# Adapting DeepMIMO for 5G Networks: A New Approach to Ray-Tracing Based Beam Prediction

Ritika[1], Aninda Paul[2], Ayush Kumar[3], Ved Prakash Meena[4]
*EC431 - 5G Communication and Network*
*Indian Institute of Information Technology Vadodara - International Campus Diu*
Emails: {202211074, 202211002, 202211008, 202211096 }@diu.iiitvadodara.ac.in

***Abstract*—Recent advances in machine learning have unlocked promising capabilities for addressing complex challenges in millimeter wave (mmWave) and massive MIMO systems. These techniques excel at modeling intricate environments and solving demanding optimization problems. However, advancing research in mmWave/massive MIMO requires a unified dataset for evaluating algorithm performance, reproducing results, establishing benchmarks, and comparing diverse approaches. In this work, we present the DeepMIMO dataset [1]—an adaptable and generic dataset specifically designed for mmWave/massive MIMO channel studies. The dataset generation framework embodies two key elements. First, the channels are derived from accurate ray-tracing data acquired from Remcom Wireless InSite [2], ensuring they capture the impact of environmental geometry, material properties, and transmitter/receiver locations—features crucial for various machine learning applications. Second, the dataset is fully parameterized, allowing researchers to adjust system and channel parameters to tailor the dataset for their specific applications. An exemplary implementation based on an outdoor scenario with 18 base stations and over one million users is discussed.**

## I. Introduction

The rapid evolution of 5G networks has driven the need for accurate channel modeling and efficient beam prediction algorithms. Machine learning techniques are increasingly employed to tackle the challenges of millimeter wave (mmWave) and massive MIMO systems. A key enabler for this research is the availability of realistic, customizable datasets. The DeepMIMO dataset, introduced by Alkhateeb et al. [1], provides a flexible platform generated using ray-tracing data from Remcom Wireless InSite [2]. In this project, the DeepMIMO dataset generation framework is implemented for an outdoor scenario with 18 base stations and over one million users.

## II. Implementation Details

### A. Dataset Generation Framework

The DeepMIMO dataset framework rests on two principal components:

- **Ray-Tracing Scenario:** The dataset is defined using outputs from ray-tracing simulations that capture the environmental geometry, material properties, and the spatial distribution of both base stations and users.
- **Parameter Set (S):** Researchers can customize the dataset by setting parameters such as the number of BS antennas, antenna spacing, system bandwidth, OFDM subcarriers, and the number of channel paths.

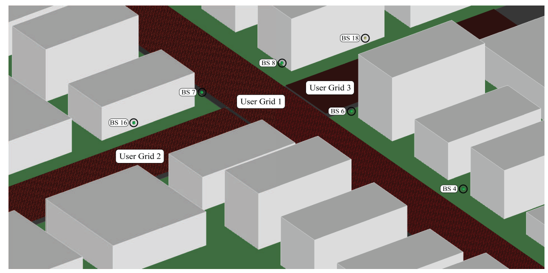


Fig. 1. A top view of the 'O1' ray-tracing scenario, showing the two streets, the buildings, the 18 base stations, and the user x-y grids. The scenario is generated using Remcom Wireless InSite [2].

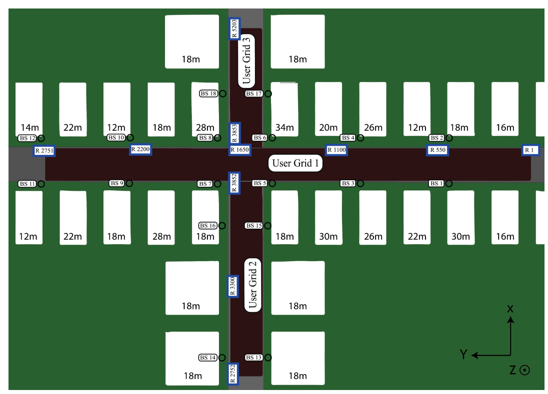


Fig. 2. A bird's-eye view of a section of the 'O1' ray-tracing scenario, showing the intersection of the two streets. The scenario is generated using Remcom Wireless InSite [2].

For this project, the following parameters were configured:

- **Active Base Stations:** [3, 4, 5, 6]
- **Active Users:** Selected rows from a single user grid for a manageable dataset size.
- **BS Antenna Configuration:** $M_x = 1$, $M_y = 32$, $M_z = 8$
- **Antenna Spacing:** 0.5 wavelengths
- **System Bandwidth:** 0.5 GHz
- **OFDM Parameters:** 1024 subcarriers, with a sampling factor of 1 and channel computation on the first 64 subcarriers.
- **Channel Paths:** 5 strongest paths per BS–user pair.

### B. Channel Generation and Preprocessing

Channel matrices are generated for each active BS–user pair using MATLAB scripts provided by the DeepMIMO framework. The channel vector for each subcarrier is modeled as

$$\mathbf{h}_{b,u,k} = \sum_{\ell=1}^{L} \sqrt{\frac{\rho_\ell}{K}}\, e^{j\left(\vartheta_{b,u,\ell} + 2\pi \frac{k}{K} \tau_{b,u,\ell} B\right)}\, \mathbf{a}\left(\phi_{b,u,\mathrm{az}}, \phi_{b,u,\mathrm{el}}\right), \tag{1}$$

where $\mathbf{a}(\cdot)$ is the array response vector that accounts for variations in angles of departure and arrival as well as path gains. This formulation ensures that the generated channels accurately represent the real-world propagation environment.

```
>> DeepMIMO_Dataset_Generator
 DeepMIMO Dataset Generation started
 Reading the channel parameters of the ray-tracing scenario O1_60
 Basestation 3
 Basestation 4
 Basestation 5
 Basestation 6
 Constructing the DeepMIMO Dataset for BS 3 - Percentage completed: 100.0
 Constructing the DeepMIMO Dataset for BS 4 - Percentage completed: 100.0
 Constructing the DeepMIMO Dataset for BS 5 - Percentage completed: 100.0
 Constructing the DeepMIMO Dataset for BS 6 - Percentage completed: 100.0
 DeepMIMO Dataset Generation completed

ans =

    56     4    32    32
```

Fig. 3. DeepMIMO dataset generation process log.

## III. Project Outputs and Results

The implementation successfully generated the DeepMIMO dataset using the specified configuration. The output is stored in a single `.mat` file named `DeepMIMO_dataset.mat`, which contains a cell array named `DeepMIMO_dataset`. This array holds the channel data and location information for each active base station–user pair.

```
>> DeepMIMO_dataset{1}.user{1}

ans = 

  struct with fields:

        channel: [56×4×32×32 double]
            loc: [242.4230 297.1710 2]
     LoS_status: 1
       distance: 192.4990
       pathloss: 112.8330
    path_params: [1×1 struct]
```

Fig. 4. DeepMIMO dataset structure preview, illustrating how each base station–user pair is stored.

The dataset structure is organized as follows:

- **Channel Matrix:** For each active base station $b$ and user $u$, the dataset entry `DeepMIMO_dataset{b}.user{u}.channel` holds an $M \times K$ matrix of channel coefficients across the selected subcarriers.
- **User Location:** The location of each active user is stored as a position vector in `DeepMIMO_dataset{b}.user{u}.loc`.

For example, if the active base stations are set as `[3, 4, 5, 6]` and the active user row range is defined from 1000 to 1500, then:

- `DeepMIMO_dataset{1}.user{1}.channel` accesses the channel between the first active BS (BS3) and the first user in row 1000.
- `DeepMIMO_dataset{1}.user{1}.loc` accesses the spatial coordinates of that user.

This output format enables easy indexing and efficient access to specific BS–user pairs, which is essential for integrating the dataset into machine learning workflows.

## IV. Learnings and Discussion

The project offered several critical insights:

- **Parameter Customization:** Flexibility in adjusting system parameters ensures that the dataset can be tailored to specific research needs.
- **Reproducibility:** Defining the dataset entirely through the ray-tracing scenario and its parameter set facilitates reproducible experiments, a key aspect for benchmarking.
- **Practical Applications:** Although not integrated into a deep learning pipeline in this project, the generated dataset paves the way for future experiments in beam prediction and other 5G network optimizations.
- **Computational Considerations:** Handling large user grids and multiple channel paths is computationally intensive; strategic approaches such as downsampling or focusing on representative subsets can mitigate these challenges.

## V. Conclusion

This work demonstrates the feasibility of adapting the DeepMIMO dataset framework for deep learning-based beam prediction in 5G networks. By leveraging customizable, ray-tracing-based channel models, the generated dataset enables effective training and benchmarking of beam prediction algorithms. The project underscores the importance of reproducible and flexible datasets in advancing research in mmWave and massive MIMO systems.

## References


[1] A. Alkhateeb, "DeepMIMO: A Generic Deep Learning Dataset for Millimeter Wave and Massive MIMO Applications," available online: http://www.DeepMIMO.net.

[2] Remcom, "Wireless InSite," available online: http://www.remcom.com/wireless-insite.

[3] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," IEEE Communications Magazine, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[4] V. Jungnickel *et al.*, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," IEEE Communications Magazine, vol. 52, no. 5, pp. 44–51, 2014.

[5] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" IEEE Transactions on Wireless Communications, vol. 15, no. 2, pp. 1293–1308, 2016.