Benchmarking Machine Learning Techniques for THz Channel Estimation Problems

Mounir Bensalem and Admela Jukan
Technische Universität Braunschweig, Germany
{mounir.bensalem, a.jukan}@tu-bs.de

Abstract—**Terahertz communication is one of the most promising wireless communication technologies for 6G generation and beyond. For THz systems to be practically adopted, channel estimation is one of the key issues. We consider the problem of channel modeling and estimation with deterministic channel propagation and the related physical characteristics of THz bands, and benchmark various machine learning algorithms to estimate THz channel, including neural networks (NN), logistic regression (LR), and projected gradient ascent (PGA). Numerical results show that PGA algorithm yields the most promising performance at SNR=0 dB with NMSE of -12.8 dB.**

I. INTRODUCTION

THz communication systems are positioned as the main driver in the evolution towards next generation wireless communications, also referred to as 6G [1]. In THz systems, operating in a frequency range 300 GHz - 10 THz and enabling unprecedented wireless transmission rates, several major physical limitations make the channel modeling and estimation fundamentally hard [2]. Recent work focused on developing systems and algorithms for estimating mmWave channels in MIMO systems [3], [4]. MIMO channel estimation is challenging due to the quantization of the linear combination of transmitted signals from multiple input antennas [5]. Using higher frequencies, and larger bandwidths needs high precision analog-to-digital-converters (ADCs), which motivates the utilization of low resolution ADCs for MIMO systems [3], and more practically 1-bit ADCs [6].

Today, significant efforts are underway to accurately model and estimate the channel with machine learning (ML). The trend of applying machine learning algorithms in order to estimate mmWave and furthermore THZ channels is rather notable in fact, also considering low resolution ADCs [3], [6]–[8]. [8] proposed a deep convolutional neural network (DCNN)-based spherical-wave CE algorithms for THz systems, the THz channel is simulated using a channel model developed for mmWave systems and adapting the frequency parameters to 0.6 THz. We refer to a some ML algorithms studied in [6] for MIMO systems, such as projected gradient ascent (PGA) and Franke-Wolfe techniques, to benchmark their performance in a THz scenario alongside with basic neural network estimator, due to their good performance for a carrier frequency of 28 GHz. Overall, ML-based channel estimation problem in THz communication systems is still a rather open research direction, with experimental setups and analytical models still few and far between.

In this paper, we benchmark different machine learning algorithms used for THz channel estimation. Our reference model is THz-MIMO channel with 1-bit ADCs, as proposed in [9] and [10] due to the limited quantization level of ultra-high speed low power ADCs. We model the THz channel using a deterministic channel propagation model from the literature [11], considering the existence molecular absorption loss and showing that data rates around 1 Tbps can be obtained with a short distance of around 1 meter. Our model advances the basic model of THz band proposed in [12], which uses an updated molecular High Resolution Transmission (HITRAN) 2012 database. We adopt an indoor scenario with reflecting walls in order to consider the line-of-sight (LoS) and several non-line-of-sight (NLoS) reflected rays. We apply the analytical model to obtain the channel state information. Using a pilot generation methods, Discrete Fourier Transform (DFT) and Zadoff-Chu (ZC) based sequence training, we generate a training dataset of pilots input and signal response at the receiver. Based on this reference model, we apply different machine learning algorithms for channel estimation, including neural networks (NN), logistic regression (LR), and PGA We chose these specific algorithms as they proved good performance for similar problems, with complexity similar to the complexity of gradient descent learning Numerical results, including analysis and simulations, show that PGA algorithm yields the most promising performance at SNR=0 dB with NMSE of -12.8 dB.

The rest of the paper is organize as follows: Section II describes the channel and system model. Section III describes the studied ML algorithms for Channel Estimation. Section IV shows the simulation results. We conclude the paper in Section V.

II. CHANNEL AND SYSTEM MODEL

Notation: $\mathbf{H}$ is a complex matrix and H denotes a real matrix. $\mathbf{H}^T$ and $\mathbf{H}^*$ represent the transpose and conjugate transpose of $\mathbf{H}$, respectively. Re($\mathbf{H}$), Im($\mathbf{H}$) are the real and imaginary part of $\mathbf{H}$. $\|\mathbf{H}\|$ is the Frobenius norm of $\mathbf{H}$. $\mathbb{1}_{[.]}$ is the indicator function.

A. THz Propagation Model

We consider a single transmitter (Tx) and a single receiver (Rx) THz transmission system, using a point-to-point line-of-sight communication scenario with omnidirectional antennas

[12], separated by a distance d. The main difference between Thz bands and other frequency bands appears from the molecular absorption loss, which varies with the signal frequency, the transmission distance and the concentration and the mixture of molecules coming across the path.

We consider the medium loss due to the vibration changes of molecules, which is denoted as the absorption loss $L_{abs}(f, d)$. This loss accurs due the transformation of a part of the wave energy into a kinetic energy, and it depends on the operating frequency the distance between the sender and receiver, and the composition of the meduim. The equation is obtained from [12] as follows:

$$L_{abs}(f, d) = e^{k(f) \cdot d} = k(f) \cdot d \cdot 10 \cdot log(e) \quad (dB) \quad (1)$$

Where f is the operating frequency, and $k(f)$ is is the overall absorption coefficient of the medium available from HITRAN database [13], which includes HITRAN line-transition parameters and cross-sections, that can be used to model signal transmission.

Similar to [12], we assume that the spreading loss, which represents the attenuation due to the expansion of a wave that propagates through the medium, is given by:

$$L_{spread}(f, d) = \left(\frac{4\pi f d}{c}\right)^2 = 20 \cdot log\left(\frac{4\pi f d}{c}\right) \quad (dB) \quad (2)$$

Where c is the speed of light in the vacuum. The received signal power spectral densit (psd) in THz band is given by:

$$P_{Rx} = \frac{P_{Tx}}{L_{abs} \cdot L_{spread}}$$
$$= P_{Tx} \cdot C \cdot f^{-2} \cdot d^{-2} e^{-k(f)d}, \quad C = \frac{c^2}{16\pi^2} \quad (3)$$

Where P_{Tx} represents the transmitted signal psd.

Molecular absorption introduces noise to the transmitted signal, which needs to be considered to evaluate the signal to noise ratio (SNR). The molecular noise psd is given by [12]:

$$P_M = \frac{P_{Tx}(f)}{L_{spread}(f, d)}(1 - e^{-k(f) \cdot d}) \quad (4)$$

Additionally, we consider the Johnson-Nyquist (JN) noise generated by thermal agitation of electrons in conductors. The psd of JN noise is given by [2]:

$$P_{JN} = \frac{hf}{(\exp(\frac{hf}{k_B T}) - 1)} \quad (5)$$

Where k_B denotes the Boltzmann constant and T denotes the temperature in Kelvin.

In order to simplify our analysis, we approximate eq. (5) using series expansion, then we keep the first order term as follows:

$$P_{JN} = \frac{hf}{(\exp(\frac{hf}{k_B T}) - 1)}$$
$$= \frac{hf}{\frac{hf}{k_B T} + \left(\frac{hf}{k_B T}\right)^2 \cdot \frac{1}{2!} + \left(\frac{hf}{k_B T}\right)^3 \cdot \frac{1}{3!} + ..} \quad (6)$$
$$\approx k_B T, \quad 1^{st} \text{ order approx.}$$

Using eq. (4) and (6), the total noise psd is given by:

$$P_N = k_B T + P_{Tx} \cdot C \cdot f^{-2} \cdot d^{-2}(1 - e^{-k(f)d}) \quad (7)$$

Therefore, the signal-to-noise ratio (SNR) at the receiver is represented as :

$$SNR = \frac{P_{Rx}}{P_N} \quad (8)$$

B. System Model

We consider a narrowband THz-MIMO system with M_t antennas at the transmitter, M_r antennas at the receiver. Our studied system is depicted in Fig. 1. We assume that the number of antennas at both transmitter and receiver are equal. The arrival paths are considered to include two main components of a THz channel: the LoS and several NLoS reflected rays caused by reflection. We assume that the scattering and diffraction can be neglected due to their negligible effect on propagation in the THz band.

Let $\mathbf{H} \in \mathbb{C}^{M_t \times M_r}$, denotes the narrowband channel matrix of THz channel impulse response. The channel matrix is expressed as follows [11]:

$$\mathbf{H} = \alpha_L G_t G_r a_r(\theta^r, \psi^r) a_t(\theta^t, \psi^t)$$
$$+ \sum_{i=1}^{N_{clu}} \sum_{l=1}^{L_{ray}^i} \alpha_{i,l}^{NL} G_t G_r a_r(\theta_{i,l}^r, \psi_{i,l}^r) a_t(\theta_{i,l}^t, \psi_{i,l}^t) P_r(f, \tau_{i,l})$$
$$(9)$$

Where α_L is the complex gains of the LOS ray component, given by $|\alpha_L|^2 = L_{spread}(f, d) L_{abs}(f, d)$. $\alpha_{i,l}^{NL}$ is the complex gains of the NLOS ray component, given by $|\alpha_{i,l}^{NL}|^2 = \mathcal{F}_{i,l}(f) L_{spread}(f, d_1 + d_2) L_{abs}(f, d_1 + d_2)$, with $\mathcal{F}_{i,l}$ is the Fresnel reflection coefficient, d_1 is the distance between the transmitter and the reflector, and d_2 is the distance between the reflector and the receiver [14]. $P_r(\tau_{i,l})$ is the pulse-shaping function at time delay $\tau_{i,l}$. G_t and G_r are the associated transmit and receive antenna gains, respectively, $a_t(.)$ and $a_r(.)$ are the antenna array response vectors at the transmitter and receiver, θ^t, ψ^t, θ^r and ψ^r are the azimuth AoD, elevation AoD, azimuth AoA, and elevation AoA of the LOS ray component, $\theta_{i,l}^t$, $\psi_{i,l}^t$, $\theta_{i,l}^r$ and $\psi_{i,l}^r$ are the corresponding parameters for the lth NLOS ray in the ith cluster.

At the receiver, after obtaining the signal the FFT is computed. Let $\mathbf{r} \in \mathbb{C}^{M_r \times 1}$ denote the unquantized received signal, and given as follows:

$$\mathbf{r} = \mathbf{H}\mathbf{x} + \mathbf{z}, \quad (10)$$

where $\mathbf{z} \in \mathbb{C}^{M_r \times 1}$ is the additive white gaussian noise (AWGN) matrix, whose samples are independent and identically distributed (i.i.d.) $\mathbf{z} \sim \mathcal{CN}(0, N_0)$, and N_0 is the noise power.

Definition 1. *An ADC is a device that applies a binary operation on a complex space by mapping its elements into binary numbers. We define the following mappings:*

- *1-bit ADC mapping:*

$$sign : \mathbb{R} \longrightarrow \mathbb{Z}$$
$$x \longmapsto \begin{cases} 1 & if \ x \geq 0 \\ -1 & otherwise \end{cases}, \quad (11)$$
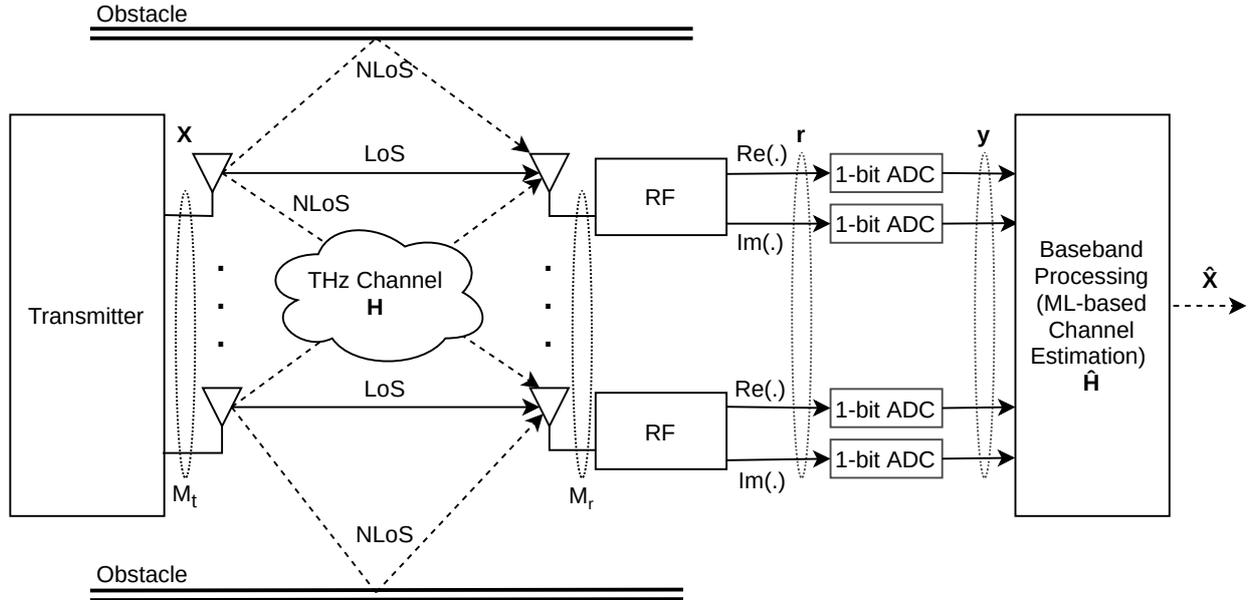
Fig. 1: THz-MIMO with M_t transmit antennas, M_r receiver antennas. Two 1-bit ADCs are used to equalize each received signal.

- *Vector ADC mapping:*
$$Sign : \mathbb{R}^n \longrightarrow \mathbb{Z}^n$$
$$(x_1, ..., x_n) \longmapsto (sign(x_1), ..., sign(x_n)), \qquad (12)$$

- *Matrix ADC mapping:*
$$Sign : \mathbb{R}^{n \times m} \longrightarrow \mathbb{Z}^{n \times m}$$
$$\{x_{i,j}\}_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}} \longmapsto \{Sign(x_{i,j})\}_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}}, \qquad (13)$$

- *Element-wise ADC quantization operator for channel response:*
$$\widehat{g} : \mathbb{C}^{n \times m} \longrightarrow \mathbb{Z}[j]^{n \times m}$$
$$a + jb \longmapsto Sign(a) + j\, Sign(b) \qquad (14)$$

We equalize the analog received signal to obtain a binary output, defined using the previous mappings as follows:
$$\mathbf{y} = \widehat{g}(\mathbf{r}) = Sign(\mathrm{Re}(\mathbf{r})) + j\, Sign(\mathrm{Im}(\mathbf{r})) \qquad (15)$$

III. ML ALGORITHMS FOR CHANNEL ESTIMATION

The aquisition of accurate CSI is crucial for THz communication systems. Channel estimation in THz systems is a challenging task due to the tiny wavelength of THz signals, the hardware limitations eg. high speed ADCs, and difficulty of estimating large amount of parameters with small size measurements. Moreover, channel estimation needs a minimum SNR value to ensure the reliability of measurments, while the bandwidth in THz systems is in the order of gigahertz, which presents a high thermal noise. As a results, traditional channel estimation techniques, which are based on approximative channel models may not be sufficant to provide an accurate channel parameters estimation. In this section we define the problem of channel estimation, and present different solutions and various ML based techniques, which learn channel parameters over any type of channel without prior assumptions.

A. Problem Definition

The goal of CE problem is the estimation of channel response parameters $\mathbf{H}$ from the received noisy symbols $\mathbf{y}_n, \forall n \in \mathcal{N}$, given perfect knowledge of the sent pilots $\mathbf{x}_n, \forall n \in \mathcal{N}$, where $\mathcal{N}$ defines the set of indexes of the training blocks, $N = |\mathcal{N}|$ is the training block size assumed to be equal to M_t. We convert the complex definition of parameters into real definition, as follows:

$$y_n = [\mathrm{Re}(\mathbf{y}_n), \mathrm{Im}(\mathbf{y}_n)] \qquad (16)$$

$$H = [\mathrm{Re}(\mathbf{H}), \mathrm{Im}(\mathbf{H})] \qquad (17)$$

$$x_n = \begin{bmatrix} \mathrm{Re}(\mathbf{x}_n) & \mathrm{Im}(\mathbf{x}_n) \\ -\mathrm{Im}(\mathbf{x}_n) & \mathrm{Re}(\mathbf{x}_n) \end{bmatrix} \qquad (18)$$

$$z = [\mathrm{Re}(\mathbf{z}), \mathrm{Im}(\mathbf{z})] \qquad (19)$$

Let y_n^e be the expected equalized binary output of received signal. In order to study a variety of ML algorithms, we define more than one loss function for CE problem, considering models from the literature and proposed variations as follows:

- The non-linear least squares minimization: it aims at minimizing the error between the expected output and the calculated one.
$$[\widehat{H}, \widehat{z}] = \arg\min_{H,z} \|y_n^e - y_n\|_2,$$
$$y_n = f(Hx_n + z), \forall n \in \mathcal{N} \qquad (20)$$

  - with linear activation function [15]: In this case, we consider $f(.)$ to be the identity function, which is equivalent in neural network terminology to a perceptron with linear activation function.
  - with tanh activation function: In this case, $f(x) = tanh(x)$ is an activation function defined on $\mathbb{R} \rightarrow$

$[-1, 1]$, which gives an output close to the $Sign$ function when the input is absolutely high, while being differentiable. The differentiability of the function is important for any neural network algorithm in order to calculate the derivative.

- The log-likelihood minimization [6], [16]: is designed to estimate a transformed varaible $\mathbf{X} = X^{Re} + jX^{Im} \in \mathbb{C}^{M_t \times M_r}$ corresponding to the matrix $\mathbf{Hx}$. The noise component is removed from the optimization because it is considered to be constant.

$$[\widehat{X^{Re}}] = \arg \min_{X^{Re}} \sum_{i=1}^{M_t} \sum_{j=1}^{M_r} \mathbb{1}_{[y_{i,j}^{Re}=1]} log(\phi(X_{i,j}^{Re}/\sigma))$$
$$+ \mathbb{1}_{[y_{i,j}^{Re}=-1]} log(1 - \phi(X_{i,j}^{Re}/\sigma)) \tag{21}$$

where $\phi(.)$ is the cumulative distribution function of the standard normal random variable, and σ represents its standard deviation.

B. Learning based Approach

Before introducing the proposed benchmark solutions, it is essential to reformulate the minimization problem in eq. (20) and eq. (21) as a binary classification problem. First, we create a training dataset of N data points $\{(x_n, y_n), \forall n \in \mathcal{N}\}$, where each data point represents one pilot. The minimization problem in eq. (20) can be seen as a minimization of the $L2$-norm loss function. It is well known that this loss function can provide stable solutions unlike $L1$-norm, but it suffers from overfitting problem. To prevent this the coefficients from fitting perfectly to the pilot training data, we introduce a regularization term to the loss function called $L2$-regularization.

$$[\widehat{H}, \widehat{z}] = \arg \min_{H,z}[\mathcal{J}(H, z)],$$
$$\mathcal{J}(H, z) = \frac{1}{N} \sum_{n=1}^{N} \|y_n^e - \tanh(Hx_n + z)\|_2 \tag{22}$$
$$+ \lambda \|H\|_2$$

where λ denotes a regularization coefficient.

1) Processing Units: The basic processing unit is called artificial neuron and it computes additive and multiplicative operations over input variables, before applying an activation function. Fig. 2 illustrates both the architecture of a general artificial neuron and the instantiated version for CE.
Let $W = (w_1, w_2, ..., w_n)$ be the weight vector of the artificial neuron, $x = (x_1, x_2, ..., x_n)$ is an input variables, and a bias value b. The process of a neuron can be represented as follows:

$$y = f(\sum_{i=1}^{n} w_i x_i + b) \tag{23}$$

where y denotes the output, and $f(.) : \mathbb{R} \longrightarrow \mathbb{R}$ represents an activation function.

Commonly, $f(.)$ is a non-linear function, such as sigmoid, hyperbolic, and rectified linear function. In CE problem, the activation function of a neuron corresponds to the $sign(.)$
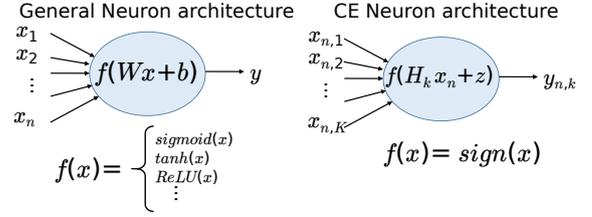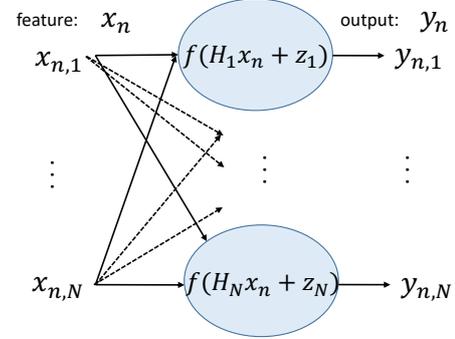

Fig. 2: Neuron architecture.

Fig. 3: NN-CE architecture.

function applied by ADC device, which fulfills the characteristics of an activation function.

2) Network Components: Generally, neurons are grouped into layers to form a multilayered network architecture, and based on the stacking configuration of processing units, we obtain different types of neural networks, such as fully-connected, convolutional, and recurrent networks. In our case, the CE problem is seen as a basic neural network architecture, which has one hidden layer with N neurons. Fig 3 shows the neural network based channel estimation NN-CE architecture, where the vectors $H_n = [h_{n,1}, ..., h_{nN}], \forall n \in \mathcal{N}$ represent the channel response vector for the receive antenna $n \in \mathcal{N}$. The vector $z = [z_1, ..., z_N]$ denotes the bias parameters of NN and it can be explained as the channel noise associated to each receive antenna $n \in \mathcal{N}$.

3) Training: After modeling the NN architecture, we need to define a loss function, in order evaluate and improve its results. The main goal of the training is to minimize the error loss function, while adjusting the weights and bias parameters, which is eventually the same goal of CE problem. The loss function was discribed in III-A.

Some optimization algorithm like Stochastic Gradient Descent (SGD) can be used to train the NN-CE, while minimizing the loss function. The optimization process updates gradually the weights representing the channel response parameters, and the bias representing the noise, in order to find the optimal solution:

$$h_{i,j} = h_{i,j} - \alpha \frac{\partial \mathcal{J}(H, z)}{\partial h_{i,j}}$$
$$z_i = z_i - \alpha \frac{\partial \mathcal{J}(H, z)}{\partial z_i} \tag{24}$$

where α is a parameter called the learning rate, which determines how much the model learns in each step. The back-

propagation algorithm is used to calculate both the weights and bias partial derivatives of the cost function.

The training steps of NN-CE can be described in two main steps: (i) the feed-forward pass, where the information cross all the network layers (one layer in our case) from the input layer to the output layer which contains the classification decision, i.e. the channel response, and (ii) the backpropagation pass, where the error generated by the NN-CE is calculated and propagated through the layers in the opposite direction, from the output layer to the input layer.

C. Benchmark Algorithms

Variety of ML algorithms can be tested for the CE problem, where the concept of iterative learning is preserved, while modifying some details in the way we update the CSI at each step. In the previous subsection, we described how NN is modeled and used for channel estimation. Similar to it we can use Logistic Regression (LR) while changing the loss function. In our paper, we consider two other learning algorithms: Projected Gradient Ascent (PGA) and Franke-Wolfe techniques, developed in [6] for low-rank mmWave MIMO channel, in order to check its performance on THz MIMO channel, which is extremely sensitive to weather condition, distance, and suffers from molecular attenuation. For PGA algorithm, the learning process is based on gradient ascent, a similar learning algorithm to gradient descent with a positive addition of the gradient at each update. PGA includes a projection step at each iteration, assuming that the matrix H has lower rank r than N, $r \ll N$, where $N = M_t = M_r$. Singular Value Decomposition (SVD) and simplex projection are used to find the closest low rank matrix to the updated estimation of H. For Franke-Wolfe algorithm, an additional step is included to the basic gradient ascent learning, where we update the channel matrix differently: instead of adding a weighted gradient matrix, we compute the top singular vector of the gradient, then we soustract it from the gradient and update the channel matrix

D. Pilot Generation Methods

In order to simulate the transmission process in our THz MIMO scenario, it is required to define a pilot generation method, regarding its role in TX-RX synchronization in wireless communication systems, and its importance in learning the CSI. Similar to [6] we adopt two methods from the literature: DFT and ZC based sequence training. DFT and ZC sequences can be pre-determined without resorting to the knowledge about the channel, which makes both methods suitable for our scenario.

IV. SIMULATION RESULTS

In the simulation, we consider a THz-MIMO system with $Mt = 16$, $M_r = 16$, and a frequency $f = 0.3$ THz. We simulate the THz channel using the proposed model in section II. The deterministic channel model proposed in [11] works in the range of frequencies between 0.1 and 1 THz, which explains our choice of the frequency $f = 0.3$ THz

in the simulation. We consider a dry air environment, where we adopt an updated molecular HighResolution Transmission (HITRAN) 2012 database. We consider the following gas composition of the air including their mixing ratios: N_2: 0.78, O_2: 0.21, CO_2: $365 \cdot 10^{-6}$, O_3: $10 \cdot 10^{-6}$, CH_4: $1.7 \cdot 10^{-6}$, H_2: $500 \cdot 10^{-9}$, N_2O: $320 \cdot 10^{-9}$, H_2O: 0.0096. We set the operating temperature equal to the standard temperature 296 K, and the pressure to 1 atm. We set the distance between the transmitter and the receiver $d = 1$ meter, where this value is set to be similar to the distance used in some experiments [17]. The azimuth AoA and AoD are generated from the uniform distribution in the range of $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. The CSI is generated using eq. (9), and the channel response at the receiver is measured as described in section II. The developed THz simulator will be documented and open source, along with the generated dataset and evaluated algorithms.

We train the proposed algorithms: Logistic Regression, NN, PGA, and Frank Wolf, with 10 channel matrices generated with a fixed seed. The number of epochs is set to 100, which ensures the convergences based on our simulations. The stopping criterion is set the same for all learning algorithms as $\epsilon = 10^{-10}$. The learning rate is intiialized with 0.01, and to $1/N_p$, where N_p is the number of pilot transmissions for PGA algorithm. The learning rate is decreasing by 0.7 for LR and NN, when the loss function decreases, and by 0.5 with other algorithms. The previous settings are empirically chosen. The pilots used for channel measurements are generated using ZC-based and the DFT-based training matrices. The performance of channel state information estimation algorithms are compared using the normalized mean squared error (NMSE) over all channel realizations, which is defined as the average of the Frobenius norm of $\|H - \beta \widehat{H}\|^2/\|H\|^2$, where β is given by $\beta = \|\widehat{H}^*H\|_1/\|\widehat{H}^*\widehat{H}\|_1$ [6].

Figure 4 shows the NMSE of the channel estimation algorithms with different number of pilots, where the SNR is set to 0 dB, and the frequency to 0.3 THz. Each algorithm is tested using two different pilot training algorithms ZC and DFT based training. We remark that for all algorithms the error decreases when we increase the number of pilot transmissions. The ZC-based training outperform DFT-based training for all simulations. PGA and Frank Wolf algorithms has almost the same performance with different number of pilot transmissions number, compared to LR and NN, which can be explained by the fact that those algorithms uses the knowledge about the low rank characteristics of the channel matrices, in order to reduce the dimentionality of the estimated channel information. Figure 5 shows that the PGA and Frank Wolf algorithms give better results than LR and NN, for different values of SNR. We can remark also that using DFT pilot generation is giving better results with NN than ZC method which is not the case with other algorithms. NN is using a basic network structure that can be improved in future work, regarding the potential of NN-based algorithms to learn complex structure. In THz transmission systems, it is challenging to achieve high SNR values, where SNR = 0 is considered as a good reachable value. Both PGA and Frank

Fig. 4: Performance of several learning algorithms with different pilot tranmissions number with $SNR = 0$ and frequency $f = 0.3\,THz$.

Fig. 5: Performance of several learning algorithms with different SNR values, with 240 pilot transmissions and frequency $f = 0.3\,THz$.

Wolf algorithms give an NMSE value around -12 dB, which is a promising result.

V. CONCLUSION

In this paper, we have modeled the channel estimation problem in THz-MIMO systems, considering a deterministic channel propagation model from the literature, and one-bit ADCs at the receiver. We have developed several machine learning algorithms to estimate the CSI, and evaluated two different pilot training techniques: DFT and ZC based training. Our results showed a comparison between few ML algorithms that are using gradient descent as a learning optimization. PGA algorithm has given the best performance at SNR=0 dB with NMSE of -12.8 dB. In a future work, we will compare the current algorithms to more ML and DL algorithms and other approaches from the literature, and consider studying the complexity and the hardware requirements.

REFERENCES

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6g wireless systems: Applications, trends, technologies, and open research problems," *IEEE network*, vol. 34, no. 3, pp. 134–142, 2019.

[2] V. Petrov, D. Moltchanov, and Y. Koucheryavy, "Interference and sinr in dense terahertz networks," in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*. IEEE, 2015, pp. 1–5.

[3] J. Mo, P. Schniter, and R. W. Heath, "Channel estimation in broadband millimeter wave mimo systems with few-bit adcs," *IEEE Transactions on Signal Processing*, vol. 66, no. 5, pp. 1141–1154, 2017.

[4] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmwave massive mimo systems," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 852–855, 2018.

[5] J. Zhang, L. Dai, X. Li, Y. Liu, and L. Hanzo, "On low-resolution adcs in practical 5g millimeter-wave massive mimo systems," *IEEE Communications Magazine*, vol. 56, no. 7, pp. 205–211, 2018.

[6] N. J. Myers, K. N. Tran, and R. W. Heath, "Low-rank mmwave mimo channel estimation in one-bit receivers," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5005–5009.

[7] M. Y. Takeda, A. Klautau, A. Mezghani, and R. W. Heath, "Mimo channel estimation with non-ideal adcs: Deep learning versus gamp," in *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2019, pp. 1–6.

[8] Y. Chen and C. Han, "Deep cnn-based spherical-wave channel estimation for terahertz ultra-massive mimo systems," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*. IEEE, 2020, pp. 1–6.

[9] A. Lozano, "1-bit mimo for terahertz channels," *arXiv preprint arXiv:2109.04390*, 2021.

[10] M. Saad, A. C. Al Ghouwaye, H. Hijazi, F. Bader, and J. Palicot, "Mimo techniques for wireless terabits systems under sub-thz channel with rf impairments," in *2020 IEEE ICC Workshops*. IEEE, 2020, pp. 1–6.

[11] A. Moldovan, M. A. Ruder, I. F. Akyildiz, and W. H. Gerstacker, "Los and nlos channel modeling for terahertz wireless communication with scattered rays," in *2014 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2014, pp. 388–392.

[12] J. M. Jornet and I. F. Akyildiz, "Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band," *IEEE Transactions on Wireless Communications*, vol. 10, no. 10, pp. 3211–3221, 2011.

[13] L. S. Rothman, I. E. Gordon, A. Barbe, D. C. Benner, P. F. Bernath, M. Birk, V. Boudon, L. R. Brown, A. Campargue, J.-P. Champion *et al.*, "The hitran 2008 molecular spectroscopic database," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 110, no. 9-10, pp. 533–572, 2009.

[14] C. Lin and G. Y. Li, "Indoor terahertz communications: How many antenna arrays are needed?" *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3097–3107, 2015.

[15] C.-J. Chun, J.-M. Kang, and I.-M. Kim, "Deep learning-based channel estimation for massive mimo systems," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1228–1231, 2019.

[16] J. Choi, J. Mo, and R. W. Heath, "Near maximum-likelihood detector and channel estimator for uplink multiuser massive mimo systems with one-bit adcs," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 2005–2018, 2016.

[17] V. Petrov, J. M. Eckhardt, D. Moltchanov, Y. Koucheryavy, and T. Kurner, "Measurements of reflection and penetration losses in low terahertz band vehicular communications," in *2020 14th European Conferenc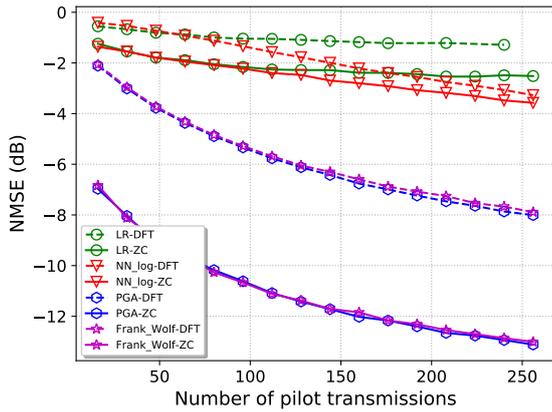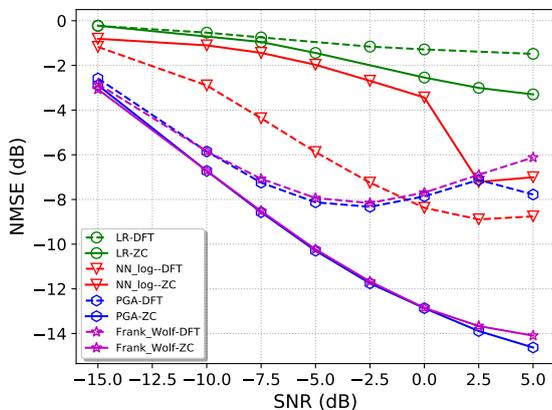e on Antennas and Propagation (EuCAP)*. IEEE, 2020, pp. 1–5.