 10th Int. Conf. Elect. Comput. Eng.*, 2018, pp. 333–336.
- [226] S. K. Das and M. F. Hossain, "A location aware communication mode selection mechanism for M2M communications over cellular networks," in *Proc. 5th Int. Conf. Adv. Elect. Eng.*, 2019, pp. 335–340.
- [227] S. K. Das and M. F. Hossain, "A distance based communication mode selection mechanism for M2M communications over cellular networks," *Wireless Pers. Commun.*, vol. 115, no. 3, pp. 2501–2514, 2020.

- [228] S. K. Das and M. F. Hossain, "A location-aware power control mechanism for interference mitigation in M2M communications over cellular networks," *Comput. Elect. Eng.*, vol. 88, 2020, Art. no. 106867.
- [229] X. Chi, X. Xu, S. Han, and J. Zhang, "Learning-based cooperative multiplexing mode selection and resource allocation for eMBB and uRLLC," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2022, pp. 836–841.
- [230] D. Wang, H. Qin, B. Song, K. Xu, X. Du, and M. Guizani, "Joint resource allocation and power control for D2D communication with deep reinforcement learning in MCC," *Phys. Commun.*, vol. 45, 2021, Art. no. 101262.
- [231] P. H. Barros, I. Cardoso-Pereira, L. Foschini, A. Corradi, and H. S. Ramos, "Load balancing in D2D networks using reinforcement learning," in *Proc. IEEE Symp. Comput. Commun.*, 2019, pp. 1–6.
- [232] S. K. Das, M. Siddikur Rahman, L. Mohjazi, M. A. Imran, and K. M. Rabie, "Reinforcement learning-based resource allocation for M2M communications over cellular networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2022, pp. 1473–1478.
- [233] Y. Liu, J. Yan, and X. Zhao, "Deep-reinforcement-learning-based optimal transmission policies for opportunistic UAV-aided wireless sensor network," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 13823–13836, Aug. 2022.
- [234] I. Ioannou, C. Christophorou, V. Vassiliou, and A. Pitsillides, "A distributed AI/ML framework for D2D transmission mode selection in 5G and beyond," *Comput. Netw.*, vol. 210, 2022, Art. no. 108964.
- [235] S. K. Das, "Design of M2M communications over cellular networks with localization-assisted power control and communication mode selection mechanism," M.S. thesis, Dept. Elect. Electron. Commun. Eng., Mil. Inst. Sci. Technol., Dhaka, Bangladesh, 2020.
- [236] S. K. Das and R. Mudi, "Range-based location estimation of machines in M2M communications over cellular networks," in *Proc. Int. Joint Conf. Comput. Intell.*, M. S. Uddin and J. C. Bansal, Eds., Singapore: Springer Nature, 2020, pp. 693–703.
- [237] H. Yang, X. Xie, and M. Kadoch, "Intelligent resource management based on reinforcement learning for ultra-reliable and low-latency IoV communication networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4157–4169, May 2019.
- [238] B. Sheen, J. Yang, X. Feng, and M. M. U. Chowdhury, "A deep learning based modeling of reconfigurable intelligent surface assisted wireless communications for phase shift configuration," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 262–272, 2021.
- [239] S. Alkaraki, A. L. Borja, J. R. Kelly, R. Mittra, and Y. Gao, "Reconfigurable liquid metal-based SIW phase shifter," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 1, pp. 323–333, Jan. 2022.
- [240] H. He, S. Jin, C.-K. Wen, F. Gao, G. Y. Li, and Z. Xu, "Model-driven deep learning for physical layer communications," *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 77–83, Oct. 2019.
- [241] C. Wang et al., "Adaptive semantic-bit communication for extended reality interactions," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 5, pp. 1080–1092, Sep. 2023.
- [242] L. U. Khan, Z. Han, D. Niyato, M. Guizani, and C. S. Hong, "Metaverse for wireless systems: Vision, enablers, architecture, and future directions," *IEEE Wireless Commun.*, vol. 31, no. 4, pp. 245–251, Aug. 2024.
- [243] L.-H. Lee et al., "All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda," 2021, arXiv:2110.05352.
- [244] S. Duan et al., "Distributed artificial intelligence empowered by endedge-cloud computing: A survey," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 591–624, Firstquarter 2023.



SREE KRISHNA DAS (Student Member, IEEE) received the B.Sc. degree in electrical and electronic engineering from Metropolitan University, Sylhet, Bangladesh, and the M.Sc. degree in electrical, electronic, and communication engineering from MIST, situated in Mirpur Cantonment, Dhaka, Bangladesh. His research interests include state-of-the-art technologies including 6G networks, integrated sensing and communications, IoT and M2M communications, reconfigurable intelligent surfaces, XL-MIMO, and federated learn

ing applications in wireless communications. He was a Member of the TPC and also a Reviewer of numerous prestigious international journals and conferences.



RATNA MUDI received the Bachelor of Science degree in computer science and engineering from the University of Asia Pacific, Dhaka, Bangladesh, and the Master of Science degree in computer science from the Jahangirnagar University, Savar, Bangladesh, respectively. From 2018 to 2022, she was a Lecturer with the Department of Computer Science and Engineering, World University of Bangladesh, Dhaka, Bangladesh. Her research interests include, but are not limited to, resource management, federated learning, M2M-IoT com-

munications, and wireless sensor networks.



MD. SIDDIKUR RAHMAN (Student Member, IEEE) received the B.Sc. degree in electrical and electronic engineering (EEE) from American International University-Bangladesh (AIUB), Dhaka, Bangladesh, in 2014. He is currently working toward the M.Sc. degree with the Department of Electrical and Electronics Engineering, Universiti Teknologi PETRONAS (UTP), Seri Iskandar, Perak, Malaysia. From 2014 to 2017, he was with the EEE Department, Ahsanullah University of Science and Technology (AUST), Dhaka, and the

Physics Department, Jagannath University (JnU), Dhaka, Bangladesh, as a Volunteer Research Assistant. Since 2017, he has been a Faculty Member with the Department of Electrical and Electronic Engineering, World University of Bangladesh (WUB), Dhaka, Bangladesh, with the rank of Lecturer. Since 2022, he has been a Graduate Research Assistant (GRA) and Teaching Assistant (TA) with the Department of Electrical and Electronics Engineering, UTP. His research interests include power quality, power electronics, power system analysis, renewable energy sources, microgrids, biomass energy, the IoT, and federated reinforcement learning for wireless communication.



KHALED M. RABIE (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in electrical and electronic engineering from the University of Manchester, Manchester, U.K., in 2011 and 2015, respectively. He is currently a Distinguished Associate Professor with the Computer Engineering Department, King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia. He was a part of several large-scale industrial projects and has authored or coauthored more than 200 journal and conference articles (mostly IEEE).

His research interests include designing and developing next-generation wireless communication systems. He is also on the Technical Program Committee (TPC) of several major IEEE conferences, such as GLOBECOM, ICC, and VTC. He was the recipient of many awards over the past few years in recognition of his research contributions including the Best Paper awards at the 2021 IEEE CITS and the 2015 IEEE ISPLC, and the IEEE ACCESS Editor of the month Award for August 2019. He is the Editor of IEEE COMMUNICATIONS LETTERS and *IEEE Internet of Things Magazine*. He is also a Fellow of the U.K. Higher Education Academy (FHEA) and Fellow of the European Alliance for Innovation (EAI).



XINGWANG LI (Senior Member, IEEE) received the M.Sc. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2010, and the Ph.D. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2015. From 2010 to 2012, he was an Engineer with Comba Telecom Ltd., Guangzhou, China. From 2017 to 2018, he spent one year as a Visiting Scholar with Queen's University Belfast, Belfast, U.K. He is currently an Associated Professor with the School of Physics

and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo China. His research interests include wireless communication, intelligent transport system, artificial intelligence, and the Internet of Things. He was also a TPC Members, such as IEEE ICC and IEEE GLOBECOM. He is on the Editorial Board of IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON VEHICULAR TECH-NOLOGY, IEEE SYSTEMS JOURNAL, IEEE SENSORS JOURNAL, and *Physical Communication*. He was a Guest Editor of the Special Issue on Computational Intelligence and Advanced Learning for Next-Generation Industrial IoT of IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING and "AI driven Internet of Medical Things for Smart Healthcare Applications: Challenges, and Future Trends" of the IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS.