Generation of Realistic OTFS Wireless Channel Datasets for Machine Learning Applications using QuaDRiGa

1. Introduction & Motivation:

The performance of next-generation wireless communication systems, such as those employing Orthogonal Time Frequency Space (OTFS) modulation, heavily relies on accurate Channel State Information (CSI). Machine learning (ML) and Artificial Intelligence (AI) offer promising avenues for enhanced channel estimation and prediction. However, training such AI models requires vast amounts of high-quality, realistic channel data. Real-world channel measurements are expensive and time-consuming to acquire. Therefore, sophisticated channel simulators like QuaDRiGa become indispensable tools for generating synthetic datasets that capture the complex dynamics of wireless channels, especially in fast-varying environments. This project focuses on building such a dataset, serving as the foundation for subsequent AI/ML research in OTFS.

2. Project Goal:

The primary goal of this project is to develop a comprehensive MATLAB-based simulation framework using QuaDRiGa to generate a diverse dataset of paired channel features (e.g., Path Loss, LOS/NLOS status, SNR, Doppler spread) and their corresponding OTFS Delay-Doppler channel coefficients. This dataset will be suitable for training and evaluating machine learning models.

3. Specific Objectives:

Upon successful completion of this project, the student will be able to:

- Familiarize themselves with the fundamentals of wireless channel modeling, OTFS modulation, and the QuaDRiGa channel simulator.
- Design and configure diverse wireless communication scenarios (e.g., urban microcell, indoor, vehicular) incorporating varying mobility, environments, and antenna configurations.
- Develop MATLAB scripts to automate channel simulations, ensuring consistency and reproducibility.

- Extract relevant channel parameters/features (e.g., Path Loss, Line-of-Sight (LOS)/Non-Line-of-Sight (NLOS) indicator, Doppler Spread, Delay Spread, Tx/Rx positions/speeds, SNR estimates) from the simulated environments.
- Accurately derive the complex-valued OTFS Delay-Doppler channel coefficients from the simulated Channel Impulse Responses (CIRs).
- Structure and store the generated paired dataset in a suitable format (e.g., .mat, .hdf5) for subsequent machine learning model training in Python.
- Perform basic statistical analysis and visualization of the generated dataset to verify its realism and diversity.

4. Methodology & Tasks:

1. Literature Review (Weeks 1-3):

- Study fundamental concepts of wireless channel propagation, multipath fading, Doppler effect, and delay spread.
- Understand the basics of OTFS modulation and its transformation from the time-delay domain to the delay-Doppler domain.
- Review QuaDRiGa documentation and examples for setting up scenarios and extracting channel information.
- Investigate existing channel measurement campaigns and dataset structures.

2. QuaDRiGa Setup & Familiarization (Weeks 4-5):

- Install and configure QuaDRiGa in MATLAB.
- Run basic examples to understand its functionality, scenario definitions, and output structures (e.g., CIRs).

3. Scenario Definition (Weeks 6-8):

- Propose at least 3-5 distinct and diverse wireless scenarios (e.g., stationary indoor, pedestrian urban, vehicular highway).
- Define key parameters for each scenario: carrier frequency, bandwidth, antenna configurations (MIMO), Tx/Rx positions and speeds, environmental characteristics.
- Specify the number of channel snapshots and simulation duration for each scenario to capture relevant time-variations.

4. MATLAB Script Development (Weeks 9-14):

- Develop modular MATLAB code to:
  - Initialize QuaDRiGa layout and configure chosen scenarios.
  - Run iterative simulations to generate a large number of channel snapshots.
  - For each snapshot:
    - Extract low-level channel features: Tx/Rx coordinates, speeds, LOS/NLOS status (if directly available), estimated Doppler spread, estimated delay spread, average path loss.
    - Generate Channel Impulse Responses (CIRs).
    - Transform the CIRs into OTFS Delay-Doppler domain channel coefficients.
    - Calculate SNR based on simulated received power and a defined noise floor.

5. Dataset Structuring & Storage (Weeks 15-16):

- Organize the extracted channel features and corresponding OTFS channel coefficients into a consistent, paired dataset structure.
- Implement efficient storage methods (e.g., MATLAB .mat files or HDF5 files) to facilitate easy loading in Python for the subsequent ML project.
- Ensure proper labeling and documentation of each data point.

6. Data Analysis & Validation (Weeks 17-18):

- Perform statistical analysis (e.g., histograms, correlations) on the generated features and coefficients.
- Visualize key channel characteristics to ensure the generated data aligns with expected physical properties for the defined scenarios.
- Identify any biases or inconsistencies in the generated dataset.

5. Expected Outcomes:

- A well-documented MATLAB codebase for generating OTFS channel datasets using QuaDRiGa.
- A comprehensive, high-fidelity dataset containing at least 1000-5000 (or more, depending on complexity) paired examples of channel features and corresponding OTFS Delay-Doppler channel coefficients.

- A written bachelor thesis detailing the methodology, simulation parameters, dataset characteristics, and analysis results.
- A presentation summarizing the project's findings.

6. Required Skills:

- Strong proficiency in MATLAB programming.
- Basic understanding of wireless communication fundamentals (propagation, fading, MIMO).
- Familiarity with signal processing concepts (FFT, time-frequency analysis).
- Ability to read and understand technical documentation.