

Per-Round Resource Optimization for ISCC-Enabled Vertical Federated Learning in Edge Perception Network

Shicheng Chu*, Xiaowen Cao*, Lifeng Xie[‡], Xianxin Song[§], and Yejun He*

*College of Electronics and Information Engineering, Shenzhen University, Shenzhen, China

[‡]Peng Cheng Laboratory, Shenzhen, China

[§]City University of Hong Kong, Hong Kong, China

Email: 2400042041@mails.szu.edu.cn, caoxwen@szu.edu.cn, xielf@pcl.ac.cn, xianxin.song@cityu.edu.hk, heyejun@126.com

Abstract—This paper investigates an integrated sensing, communication, and computation (ISCC) enabled edge perception network, where multiple distributed edge devices wirelessly sense objects and collaboratively train a learning model under a coordination of an edge server. To address the feature heterogeneity of distributed wireless sensing data at devices, a vertical federated edge learning (VFEEL) framework is adopted, while over-the-air computation (AirComp) is employed to efficiently aggregate local embeddings. Under this setup, we characterize a rigorous convergence analysis to quantify the joint impact of wireless sensing noise and AirComp aggregation errors on the VFEEL performance. Guided by this analysis, we formulate a per-round optimization problem that jointly optimizes the batch size and resource allocation under strict latency and energy constraints. Numerical results on a human motion recognition task demonstrate that the proposed scheme significantly enhances learning performance and resource efficiency compared to existing baselines.

Index Terms—Over-the-air federated edge learning, vertical federated learning, integrated sensing, communication, and computation, convergence analysis, resource allocation.

I. INTRODUCTION

Next-generation wireless networks are envisioned to support intelligent applications such as autonomous driving and digital twins, which impose stringent requirements on high-precision sensing and ultra-low-latency processing [1], [2]. Edge perception, which integrates wireless sensing with edge intelligence, is emerging as a key paradigm to meet these demands by enabling artificial intelligence (AI) models to learn from their physical surroundings [3], [4]. Wireless sensing provides a massive source of real-time data for edge intelligence. Such data are inherently distributed across multiple devices located in different positions, giving rise to a multi-view sensing scenario. In this setting, each device captures a distinct yet complementary set of features from the same target, leading to naturally feature-partitioned data.

As a typical distributed learning framework, federated edge learning (FEEL) allows collaborative training without sharing raw data, generally categorized into Horizontal FEEL (HFEEL) and Vertical FEEL (VFEEL) [5]. HFEEL refers to the case when devices hold different data samples with the

same feature space, while VFEEL is designed for scenarios where devices hold different feature subsets. Therefore, VFEEL provides a promising framework to deal with the property of multi-view sensing, as it enables collaborative model training on feature-partitioned data while preserving data privacy [5]. However, the learning process introduces considerable communication overhead due to frequent aggregation of intermediate results (i.e., embeddings).

To proceed, the integrated sensing, communication, and computation (ISCC) is proposed to leverage advanced sensing and communication techniques for efficient edge AI [6], [7]. Specifically, integrated sensing and communication (ISAC) enhances data acquisition efficiency [8], while over-the-air computation (AirComp) enables one-shot embedding aggregation, significantly reducing latency by merging communication and computation [9], [10]. Recent studies have explored the ISCC paradigm for edge AI. For example, [7] and [11] proposed a task-oriented framework that jointly optimizes sensing, computation, and communication to enhance learning performance, while [12] investigated AirComp-based ISCC schemes for efficient feature fusion and task-oriented transmission. [13] utilized a model dropout approach within an ISCC framework to reduce transmission overload. However, these works mainly focus on HFEEL. In contrast, ISCC-enabled VFEEL remains largely unexplored, and existing HFEEL analyses cannot be directly extended due to the distinct feature-partitioned data structure and embedding aggregation process in VFEEL.

To bridge this gap, we propose an ISCC framework specifically for the VFEEL architecture, where distributed devices acquire feature-partitioned data via multi-view sensing, culminating in the efficient aggregation of their computed embeddings. Our core contribution is a novel per-round resource optimization. Specifically, we first conduct a rigorous convergence analysis to characterize the joint impact of wireless sensing noise and AirComp aggregation errors on the VFEEL process. Based on this analysis, we formulate the per-round optimization problem to jointly optimize batch size, sensing power, transmission power, and denoising factor under strict latency

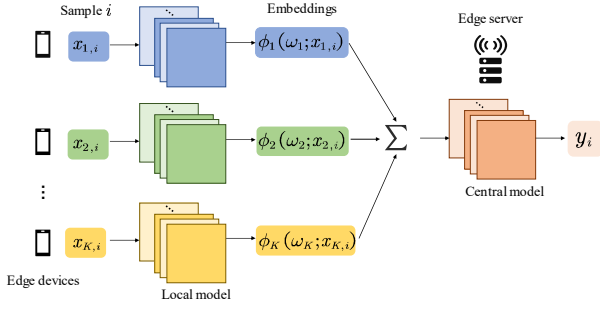


Figure 1. Illustration of ISCC-enabled VFEEL system.

and energy constraints. To solve this non-convex problem, we design an efficient alternating optimization-based algorithm. Finally, numerical results on a human motion recognition task verify that our proposed dynamic optimization scheme significantly outperforms benchmark approaches.

II. SYSTEM MODEL

We consider an ISCC-enabled VFEEL system, as illustrated in Fig. 1. In this architecture, a set of K distributed edge devices with $\mathcal{K} \triangleq \{1, \dots, K\}$, each locating at different positions and thus holding a unique feature partition of its sensing sample, collaboratively train a global machine learning model coordinated by an edge server. Each edge device is equipped with a single-antenna transceiver, allowing it to switch between wireless sensing for data acquisition and communication for embedding updates.

A. Vertical Federated Learning Process

The global learning objective is to find the optimal model parameters $\mathbf{W} = [\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_K]$ that minimize a global loss function $\mathcal{L}(\mathbf{W})$, defined as the expectation of the sample-wise loss $\ell_i(\cdot)$ over the overall clean data distribution $\mathcal{P}$:

$$\min_{\mathbf{W}} \mathcal{L}(\mathbf{W}) = \mathbb{E}_{\mathbf{x}_i \sim \mathcal{P}} \ell_i(\mathbf{w}_0; \phi_i(\mathbf{W}; \mathbf{x}_i); y_i), \quad (1)$$

where $\mathbf{x}_i = [\mathbf{x}_{1,i}, \dots, \mathbf{x}_{K,i}]$ is the i -th ideal noise-free complete data sample from distribution $\mathcal{P}$, y_i is its label, $\mathbf{w}_0$ is the global model at the edge server, $\mathbf{w}_k$ is the local model at device k , $\phi_k(\cdot)$ is the embedding function for the k -th clean feature partition $\mathbf{x}_{k,i}$, and $\phi_i(\mathbf{W}; \mathbf{x}_i) = \sum_{k \in \mathcal{K}} \phi_k(\mathbf{w}_k; \mathbf{x}_{k,i})$ represents the aggregated embedding for sample i .

The problem (1) is solved in a distributed manner using an iterative SGD algorithm. We next introduce the training procedure for any given round t :

- 1) **Sensing collection:** Each edge device k uses wireless signals to collect a batch of $B^{(t)}$ noisy data samples, denoted by $\{\hat{\mathbf{x}}_{k,i}^{(t)}\}_{i=1}^{B^{(t)}}$, where $B^{(t)}$ is the batch size.
- 2) **Local computation:** Each device k uses local model to compute the corresponding local embedding for each sample i , $\hat{\phi}_{k,i}^{(t)} \triangleq \phi_k(\mathbf{w}_k^{(t)}; \hat{\mathbf{x}}_{k,i}^{(t)})$.
- 3) **Embedding forward:** Then, all devices simultaneously transmit their embeddings $\{\hat{\phi}_{k,i}^{(t)}\}$ to the edge server for AirComp. The server receives a noisy aggregated

embedding for each sample, denoted as $\hat{\phi}_i^{(t)}$. The server then computes the sample-wise loss $\ell_i(\mathbf{w}_0^{(t)}; \hat{\phi}_i^{(t)}; y_i)$ using the aggregated embedding and the true label y_i .

- 4) **Backward propagation:** The server calculates the gradient for its own model, $\nabla(\mathbf{w}_0^{(t)})$. It also computes the partial gradient of the loss with respect to the aggregated embedding and broadcasts back to all devices.
- 5) **Local Model Update:** Using the received partial gradient, each device k computes its local model gradient $\nabla(\mathbf{w}_k^{(t)})$.

Finally, both the server and all devices update their respective models using SGD, as given by:

$$\mathbf{w}_k^{(t+1)} = \mathbf{w}_k^{(t)} - \gamma^{(t)} \nabla(\mathbf{w}_k^{(t)}), \forall k \in \{0\} \cup \mathcal{K}, \quad (2)$$

where $\gamma^{(t)}$ is the learning rate and $\nabla(\cdot)$ represents the computed stochastic gradient.

B. ISCC Model

1) **Sensing Model:** During the sensing phase of round t , each device k periodically senses the target by transmitting a *frequency-modulated continuous wave* with multiple up-chirps and receiving the echo signal. The echo signal received from each period is processed to form a single training data sample. This received signal consists of three components: the desired line-of-sight echo signal containing useful micro-Doppler information, the clutter caused by multi-hop reflective paths, and additive sensing noise. The device obtains sample $\hat{\mathbf{x}}_{k,i}^{(t)}$ by processing the echo signal through sampling, singular value decomposition (SVD) based clutter cancellation, short-time Fourier transform, and vectorization and normalization. This sensed data is noisy, given by:

$$\hat{\mathbf{x}}_{k,i}^{(t)} = \mathbf{x}_{k,i}^{(t)} + \mathbf{c}_k^{(t)} + \frac{\mathbf{u}_s^{(t)}}{\sqrt{P_{k,s}^{(t)}}}, \quad (3)$$

where $\hat{\mathbf{x}}_{k,i}^{(t)}$ is the noisy sample obtained by device k , $\mathbf{x}_{k,i}^{(t)}$ is the corresponding ideal ground-truth data sample, $\mathbf{c}_k^{(t)} \sim \mathcal{N}(0, \sigma_{k,c}^2)$ is clutter, $\mathbf{u}_s^{(t)} \sim \mathcal{N}(0, \sigma_s^2)$ is sensing noise, and $P_{k,s}^{(t)}$ is the sensing power. The per-round sensing latency is $T_{k,s}^{(t)} = B^{(t)} \tau_{k,s}$, where $B^{(t)}$ is the batch size in round t , and $\tau_{k,s}$ is the per-sample sensing time. The per-round sensing energy is $E_{k,s}^{(t)} = P_{k,s}^{(t)} B^{(t)} \tau_{k,s}$.

2) **Computation Model:** The per-round computation latency is $T_{k,c}^{(t)} = \Gamma_k B^{(t)} / f_k$. The per-round computation energy is $E_{k,c}^{(t)} = \kappa_k \Gamma_k B^{(t)} f_k^2$, where κ_k is the capacitance coefficient and Γ_k is the CPU cycles per sample.

3) **AirComp Model:** The batch embeddings vector for device k , containing $B^{(t)}$ embeddings computed from the noisy samples, is fully defined as:

$$\mathbf{e}_k^{(t)} \triangleq [\phi_k(\mathbf{w}_k^{(t)}; \hat{\mathbf{x}}_{k,1}^{(t)}), \dots, \phi_k(\mathbf{w}_k^{(t)}; \hat{\mathbf{x}}_{k,B^{(t)}}^{(t)})].$$

The received signal at the server via AirComp is the noisy superposition of all transmitted vectors:

$$\mathbf{y}^{(t)} = \sum_{k \in \mathcal{K}} g_k^{(t)} \sqrt{P_{k,c}^{(t)}} \mathbf{e}_k^{(t)} + \mathbf{n}^{(t)},$$

where $P_{k,c}^{(t)}$ is the transmission power, $g_k^{(t)}$ is the channel gain, and $\mathbf{n}^{(t)} \sim \mathcal{N}(0, \sigma_n^2)$ is the additive white Gaussian noise. The server then estimates the sum $\hat{\mathbf{e}}^{(t)} = \mathbf{y}^{(t)} / \sqrt{\alpha^{(t)}}$ using a denoising factor $\alpha^{(t)}$. The per-round communication latency is $T_t^{(t)} = \lceil DB^{(t)} / M \rceil \tau_c$, where D is the embedding dimension and M symbols are in a resource block of duration τ_c . The per-round communication energy is $E_{k,t}^{(t)} = P_{k,c}^{(t)} \tau_c$.

III. PROBLEM FORMULATION

To proceed with the convergence analysis, we adopt standard assumptions from federated learning literature.

A. Basic Assumptions

For the VFEEL model, we assume the following standard assumptions hold:

- **L-Smoothness:** The global loss function $\mathcal{L}(\mathbf{W})$ is L -smooth.
- **Unbiased Gradients:** The stochastic gradients are unbiased estimators of the true gradient, $\mathbb{E}[\hat{\nabla}(\mathbf{W})] = \nabla \mathcal{L}(\mathbf{W})$, and have bounded variance.
- **Bounded Hessian:** The Frobenius norm of the Hessian of the sample-wise loss function with respect to the local embedding is bounded by a constant $H > 0$.
- **Bounded Embedding Gradients:** The partial gradients of the embedding function $\phi_k(\cdot)$ are bounded. Specifically, there exist positive constants L_1 and L_2 such that:

$$\begin{aligned} \left\| \nabla_{\mathbf{w}_k^{(t)}} \phi_k \left(\mathbf{w}_k^{(t)}; \mathbf{x}_{k,i}^{(t)} \right) \right\|_{\mathcal{F}} &\leq L_1, \\ \left\| \nabla_{\mathbf{x}_{k,i}^{(t)}} \phi_k \left(\mathbf{w}_k^{(t)}; \mathbf{x}_{k,i}^{(t)} \right) \right\|_{\mathcal{F}} &\leq L_2. \end{aligned}$$

Here, L_1 bounds how much the embedding changes with respect to model parameter updates, and L_2 represents how sensitive the embedding is to perturbations in the input data (i.e., sensing noise).

B. Convergence Analysis

We next characterize the aggregation and sensing errors induced by AirComp and wireless sensing, respectively, using the MSE of the aggregation error as given by [7], [11]

$$\begin{aligned} \mathbb{E}[\|\tilde{\delta}^{(t)}\|^2] &= \mathbb{E}[\|\hat{\mathbf{e}}^{(t)} - \mathbf{e}^{(t)}\|^2] \\ &\leq \sum_{k \in \mathcal{K}} \left(\frac{g_k^{(t)} \sqrt{P_{k,c}^{(t)}}}{\sqrt{\alpha^{(t)}}} - 1 \right)^2 + \frac{\sigma_n^2}{\alpha^{(t)}} + \sum_{k \in \mathcal{K}} \frac{(g_k^{(t)})^2 P_{k,c}^{(t)}}{\alpha^{(t)}} \left(\sigma_{k,c}^2 + \frac{\sigma_s^2}{P_{k,s}^{(t)}} \right) L_2^2, \end{aligned}$$

where $\hat{\mathbf{e}}^{(t)}$ is the estimated aggregated embedding vector at the server and $\mathbf{e}^{(t)} \triangleq \sum_{k \in \mathcal{K}} \mathbf{e}_{k,\text{clean}}^{(t)}$ is the desired (ground-truth) aggregated embedding vector computed from clean data samples $\{\mathbf{x}_{k,i}^{(t)}\}$. These terms represent the aggregation error from signal misalignment, channel noise, and sensing noise, respectively.

Based on assumptions and the MSE of aggregation error $\mathbb{E}[\|\tilde{\delta}^{(t)}\|^2]$, we could bound the per-round loss function gap under a properly chosen learning rate, which is derived as follows.

Proposition 1. The expected per-round loss function degradation can be bounded as:

$$\begin{aligned} \mathbb{E}[\mathcal{L}(\mathbf{W}^{(t+1)})] &\leq \mathcal{L}(\mathbf{W}^{(t)}) - \beta_1 \|\nabla \mathcal{L}(\mathbf{W}^{(t)})\|^2 \\ &\quad + \beta_2 \frac{K+1}{B^{(t)}} \left(\sigma^2 + L_1^2 H^2 \mathbb{E}[\|\tilde{\delta}^{(t)}\|^2] \right), \end{aligned} \quad (4)$$

where $\beta_1 \triangleq \gamma^{(t)} - L(\gamma^{(t)})^2/2$ (assuming $\gamma^{(t)} < 2/L$) and $\beta_2 \triangleq L(\gamma^{(t)})^2/2$ are positive constants.

Proof: Please refer to Appendix A. ■

The per-round loss descent characterized in (4) serves as a theoretical foundation for resource allocation. Guided by this result, we aim to optimize the transmit power, sensing power, denoising factor, and batch size for accelerating the convergence speed which is equivalent to minimizing the MSE term in the left of (4), i.e., $\frac{1}{B^{(t)}} \mathbb{E}[\|\tilde{\delta}^{(t)}\|^2]$.

C. Per-Round Optimization Problem Formulation

To maximize the convergence speed, we aim to jointly optimize the ISCC resources under the per-round constraints on both latency and energy, by optimizing the transmission power $\{P_{k,c}^{(t)}\}$, sensing power $\{P_{k,s}^{(t)}\}$, batch size $B^{(t)}$, and denoising factor $\alpha^{(t)}$. Thus, the problem is formulated as:

$$(P1) : \min_{\substack{P_{k,c}^{(t)} \geq 0, P_{k,s}^{(t)} \geq 0, \\ B^{(t)} \in \mathbb{Z}^+, \alpha^{(t)} \geq 0}} \frac{1}{B^{(t)}} \mathbb{E}[\|\tilde{\delta}^{(t)}\|^2]$$

$$\text{s.t. } B^{(t)} \tau_{k,s} + \frac{\Gamma_k B^{(t)}}{f_k} + \left\lceil \frac{DB^{(t)}}{M} \right\rceil \tau_c \leq \Delta_k^{(t)}, \forall k \quad (5a)$$

$$P_{k,s}^{(t)} B^{(t)} \tau_{k,s} + \kappa_k \Gamma_k B^{(t)} f_k^2 + P_{k,c}^{(t)} \tau_c \leq E_k^{(t)}, \forall k \quad (5b)$$

$$P_{k,s}^{(t)} \leq P_k^{\max}, \forall k \quad (5c)$$

$$P_{k,c}^{(t)} \leq P_k^{\max}, \forall k. \quad (5d)$$

Here, (5a) denotes the per-round latency constraint ensuring the total time for sensing, computation, and communication does not exceed the budget $\Delta_k^{(t)}$. Constraint (5b) is the per-round energy budget. Constraints (5c) and (5d) represent the peak power constraints for sensing and communication, respectively. Due to the integer variable $B^{(t)}$ and deep coupling among variables, problem (P1) is a non-convex mixed-integer problem and is hard to solve optimally.

IV. RESOURCE ALLOCATION ALGORITHM

To solve the challenging non-convex problem (P1), we propose an efficient algorithm based on alternating optimization. The core idea is to decouple the optimization variables into two blocks, one associated with sensing and data acquisition ($\{P_{k,s}^{(t)}\}, B^{(t)}$) and the other with communication ($\{P_{k,c}^{(t)}\}, \alpha^{(t)}$), and optimize them iteratively until convergence.

A. Joint Optimization of Batch Size and Sensing Power

Under any given transmission power $\{P_{k,c}^{(t)}\}$ and denoising factor $\alpha^{(t)}$, we first optimize the batch size $B^{(t)}$ and the sensing powers $\{P_{k,s}^{(t)}\}$. Let $\mathcal{C}_1^{(t)} \triangleq \sum_{k \in \mathcal{K}} \left(\frac{g_k^{(t)} \sqrt{P_{k,c}^{(t)}}}{\sqrt{\alpha^{(t)}}} - 1 \right)^2 +$

$\frac{\sigma_n^2}{\alpha^{(t)}} + L_2^2 \sum_{k \in \mathcal{K}} \frac{(g_k^{(t)})^2 P_{k,c}^{(t)}}{\alpha^{(t)}} \sigma_{k,c}^2$, and the optimization problem is reformulated as

$$(P1.1) : \min_{\substack{\{P_{k,s}^{(t)} \geq 0\}, \\ B^{(t)} \in \mathbb{Z}^+}} \frac{1}{B^{(t)}} \left[C_1^{(t)} + L_2^2 \sum_{k \in \mathcal{K}} \frac{(g_k^{(t)})^2 P_{k,c}^{(t)} \sigma_s^2}{\alpha^{(t)} P_{k,s}^{(t)}} \right] \quad (5a), (5b), (5c), \text{ and } (5d).$$

Note that problem (P1.1) is non-convex due to the coupling of $B^{(t)}$ and $P_{k,s}^{(t)}$ in the objective function. Therefore, we introduce an auxiliary variable as $\mathcal{E}_{k,s}^{(t)} \triangleq P_{k,s}^{(t)} B^{(t)}$ to represent the total sensing energy investment at device k . We also relax the integer constraint on $B^{(t)}$ into a continuous variable. In this case, problem (P1.1) is rewritten as

$$\min_{\substack{B^{(t)} \geq 0, \\ \{\mathcal{E}_{k,s}^{(t)} \geq 0\}}} \frac{C_1^{(t)}}{B^{(t)}} + L_2^2 \sum_{k \in \mathcal{K}} \frac{(g_k^{(t)})^2 P_{k,c}^{(t)} \sigma_s^2}{\alpha^{(t)} \mathcal{E}_{k,s}^{(t)}} \quad (6)$$

$$\text{s.t. } \mathcal{E}_{k,s}^{(t)} \tau_{k,s} + \kappa_k \Gamma_k B^{(t)} f_k^2 \leq \tilde{E}_k^{(t)}, \forall k, t \quad (7)$$

$$\mathcal{E}_{k,s}^{(t)} \leq B^{(t)} P_{k,s}^{\max}, \forall k, t \quad (8)$$

(5a) and (5d),

where $\tilde{E}_k^{(t)} \triangleq E_k^{(t)} - P_{k,c}^{(t)} \tau_c$. Notice that problem (6) is convex. The transformed problem is now a standard convex optimization problem and can be solved efficiently using the Lagrange duality method.

Then, the optimal solution to problem (6) is given by

$$\mathcal{E}_{k,s}^{(t)*} = \sqrt{\frac{L_2^2 (g_k^{(t)})^2 P_{k,c}^{(t)} \sigma_s^2}{\lambda_k^* \alpha^{(t)} \tau_{k,s}}}, \quad B^{(t)*} = \left(\sqrt{\frac{C_1^{(t)}}{\sum_{k \in \mathcal{K}} \lambda_k^* \kappa_k \Gamma_k f_k^2}} \right)_{B_{\min}}^{B_{\max}},$$

where λ_k^* are the optimal dual variables corresponding to the energy constraints (5b), which can be obtained via the Ellipsoid method [14]. The terms $B_{\max}$ and $B_{\min}$ represent the upper and lower bounds on the batch size, derived from the relaxed latency constraint (5a) and the communication power constraint (5d), respectively, where $B_{\max} = \min_{k \in \mathcal{K}} \{\Delta_k^{(t)} / (\tau_{k,s} + \Gamma_k / f_k + D \tau_c / M)\}$ and $B_{\min} = \max_{k \in \mathcal{K}} \{P_{k,c}^{(t)} / (D P_{k,s}^{\max})\}$.

With the optimal values $\mathcal{E}_{k,s}^{(t)*}$ and $B^{(t)*}$, the optimal sensing power to problem (P1.1) is given by

$$P_{k,s}^{(t)*} = \min \left(\frac{\mathcal{E}_{k,s}^{(t)*}}{B^{(t)*}}, P_{k,s}^{\max} \right). \quad (9)$$

And the continuous batch size $B^{(t)*}$ is rounded to the nearest feasible integer to obtain the final solution.

Algorithm 1 Per-Round Resource Allocation Algorithm

- 1 **Initialize:** Feasible communication variables $\{P_{k,c}^{(t),0}\}, \alpha^{(t),0}$. Set iteration index $j = 0$.
 - 2 **Repeat:**
 - a) **Solve Subproblem 1:** Given $\{P_{k,c}^{(t),j}\}, \alpha^{(t),j}$, compute optimal batch size $B^{(t),j+1}$ and sensing powers $\{P_{k,s}^{(t),j+1}\}$.
 - b) **Solve Subproblem 2:** Given $B^{(t),j+1}, \{P_{k,s}^{(t),j+1}\}$, iteratively solve for transmission powers $\{P_{k,c}^{(t),j+1}\}$ and denoising factor $\alpha^{(t),j+1}$ using (10) and (11) until convergence.
 - c) Set $j \leftarrow j + 1$.
 - 3 **Until** the objective of (P1) converges.
 - 4 **Output:** Optimized solution $\{P_{k,c}^{(t),*}, P_{k,s}^{(t),*}, B^{(t),*}, \alpha^{(t),*}\}$.
-

B. Joint Optimization of Transmission Power and Denoising Factor

Next, with any given batch size $B^{(t)}$ and sensing powers $\{P_{k,s}^{(t)}\}$, we optimize the communication resources. This subproblem (P1.2) is formulated as:

$$(P1.2) : \min_{\substack{\{P_{k,c}^{(t)} \geq 0\}, \alpha^{(t)} \geq 0}} \frac{1}{B^{(t)}} \mathbb{E}[\|\tilde{\mathbf{d}}^{(t)}\|^2] \quad \text{s.t.} \quad (5d) \text{ and } (5b).$$

However, this problem is still non-convex due to the coupling of $\{P_{k,c}^{(t)}\}$ and $\alpha^{(t)}$ in the objective function.

We use an alternating optimization based algorithm to solve problem (P1.2). Under any given transmission powers, the optimal denoising factor $\alpha^{(t)*}$ that minimizes the MSE can be found in a closed form:

$$\alpha^{(t)*} = \left(\frac{\sum_k (g_k^{(t)})^2 P_{k,c}^{(t)} (1 + \tilde{\sigma}_k^{(t)}) + \sigma_n^2}{\sum_k (g_k^{(t)})^2 P_{k,c}^{(t)}} \right)^2, \quad (10)$$

where $\tilde{\sigma}_k^{(t)} = L_2^2 (\sigma_{k,c}^2 + \sigma_s^2 / P_{k,s}^{(t)})$.

Conversely, for a fixed $\alpha^{(t)}$, the problem of finding the optimal transmission powers $\{P_{k,c}^{(t)*}\}$ is made convex via the substitution $\hat{P}_{k,c}^{(t)} = \sqrt{P_{k,c}^{(t)}}$. Thus the solution for the transmission powers, derived from the KKT conditions of the reformulated problem, exhibits a regularized channel inversion structure that balances channel quality against the energy cost

$$P_{k,c}^{(t)*} = \left(\min \left(\frac{g_k^{(t)} \sqrt{\alpha^{(t)}}}{(g_k^{(t)})^2 (1 + \tilde{\sigma}_k^{(t)}) + \nu_k^* B^{(t)} \tau_c \alpha^{(t)}}, \sqrt{P_{k,c}^{\max} D B^{(t)}} \right) \right)^2, \quad (11)$$

where ν_k^* is the optimal dual variable for the energy constraint.

C. Overall Algorithm

The complete procedure is summarized in Algorithm 1, which iterates between solving problem (P1.1) and problem (P1.2). Since each step ensures a non-increasing objective value and the objective is lower-bounded, the algorithm is guaranteed to converge to a high-quality suboptimal solution.

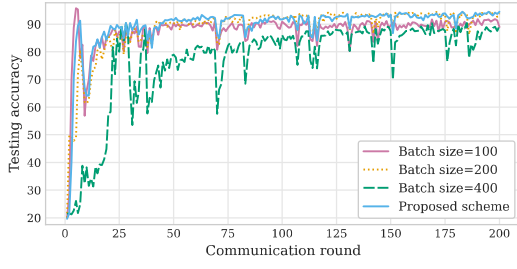


Figure 2. Learning performance of ISCC-enabled VFEEL over different batch size (proposed optimized ≈ 120).

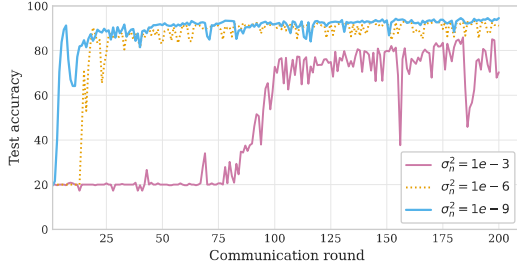


Figure 3. ISCC-enabled VFEEL performance under varying channel noise.

V. PERFORMANCE EVALUATION

A. Simulation Setup

We evaluate our proposed ISCC scheme in a VFEEL network comprising one edge server and $K = 3$ ISAC devices. The wireless channels are modeled with Rayleigh fading and a path loss of 10^{-3} . We use a public wireless sensing dataset [15] for a seven-class human motion recognition task, training a ResNet-10 model over $T = 200$ communication rounds. Unless stated otherwise, we adopt the following default parameters for each device k : total energy budget $E_k = 1000$ J and total latency budget $\Delta_k = 300$ s. These budgets are evenly distributed across all rounds. The maximum transmission power is $P_k^{\max} = 5$ W, and the maximum sensing power is $P_{k,s}^{\max} = 0.5$ W. Other parameters include: noise variances $\sigma_n^2 = \sigma_s^2 = \sigma_{k,c}^2 = 10^{-9}$ W, embedding dimension $D = 100$, per-sample CPU cycles $\Gamma_k \approx 10^7$, and CPU frequency $f_k = 2$ GHz. The learning rate is set to 0.1.

To validate the effectiveness of our proposed scheme, we compare it against the following benchmarks:

- **Fixed Power:** Sensing and communication powers are fixed at $P_k^{\max}$, while other variables are optimized.
- **Fixed Batch Size:** The batch size is fixed at $B^{(t)} = 100$, while other variables are optimized.
- **ISCC HFEEL [11]:** An existing ISCC scheme designed for horizontal federated learning, adapted to our system for comparison.

B. Results and Analysis

Fig. 2 illustrates the impact of different batch sizes on learning performance. It can be observed from the figures that our proposed scheme achieves superior performance in terms of both convergence and accuracy. Specifically, when the fixed

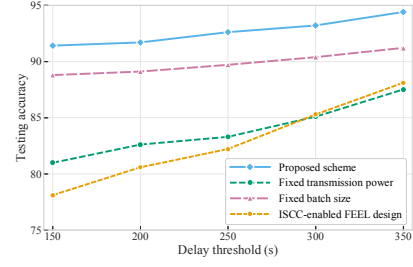


Figure 4. Testing accuracy versus uniform delay threshold.

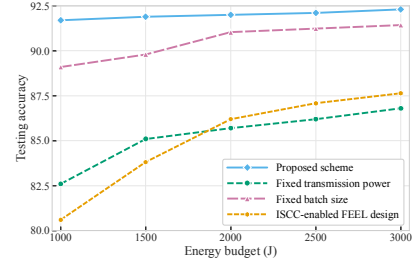


Figure 5. Testing accuracy versus uniform energy budget.

batch size is smaller than that determined by our scheme, the learning performance degrades. A slight increase in the batch size maintains a relatively stable system performance compared to our approach. However, setting the batch size too large leads to performance deterioration due to limited network resources. These experimental results are consistent with our analysis regarding the batch size.

Fig. 3 demonstrates the robustness of the parameters optimized by our proposed scheme under various channel noise variances, σ_n^2 . The figures show that the joint optimization of transmission power and the denoising factor enables the system to effectively mitigate the impact of AirComp errors. Although higher noise levels inevitably degrade performance and slow down convergence, this result highlights that adaptive resource allocation for each round is necessary to maintain reliable learning in dynamic wireless environments.

Figs. 4 and 5 show the final test accuracy achieved by different schemes under varying per-round latency and energy budgets, respectively. In both scenarios, the proposed joint optimization scheme consistently outperforms all benchmarks, demonstrating its ability to make the most efficient use of available resources. As either the latency or energy budget increases, the performance of all schemes improves, as more resources can be allocated to sensing, computation, and communication. However, the performance gains saturate when another resource becomes the bottleneck. For instance, in Fig. 4, with a very generous latency budget, the system performance becomes limited by the energy budget. Our scheme achieves the highest accuracy across all budget levels, highlighting the significant advantage of a holistic, adaptive resource allocation strategy over partial or static approaches. Notably, it also surpasses the ISCC HFEEL scheme, underscoring the VFEEL framework's strength in leveraging feature-rich, multi-view sensing data.

VI. CONCLUSION

This paper considered an ISCC enabled edge perception system, where VFEEL is adopted to deal with the multi-view sensing and AirComp is leveraged for efficient aggregation of device embeddings. By establishing a rigorous convergence analysis, we quantified the joint effects of sensing noise and AirComp aggregation errors on the learning process. Building upon this, we formulated a per-round optimization problem to jointly optimize the batch size, sensing power, transmission power, and denoising factor under both latency and energy constraints. An efficient alternating optimization-based algorithm was designed to solve this non-convex problem. Experimental results on a human motion recognition task validated the proposed design. Future work could extend this framework to decentralized VFEEL scenarios without a central server.

APPENDIX

A. Proof of the Per-Round Loss Degradation Bound

We derive the bound (4) starting from the standard descent lemma for the L -smooth loss function $\mathcal{L}(\cdot)$ and the SGD update rule (2). Taking expectation over randomness (sampling, sensing, channel) and using unbiasedness $\mathbb{E}[\langle \nabla \mathcal{L}, \hat{\nabla} \rangle] = \|\nabla \mathcal{L}\|^2$, we obtain the bound on the expected one-step loss degradation:

$$\mathbb{E}[\mathcal{L}(\mathbf{W}^{(t+1)})] \leq \mathcal{L}(\mathbf{W}^{(t)}) - \gamma^{(t)} \|\nabla \mathcal{L}\|^2 + \frac{L(\gamma^{(t)})^2}{2} \mathbb{E}[\|\hat{\nabla}\|^2]. \quad (12)$$

We bound the second moment $\mathbb{E}[\|\hat{\nabla}\|^2]$ using variance decomposition:

$$\mathbb{E}[\|\hat{\nabla}\|^2] = \mathbb{E}[\|\hat{\nabla} - \nabla \mathcal{L}\|^2] + \|\nabla \mathcal{L}\|^2. \quad (13)$$

The variance term $\mathbb{E}[\|\hat{\nabla} - \nabla \mathcal{L}\|^2]$ sums variances over server and K devices:

$$\mathbb{E}[\|\hat{\nabla} - \nabla \mathcal{L}\|^2] = \sum_{k=0}^K \mathbb{E}[\|\hat{\nabla}(\mathbf{w}_k^{(t)}) - \nabla_k \mathcal{L}\|^2]. \quad (14)$$

Each partial variance combines sampling noise (σ^2) and ISCC aggregation error ($\tilde{\delta}^{(t)}$). Following [11], [16], using Taylor expansion and bounds L_1, H :

$$\mathbb{E}[\|\hat{\nabla}(\mathbf{w}_k^{(t)}) - \nabla_k \mathcal{L}\|^2] \leq \frac{\sigma^2}{B^{(t)}} + \frac{L_1^2 H^2}{B^{(t)}} \mathbb{E}[\|\tilde{\delta}^{(t)}\|^2],$$

where the first term is from sampling, and the second is from the aggregation error. Summing over all $K+1$ models gives the total variance bound:

$$\mathbb{E}[\|\hat{\nabla} - \nabla \mathcal{L}\|^2] \leq \frac{(K+1)}{B^{(t)}} \left(\sigma^2 + L_1^2 H^2 \mathbb{E}[\|\tilde{\delta}^{(t)}\|^2] \right). \quad (15)$$

Substituting (13) and (15) into (12) yields:

$$\begin{aligned} \mathbb{E}[\mathcal{L}(\mathbf{W}^{(t+1)})] &\leq \mathcal{L}(\mathbf{W}^{(t)}) - \left(\gamma^{(t)} - \frac{L(\gamma^{(t)})^2}{2} \right) \|\nabla \mathcal{L}\|^2 \\ &\quad + \frac{L(\gamma^{(t)})^2 (K+1)}{2B^{(t)}} \left(\sigma^2 + L_1^2 H^2 \mathbb{E}[\|\tilde{\delta}^{(t)}\|^2] \right). \end{aligned}$$

This completes the proof.

ACKNOWLEDGMENT

This work is supported in part by National Key Research and Development Program of China under Grant 2023YFE0107900, the NSFC under Grants 62501407 and U2541208, the Key Program of Shenzhen Natural Science Foundation under Grant JCYJ20241202124219023, the Program of Shenzhen Key Laboratory Evaluation under Grant SYSPG20241211173908022, and the Shenzhen Science and Technology Program under Grant RCBS20231211090520032, and the Guangdong Provincial Key Laboratory of Future Networks of Intelligence under Grant 2022B1212010001.

REFERENCES

- [1] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, Aug. 2019.
- [2] Z. Wang, Y. Du, K. Wei, et al., "Vision, application scenarios, and key technology trends for 6G mobile communications," *SCIENCE CHINA Information Sciences*, vol. 65, pp. 1–27, 2022.
- [3] Y. Cui, X. Cao, G. Zhu, J. Nie, and J. Xu, "Edge perception: Intelligent wireless sensing at network edge," *IEEE Commun. Mag.*, vol. 63, no. 3, pp. 166–173, 2025.
- [4] G. Zhu, Z. Lyu, X. Jiao, P. Liu, M. Chen, J. Xu, S. Cui, and P. Zhang, "Pushing AI to wireless network edge: An overview on integrated sensing, communication, and computation towards 6G," *Sci. China Inf. Sci.*, vol. 66, no. 3, Feb. 2023. [Online]. Available: <http://dx.doi.org/10.1007/s11432-022-3652-2>
- [5] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transaction on Intelligent Systems*, vol. 10, no. 2, pp. 1–19, 2019.
- [6] Z. Feng, Z. Wei, X. Chen, H. Yang, Q. Zhang, and P. Zhang, "Joint communication, sensing, and computation enabled 6g intelligent machine system," *IEEE Network*, vol. 35, no. 6, pp. 34–42, 2021.
- [7] P. Liu, G. Zhu, S. Wang, W. Jiang, W. Luo, H. V. Poor, and S. Cui, "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.
- [8] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1728–1767, Jun. 2022.
- [9] X. Cao, Z. Lyu, G. Zhu, J. Xu, L. Xu, and S. Cui, "An overview on over-the-air federated edge learning," *IEEE Wireless Commun.*, vol. 31, no. 3, pp. 202–210, 2024.
- [10] X. Cao, G. Zhu, J. Xu, and K. Huang, "Optimized power control for over-the-air computation in fading channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7498–7513, Nov. 2020.
- [11] D. Wen, S. Xie, X. Cao, Y. Cui, J. Xu, Y. Shi, and S. Cui, "Integrated sensing, communication, and computation for over-the-air federated edge learning," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2025.
- [12] D. Wen, X. Jiao, P. Liu, G. Zhu, Y. Shi, and K. Huang, "Task-oriented over-the-air computation for multi-device edge AI," *IEEE Trans. Wireless Commun.*, vol. 23, no. 3, pp. 2039–2053, Mar. 2024.
- [13] X. Jiao, G. Zhu, W. Jiang, L. Chen, W. Luo, and D. Wen, "Sensing communication computation integration for federated edge learning with controllable model dropout," *IEEE Internet Things J.*, vol. 12, no. 12, pp. 19767–19781, Jun. 2025.
- [14] S. Boyd, "Ellipsoid method." [Online]. Available: <https://web.stanford.edu/class/ee364b/lectures/ellipsoidmethods/slides.pdf>
- [15] P. Liu, G. Zhu, W. Jiang, W. Luo, J. Xu, and S. Cui, "Vertical federated edge learning with distributed integrated sensing and communication," *IEEE Communications Letters*, vol. 26, no. 9, pp. 2091–2095, 2022.
- [16] T. J. Castiglia, A. Das, S. Wang, and S. Patterson, "Compressed-VFL: Communication-efficient learning with vertically partitioned data," in *Proc. 39th ICML*, ser. Proceedings of Machine Learning Research, vol. 162. PMLR, 17–23 Jul. 2022, pp. 2738–2766.