FLISC^3: Federated Learning-Oriented Resource Optimization in ISCC-Enabled Edge Collaborative Networks

An Du, Jie Jia, *Member, IEEE*, Schahram Dustdar, *Fellow, IEEE*, Andrea Morichetta, *Member, IEEE*, Jian Chen, *Member, IEEE*, and Xingwei Wang

***Abstract*—Federated edge learning (FEEL) greatly facilitates the development of ubiquitous intelligence by combining federated learning and edge computing. However, traditional FEEL implementations assume fixed-sized local datasets, neglecting the potential of edge devices to acquire sensory information actively. Such a simplistic scenario leads to overestimating data availability and underestimating resource utilization in networks with varying resource capacity. Moreover, the existing FEEL-oriented systems with integrated sensing, communication, and computation (ISCC) have separate-based designs, leading to an inefficient use of wireless resources. To alleviate these issues, we propose a novel FEEL-oriented ISCC framework in edge collaborative networks, by leveraging the integrated sensing and communication (ISAC) technique to achieve the dual purpose of data sensing and parameter transmission. Then, over the designed framework, we present FEEL convergence analysis under non-independent and identically distributed (non-iid) and iid data. Correspondingly, we formulate a joint beamforming and flexible time duration optimization problem to maximize the convergence speed of FEEL, subject to limited resources on the devices and requirements for data sensing and communication. To address the problem efficiently, we propose an alternative optimization framework, in which the successive convex approximation (SCA) method is adopted to solve the nonconvex beamforming design subproblem, and a low-complexity method is derived for optimal time allocation. Extensive results reveal that the proposed framework can achieve excellent performance in model training accuracy by efficiently utilizing limited resources in edge collaborative networks, under iid and non-iid data.**

***Index Terms*—Alternative optimization, beamforming, federated edge learning, integrated sensing, communication, and computation (ISCC).**

I. Introduction

THE upcoming sixth-generation (6G) wireless networks and unprecedented advances in artificial intelligence (AI) techniques are expected to provide opportunities for a multitude of applications, including smart cities, health care, and industrial Internet of Things [1]. In particular, the synergy between mobile edge computing (MEC) technologies and AI enables a major transformation of the AI paradigm from cloud AI to edge AI [2]. As such, data communication efficiency can be significantly improved, and latency-critical intelligence applications can also be supported by leveraging resources in proximity to the end devices. Nonetheless, deploying a centralized AI model at the edge is both energy- and communication-inefficient, as the edge node becomes a bottleneck; furthermore, such a configuration presents privacy-related risks [3]. Thanks to modern hardware capabilities, also small devices can compute complex logic, making distributed AI at the edge a viable solution to delivering AI services [4]. Federated learning (FL), a distributed learning paradigm, enables multiple data owners to collaborate in model training without revealing their local data, thus reducing communication overheads and avoiding privacy issues. Deploying FL in MEC drives the development of federated edge learning (FEEL), which consists of a set of resource-constrained end devices that possess personal data and participate in collaborative learning, and a server that coordinates the learning process [5]. During each training round, each participating end device downloads the global model from the server and uses its own dataset for training. After that, the server collects the new parameters from all participants and performs a global aggregation. Herein, FEEL enables the computing and networking resources dispersed on various edge devices to be leveraged for ubiquitous intelligence in the 6G network [6].

However, there are still many issues in supporting FEEL, including channel fading in the wireless system and resource

Received 16 April 2025; revised 30 September 2025; accepted 12 October 2025. Date of publication 15 October 2025; date of current version 5 February 2026. This work was supported in part by the Major Research Plan of National Natural Science Foundation of China under Grant 92167103, in part by the National Natural Science Foundation of China under Grant 62172084, Grant 62132004, Grant 62032013, Grant 62572109, and Grant 62572113, in part by the Aeronautical Science Foundation of China under Grant 20230026050001, in part by the Young and Middle-Aged Leading Talents in Technological Innovation of Shenyang under Grant RC231173, in part by the Key Research and Development Program of LiaoNing under Grant 2023JH2/101300196, in part by Shenyang Science and Technology Plan Fund Project under Grant 23-503-6-17, in part by the Fundamental Research Funds for the Central Universities under Grant N2416007, Grant N2416001, Grant N2424010-18, Grant N2416006, and Grant N2216009, and in part by China Scholarship Council under Grant 202406080078. *(Corresponding author: Jie Jia.)*

An Du, Jie Jia, Jian Chen, and Xingwei Wang are with the School of Computer Science and Engineering, the Engineering Research Center of Security Technology of Complex Network System, and the Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education, Northeastern University, Shenyang 110819, China (e-mail: 2110663@stu.neu.edu.cn; jiajie@mail.neu.edu.cn; chenjian@mail.neu.edu.cn; wangxw@mail.neu.edu.cn).

Schahram Dustdar is with the Distributed Systems Group, TU Wien, 1040 Vienna, Austria, and also with ICREA, UPF Barcelona, 08010 Barcelona, Spain (e-mail: dustdar@dsg.tuwien.ac.at).

Andrea Morichetta is with the Distributed Systems Group, TU Wien, 1040 Vienna, Austria (e-mail: a.morichetta@dsg.tuwien.ac.at).

This article has supplementary downloadable material available at https://doi.org/10.1109/TSC.2025.3622026, provided by the authors.

Digital Object Identifier 10.1109/TSC.2025.3622026

limitation of edge devices, which may cause transmission errors, model accuracy degradation, and high training costs. Therefore, efficient resource allocation is critical for the MEC platform to achieve high-performance FL. To this end, numerous FEEL architectures and communication technologies have been developed to minimize the FL loss function [7], maximize FL convergence rate [8], and maximize the transmission efficiency [9]. Unlike traditional FEEL systems that rely on existing datasets for model training, e.g., users' browsing history information and social media content, the future 6G networks are committed to creating an AI-native environment to achieve ambient intelligence [10]. Take smart inspection in the intelligent industry as an example: massive sensing nodes undertake data acquisition to offer high-accuracy and real-time status of the production lines, seamless communication helps rapid interaction in terms of data and feature information, and distributed computing power accomplishes data analysis and feature extraction. Therefore, a multi-functional platform with integrated sensing, communication, and computation (ISCC) is necessary to support the emerging 6G intelligent applications, such as the FEEL.

Taking advantage of the dual functionality design of radar sensing and data communications simultaneously, integrated sensing and communication (ISAC) is gaining popularity as a 6G network solution, as it can be implemented on devices with small size, lightweight, and low-cost [11]. Until now, to improve spectral efficiency and network performance, several works have been devoted to involving ISAC-enabled devices in ISCC networks [12], [13], [14]. However, most of the network architectures adopt the separation-based design, which aims to simultaneously satisfy the requirements of all processes in isolation. Conversely, FEEL in the future 6G era calls for deep integration in network design, in which the network functions in sensing, communication, and computation must be symbiotic to complete a common learning task [15]. Specifically, the edge devices need to sense the environment and generate data samples concurrently, update local models in parallel, and upload local models for global aggregation. The quality and quantity of acquired sensing data are restricted by the limited resources on devices and deeply affect the FEEL performance. Compared to traditional ISCC systems with fixed dataset assumptions of FEEL, the unified framework can more precisely evaluate network performance and flexibly control network resources, thereby alleviating the issues of overestimating data availability and resource wastage in networks with scarcity and sufficient resources, respectively.

To break through the concern of separating sensing information acquisition from model training, a new FEEL-oriented ISCC system needs to be investigated. Meanwhile, the system performance metrics of interest will transform into learning performance. However, there still remain many challenges to be addressed: How to unleash the potential of the ISCC system to improve FEEL performance? How to explicitly quantify the impact of such a system on FEEL performance? How to achieve efficient multidimensional network resources management to exploit the ISCC benefits? For these purposes, we propose a novel FEEL-oriented ISCC framework in edge collaborative networks (FLISC^3), in which the ISAC technology is leveraged as a spectrum-efficient solution to achieve the dual purpose of dataset collection and model transmission. In particular, each device transmits the independent radar waveform and communication symbol, forming multiple beams toward the target and the edge server, respectively. During each FEEL iteration, each device senses the target for local training and simultaneously transmits the local model-updates generated at the prior iteration, thus efficiently accelerating the training process. Within the proposed architecture, we aim to achieve an enhanced learning performance by jointly optimizing beamforming and time duration allocation. The main contributions of this paper are summarized as follows.

- To support high-performance FEEL in resource-constrained edge networks, we propose a novel FEEL-oriented framework by involving horizontal collaboration among multiple edge devices and vertical collaboration across the triple-function of sensing, communication, and computation. In addition, we derive the FEEL convergence rate both under non-independent and identically distributed (non-iid) and iid data in the FLISC^3 system. Based on the analysis, we formulate a joint resource allocation problem to maximize the sensing and communication (S&C) duration under the requirements of delay, energy consumption, and sensing performance.
- To tackle the problem efficiently, we first analyze the relations between the coupled control variables and develop an algorithm to alternatively optimize transmit precoding and time duration allocation. Specifically, given the S&C time duration, we provide semidefinite relaxation (SDR) and an iterative solution based on the successive convex approximation (SCA) method for the joint beamforming design. Then, the closed-form expression of optimal S&C time duration can be derived given the transmit precoding.
- Numerical results demonstrate that the proposed FLISC^3 can provide successful and high-performance FEEL services by unleashing the potential of the collaborative edge system, compared with the other three traditional ISCC training frameworks. Additionally, unlike the positive performance gain under iid, increased devices degrades convergence performance under severe data heterogeneity.

The remainder of the paper is organized as follows. In Section II, we provide related works on resource management for FEEL and ISCC networks, respectively. In Section III, we introduce the system model. In Section IV, we provide the expected convergence rate of FL in our ISCC system and give the problem formulation. In Section V, we provide the problem analysis and the optimization framework. In Section VI, we provide extensive simulations and contrastive analysis of numerical results. Finally, Section VII concludes this paper.

Notation: We use boldface letters in lowercase and uppercase to denote column vectors and matrices, respectively. For the complex-valued vector $\mathbf{x}$, $\|\mathbf{x}\|$ represents the corresponding Euclidean norm. Superscripts $(\cdot)^H$ and $(\cdot)^T$ represent Hermitian transpose and transpose, respectively. Moreover, $\mathrm{Tr}(\cdot)$, $\mathrm{rank}(\cdot)$, and $\det(\cdot)$ are the trace operation, rank operator, and matrix determinant, respectively. $\mathbf{I}_N$ denotes the $N \times N$ identity matrix.

$\mathbb{C}^{N\times N}$ denotes the complex space of $N \times N$. $\mathbb{E}(\cdot)$ denotes the statistical expectation.

II. Related Work

To provide the best support of FL in wireless edge networks, resource allocation for FEEL has been widely investigated, with a focus on the perspectives in communication and computation. Moreover, with the requirements of obtaining real-time environment information, wireless sensing is further considered, thus driving brand-new research in ISCC architecture design and corresponding resource management for supporting different computing tasks.

A. Resource Management for FEEL

Considering the resource bottleneck of the edge environment, a group of existing works has paid plenty of attention to improving the performance of FEEL, which mainly investigated it from two perspectives. In particular, several works focused on enhancing FL algorithms to address the challenges of FL in edge computing. For example, considering the device heterogeneity and resource scarcity, the authors in [16] and [17] proposed efficient network pruning-based FL systems, while the authors in [18] and [19] studied the joint resource allocation and device selection problem in wireless networks. Moreover, to improve model aggregate efficiency, importance-based updating and gradient quantization technologies were adopted in [8] and [20], respectively. Considering the model staleness and unacceptable waiting time caused by heterogeneous devices, the authors in [21] proposed a blockchained dual-asynchronous federated learning service model in the digital twin-empowered edge-cloud continuum. On the other hand, a wide spectrum of studies have invoked various technologies and proposed different resource management schemes to improve the MEC system, thereby offering great potential to support FL. For example, to address the straggler problem caused by user resource heterogeneity, the authors in [22] invoked a hybrid non-orthogonal multiple access (NOMA)- time division multiple access (TDMA) scheme. Moreover, over-the-air federated learning (AirFL) is developed as a promising solution to reduce communication and computation overheads as well as shorten the training latency by exploring the superposition property of wireless channels [23].

Although existing research efforts have effectively improved the performance of wireless networks as well as the efficiency of FL, they generally assumed a readily available dataset on each device and ignored the ability of devices to actively sense the environment. This cannot align with the development of future AI-native networks and may significantly affect the learning performance due to the highly coupled functions.

B. Resource Management in ISCC Network

As a potential key technology in 6G networks, ISAC has drawn significant attention due to its flexible control and efficient performance guarantee over the beamforming process. In the most recent ISCC system, edge computing is leveraged to reduce the computation burden for processing sensor data. For example, the authors in [12] considered an ISCC architecture with multiple user terminals, and each of them performs radar sensing and computation offloading simultaneously by using multiple-input and multiple-output (MIMO) arrays. The authors in [24] constructed an ISCC system and adopted the orthogonal frequency division multiplexing (OFDM)-based integrated radar and communication waveform for surroundings sensing and computation offloading. In addition, the authors in [25] introduced an ISCC-based task offloading and resource allocation framework for vehicular cooperation perception. Considering the reliability requirements of the offloading process, the authors in [13] further introduced short-packet transmissions in the MIMO-based ISCC network. Different from the assumptions that the sensing and communication are integrated on edge devices, the authors in [26] and [27] investigated an ISCC system where a triple-functional base station (BS) is invoked to perform a sensing task, and also provide communication and computing services.

However, the ISCC systems mentioned above were designed in a task-agnostic manner, which poses a great challenge in satisfying the strict requirements of emerging AI applications. For this purpose, an edge intelligence-oriented ISAC was investigated in [28] and [29]. Specifically, the authors in [28] investigated an ISAC-accelerated edge intelligence system, while the authors in [29] adopted a cooperative multi-station sensing scheme to improve model training accuracy. As for edge AI inference, the authors in [30] proposed an edge-device co-inference oriented ISCC scheme with the help of over-the-air computation (AirComp), and the authors in [31] further investigated the mode selection in edge inference to make full use of the different paradigms' benefits.

Different from the edge inference that concentrates on computation efficiency, FEEL pays more attention to the learning performance via device collaboration and resource management. However, the research on the FL-oriented ISCC system is still in its infancy, and target sensing is generally viewed as an add-on module in most existing literature. For example, the authors in [32] proposed a unified ISCC framework to complete over-the-air FL and target estimation, and the authors in [33] further conducted the FL convergence analysis by considering the downlink transmission error. To push the AI-native network, FEEL requires the support of sensing for data acquisition, thereby constituting a data delivery pipeline. For this purpose, the authors in [34] proposed an FEEL-oriented ISCC framework, where sensing and communication signals occupy orthogonal time-frequency resources. Considering the energy bottleneck of mobile devices, the authors in [35] further investigated the sustainable FL by leveraging wireless power transfer. However, the completely sequential execution of all functions inevitably causes inefficiency in resource consumption due to the lack of flexibility in the system. To this end, we investigated the joint resource optimization in FLISC^3, in which the ISAC is leveraged to enhance system performance under stringent resource constraints. To the best of our knowledge, this problem has not been yet investigated.

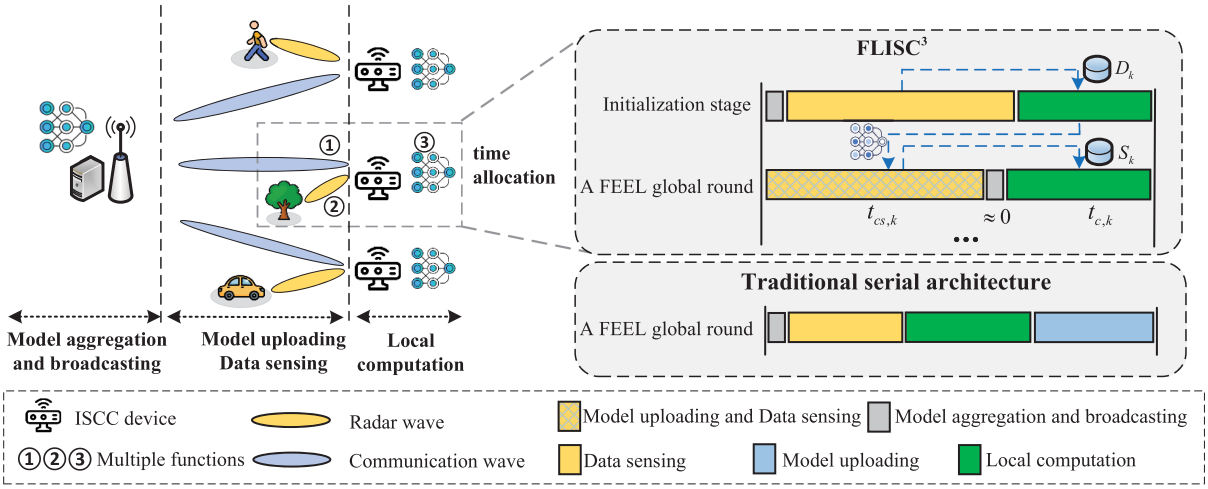


Fig. 1. Illustration of the system model and time allocation architecture for ISCC device k.

TABLE I
IMPORTANT NOTATIONS IN OUR MODEL

Symbol	Definition
$\mathcal{K}$	The set of ISCC devices, $k \in \mathcal{K}$.
N	The number of antennas on each ISCC device.
$\mathbf{x}_k$	The transmit signal vector of ISCC device k.
y	The received communication signal of BS.
$\mathbf{y}_{s,k}$	The received radar echo of ISCC device k.
$\mathbf{h}_k$	The channel vector from ISCC device k to the BS.
$\mathbf{H}_{i,k}$	The interference channel matrix from ISCC device i to k.
$\mathbf{w}_{c,k}$	The precoding vector for data communication of k.
$\mathbf{W}_{s,k}$	The precoding matrix for radar sensing of k.
$I_{s,k}$	The sensing mutual information of ISCC device k.
r_k	The communication data rate of ISCC device k.
$t_{cs,k}$	S&C time duration of device k.
$t_{c,k}$	Time duration of device k for local training .
T_k	The total time overhead of device k for one global iteration.
E_k	The energy consumption of device k for one global iteration.
ϕ_k	The required number of CPU cycles per sample.
f_k	The computing capacity of ISCC device k.
D_k	The maximum number of sensing data of device k.
S_k	The number of actual sensing data of device k.

III. SYSTEM MODEL

As shown in Fig. 1, we consider an FEEL-oriented and ISCC-enabled edge collaborative network, which incorporates a single BS and K ISCC devices indexed by $\mathcal{K} = \{1, \ldots, K\}$ to train an AI model cooperatively. The BS is equipped with a single antenna and is connected to an edge server. Each ISCC device is equipped with an antenna array with N antennas for simultaneously performing radar detection (yellow connection) and parameter uploading (blue connection). Some key parameter notations are summarized in Table I.

A. FEEL Model

The time allocation architectures of FLISC^3 and the traditional serial FEEL system are illustrated in Fig. 1, in which the training process consists of multiple communication rounds. At the initialization stage in FLISC^3, the BS broadcasts the global FL model with model parameters $\boldsymbol{g}_t$ to all ISCC devices. After receiving the global parameters, the ISCC devices leverage all power resources to sense surroundings simultaneously. The reflected signal from each target can be extracted as a set of radar data samples. Let D_k represent the maximum number of the collected samples of ISCC device k by leveraging all power resources for radar sensing. Then, each device executes local FL model updates with the acquired data and its own computing resources. Each subsequent training round mainly consists of three steps as follows.

- Simultaneous local model uploading and data sensing: Instead of executing sensing and communication sequentially in traditional design, the sensing and communication stages are merged in FLISC^3 to fully utilize the wireless signals. Specifically, each ISCC device performs communication and sensing simultaneously by sharing the same antenna array and leveraging the multi-beam-based transmitter. In particular, each device transmits the local FL model $\boldsymbol{w}_k$ updated at the last round to the BS, and transmits radar waveforms for target sensing. Then, a certain amount of radar data can be obtained by processing the received radar echo signals. Here, we denote the set of acquired sensing data by the device k as $\mathcal{S}_k = \{\boldsymbol{x}_{k,i}, y_{k,i}\}_{i=1}^{S_k}$, where S_k is the total number of samples. $\boldsymbol{x}_{k,i}$ and $y_{k,i}$ are the feature vector and label of the i_{th} sample, respectively.
- Global model aggregation and broadcast on BS: After receiving the local FL models from all ISCC devices, the global FL model on BS will be updated according to

$$\boldsymbol{g} = \sum_{k=1}^{K} \frac{S_k}{S} \boldsymbol{w}_k, \tag{1}$$

where $S = \sum_{k=1}^{K} S_k$ is the total samples perceived by all ISCC devices. Due to the abundant power resources on BS, the time and energy consumption for global parameter aggregation and broadcast are omitted [18], [34].
- Local gradient update: With the acquired data samples $\mathcal{S}_k$ and the received global model in the current round, each ISCC device can use the stochastic gradient descent (SGD) method to update its local FL model.

B. Transmit Signal Design

Considering the dual functionalities enabled by joint beamforming, the transmit signal of the ISCC device k can be represented as a weighted sum of communication symbols and radar waveforms. At the time slot n, the transmit signal is given by

$$\mathbf{x}_k[n] = \mathbf{w}_{c,k} c_k[n] + \mathbf{W}_{s,k}\mathbf{s}_k[n], \tag{2}$$

where $c_k[n]$ is the communication symbol to be transmitted to the BS, the $N \times 1$ vector $\mathbf{s}_k[n] = [s_1[n], \ldots, s_N[n]]^T$ includes N individual radar waveforms. $\mathbf{w}_{c,k} \in \mathbb{C}^{N\times 1}$ and $\mathbf{W}_{s,k} \in \mathbb{C}^{N\times N}$ are the precoders for communication and radar, respectively. To facilitate the analysis, we make the following assumptions [12], [27]: 1) Both radar and communication signals are zero-mean, temporally white, and wide-sense stationary stochastic process; 2) The communication symbols are uncorrelated with the radar waveforms, i.e., $\mathbb{E}(\mathbf{s}_k c_j^H) = \mathbf{0}_{N\times 1}, \forall k, l \in \mathcal{K}$; 3) Communication symbols satisfy $\mathbb{E}(c_k c_k^H) = 1$ and $\mathbb{E}(c_k c_j^H) = 0, \forall k \neq l$; 4) Radar waveforms for all ISCC devices are uncorrelated with each other, i.e., $\mathbb{E}(\mathbf{s}_k \mathbf{s}_k^H) = \boldsymbol{I}_N$ and $\mathbb{E}(\mathbf{s}_k \mathbf{s}_j^H) = \mathbf{0}_N, \forall k \neq l$.

Let $\mathbf{R}_k$ denote the covariance of the transmit waveform for ISCC device k, which is given by

$$\mathbf{R}_k = \mathbb{E}\left(\mathbf{x}_k[n]\mathbf{x}_k^H[n]\right) = \mathbf{w}_{c,k}\mathbf{w}_{c,k}^H + \mathbf{W}_{s,k}\mathbf{W}_{s,k}^H. \tag{3}$$

Hence, the transmit power for ISCC device k is calculated as $\mathrm{Tr}(\mathbf{R}_k)$.

C. Radar Sensing Model

To acquire sensing data samples with satisfactory quality for training, we adopt sensing mutual information (SMI) to evaluate the radar sensing performance, which aims to extract the environmental information in the target response from the received echo signal [36]. Moreover, we assume that the communication signals can be used for radar sensing since the radar receiver of each ISCC device knows exactly the transmitted communication symbols [12], [29]. For ISCC device k, considering that a single point-like target in the far-field is located at direction θ_k from it, its received radar echo is given by

$$\mathbf{y}_{s,k}[n] = \beta_k \boldsymbol{a}_k(\theta_k)\boldsymbol{a}_k^T(\theta_k)\mathbf{x}_k[n] + \sum_{i=1, i\neq k}^{K} \mathbf{H}_{i,k}\mathbf{x}_k[n] + \mathbf{v}_k[n], \tag{4}$$

where $\beta_k \in \mathbb{C}$ is the complex reflection coefficient depending on the radar cross section (RCS) and path loss between the device and target. $\boldsymbol{a}_k = \frac{1}{N}\left[1, e^{j2\pi\triangle \sin(\theta_k)/\lambda}, \ldots, e^{j2\pi(N-1)\triangle \sin(\theta_k)/\lambda}\right]^T \in \mathbb{C}^{N\times 1}$ is the array steering vector of direction, where $\triangle$ is the antenna spacing and λ denote the signal wavelength. In addition, $\mathbf{H}_{i,k} \in \mathbb{C}^{N\times N}$ is the interference channel coefficient from ISCC device i to k, and $\mathbf{v}_k[n]$ is additive zero-mean temporally-white noise with $\mathbf{v}_{s,k}[n] \sim \mathcal{CN}(0, \sigma_s^2 \mathbf{I}_N)$.

We assume that each ISCC device can accurately track the target's position. Therefore, SMI measures the information in reflection coefficient β from the reflected echo signal $\mathbf{y}_{s,k}$. With the knowledge of $\mathbf{x}_k[n]$, the SMI for ISCC device k can be given by

$$\begin{aligned} I_{s,k}[n] &= I\left(\beta_k; \mathbf{y}_{s,k}[n] | \mathbf{x}_k[n]\right) \\ &= \log_2 \det\left(\mathbf{I}_N + \mathbf{H}_k\left(\mathbf{w}_{c,i}\mathbf{w}_{c,i}^H + \mathbf{W}_{s,i}\mathbf{W}_{s,i}^H\right)\mathbf{H}_k^H \mathbf{N}_{s,k}^{-1}\right), \end{aligned} \tag{5}$$

where $\mathbf{H}_k = \beta_k \boldsymbol{a}_k(\theta)\boldsymbol{a}_k^T$. In addition, $\mathbf{N}_{s,k}^{-1}$ is the interference-plus-noise, which is expressed as

$$\mathbf{N}_{s,k} = \sum_{i\neq k}^{K} \mathbf{H}_{i,k}\left(\mathbf{w}_{c,i}\mathbf{w}_{c,i}^H + \mathbf{W}_{s,i}\mathbf{W}_{s,i}^H\right)\mathbf{H}_{i,k}^H + \sigma_s^2 \mathbf{I}_N. \tag{6}$$

D. Communication Model

Considering the model parameters of multiple ISCC devices are transmitted simultaneously via uplink, the received signal at the BS is given by

$$y[n] = \sum_{k=1}^{K} \mathbf{h}_k^T \mathbf{x}_k + v_c[n], \tag{7}$$

where $\mathbf{h}_k \in \mathbb{C}^{N\times 1}$ is the communication channel between ISCC device k and BS, and $v_c[n] \sim \mathcal{CN}(0, \sigma_c^2)$ is the additive white Gaussian noise (AWGN). According to (2), the received signal of BS from ISCC device k is given by

$$y_k[n] = \mathbf{h}_k^T \mathbf{w}_{c,k} c_k[n] + \sum_{i=1}^{K} \mathbf{h}_i^T \mathbf{W}_{s,i}\mathbf{s}_i[n] + \sum_{i=1, i\neq k}^{K} \mathbf{h}_i^T \mathbf{w}_{c,i} c_i[n] + v_c[n], \tag{8}$$

where the first part is the desired communication signal from device k, the second part is the interface radar signal from all devices, and the third part is the interference communication signal from other devices.

Correspondingly, the data communication rate of ISCC device k can be written as

$$r_k = B\log_2\left(1 + \frac{|\mathbf{h}_k^T \mathbf{w}_{c,k}|^2}{n_{c,k} + \sigma_c^2}\right), \tag{9}$$

where B is the allocated bandwidth and $n_{c,k} = \sum_{i\neq k}^{K} |\mathbf{h}_i{}^T \mathbf{w}_{c,i}|^2 + \sum_{i=1}^{K} \|\mathbf{h}_i^T \mathbf{W}_{s,i}\|^2$ is the interference signal from other devices and local sensing.

E. Delay and Energy Consumption Model

After receiving the corresponding echoes reflected from the target during a period, the echoes are utilized to generate samples for model learning. Denote the allocated time duration for simultaneous sensing and communication as $t_{cs,k}$ and the required time for generating one sample as t_0. The amount of data samples can be calculated as $S_k = \frac{t_{cs,k}}{t_0}$. Accordingly, the computation delay for training S_k data samples on ISCC device k can be written as

$$t_{c,k} = \frac{S_k \phi_k}{f_k}, \tag{10}$$

where ϕ_k is the required number of CPU cycles per sample and f_k is the available CPU cycles per second on ISCC device k. In addition, as seen in Fig. 1, we consider the synchronous FEEL

training in this paper, where the aggregation and local model training won't start until the BS receives the local models from all ISCC devices. Therefore, the total time overhead of ISCC device k for one global iteration is given by

$$T_k = \max\left\{t_{cs,k}\right\}_{k\in\mathcal{K}} + t_{c,k}. \tag{11}$$

Moreover, during the training process, the energy consumption of each ISCC device is constituted by the transmission energy consumption during radar sensing and parameter uploading, and computation energy consumption for local model training, which can be expressed as

$$E_{cs,k} = t_{cs,k}\mathrm{Tr}(\mathbf{w}_{c,k}\mathbf{w}_{c,k}^H + \mathbf{W}_{s,k}\mathbf{W}_{s,k}^H), \tag{12}$$

and

$$E_{c,k} = \kappa S_k \phi_k f_k^2, \tag{13}$$

respectively. $\kappa > 0$ is the energy efficiency parameter determined by the structure of devices. Correspondingly, the overall energy consumption of ISCC device k for one global iteration can be given by

$$E_k = E_{cs,k} + E_{c,k}. \tag{14}$$

IV. FEEL Convergence Analysis and Problem Formulation

Due to resource competition on ISCC devices during sensing, communication, and computing stages, the resource management scheme deeply impacts the acquisition of sensing data and FEEL performance. Therefore, we can always consider $S_k \ll D_k, \forall k \in \mathcal{K}$. With the local sensing data samples, each device updates the local model by using the SGD method to minimize training loss, which can be expressed as

$$\boldsymbol{w}_{k,t+1} = \boldsymbol{g}_t - \frac{\lambda}{S_k}\sum_{i=1}^{S_k}\nabla f\left(\boldsymbol{g}_t\left(\mathbf{w}_{c,k}, \mathbf{W}_{s,k}, t_{cs,k}\right), \boldsymbol{x}_{k,i}, y_{k,i}\right), \tag{15}$$

where λ is the learning rate, f is loss function, and ∇f is the gradient of f with respect to $\boldsymbol{g}_t$. In addition, the global model is updated as

$$\boldsymbol{g}_{t+1} = \frac{\sum_{k=1}^K S_k \boldsymbol{w}_{k,t+1}}{\sum_{k=1}^K S_k}. \tag{16}$$

Denote $F(\boldsymbol{g}) = \frac{1}{D}\sum_{k=1}^K\sum_{i=1}^{D_k} f(\boldsymbol{g}, \boldsymbol{x}_{k,i}, y_{k,i})$, according to (15) and (16), the global model update can be rewritten as

$$\boldsymbol{g}_{t+1} = \boldsymbol{g}_t - \lambda(\nabla F(\boldsymbol{g}) - \boldsymbol{o}), \tag{17}$$

where $D = \sum_{k=1}^K D_k$ is the maximum number of acquired data samples by using all time-frequency resources for radar sensing, and $\boldsymbol{o} = \nabla F(\boldsymbol{g}) - \frac{\sum_{k=1}^K\sum_{i=1}^{S_k}\nabla f(\boldsymbol{g}_t, \boldsymbol{x}_{k,i}, y_{k,i})}{\sum_{j=1}^K S_k}$.

A. Convergence Analysis With IID Data in Strongly Convex Case

To facilitate the analysis of the excepted convergence rate, we first make the following assumptions [7]:

- *Assumption A1:* The overall gradient of $F(\boldsymbol{g})$ is uniformly Lipschitz continuous, i.e.

$$\|\nabla F(\boldsymbol{g}_{t+1}) - \nabla F(\boldsymbol{g}_t)\| \le L\|\boldsymbol{g}_{t+1} - \boldsymbol{g}_t\|. \tag{18}$$

- *Assumption A2:* The loss function F is strongly convex with positive parameter μ:[1]

$$\|\nabla F(\boldsymbol{g}_{t+1}) - \nabla F(\boldsymbol{g}_t)\| \ge \mu\|\boldsymbol{g}_{t+1} - \boldsymbol{g}_t\|. \tag{19}$$

- *Assumption A3:* The loss function F is twice-continuously differentiable. According to (18) and (19), we have

$$\mu\boldsymbol{I} \preceq \nabla^2 F(\boldsymbol{g}) \preceq L\boldsymbol{I}. \tag{20}$$

- *Assumption A4:* We assume the loss $f(\boldsymbol{g}_t, \boldsymbol{x}_{k,i}, y_{k,i})$ calculated on the i_{th} samples and k_{th} device satisfies[2]

$$\|\nabla f(\boldsymbol{g}_t, \boldsymbol{x}_{k,i}, y_{k,i})\|^2 \le \zeta_1 + \zeta_2\|\nabla F(\boldsymbol{g}_t)\|^2, \tag{21}$$

in which ζ_1 and ζ_2 are two non-negative constants [7], [16].

Under the above assumptions, the expected convergence rate of the FEEL model satisfies the following theorem.

Theorem 1. Given the beamforming design and the S&C time duration of all ISCC devices, the expected optimality gap of FEEL with setting learning rate $\lambda = \frac{1}{L}$ is bounded by

$$\mathbb{E}\left[F(\boldsymbol{g}_{t+1}) - F(\boldsymbol{g}^*)\right] \le A^{t+1}\mathbb{E}\left[F(\boldsymbol{g}_0) - F(\boldsymbol{g}^*)\right] + \frac{2R\zeta_1}{LD^2}\frac{1-A^{t+1}}{1-A}, \tag{22}$$

where $R = (D-S)^2$ and $A = \frac{4\mu R\zeta_2}{LD^2} - \frac{\mu}{L} + 1$.

Proof. The detailed proof is shown in Appendix A. □

Proposition 1. Given the learning rate $\lambda = \frac{1}{L}$, to guarantee the convergence of FL, ζ_2 must satisfy the following condition

$$0 < \zeta_2 < \frac{D^2}{\max 4R}. \tag{23}$$

Proof. In **Theorem 1**, when $A \ge 1$, the FEEL algorithm will not converge, while when $A < 1$, we have $A^{t+1} \to 0$ as $t \to \infty$. Hence, the FL converges with convergence residual $\frac{2R\zeta_1}{LD^2(1-A)}$. To guarantee $A < 1$, $\frac{4\mu R\zeta_2}{LD^2} - \frac{\mu}{L} + 1 < 1$ needs to be ensured. Based on *Assumption A3* in (20), we can derive $\mu \le L$. Correspondingly, $\zeta_2 < \frac{D^2}{4R}$ needs to be satisfied. To achieve this goal, $\zeta_2 < \frac{D^2}{\max 4R}$ needs to be achieved. Based on (21), we have $\zeta_2 > 0$. This completes the proof. □

From Proposition 1, we see that the FEEL convergence depends on the parameters related to the approximation of $\|\nabla F(\boldsymbol{g}_t)\|^2$, and the convergence can be ensured only if ζ_2 satisfies the constraint (23). Under the iid data assumption, the divergence between local gradient and global gradient is small, thus causing a small value of ζ_2 [37]. Therefore, there always exists ζ_1 and ζ_2 that satisfy the constraint in (23) and the assumption in (21), thus we remove the constraint in the following analysis. Additionally, in **Theorem 1**, $\boldsymbol{g}_{t+1}$ is the global FL model based on the sensed data samples of all ISCC devices at learning iteration $t+1$, and $\boldsymbol{g}^*$ is the optimal global FL model obtained by leveraging the ideal datasets without wireless limitations. In (22), there exists a gap, $\frac{2R\zeta_1}{LD^2}\frac{1-A^{t+1}}{1-A}$, between $\mathbb{E}[F(\boldsymbol{g}_{t+1})]$ and $\mathbb{E}[F(\boldsymbol{g}^*)]$, which is affected by wireless communication aspects, i.e., S&C time duration allocation. Specifically, subject to the

[1] This assumption can be fulfilled by many popular learning models, e.g., least-squared SVM and the linear regression.

[2] This bounded dissimilarity assumption can be seen as a generalization of the iid assumption with bounded dissimilarity, while allowing for statistical heterogeneity. Specifically, ζ_1 captures the inherent noise of the system even when the global model converged to the global optimum, and ζ_2 reflects dissimilarity among the local gradients.

successful communication and quality-guaranteed sensing, the gap will decrease as the sensing time increases. It can also be seen that the value of A decreases as the sensing time increases, thereby accelerating the FEEL convergence rate.

B. Convergence Analysis With IID Data in Non-Convex Case

In the deep learning tasks, non-convex loss functions are generally leveraged in FEEL, such as neural networks, which makes the strong convexity assumptions in (19) and (20) ineffective. Due to the non-convexity, SGD may converge to a local minimum or saddle point. Thus, we adopt the expected gradient norm as an indicator of convergence [38]. Based on *Assumption A1* and *A4, the expected convergence rate of the FEEL model is given in the following theorem.*

Theorem 2 *Let learning rate $\lambda = \frac{1}{L}$, the expected convergence rate of FEEL is bounded by*

$$\frac{1}{T+1}\sum_{t=0}^{T}\mathbb{E}\left[\|\nabla F(\boldsymbol{g}_t)\|^2\right] \leq \frac{2L}{d(T+1)}\mathbb{E}\left[F(\boldsymbol{g}_0)-F(\boldsymbol{g}^*)\right]+\frac{4\zeta_1\overline{R}}{dD^2}, \tag{24}$$

where $d = 1-4\zeta_2$ and $\overline{R}$ is the average sensing data difference during $T+1$ iterations.

Proof. *The detailed proof is shown in Appendix B.* □

Based on **Theorem 2**, the average l_2-norm of the global gradient is bounded by two terms. The first term is affected by the gap between the initial model and the optimal model, which converges to zero as $T \to \infty$. The second term is impacted by the sensing time allocation scheme. A higher value of $\overline{R}$ will result in a larger upper bound. The result revealed in (24) is consistent with the actual situation and the conclusion in **Theorem 1**, since allocating more sensing time can obtain more sensing samples, thus decreasing $\overline{R}$ and the upper bound. Moreover, under strong convexity, **Theorem 1** shows linear convergence with a rate of $\mathcal{O}(A^T), 0 < A < 1$, which means that the expected optimality gap converges at least as fast as a geometric series with A. Differently, **Theorem 2** reveals a convergence rate of $\mathcal{O}(1/T)$ to its corresponding stationary points. To achieve ϵ-accuracy, **Theorem 1** and **Theorem 2** require $O(\log(1/\epsilon))$ and $O(1/\epsilon)$ iterations, respectively, which reflects the slower convergence rate of **Theorem 2**.

C. Convergence Analysis With Non-IID Degree

Due to the dynamic environments in practical networks and heterogeneous sensing capabilities on devices, the sensing samples may exhibit non-iid among devices. Therefore, in this subsection, we present the convergence upper bound of FEEL with non-iid degree. Besides Lipschitz Continuity *Assumption A1* in (18), we also make the following assumptions:

- *Assumption C1:* Suppose $\nabla f_k(\boldsymbol{g}_t, \xi_k)$ is the stochastic gradient of device k on the sensing samples ξ_k, where $|\xi_k| = S_k$. The gradient is a device-level unbiased gradient, i.e.
$$\mathbb{E}\left[\nabla f_k(\boldsymbol{g}_t, \xi_k)\right] = \nabla f_k\left(\boldsymbol{g}_t\right), \tag{25}$$
where $\nabla f_k(\boldsymbol{g}_t)$ is unbiased gradient of device k.
- *Assumption C2:* The gradient $\nabla f_k(\boldsymbol{g}_t, \xi_k)$ of device k has bounded variance, i.e.,
$$\mathbb{E}\left[\|\nabla f_k(\boldsymbol{g}_t, \xi_k) - \nabla f_k\left(\boldsymbol{g}_t\right)\|^2\right] \leq \sigma_k^2, \tag{26}$$
where σ_k^2 is the variance bound of device k and in inverse proportion to the data size S_k.
- *Assumption C3:* To reflect the non-iid degree for non-convex objective functions, we assume $\nabla f_k(\boldsymbol{g}_t)$ has the following bounded variance:
$$\mathbb{E}\|\nabla f_k\left(\boldsymbol{g}_t\right) - \nabla F\left(\boldsymbol{g}_t\right)\|^2 \leq \Gamma^2. \tag{27}$$

These basic assumptions are widely adopted in the theoretical convergence analysis of FL systems with non-iid data [39], [40], [41]. Based on *Assumption A1* and *Assumption C1-C3*, the expected convergence rate of the FEEL model satisfies the following theorem.

Theorem 3. Let the learning rate $\lambda \leq \frac{1}{L}$ and T is the communication round, the convergence upper bound is

$$\frac{1}{T+1}\sum_{t=0}^{T}\mathbb{E}\left[\|\nabla F(\boldsymbol{g}_t)\|^2\right] \leq \frac{2}{(T+1)\lambda}\left(F(\boldsymbol{g}_0)-\mathbb{E}\left[F(\boldsymbol{g}^*)\right]\right) + \frac{L\lambda^2}{2}\mathbb{E}\left[\sum_{k=1}^{K}p_{k,t}\sigma_k^2\right]+\frac{\lambda\Gamma^2}{2}, \tag{28}$$

where $p_{k,t}$ is the aggregation weight, and we adopt the proportion of local sensing dataset size, i.e., $p_{k,t} = \frac{S_k}{\sum_{k=1}^{K} S_k}$.

Proof. The detailed proof is shown in Appendix C. □

From (28), it is evident that the first term is mathematically guaranteed to tend to zero with a sufficiently large communication round T. In addition, **Theorem 3** qualitatively reveals the impact of sensing data samples and different levels of non-iid data on the convergence performance. Specifically, increasing the sensing data size can reduce the gradient variance bound σ^2 of device k, thereby accelerating convergence. With respect to non-iid data measured by the parameter Γ^2, it is observed that Γ^2 has a negative impact on the model utility. Additionally, the convergence rate of $\mathcal{O}(1/T)$ in **Theorem 3** also exhibits sublinear convergence, which means that it requires $O(1/\epsilon)$ to achieve ϵ-accuracy. Therefore, **Theorem 3** illustrates the slower convergence rate to stationary points, compared to the linear convergence shown in **Theorem 1**.

Our convergence analyses across three cases reveal that the assumption of objective functions and data distributions has different impacts on the FEEL convergence. A critical commonality that emerges among them is that the convergence upper bounds are influenced by wireless network factors, especially the sensed data samples. The bound can be lowered by increasing the acquired data, which is determined by the S&C time duration. However, this variable is deeply coupled with the transmit beamforming design in the FLISC[3] system for successful data transmission. Therefore, to implement FEEL over a resource-aware wireless network, joint resource management is necessary to achieve the tradeoff between the convergence performance and network constraints.

D. Problem Formulation

To minimize the loss of FEEL training in such a system by unified resource management, we formulate an optimization problem to maximize the amount of sensing data, thus minimizing the convergence gap. Meanwhile, the constraints of the resources in multiple dimensions, the SMI requirements, and the successful transmission of parameters also need to be ensured. We thus formulate the problem as

$$(\mathbf{P0}): \max_{\{\mathbf{w}_{c,k}\},\{\mathbf{W}_{s,k}\},\mathbf{t}_{cs}} \sum_{k=1}^{K} \frac{t_{cs,k}}{t_0} \tag{29}$$

$$\text{s.t.} \quad \mathbf{R}_k = \mathbf{w}_{c,k}\mathbf{w}_{c,k}^H + \mathbf{W}_{s,k}\mathbf{W}_{s,k}^H, \forall k \tag{29a}$$

$$\text{Tr}\left(\mathbf{R}_k\right) \le P_k, \forall k \tag{29b}$$

$$I_{s,k}[n] \ge I_{s,k}^{min}, \forall k, \tag{29c}$$

$$r_k t_{cs,k} \ge Z(\mathbf{W}_k), \forall k \tag{29d}$$

$$E_k \le E_k^{max}, \forall k \tag{29e}$$

$$T_k \le \tau, \forall k, \tag{29f}$$

where $\mathbf{t}_{cs} = \{t_{cs,k}\}_{k=1}^{K}$ is a set of S&C time allocation for all ISCC devices, constraint (29b) is the power constraint of each ISCC device, constraint (29c) is the required SMI level for target sensing, constraint (29d) ensures successful transmission of the local FEEL model, constraint (29e) limits the total energy consumption on each device during each global iteration, and constraint (29f) represents the overall time constraint for each device during each global iteration.

Problem $(\mathbf{P0})$ is challenging to solve due to the multiplicative terms and non-convex functions in constraints (29c), (29d),(29e), (29f), thus causing standard convex optimization techniques not to be applied directly. Nonetheless, close observation of $(\mathbf{P0})$ shows that the objective function is only related to the S&C time duration allocation and its optimal value can be obtained once the optimal joint beamforming design is obtained. To this end, we develop an efficient optimization algorithm to decouple the problem into multiple parts and solve them alternatively in the next section.

V. Proposed Algorithm Design

A. Alternative Optimization Algorithm Design

In the problem $(\mathbf{P0})$, the S&C time duration should be maximized to accelerate the FEEL convergence speed by precisely adjusting other variables. For this purpose, we first analyze the impact of $\{\mathbf{w}_c\}, \{\mathbf{W}_s\}$ on $t_{cs,k}$. It can be seen that the upper bound of $t_{cs,k}$ for each ISCC device can be derived from the constraints (29e) and (29f), which is given by

$$UpperE_k = \frac{E_k^{max} t_0}{\text{Tr}\left(\mathbf{R}_k\right) t_0 + \kappa \phi_k f_k^2}, \forall k \tag{30}$$

and

$$UpperT_k = \frac{\left(\tau - \max\left\{t_{cs,k}\right\}_{k\in\mathcal{K}}\right) t_0 f_k}{\phi_k}, \forall k, \tag{31}$$

respectively. Then, the optimal $t_{cs,k}$ can thus be expressed as

$$t_{cs,k} = \min\left\{UpperE_k, UpperT_k\right\}, \tag{32}$$

which is determined by the beamforming design, i.e., $\{\mathbf{w}_c\}$ and $\{\mathbf{W}_s\}$. Correspondingly, our problem can be transformed into maximizing the sum of (32) for all ISCC devices while ensuring the constraints (29a)–(29d). However, due to the complex objective function and variable coupling among multiple devices, the problem is still non-convex and is intractable to be solved directly. By further analysis, we find that the problem might be energy-limited or delay-limited for different settings of FEEL environment parameters and beamforming design, thus leading to different time allocation solutions. Herein, the transmit beamforming should be optimized to maximize the upper bound in (30), and S&C time duration for multiple devices needs to be optimized to maximize the upper bound in (31). By this observation, the problem $(\mathbf{P0})$ can be decoupled into two subproblems as follows.

As seen in (30), minimizing $\text{Tr}(\mathbf{R}_k)$ can directly increases $t_{cs,k}$. Therefore, given the S&C time duration, the beamforming optimization problem can be written as

$$(\mathbf{P1}): \min_{\{\mathbf{w}_{c,k}\},\{\mathbf{W}_{s,k}\}} \sum_{k=1}^{K} \text{Tr}\left(\mathbf{R}_k\right) \tag{33}$$

$$\text{s.t.} \quad (29a) - (29d). \tag{33a}$$

Considering the interdependence among devices during sensing and communication, $\{\mathbf{w}_c\}$ and $\{\mathbf{W}_s\}$ should be optimized jointly by solving problem $(\mathbf{P1})$, and the corresponding algorithm is provided in Algorithm 2.

The other subproblem is to tune up S&C time duration for each ISCC device with the given beamforming design, which can be written as

$$(\mathbf{P2}): \max_{\{t_{cs,k}\}} \sum_{k=1}^{K} t_{cs,k} \tag{34}$$

$$\text{s.t.} \quad (29d) - (29f). \tag{34a}$$

To solve the problem with complicated constraint in (29f), we propose a low-complexity algorithm to obtain the closed-form of $t_{cs,k}$, as seen in Algorithm 3.

Overall, the suboptimal solution of $(\mathbf{P0})$ can be obtained by repeating alternatively optimizing the subproblem $(\mathbf{P1})$ and $(\mathbf{P2})$ until the stopping criterion is satisfied. The details of the alternative optimization-based algorithm are concluded in Algorithm 1.

B. Transmit Beamforming Design

In problem **(P1)**, both the SMI and model transmission rate constraints are non-convex. In this subsection, we tackle the problem by adopting an SDR strategy and resorting to the SCA method based on sequential convex programming to provide a locally optimal solution. In particular, we first introduce new matrix variables $\mathbf{R}_{c,k} = \mathbf{w}_{c,k}\mathbf{w}_{c,k}^H$ and $\mathbf{R}_{s,k} = \mathbf{W}_{s,k}\mathbf{W}_{s,k}^H$ to reformulate $(\mathbf{P1})$ as the following semidefinite programming (SDP)

$$\widetilde{(\mathbf{P1})}: \min_{\{\mathbf{R}_{c,k}\},\{\mathbf{R}_{s,k}\}} \sum_{k=1}^{K} \text{Tr}\left(\mathbf{R}_{c,k} + \mathbf{R}_{s,k}\right) \tag{35}$$

$$\text{s.t.} \quad \text{Tr}\left(\mathbf{R}_{c,k} + \mathbf{R}_{s,k}\right) \le P_k, \forall k \tag{35a}$$

Algorithm 1: Alternative Optimization Algorithm for Solving **(P0)**.

1: Initialize a feasible sensing and communication time duration $t^0_{cs,k}$.
2: Initialize iteration index $i = 1$, maximum iteration number I, and solution precision ϵ.
3: **repeat**
4: Jointly optimize problem $(\mathbf{P1})$ and obtain $\mathbf{w}^i_{c,k}, \mathbf{W}^i_{s,k}$ for all ISCC devices by using Algorithm 2.
5: Solve problem $(\mathbf{P2})$ and obtain the optimal $t^i_{cs,k}$ for each ISCC device according to Algorithm 3.
6: $i = i + 1$.
7: **until** $|t^{i-1}_{cs,k} - t^i_{cs,k}| \le \epsilon, \forall k$ or $i >= I$

$$\log_2 \det\left(\mathbf{I}_N + \mathbf{H}_k\left(\mathbf{R}_{c,k} + \mathbf{R}_{s,k}\right)\mathbf{H}_k^H \mathbf{N}_{s,k}^{-1}\right) \ge I_{s,k}^{min}, \tag{35b}$$

$$B\log_2\left(1 + \frac{\mathbf{h}_k^H \mathbf{R}_{c,k}\mathbf{h}_k}{n_{c,k} + \sigma_c^2}\right) \ge \frac{Z}{t_{cs,k}}, \tag{35c}$$

$$\mathbf{R}_{c,k}, \mathbf{R}_{s,k} \in \mathcal{S}_N^+, \text{rank}(\mathbf{R}_{c,k}) = 1, \forall k, \tag{35d}$$

where the inference terms in sensing MI constraint (35b) and communication rate constraint (35c) can be rewritten as $\mathbf{N}_{s,k} = \sum_{i\neq k}^K \mathbf{H}_{i,k}(\mathbf{R}_{c,i} + \mathbf{R}_{s,i})\mathbf{H}_{i,k}^H + \sigma_s^2\mathbf{I}_N$ and $n_{c,k} = \sum_{i\neq k}^K \mathbf{h}_i^H \mathbf{R}_{c,i}\mathbf{h}_i + \sum_i^K \mathbf{h}_i^H \mathbf{R}_{s,i}\mathbf{h}_i$, respectively.

The optimization problem is still non-convex due to the rank-one constraints. Fortunately, SDP relaxation by omitting rank-one constraints and solving the remaining problem is guaranteed to have at least one optimal solution, which is rank one [28]. Therefore, the optimal solution of the relaxation problem can be leveraged to construct a feasible solution of $\widetilde{(\mathbf{P1})}$. Nonetheless, the remaining problem after removing the rank-one constraints is still challenging because of the non-convex SMI constraints. We thus adopt the SCA method to solve the problem efficiently.

We first transform the left hand of (35b) based on a series of properties and rewrite (35b) into the following difference of convex (DC) form:

$$\log_2 \det\left(\sum_{i\neq k}^K \mathbf{H}_{i,k}\left(\mathbf{R}_{c,i} + \mathbf{R}_{s,i}\right)\mathbf{H}_{i,k}^H/\sigma_s^2 + \mathbf{I}_N\right) - \log_2 \det\left(\sum_{k=1}^K \mathbf{H}_{i,k}\left(\mathbf{R}_{c,k} + \mathbf{R}_{s,k}\right)\mathbf{H}_{i,k}^H/\sigma_s^2 + \mathbf{I}_N\right) \le I_{s,k}^{min}, \tag{36}$$

in which the terms with $\log\det(\mathbf{I} + \mathbf{X})$ are concave. Note the transformation in (36) is derived sequentially according to

$$\begin{cases} \det\left(\mathbf{I} + \mathbf{AB}\right) = \det\left(\mathbf{I} + \mathbf{BA}\right), \\ \det\left(\mathbf{B}^{-1}\mathbf{A}\right) = \det\left(\mathbf{B}\right)/\det\left(\mathbf{A}\right), \\ \log\det(\mathbf{A}) - \log\det(\mathbf{B}) = \log\det(\mathbf{A}/c) - \log\det(\mathbf{B}/c). \end{cases} \tag{37}$$

Then, we adopt the first-order Taylor expansion to approximate the first term in the concave form in (36). Before this, we give the first-order Taylor expansion of $\log\det(\mathbf{I} + \mathbf{X})$ as

$$\log\det\left(\mathbf{I} + \mathbf{X}\right) \le \log\det\left(\mathbf{I} + \mathbf{X_0}\right) + \text{Tr}\left[\left(\mathbf{I} + \mathbf{X_0}\right)^{-1}\left(\mathbf{X} + \mathbf{X_0}\right)\right]. \tag{38}$$

According to (38), the nonconvex term of (36) is upper bounded by (39), shown at the bottom of the next page, where $\overline{\mathbf{Y}}_k = \sum_{i\neq k}^K \mathbf{H}_{i,k}(\overline{\mathbf{R}}_{c,i} + \overline{\mathbf{R}}_{s,i})\mathbf{H}_{i,k}^H/\sigma_s^2$. $\overline{\mathbf{R}}_{c,i}$ and $\overline{\mathbf{R}}_{s,i}$ are the fixed covariance matrixs of the i_{th} device from the optimization at last round. Based on (39), the constraint (36) can be further reorganized into (40), shown at the bottom of the next page, where $\gamma_k = \log\det(\mathbf{I}_N + \overline{\mathbf{Y}}_k) - \text{Tr}(\sum_{i\neq k}^K (\mathbf{H}_{i,k}^H(\mathbf{I}_N + \overline{\mathbf{Y}}_k)^{-1}\mathbf{H}_{i,k})(\overline{\mathbf{R}}_{c,i} + \overline{\mathbf{R}}_{s,i}))$ is constant. Correspondingly, the constraint (40) is convex since $\log\det(\mathbf{I} + \mathbf{X})$ is concave and $\text{Tr}(\mathbf{X})$ is linear.

In addition, constraint (35c) can be rewritten as the following linear form

$$\left(2^{\frac{Z}{Bt_{cs,k}}} - 1\right) n_{c,k} + \sigma_c^2 \le \mathbf{h}_k^H \mathbf{R}_{c,k}\mathbf{h}_k. \tag{41}$$

After the above transformation, the SDP relaxation for the problem $(\mathbf{P1})$ can be further represented as

$$\min_{\{\mathbf{R}_{c,k}\},\{\mathbf{R}_{s,k}\}} \sum_{k=1}^K \text{Tr}\left(\mathbf{R}_{c,k} + \mathbf{R}_{s,k}\right) \tag{42}$$

$$\text{s.t.} \quad (35a), (40), (41),$$

$$\mathbf{R}_{c,k}, \mathbf{R}_{s,k} \in \mathcal{S}_N^+, \forall k,$$

which is a convex problem and can be efficiently computed using standard convex optimization numerical techniques.

The details of the SCA-based iterative algorithm for transmit beamforming design are shown in Algorithm 2. The problem (42) is repeatedly solved in each iteration, until the increase of sum transmission power is lower than the defined solution precision. Correspondingly, we can obtain the globally suboptimal solution $\mathbf{R}^*_{c,k}, \mathbf{R}^*_{s,k}, \forall k$. Note that the SCA-based procedure is guaranteed to converge within a limited number of iterations. In particular, the optimal policy of problem (42) always exists in general, which corresponds to a bounded minimum of the objective function. Besides, the objective function value is non-increasing during the whole optimization procedure of Algorithm 2. Therefore, the convergence of the proposed algorithm can be guaranteed. Note that there is no explicit rank constraint for the covariance matrix $\mathbf{R}_{s,k}, \forall k$, while if the rank of $\mathbf{R}^*_{c,k}, \forall k$ is larger than one, a feasible solution of $(\mathbf{P1})$ must be extracted from $\mathbf{R}^*_{c,k}$. In this paper, we then use the randomized procedure in [42] to generate a feasible rank-1 solution $\mathbf{w}_{c,k}$ for all ISCC devices.

C. Optimal S&c Time Duration Allocation

With the given joint beamforming design, the problem $(\mathbf{P2})$ can be rearranged into the following concise form

$$\widetilde{(\mathbf{P2})}: \max_{\mathbf{t}_{cs}} \sum_{k=1}^K t_{cs,k} \tag{43}$$

$$\text{s.t.} \quad \frac{Z}{r_k} \le t_{cs,k} \le \frac{t_0 E_k^{max}}{t_0 \text{Tr}\left(\mathbf{R}_k\right) + \kappa\phi_k f_k^2}, \tag{43a}$$

Algorithm 2: SCA-based Beamforming Design Method.

Initialize the covariance matrices for all ISCC devices as $\mathbf{R}_{c,k} = \mathbf{0}$ and $\mathbf{R}_{s,k} = \mathbf{0}, \forall k$
Initialize solution precision ϵ.
repeat
  Initialize the covariance matrices for each ISCC devices $\overline{\mathbf{R}}_{c,k} = \mathbf{R}_{c,k}, \overline{\mathbf{R}}_{s,k} = \mathbf{R}_{s,k}, \forall k$.
  With $\overline{\mathbf{R}}_{c,k}, \overline{\mathbf{R}}_{s,k}$, calculate the objective function of (42) as $\overline{Obj}$.
  Calculate constant terms including $\gamma_k, \overline{\mathbf{Y}}_k, \forall k$.
  Solve problem (42) and update $\mathbf{R}_{c,k}, \mathbf{R}_{s,k}, \forall k$.
  With $\mathbf{R}_{c,k}, \mathbf{R}_{s,k}$, calculate the objective function of (42) as Obj.
until $|Obj - \overline{Obj}| \leq \epsilon$
$\mathbf{R}^*_{s,k} = \mathbf{R}_{s,k}$
if $\text{rank}(\mathbf{R}_{c,k}) > 1$ **then**
  $\mathbf{R}^*_{c,k} = \mathbf{R}_{c,k}, \forall k$
else
  Derive $\mathbf{R}^*_{c,k}$ from Gaussian randomization of $\mathbf{R}_{c,k}$.
end if

Algorithm 3: Optimal S&C Time Duration Allocation Method.

Rank all devices in order of decreasing computational speed. Denote the sorted list as p, the device at the i_{th} index as $p(i)$, and $\tilde{t} = \max_{k\in\mathcal{K}}\{t_{cs,k}\}$.
for $i = 1, \ldots, K$ **do**
Denote $k' = p(i)$ and calculate $\tilde{t}_{cs,k'}$ according to (44).
  if $\frac{Z}{r_{k'}} \leq \tilde{t}_{cs,k'} \leq \frac{t_0 E_k^{max}}{t_0 \text{Tr}(\mathbf{R}_{k'}) + \kappa \phi_{k'} f_{k'}^2}$ **then**
    Obtain $t^*_{cs,k'} = \tilde{t}_{cs,k'}$
    Calculate $\tilde{t}_{cs,k}$ for the remaining devices according to (45).
    The optimal S&C time duration for each device is given by $t^*_{cs,k} = \min\{\tilde{t}_{cs,k}, \frac{t_0 E_k^{max}}{t_0 \text{Tr}(\mathbf{R}_k) + \kappa \phi_k f_k^2}\}$.
    break
  else
    The optimal S&C time duration for the device k' is given by $t^*_{cs,k} = \frac{t_0 E_k^{max}}{t_0 \text{Tr}(\mathbf{R}_{k'}) + \kappa \phi_{k'} f_{k'}^2}$.
  end if
end for

$$\max\left\{t_{cs,k}\right\}_{k\in\mathcal{K}} + \frac{t_{cs,k}\phi_k}{t_0 f_k} \leq \tau, \tag{43b}$$

in which the S&C time duration is mainly impacted by the required time for processing local sensing data samples.

Lemma 1. Without considering the constraint (43a), for the device with minimum value of ϕ_k/f_k, its optimal S&C time duration is the maximized, i.e., $\arg\min_k\{\phi_k/f_k\} = \arg\max_k\{t_{cs,k}\}$.

Proof. Let $k' = \arg\max_k\{t_{cs,k}\}$ represent the device with the maximum S&C time duration, and denote $t_{cs,k'} = \tilde{t}$. According to the constraint (43b), the upper bound for device k' is $\tilde{t} \leq t_0 \frac{\tau}{1+\phi_k/t_0 f_k}$. Hence, t' can be maximized only when k' is the device with the minimum value of ϕ_k/f_k. In addition, for device $k \neq k'$, its upper bound of S&C time duration can be expressed as $t_{cs,k} \leq \frac{\tau - \tilde{t}}{\phi_k/f_k} \leq \tilde{t}$, since $\phi_k/f_k > \phi_{k'}/f_{k'}$. Therefore, the constraints (43b) can be ensured for all devices, and $\sum_{k=1}^K t_{cs,k}$ can be maximized. □

With Lemma 1, the optimal S&C time duration for all ISCC devices can be expressed as

$$\tilde{t} = \tilde{t}_{cs,k'} = t_0 \frac{\tau}{1 + \phi_k/t_0 f_k}, k' = \arg\min_k \left\{\phi_k/f_k\right\}, \tag{44}$$

and

$$\tilde{t}_{cs,k} = t_0 \frac{\tau - \tilde{t}}{\phi_k/t_0 f_k}, \forall k \neq k'. \tag{45}$$

However, the optimal S&C time duration for each device is still limited by constraint (43a). With this consideration, we propose an iteration-based algorithm for giving the optimal time duration. Specifically, we first rank all devices in order of decreasing computational speed. Then, the optimal S&C time duration can be sequentially determined based on the constraint (43a), (44), and (45). The pseudo-code is illustrated in Algorithm 3.

D. Performance Analysis

1) Convergence Analysis: As for problem ($\mathbf{P0}$), the optimal policy exists in general, which corresponds to a bounded maximum of the objective function. During the whole optimization procedure of Algorithm 1, with the initialized feasible S&C time duration $t^0_{cs,k} \leq t^*_{cs,k}$ for each ISCC device, the allocated transmit power can be minimized by leveraging Algorithm 2, thereby increasing the upper bound in (30). When the system is time-limited, the optimal S&C time duration can thus be obtained,

$$\log_2 \det\left(\sum_{i\neq k}^K \mathbf{H}_{i,k}\left(\mathbf{R}_{c,i} + \mathbf{R}_{s,i}\right)\mathbf{H}_{i,k}^H/\sigma_s^2 + \mathbf{I}_N\right) \leq \text{Tr}\left(\sum_{i\neq k}^K \left(\mathbf{H}_{i,k}^H\left(\mathbf{I}_N + \overline{\mathbf{Y}}_k\right)^{-1}\mathbf{H}_{i,k}\right)\left(\mathbf{R}_{c,i} + \mathbf{R}_{s,i}\right)\right)$$
$$+ \log_2 \det\left(\mathbf{I}_N + \overline{\mathbf{Y}}_k\right) - \text{Tr}\left(\sum_{i\neq k}^K \left(\mathbf{H}_{i,k}^H\left(\mathbf{I}_N + \overline{\mathbf{Y}}_k\right)^{-1}\mathbf{H}_{i,k}\right)\left(\overline{\mathbf{R}}_{c,i} + \overline{\mathbf{R}}_{s,i}\right)\right) \tag{39}$$

$$\text{Tr}\left(\sum_{i\neq k}^K \left(\mathbf{H}_{i,k}^H\left(\mathbf{I}_N + \overline{\mathbf{Y}}_k\right)^{-1}\mathbf{H}_{i,k}\right)\left(\mathbf{R}_{c,i} + \mathbf{R}_{s,i}\right)\right) - \log_2 \det\left(\sum_{k=1}^K \mathbf{H}_{i,k}\left(\mathbf{R}_{c,k} + \mathbf{R}_{s,k}\right)\mathbf{H}_{i,k}^H/\sigma_s^2 + \mathbf{I}_N\right) + \gamma_k \leq I_{s,k}^{min}, \tag{40}$$

TABLE II
SIMULATION PARAMETERS

Parameters	Value
Communication bandwidth B	2MHz
Noise power δ_s^2, δ_c^2	-110 dBm
Racian factor	3
Maximum transmit power of device P_k	30 dBm
Maximum CPU capacity of devices F_k	[0.5, 1] GHz
Energy consumption budget at each FEEL step	0.5 J
Delay requirement at each FEEL step	5 s
Time for generating a sensing sample t_0	10 ms
Required CPU cycles for each sample	3×10^6 cycles
Energy efficiency parameter κ	10^{-27}

which is irrelevant to the beamforming design. Otherwise, the optimal S&C time duration will increase since it is determined by (30). For the next iteration, increasing S&C time duration causes less power consumption in parameter transmission. Therefore, the value of the sum S&C time duration is non-increasing as the iterations proceed, which guarantees Algorithm 1 can converge within a limited number of iterations.

2) Complexity Analysis: The computational complexity of Algorithm 1 within each iteration mainly lies in solving the two subproblems. For the transmit beamforming design algorithm, the worst-case complexity of solving SDP in problem $\widetilde{(\mathbf{P1})}$ is $\mathcal{O}(N^7)$ [43]. Assuming that the number of iterations of the Algorithm 2 is S_ϵ, the corresponding complexity can be represented as $\mathcal{O}(S_\epsilon N^7)$. For the S&C time duration optimization algorithm, the computational complexity to determine $t_{cs,k}$ is proportional to the number of ISCC devices, thus it can be expressed as $\mathcal{O}(K)$.

VI. NUMERICAL RESULTS

In this section, we present simulations to verify the effectiveness of our proposed system design and the corresponding resource optimization method. We consider an FEEL network in a circular area with a radius of 500 m, in which a BS is located at the center and several ISCC devices are randomly deployed. The interval between adjacent antennas of each ISCC device is assumed to be half-wavelength, and the default number of transmit antennas is $N = 8$. For multiple devices radar transmit precoding, similar to the radar target model, we assume each ISCC device senses a single point-like target in the far field at the direction 10°, and the width of each ideal beam is 10°. Without loss of generality, the communication channels are assumed to be Rayleigh fading, and the path loss is modeled as $PL(dB) = 30 + 22\log_{10}(d_k)$, where d_k (in km) is the distance between ISCC device k and the BS. We set the other simulation parameters as listed in Table II except specified otherwise.

For performance evaluation, we consider the following benchmarks. Moreover, the performance of all algorithms is implemented by using the CVX Toolbox and Deep Learning Toolbox-enabled MATLAB platform on a computer with a 2.50 GHz Core i7 CPU and 64 GB RAM.

- ISCC Single: Different from the distributed FEEL system, the traditional edge intelligence system generally adopts single-node-based AI training mode. In particular, only an individual edge device is invoked to sense all targets sequentially, and the sensing data samples are leveraged for model training. Besides, this scheme adopts the advantage of the ISAC paradigm to simultaneously sense the specific target and upload the model-updates to the edge server [28].
- SCC FL: The scheme leverages the FEEL framework to collaboratively train the AI model across multiple ISCC devices. For each ISCC device, instead of parallel execution of data sensing and communication in FLISC^3, this scheme adopts the traditional serial-based architecture, in which the three stages of data sensing, local training, and parameter uploading are operated sequentially within each training round [34], [35].
- AirSCC FL: Similar to the SCC FL scheme, each ISCC device shifts between sensing and communication models in a time-division manner. Differently, during the communication process, this scheme adopts AirComp for ISCC devices to aggregate their local models, in which each subcarrier is assigned to transmit one dimension of the gradient. Meanwhile, the devices sense targets by assigning orthogonal bandwidths [44].

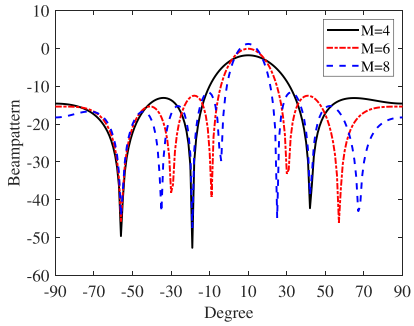


Fig. 2. Beampattern at given direction angles versus ISCC device antenna number.

Fig. 2 shows the optimized beampattern versus the direction angle under different numbers of antenna on ISCC devices. It can be seen that the beampattern is high at the direction of the main beam, while the beampattern on other direction angles is very low, which illustrates that the obtained radar beampattern has higher beam gain on the desired main beam locations while other directions are suppressed. Moreover, we observe that the performance of beampattern becomes better from the radar's viewpoint as the number of radar antennas increases. In particular, the maximum peak to sidelobe ratio decreases with the increasing number of radar antennas. Meanwhile, the main beam width also decreases with the increase of the number of radar antennas.

Fig. 3 shows the convergence behavior of Algorithm 1 with varying system parameters, including the number of ISCC devices and antennas. From the figure, the S&C time duration gradually increases and converges within a certain number of iterations, which reflects the fast convergence performance of Algorithm 1. In addition, we observe that the total S&C time duration within the FLISC^3 framework increases with the number of ISCC devices. This is mainly because an increase in the number of devices introduces more energy and power

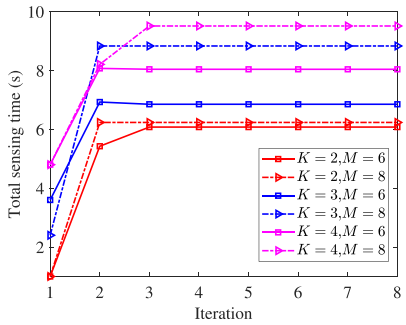
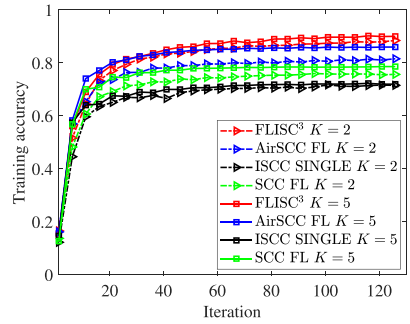


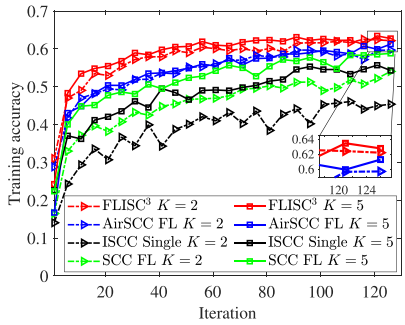
Fig. 3. Convergence performance of Algorithm 1 .

resources to enhance edge collaboration, which effectively helps increase the sum of sensing time and acquire more data samples. However, since all devices perform data sensing and parameter transmission simultaneously in FLISC^3, increasing K leads to invoking more interference during the transmission process. Therefore, with the limited network resources, the performance gain will gradually decrease with a larger value of K. Moreover, the S&C time duration increases with the antenna number on ISCC devices. This is because a larger number of antenna indicates that more data streams can be transmitted and a higher transmission rate can be achieved.

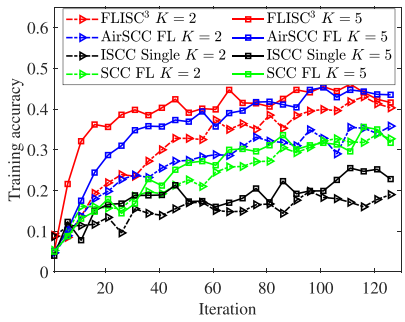
To comprehensively evaluate the learning performance of different schemes, we perform simulations on image classification tasks of varying complexity. Specifically, we employ a Multilayer Perceptron (MLP) based on the MNIST dataset [45], which contains 10 classes of hand-written digit images. Second, we adopt a widely used model, Alexnet [46], for training the CIFAR-10 dataset, which contains 10 different categories of color images of 32×32 resolution. Third, we conduct experiments on a more complex learning task, using a simplified VGG-16 [47] to adapt the 64×64 input size of TinyImagenet. Notably, deeper networks enable more powerful classification capabilities for high-resolution images, while having stricter resource requirements due to the exponential increase in network parameters. Fig. 4(a), (b), and (c) show the training accuracy of all training tasks under different schemes, respectively. All tasks are performed within the same number of iterations and network parameter settings. It can be seen that the training accuracy increases with an increase in the number of ISCC devices. In addition, all models gradually reach a convergence state under different schemes, as the number of iterations increases. Specifically, by utilizing the ISAC technique on devices and efficient resource allocation among them, FLISC^3 enables more flexible resource sharing and utilization, thus achieving superior convergence performance among all schemes. Moreover, AirSCC FL achieves better performance than SCC FL by leveraging the AirComp technique. Conversely, the ISCC Single scheme exhibits the worst performance among all schemes because of the restricted resources on a single node. We also observe that the convergence speed and the convergence error get worse for all schemes with increasing complexity of training tasks, especially in Fig. 4(c). This general degradation of accuracy is caused by the constrained environment that limits



(a) MLP with MNIST



(b) Alexnet with CIFAR-10



(c) Simplified VGG-16 with TinyImagenet

Fig. 4. Training accuracy of different system designs under different learning tasks.

both the sensing ability and data distribution in the considered devices. In particular, to guarantee the successful transmission of heavy parameters, more power resources are required, thus degrading the sensing performance and reducing sensing data samples due to the energy-limits. Besides, for the TinyImagenet dataset with more image classes, per-class sample scarcity further deteriorates learning performance, as shown in Fig. 4(c). Still, FLISC^3 outperforms the other approaches by guaranteeing faster convergence to better accuracy values. Fine-tuning the VGG-16 model is possible, but out of the scope for this work.

Next, taking the simpler handwritten digit identification task as an example, we evaluate the impact of the network parameter setting on training performance. Fig. 5 illustrates the impact of SMI requirements, i.e., $I_{s,k}^{min}$, on the performance of sensing time duration and final training accuracy, in which the threshold of SMI reflects the satisfactory quality of data samples. It can be

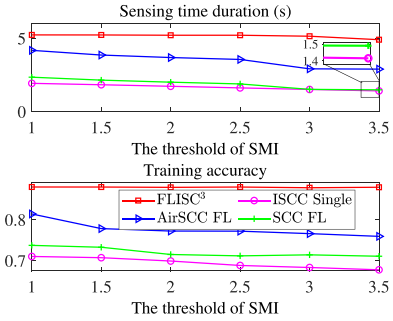


Fig. 5. Performance comparisons versus the threshold of SMI requirements.

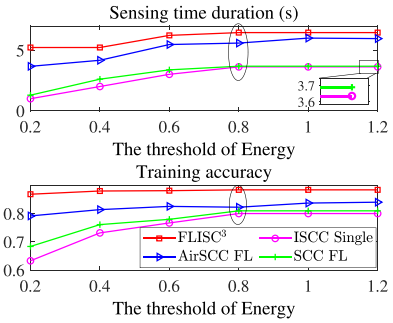


Fig. 6. Performance comparisons versus the threshold of energy resources.

seen that the S&C time duration and model training accuracy for all schemes decrease when the threshold of SMI increases. This is because higher SMI demands require more power resources for target sensing, and the upper bound of S&C time duration in (30) decreases due to the limited energy resources on each ISCC device. Moreover, based on the FEEL performance analysis in Section IV, increasing sensing time duration contributes to accelerating the convergence rate due to more sensing data samples participating in the FEEL process. We also observe that FLISC^3 can achieve higher training accuracy compared to other schemes, reflecting its robustness to acquire more sensing data through efficient device collaboration and multi-functional resource management. Meanwhile, when the threshold of SMI varies from [1,3], increasing the threshold of SMI does not result in a significant reduction of S&C time duration for FLISC^3. With this observation, we can derive that FLISC^3 can enhance FEEL performance by offering consistent performance.

Fig. 6 shows the impact of the energy resource constraint of ISCC devices on the performance of S&C time duration and training accuracy. Obviously, when $E_k^{max} < 1$, the S&C time duration and training accuracy increase with the energy constraint for all schemes, and FLISC^3 achieves superior performance. This can be explained by the fact that the S&C time duration allocation in the energy limit system is mainly determined by beamforming optimization and the system model design of FEEL. Compared to FLISC^3, AirSCC FL sacrifices system flexibility due to the design of frequency division between devices and time division between functions. Alternatively, AirSCC FL adopts AirComp-based aggregation to improve communication efficiency, thereby achieving competitive but suboptimal results.

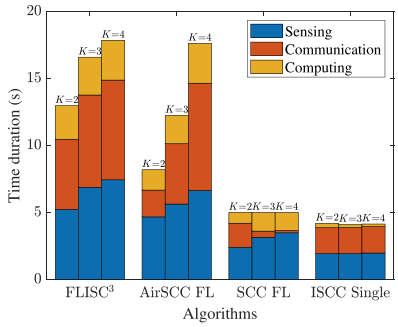
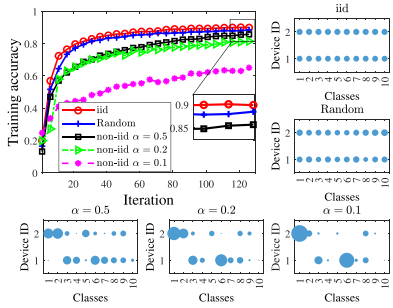


Fig. 7. Time distribution versus different K and system designs.

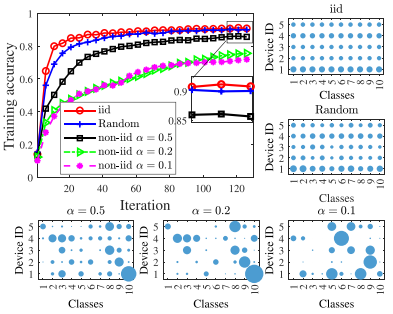
We also observe that ISCC Single performs the worst among all schemes in the energy-limited environment since it always requires higher energy consumption to sense all targets sequentially. Besides, when the energy capacity increases to 1, the allocated S&C time and achieved training accuracy will no longer vary with the energy constraints. This is because the optimal S&C time duration is fixed in the time limit system for all schemes, which is only restricted by the computing capacity on the devices and FEEL system design. Specifically, benefiting from simultaneous sensing for all ISCC devices, the achieved model accuracy of FLISC^3 is better than other schemes. Meanwhile, although the available time duration for sensing each target is the same for ISCC Single and the separate-based design of FL, including AirSCC FL and SCC FL, collaborative learning can obtain additional performance gain from device heterogeneity, since the device with more computational resources can reserve more available time for sensing.

Fig. 7 shows the distribution of time allocation during one global iteration versus different K and FEEL schemes. It can be seen that the S&C time increases as the number of devices increases for the FLISC^3 scheme, which benefits from efficient collaborative resource management among multiple devices. Compared to FLISC^3, AirSCC FL saves communication delay by leveraging the AirComp technique. However, since the data sensing and parameter transmission are operated sequentially, AirSCC FL enables inefficient utilization of signal and system resources, thus exhibiting worse sensing performance. As for the SCC FL scheme, the sensing time duration is increased as more devices participate in the learning process by increasing communication transmit power to reduce communication time as much as possible. Additionally, in the ISCC Single scheme, the device needs to sense all targets and complete local training with both the constraints of energy consumption and total time duration. Therefore, as the number of targets increases, if the available energy and time duration can be leveraged to ensure the required sensing MI and successful parameter uploading, the S&C time duration will not exhibit significant changes. Otherwise, the AI training fails once the number of targets increase to a certain number that the ISCC Single scheme cannot support.

In the above simulations, following the assumption of randomly distributed targets in the system, we assume that the class of the acquired sensing data is randomly selected from all

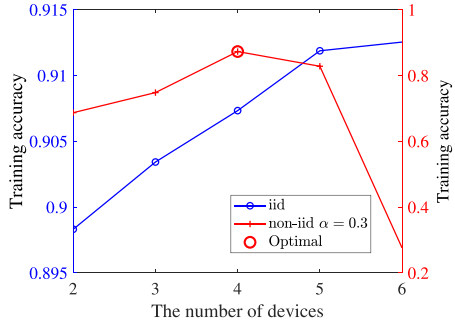


(a) $K = 2$

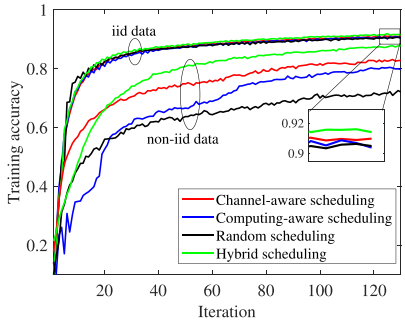


(b) $K = 5$

Fig. 8. Training accuracy of FLISC^3 on different data distributions.



(a)



(b)

Fig. 9. Evaluation of FLISC^3 in large-scale resource-limited network, where $K = 20, I_{s,k}^{min} = 1, E_k^{max} = 0.4$ (a) Training accuracy under feasible number of K; (b) Training accuracy under different device scheduling schemes.

classes. However, in practical networks, dynamic sensing environments such as target mobility greatly affect the target types that can be sensed by different devices, thereby causing the non-iid data distributions on devices. To evaluate the convergence performance of FLISC^3 under non-iid data, we conduct experiments on different levels of data distribution heterogeneity, i.e., $\alpha = \{0.5, 0.2, 0.1\}$. The Dirichlet distribution parameter α is used for controlling the degree of data heterogeneity among devices, and a lower value indicates more severe data heterogeneity among devices. Fig. 8 illustrates the training performance and data distributions of our FLISC^3 under iid (ideal), random (ours), and non-iid (practical). The scatter plot shows data distributions on the MNIST dataset, where the x-axis represents the class labels and the y-axis represents device ID. It can be seen that the best performance can be achieved under the iid assumption, in which data is uniformly distributed across labels locally on each device. However, with an increasing level of data heterogeneity, the convergence speed and model accuracy degrade as expected. This can be explained by the convergence analysis in **Theorem 3**. In addition, comparing Fig. 8(b) with Fig. 8(a), we observe that the "collective intelligence gain" introduced by increasing the number of devices, as shown in Fig. 4, vanishes when $\alpha < 0.5$. This is because higher data heterogeneity causes conflicted local updates, thereby affecting aggregation performance.

We further evaluate the performance of FLISC^3 in large-scale network, where $K = 20, I_{s,k}^{min} = 1, E_k^{max} = 0.4$, the results are shown in Fig. 9. At the beginning, the optimal number of devices in feasible space is illustrated in Fig. 9(a). Under iid data, we see that a better training performance can be achieved as the participating devices increase, and a muted growth is exhibited as K increases from 5 to 6. Recall the results in Fig. 3, this is because additional devices contribute local resources to enhance performance, which also results in intra-interference during the transmission process. When the strict constraints of sensing and communication cannot be met with constrained resources, i.e., $K > 6$, problem $(\mathbf{P0})$ is unfeasible to be solved, which means the network reaches its maximum device capacity. Under non-iid data conditions $\alpha = 0.3$, training performance peaks when $K = 4$, benefiting from additional resources and improved sensing capability, while rapidly decreases when $K > 4$ due to the higher data heterogeneity among devices. Based on Fig. 9(a), we set the optimal number of devices in iid and non-iid data $\alpha = 0.3$ as $K = 6$ and $K = 4$, respectively. This means that only a subset of devices can be scheduled in a large-scale resource-limited network.

Herein, Fig. 9(b) plots the training accuracy of FLISC^3 versus different scheduling schemes. In particular, Channel-aware scheduling: devices are scheduled based only on their channel gains; Computing-aware scheduling: devices are ranked by local computing resources, and the optimal number of devices with more resources is chosen. Random scheduling: the optimal number of devices is randomly selected from all devices; Hybrid scheduling: half of the devices with the best channel quality are first selected, from whom the optimal number with the most computing resources is chosen. From Fig. 9(b), it can be seen that channel-aware and computing-aware scheduling schemes perform better than random scheduling. Therefore, it is essential

to jointly consider the channel conditions and local computing capacity to make sure that the devices can transmit enough information and save more time for sensing more data samples. As expected, the achieved performance of the hybrid scheduling scheme outperforms others, both in iid and non-iid conditions.

VII. Conclusion

In this paper, we proposed a novel collaborative edge system that takes into account closed-loop sensing-communication-computation processes in an FEEL-oriented manner, by exploiting the multi-function integrated edge devices. Specifically, to improve system performance and avoid resource wastage, we employed ISAC technology in the system, thereby the sensing data collection and parameter transmission can execute simultaneously. Considering the deeply coupled processes and heterogeneous device resources, we further provided the theoretical analysis of the effect of resource allocation on the FEEL convergence upper bound. Then, we formulated a joint optimization problem of beamforming design and time duration allocation to maximize the convergence performance of FLISC^3, subject to multidimensional network resources and successful data transmission. To solve this non-convex problem, we designed an alternative optimization algorithm, in which an iterative solution based on the SCA method for beamforming and closed-form solutions for time allocation were proposed within each iteration. Simulation results revealed that the proposed algorithm has great convergence performance, and the proposed framework yields superior learning accuracy compared to traditional implementations of ISCC systems.

References

[1] E. Baccour et al., "Pervasive AI for IoT applications: A survey on resource-efficient distributed artificial intelligence," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 4, pp. 2366–2418, Fourthquarter 2022.

[2] 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.

[3] Y. Mao, X. Yu, K. Huang, Y.-J. A. Zhang, and J. Zhang, "Green edge AI: A contemporary survey," *Proc. IEEE*, vol. 112, no. 7, pp. 880–911, Jul. 2024.

[4] M. Le, T. Huynh-The, T. Do-Duy, T.-H. Vu, W.-J. Hwang, and Q.-V. Pham, "Applications of distributed machine learning for the internet-of-things: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 27, no. 2, pp. 1053–1100, Apr. 2025.

[5] 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.

[6] Q. Duan, J. Huang, S. Hu, R. Deng, Z. Lu, and S. Yu, "Combining federated learning and edge computing toward ubiquitous intelligence in 6 g network: Challenges, recent advances, and future directions," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 4, pp. 2892–2950, Fourthquarter 2023.

[7] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A joint learning and communications framework for federated learning over wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 269–283, Jan. 2021.

[8] 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.

[9] X. Cao, G. Zhu, J. Xu, Z. Wang, and S. Cui, "Optimized power control design for over-the-air federated edge learning," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 342–358, Jan. 2022.

[10] K. B. Letaief, Y. Shi, J. Lu, and J. Lu, "Edge artificial intelligence for 6 G: Vision, enabling technologies, and applications," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 5–36, Jan. 2022.

[11] S. Lu et al., "Integrated sensing and communications: Recent advances and ten open challenges," *IEEE Internet Things J.*, vol. 11, no. 11, pp. 19094–19120, Jun. 2024.

[12] C. Ding, J.-B. Wang, H. Zhang, M. Lin, and G. Y. Li, "Joint MIMO precoding and computation resource allocation for dual-function radar and communication systems with mobile edge computing," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2085–2102, Jul. 2022.

[13] N. Huang et al., "Mobile edge computing aided integrated sensing and communication with short-packet transmissions," *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 7759–7774, Jul. 2024.

[14] P. Liu, Z. Fei, X. Wang, J. Huang, J. Hu, and J. A. Zhang, "Joint offloading and beamforming design in integrating sensing, communication, and computing systems: A distributed approach," *IEEE Trans. Commun.*, vol. 73, no. 7, pp. 4697–4712, Jul. 2025.

[15] D. Wen, Y. Zhou, X. Li, Y. Shi, K. Huang, and K. B. Letaief, "A survey on integrated sensing, communication, and computation," *IEEE Commun. Surveys Tuts.*, vol. 27, no. 5, pp. 3058–3098, Dec. 23, 2024, doi: 10.1109/COMST.2024.3521498.

[16] J. Ren, W. Ni, and H. Tian, "Toward communication-learning trade-off for federated learning at the network edge," *IEEE Commun. Lett.*, vol. 26, no. 8, pp. 1858–1862, Aug. 2022.

[17] Z. Jiang et al., "Computation and communication efficient federated learning with adaptive model pruning," *IEEE Trans. Mobile Comput.*, vol. 23, no. 3, pp. 2003–2021, Mar. 2024.

[18] H. Sun, H. Tian, J. Zheng, and W. Ni, "Joint optimization of convergence and latency for hierarchical federated learning over wireless networks," *IEEE Wireless Commun. Lett.*, vol. 13, no. 3, pp. 691–695, Mar. 2024.

[19] W. Mao, X. Lu, Y. Jiang, and H. Zheng, "Joint client selection and bandwidth allocation of wireless federated learning by deep reinforcement learning," *IEEE Trans. Serv. Comput.*, vol. 17, no. 1, pp. 336–348, Jan./Feb. 2024.

[20] X. Hou, J. Wang, C. Jiang, Z. Meng, J. Chen, and Y. Ren, "Efficient federated learning for metaverse via dynamic user selection, gradient quantization and resource allocation," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 4, pp. 850–866, Apr. 2024.

[21] Y. Qu, S. Yu, L. Gao, K. Sood, and Y. Xiang, "Blockchained dual-asynchronous federated learning services for digital twin empowered edge-cloud continuum," *IEEE Trans. Serv. Comput.*, vol. 17, no. 3, pp. 836–849, May/Jun. 2024.

[22] D. Xu, "A novel hybrid NOMA-TDMA scheme for wireless federated learning networks," *IEEE Trans. Veh. Technol.*, vol. 74, no. 5, pp. 8448–8453, May 2025.

[23] Y. Tao et al., "Private over-the-air federated learning at band-limited edge," *IEEE Trans. Mobile Comput.*, vol. 23, no. 12, pp. 12444–12460, Dec. 2024.

[24] L. Zhao, D. Wu, L. Zhou, and Y. Qian, "Radio resource allocation for integrated sensing, communication, and computation networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 10, pp. 8675–8687, Oct. 2022.

[25] M. Dong, Y. Fu, C. Li, M. Tian, F. R. Yu, and N. Cheng, "Task offloading and resource allocation in vehicular cooperative perception with integrated sensing, communication, and computation," *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 6, pp. 8481–8496, Jun. 2025.

[26] Y. He, G. Yu, Y. Cai, and H. Luo, "Integrated sensing, computation, and communication: System framework and performance optimization," *IEEE Wireless Commun. Lett.*, vol. 23, no. 2, pp. 1114–1128, Feb. 2024.

[27] Y. Zhao et al., "Multi-functional beamforming design for integrated sensing, communication, and computation," *IEEE Trans. Commun.*, vol. 73, no. 8, pp. 6322–6336, Aug. 2025.

[28] T. Zhang, G. Li, S. Wang, G. Zhu, G. Chen, and R. Wang, "Isac-accelerated edge intelligence: Framework, optimization, and analysis," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 1, pp. 455–468, Mar. 2023.

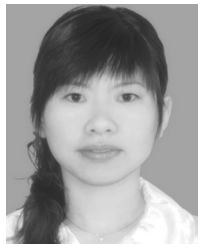
[29] N. Huang et al., "Edge intelligence oriented integrated sensing and communication: A multi-cell cooperative approach," *IEEE Trans. Veh. Technol.*, vol. 73, no. 6, pp. 8810–8824, Jun. 2024.

[30] S. Liu, D. Wen, D. Li, Q. Chen, G. Zhu, and Y. Shi, "Energy-efficient optimal mode selection for edge AI inference via integrated sensing-communication-computation," *IEEE Trans. Mobile Comput.*, vol. 23, no. 12, pp. 14248–14262, Dec. 2024.

[31] Z. Zhuang, D. Wen, Y. Shi, G. Zhu, S. Wu, and D. Niyato, "Integrated sensing-communication-computation for over-the-air edge ai inference," *IEEE Trans. Wireless Commun.*, vol. 23, no. 4, pp. 3205–3220, Apr. 2024.

[32] Q. Qi, X. Chen, A. Khalili, C. Zhong, Z. Zhang, and D. W. K. Ng, "Integrating sensing, computing, and communication in 6G wireless networks: Design and optimization," *IEEE Trans. Commun.*, vol. 70, no. 9, pp. 6212–6227, Sep. 2022.
[33] M. Du, H. Zheng, M. Gao, X. Feng, J. Hu, and Y. Chen, "Integrated sensing, communication and computation for over-the-air federated learning in 6G wireless networks," *IEEE Internet Things J.*, vol. 11, no. 21, pp. 35551–35567, Nov. 2024.
[34] P. Liu et al., "Toward ambient intelligence: Federated edge learning with task-oriented sensing, computation, and communication integration," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 1, pp. 158–172, Jan. 2023.
[35] M. Le, D. T. Hoang, D. N. Nguyen, Q.-V. Pham, and W.-J. Hwang, "Wirelessly powered federated learning networks: Joint power transfer, data sensing, model training, and resource allocation," *IEEE Internet Things J.*, vol. 11, no. 21, pp. 34093–34107, Nov. 2024.
[36] A. Liu et al., "A survey on fundamental limits of integrated sensing and communication," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 994–1034, Secondquarter 2022.
[37] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. Mach. Learn. Syst.*, 2020, pp. 429–450.
[38] J. Wang and G. Joshi, "Cooperative SGD: A unified framework for the design and analysis of communication-efficient SGD algorithms," *J. Mach. Learn. Res.*, vol. 22, no. 1, Jan. 2021.
[39] Z. Zhou, Y. Li, X. Ren, and S. Yang, "Towards efficient and stable K-asynchronous federated learning with unbounded stale gradients on non-IID data," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 12, pp. 3291–3305, Dec. 2022.
[40] L. Chen, X. Ding, Z. Bao, P. Zhou, and H. Jin, "Differentially private federated learning on non-IID data: Convergence analysis and adaptive optimization," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 9, pp. 4567–4581, Sep. 2024.
[41] X. Li, Y. Gao, Y. Deng, and X. Jiang, "Federated learning with adaptive aggregation weights for non-IID data in edge networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 11, no. 5, pp. 3425–3439, Oct. 2025.
[42] Z.-Q. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1426–1438, Aug. 2006.
[43] L. Chen, Z. Wang, Y. Du, Y. Chen, and F. R. Yu, "Generalized transceiver beamforming for DFRC with MIMO radar and MU-MIMO communication," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1795–1808, Jun. 2022.
[44] D. Wen et al., "Over-the-air federated edge learning with integrated sensing, communication, and computation," in *Proc. IEEE/CIC Int. Conf. Commun. China*, 2024, pp. 862–867.
[45] Y. LeCun and C. Cortes, "MNIST database of handwritten digits," Dec. 2024. [Online]. Available: https://service.tib.eu/ldmservice/dataset/mnist-database-of-handwritten-digits
[46] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, vol. 25, pp. 1–9.
[47] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–14.

An Du received the MS degree in computer technology in 2021 from Northeastern University, Shenyang, China, where she is currently working toward the PhD degree in computer science and technology. She is a visiting student with Distributed Systems Group, TU Wien, from 2024 to 2025. Her research interests include edge computing, edge intelligence, and network function virtualization.



Jie Jia (Member, IEEE) received the PhD degree in computer system architecture from Northeastern University, China, in 2009. She is currently a professor with the School of Computer Science and Engineering, Northeastern University. In 2016, she was a visiting research associate with the Department of Informatics, King's College London. She has authored or coauthored more than 100 technical papers on various aspects of wireless networks. Her current research interests include HetNets, IoT, and cognitive radio networks. She is a member of various international societies, such as China Computer Federation.

Schahram Dustdar (Fellow, IEEE) is currently a full professor of computer science (informatics) with a focus on internet technologies with Distributed Systems Group, TU Wien. He is the founding co-editor-in-chief of *ACM Transactions on Internet of Things* and the editor-in-chief of *Computing* (Springer). He is an associate editor of IEEE TRANSACTIONS ON SERVICES COMPUTING, IEEE TRANSACTIONS ON CLOUD COMPUTING, *ACM Computing Surveys*, *ACM Transactions on the Web*, and *ACM Transactions on Internet Technology*, and he is on the editorial board of IEEE INTERNET COMPUTING and IEEE COMPUTER. He was the recipient of multiple awards: TCI Distinguished Service Award in 2021, IEEE TCSVC Outstanding Leadership Award in 2018, IEEE TCSC Award for Excellence in Scalable Computing in 2019, ACM Distinguished Scientist in 2009, ACM Distinguished Speaker in 2021, and IBM Faculty Award in 2012.

Andrea Morichetta (Member, IEEE) received the PhD degree in electrical, electronics, and communication engineering from Politecnico di Torino. He possesses a strong international profile. He is currently a postdoctoral researcher with Distributed Systems Group, TU Wien, specializing in machine learning, distributed systems, and edge-to-cloud computing. He has collaborated with leading institutions and industry partners such as Cisco (San Jose, CA, US), AIT (Vienna, AT, Australia), Futurewei Technologies (Seattle, WA, USA), and Tsinghua University (Beijing, China). His research interests include the study of complex, large-scale distributed systems, through modular, multi-modal coordination.

Jian Chen (Member, IEEE) received the PhD degree in computer application technology from Northeastern University in 2010. He is currently a professor with the School of Computer Science and Engineering, Northeastern University. In 2016, he was a visiting research associate with the Department of Informatics, King's College London. His research interests include intelligent reconfigurable surface, D2D communication, location technology, network management, and signal and image processing.

Xingwei Wang received the BS, MS, and PhD degrees in computer science from Northeastern University, Shenyang, China, in 1989, 1992, and 1998, respectively. He is currently a professor with the College of Computer Science and Engineering, Northeastern University. He has authored or coauthored more than 100 journal articles, books and book chapters, and refereed conference papers. He was the recipient of several Best Paper Awards. His research interests include cloud computing and future internet.