

Long-Range MIMO Channel Prediction Using Recurrent Neural Networks

Wei Jiang^{*†}, Mathias Strufe^{*}, and Hans Dieter Schotten^{†*}

^{*}Intelligent Networking Department, 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Abstract—Outdated channel state information (CSI) has a severely negative impact on the performance of a wide variety of adaptive transmission systems. Channel prediction is an effective method that can directly improve the quality of CSI. To realize the full potential of adaptive systems, the prediction horizon should be long enough to at least compensate for the time delay. In this paper, therefore, we focus on the problem of long-range prediction (LRP), i.e., how to forecast fading channels as far ahead as possible. Two different LRP approaches - *Multi-Step Prediction* and *Fading Signal Processing* - are proposed for the predictors based on classical Kalman filter and recently proposed recurrent neural networks. As an application example, we present an LRP-aided transmit antenna selection system, whose performance in noisy and correlated channels is evaluated. Numerical results reveal that the RNN predictor can achieve a comparable performance with respect to the classical predictor, while avoiding its drawbacks in parameter estimation and multi-step processing.

I. INTRODUCTION

Due to feedback and processing delays, channel state information (CSI) at the transmitter might be outdated before its actual usage, especially in fast fading channels. It has been well proved that outdated CSI has a severe impact on the performance of a wide variety of adaptive wireless techniques, such as precoding in multiple-input multiple-output (MIMO) [1] and Massive MIMO, interference alignment [2], transmit antenna selection [3], transmit diversity [4], cooperative relaying [5], coordinated multi-point transmission [6], physical layer security [7], mobility management [8], etc. In the literature, a large number of algorithms and protocols have been proposed to combat outdated CSI. However, these methods either *passively* compensate for the performance loss with a cost of scarce wireless resources [9] or aim to achieve merely a portion of the full potential under imperfect CSI [10].

A technique known as channel prediction that can *actively* forecast future CSI has drawn much attention from researchers due to its potential of effectively solving this problem. Through exploiting temporal correlation, a prediction method based on a Kalman filter (KF) has been proposed [11]. It models a fading channel as an autoregressive (AR) process [12] and extrapolates future CSI using a weighted linear combination of

current and a series of past CSI. However, this method needs to estimate the maximal Doppler shift, which is by far difficult, if not impossible, in practice. The AR predictor can only enable one-step prediction. Although long-range prediction can be realized by recursively reusing predicted values at previous time instants, the problem of error propagation is raised. Recently, exploiting the capability on time-series prediction of a recurrent neural network (RNN) [13], a narrow-band channel predictor [14] and its extension for MIMO channels [15] have been proposed. In [16], the authors proposed to employ a real-valued RNN to improve prediction accuracy, and justified its performance in [17], followed by a frequency-domain predictor [18] proposed for frequency-selective MIMO channels. As a data-driven approach, the complex modeling and tedious parameter estimation process in the AR model can be totally avoided. Training a RNN does not need prior knowledge of the fading channel, and a series of channel response samples is enough, eliminating the gap between modeling and reality.

Beyond the aforementioned works, the limit on prediction horizon and the methods to realize long-range prediction (LRP) is still an open issue. In this paper, therefore, we will investigate the problem of LRP, i.e., how to forecast MIMO fading channels as far ahead as possible. Two different LRP approaches, i.e., *Multi-Step Prediction* (MSP) employing a multi-step recurrent network that can flexibly tune the number of prediction steps, and *Fading Signal Processing* (FSP) that lowers the sampling rate of a fading signal before feeding into a predictor, are proposed. To measure the prediction range in a unified way, a metric called *the prediction factor* with respect to the coherence time of channels is defined in this paper. Performance assessment of outage probability in a transmit antenna selection (TAS) system with the aid of LRP is carried out through Monte-Carlo simulations. Some representative results taking into account different influential factors, such as additive noise, inter-antenna correlation, and Doppler spectra, are illustrated.

The rest of this paper is organized as follows: Section II briefly introduces the structure of a RNN. Section III details two LRP approaches. The simulation configurations and numerical results are illustrated in Section IV, and conclusions are made in Section V.

^{*}This work was supported by the German Federal Ministry of Education and Research (BMBF) under the *TACNET 4.0* project with grant number *KIS15GT1007*.

II. STRUCTURE OF A RNN

Recurrent neural network is a popular machine learning technique that has shown great potential in time-series prediction tasks [13]. The structure of a recurrent network used to build a channel predictor is shown in Fig.1. For simplicity, only a *shallow* network is illustrated, while the number of hidden layers can be added to form a *deep* (learning) neural network. It consists of three layers: a hidden layer with N_L neurons, a layer having N_o outputs, and an input layer with N neurons including N_E external inputs and N_F feedback inputs, where $N=N_E+N_F$. Representing the external input by a vector $\mathbf{x}_e=[x_1, \dots, x_{N_E}]^T$, while the output by $\mathbf{y}=[y_1, \dots, y_{N_o}]^T$. Denoting the mapping from the output to the feedback $\mathbf{f}=[f_1, \dots, f_{N_F}]^T$ as a function Φ , we have $\mathbf{f}=\Phi(\mathbf{y})$. As a combination of \mathbf{x}_e and \mathbf{f} , the input vector is represented by $\mathbf{x}=[x_1, \dots, x_{N_E}, f_1, \dots, f_{N_F}]^T$. The behaviour of a RNN is mainly determined by the values of network weights. Each connection between the output of a neuron in the predecessor layer and the input of a neuron in the successor layer is assigned a weight. As shown in Fig.1, $w_{l,n}$ denotes the weight connecting the n^{th} input and the l^{th} hidden neuron, while $c_{m,l}$ is the weight for connecting hidden neuron l and output m , where $1 \leq n \leq N$, $1 \leq l \leq N_L$, and $1 \leq m \leq N_o$.

Constructing a $N_L \times N$ weight matrix \mathbf{W} as

$$\mathbf{W} = \begin{bmatrix} w_{1,1} & \cdots & w_{1,N} \\ \vdots & \ddots & \vdots \\ w_{N_L,1} & \cdots & w_{N_L,N} \end{bmatrix}, \quad (1)$$

the input for the hidden layer is expressed in matrix form by

$$\mathbf{z}^h = \mathbf{W}\mathbf{x} + \mathbf{b}^h, \quad (2)$$

where $\mathbf{b}^h = [b_1^h, \dots, b_{N_L}^h]^T$ denotes the vector of biases in the hidden layer. The activation function is an important feature of a neuron, having also an impact on the behaviour of a neural network. As usual, a sigmoid function is employed to deal with nonlinearity, which is defined as

$$S(x) = \frac{1}{1 + e^{-x}}. \quad (3)$$

Substituting (2) into (3), the activation vector of the hidden layer is thus

$$\mathbf{a}^h = S(\mathbf{z}^h) = S(\mathbf{W}\mathbf{x} + \mathbf{b}^h), \quad (4)$$

where $S(\mathbf{z}^h)$ means an element-wise operation for simplicity, i.e., $S(\mathbf{z}^h) = [S(z_1^h), \dots, S(z_{N_L}^h)]^T$. In analogous to (1), another weight matrix \mathbf{C} having a dimension of $N_o \times N_L$ with entries $\{c_{m,l}\}$ is introduced. Then, the activation vector for the output layer is:

$$\mathbf{y} = S(\mathbf{V}\mathbf{a}^h + \mathbf{b}^y). \quad (5)$$

The operation of a RNN predictor is divided into two phases: training and predicting. Provided a training data set, a series of continuous channel responses sampled from a real channel or synthetical model, the network feeds forward each

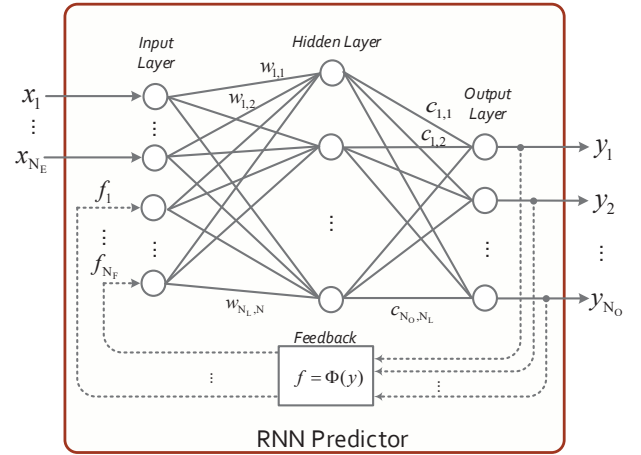


Fig. 1. Schematic diagram of a recurrent neural network.

sample and compares its resulting output with a desired value. Prediction errors are propagated back through the network, causing recursive updates of weights and biases until a certain convergence condition satisfies. Afterwards, the trained RNN is applied to predict unknown upcoming channel state based on current and a few past states. The training procedure of a network has already been well studied such as [19].

III. LONG-RANGE PREDICTION

Without loss of generality, a multi-antenna wireless system with N_t transmit and N_r receive antennas in a flat fading channel is modeled as

$$\mathbf{r}(t) = \mathbf{H}(t)\mathbf{s}(t) + \mathbf{n}(t), \quad (6)$$

where $\mathbf{r}(t)$ represents the $N_r \times 1$ received symbol vector at time t , \mathbf{s} is the $N_t \times 1$ transmit vector, \mathbf{n} stands for the vector of additive white Gaussian noise (AWGN), and $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the matrix of time-varying channel gains. The channel gain between transmit antenna n_t and receive antenna n_r is represented by a complex-valued variable $h_{n_r n_t} \in \mathbb{C}^{1 \times 1}$. Owing largely to processing and feedback delays, the CSI at the time of selecting adaptive parameters may substantially differ from the CSI at the instant of using these parameters to transmit signals, i.e., $\mathbf{H}(t) \neq \mathbf{H}(t+\tau)$, where τ denotes the delay. The outdated CSI severely deteriorates the performance of a wide variety of adaptive transmission systems [1] - [8]. The aim of LRP is to predict a fading channel as far ahead as possible where predicted CSI $\hat{\mathbf{H}}(t+\tau)$ can still provide useful information about its actual value $\mathbf{H}(t+\tau)$.

Two different LRP approaches for MIMO channels, i.e., M-SP, taking advantage of the multi-step flexibility of a RNN, and FSP that down-samples a fading signal to lower its sampling rate before feeding it into a predictor, are detailed as follows:

A. Multi-Step Prediction

Due to the similarity between RNN's structure and MIMO's channel model, which are both multiple inputs and multiple

outputs with fully weighted connections, the RNN predictor congenitally suits a multi-antenna system by tuning the number of input and output neurons in terms of the number of antennas. It is computationally efficient in comparison with predicting each subchannel independently with a separate linear filter or neural network. To adapt to the input layer, a channel matrix needs to be vectorized following:

$$\mathbf{h}_v = \vec{\mathbf{H}} = [h_{11}, h_{12}, \dots, h_{N_r N_t}]. \quad (7)$$

At time t , $\mathbf{H}(t)$ and its delays $\mathbf{H}(t-1), \dots, \mathbf{H}(t-d)$ are fed into the RNN as the external input, which in this case can be rewritten as $\mathbf{x}_e(t) = [\mathbf{h}_v(t), \mathbf{h}_v(t-1), \dots, \mathbf{h}_v(t-d)]$. The feedback function Φ can be flexibly configured to generate a small feedback including only $\hat{\mathbf{H}}(t)$, where $\mathbf{f} = [\hat{\mathbf{h}}_v(t)]$, or a larger one like $\{\hat{\mathbf{H}}(t), \hat{\mathbf{H}}(t+1), \dots\}$, depending on the trade-off between prediction accuracy and computational complexity. The output of a multi-step RNN predictor is denoted by $\hat{\mathbf{H}}(t+D)$, where D is a positive integer standing for the number of steps being predicted ahead. If $D=1$, it returns back to a one-step predictor. From the perspective of training, there is no intrinsic difference between one-step and multi-step predictors except that the desired value for calculating the propagation error shifts from $\mathbf{H}(t+1)$ to $\mathbf{H}(t+D)$, resulting in different weights and biases.

In contrast, according to [18], the AR predictor based on a Kalman filter is given by:

$$\hat{\mathbf{H}}(t+1) = \sum_{k=1}^p a_k \mathbf{H}(t-k+1), \quad (8)$$

where p is the number of filter taps and a_1, a_2, \dots, a_p are AR coefficients. It is noted that (8) can only provide one-step prediction. By employing predicted values recursively, a multi-step AR predictor can be built, i.e.,

$$\hat{\mathbf{H}}(t+D) = \sum_{k=1}^p a_k \hat{\mathbf{H}}(t-k+D) \quad (9)$$

B. Fading Signal Processing

A discrete-time channel series $\{\mathbf{H}[k] | k=1, 2, 3, \dots\}$ is a sampling of continuous-time channel response. Its sampling rate f_c is generally as same as the data symbol rate f_s for the sake of channel estimation and symbol detection, where the interval of each prediction step equals to one symbol period $T_s = 1/f_s$. However, using a RNN to predict a fading signal symbol-by-symbol is both inefficient and unnecessary. According to the Sampling Theorem [20], a discrete-time series can capture all the information of a continuous-time signal with a finite bandwidth if the sampling rate is greater than the Nyquist rate. In our case, that is to say, the original channel response can be properly recovered if the sampling rate is greater than two times the maximal Doppler shift f_d . Since f_d is generally far less than f_s , it is possible to predict a fading channel with a lower rate that is less than f_s but greater than $2f_d$, namely $2f_d < f_c < f_s$. Predicting a fading signal at a lower sampling rate means the length of each prediction step is enlarged.

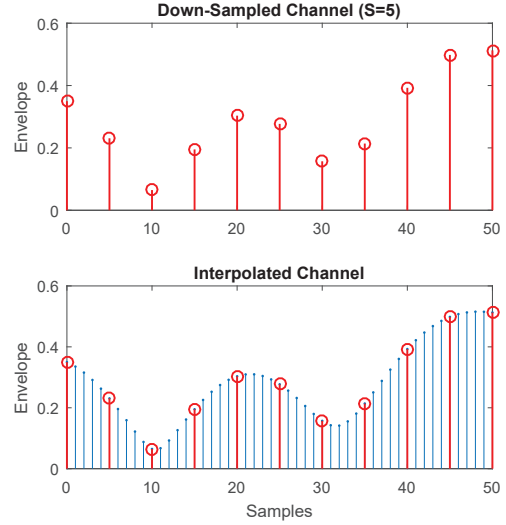


Fig. 2. A down-sampled fading signal with a factor of $S = 5$ and its interpolated version.

In addition to MSP, therefore, another LRP method can be realized by means of using a fading signal with a sampling rate as low as possible (equivalent to extend the prediction interval). To be specific, the series $\{\mathbf{H}[k] | k=1, 2, 3, \dots\}$ at a sampling rate of $f_c = f_s$ is down-sampled by a factor of S , generating a new series $\{\mathbf{H}[k'] | k'=S, 2S, 3S, \dots\}$ that has a lowered rate of $f'_c = f_s/S$. An example of down-sampling and interpolating a fading signal with a factor of $S=5$ (for a better illustration) is shown in Fig.2. Obviously, if the prediction is conducted at the lower rate, the signal at the upper of the figure, the prediction range of each step is multiplied by S , compared with the original signal. Once a RNN is trained by down-sampled signals, feeding $\mathbf{H}[k]$ and its delayed versions $\mathbf{H}[k-S], \dots, \mathbf{H}[k-dS]$ into a one-step predictor, its output is $\hat{\mathbf{H}}[k+S]$, which has a longer prediction range S times that of $\hat{\mathbf{H}}[k+1]$. It can be also derived that a D -step predictor can forecast SD symbols ahead by combining MSP with FSP.

In general, a wireless system acquires CSI by inserting comb-type pilot symbols into transmitted signals at the transmitter and conducting channel estimation at the receiver. FSP can be well combined with a pilot-assisted system, e.g., transmit antenna selection in a MIMO system. As illustrated in Fig.3, the signal transmission is organized in block-wise, where a pilot symbol is inserted in each block that has a total size of S symbols. At time t , the receiver conducts channel estimation based on the received pilot symbol, resulting in a channel estimate $\mathbf{H}[t]$. Feeding $\mathbf{H}[t]$ into a multi-step predictor, together with the estimates $\mathbf{H}[t-S], \dots, \mathbf{H}[t-dS]$ obtained in previous blocks, the predictor can extrapolate $\hat{\mathbf{H}}[t+SD]$. With a number of neighboring predicted values, making use of channel's temporal correlation, the interpolator can also provide predicted CSI at the symbol rate, i.e., $\hat{\mathbf{H}}[t+1], \hat{\mathbf{H}}[t+2], \dots, \hat{\mathbf{H}}[t+SD]$, if needed.

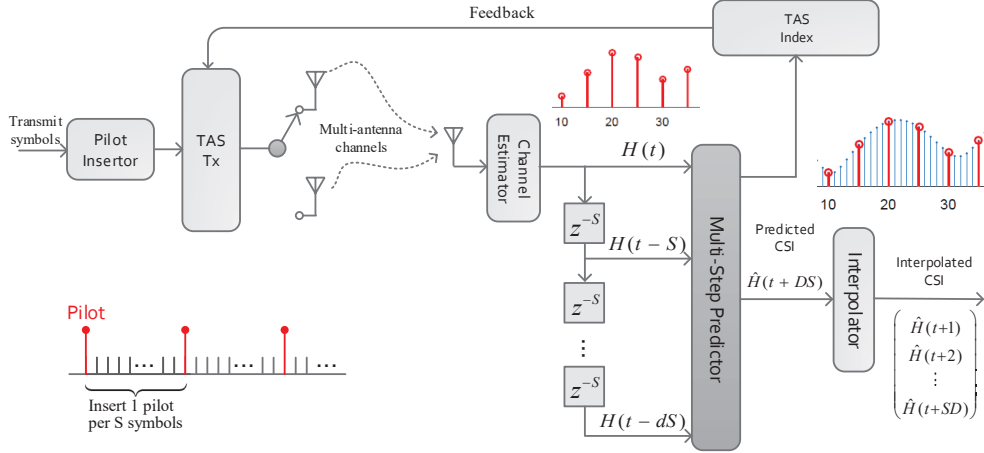


Fig. 3. Illustration of a long-range predictor applied for TAS in a MIMO system.

IV. NUMERICAL RESULTS

In this section, the performance of the long-range RNN predictor over multi-antenna channels is evaluated via Monte-Carlo simulations, in comparison with that of the predictor based on a Kalman filter. Outage probability achieved by LRP-assisted TAS in a MIMO system is illustrated. To measure the prediction range in a unified way, we define a metric called *the prediction factor*, which is the ratio of the time predicted ahead T_p and the coherence time T_c , namely

$$\eta = \frac{T_p}{T_c}. \quad (10)$$

Since $T_p = SD/f_s$ and $T_c = 1/f_d$ generally, (10) can be further rewritten as

$$\eta = \frac{SDf_d}{f_s}. \quad (11)$$

Given a flat fading Rayleigh channel with an average power gain of 0dB, where the channel coefficient h is zero-mean circularly-symmetric complex Gaussian random variable with the variance of 1, i.e., $h \sim \mathcal{CN}(0, 1)$. The symbol rate is set to $f_s = 10^5$ Hz to satisfy the flat fading assumption and the maximal Doppler frequency is $f_d = 100$ Hz to emulate fast fading environment. The signal transmission is organized in block-wise, with a block size of $S = 50$ that includes N_t antenna-specific pilot symbols inserted at the head of each block. Through the observation in the simulation, the optimal number of neurons in the hidden layer is set to $N_L = 10$, while the length of the tapped delay line is selected to $d = 3$. The simulation parameters are summarized in Table I.

To train a recurrent network, we build a training dataset that consists of a series of CSI extracted from consecutive 10^3 blocks, i.e., $\{\mathbf{H}[t] | 1 \leq t \leq 10^3\}$. The training process starts from an initial state with random weights and biases. At iteration t , feeding $\mathbf{H}[t]$ into the network, the resultant output is compared with the desired value and the prediction error $\hat{\mathbf{H}}[t+D] - \mathbf{H}[t+D]$ is back propagated to update the

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Sampling rate	$f_s = 10^5$ Hz
Down-sampling factor	$S = 50$
Max. Doppler frequency	$f_d = 100$ Hz
MIMO configuration	4×1
Channel	Rayleigh flat fading
Doppler Spectra (default)	Jakes
Neural Network	3-layer RNN
Training algorithm	Levenberg-Marquardt
Number of hidden neuron	$N_L = 10$
Input tapped-delay length	$d = 3$

weights and biases by training algorithms such as Levenberg-Marquardt [19]. This process is iteratively carried out until the convergence condition reaches. In contrast, a KF predictor does not need training. Its coefficients a_1, a_2, \dots, a_p in (9) can be figured out if f_d and f_s are known, according to [12].

Suppose the applied MIMO system is a uniform linear array having $N_t = 4$ transmit and $N_r = 1$ receive antennas, a single transmit antenna with the largest instantaneous channel gain is selected. For simplicity, the selection is carried out in block-wise, i.e., each block selects its respective transmit antenna using the CSI of its pilot symbol. The outage probability, an important performance metric over fading channels, is defined as $P(R) = \Pr\{\log_2(1 + \text{SNR}) < R\}$, where \Pr is the notation of mathematical probability and R means a target end-to-end data rate that is set to 1bps/Hz in our simulations. To decide the transmit antenna for upcoming block $t+D$, the possible selection methods are:

- The *perfect* mode that chooses the antenna according to the perfect CSI at that block, i.e., $\mathbf{H}(t+D)$, despite it is never practically available.
- The *outdated* mode in traditional TAS systems where the

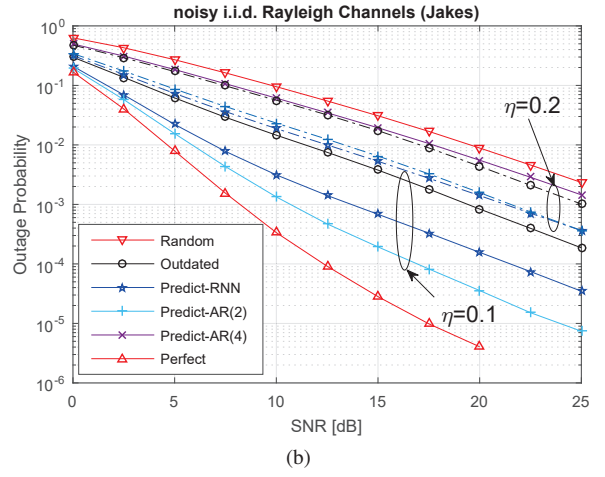
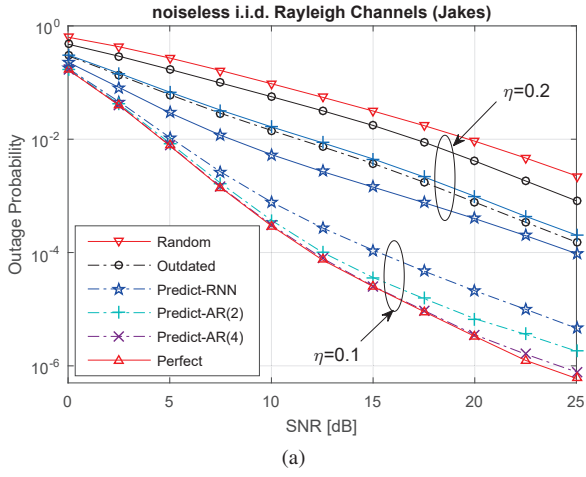


Fig. 4. Outage probabilities of the prediction-aided TAS system in (a) noiseless *i.i.d.* and (b) noisy *i.i.d.* Rayleigh channels.

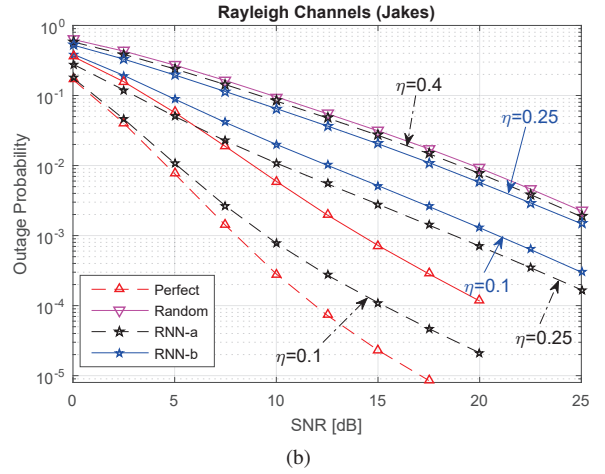
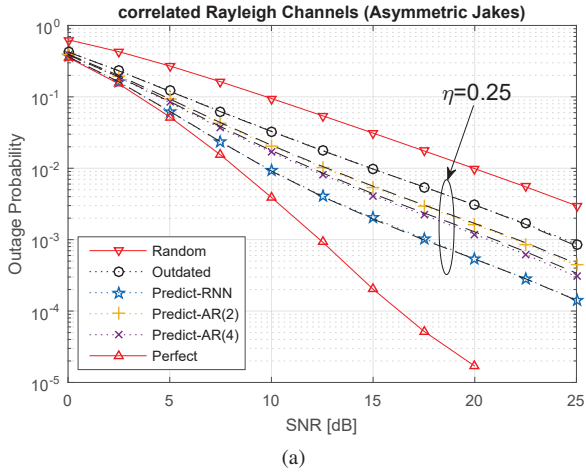


Fig. 5. Outage probabilities of the prediction-aided TAS system: (a) in noiseless correlated channels with the Doppler spectrum of *Asymmetric Jakes*, and (b) in noisy correlated (solid line) or noiseless *i.i.d.* (dashed line) *Jakes* channels.

outdated CSI $\mathbf{H}(t)$ is applied.

- The *prediction* mode makes a selection decision based on the predicted CSI $\hat{\mathbf{H}}(t+D)$ that probably approximates $\mathbf{H}(t+D)$.
- The *random* mode where the antenna is randomly selected without any consideration of CSI.

The performance assessment is first carried out in noiseless *i.i.d.* MIMO channels with $f_d=100\text{Hz}$. The predictors are tuned to the multi-step mode of $D=2$ corresponding to a prediction factor of $\eta=0.1$ in terms of (11). As illustrated in Fig.4a, the outdated CSI drastically deteriorates the performance with a high SNR loss of around 14dB in comparison with the perfect mode at outage probability of 10^{-4} . The KF predictor with filter order $p=4$ denoted by *AR(4)* is able to achieve a near optimal performance, outperforming that of the RNN predictor. The TAS system with the aid of either KF or RNN can achieve remarkable gains over the outdated mode, justified the effectiveness of the proposed LRP approaches. Further increased the prediction factor to $\eta=0.2$

by means of tuning to $D=4$, the performance gap between the outdated mode and the perfect mode enlarges to over 16dB. Two observations need to be highlighted: 1) although the performance of LRP-aided system drops when the prediction range become longer, it still has at least 5dB gain over the outdated mode; and 2) the RNN predictor is more robust in a longer range, with nearly 3dB over the KF predictors. It implies that the multi-step RNN is a better option for LRP.

In practice, the available CSI is impaired by estimation errors since additive noise cannot be avoided in the process of channel estimation. Under the assumption that the SNR of pilot symbols is $\text{SNR}_p=25\text{dB}$, the performance evaluation in noisy channels is also conducted. In the case of $\eta=0.1$, the KF predictor still outperforms the RNN predictor, while the better performance is achieved by *AR(2)*, rather than *AR(4)* in the noiseless case. As illustrated in Fig.4b, the performance of *AR(4)* is very close to that of the random mode, even slightly worse than the outdated mode. That is because the problem of error propagation becomes severe with a larger

filter order when the applied CSI contains estimation error. From this result, we can remark that the RNN is more robust against additive noise than the KF.

The purpose of Fig.5a is three fold: First, to observe the impact of antenna correlation, applying the correlation matrix recommended in 3GPP LTE standards [21] and selecting the medium correlation with coefficient of $\alpha=0.3$. Second, to observe the effect of different Doppler spectra by replacing the default *Jakes* model with *Asymmetric Jakes* because the KF predictor is specifically optimized for the *Jakes* model. As shown in the figure, the RNN predictor achieves a notably SNR gain of more than 7dB over the outdated mode in the case of $\eta=0.25$ and is superior to the KF predictor with a gain of over 3dB. Last but certainly not least is to verify the impact of different sampling rates. As given in Table I, the default is $f_s=10^5$ Hz and $S=50$, amounting to a block length of 0.5ms. Using a different rate but keeping the block length constant, e.g., $f_s=10^4$ Hz and $S=5$, it is found that the performance is independent of f_s . As shown in the figure, on each colored curves, a block dash line, obtained in the case of $f_s=10^4$ Hz and $S=5$, is overlapped. It implies that the performance of the prediction-aided TAS system is decided by the prediction range, regardless different sampling and down-sampling rates, justified that the defined prediction factor η in (11) can measure the predication range in a unified way.

Fig.5b illustrates the prediction limit of the RNN predictor in both noiseless *i.i.d.* channels, denoted by *RNN-a* in the legend with dashed curves in the figure, and noisy correlated channels, denoted by *RNN-b* with solid curves. The limit is bounded by the random mode where the LRP cannot provide any useful information to improve a bit performance over the random selection. It can be observed that around $\eta=0.4$ and $\eta=0.25$ are the limitation for noiseless *i.i.d.* and noisy correlated channels, respectively. Recalling (11), it implies that the RNN predictor is still effective until the prediction range increases beyond one fourth of the coherence time, corresponding to approximately 2.5ms in the condition of $f_d=100$ Hz, which is meaningful from the practical view in comparison with the length of a radio frame of 10ms in LTE systems for example.

V. CONCLUSION

This paper investigated the problem of long-range fading channel prediction. Two different LRP approaches - *Multi-Step Prediction* and *Fading Signal Processing* - were proposed for recurrent neural networks. A metric used to indicate the prediction range in a unified way, referred to as the prediction factor, was defined. Performance assessment in terms of outage probability for LRP-aided TAS in a MIMO system was carried out. The numerical results illustrated that the RNN predictor can achieve a comparable performance with respect to the KF predictor, while avoiding its difficulties in modeling and parameter estimation. It was also revealed that LRP can effectively combat the outdated CSI in a range of

over 25% of the coherence time, which is long enough to achieve remarkable performance gains in fast fading channels. Taking into account the moderate computational complexity of a RNN predictor (being proved in pervious works), the LRP is promising to facilitate adaptive transmission systems from the practical viewpoint.

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, pp. 2886–2901, 2008.
- [2] P. Aquilina and T. Ratnarajah, "Performance analysis of IA techniques in the MIMO IBC with imperfect CSI," *IEEE Trans. Commun.*, vol. 63, pp. 1259–1270, 2015.
- [3] 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, pp. 494–508, 2017.
- [4] E. Onggosanusi *et al.*, "Performance analysis of closed-loop transmit diversity in the presence of feedback delay," *IEEE Trans. Commun.*, vol. 49, no. 9, pp. 1618–1630, Sep. 2001.
- [5] 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.
- [6] L. Su *et al.*, "The value of channel prediction in CoMP systems with large backhaul latency," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4577–4590, Nov. 2013.
- [7] 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.
- [8] Y. Teng *et al.*, "Effect of outdated CSI on handover decisions in dense networks," *IEEE Commun. Lett.*, vol. 21, pp. 2238–2241, 2017.
- [9] W. Jiang *et al.*, "A robust opportunistic relaying strategy for co-operative wireless communications," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2642–2655, Apr. 2016.
- [10] D. J. Love *et al.*, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 1341–1365, 2008.
- [11] 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.
- [12] K. E. Baddour *et al.*, "Autoregressive modeling for fading channel simulation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1536–1276, Jul. 2005.
- [13] 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.
- [14] W. Liu *et al.*, "Recurrent neural network based narrowband channel prediction," in *Proc. IEEE Vehicular Tech. Conf. (VTC)*, Melbourne, Australia, May 2006.
- [15] 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.
- [16] W. Jiang and H. D. Schotten, "Multi-antenna fading channel prediction empowered by artificial intelligence," in *Proc. IEEE Vehicular Tech. Conf. (VTC)*, Chicago, USA, Aug. 2018.
- [17] W. Jiang *et al.*, "Neural network-based channel prediction and its performance in multi-antenna systems," in *Proc. IEEE Vehicular Tech. Conf. (VTC)*, Chicago, USA, Aug. 2018.
- [18] W. Jiang and H. Schotten, "Recurrent neural network-based frequency-domain channel prediction for wideband communications," in *Proc. IEEE Vehicular Tech. Conf. (VTC)*, Kuala Lumpur, Malaysia, Apr. 2019.
- [19] X. Fu *et al.*, "Training recurrent neural networks with the Levenberg Marquardt algorithm for optimal control of a grid-connected converter," *IEEE Trans. Neural Netw.*, vol. 26, no. 9, pp. 1900–1912, Sep. 2015.
- [20] A. V. Oppenheim and R. W. Schaffer, *Digital Signal Processing*, 1st ed. New York, US: Prentice-Hall, 1975.
- [21] *3GPP TS36.104: Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception*, Sep. 2018, v15.4.0.