Channel State Information Prediction for Adaptive Underwater Acoustic Downlink OFDMA System: Deep Neural Networks Based Approach

Lei Liu, *Student Member, IEEE*, Lin Cai, *Fellow, IEEE*, Lu Ma, *Member, IEEE*, and Gang Qiao, *Member, IEEE*

***Abstract*—In underwater acoustic (UWA) adaptive communication system, due to time-varying channel, the transmitter often has outdated channel state information (CSI), which results in low efficiency. UWA channels are much more difficult to estimate and predict than terrestrial wireless channels, given the more severe multipath environments with varying propagation speeds in different locations, non-linear propagation paths, several-order higher propagation latency, mobile transceiver and obstacles in the sea, etc. To handle the complexity, this paper proposes an efficient online CSI prediction model for UWA CSI prediction considering the complicated correlationship of UWA channels in both the time and frequency domains. This paper designs a learning model called CsiPreNet, which is an integration of a one-dimensional convolutional neural network (CNN) and a long short term memory (LSTM) network. The performance is compared with the widely used recursive least squares (RLS) predictor, the approximate linear dependency recursive kernel least-squares (ALD-KRLS), and two common conventional deep neural networks (DNN) predictors, i.e., back propagation neural network (BPNN) and LSTM network using the measured data recorded in the South China Sea. To validate the efficacy of prediction, we investigate the prediction of CSI in simulated downlink UWA orthogonal frequency division multiple access (OFDMA) systems. Specifically, the measured UWA channel is used in the OFDMA system. A joint subcarrier-bit-power adaptive allocation scheme is used for resource allocation. To further improve the performance, we develop an offline-online prediction scheme, enabling the prediction results to be more stable. Simulation and experimental results show that the CsiPreNet has superior performance than the existing solutions, thanks to its capability in capturing both the temporal and frequency correlation of the UWA CSIs.**

***Index Terms*—Underwater acoustic (UWA) channel, adaptive communication system, channel state information (CSI) prediction, deep neural network (DNN), orthogonal frequency division multiple access (OFDMA).**

I. Introduction

DUE to the low speed of sound propagation underwater (about 1500 m/s), limited bandwidth, serious multipath and Doppler effects, etc., underwater acoustic (UWA) channels are one of the most challenging communication media [1]. Specifically, UWA channels exhibit larger time dispersions (on the order of hundreds of milliseconds), several-order higher than those with terrestrial radio channels. As such, UWA channels often experience severe inter symbol interference, which requires sophisticated and computationally expensive equalization techniques [2], [3]. Moreover, small-scale phenomena, e.g., scattering caused by surface waves, also contribute to the fast temporal variability of UWA channels [4]. Time-varying multipath propagation and limited bandwidth severely affect the performance of the UWA communication system. To improve spectrum utilization, adaptive communication technology is applied to UWA communication system [5]–[10].

The performance of the adaptive communication systems depend on the knowledge of channel state information (CSI) provided by the feedback of the receiver. Specifically, CSI determines the physical-layer parameters and setting of adaptive UWA communications. For example, the transmitter needs to implement low-order modulation schemes at the physical layer in the case of poor channel status, and vice versa. Obviously, inaccurate CSI can lead to improper modulation schemes which in turn leads to low communication efficiency. Furthermore, CSI also has a significant impact on resource allocation [11]–[13] in adaptive UWA orthogonal frequency division multiple access (OFDMA) systems. Therefore, obtaining accurate CSI is important for improving the performance of the adaptive UWA communication systems.

However, because of the large propagation delay and rapid changing of UWA channel, CSI received by the transmitter is usually outdated. Especially, in the UWA communication

Manuscript received January 23, 2021; revised June 11, 2021; accepted July 13, 2021. Date of publication July 26, 2021; date of current version September 17, 2021. This work is funded in part by the National Natural Science Foundation of China (NSFC) under Grants U1806201, the Young Elite Scientists Sponsorship by CAST and NSFC under Grant 61771152 and 11774074, by the National Key Research and Development Program of China under Grant 2018YFC0308500, by the Sustainable Funding of the Key Laboratory of Underwater Acoustic Technology under Grant JCKYS2020604SSJS015, by the Special Fund Project for Fundamental Scientific Research Expenses of Central Universities under Grant 3072020CFJ0506, and by the Natural Sciences and Engineering Research Council of Canada (NSERC). The review of this article was coordinated by Dr. Namyoon Lee. *(Corresponding author: Lu Ma.)*

Lei Liu, Lu Ma, and Gang Qiao are with the Acoustic Science and Technology Laboratory, the Key Laboratory of Marine Information Acquisition and Security, Ministry of Industry and Information Technology, College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China (e-mail: heu_liulei@hotmail.com; malu@hrbeu.edu.cn; qiaogang@hrbeu.edu.cn).

Lin Cai is with the Acoustic Science and Technology Laboratory, the Key Laboratory of Marine Information Acquisition and Security, Ministry of Industry and Information Technology, College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China, and also with the Department of Electrical and Computer Engineering, University of Victoria, Victoria V8P 5C2, Canada (e-mail: cai@ece.uvic.ca).

Digital Object Identifier 10.1109/TVT.2021.3099797

networks, the central node needs to take more time to obtain the CSIs from all users. This makes the problem of outdated CSI even more severe. In recent years, many researches have studied the performance of adaptive UWA communication systems using outdated CSI. In [14], an adaptive UWA communication scheme for multiantenna transmissions based on the partial CSI was proposed. In [12], long-term statistics of CSI were considered for channel feedback to mitigate the impact of outdated CSI. In [13], the CSI used for feedback was selected between instantaneous CSI and average CSI for system performance improvement. In addition to the extensive research on how to deal with the negative impact of the outdated CSIs, it is found that channel prediction is a desirable method to fully address the problem. Since the speed of sound is very low, the ability to predict the CSI at least one transmission round ahead highly affects the performance of the adaptive communication systems. It is indeed challenging for UWA communications in the range of several kilometers, which significant limits the use of feedback.

With a reasonable tracking ability and simple design, a linear predictor applying the recursive least-squares (RLS) algorithm has been widely used for UWA channel impulse responses (CIRs) prediction [8], [15]–[18]. For seasonal UWA time-varying channels, autoregressive (AR) processs [20] and Holt-Winters [21], [24] were introduced to model the seasonal correlation. An efficient adaptive predictor operating in the delay-Doppler domain for UWA time-varying channels was proposed in [25], which does not require any prior knowledge of channel dynamic model and noise statistics.

For adaptive UWA OFDM and OFDMA systems, existing channel predictors were often realized at several significant channel taps in the time domain [8], [17]–[19]. Few researchers applied the predictor to each OFDM subcarrier in the frequency domain. This is because the variation at each subcarrier is a combined variation of multiple taps, which makes it difficult to predict using a linear predictor. Moreover, prediction on a large number of OFDM subcarriers leads to high complexity. The traditional linear predictor can not handle frequency domain prediction well. Kernel adaptive filtering (KAF) methods are rapidly gaining popularity to solve a wide variety of prediction, identification and regression problems, since they provide state-of-the-art performance in many real-world applications [22]. In particular, the kernel RLS (KRLS) algorithm based on the RLS algorithm is one of the most popular algorithms in KAF [23], and shows good performance in channel prediction [22]. The rapid development of deep neural networks (DNNs) [26], [27], [31], [32] also makes it possible to perform channel prediction in the frequency domain. As long as there are sufficient data to train the neural network, DNNs can capture the complex correlation between different data in both the time and frequency domains. In the past thirty years, DNNs have been widely applied in wireless communications [28]. Reasonable application of DNNs can effectively optimize wireless communication and network system [29]. DNNs will be an indispensable tool for the design and operation of future wireless communication networks [30]. It convinces us that DNNs will also be promising for future research in UWA communications and networks. In wireless communications, many researches focused on channel prediction based on DNNs [33]–[37]. Although some studies have applied DNNs to UWA communications [38], to the best of our knowledge, no research ever used DNNs for channel prediction in UWA communication systems.

UWA channels are much more difficult to predict than terrestrial wireless channels, given the more severe multipath environments with varying propagation speeds in different locations, non-linear propagation paths, several-order higher propagation latency, mobile transceiver and obstacles in the sea, etc. The existing CSI prediction approaches, e.g., the RLS/KRLS-based prediction methods, only consider the amplitude of the channel tap, and do not consider the variety of the phase. However, when predicting UWA channel, the variety of channel delay cannot be ignored. In addition, due to the requirement of low power consumption of UWA communication hardware design, its computing power is far less than that of wireless communication system, which makes the traditional CSI prediction approaches not fully applicable for UWA channel prediction. In UWA networks, such as the Internet of underwater things (IoUT) with extremely large-scale deployment of sensor nodes [39], the above difficulties are more serious. To handle the complexity, in this paper, we propose an DNN-based efficient online CSI prediction model for CSI prediction and study the performance according to measured UWA channels in South China Sea. To validate the efficacy of prediction, we investigate the prediction of the measured channels in the simulated UWA OFDMA system. Furthermore, an offline-online prediction scheme is developed to improve system performance. The main contributions of this paper are three-fold.

- We design a learning model called CsiPreNet for UWA CSI prediction, that is a combination of a one-dimensional convolutional neural network (CNN) and a long short term memory (LSTM) network. The proposed CsiPreNet can exploit the complicated frequency-temporal correlationship of UWA channels to conduct the CSI prediction effectively. The prediction performance is evaluated with measured channel data recorded in South China Sea. The proposed CsiPreNet has excellent performance on UWA CSI prediction, based on its ability to capture the information in both the frequency and time domains.
- We design a simulated downlink UWA OFDMA system, and investigate the impact of CSI prediction. Specifically, the measured UWA channels are used in the simulated OFDMA system as ground truth. A joint subcarrier-bit-power adaptive allocation scheme is used for resource allocation by using the predicted CSI for feedback.
- We design an offline-online prediction scheme to improve the stability of the developed learning models when applying it to the simulated OFDMA system. A large amount of historical CSI is input into the offline part to train the learning models, and the trained model is applied to the online part for CSI prediction. The predicted CSI is used as feedback for resource allocation. Furthermore, we update the training data of the offline part after several transmission rounds to ensure the stability of the learning models. Simulation results demonstrate the superior performance of the proposed offline-online prediction based on CsiPreNet model.

The rest of this paper is organized as follows. Section II introduces the related work. In Section III, we describe the downlink UWA OFDMA system, the UWA channel model, and the measured data recorded in sea test. In Section IV, we propose the BPNN, LSTM, and CsiPreNet prediction models. In Section V, we present the joint subcarrier-bit-power allocation scheme, the limited feedback for UWA OFDMA system, and the proposed offline-online prediction scheme. In Section VI, we analyze the prediction performance of the three DNN prediction models using measured CSI, and apply these three models to the simulated downlink UWA OFDMA system for performance evaluation, and then analyze the computational complexity and the running time of all models, followed by the concluding remarks and further research issues in Section VII.

II. Related Work

Due to the fast changing nature and the extremely limited bandwidth of the UWA channels, an outdated CSI can lead to severe performance degradation, so accurate channel prediction is critical. Although many scholars have studied channel prediction in terrestrial wireless communications [33]–[37], [40], [41], [43], the prediction of UWA channels has received little attention until recent years [8], [15]–[21], [24], [25].

The prediction methods are generally divided into two categories, i.e., model-based and non-model-based. The model-based methods assume that certain knowledge and variation about the channel are known, then the correctly defined model can improve the prediction accuracy. For example, the channel variations and parameters were based on an AR model and tracked by an extended Kalman filter in [40], and it assumed that the channel tap coefficients are uncorrelated. By contrast, in [41], it used the RLS algorithm with a postfilter for channel tracking, and considered the multipath correlation. If the real channel does not match the assumed model, the performance of the above-mentioned methods can not be satisfactory. On the contrary, the non-model-based adaptive prediction methods do not rely on the prior knowledge of the channel. Thus it is more suitable for the real communication systems [8], [15]–[21], [24], [25]. Generally speaking, the adaptive predictors can adaptively track channel variations and adjust itself. Least Mean Square (LMS) and RLS are two widely used adaptive algorithms for CSI prediction [8], [15]–[19], [43]. Compared with LMS algorithm, RLS algorithm has the better tracking ability at the cost of a higher computational complexity [42]. The common method using RLS predictor for time domain prediction is shown in Fig. 1 [18], [43]. However, these studies only considered the variety of channel taps and assumed that the delay of channel taps is stable for a certain period of time.

On the other hand, the rapid development of DNNs makes it possible to capture complex channel variations accurately. In wireless communications, DNNs has been used to predict channel [33]–[37]. It inspires us to apply DNNs to channel prediction in UWA communication systems, which have drastically different channel characteristics.

Obviously, combining the channel prediction with the adaptive UWA communication system can significantly improve the system performance. In [8], RLS predictor was embedded in an adaptive UWA OFDM system. In [19], a precoding based channel prediction scheme was designed for UWA OFDM system. In [25], the adaptive delay-Doppler spreading function (DDSF) prediction scheme was proposed, and used for UWA communication system under several simulated UWA channels. There is currently no deep learning based prediction model for adaptive UWA communication systems. This work is motivated to fill the gap.

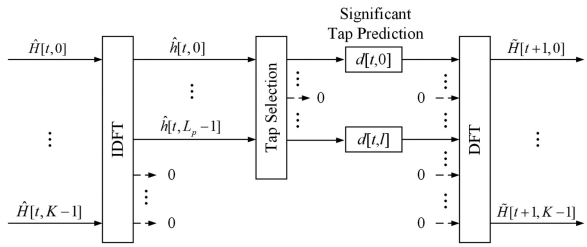

Fig. 1. Time domain predictor for sparse channel based on RLS algorithm.

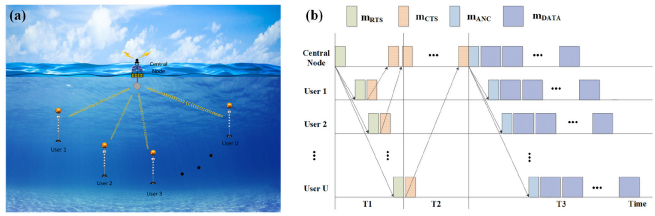

Fig. 2. System model: (a) Downlink UWA OFDMA system, (b) OFDMA handshaking process.

III. System Model and Dataset Description

The UWA channel is time-spatial correlated, but in a very complicated way. Due to the characteristics of the long propagation delay (in seconds), curved propagation ray due to different acoustic velocity in different height, relative motions and changing propagation multi-paths, lack of synchronization in UWA channels [1], the parameter design and experiment setup for UWA communication is important.

A. Downlink UWA OFDMA System

As shown in Fig. 2(a), we consider a simulated downlink UWA OFDMA system with U spatially separated users and a central node.

Consider an OFDMA setup, where a total number of K subcarriers are allocated to all users. K_u non-overlapping subcarriers are allocated to user u, where $\sum_{u=1}^{U} K_u = K$. The OFDM symbol period is T, the cyclic prefix (CP) length is T_{cp}, and the subcarrier interval is $1/T$. The center frequency is f_c, and then the subcarrier frequency $f_k = f_c + k/T, k = -K/2, \ldots, K/2 - 1$. Define $d[k]$ as the coded information on the k-th subcarrier, and then the transmitted signal is

$$x_u(t) = \mathrm{Re}\left\{\sum_{k\in S} d[k] exp(j2\pi f_k t)\right\}, t \in [0, T], \quad (1)$$

where S is the subcarrier index set, including the data subcarrier index set S_D and the comb pilot index set S_P, $S = S_D \cup S_P$. All users use the same pilot symbols for channel estimation, and the k-th subcarrier can only be allocated to one user, where $k \in S_D$. The candidate modulation schemes of $d[k]$ are binary phase shift keying (BPSK), quadrature phase-shift keying (QPSK), 8-quadrature amplitude modulation (8QAM), and 16-quadrature amplitude modulation (16QAM) with 2-D Gray mapping. In other words, for the k-th subcarrier, where $k \in S_D$, the modulation level $M_k \in \{2, 4, 8, 16\}$, and if no data are transmitted, $M_k = 1$. We assumed that the pilot symbols ($k \in S_P$) use the modulation scheme of QPSK.

Assuming that the receiver (user u) has correctly compensated for the Doppler frequency offset caused by the relative motion, the time-varying UWA multipath channel model transmitted from the central node to user u in a CP-OFDM block can be expressed as

$$h_u(\tau, t) = \sum_{p=1}^{N_{p,u}} A_{p,u}(t)\delta(\tau - \tau_{p,u}(t)), \tag{2}$$

where $N_{p,u}$ is the number of channel paths from the central node to user u; $A_{p,u}(t)$ is the attenuation coefficient of the p-th path in a CP-OFDM block; $\tau_{p,u}(t)$ is the time delay corresponding to the p-th path. Assuming that the CP length T_{cp} is longer than the maximum multipath delay, the signal received by user u is

$$y(t) = \sum_{p=1}^{N_{p,u}} A_{p,u}(t)x_u(t - \tau_{p,u}(t)) + w(t), \tag{3}$$

where $w(t)$ is the additive noise. Substituting (1) into (3), after removing CP and undergoing discrete fourier transform (DFT) transformation, the frequency domain baseband reception vector of user u on the k-th subcarrier can be obtained by

$$z_u[k] = H_u[k]d_u[k] + v_u[k], \tag{4}$$

where $z_u[k]$ and $d_u[k]$ represent the received data and the transmitted data on the k-th subcarrier allocated to user u, respectively. $H_u[k]$ represents the channel gain of user u on the k-th subcarrier. $v_u[k]$ is the frequency domain additive noise in the frequency domain. CSI can be measured in the form of Signal-to-noise ratio (SNR), which can be expressed as

$$\gamma_u[k] = \frac{|H_u[k]|^2\sigma_s^2}{\sigma_u^2}, \tag{5}$$

where σ_s^2 is the symbol energy, and σ_u^2 is the noise variance of user u.

Therefore, based on the knowledge of the transmitted and received signals, i.e., $d[k]$ and $z[k]$, the CSI can be estimated. So far, scholars have developed several CSI estimation methods, e.g., maximum likelihood (ML), minimum mean square error (MMSE), least squares (LS), and compressed sensing (CS) methods. In this paper, we use the orthogonal matching pursuit (OMP) algorithm [44] for CSI estimation, which is one of the CS methods.

Furthermore, the handshaking process of this adaptive downlink UWA OFDMA system is shown in Fig. 2(b). The CSI of each user is estimated by the request-to-send (RTS) packet, and then embedded into the clear-to-send (CTS) packet. The central node collects all users' CSIs, and performs the adaptive resource allocation scheme based on them. Furthermore, the allocation table is embedded in the announcement (ANC) packet, and broadcasted with the DATA packet to all users. Upon reception, each user decodes its own data based on the allocation table. Therefore, the CSI embedded in a CTS packet is the CSI based on the RTS packet in slot T1, which outdates while using it for resource allocation in slot T3, under time-varying UWA channel.

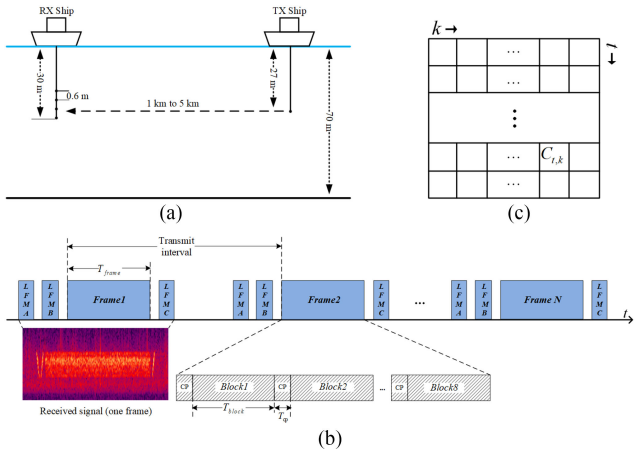

Fig. 3. Experimental setup: (a) The geometry and the setup of the communication system. (b) Structure of one frame OFDM signal. (c) Organized CSI image.

B. Experimental Dataset Description

We carried out the sea trial in Lingshui Bay, South China Sea on May 8, 2014. In the experiment, two ships were used for communication. The sea depth was 70 meters. One ship was the transmitter with a transducer deployed 27 meters depth. Another ship was the receiver with a four-element line array deployed at the depth of 30 meters, and array element spacing was 0.6 meters. Transmission distance ranged from 1 km to 5 km. The geometry of the experiment and the setup of the system are given in Fig. 3(a).

As shown in Fig. 3(b), the transmitter sends frames of CP-OFDM blocks. Each frame contains 8 OFDM blocks. One OFDM block has 681 subcarriers, which contains 595 data subcarriers and 86 pilot subcarriers, and the pilot interval is 8. CSI is estimated from the pilot subcarriers. CP length, block length and frame length are T_{cp}=25 milliseconds (ms), T_{block}=171 ms, and T_{frame}=1.568 seconds (s), respectively. The transmission interval between two adjacent frames is 4 s. The bandwidth is 4 kHz (kHz), and the central frequency is 8 kHz. LFM A signal is used for signal synchronization, while LFM B signal and LFM C signal are the same signal that are used for Doppler estimation.

The channel of each block can be estimated by the pilot signal in the block. The time interval between two channels of the same block in adjacent frames is 4 s. We record a large amount of measured UWA CSI. We organize the CSI into a two-dimensional image as shown in Fig. 3(c). One dimension is the frequency dimension, representing the k-th subcarrier, and another dimension is the time dimension, representing the time t.

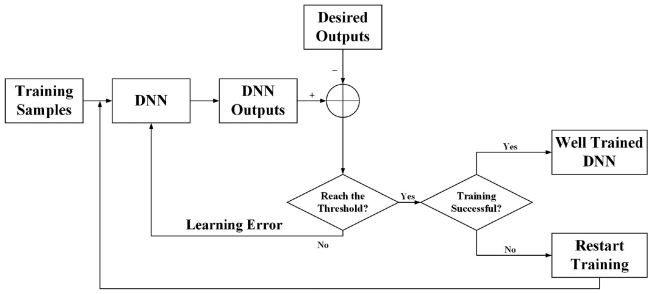

Fig. 4. Training process of DNN model.

We can also express the CSI image as a matrix

$$CSI_{img} = [\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_t, \ldots, \mathbf{C}_T]^T, \tag{6}$$

$$\mathbf{C}_t = [C_{t,1}, C_{t,2}, \ldots, C_{t,k}, \ldots, C_{t,K}], \tag{7}$$

where $C_{t,k}$ represents the CSI of the k-th subcarrier at the time t.

This paper selects two different measured channels with a transmission distance of 3 km and 5 km for analysis. In this case, the two ships anchored at a distance of 3 km and 5 km, respectively. But the ships still drift randomly in a small range with the water surface waves, resulting in time-varying of the channel. The analytical results will be presented in Section VI.

IV. Prediction Model Based on DNN

The long propagation delay makes it critical to conduct accurate channel prediction, while the complicated mobile UWA environment makes the traditional solutions less effective. Therefore, we investigate the deep-learning approach for UWA channel prediction.

In this section, we will introduce three DNN models, i.e., BPNN, LSTM, and the proposed CsiPreNet. We use these three models for CSI prediction. Fig. 4 shows the training process of a typic DNN. Generally speaking, a DNN prediction model consists of two parts, the training part and the prediction part. Before the prediction part works, the network must be trained by the training symbols in the training part. First, the training samples are input into the neurons in the hidden layers and then are propagated to the output layer. If the learning errors between the DNN output results and the desired outputs are not reached the training threshold, the learning errors will be back-propagated from the output layer to the hidden layers. Meanwhile, the weights and biases in the neurons will be updated according to the learning errors. The training part will be repeated until the training threshold are reached, such as a given number of training rounds have been completed, or a certain learning error has been met. Finally, the well-trained model is used for CSI prediction.

A. Back Propagation Neural Network (BPNN)

BPNN can be described as a multilayer perceptron (MLP) neural network with multiple hidden layers trained by the back propagation algorithm [45], [46]. Fig. 5 illustrates a typical

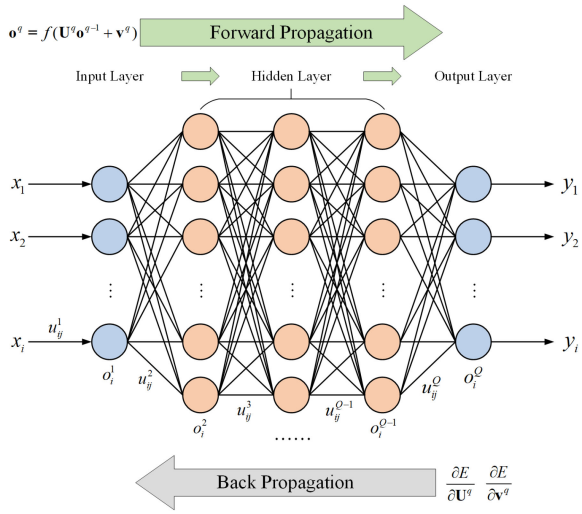

Fig. 5. Typical architecture of BPNN model.

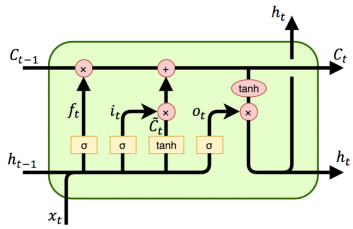

Fig. 6. Typical architecture of LSTM cell.

BPNN, which includes an input layer, multiple hidden layers, and an output layer.

Then the back propagation algorithm [45], [46] is used to update the weights $\mathbf{U}$ and biases $\mathbf{v}$ in the neurons according to the learning errors. As shown in Fig. 5, the training part of the BPNN can be summarized as the forward propagation of training symbols and the back propagation of updating errors.

The update of BPNN can be briefly summarized as follows

$$\mathbf{y} = \text{BPNN}(\mathbf{x}, P), \tag{8}$$

where BPNN$(\cdot)$ presents all propagation process of BPNN, and P represents all the parameters in the BPNN.

For the CSI prediction, we replace $\mathbf{x}$ and $\mathbf{y}$ with $\mathbf{C}_t, \mathbf{C}_{t-1}, \ldots, \mathbf{C}_{t-n}$ and $\mathbf{C}_{t+1}$ defined in (7), respectively. The prediction expression can be expressed as

$$\mathbf{C}_{t+1} = \text{BPNN}(\mathbf{C}_t, \mathbf{C}_{t-1}, \ldots, \mathbf{C}_{t-n}, P). \tag{9}$$

B. Long Short Term Memory (LSTM)

The LSTM network is explicitly designed to handle the long-term dependency problem, which is suitable for time series prediction [48]. LSTM network uses the approximate gradient calculation algorithm, which updates the weight matrices after each time step. However, the back-propagation algorithm can be used to calculate the entire gradient.

In a typical LSTM network, each cell has four gates, interacting in a special way. Fig. 6 illustrates a single LSTM memory cell.

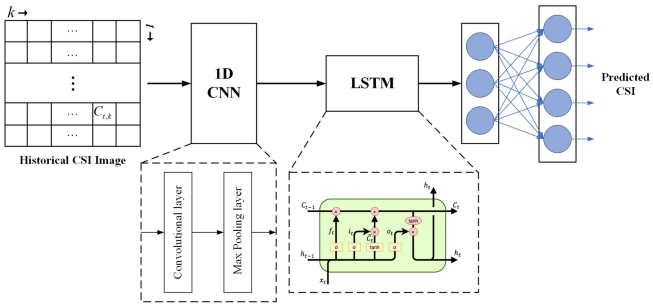

Fig. 7. Structure of CsiPreNet.

The same as the BPNN model, we can summarize the update of each LSTM unit as

$$\mathbf{h}_t = \text{LSTM}(\mathbf{h}_{t-1}, \mathbf{x}_t, P), \tag{10}$$

where LSTM$(\cdot)$ presents all propagation process of LSTM, and P refers to all the parameters in the LSTM network.

Similar to (9), the update of each LSTM unit for CSI prediction can be expressed as

$$\mathbf{h}_t = \text{LSTM}(\mathbf{h}_{t-1}, \mathbf{C}_t, P). \tag{11}$$

At last, a fully-connected softmax layer is used to generate a probability distribution of the prediction result as follows

$$\hat{\mathbf{S}}_t = \text{softmax}(\mathbf{U}\mathbf{h}_t + \mathbf{v}). \tag{12}$$

C. Csiprenet

Considering the correlations in the frequency, time, and space domains, we develop a learning model called CsiPreNet, a combination of the CNN [49] and the LSTM, to exploit the frequency-temporal relationship of the CSI. Fig. 7 shows the architecture of this learning model, which consists of a 1D CNN model, and an LSTM model. The 1D CNN model is used to extract the frequency correlation from the CSI between adjacent subcarriers, which includes a 1D convolutional layer, a 1D max-pooling layer, and a flatten layer. Followed by an LSTM model that is used to extract temporal correlation and perform prediction.

CNN has excellent performance in image processing. This inspired us to organize the raw CSI data in a similar way with the image. Specifically, the CSI on each subcarrier can act as a pixel of the image. The organized CSI image data is described in Section III-B. As shown in Fig. 7, the k-th row, t-th column pixel represents the CSI of the k-th subcarrier at time t.

To capture the frequency correlation, we propose a 1D CNN model to process the CSI first. Generally, a CNN model has several parallel filters for processing the CSI images by a set of weights. We describe the CNN model with one filter for simple. $\mathbf{U}$ is defined as the weight of a convolutional filter. The filter strides through the CSI images along one dimension to calculate the convolutional result. After a set of CSIs $\mathbf{C}$ pass the filter, the convolutional result $\mathbf{C}^*$ can be calculated as follows

$$\mathbf{C}^* = f(\mathbf{C} \otimes \mathbf{U} + \mathbf{v}), \tag{13}$$

where $\mathbf{v}$ is the bias and $f(\cdot)$ is the activation function. Then a 1D max-pooling layer is used to capture the most salient features in $\mathbf{C}^*$ as follows

$$C^* = \max(\mathbf{C}^*). \tag{14}$$

The filters go through each time line of the CSIs to obtain the feature image CSI^*_{img} as follows

$$CSI^*_{img} = [C^*_{11}, C^*_{12}, \ldots, C^*_{tn}, \ldots, C^*_{T1}, C^*_{T2}, \ldots, C^*_{Tn}], \tag{15}$$

where n is the number of filters and t is the time.

Furthermore, a flatten layer is used to combine and flatten the output results of all filters into a vector as follows

$$\mathbf{C}^*_t = \text{flatten}(C^*_{t1}, C^*_{t2}, \ldots, C^*_{tn}), \tag{16}$$

where flatten$(\cdot)$ represents the flattening process.

We use conv1D$(\cdot)$ to represent the entire 1D convolution process. Thus the 1D CNN model can be briefly described as follows

$$[\mathbf{C}^*_1, \mathbf{C}^*_2, \ldots, \mathbf{C}^*_T] = \text{conv1D}(CSI_{img}). \tag{17}$$

Then, an LSTM model is used to predict the current vector, since it is good at learning long-term dependencies. Based on (11), we can obtain the output values of the LSTM model as shown below

$$\mathbf{h}_t = \text{LSTM}(\mathbf{h}_{t-1}, \mathbf{C}^*_t, P). \tag{18}$$

Same as (12), a fully-connected softmax layer is followed

$$\hat{\mathbf{S}}_t = \text{softmax}(\mathbf{U}\mathbf{h}_t + \mathbf{v}). \tag{19}$$

To sum up, BPNN can be described as an MLP, and uses the backpropagation algorithm to update network parameters, which is the most basic and widely used DNN model. Since the ability of learning long-term dependencies, LSTM can better handle time series and solve the problem of vanishing gradient. The CsiPreNet combines the advantages of CNN in spatial information processing and the advantages of LSTM in temporal information processing. We will evaluate the performance of CsiPreNet in Section VI.

V. Adaptive Resource Allocation Based on Predicted CSI

In this section, we introduced how to combine the DNN based CSI prediction model with a real-time adaptive UWA communication system. Since the training time of DNN is long, we train the model offline, and the trained model is used for CSI prediction online. Meanwhile, a limited feedback mechanism is used to feed back the predicted CSI.

A. Joint Subcarrier-Bit-Power Allocation Scheme

Considering the computational complexity, power consumption and system performance, as well as the particularity of UWA communication environment, a joint subcarrier-bit-power allocation scheme is used for resource allocation in the proposed OFDMA system, where users take turns to be assigned an additional bit which consumes the least power at each iteration round until the data rate requirement is satisfied.

- $\delta_u[k]$ is defined as the subcarrier allocation factor, which means that the k-th subcarrier is allocated to the user u. $\delta_u[k] = 1$ represents that the k-th subcarrier is allocated to the user u, otherwise, $\delta_u[k] = 0$.
- $b_u[k]$ is defined as the bit loading factor, which represents that a certain number of bits are loaded on the k-th subcarrier for the user u, and determines the modulation constellation.
- $\beta_u[k]$ is defined as the power scaling factor, which determines the transmission power on the k-th subcarrier for user u. If there is no variation in power scaling, $\beta_u[k] = 1$.

Assuming that one data subcarrier can only be allocated to one user, we have

$$\sum_{u=1}^{U} \delta_u[k] = 1, \tag{20}$$

$$\sum_{u=1}^{U} \sum_{k \in S_D} \delta_u[k] = K_D. \tag{21}$$

All subcarriers allocated to user u can be expressed as

$$S_u = \{k | \delta_u[k] = 1, k \in S_D\}. \tag{22}$$

The data rate of user u is

$$R_u = \sum_{k \in S_D} \delta_u[k] b_u[k] = \sum_{k \in S_u} b_u[k]. \tag{23}$$

Thus the total data rate of the OFDMA system can be expressed as

$$R = \sum\nolimits_{u=1}^{U} R_u. \tag{24}$$

Consider the power scaling factor, (4) can be written as follows [9]

$$z_u[k] = H_u[k] \sqrt{\beta_u[k]} d_u[k] + v_u[k], \quad k \in S_u. \tag{25}$$

Then CSI $\gamma_u[k]$ can be obtained by (5).

The total transmission power can be expressed as

$$P_{total} = \sigma_s^2 \sum_{u=1}^{U} \sum_{k=1}^{K} \delta_u[k] \beta_u[k]. \tag{26}$$

Different modulation levels are adopted and the constellation mapping factor $g(\cdot)$ can be written as [50]

$$g(b) = \begin{cases} \frac{6}{5 \times 2^b - 4}, b = 1, 3, 5 \ldots \\ \frac{6}{4 \times 2^b - 4}, b = 2, 4, 6 \ldots \end{cases}. \tag{27}$$

The resource allocation algorithm is described in detail in Algorithm 1. The first process is to initialize the parameters to appropriate values. The power required to load the first bit on the k-th subcarrier of the user u is calculated in (28). The second process is to iteratively allocate subcarriers and bits to each user. In each iteration, each user takes turns loading one bit with the lowest power consumption to a subcarrier that is not occupied by other users. The final step is to calculate the power scaling factor, which determines the transmission power loaded on each subcarrier.

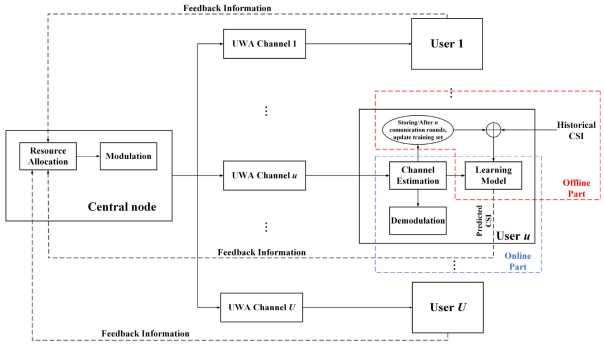

Fig. 8. Structure of online-offline prediction model in adaptive downlink UWA OFDMA system.

B. Limited Feedback for Adaptive UWA Systems

We assume that a limited-feedback channel is available for conveying information from all users to the central node. Each user has the knowledge of the CSI from the central node to itself.

We consider to feed back the CSI of K subcarriers, where $K = 681$. If we feed back the CSI of all 681 subcarriers, this will bring a large feedback overhead. If the CSI changes slowly across frequencies, we can cluster multiple adjacent subcarriers and share the same suitable CSI. In such a case, it is not necessary to feed back the CSI for each subcarrier.

We cluster 20 subcarriers into one cluster, so we only need to feed back the CSI of 34 subcarrier clusters (the last 21 subcarriers are clustered into one cluster). We calculated the average CSI of each cluster of subcarriers to represent the CSI of all subcarriers in this cluster. Hence, the total number of bits fed back can be reduced by 20 times. Moreover, a 16-order uniform quantization algorithm is used to perform CSI quantization [51].

C. Offline-Online Prediction Scheme

Because the DNN training process requires a lot of time, we need to train the model in advance, and then the trained model will be directly used for CSI prediction. Due to the rapid time-varying of UWA channel, the trained model may be outdated after several transmission rounds. To solve this problem, we develop an offline-online prediction scheme for the adaptive OFDMA system to improve the stability of the DNN models described in Section IV, as shown in Fig. 8. The offline part is to train, store, and update the historical CSI, and the online part is to integrate CSI prediction and measurement. When a CSI prediction request arrives, the online part first collects real-time received signal, then estimates the CSI, and finally inputs the CSI into the well-trained DNN models to perform the CSI prediction. In the following, we describe the offline part and online part in detail.

For the offline part, we first use historical CSI data to train the learning model, and the well-trained model will be used for CSI prediction in the online part. In the communication process of the OFDMA system, the offline part continuously stores the measured CSIs estimated by the user in the online part, and updates the training data in the learning model after several transmission rounds to retrain the model.

Algorithm 1: Joint Subcarrier-Bit-Power Allocation Scheme.

Input: CSI $\gamma_u[k]$;
Output: Subcarrier allocation factor $\delta_u[k]$, bit loading factor $b_u[k]$, power scaling factor $\beta_u[k]$;
procedure Initialization parameters
  Set R_u and P_E to the target values R_{target} and P_{target};
  Set $b_u[k]$ and $P_u[k]$ to 0;
  for u from 1 to U **do**
    for k from 1 to K_d **do**
      Calculate the power consumption of loading the first bit

$$\Delta P_u[k] = \frac{-\ln(5P_{target})}{g(1)\gamma_u[k]}; \tag{28}$$

    end for
  end for
end procedure
procedure Subcarrier allocation and bit loading
  for i from 1 to R_{target} **do**
    for u from 1 to U **do**
      Find the subcarrier with the least power consumption for user u and load one bit

$$S_u = \{k|\delta_{\bar{u}}[k] = 0, \forall \bar{u} \neq u\}; \tag{29}$$

$$\hat{k} = \{k|\min\{\Delta P_u[k], k \in S_u\}\}; \tag{30}$$

$$\delta_u[\hat{k}] = 1; b_u[\hat{k}] = b_u[\hat{k}] + 1; \tag{31}$$

$$P_u[\hat{k}] = P_u[\hat{k}] + \Delta P_u[\hat{k}]; \tag{32}$$

      Calculate the power consumption of loading an additional bit

$$\Delta P_u[\hat{k}] = \frac{-\ln(5P_{target})}{\gamma_u[\hat{k}]} \left(\frac{1}{g(b_u[\hat{k}]+1)} - \frac{1}{g(b_u[\hat{k}])} \right); \tag{33}$$

    end for
  end for
end procedure
procedure Power scaling
  for u from 1 to U **do**
    for k from 1 to K_d **do**
      Power scaling factor calculation

$$\beta_u[k] = \frac{\delta_u[k]P_u[k]}{\sum_u \sum_k \delta_u[k]P_u[k]} K_d; \tag{34}$$

    end for
  end for
  return $\{\delta_u[k], b_u[k], \beta_u[k]\}_{\forall u, \forall k}$
end procedure

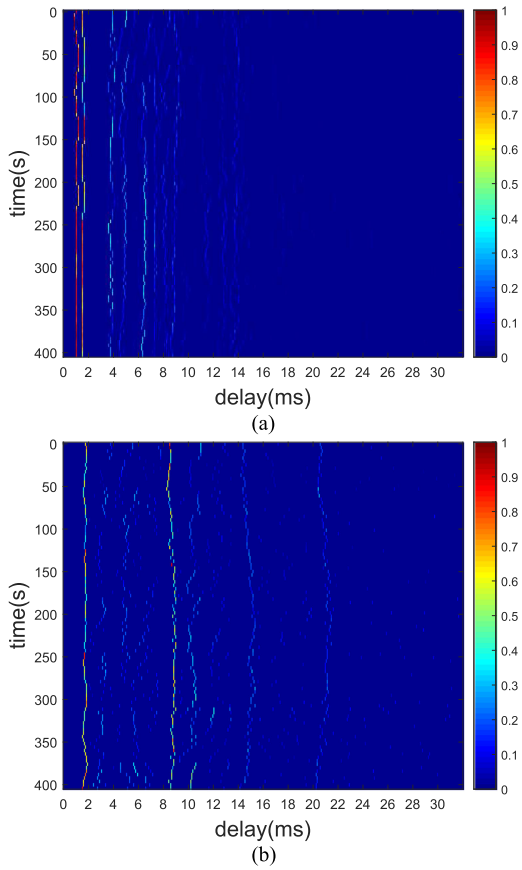

Fig. 9. Measured time-varying channel impulse response of (a) Channel A and (b) Channel B.

For the online part, each user obtains real-time received CSI through channel estimation and store these CSI in the offline part. At the same time, CSI prediction is performed by the well-trained model. Finally, the predicted CSI is fed back to the central node for resource allocation.

VI. Performance Evaluation

In this section, we present numerical results of the three proposed DNN, the RLS-based and ALD-KRLS-based predictors, simulation results on the performance of the proposed adaptive OFDMA system, and experimental results on the measured data. The numerical and experimental results are based on the channel measurement data described in Section III-B, and the simulation results are from the simulated OFDMA system designed in Section III-A. Channel measurements are used in the simulated OFDMA system, and the offline-online prediction scheme described in Section V-C is embedded in the CSI prediction model. Two different measured time-varying channels are shown in Fig. 9. Fig. 9(a) and Fig. 9(b) show the time-varying UWA channel with a transmission distance of 3 km (Channel A) and a transmission distance of 5 km (Channel B), respectively. The vertical axis is time and the horizontal axis is time delay. The amplitude and delay of channel tap are time-varying. Specifically, UWA channel impulse responses (CIRs) often have the

sparse structure in the sense that most paths can be neglected due to overly low energy leaving only few dominant paths to be considered [52].

A. Parameter Settings of the Prediction Models

We deployed the BPNN, LSTM, and proposed CsiPreNet models for training and predictive analysis. RLS predictor proposed in [8], [15]–[19] is a benchmark for performance comparison. ALD-KRLS predictor [23] is also applied to predict the UWA channel. For RLS, the order is set to 8 and the forgetting factor is set to 0.75. The sensitivity threshold ν of the ALD-KRLS is set to 0.000 001, and a Gaussian kernel with $\sigma = 5.1$ is used. The prediction models were trained according to the measured CSI. We have 10 584 samples of Channel A and 4448 samples of Channel B. Each sample has 8 data flows and 1 data label, i.e, we used 8 time-step CSIs to predict the next one. 70% of the CSI data were allocated to the training phase, and the remaining part were used for the test and generalization. All the DNN frameworks were implemented in Python, and "Keras" [53] library was used to build the three DNN models' architecture.

For the BPNN model, two hidden layers containing 64 and 256 neurons were used to extract the frequency and temporal correlation in CSI. For the LSTM model, the stacked LSTM layers were exploited for the CSI prediction, which consisted of two hidden layers, containing 64 and 256 units, respectively. A dropout of 0.5 was applied to the two hidden layers of the model to avoid over-fitting. For the CsiPreNet, two one-dimensional convolutional layers with 40 filters (filter size is 1×2) were used to extract frequency information, followed by the LSTM layers, which consisted two hidden layers for temporal prediction, and the two hidden layers contained 64 and 256 units, respectively. A dropout of 0.5 was applied to the two hidden layers of the model to avoid over-fitting. The learning rate and batch size of all DNN models were 0.0005 and 128, respectively. The loss function was the mean absolute error (MAE). The activation function and the optimization algorithm that were used in the training phase are "relu" and "adam," respectively. Based on the size of CSI images, the size of input layer and output layer for all DNN models are $8 \times 34 \times 1$ and 34×1, respectively.

The hyper-parameter optimization algorithm used in this paper is random search [54], and each hyper-parameter was tested by "RandomizedSearchCV" function in "sklearn" [55] library. Each parameter was tested with a set of different values, and the best performance value was selected.

B. Prediction Results

We put the training samples of CSI into the DNN models. After multiple rounds of training, we obtained the MAE loss and training rounds, as shown in Fig. 10.

Through comparative experiments, in both the training set and the validation set, the CsiPreNet achieved a faster convergence rate and less losses than the BPNN and LSTM models for both Channel A and B. The performance of the BPNN is the worst. In Fig. 10(a), the train loss curve decreases as the training rounds increases, but the validation loss curve firstly decreases as the training rounds increases, and reaches the lowest value when the training rounds in the range from 50 to 100, and then increases. In Fig. 10(b), the train loss curve decreases as the training rounds increases, but the validation loss curve firstly decreases as the training rounds increases, and when the training rounds exceed 100, it stays stable. Thus in order to avoid overfitting, in the following analysis, we set the training rounds to 75 for Channel A, and the training rounds to 100 for Channel B. The MAE loss of the RLS-based predictor after the iteration are 0.1287 and 0.1315 for Channel A and Channel B, respectively, higher than the three DNN prediction models.

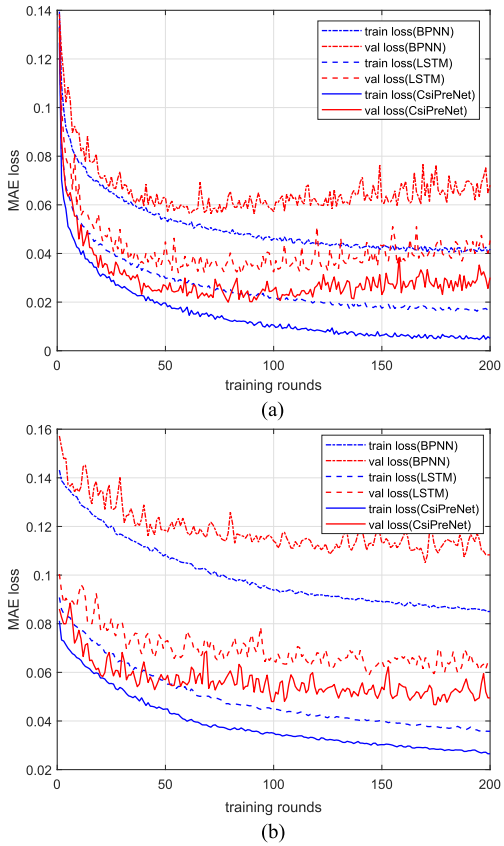

Fig. 10. MAE loss of three DNN models perform training on (a) Channel A and (b) Channel B.

Fig. 11 shows the prediction results of the CSI by applying the BPNN, LSTM, RLS, ALD-KRLS, and CsiPreNet models. Fig. 11(a) and Fig. 11(b) shows the prediction results of Channel A and Channel B, using the five prediction models. Fig. 12 shows the absolute prediction error of them. As shown in Fig. 11 and Fig. 12, the prediction results of all the prediction methods look good, and the prediction error of the CsiPreNet is slightly better than the other four models, and the RLS-based prediction model has the highest prediction error for both Channel A and Channel B. It is worth noting that the performance of ALD-KRLS method is close to that of LSTM method, and better than that of BPNN and RLS methods in both Channel A and Channel B. Moreover, it has more advantages when the amount of data is smaller. Compared with the traditional RLS algorithm, KRLS algorithm expands the processing ability of nonlinear data. Thus, KRLS algorithm has stronger processing ability for complex time-varying sequences.

TABLE I
ERROR COMPARISON OF DIFFERENT PREDICTION MODELS

Prediction Methods	Channel A					Channel B				
	RLS	*ALD-RLS*	*BPNN*	*LSTM*	*CsiPreNet*	*RLS*	*ALD-RLS*	*BPNN*	*LSTM*	*CsiPreNet*
MAPE	13.815	3.935	5.493	2.335	1.721	14.433	5.823	11.279	6.976	4.493
RMSE	0.269	0.067	0.085	0.054	0.037	0.287	0.089	0.223	0.106	0.072

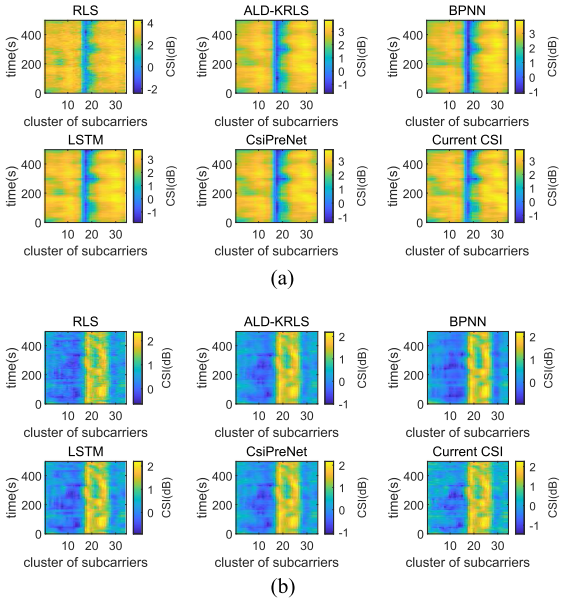

Fig. 11. Prediction result by using different DNN models for (a) Channel A and (b) Channel B.

The cumulative distribution functions (CDF) of the absolute prediction error for the five prediction models are shown in Fig. 13. The probability of a smaller prediction error is higher in the CsiPreNet compared to the other four methods.

Mean absolute percentage error (MAPE) and root mean square error (RMSE) are used measure error. The formula of MAPE and RMSE are

$$MAPE = 100 \times \frac{1}{N} \sum_{t=1}^{N} \left| \frac{y_t - \hat{y}_t}{y_t} \right|, \tag{35}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2}, \tag{36}$$

where y_t, $\hat{y}_t$, and N are the real data, the predicted one, and its length, respectively. Table I shows the prediction errors of different prediction models.

Same as the above analysis, the prediction errors of the CsiPreNet are the lowest, and the RLS-based prediction model has the highest prediction errors. No matter which DNN model is used, prediction errors of Channel A are lower than that of Channel B. This is because there are more training samples of Channel A than Channel B. Generally speaking, more training samples means a better trained model.

C. OFDMA Performance Analysis

The proposed downlink UWA OFDMA system described in Section III-A is used to evaluate the performance gain applying channel prediction. Measured Channel A and Channel B are used as the channels for two users. The offline-online prediction scheme proposed in Section V-C is also embedded in the proposed OFDMA system. Each data packet is a frame of OFDM signal and contains 8 OFDM blocks. The simulated OFDM parameters are the same as the experimental parameters described in Section III. For the three proposed DNN models, we input enough data to train the model in the offline part. Then the trained model is used for CSI prediction in the online part, and we update the training set every 5 minutes. All simulation results are statistical values after 200 transmission rounds, and the time interval between two adjacent handshaking rounds (each transmission round has three handshaking rounds, i.e., RTS, CTS, and data transmission) is 4 s.

Fig. 14 shows the RMSE of the prediction results from different predictors. The prediction accuracy of all predictors increases with SNR increases. The same as the result of the previous analysis, the CsiPreNet has the highest prediction accuracy, no matter whether Channel A or Channel B, followed by the LSTM and ALD-KRLS models. RLS has the worst prediction performance. The DNN models in Channel A perform better than Channel B, because we have more data for model training of Channel A. KRLS and RLS predictors do not require model training, it only needs to update the parameters adaptively according to the input data, so the prediction errors in Channel A and Channel B by using RLS predictor are almost the same. Thus, the gap of prediction errors between Channel A and Channel B by using the RLS predictor is small, so does the ALD-KRLS method.

Fig. 15(a) shows the BER performance of the OFDMA system by using different predictors. The fixed interlaced subcarrier allocation (modulation scheme is QPSK) and the resource allocation based on outdated CSI and current CSI are used for comparison. Specifically, the "current CSI" is an ideal benchmark (not practical) for comparison, and is defined as the CSIs at the time slot T3 in Fig. 2(b), which means that the CSIs are current and not outdated. The BER of all adaptive allocation methods is lower than that of fixed interlaced subcarrier allocation, and BER based on predicted CSI is lower than that based on outdated CSI, but higher than that based on current CSI. The BER of all DNN models is lower than that of the RLS predictor. The ALD-KRLS model has the same BER performance as the LSTM model, but the better BER performance than the BPNN model. The CsiPreNet has the lowest BER, and has 1.5 dB and 2 dB performance improvements under the case of BER$= 10e-3$, compared with the BPNN model and the RLS

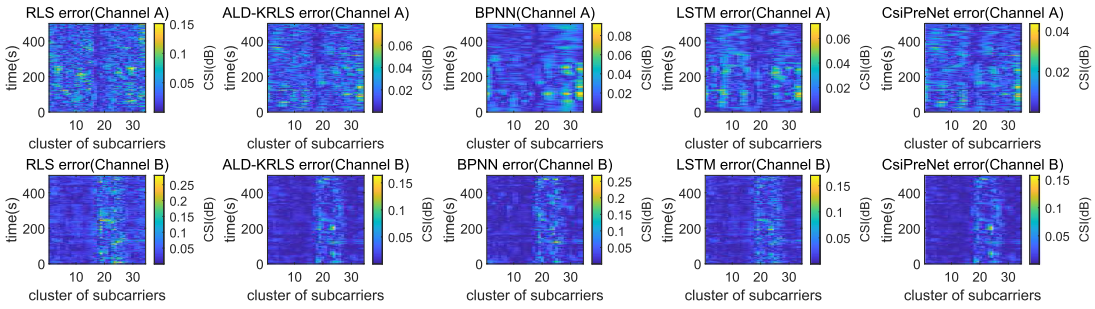

Fig. 12. Absolute prediction error.

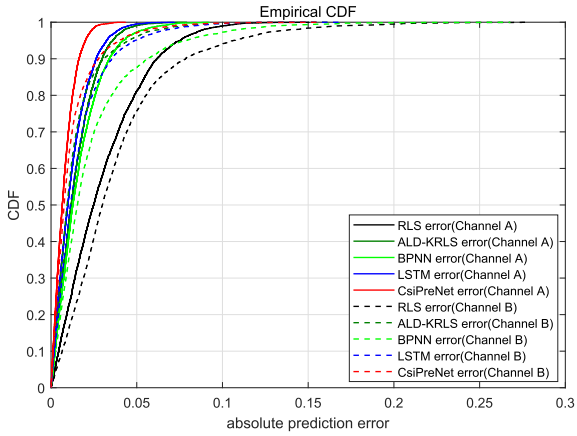

Fig. 13. CDF of the absolute prediction error for different prediction models.

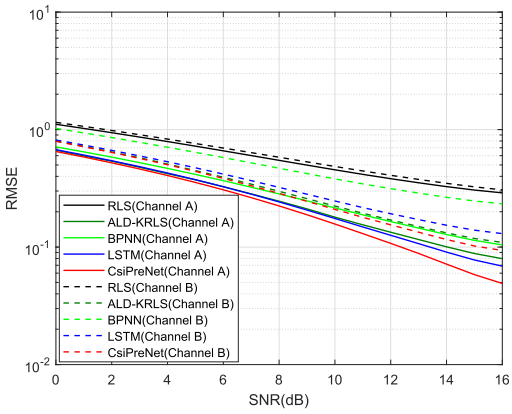

Fig. 14. RMSE of different prediction model through SNR.

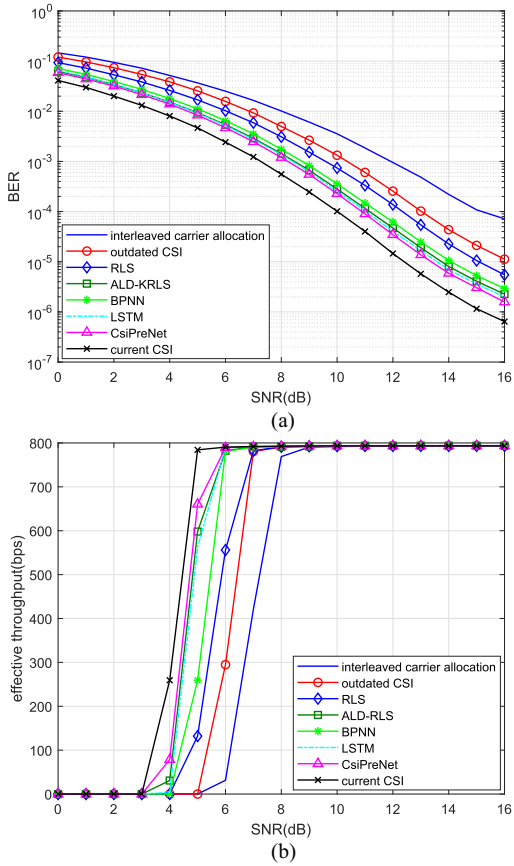

Fig. 15. (a) BER performance and (b) Effective throughput of OFDMA system based on different predictor.

predictor, respectively. But the performance gap between the CsiPreNet and the LSTM/ALD-KRLS is small.

Fig. 15(b) shows the effective throughput of the OFDMA system by using the different predictors. The effective throughput is defined as: the amount of valid data successfully sent in a unit of time. We assume that when the bit error rate exceeds 0.01, the packet will be dropped. The results are similar to the BER performance analysis. The effective throughput of all adaptive allocation methods is higher than that of fixed interlaced subcarrier allocation, and effective throughput based on predicted CSI is higher than that based on outdated CSI, but lower the that based on current CSI. The effective throughput of all DNN models is higher than that of RLS. The ALD-KRLS model has the same throughput as the LSTM model, but the higher throughput than the BPNN model. The CsiPreNet has the highest effective throughput, and has 0.8 dB and 1 dB performance improvements under the case of SNR in the range of $4-8$ dB, compared with the BPNN model and the RLS predictor, respectively. Again, the performance gap between CsiPreNet and LSTM/ALD-KRLS is also small.

D. Experimental Results

Furthermore, we use the sea trial data described in Section III-B for performance evaluation. We define the data transmitted on Channel A as Data A, and the data transmitted on Channel B as Data B. According to the outdated CSI, the current CSI, and the predicted CSI based on the RLS and DNN models, we perform adaptive subcarrier allocation (modulation scheme

TABLE II
BER COMPARISON OF DIFFERENT PREDICTION MODELS

Methods	*Interleaved*	*Outdated CSI*	*RLS*	*ALD-RLS*	*BPNN*	*LSTM*	*CsiPreNet*	*Current CSI*
BER(Uncoded)	0.0620	0.0464	0.0322	0.0198	0.0226	0.0193	0.0177	0.0107
BER(Coded)	0.0021	0.0014	0.0002	0	0	0	0	0

TABLE III
COMPUTATIONAL COMPLEXITY OF DIFFERENT PREDICTION MODELS

Methods	*RLS*	*ALD-RLS*	*BPNN*	*LSTM*	*CsiPreNet*
computing complexity in training phase	NA	NA	$M \times S \times \mathrm{O}(\mathrm{BPNN})$	$M \times S \times \mathrm{O}(\mathrm{LSTM})$	$M \times S \times \mathrm{O}(\mathrm{CsiPreNet})$
computing complexity in prediction phase	$\mathrm{O}(n_r^2)$	$\mathrm{O}(D_s^2)$	$\mathrm{O}(\mathrm{BPNN})$	$\mathrm{O}(\mathrm{LSTM})$	$\mathrm{O}(\mathrm{CsiPreNet})$
CPU running time for traninig/prediction phase	1.04s	3.19s	8.31s/0.034ms	110.44s/0.455ms	293.09s/0.981ms

is QPSK) on Data A and Data B. Moreover, fixed interleaved subcarrier allocation is used for comparison. A rate-$1/2$ 64-state convolution code is used for encoding. Finally, we demodulate the data separately on the subcarriers allocated to Data A and Data B. Table II shows the statistical results of BER based on 200 frame signals.

As Table II shown, the BER of all adaptive allocation methods are lower than that of fixed interleaved subcarrier allocation, and the allocation scheme while using DNN models, have better BER performance than using the outdated CSI and the CSI predicted by the RLS method. The ALD-KRLS model has the same BER performance as the LSTM model, but the better BER performance than the BPNN model. The CsiPreNet still has the best performance.

To sum up, the above results have shown that based on a large amount of historical data, comparing with the RLS predictor, our proposed DNN models have better performance. Due to the enhancement of nonlinear data processing ability, the ALD-KRLS method in this experiment has similar performance to the LSTM method. Especially, the CsiPreNet has the best performance, which sheds the light that the CSI prediction does not only depend on the temporal information but also the frequency information, and the CsiPreNet can well capture the correlation of information in time and frequency domain.

E. Computational Complexity Discussion

In this subsection, we compare the computational complexity of the prediction methods used in this paper. The computing complexity of LSTM layer is $\mathrm{O}(4n_{li} \times n_{lh} + 4{n_{lh}}^2 + 3n_{lh} + n_{lh} \times n_{lo})$, where n_{li}, n_{lo}, and n_{lh} are the number of inputs, outputs, and hidden layer units, respectively. The computing complexity of CNN layer is $\mathrm{O}((n_{mx} \times n_{my}) \times (n_{fx} \times n_{fy}) \times n_{ci} \times n_{co})$, where $n_{mx} \times n_{my}$ and $n_{fx} \times n_{fy}$ are the size of filter and feature map, and n_{ci} and n_{co} are the number of input and output channels. The computational complexity of the fully connected layer is $\mathrm{O}(n_{fci} \times n_{fch} + n_{fch} \times n_{fco})$, where n_{fci}, n_{fch}, and n_{fco} are the number of neurons in input layer, hidden layer, and output layer, respectively. Therefore, the computing complexity of BPNN, LSTM, and CsiPreNet are $\mathrm{O}(n_{fci} \times n_{fch} + n_{fch} \times n_{fco})$, $\mathrm{O}(4n_{li} \times n_{lh} + 4{n_{lh}}^2 + 3n_{lh} + n_{lh} \times n_{lo} + n_{fci} \times n_{fch} + n_{fch} \times n_{fco})$, and $\mathrm{O}((n_{mx} \times n_{my}) \times (n_{fx} \times n_{fy}) \times n_{ci} \times n_{co} + 4n_{li} \times n_{lh} + 4{n_{lh}}^2 + 3n_{lh} + n_{lh} \times n_{lo} + n_{fci} \times n_{fch} + n_{fch} \times n_{fco})$, respectively. In Table III, we use $\mathrm{O}(\mathrm{LSTM})$, $\mathrm{O}(\mathrm{LSTM})$ and $\mathrm{O}(\mathrm{LSTM})$ to simplify the computing complexity of the three DNN models. The computing complexity of the RLS algorithm is of the order $\mathrm{O}(n_r^2)$, where n_r is the length of the control filter. The computing complexity of the ALD-RLS algorithm is $\mathrm{O}(D_s^2)$, where D_s is the size of the dictionary.

Based on the above analysis, the computing complexity of the DNNs depends on the depth of the network and the number of neurons, while the computing complexity of the RLS algorithm depends on the length of the control filter. In the training process, all deep learning methods have high computational complexity, i.e., $M \times S \times \mathrm{O}(N)$, where M and S are the number of training rounds and training samples, and $\mathrm{O}(N)$ is the computational complexity of deep learning method. If the training process is processed offline and we only consider the prediction part ($M = 1$, $S = 1$), and even reduce some parameters (such as the number of hidden layers and neurons), the computational complexity can be reduced a lot. In this paper, we run all the models on a computer with 16 GB RAM and a 2.2 GHz Intel(R) Core(TM) i7–8750H CPU. Take Channel A as an example, the training phase (75 training rounds) of the BPNN, LSTM, and CsiPreNet models take around 8.31 s, 110.44 s, and 293.09 s, respectively, while the running time of the RLS model and ALD-KRLS model are 1.04 s and 3.19, respectively. During the prediction phase, the running time of the BPNN, LSTM, and CsiPreNet models take around 0.034 ms, 0.455 ms, and 0.981 ms, respectively. Therefore, in the case of a large number of training samples trained offline, the DNN models can be applied to the real-time UWA communication system according to the short prediction time. The intuitive computational complexity comparison is shown in Table III.

VII. CONCLUSION

In this paper, for adaptive downlink UWA OFDMA system over time-varying UWA channels, we investigate the possibility of predicting an UWA channel in the frequency domain. The

DNN models are considered to capture complex nonlinear relationships for channel prediction in the frequency domain. First, we design a learning model called CsiPreNet, which combines the CNN and LSTM networks and is used to capture the temporal correlation and frequency correlation on UWA CSIs. Second, we further develop an offline-online prediction scheme, enabling the prediction results to be more stable when applying it to the adaptive downlink UWA OFDMA system. The offline part is used to train a large number of collected CSI, and then the well-trained model is used for CSI prediction in the online model. Finally, a simulated downlink UWA OFDMA system is designed to evaluate performance. Two measured channels are used for two users in this OFDMA system. The offline-online prediction scheme is embedded in the OFDMA system. The predicted CSI is used for resource allocation. Simulation and experimental results that are obtained with measured channels and simulated OFDMA system, respectively, show that DNN prediction models can effectively estimate the future CSI, and their performance is better than the existing RLS predictor. The CsiPreNet has the best performance, because it can better capture the correlation of CSI not only in the time domin, but also in the frequency domain.

This work leads us to conclude that DNN based prediction models may be viable for reliable adaptive UWA communication systems. To the best of our knowledge, this is the first paper that designs DNN models for UWA CSI prediction links with measured experimental data. Applying the DNN models to a real-time at-sea UWA communication systems will be an important further research issue, e.g., how to store a large amount of channel data, hardware design, and the balance between the computational complexity and system performance. Considering critical CSI-relevant features (e.g., frequency band, location, temperature, etc.) in the learning process may improve the prediction accuracy, which beckons further research.

REFERENCES

[1] M. Stojanovic, "Underwater acoustic communications: Design considerations on the physical layer," in *Proc. IEEE Conf. Wireless Demand Netw. Syst. Serv.*, Garmisch-Partenkirchen, Germany, Jan. 2008, pp. 1–10.

[2] M. Stojanovic, J. Catipovic, and J. G. Proakis, "Adaptive multichannel combining and equalization for underwater acoustic communications," *J. Acoustical Soc. Amer.*, vol. 94, no. 3, pp. 1621–1631, May 1993.

[3] W. K. Lam, R. F. Ormondroyd, and J. J. Davies, "A frequency domain adaptive coded decision feedback equalizer for a broadband UWAC OFDM system," in *Proc. IEEE Oceans Conf.*, Nice, France, vol. 2, pp. 794–799, 1998.

[4] P. Qarabaqi and M. Stojanovic, "Statistical characterization and computationally efficient modeling of a class of underwater acoustic communication channels," *IEEE J. Ocean. Eng.*, vol. 38, no. 4, pp. 701–717, Oct. 2013.

[5] S. Mani, T. M. Duman, and P. Hursky, "Adaptive coding/modulation for shallow-water UWA communications," in *Proc. Acoustical Soc. Amer.*, Paris, France, Jul. 2008, pp. 4255–4260.

[6] C. Choudhuri and U. Mitra, "Capacity bounds and power allocation for underwater acoustic relay channels with ISI," in *Proc. 4th ACM Int. Workshop Underwater Netw.*, Nov. 2009, pp. 1–8.

[7] X. Huang, "Analysis and optimization of OFDM underwater acoustic communications," Ph.D. dissertation, Department of Electrical & Computer Engineering, Stevens Inst. Technol., Hoboken, NJ, USA, Mar. 2011.

[8] A. Radosevic, R. Ahmed, T. M. Duman, J. G. Proakis, and M. Stojanovic, "Adaptive OFDM modulation for underwater acoustic communications: Design considerations and experimental results," *IEEE J. Ocean. Eng.*, vol. 39, no. 2, pp. 357–370, Apr. 2014.

[9] L. Wan *et al.*, "Adaptive modulation and coding for underwater acoustic OFDM," *IEEE J. Ocean. Eng.*, vol. 40, no. 2, pp. 327–336, Apr. 2015.

[10] K. Pelekanakis and L. Cazzanti, "On adaptive modulation for low SNR underwater acoustic communications," in *Proc. OCEANS MTS/IEEE Charleston*, Charleston, SC, USA, 2018, pp. 1–6.

[11] J. Cheon and H. S. Cho, "A heuristic resource allocation method for underwater uplink OFDMA system," in *Proc. IEEE 7th Int. Conf. Ubiquitous Future Netw.*, Jul. 2015, pp. 811–813.

[12] Y. Zhang, Y. Huang, L. Wan, S. Zhou, X. Shen, and H. Wang, "Adaptive OFDMA with partial CSI for downlink underwater acoustic communications," *J. Commun. Netw.*, vol. 18, no. 3, pp. 387–396, 2016.

[13] G. Qiao, L. Liu, L. Ma, and Y. Yin, "Adaptive downlink OFDMA system with low-overhead and limited feedback in time-varying underwater acoustic channel," *IEEE Access*, vol. 7, pp. 12729–12741, 2019.

[14] S. Zhou and G. B. Giannakis, "Adaptive modulation for multiantenna transmissions with channel mean feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1626–1636, Sep. 2004.

[15] A. Radosevic, T. M. Duman, J. G. Proakis, and M. Stojanovic, "Channel prediction for adaptive modulation in underwater acoustic communications," in *Proc. IEEE OCEANS*, Spain, Santander, 2011, pp. 1–5.

[16] S. H. Huang, J. Tsao, T. C. Yang, and S. Cheng, "Model-based signal subspace channel tracking for correlated underwater acoustic communication channels," *IEEE J. Ocean. Eng.*, vol. 39, no. 2, pp. 343–356, Apr. 2014.

[17] X. Cheng, L. Yang, and X. Cheng, "Adaptive relay-aided OFDM underwater acoustic communications," in *Proc. IEEE Int. Conf. Commun.*, London, U.K., 2015, pp. 1535–1540.

[18] N. Lin, H. Sun, E. Cheng, J. Qi, X. Kuai, and J. Yan, "Prediction based sparse channel estimation for underwater acoustic OFDM," *Appl. Acoust.*, vol. 96, pp. 94–100, Sep. 2015.

[19] E. Cheng *et al.*, "Precoding based channel prediction for underwater acoustic OFDM," *China Ocean Eng.*, vol. 31, pp. 256–260, 2017.

[20] W. Sun and Z. Wang, "Online modeling and prediction of the large-scale temporal variation in underwater acoustic communication channels," *IEEE Access*, vol. 6, pp. 73984–74002, 2018.

[21] Z. Wang, C. Wang, and W. Sun, "Adaptive transmission scheduling in time-varying underwater acoustic channels," in *Proc. MTS/IEEE OCEANS*, Washington, Washington, DC, USA, 2015, pp. 1–6.

[22] S. Van Vaerenbergh and I. Santamara, "A comparative study of kernel adaptive filtering algorithms," in *Proc. IEEE Digit. Signal Process. Signal Process. Educ. Meeting*, Napa, California, USA, 2013, pp. 181–186.

[23] Y. Engel, S. Mannor, and R. Meir, "The kernel recursive least squares algorithm," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2275–2285, Aug. 2004.

[24] Y. Li, B. Li, and Y. Zhang, "A channel state information feedback and prediction scheme for time-varying underwater acoustic channels," in *Proc. Int. Conf. Intell. Transp., Big Data Smart City*, 2018, pp. 141–144.

[25] Y. Zhang, R. Venkatesan, O. A. Dobre, and C. Li, "Efficient estimation and prediction for sparse time-varying underwater acoustic channels," *IEEE J. Ocean. Eng.*, vol. 45, no. 3, pp. 1112–1125, Jul. 2020.

[26] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[27] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[28] J. Wang, C. Jiang, H. Zhang, Y. Ren, K. C. Chen, and L. Hanzo, "Thirty years of machine learning: The road to pareto-optimal wireless networks," *IEEE Commun. Surv. Tut.*, vol. 22, pp. 1472–1514, 2020.

[29] A. Zappone, M. Di Renzo, M. Debbah, T. T. Lam, and X. Qian, "Model-aided wireless artificial intelligence: Embedding expert knowledge in deep neural networks for wireless system optimization," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 60–69, Jan. 2020

[30] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?" *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 7331–7376, Oct. 2019.

[31] W. Liu *et al.*, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, Apr. 2017.

[32] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Commun. Surv. Tut.*, vol. 21, no. 3, pp. 2224–2287, Jul.-Sep. 2019.

[33] L. Liu, H. Feng, T. Yang, and B. Hu, "MIMO-OFDM wireless channel prediction by exploiting spatial-temporal correlation," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 310–319, Jan. 2014.

[34] 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, pp. 1–11, Jun. 2018.

[35] G. Liu, Y. Xu, Z. He, Y. Rao, J. Xia, and L. Fan, "Deep learning-based channel prediction for edge computing networks toward intelligent connected vehicles," *IEEE Access*, vol. 7, pp. 114487–114495, 2019.
[36] M. Arnold, S. Drner, S. Cammerer, S. Yan, J. Hoydis, and S. T. Brink, "Enabling FDD massive MIMO through deep learning-based channel prediction," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, Cannes, France, Jul. 2019, pp. 1–19.
[37] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel state information prediction for 5 G wireless communications: A deep learning approach," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 227–236, Jan.-Mar. 2020
[38] Y. Zhang *et al.*, "Deep learning based single carrier communications over time-varying underwater acoustic channel," *IEEE Access*, vol. 7, pp. 38420–38430, 2019.
[39] Z. Fang, J. Wang, C. Jiang, Q. Zhang, and Y. Ren, "AoI inspired collaborative information collection for AUV assisted Internet of underwater Things," *IEEE Internet Things J.*, to be published, doi: 10.1109/JIOT.2021.3049239.
[40] A. W. Fuxjaeger and R. A. Iltis, "Adaptive parameter estimation using parallel Kalman filtering for spread spectrum code and doppler tracking," *IEEE Trans. Commun.*, vol. 42, no. 6, pp. 2227–2230, Jul. 1994.
[41] R. Nadakuditi and J. C. Preisig, "A channel subspace post-filtering approach to adaptive least-square estimation," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1901–1914, Jul. 2004.
[42] E. Eleftheriou and D. Falconer, "Tracking properties and steady-state performance of RLS adaptive filter algorithms," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP- 34, no. 5, pp. 1097–1110, Oct. 1986.
[43] L. Changwei, H. Shujuan Shuuan, and M.Wenbo, "Adaptive prediction of channels with sparse features in OFDM systems," *Int. J. Antennas Propag.*, vol. 5, pp. 140–154, 2013.
[44] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1708–1721, Mar. 2010.
[45] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Proc. Int. Joint Conf. Neural Netw.*, 2002, pp. I:593–608.
[46] R. Jiang, X. Wang, S. Cao, J. Zhao, and X. Li, "Deep neural networks for channel estimation in underwater acoustic OFDM systems," *IEEE Access*, vol. 7, pp. 23579–23594, Feb. 2019.
[47] "Unsupervised learning and deep learning tutorial," 2011. [Online]. Available: http://udl.stanford.edu/wiki/
[48] F. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, 2000.
[49] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 25, no. 2, pp. 1106–1114, 2012.
[50] K. Cho and D. Yoon, "On the general BER expression of one-and two dimensional amplitude modulations," *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1074–1080, Nov. 2002.
[51] M. Gray and L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. IT-44, No 6, pp. 2325–2383, Oct. 1998.
[52] S. Zhou and Z. Wang, *OFDM for Underwater Acoustic Communications*. Hoboken, NJ, USA: Wiley, Mar. 2014.
[53] F. Chollet *et al.*, "Keras," 2015. [Online]. Available: https://keras.io
[54] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, no. 2, pp. 281–305, 2012.
[55] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.* 12, pp. 2825–2830, 2011.

Lei Liu (Student Member, IEEE) received the B.S. degree in 2015 in underwater acoustic engineering from Harbin Engineering University, Harbin, China, where he is currently working toward the Ph.D. degree with the College of Underwater Acoustic Engineering. From October 2019 to October 2020, he visited the University of Victoria, Victoria, BC, Canada. His current research interests include adaptive communication and communication network for time-varying underwater acoustic channels.

Lin Cai (Fellow, IEEE) received the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2002 and 2005, respectively. Since 2005, she has been an Assistant Professor and then an Associate Professor with the Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada. Her research interests include wireless communications and networking, with a focus on network protocol and architecture design supporting emerging multimedia traffic over wireless, mobile, ad hoc, and sensor networks. She was the recipient of the Natural Sciences and Engineering Research Council Discovery Accelerator Supplement Grant in 2010. She was also the recipient of the best paper awards at the 2008 IEEE International Conference on Communications and the 2011 IEEE Wireless Communications and Networking Conference. She was an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, *EURASIP Journal on Wireless Communications and Networking*, *International Journal of Sensor Networks*, and *Journal of Communications and Networks*.

Lu Ma (Member, IEEE) received the B.S. and Ph.D. degrees in signal and information processing from Harbin Engineering University (HEU), Harbin, China, in 2010 and 2016, respectively. From October 2014 to October 2015, she visited the University of Connecticut, Storrs, CT, USA. From 2016 to 2018, she was an Assistant Professor with the College of Underwater Acoustic Engineering, HEU, where she is currently an Associate Professor. Her research interests include multicarrier and multiuser communications for underwater acoustic channels.

Gang Qiao (Member, IEEE) received the B.S., M.S., and Ph.D. degrees from the College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin, China, in 1996, 1999, and 2004, respectively. He is currently a Professor and the Associate Dean with the College of Underwater Acoustic Engineering, Harbin Engineering University. In 1999, he was with the College of Underwater Acoustic Engineering, Harbin Engineering University. He has already authored or coauthored more than 80 papers and owned seven national invention patents. His current research interests include underwater communication and network, detection and positioning of Underwater Targets, and the sonar designed for small carriers. He is a Member of the Acoustical Society of China, a Member of the Youth Federation of Hei Longjiang Province, and the Vice Chairman of the Robotics Society of Hei Longjiang Province. He won the national award for the outstanding scientific and technological workers and the Science & Technology Award for Young Talents in Hei Longjiang Province.