

Deep Learning Model-Based Channel Estimation for THz Band Massive MIMO with RF Impairments

Pulok Tarafder¹, Imtiaz Ahmed¹, Danda B. Rawat¹, Ramesh Annavajjala², and Kumar Vijay Mishra³

¹Howard University, Washington, DC, USA

²University of Massachusetts, Boston, MA, USA

³United States DEVCOM Army Research Laboratory, Adelphi, MD, USA

Abstract—THz band enabled large scale massive MIMO (M-MIMO) is considered as a key enabler for the 6G technology, given its enormous bandwidth and for its low latency connectivity. In the large-scale M-MIMO configuration, enlarged array aperture and small wavelengths of THz results in an amalgamation of both far field and near field paths, which makes tasks such as channel estimation for THz M-MIMO highly challenging. Moreover, at the THz transceiver, radio frequency (RF) impairments such as phase noise (PN) of the analog devices also leads to degradation in channel estimation performance. Classical estimators as well as traditional deep learning (DL) based algorithms struggle to maintain their robustness when performing for large scale antenna arrays i.e., M-MIMO, and when RF impairments are considered for practical usage. To effectively address this issue, it is crucial to utilize a neural network (NN) that has the ability to study the behaviors of the channel and RF impairment correlations, such as a recurrent neural network (RNN). The RF impairments act as sequential noise data which is subsequently incorporated with the channel data, leading to choose a specific type of RNN known as bidirectional long short-term memory (BiLSTM) which is followed by gated recurrent units (GRU) to process the sequential data. Simulation results demonstrate that our proposed model outperforms other benchmark approaches at various signal-to-noise ratio (SNR) levels.

Index Terms—Tera Hertz Communication, 6G Wireless, Phase Noise, Hybrid-field channel, channel estimation, Ultra Massive MIMO

I. INTRODUCTION

With the recent advancements in sixth-generation (6G) communication, it has been identified that spectrum bottleneck is a key limitation of enabling increased data rates [1]. To facilitate the overgrowing data demands, the new terahertz (THz) band from 0.1 to 10 THz has been identified as the key enabler to ensure extreme data rates [2], [3]. However, one of the core challenges in THz band is its high attenuation, which is caused by high spreading loss and molecular absorption [4]. In response, massive multiple-input multiple-output (M-MIMO) technology has emerged as a promising solution, leveraging the simultaneous utilization

of hundreds of antennas at the base station to overcome the challenges such as path loss and blockage [5]. Moreover, compared to millimeter-wave (mmWave) frequencies, THz channels are extremely sparse and typically modeled as line-of-sight (LoS)-dominant [6]. THz band occupies a spectral domain that bridges the mmWave frequencies and the infrared band, and it presents a unique challenge for signal generation across both electronic and photonic devices, resulting in radio-frequency (RF) impairments [7]–[10].

Electromagnetic (EM) radiation field can typically be categorized into far-field and near-field regions, and it is expected that both of these components will be exploited in future 6G mobile networks. As a result, in the THz enabled M-MIMO systems literature, there are typically two categories of low-overhead channel estimation schemes are present, i.e., far-field channel estimation [11]–[13] and near-field channel estimation [14]. For far-field channel estimation, the channel sparsity is considered in the angle domain, where signals can only be pointed towards a specific direction [15]. Whereas, near-field channel estimation considers aperture arrays will experience spherical wavefronts [16], and the channel sparsity is in polar domain.

With a few exceptions where deep learning (DL) methodologies were used, most studies in the literature that addressed far-field, near-field, or hybrid-field ultra M-MIMO configurations [17] focused on traditional channel estimating techniques. In [13], an orthogonal matching pursuit (OMP) is used to estimate the far-field angular-domain channel. In [14], Han *et al.* exploited dedicated sparsity patterns and OMP for near-field estimation. Cui *et al.* in [18] proposed a polar-domain simultaneous OMP (PSOMP) channel estimator for near-field communication. Furthermore, a hybrid-field OMP based channel estimator was proposed in [19]. A fixed point theory based DL assisted channel estimator is presented in [20] for a hybrid-field channel model. Another DL based channel estimation algorithm is presented in [21] while focusing only on near-field channel model. In [22], a novel near-field channel estimation algorithm is presented using residual dense networks.

To the best of our knowledge, no prior work on hybrid-field channel estimation techniques have addressed RF impairment (phase noise (PN)) incorporated intelligent DL aided channel estimation algorithm for THz ultra M-MIMO (UMMIMO)

This work is funded in part by the DoD/US Army under Contract W911NF-20-2-0277 via DoD Center of Excellence in AI/ML (CoE-AIML), National Science Foundation (NSF) Project under Award 2200640, and DoD/US Army Contract W911NF-22-1-022. However, any opinion, finding, and conclusions or recommendations expressed in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the funding agencies.

communication systems. In this paper, owing to the inherent challenges constituted by the PN and the small wavelengths of high frequency THz-band, we propose a DL assisted intelligent channel estimation architecture for hybrid-field THz band UMMIMO architecture. Although our proposed scheme is constructed for THz band, the scheme can also be utilized in very high frequency mmWave band. Our proposed method captures the dynamic behaviors of PN in the hybrid-field channel model and accurately estimates the channel utilizing the powerful capabilities of bidirectional long short-term memory (BiLSTM) and gated recurrent units (GRU) combined. BiLSTM exploits the advantage of the sequential prediction tasks combined with GRU to accelerate the training process.

The remainder of the paper is organized as follows. Section II describes the THz-band system model. Section III discusses the proposed DL framework. Afterwards, section IV presents the performance analysis of the proposed framework, and some concluding remarks are summarized in section V.

II. SYSTEM MODEL

Within the framework of UMMIMO systems functioning in the THz-band, we consider a base station (BS) constructed with an ultra-large-scale array of N antenna components. The objective of using a large antenna array at BS is to facilitate downlink communication with a single-antenna user equipment (UE). Let $\hat{\mathbf{H}} \in \mathbb{C}^{1 \times N}$ denote the hybrid channel characterizing the propagation environment between the BS and the UE. To model the downlink channel estimation process, we consider the following signal representation:

$$\mathbf{y} = \hat{\mathbf{H}}\Phi + \mathbf{w}, \quad (1)$$

where $\mathbf{y} \in \mathbb{C}^{1 \times M}$ represents the received signal observations at the UE over M pilot vectors, $\Phi \in \mathbb{C}^{N \times M}$ denotes the known pilot matrix transmitted by the BS, and $\mathbf{w} \in \mathbb{C}^{1 \times M}$ denotes independent and identically distributed (i.i.d.) complex additive white Gaussian noise (AWGN) with zero mean and variance σ^2 . Note that (1) does not incorporate PN in the considered model. We leverage PN models for both transmitter and receiver and update the system model at the end of this section.

Given the knowledge of $\mathbf{y}$ and Φ , the downlink channel estimation problem entails the recovery of the channel vector $\hat{\mathbf{H}}$. In the context of UMMIMO systems, the number of antennas N at the BS is substantially large. In these systems, the spatial dynamics of signal transmission necessitate a comprehensive model that can simultaneously account for the different characteristics of far-field and near-field propagation. To this end, for our analysis, we utilize an advance channel model known as hybrid-field channel model [18].

A. Far-Field Component

The far-field component of the channel $\hat{\mathbf{H}}$ can be modeled as a sum of L_f plane waves impinging on the antenna

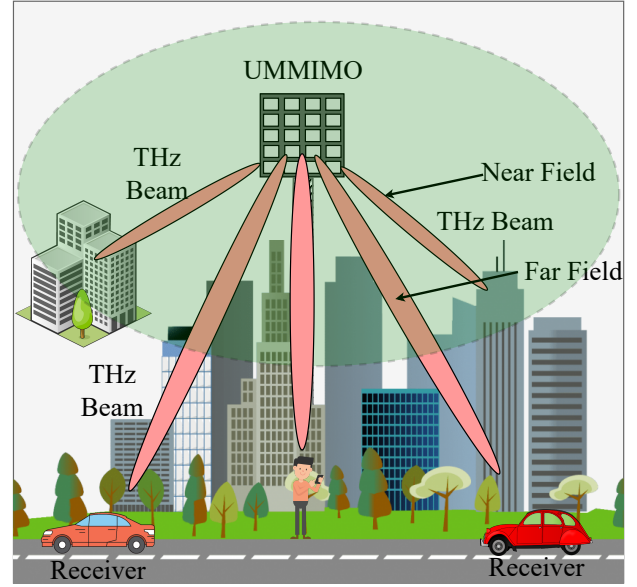


Fig. 1: THz Massive MIMO.

array from different angles. The far-field channel component, denoted by $\mathbf{H}_f \in \mathbb{C}^{1 \times N}$, can be expressed as:

$$\mathbf{H}_f = \sum_{l=1}^{L_f} \alpha_f^{(l)} \mathbf{a}_f(\theta_f^{(l)}), \quad (2)$$

where L_f is the number of far-field paths, $\alpha_f^{(l)} \in \mathbb{C}$ denotes the complex gain of the l^{th} far-field path, $\theta_f^{(l)}$ is the angle of arrival of the l^{th} far-field path, and $\mathbf{a}_f(\theta_f^{(l)}) \in \mathbb{C}^{1 \times N}$ is the far-field array steering vector, given by:

$$\mathbf{a}_f(\theta_f) = \frac{1}{\|\mathbf{a}'_f(\theta_f^{(l)})\|} \left[1, e^{j\pi \sin(\theta_f^{(l)})}, \dots, e^{j\pi(N-1) \sin(\theta_f^{(l)})} \right]^T. \quad (3)$$

B. Near-Field Component

The near-field component of the $\hat{\mathbf{H}}$ can be modeled as a sum of L_{nf} spherical wavefronts originating from different distances and angles. The near-field channel component, denoted by $\mathbf{H}_{nf} \in \mathbb{C}^{1 \times N}$, can be expressed as:

$$\mathbf{H}_{nf} = \sum_{l=1}^{L_{nf}} \alpha_{nf}^{(l)} \mathbf{a}_{nf}(\Delta_{nf}^{(l)}, \theta_{nf}^{(l)}), \quad (4)$$

where L_{nf} is the number of near-field paths, $\alpha_{nf}^{(l)}$ presents the complex gain of the l^{th} near-field path, $\Delta_{nf}^{(l)}$ denotes the distance of the l^{th} near-field path, $\theta_{nf}^{(l)}$ depicts the angle of

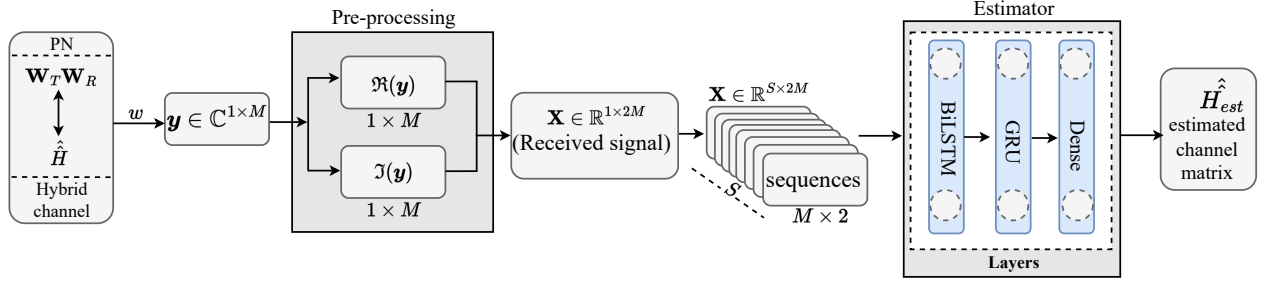


Fig. 2: Proposed deep learning model with BiLSTM-GRU.

arrival of the l^{th} near-field path, and $\mathbf{a}_{nf}(\Delta_{nf}^{(l)}, \theta_{nf}^{(l)})$ is the near-field array steering vector, given by:

$$\mathbf{a}_{nf}(\Delta_{nf}^{(l)}, \theta_{nf}^{(l)}) = \frac{1}{\Lambda} \left[\frac{\Delta_{nf}^{(l)}}{\xi_{nf}^{(1)}} e^{j \frac{2\pi F}{c} (\xi_{nf}^{(1)} - \Delta_{nf}^{(l)})}, \frac{\Delta_{nf}^{(l)}}{\xi_{nf}^{(2)}} e^{j \frac{2\pi F}{c} (\xi_{nf}^{(2)} - \Delta_{nf}^{(l)})}, \dots, \frac{\Delta_{nf}^{(l)}}{\xi_{nf}^{(N)}} e^{j \frac{2\pi F}{c} (\xi_{nf}^{(N)} - \Delta_{nf}^{(l)})} \right]^T. \quad (5)$$

Here $\Lambda = \|\mathbf{a}'_{nf}(\Delta_{nf}^{(l)}, \theta_{nf}^{(l)})\|$ and c is the speed of light. Moreover, F is the carrier frequency, $\Delta_{nf}^{(l)}$ denotes the distance from the l^{th} scatter to the center of the antenna array, and $\xi_{nf}^{(k)} \big|_{k=1}^N$ depicts the distance from the k^{th} antenna element to the corresponding scatter. Furthermore, $\xi_{nf}^{(k)} \big|_{k=1}^N$ can be expressed by:

$$\xi_{nf}^{(k)} \big|_{k=1}^N = \sqrt{(\Delta_{nf}^{(l)})^2 + (\Psi_k d)^2 - 2\Delta_{nf}^{(l)} \Psi_k d \sin(\theta_{nf}^{(l)})}, \quad (6)$$

where $\Psi_k = \frac{2k-N-1}{2}$ and d denotes the antenna spacing.

C. Hybrid Channel Model

In the hybrid-field model, the two propagation methodologies are distinguished by Rayleigh distance $\Omega = \frac{2D^2}{\lambda}$, where D is the antenna array aperture of radiation, and λ represents the wavelength. Note that Ω is the key parameter, which determines whether the channel acts as near field or far field. Near-field channel exists at a region $\Theta < \Omega$ whereas far-field exists at $\Theta \geq \Omega$. In particular,

$$\text{Channel Region}(\Theta) = \begin{cases} \text{Near-field,} & \text{if } \Theta < \Omega \\ \text{Far-field,} & \text{if } \Theta \geq \Omega \end{cases}.$$

The comprehensive hybrid channel model, encapsulated within $\hat{\mathbf{H}} \in \mathbb{C}^{1 \times N}$, emerges from the combination of the far-field and near-field components. This model is formulated through the equation:

$$\hat{\mathbf{H}} = \sum_{L_f=1}^{\gamma L} \mathbf{H}_f \big|_{\{\Theta \geq \Omega\}} + \sum_{L_{nf}=1}^{(1-\gamma)L} \mathbf{H}_{nf} \big|_{\{\Theta < \Omega\}}, \quad (7)$$

where $\gamma \in [0, 1]$ and L indicate the number of possible pathways. The crucial factor in determining the number

of far-field and near-field components in the propagation environment is γ . The considered hybrid-field channel model reduces to the conventional far-field channel model for $\gamma = 1$. On the other hand, it turns into a near-field channel model when $\gamma = 0$. In order to guarantee a consistent overall power, the hybrid channel is normalized as follows:

$$\hat{\mathbf{H}} = \hat{\mathbf{H}} \sqrt{\frac{N}{L_f + L_{nf}}}. \quad (8)$$

D. Phase Noise Incorporation

In the domain of high-frequency M-MIMO systems, the manifestation of accurately modeling RF impairments becomes complicated owing to PN being a predominant factor that significantly affects system performance. PN, originating from the inherent instability in the oscillator frequencies at both the transmitter (Tx) and receiver (Rx), manifests as random phase variations over time. This section delineates the mathematical framework adopted to integrate the PN into the hybrid channel model, thereby capturing the essence of real-world RF imperfections at THz-bands.

Let us denote θ_T and θ_R as the PN processes at the Tx and Rx, respectively. For each sample n , the Wiener process random-walk [23] PN for the Tx and Rx can be represented as:

$$\begin{aligned} \theta_T[n] &= \theta_T[n-1] + \Delta\theta_T[n] \text{ and} \\ \theta_R[n] &= \theta_R[n-1] + \Delta\theta_R[n]. \end{aligned} \quad (9)$$

Here, $\Delta\theta_T[n] \sim \mathcal{N}(0, \sigma_T^2)$ and $\Delta\theta_R[n] \sim \mathcal{N}(0, \sigma_R^2)$ represent changes in PN for the Tx and Rx, respectively. Each is normally distributed ($\mathcal{N}$) with a mean of 0 and variances σ_T^2 for the Tx and σ_R^2 for the Rx. The mathematical representation of the received signal, encompassing the influence of PN alongside impact on the channel is expressed as follows:

$$\mathbf{y} = e^{-j\theta_R} \left(\hat{\mathbf{H}} \Phi e^{j\theta_T} \right) + \mathbf{w}. \quad (10)$$

Here, $e^{j\theta_T}$ and $e^{-j\theta_R}$ represent the phase modulation and demodulation processes attributable to the Tx and Rx PNs, respectively. Note that, Eq. (10) can be further be simplified as:

$$\mathbf{y} = \mathbf{W}_T \mathbf{W}_R \hat{\mathbf{H}} \Phi + \mathbf{w}, \quad (11)$$

where, the symbol $\mathbf{y}$ represents the received signal, and $\mathbf{W}_T \in \mathbb{R}^{1 \times N}$ and $\mathbf{W}_R \in \mathbb{R}^{N \times 1}$ represent the Tx and Rx PN components, respectively.

III. PROPOSED LSTM-BASED PN CHANNEL ESTIMATION

This section describes a BiLSTM and GRU based channel estimation scheme for the considered system. Neural networks (NNs) excel at classification and recognition tasks where the output is a discrete label, and information loss within the model may not significantly impact performance. However, for data generation problems such as channel estimation, where the output is a continuous signal or waveform, information loss can lead to degraded performance [24]. In addition, the dynamic nature of PN [25] incorporated with the channel makes the channel estimation task more complicated. Conventional feed forward deep NN (DNN) based channel estimation methods do not have the capability to exploit the long term channel correlations. Techniques that prioritize information preservation are more suitable than conventional NN models optimized for classification or recognition. Our proposed BiLSTM-based approach leverages the sequential learning capabilities of LSTMs to effectively capture the temporal dependencies introduced by PN. The goal is to accurately estimate the combined effect of PN ($\mathbf{W}_T \mathbf{W}_R$) and the channel characteristics ($\hat{\mathbf{H}}$). This proposed scheme considers the influence of PN and channel properties as intertwined elements of a singular, comprehensive channel model. It is worth mentioning that the pilot signal matrix Φ is pre-designed and accurately known to both Tx and Rx.

A. Data Preprocessing

Before feeding the data into the LSTM model, a crucial preprocessing is performed for efficient training. Note that we consider a downlink transmission model with received signal processing on the UE side. At the UE, the complex received signals $\mathbf{y} \in \mathbb{C}^{1 \times M}$ are separated into its real and imaginary components at first, then we concatenated the real and imaginary parts to construct the input dataset for the proposed scheme. Given the received signal vector set denoted by $\mathbf{y} \in \mathbb{C}^{1 \times M}$, the separation and concatenation of the real and imaginary components can be mathematically represented as follows.

Let $\Re(\mathbf{y} \in \mathbb{C}^{1 \times M})$ denote the real part and $\Im(\mathbf{y} \in \mathbb{C}^{1 \times M})$ denote the imaginary part of the complex channel phenomena. The concatenated vector for each sample then becomes:

$$\mathbf{X} = [\Re(\mathbf{y} \in \mathbb{C}^{1 \times M}), \Im(\mathbf{y} \in \mathbb{C}^{1 \times M})] \in \mathbb{R}^{1 \times 2M}, \quad (12)$$

where $\mathbf{X}$ represents the new data structure after separation and concatenation of the real and imaginary components. For S number of training samples, this transformation is applied to each sample, resulting in the dimensionality transformation as follows:

$$\mathbf{y} \in \mathbb{C}^{1 \times M} \in \mathbb{C}^{S \times M} \Rightarrow \mathbf{X} \in \mathbb{R}^{S \times 2M}. \quad (13)$$

B. Proposed Model Architecture

In the construction of our proposed model as portrayed in Fig. 3, we design a DL network leveraging the robustness of BiLSTM and GRU, followed by a dense layer. LSTM is a promising denoiser in the context of channel estimation for UMMIMO at higher frequencies [26]. The hidden layers of the LSTM cell can capture the important information from the past and avoid the redundant information, thus providing a greater ability to capture the information compared to the simple RNN cell. The structure of the BiLSTM network is the combination of two LSTM networks with two different directions [27]. Therefore, the BiLSTM approach facilitates the processing of input data in both forward and reverse temporal directions, and enables a comprehensive analysis of the past and future states of the signal. This dual-directional processing capability makes BiLSTM exceptionally adept at capturing the dynamic fluctuations of PN inherent in highly noisy and attenuated THz communication environments. Afterwards, GRU compensates and retrieves the data lost during the denoising operation by BiLSTM.

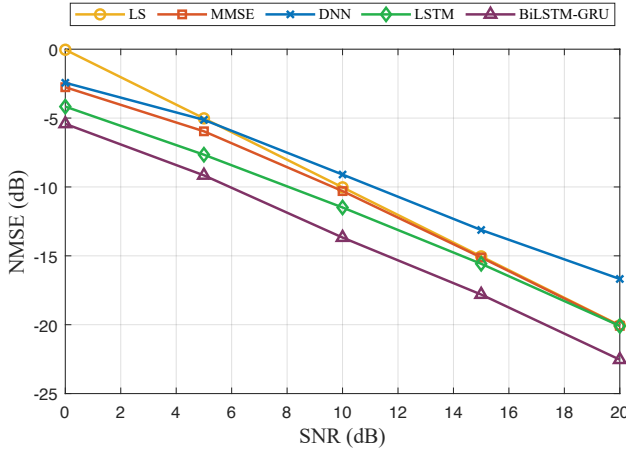
In the DL module, the model expects an input sequence of dimension $2N$. This input shape is reshaped to $2N \times 1$ for processing, implying that each sequence consists of $2N$ timesteps, with a single feature per timestep for the BiLSTM and GRU layers. The single feature at each timestep represents the concatenated real-valued signal vector for a particular Tx-Rx antenna pair. Both the BiLSTM and GRU layers are designed to process this input structure, each equipped with N LSTM and GRU cells respectively to ensure the consistent flow of information throughout the network. Finally, the dense layer with a single neuron and a linear activation function concludes the proposed DL network and maps the high-level learned features from the previous two layers and predicts a single continuous value from the sequence. It is worth noting that we trained our proposed model offline using comprehensive datasets, and subsequently evaluated its performance on live network conditions online.

IV. SIMULATION RESULTS

This section includes a comprehensive numerical study of our proposed DL framework and an assessment of its performance in relation to conventional channel estimation schemes. We create the training datasets for our simulation using Eq. (11). Additionally, Table I provides a summary of all the simulation parameters. In this work, we employ the normalized mean square error (NMSE) as the primary performance assessment metric to evaluate the accuracy of our channel estimation model at various signal-to-noise ratio (SNR) levels, where we define SNR as $1/\sigma^2$. The NMSE is used to quantify the deviation between the ground truth channel matrix $\hat{\mathbf{H}}$ and the estimated channel matrix $\hat{\mathbf{H}}_{est}$ obtained from our model. A lower NMSE value indicates a smaller deviation between the estimated channel matrix and the ground truth, signifying a more accurate channel

TABLE I: Adopted simulation parameters

Hyperparameter	Value
Linear signal-to-noise ratio (SNR)	$1/\sigma^2$
SNR (dB) range	[0 : 5 : 20]
BS antenna array N	64, 128
Total number of paths L	4
Hybrid-field path component γ	0.5
Sub-THz carrier frequency F	100 GHz
$\Delta_{nf}^{(l)} \min$	10 m
$\Delta_{nf}^{(l)} \max$	80 m
Speed of light c	$3 \times 10^8 \text{ ms}^{-1}$
Initial σ_T^2	0.1
Initial σ_R^2	0.2


 Fig. 3: NMSE vs. SNR (dB) for $N = 64$.

estimation. The NMSE estimate in decibel (dB) scale is calculated by

$$\text{NMSE} = 10 \log \left\{ \mathbb{E} \left[\frac{\|\hat{\mathbf{H}} - \hat{\mathbf{H}}_{est}\|^2}{\|\hat{\mathbf{H}}\|^2} \right] \right\}, \quad (14)$$

where $\mathbb{E}$ denotes the statistical expectation.

At first, using Eq. (11), we generate datasets for $N = 64$ and $N = 128$ separately, each having 6000 data samples, for various SNR levels i.e., 0, 5, 10, 15, and 20 dB. We then divide the generated dataset into 80% for training and 20% for testing. We utilize TensorFlow [28] backend Keras to execute the DL simulation. We train each dataset separately using Adam optimizer [29]. We tune the model with 0.001 learning rate and batch size 16, and set mean square error (MSE) as the loss function until convergence, exploiting the callbacks feature of TensorFlow. We then compare the performance of our proposed DL model based scheme with the conventional least square (LS) and minimum mean square error (MMSE) estimation schemes, DNN, and LSTM. Fig. 3 and 4

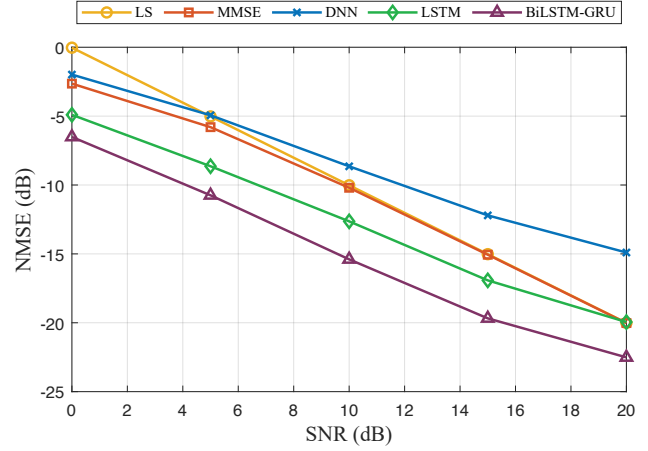
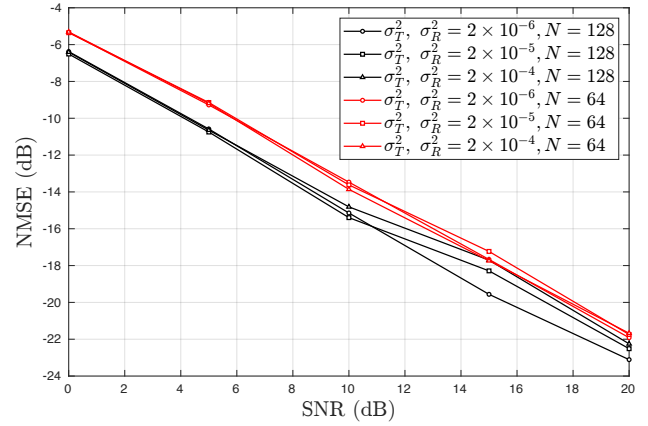

 Fig. 4: NMSE vs. SNR (dB) for $N = 128$.


Fig. 5: Performance Analysis.

demonstrate the NMSE performance comparison against LS, MMSE, DNN, and LSTM at various SNR levels, and Fig. 5 presents the performance of our DL model at multiple levels of PN variance. We plotted NMSE values at five different levels stated above and observed that our proposed scheme performs very consistently and outperforms other algorithms by large margin.

In Fig. 3, for $N = 64$, it is portrayed that our proposed model achieves an estimation NMSE of -5.42 dB at 0 dB SNR, whereas LS obtains -0.04 dB, MMSE obtains -2.76 dB, DNN obtains -2.438 dB, and standalone LSTM obtains -4.167 dB. On the other hand, at 20 dB SNR level, they obtain -20.05 dB, -20.08 dB, -16.679 dB, and -20.089 dB respectively, and our proposed model achieves -22.538 dB. In this scenario, at the very low SNR, our model outperforms LS, MMSE, DNN, and LSTM by 99%, 49%, 55%, and 23% margins respectively, whereas at very high SNR, it outperforms by 11%, 10%, 25%, and 10.8% margins respectively in terms of NMSE.

We can observe the similar trend in estimation performance for $N = 128$ in Fig. 4. At low SNR, our model achieves

−6.511 dB whereas LS, MMSE, DNN, and LSTM obtains 0.03 dB, −2.64 dB, −1.984 dB, and −4.9037 dB respectively. It is evident that the proposed scheme outperforms LS, MMSE, DNN, and LSTM by 99.5%, 59.45%, 69.52%, and 24.68% respectively in terms of NMSE. Similarly, at very high SNR, e.g., 20 dB, we outperform LS, MMSE, DNN, LSTM by 11.11%, 11.07%, 33.81%, and 11.39% respectively.

Furthermore, in Fig. 5 we also compare the performance of our model at various σ_T^2 and σ_R^2 for both $N = 128$ and 64. As we can observe from the figure that it compares the performance under three different phase noise variances, i.e. 2×10^{-6} , 2×10^{-5} , and 2×10^{-4} . Higher PN adversely impacts the stability and accuracy of the signal processing at both the Tx and Rx. For all configurations, the NMSE decreases with increasing SNR. This trend is consistent and expected, as higher SNR implies a stronger signal relative to noise, leading to more accurate channel estimation. In addition, with a larger number of antennas, ($N = 128$) the channel estimation task yields consistently better performance than with fewer antennas ($N = 64$).

V. CONCLUSIONS

In this paper, we have proposed a novel BiLSTM-GRU based channel estimation algorithm for high frequency UM-MIMO networks that operate in the THz-band. The proposed model is designed and deployed with a combination of BiLSTM, GRU, and Dense layers. The first BiLSTM layer performs the denoising of the channel and GRU layers accelerates the training while smoothing the channel matrices. We have compared our proposed algorithm with baselines LS, MMSE, LSTM, and DNN and furthermore have evaluated the performance in terms of NMSE at various PN noise variances. The results concluded that our model is robust in both high and low SNR scenarios and outperforms the baseline consistently by a large margin.

REFERENCES

- [1] K. Dovelos, M. Matthaiou, H. Q. Ngo, and B. Bellalta, "Channel estimation and hybrid combining for wideband terahertz massive MIMO systems," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 6, pp. 1604–1620, 2021.
- [2] X. Cai, X. Cheng, and F. Tufvesson, "Toward 6G with terahertz communications: Understanding the propagation channels," *IEEE Communications Magazine*, vol. 62, no. 2, pp. 32–38, 2024.
- [3] A. M. Elbir, K. V. Mishra, and S. Chatzinotas, "Terahertz-band joint ultra-massive MIMO radar-communications: Model-based and model-free hybrid beamforming," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 6, pp. 1468–1483, 2021.
- [4] Z. Sha and Z. Wang, "Channel estimation and equalization for terahertz receiver with rf impairments," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 6, pp. 1621–1635, 2021.
- [5] Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu, "5G network capacity: Key elements and technologies," *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 71–78, 2014.
- [6] A. M. Elbir, W. Shi, K. V. Mishra, and S. Chatzinotas, "Federated multi-task learning for thz wideband channel and doa estimation," in *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*. IEEE, 2023, pp. 1–5.
- [7] I. Kallfass, F. Boes, T. Messinger, J. Antes, A. Inam, U. Lewark, A. Tessmann, and R. Henneberger, "64 gbit/s transmission over 850 m fixed wireless link at 240 ghz carrier frequency," *Journal of Infrared, millimeter, and terahertz waves*, vol. 36, pp. 221–233, 2015.
- [8] I. Kallfass, I. Dan, S. Rey, P. Harati, J. Antes, A. Tessmann, S. Wagner, M. Kuri, R. Weber, H. Massler *et al.*, "Towards mmic-based 300ghz indoor wireless communication systems," *IEICE transactions on electronics*, vol. 98, no. 12, pp. 1081–1090, 2015.
- [9] J. Grzyb, P. Vazquez, B. Heinemann, and U. Pfeiffer, "A high-speed qpsk/16-qam 1-m wireless link with a tunable 220–260 ghz lo carrier in sige hbt technology," in *2018 43rd International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*. IEEE, 2018, pp. 1–2.
- [10] P. Rodríguez-Vázquez, J. Grzyb, B. Heinemann, and U. R. Pfeiffer, "A 16-qam 100-gb/s 1-m wireless link with an evm of 17% at 230 ghz in an sige technology," *IEEE Microwave and Wireless Components Letters*, vol. 29, no. 4, pp. 297–299, 2019.
- [11] X. Wei, C. Hu, and L. Dai, "Deep learning for beamspace channel estimation in millimeter-wave massive MIMO systems," *IEEE Transactions on Communications*, vol. 69, no. 1, pp. 182–193, 2020.
- [12] X. Gao, L. Dai, S. Zhou, A. M. Sayeed, and L. Hanzo, "Wideband beamspace channel estimation for millimeter-wave MIMO systems relying on lens antenna arrays," *IEEE Transactions on Signal Processing*, vol. 67, no. 18, pp. 4809–4824, 2019.
- [13] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2370–2386, 2016.
- [14] Y. Han, S. Jin, C.-K. Wen, and X. Ma, "Channel estimation for extremely large-scale massive MIMO systems," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 633–637, 2020.
- [15] H. Zhang, N. Shlezinger, F. Guidi, D. Dardari, and Y. C. Eldar, "6G wireless communications: From far-field beam steering to near-field beam focusing," *IEEE Communications Magazine*, 2023.
- [16] X. Yin, S. Wang, N. Zhang, and B. Ai, "Scatterer localization using large-scale antenna arrays based on a spherical wave-front parametric model," *IEEE Transactions on Wireless Communications*, vol. 16, no. 10, pp. 6543–6556, 2017.
- [17] Z. Hu, C. Chen, Y. Jin, L. Zhou, and Q. Wei, "Hybrid-field channel estimation for extremely large-scale massive MIMO system," *IEEE Communications Letters*, vol. 27, no. 1, pp. 303–307, 2022.
- [18] M. Cui and L. Dai, "Channel estimation for extremely large-scale MIMO: Far-field or near-field?" *IEEE Transactions on Communications*, vol. 70, no. 4, pp. 2663–2677, 2022.
- [19] X. Wei and L. Dai, "Channel estimation for extremely large-scale massive mimo: Far-field, near-field, or hybrid-field?" *IEEE Communications Letters*, vol. 26, no. 1, pp. 177–181, 2021.
- [20] W. Yu, Y. Shen, H. He, X. Yu, J. Zhang, and K. B. Letiaief, "Hybrid far-and near-field channel estimation for thz ultra-massive MIMO via fixed point networks," in *GLOBECOM 2022-2022 IEEE Global Communications Conference*. IEEE, 2022, pp. 5384–5389.
- [21] X. Zhang, Z. Wang, H. Zhang, and L. Yang, "Near-field channel estimation for extremely large-scale array communications: A model-based deep learning approach," *IEEE Communications Letters*, 2023.
- [22] H. Lei, J. Zhang, H. Xiao, X. Zhang, B. Ai, and D. W. K. Ng, "Channel estimation for XL-MIMO systems with polar-domain multi-scale residual dense network," *IEEE Transactions on Vehicular Technology*, 2023.
- [23] H. Mehrpouyan, A. A. Nasir, S. D. Blostein, T. Eriksson, G. K. Karagiannidis, and T. Svensson, "Joint estimation of channel and oscillator phase noise in MIMO systems," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4790–4807, 2012.
- [24] Y. Dong, H. Wang, and Y.-D. Yao, "Channel estimation for one-bit multi-tuser massive MIMO using conditional gan," *IEEE Communications Letters*, vol. 25, no. 3, pp. 854–858, 2020.
- [25] M. E. Rasekh, M. Abdelghany, U. Madhow, and M. Rodwell, "Phase noise in modular millimeter wave massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 20, no. 10, pp. 6522–6535, 2021.
- [26] I. Helmy, P. Tarafder, and W. Choi, "LSTM-GRU model-based channel prediction for one-bit massive MIMO system," *IEEE Transactions on Vehicular Technology*, 2023.
- [27] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.
- [28] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "{TensorFlow}: a system for {Large-Scale} machine learning," in *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, 2016, pp. 265–283.
- [29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.