**Proceedings Paper:**

Norouzi, Sajedeh, Samikwa, Eric, Rahmani, Mostafa orcid.org/0000-0002-7943-9977 et al. (2 more authors) (2025) Decentralized Federated Learning for GNN-Based Channel Estimation With DM-RS in O-RAN. In: IEEE International Conference on Machine Learning for Communication and Networking. IEEE International Conference on Machine Learning for Communication and Networking, 26-29 May 2025 IEEE , ESP

# Decentralized Federated Learning for GNN-Based Channel Estimation With DM-RS in O-RAN

Sajedeh Norouzi[1], Eric Samikwa[1], Mostafa Rahmani[2], Torsten Braun[1], Alister Burr[2]
[1]Institute of Computer Science, University of Bern, Bern, Switzerland.
[2]School of Physics, Engineering and Technology, University of York, York, England.
{sajedeh.norouzi, eric.samikwa, torsten.braun}@unibe.ch -{rahmani.mostafa, alister.burr}@york.ac.uk

***Abstract*—This paper introduces GrINet, a Graph-based neural NETwork for channel estimation and Interpolation using Demodulation Reference Signals (DM-RS) to enhance estimation accuracy in wireless communication systems. GrINet models each Resource Element (RE), the smallest resource unit in 5G New Radio, as a node in a graph, with edges connecting DM-RS-enriched nodes, enabling effective modeling and processing of complex channel conditions. Building on this, we propose decentralized Federated GrINet (FGrINet), a hierarchical framework that combines Federated Learning (FL) at the base station (BS) level with decentralized collaboration across BSs. Locally, each BS employs FL to optimize CE models using data from its connected users under unique channel conditions. Globally, BSs share and aggregate their locally trained models in a decentralized manner, enabling collaborative learning without relying on centralized orchestration. This two-tiered approach allows BSs to operate in favorable conditions to assist others, enhancing adaptability to diverse channel environments. FGrINet can be implemented as an xApp architecture, aligning with O-RAN's goals of using distributed machine learning for intelligent, real-time RAN optimization. Our simulation involving multiple BSs, diverse channel profiles, and varying user mobility demonstrates that FGrINet reduces local training time, enhances CE accuracy, and achieves low mean squared error (MSE).**

***Index Terms*—Graph Neural Network (GNN), Federated Learning (FL), Wireless Channel Estimation, Demodulation Reference Signals (DM-RS), O-RAN**

## I. Introduction

Channel Estimation (CE) is a fundamental process in modern wireless communication systems, enabling accurate signal detection and robust system optimization. It is particularly vital in 5G scenarios that demand high throughput, low latency, and reliable performance in single-user and multi-user MIMO configurations. In 5G New Radio (NR), CE is achieved through Demodulation Reference Signal (DM-RS), which utilize orthogonal frequencies and orthogonal cover codes to ensure robust pilot transmission across time and frequency resources [1]. The accuracy of CE directly impacts key network metrics such as spectral efficiency, signal robustness, and user quality of service, especially in dynamic and interference-prone environments.

Emerging architectures such as Open Radio Access Network (O-RAN) offer new opportunities to enhance CE through flexible, modular, and software-driven solutions. O-RAN disaggregates traditional RAN systems into Central Units (CUs), Distributed Units (DUs), and Radio Units (RUs), connected via open interfaces to enable multi-vendor interoperability and scalability [2]. O-RAN introduces programmable Radio Intelligent Controllers (RICs), which host xApps and rApps that leverage Artificial Intelligence (AI) and Machine Learning (ML) to dynamically optimize network operations [3]. This modular and programmable architecture is particularly suited for integrating advanced CE techniques to address challenges posed by heterogeneous and high-mobility deployments.

While traditional CE methods like Least Squares (LS) and Minimum Mean Squared Error (MSE) are computationally efficient, they often fail in highly dynamic or complex channel conditions. Recent advancements in ML have enabled data-driven approaches that significantly improve CE [4]. Techniques like Convolutional Neural Networks (CNN) [5] and Long Short-Term Memory (LSTM) networks [6] improved accuracy by leveraging the spatial and temporal patterns of wireless channels. However, these methods typically rely on grid- or sequence-based representations, which are insufficient to capture the complex relational structures in wireless channels fully. Centralized ML-based CE frameworks face challenges such as communication overhead, privacy concerns, and scalability issues, particularly in large-scale deployments.

Federated Learning (FL) addresses many of these challenges by enabling decentralized model training, where local devices or Base stations (BS) share model parameters rather than raw data, thereby preserving user privacy and reducing communication overhead [7]. Recent studies [8], [9] have proposed FL-based CE frameworks that achieve high performance while addressing privacy concerns. However, existing FL-based approaches lack mechanisms for collaborative knowledge sharing between BSs in diverse environments, limiting their adaptability to dynamic deployment scenarios. Therefore, it is essential to determine how decentralized learning frameworks can be effectively leveraged to enhance knowledge sharing across BSs while preserving privacy and ensuring adaptability to dynamic and heterogeneous environments.

The reliance on simplistic data representations significantly limits the effectiveness of existing CE methods in capturing the complex characteristics of wireless channels. Graph Neural Networks (GNN)s have emerged as a powerful tool for modeling relational and structural data. Unlike traditional grid-based approaches, GNNs represent wireless channels as graphs, where nodes and edges capture spatial and temporal relationships. This relational modeling capability makes GNNs particularly well-suited for CE [10]. However, GNNs have yet to see widespread adoption for CE. This paper addresses three key research questions:

1) How can algorithms be designed to effectively leverage graph-based representations of wireless channels?
2) How can the integration of GNNs with decentralized learning paradigms, such as FL, overcome current challenges?
3) How can collaborative learning approaches improve overall CE performance?

To address these questions, we introduce a novel framework that integrates GNNs with FL within the O-RAN architecture to enhance CE. The main contributions of this work are as follows:

1) **Graph-Based Wireless Channel Estimation & Interpolation:** We propose a graph-based representation for wireless channels, where Resource Element (RE)s are modeled as nodes, referred to as *RE* nodes throughout this paper, and each node is connected to *DM-RS* nodes with weighted edges. This approach captures complex spatial and temporal correlations, surpassing the limitations of grid- or sequence-based methods.
2) **Two-Tier Hierarchical Framework:** We propose a two-tier framework to enhance scalability and adaptability in decentralized CE. At the first tier, user-level models are aggregated locally at each BS, enabling the capture of environment-specific channel characteristics tailored to the unique conditions of each BS. At the second tier, these locally aggregated models are further combined across multiple BSs, facilitating collaborative knowledge sharing and enabling BSs in diverse or challenging conditions to benefit from the network's collective expertise.
3) **Efficient Fine-Tuning Mechanism:** To reduce computational and communication overhead, we introduce a fine-tuning mechanism where pre-trained GNNs weights at each BS are updated by received DM-RS symbols from new user devices.

Through this integration of GNNs, FL, and O-RAN, the proposed framework addresses key limitations of existing CE approaches, including their reliance on simplistic data representations, lack of scalability, and centralized architectures. Simulation results demonstrate that the framework achieves notable accuracy, robustness, and efficiency in dynamic and heterogeneous deployment scenarios, setting the stage for next-generation wireless communication systems.

Fig. 1 illustrates the architecture of the proposed system, which integrates multiple BSs operating under diverse channel models and serving users with varying mobility profiles. Users transmit DM-RS to the BSs, which use these symbols to train the Graph-based Neural Network for Channel Estimation and interpolation (GrINet) model Fig. 1A. In the first tier of the architecture, each BS performs local FL by aggregating individual models trained on its respective users' data Fig. 1B. In the second tier, these locally aggregated models are shared in a decentralized manner with other BSs for collaborative improvement Fig. 1C. This decentralized sharing eliminates the need for BSs to train GrINet from scratch for new users. Instead, BSs can update the pre-trained model using the DM-RS of the new users. The system employs a two-tier aggregation process: first, local aggregation at each BS combines the models trained for individual users. Subsequently, decentralized aggregation across BSs refines the overall model, enhancing performance and scalability.

Fig. 1A provides an overview of the GrINet architecture. In each time slot, BSs represent the received DM-RS REs as *DM-RS* nodes in a graph, while the remaining REs requiring CE are represented as RE nodes. GrINet leverages this graph-based structure by connecting each *RE* node to all *DM-RS* nodes using weights. The input to GrINet comprises features of the *DM-RS* nodes, including the channel estimates' real and imaginary components. These features are passed through three stages involving Rectified Linear Unit (ReLU) activation functions and graph convolution layers, allowing GrINet to extract enhanced feature representations. The output of GrINet is the estimated channel for all nodes in the network for a given user, representing the overall channel state.

## II. System Model

### A. Frame Structure of 5G NR

In 5G NR, downlink and uplink data are transmitted in 10 ms frames, each divided into 10 subframes of 1 ms. Subframes consist of variable slots determined by the subcarrier spacing ($\Delta f$), allowing flexibility based on cell size, latency, and interference requirements. Each slot consists of 14 Orthogonal Frequency Division Multiplexing (OFDM) symbols when using a normal Cyclic Prefix (CP) or 12 OFDM symbols when using an extended CP. RE, the smallest resource unit in 5G NR, are grouped into Physical Resource Blocks (PRB) of 12 neighboring REs per symbol. The flexible frame structure and multiple OFDM numerologies improve adaptability and performance. As shown in Fig. 2, this paper utilizes Numerology 0 with an extended CP, configured with four PRBs and 1, 2, or 4 DM-RS symbols per slot.

### B. Network Topology

This paper considers a typical wireless communication system, where multiple BSs serve single-antenna users, denoted as $\mathcal{U}_b$. During the uplink phase, each user transmits DM-RS symbols along with data to the BS. At the BS, CE is performed by training the GrINet using the received DM-RS. At each BS, GrINet is trained for individual users,

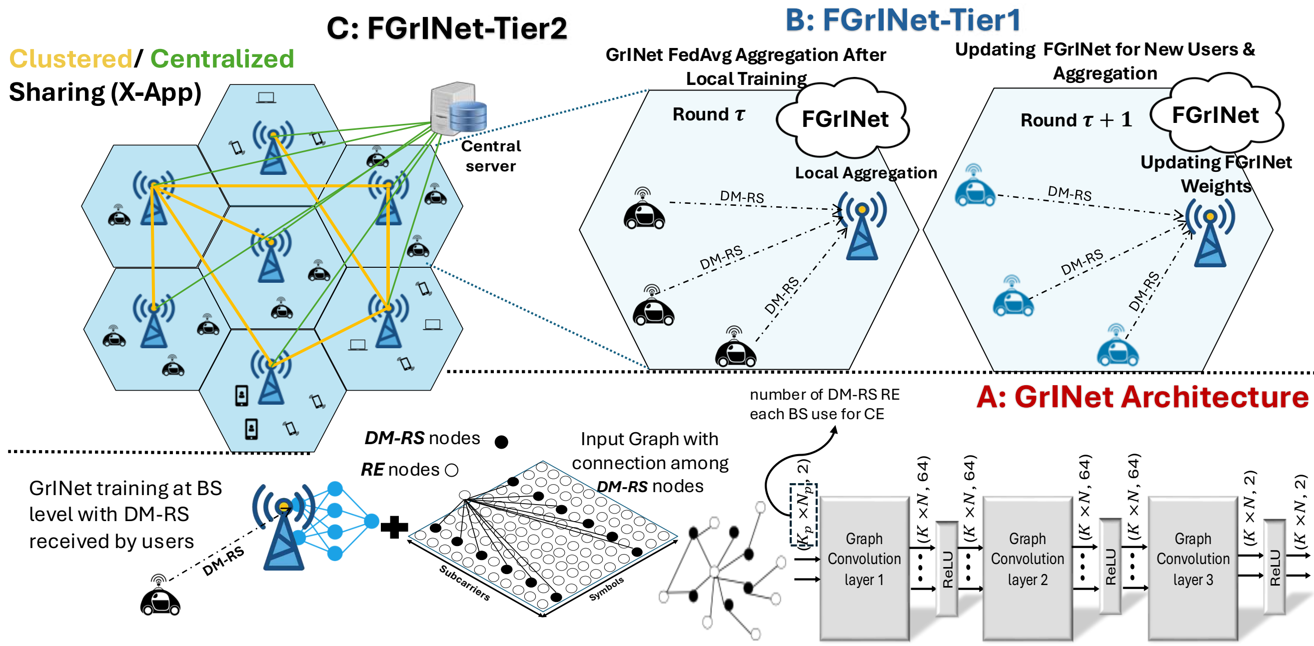


Fig. 1: System Model Overview: including GrINet architecture, GrINet local training, GrINet federated averaging at the BSs (FGrINet-T1), decentralized sharing among BSs (FGrINet-T2), its mapping to the O-RAN framework.

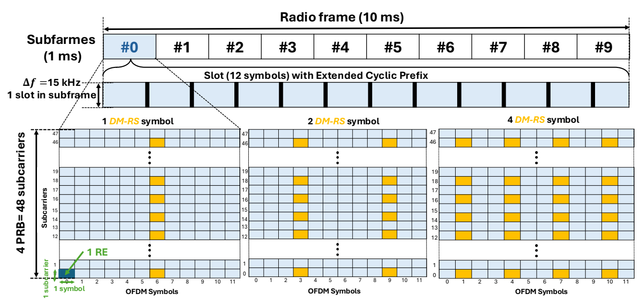


Fig. 2: Frame structure in 5G NR

and the resulting models are aggregated using Federated Averaging (FedAvg) [11] to construct a Federated GrINet (FGrINet) model. To address the heterogeneity of channel configurations, we consider a combination of channel settings across BSs, including two models that utilize different numbers of DM-RS symbols and support users with varying mobility speeds.

The network consists of multiple BSs, denoted as $b \in \mathcal{B} = \{1, 2, \cdots, B\}$, where $B = 12$ represents the total number of BSs in the system. We assume the use of 3GPP-defined Clustered Delay Line (CDL) channel models, specifically types A and C [12], to capture different propagation environments. Additionally, two user mobility patterns are considered, with speeds of 3 km/h and 300 km/h, representing low and high mobility scenarios, respectively. Users transmit a varying number of DM-RS for CE, ranging from 1 to 4 per slot, with each having a duration of $71\mu s$. This variation reflects diverse environmental and communication conditions. Let $u_{bm}$ represent the $m$-th user associated with the $b$-th BS, where $b \in \mathcal{B}$ and $m \in \mathcal{U}_b$. Here, $\mathcal{U}_b$ denotes the set of users served by BS $b$. The wireless channel for each $u_{bm}$ in the uplink transmission is represented by the channel matrix $\mathbf{H}_{u_{bm},b} \in \mathbb{C}^{K \times N}$, where $K$ is the total number of subcarriers and $N$ is the total number of symbols. The indices $k \in \mathcal{K} = \{1, 2, \cdots, K\}$ denote the subcarrier index, and $n \in \mathcal{N} = \{1, 2, \cdots, N\}$ denote the symbol index.

The estimated channel corresponds to the received signal from the user $u_{bm}$ at BS $b$ at DM-RS positions is given by $\hat{\mathbf{H}} \in \mathbb{C}^{K_p \times N_p}$, where the number of DM-RS symbols, $N_p$, varies depending on the channel configuration. We assume a half-DM-RS density in frequency, meaning that the DM-RS symbols are spread across half of the available subcarriers, specifically on every second subcarrier, i.e., $k_p = \{2, 4, \cdots, \lfloor K/2 \rfloor \cdot 2\}$. By incorporating these diverse channel models and configurations, our system effectively accounts for the heterogeneity of wireless channels. The final step of CE involves interpolation, where the channel estimates are extended to cover all subcarriers and the entire OFDM resource grid. Instead of performing simple interpolation sequentially first in the frequency domain and then in the time domain, we integrate the interpolation process directly into the GrINet framework. This integrated approach allows for more accurate and efficient CE, as detailed in the following subsection III-A.

## III. GrINet & FGrINet

### A. Graph Estimation & Interpolation Model: GrINet

This subsection introduces GrINet, by defining nodes, edges, and their associated features. With graph representation of wireless channels, estimation and interpolation can be integrated into a unified framework, improving wireless channel prediction accuracy. A graph is a triplet $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$, which includes vertices or nodes $\mathcal{V}$, edges $\mathcal{E}$, and weights $\mathcal{W}$. Each node $v \in \mathcal{V}$ corresponds to a specific combination of a subcarrier and symbol. The node $v$ is characterized by its feature vector $x_v \in \mathbb{R}^2$, where the dimensions represent the real and imaginary parts of the channel coefficient. Initially, the features of nodes corresponding to DM-RS are set to the estimated channel values. For non-*DM-RS* nodes (*RE* nodes), the feature values are initialized to zero and subsequently refined during the training process.

To establish relationships between nodes, edges are created between every node $v_i$, where $i \in \mathcal{V} = \{1, 2, \cdots, K \times N\}$ and all *DM-RS* nodes $d_j$, where, $j \in \mathcal{V_P} = \{1, 2, \cdots, K_p \times N_p\} \subset \mathcal{V}$, forming the set of edges $\mathcal{E}$. Let the Subcarrier Time Pair (S-T pair) of node $v_i$ be $(k_{v_i}, n_{v_i})$, where $k_{v_i}$ represents the subcarrier index and $n_{v_i}$ represents the symbol index. Similarly, let $(k_{d_j}, n_{d_j})$ denote the S-T pair of node $d_j$. The edge weight $w_{v_i,d_j}$ is defined as:

$$w_{v_i,d_j} = 1/\sqrt{|k_{v_i} - k_{d_j}|^2 + |n_{v_i} - n_{d_j}|^2} \tag{1}$$

If $v_i = d_j$, we assign $w_{v_i,d_j} = 1/\epsilon$, where constant $\epsilon$ is used to prevent division by zero and to ensure high-weighted self-connection for the *DM-RS* nodes. The resulting edge set $\mathcal{E}$ and weight set $\mathcal{W}$ are given by:

$$\mathcal{E} = \{(v_i, d_j) | v_i \in \mathcal{V}, d_j \in \mathcal{V_P}\} \tag{2}$$

$$\mathcal{W} = \{w_{v_i,d_j} | e_{v_i,d_j} \in \mathcal{E}\} \tag{3}$$

In this model, BSs are provided DM-RS by users for CE. After receiving these signals and estimating the channel, each BS constructs a graph $\mathcal{G}$ that represents the channel for the corresponding user requiring estimation. The features of the nodes in the graph are initialized based on initial estimation for received DM-RS. According to our approach, every *RE* node is connected to *DM-RS* nodes with weights proportional to their relative positions to the *DM-RS* nodes. Once the graph $\mathcal{G}$ is constructed, a GNN is employed to refine and predict the wireless channel for each node. Specifically, we utilize a Graph Convolutional Network (GCN)-based architecture comprising three layers. Each layer consists of three key steps:

1) Self-Loop Addition: Self-loops are added to each node, enabling nodes to aggregate information from their own features and connected nodes.
2) Degree Normalization: Let $A$ be the degree matrix, where $a_{v_i}$ is the degree of node $v_i$, including self-loops. The normalized degree matrix is computed as: $\tilde{A}_{v_i v_i} = a_{v_i}^{-\frac{1}{2}}$.
3) Aggregation and Transformation: for each node $v_i$, information can be aggregated from *DM-RS* nodes $d_j \in \mathcal{V_P}$ by: $\hat{\mathbf{x}}_{v_i} = \sum_{d_j \in \mathcal{V_P}} \tilde{A}_{v_i v_i} \tilde{A}_{x'_j} \hat{\mathbf{x}}_{d_j} W$, where $\bar{\mathbf{x}}_{v_i}$ is the updated feature for node $v_i$, $\hat{\mathbf{x}}_{d_j}$ is the feature node of $d_j$, $W$ is the learnable weight matrix for the layer.

The CE model can be described as a series of transformations consisting of three GCN layers. The first GCN layer takes input features $\hat{\mathbf{x}}_{d_j}$ corresponding to the *DM-RS* node features (which include the real and imaginary parts of the estimated channel for these nodes:

$$\hat{\mathbf{x}}_{d_j} = \begin{bmatrix} \mathrm{Re}\left(\hat{\mathbf{H}}(k_p, n_p)\right) \\ \mathrm{Im}\left(\hat{\mathbf{H}}(k_p, n_p)\right) \end{bmatrix}, \\ \forall k_p \in \{2, 4, \cdots, \lfloor K/2 \rfloor \cdot 2\}, \forall n_p \in \mathcal{N}_p, \tag{4}$$

where $k_p$ and $n_p$ denote the indices for DM-RS subcarriers and symbols, respectively. After passing through the first GCN layer, the output is a feature matrix $\bar{\mathbf{x}}^{\mathbf{1}}_{v_i}$ of dimension 64, which captures the enhanced features of the channel for each node. This is followed by a ReLU activation and a second GCN layer, producing output $\bar{\mathbf{x}}^{\mathbf{2}}_{v_i}$, which is again a 64-dimensional feature matrix. After activation, the third GCN layer produces the final output, $\bar{\mathbf{x}}^{\mathbf{3}}_{v_i}$, which represents the estimated channel for each node. The output of the third layer represents the final CE for each node, where each node's output feature dimension equals 2, matching the required dimension for CE as the real and imaginary value of the channel coefficient. The features for each node $v_i$ after the third GCN layer are written as:

$$\bar{\mathbf{x}}^3_{v_i} = \begin{bmatrix} \mathrm{Re}\left(\bar{\mathbf{H}}(k, n)\right) \\ \mathrm{Im}\left(\bar{\mathbf{H}}(k, n)\right) \end{bmatrix}, \quad \forall k \in \{1, 2, \cdots, K\}, \quad \forall n \in \mathcal{N}, \tag{5}$$

where $\bar{H}(k, n)$ is the channel estimate for the $k$-th subcarrier and $n$-th symbol as predicted by GrINet.

### B. FGrINet: Tier1 & Tier2

In our proposed two-tier framework, the first tier occurs locally at each BS, where user-specific GrINet models are aggregated to construct a FGrINet. This process ensures that each BS consolidates the knowledge gained from its associated users. In the second tier, the FGrINet models from individual BSs are shared and collaboratively refined in a decentralized manner across the network. This aggregation occurs directly between BSs without a central server. A local federated round at the BS is denoted by $t \in T$, during which BSs collaboratively train user-specific models. For a BS $b$ serving user $u_{bm}$, the GrINet model is optimized by minimizing the MSE loss, which is defined as:

$$\mathcal{L}_{u_{bm}}(\theta^t_{u_{bm}}) = \frac{1}{|D_{u_{bm}}|} \sum_{(\hat{x}_{d_j}, x_{v_i}) \in D_{u_{bm}}} \ell(f(\hat{x}_{d_j}; \theta^t_{u_{bm}}), x_{v_i}) \tag{6}$$

In this model, $f(\hat{x}_{d_j}; \theta^t_{u_{bm}})$ represents the output of GrINet for the input $\hat{x}_{d_j}$ (feature vector corresponding to DM-RS 4). The target channel value $x_{v_i}$ represents the true channel estimate at node $v_i$, encompassing all node features and capturing the entire channel. The local dataset for user $u_{bm}$, denoted by $D_{u_{bm}}$, contains the data used by the BS to train the user-specific GrINet model. At the end of each local training round, the BS aggregates the user-specific GrINet models using the FedAvg approach [11]. This aggregation results in the creation of FGrINet, which is the BS-specific model, denoted by $\theta_b^t$, the model update rule is given by:

$$\theta_b^{t+1} = \theta_b^t - \eta \sum_{u_{bm} \in \mathcal{U}_b} \frac{|D_{u_{bm}}|}{|D_b|} \nabla \mathcal{L}_{u_{bm}}(\theta_{u_{bm}}^t), \quad (7)$$

where $\eta$ is the learning rate for the aggregation step, $\nabla \mathcal{L}_{u_{bm}}(\theta_{u_{bm}}^t)$ is the gradient of the local loss function for user $u_{bm}$, $|D_b| = \sum_{u_{bm} \in \mathcal{U}_b} |D_{u_{bm}}|$ is the total dataset size for all users under BS $b$ participated in local training, $\mathcal{U}_b$ represents the set of users served by BS $b$. This update rule ensures that the BS's model is updated based on the aggregated gradient information from all associated users, with the influence of each user weighted by the size of their local dataset. The result is a model that reflects the collective learning progress of all users under BS $b$.

Once each BS has obtained its own specific model, a decentralized phase is initiated. During this phase, BSs share their updated models with one another through decentralized model sharing. The primary motivation for this sharing and aggregation is to improve model performance across BSs by leveraging the diversity in the data and varying channel conditions. As BSs exchange models, the accuracy of CE improves, and the convergence rate of subsequent local training rounds accelerates. In the next phase of local training, which corresponds to the following round of decentralized sharing, the BSs utilize the achieved FGrINet model for CE. When new DM-RS signals are received from new users, the BS uses the FGrINet model and updates it with the newly received DM-RS to perform accurate CE.

This process significantly reduces local training time, as the FGrINet model has inherited valuable information from previous rounds. In collaboration among BSs, several distinct aggregation methods are considered, which differ based on how weights are assigned to the BSs and whether model sharing is restricted to clusters or applied across the entire network. The aggregation methods are as follows:

**Weighted Clustered Sharing (W-Cl):** BSs are grouped into clusters, and weights are assigned to each BS based on the number of DM-RS used for CE. BSs with more DM-RS symbols contribute more to the aggregated model because their channel estimates are more accurate, $w_b$ represents the weight of the model from BS $b$. In each aggregation round $\tau \in \mathbb{N}_0$, BS $b$ aggregates the models of other BSs within the same cluster $\mathcal{C}_b \subseteq \mathcal{B}$, using a weighted averaging scheme. The update rule for the model at BS $b$ is expressed as:

$$\theta_b^{\tau+1} = \sum_{b' \in \mathcal{C}_b} \frac{q_{b'}}{\sum_{b'' \in \mathcal{C}_b} q_{b''}} \theta_{b'}^{\tau}, \quad (8)$$

where $q_{b'}$ is a quality metric for the model from BS $b'$. The quality metric depends on the number of DM-RS symbols each BS uses during local GrINet training. The term $\theta_b^{\tau+1}$ represents the updated model for BS $b$ after round $\tau$, while $\theta_{b'}^{\tau}$ is the model from BS $b'$ at the same round $\tau$.

**Uniform Clustered Sharing (U-Cl):** In each round $\tau \in \mathbb{N}_0$, BS $b$ aggregates the models of the BSs within the same cluster $\mathcal{C}_b \subseteq \mathcal{B}$, using uniform (same) weights. The update rule for the model at BS $b$ is expressed as:

$$\theta_b^{\tau+1} = \frac{1}{|\mathcal{C}_b|} \sum_{b' \in \mathcal{C}_b} \theta_{b'}^{\tau}, \quad (9)$$

where $|\mathcal{C}_b|$ is the number of BSs in cluster $\mathcal{C}_b$. The term $\theta_b^{\tau+1}$ represents the updated model for BS $b$ after round $\tau$, while $\theta_{b'}^{\tau}$ is the model from BS $b'$ at the same round $\tau$.

**Weighted Centralized Sharing (W-Ce):** In a weighted centralized manner, each BS aggregates models from all other BSs using a weighted averaging scheme; the update rule is:

$$\theta_b^{\tau+1} = \sum_{b' \in \mathcal{B}} \frac{q_{b'}}{\sum_{b'' \in \mathcal{B}} q_{b''}} \theta_{b'}^{\tau}, \quad (10)$$

where $\mathcal{B}$ is the set of all BSs and other parameters are the same with weighted clustered sharing.

**Uniform Centralized Sharing (U-Ce):** In a uniform centralized manner, each BS aggregates the model from all other BSs, but this time all models are given equal weight. The update rule is:

$$\theta_b^{\tau+1} = \frac{1}{|\mathcal{B}|} \sum_{b' \in \mathcal{B}} \theta_{b'}^{\tau}. \quad (11)$$

These methods outline how BSs update their models in decentralized aggregation rounds, either with or without weights, within clusters, or in a centralized manner, providing flexibility in balancing model accuracy and communication efficiency. The training process for GrINet and hierarchical FGrINet along with their interactions, is outlined in the algorithm 1. It is important to note that different users are assigned in each round.

## IV. Performance Evaluation

### A. *Channel Models for Link-Level Evaluations*

CDL models are defined to cover a wide frequency range from 0.5 GHz to 100 GHz, with a maximum bandwidth of 2 GHz. These models can be implemented through the generation of a Tapped Delay Line (TDL) model using spatial filters. To represent distinct channel profiles, three CDL models CDLA, CDLB, CDLC are used for Non-Line-of-Sight (NLOS) conditions, while CDL-D and CDL-E models are employed for Line-of-Sight (LOS) conditions. In this paper, we utilize data from 12 channel configurations, each containing 100 slots with a total of 100 users. The channel types employed are CDLA and CDLC, with two different user speeds and varying numbers of DM-RS symbols. The carrier frequency is set at 3.5 GHz, with 4 PRBs comprising 48 subcarriers and a delay spread of 300 ns. To evaluate the performance of the proposed work, we utilized a dataset generated for AI-based link-level CE research [13], following the specifications outlined in Table 7.7.1-1 and Table 7.7.1-3 of the 3GPP standard [12].

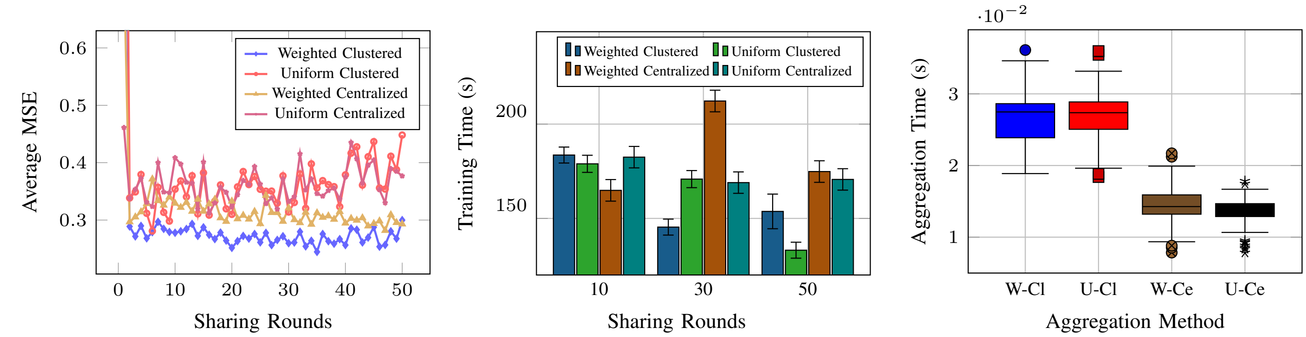


Fig. 3: (a) Average MSE of all BSs versus federated rounds, (b) Average training time across all BSs for decentralized sharing rounds 10, 30, and 50, (c) Aggregation time versus decentralized sharing rounds for each aggregation methods.

**Algorithm 1:** GrINet, FGrINet

**OFDM Uplink Establishment and Resource Grid Configuration:**
Configure uplink channels, initialize resource grid parameters.
**Network Configuration:**
Set up channel models, number of BSs $\mathcal{B}$, number of users, DMRS symbols, and their indices.
**for** *each aggregation round* $\tau \in \mathbb{N}_0$ **do**
  **for** *each local federated round* $t \in T$ **do**
    **for** *each BS* $b \in \mathcal{B}$ **do**
      **for** *each user in local training* $u_{bm} \in \mathcal{B}_b$ **do**
        **if** *FGrINet is available from previous aggregation round* **then**
          **Update FGrINet:**
          Reinitialize graph $\mathcal{G}$ with updated *DMRS* nodes features from new users 4
          Fine-tune the model with new DMRS features from the current training 6.
        **end**
        **else**
          **Train GrINet:**
          Initialize graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ with *RE* and *DMRS* nodes, edges 2, weights 3 and input the features for *DMRS* nodes 4
          Train GrINet model by minimizing MSE loss 6.
        **end**
      **end**
      **FGrINet-T1**: Federated averaging for all participants in local training.
    **end**
  **end**
  **FGrINet-T2**: BSs share their locally updated FGrINet models among themselves based on one of the four aggregation methods; W-Cl 8, U-Cl 9, W-Ce 10, U-Ce 11.
**end**

### *B. Simulation Results*

Fig. 3(a) shows the average MSE of all BSs versus decentralized sharing rounds for the four aggregation methods. The plot demonstrates weighted aggregation methods achieving significantly lower MSE values. This is because BSs using only 1 DM-RS symbol can benefit from the model estimations of other BSs, particularly those using 4 symbols for more accurate CE. The next comparison focuses on the clustered methods, highlighting that clustering BSs with similar channel profiles leads to more accurate estimations. The best performance is observed with weighted clustered sharing, while uniform centralized sharing yields the worst performance.

Fig. 3(b) shows the average training time across all BSs for specific decentralized sharing rounds, illustrating the trend of local training times. Uniform clustered sharing achieves the lowest training time, indicating that users in local training can more quickly fine-tune models aggregated using this method. Weighted clustered sharing also performs similarly, with clustering helping to optimize model training. In contrast, centralized sharing methods, both with and without weights, require more time to fine-tune the models. This is because BSs need to train for longer periods to adapt to models that aggregate information from all channel profiles, making the fine-tuning process more challenging.

Fig. 3(c) illustrates the aggregation time for models across sharing rounds. Clustered methods require slightly more time for aggregation compared to centralized learning methods. This difference arises because clustered aggregation involves multiple model specifics for each cluster, whereas centralized aggregation only requires a single model for all BSs.

## V. Conclusion

This paper presents an algorithm for CE and interpolation based on GNN, leveraging the inherent graph structure of wireless channel properties to achieve more accurate estimations. We consider various channel models, including different user speeds, and DM-RS symbol counts, and propose a decentralized, federated learning framework, FGrINet. In this scheme, BSs with better channel conditions, particularly those utilizing more DM-RS symbols for estimation, play a key role in improving overall model accuracy. Rather than retraining from scratch, BSs can fine-tune their models, leading to efficient learning. Among aggregation methods, clustered weighted aggregation in decentralized learning, which considers both model weights and clustering of BSs, delivers superior performance.


## Acknowledgment

The work presented in this paper was jointly funded by the Swiss National Science Foundation (grant number 219330) and the UK Department for Science, Innovation and Technology under project YO-RAN.

# References


[1] S. Sabapathy, J. S. Prabhu, S. Maruthu, and D. N. K. Jayakody, "Profuse channel estimation and signal detection techniques for orthogonal time frequency space in 6G epoch: A survey," *IEEE Access*, vol. 11, pp. 129 963–129 993, 2023.

[2] R. J. B. Pousa, G. T. Nguyen, R. Bassoli, and F. H. P. Fitzek, "Road to Dynamic Functional Split in Radio Access Networks," in *2024 IEEE 22nd Mediterranean Electrotechnical Conference (MELECON)*, 2024, pp. 844–849.

[3] M. Polese, L. Bonati, S. D'oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1376–1411, 2023.

[4] K. Weththasinghe, B. Jayawickrama, and Y. He, "Machine learning-based channel estimation for 5G new radio," *IEEE Wireless Communications Letters*, vol. 13, no. 4, pp. 1133–1137, 2024.

[5] A. M. Elbir and S. Coleri, "Federated learning for channel estimation in conventional and RIS-assisted massive MIMO," *IEEE transactions on wireless communications*, vol. 21, no. 6, pp. 4255–4268, 2021.

[6] S. Liu, C. Li, and J. Deng, "Real-Time Wireless Channel Prediction Based on Online Federated Learning," in *2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring)*. IEEE, 2024, pp. 1–6.

[7] V. Ardianto Nugroho and B. M. Lee, "A survey of federated learning for mmwave massive MIMO," *IEEE Internet of Things Journal*, vol. 11, no. 16, pp. 27 167–27 183, 2024.

[8] J. Kaur and M. A. Khan, "Wireless Channel Estimation using Federated Learning," in *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*. IEEE, 2024, pp. 1–6.

[9] L. Zhao, H. Xu, Z. Wang, X. Chen, and A. Zhou, "Joint channel estimation and feedback for mm-Wave system using federated learning," *IEEE communications letters*, vol. 26, no. 8, pp. 1819–1823, 2022.

[10] S. Norouzi, M. Rahmani, T. Braun, and A. Burr, "Channel Estimation in 5G NR MIMO Systems Using GraNet: A Graph Neural Network Framework," 2024. [Online]. Available: https://boris-portal.unibe.ch/handle/20.500.12422/191619

[11] E. Samikwa, A. Di Maio, and T. Braun, "DFL: Dynamic Federated Split Learning in Heterogeneous IoT," *IEEE Transactions on Machine Learning in Communications and Networking*, vol. 2, pp. 733–752, 2024.

[12] ETSI, "3GPP TR 38.901, "study on channel model for frequencies from 0.5 to 100 GHz", Release 16.1," Tech. Rep.

[13] O. R. Institute, "Wireless intelligence: Oppo research institute," https://wireless-intelligence.com/#/dataSet?id=2c92185c7e3f1aa4017e3f2c93e00001, accessed: Nov. 27, 2023.