AN OVERVIEW ON OVER-THE-AIR FEDERATED EDGE LEARNING

Xiaowen Cao, Zhonghao Lyu, Guangxu Zhu, Jie Xu, Lexi Xu, and Shuguang Cui

ABSTRACT

Over-the-air federated edge learning (Air-FEEL) has emerged as a promising solution to support edge artificial intelligence (AI) in future, beyond 5G (B5G) and 6G networks. In Air-FEEL, distributed edge devices use their local data to collaboratively train AI models while preserving data privacy, in which the over-the-air model/gradient aggregation is exploited for enhancing the learning efficiency. This article provides an overview of the Air-FEEL state-of-the-art. First, we present the basic principle of Air-FEEL, and introduce the technical challenges for Air-FEEL design due to the over-the-air aggregation errors as well as the resource and data heterogeneities at edge devices. Next, we present the fundamental performance metrics for Air-FEEL, and review resource management solutions and design considerations for enhancing the Air-FEEL performance. Finally, several interesting research directions are pointed out to motivate future work.

INTRODUCTION

Recently, edge learning or edge intelligence has been recognized as a promising technique toward beyond 5G (B5G) and 6G to enable emerging applications such as meta-verse, smart city, automated driving, and industrial automation, in which artificial intelligence (AI) is relocated from the central cloud to the network edge to allow prompt data access and exploitation with significantly reduced latency. Among a variety of edge learning paradigms, federated edge learning (FEEL) is particularly appealing, due to its advantages in efficient model training and data privacy preservation by exploiting the distributed computing and storage capabilities at network edge [1]. Essentially, FEEL is implemented based on the distributed stochastic gradient descent (SGD) over wireless networks, which aims to find optimized AI-model parameters by minimizing a properly designed loss function in an iterative manner. To be specific, at each global iteration or communication round, the edge server first broadcasts the global model to edge devices for synchronizing their local models; next, the edge devices update their respective local models by performing one or more local gradient descent iterations; finally, the edge devices upload their local gradients or models to the edge server for aggregation, thus yielding an updated global model to initiate the next-round training. Based on the distributed SGD principle, FEEL can be implemented via two typical approaches, namely the federated stochastic gradient descent (FedSGD) and federated averaging (FedAvg), in which edge devices upload the local gradients and models after one and multiple local iterations, respectively, at each communication round [2]. However, due to the frequent exchange of high-dimensional models or gradients (usually comprising millions to billions of parameters) between the edge server and edge devices, the communication cost becomes the paramount performance bottleneck for FEEL, especially when the number of participating edge devices becomes large.

To tackle this issue, over-the-air FEEL (Air-FEEL) has been proposed recently as a promising solution, in which distributed edge devices are allowed to concurrently transmit their local gradient or model updates for “one-shot” aggregation at the edge server in each communication round. In such a way, the communication and computation are seamlessly integrated to enhance the communication efficiency in FEEL, thus supporting various low-latency and communication-efficient applications, such as Internet of vehicles and intelligent community, as shown in Fig. 1. The magic behind the promising Air-FEEL applications is the so-called over-the-air computation (AirComp) technique, which utilizes the superposition property of a multiple-access channel (MAC) for distributed functional computation over the air [3]. With AirComp, the communication efficiency in Air-FEEL is enhanced as compared to the conventional FEEL based on the classic communicate-then-compute approach for gradient/model aggregation. This is because that the sharing of the whole spectrum by all devices overcomes the issue of long latency faced by massive access protocol, which enables Air-FEEL to achieve remarkable communication latency reduction up to a factor equal to the device population. Moreover, Air-FEEL is able to further enhance the data privacy by combating the potential eavesdropping attack during the gradient/model uploading process. This is due to the fact that with AirComp, potential eavesdroppers can only access to the...
aggregated updates with each local component hidden in the crowd [4]. However, it is inevitable that the random perturbation caused by the fading and noise in the aggregation process may slow down the convergence speed and thus pose a fundamental trade-off between the convergence rate and privacy.

Due to the aforementioned advantages, Air-FEEL is becoming a hot research topic in the field of edge intelligence, which has attracted growing research interests from both academia and industry. However, the successful implementation of Air-FEEL to materialize its promising gain faces various technical challenges resulting from the over-the-air aggregation errors as well as the resource and data heterogeneities at distributed edge devices. Various research efforts have been conducted in the literature to tackle these issues.

In view of the growing research interests, this article provides an overview on the state of the art of Air-FEEL. First, we introduce the basic principle of Air-FEEL, and discuss its fundamental performance metrics in terms of training loss, learning accuracy, learning latency, and energy consumption. Next, we review various solutions to enhance the Air-FEEL performance, covering power control, beamforming, user scheduling, and large-scale network optimization. Finally, we point out several interesting research directions to motivate future work on Air-FEEL. It is worth noting that in contrast to prior works [1, 3], this article provides a more profound and in-depth review for Air-FEEL in terms of performance metrics, fundamental challenges, and performance limits.

**Basic Principle of Air-FEEL**

This section introduces the basics of Air-FEEL. A typical Air-FEEL system is shown in Fig. 2, where one edge server coordinates multiple distributed edge devices to train an AI model based on the “one-shot” over-the-air model/gradient aggregation. As mentioned earlier in the introduction, over-the-air FedSGD (Air-FedSGD) and over-the-air FedAvg (Air-FedAvg) are two typical approaches for Air-FEEL implementation. In particular, at each global communication round, Air-FedSGD employs the over-the-air gradient aggregation after one local computation iteration, while Air-FedAvg performs the over-the-air model aggregation after multiple local computation iterations. With the different aggregation rules, Air-FedSGD generally needs more iterations to converge but enjoys better robustness to the non-independent and identically distributed (i.i.d.) data, while Air-FedAvg requires much less frequent model aggregation and thus lower communication cost.

The main difference of Air-FEEL from other FEEL paradigms is that, Air-FEEL exploits AirComp for one-shot over-the-air model/gradient aggregation, as explained in the following. For example, consider the basic setup with one edge server and K edge devices in an additive white Gaussian noise (AWGN) channel. Consider a particular global communication round, in which each device $k$ needs to upload its local gradient/model parameter $w_k \in \mathbb{C}^{N \times 1}$, and the desired global parameter at the edge server is $(1/K) \sum_{k=1}^{K} w_k$. In this case, after acquiring the information of its own channel coefficient $h_k$, each device $k$ multi-

FIGURE 1. Example application scenarios of Air-FEEL.
The state-of-the-art methods with proper pre- and post-processing, AirComp can go beyond averaging to compute a series of nonmonotonic functions such as geometric mean, weighted sum, polynomial, and Euclidean norm (see [3] and the references therein).

The benefits of Air-FEEL are two-fold. First, Air-FEEL is advantageous in enhancing the communication efficiency and reducing the training latency. In particular, in Air-FEEL, the over-the-air aggregation of each AI-model parameter from $K$ edge devices only requires one single channel use, thus making the required total number of resource blocks and the training latency independent from $K$. This is in sharp contrast to the conventional design with sequential “communication-then-computation,” in which the edge server needs to decode individual AI-model parameters from $K$ devices by mitigating the inter-user interference, such that the required total number of resource blocks and the training latency generally grow with respect to $K$ linearly.

Next, Air-FEEL is beneficial in better preserving data privacy. In conventional FEEL, raw data exposure to the third party is avoided by exchanging model/gradient updates; however, private data information can still be inferred from the individual updates via membership inference or update leakage attacks [4], thus causing certain level of privacy leakage. By contrast, in Air-FEEL, only the aggregate updates, instead of individual ones, can be accessed by potential eavesdroppers, thus providing double privacy protection at edge devices.

**Challenges**

Due to the employment of AirComp and FEEL, Air-FEEL faces their respective technical challenges such as CSI acquisition and synchronization for AirComp, and data heterogeneity for FEEL, which can be resolved efficiently by well-established approaches in the AirComp and FEEL literatures. Nevertheless, the successful implementation of Air-FEEL also faces two unique issues as follows.

**Over-the-Air Aggregation Errors:** In Air-FEEL, analog magnitude modulation is implemented for one-shot distributed functional computation. This, however, is vulnerable to signal distortion caused by channel fading, noise, and channel estimation errors, and thus may degrade the training performance. How to quantify the effect of such over-the-air aggregation errors on Air-FEEL performance and accordingly optimize the system design is an uncharted problem that has not been investigated in conventional FEEL and wireless communication systems.

**Resource Limitation and Heterogeneity:** While the distributed learning tasks are generally resource-consuming, the participating edge devices only have limited computation and communication resources, which the above process can be implemented over each subchannel independently. In particular, for Air-FEEL, the over-the-air aggregation of each AI-model parameter from $K$ edge devices only requires one single channel use, thus making the required total number of resource blocks and the training latency independent from $K$. This is in sharp contrast to the conventional design with sequential “communication-then-computation,” in which the edge server needs to decode individual AI-model parameters from $K$ devices by mitigating the inter-user interference, such that the required total number of resource blocks and the training latency generally grow with respect to $K$ linearly.

Next, Air-FEEL is beneficial in better preserving data privacy. In conventional FEEL, raw data exposure to the third party is avoided by exchanging model/gradient updates; however, private data information can still be inferred from the individual updates via membership inference or update leakage attacks [4], thus causing certain level of privacy leakage. By contrast, in Air-FEEL, only the aggregate updates, instead of individual ones, can be accessed by potential eavesdroppers, thus providing double privacy protection at edge devices.
resources, and their available resources may vary significantly. Such resource limitation and heterogeneity are becoming a challenging issue faced in Air-FEEL. On one hand, the over-the-air gradient/model aggregation needs to be implemented synchronously by all edge devices after their local iteration, but their local iteration latencies may differ significantly due to their distinct computation capabilities. In this case, the total computation latency of Air-FEEL is restrained by the device with the smallest computation power. On the other hand, the over-the-air aggregation errors highly depend on the channel conditions and the transmit power at distributed edge devices. In this case, devices with poor channel and/or limited transmit power become the performance bottleneck of AirComp, and increasing their transmit powers can accelerate the Air-FEEL training but at a cost of high energy consumption cost.

**FUNDAMENTAL PERFORMANCE LIMITS**

To facilitate the Air-FEEL design, it is important to understand its fundamental performance limits. This section first presents the widely adopted performance metrics in Air-FEEL, and then discusses their fundamental trade-offs under practical over-the-air aggregation errors.

**Performance Metrics**

Air-FEEL aims to train AI models. The widely adopted performance metrics include training loss and learning accuracy, learning latency, and energy consumption. It is worth noting that most performance metrics for FEEL are also applicable and important for Air-FEEL. In addition, Air-FEEL also introduces some unique issues to tackle, that is, the aggregation error and its impact on the convergence and privacy protection.

**Training Loss and Learning Accuracy:** Training loss means the loss (or objective) function value of the trained AI model, and learning accuracy describes how the trained model performs over test samples. In general, with a given AI model, lower training loss corresponds to higher learning accuracy. To analyze the convergence behavior of Air-FEEL, loss optimality gap (the gap between the current loss and the global minimum) is widely adopted in practice [4–6].

**Learning Latency:** Learning latency is defined as the wall clock time for the learning process to converge within a given accuracy. In Air-FEEL, the learning latency generally corresponds to the number of global communication rounds times per-round latency [5]. Here, the per-round latency includes the computation latency (of the slowest device) for local gradient/model-update iterations, and the communication latency for over-the-air updates aggregation from the edge devices to the edge server.

**Energy Consumption:** Energy consumption corresponds to the communication and computation energy required by both the edge devices and edge server for training the AI model. The computation energy is mainly consumed by the edge devices for locally updating AI models, and the communication energy is consumed by the edge devices and edge server to exchange their models/gradients.

Intuitively, the above three Air-FEEL performance objectives are competing, for example, to achieve higher accuracy normally requires longer learning latency and more energy consumption. Such performance trade-offs, however, are quite complicated and depend on various system design parameters, such as the numbers of global and local iterations, the transmit power at edge devices for AirComp, and their computation power for local computation. In the literature, there have been well-established analytic models on the learning latency and energy consumption for Air-FEEL [3]. It remains a fundamental problem to analytically understand the relationship between the training loss and learning accuracy versus system parameters under different AI models, especially subject to the unique over-the-air aggregation errors.

**Convergence of Air-FEEL with Aggregation Errors**

To resolve the above issue, it is essential to analyze the convergence behavior of Air-FEEL, which characterizes how the (expected) gradient norm or loss optimality gap changes over global communication rounds. However, it is difficult.
To find the exact expression for gradient norm or loss optimality gap under general AI models. To tackle this problem, their upper bounds are alternatively used to represent the convergence behavior in the FEEL literature [4], by making specific assumptions on the AI model (e.g., on the smoothness of the loss functions and Lipschitz-continuous gradients). However, due to the newly involved over-the-air aggregation errors, the convergence analysis for Air-FEEL is still challenging.

Recently, the work [6] established a convergence analysis for Air-FedSGD with gradient aggregation after one local iteration. This framework reveals the impact of the bias and MSE of the aggregation error at each global communication round $t$. First, the Air-FEEL algorithm is observed to converge toward the optimality within a certain error floor, whose value is the weighted sum of the bias $\mathbb{E}[e(t)]$ of the aggregation error. The error floor becomes zero if the over-the-air aggregation is unbiased (i.e., $\mathbb{E}[e(t)] = 0$, $\forall t$), and it becomes non-zero if it is biased (i.e., $\mathbb{E}[e(t)] \neq 0$, $\forall t$). Next, the biasedness of the over-the-air aggregation also affects the convergence behaviors of Air-FEEL. For example, enforcing unbiased aggregation for vanishing error floor may lead to the exact convergence to the optimal point, but slow down the convergence. Furthermore, it is observed that with the increasing discounting factor, the aggregation errors at the later global rounds contribute more to the optimality gap and thus are more damaging to the learning process.

The above analytical result in [6] sheds light on how the imperfect aggregate updates affect the convergence of Air-FedSGD. This result was then extended to Air-FedAvg with model aggregation after multiple local iterations [5]. It is shown that, as the number of local iterations increases, the needed number of global communication rounds may decrease, but at the cost of a large divergence among the local updates at edge devices. This thus shows the significance of judiciously designing the number of local iterations in Air-FedAvg for improving the learning performance. These analytical results build the foundation for optimizing the Air-FEEL performance, as will be elaborated in the sequel.

**Resource Management for Air-FEEL**

Based on the performance analysis above, this section reviews the state-of-the-art techniques for improving the Air-FEEL performance, which jointly optimizes the communication parameters for Air-Comp and the learning parameters.

**Power Control Optimization**

In Air-FEEL, the edge devices can adaptively control their transmit power to reduce the aggregation error for model/gradient aggregation, and accordingly enhance the learning accuracy or convergence rate. In general, there are two different transmit power control design principles. The first design principle optimizes Air-Comp independently at each global round for, for example, minimizing the communication error or MSE. For instance, the most straightforward way is for the edge devices to adopt the channel inversion power control for achieving the perfect signal magnitude alignment. This, however, is highly sub-
optimal in MSE minimization. This is because in this case, the aligned signal magnitude is compromised by the device with the worst channel conditions, which may significantly weaken the aggregate signal strength and equivalently amplify the noise power, especially when some devices are in deep fading. To deal with this issue, the optimal power control to minimize the AirComp MSE was developed in [3, 7], which exhibits a threshold-based structure depending on the channel conditions at different edge devices. For a certain device, if the channel power gain is greater than a certain threshold, then a channel-inversion-like power control is employed at that device for signal magnitude alignment; otherwise, the full power transmission is performed at that device to mitigate the noise.

Instead of minimizing the aggregation MSE in each round independently, the second design principle jointly optimizes the power allocation over different global rounds to maximize the convergence rate in terms of the loss optimality gap (e.g., in the upper frame in Fig. 4). For instance, the work [6] investigated the power allocation problem in Air-FedAvg, in which two joint power control designs with biased and unbiased aggregations are developed, respectively. As shown in Fig. 5, both joint power control designs achieve faster convergence and lower loss value compared with the fixed power transmission and independent power control for MSE minimization. This thus shows the benefit of power control optimization in accelerating the learning convergence. As compared with independent power control for MSE minimization, the joint power control designs capture the fact that the loss optimality gap is more sensitive to the aggregation errors at latter rounds. Figure 5 also compares the loss value and learning accuracy of the two joint power control designs. It is shown that the design under the unbiased aggregation achieves lower optimality gap than that under the biased aggregation when the number of communication rounds is sufficiently large. This coincides with the previous discussion in that the Air-FedSGD algorithm converges to the optimal point with unbiased gradient aggregation. The convergence-aware joint power control design was then extended to the Air-FedAvg system in [5].

**Beamforming Design**

Multi-antenna beamforming is an efficient solution to further improve the Air-FEEL performance by enhancing both aggregation efficiency and aggregation accuracy, via exploiting the spatial multiplexing and spatial diversity gains, respectively. On one hand, by deploying multiple antennas at both the edge devices and the edge server, multiple AI-model parameters or gradient elements can be simultaneously aggregated over one single time-frequency block. On the other hand, via proper transmit and/or receive beamforming, multiple signal copies from different antennas can be properly combined to further reduce the aggregated errors. For instance, existing works in [3, 8] developed the joint design of transmit beamforming at the edge devices and aggregation beamforming at edge server for substantially enhancing the aggregation performance, which can be directly used for enhancing Air-FEEL performance.

More recently, reconfigurable intelligent surface (RIS) has emerged as a candidate technique to enhance the spectrum and energy efficiencies of 6G networks [9]. RIS is composed of a large number of low-cost and passive reflecting elements, whose amplitudes and phases can be adjusted to properly control the propagation environment to facilitate wireless transmissions. In particular, RIS can assist the edge devices with poor channel conditions to enhance their signal strengths toward the edge server, thus enhancing the aggregated signal magnitude level for minimizing the aggregation error. For RIS-empowered Air-FEEL, it is critical to jointly control the phase shifts of reflecting elements at RIS, the transmit beamforming at edge devices, and the receive beamforming at edge server for enhancing the learning performance [9].

**User Selection/Scheduling**

Use selection and scheduling is another efficient technique to enhance the Air-FEEL performance by resolving the resource heterogeneity issue via dropping the edge devices with bad channels and/or limited communication and computation resources. A new trade-off arises in scheduling edge devices for optimizing Air-FEEL performance. In particular, allowing more devices to participate in Air-FEEL can aggregate more training data for accelerating the convergence, which, however, may lead to larger aggregation errors as more devices with poorer channels are involved. Such trade-offs were first studied in [10] to regulate the device population in the scenario when the edge devices are i.i.d. distributed in a certain area, in which the devices with weak channel conditions were excluded. Next, the joint design of device scheduling and receive beamforming was presented in [8], in which the devices with weak signal strengths after receive beamforming were dropped from the training process. Moreover, the work [11] investigated the device scheduling by further...
considering their diverse energy constraints and computation capabilities, in which an energy-aware dynamic device scheduling algorithm was proposed based on Lyapunov optimization.

**Model/Gradient Compression Based on Sparsification**

Compression is useful in reducing the data dimensions to be exchanged, and thus is important to accelerate the communication process in Air-FEEL. However, different from FEEL that can apply conventional digital compression techniques, Air-FEEL needs to implement new compression methods based on analog modulated model/gradient parameters. For example, the authors in [12] exploited the sparsity of the gradient/model updates to compress the local gradient estimates to the dimension of available channel uses.

In this design, edge devices first sparsify the gradient estimates by setting elements with small absolute values to zero, and then projected the processed gradient estimates into a low-dimensional vector by applying a synchronized pseudo-random measurement matrix. This thus avoids the additional transmission of index information of zero elements in gradient vector. Furthermore, the work [13] considered gradient sparsification together with an error-feedback mechanism, in which the common dimensions can be selected uniformly via synchronized pseudo-random number generators and the sparsification errors are added back to the next-round gradient update to alleviate the possible gradient error. Under this setup, the adaptive power control was implemented to combat the aggregation errors over time-varying channels.

**Future Research Directions**

Besides the above resource management designs, there also remain various new design issues that are important for the success of Air-FEEL in 6G, but not well investigated yet. In this section, we discuss some of these issues to motivate future work.

**Secure Design Against Jamming and Spoofing**

The analog over-the-air aggregation in Air-FEEL makes it vulnerable to jamming or spoofing attacks due to the lack of coding and encryption. For instance, some corrupted devices due to malfunction or malicious adversary (a.k.a. Byzantine devices) may send fault or manipulated messages to the edge server to attack Air-FEEL for preventing the convergence or leading to a counterfeit model. In this case, how to detect or even correct the manipulated abnormal model/gradient updates by malicious devices (i.e., poisoning attacks) on global model training is important. This task, however, is particularly challenging, due to the simultaneous model/gradient transmission over the air.

To tackle the above issue, the authors in [14] proposed to use the Weiszfeld algorithm to obtain the smoothed geometric median of model/gradient for resisting the Byzantine attack. In this design, the edge devices are divided into several groups, based on which the devices within each group can employ AirComp for computing the geometric median of their local models/gradients, but those in different groups need to transmit over orthogonal time-frequency blocks. Based on such an over-the-air geometric median aggregation scheme, it is shown that when the number of Byzantine attacks
is lower than half of the number of groups or individual transmissions, the aggregation values are ensured to achieve a well approximation to the mean of updates of honest devices. In essence, the design in [14] combats against the Byzantine attacks at the cost of more communication costs. How to further improve such design and study new approaches to combat other attacks still requires more efforts from our community.

**AIR-FEEL OVER LARGE-SCALE NETWORKS**

The success of FEEL needs massive training data. It is thus desirable to involve more edge devices over large-scale networks for AIR-FEEL. Toward this end, hierarchical AIR-FEEL and device-to-device (D2D) AIR-FEEL can be two viable architectures, as depicted in the two lower sub-figures in Fig. 6. In particular, the hierarchical AIR-FEEL architecture may correspond to a multi-cell network with multiple edge servers and one (or more) cloud server at the core network. At the lower level, the edge server may aggregate local models/gradadients from edge devices over the air; at the upper level, the cloud server performs global aggregation of regional updates from edge servers via wired or wireless backhauls. How to associate edge devices with edge servers, and how to determine the numbers of local computation iterations at the lower and upper levels are new problems to tackle. These problems, however, are difficult, especially when the distributed data samples follow varied distribution at edge devices.

On the other hand, in the D2D AIR-FEEL without centralized servers, all participating edge devices are divided into multiple clusters, such that each device is associated with a cluster head via D2D communications [15]. In this system, edge devices within the same cluster first simultaneously send the updates to the associated cluster head, and then the head broadcasts the aggregated (cluster) signal to all devices to achieve consensus. In this case, how to cluster the edge devices and how to determine cluster heads are important problems.

**AIR-FEEL WITH NEW ENERGY TECHNOLOGIES**

AI model training is generally energy consuming, and the resultant energy cost and carbon emission are becoming serious problems. While conventionally resource management solutions are implemented at the energy demand side, exploiting new energy technologies such as energy harvesting (from, e.g., solar, wind, and ambient radio signals) and smart grids at the energy supply side is becoming a viable new solution to resolve the above problem, as shown in the upper-left subfigure in Fig. 6. However, the energy harvesting processes are random and intermittent in nature and the smart grids enable various new features such as dynamic energy pricing and two-way energy trading and sharing. Sophisticated joint energy and AIR-FEEL design is desirable for future research.

In addition, wireless power transfer (WPT) is another powering technique over wireless networks, which can be utilized for the edge server to wirelessly charge edge devices to enable their sustainable and even battery-free operation. For such wireless powered AIR-FEEL systems, how to jointly optimize the WPT transmission and the AIR-FEEL operation is a new technical issue.

**AIR-FEEL WITH INTEGRATED SENSING AND COMMUNICATIONS**

Integrated sensing and communications (ISAC) has been recognized as an enabling 6G technology at the physical layer to enable wireless sensing over cellular networks for supporting new intelligent applications. ISAC integrates both functions of data collection and transmission, and its combination with AIR-FEEL can be exploited to provide an ultra-low-latency “sensing-communication computation” solution for edge intelligence. For example, as shown in the upper-right subfigure in Fig. 6 distributed edge devices can simultaneously transmit probing signals to detect the target and data symbols to the edge server for update aggregation via AirComp. However, the implementation of integrated AIR-FEEL and ISAC also faces many challenges. First, since such networks need to support wireless sensing and update aggregation at the same time, the already scarce spectrum resources may become even more strained. Next, different types of edge devices generate different data modalities, and thus how to apply multi-modal sensing data to improve the performance of AIR-FEEL deserves more in-depth investigation.

**CONCLUSION**

This article presented an overview on AIR-FEEL that exploits the advanced over-the-air data aggregation or AirComp to enhance the communication efficiency for FEEL toward edge intelligence. We discussed the basic principles, design challenges, advanced techniques, and various new research issues of AIR-FEEL. It is our hope that this article can provide new insights on this interesting research topic, and motivate more interdisciplinary research from the communities of wireless communications, machine learning, and computing.

**ACKNOWLEDGMENT**

The work was supported in part by NSFC with Grant No. 62293482, the Basic Research Project HZQB-KCZYX-2021067 of Hetao Shenzhen-HK S&T Cooperation Zone, the National Key R&D Program of China with grant No. 2018YFB1800800, the Shenzhen Outstanding Talents Training Fund 2020020, the Guangdong Research Projects No. 2017ZT07X152 and No. 2019CX01X104, the Guangdong Provincial Key Laboratory of Future Networks of Intelligence (Grant No. 2022B1212010001), the Shenzhen Key Laboratory of Big Data and Artificial Intelligence (Grant No. ZDSYS20170725140905), and the Key Area R&D Program of Guangdong Province with grant No. 2018B030338001, the NSFC with grants Nos. U20031208, 92267202 and 62001310, the Shenzhen Fundamental Research Program under grant No. JCYJ20210324133405015, the Guangdong Basic and Applied Basic Research Foundation under Grant 2022A1515010109, and the Internal Project fund from Shenzhen Research Institute of Big Data under Grants J00120230001.

**REFERENCES**

[1] M. Chen et al., “Distributed Learning in Wireless Networks: Recent Progress and Future Challenges,” IEEE JSAC, vol. 39, no. 12, Dec. 2021, pp. 3579–3605.

[2] B. McMahan et al., “Communication-Efficient Learning of Deep Networks From Decentralized Data,” Proc. Int’l Conf. Artificial Intelligence and Statistics, vol. 54, Fort Lauderdale,
Future Network of Intelligence Institute, The Chinese University of Hong Kong, Shenzhen, China, and a visiting scholar at the Shenzhen Research Institute of Big Data, Shenzhen, China. His research interests include data driven large-scale system control and resource management, large dataset analysis, IoT system design, energy-harvesting-based communication system design, and cognitive network optimization. He is currently serving as an Area Editor for IEEE Signal Processing Magazine and an Associate Editor for several IEEE Transactions.

Guangchu Zhu [M] (gxzhu@sribd.cn) received a Ph.D. degree in electrical and electronic engineering from The University of Hong Kong in 2019. He is now a research scientist at the Shenzhen research institute of big data, associated with the Chinese University of Hong Kong, Shenzhen. His research interests include edge intelligence, machine learning in wireless communication, federated learning, integrated sensing and communication, and edge intelligence.

Jie Xu [SM] (jxue@cuhk.edu.cn) is currently an Associate Professor with the School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen, China. His research interests include wireless communications, wireless information and power transfer, UAV communications, edge computing and intelligence, and integrated sensing and communication (ISAC). He served or is serving as an Editor of the IEEE Trans. Wireless Communications, IEEE Trans. Communications, IEEE Wireless Commun. Letters, and J. Communications and Information Networks.

Lei Xu [SM] (davidleix@hotmail.com) received an M.S. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2009, and the Ph.D. degree from the Queen Mary University of London, London, U.K., in 2013. He is currently a Senior Engineer with Research Institute, China Unicom delegate in ITU, ETSI, 3GPP, and CCSA. His research interests include big data, self-organizing networks, satellite system, and radio resource management in wireless system.

Shuguang Cui [F] (shuguangcui@cuhk.edu.cn) is the Chair Professor of the Chinese University of Hong Kong at Shenzhen and the Vice Director at Shenzhen Research Institute of Big Data. His current research interests focus on data driven large-scale system control and resource management, large dataset analysis, IoT system design, energy-harvesting-based communication system design, and cognitive network optimization. He is currently serving as an Area Editor for IEEE Signal Processing Magazine and an Associate Editor for several IEEE Transactions.