EPFL-REMNet: Efficient Personalized Federated Digital Twin Towards 6G Heterogeneous Radio Environment

Peide Li*, Liu Cao*, Lyutianyang Zhang†, Dongyu Wei‡, Ye Hu§, Qipeng Xie ¶
*Department of Electronic Information Engineering, City University of Hong Kong (Dongguan), Dongguan, China
†The School of Microelectronics and Communication Engineering, Chongqing University, Chongqing, China
‡Department of Electrical and Computer Engineering, University of Miami, FL, USA
§Department of Industrial and Systems Engineering, University of Miami, FL, USA
¶Information Hub, Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China
Emails: {peide.li, liu.cao}@cityu-dg.edu.cn, zhanglyutianyang@cqu.edu.cn, {dongyu.wei,yehu}@miami.edu, qxieaf@connect.ust.hk

***Abstract*—Radio Environment Map (REM) is transitioning from 5G homogeneous environments to B5G/6G heterogeneous landscapes. However, standard Federated Learning (FL), a natural fit for this distributed task, struggles with performance degradation in accuracy and communication efficiency under the non-independent and identically distributed (Non-IID) data conditions inherent to these new environments. This paper proposes EPFL-REMNet, an efficient personalized federated framework for constructing a high-fidelity digital twin of the 6G heterogeneous radio environment. The proposed EPFL-REMNet employs a "shared backbone + lightweight personalized head" model, where only the compressed shared backbone is transmitted between the server and clients, while each client's personalized head is maintained locally. We tested EPFL-REMNet by constructing three distinct Non-IID scenarios (light, medium, and heavy) based on radio environment complexity, with data geographically partitioned across 90 clients. Experimental results demonstrate that EPFL-REMNet simultaneously achieves higher digital twin fidelity (accuracy) and lower uplink overhead across all Non-IID settings compared to standard FedAvg and recent state-of-the-art methods. Particularly, it significantly reduces performance disparities across datasets and improves local map accuracy for long-tail clients, enhancing the overall integrity of digital twin.**

***Index Terms*—Federated Learning, 6G, Heterogeneous Radio Environment Map, Digital Twin, Non-IID.**

I. Introduction

The construction of a real-time, high-fidelity Radio Environment Map (REM) is emerging as a cornerstone technology for the intelligent operation and optimization of 5G-Advanced and future 6G networks. A REM functions as a digital twin of the radio environment, mapping key performance indicators like received power and path loss to replace costly ray-tracing simulations with a dynamic, data-driven model [1]. Recent academic and industrial efforts, such as challenges focused on path loss estimation, have helped standardize datasets and benchmarks, fostering collaborative progress [2]. In practice, the measurement data required to build this digital twin is naturally distributed across end-user devices and edge servers. This distribution, coupled with privacy regulations and uplink bandwidth constraints, makes centralized training untenable. Federated Learning (FL), with its principle of "training locally while building a shared model," offers a compelling paradigm for this task, enabling the collaborative construction of a network-wide digital twin from decentralized data sources [3].

Despite its promise, deploying FL for REMs immediately confronts two formidable challenges. The first is *extreme statistical heterogeneity*. 6G wireless signal propagation is strongly correlated with geography; diverse urban topographies: building densities, and material compositions create vast differences in data distributions across clients [4], [5]. This location-dependent, structural Non-IID nature of the data means a single global model often fails to generalize to specific local environments, particularly for "long-tail" clients with unique propagation characteristics. The second challenge is the *severe communication bottleneck*. In edge-centric wireless networks, the uplink is typically the most constrained resource. The iterative optimization process of FL requires frequent exchanges of high-dimensional model parameters or gradients, imposing significant burdens on network bandwidth and operational costs, thereby hindering large-scale, continuous training [6].

Existing research has addressed the above issues in two separate directions: Personalized Federated Learning (PFL) is used to tackle the heterogeneity, for example, splitting models into a shared backbone and local heads (e.g., FedRep [7]), applying personalized regularization (e.g., Ditto [8]), or using local normalization layers to mitigate feature shift (e.g., FedBN [9]). While being effective from the adaptation perspective, these methods do not inherently reduce communication costs. Besides, the communication-efficient FL focuses on minimizing data transmission through techniques such as sparsification with error feedback [10], quantization [11], and reduced communication frequency [12]. However, under strong Non-IID conditions, aggressive compression can discard fine-grained gradient information crucial for local adaptation, leading to instability and slower convergence. This suggests that person-

This work was supported by the Youth Innovation Talent Project of Guangdong Provincial Universities (Grant No. 2025KQNCX17).

alization and communication efficiency are not independent problems. A robust personalization strategy can make the global aggregation task more amenable to compression, while a well-calibrated compression scheme can, in turn, regularize the shared model to benefit personalization [13]–[17].

Motivated by the aforementioned issues, we propose an Efficient PFL(EPFL)- REMNet, a framework that synergistically integrates both concepts. The key contributions are summarized as follows:

- **A Novel Co-Design Framework:** We deeply couple a PFL architecture ("shared backbone + local personalized heads") with a hybrid communication compression pipeline (Top-K sparsification, error feedback, 8-bit quantization, and periodic updates). This design achieves both bandwidth savings and stable convergence by ensuring a proper division of labor between transferability and adaptability.
- **A Comprehensive Evaluation Methodology:** We construct light, medium, and heavy Non-IID scenarios to mirror real-world data heterogeneity. Beyond standard RMSE/MAE metrics, we introduce fairness and robustness indicators to quantify both model utility and engineering trustworthiness.
- **State-of-the-Art Performance:** Across all heterogeneity levels, EPFL-REMNet demonstrates significant advantages over standard FedAvg [18] and recent baselines like FedFly [19] and PFedgb [20]. Compared to uncompressed PFL, it reduces cumulative communication overhead by over 97% with negligible accuracy loss, while simultaneously improving performance for long-tail clients.
- **A Practical Multi-Objective Optimization Paradigm:** We frame the construction of the REM-based digital twin as a trade-off between accuracy and communication cost, and we map the Pareto frontier. This provides an actionable process for hyperparameter selection, enabling deployment configurations that meet specific operational constraints, such as minimizing communication cost for a target digital twin fidelity.

The rest of this paper is organized as follows: Sec II details the system architecture and problem formulation of the EPFL-REMNet framework while Sec III provides the corresponding federated optimization algorithm. The experimental design used to rigorously evaluate EPFL-REMNet, including the dataset, heterogeneity scenario construction, evaluation metrics, and baseline methods are detailed in Sec IV. Finally, section V draws the conclusions for this paper.

II. The EPFL-REMNet Framework

A. System Architecture

We consider a 6G heterogeneous radio environment scenario (including significantly distinct light, medium, and heavy radio environment areas that are Non-IID) with N clients (e.g., edge servers representing distinct geographical areas) and M base stations (BSs). Each client $i = \{1, 2, ..., N\}$ possesses a local dataset $\mathcal{D}_i$.

$$\mathcal{D}_i = \{(\boldsymbol{X}_j, \boldsymbol{y}_j)\}_{j=1}^{|\mathcal{D}_i|}, \tag{1}$$

where the input feature vector $\boldsymbol{X}_j \in \mathbb{R}^{2+P}$ comprises 2-dimension spatial coordinates and P environmental auxiliary features. The output label $\boldsymbol{y}_j \in \mathbb{R}^M$ represents the signal strength from M BSs:

$$\boldsymbol{y}_j = [s_1, s_2, ..., s_M]^T. \tag{2}$$

To handle data heterogeneity, each client performs local preprocessing. Spatial coordinates are normalized to the range $[0, 1]$ using min–max scaling. The signal strength labels are standardized using the client's local mean μ_i and standard deviation σ_i via z-score normalization:

$$\hat{\boldsymbol{y}} = \frac{\boldsymbol{y} - \boldsymbol{\mu}_i}{\sigma_i}. \tag{3}$$

As Fig. 1 shows, the proposed EPFL-REMNet's architecture embodies the principle of decoupling generalizable knowledge from local specificities.

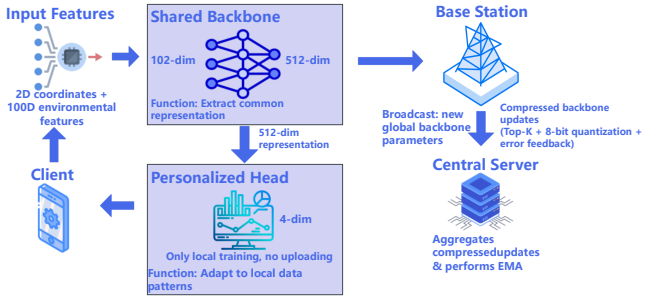

Fig. 1: EPFL-REMNet System Architecture.

Shared Backbone Network ($f_{\boldsymbol{\theta}} : \mathbb{R}^{2+P} \to \mathbb{R}^{512}$): This component is responsible for extracting high-level, transferable representations from the input features and is shared among all clients. It consists of a three-layer Multi-Layer Perceptron (MLP) that maps the input to a 512-dimensional latent representation $\boldsymbol{z} \in \mathbb{R}^{512}$. Its structure is

$$z = f_{\boldsymbol{\theta}}(x) = \mathcal{N}_3\Big(\mathcal{A}\big(\boldsymbol{W}_3 \cdot \mathcal{N}_2(\mathcal{A}(\boldsymbol{W}_2 \cdot \mathcal{N}_1(\mathcal{A}(\boldsymbol{W}_1 x))))\big)\Big), \tag{4}$$

where $\boldsymbol{W}_1, \boldsymbol{W}_2, \boldsymbol{W}_3$ are weight matrices, $\mathcal{A}(\cdot)$ is the SiLU activation function, and $\mathcal{N}_n(\cdot)$ denotes LayerNorm of the nth layer. The use of LayerNorm is crucial for stabilizing the training process, especially under FL scenarios with heterogeneous data distributions.

Personalized Head Network ($g_{\boldsymbol{\phi}_i} : \mathbb{R}^{512} \to \mathbb{R}^M$): This lightweight network takes the latent representation $\boldsymbol{z}$ from the backbone and maps it to the final M-dimension signal strength estimation, which constitutes a data point within the local REM. Each client i exclusively trains and maintains its own head $g_{\boldsymbol{\phi}_i}$, allowing it to specialize in the unique radio characteristics of its local environment, thereby ensuring a more

accurate local representation within the global digital twin. We consider a two-head architecture:

- **Single-layer head:** A simple linear layer,

$$g_{\boldsymbol{\phi}_i}(\boldsymbol{z}) = \boldsymbol{W}_{\text{out}}^{(i)}\boldsymbol{z} + \boldsymbol{b}_{\text{out}}^{(i)} \tag{5}$$

- **Two-layer head:** An MLP with one hidden layer for greater non-linear capacity,

$$g_{\boldsymbol{\phi}_i}(\boldsymbol{z}) = \boldsymbol{W}_2^{(i)}\sigma\left(\boldsymbol{W}_1^{(i)}\boldsymbol{D}(\boldsymbol{z})\right) + \boldsymbol{b}_2^{(i)} \tag{6}$$

  Where $\boldsymbol{D}(\cdot)$ denotes the dropout operation. During training, it randomly drops out the neuron outputs of part of the input feature $\boldsymbol{z}$ to suppress model overfitting.

B. Problem Formulation

The system's objective is to minimize the weighted average of a loss function over all N clients, which achieves the high-fidelity digital twin per BS radio environment. The model is decomposed into a shared backbone network $f_{\boldsymbol{\theta}}$ with parameters $\boldsymbol{\theta}$ and client-specific personalized head networks $g_{\boldsymbol{\phi}_i}$ with parameters $\boldsymbol{\phi}_i$ of client i. The global optimization problem is

$$\min_{\boldsymbol{\theta},\{\boldsymbol{\phi}_i\}_{i=1}^N} \sum_{i=1}^{N} \frac{|\mathcal{D}_i|}{\sum_{j=1}^{N}|\mathcal{D}_j|}\, \mathbb{E}_{(x,y)\sim\mathcal{D}_i}\left[\mathcal{L}\big(g_{\boldsymbol{\phi}_i}(f_{\boldsymbol{\theta}}(x)), \hat{\boldsymbol{y}}\big)\right], \tag{7}$$

where $\mathcal{L}(\cdot)$ is the Huber loss, chosen for its robustness to outliers by combining the advantages of Mean Squared Error (MSE) and Mean Absolute Error (MAE).

III. Federated Optimization and Communication Method

The training process of EPFL-REMNet is governed by a carefully designed protocol that co-optimizes for accuracy and communication efficiency,

In a communication round t, the process is as follows:

1) **Server Broadcast:** The server sends the current global backbone parameters $\boldsymbol{\theta}^t$ to a selected subset of clients.
2) **Local Training:** Each participating client i synchronizes its local backbone to $\boldsymbol{\theta}^t$. It then performs E epochs of local training on its full model (backbone and head) using its local dataset $\mathcal{D}_i$. The personalized head parameters $\boldsymbol{\phi}_i$ are updated and stored locally, while the update to the backbone, $\boldsymbol{\Delta}_i^t$, is computed.
3) **Update Compression:** To minimize uplink communication, the update $\boldsymbol{\Delta}_i^t$ undergoes a three-stage compression pipeline [21]:
   - **Error Feedback and Accumulation:** The current update $\boldsymbol{\Delta}_i^t$ is adjusted by incorporating the *sparsification error* $\boldsymbol{e}_i^{t-1}$, which represents the residual information from the previous round that was not transmitted due to sparsification. Formally:

$$\boldsymbol{u}_i^t = \boldsymbol{\Delta}_i^t + \boldsymbol{e}_i^{t-1}. \tag{8}$$

     This mechanism compensates for the information loss introduced by sparsification and helps maintain stable convergence.

Algorithm 1: EPFL-REMNet Training Algorithm

Input: Global round t, global shared feature parameters $\boldsymbol{\theta}^t$, client sampling set S_t, local epochs E, communication period R
Server executes:
Broadcast $\boldsymbol{\theta}^t$ to all clients $i \in S_t$;
Initialize empty set of updates $\mathcal{U} = \emptyset$;
for *client* $i \in S_t$ *in parallel* **do**
  $\boldsymbol{\Delta}_i^t \leftarrow$ ClientUpdate$(i, \boldsymbol{\theta}^t)$;
  if $\boldsymbol{\Delta}_i^t$ *is not null* **then**
    Add $\boldsymbol{\Delta}_i^t$ to $\mathcal{U}$;
  end
end
$\boldsymbol{\theta}^{t+1} \leftarrow \boldsymbol{\theta}^t + \frac{1}{|\mathcal{U}|}\sum_{\boldsymbol{\Delta}\in\mathcal{U}} \boldsymbol{\Delta}$;
ClientUpdate$(i, \boldsymbol{\theta}^t)$**:**
$\boldsymbol{\theta}_i \leftarrow \boldsymbol{\theta}^t$;
for $e = 1$ *to* E **do**
  Train $(f_{\boldsymbol{\theta}_i}, g_{\boldsymbol{\phi}_i})$ on local data $\mathcal{D}_i$;
end
$\boldsymbol{\Delta}_i^t \leftarrow \boldsymbol{\theta}_i - \boldsymbol{\theta}^t$;
if $t \pmod R == 0$ **then**
  $\boldsymbol{u}_i^t \leftarrow \boldsymbol{\Delta}_i^t + \boldsymbol{e}_i^{t-1}$; // Error feedback
  $\tilde{\boldsymbol{u}}_i^t \leftarrow \mathcal{T}_K(\boldsymbol{u}_i^t, K)$; // Sparsification
  $\boldsymbol{e}_i^t \leftarrow \boldsymbol{u}_i^t - \tilde{\boldsymbol{u}}_i^t$; // Cache error
  $\hat{\boldsymbol{u}}_i^t \leftarrow$ Quantize$(\tilde{\boldsymbol{u}}_i^t)$; // Quantization
  return Decode$(\hat{\boldsymbol{u}}_i^t)$ to server;
end
else
  return null;
end

   - **Top-K Sparsification:** The client retains only the K elements of $\boldsymbol{u}_i^t$ with the largest absolute values and sets the remaining ones to zero. The new sparsification error is defined as

$$\boldsymbol{e}_i^t = \boldsymbol{u}_i^t - \mathcal{T}_K(\boldsymbol{u}_i^t, K), \tag{9}$$

     where $\mathcal{T}_K(\cdot, K)$ denotes the operation that preserves the K largest-magnitude entries and zeroes out the rest. This residual $\boldsymbol{e}_i^t$ is cached for the next round.
   - **8-bit Symmetric Quantization:** The non-zero elements of the sparse vector are quantized to 8-bit integers. The client transmits only the indices of these elements, their 8-bit values, and a single scaling factor for reconstruction.
4) **Periodic Synchronization:** The compression and upload steps are performed only once every R rounds. In the intermediate $R-1$ rounds, clients train locally without communicating with the server, drastically reducing the total number of communication rounds.
5) **Server Aggregation:** The server collects the compressed updates, de-quantizes and decodes them, and aggregates them via averaging to update the global backbone model to $\boldsymbol{\theta}^{t+1}$. Optionally, the server can maintain an Exponential Moving Average (EMA) of the global parameters to further enhance training stability.

In summary, the EPFL-REMNet training algorithm as outlined in Algorithm 1 for a single round.

TABLE I: Overall Performance Comparison Across Three Non-IID Scenarios.

Scenario	Method	RMSE (micro) ↓	RMSE (macro) ↓	MAE (macro) ↓	Communication (MB) ↓
Light	FedAvg	6.66	6.63	5.22	86689.5
	FedFly	7.73	9.94	8.21	182111.6
	PFedgb	7.51	7.37	5.89	20610.4
	EPFL-REMNet	**3.77**	**3.75**	**1.83**	**13437.6**
Medium	FedAvg	4.10	3.96	3.20	86689.5
	FedFly	4.71	4.71	3.77	182111.6
	PFedgb	5.16	5.03	4.02	20610.4
	EPFL-REMNet	**1.37**	**1.33**	**0.88**	**7043.2**
Heavy	FedAvg	6.56	6.27	4.91	86689.5
	FedFly	4.57	4.58	3.88	182111.6
	PFedgb	5.67	5.63	4.76	20610.4
	EPFL-REMNet	**2.07**	**2.02**	**1.40**	**21630.7**

IV. Experimental Evaluation

A. Experimental Setup

Dataset and Preprocessing: We use the public RadioMapSeer dataset [22], which provides rasterized path loss and received power maps for multiple base stations. We developed a custom script to process these maps into a unified format. For each data point, we extract the normalized 2D coordinates and the signal strengths for four base stations (1, 2, 3, 4), which are mapped from pixel values to a dB scale. These are combined with P =100 environmental features to form a 102-dimensional input vector, with the signal strengths as the label.

Non-IID Scenario Generation: To simulate real-world statistical heterogeneity, we partition the data into three scenarios based on signal strength variability. We define a heterogeneity metric for each location as $H = \text{std}(s_1, s_2, s_3, s_4)$. Using the 33rd and 66th percentiles of H (q_{33}, q_{66}), we create three scenarios:

- **Light Non-IID:** $H \leq q_{33}$. Characterized by smooth signal transitions and low variability between base stations.
- **Medium Non-IID:** $q_{33} < H \leq q_{66}$. Represents areas with moderate signal fluctuations and increasing data distribution disparity.
- **Heavy Non-IID:** $H > q_{66}$. Corresponds to complex terrains like urban canyons with sharp signal drops and highly skewed data distributions.

Within each scenario, we partition the map into a 10×9 grid to create 90 virtual clients. Each client's dataset is primarily composed of samples from its corresponding grid cell, with a small number of samples from neighboring cells, thus preserving geographical locality.

Evaluation Metrics:

- **Accuracy:** Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), reported as both micro-average (computed over all test samples) and macro-average (averaged over per-client results).
- **Fairness and Robustness:** Per-base-station RMSE to assess performance consistency across different propagation conditions.
- **Communication Cost:** Total cumulative uplink communication volume in Megabytes (MB) from the start to the end of training.

Baseline Methods:

- **FedAvg** [18]: The standard federated averaging algorithm, training a single global model.
- **FedFly** [19]: An edge FL framework designed for mobile environments, representing a standard edge protocol.
- **PFedgb** [20]: A PFL method based on Gaussian processes that uses a shared deep kernel for personalization.
- **EPFL-REMNet (Ours)**: The proposed framework combining personalization and communication efficiency.

B. Main Results: Performance Across Heterogeneity Levels

Table I summarizes the core performance metrics for all methods across the three Non-IID scenarios. Several clear trends emerge from the data.

Impact of Non-IID and Necessity of Personalization: As the degree of Non-IID increases from light to heavy, the performance of non-personalized methods like FedAvg deteriorates significantly, with its macro RMSE reaching 6.27 in the heavy scenario. This strongly validates that for tasks with high geographical correlation, such as constructing a radio environment digital twin, personalization is not an option but a necessity to accurately capture local environmental nuances.

Effectiveness of EPFL-REMNet's Personalization Strategy: Among personalized methods, EPFL-REMNet substantially outperforms PFedgb. In the heavy scenario, EPFL-REMNet's macro RMSE (2.02) is approximately 64% lower than that of PFedgb (5.63). This indicates that the "shared backbone + private head" architectural split is a more effective personalization strategy for this task than sharing a non-parametric kernel.

Overall Superiority of EPFL-REMNet: The EPFL-REMNet framework achieves the highest digital twin fidelity (i.e., mapping accuracy) in all scenarios, with the performance gap widening as heterogeneity increases. In the medium scenario, its macro RMSE (1.33) is 66% lower than that of the next best method, FedAvg (3.96). Crucially, it achieves this superior fidelity with a communication cost far below that of FedAvg and FedFly, demonstrating an exceptional fidelity-communication trade-off.

C. Convergence and Communication Dynamics

Fig. 2 would visually represent the convergence behavior and communication costs. The plots would show EPFL-REMNet's

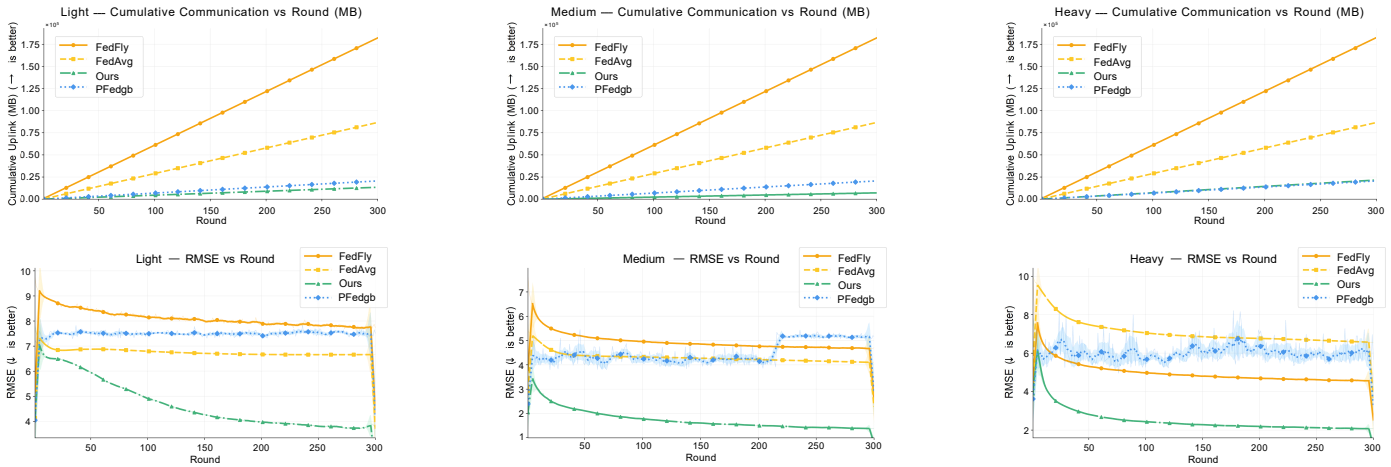

Fig. 2: Convergence curves and cumulative communication volume curves of each method in three-level Non-IID scenarios.

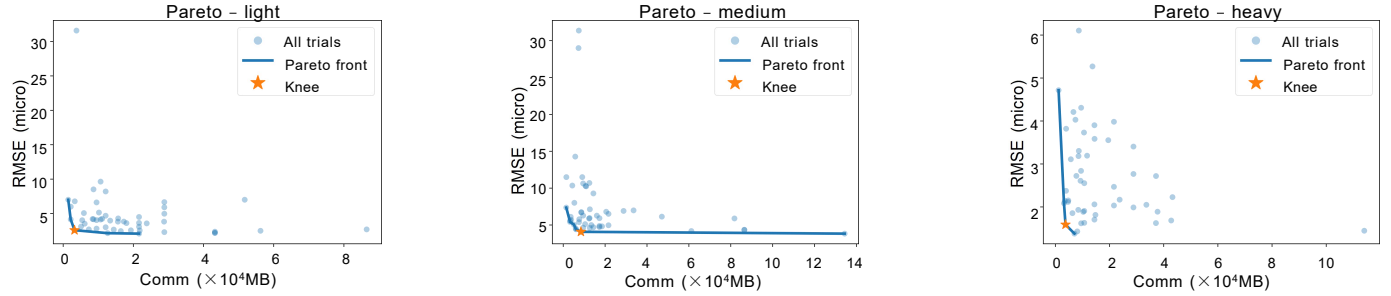

Fig. 3: Precision-communication Pareto frontier graph of EPFL-REMNet in three-level Non-IID scenarios.

macro RMSE curve rapidly descending to a much lower error floor compared to all baselines, indicating faster convergence and higher final accuracy. Concurrently, its cumulative communication curve would exhibit the flattest slope, visually confirming the high efficiency of its communication protocol. In contrast, baselines like FedAvg would show convergence to higher error levels with steeply rising communication costs.

D. Accuracy-Communication Pareto Frontier Analysis

Plotting the final macro RMSE against the total communication cost for all methods (as would be done in Figure 3) provides a clear visualization of the accuracy-communication trade-off. In such a plot, EPFL-REMNet would consistently occupy the bottom-left region, establishing a new and dominant Pareto frontier. This demonstrates that it offers a fundamentally better trade-off than the alternatives, achieving higher accuracy for a given communication budget or requiring less communication for a given accuracy target.

E. Fairness and Robustness Analysis

Table II disaggregates the RMSE by base station, providing insight into the model's robustness across different signal conditions. The data shows that while methods like FedAvg and FedFly exhibit significant performance variance across base stations and scenarios (e.g., FedFly's RMSE for base station 2 in the light scenario is 13.73), EPFL-REMNet maintains remarkably stable and low errors across all base stations and all levels of heterogeneity. The range of RMSE values for EPFL-REMNet within any given scenario is extremely narrow (e.g., 0.29 dB in the light scenario), and the variation across scenarios is also minimal. This indicates that the personalization mechanism not only improves average performance but also ensures fairness and consistency, making the system more reliable for practical deployment. Visualizing the distribution of RMSEs across all 90 clients via a box plot would further show that EPFL-REMNet has a much smaller inter-quartile range, confirming its ability to significantly uplift the performance of the worst-performing "long-tail" clients.

TABLE II: RMSE (Per BS) Analysis for Fairness and Robustness.

Scenario	Method	RMSE_A	RMSE_B	RMSE_C	RMSE_D
Light	FedAvg	6.73	6.64	7.02	6.12
	FedFly	10.73	13.73	7.85	7.42
	PFedgb	7.54	7.34	7.68	6.90
	EPFL-REMNet	**3.75**	**3.83**	**3.87**	**3.58**
Medium	FedAvg	4.26	3.50	4.11	3.95
	FedFly	4.66	4.66	4.76	4.75
	PFedgb	5.30	4.68	5.23	4.89
	EPFL-REMNet	**1.29**	**1.24**	**1.38**	**1.39**
Heavy	FedAvg	6.58	6.11	6.11	6.26
	FedFly	4.47	4.64	4.60	4.56
	PFedgb	5.63	5.27	5.25	5.80
	EPFL-REMNet	**2.08**	**1.77**	**2.18**	**2.04**

F. Ablation Study: Deconstructing EPFL-REMNet

To quantify the contribution of each component in EPFL-REMNet, we conducted an ablation study in the challenging heavy Non-IID scenario. The results are presented in Table III.

The results reveal several key findings. First, removing the personalized heads (w/o split-head) is the most detrimental

TABLE III: Performance Breakdown of EPFL-REMNet Variants Across Three Non-IID Scenarios.

Scenario	Variant (Method)	Micro RMSE ↓	Macro RMSE ↓	Comm (MB) ↓	ΔRMSE vs Full (%) ↓	ΔComm vs Full (%) ↓
Light	w/o split-head	5.6414	5.6115	2162.4	312.17% / 323.13%	-69.30%
	w/o periodic sync	6.6307	6.5987	2332.8	384.44% / 397.56%	-64.04%
	w/o Top-K	3.7590	3.7417	10811.8	174.64% / 182.14%	53.51%
	w/o quantization	5.4228	5.3963	21620.1	296.19% / 307.05%	206.96%
	w/o EMA	5.5410	5.5129	3458.8	305.00% / 315.70%	-50.89%
	Full	**1.3687**	**1.3262**	**7043.2**	**0.0%**	**0.0%**
Medium	w/o split-head	2.4123	2.3383	2162.4	-36.04% / -37.73%	-83.90%
	w/o periodic sync	4.0384	3.9253	2332.8	7.07% / 4.54%	-82.63%
	w/o Top-K	1.5782	1.5362	10811.8	-58.16% / -59.09%	-19.54%
	w/o quantization	2.4224	2.3473	21620.1	-35.77% / -37.49%	60.90%
	w/o EMA	2.4751	2.4005	3458.8	-34.38% / -36.07%	-74.26%
	Full	**3.7717**	**3.7548**	**13437.6**	**0.0%**	**0.0%**
Heavy	w/o split-head	6.2863	6.0254	2332.8	203.27% / 198.60%	-89.22%
	w/o periodic sync	2.4999	2.4311	10811.8	20.61% / 20.48%	-50.02%
	w/o Top-K	4.5657	4.4361	21620.1	120.27% / 119.84%	-0.05%
	w/o quantization	4.5772	4.4326	3458.8	120.82% / 119.66%	-84.01%
	w/o EMA	2.7651	2.6893	2162.4	33.40% / 33.27%	-90.00%
	Full	**2.0728**	**2.0179**	**21630.7**	**0.0%**	**0.0%**

change, causing the macro RMSE to skyrocket by nearly 200%. This confirms that personalization is the foundational and most critical element for tackling data heterogeneity.

A second, and highly significant, finding is the catastrophic performance degradation observed when removing the communication compression components. Disabling Top-K sparsification or 8-bit quantization increases the macro RMSE by approximately 120%. This counter-intuitive result suggests that these compression techniques do more than just save bandwidth; they play a crucial, positive role in the learning process itself, a point we will analyze further in the Discussion section. Finally, removing periodic synchronization (i.e., communicating more frequently) and EMA smoothing also leads to notable performance drops (20-33%), underscoring their importance for stabilizing training. The study confirms that the components of EPFL-REMNet are complementary and their synergy is key to achieving the final "high-accuracy, low-communication" goal.

V. Conclusion

This paper proposed EPFL-REMNet, an efficient personalized federate framework for 6G heterogeneous radio environment digital twin. By synergistically co-designing a "shared backbone + local personalized heads" architecture with novel communication efficiency pipelines—including Top-K sparsification, error feedback, quantization, and periodic updates, our framework effectively addresses the dual challenges of strong Non-IID data and constrained uplink bandwidth, which makes the digital twin equitable and robust across the entire 6G network. Our results highlight the necessity of co-optimizing personalization and communication efficiency and confirm that aggregating only the backbone updates is an effective strategy for maintaining training stability under high compression ratios.

References

[1] J.-H. Lee, J. Lee, S.-H. Lee, and A. F. Molisch, "A scalable and generalizable pathloss map prediction," *IEEE Transactions on Wireless Communications*, 2024.

[2] Ç. Yapar *et al.*, "Overview of the first pathloss radio map prediction challenge," *IEEE Open Journal of Signal Processing*, 2024.

[3] J. Lee, F. Solat, T. Y. Kim, and H. V. Poor, "Federated learning-empowered mobile network management for 5G and beyond networks: From access to core," *IEEE Communications Surveys & Tutorials*, vol. 26, no. 3, pp. 2176–2212, 2024.

[4] I. Fine *et al.*, "Fine-tuning approach to configuration and data selection for path loss prediction in different geographical environments," in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2021.

[5] G. C. Author, "Geographical clustering of path loss modeling for wireless emulation in various environments," in *Conference on Wireless Emulation*, 2018.

[6] A. Spiridonoff, K. Mishchenko, and P. Richtárik, "From local SGD to one-shot averaging," in *Advances in Neural Information Processing Systems*, 2021.

[7] L. Collins, H. Hassani, S. M. M. Tabei, and A. Atkeson, "FedRep: Federated Representation Learning," in *International Conference on Machine Learning*. PMLR, 2021.

[8] T. Li, S. Hu, A. Beirami, and V. Smith, "Ditto: Fair and Robust Federated Learning Through Personalization," in *International Conference on Machine Learning*. PMLR, 2021.

[9] X. Li, M. Zhang, and J. Luo, "FedBN: Federated Learning on Non-IID Features via Local Batch Normalization," in *International Conference on Learning Representations*, 2021.

[10] P. Richtárik *et al.*, "EF21: Better Error Feedback for Over-Parameterized Models," in *Advances in Neural Information Processing Systems*, 2021.

[11] E.-l. Sun, H. Luo, and F. Koushanfar, "Low-Precision Local Training is Sufficient for Federated Learning," in *Advances in Neural Information Processing Systems*, 2024.

[12] D. Gorbunov, D. Kovalev, and P. Richtárik, "MARINA: Faster Non-Convex Distributed Learning with Communication Compression," in *Advances in Neural Information Processing Systems*, 2021.

[13] C. Liu, Z. Xue, C. Liao, J. Kang, and G. Han, "DRL-enhanced vehicular edge caching addressing content dynamics and complex intersections," *IEEE Internet of Things Journal*, vol. 12, no. 2, pp. 1732–1745, 2025.

[14] B. Hu, W. Zhang, Y. Gao, J. Du, and X. Chu, "Multi-agent deep deterministic policy gradient-based computation offloading and resource allocation for isac-aided 6g v2x networks," *IEEE Internet of Things Journal*, 2024.

[15] L. Cao, S. Roy, and H. Yin, "Resource allocation in 5G platoon communication: Modeling, analysis and optimization," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 4, pp. 5035–5048, 2022.

[16] H. Yin, S. Roy, and L. Cao, "Routing and resource allocation for iab multi-hop network in 5G advanced," *IEEE Transactions on Communications*, vol. 70, no. 10, pp. 6704–6717, 2022.

[17] L. Zhang, L. Cao, D. Wei, M. Chen, Z. Chen, and S. Cui, "Cross-layer channel sounding optimization towards next-gen Wi-Fi: From model driven to data driven," *IEEE Transactions on Wireless Communications*, 2025.

[18] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*. PMLR, 2017.

[19] V. Varyam, A. Varghese, R. Du, H. Casanova, and L. Peterson, "Towards

migration in edge-based distributed federated learning," *IEEE Communications Magazine*, 2022.
[20] A. Dmitriev and P. Richtárik, "pFedGP: Personalized Federated Learning with Gaussian Processes," in *Advances in Neural Information Processing Systems*, 2021.
[21] D. Wei, X. Xu, S. Mao, and M. Chen, "Optimizing communication and device clustering for clustered federated learning with differential privacy," *IEEE Transactions on Mobile Computing*, pp. 1–14, 2025, early access.
[22] D. Author, "Dataset of pathloss and toa radio maps with localization application," *IEEE Dataport*, 2021.