Recurrent Neural Network-based Channel Prediction in mMIMO for Enhanced Performance in Future Wireless Communication

Joel Poncha Lemayian
dept. Electrical and Computer Engineering
Antalya Bilim University
Antalya, Turkey
lab: WISLAB@Teleng for Wireless Research
e-mail: lemayian.joel@std.antalya.edu.tr

Jehad M. HAMAMREH
dept. Electrical and Electronics Engineering
Antalya Bilim University
Antalya, Turkey
lab: WISLAB@Teleng for Wireless Research
e-mail: jehad.hamamreh@antalya.edu.tr

***Abstract*—Massive MIMO (mMIMO) has been classified as one of the high potential future wireless communication technologies due to its unique abilities such as high user capacity, increased spectral density, and diversity. Due to the exponential increase of connected devices, these properties are critical for the current 5G-IoT era and future telecommunication networks. However, outdated channel state information (CSI) causes major performance degradation in mMIMO systems. Nevertheless, channel prediction using neural networks (NN) has gained tremendous attention as a way of mitigating outdated CSI. Hence, combined mMIMO and NN-based channel prediction is a revolutionary technology of future wireless communications. In this work, we review the current recurrent neural network-based (RNN-based) mMIMO channel prediction schemes and propose a low complexity, low cost channel prediction scheme.**

***Index Terms*—mMIMO, RNN, Artificial Intelligence, CSI.**

I. Introduction

Multiple input multiple output (MIMO) employs multiple antennas at the transmitter and/or receiver. This technology has highly desired properties such as high throughput, high spectral efficiency, and multiplexing gains [7]. MIMO has evolved from a mere research concept to a real-world application and has been integrated into state-of-the-art wireless network standards such as IEEE 802.11n, 3GPP long-term evolution (LTE) and LTE- Advanced (E-UTRA) [8]. In a massive MIMO (mMIMO) system, the number of antennas in MIMO increases to hundreds. mMIMO has been classified as one of the high potential wireless communication technologies with the ability to have high user capacity [1], which is a key requirement for 5G-IoT and beyond technologies.

mMIMO is affected by outdated channel state information (CSI), which occurs when the information obtained about the channel changes before data transmission [1]. [2] studies the effect of outdated CSI on mMIMO and demonstrate how capacity is lost due to varying channel and hence resulting in performance degradation.

This work is funded by the scientific and technological research council of Turkey (TÜBITAK) under grand 119E392. All the codes used can be found at researcherstore.com

Neural Networks (NN) as an AI technique, is an effective recently proposed model for combating outdated CSI without wasting resources [4]. It is highly valued because it can avoid parameter estimation due to its data-driven nature. Channel prediction is viewed as a revolutionary future technology and hence it has attracted the attention of many researchers [**?**]. Additionally, instantaneously selecting transmit parameters (enabled by channel prediction) such as transmit power, coding rate, transmit antennas, and carrier frequency, depending on the instantaneous condition of the channel using channel prediction will tremendously improve the performance of adaptive wireless communications systems (Which are the future of effective wireless communication).

Recurrent neural networks (RNNs) are very effective when processing time series data. Since channel response data is closely related to time series data, we look at mMIMO channel prediction using RNNs as a technology with great future potential that will have a major impact in wireless technology. Moreover, we compare performance metrics of conventional CSI prediction processes with RNN-based prediction. lastly, an RNN model utilizing LSTMs is designed. The novelty of this work is enumerated below:

1) Propose a low cost mMIMO RNN-based CSI predictor.
2) Provide quick answers about RNN-based mMIMO CSI prediction.
3) Demonstrate performance metrics between conventional CSI predictors and RNN-based, in terms of complexity and cost.

The remainder of this work is organised as follows. We discuss the structure of RNNs in section II. mMIMO system model is presented in section III. In section IV, we review some common non-AI CSI prediction models. Section V talks about RNN-bases mMIMO CSI prediction. Some results and conclusion are given in sections VI and VII respectively.

II. Recurrent Neural Networks (RNNs)

In this section, we will discuss the RNN model which is a powerful machine learning technique that has shown great potential in predicting time series data. RNNs are superior because they not only use training data for learning but also learn from historical data of past events. There are different models of RNNs, Fig. 1 depicts the commonly known Jordan network. A simplified RNN network consists of an input layer with N_i neurons, a hidden layer with N_h neurons, and an output layer with N_o outputs. Each connection between the input layer, the hidden layer, and the output layer is assigned a weighted value. Let w_{ln} denote the weight between the l^{th} input and the n^{th} hidden neuron. And v_{ol} represent the weight of the l^{th} hidden neuron and the o^{th} output neuron. Such that $1 \leq n \leq N_i$, $1 \leq l \leq N_h$, and $1 \leq o \leq N_o$. Therefore, we can construct an $N_h \times N_i$ matrix $\mathbf{W}$ of weights. The input

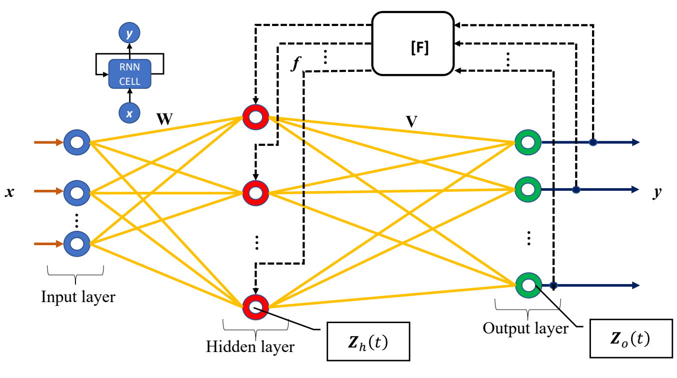

Fig. 1. Recurrent Neural Network system.

activation vector is denoted as $\mathbf{x}(t) = [x_1(t), \ldots, x_{(N_i)}(t)]^T$ while the recurrent or feedback component is modeled as $\mathbf{f}(t) = [f_1(t), \ldots, f_{(N_i)}(t)]^T$. Therefore, the input to the hidden layer can be represented by the following equation:

$$\mathbf{z}_h(t) = \mathbf{W}\mathbf{x}(t) + \mathbf{f}(t) + \mathbf{b}_h, \tag{1}$$

where $b_h = [b_1^h, \ldots, b_{(N_h)}^h]^T$ represents the bias in the hidden layer. In addition, we use a matrix $\mathbf{F}$ to map the previous output $y(t-1) = [y_1(t-1), \ldots, y_{N_o}(t-1)]^T$ to the current input component. Hence, $\mathbf{f}(t) = \mathbf{F}\mathbf{y}(t-1)$.. The behavior of a neuron network is determined by an activation function (AF). The activation function introduces some nonlinearity characteristics to the system. Hence enabling it to solve complex problems. Common activation functions include liner, rectified liner hyperbolic tangent, and leaky rectified liner. In this work, we use the sigmoid function to introduce nonlinearity. The sigmoid function is defined as:

$$S(x) = \frac{1}{1+e^x}. \tag{2}$$

Substituting equations (1) and $\mathbf{f}(t)$ into equation (2), we get the following equation:

$$\mathbf{h}(t) = S(\mathbf{Z}_h(t)) = S(\mathbf{w}\mathbf{x}(t) + \mathbf{F}\mathbf{y}(t-1) + \mathbf{b}_h). \tag{3}$$

Executing equation (3) is an element wise operation. That is, $S(\mathbf{Z}_h) = [S(z_1), \ldots, S(z_{N_h})]^T$. Consequently, we will have matrix $\mathbf{V}$ with dimensions $N_o \times N_h$. The input to the output layer therefore becomes, $\mathbf{z}_o(t) = \mathbf{V}\mathbf{h}(t) + \mathbf{b}_y$ where $\mathbf{b}_y$ is the bias matrix for the output layer. The output equation is derived as:

$$\mathbf{y}(t) = S(\mathbf{z}_o(t)) = S(\mathbf{V}\mathbf{h}(t) + \mathbf{b}_y). \tag{4}$$

However, RNNs suffer from vanishing and exploding gradient problems. The vanishing gradient problem is experienced during back-propagation. This is where the partial derivative of the loss function with respect to the current weight progressively diminishes during back-propagation and hence has no effect on the weights when performing gradient descent. On the other hand, the exploding gradient is experienced when large gradient errors accumulate causing large updates on the network weights during training. LSTM (long short term memory) networks are special types of RNNs that use gates to overcome the problems experienced by conventional RNNs. The gates in LSTM networks include input, output, and forget gates, they facilitate better control of gradient flow and prevention of long-range dependencies. In this work, we utilize LSTMs shown in Fig. 2. Moreover, we use dropout regularization to prevent overfitting.

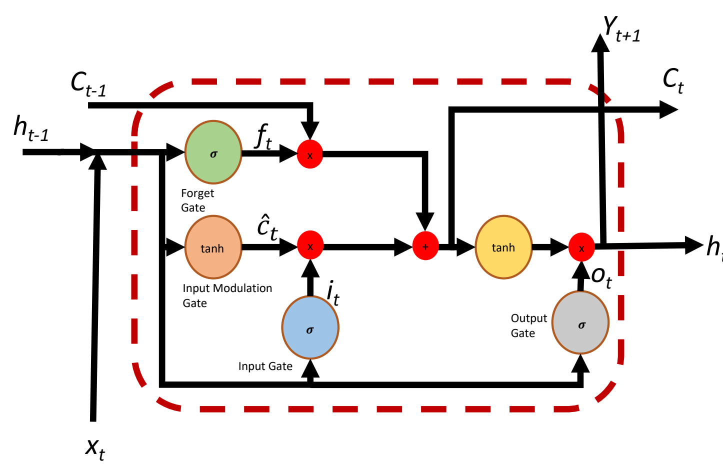

Fig. 2. Long Short Term Memory (LSTM) unit cells.

Equations that describe the functionality of the LSTM unit cell are given below:

$$i_t = \sigma(W^{(i)}x_t + U^{(i)}h_{t-1}), \tag{5}$$

$$f_t = \sigma(W^{(f)}x_t + U^{(f)}h_{t-1}), \tag{6}$$

$$o_t = \sigma(W^{(o)}x_t + U^{(o)}h_{t-1}), \tag{7}$$

$$\hat{c}_t = tanh(W^{(c)}x_t + U^{(c)}h_{t-1}), \tag{8}$$

$$c_t = f_t \bigcirc c_{t-1} + i_t \bigcirc \hat{c}_t, \tag{9}$$

$$h_t = o_t \bigcirc tanh(c_t). \tag{10}$$

$\bigcirc$ denotes element-wise multiplication. W denotes the recurrent connection between the previous and the current hidden layers. U is the weight matrix between the current input and the hidden layer. $\hat{C}$ is a hidden state candidate, and it is calculated based on the current input and the previous hidden state. C is the internal memory of the cell, which is the sum of $previous\_memory \bigcirc forget\_gate$ and $newly\_computed\_hidden\_state \bigcirc input\_gate$.

The first operation done in an LSTM layer is to decide whether the information is kept or discarded. This operation is done in the forget gate. A number between 0 and 1 is produced, where 1 means keep the information, while 0 means completely forget. The next step is to decide which new information will be kept in the cell state. This process is done in two parts, first, a sigmoid function called the input gate determines which values will be updated, then a $tanh$ layer determines candidate cell states to be added. Next, the previous cell state C_{t-1} needs to be updated, that is $C_t = f_t \bigcirc C_{t-1} + i_t \bigcirc \hat{C}_t$. where f_t makes the old state forget information discarded in the forget gate. Finally, the output of the cell is modeled by the output gate which is a sigmoid layer that determines which part of the cell state will be output. The cell state is passed through a layer of $tanh$ which forces the cell state to be between -1 and 1, then the output of the output gate is multiplied by the output of the $tanh$ layer to get the output of the cell. In the next section, the mMIMO system model is described for easy understanding of channel prediction in RNNs.

III. mMIMO SYSTEM MODEL

In a time-varying mMIMO system, there are M_T transmitter antennas and M_R receiver antennas. Where, at any instance $M_T \leq M_R$. Each antenna transmits N time slots at a given transmission, where $N \geq M_T$. Considering an instantaneous signal transmit and receive in mMIMO, the base-band receiver can be shown as:

$$\mathbf{r}(t) = \mathbf{H}(t)\mathbf{s}(t) + \mathbf{n}(t), \tag{11}$$

where $\mathbf{r}(t) = [r_1(t), \ldots, r_{N_r}(t)]^T$ is an $N_r \times 1$ matrix of the receive signal at time t (N_r is the number of receive antennas) and $\mathbf{s}(t) = [s_1(t), \ldots, s_{N_r}(t)]^T$ is an $N_t \times 1$ matrix of the transmit signal at time t (N_t is the number of transmit antennas). $\mathbf{H}(t) = [h_{n_r n_t}(t)]_{N_r \times N_t}$ is the matrix of continues channel impulse response and $h_{n_r n_t} \in \mathbb{C}^{1\times 1}$ is the flat fading channel gain between transmitter antenna n_t and receiver antenna n_r. Moreover, $1 \leq n_r \leq N_r$ and $1 \leq n_t \leq N_t$. Due to the multipath fading, feedback and processing delays the obtained CSI at the transmitter may be outdated before it can be used. That is, $\mathbf{H}(t) \neq \mathbf{H}(t+\tau)$, consequently, this will lead to performance degradation of adaptive communication systems [8]. The goal of channel prediction is to estimate $\mathbf{H}(t+\tau)$ at time t to be as close as possible to the actual value at $(t+\tau)$. That is $\mathbf{H}(t+\tau) \rightarrow \hat{\mathbf{H}}(t+\tau)$. Closely related MIMO prediction techniques are briefly discussed in the subsequent section.

IV. RELATED PREDICTOR MODELS

Apart from RNN-base predictors, several other methods have been proposed for mMIMO CSI prediction. Parametric and autoregressive channel prediction models are the most popular techniques [7]. In this section, the models are briefly discussed to provide the reader with a clear distinction to the proposed work.

A. parametric predictor model

As stated by [5], [6], a single antenna channel is represented by overlaying a set of complex sinusoids in popular multipath fading models.

$$h(t) = \sum_{p=1}^{P} \alpha_p e^{j(\omega_p t + \phi_p)}, \tag{12}$$

where α_p is the complex amplitude, ω_p is the Doppler frequency shift in radians of the p^{th} sinusoid, and ϕ_p is the phase. $j^2 = -1$ denotes complex units and P represents the total number of scattered sinusoids. The single-antenna system depicted in equation (12) can be modeled to represent a MIMO propagation model shown by equation (13), by introducing spatial dimension parameters.

$$\mathbf{H}(t) = \sum_{p=1}^{P} \alpha_p \mathbf{a_r}(\theta_p) \mathbf{a}_t^T(\psi_p) e^{j(\omega_p t + \phi_p)}, \tag{13}$$

where θ_p and ψ_p are the angle of arrival (AOA) and angle of departure (AOD) respectively. $\mathbf{a_r}$ and $\mathbf{a_t}$ are the response vector of the receiver and transmitter antenna arrays respectively. $\mathbf{a}$ can be represented as a uniform linear array (ULA) with M equally spaced antenna elements as follows:

$$\mathbf{a}(x) = [1, e^{-j(\frac{2\pi}{\lambda})dsin(x)}, ..., e^{-j(\frac{2\pi}{\lambda})(M-1)dsin(x)}]^T, \tag{14}$$

where x can either be the angle of arrival or departure, λ is the wavelength of the sub-carrier frequency, and d is the distance between antennas. According to [7], multipath parameters change slowly compared to the channel fading rate. Therefore, future CSI up to a certain period can be obtained by simply extrapolating the known multipath parameters. Hence, channel prediction in mMIMO using equation (13) is reduced to parameter prediction. That is, a parameter prediction model to predict the total number of scatters, the angle of arrival and departure, and the Doppler shift for each path (i.e. $\{\hat{\alpha}_p, \hat{\omega}_p, \hat{\theta}_p, \hat{\psi}_p\}_{p=1}^{\hat{P}}$). Classical algorithims such as multiple signal classification (MUSIC) and estimation of signal parameters by rotational invariance techniques (ASPIRIT) are used to get calculate the parameters. The details of the calculations can be obtained from. Having calculated all the parameters, a parametric model can the be build to $\hat{\mathbf{H}}$ (predicted channel) by substituting $\{\hat{\alpha}_p, \hat{\omega}_p, \hat{\theta}_p, \hat{\psi}_p\}_{p=1}^{\hat{P}}$ to equation 13.

B. Autoregressive (AR) predictor model

The mMIMO time varying channel can alternatively be formulated using an autoregressive process where Kalman filters (KF) are used to compute AR coefficients used to build a liner predictor used to predict future CSI using current and past CSI [7]. the AR predictor for mMIMO can be represented as:

$$\hat{\mathbf{H}}(t-1)=\sum_{p=1}^{P}\mathbf{A}_p \bigcirc \mathbf{H}[t-p+1], \qquad (15)$$

where $\mathbf{A}_p = \{a^p_{n_r n_t}\}$ is an $n_r \times n_t$ AR coefficient matrix, such that $a^p_{n_r n_t}$ is the p^{th} AR coefficient of the channel between transmitter n_t and receiver n_r. Other predictors proposed in literature include maximum-likelihood (ML) estimation, least-square (LS) estimation, and minimum-mean-square-error (MMSE) estimation

CSI prediction using the above models is faced by a number of challenges such as high complexity, low accuracy, lack of generality, single-step prediction limitation, and unreliability, hence such methods are only suitable for small scale estimation [3]. CSI prediction process in RNNs is summarised in the subsequent section.

V. MMIMO CSI PREDICTION PROCESS USING RNNS

The data set $\mathbf{h}$ is divided into two, training (60%) and testing/prediction (40%) data sets. The training data set is used to train the model through backpropagation (BP). Referring to Fig. 2, the model takes in the input $\mathbf{x} = \mathbf{h}(t)$ and the preferred output $\mathbf{y} = \mathbf{h}(t + D)$, where D is the steps to be predicted into the future. The model then calculates the cost $\mathbf{C}$, using mean squared error (MSE):

$$\mathbf{C}=\frac{1}{T}\sum_{t=1}^{T}||\hat{\mathbf{h}}(t+D)-\mathbf{h}(t+D)||^2, \qquad (16)$$

T is the is the total number of channel samples. The error is then propagated back through the network, causing the weights to iteratively adjust through gradient descent until convergence is obtained and a minimum cost is achieved. Initial parameters such as C_{t-1} and h_{t-1} are generally initialised with zero values, while the weights are randomly initialized. The testing set is then used to measure the accuracy of the network. A 2-layer mMIMO RNN-based model predictor utilizing LSTM was trained using 200,000 samples of $\mathbf{h}$ from 128 antenna generated with Rayleigh distribution.

VI. RESULTS

Fig.3 shows the validation and training loss after 10 epochs with a batch of 500. It can be observed that both training loss and validation loss gradually converge. This indicates that the model is neither over-fitting nor under-fitting. In this case, the accuracy of the model can be improved by increasing the number of epoch since the losses are still on a downwards trend.

Moreover, Fig.4 shows the prediction of the real part of the channel. Similarly, the imaginary part of the channel is fed to the LSTM predictor to obtain a prediction. The results show that the model can estimate the channel's characteristics with minimum complexity compared to other methods mentioned in section IV. It can also be observed from the figure that the accuracy of the predictor is not 100%. To improve the accuracy of the predictor, training data and iterations can be increased as well as LSTM layers or using different regularizers.

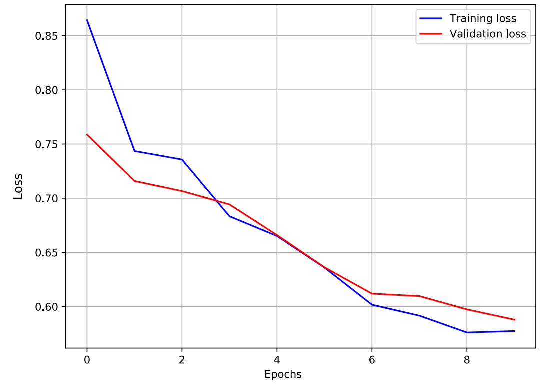

Fig. 3. mMIMO training and validation loss.

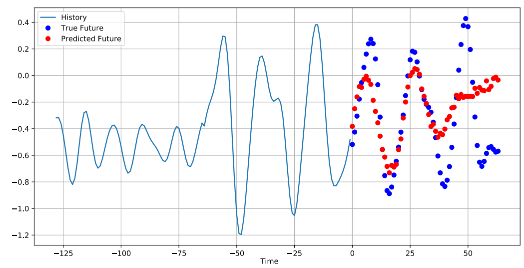

Fig. 4. Real part prediction of the massive MIMO channel.

VII. CONCLUSION

Considering that mMIMO is a technology that has shown great potential in enhancing wireless communication in the future, this work has demonstrated that RNN-based CSI prediction is the ideal technology to boost the performance of mMIMO systems by lowering complexity and increasing accuracy. Therefore, we intent to further this work by perfecting mMIMO CSI prediction using different configurations of RNNs networks utilizing LSTMS or GRUs and improving the accuracy of the designed predictor.

REFERENCES

[1] Jiang W, Schotten HD. "Deep Learning for Fading Channel Prediction." IEEE Open Journal of the Communications Society 1 (2020): 320-332.

[2] Truong KT, Heath RW. "Effects of channel aging in massive MIMO systems." Journal of Communications and Networks 15, no. 4 (2013): 338-351.

[3] Luo C, Ji J, Wang Q, Chen X, Li P. "Channel state information prediction for 5G wireless communications: A deep learning approach." IEEE Transactions on Network Science and Engineering (2018).

[4] Hamamreh JM, El-sallabi H, Abdallah M, Qaraqe K. "Advance in Adaptive Modulation for Fading Channels." (2013).

[5] Hamamreh JM, Hajar A, Abewa M. "Orthogonal frequency division multiplexing with subcarrier power modulation for doubling the spectral efficiency of 6G and beyond networks." Transactions on Emerging Telecommunications Technologies 31, no. 4 (2020): e3921.

[6] aradat AM, Hamamreh JM, Arslan H. "OFDM With Hybrid Number and Index Modulation." IEEE Access 8 (2020): 55042-55053.

[7] Jiang W, Schotten HD. "Neural Network-Based Fading Channel Prediction: A Comprehensive Overview." IEEE Access 7 (2019): 118112-118124.

[8] Zheng J, Rao BD, "Capacity analysis of MIMO systems using lim- ited feedback transmit precoding schemes," IEEE Trans. Signal Process., vol. 56, no. 7, pp. 2886–2901, Jul. 2008.