Deep Learning Approach to Channel Sensing and Hybrid Precoding for TDD Massive MIMO Systems

Kareem M. Attiah, Foad Sohrabi, and Wei Yu

Department of Electrical and Computer Engineering

University of Toronto, Toronto, Ontario M5S 3G4, Canada

Emails: {kattiah, fsohrabi, weiyu}@ece.utoronto.ca

Abstract—This paper proposes a deep learning approach to channel sensing and downlink hybrid analog and digital beamforming for massive multiple-input multiple-output systems with a limited number of radio-frequency chains operating in the time-division duplex mode at millimeter frequency. The conventional downlink precoding design hinges on the two-step process of first estimating the high-dimensional channel based on the uplink pilots received through the channel sensing matrices, then designing the precoding matrices based on the estimated channel. This two-step process is, however, not necessarily optimal, especially when the pilot length is short. This paper shows that by designing the analog sensing and the downlink precoding matrices directly from the received pilots without the intermediate channel estimation step, the overall system performance can be significantly improved. Specifically, we propose a channel sensing and hybrid precoding methodology that divides the pilot phase into an analog and a digital training phase. A deep neural network is utilized in the first phase to design the uplink channel sensing and the downlink analog beamformer. Subsequently, we fix the analog beamformers and design the digital precoder based on the equivalent low-dimensional channel. A key feature of the proposed deep learning architecture is that it decomposes into parallel independent single-user DNNs so that the overall design is generalizable to systems with an arbitrary number of users. Numerical comparisons reveal that the proposed methodology requires significantly less training overhead than the channel recovery based counterparts, and can approach the performance of systems with full channel state information with relatively few pilots.

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) communication has been widely recognized as one of the key enablers of millimeter-wave (mmWave) systems in the next generations of cellular networks owing to its excellent capabilities to combat severe path loss [1]. The use of massive MIMO technology in time-division duplex (TDD) operation is particularly attractive since channel reciprocity, when coupled with uplink channel estimation, allows for system scalability with respect to the number of antennas at the base station (BS) [2]. One of the key implementation issues for the massive MIMO technology is the high power expenditure and high cost associated with radio-frequency (RF) components, which render the deployment of the conventional fully-digital beamforming architecture practically infeasible. To alleviate this hardware limitation, the so-called hybrid beamforming architecture is proposed in [3], [4], in which the fully-digital beamformer is replaced by an analog beamformer, implemented using low-cost analog phase shifters, followed by a low-dimensional digital beamformer.

The design of hybrid precoders is in general a challenging task even under the perfect channel state information (CSI) assumption, because implementing the analog precoder via phase shifter array imposes a non-convex constant modulus constraint on the design parameters, which makes the optimization of precoders a computationally demanding task. More importantly, the hybrid architecture also severely limits the uplink channel sensing capability. Nevertheless, most current state-of-the-art hybrid precoding methods still follow the conventional communication system design methodology of decomposing the precoding process into the two separate components: (i) channel sensing and estimation, and (ii) downlink precoding design. In the channel estimation step, the spatial correlation of the mmWave channel is often exploited [5]–[8]. In particular, as the mmWave channels possess sparse representation in the angular domain, existing channel estimation methods typically use compressed sensing techniques to recover the parameters of the channel model based on, for example, the mean squared error metric. Subsequently, the BS then constructs the analog and digital precoding matrices assuming that the estimated CSI is perfect [9]–[11].

A key observation of this paper is that the above separated channel estimation and hybrid precoding approach is not necessarily optimal, especially when the pilot sequences are not sufficiently long to ensure the accurate recovery of the CSI. In this case, only a noisy estimate of the CSI is available at the BS, and the existing precoding methods may not provide good performance. The main contribution of this paper is to show that instead of the conventional two-step design that aims to reconstruct the channel with respect to some arbitrary distance metric (e.g., mean squared error), which may not necessarily be a good choice for designing the rate-maximizing hybrid precoders, it is possible to bypass the channel estimation and to directly design the precoding matrices from the baseband received pilots to obtain significant improvements in the overall system performance.

This paper proposes to use a deep learning methodology to directly design the channel sensing and hybrid precoding matrices based on the received pilots. However, because the hybrid precoder is composed of the analog and the digital components, directly designing both at the same time is

highly nontrivial to do in a data-driven fashion. The training complexity and the lack of generalizability with respect to the number of users render such a design impractical. In this paper, we present an alternative learning-based strategy that overcomes the preceding limitations but still bypasses model-based explicit channel estimation. To this end, we propose to divide the uplink pilot training into two separate phases where in the first phase we employ a deep neural network (DNN) that incorporates the design of the sensing matrices and directly maps the baseband pilot observations into the analog precoding matrix. In the second phase, we make use of the observation that once the analog beamforming matrix is determined, the end-to-end hybrid precoding system can be transformed into a low-dimensional fully-digital system, thereby allowing the digital precoding matrix to be designed. Numerical simulations indicate that such a strategy is advantageous, especially in the short pilot-length regime, and can achieve significantly better performance relative to the conventional channel recovery based schemes.

It is worth noting that a number of existing works have already proposed the use of DNNs in mmWave massive MIMO systems with hybrid architecture to replace key design components, e.g., channel estimation [12], precoding design with perfect/imperfect CSI [13]–[15], or both [16]. However, all of these works adopt the traditional separation of channel estimation and precoding and do not exploit the full potential of DNNs to design the end-to-end downlink precoding system.

II. SYSTEM MODEL

A. Downlink Transmission

We consider a TDD massive MIMO system operating in a block-fading frequency-flat mmWave environment in which a BS with M antennas and N_{RF} RF chains serves K single-antenna users. Using $\mathbf{s} \in \mathbb{C}^K$ to denote the collection of intended symbols to be sent to the users in the downlink, the complex baseband received signal at the k-th user is given by:

$$
\begin{aligned}
y_k &= \mathbf{h}_k^H \mathbf{V} \mathbf{s} + n_k, \\
&= \mathbf{h}_k^H \mathbf{v}_k s_k + \sum_{j \neq k} \mathbf{h}_k^H \mathbf{v}_j s_j + n_k, \quad (1)
\end{aligned}
$$

where $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_K] \in \mathbb{C}^{M \times K}$ is the downlink precoding matrix, $\mathbf{h}_k \in \mathbb{C}^M$ is the vector of complex downlink channel gains from the BS to user k, and $n_k \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise at user k. In addition, we assume that the BS employs a hybrid beamforming architecture with limited number of RF chains, N_{RF}, where $K \leq N_{\text{RF}} \leq M$. Thus, the overall precoding matrix must be of the form:

$$
\mathbf{V} = \mathbf{V}_{\text{RF}} \mathbf{V}_{\text{D}}, \quad (2)
$$

where $\mathbf{V}_{\text{D}} \in \mathbb{C}^{N_{\text{RF}} \times K}$ is the low-dimensional digital precoding matrix and $\mathbf{V}_{\text{RF}} \in \mathbb{C}^{M \times N_{\text{RF}}}$ is the analog precoding matrix of phase shifters whose entries satisfy a constant modulus constraint, i.e., $[\mathbf{V}_{\text{RF}}]_{ij} = e^{\imath \phi_{ij}}$ where $\imath$ is the imaginary unit. Finally, we impose a power constraint on the transmitted signal by requiring $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_K$ and $\|\mathbf{V}\|_F^2 \leq P_{\text{D}}$.

B. Channel Model

We consider a mmWave propagation environment with limited number of scatterers. Accordingly, the k-th user channel vector $\mathbf{h}_k$ follows a sparse channel model consisting of L_p dominant paths:

$$
\mathbf{h}_k = \frac{1}{\sqrt{L_p}} \sum_{\ell=1}^{L_p} \alpha_{\ell,k} \mathbf{a}_t(\theta_{\ell,k}), \quad (3)
$$

where $\alpha_{\ell,k} \sim \mathcal{CN}(0,1)$ is the complex gain of the ℓ-th path between the BS and user k, $\theta_{\ell,k}$ is the corresponding angle of departure (AoD) and $\mathbf{a}_t(\cdot)$ is the array response vector. Assuming a uniform linear array with M antenna elements at the BS, we have:

$$
\mathbf{a}_t(\theta) = \left[1, e^{\imath \frac{2\pi}{\lambda} d \sin(\theta)}, \ldots, e^{\imath \frac{2\pi}{\lambda} d(M-1)\sin(\theta)}\right]^T, \quad (4)
$$

where λ is the wavelength and d is the antenna spacing.

C. Uplink Pilot-Based Channel Sensing

We assume that BS has no prior knowledge of the channel and must learn an estimate of the CSI (either explicitly or implicitly) through uplink pilot training, based on channel reciprocity in a TDD operation [17]. To this end, we consider an uplink pilot training phase consisting of L time frames, where the users send K mutually orthogonal pilot sequences repeatedly in each time frame. Because of the limited RF chain constraint, the BS must sense the channel using analog sensing (or combining) matrices of form:

$$
\mathbf{W}_{\text{RF}}^{(\ell)} \in \mathbb{C}^{N_{\text{RF}} \times M}, \qquad \ell \in \{1, \ldots, L\}, \quad (5)
$$

whose entries must satisfy the constant modulous constraint $[\mathbf{W}_{\text{RF}}^{(\ell)}]_{ij} = e^{\imath \psi_{ij}}$.

Let $\mathbf{X} \in \mathbb{C}^{K \times K}$ be the unitary matrix whose rows correspond to the orthogonal pilots sent by the users, the $N_{\text{RF}} \times K$ received signal at the BS in the ℓ-th time frame is given by:

$$
\mathbf{R}^{(\ell)} = \sqrt{P_{\text{U}}} \mathbf{W}_{\text{RF}}^{(\ell)} \mathbf{H} \mathbf{X} + \mathbf{W}_{\text{RF}}^{(\ell)} \mathbf{N}^{(\ell)}, \quad \ell \in \{1, \ldots, L\}, \quad (6)
$$

where $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K] \in \mathbb{C}^{M \times K}$ and $\mathbf{N}^{(\ell)} = \left[\mathbf{n}_1^{(\ell)}, \ldots, \mathbf{n}_K^{(\ell)}\right] \in \mathbb{C}^{M \times K}$ are the channel and noise matrices for all users, and P_{U} is the allocated power for each user in a single time frame. Typically, $N_{\text{RF}} \ll M$, hence we need multiple time frames (i.e., $L > 1$) to capture the user channels.

Since the pilots are orthogonal, we have $\mathbf{X}\mathbf{X}^H = \mathbf{I}_K$, so the BS can right-multiply the received signal by $\mathbf{X}^H$ to obtain:

$$
\tilde{\mathbf{Y}}^{(\ell)} = \sqrt{P_{\text{U}}} \mathbf{W}_{\text{RF}}^{(\ell)} \mathbf{H} + \mathbf{Z}^{(\ell)}, \quad \ell \in \{1, \ldots, L\}, \quad (7)
$$

where $\mathbf{Z}^{(\ell)} = \mathbf{W}_{\text{RF}}^{(\ell)} \mathbf{N}^{(\ell)} \mathbf{X}^H$ is the corresponding noise matrix in the ℓ-th time frame. Denoting the k-th column of $\tilde{\mathbf{Y}}^{(\ell)}$ by $\tilde{\mathbf{y}}_k^{(\ell)}$, we can write:

$$
\tilde{\mathbf{y}}_k^{(\ell)} = \sqrt{P_{\text{U}}} \mathbf{W}_{\text{RF}}^{(\ell)} \mathbf{h}_k + \mathbf{z}_k^{(\ell)}, \quad \ell \in \{1, \ldots, L\}. \quad (8)
$$

The overall $N_{\text{RF}}L$ received vector (after multiplying by $\mathbf{X}^H$) due to the k-th user transmissions in the entire pilot transmission phase is obtained by vertically stacking $\tilde{\mathbf{y}}_k^{(\ell)}$:

$$\tilde{\mathbf{y}}_k \triangleq \begin{bmatrix} \tilde{\mathbf{y}}_k^{(1)} \\ \vdots \\ \tilde{\mathbf{y}}_k^{(L)} \end{bmatrix} = \sqrt{P_U} \begin{bmatrix} \mathbf{W}_{\text{RF}}^{(1)} \\ \vdots \\ \mathbf{W}_{\text{RF}}^{(L)} \end{bmatrix} \mathbf{h}_k + \mathbf{z}_k, \quad (9)$$

where $\mathbf{z}_k \in \mathbb{C}^{N_{\text{RF}}L}$ is the corresponding noise vector.

D. Design Objective

The channel sensing and hybrid precoding problem can now be stated as follows. Assuming the channel model as described by (3), we seek to determine the sensing matrices $\{\mathbf{W}_{\text{RF}}^{(\ell)}\}$, so that the analog and digital precoding matrices $\mathbf{V}_{\text{RF}}$ and $\mathbf{V}_{\text{D}}$ can be designed as a function of the received pilot signals $\tilde{\mathbf{y}}_k$. More specifically, we choose the downlink sum rate as the overall objective function, i.e.,

$$R = \sum_k R_k, \quad (10)$$

where R_k is the achievable rate at user k as given by:

$$R_k = \log_2 \left(1 + \frac{|\mathbf{h}_k^H \mathbf{v}_k|^2}{\sum_{j \neq k} |\mathbf{h}_k^H \mathbf{v}_j|^2 + \sigma^2} \right). \quad (11)$$

Note that the dependence of the sum rate expression on the choice of the sensing matrices $\{\mathbf{W}_{\text{RF}}^{(\ell)}\}$ is implicit. In particular, from (9) we observe that the design of the sensing matrices $\{\mathbf{W}_{\text{RF}}^{(\ell)}\}$ is crucial as they serve the important role of summarizing the information about the user channels, and hence can directly affect the quality with which we construct the hybrid precoding matrices.

III. PROPOSED PRECODING SCHEME

A. Motivation

The conventional communication system design typically relies on analytic methods to separately carry out the channel estimation and precoding design steps. In such schemes, some channel model is typically assumed; the model parameters are then estimated according to some distance metric. The issue is, however, that this distance metric may not exactly match the ultimate goal of maximizing the overall system performance. We argue that this possible mismatch is a key limitation of the conventional system design, but is otherwise necessary to ensure that the conventional design problem is tractable.

The advent of data-driven techniques now makes a different paradigm possible. It is no longer necessary to adhere to the above separation, i.e., it may be beneficial to bypass the CSI estimation altogether and to directly design the hybrid precoders from the baseband received pilots. In other words, because of the power of DNN as a universal function approximator, it is now possible to pursue an end-to-end design encompassing both channel sensing/estimation and hybrid precoding at the same time. More precisely, the design problem

becomes that of learning the sensing matrices $\{\mathbf{W}_{\text{RF}}^{(\ell)}\}$ as well as the direct mapping $\mathcal{F}(\cdot)$ given by:

$$(\mathbf{V}_{\text{RF}}^*, \mathbf{V}_{\text{D}}^*) = \mathcal{F}(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_K), \quad (12)$$

where $\tilde{\mathbf{y}}_k$'s are the received pilots, and $\mathbf{V}_{\text{RF}}^*$ and $\mathbf{V}_{\text{D}}^*$ are the rate-maximizing precoding matrices. Once again, we emphasize that the dependence of $\tilde{\mathbf{y}}_k$ on the sensing matrices $\{\mathbf{W}_{\text{RF}}^{(\ell)}\}$ following from (9).

A naive implementation of this approach is, however, hindered by complexity. Indeed, the multiplicative interaction between the analog and digital precoding matrices renders the training of this DNN rather unwieldy even for a small number of users. Furthermore, the multi-user system described in Section II-A is expected to serve a varying number of users. However, for a DNN to output the digital precoding matrix, it must have an output dimension equal to the number of users. To support such dynamic operation, it is desirable to design the DNN on a per-user basis. To this end, we develop a novel DNN architecture below. The proposed architecture bears some similarity to the one in [16], but with the key difference that our approach bypasses the intermediate channel estimation step.

B. Learning Approach to Hybrid Precoding

The proposed learning-based scheme decouples the design of the analog and digital precoding matrices. Specifically, we propose to divide the overeall pilot training phase of L frames into an analog training phase of L_a frames and a digital training phase of L_d frames with $L = L_a + L_d$. In the analog training phase, the baseband signals received in the first L_a frames are mapped directly into the analog precoding matrix using a DNN. In the digital training phase, the analog beamformers are fixed and the digital precoding matrix is obtained based on the pilot signals received in the remaining L_d frames.

1) Analog Training Phase: Let us denote the received pilot from the k-th user in the first L_a time frames by $\tilde{\mathbf{y}}_k^a \triangleq \left[\left(\tilde{\mathbf{y}}_k^{(1)} \right)^T, \ldots, \left(\tilde{\mathbf{y}}_k^{(L_a)} \right)^T \right]^T$, then by (9) we may write:

$$\tilde{\mathbf{y}}_k^a = \sqrt{P_U} \mathbf{W}_{\text{RF}}^a \mathbf{h}_k + \mathbf{z}_k^a, \quad (13)$$

where $\mathbf{W}_{\text{RF}}^a \triangleq \left[\left(\mathbf{W}_{\text{RF}}^{(1)} \right)^T, \ldots, \left(\mathbf{W}_{\text{RF}}^{(L_a)} \right)^T \right]^T$. In the analog training phase, we aim to design an DNN to learn the sensing matrix $\mathbf{W}_{\text{RF}}^a$ as well as the (direct) mapping from $\{\tilde{\mathbf{y}}_k^a\}_{k=1}^K$ to $\mathbf{V}_{\text{RF}}$. However, instead of using an overall DNN for all K users, we propose the following architecture that decomposes the overall DNN into K parallel *single-user DNNs* that learn the sensing matrix $\mathbf{W}_{\text{RF}}^a$ as well as K parallel mappings $\mathcal{G}^{(k)}(\cdot)$ from $\tilde{\mathbf{y}}_k^a$ to the k-th column of $\mathbf{V}_{\text{RF}}$, i.e.,

$$\mathbf{v}_{\text{RF}}^{(k)} = \mathcal{G}^{(k)}(\tilde{\mathbf{y}}_k^a), \quad \forall k \in \{1, \ldots, K\}. \quad (14)$$

The idea is that the analog precodering matrix $\mathbf{V}_{\text{RF}}$ consists of the analog precoders $\mathbf{v}_{\text{RF}}^{(k)}$ for each of the K users, which can be learned separately. This works when the system is fully loaded with $N_{\text{RF}} = K$, which we assume in this section.

Ideally, we would like to learn the mappings $\mathcal{G}^{(k)}(\cdot)$ to maximize the system sum rate objective. But the sum rate depends on the digital beamformers, which are not yet designed. So instead, we approximate each user's achievable rate based on its analog precoder only, without taking interference into account. This can be justified, because the subsequent digital phase will use zero-forcing (ZF) precoder (with equal power allocation) to eliminate interference. In this case, the loss function becomes

$$\mathcal{L} = -\sum_{k=1}^{K} \log_2 \left(1 + \frac{P_{\mathrm{D}}}{MK\sigma^2} \left| \mathbf{h}_k^H \mathbf{v}_{\mathrm{RF}}^{(k)} \right|^2 \right). \qquad (15)$$

Notice that the above loss function consists of K independent terms, each of which is a function of the output of k-th user's $\mathcal{G}^{(k)}(\cdot)$. A key consequence is that the overall training process can now be decomposed into learning K independent single-user mappings.

We can also incorporate the training of the channel sensing matrix $\mathbf{W}_{\mathrm{RF}}^a$ into the DNN architecture. As can be seen from (13), $\mathbf{W}_{\mathrm{RF}}^a$ can actually be incorporated as a trainable linear layer in each single-user DNN. Note that the sensing matrix is common to all users. So by tying the parameters of this linear layer together across the single-user DNNs, a common channel sensing matrix can be readily designed.

To further facilitate training, we can also share the parameters of the mappings $\mathcal{G}^{(k)}(\cdot)$ across the users, call it $\mathcal{G}(\cdot)$. In this case, all the single-user DNNs now have identical structure, i.e., with the user channel as input and the corresponding analog precoder as output, and have the same loss function and same parameters. As a consequence, training the overall DNN using a batch of independent channel realizations is now equivalent to training just one single-user DNN with K times as many channel samples. The key advantage is that by employing K identically trained single-user DNNs, we now arrive at a *generalizable* architecture suitable for systems with an arbitrary number of users.

2) Digital Training Phase: Once $\mathbf{V}_{\mathrm{RF}}$ is determined, we proceed to find $\mathbf{V}_{\mathrm{D}}$. A key insight is that due to channel reciprocity, the designed downlink analog precoding matrix must also be a suitable uplink channel sensing matrix. Further, if we set the analog sensing matrix to be the Hermitian of the analog precoding matrix, i.e., $\mathbf{W}_{\mathrm{RF}} = \mathbf{V}_{\mathrm{RF}}^H$, then the end-to-end low-dimensional fully-digital system would also have uplink-downlink reciprocity. This low-dimensional channel can be obtained by the additional L_d pilots.

More precisely, the downlink equivalent low-dimensional channel $\mathbf{H}_{\mathrm{eq}} \in \mathbb{C}^{N_{RF} \times K}$, as given by (1) and (2), is:

$$\mathbf{H}_{\mathrm{eq}} = \mathbf{V}_{\mathrm{RF}}^H \mathbf{H}. \qquad (16)$$

We can estimate $\mathbf{H}_{\mathrm{eq}}$ in the uplink by setting

$$\mathbf{W}_{\mathrm{RF}}^{(\ell)} = \mathbf{V}_{\mathrm{RF}}^H, \quad \forall \ell \in \{L_a + 1, \ldots, L\}, \qquad (17)$$

and by observing from (7), i.e.,

$$\tilde{\mathbf{Y}}^{(\ell)} = \sqrt{P_{\mathrm{U}}} \overbrace{\mathbf{V}_{\mathrm{RF}}^H \mathbf{H}}^{\mathbf{H}_{\mathrm{eq}}} + \mathbf{Z}^{(\ell)}, \quad \ell \in \{L_a + 1, \ldots, L\}, \qquad (18)$$
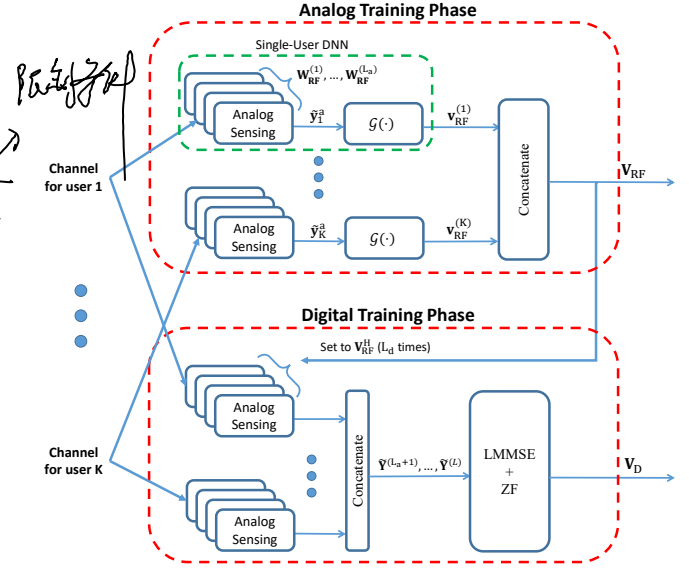

Fig. 1: Block diagram of the proposed precoding approach for designing the analog and digital precoding matrices.

that $\tilde{\mathbf{Y}}^{(L_a+1)}, \ldots, \tilde{\mathbf{Y}}^{(L)}$ are exactly the repeated transmissions of the pilots through the above equivalent channel. We can now derive the linear minimum mean squared error (LMMSE) estimator of $\mathbf{H}_{\mathrm{eq}}$. Assuming that the BS antenna elements see uncorrelated channels (i.e., $\mathbb{E}[\mathbf{h}_k \mathbf{h}_k^H] = \mathbf{I}_M, \forall k$), the LMMSE estimator for the equivalent channel is given by:

$$\hat{\mathbf{H}}_{\mathrm{eq}} = \frac{\sqrt{P_{\mathrm{U}}}}{P_{\mathrm{U}} L_d + \sigma^2} \sum_{\ell=L_a+1}^{L} \tilde{\mathbf{Y}}^{(\ell)}. \qquad (19)$$

Finally, we apply conventional linear precoding schemes such as ZF to determine the digital precoder:

$$\mathbf{V}_{\mathrm{D}} = \gamma_{\mathrm{ZF}} \hat{\mathbf{H}}_{\mathrm{eq}} (\hat{\mathbf{H}}_{\mathrm{eq}}^H \hat{\mathbf{H}}_{\mathrm{eq}})^{-1}, \qquad (20)$$

where γ_{ZF} is a scalar ensuring that the power constraint is satisfied. Fig. 1 illustrates the proposed scheme.

Note that the proposed architecture overcomes the limitations discussed in the preceding section, specifically:

- The proposed approach determines the hybrid precoding matrices from the received pilots without having to estimate the high-dimensional channel first.
- As compared to the direct design in (12), the mapping $\mathcal{G}(\cdot)$ is now significantly simpler to learn.
- The output dimension of the single-user DNN is not a function of the number of users, so the proposed scheme is generalizable to systems with any number of users.

C. Single-User DNN Architecture

We now discuss the implementation details of the single-user DNN that learns the mapping $\mathcal{G}(\cdot)$ and the sensing matrix $\mathbf{W}_{\mathrm{RF}}^a$. In particular, the proposed architecture, shown in Fig. 2 for the k-th user, comprises a sensing layer, intermediate layers, and a phase mapping layer. In the offline training of the DNN, the channel and noise realizations $\{\mathbf{h}_k, \mathbf{n}_k\}$ are fed into the first layer (i.e., the sensing layer) and the corresponding
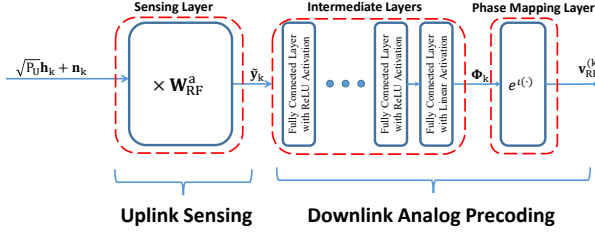
Fig. 2: The architecture of the proposed single-user DNN for uplink sensing and downlink analog precoding design.

Fig. 3: Performance comparison of the proposed approach against existing schemes for different values of L when $K = N_{\text{RF}} = 4$.

analog precoding vector $\mathbf{v}_{\text{RF}}^{(k)}$ is produced by the final layer (i.e., the phase mapping layer). After training, we use K identical single-user DNNs for operation wherein the sensing layer is used as the analog sensing matrices in the uplink pilot phase and the received pilot signal $\tilde{\mathbf{y}}_k^a$ is fed into the intermediate layers to produce the downlink analog precoder.

1) Sensing Layer: The sensing matrices can be seen as a trainable linear layer of the DNN with the channel realizations of the k-th user plus the noises as the input. Indeed, it can be seen from (13) that $\mathbf{W}_{\text{RF}}^a$ are just weights in linear neural network layers with the additional unit modulus constraint on the entires. Incidentally, existing deep learning libraries (e.g., TensorFlow [18]) allow the creation of customized layers with trainable parameters. In this case, we take the trainable parameters to be the phases of the sensing matrix $\mathbf{W}_{\text{RF}}^a$.

2) Intermediate Layers: The input to the second stage is the baseband received pilot vector $\tilde{\mathbf{y}}_k^a$, whereas the output of this stage is the vector containing the phases of the corresponding analog precoding vector, denoted by $\mathbf{\Phi}_k$. We consider an intermediate stage consisting of R layers, where each layer is fully connected with rectified linear unit (ReLU) activation, except for the last layer which has linear activation. Mathematically, the output of the intermediate layers is:

$$\mathbf{\Phi}_k = \mathbf{W}_R \mathcal{R}\left(\ldots \mathcal{R}\left(\mathbf{W}_1 \tilde{\mathbf{y}}_k^a + \mathbf{b}_1\right)\ldots\right) + \mathbf{b}_R, \qquad (21)$$

where $\mathcal{R}(\cdot) = \max(\cdot, 0)$ is the ReLU activation function and $\{\mathbf{W}_r, \mathbf{b}_r\}$ is the set of trainable parameters (i.e., weights and biases) for layer r.

3) Phase Mapping Layer: The output of this layer is the analog precoding vector for user k, hence we need to enforce the constant modulus constraint. To accomplish this, we adopt the idea in [13] where the output of the intermediate layers in (21) the phases of the analog combining vector. In this case, this layer applies the entry-wise transformation $[\mathbf{v}_{\text{RF}}^{(k)}]_j = e^{\imath[\mathbf{\Phi}_k]_j}$, where $[\cdot]_j$ denotes the j-th entry of a vector. Such a transformation layer can be implemented using the so-called *Lambda layer* in existing deep learning libraries.

4) Loss Function: As explained earlier, the loss function for the k-th single-user DNN can be taken as the approximate single-user achievable rate

$$\mathcal{L}^{(k)}(\mathbf{h}_k, \mathbf{v}_{\text{RF}}^{(k)}) = -\log_2\left(1 + \frac{P_{\text{D}}}{MK\sigma^2}\left|\mathbf{h}_k^H \mathbf{v}_{\text{RF}}^{(k)}\right|^2\right). \qquad (22)$$
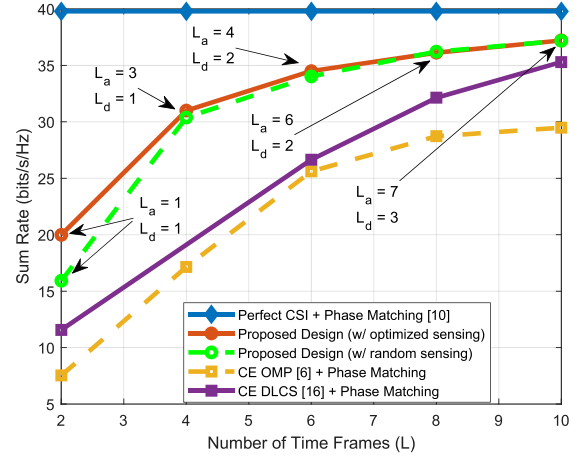
IV. NUMERICAL RESULTS

In this section, we compare the performance of the proposed hybrid precoding approach against several existing schemes in the literature. To this end, we consider a single-cell massive MIMO system in TDD operation where the BS is equipped with uniform linear array with $M = 64$ transmit antennas with half-wavelength antenna separation. The channel for each user follows a sparse channel model introduced in Section II-A with $L_p = 6$, and the path AoDs follow a uniform distribution in the interval $[0, 2\pi]$. Finally, we set both the uplink and downlink signal-to-noise ratios to be SNR $= 10 \log P/\sigma^2 = 20$ dB, where $P = P_{\text{U}} = P_{\text{D}}$. In our simulations, we consider comparisons against the following schemes:

1) Hybrid design with full CSI: This design follows the work in [10] which assumes that the channel is perfectly known to the BS and is included to serve as the baseline for the other schemes. In this design, the analog precoding matrix is obtained by matching to the phases of the channel matrix (i.e., $[\mathbf{V}_{\text{RF}}]_{ij} = [\mathbf{H}]_{ij}/|[\mathbf{H}]_{ij}|$), whereas the digital precoding matrix is obtained as the ZF of the equivalent channel.

2) Hybrid design with imperfect CSI using orthogonal matching pursuit (OMP) [6]: In this design, the parameters of the high-dimensional channel of each user are estimated using the OMP algorithm and the precoding matrices are subsequently determined from the recovered channel using the scheme in [10] explained above.

3) Hybrid design with imperfect CSI using deep learning compressive sensing (DLCS) [16]: In an attempt to overcome the greedy manner in which the OMP resolves the channel paths, the authors of [16] develop an OMP-inspired DNN architecture to recover the channel parameters. We use the same DNN architecture here to estimate the channel and design the precoding matrices from the recovered channel according to the design scheme in [10].

We implement the proposed DNN in Fig. 2 and the DLCS using TensorFlow. For the proposed DNN, we set the number of intermediate layers as $R = 3$, with dense layers of widths 2048, 1024, and 512, respectively. For faster convergence, each
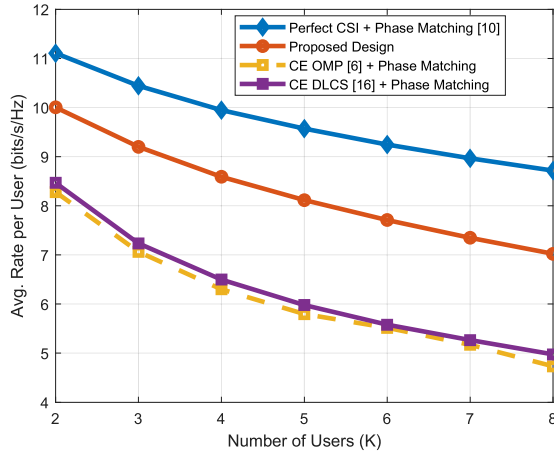
Fig. 4: Generalizability with respect to the number of users for the proposed scheme. Here, we set $N_{\mathrm{RF}} = 8$ and vary K from 2 to 8 users.

dense layer is preceded by a batch normalization layer. In addition, for the first layer we initially set the phases of the analog precoding matrices at random according to a uniform distribution in the interval $[0, 2\pi]$. For the DLCS, we consider a 4-layer dense network of widths 1024, 512, 256, and 128. The training of both networks is performed in batches and over several epochs. Each batch contains 500 samples and there are 20 batches per epoch. We fix the channels and noise distributions in the uplink training phase so we can generate as many data samples as needed for training the DNNs.

In Fig. 3, we plot the sum rate against the number of pilot frames L for different schemes with $N_{\mathrm{RF}} = K = 4$. For the proposed approach, L_a and L_d are indicated by arrows. We see that the proposed scheme significantly outperforms both the OMP and DLCS approaches. For example, the proposed approach achieves over 90% of the total sum rate of the full CSI systems at $L = 8$, as compared to $L = 10$ for the DLCS, thereby indicating over 20% saving in pilot overhead relative to the conventional schemes. This supports our main claim that the separation of channel estimation and precoding design is not optimal in hybrid systems with limited training overhead. We note here that in the limit of long pilot length, however, the conventional separate scheme can do reasonably well. Finally, we observe that the effect of optimizing the channel sensing matrix is noticeable only at a very short pilot length.

Next, we study the generalizability with respect to the number of users. To this end, we set $L = 3$, $N_{\mathrm{RF}} = 8$, and K is varied from 2 to 8. For the proposed scheme, we set $L_a = 2$ and $L_d = 1$, and design a single-user DNN to use for the $K = 8$ case. If $K < 8$, we make a simplifying choice of utilizing only K RF chains for downlink precoding. To determine the analog precoder, we use $K < 8$ copies of the single-user DNNs. Note that this simplifying choice must also be made for the conventional schemes to work. In Fig. 4, we plot the average rate per user against the number of users for all four schemes. We observe that our approach is able to support generalizable operation with respect to the number of users while maintaining significantly better performance than the channel recovery based approaches.

V. Conclusion

This paper addresses the design of the analog and digital precoding matrices in a mmWave TDD massive MIMO system. The proposed learning-based precoding strategy is generalizable for any number of users and overcomes the limitations of the existing schemes by constructing the precoding matrices directly from the received pilots without the intermediate step of estimating the high-dimensional channel. Numerical evaluations indicate significant gains in spectral efficiency relative to the channel recovery based schemes.

References

[1] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[3] X. Zhang, A. F. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Oct. 2005.

[4] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[5] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[6] J. Lee, G. 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, June 2016.

[7] C. Hu, L. Dai, T. Mir, Z. Gao, and J. Fang, "Super-resolution channel estimation for mmWave massive MIMO with hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8954–8958, Sept. 2018.

[8] F. Bellili, F. Sohrabi, and W. Yu, "Generalized approximate message passing for massive MIMO mmWave channel estimation with Laplacian prior," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3205–3219, May 2019.

[9] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Jan. 2016.

[10] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.

[11] A. Li and C. Masouros, "Hybrid precoding and combining design for millimeter-wave multi-user MIMO based on SVD," in *IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.

[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] T. Lin and Y. Zhu, "Beamforming design for large-scale antenna arrays using deep learning," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 103–107, Jan. 2020.

[14] A. M. Elbir, "CNN-based precoder and combiner design in mmWave MIMO systems," *IEEE Commun. Lett.*, vol. 23, no. 7, pp. 1240–1243, July 2019.

[15] A. M. Elbir and A. K. Papazafeiropoulos, "Hybrid precoding for multiuser millimeter wave massive MIMO systems: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 552–563, Jan. 2020.

[16] W. Ma, C. Qi, Z. Zhang, and J. Cheng, "Sparse channel estimation and hybrid precoding using deep learning for millimeter wave massive MIMO," *IEEE Trans. Commun.*, vol. 68, no. 5, pp. 2838–2849, Jan. 2020.

[17] G. S. Smith, "A direct derivation of a single-antenna reciprocity relation for the time domain," *IEEE Trans. Antennas Propag.*, vol. 52, no. 6, pp. 1568–1577, June 2004.

[18] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," Mar. 2016. [Online]. Available: https://arxiv.org/abs/1603.04467