Deep Learning for Fading Channel Prediction

Wei Jiang, *Senior Member, IEEE*, and Hans Dieter Schotten, *Member, IEEE*

Channel state information (CSI), which enables wireless systems to adapt their transmission parameters to instantaneous channel conditions and consequently achieve great performance boost, plays an increasingly vital role in mobile communications. However, getting accurate CSI is challenging due mainly to rapid channel variation caused by multi-path fading. The inaccuracy of CSI imposes a severe impact on the performance of a wide range of adaptive wireless systems, highlighting the significance of channel prediction that can combat outdated CSI effectively. The aim of this article is to shed light on the state of the art in this field and then go beyond by proposing a novel predictor that leverages the strong time-series prediction capability of deep recurrent neural networks incorporating long short-term memory or gated recurrent unit. In addition to an analytical comparison of computational complexity, performance evaluation in terms of prediction accuracy is carried out upon multi-antenna fading channels. Numerical results reveal that deep learning brings a notable performance gain compared with the conventional predictors built on shallow recurrent neural networks.

***Index Terms*—5G, Artificial Intelligence, channel prediction, channel state information, deep learning, GRU, LSTM, machine learning, multi-antenna system, recurrent neural network.**

I. Introduction

In wireless communications, the form of signal transmission can be basically categorized into two types: open-loop and closed-loop. The former is blind to channel conditions, whereas the latter has to get some channel knowledge so as to realize efficient transmission, known as adaptive transmission system (ATS) or adaptive signaling [1]. With the assist of channel state information (CSI), the transmitter is able to adaptively choose its parameters such as the transmit power, constellation size, coding rate, transmit antenna, and precoding codeword to achieve great performance. In a frequency-division duplex (FDD) system, the CSI is obtained by estimating received reference signals at the receiver and then fed back to the transmitter. Due to the feedback delay, it tends to become outdated under the rapid channel variation caused by multi-path fading. Although a time-division duplex (TDD) system avoids the necessity of feedback by means of channel reciprocity [2], the processing delay still raises inaccuracy, especially in high mobility or high frequency band scenarios. It has been extensively recognized that the outdated CSI imposes overwhelming performance deterioration on a wide variety of wireless systems, such as multi-input multi-output (MIMO), massive MIMO (mMIMO), multi-user scheduling, interference alignment (IA), beamforming (BF), transmit antenna selection (TAS), closed-loop transmit antenna diversity (TAD), opportunistic relaying (OR), coordinated multi-point (CoMP), orthogonal frequency-division multiplexing (OFDM), mobility/resource management (MRM), and physical layer security (PLS). Nowadays, some of 5G usage scenarios, e.g., millimeter wave (having shorter wavelength) [3], high-speed train, vehicle-to-x, and unmanned aerial vehicle [4] (with higher moving speed), suffer from faster channel fading, further highlighting the problem of outdated CSI.

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

The authors are with the Intelligent Networking Research Group, German Research Center for Artificial Intelligence (DFKI), 67663 Kaiserslautern, Germany, and also with the Department of Electrical and Computer Engineering, Technische Universität Kaiserslautern, 67663 Kaiserslautern, Germany, e-mail: (wei.jiang@dfki.de; schotten@eit.uni-kl.de).

To deal with it, researchers have proposed a large number of algorithms and protocols, which either passively compensate for the performance loss with a cost of scarce wireless resources [5] or aim to achieve merely part of the full potential under the assumption of imperfect CSI [6]. In contrast, an alternative technique, referred to as channel prediction [7], provides an efficient solution that can improve the accuracy of CSI directly without spending extra radio resources. Its key idea is to forecast future CSI in advance with a time span that counteracts the induced delay. Through modeling a wireless channel into a set of radio propagation parameters, two statistical prediction approaches - Auto-Regressive (AR) and Parametric Model (PM) - have been proposed [8]. However, the modelling is fossilized, leading to a gap between these models and real channels, and - in addition - the parameter estimation process relying on complex algorithms such as MUSIC and ESPRIT [9] is tedious, harmed its applicability in practical systems. In 2016 when AlphaGo [10], a deep learning (DL) computer program developed by Google, achieved a historic victory versus a human champion in the game of Go, an exploration on Artificial Intelligence (AI) was triggered almost in all scientific and engineering branches [11], [12]. As a classical AI technique, neural networks (NNs) can avoid the parameter estimation thanks to its data-driven nature, and therefore attracts the interest from researchers in the field of channel prediction. Making use of its capability on time-series prediction [13], a number of predictors [14] mostly based on recurrent neural networks (RNNs) were proposed, as surveyed in the next section.

By far the up-to-date prediction approaches are still limited to *shallow* neural networks, to the best knowledge of the authors, deep learning with advanced recurrent structures such as long short-term memory (LSTM) or gated recurrent unit (GRU) is still untouched in this field. This article first conducts a survey in order to shed light on the state of the art and then go beyond by proposing a novel predictor leveraging the strong time-series prediction capability of deep learning. The mechanism of a deep RNN that incorporates LSTM or GRU memory cells is presented, followed by an analytical comparison of computational complexity. Furthermore, performance

assessment in terms of prediction accuracy is carried out upon multi-antenna fading channels, taking into account a number of factors such as additive noise, mobility (indicated by the Doppler frequency shift), MIMO scale, activation function, and number of hidden neurons. The main contributions of this article can be listed as follows:

- A survey, summarizing the impact of outdated CSI on the performance of a wide rage of wireless systems and the existing prediction schemes based on statistical modeling and neural networks, is provided.
- A novel MIMO channel predictor built on a deep recurrent neural network that incorporates LSTM or GRU memory cells is proposed.
- The computational complexity of predictors based on RNN, LSTM, and GRU with different hidden layers is comparatively analyzed.
- Simulation evaluation in terms of prediction accuracy in multi-antenna fading channels is conducted and some representative numerical results are explained.

The rest of this article is organized as follows: Section II surveys the performance deterioration due to the outdated CSI, and the existing prediction schemes. Section III introduces a deep recurrent network with LSTM or GRU hidden layers and presents the principle of the proposed DL predictor. Section IV analyzes the complexity of different predictors. Section V details the simulation configuration and presents the numerical results. Finally, Section VI closes this article with our remarks.

TABLE I
LIST OF ABBREVIATIONS IN ALPHABETICAL ORDER.

Acronym	Definition
5G	Fifth Generation
AI	Artificial Intelligence
AR	Auto-Regressive
ATS	Adaptive Transmission System
BER	Bit Error Rate
BF	Beam-Forming
CNN	Convolutional Neural Network
CSI	Channel State Information
CoMP	Coordinated Multi-Point
DL	Deep Learning
FC	Fully Connected
FDD	Frequency-Division Multiplexing
GRU	Gated Recurrent Unit
IA	Interference Alignment
LSTM	Long Short-Term Memory
LTE	Long-Term Evolution
MIMO	Multi-Input Multi-Output
mMIMO	massive MIMO
MRM	Mobility/Resource Management
MSE	Mean Squared Error
MU-MIMO	Multi-User MIMO
NN	Neural Network
OFDM	Orthogonal Frequency-Division Multiplexing
OR	Opportunistic Relaying
PLS	Physical Layer Security
PM	Parametric Model
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
SISO	Single-Input Single-Output
SNR	Signal-to-Noise Ratio
TAD	Transmit Antenna Diversity
TAS	Transmit Antenna Selection
TDD	Time-Division Multiplexing

II. A SURVEY ON PREVIOUS WORKS

It is well known that timely adaption of transmission parameters in wireless systems can yield large performance improvement. From a practical point of view, however, the channel condition at the time of selecting parameters may substantially differ from the condition at the instant of using these parameters to transmit due to multi-path fading. Using an outdated version of CSI rather than its actual value severely deteriorates system performance, which has been extensively reported in the literature. Some representative works are summarized below and also listed in Table II.

Li *et al.* investigate the performance degradation of a MIMO maximal ratio combining system in the presence of feedback delay and channel estimation error [15], while the impact of feedback delay on spectral efficiency and bit error rate (BER) of MIMO is presented in [16]. Truong *et al.* discuss the effect of channel aging in massive MIMO systems, where the analytical results show how capacity is lost due to time variation in the channel [17]. The authors of [18] focus on the challenge that delayed CSI severely degrades the throughput of multi-user multi-input multi-output (MU-MIMO) transmission and therefore results in a preference for single-user transmission. In [19], Yu *et al.* deliver unified analyses of BER and outage probability for MU-MIMO systems with imperfect CSI caused by feedback over Rayleigh fading channels. Kim *et al.* compare the effects of outdated CSI and channel estimation error on the outage performance of cooperative distributed beamforming. It is shown in [20] that outdated CSI seriously degrades the performance, and therefore the outage probability is bounded and no diversity is achieved. Conversely, channel estimation errors merely cause slight performance degradation, and full diversity order is still achievable. Aquilina and Ratnarajah analyse the performance of interference alignment under imperfect CSI [21], and point out that the full degree-of-freedom gain promised by IA cannot be achieved if the CSI imperfection exists. In [22], Yu *et al.* give a unified error analysis of multi-antenna systems with orthogonal space-time block coding using transmit antenna selection over Nakagami-m channels in the presence of feedback delay and channel estimation error. The authors of [23] derive closed-form expressions of the average BER for closed-loop transmit diversity in a time-selective Rayleigh fading channel containing feedback delay, demonstrating that the advantage of closed loop over open loop disappears due to the CSI imperfection. Vicario *et al.* study the impact of outdated CSI on opportunistic relaying by analyzing the outage probability and diversity order of decode-and-forward OR in the presence of feedback delay [24]. The analytical and numerical results

TABLE II
SUMMARY OF THE IMPACT OF OUTDATED CSI AND THE EXISTING PREDICTIVE SCHEMES.

Domain	Subject	Reference
Impact of Outdated CSI	MIMO	[15], [16], [30]
	mMIMO	[17], [31], [32]
	MU-MIMO	[18], [19], [33]
	BF	[20], [34], [35]
	IA	[21], [36], [37]
	TAS	[22], [38], [39]
	TAD	[23], [40], [41]
	OR	[24], [42], [43]
	CoMP	[25], [44], [45]
	OFDM	[26], [27], [46]
	MRM	[28], [47], [48]
	PLS	[29], [49], [50]
Statistical Prediction	AR	[51]–[54]
	PM	[55]–[57]
	Application	[58]–[61]
AI Prediction	RNN	[62]–[69]
	NN	[70]–[72]
	LSTM	[73], [74]
	DL	[75]–[78]

show that the diversity order of the OR system is reduced to 1, i.e., no diversity, when CSI is outdated, being independent of the level of CSI accuracy. The performance of CoMP systems heavily depends on the feedback quality and channel imperfection. In [25], therefore, the authors focus on the impact of quantized and delayed CSI on the average achievable rate of joint transmission and coordinated beamforming. Wang *et al.* report the effect of imperfect CSI on the performance of radio resource management for an OFDM system, identifying that the imperfect CSI yields an error floor [26]. Guharoy and Mehta develop a comprehensive analysis for OFDM system performance in the presence of feedback delays in [27], which augues that small feedback delays markedly degrade throughput and increase outage probability even at low vehicular speeds. The effect of outdated CSI on handover decisions in dense networks is studied in [28], drawing a conclusion that handover decisions are very sensitive to the accuracy of CSI. Last but not least, Yang *et al.* [29] identify a tight asymptotic lower bound of the ergodic secrecy capacity under imperfect CSI that includes both outdated CSI due to the transmission and processing delay, and channel estimation errors, proving that the imperfect CSI greatly reduces the performance of physical layer security systems.

Compared with the traditional approaches against outdated CSI, channel prediction is more efficient since it improves the accuracy of CSI directly without spending scarce radio resources. Relying on statistical modeling, the conventional prediction schemes, mainly including AR and PM, were designed. By exploiting temporal correlation, such as [51]–[54], AR models the impulse response of a time-varying channel as an autoregressive process and employs a Kalman filter to estimate AR coefficients. As illustrated in [55]–[57], PM decomposes a fading channel into a superposition of a finite number of complex sinusoids, each of which has its respective amplitude, Doppler shift, and phase, and a channel can be extrapolated from propagation parameters. The rationale is based on an observation that these parameters change slowly in comparison with the fading rate. Additionally, some articles also focus on the application of channel prediction in dedicated wireless aspects. In [58], for example, Svantesson and Swindlehurst identify the performance bound for MIMO channel prediction and verify it through numerical evaluation. Zhou and Giannakis [59] present a predictor based on pilot symbol assisted modulation for MIMO channels and then analyze the impact of prediction error on the BER of a transmit-beamformer. Su *et al.* discuss whether prediction is useful for CoMP systems with backhaul latency in time-varying channels and claim that it provides much higher gain over channel estimation [60]. The performance degradation of TAS systems due to outdated channel knowledge is analytically determined and a predictive scheme is designed to mitigate delay-induced degradation in [61].

As mentioned previously, AI shows great potential since it provides an alternative solution that replaces statistical modeling with data-driven approaches, therefore, attracting the interest of researcher in this field. Making use of its capability on time-series prediction, a recurrent neural network is first employed to predict narrowband single-input single-output (SISO) channels in [62] and is further extended to MIMO flat-fading channels by [63], [64]. To tackle the frequency selectivity in wideband communications, Ding and Hirose [70] use a complex-valued NN to predict frequency-domain channel characteristics, while a frequency-domain RNN predictor that deal with a frequency-selective MIMO channel as a set of parallel flat-fading sub-carriers is proposed in [65]. Reference [71] delivers a predictive method by means of combining a multi-layer complex-valued NN with the chirp Z-transform. To lower complexity, the authors of [66] propose real-valued RNNs to implement multi-step predictors for long-range prediction [67], and further verify its effectiveness in a TAS system [68]. In [69], a RNN predictor is presented, with an emphasis on different training algorithms. Sui *et al.* deliver a jointly optimized extreme learning machine for short-term prediction in [72], while Tong and Sun explore the application of LSTM for long-term prediction in their work [73]. In [75], the authors argue for applying conventional neural network (CNN) to extract CSI pattern and present a CNN-RNN architecture for CSI aging. Mehrabi *et al.* build a decision-directed estimation with deep feedforword neural network based channel prediction for MIMO transmission [76]. In addition to channel prediction, AI methods are also used to tackle classical channel estimation problem, such as [77], [79]. The novelty of the predictor proposed in this article is mainly that it makes use of deep recurrent networks with LSTM or GRU layers, which is not explored in this field until now.

In contrast to the aforementioned "narrow-sense" prediction that focuses on forecasting future CSI from its current and past values at the same frequency, a few recent works discuss an inter-band idea to predict downlink CSI from uplink channel knowledge. Wang *et al.* [74] deliver a hybrid of a convolutional

neural network and long short-term memory to extract the downlink CSI according to that of uplink channels assuming strong channel correlation. Arnold *et al.* present a deep-learning based extrapolation approach that infers the downlink CSI by solely observing uplink CSI on an *adjacent* frequency band [78]. Because of severe signal attenuation in radio channels, a transceiver suffers from strong self-interference between its emitting signals in the downlink and incoming signals in the uplink. Hence, a FDD system must set an enough-wide guard band between its paired carrier frequencies. For example, LTE band 1 (uplink: 1920 to 1980MHz, downlink: 2110 to 2170MHz) has a duplex spacing of 190MHz. From the practical perspective, these inter-band predictive methods, relying on the frequency correlation among adjacent channels, are infeasible.

III. Deep Learning based Channel Prediction

This section first introduces the principle of a deep recurrent neural network incorporating LSTM or GRU layers, and then illustrates the proposed DL-based channel predictor through a prediction-aided adaptive MIMO system.

A. Deep RNN

Differing from feedforward neural networks, recurrent neural networks have self-connections, feeding the activation from the previous time step back to the network as input for the current time step. As a class of classical neural networks, RNN is good at processing data sequences through storing indefinite historical information in its internal state, exhibiting great potential in time-series prediction [13]. However, it suffers from the gradient exploding and vanishing problems with the gradient-based back-propagation through time training technique, where a back-propagated error signal is apt to be very large, leading to oscillating weights, or tends to zero that implies a prohibitively long training time or training does not work at all.

To this end, Hochreiter and Schmidhuber designed an elegant RNN structure - long short-term memory - in 1997 in their pioneer work of [80]. The key innovation of LSTM to deal with long-term dependency is the introduction of special units called memory cells in the recurrent hidden layer and multiplicative gates that regulate the information flow. In the original structure of LSTM, each memory block contains two gates: an *input* gate protecting the memory contents stored in the cell from perturbation by irrelevant interference, and an *output* gate that controls the extent to which the memory information applied to generate the output activation. To address a weakness of LSTM, namely the internal state grows indefinitely and eventually cause the network to break down when processing continual input streams that are not segmented into subsequences, a *forget* gate is added [81]. It scales the internal state of the memory cell before cycling back through self recurrent connections. Although its history is not long, LSTM has been applied successfully to sequence prediction and labeling tasks. It has already gotten the state-of-the-art technological results in many fields such as machine translation, speech recognition, and handwriting recognition, and has also achieved a great commercial success, justified by many unprecedented intelligent services such as Google Translate and Apple iPhone Siri.

Like a deep RNN consisting of multiple recurrent hidden layers, a deep LSTM network is built by stacking multiple LSTM layers. Without loss of generality, Fig. 1 shows an example of a deep LSTM network that consists of an input layer, three hidden layers, and an output layer. At an arbitrary time step, as illustrated in the left part of Fig. 1, a data vector $\mathbf{x}$ goes through the input feedforward layer to get $\mathbf{d}^{(1)}$, which is the activation for memory cells in the first hidden layer. Along with the recurrent unit feeding back from the previous time step, $\mathbf{d}^{(2)}$ is generated and then forwarded to the second hidden layer. This recursive process continues until the output layer gets $\mathbf{y}$ according to $\mathbf{d}^{(4)}$. Unrolling the network through time, as illustrated in the right part of Fig. 1, the memory block at the l^{th} hidden layer has two internal states at time step $t-1$, i.e., the short-term state $\mathbf{s}_{t-1}^{(l)}$ and the long-term state $\mathbf{c}_{t-1}^{(l)}$. Traversing the memory cells from the left to the right, $\mathbf{c}_{t-1}^{(l)}$ first throws away some old memories at the forget gate, integrates new information selected by the input gate, and then sends out as the current long-term state $\mathbf{c}_{t}^{(l)}$. The input vector $\mathbf{d}_{t}^{(l)}$ and the previous short-term memory $\mathbf{s}_{t-1}^{(l)}$ are fed into four different fully connected (FC) layer, generating the activation vectors of gates:

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

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

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

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

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

Dropping some old memories at the forget gate, and adding some new information selected from current memory input that is defined as

$$\mathbf{g}_t^{(l)} = \sigma_h\left(\mathbf{W}_g^{(l)}\mathbf{d}_t^{(l)} + \mathbf{U}_g^{(l)}\mathbf{s}_{t-1}^{(l)} + \mathbf{b}_g^{(l)}\right), \quad (5)$$

the previous long-term memory $\mathbf{c}_{t-1}^{(l)}$ is thus transformed into

$$\mathbf{c}_t^{(l)} = \mathbf{f}_t^{(l)} \otimes \mathbf{c}_{t-1}^{(l)} + \mathbf{i}_t^{(l)} \otimes \mathbf{g}_t^{(l)}, \quad (6)$$

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

$$\sigma_h(x) = \frac{e^{2x}-1}{e^{2x}+1}. \quad (7)$$

In addition to sigmoid and tanh, there are other commonly used activation functions, e.g., the *rectified linear unit* (ReLU) that can be written as

$$\sigma_r(x) = max(0, x), \quad (8)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2020.2982513, IEEE Open Journal of the Communications Society

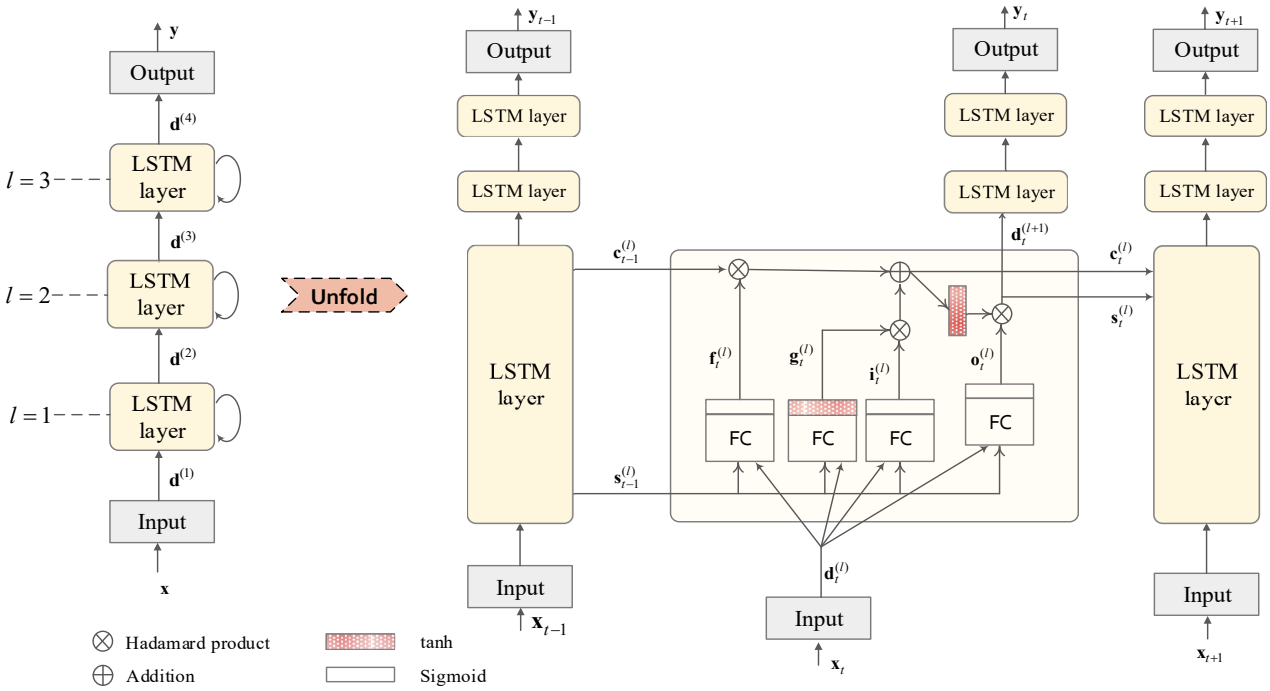


Fig. 1. Schematic diagram of a three-hidden-layer deep LSTM network.

which returns 0 if it receives any negative input, but for any positive value x it returns that value back. Further, $\mathbf{c}_t$ passes through the *tanh* function and then is filtered by the output gate to produce the current short-term memory, as well as the output for this memory block, i.e.,

$$\mathbf{s}_t^{(l)} = \mathbf{d}_t^{(l+1)} = \mathbf{o}_t^{(l)} \otimes \sigma_h\left(\mathbf{c}_t^{(l)}\right). \tag{9}$$

Since the advent of LSTM, its original structure continues to evolve. Cho *et al*. [82] proposed a simplified version with fewer parameters in 2014, known as GRU, which exhibits even better performance over LSTM on certain smaller data sets. In a GRU memory block, the short- and long-term states are merged into a single one, and a single gate $\mathbf{z}_t^{(l)}$ is used to replace the forget and input gates, namely,

$$\mathbf{z}_t^{(l)} = \sigma_g\left(\mathbf{W}_z^{(l)}\mathbf{d}_t^{(l)} + \mathbf{U}_z^{(l)}\mathbf{s}_{t-1}^{(l)} + \mathbf{b}_z^{(l)}\right). \tag{10}$$

The output gate is removed, but an intermediate state $\mathbf{r}_t^{(l)}$ is newly introduced, i.e.,

$$\mathbf{r}_t^{(l)} = \sigma_g\left(\mathbf{W}_r^{(l)}\mathbf{d}_t^{(l)} + \mathbf{U}_r^{(l)}\mathbf{s}_{t-1}^{(l)} + \mathbf{b}_r^{(l)}\right). \tag{11}$$

Likewise, the hidden state at the previous time step transverses the memory cells, drops some old memory, and loads some now information, resulting in the current state:

$$\begin{aligned}\mathbf{s}_t^{(l)} &= (1-\mathbf{z}_t^{(l)}) \otimes \mathbf{s}_{t-1}^{(l)} \\ &+ \mathbf{z}_t^{(l)} \otimes \sigma_h\left(\mathbf{W}_s^{(l)}\mathbf{d}_t^{(l)} + \mathbf{U}_s^{(l)}(\mathbf{r}_t^{(l)} \otimes \mathbf{s}_{t-1}^{(l)}) + \mathbf{b}_s^{(l)}\right).\end{aligned} \tag{12}$$

B. DL-based Predictor

To shed light on the operation of the DL predictor, a point-to-point flat-fading MIMO system with N_t transmit and N_r receive antennas is studied. Its transmission is modelled as

$$\mathbf{r}[t] = \mathbf{H}[t]\mathbf{s}[t] + \mathbf{n}[t], \tag{13}$$

where $\mathbf{r}[t]$ and $\mathbf{s}[t]$ denote the received and transmitted signal vectors at time step t, respectively, $\mathbf{n}[t]$ is the vector of additive noise, and $\mathbf{H}[t]$ represents an $N_r \times N_t$ channel matrix, whose (n_r, n_t)-entry $h_{n_r n_t}$ is the complex-valued gain of the channel between transmit antenna n_t and receive antenna n_r. As illustrated in Fig.2, the transmitter requires CSI feedback so as to adapt its transmission parameters to a time-varying channel. Due to feedback delay τ, when the transmitter uses $\mathbf{H}[t]$ to choose parameters, the instantaneous channel gain already changes to $\mathbf{H}[t+\tau]$. It is probably that $\mathbf{H}[t] \neq \mathbf{H}[t+\tau]$, especially in high mobility environment. Outdated CSI imposes severe performance loss on a wide variety of adaptive wireless techniques. Hence, it is worth conducting channel prediction at the receiver to obtain predicted CSI $\hat{\mathbf{H}}[t+D]$, where $D \geqslant \tau$, to counteract the effect of feedback delay, or equivalently, performing prediction at the transmitter.

As illustrated in Fig.2, the instantaneous channel matrix $\mathbf{H}[t]$ is estimated at the receiver, which is fed into the predictor, rather than being fed back to the transmitter directly, as does a typical adaptive MIMO system. Based on an observation that the magnitude of channel gain, i.e., $|h_{n_r n_t}|$, is already enough for most of adaptive tasks, the prediction is applied on such real-valued channel data. In order to adapt the input layer of a neural network, we add a data pre-processing layer in the predictor, where a channel matrix is transferred into a vector

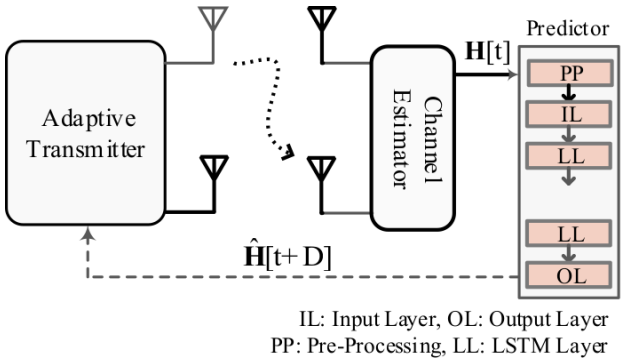


Fig. 2. Block diagram of an adaptive MIMO system integrating a DL predictor with LSTM layers.

of channel magnitudes, like:

$$\mathbf{H}[t] \rightarrow [|h_{11}[t]|, |h_{12}[t]|, ..., |h_{N_r N_t}[t]|]^T. \quad (14)$$

Replacing $\mathbf{x}_t$ in Fig.1 with this channel vector and going through a number of hidden layers, the output layer generates $\hat{\mathbf{H}}[t+D]$, which is the D-step prediction. In addition to the magnitude, the predictor can also be applied to process complex-valued channel gains. To this end, the predictor generally needs to be built on a complex-valued deep neural network, which is not well implemented in current AI software tools. Instead, a real-valued network is applied to predict the real and imaginary parts of channel gains, where the data preprocessing layer transforms $\mathbf{H}[t]$ into

$$[\Re(h_{11}[t]), ..., \Re(h_{N_r N_t}[t]), \Im(h_{11}[t]), ..., \Im(h_{N_r N_t}[t])]^T. \quad (15)$$

where $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary units, respectively. Feeding this vector into the predictor at time t, $\hat{\mathbf{H}}[t+D]$ can be obtained after simply combining the predicted real and imaginary units.

In addition to flat fading channels, the predictor can also be applied for frequency-selective channels simply by means of converting it into a set of N narrow-band sub-carriers through, for example, the OFDM modulation [83]. At the n^{th} sub-carrier, the signal transmission is represented by

$$\tilde{\mathbf{r}}_n[t] = \tilde{\mathbf{H}}_n[t]\tilde{\mathbf{s}}_n[t] + \tilde{\mathbf{n}}_n[t], \;\; n = 0, 1, \ldots, N-1, \quad (16)$$

where $\tilde{\mathbf{r}}_n[t]$ represents N_r received symbols for sub-carrier n at time t, $\tilde{\mathbf{s}}_n[t]$ corresponds to N_t transmit symbols, $\tilde{\mathbf{n}}[t]$ is a vector of additive noise. $\tilde{\mathbf{H}}_n[t]=[\tilde{h}_n^{n_r n_t}[t]]_{N_r \times N_t}$ denotes the frequency-domain channel matrix, where $1 \leqslant n_r \leqslant N_r$, $1 \leqslant n_t \leqslant N_t$, and $\tilde{h}_n^{n_r n_t} \in \mathbb{C}^{1\times 1}$ stands for the channel frequency response on sub-carrier n between transmit antenna n_t and receive antenna n_r. The prediction process on each sub-carrier is as same as that of the flat fading channel.

IV. Computational Complexity

The number of parameters including weights and biases for a shallow RNN consisting of an input, a hidden, and an output layer is calculated as follows:

$$N_{rnn} = n_i \times n_c + n_c \times n_c + n_c \times n_o + n_c + n_o, \quad (17)$$

where n_c, n_i and n_o denote the number of hidden neurons, input units, and output units, respectively. The computational complexity of learning models per parameter and time step under the typical stochastic gradient descent training is $\mathcal{O}(1)$. Consequently, the complexity per time step in the training phase can be measured by $\mathcal{O}(N_{rnn})$. During the predicting phase, each weight or bias per time step requires one complex-valued multiplication, amounting to the complexity per prediction of $\mathcal{O}(N_{rnn})$. According to (1)-(12), the number of parameters for a shallow LSTM and GRU network can be derived, which are

$$N_{lstm} = 4(n_i \times n_c + n_c \times n_c + n_c) + n_c \times n_o + n_o, \quad (18)$$

and

$$N_{gru} = 3(n_i \times n_c + n_c \times n_c + n_c) + n_c \times n_o + n_o. \quad (19)$$

Likewise, the complexity of LSTM and GRU can be indicated by $\mathcal{O}(N_{lstm})$ and $\mathcal{O}(N_{gru})$, respectively.

Suppose a deep recurrent network consists of an input layer, L hidden layers, and an output layer, where the number of input units, output units, and hidden neurons or memory cells are n_i, n_o, and n_c^l, $l = 1, \ldots, L$, respectively. Thus, the number of parameters for a deep RNN can be computed as

$$\begin{aligned} N_{drnn} &= n_i \times n_c^1 + n_c^1 \times n_c^1 + n_c^1 \\ &\quad + \sum_{l=2}^{L} \left(n_c^{l-1} \times n_c^l + n_c^l \times n_c^l + n_c^l\right) \\ &\quad + n_c^L \times n_o + n_o. \end{aligned} \quad (20)$$

From the structure of a deep LSTM network as shown in Fig. 1, we can calculate its number of parameters as follows:

$$\begin{aligned} N_{dlstm} &= 4(n_i \times n_c^1 + n_c^1 \times n_c^1 + n_c^1) \\ &\quad + \sum_{l=2}^{L} 4\left(n_c^{l-1} \times n_c^l + n_c^l \times n_c^l + n_c^l\right) \\ &\quad + n_c^L \times n_o + n_o. \end{aligned} \quad (21)$$

To derive that value for a deep GRU network, it is straightforward to replace the factor of 4 in (21) with 3. That is,

$$\begin{aligned} N_{dgru} &= 3(n_i \times n_c^1 + n_c^1 \times n_c^1 + n_c^1) \\ &\quad + \sum_{l=2}^{L} 3\left(n_c^{l-1} \times n_c^l + n_c^l \times n_c^l + n_c^l\right) \\ &\quad + n_c^L \times n_o + n_o. \end{aligned} \quad (22)$$

To utilize a concrete case for comparison, we assume that the predictors have the same total number of 30 hidden neurons or memory cells (deep neural networks equally distribute their neurons among different hidden layers) and prediction is conducted upon a SISO channel. Then, the number of network parameters for different predictors are obtained and listed in Table III. It is interesting to point out that a deeper network corresponds to lower complexity under the same number of hidden neurons, as visualized in Fig. 3.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2020.2982513, IEEE Open Journal of the Communications Society

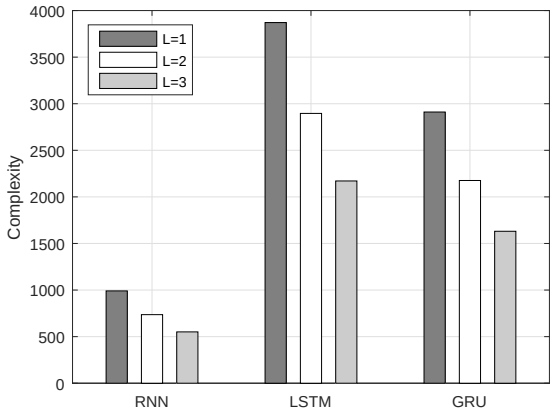


Fig. 3. Comparison on computational complexity.

TABLE III
THE NUMBER OF PARAMETERS (WEIGHTS AND BIASES) FOR DIFFERENT RECURRENT NETWORKS IN SISO CHANNELS WITH $N_c = 30$.

Num. of Layers (L)	**RNN**	**LSTM**	**GRU**
1	991	3871	2911
2	736	2896	2176
3	551	2171	1631

V. PERFORMANCE EVALUATION

A. Simulation Configuration

A performance comparison between the proposed deep learning predictors and the previous predictors based on shallow recurrent networks is conducted. Some representative numerical results in terms of prediction accuracy are reported in this section. We apply independent and identically distributed flat-fading MIMO channels having 4 transmit antennas and a single receive antenna as the baseline. Each subchannel follows the Rayleigh distribution with an average power gain of 0dB, where its channel gain h is zero-mean circularly-symmetric complex Gaussian random variable with a variance of 1, i.e., $h{\sim}\mathcal{CN}(0,1)$. The maximal Doppler frequency shift is set to f_d=100Hz, emulating fast fading environment. To acquire training and testing data, continuous-time channel responses are sampled with a rate of f_s=1KHz, adhering to the assumption of flat fading. The data set contains a series of 10^4 consecutive CSI $\{\mathbf{H}[t]\,|t{=}1,2,\ldots,10000\}$, with a time interval of $T_s{=}1\text{ms}$. In our simulation, 75% of the data is allocated for training and the remaining 25% is testing data.

TABLE IV
SIMULATION CONFIGURATION

Parameters	**Values**
Sampling rate	$f_s = 1\text{KHz}$
Max. Doppler shift	$f_d = 100\text{Hz}$
Channel model	Rayleigh fading
Doppler spectrum	Jakes's model
MIMO size	4×1
Dataset size	10^4
Deep learning	L=2/3 GRU netwok
Training algorithm	Adam optimizer
Batch size	256
Cost function	MSE
(default) Prediction length	1ms
(default) Actuation function	$\tanh$
(default) Hidden neurons	$N_c = 30$

TABLE V
COMPARISON OF MSE RESULTS WITH DIFFERENT NUMBER OF HIDDEN NEURONS.

N_c	$DL^{(3)}$	$DL^{(2)}$	GRU	$LSTM$	RNN
10	0.0021	0.0013	0.0021	0.0021	0.0028
20	9.45e-4	0.0011	0.0016	0.0013	0.0017
30	8.81e-4	**6.52e-4**	**0.0011**	0.0013	0.0022
40	9.97e-4	7.79e-4	0.0011	0.0015	0.0018
50	7.72e-4	7.32e-4	0.0012	0.0013	**0.0015**
60	7.88e-4	8.76e.4	0.0014	0.0013	0.0015
70	6.90e-4	6.62e-4	0.0013	**0.0012**	0.0015
80	7.35e-4	9.07e-4	0.0012	0.0014	0.0018
90	**6.39e-4**	7.40e-4	0.0013	0.0012	0.0020
100	8.19e-4	8.80e-4	0.0018	0.0015	0.0022

A conventional RNN predictor, denoted by *RNN* in the legend of the figures, is exactly a shallow network consisting of an input, an output, and a hidden layer. If this hidden layer is replaced with LSTM or GRU block, the network is transformed into a shallow LSTM or GRU network, respectively, marked by *LSTM* and *GRU* in the legend. The default number of hidden neurons or memory cells is set to 30, and $\tanh$ is applied as the default activation function. Without loss of generality, two GRU networks with two and three hidden layers are employed in our simulation as the representative setup of the proposed DL predictor, notated by $DL^{(2)}$ and $DL^{(3)}$, respectively. For a fair comparison, the total number of hidden neurons among all hidden layers is also 30.

A training process starts from an initial state where all weights and biases are randomly selected. Recalling the structure of the deep recurrent network in Fig. 1, the input of the predictor is $\mathbf{x}_t{=}\mathbf{H}[t]$ and the output is its D-step ahead prediction, i.e., $\mathbf{y}_t{=}\hat{\mathbf{H}}[t{+}D]$. Mean squared error (MSE), a metric for measuring prediction accuracy, is also chosen as the cost function here for training. It is written as

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T}\left\|\hat{\mathbf{H}}[t+D] - \mathbf{H}[t+D]\right\|^2, \qquad (23)$$

where T is the total number of channel samples used for evaluation, $\hat{\mathbf{H}}[t{+}D]$ denotes the predicted CSI, $\mathbf{H}[t{+}D]$ stands for its desired value, and $\|\cdot\|$ notates the Frobenius norm of a matrix. Using the *batch* training method, a batch of 256 samples is fed into the network at each epoch, the resultant outputs are compared with the desired values and the error signals are propagated back through the network to update the weights and biases by training algorithms such as the Adam optimizer used in our simulation. The training process is iteratively carried out until the network reaches a certain

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2020.2982513, IEEE Open Journal of the Communications Society

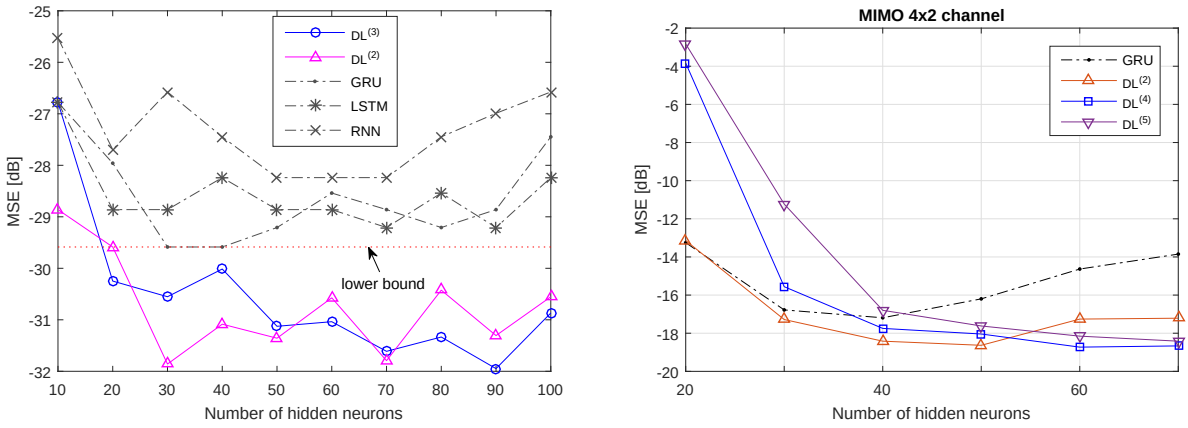


Fig. 4. Prediction accuracy indicated by MSE (in dB) as a function of the number of hidden neurons.

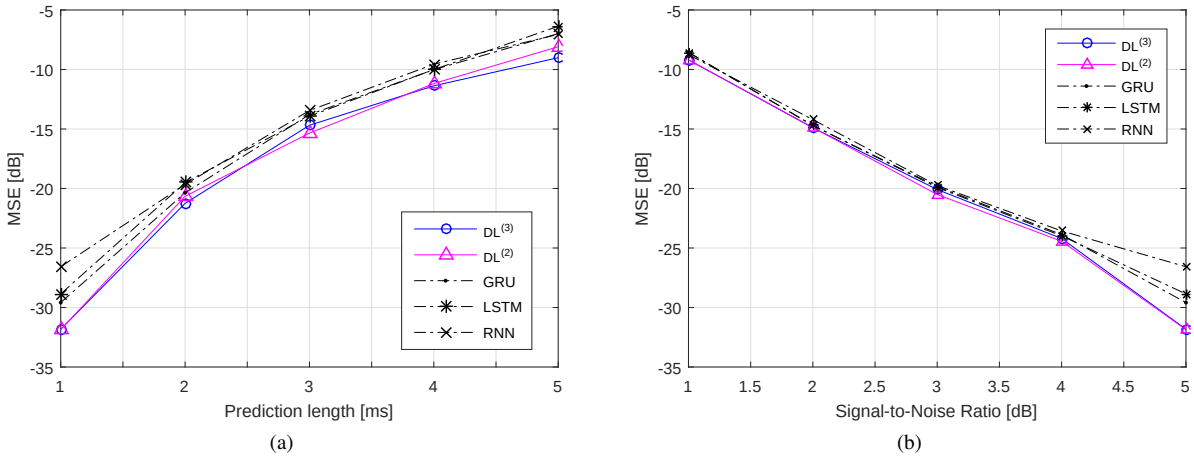


Fig. 5. Performance comparisons in terms of: (a) the length of prediction and (b) the strength of additive noise.

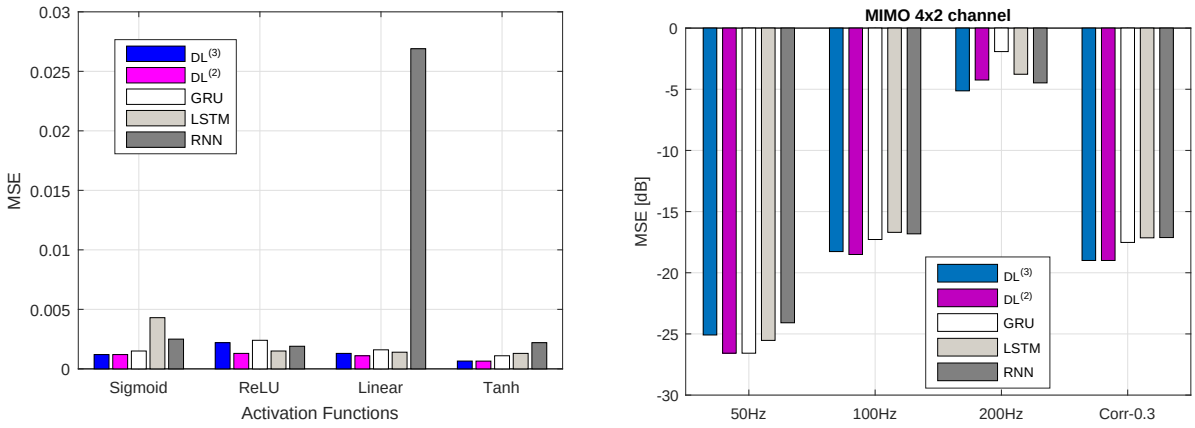
convergence condition. Once it completes, the trained network can be employed to predict future CSI.

B. Prediction Accuracy

The prediction accuracy of the proposed DL predictors as a function of the number of hidden neurons is evaluated and compared with those of three predictors based on shallow NNs. Starting from N_c=10 hidden neurons, as summarized in Table V, the shallow LSTM and GRU networks achieve the same result of MSE=0.0021. The 2-hidden-layer DL network notably outperforms the conventional predictors with an MSE of 0.0013 under the same number of hidden neurons. The 3-hidden-layer DL network performs not so well at this point, getting a result of 0.0021 too, because distributing too few (only 10) neurons in three layers. After that, it will clearly outperform the conventional predictors. With the growth of N_c, the performance of predictors improves with strengthened network capability, until a turn point, at which the network saturates and the overfitting problem appears. For instance, the best result of the $DL^{(2)}$ predictor occurs at N_c=30, rather than N_c=100. Regardless the number of neurons, the optimal results of RNN, LSTM, and GRU are 0.0015, 0.0012, and 0.0011, respectively, which are their performance limits, as indicated by the line of *lower bound* in Fig. 4a. Note that Fig. 4 and the following Fig. 5 display MSE values in decibels (dB) for a clear illustration, calculating by MSE_{dB}=$10\log_{10}(MSE)$. The deep learning can break this limit effectively, as shown in this figure, the performance of two DL predictors is clearly below this lower bound except the case at N_c=10. The best performance of $DL^{(2)}$ and $DL^{(3)}$ are $6.52\times10e^{-4}$ and $6.39\times10e^{-4}$, respectively, with a substantial improvement over the shallow NNs. It is noted that a simplified notation is used in Table V due to space limitation, where for instance $6.52\times10e^{-4}$ is written as $6.52e^{-4}$. In summary, deep learning can improve the accuracy of channel prediction by increasing hidden layers, instead of stacking a large number of neurons within a single layer. As we analyzed in the previous section, the complexity of a deep network is less than a shallow network under the same number of hidden neurons. Hence, we can remark that a deep network is not only effective (higher accuracy) but also efficient (lower complexity) over a shallow network.

To provide a hint about how to determine network hyper-parameters, i.e., the number of hidden layers and the number of neurons, performance evaluation of different kinds of deep

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2020.2982513, IEEE Open Journal of the Communications Society



(a) (b)

Fig. 6. Performance comparisons with respect to: (a) activation function and (b) mobility and correlation.

RNNs over MIMO fading channels with a typical array scale of 4×2 is carried out. According to the results in Fig.4a, we know that GRU is the best among all shallow neural network based predictors, which is therefore selected as the benchmark for comparison here. As shown in Fig.4b, GRU receives its best performance with an MSE of 0.0191 when its single hidden layer has 60 neurons. At the same point, $DL^{(2)}$, i.e., a two-hidden-layer deep RNN with 30 neurons at either layer, as well as $DL^{(4)}$, outperform GRU, lowering the MSE to 0.0144. Due to the problem of overfitting, the prediction accuracy of GRU drops remarkably with the number of hidden neurons increases, reaching 0.0412 at 120 neurons. In contrast, the DL predictors can further improve performance with more neurons, where the result of 0.0134 is achieved by $DL^{(4)}$ with 25 neurons at each hidden layers, which is determined as the best predictor for 4×2 MIMO channels.

The MSE results of the predictors in terms of different prediction lengths are illustrated in Fig. 5a. To begin with one-step ahead mode, i.e., D=1, which corresponds to a prediction length of 1ms as the sampling rate is $f_s = 1\text{KHz}$. The proposed DL predictors clearly outperform three conventional ones, reaping performance gains from 3 to 5dB. Increasing D from 1 up to 5 incrementally, the results on longer ranges from 1ms to 5ms are obtained. The longer prediction length, the worse prediction accuracy, because channel's temporal correlation weakens. Under different lengths, deep learning can receive a gain from 1 to 2dB compared with shallow networks. It is worth emphasizing again that the gain is achieved under the identical number of hidden neurons, implying less complexity in deep networks. In addition, the effect of noise is observed and illustrated in Fig. 5b. The horizonal axis of the figure is the signal-to-noise ratio (SNR) of the channel samples, where the rightmost "infinity" corresponds to an extremely large SNR, i.e., noiseless or noise-free. In noisy channels, three conventional predictors have approximately the same (bad) results. It implies that an effort to improve accuracy against noise by using LSTM or GRU to enhance the capability of a shallow network is marginal. In contrast, deep learning performs a bit well and set a clear performance border with others, exhibiting a gain of nearly 1dB.

Performance comparisons with respect to different activation functions and mobility/correlation are also carried out. As illustrated in Fig. 6a, the recurrent networks can collectively achieve their best performance by using $\tanh$, which is the default activation function in our simulation. In particular, the prediction accuracy of the conventional RNN with the linear activation function is obviously weak, because it cannot deal with nonlinearity. This case, LSTM and GRU boost the performance with an order of magnitude, dropping the MSE value from over 0.025 to less than 0.0025. As we can observe from the figure, deep learning has the best performance under all kinds of activation functions. Fig.6b compares the performance over 4×2 MIMO channels under different mobility and antenna correlation. For a clear illustration, the MSE values are represented in dB in this figure. The deep learning achieves higher prediction accuracy in all three mobility scenarios indicated by the Doppler shift of 50, 100, and 200Hz. In correlated MIMO channels with correlation coefficient of 0.3, the deep learning still clearly outperforms the existing shallow networks.

Last but not least, we explore the impact of recurrent cell structure on the performance. Using the GRU cell described in (12) as an example, the coefficients are first swapped to build a new cell, which is expressed by

$$\begin{aligned}\mathbf{s}_t^{(l)} &= \mathbf{z}_t^{(l)} \otimes \mathbf{s}_{t-1}^{(l)} \\ &\quad + (1-\mathbf{z}_t^{(l)}) \otimes \sigma_h\left(\mathbf{W}_s^{(l)}\mathbf{d}_t^{(l)} + \mathbf{U}_s^{(l)}(\mathbf{r}_t^{(l)} \otimes \mathbf{s}_{t-1}^{(l)}) + \mathbf{b}_s^{(l)}\right).\end{aligned}$$

This customized GRU with 60 hidden neurons in a single hidden layer achieves an MSE of 0.0161 over 4×2 MIMO channels, approaching closely that of the original GRU, i.e., 0.016. That is because such swapping does not change the structure and the new cell adopts itself by generating different weight matrices and biases in order to get the same objective (minimizing MSE). Additionally, we change its structure by nulling a branch, i.e., replacing $(1-\mathbf{z}_t^{(l)})$ in (12) with 0, resulting in

$$\mathbf{s}_t^{(l)} = \mathbf{z}_t^{(l)} \otimes \sigma_h\left(\mathbf{W}_s^{(l)}\mathbf{d}_t^{(l)} + \mathbf{U}_s^{(l)}(\mathbf{r}_t^{(l)} \otimes \mathbf{s}_{t-1}^{(l)}) + \mathbf{b}_s^{(l)}\right). \tag{24}$$

It slightly outperforms the original GRU with the resulting MSE of 0.014. It seems that we can derive a customized recurrent network dedicated for channel prediction from the general GRU cell. Last, we continue to nulling another branch to get

$$\mathbf{s}_t^{(l)} = (1 - \mathbf{z}_t^{(l)}) \otimes \mathbf{s}_{t-1}^{(l)},$$

which breaks the structure of the original cell drastically and the input data is omitted, leading to the MSE of 0.9814, tens of times higher than the original one.

VI. Conclusions

This article first surveyed the impact of outdated CSI on the performance of adaptive transmission systems, and identified the state of the art in the area of multi-path fading channel prediction. On top of this, a novel channel predictor empowered by a deep recurrent neural network integrating long short-term memory or gated recurrent unit were proposed. In addition to an analytical comparison on computational complexity, the evaluation of prediction accuracy taking into account a number of affecting factors such as prediction range, additive noise, number of hidden neurons, activation function, and mobility were conducted. The numerical results in terms of mean squared error revealed that deep learning can achieve remarkably better performance. From the perspective of performance and complexity, we can remark that a deep network is not only effective (higher accuracy) but also efficient (lower complexity) over a shallow network, opening the possibility of its application into a wide variety of wireless systems. The positive outcomes reported in this article could encourage a further exploration of deep learning not only in fading channel prediction but also other aspects of wireless communications.

References

[1] H. Zheng *et al.*, "Adaptive signaling based on statistical characterizations of outdated feedback in wireless communications," *Proceedings of the IEEE*, vol. 95, no. 12, pp. 2337–2353, Dec. 2007.

[2] F. Gfeller and W. Hirt, "A robust wireless infrared system with channel reciprocity," *IEEE Commun. Mag.*, vol. 36, no. 12, pp. 100–106, Dec. 1998.

[3] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[4] Y. Zeng *et al.*, "Wireless communications with unmanned aerial vehicles: opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, May 2016.

[5] 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.

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

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

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

[9] W. Gardner, "Simplification of MUSIC and ESPRIT by exploitation of cyclostationarity," *Proceedings of the IEEE*, vol. 76, no. 7, pp. 845–847, Jul. 1988.

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

[11] W. Jiang, M. Strufe, and H. Schotten, "Experimental results for artificial intelligence-based self-organized 5G networks," in *Proc. IEEE PIMRC'17*, Montreal, QC, Canada, Oct. 2017.

[12] W. Jiang *et al.*, "Intelligent network management for 5G systems: The SELFNET approach," in *Proc. European Conf. on Net. and Commun. (EUCNC)*, Oulu, Finland, Jun. 2017, pp. 109–113.

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

[15] M. Li *et al.*, "Performance analysis of MIMO MRC systems with feedback delay and channel estimation error," *IEEE Trans. Veh. Technol.*, vol. 65, no. 2, pp. 707–717, Feb. 2016.

[16] H. Nguyen *et al.*, "Capacity and performance of MIMO systems under the impact of feedback delay," in *Proc. IEEE PIMRC'04*, Barcelona, Spain, Sep. 2004.

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

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

[19] X. Yu *et al.*, "Unified analysis of multiuser scheduling for downlink MIMO systems with imperfect CSI," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1344–1355, Mar. 2014.

[20] J. Kim *et al.*, "Cooperative distributed beamforming with outdated CSI and channel estimation errors," *IEEE Trans. Commun.*, vol. 62, no. 12, pp. 4269–4280, 2014.

[21] 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.

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

[23] E. N. 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.

[24] 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.

[25] D. Jaramillo-Ramirez *et al.*, "Coordinated multi-point transmission with imperfect CSI and other-cell interference," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 1882–1896, Apr. 2015.

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

[27] S. Guharoy and N. B. Mehta, "Joint evaluation of channel feedback schemes, rate adaptation, and scheduling in OFDMA downlinks with feedback delays," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1719–1731, May 2013.

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

[29] T. Yang *et al.*, "Secure massive MIMO under imperfect CSI: Performance analysis and channel prediction," *IEEE Trans. on Info. Foren. and Secu.*, vol. 14, no. 6, pp. 1610–1623, Jun. 2019.

[30] 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.

[31] S. Silva *et al.*, "Performance analysis of massive MIMO two-way relay networks with pilot contamination, imperfect CSI, and antenna correlation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 4831–4842, Jun. 2018.

[32] J. Liu *et al.*, "Understanding the impacts of limited channel state information on massive MIMO cellular network optimization," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 8, pp. 1715–1727, Aug. 2017.

[33] O. Tervo *et al.*, "The effect of imperfect CSI in a multi-cell multi-user MIMO system," in *Proc. IEEE Vehicular Tech. Conf. (VTC'14 Spring)*, Seoul, South Korea, May 2014.

[34] Y. Isukapalli *et al.*, "Performance analysis of transmit beamforming for MISO systems with imperfect feedback," *IEEE Trans. Commun.*, vol. 57, no. 1, pp. 222–231, Jan. 2009.

[35] Z. Hu *et al.*, "Performance analysis of collaborative beamforming with outdated CSI for multi-relay spectrum sharing networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 11 627–11 641, Dec. 2018.

[36] X. Chen and C. Yuen, "On interference alignment with imperfect CSI: Characterizations of outage probability, ergodic rate and SER," *IEEE Trans. Veh. Technol.*, vol. 65, no. 1, pp. 47–58, Jan. 2016.

[37] H. S. Kang *et al.*, "The degrees of freedom of the interference channel with a cognitive relay under delayed feedback," *IEEE Trans. Inf. Theory*, vol. 63, no. 8, pp. 5299–5313, Aug. 2017.

[38] A. F. Coskun and O. Kucur, "Performance analysis of maximal-ratio transmission/receive antenna selection in Nakagami-m fading channels with channel estimation errors and feedback delay," *IEEE Trans. Veh. Technol.*, vol. 61, no. 3, pp. 1099–1108, Mar. 2012.

[39] G. Lei *et al.*, "Effect of imperfect CSI on STBC-MISO system via antenna selection," *IET Signal Process.*, vol. 10, no. 2, pp. 115–124, Mar. 2016.

[40] J. Choi, "Performance limitation of closed-loop transmit antenna diversity over fast Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 51, no. 4, pp. 771–775, Jul. 2002.

[41] A. Heidari and A. K. Khandani, "Closed-loop transmit diversity with imperfect feedback," *IEEE Trans. Wireless Commun.*, vol. 9, no. 9, pp. 2737–2741, Sep. 2010.

[42] M. Chen *et al.*, "Opportunistic multiple relay selection with outdated channel state information," *IEEE Trans. Veh. Technol.*, vol. 61, no. 3, pp. 1333–1345, Mar. 2012.

[43] S. Kim *et al.*, "Performance analysis of opportunistic relaying scheme with outdated channel information," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 538–549, Feb. 2013.

[44] Q. Cui *et al.*, "Evolution of limited-feedback CoMP systems from 4G to 5G: CoMP features and limited-feedback approaches," *IEEE Veh. Technol. Mag.*, vol. 9, no. 3, pp. 94–103, Sep. 2014.

[45] P. Rost, "Robust and efficient multi-cell cooperation under imperfect CSI and limited backhaul," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1910–1922, Apr. 2013.

[46] S. Ye *et al.*, "Adaptive OFDM systems with imperfect channel state information," *IEEE Trans. Wireless Commun.*, vol. 5, no. 11, pp. 3255–3265, Nov. 2006.

[47] L. Liang *et al.*, "Spectrum and power allocation for vehicular communications with delayed CSI feedback," *IEEE Wireless Commun. Lett.*, vol. 6, no. 4, pp. 458–461, Aug. 2017.

[48] K. Choi and H. Liu, "Power allocation for distributed transmit diversity with feedback loop delay," *IEEE Trans. Commun.*, vol. 59, no. 1, pp. 52–58, Jan. 2011.

[49] 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.

[50] L. Wang *et al.*, "Joint relay and jammer selection improves the physical layer security in the face of CSI feedback delays," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6259–6274, Aug. 2016.

[51] T. Eyceoz *et al.*, "Deterministic channel modeling and long range prediction of fast fading mobile radio channels," *IEEE Commun. Lett.*, vol. 2, no. 9, pp. 254–256, 1998.

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

[53] 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.

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

[55] R. O. Adeogun *et al.*, "Parametric channel prediction for narrowband mobile MIMO systems using spatio-temporal correlation analysis," in *Proc. IEEE Vehicular Tech. Conf. (VTC)*, Las Vegas, USA, Sep. 2013.

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

[57] ——, "Parametric channel prediction for narrowband MIMO systems using polarized antenna arrays," in *Proc. IEEE Vehicular Tech. Conf. (VTC)*, Seoul, South Korea, May 2014.

[58] 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, 2006.

[59] S. Zhou and G. Giannakis, "How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over Rayleigh MIMO channels?" *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1285–1294, 2004.

[60] 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.

[61] S. Prakash and I. McLoughlin, "Effects of channel prediction for transmit antenna selection with maximal-ratio combining in Rayleigh fading," *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2555–2568, Jul. 2011.

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

[63] C. Potter *et al.*, "MIMO channel prediction using recurrent neural networks," in *Proc. Intl. Telemetering Conf.*, San Diego, California, Oct. 2008.

[64] G. Routray and P. Kanungo, "Rayleigh fading MIMO channel prediction using RNN with genetic algorithm," in *Proc. Intl. Conf. on Compu. Intellig. and Inform. Techno.*, Pune, India, Nov. 2011.

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

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

[67] ——, "Long-range fading channel prediction using recurrent neural network," in *Proc. of IEEE CCNC*, Las Vegas, USA, Jan. 2020.

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

[69] C. Potter *et al.*, "MIMO beam-forming with neural network channel prediction trained by a novel PSO-EA-DEPSO algorithm," in *Proc. IEEE Intl. Joint Conf. on Neural Networks (IJCNN)*, Hong Kong, China, Jun. 2008.

[70] T. Ding and A. Hirose, "Fading channel prediction based on complex-valued neural networks in frequency domain," in *Proc. Intl. Symp. on Electo. Theory*, Hiroshima, Japan, May 2013.

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

[72] Y. Sui *et al.*, "Jointly optimized extreme learning machine for short-term prediction of fading channel," *IEEE Access*, vol. 6, pp. 49 029–49 039, Sep. 2018.

[73] X. Tong and S. Sun, "Long short-term memory network for wireless channel prediction," in *Proc. Intl. Conf. on Signal and Inform. Process.*, Chongqing, China, Sep. 2017.

[74] J. Wang *et al.*, "UL-CSI data driven deep learning for predicting DL-CSI in cellular FDD systems," *IEEE Access*, vol. 7, pp. 96 105–96 112, Jul. 2019.

[75] J. Yuan *et al.*, "Machine learning-based channel estimation in massive MIMO with channel aging," in *Proc. of IEEE SPAWC*, Cannes, France, Jul. 2019.

[76] M. Mehrabi *et al.*, "Decision directed channel estimation based on deep neural network k-step predictor for MIMO communications in 5G," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, pp. 2443–2456, Nov. 2019.

[77] H. Ye *et al.*, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018.

[78] M. Arnold *et al.*, "Enabling FDD massive MIMO through deep learning-based channel prediction," *preprint arXiv:1901.036*, Jan. 2019.

[79] Y. Liao *et al.*, "ChanEstNet: A deep learning based channel estimation for high-speed scenarios," in *Proc. of IEEE Intl. Commu. Conf. (ICC)*, Shanghai, China, May 2019.

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

[81] F. A. Gers *et al.*, "Learning to forget: Continual prediction with LSTM," *Neural Computation*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000.

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

[83] 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. United Kindom: John Wiley&Sons and IEEE Press, 2016, ch. 3.

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 Beijing University of Posts and Telecommunications, Beijing, China, in 2008.

From 2008 to 2012, he was a Research Engineer with the 2012 Laboratory, HUAWEI Technologies. From 2012 to 2015, he was a Post-doctoral Research Fellow with Institute of Digital Signal Processing, University of Duisburg-Essen, Germany. Since 2015, he has been a Senior Researcher and Project Manager with German Research Center for Artificial Intelligence (DFKI), which is the biggest European AI research institution and is the birthplace of "Industry 4.0" strategy. Meanwhile, he is 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 papers, holds around 30 granted patents, and participated in a number of research projects, such as EU ABSOLUTE, 5G COHERENT, 5G SELFNET, and German BMBF TACNET4.0 and KICK. His research interests includes B5G/6G and Artificial Intelligence.

Dr. Jiang served as a Vice Chair of the Special Interest Group (SIG) "*Cognitive Radio in 5G*" within the Technical Committee on Cognitive Networks (TCCN) of IEEE Communication Society. He is the Guest Editor for a special issue "Computational Radio Intelligence: A Key for 6G Wireless" in ZTE Communications Journal (Dec. 2019) and is an Associate Editor for IEEE Access.

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 R&D in research and standardization in the area of mobile communications. From 2003 to 2007, he worked for Qualcomm in research and standardization. He became manager of a R&D group, Research Coordinator for Qualcomm Europe, and Director for Technical Standards. In 2007, he accepted the offer to become the full professor and director of the institute for Wireless Communications and Navigation at the University of Kaiserslautern. In 2012, he - in addition - became scientific director of the German Research Center for Artificial Intelligence (DFKI) and head of the department for Intelligent Networks. Professor Schotten served as dean of the department of Electrical Engineering of the University of Kaiserslautern from 2013 until 2017. Since 2018, he is chairman of the German Society for Information Technology and member of the Supervisory Board of the VDE. He is the author of more than 200 papers, filed 13 patents and participated in 30+ European and national collaborative research projects incl. the EU projects WINNER II, C-MOBILE, C-CAST, METIS, METIS II, 5G NORMA, 5G MONARH, 5G SELFNET, and the ITN 5G AURA.