IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ARTIFICIAL INTELLIGENCE FOR PHYSICAL-LAYER WIRELESS COMMUNICATIONS

Received August 7, 2019, accepted August 22, 2019, date of publication August 26, 2019, date of current version September 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2937588

Neural Network-Based Fading Channel Prediction: A Comprehensive Overview

WEI JIANG, (Senior Member, IEEE), AND HANS D. SCHOTTEN, (Member, IEEE)
[1]German Research Center for Artificial Intelligence (DFKI), 67663 Kaiserslautern, Germany
[2]Department of Electrical and Computer Engineering, University of Kaiserslautern, 67663 Kaiserslautern, Germany

Corresponding author: Wei Jiang (wei.jiang@dfki.de)

This work was supported by the German Federal Ministry of Education and Research (BMBF) through the TACNET 4.0.

ABSTRACT By adapting transmission parameters such as the constellation size, coding rate, and transmit power to instantaneous channel conditions, adaptive wireless communications can potentially achieve great performance. To realize this potential, accurate channel state information (CSI) is required at the transmitter. However, unless the mobile speed is very low, the obtained CSI quickly becomes outdated due to the rapid channel variation caused by multi-path fading. Since outdated CSI has a severely negative impact on a wide variety of adaptive transmission systems, prediction of future channel samples is of great importance. The traditional stochastic methods, modeling a time-varying channel as an autoregressive process or as a set of propagation parameters, suffer from marginal prediction accuracy or unaffordable complexity. Taking advantage of its capability on time-series prediction, applying a recurrent neural network (RNN) to conduct channel prediction gained much attention from both academia and industry recently. The aim of this article is to provide a comprehensive overview so as to shed light on the state of the art in this field. Starting from a review on two model-based approaches, the basic structure of a recurrent neural network, its training method, RNN-based predictors, and a prediction-aided system, are presented. Moreover, the complexity and performance of predictors are comparatively illustrated by numerical results.

INDEX TERMS 5G, artificial intelligence, back-propagation, channel prediction, channel state information, MIMO, OFDM, recurrent neural network, transmit antenna selection.

I. INTRODUCTION

By adapting radio transmission parameters, e.g., the constellation size, coding rate, transmit power, precoding codeword, time and frequency resource block, transmit antennas, and relays, to instantaneous channel conditions, adaptive wireless systems can potentially aid the achievement of great performance. To fully realize this potential, accurate channel state information (CSI) is required at the transmitter. In a frequency-division duplex system, the CSI is estimated at the receiver and then fed back to the transmitter, where the obtained CSI might be already outdated before its actual usage owing mainly to the feedback delay. Although a time-division duplex system can take advantage of channel reciprocity to avoid feedback, the processing delay still leads to inaccurate CSI, especially in high mobility scenarios.

It has been extensively proved that the outdated CSI severely deteriorates the performance of a wide variety of adaptive transmission techniques, including but not limited to precoding [1] and multi-user scheduling [2] in multiple-input multiple-output (MIMO) systems, massive MIMO [3], beam-forming [4], interference alignment [5], closed-loop transmit diversity [6], transmit antenna selection [7], opportunistic relaying [8], orthogonal frequency-division multiplexing (OFDM) [9], coordinated multi-point transmission [10], mobility management [11], and physical layer security [12]. In the era of the fifth generation (5G) system, this problem will become more serious. On the one hand, new applications and services such as Internet of Things, Tactile Internet, virtual and augmented reality, networked drones, and autonomous driving impose a great demand for extremely high-speed, ultra reliable, ubiquitous, and secure wireless connectivity, where adaptive transmission techniques are envisioned to play more critical roles, further emphasizing the importance of accurate CSI. On the other hand, the fluctuation of a fading channel will speed up if the velocity of moving objects increases or the wavelength of radio signals decreases according to the Doppler effect of electromagnetic

The associate editor coordinating the review of this article and approving it for publication was Guan Gui.

radiations [13]. Some of the 5G deployment scenarios, e.g., millimeter waves (having shorter wavelength), unmanned aerial vehicles and high-speed trains (with higher moving speed), suffer from faster fading channels that fluctuate more rapidly, leading to worse availability of accurate CSI, if not impossible.

To cope with the outdated CSI, a large number of mitigation algorithms and protocols have been proposed in the literature. These can be mainly categorized into two classes: *passive* methods [14] that compensate for the performance loss passively with a cost of scarce wireless resources (frequency, time, power, etc.) and *suboptimal* methods aiming to achieve merely a portion of the full performance potential under the assumption of imperfect CSI, e.g., the techniques based on limited feedback [15]. In contrast, an alternative technique referred to as channel prediction [16] provides an efficient approach to improve the quality of CSI directly without spending extra wireless resources, and therefore attracted much attention from researchers. Through statistical modeling of wireless channels, two classical model-based prediction, namely parametric model [17] and autoregressive (AR) model [18], have been developed. The former assumes that a fading channel is a superposition of a finite number of complex sinusoids, and its parameters, i.e., amplitude, angles of arrival and departure, Doppler shift, and the number of scattering sources, vary much slowly relative to channels' fluctuation rate and can be estimated accurately. However, the estimation process is tedious, while estimated parameters will quickly expire once the channel changes and therefore need to be re-estimated iteratively, leading to high computational complexity. The AR model approximates a fading channel as an AR process and extrapolates the future CSI using a weighted linear combination of past and current CSI. The AR model is computationally simpler, but it is vulnerable to impairments such as additive noise [19], making it unattractive in practice.

In March 2016 when AlphaGo, a computer program developed by Google DeepMind [20], achieved an overwhelming victory versus a human champion in the game of Go, the passion for exploring Artificial Intelligence (AI) technology was sparked almost in all scientific and engineering branches. Actually, the wireless research community started to apply AI to solve communication problems long ago. Especially in the recent period, with the development of deep neural networks, the application of AI in wireless communications was booming, such as channel estimation [21], resource allocation [22], modulation recognition [23], multi-user detection [24], multiple access [25], beam-forming [26], and autonomous network management [27]. There are a number of different neural network structures, among which recurrent neural network (RNN) has a strong capability on time-series prediction [28]. Reference [29] first proposed an application of a RNN to build a predictor for narrow-band single-antenna channels and was further extended to MIMO channels in [30]. The authors of [31] proposed to employ a real-valued RNN to implement a multi-step predictor and further verified its effectiveness in a MIMO system [32]. The feasibility of applying a deep neural network was also studied in [33]. Recently, a frequency-domain RNN predictor reported in [34] further extended the application range from frequency-flat to frequency-selective MIMO channels. Neural networks, as a data-driven approach, can totally avoid the tedious estimation process of propagation parameters in model-based methods, and can enable great flexibility of adapting to different prediction scenarios.

Beyond the aforementioned works, a comprehensive overview that can facilitate readers to quickly grasp the main ideas and catch up with the state of the art of this promising area is of great worth but unfortunately still missing. The aim of this article is to fill this gap in time. In contrast to [35] that proposed a hybrid of convolutional neural network and long short-term memory to get the CSI of downlink channels according to that of uplink channels, this article focuses on predicting future CSI from its past values at the same frequency/sub-carrier. Starting from a brief review on AR and parametric models, the basic structure of a recurrent neural network, its training method based on the back-propagation algorithm, and several variants of RNN predictors applied for different scenarios ranging from frequency-flat single-antenna channels to frequency-selective MIMO channels, are presented. To exemplify the applicability, prediction-aided transmit antenna selection (TAS) in a MIMO-OFDM system is depicted. Performance assessment in fading channels specified by 3GPP models, taking into account the factors such as the Doppler shift, spatial correlation, additive noise, and interpolation error, is carried out, followed by comparisons on computational complexity.

The rest of this paper is organized as follows: Section II reviews two model-based prediction schemes. The structure of a recurrent neural network and the back-propagation training algorithm are introduced in Section III. Section IV presents the variants of RNN predictors, followed by an illustration of a prediction-aided MIMO-OFDM system in Section V. Section VI compares the complexity and performance for different predictors by numerical results. Finally, concluding remarks are made in Section VII.

Notations: Throughout this article, vectors are denoted by bold lower case letters and matrices are bold upper case letters. For the operation of matrices, $(\cdot)^T$ and $(\cdot)^H$ notate the transpose and Hermitian transpose, respectively, $\|\cdot\|$ stands for the Euclidean norm, and $\odot$ marks the Hadamard product. $\mathbf{H}$ and $\tilde{\mathbf{H}}$ represent time-domain and frequency-domain channel responses, respectively, while $\hat{\mathbf{H}}$ is a predicted value. The bracket $(\cdot)$ indicates continuous-time signals and the square bracket $[\cdot]$ associates with discrete-time sequences.

II. MODEL-BASED CHANNEL PREDICTION

According to traditional statistical methodology, a fading channel can be modelled as a number of propagation parameters and the channel prediction is actually a problem of parameter estimation. Given the knowledge of current and several past channel impulse responses, these parameters can

be estimated and then future CSI is extrapolated. Existing model-based prediction approaches are mainly categorized into two classes: parametric and AR models, which are briefly reviewed as follows.

A. SYSTEM MODEL

This article focuses on point-to-point MIMO systems consisting of a single transmitter and a single receiver, whereas multi-user MIMO is not considered taking into account the state of the art in this field. A narrow-band MIMO system with N_t transmit and N_r receive antennas can be modeled by

$$\mathbf{r}(t) = \mathbf{H}(t)\mathbf{s}(t) + \mathbf{n}(t), \tag{1}$$

where $\mathbf{r}(t)=\left[r_1(t), \ldots, r_{N_r}(t)\right]^T$ denotes the $N_r \times 1$ vector of received signals at time t, $\mathbf{s}(t)=\left[s_1(t), \ldots, s_{N_t}(t)\right]^T$ is the $N_t \times 1$ vector of transmitted signals, $\mathbf{n}(t)$ stands for the vector of additive white noise, $\mathbf{H}(t)=\left[h_{n_r n_t}(t)\right]_{N_r \times N_t}$ is the matrix of continuous-time channel impulse responses, and $h_{n_r n_t} \in \mathbb{C}^{1\times 1}$ represents the gain of the flat fading channel between transmit antenna n_t and receive antenna n_r, where $1 \leqslant n_r \leqslant N_r$ and $1 \leqslant n_t \leqslant N_t$. Due to feedback and processing delays, the obtained CSI at the transmitter may be outdated before its actual usage, i.e., $\mathbf{H}(t) \neq \mathbf{H}(t+\tau)$, resulting in severe performance degradation of adaptive transmission systems [1] - [12]. The aim of channel prediction is to extrapolate a predicted value $\hat{\mathbf{H}}(t+\tau)$ at time t that approximates its actual value at the upcoming time $t+\tau$ as close as possible, namely $\hat{\mathbf{H}}(t+\tau) \rightarrow \mathbf{H}(t+\tau)$.

B. PARAMETRIC MODEL

The commonly used multi-path fading model [13] is to represent a single-antenna channel as the superposition of a finite number of complex sinusoids:

$$h(t) = \sum_{p=1}^{P} \alpha_p e^{j(\omega_p t+\phi_p)}, \tag{2}$$

where α_p, ϕ_p, and ω_p denote the complex amplitude, phase, and radian Doppler frequency shift of the p^{th} scattering source, respectively, $j^2=-1$ stands for the imaginary unit, and P is the total number of scatters.

Introducing the parameters of spatial dimension [17], the single-antenna model in (2) can be extended to model a MIMO propagation channel, i.e.,

$$\mathbf{H}(t) = \sum_{p=1}^{P} \alpha_p \mathbf{a}_r(\theta_p)\mathbf{a}_t^T(\psi_p)e^{j(\omega_p t+\phi_p)}, \tag{3}$$

where θ_p stands for the angle of arrival (AOA), ψ_p the angle of departure (AOD), $\mathbf{a}_r$ represents the response vector of the receive antenna array, while $\mathbf{a}_t$ for the transmit antenna array. Using a uniform linear array (ULA) with M equally spaced elements as an example, its steering vector is defined as

$$\mathbf{a}(x) = \left[1, e^{-j\frac{2\pi}{\lambda}d\sin(x)}, \ldots, e^{-j\frac{2\pi}{\lambda}(M-1)d\sin(x)}\right]^T, \tag{4}$$

where x can be replaced with the angle of arrival or departure, d is the inter-antenna spacing, and λ denotes the wavelength of carrier frequency. Based on an observation that the multi-path parameters change slowly in comparison with the fading rate of channels, the future CSI within a certain range can be extrapolated if these parameters are known. Hence, predicting a MIMO channel in terms of (3) is essentially transferred into a problem of parameter estimation. In other words, a parametric prediction model is built by estimating the number of scatters $\hat{P}$, and amplitude, Doppler shift, AOA, and AOD for each path, i.e., $\{\hat{\alpha}_p, \hat{\omega}_p, \hat{\theta}_p, \hat{\psi}_p\}_{p=1}^{\hat{P}}$. The parameter estimation follows this procedure:

1) Given K known discrete-time channel samples $\{\mathbf{H}[k]\,|k=1, \ldots, K\}$, sampled from continuous-time channel response $\mathbf{H}(t)$, a sufficiently large matrix exhibiting the required translational invariance structure in all dimensions is formed. According to [36], an block-Hankel matrix with a dimension of $N_r Q \times N_t S$ is given by

$$\hat{\mathbf{D}} = \begin{bmatrix} \mathbf{H}[1] & \mathbf{H}[2] & \cdots & \mathbf{H}[S] \\ \mathbf{H}[2] & \mathbf{H}[3] & \cdots & \mathbf{H}[S+1] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}[Q] & \mathbf{H}[Q+1] & \cdots & \mathbf{H}[K] \end{bmatrix}, \tag{5}$$

where Q is the size of Hankel matrix and $S=K-Q+1$. Using (5), a spatio-temporal covariance matrix containing the temporal and spatial correlation can be calculated as

$$\hat{\mathbf{C}} = \frac{\hat{\mathbf{D}}\hat{\mathbf{D}}^H}{N_t S}, \tag{6}$$

where $(\cdot)^H$ denotes the Hermitian conjugate transpose.

2) Then, the number of dominant scattering sources can be estimated using the minimum description length (MDL) criterion (see [37]), which is written as

$$\hat{P} = \arg \min_{z=1,\ldots,(N_r Q-1)} \left[S \log(\lambda_z) + \frac{(z^2+z)\log S}{2} \right], \tag{7}$$

where λ_z is the z^{th} eigenvalue of $\hat{\mathbf{C}}$.

3) Making use of classical algorithms, such as MUltiple SIgnal Classification (MUSIC) and Estimation of Signal Parameters by Rotational Invariance Techniques (ESPRIT), $\{\hat{\theta}_p, \hat{\psi}_p, \hat{\omega}_p\}_{p=1}^{\hat{P}}$ can be jointly estimated by further exploiting the invariance structure in $\hat{\mathbf{C}}$. For simplicity, the details of the calculation process are omitted but can refer to [17].

4) Given $\{\hat{\theta}_p, \hat{\psi}_p, \hat{\omega}_p\}_{p=1}^{\hat{P}}$, $\{\hat{\alpha}_p\}_{p=1}^{\hat{P}}$ is then calculated.

Substituting all estimates into (3), a parametric channel predictor is obtained

$$\hat{\mathbf{H}}(\tau) = \sum_{p=1}^{\hat{P}} \hat{\alpha}_p \mathbf{a}_r(\hat{\theta}_p)\mathbf{a}_t^T(\hat{\psi}_p)e^{j(\hat{\omega}_p \tau+\phi_p)}, \tag{8}$$

where τ denotes a time range for which the CSI is to be predicted.

From the estimation procedure, it is concluded that the generality and applicability of this method is highly constrained. For instance, if a different array is applied, (3) or (4) should be modified accordingly, such as [38], where a set of polarization parameters is additionally employed for modelling a polarized array. The propagation models or steering vectors for some irregular arrays are probably intractable, which limits the applicability of this method. Moreover, the estimation process is tedious and the computational complexity is high due to the manipulation of high-dimensional matrices, as analyzed in [38]. Last but not the least, the obtained estimates expire quickly with the change of mobile propagation environments, especially in high mobility. According to [39], the region of mobile movements where the propagation parameters are assumed constant is up to 50 wavelengths. It implies that the estimation process needs to be periodically and frequently conducted, resulting in extreme high complexity, which is unattractive from the viewpoint of practical implementation.

C. AUTOREGRESSIVE MODEL

Alternatively, the discrete-time impulse response of a time-varying channel can be modelled as an autoregressive process and a Kalman filter (KF) is utilized to estimate AR coefficients so as to build a linear predictor, which extrapolates future CSI by combining weighted current and a series of past CSI [40]–[43]. According to [18], a complex AR process of order p denoted by AR(p) can be generated via a time-domain recursion

$$x[n] = \sum_{k=1}^{p} a_k x[n-k] + w[n], \tag{9}$$

where $w[n]$ denotes zero mean complex Gaussian noise with the variance of σ_p^2, and $\{a_1, a_2, \ldots, a_p\}$ are AR coefficients. The corresponding power spectral density (PSD) of the AR(p) process is

$$S_{xx}(f) = \frac{\sigma_p^2}{\left|1 + \sum_{k=1}^{p} a_k e^{-2\pi jfk}\right|^2}. \tag{10}$$

For Rayleigh multi-path channels, the theoretical PSD associated with either in-phase or quadrature part of a fading signal has an U-shaped band-limited form

$$S(f) = \begin{cases} \frac{1}{\pi f_d \sqrt{1-\left(\frac{f}{f_d}\right)^2}}, & |f| \leqslant f_d \\ 0, & f > f_d \end{cases} \tag{11}$$

where f_d is the maximum Doppler shift in Hertz. For the purpose of digital processing, the discrete-time autocorrelation function is provided by

$$R[n] = J_0(2\pi f_m |n|), \tag{12}$$

where $f_m{=}f_d T_s$ indicates normalized f_d by the signal sampling rate $f_s{=}1/T_s$. Theoretically, an arbitrary spectrum can be closely approximated by an AR process with sufficiently large order. The relationship between a desired $R[n]$ and AR coefficients can be given in matrix form by

$$\mathbf{v} = \mathbf{R}\mathbf{a}, \tag{13}$$

where

$$\mathbf{R} = \begin{bmatrix} R[0] & R[-1] & \cdots & R[-p+1] \\ R[1] & R[0] & \cdots & R[-p+2] \\ \vdots & \vdots & \ddots & \vdots \\ R[p-1] & R[p-2] & \cdots & R[0] \end{bmatrix}, \tag{14}$$

$$\mathbf{a} = \begin{bmatrix} a_1 & a_2 & \cdots & a_p \end{bmatrix}^T, \tag{15}$$

$$\mathbf{v} = \begin{bmatrix} R[1] & R[2] & \cdots & R[p] \end{bmatrix}^T, \tag{16}$$

and

$$\sigma_p^2 = R[0] + \Sigma_{k=1}^{p} a_k R[k]. \tag{17}$$

Substituting the value of f_m into (12), the desired autocorrelation sequence $R[0], \ldots, R[p]$ are obtained [18]. Thus, $a_1, a_2, \ldots, a_p$ can be computed by solving the set of p Yule-Walker equations in (13) through calculating the inverse matrix $\mathbf{R}^{-1}$, which requires p^3 times complex multiplications, marked by $\mathcal{O}(p^3)$. The AR predictor for narrow-band single-input single-output (SISO) channels is built:

$$\hat{h}[t+1] = \sum_{k=1}^{p} a_k h[t-k+1], \tag{18}$$

where one-step ahead prediction can be obtained, in comparison with the continuous-time prediction provided by (8). By processing a MIMO channel as a set of parallel SISO channels, ignoring the spatial dimension of arrays, (18) is able to extend to MIMO channels:

$$\hat{\mathbf{H}}[t+1] = \sum_{k=1}^{p} \mathbf{A}_k \odot \mathbf{H}[t-k+1]. \tag{19}$$

where $\odot$ denotes the Hadamard product that is the element-wise multiplication of two matrices with same dimension, and the coefficient matrix $\mathbf{A}_k$ is defined as

$$\mathbf{A}_k = \begin{bmatrix} a_{11}^k & a_{12}^k & \cdots & a_{1N_t}^k \\ a_{21}^k & a_{22}^k & \cdots & a_{2N_t}^k \\ \vdots & \vdots & \ddots & \vdots \\ a_{N_r1}^k & a_{N_r2}^k & \cdots & a_{N_rN_t}^k \end{bmatrix}, \tag{20}$$

and the entry $a_{n_r n_t}^k$ represents the k^{th} coefficient of the AR filter for the subchannel between transmit antenna n_t and receive antenna n_r.

In comparison with the parametric model, the number of parameters required to estimate in the AR model is substantially reduced to only one, i.e., f_d. The estimation process from (12) to (16) is notably simpler than that of the parametric model. Despite its simplicity, the AR model has some limitations: 1) It can only deal with the temporal correlation of a single-antenna channel, whereas the spatial dimension of antenna array is not able to exploit. 2) The predictor

in (18) and (19) only enable one-step prediction $\hat{\mathbf{H}}[t+1]$ rather than multi-step $\hat{\mathbf{H}}[t+D]$. Although long-range prediction can be realized by recursively reusing predicted values at previous time instants, the problem of error propagation is raised. 3) Until now, reliable and accurate estimation for f_d is still difficult from the practical perspective.

The authors of [44] provided a performance bound for prediction of MIMO channels and revealed that the AR model is still significantly far away from that bound, indicating that other more efficient prediction schemes probably exist and a further exploration is of theoretical and practical interest, which will be the aim of RNN-based prediction introduced in the next section.

III. RECURRENT NEURAL NETWORK

Recurrent neural network is a class of machine learning that has shown great potential in the field of time-series prediction [28]. Unlike a feed-forward network that only learns from training data, a recurrent neural network can also use its memory of past states to process sequences of inputs. RNN has several variants, among which a Jordan network is currently used to build a channel predictor, as illustrated in Fig. 1. Basically, a simple network consists of three layers: an input layer with N_i neurons, a hidden layer with N_h neurons, and a layer having N_o outputs. Each connection between the activation of a neuron in the predecessor layer and the input of a neuron in the successor layer is assigned a weight. Let w_{ln} denote the weight connecting the n^{th} input and the l^{th} hidden neuron, while v_{ol} is the weight for hidden neuron l and output o, where $1 \leq n \leq N_i$, $1 \leq l \leq N_h$, and $1 \leq o \leq N_o$. Constructing a $N_h \times N_i$ weight matrix $\mathbf{W}$ as

$$\mathbf{W} = \begin{bmatrix} w_{11} & \cdots & w_{1N_i} \\ \vdots & \ddots & \vdots \\ w_{N_h 1} & \cdots & w_{N_h N_i} \end{bmatrix}, \tag{21}$$

and denoting the activation vector of the input layer and the recurrent component (feedback) at time step t as $\mathbf{x}(t) = \left[x_1(t), ..., x_{N_i}(t)\right]^T$ and $\mathbf{f}(t) = \left[f_1(t), ..., f_{N_h}(t)\right]^T$, respectively, the input for the hidden layer is expressed in matrix form by

$$\mathbf{z}_h(t) = \mathbf{W}\mathbf{x}(t) + \mathbf{f}(t) + \mathbf{b}_h, \tag{22}$$

where $\mathbf{b}_h = \left[b_1^h, ..., b_{N_h}^h\right]^T$ denotes the vector of biases in the hidden layer. Using a matrix $\mathbf{F}$ to represent the mapping from the output at the previous time step, i.e., $\mathbf{y}(t-1) = \left[y_1(t-1), ..., y_{N_o}(t-1)\right]^T$, to the recurrent component, we have

$$\mathbf{f}(t) = \mathbf{F}\mathbf{y}(t-1). \tag{23}$$

The behaviour of a neural network depends on activation functions, typically falling into the following categories: linear, rectified linear, threshold, sigmoid, and tangent. In general, a sigmoid function is employed to deal with nonlinearity,

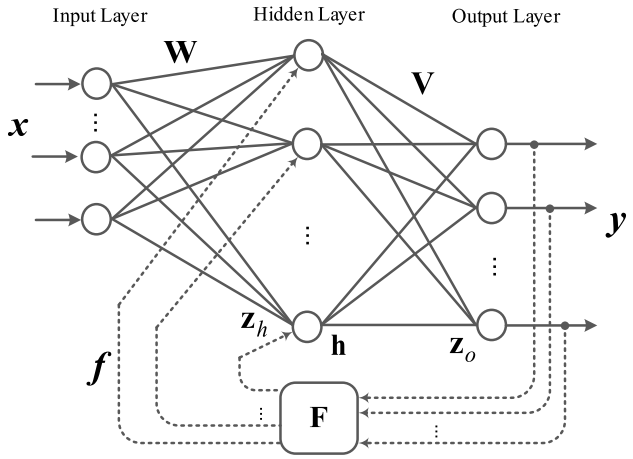


FIGURE 1. Typical structure of a recurrent neural network used for channel prediction.

which is defined as

$$S(x) = \frac{1}{1 + e^{-x}}. \tag{24}$$

Substituting (22) and (23) into (24), the activation vector of the hidden layer is thus

$$\mathbf{h}(t) = S\left(\mathbf{z}_h(t)\right) = S\left(\mathbf{W}\mathbf{x}(t) + \mathbf{F}\mathbf{y}(t-1) + \mathbf{b}_h\right), \tag{25}$$

where $S(\mathbf{z}_h)$ means an element-wise operation for simplicity, i.e., $S(\mathbf{z}_h) = \left[S(z_1), ..., S(z_{N_h})\right]^T$. In analogous to (21), another weight matrix $\mathbf{V}$ having a dimension of $N_o \times N_h$ with entries $\{v_{ol}\}$ is introduced. Then, the input for the output layer is $\mathbf{z}_o(t) = \mathbf{V}\mathbf{h}(t) + \mathbf{b}_y$, where $\mathbf{b}_y$ is the vector of biases in the output layer, resulting in an output vector:

$$\mathbf{y}(t) = S\left(\mathbf{z}_o(t)\right) = S\left(\mathbf{V}\mathbf{h}(t) + \mathbf{b}_y\right). \tag{26}$$

Like other data-driven AI techniques, the operation of a RNN is categorized into two phases: training and predicting. The training of a neural network is typically based on a fast algorithm known as Back-Propagation (BP). Provided a training dataset, the network feeds forward input data and compares the resulting output $\mathbf{y}$ against the desired value $\mathbf{y}_0$. Measured by a cost function, e.g., $C = \| \mathbf{y}_0 - \mathbf{y} \|^2$, prediction errors are propagated back through the network, causing iteratively updating of weights and biases until a certain convergence condition reaches. To provide an initial impression of this process, the BP algorithm in combination with gradient descent learning for a feed forward network is briefly depicted:

Start from an initial network state where $\{\mathbf{W}, \mathbf{V}, \mathbf{b}_h, \mathbf{b}_y\}$ are randomly set.

1) Input a training example $(\mathbf{x}, \mathbf{y}_0)$.
2) **Feed-Forward**: For the hidden layer, its input $\mathbf{z}_h$ and activation $\mathbf{h}$ can be computed using (22)[1] and (25),

[1]For simplicity, the BP algorithm applied in a feed forward network without a recurrent component is illustrated. Hence, the exact equation to calculate the input is $\mathbf{z}_h(t) = \mathbf{W}\mathbf{x}(t) + \mathbf{b}_h$.

respectively. Also, $\mathbf{z}_o$ and $\mathbf{y}$ in (26) for the output layer are obtained.

3) Compute the output error $\mathbf{e}^y = \left[e_1^y, ..., e_{N_o}^y\right]^T$ as:

$$\mathbf{e}^y = \nabla_y C \odot S'(\mathbf{z}_o), \tag{27}$$

where ∇ notates a vector whose entries are partial derivatives, namely $\nabla_y C = \left[\frac{\partial C}{\partial y_1}, ..., \frac{\partial C}{\partial y_{N_o}}\right]^T$. In addition, $S'(\mathbf{z}_o)$ stands for the derivative of the activation function with respect to its corresponding input $\mathbf{z}_o$. Given a sigmod function in (24), for instance, we have $S'(\mathbf{z}_o) = \frac{\partial S(\mathbf{z}_o)}{\partial \mathbf{z}_o} = S(\mathbf{z}_o)(1 - S(\mathbf{z}_o))$.

4) **Back-Propagation**: $\mathbf{e}^y$ is propagated back to the hidden layer to derive the error vector there, i.e.,

$$\mathbf{e}^h = \mathbf{V}^T \mathbf{e}^y \odot S'(\mathbf{z}_h). \tag{28}$$

5) **Gradient Descent**: With the back-propagated errors, the weights and biases are able to be updated according to the following rules:

$$\begin{cases} \mathbf{W} = \mathbf{W} - \eta \mathbf{e}^h \mathbf{x}^T \\ \mathbf{V} = \mathbf{V} - \eta \mathbf{e}^y \mathbf{h}^T \\ \mathbf{b}_h = \mathbf{b}_h - \eta \mathbf{e}^h \\ \mathbf{b}_y = \mathbf{b}_y - \eta \mathbf{e}^y, \end{cases} \tag{29}$$

where η stands for the learning rate.

The weights and biases are iteratively updated until the cost function is below a predefined threshold or the number of epochs reaches its maximal value. Once the training process completed, the trained network can be used to process upcoming samples. The training of a RNN typically use a variant of the BP algorithm, known as back-propagation through time (BPTT). It requires to unfold a recurrent neural network in time steps to form a pseudo feed-forward network, where the BP algorithm is applicable. Upon this, other more advanced or efficient approaches such as real-time recurrent learning and extended Kalman filtering have been designed.

IV. RNN-BASED CHANNEL PREDICTION

Observing the MIMO channel model and the structure of a neural network, high similarity that both have multiple inputs and outputs with fully weighted connections can be found. A neural network well suits to process MIMO channels by adapting the number of input and output neurons with respect to the number of transmit and receive antennas. A RNN predictor is quite flexibly to be configured to forecast channel response or envelope on demand in either frequency-flat or frequency-selective fading channels. In this section, the discussion starts from the simplest case that applies a RNN to predict a flat fading channel in a SISO system, then extends step by step until a frequency-domain predictor for frequency-selective MIMO channels.

A. FLAT FADING CHANNEL PREDICTION

1) CHANNEL GAIN PREDICTION BY A COMPLEX-VALUED RNN

To begin with, consider a discrete-time baseband equivalent model for a flat fading SISO channel:

$$r[t] = h[t]s[t] + n[t]. \tag{30}$$

The aim of RNN predictor is to get a predicted value $\hat{h}[t+\tau]$ that is as close as possible to its actual value $h[t+\tau]$. To deal with complex-valued channel gains, a network with complex-valued weights called a complex-valued RNN hereinafter is needed [29], [30]. At time t, $h[t]$ is obtained through channel estimation, while a series of d past values $h[t-1], h[t-2], ..., h[t-d]$ can be memorized simply through a tapped delay line. These $d+1$ channel gains are fed into the RNN as the input, i.e.,

$$\mathbf{x}[t] = [h[t], h[t-1], \ldots, h[t-d]]^T. \tag{31}$$

In together with the delayed feedback, the prediction of a future channel gain $\mathbf{y}[t] = \left[\hat{h}[t+1]\right]^T$ is obtained.

The extension of this predictor to flat fading MIMO channels is straightforward. To adapt to the input layer of a RNN, channel matrices are required to be vectorized into a $N_r N_t \times 1$ vector, as follows:

$$\mathbf{h}[t] = \vec{\mathbf{H}}[t] = \left[h_{11}[t], h_{12}[t], ..., h_{N_r N_t}[t]\right]. \tag{32}$$

Together with a number of d past values $\mathbf{H}[t-1], \ldots, \mathbf{H}[t-d]$, the input of RNN this case is

$$\mathbf{x}[t] = [\mathbf{h}[t], \mathbf{h}[t-1], \ldots, \mathbf{h}[t-d]]^T, \tag{33}$$

resulting in a predictive value $\mathbf{y}[t] = \hat{\mathbf{h}}^T[t+1]$, which can be transformed to a predicted matrix $\hat{\mathbf{H}}[t+1]$.

2) CHANNEL GAIN PREDICTION BY A REAL-VALUED RNN

In comparison with a complex-valued RNN, a recurrent neural network with real-valued weights called a real-valued RNN has lower complexity and higher prediction accuracy, whereas it can only deal with real-valued data. Fortunately, a complex-valued channel gain can be decomposed into two real values, namely $h = h^r + jh^i$. Hence, a real-valued RNN was proposed in [31] to build a simpler predictor with higher accuracy by means of decoupling the real and imaginary parts. Without a necessity of using two RNNs, the real and imaginary parts can be processed jointly in a single predictor. In this case, the input of the network is

$$\mathbf{x}[t] = \left[h^r[t], h^i[t], \ldots, h^r[t-d], h^i[t-d]\right]^T, \tag{34}$$

generating an output $\mathbf{y}[t] = \left[\hat{h}^r[t+1], \hat{h}^i[t+1]\right]^T$ that synthesizes to a predicted channel gain $\hat{h}[t+1] = \hat{h}^r[t+1] + j\hat{h}^i[t+1]$. Similarly, $\mathbf{H}[t]$ is decomposed into

$$\mathbf{H}[t] = \mathbf{H}_R[t] + j\mathbf{H}_I[t], \tag{35}$$

where $\mathbf{H}_R = \Re(\mathbf{H}) = [h^r_{n_r n_t}]_{N_r \times N_t}$ denotes a matrix composed by the real parts of channel gains and

$\mathbf{H}_I = \Im(\mathbf{H}) = [h^i_{n_r n_t}]_{N_r \times N_t}$ is the imaginary counterpart. Like (32), these matrices are vectorized

$$\mathbf{h}_r[t] = \vec{\mathbf{H}}_R[t] = \left[h^r_{11}[t], h^r_{12}[t], ..., h^r_{N_r N_t}[t]\right]. \quad (36)$$

Feeding

$$\mathbf{x}[t] = [\mathbf{h}_r[t], \mathbf{h}_i[t], ..., \mathbf{h}_r[t-d], \mathbf{h}_i[t-d]]^T \quad (37)$$

into the network, the resulting output is written as $\mathbf{y}[t] = \left[\hat{\mathbf{h}}_r[t+1], \hat{\mathbf{h}}_i[t+1]\right]^T$, which can be transformed into $\hat{\mathbf{H}}_R[t+D]$ and $\hat{\mathbf{H}}_I[t+D]$. Then, a predicted matrix is reaped simply by $\hat{\mathbf{H}}[t+1] = \hat{\mathbf{H}}_R[t+1] + j\hat{\mathbf{H}}_I[t+1]$.

3) CHANNEL ENVELOPE PREDICTION

Many adaptive transmission systems only need to know the envelope of channel response $|h|$, rather than a complex-valued gain h. Therefore, a real-valued RNN can be directly applied, which in turn can lower complexity, speed up training process, and improve prediction accuracy, in comparison with predicting channel gains. The channel envelope at time t denoted by $|h[t]|$ is known, with a number of d past values $|h[t-1]|, |h[t-2]|, \ldots, |h[t-d]|$, the input in this case is written as

$$\mathbf{x}[t] = [|h[t]|, |h[t-1]|, \ldots, |h[t-d]|]^T, \quad (38)$$

which generate $|\hat{h}[t+1]|$ through the network. Further, let $\mathbf{Q}[t] = \left[|h_{n_r n_t}[t]|\right]_{N_r \times N_t}$ denotes a matrix, in which the $(n_r, n_t)^{th}$ entry is the envelope of $h_{n_r n_t}[t]$ in $\mathbf{H}[t]$. Likewise, $\mathbf{Q}[t]$ is vectorized as

$$\mathbf{q}[t] = \vec{\mathbf{Q}}[t] = \left[|h_{11}[t]|, |h_{12}[t]|, ..., |h_{N_r N_t}[t]|\right]. \quad (39)$$

With the input $\mathbf{x}[t] = [\mathbf{q}[t], \mathbf{q}[t-1], ..., \mathbf{q}[t-d]]^T$, the prediction $\mathbf{y}[t] = \hat{\mathbf{q}}[t+1]$ is got and further transformed into $\hat{\mathbf{Q}}[t+1]$.

4) MULTI-STEP PREDICTION

By far the predictor is only tuned to forecast one-step ahead, namely $\hat{\mathbf{H}}[t+1]$, whereas such a prediction length is probably too short to satisfy the requirement of adaptive transmission systems. Hence, long-range prediction enabled by a multi-step predictor is of great interest. Making full use of the flexible structure of neural networks, the output at time step t can be tuned to $\hat{\mathbf{H}}[t+D]$, where D is an positive integer standing for the number of steps being predicted ahead. It returns back to the pervious one-step predictor if $D=1$. From the perspective of training, there is no intrinsic difference between one-step and multi-step prediction. The only required modification is that the desired value for calculating the prediction error in the training process is shifted from $\mathbf{H}[t+1]$ to $\mathbf{H}[t+D]$, resulting in different weights and biases.

B. FREQUENCY-SELECTIVE MIMO PREDICTION

To begin with, let us consider the discrete-time model for a frequency-selective SISO system:

$$r[t] = \sum_{l=0}^{L-1} h_l[t] s[t-l] + n[t], \quad (40)$$

where s and r denote the transmitted and received symbol, respectively, $h_l[t]$ stands for the l^{th} tap for a time-varying channel filter at time t, and n is additive noise. Dropped time index for simplicity, a frequency-selective channel is modeled as a linear L-tap filter

$$\mathbf{h} = [h_0, h_1, \ldots, h_{L-1}]^T. \quad (41)$$

It can be converted into a set of N orthogonal narrow-band channels known as sub-carriers through the OFDM modulation [45], which is represented by

$$\tilde{r}_n[t] = \tilde{h}_n[t]\tilde{s}_n[t] + \tilde{n}_n[t], \quad n = 0, 1, \ldots, N-1, \quad (42)$$

where $\tilde{s}_n[t]$, $\tilde{r}_n[t]$, and $\tilde{n}_n[t]$ stand for the transmitted signal, received signal, and noise, respectively, at sub-carrier n. According to the picket fence effect in discrete Fourier transform (DFT) [46], the frequency response of the channel filter denoted by $\tilde{\mathbf{h}} = [\tilde{h}_0, \tilde{h}_1, \ldots, \tilde{h}_{N-1}]^T$ is the DFT of $\mathbf{h}' = [h_0, h_1, \ldots, h_{L-1}, 0, \ldots, 0]^T$ that pads $\mathbf{h}$ in (41) with $N-L$ zeros at the tail.

Extending (42) to a multi-antenna system is straightforward though a MIMO-OFDM system that is modeled as

$$\tilde{\mathbf{r}}_n[t] = \tilde{\mathbf{H}}_n[t]\tilde{\mathbf{s}}_n[t] + \tilde{\mathbf{n}}_n[t], \quad n = 0, 1, \ldots, N-1, \quad (43)$$

where $\tilde{\mathbf{s}}_n[t]$ represents $N_t \times 1$ transmit symbol vector on sub-carrier n at time t, $\tilde{\mathbf{r}}_n[t]$ is $N_r \times 1$ received symbol vector, and $\tilde{\mathbf{n}}[t]$ is the vector of additive noise. The subchannel between transmit antenna n_t and receive antenna n_r is equivalent to a frequency-selective SISO channel, denoted by a channel filter $\mathbf{h}^{n_r n_t} = [h_0^{n_r n_t}, h_1^{n_r n_t}, \ldots, h_{L-1}^{n_r n_t}]^T$. Likewise, the frequency response of this filter can be obtained by by conducting DFT, that is $\tilde{\mathbf{h}}^{n_r n_t} = [\tilde{h}_0^{n_r n_t}, \tilde{h}_1^{n_r n_t}, \ldots, \tilde{h}_{N-1}^{n_r n_t}]^T$. Then, the channel matrix on sub-carrier n can be notated as $\tilde{\mathbf{H}}_n[t] = \left[\tilde{h}_n^{n_r n_t}[t]\right]_{N_r \times N_t}$.

Fig. 2 illustrates the schematics of a RNN predictor for frequency-selective fading MIMO channels [34]. The main idea is to convert a frequency-selective channel into a set of orthogonal flat fading sub-carriers, and then utilize a frequency-domain predictor to forecast the frequency response on each sub-carrier. At time t over sub-carrier n, as shown in Fig. 2, $\tilde{\mathbf{H}}_n[t]$, as well as its d-step delays $\tilde{\mathbf{H}}_n[t-1], ..., \tilde{\mathbf{H}}_n[t-d]$, are fed into the RNN. A Matrix-to-Vector (M2V) module vectorizes these matrices, dropped the time index for simplicity, following

$$\tilde{\mathbf{h}}_n = \mathrm{vec}\left(\tilde{\mathbf{H}}_n\right) = [\tilde{h}_n^{11}, \tilde{h}_n^{12}, ..., \tilde{h}_n^{N_r N_t}]. \quad (44)$$

The RNN outputs a D-step prediction, i.e., $\hat{\mathbf{h}}_n[t+D] = [\hat{h}_n^{11}[t+D], ..., \hat{h}_n^{N_r N_t}[t+D]]^T$, transforming into a predicted matrix $\hat{\mathbf{H}}_n[t+D]$ via a Vector-to-Matrix (V2M) module.

Although the prediction is conducted at sub-carrier level, we do not need to deal with all N sub-carriers taking into account channel's frequency correlation. Integrated with a pilot-assisted system, only predicting the CSI on sub-carriers carrying pilot symbols is enough. Suppose one pilot is inserted uniformly every N_P sub-carriers, amounts to a total

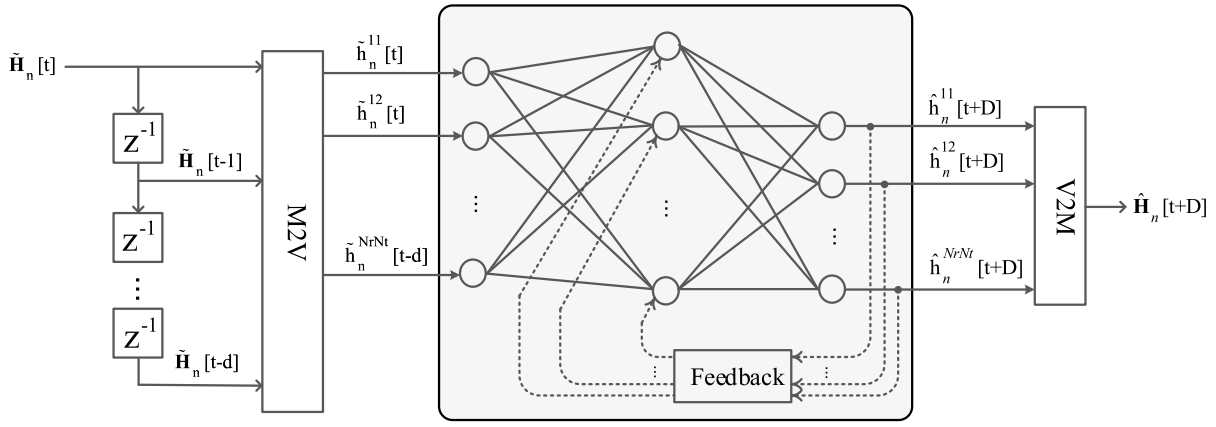


FIGURE 2. Schematics of the RNN-based predictor for frequency-selective fading MIMO channels.

of $P=\lceil \frac{N}{N_P} \rceil$ pilot sub-carriers, where $\lceil \cdot \rceil$ denotes a ceiling function. Given predicted values $\hat{\mathbf{H}}_p[t+D]$, $p=1, \ldots, P$, frequency-domain interpolation can be applied to get the predicted values on all sub-carriers $\hat{\mathbf{H}}_n[t+D]$, $n=0, \ldots, N-1$.

V. PREDICTION-AIDED MIMO-OFDM

Transmit antenna selection in a multi-antenna system has been widely recognized as a cost-effective approach due to the reduction of required number of high-power amplifiers in radio frequency chains. To select out appropriate antennas, instantaneous CSI at the transmitter is mandatory, where channel prediction can play an important role. In order to further shed light on the RNN predictor, prediction-aided TAS in a MIMO-OFDM system with N_t transmit and N_r receive antennas is presented as an application example.

As illustrated in Fig. 3, the fast Fourier transform (FFT) demodulator and inverse FFT (IFFT) modulator, with the aid of cyclic prefix (CP), convert a frequency-selective channel into N sub-carriers, where a payload of M data symbols denoted by $\mathbf{d}=[d_1, d_2, \ldots, d_M]^T$ is carried. Without the consideration of null sub-carriers reserved for out-of-band radiation suppression [47] and direct current, we can assume that the remaining $P=N-M$ sub-carriers are used for comb-type pilot symbols $\mathbf{p}=[p_1, p_2, \ldots, p_P]^T$, inserting uniformly every N_P sub-carriers. At time t, $\mathbf{d}^{(t)}$ and $\mathbf{p}^{(t)}$ are multiplexed and transmitted in one OFDM symbol. The receiver obtains $\tilde{\mathbf{H}}_p[t]$ through estimating $\mathbf{p}^{(t)}$. Taking advantage of frequency correlation, an interpolator can recover channel responses across the whole bandwidth including all data and pilot sub-carriers, i.e., $\tilde{\mathbf{H}}_n[t]$, $n=0, 1, \ldots, N-1$.

The traditional TAS system directly uses $\tilde{\mathbf{H}}_n[t]$ to make decisions, as marked by the line with a cross originated from the channel estimator in the figure. There are two kinds of selection strategies, i.e., bulk or per-tone, as analyzed in [48]. Without loss of generality, we adopt the latter in this article for simplicity, i.e., each sub-carrier chooses the best antenna individually instead of the same selection for all sub-carriers.

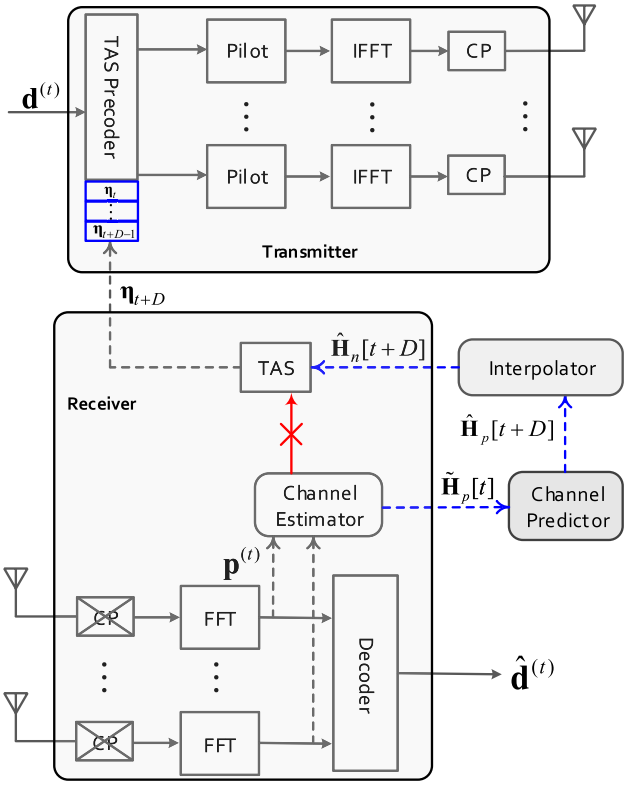


FIGURE 3. Illustration of prediction-aided TAS in a MIMO-OFDM system.

Mathematically, the traditional TAS system follows

$$\eta_n[t] = \arg\max_{1 \leqslant n_t \leqslant N_t} \left\| \tilde{\mathbf{h}}_n^{n_t}[t] \right\|, \tag{45}$$

where $\eta_n[t]$ stands for the index of the selected antenna at time t upon sub-carrier n, $\tilde{\mathbf{h}}_n^{n_t}[t]$ is the n_t^{th} column vector of $\tilde{\mathbf{H}}_n[t]$, and $\| \cdot \|$ denotes the Euclidean norm of a vector. The receiver feeds a set of selected antenna indices for all data sub-carriers $\boldsymbol{\eta}^t=\{\eta_n[t] \mid 0 \leqslant n \leqslant N-1, n \neq p\}$ back to the transmitter through a feedback channel. Assuming that the processing and feedback delays can be absorbed by the time

TABLE 1. Simulation parameters.

Parameters	Values
Sampling rate	$f_s = 1$MHz
Max. Doppler shifts	$f_d = 70$ or 300Hz
MIMO	4×1 ULA
Channel models	3GPP ETU and EVA [49]
Correlation coeff.	$\alpha = 0$(low), 0.3(medium), 0.9(high)
Correlation matrix	$\begin{bmatrix} 1 & \alpha^{\frac{1}{9}} & \alpha^{\frac{4}{9}} & \alpha \\ (\alpha^{\frac{1}{9}})^* & 1 & \alpha^{\frac{1}{9}} & \alpha^{\frac{4}{9}} \\ (\alpha^{\frac{4}{9}})^* & (\alpha^{\frac{1}{9}})^* & 1 & \alpha^{\frac{1}{9}} \\ \alpha^* & (\alpha^{\frac{4}{9}})^* & (\alpha^{\frac{1}{9}})^* & 1 \end{bmatrix}$
DFT size	$N = 64$
Neural Network	3-layer RNN
Training	Levenberg-Marquardt algorithm [50], Cross-validation, 15% validation data Max. number of epoches = 1000 Loss function = MSE Dynamic learning rate, initialed with 0.1
Tapped delay	$d = 3$
Hidden neurons	$N_H = 10$
Prediction Length	1.0ms in EVA70 0.25ms in ETU300

gap of D OFDM symbols, the transmit antennas selected in terms of $\boldsymbol{\eta}^t$ are applied for transmitting signals at time $t+D$.

Due to the channel fading, $\tilde{\mathbf{H}}_n[t]$ is outdated and may differ substantially from the actual CSI $\tilde{\mathbf{H}}_n[t+D]$ at the instant of signal transmission, leading to notable performance degradation [7]. Due mainly to additive noise and delay, $\tilde{\mathbf{H}}_n[t+D]$ is impossible to obtain in practical systems. But if a selection decision can be made according to predicted CSI $\hat{\mathbf{H}}_n[t+D]$ that approximates to $\tilde{\mathbf{H}}_n[t+D]$, the performance can be improved. At time t, as depicted in Fig. 3, $\tilde{\mathbf{H}}_p[t]$ is fed into the predictor to get $\hat{\mathbf{H}}_p[t+D]$. A frequency-domain channel interpolator is applied to recover the CSI across all sub-carriers $\hat{\mathbf{H}}_n[t+D]$ so as to replace the outdated CSI $\tilde{\mathbf{H}}_n[t]$ in the traditional TAS system. Then, the best antenna for time $t+D$ upon sub-carrier n is selected in advance as

$$\hat{\eta}_n[t+D] = \arg\max_{1 \leqslant n_t \leqslant N_t} \left\| \hat{\mathbf{h}}_n^{n_t}[t+D] \right\|, \qquad (46)$$

where $\hat{\mathbf{h}}_n^{n_t}[t+D]$ is the n_t^{th} column vector of $\hat{\mathbf{H}}_n[t+D]$. The index vector $\boldsymbol{\eta}^{t+D}$ is fed back to and is buffered at the transmitter before its actual transmission time $t+D$.

VI. PERFORMANCE AND COMPLEXITY

Monte-Carlo simulations are carried out to comparatively evaluate the performance of the predictors. In this section, some representative numerical results in terms of outage probabilities of prediction-aided TAS in a MIMO-OFDM system and the prediction accuracy measured by Mean Squared Error (MSE) are presented, together with the comparison on computational complexity.

A. PERFORMANCE

In a 4×1 ULA system, 3GPP Extended Pedestrian A (EVA) and Extended Typical Urban (ETU) models with maximal Doppler shifts of f_d=70Hz and 300Hz, respectively, also notated as EVA70 and ETU300, are applied to generate channel samples. Using a signal bandwidth (or sampling rate) of 1MHz, a frequency-selective channel is converted into N=64 sub-carriers via the OFDM modulation, resulting in a sub-carrier spacing around $\triangle f$=15KHz that is compliant with 3GPP LTE standard. From the observations in simulations, a 3-layer RNN with N_H=10 hidden neurons and d=3 tapped delay is adopted. To train this network, a data set consisting of a series of channel samples during a length equivalent to 10 channel's coherence time is extracted. Starting from an initial state with random values, the weights and biases are iteratively updated by Levenberg-Marquardt [50] algorithm. The train can be conducted in an off-line manner or rely on other computation-intensive nodes, such as mobile edge computing platform, and then the trained RNN is deployed online for predicting instantaneous channels. In contrast, the AR model does not need a training process. Given the value of f_d, the filter coefficients in (19) can be figured out. The simulation parameters are summerized in Table 1.

The outage probability defined as

$$\mathrm{P(R)}=\Pr\{\log_2(1+\mathrm{SNR})<\mathrm{R}\}, \qquad (47)$$

where Pr is the notation of mathematical probability and R means a target end-to-end data rate that is set to 1bps/Hz as usual, is employed to measure the performance of the prediction-aided MIMO-OFDM system. The following four different TAS strategies are compared:

- The *perfect* mode where the best antenna for sub-carrier n at time $t+D$ is chosen according to the perfect CSI $\tilde{\mathbf{H}}_n[t+D]$, despite it never exists in practice owing to delay and noise.
- The *outdated* mode in a traditional TAS system, making a selection decision based on the outdated CSI $\tilde{\mathbf{H}}_n[t]$.
- The *prediction* mode takes advantage of the predicted CSI $\hat{\mathbf{H}}_n[t+D]$ that may closely approximate $\tilde{\mathbf{H}}_n[t+D]$.
- The *random* mode that randomly selects an antenna without any consideration of CSI.

The assessment is first carried out in independent and identically distributed (*i.i.d.*) noiseless EVA70 channels, where channel samples used to train the predictors are obtained without the impose of additive noise. The predictors are set to a multi-step mode of D=16, corresponding to a prediction length of around 1ms. Fig. 4a compares outage probabilities for the various selection strategies in the MIMO-OFDM system. The outdated CSI substantially degrades the system performance, with a relative SNR loss of around 5.5dB at $\mathrm{P(R)}=10^{-3}$ with respect to the perfect mode. The AR model with p=1 denoted by $AR(1)$ achieves the optimal performance as same as that of the perfect mode, followed by the RNN that have a prediction gain of over 3dB compared with the outdated mode. Although $AR(1)$ outperforms the

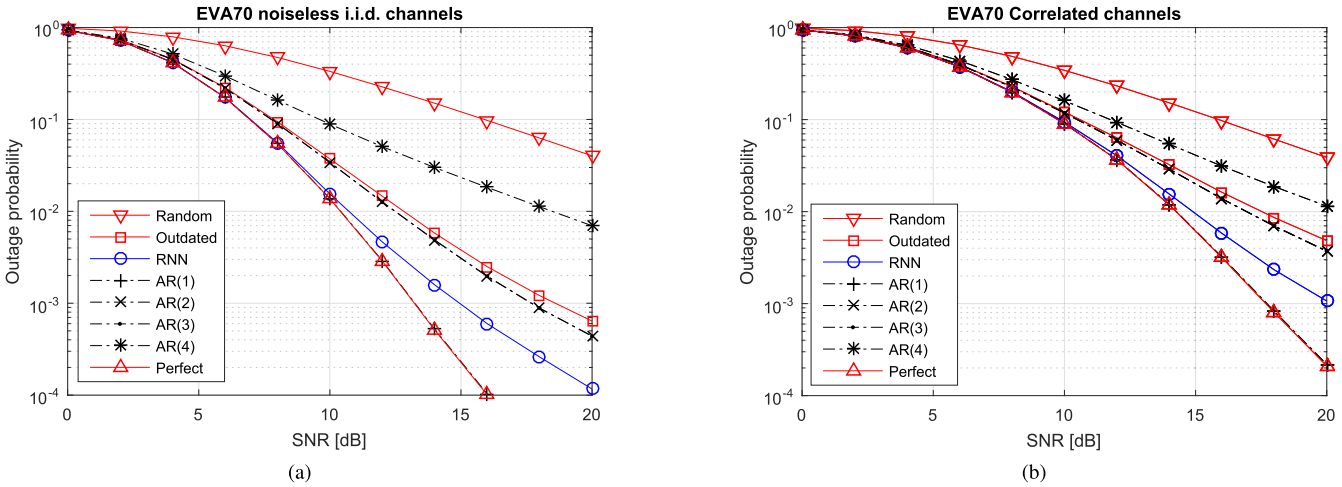


FIGURE 4. Performance comparisons in EVA70 (a) i.i.d. and (b) correlated channels.

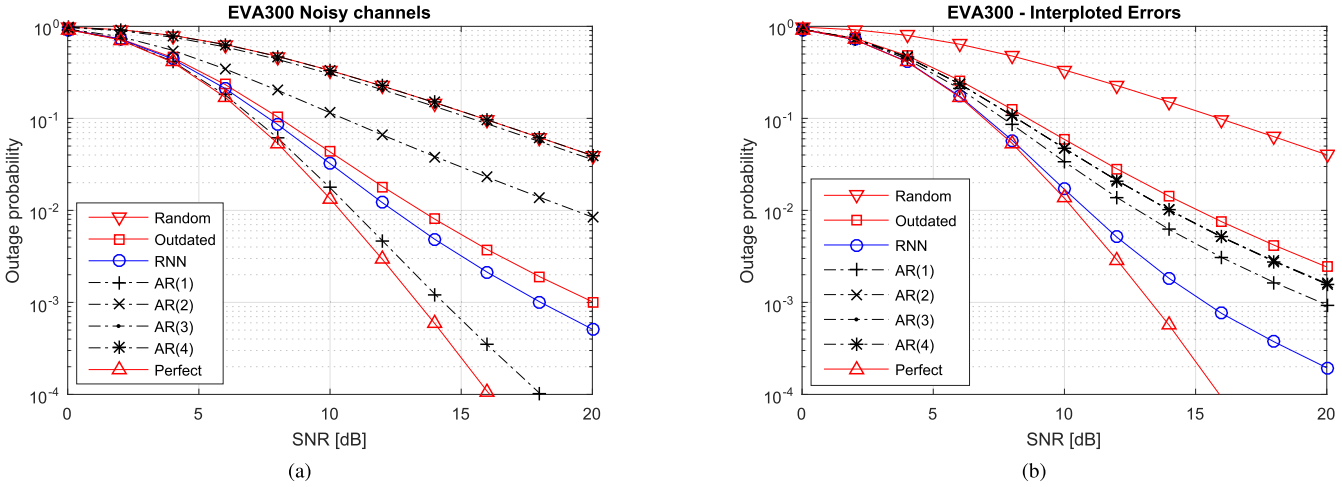


FIGURE 5. Performance comparisons in ETU300 channels taking into account the effects of (a) additive noise and (b) interpolated errors.

RNN, the performance of other AR predictors are quite differentiated. $AR(2)$ and $AR(3)$ have identical performance that is comparable to that of the outdated mode, while the $AR(4)$ got a very worse result, with an SNR loss on the order of 10dB. To check the effect of channel correlation, we apply the matrix (see Table 1) recommended in 3GPP LTE standard [49] to generate correlated channel samples. Under the medium correlation indicated by α=0.3, the curves of outage probabilities for various selection strategies are comparatively drawn in Fig. 4b. In contrast to the results in *i.i.d.* channels illustrated in Fig. 4a, the system performance degrades collectively. That is because the available spatial diversity gain vanishes gradually with the increase of channel correlation, independently of the application of predictors. Actually, the correlation has no impact on the relativity of superiority and inferiority among the predictors.

In practice, the obtained CSI is impaired by estimation errors because additive noise cannot be avoided. Under the working assumption that the signal-to-noise ratio (SNR) on pilot symbols is 20dB, the simulations in noisy ETU300 channels are also conducted. To adapt to faster channel fluctuation, the number of prediction steps is reduced to D=4, corresponding to a prediction length of 0.25ms. The results reveal that noise has a notable impact on the system performance, especially when the predictor is based on a high order AR model. As shown in Fig. 5a, the curve of $AR(4)$ is overlapped with that of the random mode, while $AR(3)$ also approximates them. That is to say, the AR model suffers severely from the problem of error propagation. In contrast, the RNN performs in a more stable manner and is more robust against noise compared to $AR(2)$ and $AR(3)$, in comparison with their behaviours in noiseless channels in Fig. 4a. It still

TABLE 2. Comparisons on MSE (dB).

	RNN	AR(1)	AR(2)	AR(3)	AR(4)
Noiseless	-44.8	-40.2	-30.7	-30.6	-30.3
Correlated	-39.2	-40.5	-31.0	-31.0	-30.7
Noisy	-25.7	-25.8	-9.8	3.5	15.3
Interpolated	-40.1	-18.6	-18.4	-18.4	-18.4

receives a prediction gain of over 2dB at outage probabilities on the order of 10^{-3} compared with the outdated mode. In addition to estimation error due to additive noise, the interpolation error that is defined as the difference between the perfect CSI and the interpolated CSI should be taken into account. Without loss of generality, the performance results for the MIMO-OFDM system having a pilot insertion interval of N_P=4 are illustrated in Fig. 5b. The interpolation has a negligible impact on the performance of the RNN predictor, while the AR predictors are substantially affected. The RNN remarkably outperforms the AR predictors with an SNR gain of at least 5dB at P(R)=10^{-3} and has a loss of nearly 2dB compared with the perfect mode.

In addition to outage probability, the prediction accuracy is also compared in terms of MSE, which is defined as

$$MSE = \frac{1}{N}\sum_{n=1}^{N}\left\|\hat{\mathbf{H}}[n]-\mathbf{H}[n]\right\|^2, \tag{48}$$

where N is the total number of data samples used for evaluation, $\hat{\mathbf{H}}[n]$ denotes the predicted channel matrix at time step n, and $\mathbf{H}[n]$ stands for its actual value. The MSE results of the RNN and the AR models with a filter length of $p = 1, 2, 3,$ and 4 in EVA70 channels are obtained, as listed in Table 2. The RNN has the best accuracy in *noiseless* and *interpolated* cases, while also achieving sub-optimal accuracy that very approaches the best values in *noisy* and *correlated* cases. Fig. 6 visualizes the MSEs of the RNN, and selects the best (indicated by *AR-Min*) and the poorest (*AR-Max*) results among AR models per case. For a clear illustration, the decibel values are used in the table and the figure, calculating by MSE_{dB}=$10\log_{10}(MSE)$.

In a nutshell, it can be concluded that the RNN predictor is effective to combat the outdated CSI in both independent and correlated channels, and shows strong robustness against additive noise and interpolation errors.

B. COMPUTATIONAL COMPLEXITY

Last but not least, the computational complexity of the predictors is assessed. As usual, the number of complex multiplications is used as a metric. There are two differentiated phases: training/parameters' estimation and predicting, thus the evaluation of their complexity are separated into two parts accordingly. From (21)-(26), we know that the hidden and output layer of a neural network need $(N_i$+$N_o)N_h$ and N_oN_h times complex multiplications per prediction, respectively, amounting to a total number of $N_h(N_i$+$2N_o)$. The number of

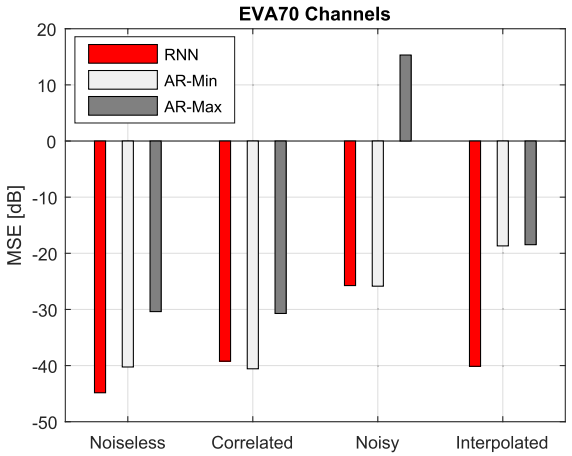


FIGURE 6. MSE comparisons between the RNN and selected AR models in noiseless, noisy, correlated, and interpolated EVA70 channels.

TABLE 3. Comparison on Computational Complexity.

Predictor	Training	Predicting
RNN	$\mathcal{O}(n_p n_s \kappa\mu)$	$\mathcal{O}(\kappa\mu)$
AR model	$\mathcal{O}(p^3\mu)$	$\mathcal{O}(p\mu)$
Parametric model	[38]	$\mathcal{O}(P\mu)$

input and output neurons is decided by the number of MIMO subchannels N_rN_t, we have N_i=$(d$+$1)N_rN_t$ and N_o=N_rN_t. Hence, each prediction needs $(d$+$3)N_hN_rN_t$ multiplications. Using μ=N_rN_t and κ=dN_h to denote the size of a MIMO system and the scale of a neural network, respectively. Then, the complexity of the RNN can be indicated by $\mathcal{O}(\kappa\mu)$. Looking at (19), we can know that the AR model requires pN_rN_t times multiplications for one prediction, marked as $\mathcal{O}(p\mu)$. Since a small filter order is generally enough and thus κ>p, the AR predictor is computationally simpler than the RNN predictor. Similarly, it can be derived from (8) that the complexity of the parametric model is $\mathcal{O}(P\mu)$, which is comparable with the AR model.

More complex part is the training or parameter estimation phase. During a training, the error back propagation through a neural network is quite similar to its feed forward process, corresponding to $\mathcal{O}(\kappa\mu)$. The total complexity is also related to the number of training samples n_s and the times of epoches n_p, i.e., $\mathcal{O}(n_pn_s\kappa\mu)$. To calculate p coefficients for a MIMO sub-channel, the AR model has to make p^3 times multiplications by solving the Yule-Walker equations, amounting to $\mathcal{O}(p^3\mu)$ for μ sub-channels in a MIMO system. The complexity of the parametric model is provided in [38]. Since a large number of propagation parameters need to be estimated periodically, the complexity of the parametric model is extremely high. In contrast, the weights for a neural network and the AR coefficients do not need to be updated frequently, which in turn drastically lower their complexity. The complexity of the predictors is summarized in Table 3.

VII. CONCLUSION

This paper provided a comprehensive overview on neural network-based predictors for multi-antenna channels. After a review on two statistical approaches - AR and parametric models - the structure of a recurrent neural network, its back-propagation training algorithm, and the principle of RNN predictors were introduced. Further, a prediction-aided MIMO-OFDM system that can improve the correctness of selecting transmit antennas was illustrated as an application example. Performance assessment in multi-path fading environment specified by 3GPP EVA and ETU channel models, taking into account the influential factors such as the Doppler shift, spatial correlation, additive noise, and interpolation error, was carried out. Numerical results justified the effectiveness of the RNN predictor to combat the outdated CSI, revealed its robustness against additive noise and interpolation errors, and shown its moderate computational complexity. This predictor is quite flexible to apply for both frequency-flat and frequency-selective MIMO fading channels in a wide variety of wireless adaptive transmission systems.

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, Jul. 2008.

[2] Q. Wang, L. J. Greenstein, L. J. Cimini, D. S. Chan, and A. Hedayat, "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, Jun. 2014.

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

[4] J. B. Kim, J. W. Choi, and J. M. Cioffi, "Cooperative distributed beamforming with outdated CSI and channel estimation errors," *IEEE Trans. Commun.*, vol. 62, no. 12, pp. 4269–4280, Dec. 2014.

[5] 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, Apr. 2015.

[6] E. N. Onggosanusi, A. Gatherer, A. G. Dabak, and S. Hosur, "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.

[7] X. Yu, W. Xu, S.-H. Leung, and J. Wang, "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, Jan. 2018.

[8] J. L. Vicario, A. Bel, J. Lopez-Salcedo, and G. Seco, "Opportunistic relay selection with outdated CSI: Outage probability and diversity analysis," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2872–2876, Jun. 2009.

[9] Z. Wang, L. Liu, X. Wang, and J. Zhang, "Resource allocation in OFDMA networks with imperfect channel state information," *IEEE Commun. Lett.*, vol. 18, no. 9, pp. 1611–1614, Sep. 2014.

[10] L. Su, C. Yang, and S. Han, "The value of channel prediction in CoMP systems with large backhaul latency," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4577–4590, Nov. 2013.

[11] Y. Teng, M. Liu, and M. Song, "Effect of outdated CSI on handover decisions in dense networks," *IEEE Commun. Lett.*, vol. 21, no. 10, pp. 2238–2241, Oct. 2017.

[12] A. Hyadi, Z. Rezki, and M.-S. Alouini, "An overview of physical layer security in wireless communication systems with CSIT uncertainty," *IEEE Access*, vol. 4, pp. 6121–6132, 2016.

[13] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[14] W. Jiang, T. Kaiser, and A. J. H. Vinck, "A robust opportunistic relaying strategy for co-operative wireless communications," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2642–2655, Apr. 2016.

[15] D. J. Love, R. W. Heath, Jr., V. K. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

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

[17] R. O. Adeogun, P. D. Teal, and P. A. Dmochowski, "Extrapolation of MIMO mobile-to-mobile wireless channels using parametric-model-based prediction," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4487–4498, Oct. 2015.

[18] K. E. Baddour and N. C. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1650–1662, Jul. 2005.

[19] W. Jiang and H. D. Schotten, "A comparison of wireless channel predictors: Artificial Intelligence versus Kalman filter," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, May 2019, pp. 1–6.

[20] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[21] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and doa estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.

[22] M. Liu, T. Song, J. Hu, J. Yang, and G. Gui, "Deep learning-inspired message passing algorithm for efficient resource allocation in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 641–653, Jan. 2019.

[23] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, Apr. 2019.

[24] K. P. Bagadi and S. Das, "Neural network-based adaptive multiuser detection schemes in SDMA–OFDM system for wireless application," *Neural Comput. Appl.*, vol. 23, nos. 3–4, pp. 1071–1082, Sep. 2013.

[25] C. V. R. Kumar and K. P. Bagadi, "MC–CDMA receiver design using recurrent neural networks for eliminating multiple access interference and nonlinear distortion," *Int. J. Commun. Syst.*, vol. 30, no. 16, May 2017, Art. no. e3328.

[26] H. Huang, W. Xia, J. Xiong, J. Yang, G. Zheng, and X. Zhu, "Unsupervised learning-based fast beamforming design for downlink MIMO," *IEEE Access*, vol. 7, pp. 7599–7605, 2018.

[27] W. Jiang, M. Strufe, and H. D. Schotten, "Experimental results for artificial intelligence-based self-organized 5G networks," in *Proc. IEEE PIMRC*, Montreal, QC, Canada, Oct. 2017, pp. 1–6.

[28] J. T. Connor, R. D. Martin, and L. E. Atlas, "Recurrent neural networks and robust time series prediction," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 240–254, Mar. 1994.

[29] W. Liu, L.-L. Yang, and L. Hanzo, "Recurrent neural network based narrowband channel prediction," in *Proc. IEEE 63rd Veh. Technol. Conf. (VTC)*, Melbourne, VIC, Australia, May 2006, pp. 2173–2177.

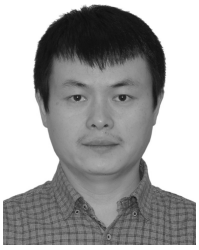
[30] T. Ding and A. Hirose, "Fading channel prediction based on combination of complex-valued neural networks and chirp Z-transform," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 9, pp. 1686–1695, Sep. 2014.

[31] W. Jiang and H. D. Schotten, "Multi-antenna fading channel prediction empowered by artificial intelligence," in *Proc. IEEE 88th Veh. Technol. Conf. (VTC-Fall)*, Chicago, IL, USA, Aug. 2018, pp. 1–6.

[32] W. Jiang and H. D. Schotten, "Neural network-based channel prediction and its performance in multi-antenna systems," in *Proc. IEEE 88th Veh. Technol. Conf. (VTC-Fall)*, Chicago, IL, USA, Aug. 2018, pp. 1–6.

[33] R.-F. Liao, H. Wen, J. Wu, H. Song, F. Pan, and L. Dong, "The Rayleigh fading channel prediction via deep learning," *Wireless Commun. Mobile Comput.*, vol. 2018, Jul. 2018, Art. no. 6497340.

[34] W. Jiang and H. D. Schotten, "Recurrent neural network-based frequency-domain channel prediction for wideband communications," in *Proc. IEEE 89th Veh. Technol. Conf. (VTC-Spring)*, Kuala Lumpur, Malaysia, Apr./May 2019, pp. 1–6.
[35] J. Wang, Y. Ding, S. Bian, Y. Peng, M. Liu, and G. Gui, "UL-CSI data driven deep learning for predicting DL-CSI in cellular FDD systems," *IEEE Access*, vol. 7, pp. 96105–96112, 2019.
[36] R. O. Adeogun, P. D. Teal, and P. A. Dmochowski, "Parametric channel prediction for narrowband mobile MIMO systems using spatio-temporal correlation analysis," in *Proc. IEEE 78th Veh. Technol. Conf. (VTC Fall)*, Las Vegas, NV, USA, Sep. 2013, pp. 1–5.
[37] L. Huang, T. Long, E. Mao, and H. C. So, "MMSE-based MDL method for accurate source number estimation," *IEEE Signal Process. Lett.*, vol. 16, no. 9, pp. 798–801, Sep. 2009.
[38] R. O. Adeogun, P. D. Teal, and P. A. Dmochowski, "Parametric channel prediction for narrowband MIMO systems using polarized antenna arrays," in *Proc. IEEE 79th Veh. Technol. Conf. (VTC Spring)*, Seoul, South Korea, May 2014, pp. 1–5.
[39] P. Kyösti, J. Meinilä, L. Hentilä, X. Zhao, T. Jämsä, C. Schneider, M. Narandzić, M. Milojević, A. Hong, J. Ylitalo, V.-M. Holappa, M. Alatossava, R. Bultitude, Y. de Jong, and T. Rautiainen, *WINNER II D1.1.2 v1.2 Part I—Channel Models*, Standard IST-4-027756, Sep. 2007.
[40] T. Eyceoz, A. Duel-Hallen, and H. Hallen, "Deterministic channel modeling and long range prediction of fast fading mobile radio channels," *IEEE Commun. Lett.*, vol. 2, no. 9, pp. 254–256, Sep. 1998.
[41] A. Duel-Hallen, S. Hu, and H. Hallen, "Long-range prediction of fading signals," *IEEE Signal Process. Mag.*, vol. 17, no. 3, pp. 62–75, May 2000.
[42] W. Peng, M. Zou, and T. Jiang, "Channel prediction in time-varying massive MIMO environments," *IEEE Access*, vol. 5, pp. 23938–23946, 2017.
[43] 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.
[44] T. Svantesson and A. L. Swindlehurst, "A performance bound for prediction of MIMO channels," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 520–529, Feb. 2006.
[45] W. Jiang and T. Kaiser, "From OFDM to FBMC: Principles and comparisons," in *Signal Processing for 5G: Algorithms and Implementations*, F. L. Luo and C. Zhang, Eds. London, U.K.: Wiley, 2016, ch. 3.
[46] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*, 1st ed. New York, NY, USA: Prentice-Hall, 1975.
[47] W. Jiang and M. Schellmann, "Suppressing the out-of-band power radiation in multi-carrier systems: A comparative study," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Anaheim, CA, USA, Dec. 2012, pp. 1477–1482.
[48] H. Zhang and R. U. Nabar, "Transmit antenna selection in MIMO-OFDM systems: Bulk versus per-tone selection," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Beijing, China, May 2008, pp. 4371–4375.
[49] *Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) Radio Transmission and Reception, v15.4.0*, document 3GPP TS36.104, Sep. 2018.
[50] X. Fu, S. Li, M. Fairbank, D. C. Wunsch, and E. Alonso, "Training recurrent neural networks with the Levenberg–Marquardt algorithm for optimal control of a grid-connected converter," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 9, pp. 1900–1912, Sep. 2015.



WEI JIANG (M'09–SM'19) received the B.S. degree in electrical engineering from Beijing Information Science and Technology University, Beijing, China, in 2002, and the Ph.D. degree in computer science from the Beijing University of Posts and Telecommunications, Beijing, in 2008.

From 2008 to 2012, he was a Research Engineer with the 2012 Laboratory, HUAWEI Technologies. From 2012 to 2015, he was a Postdoctoral Research Fellow with the Institute of Digital Signal Processing, University of Duisburg-Essen, Germany. Since 2015, he has been a Senior Researcher and a Project Manager with the German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany. He has been a Senior Lecturer with the Department of Electrical and Computer Engineering, University of Kaiserslautern, Germany. He is the author of three book chapters and over 50 conference and journal articles, holds around 30 granted patents, and participated in a number of research projects, such as FP7 ABSOLUTE, H2020 5G COHERENT, 5G SELFNET, and German BMBF TACNET4.0. His research interests include digital signal processing, MIMO-OFDM techniques, opportunistic relaying, SDN/NFV, fading channel prediction, intelligent network management, deep learning, and neural networks.

Dr. Jiang has served as a Vice Chair of the Special Interest Group (SIG) "Cognitive Radio in 5G" within the Technical Committee on Cognitive Networks (TCCN) of the IEEE Communication Society.

HANS D. SCHOTTEN (S'93–M'97) received the Diploma and Ph.D. degrees in electrical engineering from the RWTH Aachen University of Technology, Aachen, Germany, in 1990 and 1997, respectively.

From 1999 to 2003, he worked for Ericsson Corporate Research and Development in research and standardization in the area of mobile communications. From 2003 to 2007, he worked for Qualcomm in research and standardization. He became the Manager of the Research and Development Group, a Research Coordinator for Qualcomm Europe, and the Director for Technical Standards. In 2007, he accepted the offer to become a Full Professor and the Director of the institute for Wireless Communications and Navigation, University of Kaiserslautern. In 2012, he, in addition, became the Scientific Director of the German Research Center for Artificial Intelligence (DFKI) and the Head of the Department for Intelligent Networks. He has served as the Dean of the Department of Electrical Engineering, University of Kaiserslautern, from 2013 to 2017. He is the author of more than 200 articles, filed 13 patents, and participated in more than 30 European and national collaborative research projects, including the EU projects WINNER II, C-MOBILE, C-CAST, METIS, METIS II, 5G NORMA, 5G MONARH, 5G SELFNET, and ITN 5G AURA. Since 2018, he has been the Chairman of the German Society for Information Technology and a member of the Supervisory Board of the VDE.