PREDICTION OF 5G NEW RADIO WIRELESS CHANNEL PATH GAINS AND DELAYS USING MACHINE LEARNING AND CSI FEEDBACK

Ben Earle
Ala'a Al-Habashna
Gabriel Wainer

System and Computer Engineering
Carleton University
1125 Colonel By Drive, Ottawa, ON, CANADA
benearle@cmail.carleton.ca,
{alaaalhabashna, gwainer}@sce.carleton.ca

Xingliang Li
Guoqiang Xue

Ericsson Canada
349 Terry Fox Drive, Ottawa, ON, CANADA
{xingliang.li, guoqiang.xue}@ericsson.com

ABSTRACT

Next generation wireless communication systems use massive Multi Input Multi Output (m-MIMO) antenna arrays for their enhanced beamforming capabilities. Providing accurate Channel State Information (CSI) is vital for optimizing m-MIMO communication systems. The complexity of channel reconstruction grows exponentially with the number of antennas, causing traditional methods to become increasingly complicated. Machine-learning techniques can be a useful alternative for channel reconstruction using partial CSI feedback. This paper presents the results of a simulation study built using the MATLAB 5G Toolbox and a neural network trained using the simulated data. The simulator emulates a 5G channel to generate its path delays and gains, and the realistic CSI feedback. This data was used to train and test a neural network to estimate the dominant path gains and delays. The models showed promising results while operating on limited CSI data.

Keywords: Machine Learning, Neural Networks, 5G, Channel Reconstruction, Simulation.

1 INTRODUCTION

Telecommunication systems demand constant innovation, and the requirements grow exponentially with each generation of the wireless technology. One of the ways the 5^{th} Generation (5G) wireless systems respond to these growing requirements is using massive Multi-Input Multi-Output (m-MIMO) antennas (Dahlman, Parkvall, and Skold 2018). These systems have the potential to improve the channel bandwidth, coverage, and capacity through beamforming and spatial multiplexing. To take advantage of these perks, the systems require accurate Channel State Information (CSI), which are metrics that describe how the channel will affect the signal. However, increasing the number of antennas increases the complexity and overhead when measuring the CSI, known as channel sounding (Mawatwal, Sen, and Roy 2020). These computations are built using a matrix of complex numbers, called the channel matrix, in which each element includes a gain and angle which, when multiplied by the transmitted signal will estimate the received signal (without any noise). The channel matrix grows exponentially with the addition of new antennas, further complicating channel reconstruction algorithms that try to rebuild the channel matrix based on CSI and/or received pilots. Classical algorithms will fail to meet the real-time constraints in these massive antenna systems (Li et al. 2020).

ANNSIM '21, July 19-22, 2021, Fairfax, VA, USA; ©2021 Society for Modeling & Simulation International (SCS)

Machine Learning (ML) is an increasingly popular approach for improving the accuracy, feedback overhead, and runtime of channel reconstruction (Ye, Li, and Juang 2018). 5G wireless communication systems are complex, making it difficult to implement and maintain new algorithms. Additionally, classical algorithms such as those described in (Han et al. 2019) and (Liu, Lau, and Dai 2106) have very high computational complexity. Alternatively, data-driven models eliminate the analytical complexity in the conventional methods above. ML can easily identify trends in multi-dimensional data in ways that humans cannot. Although, training a model is computationally taxing and requires a lot of data, once the model has been trained, it is very fast to execute (Li et al. 2020). ML solutions are often faster to execute and with equivalent performance, if not better, than existing solutions.

This research proposes a method for channel reconstruction from limited CSI using NN. The data was generated using a simulator built with the MATLAB 5G Toolbox (MathWorks 2020). The MATLAB 5G toolbox provides a library of functions and examples for simulating the 5G NR wireless channels. This includes various physical layer models to simulate the functionality of the transmitter and receiver (e.g., modulation and encoding) as well as the different effects of the wireless channel. The data is generated by simulating a wireless channel and calculating the CSI, which is recorded along with the path gains and angles. A path is a physical route through the channel traveled by the wireless signal; a path is characterized by its delay, angle, and gain. A combination of the 5G toolbox functions and custom functions calculate the CSI, including the Channel Quality Indicator (CQI), Received Signal Strength Indicator (RSSI), Precoding Matrix Indicator (PMI), and Rank Indicator (RI). These metrics are commonly calculated in practice and will be used as the input features for the ML model. The data from multiple simulations is aggregated and stored to be used to train a Neural Network (NN) in the statistical programing language R. The NN models predict the dominant path gains and delays from partial CSI feedback. Many different NN were tested using different combinations of the CSI to determine which would give optimal results. The predicted path gains and delays can be used to reconstruct the channel. Results show that the accuracy of the NN is promising while operating with limited CSI.

The remainder of this paper is organized as follows. First, the background section defines the wireless channel model, MIMO systems, describe the CSI used in the NN and other channel reconstruction techniques. Section 3 explains the experiment and the NN model used. Section 4 describes the simulator, along with its configurable components and outputs. Section 5 presents the output of the simulator and the trained NN's channel path predictions. Finally, Section 6 concludes the paper.

2 BACKGROUND AND RELATED WORK

This section provides an overview of the wireless communication and ML topics involved in this work. It includes a definition of wireless channels, beamforming and m-MIMO antennas, CSI, Orthogonal Frequency Division Multiplexing (OFDM) transmission grids, and ML.

2.1 Wireless Channels

A channel is a physical transmission medium. In the case of wireless communication, the channel is the effects of the environment on the radio waves as they travel between transmitter and receiver. Wireless channels can be impacted by many factors: pathloss, shadowing and multipath effects. The pathloss is the attenuation due to the distance between the transmitter and the receiver. Free-Space Path Loss (FSPL) is the simplest form of pathloss, shown in the following equation (Tse and Viswanat 2004),

$$FSPL = \left(\frac{4\pi df}{c}\right)^2 \tag{1}$$

where d is the distance, f is the carrier frequency, and c is the speed of light. The attenuation increases exponentially with distance. The channel's effect on the transmitted signal is represented as a matrix, if known, the transmitter and receiver can work around the physical limitations in the channel to communicate more effectively. $Y = H * S + Z$ shows how the received signal is related to the transmitted signal. Let N

be the number of subcarriers. M_{TX} and M_{RX} are the number of antennas on the transmitter and receiver, respectively. S is the transmitted signal which has the dimension $[M_{TX} \times N]$. H is the channel matrix and has dimension $[M_{RX} \times M_{TX}]$. Z is the additive noise, and Y is the received signal, both with dimension $[M_{RX} \times N]$. A wireless channel will apply some form of attenuation and phase shift to each of the transmitted symbols.

Wireless signals are also impacted by objects in their path, causing them to experience reflection, refraction, and dispersion. Shadowing is the attenuation caused by obstacles in the path between the transmitter and receiver. Signals are not limited to following a single path to the receiver, which also affects the channel; often referred to as multipath effects. Each path will experience a different channel, as they may have different obstacles and/or distances. This will cause the signals to take a different amount of time to propagate to the receiver resulting in a different delay, phase, and gain for each of these paths. When adding signals that are out of phase, they can either add constructively or destructively. Moving the receiver slightly can change the relative phase between signals causing a significant change in received power. The channel matrix can be reconstructed with knowledge of the path gains, angles, and delays; shown in (Li et al. 2020) and (Han et al. 2019).

2.1.1 Tapped Delay Line Channel

Tapped Delay Line (TDL) is a multipath channel model defined by the 3rd Generation Partnership Project (3GPP) for link level simulations. The MATLAB 5G Toolbox implements a TDL channel model following the standard definitions found in (3GPP 2020). This channel has five premade profiles labeled TDL-A through TDL-E. The A-C models are for Non-Line Of Sight (NLOS) applications, whereas D and E are for Line Of Sight (LOS). A TDL channel profile has a predetermined path, and each of the delays and gains are constant for that profile. However, the channel profiles can be scaled in delay spread and doppler shift.

The Delay Spread of a channel is the difference between the propagation delays on each channel path. The Delay Spread of a channel is the difference between propagation delays on each channel path. Delay spread is inversely related with the coherence bandwidth, which defines the range of frequencies which have similar fading patterns. The longer the delay spread the more frequency selective fading a channel will experience and the more each transmitted symbol will interference with the future symbols; known as inter symbol interference (Linnartz, Delay Spread 2021). The Delay Spread is offered as a parameter to the TDL channel to scale the path delays in its default profiles. 3GPP suggest reasonable values for delay spreads, shown in table 1. They suggest using a shorter delay spread of LOS applications and longer delay spreads for NLOS paths. Indoor applications also typically have shorter delay spreads compared to an urban macro cell which will have very long delays.

Table 1: Delay Spread Range (3GPP 2020).

Model	Delay Spread
Very short	10 ns
Short	30 ns
Nominal	100 ns
Long	300 ns
Very long	1000 ns

The Doppler Shift represents a change in the frequency of a wave due to the relative velocity between the wave's source and its receiver (Tse and Viswanat 2004). This effect applies to both electromagnetic and sound waves (this is clearly heard in the noise that vehicles make when driving by quickly: when they are approaching, the sound waves are compressed, resulting in a higher frequency noise, and once it passes the observer, the sound waves are stretched, resulting in lower frequency noise). In the case of wireless systems, there is a Doppler Shift when the User Equipment (UE) is in motion relative to the Base Station (BS). The following equation shows how the doppler shift is calculated.

$$doppler\ shift = \frac{f \cdot v}{c} \tag{2}$$

The carrier frequency is f, the velocity of the UE is v, and the speed of light is c. This equation is used to calculate a reasonable doppler shift for the simulation. The relative velocity depends on the angle the receiver's direction of travel, in multipath channels each path may arrive with a different angle. This means that the Doppler Shift for each path will be different. The TDL channel model uses a specified maximum Doppler Shift to calculate the Doppler Spread, which accounts for the difference in the Doppler Shifts between paths (Linnartz, 2021).

2.1.2 Channel State Information

Reference signals known to both the transmitter and receiver are sent through the channel and analyzed by the receiver, which can predict the channel effects based on the difference between the known signal and the received signal. The receiver uses the reference signal to calculate a variety of CSI which is sent back to the transmitter and used to optimize the communications. The CSI metrics considered in this research are SNR, RSSI, CQI, PMI, and RI. Signal to Noise Ratio (SNR) describes how much more powerful the signal is compared to the background noise. The higher the SNR, the easier it is to distinguish the signal from the noise. The transmitter uses the SNR to determine which modulation scheme to use, higher SNR will support more bits per second you have not said much about the "symbol". RSSI simply returns the signal power measured at the receiver in negative dB, it typically ranges from 0-127. The CQI is an index into a list of SNR ranges. The SNR is measured and compared to a list of 15 ranges, the CQI is the index to the range that matches the SNR. The higher the CQI, the better the SNR range. PMI is used in MIMO systems; it is an index into a list of potential precoding matrices. The matrix selected will be multiplied by the signals to be transmitted, mapping and scaling them for the appropriate antennas. The PMI is related to the angles of the channel and is used for beamforming. Finally, RI is related to the rank of the channel matrix, if the RI is only rank 1 that means there is one independent path through the signal. The RI is used to know how many independent paths there are between the transmitter and receiver. If paths are independent, then the system can transmit different data streams over them and still recover the data. The transmitter will consider the PMI and RI together to choose the appropriate precoder matrix.

2.1.3 Orthogonal Frequency Division Multiplexing (OFDM)

OFDM is the most common modulation scheme in modern communication systems. It splits the available frequency bandwidth into many narrow band subcarriers and transmits many symbols in parallel at a lower symbol rate. This approach avoids the extra complications caused by high data rates while maintaining the same throughput, achieving high spectral efficiency. An OFDM grid is used when scheduling OFDM transmissions. The grid breaks the available bandwidth into subcarriers and resource blocks, which are a collection of subcarriers (Zaidi et al. 2018). In the time domain each element represents an OFDM symbol and a collection of 14 symbols makes a slot. A single element is called a resource element, although blocks and slots are the smallest schedulable units. Each resource element will have a different channel effect on each of the transmit antennas. The channel matrix will be four dimensioned: frequency subcarrier × time slot × receive antenna × transmit antenna. The OFDM grid will be populated with various symbols, such as, reference, control, or data signals. The UE and the BS will agree on which symbols will be used for what transmission in advance.

2.2 Machine Learning Applications in 5G Systems

Machine Learning (ML) models are adaptive algorithms that make predictions based on a set of input features. A feature is any combination of measurements on the object of study. Due to the complex nature of 5G wireless systems, ML based approaches have become increasingly popular (Luo et al. 2020). A ML model must be "trained" using example data, including sample inputs and desired outputs. During training, the model tunes its internal parameters to increase the likelihood of correctly predicting the desired outputs. Next, it is important to test the model's accuracy. This is done by providing the model with inputs it has never seen and comparing the predictions with the expected outputs. There are two types of ML problems: classification and regression. Classification problems focus on fitting the observed feature set into discrete classes. Whereas regression problems involve trying to predict continuous values.

A Neural Network (NN) is a common non-linear statistical model used for ML. The NN is built as layers of nodes converting inputs into outputs; each NN consists of an input layer, one or more hidden layers, and an output layer. The number of hidden layers and the number of nodes in the hidden layers are parameters that can be tuned (Hastie, Tibshirani, and Friedman 2017). A NN with more than one hidden layer is known as a Deep Neural Networks (DNN). The Universal Approximation Theorem proved that an NN with one hidden layer can approximate any continuous function (Winkler and Le 2016). Therefore, any resulting DNN could be recreated with a single hidden layer if given enough nodes.

3 EXPERIMENT STRUCTURE

In this study, a simulator for a 5G NR wireless system is built to generate datasets for training a NN that will predict path components using limited CSI. The simulator uses MATLAB's 5G toolbox for the channel model, signal modulation, frame structure, OFDM grid population, and some of the CSI calculations. The inputs to the simulator include channel model, carrier, frame count, doppler shift, and delay spread. The simulator takes the many parameters as inputs and is run in two ways: random input data within the acceptable range, or equally spaced data spanning the input space. The dataset generated for training spans the input space with finely grained steps and a high SNR ratio (20dB). The test set uses the random parameter generation to create unpredictable data for evaluating the model. The simulator outputs the calculated CSI, and path gains and delays present in the channel. The channel model used had a variable number of paths, with a maximum of 24. The simulation logs all the inputs and outputs for each frame in a csv file, padding the entries with less than 24 paths with delays of 0s and gains of -1000dB. Section 4 contains an in-depth description of the simulator and section 5 discusses the generated dataset.

The simulated datasets were used as training and testing data for ML models. The models' input features were the measured CSI. The model was trained to predict the path gains and delays. Once trained the models were run on the test data without knowledge of the path gains and delays. The predicted results were compared to the true values and two performance metrics were considered: Root Mean Squared Error (RMSE) and the Coefficient of Determination (R^2). The models tried to predict 24 path components for each set of CSI, although only the 10 paths with the highest power were considered when calculating performance metrics. Different combinations of CSI were tested to determine which features were best. Additionally, NN with a varying number of nodes and hidden layers were tested. The NN with a single hidden layer and 40 hidden nodes was found to perform the best. The single hidden layer is further justified by the Universal Approximation Theorem. Results for the accuracy of the NN are presented in Section 5.

4 SIMULATOR DESIGN AND IMPLEMENTATION

The channel simulator was developed to generate realistic test data to be used in this study. The channel simulator was programmed using MATLAB's 5G toolbox, which provides support for modeling, simulation, and verification of NR communication systems (MathWorks 2020). Two scripts were developed, the first simulates downlink CSI-RS, and the second is for uplink SRS symbols. This work focused on the data generated in the downlink direction. The remainder of this section will include design decisions, inputs, outputs.

4.1 Simulator Architecture

The simulator emulates a 5G wireless system transmitting reference signals and calculating the CSI. An overview of the simulator is shown in Figure 1.

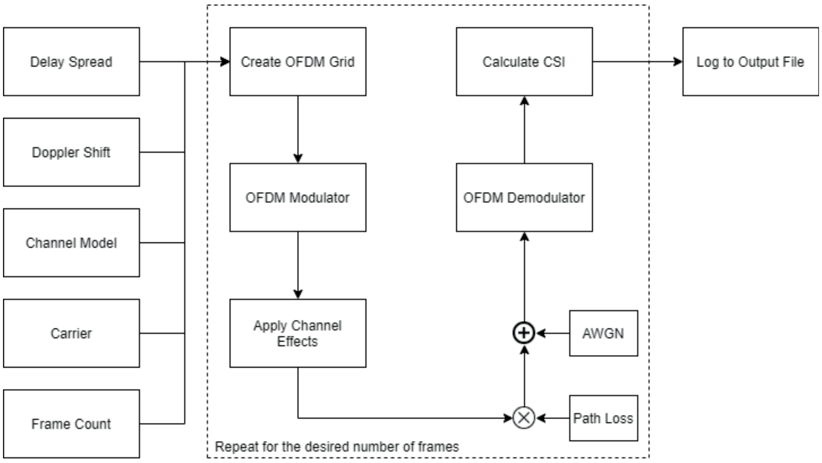

Figure 1: Simulator Block Diagram.

The simulation starts by creating an OFDM grid populated with reference symbols. The grid of symbols is modulated using MATLAB 5G Toolbox functions based on the configuration in the carrier object. The modulated signal is passed through the channel, resulting in an ideal received signal. The Free Space Pathloss equation shown in (1) is then applied to the signal to scale down the received power prior to adding Additive White Gaussian Noise. The received signal is then demodulated and used to calculate CSI. The channel conditions and resulting CSI are recorded and logged to the output file. This process is repeated for as many frames as desired.

4.2 Channel Parameters

The script generates channel parameters in two ways: the first using a desired range and steps for each parameter, simulating all their combinations; the second generating a preset amount randomly within the range. The first method is to be used when generating training data, this way the training data spans the set of all possible channel parameters. While generating test datasets, the random distributions in the second method may better represent real channel conditions. The random parameter generation is limited to a uniform distribution.

The simulated datasets have variable doppler shift, delay spread, and channel profiles. Currently the datasets focus on maximum doppler shifts in the range of 10Hz to 220Hz. This range was derived using equation (2). A theoretical maximum speed (v) of 130km/h and a carrier frequency (f) of 2GHz was used. The range of delay spreads considered started at 10ns and went to 1000ns. The range was taken from table 1 as we wanted our training data to cover all reasonable delay spreads from (3GPP 2020). The script will simulate each TDL profile (A-E) for each of the given channel spreads and shifts.

4.3 Dataset Output

The script aggregates the results from each simulation and saves them in excel spreadsheet. The headings for the dataset are shown in Table 2, along with a description of their value. The italic characters in the column names represent variables. The ‘x’ variable is the number of frequency ranges, the script simulates a system with 52 subcarriers. There are 4 subcarriers per frequency range, which goes from 0 – 12 (13x4 =

52 subcarriers). The 'a' and 'b' represent transmit and receive antennas respectively, and this variable is used for the CQI for a given antenna pairing. The 'p' is for path number, different channel variants will have a different number of paths, typically ranging from 10-25. Finally, the 'y' is for the receive-antenna number.

Table 2: Dataset Column Descriptions.

Column Name	Description
Frame	Fame number of the CSI reading
Slot	Slot number of the CSI reading
Channel	Channel model (TDL-A, CDL-C, etc...)
Tx_Ant	Number of transmit antennas
Rx_Ant	Number of receive antennas
Doppler_Shift	The maximum Doppler shift (Hz)
Delay_Spread	Desired delay spread (s)
RI_FR*x*	Rank Indicator for each Frequency Range
PMI_FR*x*	Precoder Matrix Indicator for each Frequency Range
CQI[*ab*]_FR*x*	Channel Quality Indicator for each antenna pair (inside []) and Frequency Range.
Path_Delay*p*	The path delays extracted from the channel model. The number of paths is not uniform, if the path does not exist the delay value assigned is 0.
Path_Avg_Gain*p*	The path gains extracted from the channel model. The number of paths is not uniform, if the path does not exist the gain value assigned is -1000 db.
RSRP*y*	Row vector of reported RSRP values for all CSI-RS resources, where each element represents the maximum RSRP on each receive antenna.
RSSI*y*	Row vector of reported RSSI values for all CSI-RS resources, where each element represents the maximum RSRI on each receive antenna.
RSRQ*y*	Row vector of reported RSRQ values for all CSI-RS resources, where each element represents the maximum RSRQ on each receive antenna.

5 SIMULATIONS AND OBTAINED RESULTS

This section starts with the simulation parameters and the obtained results. Thereafter, is a discussion of the developed models for channel reconstruction from CSI feedback. Finally, the results from the NN estimation of channel path gains and delays are presented.

The simulations use the TDL channel with varying profiles available in the MATLAB 5G toolbox. Each of the preset channel profiles were considered TDL-A to TDL-E, to capture the performance in a variety of channel conditions. Furthermore, the doppler shift and delay spread were varied across their range in fine grained steps. The doppler shift started as low as 30Hz, increasing by 10Hz steps until reaching the maximum value of 220Hz. The delay spread started at 10ns, increased by 10ns steps all the way to 1000ns. Each combination of the channel profiles, delay spreads, and doppler shifts was simulated for 18 frames containing reference signals. The channel state information was calculated and logged along with the simulator parameters and path components.

After collecting the data from the simulations, the data is cleaned and organized in a data frame that will be used to build the model. Tables 3 and 4 show a sample of the data frame generated. After preparing and cleaning the data frame, it is used to build a supervised learning model to predict the delays and gains of the different channel paths from the CSI feedback. NNs were used for this purpose. As can be seen from Tables 3 and 4, there are 24 paths of the used channel models, each has their own average gains and delays.

Table 3: Sample of the data frame (part-1).

RI_FR0		RI_FR12	PMI_FR0		PMI_FR12	CQI[11]_FR0		CQI[22]_FR12	Path 0 Delay		Path 23 Delay
2		2	1		0	4		5	0		0
2	…	2	1	…	0	3	…	7	0	…	0
2		2	1		1	0		8	0		0
2		2	1		1	0		9	0		0
2		2	1		0	0		8	0		0
2		2	1		0	3		7	0		0
2		2	1		0	3		7	0		0
2		2	1		1	3		8	0		0
2		2	1		1	2		6	0		0

Table 4: Sample of the data frame (part-2).

Path 0 Gain		Path 23 Gain	RSRP1 (dBm)	RSRP2 (dBm)	RSSI1	RSSI2	RSRQ1	RSRQ2
-13.4		-1000	-86.7729	-93.3184	3.29E-11	5.27E-11	0.91589	0.45948
-13.4	…	-1000	-85.7395	-92.8659	1.56E-10	4.03E-11	0.88646	0.66748
-13.4		-1000	-85.333	-91.9726	1.71E-10	4.55E-11	0.89134	0.72489
-13.4		-1000	-85.3428	-91.5111	1.70E-10	5.37E-11	0.89141	0.68337
-13.4		-1000	-85.7687	-90.747	1.55E-10	7.53E-11	0.88826	0.58172
-13.4		-1000	-86.7355	-90.2949	1.25E-10	9.41E-11	0.88201	0.51643
-13.4		-1000	-87.424	-90.1842	1.07E-10	9.95E-11	0.87838	0.50074
-13.4		-1000	-89.2049	-90.1273	7.16E-11	1.00E-10	0.87263	0.50515
-13.4		-1000	-91.4055	-90.1683	4.29E-11	8.86E-11	0.87791	0.56471

Tables 3 and 4 also show the different CSI feedback values used to build the model. First, when it comes to RI, there are 13 different values in each row. Each value is for a different subchannel. There are also 13 different PMI values in each row (one for each subchannel). Regarding CQI, there are 52 values in each row. As 2×2 MIMO is used in the simulations, we have 4 different groups of CQI values, i.e., one group for each input/output combination. In each group, there are 13 different values (one for each channel). There are also 2 values of RSRP, RSSI, and RSRQ values in each row (one for each antenna pairing).

As mentioned above, the goal is to build a model to estimate the gains and delays of the different channel paths from the CSI feedback. Many models were tested with different combinations of CSI values and the best results were obtained by training the model with all the RSSI, RSRQ, and CQI values as the dependent variables. The output of the model (independent variables) is the gains and delays for all the paths. Additionally, testing was done to determine the optimal depth and width of the hidden layers. It was concluded that one hidden layer performed best with 40.

The testing results from the trained model are shown in Figure 2. The true values (on the y-axis) are plotted against the predicted values (on x-axis) for the path delays. The image on the left is for the second path while the one on the right is for the third. Path one is omitted as this is the reference path and its delay relative to itself is always zero. The $y = x$ line in the figures shows where all the points would lie if the classifier predicted each value perfectly (true value = predicted value). As can be seen in Figure 2, there is a strong correlation between the actual and predicted values of the delays, shown by most of the points lying close to the $y = x$ line. Similar results are obtained for the remaining paths. The average RMSE of delay predictions of the first 10 paths is 9.599391e-08 seconds. Additionally, the R^2 value for the delay predictions was 0.9664731, meaning almost all the variance was accounted for in the model.

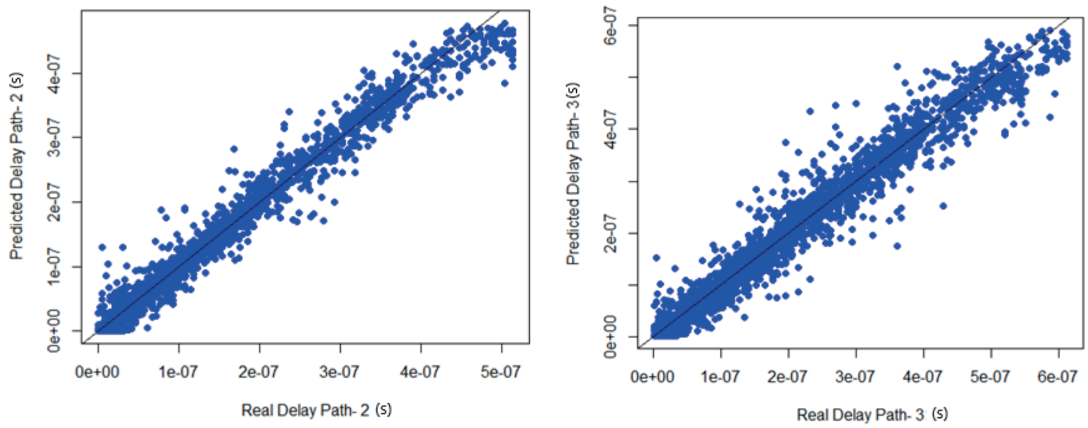

Figure 2: Real vs. predicted delays of the second (left) and third (right) path.

Figure 3 shows the true values plotted against the predicted values for the path gains in the first and second paths. Although the figures show that some of the predicted values deviate from the actual values, further statistical analysis show that there are few outliers, and most of the predicted results are close to the true values. Results show that the average RMSE of gain predictions of the first 10 paths is 0.9862863 dB. This means that out of the total range of about 20 dB, the average error of estimated path gain values is less than 1 dB. The R^2 value for the gain predictions was 0.9818478, again showing that almost all the variance was accounted for in the model.

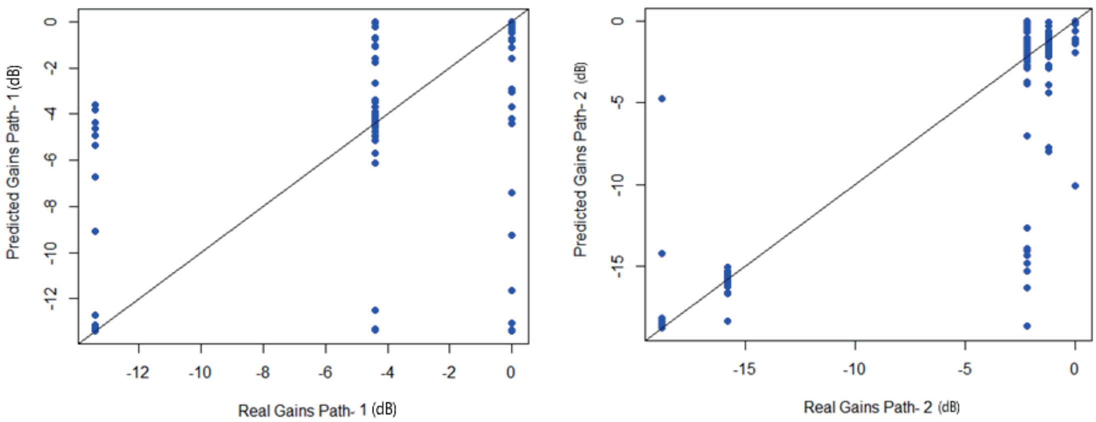

Figure 3: Real vs. predicted gains of the first (left) and second (right) path.

The results show that the path gains and delays can be predicted by partial CSI feedback. The models above were trained on a laptop with Quad-Core processor and 16GB RAM. More accurate results can be obtained with bigger data sets that can be trained with more powerful computational resources (e.g., on the cloud).

6 CONCLUSION

We presented a simulator developed using the MATLAB's 5G Toolbox and a Neural Network (NN) to predict 5G NR channel path gains and delays from limited CSI. Various combinations of channel conditions were simulated to generate a dataset of CSI, path gains and delays. The TDL channel model was used, along with free space pathloss and AWGN to emulate a real system. The CSI considered in this research was RI, RSSI, CQI, and PMI. The simulator was used to generate several datasets which were used for training and testing NNs in R. The ML models used the CSI as input features and predicted the gain and delay for each of the paths. These values were predicted since they can be used to reconstruct the channel matrix. Many NN were tested with varying input features, hidden layer depth and width; only the best performing model was presented here. The RMSE and R^2 values were compared, and the best results came from using all the CSI with a single hidden layer and 40 hidden nodes. The NN showed promising accuracy considering the limited CSI used for training. In future work we will present the hyperparameter optimization, comparing the developed models and expanding to uplink channel sounding. Additionally, validation will be done using over the air data.

REFERENCES

3GPP. 2020. "Study on Channel Model for Frequencies from 0.5 to 100 GHz". Technical Report 38.901, 3rd Generation Partnership Project.

Dahlman, E., Parkvall, S., and Skold, J. 2018. *The Next Generation Wireless Access Technology*. London: Elsevier Ltd.

Han, Y., Hsu, T.-H., Wen, C.-K., Wong, K.-K., and Jin, S. 2019. "Efficient Downlink Channel Reconstruction for FDD Multi-Antenna Systems". *IEEE Transactions on Wireless Communications*, pp. 3161 - 3176.

Hastie, T., Tibshirani, R., and Friedman, J. 2017. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag.

Li, M., Han, Y., Li, X., Wen, C.-K., and Jin, S. 2020. "Deep Learning Based Fast Downlink Channel Reconstruction for FDD Massive MIMO Systems". 2020 In *IEEE Wireless Communications and Networking Conference*, pp. 1-6. Seoul, Korea, IEEE Electronic Library.

Linnartz, J. P. 1995. *Doppler spectrum*. www.wirelesscommunication.nl/reference/chaptr03/fading/doppler.html. Accessed 2021, Mar. 22.

Liu, A., Lau, V., and Dai, W. 2016. "Joint Burst LASSO for Sparse Channel Estimation in Multi-User Massive MIMO". In *IEEE International Conference on Communications (ICC)*. Kuala Lumpur, Malaysia, IEEE Electronic Library.

Luo, C., Ji, J., Wang, Q., Chen, X., and Li, P. 2020. "Channel State Information Prediction for 5G Wireless Communications: A Deep Learning Approach". *IEEE transactions on network science and engineering* vol. 7, no. 1, pp. 227–236.

MathWorks. 2020. "5G Toolbox". www.mathworks.com/products/5g.html. Accessed 2020, Dec 18.

Mawatwal, K., Sen, D., and Roy, R. 2020. "Performance Analysis of a SAGE-Based Semi-Blind Channel Estimator for Pilot Contaminated MU Massive MIMO Systems". *IEEE Access*, vol. 8, pp. 46682-46700.

Ryu, J. 2020. *LTE Basic Procedure: Precoding*. www.sharetechnote.com/html/BasicProcedure_LTE_PHY_Precoding.html. Accessed 2020, Dec 18

Tse, D., and Viswanat, P. 2004. *Fundamentals of Wireless Communication*. Cambridge, United Kingdom: Cambridge University Press.

Winkler, D. A., and Le, T. C. 2016. "Performance of Deep and Shallow Neural Networks, the Universal Approximation Theorem, Activity Cliffs, and QSAR". *Molecular Informatics* vol. 36, no. 1-2.

Ye, H., Li, G. Y., and Juang, B. 2018. "Power of Deep Learning for Channel Estimation and Signal Detection in OFDM Systems". *IEEE Wireless Communications Letters*, pp. 114-117.

Zaidi, A., Athley, F., Medbo, J., Gustavsson, U., Durisi, G., & Chen, X. 2018. *5G Physical Layer: Principles, Models and Technology Components*. London: Academic Press.

AUTHOR BIOGRAPHIES

BEN EARLE is a PhD student in Electrical and Computer Engineering at Carleton University under the supervision of Dr. Gabriel Wainer and Dr. Ala'a Al-Habashna. His research interests include wireless communication, embedded systems, modeling and simulation, and machine learning. His email address is benearle@cmail.carleton.ca.

ALA'A AL-HABASHNA is an Adjunct Research Professor at Carleton University. His research interests include 5G wireless communication systems, internet of things, machine learning, computer vision, and modeling and simulation. His email address is alaaalhabashna@sce.carleton.ca and his page is https://carleton.ca/sce/people/al-habashna/

GABRIEL WAINER is Professor in the Department of Systems and Computer Engineering, Carleton University (Ottawa, ON, Canada). His current research interests are related with modeling methodologies and tools, parallel/distributed simulation, and real-time systems. He is a Fellow of SCS. His e-mail is gwainer@sce.carleton.ca. His website is www.sce.carleton.ca/faculty/wainer.

XINGLIANG LI is a Developer, Team Leader at Ericson. He received his PhD is in Telecommunications. His expertise is in algorithm development for 5G and LTE systems. His research is focused on performance verification, channel modeling, and beam management. His email is xingliang.li@ericsson.com.

GUOQIANG XUE is a manager of network system verification at Ericsson. His PhD is in Electrical Engineering. His team is specialized in 5G performance and capacity verification. Guoqiang is an expert in MIMO and beamforming testing and verification. He has filed many IPRs on the topics of MIMO testing and CDMA data transmission. His email is guoqiang.xue@ericsson.com.