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

Qin Yi†, Ping Yang‡, Senior Member, IEEE, Zilong Liu§, Senior Member, IEEE, Yiqian Huang†, and Saviour Zammit*, Senior Member, IEEE

†National Key Laboratory of Wireless Communications, University of Electronic Science and Technology of China, Sichuan, China.
‡School of Computer Science and Electronics Engineering, University of Essex, Colchester, United Kingdom.
§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 modulation 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...
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 $N_{RF}$ radio frequency (RF) chains, and we assume $N_{RF} = K$. 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

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

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

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

where $\lambda$ and $d$ denote the wavelength of the signal and the distance between antenna elements, respectively, and $\mathbf{u} = [0, 1, \ldots, N_t - 1]^T$. $\varphi_{k,p}$ 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

$$\mathbf{h}_k = \mathbf{U} \hat{\mathbf{h}}_k$$

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

$$\mathbf{U} = \begin{bmatrix} \hat{\mathbf{a}}(\varphi_1), \hat{\mathbf{a}}(\varphi_2), \ldots, \hat{\mathbf{a}}(\varphi_{N_t}) \end{bmatrix},$$

where

$$\hat{\mathbf{a}}(\varphi_a) = \frac{1}{\sqrt{N_t}} [1, e^{-j \pi \varphi_a}, \ldots, e^{-j \pi (N_t-1) \varphi_a}]^T,$$

$$\varphi_a = -1 + \frac{2a - 1}{N_t}, a = 1, 2, \ldots, N_t.$$

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

$$\tilde{\mathbf{r}}_k[q] = \mathbf{h}_k^T \mathbf{F}_{RF}[q] \mathbf{F}_{BB}[q] \mathbf{P} + \tilde{\mathbf{n}}_k[q],$$

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

$$\hat{\mathbf{r}}_k[q] = \tilde{\mathbf{r}}_k[q] \mathbf{P}^H = \mathbf{h}_k^T \mathbf{F}_{RF}[q] \mathbf{F}_{BB}[q] + \tilde{\mathbf{n}}_k[q],$$

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

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

$$\mathbf{y}_k = \mathbf{F}^T \mathbf{h}_k + \mathbf{n}_k,$$

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

$$\mathbf{y}_k = \mathbf{W} \hat{\mathbf{h}}_k + \mathbf{n}_k,$$

where $\mathbf{W} = \mathbf{F}^T \mathbf{U} \in \mathbb{C}^{S \times N_t}$ 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

$$\mathbf{y} = \mathbf{W} \hat{\mathbf{h}} + \mathbf{n}.$$

We let $\hat{\mathbf{h}}$ denote an estimate of $\mathbf{h}$. Since the sparse property of $\mathbf{h}$, the beamspace channel estimation problem in (11) can be formulated as a sparse signal recovery problem. Utilizing the CS techniques, $\hat{\mathbf{h}}$ 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 $\mathbf{h}^l \in \mathbb{C}^{N_t \times 1}$, $\mathbf{v}^l \in \mathbb{C}^{S \times 1}$, and $\mathbf{y} \in \mathbb{C}^{S \times 1}$, where $\mathbf{h}^l$ 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
The signal as follows:

$$
\hat{h}^{l+1} = \eta_{st}(\hat{y}^l; \gamma^l),
$$

(12)

$$
v^{l+1} = y - W^l \hat{h}^{l+1} + e^{l+1} v^l,
$$

(13)

for \( l = 0, 1, \ldots, L - 1 \), where

$$
\hat{y}^l = \hat{h}^l + \hat{W}^l v^l,
$$

(14)

$$
\gamma^l = \frac{\chi}{\sqrt{S}} ||v^l||_2,
$$

(15)

$$
W^l = \delta^l W^l,
$$

(16)

$$
\epsilon^{l+1} = \frac{1}{S} ||\hat{h}^{l+1}||_0,
$$

(17)

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

$$
[\eta_{st}([\hat{y}^l]_i; \gamma^l)]_i = \max\left\{ ([\hat{y}^l]_i - \gamma^l, 0) e^{j\omega^l i} \right\},
$$

(18)

and \( [\hat{y}^l]_i \) denotes the size of the global dataset. \( \hat{f}_{LAMP}(\cdot; \cdot) \) is 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 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 unifies 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 on users with datasets. Therefore, the optimization problem can be written as

$$
\min_{\Theta} \mathcal{L}(\bar{h}(\cdot), \tilde{h}(\cdot)) = \frac{1}{D} \sum_{i=1}^{D} L(h^i, \tilde{h}^i),
$$

(21)

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

Fig. 1. LAMP network structure. The network consists of \( L \) cascading layers, and each layer has the same structure.
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 \( \Theta(0) = \{ \hat{W}^{l}(0), \delta^{l}(0), \chi^{l}(0) \}_{l=0}^{L-1} \) of the BS as
\[
\left\{ \begin{array}{l}
\hat{W}^{l}(0) = \nu^{-1}W^{H}(WW^{H} + I_{S})^{-1}, \\
\delta^{l}(0) = 1, \\
\chi^{l}(0) = 1,
\end{array} \right.
\]
for \( l = 0, 1, \ldots, L - 1 \), where \( \hat{W}^{l}(0) \) satisfies
\[
\text{tr}\left(W\hat{W}^{l}(0)\right) = N_{t},
\]
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
\[
\Theta_{k}(t + 1) = \arg\min_{\Theta_{k}} \mathcal{F}_{k}(\Theta_{k}(t)),
\]
where \( \Theta_{k}(t) = \{ \hat{W}_{k}^{l}(t), \delta_{k}^{l}(t), \chi_{k}^{l}(t) \}_{l=0}^{L-1} \) denotes the local network parameters at the \( t \)th iteration of the \( k \)th user and \( \Theta(t) = \{ \hat{W}^{l}(t), \delta^{l}(t), \chi^{l}(t) \}_{l=0}^{L-1} \) 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
\[
\Theta(t + 1) = \frac{1}{D} \sum_{k = 1}^{K} (D_{k}\Theta_{k}(t + 1)),
\]
Thus, the parameters for each layer in the global LAMP network can be calculated as
\[
\hat{W}^{l}(t + 1) = \frac{1}{D} \sum_{k = 1}^{K} (D_{k}\hat{W}_{k}^{l}(t + 1)),
\]
\[
\delta^{l}(t + 1) = \frac{1}{D} \sum_{k = 1}^{K} (D_{k}\delta_{k}^{l}(t + 1)),
\]
\[
\chi^{l}(t + 1) = \frac{1}{D} \sum_{k = 1}^{K} (D_{k}\chi_{k}^{l}(t + 1)),
\]
for \( l = 0, 1, \ldots, L - 1 \). 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:** \( D_{k} \): training dataset, \( T \): the number of iterations.

**Output:** Global LAMP network parameters \( \Theta(T) \).

1. **BS executes:**
2. Initialize global network parameters \( \Theta(0) \) from (25).
3. Distribute \( \Theta(0) \) to all users.
4. for each \( t = 0, 1, \ldots, T - 1 \) do
   5. for each \( k \in \{1, 2, \ldots, K\} \) in parallel do
      6. \( \Theta_{k}(t + 1) \leftarrow \text{UserUpdate}(k, \Theta(t)) \).
   7. end for
5. for each \( k \in \{1, 2, \ldots, K\} \) do
   8. Compute \( \Theta_{k}(t + 1) \) based on (27).
   9. Send \( \Theta_{k}(t + 1) \) to all users.
10. end for
11. UserUpdate\((k, \Theta(t))\) : // Run on user \( k \)
12. end for
13. Upload \( \Theta_{k}(t + 1) \) to the BS.

We can obtain the optimal LAMP network parameters \( \Theta(T) = \{ \hat{W}^{l}(T), \delta^{l}(T), \chi^{l}(T) \}_{l=0}^{L-1} \) by performing (ii) and (iii) several times. The detailed process 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-
channel 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 \( P = P_k = 5 \), for \( k = 1, 2, \ldots, K \). 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:

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

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 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 \( P = 5 \), 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.
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.

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-LAMP, CML-LAMP and proposed FL-LAMP schemes are determined by the LAMP network structure, and they have the same complexity \( O(\text{LSN}_t) \). In addition, the computational complexity of the OMP and AMP schemes are represented as \( O(\text{JSN}_t) + O(J^3S) \). 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 \( S N_L+2L \) (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-

\[ \text{REFERENCES} \]


