Deep Learning Assisted Channel Estimation for Cell-Free Distributed MIMO Networks

Imtiaz Ahmed[†], Md. Zoheb Hasan[††], Ahmed Rubaai[†], Kamrul Hasan[††††], Cong Pu[†††], and Jeffrey H. Reed[††]
[†]Howard University, Washington, DC, USA,
[††]Virginia Polytechnic Institute and State University, Blacksburg, VA, USA,
[††††]Tennessee State University, Nashville, TN, USA,
[†††]Oklahoma State University, Stillwater, OK, USA
Emails: imtiaz.ahmed@howard.edu[†], mdzoheb@vt.edu[††],
arubaai@howard.edu[†], mhasan1@tnstate.edu[††††], cong.pu@ieee.org[†††], reedjh@vt.edu[††]

***Abstract*—Pilot contamination poses a critical challenge for channel estimation in dense cell-free (CF) distributed multiple-input multiple-output (CF-DMIMO) wireless networks. State-of-the-art channel estimation schemes require inversion of a high-dimensional channel covariance matrix, which is practically infeasible for dense CF-DMIMO networks owing to the requirement of large storage and high dimensional computational complexity. In this work, we investigate channel estimation problem for a CF-DMIMO network, where both terrestrial and aerial users are jointly supported by distributed access points. We formulate the problem of estimating channel coefficients from the received in-phase/quadrature (I/Q) samples as a non-linear regression problem and propose two deep-learning aided channel estimation schemes for the considered network, namely, deep model-agnostic neural network (DMANN) and deep successive contamination cancellation (DSCC) schemes. Compared to the state-of-the-art channel estimation schemes for CF-DMIMO networks, the proposed schemes (i) tackle the unavoidable pilot contamination issue in dense CF-DMIMO networks while estimating the channel gains for both terrestrial and aerial users; (2) does not require prior knowledge of signal-to-noise ratios; and (3) works well in the presence of non-Gaussian correlated noise. Simulation results demonstrate the effectiveness of the proposed schemes over state-of-the-art channel estimation schemes in various use cases of the CF-DMIMO networks.**

***Index Terms*—Cell free massive multiple input multiple output, channel estimation, deep learning, pilot contamination.**

I. Introduction

The cell-centric paradigm of wireless networks, which is limited by inter-cell interference, is evolving toward a ubiquitous cell-free (CF) architecture that is more user-centric and robust to interference while providing users with macro-diversity [1]. In the CF distributed multiple input multiple output (CF-DMIMO) systems, many geographically distributed access points (APs) employing single or multiple antennas simultaneously serve a limited number of user equipment (UE) in a time-division duplex (TDD) scheme with the help of a fronthaul network and a central processing unit (CPU) operating in the same time-frequency resource. The CPU transmits data and resource control information to the APs through the downlink, while the APs transmit data received from the UEs using the uplink to the CPU via the fronthaul connection. CF-DMIMO essentially integrates the best attributes of ultra-dense networks, coordinated multi-point transmission, and cellular MIMO, and achieves improved spectral efficiency and transmission reliability by leveraging robust channel estimation and favorable propagation characteristics arising from the exploitation of a large number of propagation paths [2]. CF-DMIMO is therefore a key enabler of the emerging high-throughput, ultra-reliable, and low-latency applications of sixth generation (6G) networks [3].

Besides the conventional terrestrial user equipments (TUEs), 6G cellular network is envisioned to support drones equipped with communication transceivers and sensors, known as aerial user equipments (AUEs), to facilitate surveillance, remote sensing and environmental monitoring, real-time multimedia streaming, and intelligent transportation [4]. In particular, serving AUEs from the conventional base station or terrestrial APs along with terrestrial UEs (TUEs) not only efficiently uses the bandwidth for both terrestrial and aerial UEs but also offers several advantages to the AUEs, including, improved channel capacity, ubiquitous coverage, and beyond visual line-of-sight (VLOS) range of drone operations [5]. However, it is challenging to transmit high data rate traffic from AUEs, such as streaming videos, since AUEs can create strong interference at several neighboring APs due to AUEs' high elevations and increased likelihood of establishing LOS paths with a large footprints. Such high interference can significantly reduce the achievable data rate of the terrestrial UEs, making coexistence of TUEs and AUEs a non-trivial problem. CF-DMIMO system is a promising solution to address the coexistence issues of AUEs and TUEs, thanks to its macro-diversity gain, centralized signal processing, and interference suppression capability.

To maximize the processing gain and interference-suppressing capability of a massive MIMO (MMIMO) network, accurate channel estimation is required. Note that in any cellular network, the length of pilot signals is limited and the number of orthogonal pilots is also finite. Essentially, pilot-based channel estimation in large-scale MMIMO networks often exhibits pilot contamination due to the fact of reusing pilots among multiple antennas [6] or multiple UEs [7]. We emphasize that the channel estimation in the presence of pilot contamination is an intricate optimization problem. Conventional optimization methods require inversion of the entire covariance matrix of the channels to solve such a channel estimation problem, which requires large signaling overhead, high computational complexity, and large storage. Essentially, conventional optimization methods are practically infeasible for channel estimation in large-scale MMIMO net-

979-8-3503-3667-2/23/$31.00 ©2023 IEEE

works, specifically when rapid and frequent estimations of channel are required due to small channel coherence time. To address this issue, several deep-learning (DL) enabled channel estimation methods were developed [8], [9]. However, these studies considered co-located MMIMO (C-MMIMO) networks that is fundamentally different from CF-DMIMO networks. More specifically, different propagation characteristics exhibited by the distributed antennas in CF-DMIMO systems is not appropriately captured by the DL models trained for C-MMIMO networks. In [10], a deep unsupervised pilot signals' power allocation method was developed for minimizing mean square (MSE) of a linear minimum MSE (LMMSE) based channel estimation in the distributed MIMO systems. However, it will be shown in our numerical results that the channel estimation accuracy of LMMSE based channel estimator is notably compromised in the presence of pilot contamination. In [11], a DL based channel estimator was developed for CF-DMIMO networks by leveraging a denoising convolutional network. However, the channel estimator developed in [11] requires prior knowledge of signal-to-noise ratio (SNR) of the channels to be estimated, which adds additional computational complexity. As a result, there is a need of near-optimal yet computationally-inexpensive DL-based channel estimators for practical CF-DMIMO networks.

In this paper, we investigate the channel estimation problem for a generalized CF-DMIMO network where both TUEs and AUEs are simultaneously supported by the distributed APs over the same spectrum. We consider a dense network setting where the numbers of AUEs and TUEs are much larger than the available orthogonal pilots, and therefore, employ multiplexing of multiple AUEs and TUEs over the same pilot sequences. However, the resultant pilot contamination severely affects the accuracy of channel estimation. To address this issue, we propose two DL-based channel estimation methods to estimate the channel coefficients directly from the received in-phase/quadrature (I/Q) samples. More specifically, in the first method, we employ a single deep neural network to estimate complex channel gains from I/Q samples whereas in the second method, we merge both DL and successive interference cancellation to improve the accuracy of channel estimation. Compared to state-of-the-art channel estimation schemes, our proposed DL methods have the following attributes. First, our proposed DL methods improve channel estimation accuracy in severe pilot contamination without prior knowledge of channel SNRs. Second, our proposed method does not require estimation and inversion of the high-dimensional channel correlation matrix, thus alleviates both computational complexity and signaling overhead. Finally, the proposed methods also efficiently estimate channels in the presence of non-Gaussian correlated noise, and consequently, the proposed channel estimation methods can be useful in many emerging applications of CF networks, such as industrial Internet-of-Things (I-IoT), smart grid, and cyber-physical systems where non-Gaussian noise can appear. We also emphasize that the proposed DL approaches are applicable to conventional CF-DMIMO networks with only TUEs. Simulation results confirm that our proposed DL approaches achieve near-optimal performance and outperform the conventional channel estimation methods.

The rest of the paper is organized as follows. Section II

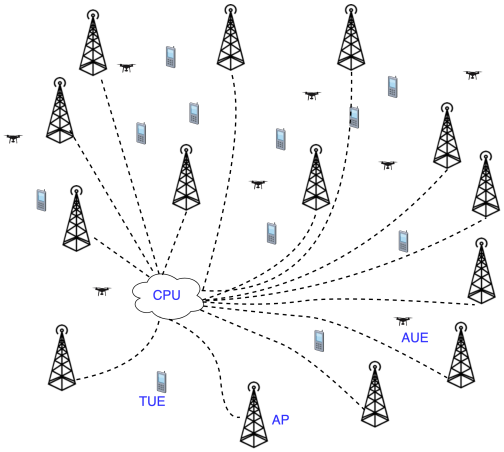

Fig. 1: CF-DMIMO System Model.

describes the considered system model. The proposed data driven estimation schemes are discussed in Section III. The simulation results are presented in Section IV and Section V concludes the paper.

II. System Model

We consider a uplink CF-DMIMO system, where L APs are uniformly distributed over a large geographical area to provide coverage to K_t and K_a single-antenna TUEs and AUEs, respectively. Each AP is equipped with N antennas; hence, $T = NL$ represents the total number of AP antennas in the considered CF network. We assume $T \gg K$, where $K = K_t + K_a$ to provide the network with large degrees-of-freedom (DoF) so that the UEs can be segregated in space by applying signal processing techniques on transmit and receive signals. We intend to keep N small and hence $L \gg K$ holds to realize CF communications. Each UE is communicated via a subset of APs, and an AP may serve more than one UEs simultaneously via appropriate AP cooperation while satisfying the network quality-of-service (QoS) requirements. We assume that the APs leverage sub-6 GHz spectrum to communicate with UEs; however, this assumption can be extended to millimeter-wave (mmW) and tera-hertz (THz)-band transmissions by appropriate modifications of the considered system model. A high capacity ideal fronthaul communication link, such as optical fiber, is deployed between each AP and CPU. The CPU is usually hosted in the edge cloud platform, and thereby provides high performance computing and centralized signal processing for a large number of UEs' data.

Each UE transmits a pilot sequence $\boldsymbol{\phi}_q$, $q \in \{1, 2, \cdots, \tau_p\}$ of τ_p samples that satisfy

$$\boldsymbol{\phi}_{q_1}^H \boldsymbol{\phi}_{q_2} = \begin{cases} \tau_p & q_1 = q_2 \\ 0 & q_1 \neq q_2 \end{cases} \tag{1}$$

and $||\boldsymbol{\phi}_q||^2 = \tau_p$, where $(\cdot)^H$ denotes the Hermitian operation. Moreover, we assume $\tau_p \leq \tau_c$, where τ_c is the length (in samples) of coherence block. In general, $K \gg \tau_c$ for a CF configuration. Therefore, τ_p orthogonal pilot sequences cannot completely eliminate the pilot contamination for the CF system. As more than one TUE or AUE can be assigned with the same pilot sequence, let us denote the index of the pilot sequence assigned to UE k, $k = 1, 2, \cdots, K$ by

$q_k \in \{1, 2, \cdots, \tau_p\}$ and define $\mathcal{S}_{q_k}$ as the set of UEs that use the same pilot sequence q_k. We define P_k as the transmit power for UE $k \in \{1, 2, \cdots, K\}$ during the uplink pilot transmission phase. The signal received at AP $l \in \{1, 2, \cdots, L\}$ during the pilot transmission phase is represented as follows:

$$\boldsymbol{Y}_l = \sum_{k_a=1}^{K_a} \sqrt{P_{k_a}} \boldsymbol{h}_{k_a l} \boldsymbol{\phi}_{q_{k_a}}^{\mathrm{T}} + \sum_{k_t=1}^{K_t} \sqrt{P_{k_t}} \boldsymbol{h}_{k_t l} \boldsymbol{\phi}_{q_{k_t}}^{\mathrm{T}} + \boldsymbol{N}_l, \quad (2)$$

where $\boldsymbol{N}_l \in \mathcal{C}^{N \times \tau_p}$ is the noise at AP l with complex-valued and independent and identically distributed (i.i.d) elements with zero-mean and variance σ_n^2, and $(\cdot)^{\mathrm{T}}$ presents the transpose operation. Here, $\boldsymbol{h}_{k_a l} = [h_{k_a l}^1, h_{k_a l}^2, \cdots, h_{k_a l}^N]^T$ presents channel vector for the link between AUE $k_a \in \{1, 2, \cdots, K_a\}$ and AP $l \in \{1, 2, \cdots, L\}$. Likewise, $\boldsymbol{h}_{k_t l} = [h_{k_t l}^1, h_{k_t l}^2, \cdots, h_{k_t l}^N]^T$ presents channel vector for the link between TUE $k_t \in \{1, 2, \cdots, K_t\}$ and AP $l \in \{1, 2, \cdots, L\}$. The channel gain $h_{k_i l}^n$, for $i \in \{a, t\}$ and $n \in \{1, 2, \cdots, N\}$ contains both small and large scale fading components depending on the propagation environments. The ensuing analysis is valid for any suitable channel models of AUEs and TUEs. For simplicity, we consider that channel gains are temporally uncorrelated. The baseband signal received at CPU from AP l over ideal fronthaul link can be represented as

$$\boldsymbol{Z}_l = \boldsymbol{Y}_l + \boldsymbol{W}_l, \quad (3)$$

where $l \in \{1, 2, \cdots, L\}$. Here, $\boldsymbol{W}_l$ presents a quantization noise matrix, where each element is complex-valued and i.i.d. and can be represented with zero-mean and variance σ_w^2.

III. Data Driven Channel Estimation

In order to estimate the channel $\boldsymbol{h}_{k_i l}$, $i \in \{a, t\}$ based on $\boldsymbol{Z}_l$, the received signals from the TUEs and AUEs reusing the pilot sequence q_k are extracted as $\boldsymbol{z}_{q_k l} = \boldsymbol{Z}_l \boldsymbol{\phi}_{q_{k_i}}^* / \sqrt{\tau_p}$. Hence, $\boldsymbol{z}_{q_k l}$ is expressed as

$$\boldsymbol{z}_{q_k l} = \sqrt{P_k \tau_p} \boldsymbol{h}_{k_i l} + \sum_{m \in \mathcal{S}_{q_k}, m \neq k_i} \sqrt{P_m \tau_p} \boldsymbol{h}_{k_m l} + \boldsymbol{n}_{q_k l} + \boldsymbol{w}_{q_k l}. \quad (4)$$

The estimation of channel gains, $\boldsymbol{h}_{k_i l}$, $\forall i \in \{a, t\}$ and $l \in \{1, 2, \cdots, L\}$, from $\boldsymbol{z}_{q_k l}$ is formulated as a non-linear regression problem, $\hat{\boldsymbol{h}}_{k_i l} = \mathcal{F}(\boldsymbol{z}_{q_k l})$, where $\mathcal{F}(\cdot)$ is the channel estimation function. Leveraging universal approximation theorem, $\mathcal{F}(\cdot)$ can be approximated by a parameterized neural network (NN). Towards this end, we propose two data-driven channel estimation approaches, namely, a) deep model-agnostic neural network (DMANN) and b) deep successive contamination cancellation (DSCC). In both schemes, the signals received at the CPU are exploited for joint channel estimation of all the communication links between APs and TUEs/AUEs. Therefore, the burden of computational complexity is transferred from all APs to CPU, where NN(s) is trained offline on the computationally intensive powerful servers. The trained model is deployed online for joint estimation of all the corresponding channels at CPU. It is worth pointing out that DMANN scheme aims at estimating multiple channels simultaneously using a single deep NN. Meanwhile, DSCC scheme successively estimates the channels of the participating UEs sharing the same pilot signals by iteratively cancelling interference caused by pilot contamination. Essentially, DMANN scheme requires a straightforward implementation whereas DSCC scheme is a high-performance yet more complex channel estimation scheme for CF-DMIMO networks.

A. DMANN Based Channel Estimation

1) NN Model: DMANN scheme is trained by considering $\boldsymbol{z}_{q_k l}$ and $\boldsymbol{h}_{k_m l}$, $m \in \mathcal{S}_{q_k}$, as the model's input and output in a supervised learning approach for estimating the channels between the l-th AP and the k_m-th UE, where $l \in \{1, 2, \cdots, L\}$ and $k_m \in \mathcal{S}_{q_k}$. A fully-connected deep NN is created with I_M features, O_M labels, G_M hidden layers, and N_g neurons in each hidden layer $g \in \{1, 2, \cdots, G_M\}$. We set $I_M = 2N$ and $O_M = 2\lceil K/\tau_p \rceil N$. Here, $\lceil \cdot \rceil$ denotes the ceiling operation. The reason for incorporating a scaling factor 2 (in defining I_M and O_M) is that we separate the real and imaginary parts of the signals involved in input (received signals) and output (channel gains).

2) Offline Training: We consider U and V as the total numbers of epoch for training and the numbers of batches in each training epoch, respectively. Let us denote $\mathcal{Z}^u = [\mathrm{Re}\{\boldsymbol{z}_{q_k l}^u\}, \mathrm{Im}\{\boldsymbol{z}_{q_k l}^u\}]^T$, for $q_k \in \{1, 2, \cdots, \tau_p\}$, $k \in \mathcal{S}_{q_k}$, and $l \in \{1, 2, \cdots, L\}$ as the input training symbols for epoch $u \in \{1, 2, \cdots, U\}$, where $\mathrm{Re}(x)$ ($\mathrm{Im}(x)$) represents the real (imaginary) part of complex variable x. Likewise, the corresponding channel gains $\mathcal{H}^u = [\mathrm{Re}\{\boldsymbol{h}_{k_m l}^u\}, \mathrm{Im}\{\boldsymbol{h}_{k_m l}^u\}]^T$, for $m \in \mathcal{S}_{q_k}$ are set as the output training symbols for epoch $u \in \{1, 2, \cdots, U\}$. It is worth mentioning that the training symbols can be generated using Monte Carlo simulations by leveraging (4) and the accurate statistical distributions of the underlying random variables or by the data collected from controlled experiments in lab environment or in live networks.

We define the set of parameters for the deep NN block as $\boldsymbol{\eta}_g = \{\boldsymbol{\alpha}_g, \boldsymbol{\beta}_g\}$ for a given hidden-layer $g \in \{1, 2, \cdots, G_M\}$, where $\boldsymbol{\alpha}_g$ and $\boldsymbol{\beta}_g$ denote the weights and bias factors, respectively. The estimation of $\mathcal{H}^u$ is therefore obtained as

$$\hat{\mathcal{H}}^u = f_{G_M}\left(\boldsymbol{\alpha}_{G_M} f_{G_M - 1}\left(\cdots f_1\left(\boldsymbol{\alpha}_1 \mathcal{Z}^u + \boldsymbol{\beta}_1\right) \cdots\right) + \boldsymbol{\beta}_{G_M}\right) \quad (5)$$

where $f_g(\cdot)$ is the activation function at the g-th hidden layer, $\forall g$. The optimal set of parameters of the DMANN model are obtained as

$$\boldsymbol{\eta}^* = \arg\min_{\eta} \mathbb{E}\left\{ \left|\left| H_u - \hat{\mathcal{H}}^u \right|\right|^2 \right\}, \quad (6)$$

where $\mathbb{E}\{\cdot\}$ and $||\cdot||$ represent the statistical expectation and L_2-norm operations, respectively. We deploy the hyperbolic tangent activation function at the output layer for the considered regression problem [12], and optimize (6) with the assistance of the training datasets $\mathcal{Z}_u$ and $\mathcal{H}_u$ while using stochastic gradient decent approach and the back propagation algorithms (e.g., Adam optimizer) [12].

3) Online Estimation: The trained DMANN model is deployed online for real-time channel estimation. In particular, $\{\boldsymbol{z}_{q_k l}[n]\}$, where $n \in \{1, 2, \cdots\}$, is fed online to the trained DMANN block and the estimated channel gains $\{\hat{\boldsymbol{h}}_{k_m l}[n]\}$ of all the available links of the considered network are obtained at the output layer of the trained DMANN block.

B. DSCC Based Channel Estimation

This data driven scheme leverages the idea behind successive interference cancellation (SIC) and sequentially removes the pilot contamination while exploiting the knowledge of

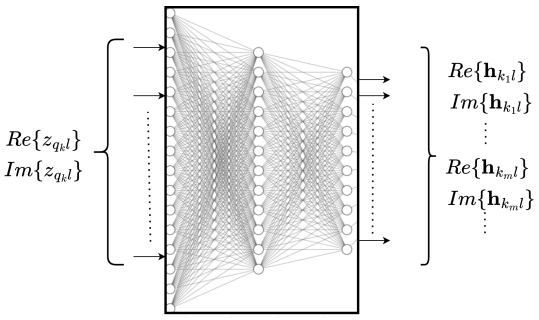

Fig. 2: DMANN Channel Estimation Model.

UEs responsible for pilot contamination. Unlike DMANN, we consider $J = \lceil K/\tau_p \rceil$ deep NN blocks for DSCC scheme enabled channel estimation. A fully-connected network is created for each deep NN block with I_S features, O_S labels, G_S hidden layers, and N_g neurons in each hidden layer $g \in \{1, 2, \cdots, G_S\}$, where we set $I_S = 2N$ and $O_S = 2N$ for each NN.

1) Offline Training: Defining the NN block for DSCC as $\mathcal{D}_j$, $j \in \{1, 2, \cdots, J\}$, the input and output training symbols in a given epoch $u \in \{1, 2, \cdots, U\}$ can be denoted as $[\mathrm{Re}\Big(\boldsymbol{z}^u_{q_k l} - \sum_{\tilde{j}=1}^{j-1} \boldsymbol{h}^u_{k_{\tilde{j}} l}\Big), \mathrm{Im}\Big(\boldsymbol{z}^u_{q_k l} - \sum_{\tilde{j}=1}^{j-1} \boldsymbol{h}^u_{k_{\tilde{j}} l}\Big)]^T$ and $[\mathrm{Re}\{\boldsymbol{h}^u_{k_j l}\}, \mathrm{Im}\{\boldsymbol{h}^u_{k_j l}\}]^T$, respectively. We follow the similar steps in setting the number of parameters for each deep NN as described for DMANN scheme.

2) Online Estimation: The trained D_j, $\forall j \in \{1, 2, \cdots, J\}$, are fed with $\{\boldsymbol{z}_{q_k l}[n]\}$, $n \in \{1, 2, \cdots\}$ online and yield the estimated channel gains $\{\hat{\boldsymbol{h}}_{k_m l}[n]\}$.

C. Computational Complexity

The offline training computational complexity of DMANN depends on the complexity of matrix multiplications in (5) for both forward and backward propagation, and it is obtained as $\mathcal{O}\left(2\left(2NN_1 + 2\lceil K/\tau_p \rceil NN_{G_M} + \sum_{g=2}^{G_M-1} N_{g-1}N_g\right)UV\right)$. Likewise, the computational complexity of DSCC in the (offline) training phase is obtained as $\mathcal{O}\left(2J\left(2NN_1 + 2NN_{G_S} + \sum_{g=2}^{G_S-1} N_{g-1}N_g\right)UV\right)$. Meanwhile, the computational complexity of both schemes during testing or real-time communications phase depends only on the forward propagation. Thus, the testing computational complexity of DMANN is obtained as $\mathcal{O}\left(2NN_1 + 2\lceil K/\tau_p \rceil NN_{G_M} + \sum_{g=2}^{G_M-1} N_{g-1}N_g\right)$. Similarly, the testing computational complexity of DSCC is obtained as $\mathcal{O}\left(J\left(2NN_1 + 2NN_{G_S} + \sum_{g=2}^{G_S-1} N_{g-1}N_g\right)\right)$. The time complexity of the trained NNs in both DMANN and DSCC schemes is linearly increased with L.

IV. Simulation Results

In this section, we evaluate the performance of the DMANN and DSCC channel estimation schemes for different use cases of the considered CF architecture by Monte Carlo simulations. We adopt a normalized mean-square error (NMSE) performance metric to represent the accuracy of channel estimation schemes. To compare the comparative performances, we consider two baseline schemes; a) least square (LS) and b) linear

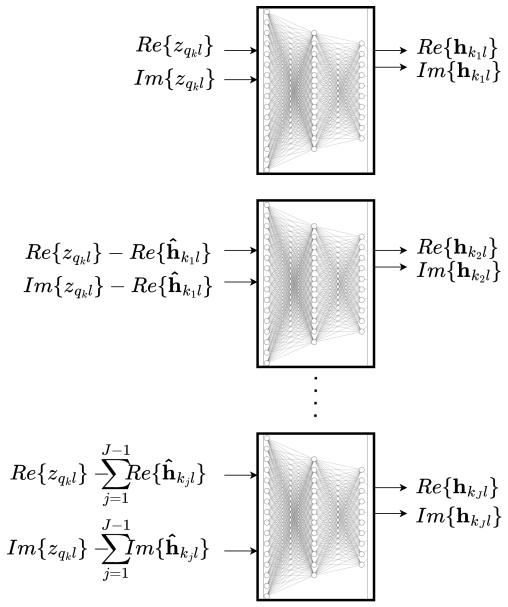

Fig. 3: DSCC Channel Estimation Model.

minimum mean square error (LMMSE) schemes. It is worth mentioning that LS and LMMSE schemes exhibit low complexity and high estimation accuracy, respectively. Throughout this section, we generate 10^6 realizations of dataset for training the proposed deep NNs. Moreover, another 10^6 realizations of dataset are produced to evaluate the NMSE performances of the proposed and baselines estimation schemes. For both training and testing dataset generation processes, we simulate a CF-DMIMO network with the following setting. In particular, we consider $N = 1$ and $\tau_c = 200$ all throughout the simulations with different L, K, and τ_p. We assume that APs are randomly deployed in an urban environment (with a path loss exponent of 3.67, which can be different in different environmental settings) with independent and uniform distribution to serve TUEs and AUEs jointly. Therefore, the large scale path loss can be expressed as $30.5 + 36.7 \log_{10}(d_e/1 \text{ km})$ as a function of effective distance d_e, when the operating carrier frequency is 2 GHz [13, Eq. (2.16)]. Here, $d_e = \sqrt{d_h^2 + d_v^2}$, where d_h and d_v represent the horizontal and vertical distances, respectively between two communication nodes, e.g., AP and TUE, AP and AUE, etc. We consider that all the APs are 10m above the ground level, on average. The AUEs are assumed to be hovering above the ground level with a mean height of 30m. The TUEs and AUEs are randomly deployed in a network area. The system operates over 20 MHz bandwidth and the noise floors at APs and CPU are -96 dBm and -126 dBm, respectively. The small-scale channel fading gain between the APs and TUEs follows i.i.d. Rayleigh distribution with zero mean and unit variance. Furthermore, the channel fading gain between the APs and AUEs follows i.i.d. Rician distribution with a Rice factor of 10 dB.

We leverage the DL toolbox of MATLAB to create the NN models, train them, and evaluate the trained (inference) models for the proposed DMANN and DSCC schemes. We consider three hidden-layers in each of the considered NN with 200 neurons in each hidden-layer. For all the demonstrated results, we incorporate Adam optimizer and set 1000 epochs with 64

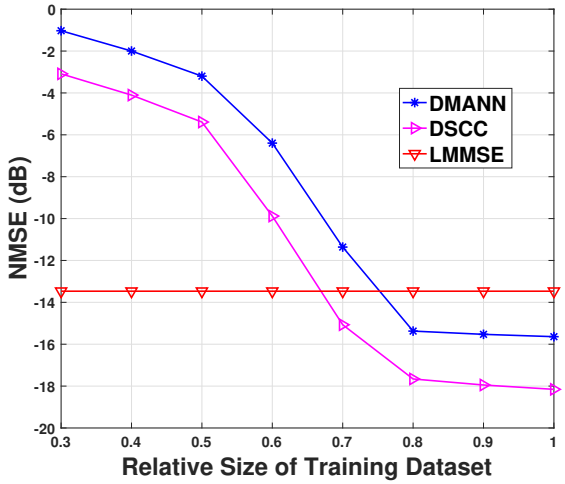

Fig. 4: NMSE vs. relative size of training dataset.

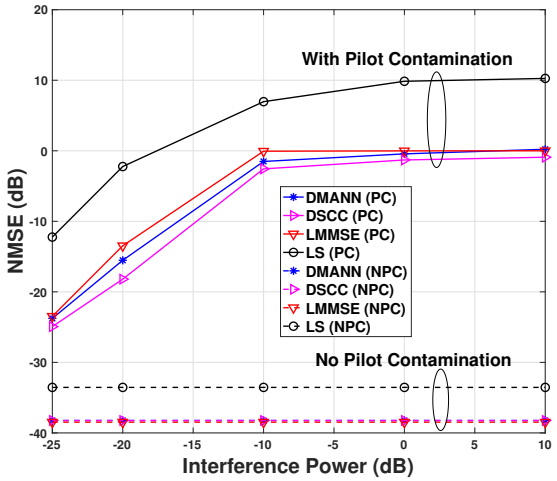

Fig. 5: NMSE vs. interference power.

batches in each epoch [14].

A. Impact of Training Dataset on Proposed Schemes

In Fig. 4, we demonstrate the effect of training dataset in the proposed DMANN and DSCC schemes. For this purpose, we set $L = 10$, $K = 3$ ($K_t = 2$ and $K_a = 1$), and $\tau_p = 1$. The transmit powers of TUE-1, TUE-2, and AUE are set to 0 dB, -20 dB, and -20 dB, respectively. We assume that the underlying noises at APs and CPU are Gaussian. All the participating UEs are 10m apart from AP on the horizontal plane. We evaluate the performance of channel estimation schemes for TUE-1 at CPU in the presence of TUE-2 and AUE, where all the users share the same pilot signal. In this figure, we show the NMSE performances of DMANN and DSCC schemes as a function of relative size (normalized) of training datasets, where the maximum size of the dataset is 10^6. Moreover, we also include the NMSE performance of LMMSE scheme, which is not a function of the training dataset. We observe that the NMSE of both DMANN and DSCC schemes decreases with the increasing size of training data set. Such an observation exemplifies the importance of accurate dataset-size to achieve competitive performances of the proposed data-driven schemes.

B. Comparison with Conventional Estimation Schemes

In Fig. 5, we show the effectiveness of pilot decontamination by the proposed channel estimation algorithms. We set $L = 10$ and $K = 3$ ($K_t = 2$ and $K_a = 1$). We denote the considered user equipment as TUE-1, TUE-2, and AUE. The transmit powers of TUE-2 and AUE are set same (at the transmitters) and considered as the interference power $\sigma_{\mathcal{I}}^2$ while estimating channels for TUE-1. The underlying noises at APs and CPU follow Gaussian distribution. We assume that the horizontal distances from AP and TUE-1, TUE-2, and AUE are 2m, 10m, and 3m, respectively. We demonstrate the average NMSE (averaged over all the considered APs) of the proposed and baseline channel estimation schemes at the CPU for TUE-1 as a function of $\sigma_{\mathcal{I}}^2$. We consider two scenarios (Scenario A and Scenario B). In Scenario A, we consider $\tau_p = 3$; hence, the pilot signals are judiciously allocated among all the users to make sure no pilot signals are shared between two or more users. Scenario B considers that a single pilot signal ($\tau_p = 1$) is shared among TUE-1, TUE-2, and AUE for channel estimation purpose. In Fig. 5, Scenarios A and B are denoted by phrases "NPC" and "PC", respectively. Our objective is to investigate the performance of channel estimation by one of the TUEs in the absence and presence of pilot contamination and thereby to demonstrate the effectiveness of the proposed DL schemes compared to conventional LS and LMMSE algorithms.

Fig. 5 demonstrates that NMSE for TUE-1 increases with increasing interference power from TUE-2 and AUE. We observe that the proposed DSCC scheme outperforms other considered schemes (LMMSE, DMANN, and LS) for Scenario B. This is because that the SIC approach in DSCC scheme systematically addresses the pilot contamination introduced by co-channel interference and nullify them. More specifically, the presence of LOS path between the AUE and AP results in the strong air-to-ground interference at the APs, leading to severe pilot contamination. We emphasize that the LMMSE does not have any interference rejection capability, and consequently, the co-channel interference introduced pilot contamination significantly degrades the LMMSE scheme's channel estimation accuracy compared to efficiently learned DSCC scheme. In contrast to Scenario B, we observe that the NMSE performance of DSCC and DMANN schemes are similar to LMMSE scheme for Scenario A, where there is no pilot contamination. Such an observation is expected since the LMMSE channel estimator is optimal over Gaussian distributed noise without any pilot contamination. However, LMMSE scheme not only requires the covariance matrix of the underlying MIMO channels to be known a priori but also has a high computational complexity of $\mathcal{O}\left(N^3L^3\right)$ due to the inversion of covariance matrix. In contrast, the real-time computational complexity of DL-based schemes is linearly scaled with L. In essence, our proposed DL approaches are efficient compared to both LMMSE and LS channel estimation schemes regardless of the presence of pilot contamination in the system.

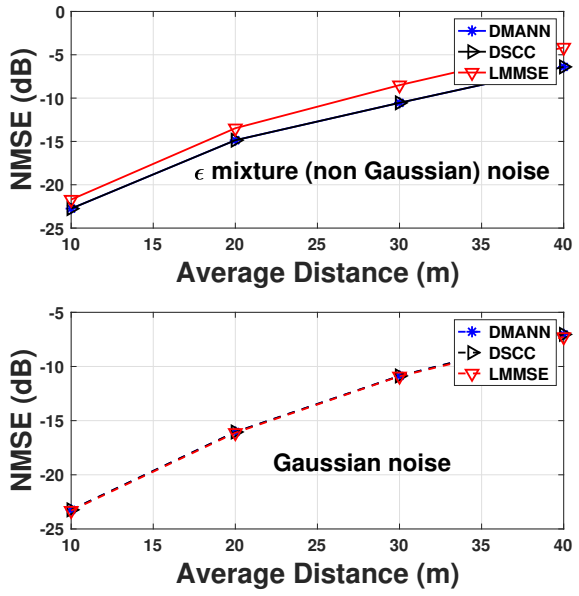

Fig. 6: NMSE vs. average distance between TUE and AP.

C. Performance in Non-Gaussian Noise

In Fig. 6, we demonstrate the performance of the proposed DMANN and DSCC schemes in the presence of non-Gaussian noise and interference. We consider two cases (Case 1 and Case 2) in this experiment. In Case 1, we assume that the APs are corrupted with ϵ-mixture Gaussian noise, which follows a heavier than Gaussian tail [15], [16]. In Case 2, we assume that all APs are corrupted with additive white Gaussian noise (AWGN). We set parameters $\epsilon = 0.10$ and $\kappa = 50$ for ϵ-mixture noise. It is worth mentioning that ϵ represents the fraction of the transmission duration when the impulsive noise component is active and κ denotes the ratio of the variances of impulsive to Gaussian noise components. Therefore, the smaller the ϵ and higher the κ, the more impulsive the noise is. In this figure, we consider $L = 10$, $K = 2$ ($K_t = 1$ and $K_a = 1$), and $\tau_p = 2$ (hence no pilot contamination) and demonstrate the NMSE to estimate the channel gain for TUE averaged over all the communication links. We set the average distance between TUE and all APs as 10m and change the average distance between APs and AUE to demonstrate NMSE performance of the proposed and LMMSE schemes. The performances of Case 1 and Case 2 are shown in the upper and lower sub-figures of Fig. 6, respectively.

We emphasize that channel estimation in non-Gaussian noise is a strong non-linear non-convex optimization problem, and cannot be optimally solved via LS or LMMSE methods. In this context, the proposed data driven approaches efficiently learn the system dynamics, thanks to the high-dimensional non-linearity of the NN. Essentially, in the presence of non-Gaussian noise, both DMANN and DSCC schemes noticeably outperform LMMSE scheme as depicted by the first sub-figure of Fig. 6. In contrast, since LMMSE is optimal for Gaussian noise, the NMSE of the proposed DL approaches and LMMSE coincide with each other as depicted from the second sub-figure of Fig. 6. Overall, the proposed DL schemes are efficient in estimating channels of the CF architecture in the presence of pilot contamination and/or non-Gaussian noise with low run-time complexity.

V. Conclusion

We developed two DL-enabled channel estimation schemes for CF-DMIMO networks. The NNs of the proposed schemes are trained offline with a large dataset considering different use-cases. The trained DL models are deployed online for real-time channel estimation and do not require the covariance matrix of high-dimensional MIMO channels and channel SNRs to be known as a priori. In the absence pilot contamination with Gaussian noise, the proposed DMANN and DSCC schemes achieve same NMSE performances similar to the conventional LMMSE scheme with low computational complexity, whereas the DSCC scheme outperforms LMMSE scheme in the presence pilot contamination. The proposed DL schemes also noticeably outperform the LMMSE scheme when the underlying noise is non-Gaussian.

References

[1] Ö. T. Demir, E. Björnson, and L. Sanguinetti, "Foundations of user-centric cell-free massive MIMO," *Foundations and Trends in Signal Processing*, vol. 14, no. 3-4, pp. 162–472, 2021.
[2] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1834–1850, 2017.
[3] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 77–90, 2019.
[4] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *IEEE communications surveys & tutorials*, vol. 21, no. 3, pp. 2334–2360, 2019.
[5] M. Z. Hassan, G. Kaddoum, and O. Akhrif, "Interference management in cellular-connected internet-of-drones networks with drone-pairing and uplink rate-splitting multiple access," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 1–19, 2022.
[6] S. Gao, P. Dong, Z. Pan, and G. Y. Li, "Deep learning based channel estimation for massive MIMO with mixed-resolution ADCs," *IEEE Communications Letters*, vol. 23, no. 11, pp. 1989–1993, 2019.
[7] H. Hirose, T. Ohtsuki, and G. Gui, "Deep learning-based channel estimation for massive MIMO systems with pilot contamination," *IEEE Open Journal of Vehicular Technology*, vol. 2, pp. 67–77, 2020.
[8] C.-J. Chun, J.-M. Kang, and I.-M. Kim, "Deep learning-based channel estimation for massive MIMO systems," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1228–1231, 2019.
[9] ——, "Deep learning-based joint pilot design and channel estimation for multiuser MIMO channels," *IEEE Communications Letters*, vol. 23, no. 11, pp. 1999–2003, 2019.
[10] J. Xu, P. Zhu, J. Li, and X. You, "Deep learning-based pilot design for multi-user distributed massive MIMO systems," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1016–1019, 2019.
[11] Y. Jin, J. Zhang, S. Jin, and B. Ai, "Channel estimation for cell-free mmWave massive MIMO through deep learning," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 10 325–10 329, 2019.
[12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
[13] Ö. T. Demir, E. Björnson, L. Sanguinetti *et al.*, "Foundations of user-centric cell-free massive MIMO," *Foundations and Trends® in Signal Processing*, vol. 14, no. 3-4, pp. 162–472, 2021.
[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[15] I. Ahmed, A. Nasri, R. Schober, and R. K. Mallik, "Asymptotic performance of generalized selection combining in generic noise and fading," *IEEE transactions on communications*, vol. 60, no. 4, pp. 916–922, 2012.
[16] I. Ahmed, A. Nasri, D. S. Michalopoulos, R. Schober, and R. K. Mallik, "Relay subset selection and fair power allocation for best and partial relay selection in generic noise and interference," *IEEE Transactions on Wireless Communications*, vol. 11, no. 5, pp. 1828–1839, 2012.