Recurrent Neural Networks with Long Short-Term Memory for Fading Channel Prediction

Wei Jiang*† and Hans D. Schotten†*
*German Research Center for Artificial Intelligence (DFKI)
Trippstadter Street 122, Kaiserslautern, 67663 Germany
†Institute for Wireless Communication and Navigation, University of Kaiserslautern
Building 11, Paul-Ehrlich Street, Kaiserslautern, 67663 Germany

***Abstract*—With the aid of accurate channel state information (CSI) at the transmitter, a wireless system can receive great performance by adaptively selecting its transmission parameters. However, the CSI becomes outdated quickly due to the rapid channel variation caused by multi-path fading, leading to severe performance degradation. Such an impact is applicable on a wide variety of adaptive transmission systems, the fading channel prediction that can combat the outdated CSI is therefore of great significance. The aim of this paper is to propose two novel recurrent neural network (RNN)-based predictors, leveraging the strong time-series prediction capability of long short-term memory or gated recurrent unit. Performance evaluation is conducted and the results in terms of prediction accuracy verify that the proposed predictors notably outperform the conventional RNN predictor.**

I. Introduction

Given accurate channel state information (CSI) at the transmitter, a wireless system is capable of adaptively selecting its transmission parameters, such as the constellation size, transmit power, coding rate, precoding codeword and transmit antennas, to realize great performance gains. In most systems, the CSI is estimated at the receiver and then fed back to the transmitter. However, the obtained CSI is probably outdated at the time of using the selected parameters to transmit. Although a time-division duplex system can avoid the feedback delay through channel reciprocity, the processing delay still raises inaccuracy, especially in high mobility or high frequency band scenarios. It has been well recognized that the outdated CSI has an overwhelming impact on the performance of a wide variety of adaptive transmission systems, spanning from multiple-input multiple-output (MIMO) [1], multi-user MIMO [2], massive MIMO [3], interference alignment [4], antenna selection [5], relaying [6], resource allocation [7], mobility management [8], to physical layer security [9]. More details can refer to [10].

In order to combat the outdated CSI, the researchers proposed a large number of mitigation algorithms and protocols, which either passively compensate for the performance loss with a cost of scarce wireless resources [11] or aim to achieve merely part of the full potential under imperfect CSI [12]. In contrast, an alternative technique referred to as channel prediction [13] provides an efficient approach to improve the accuracy of CSI directly without spending extra wireless resources. Through statistically modeling a wireless channel as a set of radio propagation parameters, two conventional prediction approaches, namely parametric model [14] and autoregressive model [15], [16], have been developed. However, there exists a gap between the model and reality, and additionally the parameter estimation relying on complex algorithms such as MUSIC is tedious, making their application in practical systems infeasible [17]. Due to the data-driven nature that can avoid the tedious estimation process of propagation parameters, neural networks [18] attracted the focus of researchers in this field. Making use of its capability of time-series prediction [19], a recurrent neural network (RNN) was firstly proposed in [20] to build a predictor for narrow-band single-antenna channels and was further extended to MIMO channels by [21], followed by a frequency-domain RNN predictor [22] suitable for application in frequency-selective MIMO channels. The authors of [23] proposed to utilize a real-valued RNN to implement a multi-step predictor and further verified its effectiveness in a MIMO system [24]. Furthermore, the feasibility of applying deep learning was also preliminarily studied [25] and an overview of RNN-based prediction methods was provided in [26].

Going beyond the conventional RNN predictor, this paper aims to build two novel recurrent networks, leveraging long short-term memory (LSTM) and gated recurrent unit (GRU), respectively. To the best knowledge of the authors, it is the first time to apply LSTM and GRU for predicting multi-path fading channels. Performance assessment in terms of prediction accuracy is carried out in MIMO fading environment, taking into account a number of influential factors such as additive noise, the Doppler shift, the scale of MIMO, activation function, and the number of hidden neurons. Some representative numerical results are provided, verifying that the proposed predictors can achieve a remarkable performance improvement. The rest of this paper is organized as follows: Section II introduces the principles of LSTM and GRU. Section III describes the simulation setup and presents the results. Finally, Section IV closes this paper with our remarks.

II. An overview of LSTM and GRU

Recurrent neural network is a class of machine learning technique that shows good potential in processing time-series

*This work was supported by German Federal Ministry of Education and Research (BMBF) through TACNET4.0 and KICK projects.

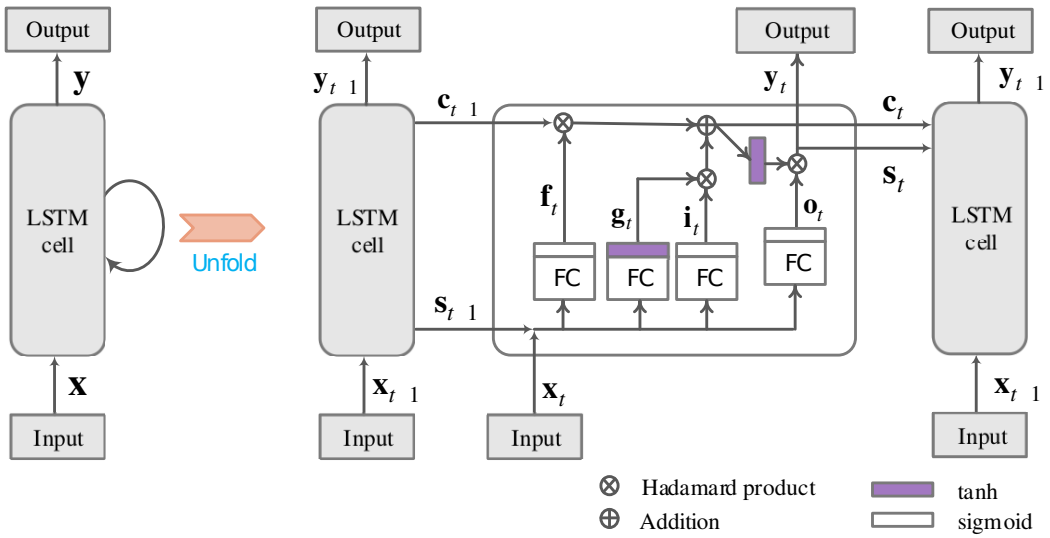

Fig. 1. Schematics of an RNN network with an LSTM memory cell.

data by storing the historical input information in its hidden state [19]. However, it suffers from the problems of gradient exploding or vanishing during its training, where the back-propagated error signals either tend to infinity, leading to oscillating weights, or tend to zero that means a prohibitive training time or the training does not work at all. To tackle this problem, Hochreiter and Schmidhuber proposed the LSTM model in 1997 [27], which was gradually improved over the years. Despite of its short history after being proposed, LSTM already achieved a great success in both academia and commercial world, demonstrated by many unprecedented intelligent services such as Google Translate and Apple iPhone Siri.

The key innovation to enlarge LSTM's capability to deal with long-term dependencies is the introduction of a memory cell and multiplicative gates that regulate the pattern of information flow. The *input* gate protects the memory contents stored in the cell from perturbation by irrelevant inputs, the *forget* gate decides how much the historical information remains in the cell, and the *output* gate controls the extent to which the memory information applied to generate the output activation of the LSTM cell. Unlike a feed-forward network where activations flow only in one direction from the input to the output layer, as illustrated in the left part of Fig.1, an LSTM network has a feedback loop that allows self-recurrent connections between the neurons within the hidden layer. At each time step, the cell generates the output $\mathbf{y}$ according to the input $\mathbf{x}$, along with the hidden state feeding back from the previous time step. Unrolling the network through time, as illustrated in the right part of Fig.1, the memory cell derives two internal states - the short-term state $\mathbf{s}_{t-1}$ and the long-term state $\mathbf{c}_{t-1}$ - at time step $t-1$. Traversing the cell from the left to the right, $\mathbf{c}_{t-1}$ first throws away some old memories at the forget gate, integrates some new memories selected by the input gate, and then sends out as the long-term state $\mathbf{c}_t$ for the next time step $t+1$. The current input vector $\mathbf{x}_t$ and the previous short-term memory $\mathbf{s}_{t-1}$ are fed into four different fully connected (FC) layer, generating the activation vectors of gates:

$$\mathbf{f}_t = \sigma_g\left(\mathbf{W}_f\mathbf{x}_t + \mathbf{U}_f\mathbf{s}_{t-1} + \mathbf{b}_f\right), \tag{1}$$

$$\mathbf{i}_t = \sigma_g\left(\mathbf{W}_i\mathbf{x}_t + \mathbf{U}_i\mathbf{s}_{t-1} + \mathbf{b}_i\right), \tag{2}$$

$$\mathbf{o}_t = \sigma_g\left(\mathbf{W}_o\mathbf{x}_t + \mathbf{U}_o\mathbf{s}_{t-1} + \mathbf{b}_o\right), \tag{3}$$

where $\mathbf{W}$ and $\mathbf{U}$ are weight matrices for the FC layers, $\mathbf{b}$ stands for bias vectors, the subscripts f, i, and o associate with the forget, input, and output gate, respectively, and σ_g represents the *sigmoid* activation function, defining by

$$\sigma_g(x) = \frac{1}{1+e^{-x}}. \tag{4}$$

Dropping some old memories and adding some new ones, the previous long-term memory $\mathbf{c}_{t-1}$ is transferred to

$$\mathbf{c}_t = \mathbf{f}_t \otimes \mathbf{c}_{t-1} + \mathbf{i}_t \otimes \sigma_h\left(\mathbf{W}_g\mathbf{x}_t + \mathbf{U}_g\mathbf{s}_{t-1} + \mathbf{b}_g\right), \tag{5}$$

where $\otimes$ denotes the Hadamard product (element-wise multiplication) for matrices, and σ_h is the *hyperbolic tangent* function denoted by $\tanh$, defining by

$$\sigma_h(x) = \frac{e^{2x}-1}{e^{2x}+1}. \tag{6}$$

Then, the long-term memory $\mathbf{c}_t$ passes through the hyperbolic tangent function that is further filtered by the output gate to produce the short-term memory, as well as the output, i.e.,

$$\mathbf{s}_t = \mathbf{y}_t = \mathbf{o}_t \otimes \sigma_h\left(\mathbf{c}_t\right). \tag{7}$$

Since the emergence of LSTM, its original architecture continues to evolve. Cho et al. [28] proposed a simplified version with fewer parameters in 2014, known as GRU, which exhibits even *better performance over LSTM on certain smaller data sets*. As shown in Fig.2, the short- and long-term states

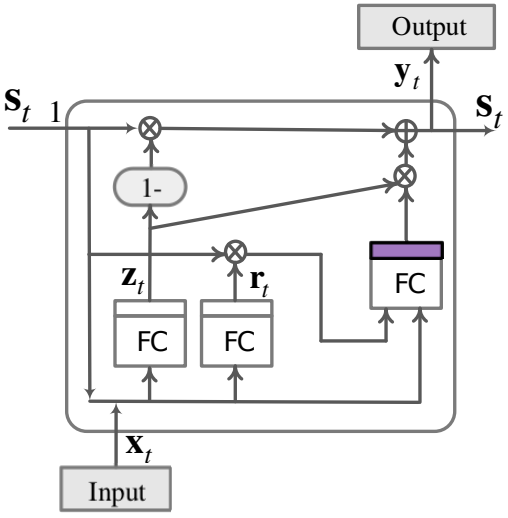

Fig. 2. Illustration of a GRU memory cell.

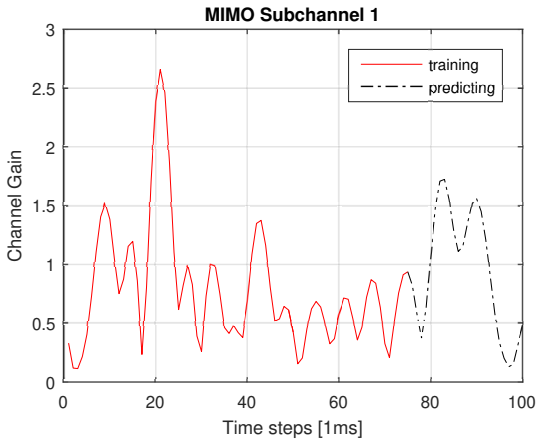

Fig. 3. An exemplified illustration of channel data used for training and testing.

are merged into a single one in GRU, and a single gate $\mathbf{z}_t$ is used to replace the forget and input gates, namely,

$$\mathbf{z}_t = \sigma_g\left(\mathbf{W}_z\mathbf{x}_t + \mathbf{U}_z\mathbf{s}_{t-1} + \mathbf{b}_z\right). \tag{8}$$

The output gate is removed, but an intermediate state $\mathbf{r}_t$ is applied, i.e.,

$$\mathbf{r}_t = \sigma_g\left(\mathbf{W}_r\mathbf{x}_t + \mathbf{U}_r\mathbf{s}_{t-1} + \mathbf{b}_r\right). \tag{9}$$

Likewise, the hidden state at the previous time step transverses the memory cell, drops some old memory, and loads some new information, resulting in the current state that is given by

$$\begin{aligned}\mathbf{s}_t = \mathbf{y}_t =&(1-\mathbf{z}_t)\otimes\mathbf{s}_{t-1}\\ &+\mathbf{z}_t\otimes\sigma_h\left(\mathbf{W}_s\mathbf{x}_t + \mathbf{U}_s(\mathbf{r}_t\otimes\mathbf{s}_{t-1}) + \mathbf{b}_s\right).\end{aligned} \tag{10}$$

III. Performance Evaluations

A performance comparison between the proposed LSTM/GRU predictors and the previous RNN predictor is conducted and some representative results are provided in this section. We apply independent and identically distributed (*i.i.d*) flat-fading MIMO channels having 4 transmit antennas and a single receive antenna as the baseline for comparing the performance. Each subchannel follows the Rayleigh distribution with an average power gain of 0dB, where its channel gain h is zero-mean circularly-symmetric complex Gaussian random variable with the variance of 1, i.e., $h\sim\mathcal{CN}(0,1)$. The maximal Doppler frequency shift is set to f_d=100Hz for emulating fast time-varying environment. To build a data set, continuous-time channel responses are sampled with a sampling rate of 1KHz, satisfying a flat fading assumption. The data set contains a series of 10^4 consecutive CSI $\{\mathbf{H}[t]\,|t{=}1,2,\ldots,10000\}$, with an interval of each time step T_s=1ms. In our simulation, 75% of the data is allocated for training and the remaining 25% is used for testing. As an example, Fig.3 visualizes a small piece of the collected channel data over MIMO subchannel 1.

TABLE I
Simulation Parameters

Parameters	Values
Sampling rate	$f_s = 1$KHz
Maximum Doppler shift	$f_d = 100$Hz
Channel model	Rayleigh fading
Doppler spectrum	Jakes's model
MIMO size	4×1
Batch size	256
Dataset size	10^4
Training algorithm	Adam optimizer
Cost function	MSE
(default) Prediction length	1ms
(default) Actuation function	tanh
(default) Hidden neurons	10

A 3-layer recurrent network, including an input layer, an output layer, and a hidden layer with an LSTM or GRU memory cell, is employed. The default number of hidden neurons is 10, while the activation function of hyperbolic tangent is applied at each hidden neuron as default. A training process starts from an initial state where all weights and biases are randomly selected. At time step t, recalling Fig.1 and 2, the input of the memory cell is $\mathbf{x}_t = \mathbf{H}[t]$ and the output is its D-step ahead prediction, i.e., $\mathbf{y}_t = \hat{\mathbf{H}}[t+D]$. Mean squared error (MSE), the metric for measuring the prediction accuracy, is chosen as the cost function during the training phase. It is defined as

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T}\left\|\hat{\mathbf{H}}[t+D] - \mathbf{H}[t+D]\right\|^2, \tag{11}$$

where T is the total number of channel samples used for evaluation, $\hat{\mathbf{H}}[t+D]$ denotes the predicted channel matrix at time t, $\mathbf{H}[t+D]$ stands for its actual value, and $\|\cdot\|$ notates the Frobenius norm of a matrix. Using the *batch* training method, a batch of 256 samples are fed into the recurrent network at each epoch, the resultant outputs are compared with the desired values and the error signals are propagated back through the

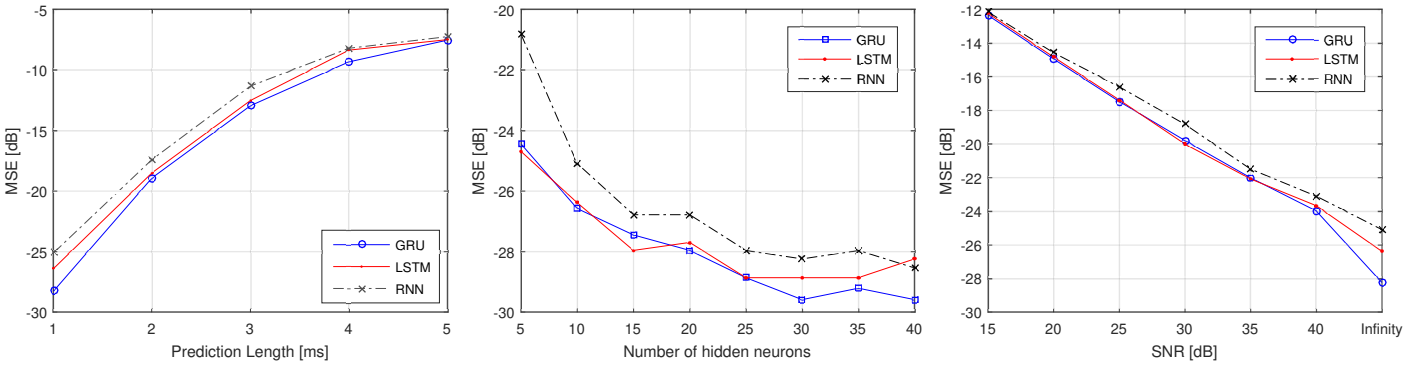

Fig. 4. Performance comparisons with respect to: (a) prediction time length; (b) the number of hidden neurons; and (c) the strength of additive noise.

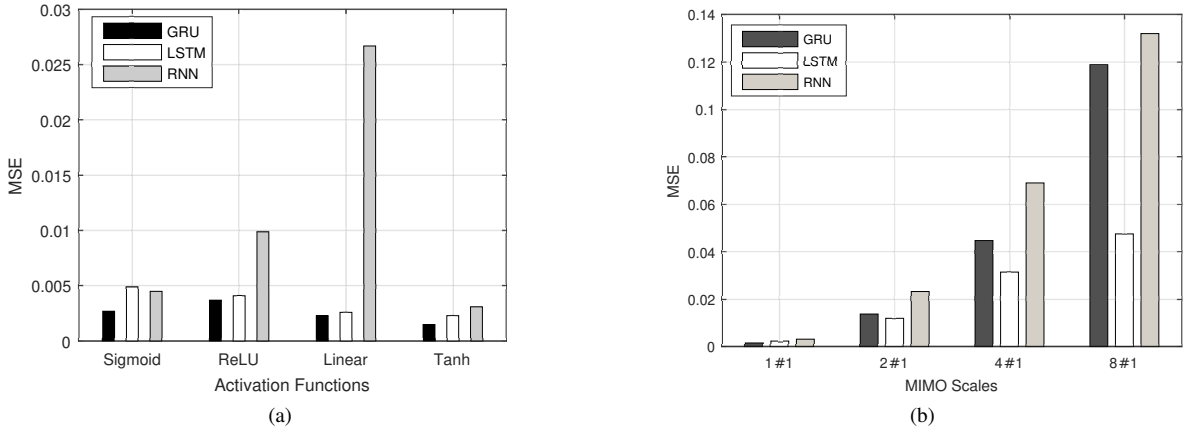

Fig. 5. Performance comparisons with respect to: (a) activation function and (b) the size of MIMO.

network to update the weights and biases by training algorithms such as the Adam optimizer used in our simulation. The training process is iteratively carried out until the network reaches a certain convergence condition. Once it completes, the trained recurrent network can be employed to predict the CSI at the next time steps.

The performance comparison of three predictors in terms of different prediction lengths is first conducted. Starting from the one-step ahead mode, i.e., D=1, which corresponds to a prediction length of 1ms according to the channel sampling rate of f_s = 1KHz. As shown in Fig.4a, the two proposed predictors clearly outperform the conventional one, where GRU achieves the best performance that is approximately 3dB better than that of RNN. Note that Fig.4 displays the MSE values in decibels for a clear illustration, calculating by MSE_{dB}=$10\log_{10}(MSE)$. Increasing D from 1 up to 5 incrementally, the results on longer lengths ranging from 1ms to 5ms are obtained. The prediction accuracy becomes worse with the growing prediction length, because channel's temporal correlation weakens. Except the case of 5ms, which seems to approach the upper limitation of channel coherence time given a maximal Doppler shift of f_d=100Hz, the proposed two predictors are better than the conventional RNN predictor in spite of the change of prediction length. The impact of different number of hidden neurons on the performance is also investigated. At the small number of 5 neurons, as illustrated in Fig.4b, LSTM and GRU achieve a similar MSE that outperforms RNN with a significant gain of more than 3dB. When the number of hidden neurons grows, the capability of predictors raises and the accuracy improves accordingly. Although LSTM performs a bit unstably, two new predictors can achieve notable performance gains. In particular, GRU performs more stably and has the gain of at least 1dB over RNN. In addition, the effect of noise is observed and illustrated in Fig.4c. The horizontal axis of the figure is the signal-to-noise ratio (SNR) of the channel samples, where the rightmost "infinity" corresponds to an extremely large SNR, meaning noiseless or noise-free. At the lower end with the SNR of 15dB, three predictors have approximately the same (worse) MSE result. That is because the effect of noise is too strong, leading to large prediction errors that overwhelm the benefits from improving the capability of predictors. Started from the SNR of 20dB, the MSE curves are separable, where the proposed predictors can effectively improve the prediction accuracy with a gain of around 1dB.

The performance comparison with respect to different activation functions and different MIMO configurations are further carried out. As illustrated in Fig.5a, the recurrent networks can collectively achieve their best performance by using $\tanh$, which is the default activation function in our simulation. But our focus is on the largest performance gain brought by the new predictors, rather than the differences exhibited among activation functions. It appears in the linear case, where the prediction accuracies of LSTM and GRU are boosted with an order of magnitude, dropping the MSE from over 0.025 to approximately 0.0025. GRU remarkably outperforms RNN in all activation functions and is also better than LSTM slightly. Looking finally at the impact of MIMO scale, as shown in Fig.5b, the performance degrades with the increasing number of antennas. That is because the neural networks have to deal with higher dimension input data, bringing burden to both the training and predicting. In the cases of 2×1 to 8×1 MIMO, LSTM provides the best prediction accuracy because its stronger power matches such higher demand.

In a nutshell, it can be concluded from the numerical results shown in Fig.4 and Fig.5 that the proposed two predictors can effectively improve the accuracy of multi-path fading channel prediction, especially the GRU predictor that reaps a significant gain in all investigated simulation scenarios.

IV. Conclusions

In this paper, two novel predictors empowered by recurrent neural networks with long short-term memory and gated recurrent unit, respectively, were proposed for multi-path fading channels in wireless communications. The performance assessment taking into account a number of influential factors, i.e., the prediction length, the number of hidden neurons, additive noise, activation function, and MIMO scale, was carried out. The numerical results in terms of mean squared error verified that the proposed predictors can bring a significant improvement in prediction accuracy, compared with the conventional RNN predictor. The positive outcomes in this paper encourages a further exploration on the application of machine learning techniques in fading channel prediction. As the follow-up work, we will not only apply advanced learning techniques, such as deep neural networks, but also design new dedicated algorithms, to further improve predication accuracy.

References

[1] J. Zheng and B. D. Rao, "Capacity analysis of MIMO systems using limited feedback transmit precoding schemes," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2886–2901, 2008.

[2] Q. Wang *et al.*, "Multi-user and single-user throughputs for downlink MIMO channels with outdated channel state information," *IEEE Wireless Commun. Lett.*, vol. 3, no. 3, pp. 321–324, 2014.

[3] K. T. Truong and R. W. Heath, "Effects of channel aging in Massive MIMO systems," *Journal of Commun. and Net.*, vol. 15, no. 4, pp. 338–351, Sep. 2013.

[4] P. Aquilina and T. Ratnarajah, "Performance analysis of IA techniques in the MIMO IBC with imperfect CSI," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1259–1270, 2015.

[5] X. Yu *et al.*, "Unified performance analysis of transmit antenna selection with OSTBC and imperfect CSI over Nakagami-m fading channels," *IEEE Trans. Veh. Technol.*, vol. 67, no. 1, pp. 494–508, 2017.

[6] J. L. Vicario *et al.*, "Opportunistic relay selection with outdated CSI: Outage probability and diversity analysis," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2872–2876, Jun. 2009.

[7] Z. Wang *et al.*, "Resource allocation in OFDMA networks with imperfect channel state information," *IEEE Commun. Lett.*, vol. 18, no. 9, pp. 1611–1614, 2014.

[8] Y. Teng *et al.*, "Effect of outdated CSI on handover decisions in dense networks," *IEEE Commun. Lett.*, vol. 21, no. 10, pp. 2238–2241, 2017.

[9] A. Hyadi *et al.*, "An overview of physical layer security in wireless communication systems with CSIT uncertainty," *IEEE Access*, vol. 4, pp. 6121–6132, 2016.

[10] W. Jiang *et al.*, "Neural network based wireless channel prediction," in *Machine Learning for Future Wireless Communications*, F. L. Luo, Ed. United Kindom: John Wiley&Sons and IEEE Press, Dec. 2019, ch. 16.

[11] ——, "A robust opportunistic relaying strategy for co-operative wireless communications," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2642–2655, Apr. 2016.

[12] D. J. Love *et al.*, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, 2008.

[13] A. Duel-Hallen, "Fading channel prediction for mobile radio adaptive transmission systems," *Proceedings of the IEEE*, vol. 95, no. 12, pp. 2299–2313, Dec. 2007.

[14] R. O. Adeogun *et al.*, "Extrapolation of MIMO mobile-to-mobile wireless channels using parametric-model-based prediction," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4487–4498, 2014.

[15] A. Duel-Hallen *et al.*, "Long-range prediction of fading signals," *IEEE Signal Process. Mag.*, vol. 17, no. 3, pp. 62–75, May 2000.

[16] J.-Y. Wu and W.-M. Lee, "Optimal linear channel prediction for LTE-A uplink under channel estimation errors," *IEEE Trans. Veh. Technol.*, vol. 62, no. 8, pp. 4135–4142, Oct. 2013.

[17] W. Jiang *et al.*, "A comparison of wireless channel predictors: Artificial Intelligence versus Kalman filter," in *Proc. of IEEE Intl. Commu. Conf. (ICC)*, Shanghai, China, May 2019.

[18] D. Silver *et al.*, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, pp. 484–489, Jan. 2016.

[19] J. Connor *et al.*, "Recurrent neural networks and robust time series prediction," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 240–254, Mar. 1994.

[20] W. Liu *et al.*, "Recurrent neural network based narrowband channel prediction," in *Proc. IEEE Vehicular Tech. Conf. (VTC)*, Melbourne, Australia, May 2006.

[21] K. T. Truong and R. W. Heath, "Fading channel prediction based on combination of complex-valued neural networks and chirp Z-transform," *IEEE Trans. Neural Netw.*, vol. 25, no. 9, pp. 1686–1695, Sep. 2014.

[22] W. Jiang and H. D. Schotten, "Recurrent neural network-based frequency-domain channel prediction for wideband communications," in *Proc. IEEE Vehicular Tech. Conf. (VTC)*, Kuala Lumpur, Malaysia, Apr. 2019.

[23] W. Jiang *et al.*, "Multi-antenna fading channel prediction empowered by artificial intelligence," in *Proc. IEEE Vehicular Tech. Conf. (VTC)*, Chicago, USA, Aug. 2018.

[24] W. Jiang and H. Schotten, "Neural network-based channel prediction and its performance in multi-antenna systems," in *Proc. IEEE Vehicular Tech. Conf. (VTC)*, Chicago, USA, Aug. 2018.

[25] R.-F. Liao *et al.*, "The Rayleigh fading channel prediction via deep learning," *Wireless Commu. and Mobile Compu.*, Jul. 2018.

[26] W. Jiang and H. D. Schotten, "Neural network-based fading channel prediction: A comprehensive overview," *IEEE Access*, vol. 7, pp. 118 112–118 124, Aug. 2019.

[27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Dec. 1997.

[28] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *preprint arXiv:1406.1078*, Jun. 2014.