HAL
open science

Federated Channel-Beam Mapping: from sub-6GHz to mmWave

Irched Chafaa, Romain Negrel, Elena Veronica Belmega, Merouane Debbah

To cite this version:

Irched Chafaa, Romain Negrel, Elena Veronica Belmega, Merouane Debbah. Federated Channel-Beam Mapping: from sub-6GHz to mmWave. IEEE Wireless Communications and Networking Conference, Mar 2021, Nanjing (on line), China. 10.1109/WCNCW49093.2021.9420006 . hal-03153692

HAL Id: hal-03153692

https://hal.science/hal-03153692

Submitted on 26 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Federated Channel-Beam Mapping: from sub-6GHz to mmWave

Irched Chafaa[⋆†], Romain Negrel[§], E.Veronica Belmega[⋆], and Mérouane Debbah[†‡]

[⋆]*ETIS UMR8051, CY Cergy Paris Université, ENSEA, CNRS, Cergy, France*
[†]*L2S, UMR 8506, Université Paris-Saclay, CentraleSupélec, CNRS, Gif-sur-Yvette, France*
[§] *ESIEE Paris, Université Gustave Eiffel, Marne-la-Vallée, France*
[‡] *Lagrange Mathematical and Computing Research Center, Paris, France*
irched.chafaa@ensea.fr, romain.negrel@esiee.fr, belmega@ensea.fr, merouane.debbah@centralesupelec.fr

***Abstract*—Accurate beamforming is a critical challenge for mmWave communications. Because of the large training overhead of beam training at high frequencies, it becomes relevant to exploit the available knowledge at sub-6GHz to predict the mmWave beamforming vectors using deep learning tools. In addition, fully centralized learning (CL) approaches require training over all the users data, rising major issues in terms of signaling and computational cost. To address these issues, we propose a federated learning (FL) scheme in a wireless network composed of multiple communicating links (access points – users) to predict directly the downlink mmWave beamforming vectors from the uplink sub-6GHz channels. The access points train their local deep neural networks using local data and only share their model parameters to obtain an average global one, which improves the quality of their prediction in terms of data rate. Our experiments demonstrate the potential and robustness of our proposed scheme especially under difficult conditions, performing close to the fully centralized one. When the training data is scarce, the relative gain of our scheme can reach up to 50% compared to a fully distributed one. Remarkably, our scheme can even outperform the fully centralized one when the quality of the training data is poor, enjoying a relative gain of up to 14%.**

***Index Terms*—mmWave beamforming, deep neural networks, distributed learning, federated learning**

I. Introduction

[1]The mmWave spectrum, which usually refers to frequencies higher than 28 GHz, has been identified as a key enabler for 5G and beyond [1] to relieve the already crowded sub-6GHz spectrum. However, the signal propagation at high frequencies suffers from strong power attenuation, mainly due to high path loss, blockage effect and channel sparsity [2]. Therefore, high-gain directional beams have to be employed, using large antenna arrays combined with beamforming techniques [3], to counter this challenge.

Such directional communications require a beam training phase to steer the transmitter's beam in the direction of the receiver, before the data transmission. This issue has largely been studied in the literature [4]–[7]. Nevertheless, most of the existing work is either based on classical methods (e.g., exhaustive search, mmWave channel estimation, etc.) or on reinforcement learning, which suffer from large training overhead especially for online and mobile applications with large antenna arrays.

The main goal of this paper is to propose an alternative data-oriented beamforming method based on deep neural networks, which takes advantage of the channel knowledge at sub-6GHz. The idea is to exploit existing and mature sub-6GHz channel estimation technologies (e.g., via pilot symbols) to perform beamforming in the mmWave band. This can be done by *learning* the mapping between the uplink sub-6GHz channel state and the downlink beamforming scheme at mmWave frequencies. Of course, deriving an analytical model of this mapping is out of the question because of its complex dependencies on numerous physical parameters of the wireless environment, such as: users location, reflectors, building materials, etc. Instead, the recent data-driven machine learning tools become indisputable. More specifically, due to their powerful generalization ability (provided appropriate training data) [8], deep learning methods are perfectly suitable to identify this mapping in complex systems (e.g., multi-cell networks), even in a distributed manner as will be shown below.

The recent success of deploying AI-based (artificial intelligence) mMIMO (massive multiple-input multiple-output) base stations [9], which use machine learning to optimize their resources and shape their beams at the sub-6GHz band (e.g., Huawei smart mMIMO [10]), provide further motivation of deep learning methods in this context.

Relevant work: The idea of mapping sub-6GHz channels to mmWave channels was introduced in [11], in which a neural network is employed to map channel states between two close sub-6GHz frequencies: 2.4 GHz and 2.5 GHz. In [12], a different neural network is trained to map sub-6GHz channels to beamforming vectors at mmWave for a single transmitter-receiver link. The proposed approach in [12] is based on a classification problem, in which the beamforming vectors are selected from a predefined and discrete set of vectors (which divide the angular domain into a certain number of directions), yielding a sub-optimal solution. The authors in [13] propose a federated learning (FL) framework to map the mmWave channels into analog beamformers in a multi-user downlink network. However, the proposed learning model requires the mmWave channel matrices as inputs, which are

[1]This is a draft copy of the accepted paper for 2021 IEEE WCNC workshop on Distributed Machine Learning

much more difficult to estimate and require larger training overheads compared to sub-6GHz channels.

Our contributions: In this paper, we propose a federated learning scheme to predict the mmWave beamforming vectors exploiting sub-6GHz channels for a network composed of multiple access point – user links. We design a deep neural network, at each access point, that takes the sub-6GHz channels as input and outputs directly the corresponding mmWave beamforming vector. Furthermore, we propose a distributed federated learning scheme to predict the beamforming vectors locally at each access point, without the need of uploading the local data to a central hub. The latter has three major advantages: the computation load is distributed to the edge of the network as opposed to a centralized cloud-based approach; users only share their neural network parameters and not their data, which highly reduces the signaling load, and protects the data of the individual users, while users still sharing their acquired knowledge of the environment to improve the quality of their predictions.

Compared to [12], our novel contributions are multi-fold. First, we formulate the problem as a regression and not a classification, which means that our predicted beam can point at any direction predicted at the output of the neural network (better angular resolution) and does not suffer from the output discretization sub-optimality. Second, we employ a less-complex network architecture with less parameters compared to [12]. Third, we consider multiple links as opposed to the single link case in [12] and propose a distributed learning approach based on federated learning to improve the quality of the local prediction. The parameters of the local models are averaged to construct a global model for beam prediction for all access points (following the federated learning concept), which protects the local data while pooling on the knowledge acquired by other learning nodes.

Compared to [13], our proposed beamforming scheme relies only on the channel state estimation at sub-6GHz, which is much easier to acquire based on existing technology as opposed to the mmWave channel estimation.

At last, we evaluate our proposed federated learning scheme via numerous experiments – exploiting the *DeepMIMO* [14] available dataset – and compare it to a fully centralized and distributed benchmarks. Our results show the high potential and robustness of the proposed distributed FL-based scheme, especially in the case of scarce training data and low quality uplink channels. When the training data is scarce at each access point, the relative sum rate gain of our federated learning scheme can reach up to 50% compared to a fully distributed one (in which the access points do not cooperate at all). Quite remarkably, our scheme outperforms the fully centralized one when the uplink sub-6GHz SNR is low (less than 5 dB), enjoying a relative sum rate gain of up to 14%.

II. System Model and Problem Formulation

We consider a wireless network composed of multiple access point – user links, for instance multiple short-range access points and their users. It is common to use this kind of model to describe a cellular network of access points serving a single user each [15]. Each access point is equipped with a sub-6GHz receive-array of M antennas (uplink) and a mmWave transmit-array of N antennas (downlink). For simplicity, we assume that each access point serves a unique user equipped with a single mmWave receive antenna and a single sub-6GHz transmit antenna. The communication in each link is performed via multiple carrier frequencies, i.e., orthogonal frequency-division multiplexing (OFDM).

We do not consider explicitly the interference between the multiple links in this work. In the mmWave band (downlink), the interference will not be a major issue because of the highly directive beams. But the multi-user interference will impact the sub-6GHz uplink and, hence, the quality of the estimated channels. Here, even though the interference is not made explicit, we do study the impact of poor quality estimated uplink channels on our proposed method, being relevant when the interference is treated as noise.

To simplify the presentation of the underlying signal model but with no loss of generality, we focus below on a focal link between an access point and its intended user. The access point aims at predicting the downlink mmWave beamforming vector for its user based on the uplink received signal at the sub-6GHz band, which can be written in the ℓth subcarrier with $\ell \in \{1, \ldots, L\}$ as follows:

$$\mathbf{y}^{\mathrm{UL}}[\ell] = \mathbf{h}^{\mathrm{UL}}[\ell]\ x^{\mathrm{UL}}[\ell] + \mathbf{n}^{\mathrm{UL}}[\ell], \tag{1}$$

where $\mathbf{h}^{\mathrm{UL}}[\ell] \in \mathbb{C}^{M\times 1}$ is the sub-6GHz uplink channel; $x^{\mathrm{UL}}[\ell]$, $\mathbb{E}[|\ x^{\mathrm{UL}}[\ell]\ |^2] = P^{\mathrm{UL}}/L$ is the uplink pilot symbol with P^{UL} denoting the uplink transmit power and $\mathbf{n}^{\mathrm{UL}}[\ell]$ is the additive Gaussian noise vector at sub-6GHz.

In the downlink, the analog transceiver is used to transmit data to the user using the mmWave antenna array. The received signal, in the ℓth subcarrier, can be written as:

$$y^{\mathrm{DL}}[\ell] = \mathbf{h}^{\mathrm{DL}^{\dagger}}[\ell]\ \mathbf{f}\ x^{\mathrm{DL}}[\ell] + n^{\mathrm{DL}}[\ell] \tag{2}$$

where $\mathbf{h}^{\mathrm{DL}}[\ell] \in \mathbb{C}^{N\times 1}$ is the mmWave downlink channel; $\mathbf{f} \in \mathbb{C}^{N\times 1}$; $\|\mathbf{f}\|^2 = 1$ is the downlink beamforming vector; $x^{\mathrm{DL}}[\ell]$, $\mathbb{E}[|\ x^{\mathrm{DL}}[\ell]\ |^2] = P^{\mathrm{DL}}/L$ is the transmitted symbol with P^{DL} representing the downlink transmit power; $n^{\mathrm{DL}}[\ell] \sim \mathcal{N}(0, (\sigma^{\mathrm{DL}})^2)$ is the additive Gaussian noise at mmWave.

Main objective: In this paper, we aim to exploit the sub-6GHz uplink signal to predict directly the downlink mmWave beamforming vector at each access point in a distributed manner, by learning the complex and highly non-linear mapping between $\{\mathbf{h}^{\mathrm{UL}}[\ell]\}_{\ell=1}^{L}$ and $\mathbf{f}$. Our two major contributions consist in: (i) leveraging the available uplink channel knowledge at sub-6GHz (easier to obtain and with less overhead compared to the uplink mmWave channel) to predict a suitable mmWave beamforming vector for data transmission in a multiple communicating links setting; and, (ii) predict the beamforming vectors locally in a distributed manner via federated learning.

For this, we employ a supervised learning approach, in which the dataset samples (i.e., the sub-6GHz and mmWave

channels as detailed in the next section) are built with the *Deep MIMO* simulator [14]. The resulting datasets rely on accurate 3D ray-tracing to construct the complex entries of the channels. The sub-6GHz channels capture the propagation physical characteristics of the wireless environment, such as: the geometry (reflection, blockage), the frequency, the surrounding materials, AoA, etc. All these characteristics make the sub-6GHz channels both relevant and essential for the proposed learning model to be able to predict the downlink beams at the mmWave band.

III. Distributed Beam Prediction

In this section, we describe in details our proposed federated learning approach for distributed mmWave beam prediction, in which the prediction of the beamforming vectors is distributed among the different access points. The local learning models, one at each access point, cooperate following the federated learning framework to learn a global mapping from sub-6GHz channels to mmWave beamforming vectors. The federated learning approach preserves the privacy of the local data by exchanging only the local model parameters instead of the actual data (as in fully centralized learning schemes), requires less signaling, and also offers the possibility of parallel computation speeding up the training and splitting the computational load.

Our proposed beam prediction scheme requires an iterative training phase. At each training epoch, the available local datasets are used to train a local neural network to predict the corresponding downlink beamforming vector $\mathbf{f}$ (output of the deep learning model), from the local uplink sub-6GHz channel vectors $\{\mathbf{h}^{\mathrm{UL}}[\ell]\}_{\ell=1}^{L}$. Then, the local model parameters are uploaded to a server for aggregation (a simple average operation). Finally, the global model parameters are downloaded by each access point to train it again on each of their local data during the next training epoch, until the end of the training phase.

After the training phase, the final global model is exploited locally at each access point to predict the mmWave beamforming vectors by feeding it with a locally estimated sub-6GHz uplink channel. Hence, our proposed mmWave beamforming solution frees the system from the large online training overhead related to the mmWave beam training or channel estimation. The different steps of the proposed distributed beam prediction method are illustrated in Fig. 1.

A. Learning dataset

Our learning method is supervised and relies on the available *Deep MIMO* dataset [14] composed of channel pairs $(\{\mathbf{h}[\ell]^{\mathrm{UL}}\}_{\ell=1}^{L}, \{\mathbf{h}[\ell]^{\mathrm{DL}}\}_{\ell=1}^{L})$ generated for different samples of the user geographical positions around the fixed access points, as detailed in Sec. IV.

The uplink channels provide a multipath signature at sub-6GHz, which capture wireless environment knowledge that is invariant with the frequency band (e.g., geometry of the various obstacles and buildings, higher order channel statistics,

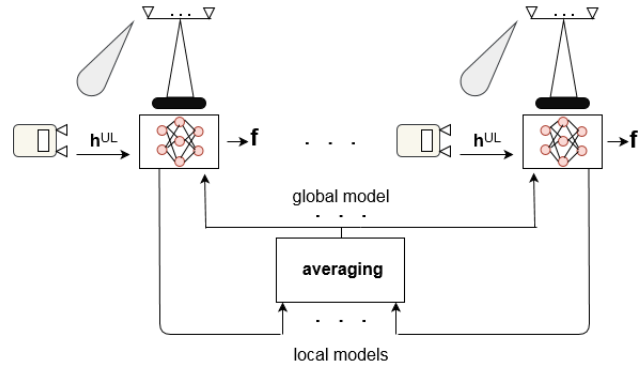

Fig. 1. Federated learning for mmWave beam prediction based on sub-6GHz channels.

etc.), remaining valid at mmWave and impacting the beamforming operation. Deep learning methods are suitable tools to learn the complex relationship between sub-6GHz uplink channels and mmWave downlink beamforming vectors. Moreover, estimating the sub-6GHz uplink channels requires less training overhead and involves an already acquired technology compared to mmWave channels.

Each complex entry of the channel vectors $\mathbf{h}^{\mathrm{UL}}[\ell]$ is decomposed into real and imaginary parts, which are stacked into a $2LM$ real valued vector containing the uplink channel information of all L subcarriers corresponding to the first component of an element in the dataset. This operation is necessary since the employed neural network does not support operations on complex inputs. The input data is also normalized before feeding it to the neural network. Similarly, the mmWave channels $\mathbf{h}^{\mathrm{DL}}[\ell]$ are partitioned into real and imaginary parts to form one real valued vector of dimension $2LN$ as the second component of an element in the dataset. The dataset is divided into a training dataset (80% of the total size) and a test set (the remaining 20%). The training dataset is also split into a training set (85% of its size) and a validation set (the remaining 15% of the training dataset).

B. Network architecture

The architecture of a local neural network is represented in Fig. 2. The uplink sub-6GHz channel vector of dimension $2LM$ represents the input of a local deep neural network. Each local network comprises 4 hidden fully-connected layers of $1024, 2048, 2048, 1024$ neurons respectively [2] with rectified linear unit (ReLu) as an activation function. Every layer employs an $\mathbf{L}2$-norm regularization with weight decay of 10^{-7}. The output layer is a fully-connected one of size $2N$, followed by a normalization layer. It provides directly the real and imaginary parts of the mmWave beamforming vector $\mathbf{f}$, which has a unit norm ($\|\mathbf{f}\|^2 = 1$), due to hardware constraints.

We choose fully-connected neural networks (where each neuron of one layer is connected to all the neurons of the

[2]The number of layers and neurons is chosen as a result of empirical trials. We choose the architecture that best tradeoffs the training and validation losses.

preceding and following layers) because they are *structure agnostic*, making no particular assumptions about the inputs and having a general purpose. Furthermore, such a network guarantees the flow of information between the inputs and outputs of each layer, which makes it able to capture any kind of dependencies between the layers: the final output can be truly based on the whole information captured by the input. This makes fully-connected networks suitable for our problem, since we do not have specific knowledge about the complex relationship between sub-6GHz channels and mmWave beamforming vectors, coupled with the numerous wireless parameters impacting it.

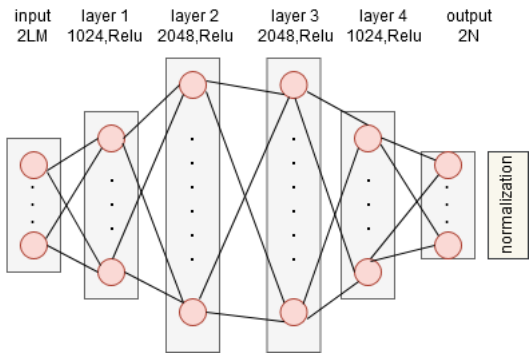

Fig. 2. Architecture diagram of the fully-connected neural network employed at each access point.

C. Loss function

Our objective is to predict beamforming vectors for an OFDM-based system of L subcarriers that maximize the data rate. For this, we define the following loss function $\mathcal{L}$, which will be minimized during the training phase:

$$\mathcal{L} = -\frac{1}{\mathcal{B}} \sum_{i=1}^{\mathcal{B}} \mathcal{R}_i, \tag{3}$$

where $\mathcal{B}$ is the size of the mini-batch and $\mathcal{R}_i$ is the average data rate over the L subcarriers for the ith sample: $(\{\mathbf{h}_i[\ell]^{\text{UL}}\}_{\ell=1}^{L}, \{\mathbf{h}_i[\ell]^{\text{DL}}\}_{\ell=1}^{L})$ of the mini-batch, and which can be written as

$$\mathcal{R}_i = \frac{1}{L} \sum_{\ell=1}^{L} \log_2 \left(1 + \frac{P^{\text{DL}}}{L(\sigma^{\text{DL}})^2} \mid \mathbf{h}_i^{\text{DL}\dagger}[\ell] \ \mathbf{f}_i \mid^2 \right), \tag{4}$$

with $\mathbf{f}_i$ denoting a normalized beamforming vector predicted by the local neural network for the ith sample (the output elements of the neural network are re-shaped into an N-dimension complex vector).

Notice that minimizing this loss function is equivalent to maximizing the achievable rate. In other words, the local neural network is trained, by minimizing the loss function using its local dataset, to predict directly downlink beamforming vectors (from uplink sub-6GHz channels) which maximize the average sum rate over the L subcarriers.

The above loss function represents a notable difference compared to existing works, in which first a loss function based on some average prediction error between the predicted vectors and the ground truth is minimized, and then the performance of the prediction is evaluated in terms of its communication performance. Here, we have chosen to train our neural network to maximize directly the communication data rate and skip the intermediary step. The first obvious advantage is that, by choosing a well-suited communication loss as opposed to a generic data-driven one, the communication performance of our resulting method can only be improved. Second, in our problem, such a data-oriented prediction error cannot be computed given that the available *Deep MIMO* dataset is only composed of channel pairs $(\{\mathbf{h}[\ell]^{\text{UL}}\}_{\ell=1}^{L}, \{\mathbf{h}[\ell]^{\text{DL}}\}_{\ell=1}^{L})$ and does not contain the corresponding optimal beamforming vectors $\mathbf{f}$. At last, creating a dataset composed of pairs of the type $(\{\mathbf{h}[\ell]^{\text{UL}}\}_{\ell=1}^{L}, \mathbf{f})$ [12] is quite problematic since there are an infinite number of optimal vectors $\mathbf{f}$ maximizing the data rate (since it is invariant to a multiplication of $\mathbf{f}$ by a complex value of unit-norm). In such a case, an arbitrary selection would have to be made, which might hinder the generalization capability of the neural network.

IV. Numerical results

We evaluate here the performance of our proposed distributed mmWave beamforming based on federated learning in terms of the sum rate, which is computed as the average of the rates of all links. We compare our proposed method to the following benchmarks:

- **centralized learning (CL)**, performed by a central authority using a similar neural network and having access to the datasets of all access points;
- **perfect downlink CSI** (channel state information), an ideal scenario in which the mmWave channels are assumed known and used to construct the beamforming vectors for each subcarrier such that $\mathbf{f}^{\star}[\ell] = \dfrac{\mathbf{h}^{\text{DL}}[\ell]}{||\mathbf{h}^{\text{DL}}[\ell]||}$ to maximize the received power at the receiver;
- **individual learning** or the fully distributed scheme, where each access point trains its neural network using its local dataset independently without with communicating the others.

The presented results are obtained after 100 training epochs using the adaptive moment estimation (ADAM) optimizer [16] with a learning rate of 10^{-4} and a batch size of $\mathcal{B} = 256$ samples. For our FL system, 100 epochs of training represent 100 epochs at each access point. The different learning models are implemented and trained using *TensorFlow*.

The local dataset for each access point is built using *DeepMIMO* [14], which employs the accurate 3D ray-tracing simulator *Wireless Insite* [17] to generate the uplink and downlink channels. In our simulations, we use the outdoor ray-tracing scenario 'O1' [14], which is available at 3.5 GHz and 28 GHz frequencies. In the outdoor setting, the area size is 600m $\times$ 400m, in which 4 fixed access points (BS1, BS4, BS6 and BS7 in [14]) are deployed with $P^{\text{DL}} = 34$ dBm, $M = 4$, $N = 64$ and $L = 32$. The access points communicate with

their corresponding users, whose position inside of the area is sampled every 20 cm in the specified 2D grid (we consider rows $1-599$, $600-1200$, $1201-1550$ and $1551-2200$ for each link resp. as in [14]). The bandwidth equals 0.5 GHz for the mmWave downlink and 0.02 GHz for the sub-6GHz uplink. The number of channel paths in the simulator is set to 5 for the mmWave channel and 15 for the sub-6GHz one. Both transceivers employ antenna arrays with $\lambda/2$ spacing between their elements. The noise spectral density is -174 dBm/Hz. The total number of samples in each of the four local datasets are: 108 419, 108 781, 63 350 and 117 650.

A. Training performance

In Fig. 3, we evaluate the evolution of the sum rate, computed on the training and validation sets, as a function of the learning epochs for both CL and our method denoted by FL (from federated learning). First, notice that the values of the sum rate and, hence, also the values of the loss function, on the training and validation sets are close for CL and FL. This implies that the proposed neural network architecture converges fast and fits well the data without underfitting or overfitting. Moreover, the achievable sum rate by the FL scheme approaches the centralized one, without the need to share the local data with the central entity.

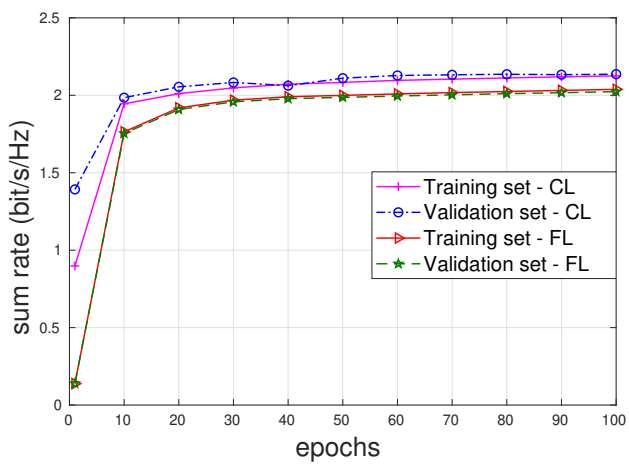

Fig. 3. Sum rate as a function of the training epochs, evaluated on the training and the validation sets. Our neural network is capable of learning without underfitting or overfitting.

B. Prediction (generalization) performance

In Fig. 4, we plot the empirical cumulative distribution function (CDF) [3] of the average rate over the test set for the CL and FL methods. Each scheme is used to predict the downlink mmWave beamforming vectors from the uplink channels of the test set (these samples have not been seen by our neural networks during the training phase) and then evaluate the achievable rate. We also compare the results with the perfect downlink CSI case on the same test sets.

[3]The CDF represents the cumulative empirical frequency of a given rate value obtained by the predicted beamforming vectors over the test set, useful to illustrate the distribution of the different achievable rates

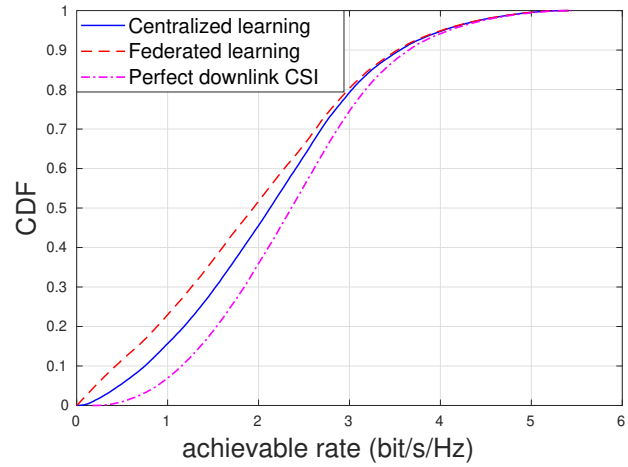

Fig. 4. Empirical CDF of the achievable rate over the test set. The CDF performance of our FL method is close to the CL one.

We can see that our FL scheme performs close to the CL one and is not far from the ideal case, illustrating the generalization capability of our method. This implies that, our proposed FL beam prediction scheme can achieve almost the same performance as CL, while requiring lower signaling, preserving the privacy of the local data, distributing the computation load to the edge of the network, etc.

C. Impact of the training set size

Fig. 5 shows the sum rate over the test set as a function of the training set size. The different sizes are relative to the total training set size of each access point. On the one hand, we can see that increasing the training set size yield better predictions of the mmWave beamforming vectors for all schemes. On the other hand, Fig. 5 shows the effectiveness of our FL distributed scheme for small sizes of local datasets. Indeed, sharing the local models to build a global one for all access points helps to overcome the scarcity of local data and offers higher sum rates compared to the individual learning scheme. For the minimum training set size, the relative gain in terms of sum rate is 50% compared to the individual scheme. Of course, as the local training sets sizes increase, the utility of exchanging the local models decreases.

D. Impact of the uplink sub-6GHz channel quality

In Fig. 6, we plot the performance in terms of the sum rate over the test set, when the uplink channels are contaminated with different levels of noise, as function of the uplink sub-6GHz SNR. We can see that our federated scheme outperforms CL and individual learning at low SNR (less than 5 dB). For an SNR equal to -10 dB, the relative gain is of 14% compared to CL. This robustness can be explained by the averaging step of the FL scheme, which acts as a regularization and noise smoothing operation, thus improving its generalization performance in the high noise regime.

E. Regression gain in the single link case

We compare here our regression-based scheme with the classification one proposed in [12] for the *single link* case.

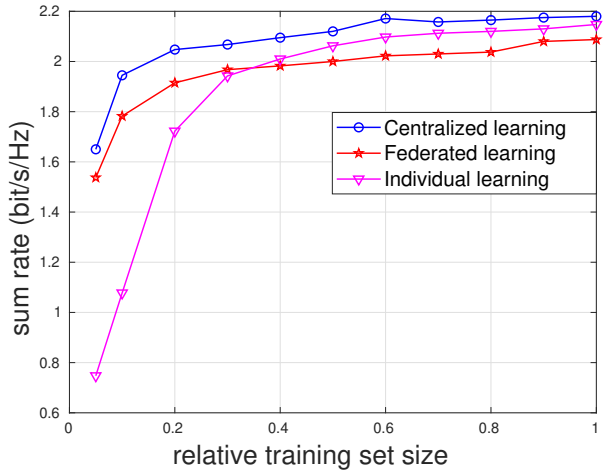

Fig. 5. Impact of the training set size on the prediction performance. When the available training data is scarce, our FL scheme outperforms the individual learning and approaches CL.

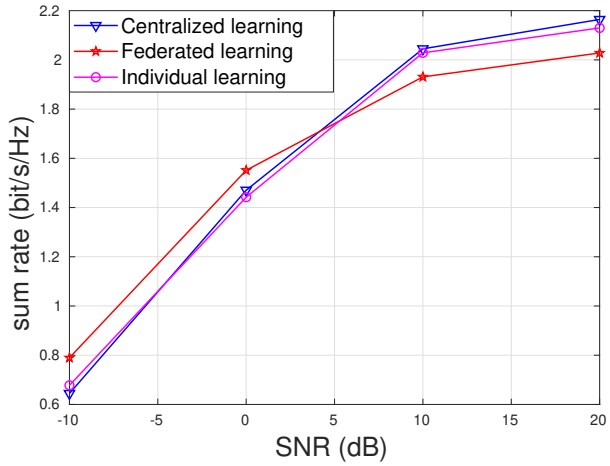

Fig. 6. Impact of the sub 6-GHz channels quality on the prediction performance over the test set. High noise levels lead to a degradation of the sum rate, but our FL scheme is more robust than CL and individual learning.

In [12], the mmWave beamforming vector $\mathbf{f}_k, k \in \{1, .., 64\}$ is chosen from a predefined discrete codebook according to the predicted index k at the output of their proposed deep neural network. Here, for comparison reasons, we evaluate our local neural network in the same setting and with the same simulation parameters as in [12]. The average rate over the test set, obtained by the predicted beams of our approach is higher than the one obtained by the method in [12] and closer to the optimal case of perfect downlink CSI as illustrated in table I. This adds to the fact that our proposed neural network employs less neurons per hidden layers (6144 neurons) compared to [12] (10240 neurons).

	Our method	Method in [12]	Perfect CSI
Average rate (bit/s/Hz)	2.21	1.86	2.39

TABLE I
AVERAGE RATE OVER THE TEST SET IN THE SINGLE LINK CASE.

V. CONCLUSIONS AND PERSPECTIVES

In this paper, we have proposed a federated deep learning scheme for mapping the sub-6GHz uplink channels to downlink mmWave beamforming vectors. Our distributed scheme is evaluated in a network composed of multiple access point – user links. The federated learning scheme provides a global learning model for all access points without exchanging their local data, but only their local models, which reduces the data-exchange overhead and distributes the computational load. The performance of our federated learning scheme approaches the centralized one in terms of sum rate, and even outperforms it when the available training data is of low quality (high noise levels). When the training data is scarce, our scheme provides relative gains of up to 50% compared to a fully distributed one. Future work may include the impact of inter-link interference, multiple users per access point, asynchronous federated learning, etc., all relevant issues in highly dense networks.

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, 2013.

[2] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of millimeter wave communications for fifth-generation 5G wireless networks with a focus on propagation models," *IEEE Trans. Antennas Propag*, vol. 65, no. 12, pp. 6213–6230, 2017.

[3] X. Wang, L. Kong, F. Kong, F. Qiu, M. Xia, S. Arnon, and G. Chen, "Millimeter wave communication: A comprehensive survey," *IEEE Commun. Surv. Tutor*, vol. 20, no. 3, pp. 1616–1653, 2018.

[4] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun*, vol. 32, no. 6, pp. 1164–1179, 2014.

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

[6] D. E. Berraki, S. M. Armour, and A. R. Nix, "Application of compressive sensing in sparse spatial channel recovery for beamforming in mmWave outdoor systems," in *IEEE WCNC*, 2014, pp. 887–892.

[7] I. Chafaa, E. V. Belmega, and M. Debbah, "One-bit feedback exponential learning for beam alignment in mobile mmwave," *IEEE Access*, vol. 8, pp. 194 575–194 589, 2020.

[8] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?" *IEEE Trans. Commun*, vol. 67, no. 10, pp. 7331–7376, 2019.

[9] "AI-based massive MIMO," www.analyticsindiamag.com/ai-behind-huawei-mass-mimo-optimisation-5g-india.

[10] "Huawei smart mMIMO," www.huawei.com/en/news/2019/2/huawei-fdd-massive-mimo-solutions.

[11] M. Alrabeiah and A. Alkhateeb, "Deep learning for TDD and FDD massive MIMO: Mapping channels in space and frequency," in *53rd IEEE ACSSC*, 2019, pp. 1465–1470.

[12] ——, "Deep learning for mmwave beam and blockage prediction using sub-6 GHz channels," *IEEE Trans. Commun*, vol. 68, no. 9, pp. 5504–5518, 2020.

[13] A. M. Elbir and S. Coleri, "Federated learning for hybrid beamforming in mm-wave massive mimo," *IEEE Communications Letters*, vol. 24, no. 12, pp. 2795–2799, 2020.

[14] A. Alkhateeb, "Deepmimo: A generic deep learning dataset for millimeter wave and massive mimo applications," *arXiv preprint arXiv:1902.06435*, 2019.

[15] K. Shen and W. Yu, "Fractional programming for communication systems — part I: Power control and beamforming," *IEEE Trans. Signal Process*, vol. 66, no. 10, pp. 2616–2630, 2018.

[16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[17] Remcom, "Wireless Insite," http://www.remcom.com/wireless-insite.