

# Model-Driven Federated Learning for Channel Estimation in Millimeter-Wave Massive MIMO Systems

Qin Yi<sup>†</sup>, Ping Yang<sup>†</sup>, *Senior Member, IEEE*, Zilong Liu<sup>§</sup>, *Senior Member, IEEE*, Yiqian Huang<sup>†</sup>,  
and Saviour Zammit<sup>\*</sup>, *Senior Member, IEEE*

<sup>†</sup>National Key Laboratory of Wireless Communications,  
University of Electronic Science and Technology of China, Sichuan, China.

<sup>§</sup>School of Computer Science and Electronics Engineering, University of Essex, Colchester, United Kingdom.

<sup>\*</sup>Department of Communications and Computer Engineering, University of Malta, Msida, Malta.

Emails: yiqin@std.uestc.edu.cn, yang.ping@uestc.edu.cn, zilong.liu@essex.ac.uk, yqhuang@std.uestc.edu.cn,  
saviour.zammit@um.edu.mt

**Abstract**—This paper investigates the model-driven federated learning (FL) for channel estimation in multi-user millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) systems. Firstly, we formulate it as a sparse signal recovery problem by exploiting the beamspace domain sparsity of the mmWave channels. Then, we propose an FL-based learned approximate message passing (LAMP) channel estimation scheme, namely FL-LAMP, where the LAMP network is trained by an FL framework. Specifically, the base station (BS) and users jointly train the LAMP network, where the users update the local LAMP network parameters by local datasets consisting of measurement signals and beamspace channels, and the BS calculates the global LAMP network parameters by aggregating the local network parameters from all the users. The beamspace channel can thus be obtained in real time from the measurement signal based on the parameters of the trained LAMP network. Simulation results demonstrate that the proposed FL-LAMP scheme can achieve better channel estimation accuracy than the existing orthogonal matching pursuit (OMP) and approximate message passing (AMP) schemes, and provides satisfactory prediction capability for multipath channels.

**Index Terms**—Channel estimation, federated learning, massive MIMO, model-driven.

## I. INTRODUCTION

IN millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) systems, the base station (BS) needs to obtain downlink channel state information (CSI) for beamforming, signal detection, adaptive coding and modulations to enhance system performance [1]. Due to the large antenna arrays in mmWave communications, the traditional least squares (LS) and minimum mean square error (MMSE) channel estimation algorithms require a large amount of pilot overhead. By leveraging the sparse mmWave channels in the beamspace domain with hybrid beamforming [2], several beamspace channel estimation schemes have been proposed, such as orthogonal matching pursuit (OMP) [3], simultaneous weighted OMP (SWOMP) [4], and approximate message passing (AMP) [5]. These compressive sensing (CS)-based

methods are able to estimate the channel at low pilot overhead by exploiting the sparsity of the beamspace channel.

Deep learning (DL) methods are powerful tools to handle large amounts of data and solve complex nonlinear problems. Recently, they have been widely introduced in beamforming [6], antenna selection [7], and signal detection [8] for wireless communications. Data-driven DL has been applied to channel estimation in MIMO systems [9]–[11], where the network model is obtained by training a large dataset. Specifically, [9] proposed a deep neural network (DNN)-based channel estimation algorithm for doubly selective fading channels. Based on this, convolutional neural network (CNN)-based channel estimation methods were studied in [10] and [11] to reduce the neural network complexity. [12] and [13] proposed model-driven DL-based channel estimation by adding and optimizing learnable parameters to traditional algorithms.

Note that most of the above DL-based approaches are based on centralized machine learning (CML). In CML-based model training, the BS needs to collect local datasets from all users, bringing large latency and huge transmission overhead due to the limited communication resources in the system. In addition, the direct transmission of data pays a serious price of the data privacy and data security of users. To address these challenges, federated learning (FL) has been introduced in wireless communications [14]. In particular, in multi-user MIMO systems, the data-driven FL has been applied for channel estimation [15] and hybrid beamforming [16]. Against a rich body of literature, this paper presents the first work for the use of model-driven FL in channel estimation. Compared to existing data-driven FL schemes, model-driven FL utilizes known domain knowledge to reduce the dependence of network parameters on the user's local dataset, while retaining the advantages of traditional mathematical models and data-driven FL.

In this paper, we propose an FL-based learned approximate message passing (LAMP) channel estimation scheme for multi-user mmWave massive MIMO systems, called FL-LAMP. The channel estimation is formulated as a sparse signal recovery problem by exploiting the sparsity of beamspace channels. Then, we employ the LAMP network to recover

This work is supported by the National Key R&D Program of China under Grant 2023YFE0115100, the Sichuan Science and Technology Program under Grant 2023NSFSC0457 and the Fundamental Research Funds for the Central Universities under Grant 2242022k60006.

The corresponding author: Ping Yang.

the high-dimensional sparse beamspace channel from the low-dimensional measurement signal, thereby improving performance and reducing the pilot overhead. To reduce transmission overhead and protect data privacy, we further propose an FL approach to optimize the LAMP network parameters. Simulation results demonstrate that our proposed FL-LAMP scheme enjoys smaller transmission overhead while maintaining satisfactory channel estimation performance.

The rest of the paper is organized as follows. In Section II, we introduce the system model and formulate the problem. In Section III, we propose an FL-LAMP channel estimation scheme. The simulation results are presented in Section IV and Section V concludes this paper.

## II. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, we first introduce the beamspace channel model of multi-user mmWave massive MIMO system. Then, the beamspace channel estimation problem is formulated as a sparse signal recovery problem.

### A. Beamspace Channel

We consider a downlink multi-user mmWave massive MIMO system with one BS and  $K$  users, each user has single antenna. The BS is equipped with  $N_t$  antennas and  $N_{\text{RF}}$  radio frequency (RF) chains, and we assume  $N_{\text{RF}} = K$ . Hybrid beamforming is adopted, and the BS communicates with each user via only one stream.

Since mmWave channels are considered to have limited scattering, we adopt the widely used geometric channel model [17]. The channel vector  $\mathbf{h}_k \in \mathbb{C}^{N_t \times 1}$  between the BS and the  $k$ th user is given as

$$\mathbf{h}_k = \sqrt{\frac{N_t}{P_k}} \sum_{p=1}^{P_k} \beta_{k,p} \mathbf{a}(\varphi_{k,p}), \quad (1)$$

where  $P_k$  refers to the number of channel paths of the  $k$ th user,  $\beta_{k,p} \sim \mathcal{CN}(0, 1)$  represents the complex gain of the  $p$ th path. We consider the BS is equipped with uniform linear arrays (ULAs), the array steering vector  $\mathbf{a}(\varphi_{k,p})$  can be written as

$$\mathbf{a}(\varphi_{k,p}) = \frac{1}{\sqrt{N_t}} \left[ e^{-j\frac{2\pi d}{\lambda} \mathbf{u} \sin(\varphi_{k,p})} \right], \quad (2)$$

where  $\lambda$  and  $d$  denote the wavelength of the signal and the distance between antenna elements, respectively, and  $\mathbf{u} = [0, 1, \dots, N_t - 1]^T$ .  $\varphi_{k,p}$  indicates the  $p$ th path's angle of departure of the BS and the  $k$ th user, which is uniformly distributed over  $[-\frac{\pi}{2}, \frac{\pi}{2}]$ .

It is worth noting that the mmWave channels are sparse in the beamspace domain [2]. We define

$$\bar{\mathbf{h}}_k = \mathbf{U} \mathbf{h}_k \quad (3)$$

as a beamspace channel vector, where  $\mathbf{U} \in \mathbb{C}^{N_t \times N_t}$  represents discrete Fourier transform (DFT) matrix satisfying  $\mathbf{U} \mathbf{U}^H = \mathbf{I}_{N_t}$  [18]. The matrix  $\mathbf{U}$  can be expressed as

$$\mathbf{U} = [\bar{\mathbf{a}}(\varphi_1), \bar{\mathbf{a}}(\varphi_2), \dots, \bar{\mathbf{a}}(\varphi_{N_t})], \quad (4)$$

where

$$\bar{\mathbf{a}}(\varphi_a) = \frac{1}{\sqrt{N_t}} \left[ 1, e^{-j\pi\varphi_a}, \dots, e^{-j\pi(N_t-1)\varphi_a} \right]^T, \quad (5)$$

$$\varphi_a = -1 + \frac{2a-1}{N_t}, a = 1, 2, \dots, N_t. \quad (6)$$

### B. Problem Formulation

For downlink channel estimation, the BS transmits pilot signals to  $K$  users over  $Q$  time slot. Denote the pilot signal by  $\mathbf{P} \in \mathbb{C}^{K \times K}$ , and the measurement signal in the  $q$ th time slot at the  $k$ th user can be written as

$$\bar{\mathbf{r}}_k[q] = \mathbf{h}_k^T \mathbf{F}_{\text{RF}}[q] \mathbf{F}_{\text{BB}}[q] \mathbf{P} + \bar{\mathbf{n}}_k[q], \quad (7)$$

where the analog precoding matrix and digital precoding matrix are denoted by  $\mathbf{F}_{\text{RF}}[q] \in \mathbb{C}^{N_t \times K}$  and  $\mathbf{F}_{\text{BB}}[q] \in \mathbb{C}^{K \times K}$ , respectively, for  $q = 1, 2, \dots, Q$ .  $\bar{\mathbf{n}}_k[q] \in \mathbb{C}^{1 \times K}$  refers to noise vector, and each entry of  $\bar{\mathbf{n}}_k[q]$  obeys  $\mathcal{CN}(0, \sigma^2)$ . Then, the measurement signal is post-processed by multiplying it by  $\mathbf{P}^H$ , i.e.,

$$\tilde{\mathbf{r}}_k[q] = \bar{\mathbf{r}}_k[q] \mathbf{P}^H = \mathbf{h}_k^T \mathbf{F}_{\text{RF}}[q] \mathbf{F}_{\text{BB}}[q] + \tilde{\mathbf{n}}_k[q], \quad (8)$$

where we assume that  $\mathbf{P} \mathbf{P}^H = \mathbf{I}_K$  and  $\tilde{\mathbf{n}}_k[q] = \bar{\mathbf{n}}_k[q] \mathbf{P}^H$ .

After  $Q$  time slots of pilot transmission, the overall measurement signal can be expressed as

$$\mathbf{y}_k = \mathbf{F}^T \mathbf{h}_k + \mathbf{n}_k, \quad (9)$$

where  $\mathbf{y}_k = [\tilde{\mathbf{r}}_k[1], \tilde{\mathbf{r}}_k[2], \dots, \tilde{\mathbf{r}}_k[Q]]^T \in \mathbb{C}^{S \times 1}$ ,  $\mathbf{F} = [\mathbf{F}_{\text{RF}}[1] \mathbf{F}_{\text{BB}}[1], \mathbf{F}_{\text{RF}}[2] \mathbf{F}_{\text{BB}}[2], \dots, \mathbf{F}_{\text{RF}}[Q] \mathbf{F}_{\text{BB}}[Q]] \in \mathbb{C}^{N_t \times S}$ ,  $\mathbf{n}_k = [\tilde{\mathbf{n}}_k[1], \tilde{\mathbf{n}}_k[2], \dots, \tilde{\mathbf{n}}_k[Q]]^T \in \mathbb{C}^{S \times 1}$ , and  $S = QK$ . Based on (3), we can obtain

$$\mathbf{y}_k = \mathbf{W} \bar{\mathbf{h}}_k + \mathbf{n}_k, \quad (10)$$

where  $\mathbf{W} = \mathbf{F}^T \mathbf{U}^H \in \mathbb{C}^{S \times N_t}$  is the measurement matrix.

Due to the fact that the BS adopts orthogonal pilots to estimate channels for  $K$  users, the channel estimation method is the same for all users, and the subscript  $k$  in (10) can be omitted and (10) is rewritten as

$$\mathbf{y} = \mathbf{W} \bar{\mathbf{h}} + \mathbf{n}. \quad (11)$$

We let  $\hat{\bar{\mathbf{h}}}$  denote an estimate of  $\bar{\mathbf{h}}$ . Since the sparse property of  $\bar{\mathbf{h}}$ , the beamspace channel estimation problem in (11) can be formulated as a sparse signal recovery problem. Utilizing the CS techniques,  $\bar{\mathbf{h}}$  can be reliably solved from the measurement signal  $\mathbf{y}$  with a low pilot overhead. Several studies have been applied to solve (11) by employing greedy iterative algorithms, such as OMP [3] and SWOMP [4]. However, these algorithms find the best sparse approximate solution by gradually increasing the number of nonzero elements in the beamspace channel in an iterative manner, and cannot achieve satisfactory channel estimation performance. Therefore, in the following we will propose an FL-LAMP channel estimation scheme to estimate beamspace channels.

## III. MODEL-DRIVEN FL FOR CHANNEL ESTIMATION

In this section, we introduce the LAMP network and present the proposed FL-LAMP channel estimation scheme.

### A. LAMP Network

LAMP constructs each iteration of the AMP algorithm as a neural network, which consists of  $L$  layers of the same structure, as shown in Fig. 1. In Fig. 1, the inputs of the  $l$ th layer LAMP network are  $\hat{\bar{\mathbf{h}}}^l \in \mathbb{C}^{N_t \times 1}$ ,  $\mathbf{v}^l \in \mathbb{C}^{S \times 1}$ , and  $\mathbf{y} \in \mathbb{C}^{S \times 1}$ , where  $\hat{\bar{\mathbf{h}}}^l$  and  $\mathbf{v}^l$  are the outputs of the previous  $(l-1)$ th layer, and  $\mathbf{y}$  is the measurement signal in (11). Specifically, the  $l$ th layer of the LAMP network process the

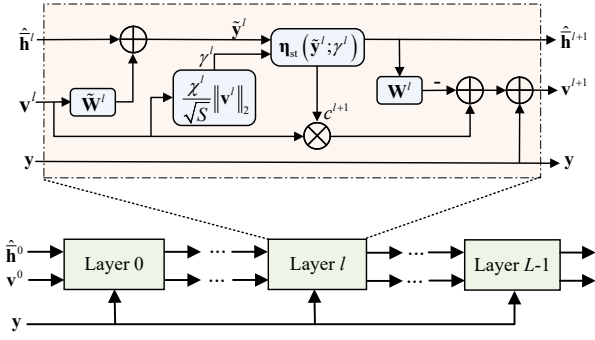


Fig. 1. LAMP network structure. The network consists of  $L$  cascading layers, and each layer has the same structure.

signal as follows:

$$\hat{\mathbf{h}}^{l+1} = \eta_{\text{st}}(\tilde{\mathbf{y}}^l; \gamma^l), \quad (12)$$

$$\mathbf{v}^{l+1} = \mathbf{y} - \mathbf{W}^l \hat{\mathbf{h}}^{l+1} + c^{l+1} \mathbf{v}^l, \quad (13)$$

for  $l = 0, 1, \dots, L-1$ , where

$$\tilde{\mathbf{y}}^l = \hat{\mathbf{h}}^l + \tilde{\mathbf{W}}^l \mathbf{v}^l, \quad (14)$$

$$\gamma^l = \frac{\chi^l}{\sqrt{S}} \|\mathbf{v}^l\|_2, \quad (15)$$

$$\mathbf{W}^l = \delta^l \mathbf{W}, \quad (16)$$

$$c^{l+1} = \frac{1}{S} \left\| \hat{\mathbf{h}}^{l+1} \right\|_0, \quad (17)$$

and the inputs of the 0th layer are  $\hat{\mathbf{h}}^0 = \mathbf{0}$  and  $\mathbf{v}^0 = \mathbf{y}$ .  $\hat{\mathbf{h}}^l$  and  $\mathbf{v}^{l+1}$  refer to the estimated beamspace channel vector and residual measurement error vector of the output of the  $l$ th layer LAMP network, respectively.  $\eta_{\text{st}}(\cdot; \cdot) : \mathbb{C}^{N_t \times 1} \rightarrow \mathbb{C}^{N_t \times 1}$  denotes the soft threshold shrinkage function, which is a nonlinear element-wise operation. For the  $i$ th entry  $[\tilde{\mathbf{y}}^l]_i$  of the input vector  $\tilde{\mathbf{y}}^l$ ,  $\eta_{\text{st}}(\cdot; \cdot)$  can be expressed as

$$[\eta_{\text{st}}(\tilde{\mathbf{y}}^l; \gamma^l)]_i = \max(|[\tilde{\mathbf{y}}^l]_i| - \gamma^l, 0) e^{j\omega^{l,i}}, \quad (18)$$

where  $[\tilde{\mathbf{y}}^l]_i = |[\tilde{\mathbf{y}}^l]_i| e^{j\omega^{l,i}}$  and  $e^{j\omega^{l,i}}$  is the phase of complex-valued element  $[\tilde{\mathbf{y}}^l]_i$ . The term  $c^{l+1} \mathbf{v}^l$  in (13) is called Onsager Correction, which is introduced into the conventional AMP algorithm to accelerate the convergence [12]. The  $l$ th layer of the LAMP network contains trainable parameters  $\Theta^l = \{\tilde{\mathbf{W}}^l, \delta^l, \chi^l\}$ , and  $\Theta = \{\Theta^l\}_{l=0}^{L-1}$  represents the overall trainable parameters. The output of the LAMP network is given as

$$\hat{\mathbf{h}}^L = f_{\text{LAMP}}(\mathbf{y}; \Theta), \quad (19)$$

where  $f_{\text{LAMP}}(\cdot)$  refers to the underlying mapping performed by the LAMP network. By optimizing these parameters in the training process, the channel estimation performance can be improved. In the following, we present the FL-based model training method to optimize the trainable parameters.

### B. Proposed FL-LAMP Channel Estimation Scheme

In this subsection, we propose the FL-LAMP scheme for channel estimation, which consists of generating training datasets, LAMP network training and prediction phases. In the training phase, the parameters of the LAMP network are opti-

mized by minimizing the loss function using a large amount of training data. In the prediction phase, the estimated beamspace channel vector are obtained by feeding the measurement signal into the trained LAMP network. Next, we introduce these three phases in detail.

1) *Training Data Generation*: Before training the LAMP network model, we generate the dataset of  $\mathbf{y}_k$  and  $\mathbf{h}_k$  for the  $k$ th user based on the simulated mmWave channel environment. Specifically, we generate the channel vector  $\mathbf{h}_k$  based on the mmWave channel model in (1), and then compute the measurement signal  $\mathbf{y}_k$  according to (10). The local dataset for the  $k$ th user can be represented as  $\mathcal{D}_k = \{(\mathbf{y}_k^{(i)}, \bar{\mathbf{h}}_k^{(i)})\}_{i=1}^{D_k}$ , where  $D_k$  refers to the size of the local dataset.

2) *Conventional CML-Based LAMP Network Training*: In CML-based network training [12], [13], the BS collects the local dataset of all users to obtain the global dataset  $\mathcal{D} \triangleq \mathcal{D}_1 \cup \mathcal{D}_2 \cup \dots \cup \mathcal{D}_K$ . Then, the global parameters of the LAMP network are trained by minimizing the loss function, which can be expressed as

$$\min_{\Theta} \mathcal{F}(\Theta) = \frac{1}{D} \sum_{j=1}^D \mathcal{L}(\hat{\mathbf{h}}^{(j)}, \bar{\mathbf{h}}^{(j)}), \quad (20)$$

where

$$\hat{\mathbf{h}}^{(j)} = f_{\text{LAMP}}(\mathbf{y}^{(j)}; \Theta), \quad (21)$$

$D = \sum_{k=1}^K D_k$  represent the size of the global dataset,  $\mathcal{L}(\cdot)$  denotes the loss function, which is given by

$$\mathcal{L}(\hat{\mathbf{h}}^{(j)}, \bar{\mathbf{h}}^{(j)}) = \frac{\left\| \hat{\mathbf{h}}^{(j)} - \bar{\mathbf{h}}^{(j)} \right\|_2^2}{\left\| \bar{\mathbf{h}}^{(j)} \right\|_2^2}, \quad (22)$$

$\hat{\mathbf{h}}^{(j)}$  and  $\bar{\mathbf{h}}^{(j)}$  refer to the predicted output and target of the  $j$ th data sample in the global dataset  $\mathcal{D}$ , respectively.

To efficiently solve (20), we employ the adaptive moment estimation (Adam) method to iteratively update the network parameters.

3) *Proposed FL-Based LAMP Network Training*: The conventional CML-based network training method requires the BS to collect channel data and measurement signals from all users to train to get LAMP network parameters, which brings in significant transmission overhead. FL decouples network training from the requirement for direct access to the original training data, where the BS unites all users to train a shared network without the data leaving the local users [14]. In contrast to CML, FL decentralizes the training process of the network to users with datasets. Therefore, the optimization problem can be written as

$$\min_{\Theta} \mathcal{F}(\Theta) \triangleq \sum_{k=1}^K \frac{D_k}{D} \mathcal{F}_k(\Theta), \quad (23)$$

where the local loss function of the  $k$ th user is expressed as

$$\mathcal{F}_k(\Theta) = \frac{1}{D_k} \sum_{i=1}^{D_k} \mathcal{L}(\hat{\mathbf{h}}_k^{(i)}, \mathbf{h}_k^{(i)}), \quad (24)$$

$\hat{\mathbf{h}}_k^{(i)}$  and  $\mathbf{h}_k^{(i)}$  refer to the predicted output and target of the  $i$ th data sample in the local dataset  $\mathcal{D}_k$ , respectively.

As shown in Fig.2, (23) is solved by the users updating the

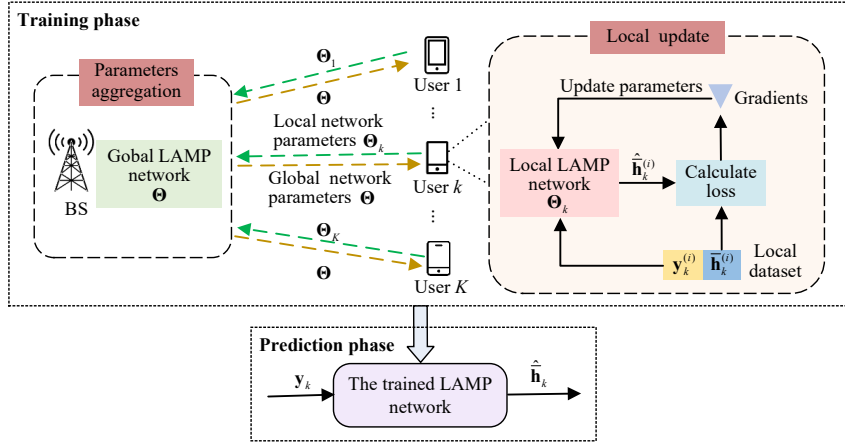


Fig. 2. Block diagram of the proposed FL-LAMP channel estimation scheme in multi-user mmWave MIMO systems: training and prediction phases.

local network parameters and the BS aggregating the network parameters, the detailed process is as follows:

(i) **Initialize network parameters.** Initialize the LAMP network parameters  $\Theta(0) = \{\tilde{\mathbf{W}}^l(0), \delta^l(0), \chi^l(0)\}_{l=0}^{L-1}$  of the BS as

$$\begin{cases} \tilde{\mathbf{W}}^l(0) = v^{-1} \mathbf{W}^H (\mathbf{W} \mathbf{W}^H + \mathbf{I}_S)^{-1}, \\ \delta^l(0) = 1, \\ \chi^l(0) = 1, \end{cases} \quad (25)$$

for  $l = 0, 1, \dots, L-1$ , where  $\tilde{\mathbf{W}}^l(0)$  satisfies

$$\text{tr}(\mathbf{W} \tilde{\mathbf{W}}^l(0)) = N_t, \quad (26)$$

and  $\text{tr}(\cdot)$  indicates the trace of the matrix. Then, the BS distributes the initialized network parameters to each user.

(ii) **The users update the local network parameters.** Based on the received global network parameters  $\Theta(t)$ , each user updates its own network parameters using its own dataset to minimize the local loss function, and the updated local network parameters for the  $k$ th user can be represented as

$$\Theta_k(t+1) = \arg \min \mathcal{F}_k(\Theta(t)), \quad (27)$$

where  $\Theta_k(t) = \{\tilde{\mathbf{W}}_k^l(t), \delta_k^l(t), \chi_k^l(t)\}_{l=0}^{L-1}$  denotes the local network parameters at the  $t$ th iteration of the  $k$ th user and  $\Theta(t) = \{\tilde{\mathbf{W}}^l(t), \delta^l(t), \chi^l(t)\}_{l=0}^{L-1}$  refers to the global network parameters at the  $t$ th iteration. Similar to the CML algorithm, the users perform the Adam method to update the local network parameters to obtain  $\Theta_k(t+1)$ .

(iii) **The BS aggregates the network parameters.** The updated local network parameters at each user are sent to the BS via wireless links, and the BS update the global network parameters as

$$\Theta(t+1) = \frac{1}{D} \sum_{k=1}^K (D_k \Theta_k(t+1)). \quad (28)$$

Thus, the parameters for each layer in the global LAMP network can be calculated as

$$\tilde{\mathbf{W}}^l(t+1) = \frac{1}{D} \sum_{k=1}^K (D_k \tilde{\mathbf{W}}_k^l(t+1)), \quad (29)$$

$$\delta^l(t+1) = \frac{1}{D} \sum_{k=1}^K (D_k \delta_k^l(t+1)), \quad (30)$$

$$\chi^l(t+1) = \frac{1}{D} \sum_{k=1}^K (D_k \chi_k^l(t+1)), \quad (31)$$

for  $l = 0, 1, \dots, L-1$ . After that, the updated global network parameters  $\Theta(t+1)$  are distributed to all users via the downlink for next network update.

---

**Algorithm 1:** Proposed FL-based LAMP Network Training Method

---

- Input:**  $\mathcal{D}_k$ : training dataset,  $T$ : the number of iterations.  
**Output:** Global LAMP network parameters  $\Theta(*) = \Theta(T)$ .
- 1: **BS executes:**
  - 2: Initialize global network parameters  $\Theta(0)$  from (25).
  - 3: Distribute  $\Theta(0)$  to all users.
  - 4: **for** each  $t = 0, 1, \dots, T-1$  **do**
  - 5:   **for** each  $k \in \{1, 2, \dots, K\}$  in parallel **do**
  - 6:      $\Theta_k(t+1) \leftarrow \text{UserUpdate}(k, \Theta(t))$ .
  - 7:   **end for**
  - 8:   Calculate  $\Theta(t+1)$  according to (28).
  - 9:   **Send**  $\Theta(t+1)$  to all users.
  - 10: **end for**
  - 11: **UserUpdate**( $k, \Theta(t)$ ): // Run on user  $k$
  - 12:   Compute  $\Theta_k(t+1)$  based on (27).
  - 13:   **Upload**  $\Theta_k(t+1)$  to the BS.
- 

We can obtain the optimal LAMP network parameters  $\Theta(*) = \{\tilde{\mathbf{W}}^l(*), \delta^l(*), \chi^l(*)\}_{l=0}^{L-1}$  by performing (ii) and (iii) several times. The detailed steps of the FL-based LAMP network training method are summarized in **Algorithm 1**.

4) **LAMP Network Prediction:** After the LAMP network parameters are optimized, the trained LAMP network can be deployed to estimate the beamspace channel in mmWave systems in real time from the measurement signal. The proposed FL-LAMP channel estimation scheme is summarized in **Algorithm 2**.

#### IV. SIMULATION RESULTS

In this section, we first evaluate the normalized mean square error (NMSE) performances of the proposed FL-LAMP chan-

---

**Algorithm 2: Proposed FL-LAMP Channel Estimation Scheme**


---

**Input:**  $y_k$ : measurement signal,  $\Theta(\cdot)$ : the parameters of the trained LAMP network.

**Output:** Predicted beamspace channel vector  $\hat{\mathbf{h}}_k = \hat{\mathbf{h}}_k^L$ .

- 1: Initialization:  $\hat{\mathbf{h}}_k^0 = \mathbf{0}$  and  $\mathbf{v}_k^0 = y_k$ .
  - 2: **for** each  $l = 0, 1, \dots, L - 1$  **do**
  - 3:  $\tilde{\mathbf{y}}_k^l = \hat{\mathbf{h}}_k^l + \tilde{\mathbf{W}}^l(\cdot) \mathbf{v}_k^l$ .
  - 4:  $\gamma_k^l = \frac{\chi^l(\cdot)}{\sqrt{S}} \|\mathbf{v}_k^l\|_2$ .
  - 5:  $\hat{\mathbf{h}}_k^{l+1} = \eta_{\text{st}}(\tilde{\mathbf{y}}_k^l; \gamma_k^l)$ .
  - 6:  $c_k^{l+1} = \frac{1}{S} \left\| \hat{\mathbf{h}}_k^{l+1} \right\|_0$ .
  - 7:  $\mathbf{v}_k^{l+1} = y_k - \delta^l(\cdot) \mathbf{W} \hat{\mathbf{h}}_k^{l+1} + c_k^{l+1} \mathbf{v}_k^l$ .
  - 8:  $l \leftarrow l + 1$ .
  - 9: **end for**
- 

nel estimation scheme. In our simulations, the BS is equipped with  $N_t = 256$  antennas serving  $K = 4$  users with single antenna, and the number of paths is set to be  $\bar{P} = P_k = 5$ , for  $k = 1, 2, \dots, K$ . For the pilot transmission, the number of time slots is set to be  $Q = 128/K$ , i.e.,  $S = QK = 128$  is satisfied in different number of users. We generate  $8 \times 10^4/K$  and  $5000/K$  samples for each user as the training and the test dataset, respectively. Therefore, the total numbers of training samples and test samples are  $8 \times 10^4$  and  $5000$ , respectively. In FL-LAMP, the training dataset is equally distributed to all users and the test dataset is placed at BS to evaluate global model performance. We adopt the Adam optimizer to train the LAMP network by Pytorch 1.9.0, the learning rate is 0.001, and mini-batch size is 256.

Fig. 3 shows the channel estimation performance of the proposed FL-LAMP scheme with the existing OMP [3], AMP [5], and CML-tied LAMP [12] schemes. The performance is measured by the NMSE which is formally defined below:

$$\text{NMSE} = 10 \log_{10} \left[ \mathbb{E} \left( \frac{1}{K} \sum_{k=1}^K \frac{\|\hat{\mathbf{h}}_k - \bar{\mathbf{h}}_k\|_2^2}{\|\bar{\mathbf{h}}_k\|_2^2} \right) \right]. \quad (32)$$

For the OMP scheme, we set the sparsity of the beamspace channel vector as  $J = 22$ . For the AMP method, the number of iterations is set as  $I = 10$  and the shrinkage parameter as  $\chi = 1.1402$  for each iteration. For the CML-tied LAMP and proposed FL-LAMP schemes, we consider the number of LAMP network layers is  $L = 8$  and  $L = 6$ , respectively.

It is seen from Fig. 3 that the FL-LAMP scheme has better channel estimation performance than existing schemes. When SNR = 25 dB, the proposed FL-LAMP scheme with  $K = 4$  has 74.49%, 58.42% and 17.21% performance improvements over the OMP, AMP and CML-tied LAMP schemes, respectively, while the proposed FL-LAMP scheme with  $K = 12$  has 68.90%, 53.35% and 13.46% performance improvements over the OMP, AMP and CML-tied LAMP schemes, respectively. The proposed FL-LAMP scheme can achieve better NMSE performance because it learns more network parameters than the CML-tied LAMP scheme. When SNR = 25 dB, the FL-LAMP scheme with  $K = 4$  has 3.62% performance loss

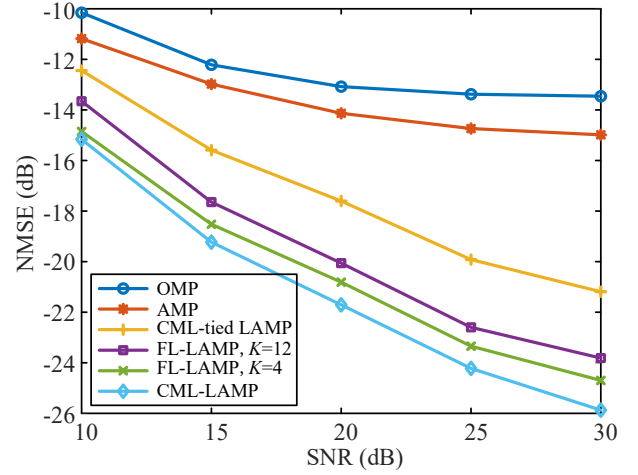


Fig. 3. Comparisons of NMSE performance in terms of SNR for different schemes.

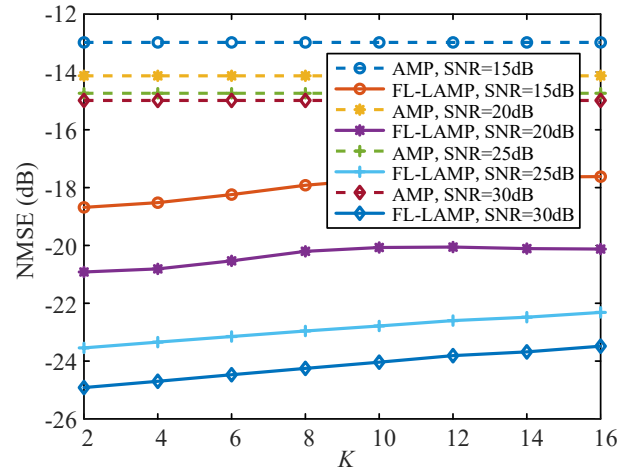


Fig. 4. Comparisons of NMSE performance in terms of the number of users for the proposed FL-LAMP scheme.

over the CML-LAMP scheme since CML-LAMP scheme can directly access the entire dataset to obtain global network parameters, while the FL-LAMP scheme obtains global network parameters by aggregating local network parameters from multiple users, and distributed network training and multi-user parameter aggregation bring partial performance loss.

As shown in Fig. 4, we compare the NMSE performance for the proposed FL-LAMP scheme in terms of the number of users. The performance of the FL-LAMP scheme decreases slightly as the number of users increases. When  $K = 16$ , the proposed FL-LAMP scheme improves the performance over the AMP scheme by 35.77%, 42.40%, 51.43% and 56.73% at SNR = 15 dB, SNR = 20 dB, SNR = 25 dB and SNR = 30 dB, respectively.

In the following, we evaluate the robustness of the proposed FL-LAMP channel estimation scheme in Fig. 5. Note that the LAMP network parameters are obtained by training the channel samples with multipath number  $\bar{P} = 5$ , and the trained network parameters are adopted to predict the channel samples with different number of paths. As can be seen in Fig. 5, the FL-LAMP scheme can robustly predict multipath channels with different number of paths without training the entire network.



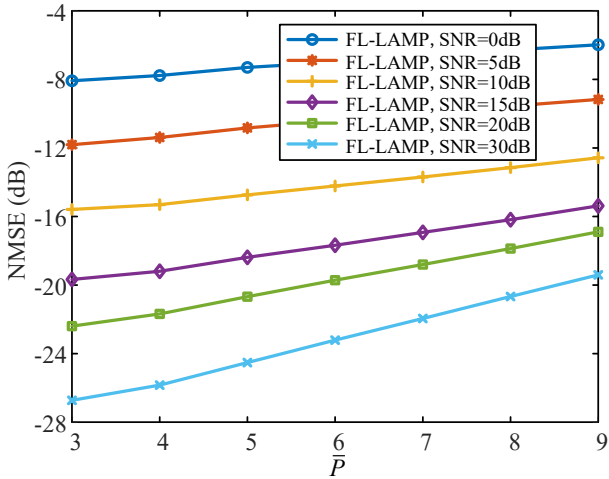


Fig. 5. Comparisons of NMSE performance in terms of the number of multipath for the proposed FL-LAMP scheme ( $K = 4$ ).

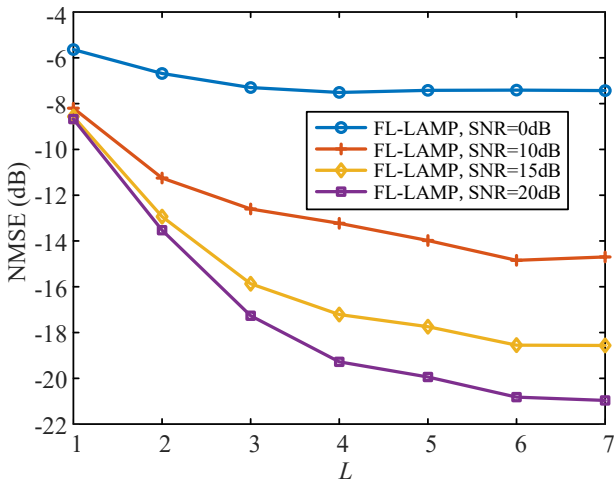


Fig. 6. Comparisons of NMSE performance in terms of the number of layers for the LAMP network ( $K = 4$ ).

To investigate the convergence of the proposed FL-LAMP scheme, the performance of LAMP network with different number of layers is shown in Fig. 6. From Fig. 6, it is shown that the LAMP network can reach convergence about at layer  $L = 6$ .

We further compare the computational complexity and transmission of the proposed FL-LAMP scheme with other existing schemes. The computational complexity of the CML-tied LAMP, CML-LAMP and proposed FL-LAMP schemes are determined by the LAMP network structure, and they have the same complexity  $O(LSN_t)$ . In addition, the computational complexity of the OMP and AMP schemes are represented as  $O(ISN_t)$  and  $O(JSN_t) + O(J^3S)$ . The CML-LAMP scheme requires the users upload the local training dataset including  $D(N_t + S)$  (i.e.,  $3.072 \times 10^7$ ) data symbols, while the proposed FL-LAMP scheme only requires the users transmit network parameters including  $SN_tL + 2L$  (i.e., 196620) data symbols. Therefore, the proposed FL-LAMP scheme can achieve satisfactory NMSE performance with lower transmission overhead.

## V. CONCLUSION

In this paper, we have proposed a model-driven FL based channel estimation scheme termed FL-LAMP for multi-

user mmWave massive MIMO systems. The proposed FL-LAMP scheme exploits the beamspace domain sparsity of the mmWave channels, which can greatly reduce the pilot overhead. We have compared the proposed FL-LAMP scheme with the existing works in terms of NMSE. The simulation results show that the proposed FL-LAMP scheme has better channel estimation performance than the existing OMP and AMP schemes. In addition, the performance of the proposed FL-LAMP scheme is close to that of the CML-LAMP scheme with less transmission overhead. The proposed FL-LAMP scheme enables flexible tradeoffs among the NMSE, the transmission overhead and the security performance metrics.

## REFERENCES

- [1] W. Shen, L. Dai, Y. Shi, B. Shim, and Z. Wang, "Joint channel training and feedback for FDD massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8762-8767, Oct. 2016.
- [2] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeterwave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259-1262, Jun. 2016.
- [3] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370-2386, Jun. 2016.
- [4] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946-2960, May. 2018.
- [5] X. Wu, L. Gu, W. Wang, and X. Gao, "Pilot design and AMP-based channel estimation for massive MIMO-OFDM uplink transmission," in *Proc. IEEE 27th Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Valencia, Spain, Sep. 2016, pp. 1-7.
- [6] H. Huang, Y. Peng, J. Yang, W. Xia, and G. Gui, "Fast beamforming design via deep learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 1065-1069, Jan. 2020.
- [7] A. M. Elbir and K. V. Mishra, "Joint antenna selection and hybrid beamformer design using unquantized and quantized deep learning networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1677-1688, Mar. 2020.
- [8] D. Lin, S. Hu, W. Wu W, *et al.* "Few-shot RF fingerprinting recognition for secure satellite remote sensing and image processing," *Science China Information Sciences*, vol. 66, no. 8, pp.189304, 2023.
- [9] Y. Yang, F. Gao, X. Ma, and S. Zhang, "Deep learning-based channel estimation for doubly selective fading channels," *IEEE Access*, vol. 7, pp. 36579-36589, 2019.
- [10] Y. Jin, J. Zhang, S. Jin, and B. Ai, "Channel estimation for cell-free mmWave massive MIMO through deep learning," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 10325-10329, Oct. 2019.
- [11] P. Dong, H. Zhang, G. Y. Li, I. Gaspar, and N. NaderiAlizadeh, "Deep CNN-based channel estimation for mmWave massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 989-1000, Sep. 2019.
- [12] M. Borgerding, P. Schniter, and S. Rangan, "AMP-inspired deep networks for sparse linear inverse problems," *IEEE Trans. Signal Process.*, vol. 65, no. 16, pp. 4293-4308, Aug. 2017.
- [13] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Commun. Lett.*, vol. 7, no. 5, pp. 852-855, Oct. 2018.
- [14] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation opportunities and challenges," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 46-51, Jun. 2020.
- [15] A. M. Elbir and S. Coleri, "Federated learning for channel estimation in conventional and RIS-assisted massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4255-4268, Jun. 2022.
- [16] A. M. Elbir and S. Coleri, "Federated learning for hybrid beamforming in mm-Wave massive MIMO," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2795-2799, Dec. 2020.
- [17] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481-6494, Nov. 2015.
- [18] X. Gao, L. Dai, S. Han, C.-L. I, and X. Wang, "Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6010-6021, Sep. 2017.