

# Machine Learning for CSI Recreation Based on Prior Knowledge

Brenda Vilas Boas<sup>1,2</sup>, Wolfgang Zirwas<sup>1</sup>, Martin Haardt<sup>2</sup>

<sup>1</sup>Nokia, Germany

<sup>2</sup>Ilmenau University of Technology, Germany

brenda.vilas\_boas@nokia.com, wolfgang.zirwas@nokia-bell-labs.com, martin.haardt@tu-ilmenau.de

**Abstract**—Knowledge of channel state information (CSI) is fundamental to many functionalities within the mobile wireless communications systems. With the advance of machine learning (ML) and digital maps, i.e., digital twins, we have a big opportunity to learn the propagation environment and design novel methods to derive and report CSI. In this work, we propose to combine untrained neural networks (UNNs) and conditional generative adversarial networks (cGANs) for MIMO channel recreation based on prior knowledge. The UNNs learn the prior-CSI for some locations which are used to build the input to a cGAN. Based on the prior-CSIs, their locations and the location of the desired channel, the cGAN is trained to output the channel expected at the desired location. This combined approach can be used for low overhead CSI reporting as, after training, we only need to report the desired location. Our results show that our method is successful in modelling the wireless channel and robust to location quantization errors in line of sight conditions.

**Index Terms**—Channel estimation, channel prediction, UNN, cGAN, digital twin.

## I. INTRODUCTION

Machine learning (ML) for physical layer applications are gaining momentum in standardization bodies, such as 3GPP and O-RAN [1]. Combining ML capabilities with virtual representations of the real world, i.e., a digital twin environment, enables a variety of possi-

bilities for wireless network planning, deployment and management. In order to leverage the potential of a digital twin of the environment, full knowledge of the channel state information (CSI) is desired such that most of the real propagation effects can be represented. Here, we propose to combine two ML methods for channel recreation/estimation which minimize the overall complexity of the neural networks (NNs), reduce training time, and enable low CSI reporting overhead. In contrast to state of the art, our solution does not rely on multimodal data, such as lidar [2] or environment images [3], which allow us to reduce the complexity of our NN architectures.

Untrained neural networks (UNNs) were first proposed in [4] to solve inverse problems, such as denoising. The term ‘untrained’ refers to the method characteristic of avoiding a huge data collection phase as the updates of the gradient descent is for a single image measurement. The *deep decoder* architecture as proposed in [5] simplifies the structure of a UNN, making it underparameterized. For wireless communications, this means we can fit a UNN to directly estimate the wireless channel based on a small noisy measurement campaign, i.e., a few time snapshots, without the need of noiseless labels. The work in [6] has proposed the use of UNNs for MIMO

channel estimation under pilot contamination. Despite the limitation to statistical channel models, UNNs could reduce the noise level of the measured signal.

The simplicity of UNNs comes at the cost of lack of generalization. Since there is no dataset collection for weights update, iterating the gradient descent is always needed when a new set of channel measurements is acquired. In our recent work [7], we have proposed to use cGAN for channel estimation in MIMO arrays with mixed radio frequency chains where part of the array had antenna elements turned-off. Our results demonstrated the good generalization capability of cGANs.

Motivated by the generalization capabilities of cGAN and the underparameterization of UNNs, we propose to combine them for MIMO channel estimation/prediction within a propagation area. Hence, the UNNs are used to generate prior CSI for a set of locations. Then, the cGAN uses the prior-CSIs together with their locations to recreate the CSI in a desired location. After deriving the weights for all the ML models, only the target location needs to be reported. Therefore, our solution enables low CSI reporting overhead between user equipment (UE) and base station (BS). Moreover, our approach can be used to add to the digital twin the small scaling fading characteristics of the wireless channels.

In this paper, Section II presents our geometrical propagation environment and the channels considered, Section III introduces our proposed method, Section IV presents details about our UNN for prior knowledge CSI estimation, Section V shows the processing performed at the cGAN for CSI recreation using location and prior-CSI, Section VI presents our results, and Section VII concludes our paper.

Regarding the notation,  $a$ ,  $\mathbf{a}$ ,  $\mathbf{A}$  and  $\mathcal{A}$  represents, respectively, scalars, column vectors, matrices and  $D$ -dimensional tensors. The superscript  $T$ , denotes transposition. For a tensor  $\mathcal{A} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_D}$ ,  $M_d$

refers to the tensor dimension on the  $d^{\text{th}}$  mode. A  $d$ -mode unfolding of a tensor is written as  $[\mathcal{A}]_{(d)} \in \mathbb{C}^{M_d \times M_{d+1} \dots M_D M_1 \dots M_{d-1}}$  where all  $d$ -mode vectors are aligned as columns of a matrix. The  $d$ -mode vectors of  $\mathcal{A}$  are obtained by varying the  $d^{\text{th}}$  index from 1 to  $M_d$  and keeping all other indices fixed. Moreover,  $\mathcal{A} \times_d \mathbf{U}$  is the  $d$ -mode product between a  $D$ -way tensor  $\mathcal{A} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_D}$  and a matrix  $\mathbf{U} \in \mathbb{C}^{J \times M_d}$ . The  $d$ -mode product is computed by multiplying  $\mathbf{U}$  with all  $d$ -mode vectors of  $\mathcal{A}$ . In addition,  $\mathcal{A} \sqcup_d \mathcal{B}$  denotes the concatenation of  $\mathcal{A}$  and  $\mathcal{B}$  among the  $d^{\text{th}}$  mode. The concatenation  $\sqcup_d$  operation also applies to matrices.

## II. PROPAGATION ENVIRONMENT

In this work, we consider an urban environment with a fixed base station (BS) equipped with an uniform rectangular array (URA) containing  $N_{\text{ant}}$  antenna elements, moving user equipment (UEs) with single antennas, operating with  $N_{\text{sub}}$  OFDM subcarriers, and collecting  $N_{\text{sp}}$  time snapshots. This scenario was modeled with IImProp, a geometry based channel simulator developed at Ilmenau University of Technology [8]. Figure 1 presents the urban environment with the BS represented by a red circle, buildings in blue squares, scatters in green circles, and three UEs moving in a linear trajectory towards the BS.

From IImProp, we collect the channels used as ground truth values  $\mathcal{H}_{\text{sim}}^C \in \mathbb{C}^{N_{\text{sp}} \times N_{\text{sub}} \times N_{\text{ant}}}$  and their relative locations  $\Gamma = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\} \in \mathbb{R}^{N_{\text{sp}} \times 3}$  to the BS. The three UEs shown in Figure 1 are used to derive the noisy channel measurements  $\mathcal{H}_{\text{mes}}^C \in \mathbb{C}^{N_{\text{sp}} \times N_{\text{sub}} \times N_{\text{ant}}}$  as

$$\mathcal{H}_{\text{mes}}^C = \mathcal{H}_{\text{sim}}^C + \mathcal{N}, \quad (1)$$

where  $\mathcal{N} \in \mathbb{C}^{N_{\text{sp}} \times N_{\text{sub}} \times N_{\text{ant}}}$  is a zero mean circularly symmetric complex Gaussian noise process. The  $\mathcal{H}_{\text{mes}}^C$  are further used to estimate the prior-CSIs. Moreover,

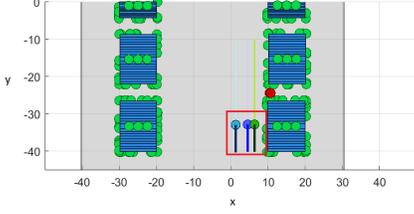


Fig. 1. Propagation scenario simulated in IlmProp, a geometric based channel simulator developed at Ilmenau University of Technology. The BS is fixed and represented by a red circle. The displayed UEs and trajectories are used to compute the prior-CSIs. The red square represents the study area (100 m<sup>2</sup>), to which we aim to reconstruct the CSIs.

the red square indicates the study area where the  $\mathcal{H}_{\text{sim}}^C$  are collected for CSI recreation.

### III. CSI RECREATION WITH PRIOR KNOWLEDGE

Here, we propose a ML framework to recreate CSI in a desired location based on prior knowledge of CSI at neighboring UEs. Figure 2 summarizes our proposed method for CSI recreation which we divide in two parts, the first in blue and the second in purple. The first ML instance aims to find the prior-CSI  $\mathcal{H}_p \in \mathbb{R}^{N_{\text{sp}} \times N_{\text{sub}} \times 2N_{\text{ant}}}$  based on the measured channels  $\mathcal{H}_{\text{mes}}^C$ . We employ a UNN for this purpose where each UNN estimates  $\mathcal{H}_p$  the channel of a single UE over multiple time snapshots. Even though UNNs are low complex structures, deriving one UNN model for each possible location in a propagation environment is unfeasible. Therefore, we propose to use a second ML instance based on a cGAN due to its generalization capabilities. The second ML instance is trained to compute the recreated-CSI  $\mathcal{H}_r \in \mathbb{R}^{(S+1) \times N_{\text{sub}} \times (2N_{\text{ant}}+1)}$  in the targeted location  $\Gamma_r \in \mathbb{R}^{1 \times 3}$  based on the knowledge of a sub-set of  $S$  selected prior-CSIs and their respective locations  $\Gamma_p \in \mathbb{R}^{S \times 3}$ .

Since UNNs do not need ‘labels’ to find their best weights, we can perform a small measurement campaign and use the UNN-estimated channels as conditional

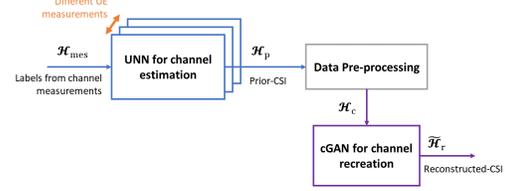


Fig. 2. Schematic of our proposed ML solution for CSI recreation combining channel measurements and simulations. In blue, we present the first ML part where the UNNs are iterated to perform channel estimation. Each UE measurement campaign derive one UNN that represents  $N_{\text{sp}}$  locations. In purple, we show the second ML part where a cGAN is trained with knowledge of the prior-CSIs, the simulated desired channels and their locations. The cGAN output is the recreated desired channel.

input to the cGAN. In a ‘day-zero’ operation where not many CSI measurements are available, the cGAN can be trained with target channels derived from simulations or a digital twin and the prior-CSIs from the UNNs are responsible to adjust the model to real propagation conditions. In the long run, we could update the cGAN model based on collected real world measurements. In this scenario, the availability of priors at the conditional input reduce the complexity of the NN structure and its training time. In the following sections, we explain in details how each part of the algorithm is trained.

Due to the low computational complexity of UNNs, the UE can derive the UNN weights and send it to the BS. Then, the BS is able to reconstruct the prior-CSIs and can train a cGAN to recreate the CSI at a desired location, which is different from the prior-CSI locations. After deriving the cGAN optimal weights, the BS can send the cGAN model to the UE for the purpose of reliability. Hence, the UE is able to identify when the BS will fail on its CSI recreation and may trigger a correction procedure.

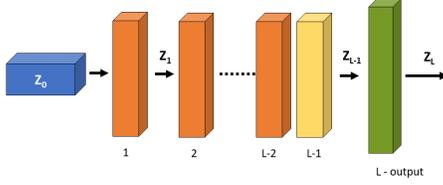


Fig. 3. General layer structure of a UNN  $P$  used to estimate the prior-CSIs  $\mathcal{H}_p = P(\mathcal{K}^*, \mathcal{Z}_0)$ . There are  $L$  layers,  $L - 2$  inner layers in orange, one pre-output layer in yellow, and one output layer in olive. In blue, we represent  $\mathcal{Z}_0$  the random input tensor.

#### IV. ESTIMATION OF PRIOR-CSI WITH UNNS

The underparametrization of deep decoder [5] and its capability to optimize noisy measurements has motivated us to employ UNNs as our channel estimator for prior-CSIs. Since we do not need the true-labels, the channel measurements can be acquired during measurement campaigns which will then represent the real world propagation environment for specific locations. The following subsections present the data pre-processing, the UNN architecture and how the gradient descent is used to update the UNN weights.

##### A. Data pre-processing for UNN

The input signal to an UNN is a random noise seed  $\mathcal{Z}_0 \in \mathbb{R}^{k_1 \times b \times c}$ , where  $b = N_{\text{sp}}/2^{L-2}$ ,  $c = N_{\text{sub}}/2^{L-2}$ ,  $k_1$  is the number of filters in the first layer, and  $L$  is the number of layers. The input tensor  $\mathcal{Z}_0$  is drawn from a uniform distribution  $U(-a, +a)$  defined on the interval  $[-a, +a]$  and kept fixed during the iterations to update the gradient descent. The measured channel  $\mathcal{H}_{\text{mes}}^C$  from Equation 1 is preprocessed as

- Each time snapshot within  $\mathcal{H}_{\text{mes}}^C$  is normalized by its Frobenius norm, and then multiplied by a scaling factor to ease convergence.
- $\mathcal{H}_{\text{mes}}^C \in \mathbb{C}^{N_{\text{sp}} \times N_{\text{sub}} \times N_{\text{ant}}}$  is rearranged by concatenating  $\Re\{\mathcal{H}_{\text{mes}}^C\}$  and  $\Im\{\mathcal{H}_{\text{mes}}^C\}$  in the dimension corresponding to the antenna elements.

After those operations,  $\mathcal{H}_{\text{mes}} \in \mathbb{R}^{N_{\text{sp}} \times N_{\text{sub}} \times 2N_{\text{ant}}}$  is directly used to compute the cost function.

##### B. UNN architecture

A UNN is a composition of  $L$  layers where there are  $(L-2)$  inner layers, one pre-output layer ( $L-1$ ) and one output layer ( $L$ ), according to the *deep decoder* architecture [5]. Figure 3 shows a generic organization of those layers, the random noise seed  $\mathcal{Z}_0$  in blue, the inner layers in orange, the pre-output layer in yellow, and the output layer in olive. All the layer types contain convolutional filters  $\mathcal{W}_l \in \mathbb{R}^{1 \times 1 \times k_{l-1} \times k_l}$  where  $l = \{1, 2, \dots, L\}$ ,  $k_{l-1}$  and  $k_l$  are hyper-parameters which define the number of filters on the respective  $(l-1)^{\text{th}}$  and  $l^{\text{th}}$  layers. However, the types of layers differ with respect to the upsampling computation and the operation of batch normalization (BatchNorm) operation [9]. The inner layers contain linear and non-linear operations. First, there is a convolutional filter  $\mathcal{W}_l$  which weights are updated by the gradient descent. Second, there is a fixed bilinear upsampling operation, where  $\mathbf{A}_l \in \mathbb{R}^{2^l b \times 2^{l-1} b}$  and  $\mathbf{C}_l \in \mathbb{R}^{2^l c \times 2^{l-1} c}$  are the linear upsampling matrices in the subcarrier and time snapshots dimensions, respectively. Third, the rectifier linear unit (ReLU) activation function is applied, and a batch normalization is computed with trainable parameters  $\mathbf{R}_l = [\gamma_l, \beta_l] \in \mathbb{R}^{k_l \times 2}$ , where  $\gamma_l \in \mathbb{R}^{k_l \times 1}$  and  $\beta_l \in \mathbb{R}^{k_l \times 1}$  are the mean and variance correction factors, respectively, for the coefficients in the  $l^{\text{th}}$  layer. For example, the output of the first inner layer  $\mathcal{Z}_1$  can be written as

$$\mathcal{Z}_1 = \text{BatchNorm}(\text{ReLU}(\mathcal{Z}_0 \times_3 [\mathcal{W}_1]_{(4)} \times_1 \mathbf{A}_1 \times_2 \mathbf{C}_1^T)), \quad (2)$$

where  $[\mathcal{W}_1]_{(4)}$  is the 4-mode unfolding of the convolutional filters operating at the antenna elements dimension. The pre-output layer differs from the inner layers

because it does not apply upsampling. Hence, it can be written as

$$\mathcal{Z}_{L-1} = \text{BatchNorm}(\text{ReLU}(\mathcal{Z}_{L-2} \times_3 [\mathcal{W}_{L-1}]_{(4)})). \quad (3)$$

Next, the output layer is used to adjust the number of filters of the pre-output layer to the size expected at the output  $k_L = 2N_{\text{ant}}$  as

$$\mathcal{Z}_L = \text{TanH}(\mathcal{Z}_{L-1} \times_3 [\mathcal{W}_L]_{(4)}), \quad (4)$$

where  $\mathcal{W}_L \in \mathbb{R}^{1 \times 1 \times k_{l-1} \times 2N_{\text{ant}}}$ , and TanH is the hyperbolic tangent activation function. Since the upsampling operations are pre-defined, the trainable parameters relates to the convolutional filters  $\mathcal{W}_l$  and the regularization parameters  $\mathbf{R}_l$  of the batch normalization operation. Therefore,  $\mathcal{K}_l = \{\mathcal{W}_l, \mathbf{R}_l\}$  is the set of trainable parameters of the  $l^{\text{th}}$  layer, and  $\mathcal{K}$  refers to all trainable parameters of the  $L$  layers.

### C. Updating the weights of an UNN

Here, we refer to the UNN as a model  $P: \mathbb{R}^N \rightarrow \mathbb{R}^{N_{\text{sub}}N_{\text{sp}}2N_{\text{ant}}}$  where  $N < N_{\text{sub}}N_{\text{sp}}2N_{\text{ant}}$  is the total number of parameters. The UNN  $P$  performs the mapping operation  $\mathcal{Z}_L = P(\mathcal{K}, \mathcal{Z}_0)$ , where  $\mathcal{Z}_0$  is the random noise seed, and  $\mathcal{K}$  is the tensor of weights that represents all the UNN trainable parameters.

The cost function is the mean square error (MSE), calculated as

$$\mathcal{L}(\mathcal{K}) = \mathbb{E}\{\|P(\mathcal{K}, \mathcal{Z}_0) - \mathcal{H}_{\text{mes}}\|_F^2\}. \quad (5)$$

The gradient descent is updated as in supervised learning, performing  $I$  gradient iterations until the optimum parameters are found, such that

$$\mathcal{K}^* = \arg \min_{\mathcal{K}} \mathcal{L}(\mathcal{K}), \text{ and } \mathcal{H}_p = P(\mathcal{K}^*, \mathcal{Z}_0) \quad (6)$$

is the channel estimation of the prior-CSI. From the loss function, we observe that the prior-CSI  $\mathcal{H}_p$  derived by the UNN  $P$  is specific to  $\mathcal{H}_{\text{mes}}$ . Hence, the model

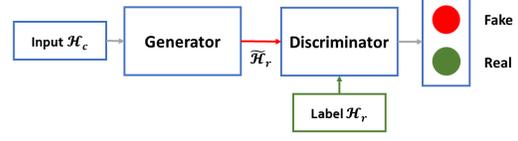


Fig. 4. Conditional GAN, two NNs play a minmax game where the generator tries to fool the discriminator. The discriminator should classify  $\tilde{\mathcal{H}}_r$  as a fake sample, while  $\mathcal{H}_r$  is classified as a real sample. The generator fools the discriminator when  $\tilde{\mathcal{H}}_r$  is classified as real.

$P$  does not directly generalize for other channels, it is specific to the  $\mathcal{H}_{\text{mes}}$  considered during gradient updates.

## V. CSI-RECREATION WITH cGAN

In our previous work [7], we have applied cGAN for the purpose of channel estimation within mixed-resolution radio-frequency chains. Here, we use the same operational principle of image to image translation [10]. However, we aim to reconstruct CSI based on the knowledge of its location and a set of prior-CSIs. Different from UNNs, cGAN requires data collection and training. Nonetheless, it has great generalization capabilities [7]. In the following subsections, we present the dataset preprocessing, our cGAN architecture, and the adversarial training.

### A. Dataset preprocessing for cGAN

In this section, we present how we construct the signals to train the cGAN: the conditional input  $\mathcal{H}_c$ , the label  $\mathcal{H}_r$ , and the generator output  $\tilde{\mathcal{H}}_r$ .

The conditional input to our cGAN is derived from the prior-CSIs  $\mathcal{H}_p$ , their locations  $\Gamma_p$ , and the target location  $\Gamma_r$  where we aim to recover the CSI of a certain UE. Each  $\mathcal{H}_p \in \mathbb{R}^{N_{\text{sub}} \times N_{\text{sp}} \times 2N_{\text{ant}}}$  estimated by a UNN has CSI for  $N_{\text{sp}}$  different locations. Therefore, if  $N_{\text{UE}}$  UNNs are used to estimate the prior-CSIs  $\mathcal{H}_p^{N_{\text{UE}}} \in \mathbb{R}^{N_{\text{UE}}N_{\text{sp}} \times N_{\text{sub}} \times 2N_{\text{ant}}}$ , there are  $N_{\text{UE}}N_{\text{sp}}$  CSI-location pairs  $\{\mathbf{H}_{pj}, \Gamma_{pj}\}$  available, where  $\mathbf{H}_{pj} = \mathcal{H}_p^{N_{\text{UE}}}(j, :, :)$

and  $j = \{1, 2, \dots, N_{\text{UE}}N_{\text{sp}}\}$ . From the available CSI-location pairs, a sub-set of  $S$  CSI-location pairs is selected according to their minimum Euclidean distance to  $\Gamma_r$ . The  $S$  selected prior-CSIs  $\mathcal{H}_p^S \in \mathbb{R}^{S \times N_{\text{sub}} \times 2N_{\text{ant}}}$  are concatenated in the first dimension and ordered according to the minimum Euclidean distance to the target location  $\Gamma_r$ . The target location vector  $\Gamma_r \in \mathbb{R}^{1 \times 3}$  and the prior location matrix  $\Gamma_p^S \in \mathbb{R}^{S \times 3}$  are extended by repeating their coordinates until  $\Gamma_p^S \in \mathbb{R}^{S \times N_{\text{sub}}}$  and  $\Gamma_r \in \mathbb{R}^{1 \times N_{\text{sub}}}$ . Hence, the complete location matrix is formed as  $\mathbf{H}_{\text{LOC}} = [\Gamma_r \sqcup_1 \Gamma_p^S] \in \mathbb{R}^{(S+1) \times N_{\text{sub}}}$ . Finally, the conditional input to the cGAN is constructed as

$$\mathcal{H}_c = [\mathbf{H}_N \sqcup_1 \mathcal{H}_p^S \sqcup_3 \mathbf{H}_{\text{LOC}}] \in \mathbb{R}^{(S+1) \times N_{\text{sub}} \times (2N_{\text{ant}}+1)}, \quad (7)$$

where  $\mathbf{H}_N \in \mathbb{R}^{N_{\text{sub}} \times 2N_{\text{ant}}}$  is a matrix of random values drawn from a Gaussian distribution. The desired channel  $\mathbf{H}_r$  is recreated in the  $\mathbf{H}_N$  position at the generator output.

Ideally, the true recreated CSI  $\mathcal{H}_r \in \mathbb{R}^{(S+1) \times N_{\text{sub}} \times (2N_{\text{ant}}+1)}$  is found in the output of the cGAN. Hence, each  $d$  time snapshot in  $\mathbf{H}_{\text{sim}}(d) = \mathcal{H}_{\text{sim}}^C(d, :, :)$  is used as ground truth value for  $\mathbf{H}_r$ , the CSI reconstructed at the target location  $\Gamma_r$ .

The pre-processing of the labels for cGAN include

- Normalization of  $\mathbf{H}_{\text{sim}}(d) \in \mathbb{C}^{N_{\text{sub}} \times N_{\text{ant}}}$  by its Frobenius norm, and multiplication by a scaling factor.
- $\mathbf{H}_r = [\Re\{\mathbf{H}_{\text{sim}}(d)\} \sqcup_2 \Im\{\mathbf{H}_{\text{sim}}(d)\}] \in \mathbb{R}^{N_{\text{sub}} \times 2N_{\text{ant}}}$  is the target real valued CSI at location  $\Gamma_r$ .

Finally, the label is constructed as

$$\mathcal{H}_r = [\mathbf{H}_r \sqcup_1 \mathcal{H}_p^S \sqcup_3 \mathbf{H}_{\text{LOC}}] \in \mathbb{R}^{(S+1) \times N_{\text{sub}} \times (2N_{\text{ant}}+1)}, \quad (8)$$

where the prior-CSIs  $\mathcal{H}_p^S$ , the location matrix  $\mathbf{H}_{\text{LOC}}$  and the recreated CSI  $\mathbf{H}_r$  form the desired output. We refer to the generator output as  $\tilde{\mathcal{H}}_r = G(\mathcal{H}_c) \in$

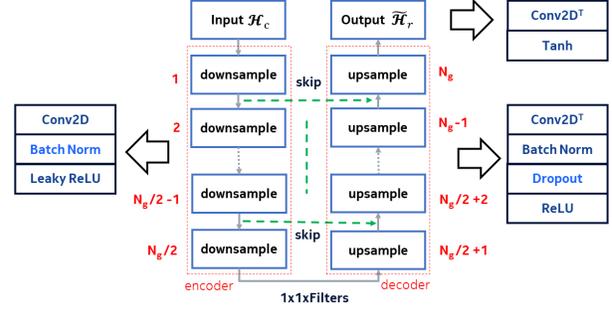


Fig. 5. U-Net architecture deployed as the generator including encoder and decoder pipeline and numbering for skip connections.

$\mathbb{R}^{(S+1) \times N_{\text{sub}} \times (2N_{\text{ant}}+1)}$ , where  $G$  is the generator mapping function that tries to approximate the label  $\mathcal{H}_r$ . The discriminator  $D$  is a classifier for which the inputs and labels are, respectively,  $\mathcal{H}_r \rightarrow \{\text{true}\}$  and  $\tilde{\mathcal{H}}_r \rightarrow \{\text{fake}\}$ .

## B. Adversarial Network Architecture

Figure 4 shows the interconnection between generator and discriminator NNs for the adversarial training. Here, the generator NN consists of an U-shaped deep NN (U-Net) which has two paths for the flow of information between blocks: the encoder-decoder path and the skip connections path, see Figure 5. The discriminator NN consists of a Patch-NN [10] where the input is reduced to a patch of arbitrary size; then, each coefficient of the patch is classified as real or fake.

Figure 5 shows the U-Net architecture employed for the generator, the  $N_g/2$  downsample blocks for the encoder and the  $N_g/2$  upsample blocks for the decoder, where  $N_g$  is the total number of processing blocks. Each downsample block consists of one convolutional 2-dimensional layer (Conv2D), one batch normalization layer (BatchNorm), and a leaky rectifier linear unit (LeakyReLU) activation function, where  $y = x$  for  $x > 0$ , and  $y = 0.3x$  for  $x < 0$ . Each upsample block consists of one transposed convolutional 2-dimensional layer (Conv2D<sup>T</sup>), followed by BatchNorm and ReLU as

activation function. The skip connections path happens between the output of the  $n^{\text{th}}$  downsample block and the output of the  $(N_g - n)^{\text{th}}$  upsample block, where  $n = [1, 2, \dots, (N_g/2 - 1)]$ . Those skip connections provide more information to the decoder block [10] since the input to each upsample block is the concatenation  $\mathcal{X}_{(N_g-n+1)} = [\mathcal{Y}_n \sqcup_3 \mathcal{Y}_{(N_g-n)}]$ , where  $\mathcal{Y}_n$  is the output of the  $n^{\text{th}}$  block.

For the discriminator NN, we employ a Patch-NN [10]. First, downsampling blocks are used to reduce the dimensionality of the input signal to some patch of arbitrary size. Second, the patch is processed by a sequence of convolutional layers (Conv2D + BatchNorm + LeakyReLU and Conv2D + Linear). Then, the discriminator is trained to classify each patch coefficient as real or fake. Implementations details are provided in Section VI.

### C. Optimization with cGAN

As shown in Figure 4, in cGAN there are two NNs playing a minmax game where the generator  $G : \{\mathcal{H}_c\} \rightarrow \mathcal{H}_r$  tries to fool the discriminator  $D : \{\tilde{\mathcal{H}}_r\} \rightarrow \{\text{true}\}$ , and it is conditional because some prior knowledge is provided. Mathematically, the optimization objective of a cGAN has two terms

$$G^* = \arg \min_G \max_D \mathcal{L}_{\text{cGAN}}(G, D) + \alpha \mathcal{L}_{L_2}, \quad (9)$$

where  $\mathcal{L}_{\text{cGAN}}(G, D)$  is the adversarial loss,  $\mathcal{L}_{L_2}$  is the  $L_2$  loss, and  $\alpha$  is the weighting factor [10]. The adversarial loss is computed as

$$\begin{aligned} \mathcal{L}_{\text{cGAN}}(G, D) = & \mathbb{E}[\log D(\mathcal{H}_r)] + \\ & \mathbb{E}[\log(1 - D(\tilde{\mathcal{H}}_r))], \end{aligned} \quad (10)$$

where the generator  $G$  learns to map the input data  $\mathcal{H}_c$  to the output data  $\mathcal{H}_r$  such that  $\tilde{\mathcal{H}}_r = G^*(\mathcal{H}_c)$ , and the discriminator  $D$  tries to recognize the channels generated by  $G$ . In order to have the generated output wireless

channels  $\tilde{\mathcal{H}}_r$  close to the wireless channel labels  $\mathcal{H}_r$ , a weighted  $L_2$  loss

$$\mathcal{L}_{L_2}(G) = \mathbb{E}[\|\mathcal{H}_r - G(\mathcal{H}_c)\|_F] \quad (11)$$

is included as a regularization term. In [11] a  $L_2$  reconstruction loss is also proposed, but 0/1 masks are used to restrict the  $L_2$  loss only to the signal we aim to estimate. Note that, at the cGAN output, we have the conditional information as well as the desired signal. Hence, we study the feasibility of a reconstruction loss  $\mathcal{L}_{\text{rec}}$  defined as

$$\mathcal{L}_{\text{rec}} = \mathbb{E}[\|\mathbf{H}_r - \tilde{\mathbf{H}}_r\|_F] + \mathbb{E}[\|\mathbf{H}_{\text{LOC}} - \tilde{\mathbf{H}}_{\text{LOC}}\|_F], \quad (12)$$

where the MSE of the target channel and the location matrix are considered. Then,  $\mathcal{L}_{\text{rec}}$  substitutes  $\mathcal{L}_{L_2}$  in Equation 10.

The generator and discriminator NNs are trained together in each epoch. For testing, or inference, only the generator architecture is used. Therefore, only knowledge of  $\mathcal{H}_c$  is needed. In practice, at inference time, we are able to estimate/predict a channel based on its location and the prior-knowledge provided by the UNNs.

## VI. SIMULATIONS AND RESULTS

Figure 1 presents our propagation environment simulated at IlmProp where there are 3 UEs used to derive  $\mathcal{H}_{\text{mes}}^C$  for prior-CSI estimation by the UNNs, and the red square indicates our study area where  $\mathcal{H}_{\text{sim}}^C$  is collected for training the cGAN to recreate the CSIs. Table I presents the simulation parameters set at IlmProp. A total of seven simulation campaigns were performed, changing the number of UEs as well as their trajectories, and including or removing scatters. The dataset for prior-CSI estimation with UNNs has 384 channel samples, and the dataset for cGAN has 4492 channel samples. We use the normalized squared error (NSE)  $\text{NSE} = \frac{\|\mathbf{B} - \tilde{\mathbf{B}}\|_F^2}{\|\mathbf{B}\|_F^2}$  as our performance metric for CSI recreation.

TABLE I  
SIMULATION PARAMETERS FOR ILMPROP.

Parameter	$\mathcal{H}_{\text{mes}}^C$	$\mathcal{H}_{\text{sim}}^C$
Carrier frequency	2.6 GHz	
Bandwidth	20 MHz	
$N_{\text{sub}}$	64	
UE velocity	1 m/s	0.1 to 5 m/s
$N_{\text{sp}}$	128	74 to 174
Total of UEs	3	many
$N_{\text{ant}}$	36	

TABLE II  
DESCRIPTION OF THE U-NET DEPLOYED AS GENERATOR NN.

	Block	$N_{\text{filter}}$	Filter size	Stride	Padding	BatchNorm	Dropout	Activation
1	downsample	64	[3,3]	[1,1]	Yes	No	No	LeakyReLU
2	downsample	64	[2,4]	[1,2]	Yes	Yes	No	LeakyReLU
3	downsample	64	[2,4]	[1,2]	Yes	Yes	No	LeakyReLU
4	downsample	64	[2,4]	[1,2]	Yes	Yes	No	LeakyReLU
5	downsample	64	[2,4]	[1,2]	Yes	Yes	No	LeakyReLU
6	downsample	128	[4,4]	[1,1]	No	Yes	No	LeakyReLU
7	upsample	64	[4,4]	[2,2]	No	Yes	Yes	ReLU
8	upsample	64	[2,4]	[1,2]	Yes	Yes	Yes	ReLU
9	upsample	64	[2,4]	[1,2]	Yes	Yes	Yes	ReLU
10	upsample	64	[2,4]	[1,2]	Yes	Yes	No	ReLU
11	upsample	64	[2,4]	[1,2]	Yes	Yes	No	ReLU
12	upsample	64	[3,3]	[1,1]	Yes	Yes	No	ReLU
13	output	73	[3,3]	[1,1]	Yes	No	No	TanH

TABLE III  
DESCRIPTION OF THE PATCH-NET DEPLOYED AS DISCRIMINATOR NN.

Block	$N_{\text{filter}}$	Filter size	Stride	Padding	BatchNorm	Activation
downsample	128	[3,3]	[1,1]	Yes	Yes	LeakyReLU
downsample	128	[2,4]	[1,2]	Yes	Yes	LeakyReLU
zero padding 2D	-	-	-	Yes	-	-
Conv2D	256	[3,3]	[1,1]	No	Yes	LeakyReLU
zero padding 2D	-	-	-	Yes	-	-
Conv2D	1	[3,3]	[1,1]	No	No	Linear

First, we define a UNN structure  $P$  that is used to derive the optimal weights for all the three UEs in Figure 1. The  $\mathcal{H}_{\text{mes}}^C$  for each UE has  $N_{\text{sp}} = 128$  time snapshots and we design a UNN capable to estimate chunks of 64 time snapshots. Therefore, we need to derive 2 sets of trainable parameters  $\mathcal{K}^*$  for each UE. In total, there are 6 different UNN set of weights  $\mathcal{K}^*$  to estimate all the prior-CSIs. For the UNN mapping structure  $P$ , we choose to have four inner layers, each with  $k_{1:L-2} = 64$  filters and both upsampling matrices,  $\mathbf{A}_l$  and  $\mathbf{C}_l$ , activated in all inner layers. In addition, we use one pre-output layer with  $k_{L-1} = 64$  filters and one output layer with  $k_L = 72$  convolutional filters. Therefore, there are  $L = 6$  layers and the random noise seed

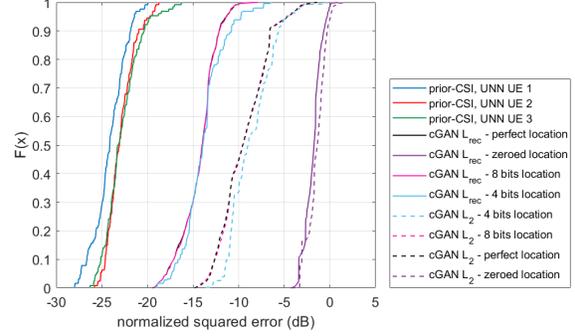


Fig. 6. Cumulative distribution function of the normalized squared error for UNN and cGAN results for CSI recreation. The cGAN trained with  $\mathcal{L}_{\text{rec}}$  ( $L_{\text{rec}}$ ) has best performance.

$\mathcal{Z}_0 \in \mathbb{R}^{4 \times 4 \times 64}$  is drawn from a uniform distribution as  $U(-0.15, +0.15)$  where  $k_0 = 64$ . After setting the UNN structure, the trainable parameters  $\mathcal{K}$  are initialized from random values and  $I = 25000$  gradient updates are performed to find the best  $\mathcal{K}^*$  for each UE, separately. Our design choices for  $k$  and  $I$ , the number of filters in each layer and the number of iterations respectively, were taken according to the estimated SNR at the UNN output. The UNN output SNR had to be at least the same as the SNR of the measured channel. Figure 6 shows the cumulative distribution function (CDF) of the NSE for the UNN-estimator in blue, red and green curves for the 3 UEs with 128 time snapshots each, where  $\text{SNR} = 20$  dB for the measured channel. The presented UNN architecture is capable to recreate a sequence of 64 time snapshots collected from a single UE. This structure contains 25,728 trainable parameters which correspond to 17.45% of the coefficients in a channel measurement  $\mathcal{H}_{\text{mes}}^C \in \mathbb{C}^{64 \times 64 \times 36}$ . Therefore, UNNs provide means to compress the CSI with a high probability of about 90% to recover it with better SNR than the measured one.

After estimating the channels for the three UEs, we have a pool of 384 prior-CSIs together with their location  $\Gamma_p$ . Here, we use the location coordinates provided by

the IlmProp simulator. In a practical implementation, we could derive the location from the estimated prior-CSIs by Unitary Tensor ESPRIT [12], for instance. A subset of  $S = 3$  CSI-location pairs are selected according to their minimum Euclidean distance to the location  $\Gamma_r$  where we aim to recreate the CSI. Hence, the input to our cGAN is  $\mathcal{H}_c \in \mathbb{R}^{4 \times 64 \times 73}$ , where 64 is the number of subcarriers. The architecture details of our generator and discriminator are presented in Table II and Table III, respectively. The adversarial training runs for 150 epochs with 60% of the simulated channels used for training and 40% for testing. Figure 6 presents the CDF curves of the NSE of the channels estimated by the cGAN with prior-knowledge and desired location as input. The black curves are the NSE for the cGANs trained with  $\mathbf{H}_{\text{LOC}}$  with infinity precision, where the full line and dashed line are for the cGANs trained with  $\mathcal{L}_{\text{rec}}$  and  $\mathcal{L}_{L_2}$  as regularization term, respectively. As we can see in Figure 6, the cGAN constrained to  $\mathcal{L}_{\text{rec}}$  achieves better performance with a 90% NSE of about  $-12.5$  dB.

Regarding state of the art, our approach provides much higher performance if compared to the 6 dB reported in [3]. Moreover, our cGAN architecture is less complex since we just use 13 layers for the U-net at the generator while [3] has reported 28 layers to process images of the environment map and output the wireless channel. The generator described in Table II has a total of 789,786 trainable parameters. Our cGAN takes about 6 hours to train in a computer with 16 GB of RAM and a GPU with 2 GB of dedicated memory.

At inference time, we test the sensitivity of the trained cGAN to a shut-down of  $\mathbf{H}_{\text{LOC}}$ . We set  $\mathbf{H}_{\text{LOC}} = 0$  and compute the NSE at the cGAN output. This sensitivity curve is plotted in purple at Figure 6. The cGAN performance decays to nearly 0 dB, which inform us that the architecture is not neglecting the location matrix. Next, we test the cGAN sensitivity to location

quantization errors. The locations are quantized using a uniform quantizer with 8 and 4 bits, their NSE are plotted in pink and cyan at Figure 6. There is very little performance loss even for 4 bits location quantization, which introduces errors in the range of 1 to 4 times the channel wavelength. The cGAN is robust to those errors because it is closely modeling the channels statistical distribution. Our analyzes of the dataset has shown that the channels are mostly line of sight (LOS) and frequency-flat for the defined study area. Therefore, the good statistical behavior of the dataset makes the CSI-recreation task quite easy for the cGAN. Such robustness against location errors may not be experienced in non-line of sight scenarios.

## VII. CONCLUSION

In this paper we propose to combine UNNs with cGAN to reconstruct wireless channels within an area in the propagation environment. The channel is reconstructed based on prior-CSIs from UNNs and the location where we aim to reconstruct the channel. Our method performs better than state of the art solutions, is much less complex and requires only some hours of training. Our results show that the cGAN performs well in a propagation scenario with mostly LOS channels. Future work is related to further evaluation of quantization effects as well as advanced methods to improve the reconstructed CSI beyond this first ML inference.

## ACKNOWLEDGEMENT

This research was partly funded by German Ministry of Education and Research (BMBF) under grant 16KIS1184 (FunKI).

## REFERENCES

- [1] 3GPP, "Advanced plans for 5G;" July 2021. [Online]. Available: [https://www.3gpp.org/news-events/2210-advanced\\_{\\_}{5g}](https://www.3gpp.org/news-events/2210-advanced_{_}{5g})

- [2] M. Dias, A. Klautau, N. González-Prelcic, and R. W. Heath, "Position and lidar-aided mmwave beam selection using deep learning," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2019, pp. 1–5.
- [3] V. V. Ratnam, H. Chen, S. Pawar, B. Zhang, C. J. Zhang, Y.-J. Kim, S. Lee, M. Cho, and S.-R. Yoon, "Fadenet: Deep learning-based mm-wave large-scale channel fading prediction and its applications," *IEEE Access*, vol. 9, pp. 3278–3290, 2021.
- [4] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9446–9454.
- [5] R. Heckel and P. Hand, "Deep decoder: Concise image representations from untrained non-convolutional networks," in *International Conference on Learning Representations*, 2019. [Online]. Available: <https://openreview.net/forum?id=rylV-2C9KQ>
- [6] E. Balevi, A. Doshi, and J. G. Andrews, "Massive MIMO channel estimation with an untrained deep neural network," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2079–2090, 2020.
- [7] B. V. Boas, W. Zirwas, and M. Haardt, "Two-step machine learning approach for channel estimation with mixed resolution rf chains," in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2021, pp. 1–6.
- [8] G. Del Galdo, M. Haardt, and C. Schneider, "Geometry-based channel modelling of MIMO channels in comparison with channel sounder measurements," *Advances in Radio Science*, vol. 2, no. BC, pp. 117–126, 2005.
- [9] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. PMLR, 2015, pp. 448–456.
- [10] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- [11] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2536–2544.
- [12] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-order SVD-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3198–3213, 2008.