An Attention-based AI Model for 3D Beam Prediction in THz Unmanned Aerial Vehicle Communication

Muhammad Enayetur Rahman
Electrical and Computer Engineering
Old Dominion University
Norfolk, VA, 23529, USA
mrahm011@odu.edu

Md. Shirajum Munir
School of Cybersecurity
Old Dominion University
Norfolk, VA, 23529, USA
mmunir@odu.edu

Sachin Shetty
Electrical and Computer Engineering
Old Dominion University
Norfolk, VA, 23529, USA
sshetty@odu.edu

***Abstract*—In this study, we considered the Terahertz (THz) Unmanned Aerial Vehicle (UAV) communication scenario, where we are dealing with the problem of beam prediction. UAVs have emerged as a critical component in evolving sixth-generation (6G) wireless networks, bringing seamless communication and extensive coverage. The usage of the THz spectrum allows for high-speed data throughput for the envisioned 6G era. However, the dynamic mobility of UAVs exacerbates the inherent challenges of THz communications, such as channel impairments and obstruction-related effects. We highlight the potential of Reconfigurable Intelligent Surface (RIS), which offers a flexible technique to boost coverage by adapting to channel fluctuations. We propose a unique deep learning paradigm based on self-attention networks in integrating RIS technology with THz UAV communication. In particular, we design and develop an attention-based mechanism that can predict the ideal serving beam for UAVs. This is based on sequential communication data such as UAV's beam trajectories, accurate position information, and Line of Sight (LoS) information. The proposed attention-based beam prediction mechanism performs well compared to the baseline Long Short-Term Memory (LSTM) and Gated recurrent units (GRUs) models, indicating superior performance and effectiveness. The findings, supported by comprehensive evaluation criteria such as Root-Mean-Square Error (RMSE) and Mean Absolute Error (MAE), demonstrate our proposed model's competency in capturing intricate temporal relationships, outperforming its competitors in predicting reliability.**

***Index Terms*—Terahertz, 6G, Beam prediction, RIS, Drone, Unmanned Aerial Vehicle communication.**

I. Introduction

A. Motivation and Challenges

Terahertz (THz) communication is an emergent wireless communication technology that works within microwave and infrared wavelengths, generally ranging from 0.1 to 10 THz [1]. Although the THz spectrum theoretically spans 300 GHz to 10 THz, researchers have chosen to designate uses beyond 100 GHz as THz communications. This boundary distinguishes them from the 5G mmWave bands, which are designated below this range [2]. This largely untapped frequency spectrum has enormous promise for high-speed, high-capacity data transmission, facilitating many applications such as wireless networks, the Internet of Everything (IoE), immersive Virtual Simulation (VS), and Enhanced Reality (ER) for sensing and visualization. The THz frequency band within wireless communications is an attractive choice for significantly increasing capacity by an order of magnitude. [3]. THz communication has a lot of advantages over other wireless systems. In Next Generation (NextG) wireless communication, ultra-high data rate applications are predicted to have high demand in the near future. To support these high-demand applications, the THz wireless band is an ideal candidate.

Unlike traditional radio frequency (RF) communication technologies, the THz frequency range offers vast available bandwidth, allowing the rapid transmission of enormous data volumes. High-definition video transmission, real-time multimedia streaming, and other resource-intensive applications are made possible by the broad bandwidth, which may handle data rates of tens or even hundreds of gigabits per second for a solitary wireless connection [4]. In [5], authors developed a lab-based real-time fiber optic/THz-wireless system and a 500-meter outdoor THz-wireless link, both achieving 100 Gb/s data rates. Because of THz's superior spatial resolution due to it's short wavelengths, object tracking and localization and tracking are made possible in this band. This is particularly beneficial for applications like driverless cars, smart cities, and indoor positioning systems [6].

However, there are several challenges and obstacles of THz drone communication. Due to air and molecular absorption noise [7], THz bands severely suffers from signal attenuation. This results limiting it's communication range. Moreover, THz waves are extremely sensitive to physical barriers including humans, leading to non-line-of-sight (NLoS) scenarios that make the communication process even more complex. In contrast to microwave and millimeter-wave frequencies, the primary challenges associated with the THz spectrum involve significant signal attenuation and restricted mobility support in environments with obstructions. As a result, network coverage

This work is supported in part by DoD Center of Excellence in AI and Machine Learning (CoE-AIML) under Contract Number W911NF-20-2-0277 with the U.S. Army Research Laboratory, National Science Foundation under Grant No. 2219742 and Grant No. 2131001.

is severely reduced. To improve 6G network coverage and address obstructions, exploring alternative techniques such as advanced signal processing, dynamic frequency allocation, AI-driven network optimization are studied. In 6G communication, the Reconfigurable Intelligent Surface (RIS) is introduced as a technology to augment the capabilities of existing THz communication systems [8]. In order to counteract signal attenuation, increase coverage in NLoS circumstances, and boost overall system capacity, RIS, which consists of passive devices that reflect and alter electromagnetic waves, can be used in THz communication systems.

Incorporating RIS and drones in THz communication poses other challenges. For example, desired Beamforming and Beam prediction patterns requires sophisticated algorithms for the dynamically changing environments. Channel estimation is also challenging due to the same reason. On the other hand, adding drones in THz scenario is highly challenging due to drone's high mobility and dynamics environment nature. As a result, accurate tracking and beam steering is a huge challenge in this scenario.

The aim of this research is to examine the effectiveness of using a time series-based transformer Deep Learning beam prediction method in THz Unmanned Aerial Vehicle (UAV) communication systems. Specifically, we are exploring the potential of utilizing a RIS in conjunction with multiple drones in a dynamic environment. This combination of RIS and beam prediction algorithm offers an innovative solution to the challenges faced in THz communication. Our goal is to showcase the benefits of Deep Learning-based beam prediction in enhancing THz drone communication by testing the proposed method in a real-world scenario. This study's results will provide insights into the benefits, challenges, and practical aspects of incorporating beam prediction algorithms in THz drone communication systems, including RIS.

B. Contribution

- First, we proposed and developed an Attention-based multivariate Deep Learning (DL) model which can predict the next sequence of 3D RIS beams for the user. It predicts the next sequence of beam and location information.
- Second, we comprehensively compared our innovative deep learning model and the standard Long Short-Term Memory (LSTM) networks and Gated Recurrent Unit (GRU) models. This evaluation thoroughly assessed these models' Root-Mean Squared Error (RMSE) and Mean Absolute Error (MAE) values.. Our attention-based proposed deep learning model performs similarly to the other state-of-the-art models.
- Third, our proposed model's total number of parameters is less than the baseline LSTM's and GRU's total. Our proposed model learns and stabilises faster than the LSTM and GRU models.

II. Related Work

In 6G THz communication, making early beam data direction predictions is crucial because it reduces the time and resources needed for beam training, leading to faster communication. The authors [9] introduced a sophisticated deep learning technique that uses a neural network called the gated recurrent unit (GRU) to predict which base station or RIS each drone should connect to and which data beam it should use. A novel approach was created by authors [10] to overcome THz hybrid beamforming system obstacles, such as 3D angle estimate with a precision of millidegrees and reducing the overhead associated with beam tracking. Their algorithm is an off-grid ultra-high-resolution direction-of-arrival (DoA) estimate technique that aligns with the array-of-subarrays (AoSA) architecture. When adjusting the narrow beams of antenna arrays, there is typically a large overhead for beam training that increases with the number of antennas. To address this issue, authors [11] propose a multi-modal machine learning strategy that uses location and visual (camera) data from the wireless communication environment to anticipate beams fast. To validate their proposed method, the authors created a real-world multi-modal vehicular dataset that comprises position, camera vision, and mmWave beam data. Recently, transformer-based [12] architecture has become popular day by day in solving many wireless communication problems. For example, Massive MIMO needs to use phased array antennae, which is limited to antenna spacing. To recover this problem, authors [13] developed an optimization problem using Holographic MIMO to maximize the channel capacity. They utilized a transformer-based framework to allocate power to the proper users for beamforming. In [14], authors developed a masked autoencoder and deep reinforcement learning-based RIS-enabled system to learn pilot signal allocation policies, which can use reduced pilot signal. Their work also enhances the channel estimation performance.

III. System Model and Problem Formulation

We are considering a scenario for multiple outside drone users with a Base Station (BS) and a RIS. The signal from the base station to the drone user comes from a multipath environment. BS serves as the central hub for the scenario. The BS manages all signals and control information, including the bandwidth of the RIS and UEs. Additionally, it processes communication and AI-assisted algorithms. The RIS remain stationary and are positioned above the ground at a fixed location. These surfaces are always linked to the BS, and there is a clear Line of Sight (LoS) between them. The RIS can be considered an Aerial Base Station (AEB), a central point linking the backhaul, which connects to the BS, and the access network, which connects to the drone users. Essentially acting as a hub, the RIS allows seamless communication between the two networks, optimizing data flow and ensuring reliable connectivity for drone users. Regardless of whether there is a direct LoS between the drone users and the BS, all drone users have a clear LoS with the RIS. In this study, we will be referring to drones or flying objects as users. These users can

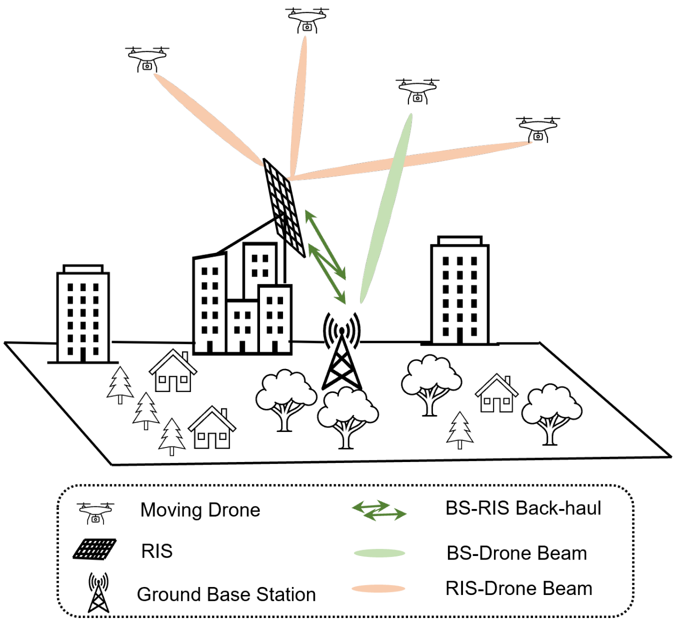

Fig. 1. System Model for Outside THz Drones with RIS.

connect to the Base Station either directly with Line of Sight (LoS) or indirectly through RIS in non-line of sight situations. The RISs are connected to the ground Base Station.

A. System and Channel Model

Our system model is depicted in Fig 1. The base station has N_B number of antenna array elements with beamforming vector $\mathbf{b} \in \mathbb{C}^{N_B \times 1}$. A predetermined beam codebook $\mathcal{P}$ of size N_{CB} is used to choose the beamforming vectors. When the direct line of sight (LoS) connection between the base station and the UAV drone is disrupted, the drone is assisted by a RIS equipped with N_{RIS} antennas, ensuring continued service provision.

Let, $\alpha_{T,k}$ and $\alpha_{T,k}^T$ are the uplink and downlink of the transmitter channel coefficients to the RIS respectively. We also denote, $\alpha_{R,k}$ and $\alpha_{R,k}^T$ for the corresponding uplink and downlink receiver channel coefficients of the RIS. We find the received signal strength is given by:

$$r_k = \alpha_{R,k}^T \mathbf{\Phi} \alpha_{T,k} d + u, \tag{1}$$

$$r_k = \left(\alpha_{R,k}^T \odot \alpha_{T,k}\right)^T \phi d + u, \tag{2}$$

Where $d \in \mathbb{C}$ is the data symbol that satisfies $\mathbb{E}\left[|d|^2\right] = P_T$, P_T is the total transmit power, $u \sim \mathcal{M}_{\mathbb{C}}\left(0, \omega^2\right)$, and $\Phi \in \mathbb{C}^{I \times I}$ serves as the interaction matrix, depicting how the incident signal from the transmitter interacts with the RIS. The interaction matrix, $\mathbf{\Phi} = \text{diag}(\phi)$ where ϕ represents a diagonal matrix. The diagonal pattern emerges due to the RIS's operation, where each element represented by $e, e = 1, 2, \ldots, E$ reflects the incoming signal after being modified by a specific interaction factor denoted as $[\phi]_e$. The interaction factor $[\phi]_e$ is determined as $[\phi]_e = e^{j\psi_e}$, given that the RIS elements are exclusively constructed using phase shifters. This interaction reflection beamforming vector, denoted as ϕ, is chosen from a predefined set known as the reflection beamforming codebook $\mathcal{B}$.

In this study, we use a W_c clustered wideband geometric THz channel model. One ray from each cluster w, $w = \{1, \ldots, W_c\}$ is assumed to contribute, with each ray having a time delay $\tau_w \in \mathbb{R}$, azimuth-elevation angles of arrival (AoA) denoted by (θ_w, ϕ_w). The complex path gain which also accounts for path loss can be represented by α_w. Let $p_{\text{rc}}(\tau)$ stand in for a pulse shaping function for seconds based T-spaced signaling. The delay of channel between user and base station is δ is represented by the following equation.

$$\mathbf{h}_\delta = \sum_{w=1}^{W_c} \alpha_w p_{\text{rc}}\left(\delta T_s - \tau_w\right) A_{\text{rv}}\left(\theta_w, \phi_w\right), \tag{3}$$

where $A_{\text{rv}}\left(\theta_w, \phi_w\right)$ is the base station's array response vector at the AoA (θ_w, ϕ_w). Channel coefficient vector at subcarrier k, $\mathbf{h}_k$, can be expressed below in frequency domain as:

$$\mathbf{h}_k = \sum_{t_p=0}^{T_p-1} \mathbf{h}_\delta e^{-j\frac{2\pi k}{K}d}. \tag{4}$$

Here, T_p is the total number of taps and K is the total number of subcarriers. In the context of a block fading channel model, $\{\mathbf{h}_{n,k}\}_{k=1}^K$ are taken to be constant during the T_C channel coherence period [15].

B. Problem Formulation

The main focus of our study is to predict RIS beamforming trajectories. To achieve this, we create an attention based DL model to effectively direct the future RIS Beam. The base station adjusts its beam $\mathbf{b}$ to match the user's movement and ensure a stable connection. The duration of beam coherence can change based on variables such as the user's velocity and the base station antenna number. The beam used to communicate with the drone at a specific moment is denoted as $\mathbf{bs}^{(t)}$, where t represents the beam coherence time. A sequence of beams is established based on this t-step.

$$\mathcal{B}_t = \left\{\mathbf{bs}^{(1)}, \mathbf{bs}^{(2)}, \ldots, \mathbf{bs}^{(t)}\right\}. \tag{5}$$

In a similar vein, when considering the RIS, the beamforming vector at time t as $\phi^{(t)}$, the sequence of steps can be expressed as:

$$\mathcal{F}_t = \left\{\Phi^{(1)}, \Phi^{(2)}, \ldots, \Phi^{(t)}\right\}. \tag{6}$$

When beam $\mathbf{b}^{(t)}$ was selected, let $\mathbf{l}^{(t)}$ denote the drone location at time t (i.e.,). Then, we define the location sequence of t time as:

$$\mathcal{L}_t = \left\{\mathbf{l}^{(1)}, \mathbf{l}^{(2)}, \ldots, \mathbf{l}^{(t)}\right\}. \tag{7}$$

We can use $d^{(t)}$ to indicate whether there is a Line of Sight (LoS) connection between the base station and the drone user at time step t:

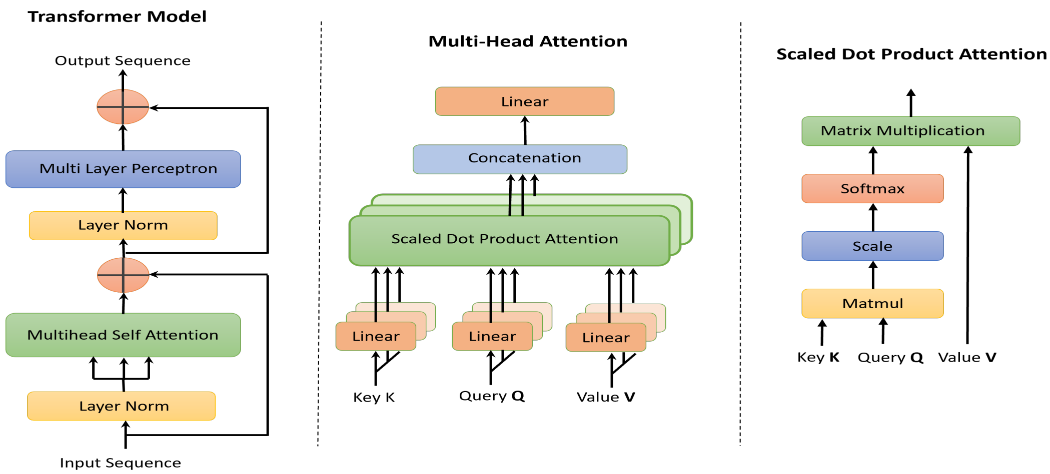

Fig. 2. Attention Based Beam Prediction.

$$\mathcal{D}_t = \left\{ d^{(1)}, d^{(2)}, \ldots, d^{(t)} \right\}. \tag{8}$$

We show how to solve the problem of directing the future beam at time $t+1$ based on the given beam sequence B_t. To achieve this, we utilize a mutivariate attention-based deep learning system that maps the beam sequence from B_t to B_{t+1}. Our system can accurately predict the appropriate communication channel and beam to use, and we explain how to train a deep neural network to achieve this high level of accuracy.

IV. Solution Approach

In this study, we developed and implemented a multivariate time-series attention based transformer network to predict the RIS beam trajectories. The Transformer [12] leverages the attention mechanism for feature extraction. An attention-based (transformer) beam prediction model architecture is shown in fig. 2. Specifically, this entails computing correlations among input data in the attention layer (fig. 2 right) and applying a multiplication step to produce a weighted input matrix. The attention layer entails three different embedding matrices, i.e., Query $\mathbf{Q}$, Key $\mathbf{K}$, and value $\mathbf{V}$. Query $\mathbf{Q}$ and Key $\mathbf{K}$ matrices comprise the features of the input sequence data. A self-attention score can be obtained at the final step of attention layer. In contrast, multi-head self-attention improves the model's performance by concurrently tackling multiple input sequence elements. By drawing upon correlations across all elements in input sequences for weight generation, the Transformer adeptly captures both local and global features that display correlation. The Transformer model is considered to be a highly effective machine learning model for analyzing sequential and time series data. This is primarily attributed to its remarkable capability to detect correlations between distant components, process data in parallel, and leverage an attention mechanism to retain positional information without necessitating explicit order analysis.

We proposed a novel deep learning model based on the transformer self-attention architecture which aims to predict regression time-series data. Our proposed model is designed to process sequences of data with self-attention mechanism and convolutional layers to capture temporal patterns in the data sequences. The model architecture comprises encoder blocks which includes normalization for stable learning, multi-head self attention, dropout for regularization. A feed-forward network incorporated with convolutional layers to extract local patterns is introduced. Global average pooling is used where the important features are extracted for analysis. After that, a multi-layer perceptron (MLP) layers with $ReLU$ activation is used to get a high order relationships within the sequence. Lastly, a dense layer as a output layer delivers the model's prediction for a continuous output values.

V. Experimental Analysis and Discussion

In this section, we thoroughly evaluated and compared our proposed Deep Learning model for 6G UAV with traditional LSTM or GRU models.

A. Experimental Settings

To access our DL-Attention based model, we used the scenario using the DeepMIMO framework [16]. In this framework, we used a drone scenario of a downtown area. The setup includes a ground base station positioned 6 meters above the ground and a hovering RIS located 80 meters high. Additionally, there is a 3D drone grid, consisting of four parallel 2D grids at varying heights, with the base of the grid placed 40 meters above the ground. The grid points are positioned at regular intervals of 0.81 meters in the horizontal (x and y) directions and 0.8 meters in the vertical (z) direction. The point where these axes intersect is enclosed by buildings of varying heights but uniform base dimensions. The drones follow random waypoint trajectories as defined in the random waypoint model. For further information can be found in [16].

The DeepMIMO script generates a dataset for the 6G drone scenario. It includes channel and beam information between the ground base station, RIS, and drones. Input parameters are listed in Table I. We created an experimental dataset by generating drone trajectories using Table I.

TABLE I
DATASET GENERATION INPUT PARAMETERS.

Parameter	Value
First, Last User Row	1, 496 [9]
BS Antenna quntity	64, 1, 1 [9]
RIS Antenna quntity	256, 1, 1 [9]
Space between Antennas	0.5 [9]
Center Frequency [GHz]	200 [9]
Bandwidth [GHz]	1 [9]
OFDM Subcarriers	512 [9]
Sampling Factor (OFDM)	1 [9]
OFDM range	1 [9]
Path count	1 [9]

The dataset includes beam vector information such as arrival/departure angles, phase, time of arrival, and power. We used beam information from the RIS, base station, and drone sequence numbers. If there was no communication path or blockage with the base station, we set the beam information to zero for that drone position. Finally, we combined the drone location vector, base station and RIS line of sight, and the number of paths from the base station. We created a dataset with 10-step time sequences for drone trajectories. It includes base station and RIS beamforming vectors, sequence number, number of paths, 3D drone positions, and LoS from the BS and RIS. The dataset has 10,000 rows. We used 21 feature vectors as input and predicted 1 row of vectors for each 10-step sequence. The experimental dataset was randomized and split into train, validation, and testing subsets, with a split ratio of 20% for testing and 20% for validation.

The features in our dataset have different scales, and no extreme outliers exist. We transform the dataset features into a range of 0 to 1. The dataset contains different scales for different features, and we want to transform them without changing their relationship. The following formula is used to transform the features:

$$XScaled = (X - XMin)/(XMax - XMin)$$

We apply feature scaling to the input dataset, which will help our machine-learning model perform better. In our dataset, we fit the scaler using the training data to learn the scaling parameters. Then, we apply the same learned parameters on the training, validation and testing dataset to ensure that all datasets have the same standardized scale. We only fit the scaling on the training data to prevent data leakage and support a more realistic model evaluation. Our proposed model's parameters are listed in Table II.

We compared our proposed model with LSTM and GRU models. LSTM/GRU models are developed for a sequential neural network architecture, which takes the input sequences with a length of 10 and 21 features per time step. To capture sequential patterns, the models include LSTM/GRU 32 units. A dropout layer with a rate of 0.2 is added to mitigate the overfitting, followed by a fully connected dense layer with 16 neurons and a $ReLU$ activation function to introduce non-linearity. Another dropout layer is added for further regularization.

TABLE II
PROPOSED MODEL'S PARAMETERS.

Parameter	Value
Head Size	4
Number of Heads	4
Feed Forward Dimension	16
Number of Transformer Blocks	2
MLP Units	32
MLP Dropout	0.3
Dropout	0.3
Batch Size	32
Optimizer	Adam (LR=0.001)
Loss Function	Mean Squared Error
Evaluation Metric	Root Mean Squared Error
Epochs	200

Finally, a dense layer with 21 neurons and a linear activation function is added for the regression tasks. For a fair evaluation of our proposed model, we tried to build the LSTM/GRU models with the same number of total parameters, the same Adam optimizer with the same learning rate (0.001) and the same number of training (200) epochs.

B. Result Analysis

The LSTM model has a total parameter of 7,797, GRU has a total of 6,165, and our proposed model has a total of 5,457. We can see from Fig. 3 that even though our proposed model has much fewer parameters than the LSTM/GRU models, the Root Mean Squared Error (RMSE) vs. Epoch plot depicts that our proposed model learns more rapidly than the LSTM/GRU models.

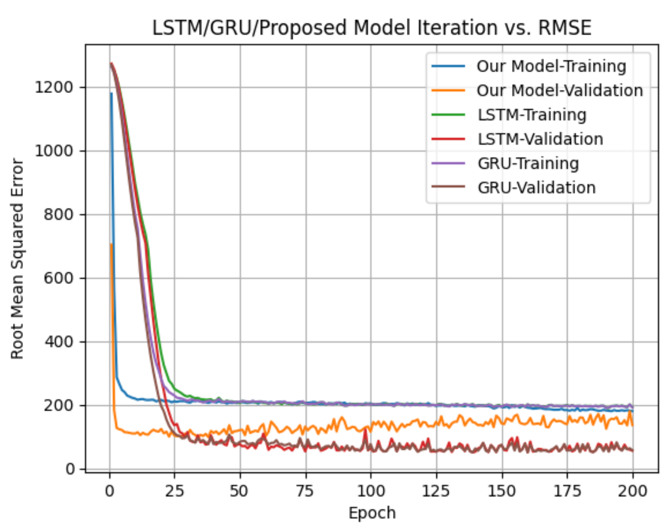

Fig. 3. Iteration vs RMSE plot for RIS Beam.

By the 10th epoch, the training RMSE is almost stable, while it takes more than the 27th epoch for LSTM/GRU. LSTM/GRU performs almost similarly. The training curves have higher RMSE values than the validation curves because there are dropout layers. However, among these models, the validation curve of our proposed model at epoch 200 merges with the training curve, which indicates a relatively better model than the other two models.

TABLE III
MODEL PERFORMANCE ON RIS'S BEAM.

Model	MAE	MSE	RMSE
LSTM	22.94199	1887.49869	43.44535
GRU	22.95527	1887.95526	43.45061
Proposed Model	32.21674	2609.64263	51.08466

We evaluate the model performance by calculating Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), which are mentioned in Table III. Our proposed model performs well enough compared to classical LSTM and GRU models.

Based on the provided data, it is apparent that the Transformer model has much higher MAE and RMSE values in comparison to the LSTM and GRU models. It might happen for several reasons.

- If the data has strong temporal pattern Transformer model might not be best fit for all sequential tasks while traditional LSTM/GRU might perform well in this case.
- If the data has short-term patterns and dependencies, then LSTM/GRU perform well rather than long-term dependencies, whereas Transformer performs well.
- Transformer models are often larger and more complex than LSTM and GRU models. They require a substantial amount of data to learn patterns effectively. We are only considering 10,000 data here.

VI. CONCLUSION

In this study, we predict the future beam of the RIS in a 6G drone communication scenario for different drone users. Our objective was to enhance the reliability of the communication between the drone user and the base station by predicting the future beam prediction. Our proposed model is based on the transformer architecture and thoroughly compared with the classical LSTM and GRU models. Our dataset comprises the drone trajectories that include beamforming vectors of both the base station and RIS, 3-D positional coordinates of drones, and Line-of-Sight (LoS) information from both the base station and RIS, as well as sequence numbers and number of paths. This large and comprehensive dataset represents drone spatial movement and communication information, which is critical for the 6G drone scenario. Our proposed model is based on the transformer architecture, which can capture intricate temporal dependencies within the sequential data. The proposed model consistently compares favorably to LSTM and GRU models, as shown by evaluating its performance using the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) metrics.

REFERENCES

[1] Z. Chen, C. Han, Y. Wu, L. Li, C. Huang, Z. Zhang, G. Wang, and W. Tong, "Terahertz wireless communications for 2030 and beyond: A cutting-edge frontier," *IEEE Communications Magazine*, vol. 59, no. 11, pp. 66–72, 2021.

[2] H. Sarieddeen, N. Saeed, T. Y. Al-Naffouri, and M.-S. Alouini, "Next generation terahertz communications: A rendezvous of sensing, imaging, and localization," *IEEE Communications Magazine*, vol. 58, no. 5, pp. 69–75, 2020.

[3] R. Piesiewicz, T. Kleine-Ostmann, N. Krumbholz, D. Mittleman, M. Koch, J. Schoebel, and T. Kurner, "Short-range ultra-broadband terahertz communications: Concepts and perspectives," *IEEE Antennas and Propagation Magazine*, vol. 49, no. 6, pp. 24–39, 2007.

[4] I. F. Akyildiz, J. M. Jornet, and C. Han, "Terahertz band: Next frontier for wireless communications," *Physical communication*, vol. 12, pp. 16–32, 2014.

[5] C. Castro, R. Elschner, T. Merkle, C. Schubert, and R. Freund, "Experimental demonstrations of high-capacity thz-wireless transmission systems for beyond 5g," *IEEE Communications Magazine*, vol. 58, no. 11, pp. 41–47, 2020.

[6] K. M. S. Huq, S. A. Busari, J. Rodriguez, V. Frascolla, W. Bazzi, and D. C. Sicker, "Terahertz-enabled wireless system for beyond-5g ultra-fast networks: A brief survey," *IEEE network*, vol. 33, no. 4, pp. 89–95, 2019.

[7] J. Kokkoniemi, J. Lehtomäki, and M. Juntti, "A discussion on molecular absorption noise in the terahertz band," *Nano communication networks*, vol. 8, pp. 35–45, 2016.

[8] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE access*, vol. 7, pp. 116 753–116 773, 2019.

[9] N. Abuzainab, M. Alrabeiah, A. Alkhateeb, and Y. E. Sagduyu, "Deep learning for thz drones with flying intelligent surfaces: Beam and handoff prediction," in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2021, pp. 1–6.

[10] Y. Chen, L. Yan, and C. Han, "Millidegree-level direction-of-arrival (doa) estimation and tracking for terahertz wireless communications," in *2020 17th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2020, pp. 1–9.

[11] G. Charan, A. Hredzak, C. Stoddard, B. Berrey, M. Seth, H. Nunez, and A. Alkhateeb, "Towards real-world 6g drone communication: Position and camera aided beam prediction," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, 2022, pp. 2951–2956.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[13] A. Adhikary, A. D. Raha, Y. Qiao, M. S. Munir, K. T. Kim, and C. S. Hong, "Transformer-based communication resource allocation for holographic beamforming: A distributed artificial intelligence framework," in *2023 24st Asia-Pacific Network Operations and Management Symposium (APNOMS)*, 2023, pp. 243–246.

[14] K. Kim, Y. K. Tun, M. S. Munir, W. Saad, and C. S. Hong, "Deep reinforcement learning for channel estimation in ris-aided wireless networks," *IEEE Communications Letters*, vol. 27, no. 8, pp. 2053–2057, 2023.

[15] V. Va, J. Choi, T. Shimizu, G. Bansal, and R. W. Heath, "Inverse multipath fingerprinting for millimeter wave v2i beam alignment," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4042–4058, 2017.

[16] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," in *Proc. of Information Theory and Applications Workshop (ITA)*, San Diego, CA, Feb 2019, pp. 1–8.