DeepMIMO: A Generic Deep Learning Dataset for Millimeter-Wave and Massive MIMO Applications to Vehicular Communications

By

Falade Oluwaseun

[redacted]

INTRODUCTION

Millimeter-wave(mmWave) and massive MIMO are key enabling technologies for current and future wireless systems.

Massive MIMO employs large numbers of antennas, while mm-Wave refers to higher frequency radio bands ranging from 24GHz to 40GHz, designed to offer much faster data speeds to support highly mobile users (Humans or vehicles).

AIM: Why Generic Deep Learning?

The main aim of this project is to reproduce the results of the other papers and then compare the different algorithms based on the common data. Generic deep learning helps us access a central dataset to evaluate the performance of the machine learning algorithms.

We apply deep learning tools to use several features of the environment and user setups to learn how to use them to predict mmWave and massive MIMO channels/beams. We can have control over the system setup and the antenna configuration. We can also reproduce results, by stating the parameters set, S, and the adopted ray-tracing scenario, 'R', to completely define the generated dataset.[3]

REVIEW OF THE EXISTING LITERATURE

The current challenges and advances in MIMO and mm-wave applications of vehicles are based on certain factors that have proven to affect the communications between vehicles and also future advances. This introduces the need to properly estimation of Vehicular channel characteristics and ensures proper modeling based on assumptions

The channel modeling technique considered for the analysis of the vehicular radio channel is the basis on which relevant channel propagation characteristics are obtained.

Early design of vehicular channels

The Stochastic and geometry-based models were used for the design and comparison of systems. They analyze the important properties of channels with low computational complexity.[13]

Deterministic modeling was used for network planning and system deployment. Furthermore, a huge database of mm-wave V2V environmental characteristics was still needed[14]

However, past models do not take into account the effects of:

- reflection, diffraction
- transmission through obstacles
- physical properties of vehicles themselves (carrying the antennas).

The current design of vehicular channels

An updated channel models classification was created which took into consideration the effects mentioned previously. They are:

- Empirical models
- Non-geometry-based stochastic models.
- Geometry-based stochastic models.
- Geometry-based deterministic channel models.

Empirical Models

Empirical models use the collection of observations, usually obtained during measurement campaigns, in order to identify singular patterns of the environment that it is intended to represent. As a result, these models are very simple, with few parameters, and thus, easy to use. Despite these advantages, its effectiveness is directly related to environments similar to those where measurements were performed. They are very simple with very few parameters to take into consideration. Since these models cannot determine situations of specific environment characteristics and have limited accuracy, their use in vehicular communication systems is very restricted[15].

Non-Geometry Stochastic Models (NGSM)

Non-geometry-based Stochastic models are used to predict modeled system values. They consider random elements occurring within the environment. They are designed to simulate uncertainty in different scenarios. These models can describe propagation channels in different scenarios. This technique uses the Tapped-delay line (TDL) model, which represents the channel by means of echoes. The average power and delay are known and described by the power delay profile (PDP). They are quite simple and computationally lighter. These models can be relatively versatile to describe propagation channels in different environments, but still computationally expensive[13].

Geometry Based Stochastic Models (GBSM)

Geometry-based stochastic models (GBSM) use a simplified distribution of scatterers around the transceivers to identify and emulate the real environment statistics. The model adopts a cluster-based structure, fine-tuning the cluster's parameters (number, center position, intra-cluster non-isotropic scattering degree) and can be adjusted to a variety of scenarios, reflecting more precisely the shadowing effect. However, these models cannot simulate the effects on the elevation plane, which may considerably decrease the scenario realism

Geometry Based Deterministic Propagation Models (GBDM)

Geometry-based deterministic propagation models (GBDM) characterize the channel communication using detailed scenario information. They are site-specific methods. Deterministic models are generally used in small areas where high precision is needed. Deterministic modeling allows consideration of all the elements within the environment. Deterministic models are based mainly on ray optics (ray-tracing techniques). RT methods use exhaustive simulations with the assistance of numerical methods. Due to the high computational load in their implementation, they are not related to large areas such as vehicular communications.

Over time the channel modeling has been really expensive to compute, which motivates the need for machine learning to be able to predict the values used for the channel modeling techniques.

The methodology used in machine learning is called beam training. Beam training is a function of the environment setup (user/Base Station locations, room furniture, street buildings, and trees, etc.).This method helps to characterize the user locations and environment set up in the learning models at the Base-stations.

These functions, though, are difficult to characterize by closed-form equations, as they generally convolve many parameters and are unique for every environment setup [6]. In our coordinated beamforming application, the correlation between the received signals at the same Base-station may carry important information that will be lost if a per-carrier normalization is adopted. Similarly, the correlation between the signals received at different Base-stations from the same user may carry some information about the relative location and multi-path patterns for this user and every BS. This information will be distorted when using a per-Base-station normalization. Further, the correlation between the joint multi-path patterns at the N Base stations for different user locations may carry relevant information, which will be lost when using a per-sample normalization. Therefore, it is intuitive to adopt a per-dataset normalization in our coordinated beamforming application to avoid losing any information that could be useful for the learning model. [2]

DATASET DESCRIPTION

This dataset scenario has 18 base stations and more than one million users, which generates a sufficiently large dataset for several mmWave/massive MIMO. We use the DeepMIMO dataset to construct the inputs/outputs of the machine learning model and generate the mmWave beam prediction results using a machine learning code. The channels generated using ray-tracing simulations capture the geometry-based characteristics, such as the correlation between the channels at different locations, and the dependence on the materials of the various elements of the environment, e.t.c. [3]

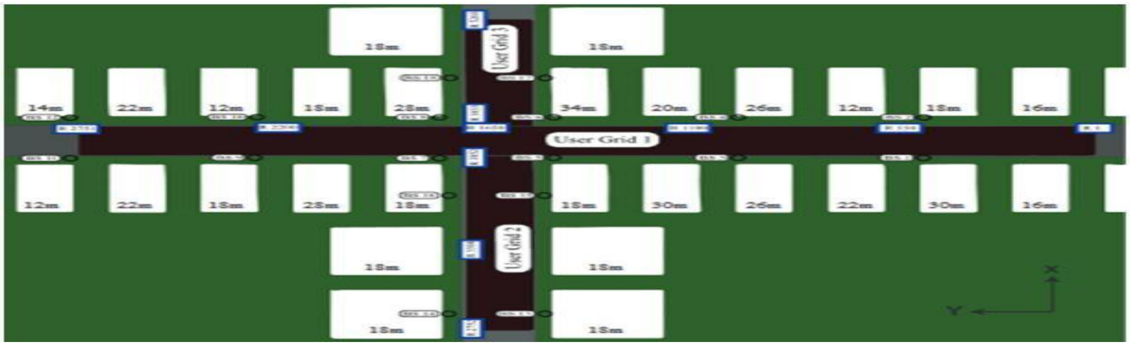

Image source: https://www.semanticscholar.org/paper/DeepMIMO%3A-A-Generic%Deep-Learning-Dataset-for-Wave%Alkhateeb/d1d8980d04d411c314910d1926f3bbaac46c2197/figure/1

METHODOLOGY

Step1: I downloaded the Ray tracing dataset the files necessary to create the source code algorithm folder and scenario (60GHz - 01-scenario) available on the authors of Deepmimo.net

Step2: I Unzipped the DeepMimo code and DL beamforming folders and placed the "01" scenario folder inside the "Ray Tracing" folder of the Master code folder directory.

Step 3: I ran MATLAB and set the MATLAB path to be able to "see" the DeepMimo code (Mat Function) and Ray Tracing scenarios folder.

Step 4: Set the following Channel parameters stated by the authors of Deepmimo for the "01" scenario which are:

DeepMIMO Dataset Parameter	Value
Active BSs	3, 4, 5, 6
Active users	From row R1000 to row R1300
Number of BS Antennas	$M_x = 1, M_y = 32, M_z = 8$
Antenna spacing	d=0.5
System bandwidth	B=0.5 GHz
Number of OFDM subcarriers	1024
OFDM sampling factor	1
OFDM limit	64
Number of paths	5

Table source: https://www.semanticscholar.org/paper/DeepMIMO%3A-A-Generic-DeepLearning-Dataset-for-WaveAlkhateeb/d1d8980d04d411c314910d1926f3bbaac46c2197#extracted

Step 5: After setting up the channel parameters, I ran the Dataset generation code. A file called Dataset.mat was produced.

Step 6: Once you have created the dataset, run the DL model beamforming.mat code to generate the DL model (DLinput.mat and DLoutput.mat).

Step 7: The neural network model has 6 fully connected layers. The fully-connected layers use ReLU activation units and every layer is followed by a drop-out regulation layer with a dropout rate of 0.5%, batch size of 100. Tensorflow framework inside google colab was used to build, train and test your model. I ran DL model python code (DLinput.mat and DLoutput.mat) placed in the DLCB_code_output folder.

Step 8: The Deep learning result files were converted to a .mat and sent to Matlab to be read by the figure_generator.m source code file. Before running the figure_generator.m code, I made sure that the DLCB_code_output folder had the result files inside and the directory was visible to MATLAB before running.

RESULTS

MATLAB figure generator code plotted my results in a graph and this is shown below.

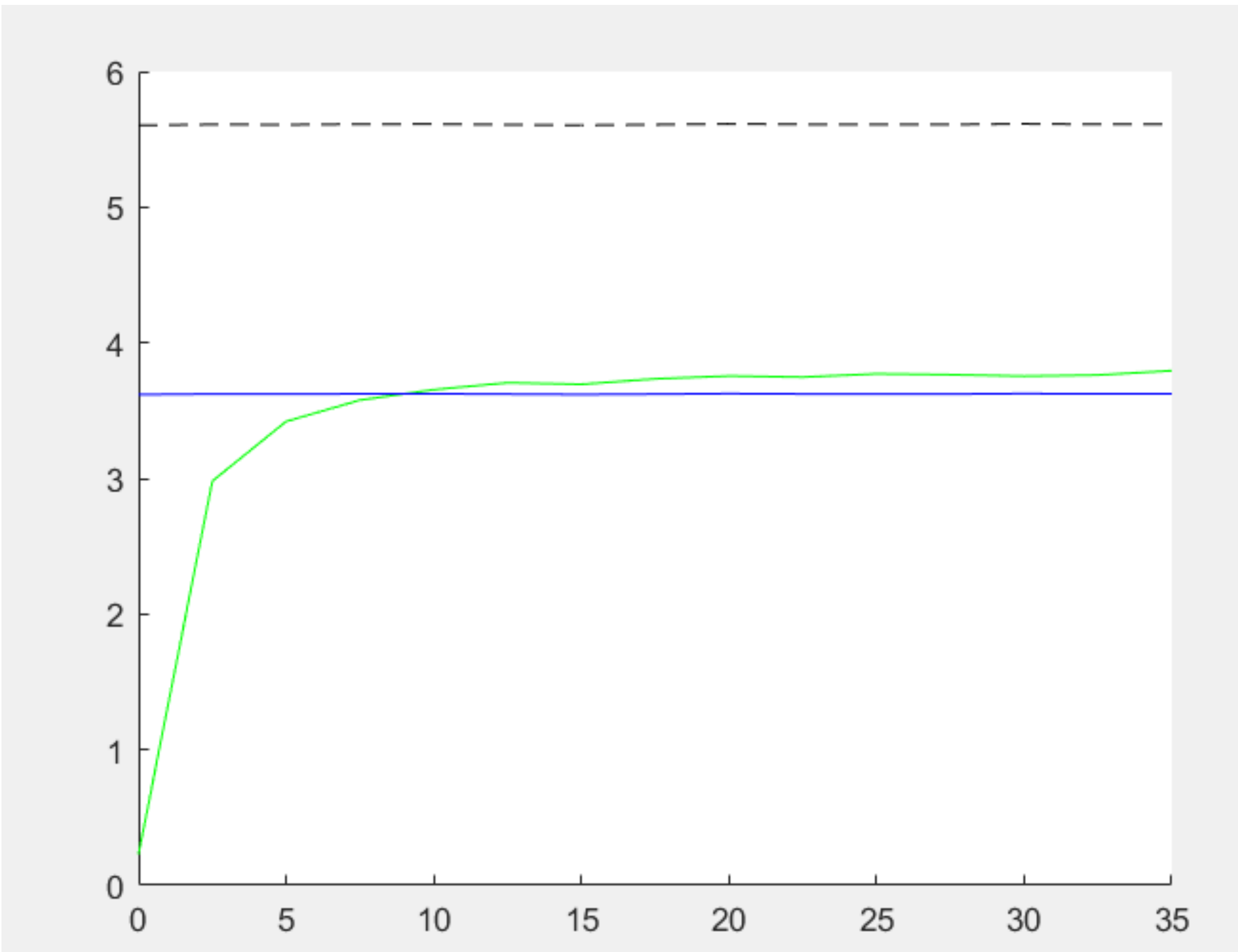

Figure 2:The x-axis is the size of the dataset, as it increases the model becomes more accurate and is able to reach and surpass the baseline coordinated beamforming solution.

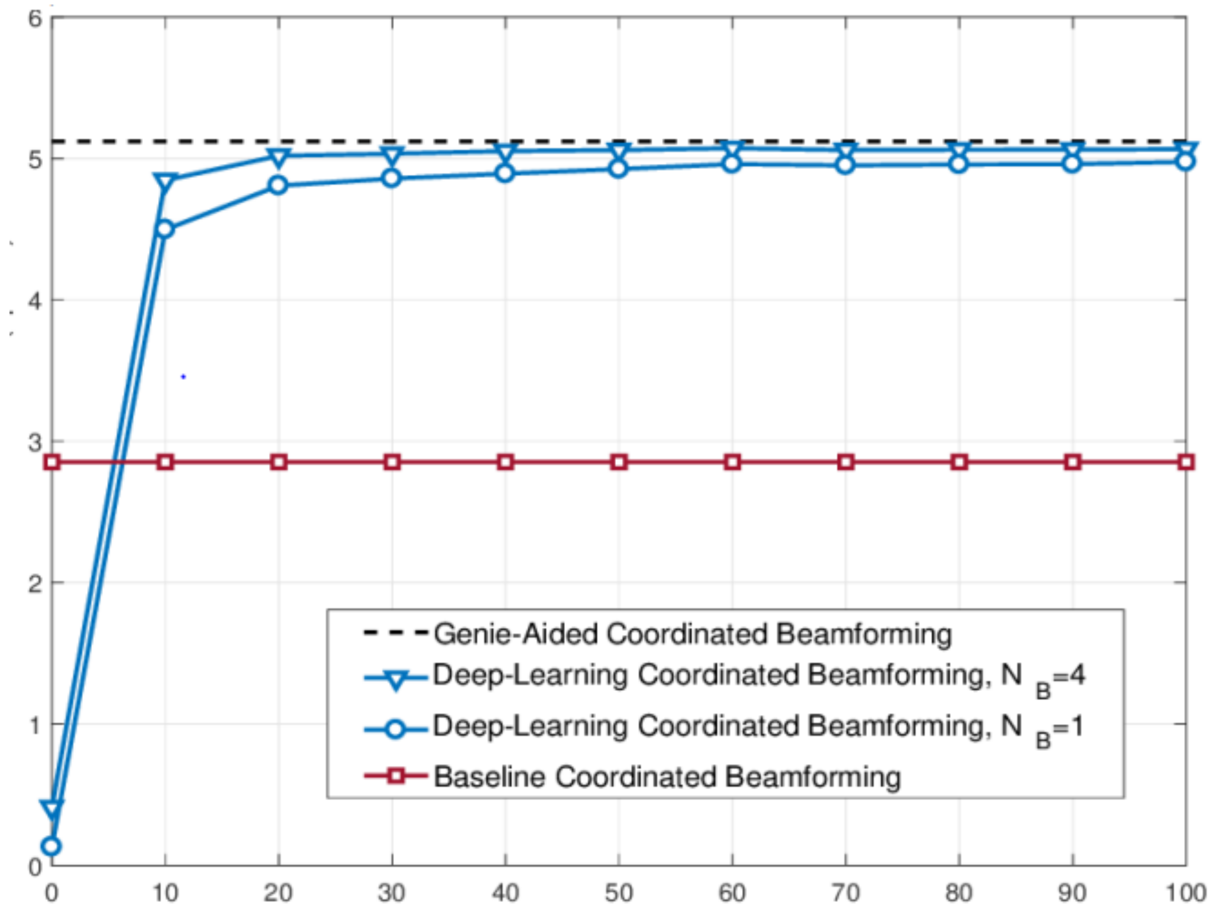

Image source: https://www.semanticscholar.org/paper/Deep-Learning-Coordinated-Beamforming-for-Wave-AlkhateebAlex/cc779ba71eabdf68944051eea852a24eaf79092c

Figure 3: This graph indicates that it would be right alongside these two beamforming solutions, similar in result. The two solutions on the graph belonging to the author and the reproduced graph from this project indicate deep learning efficiency can reach the highest achievable rate compared to the baseline solution [2].

The deep learning coordinated beamforming algorithm adopts a supervised learning model to learn the mapping between the OFDM Omni-received sequence at several Base-Stations ([3 4 5 6]) and the beamforming vector at every one of them. Every data point in the dataset that trains this deep learning model consists then of (i) the input which is the Omni-received OFDM sequence at the BSs, and (ii) the output which is the achievable rate of the candidate beamforming vectors.[3]

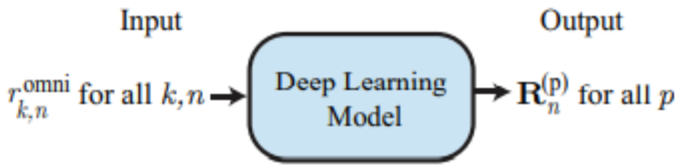

Image source: Alkhateeb, Ahmed. "DeepMIMO: A generic deep learning dataset for millimeter-waveand massive MIMO applications." *arXiv preprint arXiv:1902.06435* (2019).

Fig 4: Supervised deep learning model learns the mapping from the Omni-received sequence collected from several Base-Stations[3].

The DeepMIMO dataset generation framework constructs the MIMO channels based on ray-tracing data obtained from the accurate ray-tracing simulation, Remcom Wireless InSite. The DeepMIMO channels, therefore, capture the dependence on the various elements of the environment such as the scatterers geometry and transmitter/receiver locations, which is important for machine learning research. Further, the DeepMIMO dataset was designed to be generic, which enables the researcher to generate the dataset based on adjustable system/channel parameters. [3]

CONCLUSION

For every transmitter-receiver pair, this ray-tracing simulation shoots hundreds of rays in all directions from the transmitter and records the strongest 25 paths from those that made their ways to the receiver, where the strongest paths are the paths with the highest receive power. This would help create a report on the effects of these parameters in the set S, such as the antenna configuration and orientation on the ray-tracing scenarios.[3]

The deep-learning coordinated beamforming solution works efficiently with large antenna arrays. This is a key advantage of our developed deep-learning-based solution over traditional mmWave channel training/estimation techniques such as analog beam training. The drawback of this is that beam training takes computation time and makes it less efficient in highly mobile applications.[4]

Overall, the coordinated beamforming solution has a lot of room for efficiency improvements, which is a motivation for further research in other Neural network model architectures for the mmWave and massive MIMO systems[2].

REFERENCE

1. Karedal, Johan, Fredrik Tufvesson, Taimoor Abbas, Oliver Klemp, Alexander Paier, Laura Bernadó, and Andreas F. Molisch. "Radio channel measurements at street intersections for vehicle-to-vehicle safety applications." In *2010 IEEE 71st Vehicular Technology Conference*, pp. 1-5. IEEE, 2010.tocan
2. Alkhateeb, Ahmed, Sam Alex, Paul Varkey, Ying Li, Qi Qu, and Djordje Tujkovic. "Deep learning coordinated beamforming for highly-mobile millimeter wave systems." *IEEE Access* 6 (2018): 37328-37348.
3. Alkhateeb, Ahmed. "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications." *arXiv preprint arXiv:1902.06435* (2019).
4. Singh, Manjit, and Hisham A. Kholidy. "? Generic Datasets, Beamforming Vectors Prediction of 5G Cellular Networks." (2020).
5. Viriyasitavat, Wantanee, Mate Boban, Hsin-Mu Tsai, and Athanasios Vasilakos. "Vehicular communications: Survey and challenges of channel and propagation models." *IEEE Vehicular Technology Magazine* 10, no. 2 (2015): 55-66.
6. MacHardy, Zachary, Ashiq Khan, Kazuaki Obana, and Shigeru Iwashina. "V2X access technologies: Regulation, research, and remaining challenges." *IEEE Communications Surveys & Tutorials* 20, no. 3 (2018): 1858-1877.
7. Liang, Le, Haixia Peng, Geoffrey Ye Li, and Xuemin Shen. "Vehicular communications: A physical layer perspective." *IEEE Transactions on Vehicular Technology* 66, no. 12 (2017): 10647-10659.
8. Rodríguez-Corbo, Fidel Alejandro, Leyre Azpilicueta, Mikel Celaya-Echarri, Ana Vazquez Alejos, and Francisco Falcone. "Propagation Models in Vehicular Communications." *IEEE Access* 9 (2021): 15902-15913.
9. Viriyasitavat, Wantanee, Mate Boban, Hsin-Mu Tsai, and Athanasios Vasilakos. "Vehicular communications: Survey and challenges of channel and propagation models." *IEEE Vehicular Technology Magazine* 10, no. 2 (2015): 55-66.

10. Eldowek, Basim Mohammed, Nariman Abd El-Salam Bauomy, El-Sayed M. El-Rabaie, Saied M. Abd El-Atty, and Fathi E. Abd El-Samie. "A survey of 5G millimeter wave, massive multiple-input multiple-output, and vehicle-to-vehicle channel measurements and models." *International Journal of Communication Systems* 34, no. 16 (2021): e4830.
11. Viriyasitavat, Wantanee, Mate Boban, Hsin-Mu Tsai, and Athanasios Vasilakos. "Vehicular communications: Survey and challenges of channel and propagation models." *IEEE Vehicular Technology Magazine* 10, no. 2 (2015): 55-66.
12. Renaudin, Olivier, Veli-Matti Kolmonen, Pertti Vainikainen, and Claude Oestges. "Car-to-car channel models based on wideband MIMO measurements at 5.3 GHz." In *2009 3rd European Conference on Antennas and Propagation*, pp. 635-639. IEEE, 2009.
13. Chebil, Jalel, Hanene Zormati, and Jamel Belhadj Taher. "Geometry-Based Channel Modelling for Vehicle-to-Vehicle Communication: A Review." *International Journal of Antennas and Propagation* 2021 (2021).
14. He, Ruisi, Christian Schneider, Bo Ai, Gongpu Wang, Zhangdui Zhong, Diego A. Dupleich, Reiner S. Thomae, Mate Boban, Jian Luo, and Yunyong Zhang. "Propagation channels of 5G millimeter-wave vehicle-to-vehicle communications: Recent advances and future challenges."*IEEE vehicular technology magazine* 15, no. 1 (2019): 16-26.

    Renaudin O, Kolmonen VM, Vainikainen P, Oestges C. Car-to-car channel models based on wideband MIMO measurements at 5.3 GHz. In 2009 3rd European Conference on Antennas and
15. Renaudin O, Kolmonen VM, Vainikainen P, Oestges C. Car-to-car channel models based on wideband MIMO measurements at 5.3 GHz. In 2009 3rd European Conference on Antennas and