MANAGING HARDWARE IMPAIRMENTS IN HYBRID MILLIMETER WAVE MIMO SYSTEMS: A DICTIONARY LEARNING-BASED APPROACH

Joan Palacios[†], *Nuria González-Prelcic*[‡] *and Joerg Widmer*[†]

[†] IMDEA Networks, Spain
[‡]The University of Texas at Austin, USA ,and the University of Vigo, Spain

ABSTRACT

Compressed sensing-based strategies have been derived in prior work to reduce training overhead when estimating the high dimensional millimeter wave MIMO channel. These techniques rely on a channel model based on a sparsifying dictionary which does not account for hardware impairments such as calibration errors, mutual coupling effects, or manufacturing errors in the inter-spacing between the array elements. In this paper, we propose a learning strategy for the sparsifying dictionary that considers a channel model with hardware impairments, embedding these effects into the dictionary itself. This way, a sparser representation of the channel can be obtained even when considering realistic implementations of the antenna array and the radio frequency chains. Numerical simulations with different system configurations and parameters of the hardware impairments, show the effectiveness of the proposed dictionary learning algorithm for channel estimation at millimeter wave frequencies with hybrid MIMO architectures.

I. INTRODUCTION

Compressive channel estimation strategies exploiting channel sparsity at millmeter wave (mmWave) frequencies have been proposed in prior work [1]–[4] to reduce training overhead when operating with large arrays and hybrid architectures. These works assume sparsifying dictionaries for the channel matrices built from the steering vectors for the transmitter and receiver evaluated on a grid of possible angles of departure and arrival. This model only works well when ideal hardware is assumed. When considering hardware imperfections such as calibration errors, disturbances in the spaces between antenna elements or mutual coupling, the ideal dictionaries cannot provide sparse channel representations in the virtual domain. This motivates the need of dictionary learning (DL) strategies to estimate the sparsifying dictionaries when hardware imperfections are considered.

Channel estimation based on DL was first proposed in [5] and [6] for a sub-6 GHz massive MIMO system with single antenna users. A DL strategy for millimeter wave MIMO systems based on hybrid architectures with hardware impairments was derived in [7]. The proposed method outperforms channel estimation strategies based on ideal dictionaries, but the performance gap in spectral efficiency with respect to assuming perfect knowledge of the actual dictionaries is still significant.

In this paper, we propose a new DL method that, instead of alternatively iterating over the channel and dictionary estimates to achieve convergence as in [7], obtains a least squares (LS) channel estimate. This estimate is used to extract candidate words for the transmit and receive dictionaries embedding hardware imperfections. The candidate words are then used to build the dictionaries using a k-medoid clustering method. Due to the high quality of the channel estimate used to generate the candidate words and the effective screening and clustering process used to build the dictionaries, the new DL method leads to channel estimates that provide near optimal spectral efficiency when evaluated under a realistic mmWave MIMO-OFDM link.

II. SYSTEM MODEL

We consider a MIMO-OFDM mmWave system based on a hybrid architecture at both sides of the communication link. The transmitter is equipped with N_t antennas and L_t RF chains, while the receiver operates with N_r antennas and L_r RF chains. The number of data streams to be transmitted simultaneously employing N_c subcarriers is N_s $(\leq \min\{L_\text{t}, L_\text{r}\})$. The channel is modeled as frequency selective, and the hybrid precoder and combiner are denoted as $\mathbf{F}[k]$ and $\mathbf{W}[k]$. We assume that during the channel estimation stage the system employs a predefined sequence of training precoders and combiners known at both ends.

The frequency selective mmWave channel is characterized with a clustered channel model with N_p clusters, N_ray rays, and a delay tap length N_tap. For the d-th delay tap, the channel matrix is denoted as $\boldsymbol{H}_d \in \mathbb{C}^{N_\text{r}\times N_\text{t}}$, $d \in \mathcal{I}(N_\text{tap})$. This matrix includes models for hardware imperfections as follows. Let $\theta_{l,k}$ and $\phi_{l,k}$ denote the angles of departure (AoD) and arrival (AoA) for the kth-ray in the lth-cluster. The steering vectors for both the transmitter and receiver, denoted as $\tilde{\mathbf{a}}_\text{T}(\theta_{l,k})$ and $\tilde{\mathbf{a}}_\text{R}(\phi_{l,k})$, consider inter-antenna spacing errors denoted as $\epsilon_{\text{t},1}, \ldots, \epsilon_{\text{t},N_\text{t}-1}$ and $\epsilon_{\text{r},1}, \ldots, \epsilon_{\text{r},N_\text{r}-1}$. Thus, assuming for example a uniform linear array (ULA), the array steering vector at the receiver can be written as $\tilde{\mathbf{a}}_{\mathrm{R}}(\phi_{l,k}) = \frac{1}{\sqrt{N_{\mathrm{r}}}}[1, e^{-2\pi\frac{d+\epsilon_{\mathrm{r},1}}{\lambda}\sin(\phi_{l,k})}, \ldots, e^{-2\pi\frac{(N_{\mathrm{r}}-1)d+\epsilon_{\mathrm{r},1}}{\lambda}\sin(\phi_{l,k})i}]^{\mathrm{T}}$; a similar expression can be derived for the transmit array steering vectors. Matrix $\mathbf{\Gamma}_{\mathrm{R}} = diag\{g_{\mathrm{r},1}e^{\nu_{\mathrm{r},1}i}, \ldots, g_{\mathrm{r},N_{\mathrm{r}}}e^{\nu_{\mathrm{r},N_{\mathrm{r}}}i}\}$ represents the antenna gain and phase error matrices at the receiver, while $\mathbf{\Gamma}_{\mathrm{T}}$ includes this effect at the transmit side. $\mathbf{C}_R \in \mathbb{C}^{N_{\mathrm{r}}\times N_{\mathrm{r}}}$ and $\mathbf{C}_T \in \mathbb{C}^{N_{\mathrm{t}}\times N_{\mathrm{t}}}$ are the antenna coupling matrices at the transmit and receive side. Taking into account all these impairments, the channel matrix for the d-th delay tap can be written as

$$\boldsymbol{H}_d = \sqrt{\frac{N_{\mathrm{t}}N_{\mathrm{r}}}{N_{\mathrm{p}}N_{\mathrm{ray}}}} \sum_{\ell=1}^{N_{\mathrm{p}}} \sum_{k=1}^{N_{\mathrm{ray}}} \alpha_{\ell,k} p_{\mathrm{rc}}(dT_{\mathrm{s}} - \tau_\ell) \mathbf{C}_{\mathrm{R}}\mathbf{\Gamma}_{\mathrm{R}}\tilde{\mathbf{a}}_{\mathrm{R}}(\phi_{\ell,k})(\mathbf{C}_{\mathrm{T}}\mathbf{\Gamma}_{\mathrm{T}}\tilde{\mathbf{a}}_{\mathrm{T}}(\theta_{\ell,k}))^*, \tag{1}$$

Defining $\tilde{\mathbf{A}}_{\mathrm{R}} \triangleq [\tilde{\mathbf{a}}_{\mathrm{R}}(\phi_{1,1}), \ldots, \tilde{\mathbf{a}}_{\mathrm{R}}(\phi_{N_{\mathrm{p}},N_{\mathrm{ray}}})]$, $\tilde{\mathbf{A}}_{\mathrm{T}} \triangleq [\tilde{\mathbf{a}}_{\mathrm{T}}(\theta_{1,1}), \ldots, \tilde{\mathbf{a}}_{\mathrm{T}}(\theta_{N_{\mathrm{p}},N_{\mathrm{ray}}})]$, and $\mathbf{\Delta}_d \triangleq \sqrt{\frac{N_{\mathrm{t}}N_{\mathrm{r}}}{N_{\mathrm{p}}N_{\mathrm{ray}}}}\,\mathrm{diag}\{\alpha_{1,1}p_{\mathrm{rc}}(dT_{\mathrm{s}}-\tau_1), \ldots, \alpha_{N_{\mathrm{p}},N_{\mathrm{ray}}}p_{\mathrm{rc}}(dT_{\mathrm{s}}-\tau_{N_{\mathrm{p}}})\}$ in $\mathbb{C}^{N_{\mathrm{p}}N_{\mathrm{ray}}\times N_{\mathrm{p}}N_{\mathrm{ray}}}$ as the diagonal matrix that contains the channel coeffients, the channel matrix $\boldsymbol{H}_d$ can also be expressed in a more compact way as

$$\boldsymbol{H}_d = \mathbf{C}_{\mathrm{R}}\mathbf{\Gamma}_{\mathrm{R}}\tilde{\mathbf{A}}_{\mathrm{R}}\mathbf{\Delta}_d\tilde{\mathbf{A}}_{\mathrm{T}}^*\mathbf{\Gamma}_{\mathrm{T}}^*\mathbf{C}_{\mathrm{T}}^*. \tag{2}$$

In the frequency domain, the channel matrix for every subcarrier is

$$\mathbf{H}[k] = \mathbf{C}_{\mathrm{R}}\mathbf{\Gamma}_{\mathrm{R}}\tilde{\mathbf{A}}_{\mathrm{R}}\underbrace{\left(\sum_{d=0}^{N_{\mathrm{tap}}-1}\mathbf{\Delta}_d e^{-\mathrm{j}\frac{2\pi kd}{N_{\mathrm{c}}}}\right)}_{\mathbf{\Delta}[k]}\tilde{\mathbf{A}}_{\mathrm{T}}^*\mathbf{\Gamma}_{\mathrm{T}}^*\mathbf{C}_{\mathrm{T}}^*.$$

Considering the expression above, it is clear that new sparsifying dictionaries going beyond a set of steering vectors evaluated on a grid of AoA/AoD are needed. With the new generalized dictionaries accounting for hardware impairments, the channel matrix in the frequency domain can be approximated as

$$\mathbf{H}[k] \approx \mathbf{D}_{\mathrm{R}}\mathbf{\Omega}[k]\mathbf{D}_{\mathrm{T}}^*, \tag{3}$$

where $\mathbf{\Omega}[k]$ is a $G_{\mathrm{r}}\times G_{\mathrm{t}}$ sparse matrix containing the channel gains in a new representation of the channel that embeds the impairments into the generalized sparsifying dictionaries $\mathbf{D}_{\mathrm{T}}$ and $\mathbf{D}_{\mathrm{R}}$.

III. DICTIONARY LEARNING WITH HARDWARE IMPAIRMENTS

Our goal is to derive a strategy to obtain the matrices $\mathbf{D}_{\mathrm{T}}$ and $\mathbf{D}_{\mathrm{R}}$ in (3) from a set of measurements of the received signal, by using a dictionary learning approach well suited to this particular application. The proposed approach can be decomposed into three stages as follows. First, we design the channel measurements and the channel estimation strategy from these measurements. Second, the channel estimates are turned into word candidates for the generalized sparsifying dictionaries. Finally, the last stage screens and clusters the word candidates into a dictionary. Notice that dictionary learning process is performed off-line, so the computational complexity or the number of required measurements is not an issue. The following sections describe the details of these three stages.

III-A. Channel measurements and estimation

Since the impairments are unknown, the system does not know the real radiation properties of the transmit or receive antennas. Therefore, compressive channel estimation strategies based on dictionaries built from oversampled steering vectors [3], [4] cannot be used. Compressive strategies were proposed in the previous literature to reduce the number of required channel measurements. However, we propose a dictionary learning process to be performed off-line, so alternative approaches that need from a high number of measurements when antenna arrays are large are also possible. Thus, we propose a channel estimation strategy based on minimizing the minimum squared error in the reconstructed channel to obtain the initial channel estimates. We choose frequency flat precoders and combiners during training, $\mathbf{F}$ and $\mathbf{W}$, as a concatenation of Hadamard matrices, to minimize the number of measurements while keeping the rank completeness and adding orthogonality. Concatenation acts as spreading to increase the effective SNR.

The channel estimate can then be computed as the solution to

$$\arg\min_{\mathbf{H}[k]} \|\mathbf{Y}[k] - \mathbf{W}^H\mathbf{H}[k]\mathbf{F}\|_{\mathrm{F}}^2, \tag{4}$$

where $\mathbf{Y}[k]$ contains the received signal column vectors for subcarrier k for the different training symbols. We can easily prove that the solution to this expression is

$$\hat{\mathbf{H}}[k] = (\mathbf{W}\mathbf{W}^H)^{-1}\mathbf{W}\mathbf{Y}[k]\mathbf{F}^H(\mathbf{F}\mathbf{F}^H)^{-1}. \tag{5}$$

III-B. Word extraction

The candidate words to build the dictionary are extracted from the channel estimate. With this aim, we arrange first the channel matrix per subcarrier into a single matrix that we can easily manipulate. Then, we extract the distinguishable paths and decompose them into steering vectors. The final step is a screening process to select candidate words using a given quality metric. The next paragraphs provide the details of this process.

Our formulation is based on the idea of $\mathbf{H}[k]$ being decomposable into paths in three dimensions, that is

$$[\mathbf{H}[k]]_{a,b} = \sum_{l=1}^{L}[\mathbf{a}_{\mathrm{Rx}}(\theta_l)]_a[\mathbf{a}_{\mathrm{Tx}}(\phi_l)]_b^*[\mathbf{r}_l]_k,$$

where $\mathbf{a}_{\mathrm{Rx}}(\theta)$, $\mathbf{a}_{\mathrm{Tx}}(\phi)$ are the steering vectors for reception and transmission, respectively, and $\mathbf{r}_l$ is the channel frequency response for path l. Paths are distinguishable in reception, transmission or frequency response whenever $\mathbf{a}_{\text{Rx}}(\theta_l)^H \mathbf{a}_{\text{Rx}}(\theta_{l'})$, $\mathbf{a}_{\text{Tx}}(\phi_l)^H \mathbf{a}_{\text{Tx}}(\phi_{l'})$ or $\mathbf{r}_l^H \mathbf{r}_{l'}$ are close to 0.

To include all the information about the paths into a single matrix we define the joint steering vector

$$\mathbf{a}_{\text{Rx-Tx}}(\theta, \phi) = \mathcal{V}(\mathbf{a}_{\text{Rx}}(\theta)\mathbf{a}_{\text{Tx}}(\phi)^*), \tag{6}$$

so that the vectorized channel matrix per subcarrier can be easily written as

$$\mathbf{h}[k] = \mathcal{V}(\mathbf{H}[k]) = \sum_{l=1}^{L} \mathbf{a}_{\text{Rx-Tx}}(\theta_l, \phi_l)[\mathbf{r}_l]_k.$$

From the definition of $\mathbf{h}[k]$, we can write the channel in the frequency domain, for all the subcarriers, as

$$\begin{aligned}\mathbf{H}_{\text{F}} &= [\mathbf{h}[1], \mathbf{h}[2], \ldots, \mathbf{h}[K]] \\ &= [\mathbf{a}_{\text{Rx-Tx}}(\theta_1, \phi_1), \mathbf{a}_{\text{Rx-Tx}}(\theta_2, \phi_2), \ldots][\mathbf{r}_1, \mathbf{r}_2, \ldots]^T.\end{aligned}$$

Now, from the expression of the joint steering vector in (6), we can define the concept of paths distinguishable in space when $\mathbf{a}_{\text{Rx-Tx}}(\theta_l, \phi_l)^H \mathbf{a}_{\text{Rx-Tx}}(\theta_{l'}, \phi_{l'}) = (\mathbf{a}_{\text{Rx}}(\theta_l)^H \mathbf{a}_{\text{Rx}}(\theta_{l'}))(\mathbf{a}_{\text{Tx}}(\phi_l)^H \mathbf{a}_{\text{Tx}}(\phi_{l'}))^*$ is close to 0. This means that a path is distinguishable in space whenever it is in reception or transmission. Under the condition of distinguishable paths in space and frequency response, we have that the matrices $[\mathbf{a}_{\text{Rx-Tx}}(\theta_1, \phi_1), \mathbf{a}_{\text{Rx-Tx}}(\theta_2, \phi_2), \ldots, \mathbf{a}_{\text{Rx-Tx}}(\theta_L, \phi_L)]$ and $[\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_L][$ defined by these paths are going to be close to a scaled version of the left and right single vectors of the SVD decomposition of matrix $\mathbf{H}_{\text{F}} = \mathbf{U}_{\text{F}}\mathbf{S}_{\text{F}}\mathbf{V}_{\text{F}}$. In other words

$$\mathbf{a}_{\text{Rx-Tx}}(\theta_l, \phi_l) \simeq [\mathbf{U}_{\text{F}}]_{:,l'}[\mathbf{U}_{\text{F}}]_{:,l'}^H \mathbf{a}_{\text{Rx-Tx}}(\theta_l, \phi_l), \tag{7}$$

or equivalently,

$$\frac{\mathbf{a}_{\text{Rx-Tx}}(\theta_l, \phi_l)}{\|\mathbf{a}_{\text{Rx-Tx}}(\theta_l, \phi_l)\|} \simeq [\mathbf{U}_{\text{F}}]_{:,l'}\alpha_l \tag{8}$$

for an unknown $|\alpha_l| = 1$.

Notice that $[\mathbf{U}_{\text{F}}]_{:,l'}$ has all the spatial information about path l. To decompose this joint vector into reception and transmission steering vectors, we have to rearrange the vector back into a matrix $\mathbf{H}_{\text{W}}^{l'}$, such that $[\mathbf{U}_{\text{F}}]_{:,l'} = \mathcal{V}(\mathbf{H}_{\text{W}}^{l'})$. Plugging (6) into (8) we get

$$\frac{1}{\|\mathbf{a}_{\text{Rx}}(\theta_l)\|\|\mathbf{a}_{\text{Tx}}(\phi_l)\|}\mathbf{a}_{\text{Rx}}(\theta_l)\mathbf{a}_{\text{Tx}}^H(\phi_l) \simeq \mathbf{H}_{\text{W}}^{l'}\alpha_l. \tag{9}$$

This means that $\mathbf{H}_{\text{W}}^{l'}$ is close to a rank one matrix, so it is possible to obtain the singular value decomposition $\mathbf{H}_{\text{W}}^{l'} = \mathbf{U}_{\text{W}}^{l'}\mathbf{S}_{\text{W}}^{l'}(\mathbf{V}_{\text{W}}^{l'})^H$, with $\mathbf{S}_{\text{W}}^{l'}$ having its diagonal values sorted in decreasing order. This way, the steering vectors can be obtained as the strongest singular vectors

$$\begin{aligned}\frac{\mathbf{a}_{\text{Rx}}(\theta_l)}{\|\mathbf{a}_{\text{Rx}}(\theta_l)\|} &\simeq [\mathbf{U}_{\text{W}}^{l'}]_{:,1}\gamma_l, \\ \frac{\mathbf{a}_{\text{Tx}}(\phi_l)}{\|\mathbf{a}_{\text{Tx}}(\phi_l)\|} &\simeq [\mathbf{V}_{\text{W}}^{l'}]_{:,1}\rho_l.\end{aligned} \tag{10}$$

These steering vectors constitute the initial candidate words that have to be screened using a given quality metric. In this case, it is expected that $\mathbf{H}_{\text{W}}^{l'}$ to be close to a rank one matrix. A good indicator of this is the value $[\mathbf{S}_{\text{W}}^{l'}]_{1,1} \in [\sqrt{\frac{1}{\min(N_r,N_t)}}, 1]$, where $[\mathbf{S}_{\text{W}}^{l'}]_{1,1} = 1$ means that $\mathbf{H}_{\text{W}}^{l'}$ is rank one and $[\mathbf{S}_{\text{W}}^{l'}]_{1,1} = \sqrt{\frac{1}{\min(N_r,N_t)}}$ means that $\mathbf{H}_{\text{W}}^{l'}$ is proportional to an orthonormal matrix. The described process will only work for the strongest distinguishable in space-frequency response paths, as they dominate the channel structure while having no interaction among them. Therefore, our screening strategy will be to consider the strongest paths such that they are close enough to a rank one matrix by a threshold value $[\mathbf{S}_{\text{W}}^{l'}]_{1,1} > \tau_{\text{P}}$. This word extraction process and screening process is summarized in algorithm 1.

Algorithm 1 Word extraction from channel estimate.

```
INPUT: Ĥ, P_Rx, P_Tx
Rearrange H into H_F
Decompose H_F = U_F S_F V_F
for l' ≤ min(N_r, N_t) do
    if [S_F]_{l',l'} > τ_P then
        Rearrange [U_F]_{:,l'} into H_W^{l'}
        Decompose H_W^{l'} = U_W^{l'} S_W^{l'} (V_W^{l'})^H
        if [S_W^{l'}]_{1,1} > τ_P then
            Add [U_W^{l'}]_{:,1} to P_Rx
            Add [V_W^{l'}]_{:,1} to P_Tx
        else
            Break for loop
        end if
    end if
end for
```

III-C. Dictionary creation

At this point, the candidate steering vectors for both receiver ($\mathcal{P}_{\text{Rx}}$) and transmitter ($\mathcal{P}_{\text{Tx}}$) have been extracted. The dictionaries for transmission and reception will be created from these candidate words with a similar procedure, so we drop the Tx or Rx subindex from now on in our notation. The dictionary $\mathcal{D}$ must be a representative set of the set of words $\mathcal{P}$ that were extracted, without falling into being too large. Therefore, we constrain the dictionary size to a given number of words N_{W}. To decide which words will be included in $\mathcal{D}$ and how representative they are in $\mathcal{P}$, we need to define a distance between words. Note that the extracted words are subject to a phase uncertainty due to the extraction process, so that $\mathbf{w} \in \mathcal{P}$ is representing the set $\mathbf{w} * \alpha$, for $\alpha \in \mathbb{T}$.

We consider the distance between words as the Euclidean distance for sets, defined as $d(\mathbf{w}_1, \mathbf{w}_2) = \min_{\alpha_1,\alpha_2} \|\mathbf{w}_1\alpha_1 -$

$\mathbf{w}_2\alpha_2\|$. This expression can be computed as

$$\begin{aligned} &\min_{\alpha_1,\alpha_2} \sqrt{\|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2 - 2\mathcal{R}(\alpha_1^*\alpha_2\mathbf{w}_1^H\mathbf{w}_2)} \\ &= \sqrt{1 + 1 - 2\max_{\alpha_1,\alpha_2} \mathcal{R}(\alpha_1^*\alpha_2\mathbf{w}_1^H\mathbf{w}_2)} \\ &= \sqrt{2 - 2|\mathbf{w}_1^H\mathbf{w}_2|}. \end{aligned} \tag{11}$$

This distance will be the inout to a cluster cover finding algorithm, that we select as k-medoids [8], which will identify N_{W} cluster representatives. To make sure that no spurious data came from the previous step, we will add a simple validation algorithm to the cluster representatives. Thus, for a word to be included in the dictionary, it needs to have been obtained from a minimum amount of independent measurements τ_{M}, this is, having a cluster size of at least τ_{M}. The dictionary composition process described in this subsection is summarized in algorithm 2.

Algorithm 2 Dictionary composition from words

```
INPUT: P
D ← ∅
Use k-medioids to extract N_W clusters {C_i} from P
for i ≤ N_W do
    if |C_i| > τ_M then
        Add C_i 's representative to D
    end if
end for
```

IV. RESULTS

For simulation purposes we assume a MIMO-OFDM hybrid mmWave link using uniform linear arrays at the transmitter and the receiver sides, with $N_{\mathrm{t}} = N_{\mathrm{r}} = 16$ antennas, $L_{\mathrm{t}} = L_{\mathrm{r}} = 2$ RF chains and 128 OFDM subcarriers. The number of bits used to control the phase shifters in the analog precoding/combing networks is 2. The SNR is set to 0 dB, and a spreading factor of 10 is used during the off-line dictionary learning stage. The sampling frequency is 1760 MHz. The grid size used to generate the dictionaries is 1024.

We start by evaluating the proposed dictionary learning method using two metrics. First we consider the Euclidean distance of an estimated word $w \in \mathcal{D}$ to the actual normalized dictionary $\bar{\mathcal{D}}$ according to the distance defined in (11)

$$d(\mathbf{w}, \bar{\mathcal{D}}) = \min_{\bar{\mathbf{w}} \in \bar{\mathcal{D}}} (d(\mathbf{w}, \bar{\mathbf{w}})). \tag{12}$$

The second metric is the counterpart to the first one, the Euclidean distance of an actual normalized word to the estimated dictionary

$$d(\bar{\mathbf{w}}, \mathcal{D}) = \min_{\mathbf{w} \in \mathcal{D}} (d(\bar{\mathbf{w}}, \mathbf{w})). \tag{13}$$

The channel parameters L, τ_l, ϕ_l and θ_l are generated using the 5G-NR statistical channel model implemented in Quadriga [9]. The number of delay taps is set to 16. The evaluation is performed by means of a Monte Carlo simulation with 1024 independent realizations. Each realization generates realistic antenna radiation properties and other impairments according to the model and parameters described in [10].

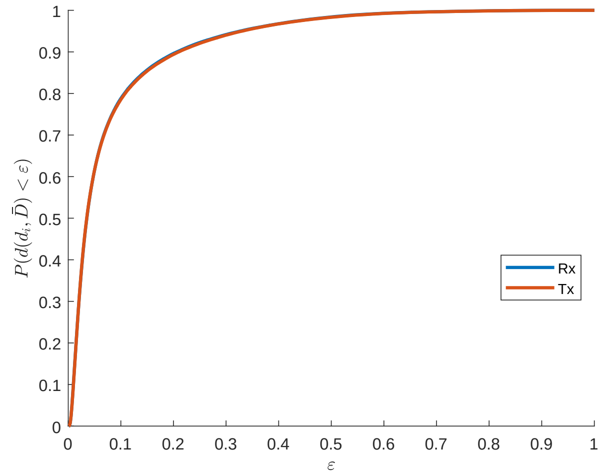

Fig. 1: Distance of the estimated codebook words to the CDF of the actual normalized one.

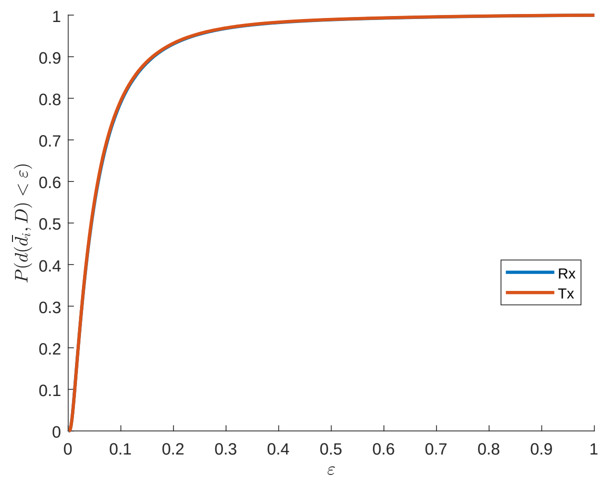

Fig. 2: Distance of the real normalized codebook words to the CDF of estimated one.

Fig. 1 shows the CDF of the Euclidean distance between the estimated dictionary words to the actual dictionary (after normalization). The accuracy of the estimated dictionary is very high, given the similarity between the actual words and the estimated ones.

Fig. 2 shows the CDF of the Euclidean distance between the actual normalized dictionary words to the estimated one. These plots show that the actual dictionary is well covered, since the actual codewords are very close to the estimated dictionary.

After evaluating the accuracy of the estimated dictionary, we consider now a metric directly related to the application of the dictionary for channel estimation at mmWave fre-

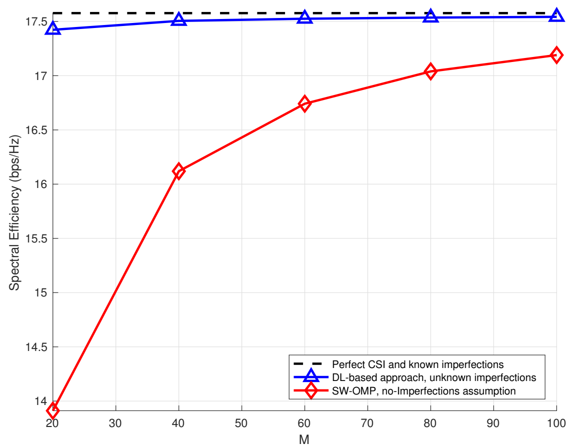

Fig. 3: Average distance of the real normalized codebook words to the estimated one CDF per angle.

quencies. We evaluate the quality of the estimated channel using the achieved spectral efficiency shown in Fig. 3. We consider the compressive channel estimation algorithm SWOMP proposed in [4], and compare the spectral efficiency that can be achieved when to estimate the channel we use the sparsifying dictionaries learnt by our proposed method (blue curve) and also the dictionaries based on the extended virtual channel model defined in [4] (red curve), that neglect the impairments. It can be clearly observed that the performance of SWOMP is near optimal when the learnt dictionaries are assumed for the channel estimation stage. When neglecting the impairments and considering ideal dictionaries the performance loss is significant, in the order of 1 bps/Hz for 50 training symbols or 2 bps/Hz for 35 training symbols. This proves the effectiveness of our off-line dictionary learning method to compensate for the hardware impairments during the channel estimation stage.

V. CONCLUSIONS

We proposed a DL method to obtain sparsifying dictionaries for channel sparse representation in hybrid wideband mmWave MIMO systems that account for hardware impairments. A state of the art channel estimation strategy for frequency selective mmWave channels, that relies on the sparse assumption for the channel, was tested using ideal dictionaries and the ones obtained with our proposed approach. Considering channel realizations that account for hardware imperfections, the achieved spectral efficiency of the mmWave MIMO link obtained with channel estimates based on the learnt dictionaries, is very close to the one obtained when perfect channel knowledge is assumed.

VI. REFERENCES

[1] P. Schniter and A. Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, 2014, pp. 273–277.

[2] R. M. Rial, C. Rusu, A. Alkhateeb, N. G. Prelcic, and R. W. Heath Jr., "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?," *IEEE Access*, , no. 99, Jan 2016.

[3] K. Venugopal, A. Alkhateeb, N. González-Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sept. 2017.

[4] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946–2960, May 2018.

[5] Y. Ding and B. D. Rao, "Dictionary learning-based sparse channel representation and estimation for FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5437–5451, Aug. 2018.

[6] Y. Ding and B. D. Rao, "Compressed downlink channel estimation based on dictionary learning in FDD massive MIMO systems," in *Proc. of IEEE GLOBECOM'15*, Dec. 2015, pp. 1–6.

[7] Hongxiang Xie, Javier Rodrguez-Fernndez, and Nuria Gonzlez-Prelcic, "Dictionary learning for channel estimation in hybrid frequency-selective mmwave mimo systems," 2019.

[8] Hae-Sang Park and Chi-Hyuck Jun, "A simple and fast algorithm for k-medoids clustering," *Expert Systems with Applications*, vol. 36, no. 2, Part 2, pp. 3336 – 3341, 2009.

[9] S. Jaeckel, L. Raschkowski, K. Brner, and L. Thiele, "QuaDRiGa: A 3-D Multi-Cell Channel Model With Time Evolution for Enabling Virtual Field Trials," *IEEE Trans. on Antennas and Propagation*, vol. 62, no. 6, pp. 3242–3256, June 2014.

[10] Michael Eberhardt, Philipp Eschlwech, and Erwin Biebl, "Investigations on antenna array calibration algorithms for direction-of-arrival estimation," *Advances in Radio Science: ARS*, vol. 14, pp. 181–190, 2016.