IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 25, 2020, accepted July 8, 2020, date of publication July 20, 2020, date of current version July 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3010359

Online LSTM-Based Channel Estimation for HF MIMO SC-FDE System

ZHIYONG WANG[1], FANGLING PU[1], (Member, IEEE), XIAOSHI YANG[1], NING CHEN[1], YONGMIN SHUAI[2], AND RUI YANG[2]

[1]School of Electronic Information, Wuhan University, Wuhan 430072, China
[2]Wuhan Maritime Communication Research Institute, Wuhan 430079, China

Corresponding author: Fangling Pu (flpu@whu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB2100503, and in part by the Project Fund of the Wuhan Maritime Communication Research Institute under Grant 2017J-12.

ABSTRACT The long distance high frequency (HF) communication suffers from time varying multi-path fading. Multiple-Input Multiple-Output (MIMO) and Single-Carrier Frequency Domain Equalization (SC-FDE) have been introduced to HF communication system to combat fading. Because accurate channel estimation is essential for system operation, an architecture of signal flow chart has been designed for the HF MIMO SC-FDE system. In the architecture, an online Long Short-Term Memory (LSTM) estimator is proposed. Different from the channel responses time series created by the LSTM training and prediction loop, a corrected channel responses that are obtained through the received data symbols and the restored transmitting data symbols make up the channel responses time series. In order to evaluate the performance of channel estimators, a simulation system has been built. The uncorrelated and correlated channels are simulated referring to International Telecommunication Union (ITU)-R F.1487 standard and Kronecker model. The simulation results demonstrate that the online LSTM estimator outperforms Least Square (LS) and Recursive Least Square (RLS) estimators in terms of Bits Error Rate (BER) and Mean Square Error (MSE). The online LSTM estimator is capable of tracking the time varying HF MIMO channels. It has potentiality in actual long distance HF communication.

INDEX TERMS HF MIMO SC-FDE, online LSTM, channel estimator.

I. INTRODUCTION

High frequency (HF) communication which range of radio frequency is between 3 and 30 MHz has been widely utilized in long-distance military and civil communications. Besides its low cost and simple operation, the main advantage of HF communication is that it is not easy to be destroyed by wars and disasters owing to its propagation through the ionosphere. However, the HF channel not only presents a limited bandwidth and moderate data transfer rate, but also exhibits time-varying nature that is made by the fluctuation of ionosphere. The HF channel can provide a narrow bandwidth allocation and exhibits multipath effect and time-varying nature. The barriers to applying HF communication are its low communication rate and poor reliability.

Multiple-Input Multiple-Output (MIMO) technology has been introduced into HF communication systems to suppress the effect of multipath and increase the data rate. Compared with the single input single output (SISO), MIMO achieves higher data rate, better Bit Error Rate (BER) and wider coverage. The combination of MIMO and HF has become the development trend of the next generation HF communication system [1]. Based on polarization diversity and radiation pattern diversity, a heterogeneous array technology has been proposed to implement a HF MIMO system in [2], [3]. Their results showed that MIMO technology can effectively improve the channel capacity of HF communication. The reliability of HF MIMO was investigated in [4], [5] [6]. The experiment results demonstrated that the MIMO technology could improve the BER performance of HF communication. In the process of implementing HF MIMO, Single-Carrier Frequency Domain Equalization (SC-FDE) and Orthogonal Frequency Division Multiplexing (OFDM) are the two techniques to mitigate the multipath interference of HF channel. The conclusion that SC-FDE has the similar performance and complexity as OFDM was proved in [7], [8]. Besides,

The associate editor coordinating the review of this manuscript and approving it for publication was Emre Koyuncu.

SC-FDE has lower Peak-to-Average Power Ratio (PAPR) and it is less sensitive to frequency offset [7]. SC-FDE is considered more suitable for HF communication. It has been successfully applied to HF MIMO system [6], [9], [10]. We combined MIMO with SC-FDE for HF communication.

Accurate channel state information (CSI) is of crucial importance for precoding the spatial stream at transmitters and detecting and decoding data at receivers. Channel estimation is a crucial activity in the HF MIMO SC-FDE system. When the transmitted signals arrive at the receiver antennas, the equalization method to mitigate the effect of multipath fading depends on the accurate estimation of CSI. The decoding error of space-time block codes (STBC) will occur if the accuracy of channel estimation is too low. If the CSI is fed back to transmitter, adaptive transmission control and pre-equalization can be implemented to further improve the communication. Therefore, the study of HF MIMO channel estimation is of great significance.

The MIMO channel estimation methods can be divided into two categories, traditional methods and machine learning methods. Least Square (LS) [2], [10], Minimal Mean Square Error (MMSE) [11], time domain correlation estimation [12] and adaptive filtering [13] belong to traditional methods. Except adaptive filtering, the other methods are presented under the assumption that the channel is static or quasi-static. LS estimator has low complexity and is easy to realize. But LS estimator is poor at low signal-to-noise ratio (SNR). MMSE estimator can provide better estimation performance at low SNR. However, the numerical instability and high computational complexity restrict its application. A time-domain estimation method based on circular orthogonal pilot sequence was proposed in [12]. The method can eliminate the interference from other antennas by circular orthogonal sequence and improve SNR by extending sequence length. LS, MMSE and time domain correlation estimation all decline in time-varying channels. Adaptive filtering is well-known approach that was proposed for time-varying channels. The convergence speed and computational complexity impede the estimation accuracy promotion and its application. Because the HF channels are multipath and time varying channels, the promotion abilities of traditional channel estimation methods are limited.

Machine learning techniques have been introduced to the MIMO communication systems with the aim of making the communication smarter. Intelligence will be one of the characteristics of future HF communication [14]. Three modes for applying neural network in wireless communication have been proposed. First, the responses of pilot channels are obtained by traditional channel estimation methods, then the estimated results can be optimized by neural network [15], [16] [17], [18]. In [18], the initial channel response was estimated by LS, then the other channel responses were interpolated and the noise was suppressed by image super-resolution (SR) and image restoration (IR) algorithms. Second, the neural network was used to map the relationship between the receiving symbol and transmitting symbol. The trained network can be utilized to decode the received signal directly [19], [20]. The channel estimation was implicit, and the real channel response could not be obtained. Third, the neural network was used to get the value of channel response directly [21], [22]. All of the results show that the neural network methods can improve the accuracy of channel estimation and the communication quality. However, when the neural network based methods were applied to real communication, their performance declined. Because the neural networks are trained mainly on channel model, they are optimal for channel model rather than real channel [23]. Most time, the channel model does not match with the real channel. So the pre-trained network can not work well.

Online training network based on real channel can overcome the weakness of the pre-trained network [24]. Online learning does not depend on the entire training data set. It can adapt dynamically to new patterns in the data. In the initial phase, network is trained by small dataset from real channel. Then the network model is retrained and optimized continuously by the constantly arriving data slices. The network model can be continuously optimized [25]. Online long short-term memory (LSTM) has been demonstrated that it can accurately predict the energy consumption of smart grid [25]. Online LSTM also succeeded in exchange rate prediction and stock price prediction [26], [27]. The examples show that significant improvements can be obtained by online LSTM on one-dimensional time series prediction, such as channel response estimation.

We designed a HF MIMO SC-FDE communication system. The channel estimation is essential to restore the transmitted data symbols at the receivers. Our contributions involve two parts.

1) An online LSTM estimator is proposed to estimate the channel responses of the received data symbols. Different from the channel responses time series that are created by the LSTM training and prediction loop, the channel responses series of the online LSTM estimator are composed of the corrected channel responses that are obtained through the received data symbols. The transmitting data symbols are restored in terms of the communication procedure of the HF MIMO SC-FDE system. The online LSTM estimator suppresses the accumulation error that is generated by the training and prediction loop of the LSTM estimation method. It can track the time varying HF MIMO channels.
2) We designed the architecture of the HF MIMO SC-FDE communication system and the signal flow chart. The uncorrelated channels are simulated according to the International Telecommunication Union (ITU)-R F.1487 standard. The correlated channels are simulated with regard to both ITU-R F.1487 standard and Kronecker model. The Zadoff-chu sequence [28] is used to create orthogonal pilots between a pair of transmitters. The channel estimation method based on the circular orthogonal sequence is applied to the architecture to

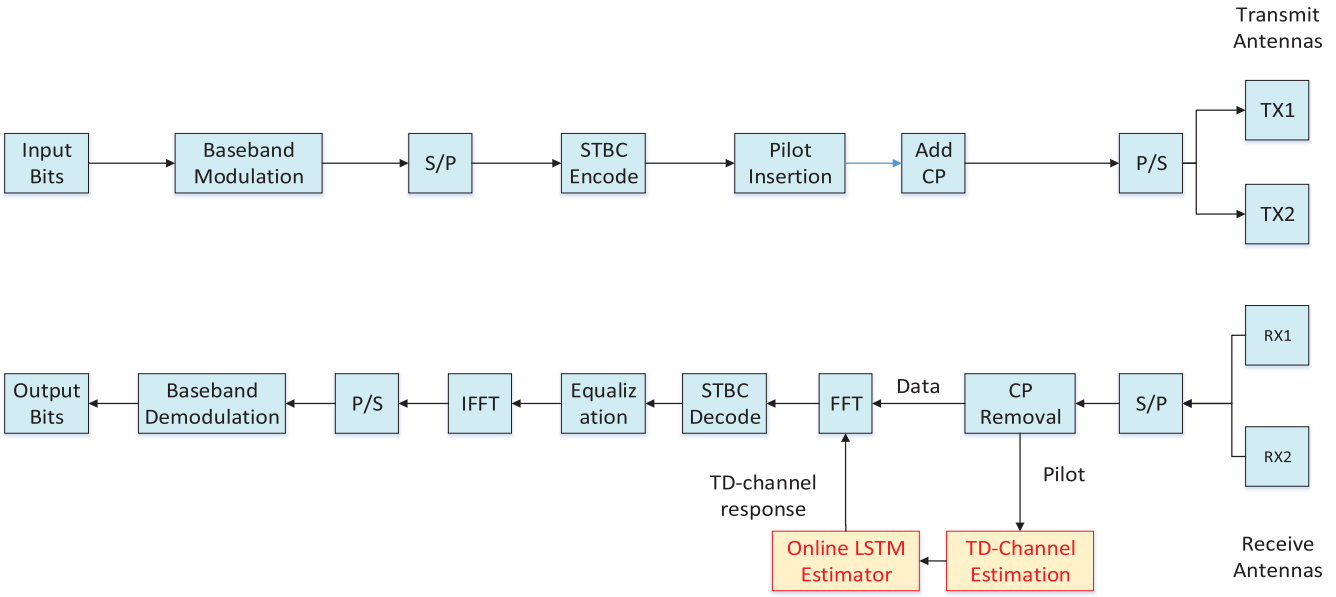


FIGURE 1. Architecture of HF MIMO SC-FDE System.

obtain the channel responses of the pilots for the online LSTM estimator training.

We built a simulation system. The online LSTM estimator has been compared with LS and RLS estimators. Simulation results show that the online LSTM estimator outperforms LS and RLS estimator in terms of BER and MSE. The remains of the paper are organized as follows. Section 2 depicts the MIMO SC-FDE system model and HF MIMO channel model. Section 3 designs the online LSTM channel estimator. Section 4 presents the experimental results and performance analysis. Finally, section 5 summaries the paper.

II. HF MIMO SC-FDE SYSTEM

We built a HF MIMO SC-FDE simulation system. The space-time block code has been applied to encode the signals sent by the transmitting antenna. The channels of HF MIMO SC-FDE system are modeled according to ITU-RF1487.

A. HF MIMO SC-FDE SYSTEM

We designed HF MIMO SC-FDE system shown in Fig. 1. At transmitter, the information bits are mapped into complex signals by baseband modulation. Data blocks are produced through serial-to-parallel conversion of the complex signals. K symbols make up a modulated data block. The data blocks of MIMO transmission are generated through coding the symbols of each modulated data block in terms of Alamouti STBC. A cyclic prefix (CP) with P length is inserted into the front of each data block to combat the multipath fading and suppress the inter-block interference (IBI). So, a data block is composed of Alamouti coding data and CP. As shown in Fig. 2, a number of data blocks form a data group. A pilot group that is composed of several pilot blocks is inserted in front of a data group. The pilot blocks are created in terms

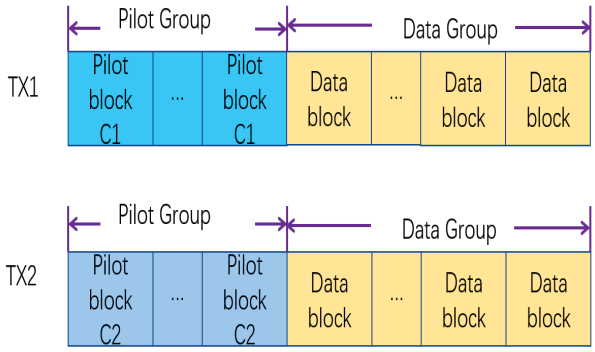


FIGURE 2. Frame Structure for HF MIMO System.

of Zadoff-chu sequence. The pilot group is configured with regard to the number of transmitter antennas. The pilot group is used for online channel estimator and channel track. The configured frame as shown in Fig. 2 is converted from parallel to serial. The serial signals undergo filtering, up-sampling, converting from digital to analogue (D/A), then, are sent out from the transmitting antennas. When the transmitted signals arrive at receiver antennas, they undergo analogue-to-digital conversion, down-sampling, filtering and serial-to-parallel conversion to obtain the receiver data frame that matches the transmitter frame. The channel responses are firstly estimated through the pilot group. The CP in each receiver data block is removed. The channel responses of data blocks are estimated through online LSTM. The estimated channel responses, and data blocks are inputted to FFT (Fast Fourier Transform) module. Then, the channel equalization and Alamouti STBC decoding are implemented in frequency domain. After that, the decoded data are processed by IFFT (inverse fast Fourier transform), parallel-to-serial conversion,

and baseband demodulation to recover the original information bits.

The construction of data blocks are shown in Fig. 3 [29]. Each data block consists of two parts, the encoded data and CP. Denote the k-th symbols of the m-th transmitted data block from antenna i by $s_i^{(m)}(k)$. According to Alamouti STBC, pairs of data between two consecutive blocks are generated by formula (1), where $m = 0, 2, 4, \cdots$. $()^*$ represents complex conjugation and $()_K$ denotes module-K operation.

$$\begin{aligned} s_1^{m+1}(k) &= -s_2^{*(m)}((-k)_K) \\ s_2^{m+1}(k) &= s_1^{*(m)}((-k)_K) \end{aligned} \tag{1}$$

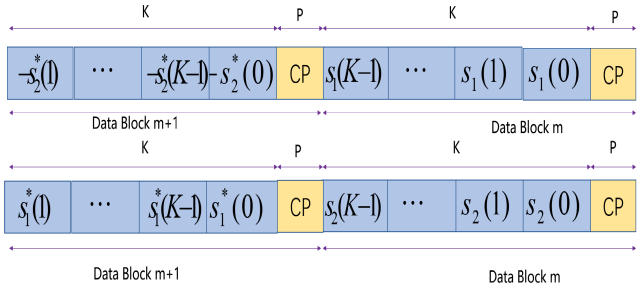


FIGURE 3. Data block format for SC-FDE STBC.

Assume that the received signals are well synchronized. After the CP is removed, the two consecutive data blocks at time $n = m, m+1$ are represented as

$$\boldsymbol{y}^n = \boldsymbol{H\_m}_1^{(n)}\boldsymbol{s}_1^{(n)} + \boldsymbol{H\_m}_2^{(n)}\boldsymbol{s}_2^{(n)} + \boldsymbol{v}^{(n)}, n = m, m+1 \tag{2}$$

where v is additive white Gaussian noise(AWGN) samples vector, $\boldsymbol{H\_m}_1^{(n)}$ is the circulant channel matrix between the antenna 1 and the receiver antenna, and $\boldsymbol{H\_m}_2^{(n)}$ is the circulant channel matrices between antenna 2 and the receiver antenna. The circulant channel matrices $\boldsymbol{H\_m}_i^{(n)}$ are created in terms of the estimated channel impulse response $\boldsymbol{h}_i^{(n)} = [h_i^{(n)}(0) \quad h_i^{(n)}(1) \quad \cdots \quad h_i^{(n)}(L)]$, where L is channel response length. The circular convolution of two blocks is converted to matrix and vector multiplication by $\boldsymbol{H\_m}_i^{(n)}$. $\boldsymbol{H\_m}_i^{(n)}$ can be written as

$$\begin{pmatrix} h_i^{(n)}(0) & 0 & \cdots & h_i^{(n)}(L) & \cdots & h_i^{(n)}(1) \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ h_i^{(n)}(L-1) & \cdots & h_i^{(n)}(0) & 0 & \cdots & h_i^{(n)}(L) \\ h_i^{(n)}(L) & h_i^{(n)}(L-1) & \cdots & h_i^{(n)}(0) & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & h_i^{(n)}(L) & h_i^{(n)}(L-1) & \cdots & h_i^{(n)}(0) \end{pmatrix} \tag{3}$$

Then the relationship between the received data $\boldsymbol{y}^{(n)}$ and the estimated circulant channel matrix are mapped in frequency domain by FFT.

$$\boldsymbol{Y}^{(n)} = \boldsymbol{W}\boldsymbol{y}^{(n)} = \boldsymbol{D}_1^{(n)}\boldsymbol{S}_1^{(n)} + \boldsymbol{D}_2^{(n)}\boldsymbol{S}_2^{(n)} + \boldsymbol{V}^{(n)} \tag{4}$$

where $\boldsymbol{S}_i^{(n)} = \boldsymbol{W}\boldsymbol{s}_i^{(n)}$ and $\boldsymbol{V}^{(n)} = \boldsymbol{W}\boldsymbol{v}^{(n)}$. $\boldsymbol{W}$ is discrete Fourier transform matrix and its elements are described by

$$\boldsymbol{W}(i,j) = \frac{1}{\sqrt{K}} e^{\frac{-j2\pi ij}{K}} \tag{5}$$

$\boldsymbol{D}_1^{(n)}$ and $\boldsymbol{D}_2^{(n)}$ are diagonal matrices, which are calculated as $\boldsymbol{D}_i^{(n)} = \boldsymbol{W}\boldsymbol{H\_m}_i^{(n)}\boldsymbol{W}^*$. According to properties of digital Fourier Transform (DFT) and formula (1), we have

$$\begin{aligned} S_1^{(m+1)}(k) &= -S_2^{*(m)}(k) \\ S_2^{(m+1)}(k) &= S_1^{*(m)}(k) \end{aligned} \tag{6}$$

where $k = 0, 1, \cdots, K-1$ and $m = 0, 2, 4, \cdots$. Combining formula (4) and (6), we have

$$\begin{aligned} \boldsymbol{Y} &= \begin{pmatrix} \boldsymbol{Y}^{(m)} \\ \bar{\boldsymbol{Y}}^{(m+1)} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{D}_1 & \boldsymbol{D}_2 \\ \boldsymbol{D}_2^* & -\boldsymbol{D}_1^* \end{pmatrix} \begin{pmatrix} \boldsymbol{S}_1^{(m)} \\ \boldsymbol{S}_2^{(m)} \end{pmatrix} + \begin{pmatrix} \boldsymbol{V}^{(m)} \\ \bar{\boldsymbol{V}}^{(m+1)} \end{pmatrix} \\ &= \boldsymbol{D}\boldsymbol{S} + \boldsymbol{V} \end{aligned} \tag{7}$$

where $\boldsymbol{D} = \begin{pmatrix} \boldsymbol{D}_1 & \boldsymbol{D}_2 \\ \boldsymbol{D}_2^* & -\boldsymbol{D}_1^* \end{pmatrix}$, $\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_1^{(m)} \\ \boldsymbol{S}_2^{(m)} \end{pmatrix}$ and $\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}^{(m)} \\ \boldsymbol{V}^{*(m+1)} \end{pmatrix}$. And $\bar{()}$ denotes complex conjugation. Then we can get $\boldsymbol{S}_1^{(m)}$ and $\boldsymbol{S}_2^{(m)}$ from the received data by solving the following linear equations

$$\widetilde{\boldsymbol{Y}} = \boldsymbol{D}^*\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{D}^\circ & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}^\circ \end{pmatrix} \begin{pmatrix} \boldsymbol{S}_1^{(m)} \\ \boldsymbol{S}_2^{(m)} \end{pmatrix} + \widetilde{\boldsymbol{V}} \tag{8}$$

where $\boldsymbol{D}^\circ = |\boldsymbol{D}_1(i,i)|^2 + |\boldsymbol{D}_2(i,i)|^2$ and $\widetilde{\boldsymbol{V}} = \boldsymbol{D}^* * \boldsymbol{V}$. We apply MMSE criterion to implement frequency equalizations as formula (9)

$$\hat{\boldsymbol{S}} = (\boldsymbol{D}^*\boldsymbol{D} + \frac{\sigma_v^2}{\sigma_s^2}\boldsymbol{I}_{2K})^{-1}\boldsymbol{D}^*\boldsymbol{Y} \tag{9}$$

where $\boldsymbol{I}_{2K}$ is a $2K * 2K$ unit matrix, σ_v^2 and σ_s^2 are the noise power and signal power respectively. At last, $\hat{\boldsymbol{S}}$ is transformed to its time domain counterpart by IFFT. The original information bits can be obtained through baseband demodulation.

B. HF MIMO CHANNEL MODEL

We modeled the HF propagation that is reflected by ionosphere over long distance. Because the ionosphere is a layered, heterogeneous, anisotropic, dispersive and random time-varying medium, the HF channel experiences time varying multipath and Doppler frequency shift. The HF MIMO channels vary in both time domain and frequency domain.

Watterson [30] model and Institute for Telecommunication Sciences (ITS) model [31] are the two HF channel models that are widely used. The Watterson channel model consists of a tap delay line, where each tap corresponds to an analytic propagation path. Each tap has a double Gaussian Doppler spectrum that consists of two Gaussian functions in the frequency domain. The Watterson model is suitable for narrowband

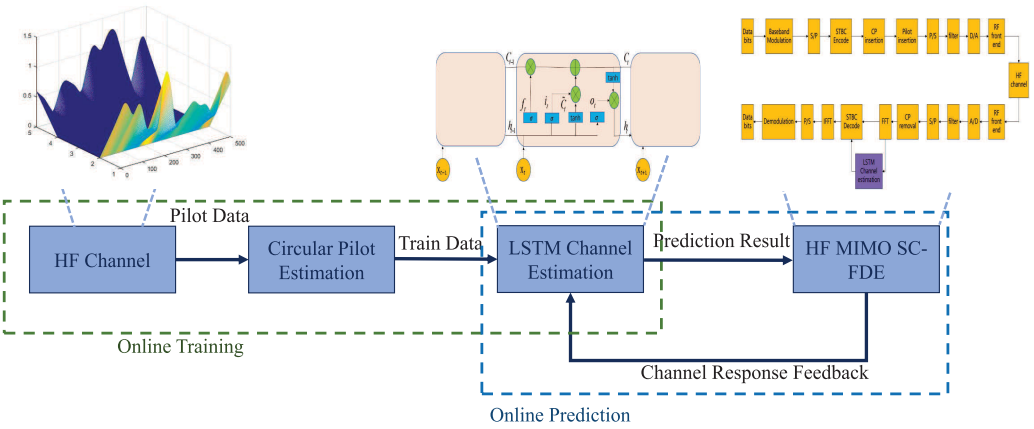


FIGURE 4. Online LSTM Channel Estimation Structure.

(< 12KHz) and short times (<10min) channels. The American Institute of Telecommunications Science proposed a broadband HF channel simulator that is known as the ITS model in 1997. However, the ITS model is not recommended by the ITU (International Telecommunication Union), because there exist issues that have not been resolved. A series of simplified channel models were specified by Recommendation ITU-R F.1487 for quantitative testing of high-frequency modems [32]. There are two independent fading paths with the same power in these models. The fading process has Rayleigh-distributed envelope. The recommended models are widely used in the research of HF simulation. We adopt the recommended models, ITU-R F.1487, in our HF MIMO channels simulation.

In the modeling of HF MIMO channels, the correlation among antennas is considered in the Kronecker model [11]. If a linear array antenna with interval 75 meters, correlation among antennas can be ignored [5]. The channel between a pair of transmitter and receiver can be modeled by the recommended Watterson model. If the correlation among antennas exists, the HF MIMO channel model should multiply the correlation matrix.

We generate the uncorrelated HF MIMO channels in terms of ITU-R F.1487. The HF MIMO channels with correlation among antennas are generated by both ITU-R F.1487 model and Kronecker model. The channels with correlation is modeled as follow

$$\boldsymbol{H}_{MIMO} = \boldsymbol{\Theta}_{\boldsymbol{Rx}}^{\frac{1}{2}} \boldsymbol{H}_{ITU} \boldsymbol{\Theta}_{\boldsymbol{Tx}}^{\frac{1}{2}} \tag{10}$$

where $\boldsymbol{H}_{MIMO}$ represents MIMO channel and $\boldsymbol{H}_{ITU}$ represents independent identically distributed ITU channel model. $\boldsymbol{\Theta}_{\boldsymbol{Rx}}$ and $\boldsymbol{\Theta}_{\boldsymbol{Tx}}$ denote spatial correlation matrices at the receiver and transmitter, respectively. The spatial correlation matrices are modeled as formula (11), (12) and (13).

$$\boldsymbol{\Theta}_{Rx} = \begin{pmatrix} \rho_{1,1} & \cdots & \rho_{1,Nr} \\ \vdots & \ddots & \vdots \\ \rho_{Nr,1} & \cdots & \rho_{Nr,Nr} \end{pmatrix} \tag{11}$$

$$\boldsymbol{\Theta}_{Tx} = \begin{pmatrix} \rho_{1,1} & \cdots & \rho_{1,Nt} \\ \vdots & \ddots & \vdots \\ \rho_{Nt,1} & \cdots & \rho_{Nt,Nt} \end{pmatrix} \tag{12}$$

$$\rho_{i,j} = J_0(\alpha d_{i,j}) \tag{13}$$

where ρ is spatial correlation matrix elements, J_0 represents the zero-order Bessel function, α is wave number and $d_{i,j}$ is the normalized distance between transmit(receive) antenna i and transmit(receiver) antenna j [33].

III. CHANNEL ESTIMATION BASED ON ONLINE LSTM

The framework of online LSTM estimator is introduced in Fig. 4. The online LSTM estimator is composed of two stages, online training and online prediction. During the online training, the channel response between one pair of transceiver antennas is estimated through the circular orthogonal sequences based channel estimation [12] in terms of the pilot block that is set for the corresponding transmitter antenna. The LSTM network is trained by the pilot blocks and the estimated channel response. During the online prediction, the channel responses of data block are estimated by LSTM framework based on the trained LSTM network and the feedback from previously estimated channel responses.

A. CIRCULAR ORTHOGONAL SEQUENCES

The circular orthogonal sequences based channel estimation [12] instead of LS estimator has been utilized for the channel responses estimation of pilot block of each antenna. The circular orthogonal sequences based channel estimation performs better than LS estimator. The configuration of pilots for HF MIMO is shown in Fig. 2. The pilot group of each transmitter is composed of pilot blocks. The number of pilot blocks corresponds to the convergence rate of online LSTM network training. The rule for the pilot sequence construction of all transmitters is the same. The pilots of one transmitter should be orthogonal to that of another transmitter. Although the configuration of pilot sequences increases the system overhead, it enhances the estimation accuracy of

channel responses from different antennas through correlation operations.

Assume N_t transmitters. The pilot blocks, such as $C1$ and $C2$, are constructed according to Zadoff-chu sequence. The Zadoff-chu sequence has cross-correlation and auto-correlation properties that are required by the circular orthogonal sequences based channel estimation. The Zadoff-chu sequence with length Q is generated as formula (14)

$$\boldsymbol{c}(q) = \begin{cases} e^{j\pi rq^2/Q}, & for\ even Q \\ e^{j\pi rq(q+1)/Q}, & for\ odd Q, \end{cases} gcd(Q, r) = 1 \quad (14)$$

where q denotes the q-th symbol of the sequence, r is an integer coprime of Q. And the Zadoff-chu sequence of i-th transmit antenna can be generated through circular shifting as follow

$$\boldsymbol{c}_i = \boldsymbol{c}(q + (i-1)\epsilon)_Q \quad i = 1, 2, \cdots, N_t \quad (15)$$

where ϵ is the shift step. Then the pilot block is constructed by insert cyclic prefix and cyclic postfix. The configuration of the pilot block is shown in Fig. 5. The last P symbols of the Zadoff-chu sequence are duplicated and inserted in the front of the Zadoff-chu sequence as cyclic prefix. The first P ymbols are duplicated and added as cyclic postfix. In order to suppress inter symbol interference (ISI), the length of cyclic prefix P must be more than maximum delay spread L of channel. And the shift step of formula (15) must meet $\epsilon \geq P$, so that the sequences between different pilot blocks keep orthogonality. The length of Zadoff-chu sequence must meet $Q \geq N_t P$, where N_t is the number of transmitting antennas.

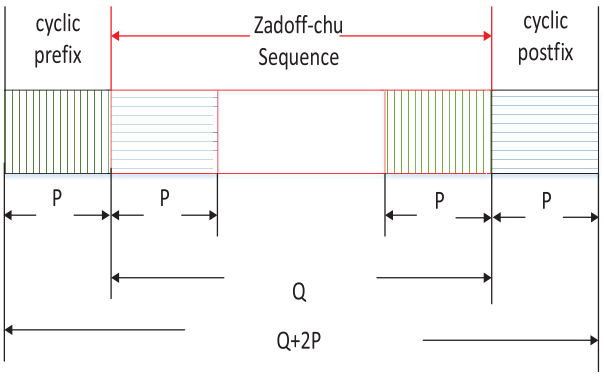


FIGURE 5. Circular Orthogonal Pilot Sequence Structure.

At receiver, $\boldsymbol{y}_j$ represents the signal of j-th receiver antenna after removing the cyclic prefix, and the channel response from transmit antenna i to receive antenna j can be estimated by formula (17)

$$\hat{\boldsymbol{h}}_{ij}(l) = (\frac{1}{Q}) \sum_{q=0}^{Q-1} \boldsymbol{c}_i(q)\boldsymbol{y}_j(q+l), \quad l = 1, 2, \cdots, L \quad (16)$$

B. ONLINE LSTM CHANNEL ESTIMATION

1) NETWORK STRUCTURE

The LSTM network structure of online training and prediction is shown in Fig. 6. It involves three parts: input layer, hidden layer and output layer. The input layer accepts the input data. The hidden layer is composed of the LSTM cells. The output layer provides the prediction results. Each LSTM cell contains input gate, forget gate and output gate [34]. The input gate $\boldsymbol{i}_t$ controls whether the new information can be stored in the cell state and prevents unwanted information from entering the memory unit. The forget gate $\boldsymbol{f}_i$ decides whether discards previous step information from the cell status $\boldsymbol{C}_{i-1}$. The forget gate and the input gate work together to update the state of the memory cell $\boldsymbol{C}_i$. The output gate $\boldsymbol{o}_t$ decides what information will be outputted. The data flow of LSTM network follows the following formulas

$$\begin{aligned} \boldsymbol{f}_i &= \sigma(\boldsymbol{W}_f[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_f) \\ \boldsymbol{i}_t &= \sigma(\boldsymbol{W}_i[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_i) \\ \widetilde{\boldsymbol{C}}_t &= \tanh(\boldsymbol{W}_C[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_C) \\ \boldsymbol{C}_t &= \boldsymbol{f}_t * \boldsymbol{C}_{t-1} + \boldsymbol{i}_t * \widetilde{\boldsymbol{C}}_t \\ \boldsymbol{o}_t &= \sigma(\boldsymbol{W}_o[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_o) \\ \boldsymbol{h}_t &= \boldsymbol{o}_t * \tanh(\boldsymbol{C}_t) \end{aligned} \quad (17)$$

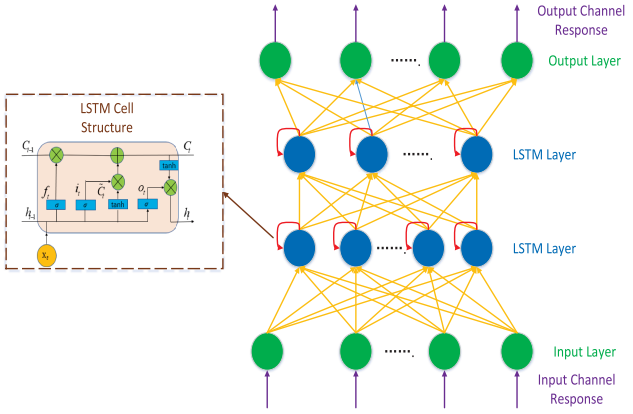


FIGURE 6. Online LSTM Network Structure.

where $\boldsymbol{C}_t$ and $\boldsymbol{h}_t$ denote the cell state and cell output respectively at current time. $\boldsymbol{C}_{t-1}$ is the cell state and $\boldsymbol{h}_{t-1}$ is the cell output at previous time. $\boldsymbol{W}_i$,$\boldsymbol{W}_f$,$\boldsymbol{W}_o$ and $\boldsymbol{W}_C$ are the weight matrix of input gate, forget gate, output gate and cell vectors respectively. $\boldsymbol{b}_i$,$\boldsymbol{b}_f$,$\boldsymbol{b}_o$ and $\boldsymbol{b}_C$ are the bias vector of input gate, forget gate, output gate and cell vectors respectively.

2) LSTM BASED CHANNEL ESTIMATION

After getting channel responses through circular orthogonal pilots, we use them to train the LSTM network. The LSTM based channel estimation process is shown in Fig. 8(a). Assume the time series channel responses are $\big(\boldsymbol{h}(1)\ \boldsymbol{h}(2)\ \cdots\ \boldsymbol{h}(n-1)\ \boldsymbol{h}(n)\big)$, where the vector $\boldsymbol{h}(n)$ represents the channel responses at different time. $\boldsymbol{h}(n)$ is predicted in terms of $\big(\boldsymbol{h}(1)\ \boldsymbol{h}(2)\ \cdots\ \boldsymbol{h}(n-1)\big)$. The procedure of the

time series sliding for the LSTM network training is shown in Fig. 7. Input sequences of Fig. 7 mean the set of channel responses that are inputted to the LSTM network, and the output data denote the output label. Assume the slide windows length is t. The training starts on the vector window that is composed of the first t vectors of the time series channel responses. The first $t-1$ vectors are LSTM network input, and the last vector is output label. Then, the vector window is moved one time step forward when a new channel response comes. The LSTM network is trained again by the vectors of the shifted forward window. The process is repeated until the LSTM network converges.

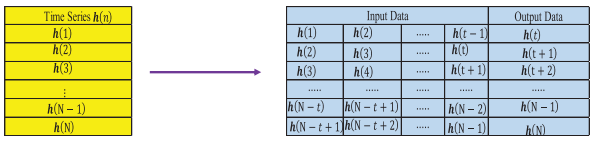
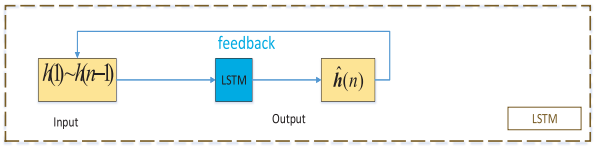


FIGURE 7. The Transforming Method from Time Series to Supervised Learning.

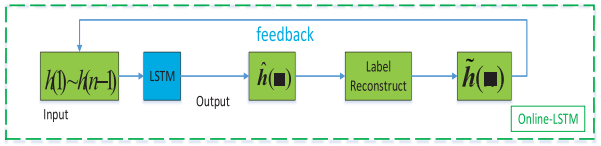
After the LSTM network is trained by the pilot block, it is applied to predict the channel responses of each symbol in the data block. Like the iterative training process, the predicted channel responses are re-inputted to the training process to update the LSTM network. The updated LSTM network is used to predict the next symbol channel responses. The training-prediction loop is iterated until the communication of all the symbols of the data blocks finishes. The main risk of this LSTM based channel estimation is that the prediction error may be accumulated in the iterative loop [35]. The deviation between the estimated channel responses of the symbols at the end of the data sequence and the real channel responses may be large.

3) PROPOSED ONLINE LSTM CHANNEL ESTIMATOR

To suppress the accumulation error of the online training-prediction loop, we adapt the online LSTM channel estimator. As shown in Fig. 8(b), online LSTM estimator uses the predicted channel response vector $\hat{\boldsymbol{h}}(n)$ to reconstruct the label and online learning. Lastly, the regenerated channel response vector $\tilde{\boldsymbol{h}}(n)$ eliminates the accumulation error.



(a) Prediction Process of LSTM Estimator



(b) Prediction Process of Online LSTM Estimator

FIGURE 8. Comparison of LSTM Estimator and Online LSTM Estimator.

We utilize two transmitters and one receiver to illustrate the revised process in Fig. 9. Assume the channel response length between each pair of transmitter and receiver pairs is L. $\hat{\boldsymbol{h}}_{11}(n) = \left[\hat{h}^1_{11}(n)\ \hat{h}^2_{11}(n)\ \cdots\ \hat{h}^L_{11}(n)\right]$ denotes the estimated channel responses vector from the transmitter 1 to the receiver at time n corresponding to data block n (for $n = 0, 2, 4, \cdots$), and $\hat{\boldsymbol{h}}_{21}(n) = \left[\hat{h}^1_{21}(n)\ \hat{h}^2_{21}(n)\ \cdots\ \hat{h}^L_{21}(n)\right]$ are the channel responses vector from the transmitter 2 to the receiver. The two estimated channel responses by LSTM are used for channel equalization and the transmitted data recovery. The corresponding frequency domain channel responses, $\hat{\boldsymbol{H}}_{11}$ and $\hat{\boldsymbol{H}}_{21}$ are obtained by FFT. $\hat{\boldsymbol{H}}_{11}$ and $\hat{\boldsymbol{H}}_{21}$ are rewritten as diagonal matrix $\boldsymbol{D}_{11} = diag(\hat{\boldsymbol{H}}_{11})$ and $\boldsymbol{D}_{21} = diag(\hat{\boldsymbol{H}}_{21})$. Alamouti decoding and channel equalization in SC-FDE system are computed as formula (18)

$$\begin{pmatrix} \widetilde{\boldsymbol{D}} & \boldsymbol{0} \\ \boldsymbol{0} & \widetilde{\boldsymbol{D}} \end{pmatrix} \begin{pmatrix} \widetilde{\boldsymbol{S}}_1(n) \\ \widetilde{\boldsymbol{S}}_2(n) \end{pmatrix} = \begin{pmatrix} \boldsymbol{D}_{11} & \boldsymbol{D}_{21} \\ \boldsymbol{D}^*_{21} & -\boldsymbol{D}^*_{11} \end{pmatrix} \begin{pmatrix} \boldsymbol{Y}(n) \\ \bar{\boldsymbol{Y}}(n+1) \end{pmatrix}$$
$$\widetilde{\boldsymbol{D}} = |\boldsymbol{D}_{11}|^2 + |\boldsymbol{D}_{21}|^2 \tag{18}$$

where $\boldsymbol{Y}(n)$ and $\boldsymbol{Y}(n+1)$ are the received data blocks at time n and $n+1$, $\bar{()}$ denotes complex conjugation and $()^*$ denotes complex conjugation transposition. The output frequency domain vector $\widetilde{\boldsymbol{S}}$ is converted to its time domain counterpart $\widetilde{\boldsymbol{s}}$ by IFFT. $\widetilde{\boldsymbol{s}}$ is the vector of the restored transmitted symbols.

The restored transmitted bits data, and the received blocks $\boldsymbol{Y}(n)$ and $\boldsymbol{Y}(n+1)$ are re-computed according to the transmission process of HF MIMO SC-FDE with the aim of revising the estimated channel responses. Bit data are re-modulated and encoded by Alamouti STBC. The encoded symbols are transformed to frequency domain by FFT. $\boldsymbol{Tx}_1$ and $\boldsymbol{Tx}_2$ denote the results of FFT. And they make up diagonal matrices $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. The new frequency domain channel responses at time n are computed as formula (19). The new time domain channel response $\widetilde{\boldsymbol{h}}_{11}(n)$ and $\widetilde{\boldsymbol{h}}_{21}(n)$ are obtained by IFFT.

$$\begin{pmatrix} \widetilde{\boldsymbol{X}} & \boldsymbol{0} \\ \boldsymbol{0} & \widetilde{\boldsymbol{X}} \end{pmatrix} \begin{pmatrix} \widetilde{\boldsymbol{H}}_{11}(n) \\ \widetilde{\boldsymbol{H}}_{21}(n) \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^*_1 & -\boldsymbol{X}_2 \\ \boldsymbol{X}^*_2 & \boldsymbol{X}_1 \end{pmatrix} \begin{pmatrix} \boldsymbol{Y}(n) \\ \boldsymbol{Y}(n+1) \end{pmatrix}$$
$$\widetilde{\boldsymbol{X}} = |\boldsymbol{X}_1|^2 + |\boldsymbol{X}_2|^2$$
$$\boldsymbol{X}_1 = diag(Tx1)$$
$$\boldsymbol{X}_2 = diag(Tx2) \tag{19}$$

It is found that the new channel response $\widetilde{\boldsymbol{h}}_{ij}(n)$ is different from the estimated channel response $\hat{\boldsymbol{h}}_{ij}(n)$. $\widetilde{\boldsymbol{h}}_{ij}(n)$ is considered closer to the real channel response than $\hat{\boldsymbol{h}}_{ij}(n)$, because it is computed through the received symbols and the transmitted symbols that is restored based on $\hat{\boldsymbol{h}}_{ij}(n)$. $\widetilde{\boldsymbol{h}}_{ij}(n)$ instead of $\hat{\boldsymbol{h}}_{ij}(n)$ is sent to the online training process as output label to suppress the accumulation error. It is also added to the channel response sequence. Then, because the Alamouti STBC requires that the channels are fixed over two consecutive blocks, i.e. $\widetilde{\boldsymbol{h}}_{ij}(n) = \widetilde{\boldsymbol{h}}_{ij}(n+1)$, the channel response window advances by one step for next channel response estimation at time $n+2$. Through the revised process shown in Fig. 9, the accumulation error caused by the typical online LSTM training-prediction process can be reduced. The channels of HF MIMO SC-FDE can be well tracked.

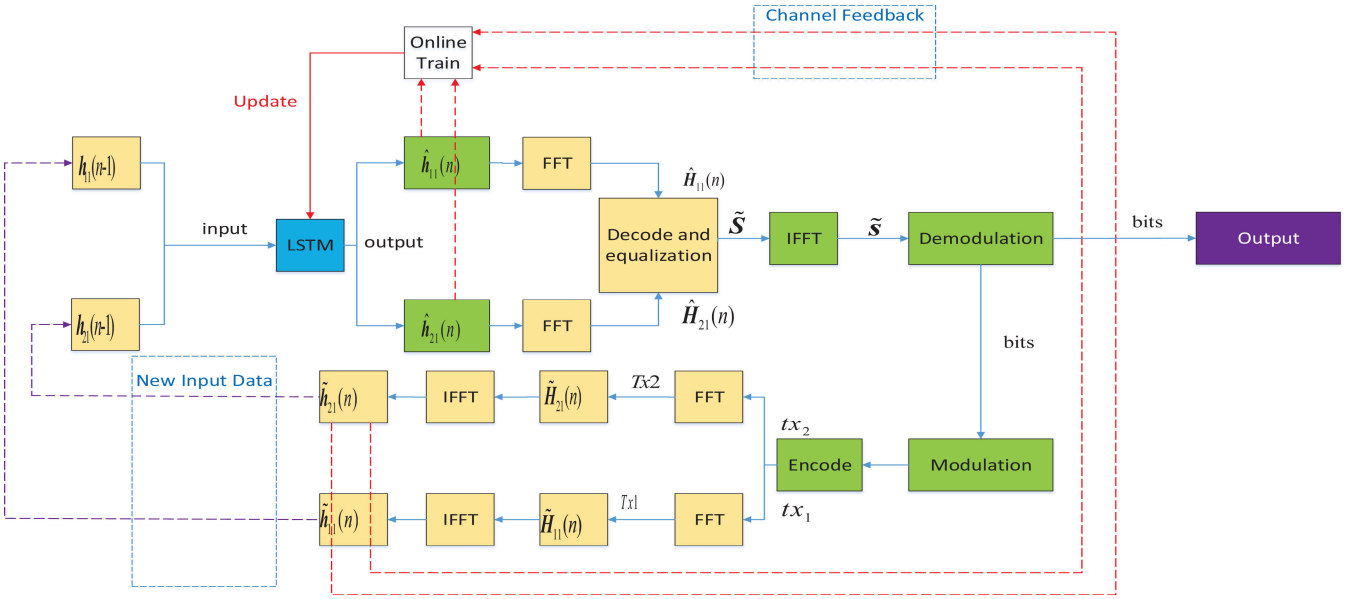


FIGURE 9. Proposed LSTM estimator algorithm diagram.

IV. EXPERIMENT RESULTS

We tested our methods on simulated system. Two kinds of RF MIMO channel models, uncorrelated and correlated, are tested in the simulation system. We also tested LS and RLS channel estimator as typical traditional methods to demonstrate the performance of our method.

A. SIMULATION EXPERIMENTS ON NON-CORRELATION CHANNELS

1) PAREMETERS SETTING

The 2×2 HF MIMO SC-FDE system parameters are listed in TABLE 1. The 2×2 HF MIMO uncorrelated channels are simulated referring to the medium latitudes channel of ITU-R F.1487 standard. This channel model simulates two multipaths with varying time delays and Gaussian Doppler power spectrum. The parameters of system frame structure are shown in Fig. 10. The bandwidth of the system is 3000Hz, which is less than the recommended bandwidth of ITU-R F.1487, 12KHz. The symbols are delivered at a rate of 4000 bauds. As shown in Fig. 10, a data block consists of 128 symbols with QPSK modulation. The CP length is set as 16. The duration of the CP is $16/2000 = 8ms$, which is more than time delay 2ms. So, the multipath interference between two consecutive data blocks can be ignored.

In order to implement online LSTM estimator, we need to transmit a group of pilot blocks for pre-training network. We constructed the pilot sequence according to section III.A. A pilot block is Zadoff-chu sequence with length of 32, and the length of both cycle prefix and postfix is the same as data block. The shift step length ϵ is equal to 16. Two circular orthogonal pilot sequences are repeated 200 times. The pilot blocks are used for channel estimation at the receiver, and the estimated channel responses are used to trains the LSTM network. Pilot blocks are followed by data blocks and the number of data blocks is set as 10000. The data blocks will be decoded using the channel response estimated by online LSTM network at the receiver.

TABLE 1. HF SC-FDE system parameter setting.

Parameter	Value
system parameter setting	3000Hz
Roll off factor	0.5
Baud rate	2000Baud/s
Sample rate	2000Hz
CP length	16
FFT	128
Pilot length	32 Zadoff-chu sequence
Frame length	200 pilot blocks+10000data blocks
Modulation	QPSK
Carrier frequency	13.86MHz
Channel model	medium latitudes quiet condition(HF-MQ): delay:0.5ms, doppler shift:0.1Hz medium latitudes moderate condition: delay(HF-MM):1ms, doppler shift:0.5Hz medium latitudes disturbed conditions(HF-MD): delay:2ms, doppler shift:1Hz

The LSTM network has one input layer, two hidden layers, one output layer. The LSTM network is trained with the aim of minimizing the mean square error (MSE) between the predicted response and the labeled response. The adaptive moment estimator (Adam) optimizer is employed for LSTM network training. Learning rate and batch size are set as 0.1 and 1, respectively. The parameters set for LSTM are listed in TABLE 2. The 200 pilot blocks are used as the training set, and the subsequent 10,000 data blocks are used as the test set. In the online learning process, we adopt the

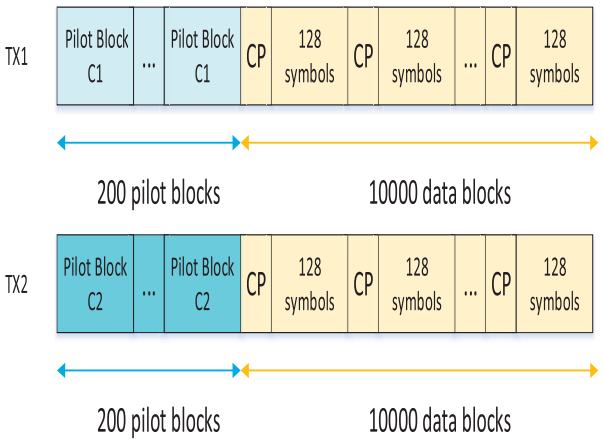


FIGURE 10. Frame Parameter of Online LSTM Estimator.

TABLE 2. online LSTM parameter setting.

Parameter	Value
Loss	MSE
Optimizer	Adam
Time Step	16
Input Dimension	8
Output Dimension	8
LSTM neurons	8
LSTM neurons	200 pilot blocks
Testing Set	10000 data blocks

mean-square error loss function shown in formula (20)

$$Loss_{MSE}(\boldsymbol{\theta}) = \frac{1}{F}\left(\sum_{n=1}^{F}\left(f(\hat{\boldsymbol{h}}(n), \boldsymbol{\theta}) - \boldsymbol{h}(n)\right)^2\right) \quad (20)$$

where F denotes the total number of samples. $f()$ is the transformation formula and $\boldsymbol{\theta}$ is all parameters of LSTM network. $f(\hat{\boldsymbol{h}}(n), \boldsymbol{\theta})$ is output of network and $\boldsymbol{h}(n)$ is actual channel response at time n. The optimal network parameters $\boldsymbol{\theta}$ are calculated with regard to the online gradient descent (OGD) [36] on the loss function. The model parameters are updated by formula (21)

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - \eta \partial Loss_{MSE}(\boldsymbol{\theta}_n) \quad (21)$$

where η is learning rate. The LSTM network is trained on an Intel (R) Core (TM) i7-7700HQ @ 2.80GHz CPU computer using PYTORCH. After training 200 pilot blocks, the LSTM network converges.

2) EXPERIMENT METHODS FOR COMPARISON

To demonstrate the performance of the proposed online LSTM estimator, LS and RLS channel estimator are also tested in our simulation environment. BER and MSE are the two parameters to evaluate the performance of the compared experimental methods.

- Ideal channel estimator
  Assume the channel responses are known at receiver. Then, the transmitted 9000 data blocks are recovered directly. The ideal channel estimator is set as a criterion for evaluating other methods.
- LS channel estimator
  LS channel estimator was tested. The channel responses of pilot blocks are estimated by LS method, and the channel responses of a data blocks are obtained through spline interpolation. The number of pilot blocks corresponds to the number of transmitter antennas. As an example, we assume there are two transmitters. The frame structure of LS method is shown in Fig. 11. In order to distinguish channel response of different channel, the pilot group includes two pilot blocks. And the first pilot block is Zadoff-chu sequence and the second pilot block is set to zero in transmitter TX1. The rule of pilot sequence construction for transmitter TX2 is opposite to that of the transmitter TX1. In our experiment, 9000 data blocks are transmitted. The pilot blocks were inserted every 40 data blocks. The total number of pilot blocks exceed that of the online LSTM estimator.

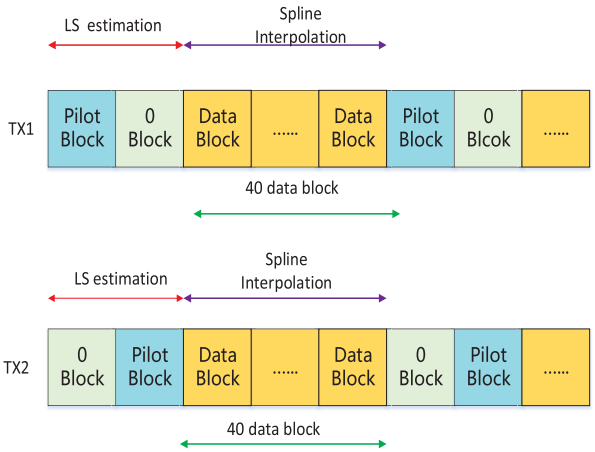


FIGURE 11. Frame Structure of LS-Interpolation.

- RLS channel estimator
  An improved Recursive Least Squares (RLS) algorithm in [37] was tested in our simulation. The traditional RLS algorithm is combined with the STBC structure. The performance of the improved RLS is the same as RLS, while the complexity is close to Least Mean Squares(LMS). The method is divided into two parts, training mode and data mode. The training mode uses pilot blocks to train the estimator until convergence. After convergence, the estimator becomes a data mode, which can decode the received data and track channel changes.
  The value of the forgetting factor λ affects the performance of RLS estimator. RLS estimator has better performance for stationary channels when λ is close to 1. However, λ is set smaller than 1 for non-stationary channels [37]. We set the forgetting factor as 0.99 when channels are assumed as stationary channels. Considering the non-stationary channels that are composed of HF-MM channel, HF-MQ channel or HF-MD channel,

we set the forgetting factor value range of [0.6, 0.99] with the increase step length of 0.01. The corresponding relationship between the data frames and the channel models is shown in Fig. 19 and we set $SNR = 20dB$. We computed the influence of different forgetting factor on BER through experiments. The experimental results are shown in the Fig. 12. When the forgetting factor is 0.72, the RLS estimator obtains the lowest BER. In the non-stationary channel experiments, λ is 0.72. In addition, we also considered using an RLS estimator with adaptive forgetting factor in non-stationary channel experiments. The algorithm proposed in [38] is used to implement an RLS filter with adaptive forgetting factor. The adaptive forgetting factor is computed as formula(22),(23) and (24)

$$\lambda_{min} = 0.72 \tag{22}$$

$$\lambda\,(n) = \lambda_{min} + (1 - \lambda_{min}) * 2^{L(n)} \tag{23}$$

$$L(n) = -round(e^2(n)) \tag{24}$$

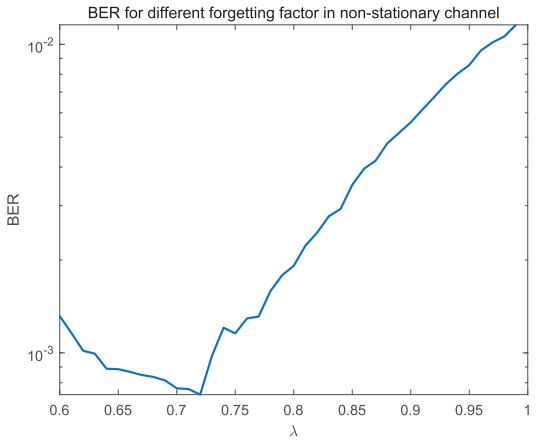


FIGURE 12. BER and forgetting factor in non-stationary channel.

where $e(n)$ is error signal and $round(e^2(n))$ represents the smallest positive integer which close to the $e^2(n)$. In non-stationary channels, when $e(n)$ is tending to infinity, $\lambda(n)$ is tending to λ_{min}. In stationary channel, when $e(n)$ is tending to 0, $\lambda(n)$ is tending to 1 [38].

Convergence is another factor that represents RLS estimator performance. Before application RLS for HF SC-FDE channel estimation, 90 blocks of training data were used to detect the appropriate length of training in transceiving process of HF SC-FDE system. The relationship between training blocks and MSE performance was computed and shown in the Fig. 13. After 6 iterations, the RLS estimator converged. Therefore, we used 6 blocks to train the RLS estimator in the transceiving process of HF SC-FDE system. Each training block contains 128(FFT points) QPSK symbols. We retrained the RLS estimator by adding 6 training blocks every 34 data blocks as described in [37] to prevent the divergence of the estimator.

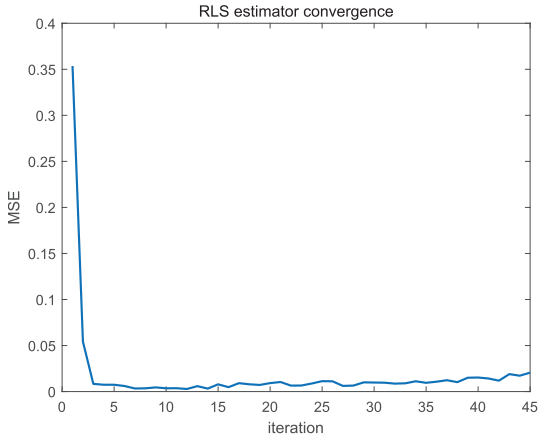


FIGURE 13. RLS Covergence.

3) SIMULATION RESULT

- Comparison of Online LSTM and LSTM
  First, we compared the performance of the online LSTM estimator and LSTM channel estimator. We simulated a HF MIMO SC-FDE system with two transmitters and two receivers. There are four channels in the system. We used medium latitudes moderate condition channel of ITU-R F.1487 standard channel model to simulate the four channels. Each channel has two multipath. The results of LSTM estimator are shown in the Fig. 14. The blue curves are the real channel response amplitude of the four channels. The red curves are the estimated results. It can be seen from Fig. 14 that the first 50 channel responses estimated by the LSTM estimator are close to the real channel responses. However, the error increases with time and the performance of LSTM estimator also decrease with time.

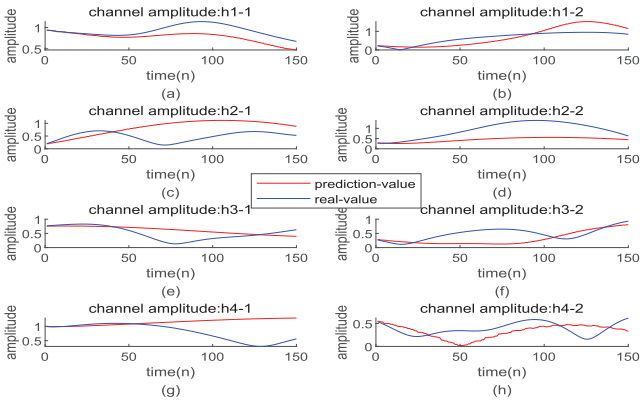


FIGURE 14. Comparison Channel Amplitude of LSTM Estimator.

Fig. 15 shows the all channel response results of the proposed online LSTM channel estimator. It can be seen that the estimated response of online LSTM estimator is close to the true channel response in more range. The presented online LSTM estimator is better than LSTM estimator.

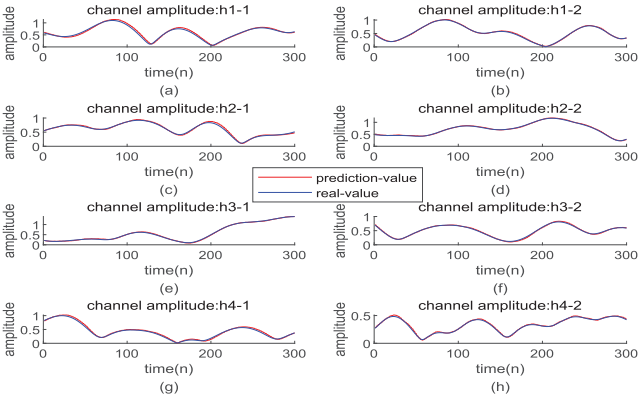


FIGURE 15. Comparison Channel Amplitude of Online LSTM Estimator.

- Comparison of Online LSTM to LS and RLS estimators
The online LSTM estimator is compared with LS and RLS estimators in terms of BER and MSE. All the experiments were established on the 2*2 HF MIMO channels of ITU-R F.1487 standard. Fig. 16 shows the results on medium latitudes moderate condition channels, and Fig. 17 exhibits the results of the medium latitudes disturbed conditions channels. The channel conditions of the latter are worse because of the larger multipath delay and doppler shift. It can be seen that the BER of the online LSTM estimator is close to the BER of the ideal channel estimator. When BER is on the 10^{-6} order of magnitude, online LSTM estimator needs $SNR = 16dB$ at receivers, while LS-Spline and RLS estimators need $SNR = 22dB$ and $SNR = 24dB$ respectively. The MSE of the online LSTM estimator is similar to the LS-Interpolation and RLS when the SNR is less than $2dB$. The reason is that the reconstructed label of the online estimator is interfered by noise. The MSE of the online LSTM estimator decreases rapidly as the SNR increases. The MSE of the online LSTM estimator is the lowest.

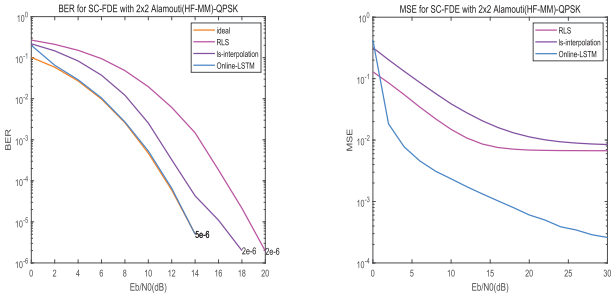


FIGURE 16. BER and MSE Comparison of Different Estimator for HF-MM.

The time varying channel tracking by different channel estimators was tested. The 2*2 channels are also simulated referring to ITU-R F.1487 standard. The transmitted symbols are modulated by QPSK. The SNR at receivers are set as 10dB. The amplitude of channel responses from the first 1000 points are drawn in Fig. 18, where $ha - b$ represent bth multipath of ath channel. The MSE values are shown in table 4.3. According to the TABLE 3, online LSTM estimator is better than LS estimator and RLS estimator. The curve of the online LSTM estimator in Fig. 18 is closer to that of the real channel responses. According to the MSE performance, online LSTM estimator is better than LS estimator and RLS estimator.

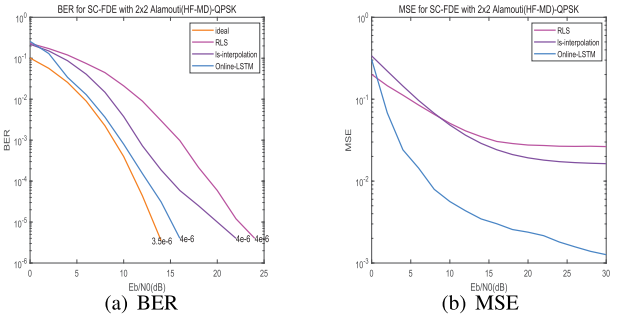


FIGURE 17. BER and MSE Comparison of Different Estimator for HF-MD.

TABLE 3. MSE comparison.

Channel	Estimator	MSE
h1-1	Online LSTM	0.001765
	LS	0.0021306
	RLS	0.0037913
h1-2	Online LSTM	0.0010272
	LS	0.0020705
	RLS	0.0031526
h2-1	Online LSTM	0.0014343
	LS	0.0022384
	RLS	0.0037295
h2-2	Online LSTM	0.0009898
	LS	0.0018242
	RLS	0.0032809
h3-1	Online LSTM	0.0009694
	LS	0.0011364
	RLS	0.0024252
h3-2	Online LSTM	0.0012644
	LS	0.0017063
	RLS	0.0030344
h4-1	Online LSTM	0.0010304
	LS	0.0014693
	RLS	0.0028646
h4-2	Online LSTM	0.0008425
	LS	0.0014721
	RLS	0.0023079

- Evaluation generalization capability
In order to show the generalization capability of online LSTM estimator, we implemented experiments on non-stationary channels. As shown in Fig. 19, we generated six channels for six consecutive data frames as shown in Fig. 19. Each of channels was selected randomly from the HF-MQ channel, HF-MM channel or HF-MD channel. Each time varying channel was simulated with one alternative set of channel parameters that involve different multipath delay and Doppler shift.

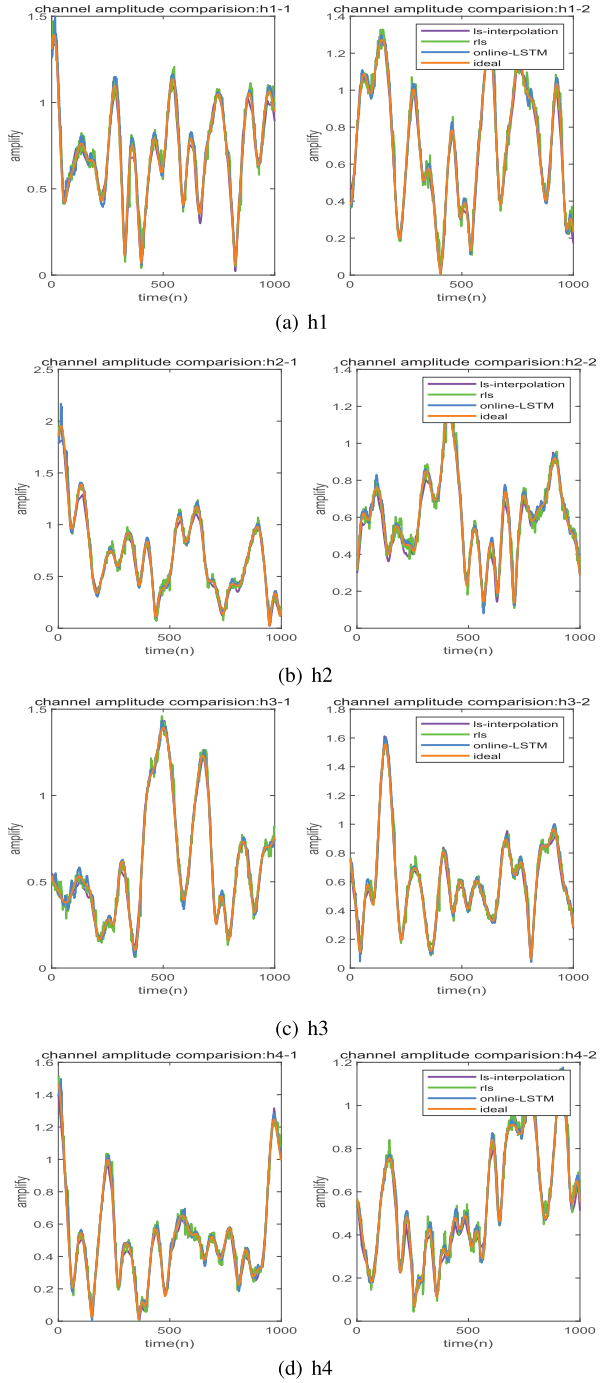


FIGURE 18. Channel Tracking of Different Estimators on QPSK data.

Each channel corresponds to a frame of data. Under the condition, we compared the performance of online LSTM with LSTM, LS and RLS estimators.

Firstly, we compared the performance of LSTM and online LSTM estimator at the same SNR. Each data frame contains only 200 data blocks. The result of LSTM estimator is shown in the Fig. 20. The LSTM estimator can only track channel changes over part of the time. Due to accumulated errors, LSTM prediction

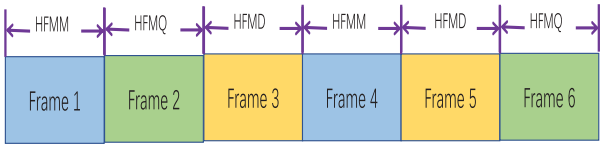


FIGURE 19. Data Frame for Non-Stationary Channel.

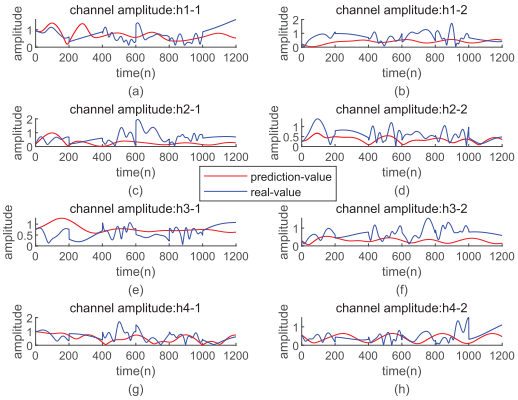


FIGURE 20. Channel Amplitude of LSTM Estimator.

results will gradually deviate from the actual channel. Lack of online learning makes LSTM difficult to deal with the changes of channel parameters in time. The BER of the LSTM estimator remained 0 at the first 33 data blocks, but then BER performance deteriorated, and it reached $BER = 0.37$ when the transmission of 200 blocks was finished. LSTM estimator can not keep good tracking performance in non-stationary channels.

Fig. 21 shows the channel track result of online LSTM estimator. Even under non-stationary channel conditions, online LSTM estimator can well track the channel change. The BER kept 0 during the transmission of the 200 data blocks.

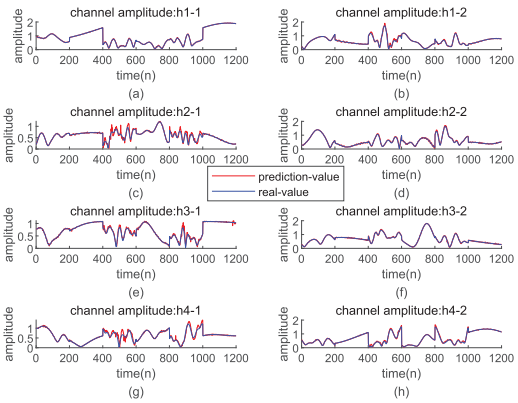


FIGURE 21. Channel Amplitude of online LSTM Estimator.

Then we compared the BER performance of online LSTM, LS and RLS estimator. Fig. 22 shows that

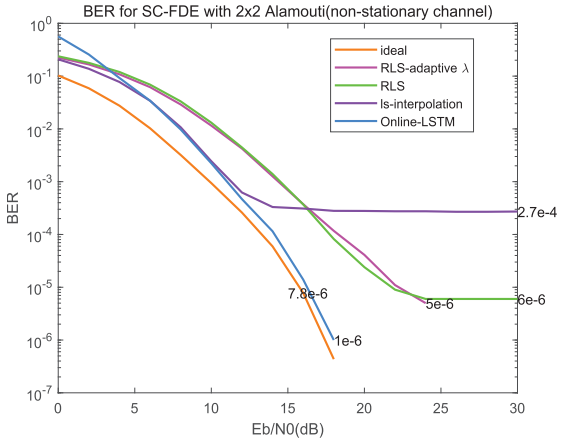


FIGURE 22. BER Comparison of Different Estimators for Non-stationary Channels.

the online LSTM estimator has best performance on non-stationary channels. The BER curve of LS estimator appears flat since the SNR is 16dB, and the improvement is limited as the SNR increases. Both RLS estimator and RLS estimator with adaptive forgetting factor have been tested. Their BER curves are similar when the SNR is no more than 22dB. When the SNR is more than 22, the BER curve of RLS estimator remains flat, while the BER of adaptive RLS estimator reach 0. When BER is on the 10^{-6} order of magnitude, online LSTM estimator needs $SNR = 18dB$, while RLS estimators with adaptive forgetting factor need $SNR = 24dB$. The experiment results on changed channel models also demonstrate that the online LSTM estimator has generalization capability.

- Transmission rate comparison
  We compared the efficient transmission rate of different channel estimator. The transmission rate can be calculated as

$$v\_tr = Baud * order * \frac{Nd}{Nd + Np} \tag{25}$$

where v_tr, $Badu$, Nd and Np represent transmission rate, baud rate, number of data and number of pilot, respectively. According the formula(25) and frame structure, we can calculate the transmission rate as shown in TABLE 4. All channel estimator have the same Baud rate, but online LSTM estimator adopts the circular orthogonal sequences as pilot. It does not need extra pilot blocks to distinguish channel response of different path compared with LS estimator. Although pilot blocks are inserted into frame header by the online LSTM estimator, it does not need to insert pilot blocks repeatedly such as LS estimator, because this estimator can predict next time channel response. On the whole, online LSTM estimator can improve the BER and MSE performance of HF MIMO SC-FDE system. Meanwhile, it can improve the transmission efficiency when transmitting mass data.

TABLE 4. transmission rate comparison.

Estimator	Transmission Rate
Online LSTM	3.524kbps
LS	3.386kbps
RLS	3.022kbps

- Complexity Comparison
  Complexity is one of the important indicators to measure the performance of channel estimators. We selected time and space consumption to compare the complexity among online LSTM, LS and RLS estimators. For the LS estimator, we need K times of complex division to obtain the channel frequency domain responses of the pilot position. Then the channel frequency domain responses of the N_B data blocks are obtained by interpolation. So the time and space complexity of the LS estimator depends on the product of the number of FFT points K and the number of transmitted data blocks N_B. For the RLS estimator, diagonal matrix is introduced into recursion formula, and complex matrix inversion is not needed due to the combination of RLS and STBC coding properties [37]. The time consumption of RLS estimator mainly depends on the product of the number of FFT points K and the number of transmitted data blocks N_B. Unlike LS, RLS estimator also needs to calculate weight coefficient vectors, error vectors, gain vectors, etc. So its time complexity is higher than LS estimator. But for space complexity, the RLS estimator depends only on the number of FFT points K because RLS is a recursive algorithm. The results of the last calculation instead of all the FFT results such as LS estimator should be saved to deduce the current results. So, when the number of data blocks N_B is large, the advantage of the RLS estimator to the LS estimator on space complexity will be displayed. For the online LSTM estimator, the circular orthogonal sequence is required to estimate the channel response at the pilot position firstly. Because this is the time-domain channel response, the required time and space both depend on the channel response length L. In general, the channel response length is less than the number of FFT points K. Then the estimated channel responses will be used to train the LSTM network. The time and space complexity of training part depends on the LSTM network size, such as input and output dimensions, the number of hidden layer neurons and time step. Finally, during the online prediction stage, the time complexity of online LSTM estimator depends on network size and the number of data blocks N_B. Because online LSTM is also a recursive algorithm in nature, no additional storage space is consumed in this stage. In summary, the time complexity of online LSTM estimator depends on network size, channel response length L and number of data blocks N_B, while the space complexity mainly depends on network size and channel response length L. The LSTM estimator is similar to

the online LSTM estimator, but the main difference is the traditional LSTM does not have the online learning process at prediction stage. So the time complexity of LSTM estimator is less than the online LSTM. However, the time cost of online LSTM estimator is a worthwhile. Because it improved BER performance and channel tracking performance dramatically.

In addition, in order to intuitively compare the complexity, we counted the time and space consumed by the three estimators in the experiment just like in [39]. We use three channel estimators on the same computer to estimate the channel response of a frame of data. Except for the estimator, the other conditions remain the same. The experiment results are shown in Table 5. LS estimator needs 12.465000s and 74.6537M bytes for a frame data. RLS estimator needs 47.104000s and 0.7025M bytes for a frame data. For online LSTM estimator, the time consumption of the three stages was counted respectively. The time domain estimated consumption time of 200 pilot blocks was 0.031s, the LSTM network was trained to consume 1.73s, and the online estimated consumption of 10000 data blocks was 37.85s. Online LSTM estimator needs 40.6024s and 0.3544M bytes for a frame data. In terms of time consumption, LS estimator consumes the least time. RLS and online LSTM estimator consume the similar time, but they are all 4 times as much as LS estimator. In terms of space consumption, online LSTM estimator consumes the least space. However, LS estimator consumes the most space because it needs to store the channel frequency domain response of all data.

TABLE 5. Complexity comparison.

Estimator	Time consume	Memory consume
LS	12.465000s	74.6537MB
RLS	47.104000s	0.7025MB
Online LSTM	0.031+1.73+37.85=40.6024s	0.3544MB

B. EXPERIMENTS ON CORRELATED MIMO CHANNELS

We applied the established 2*2 HF MIMO SC-FDE system on the correlated channels. Except the channel model, the other parameters are the same as that in TABLE 1. The correlated channels are built with regard to [10]. We used linear antenna arrays to realize the MIMO as shown in Fig. 23. The interval between adjacent antennas was set as 3 meters, 5 meters and 7 meters, respectively. The corresponding correlation coefficients were set as 0.82, 0.54, and 0.21 based on formula (11), (12) and(13).

The BER of the online LSTM estimator on the correlated channels is shown on Fig. 24. The lower correlation, the lower BER. We also compared the online estimator with other methods on the correlated channels. The correlation coefficient was set as 0.54, and the data were modulated by QPSK. Fig. 25 shows that when the SNR is lower than 10dB, the BER

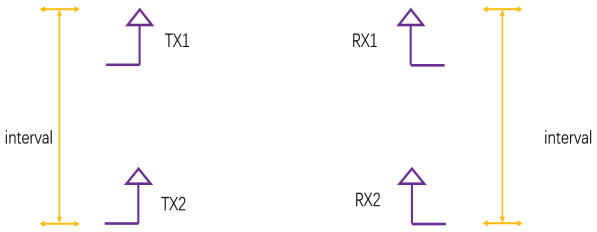


FIGURE 23. Linear Antenna Array of HF MIMO System.

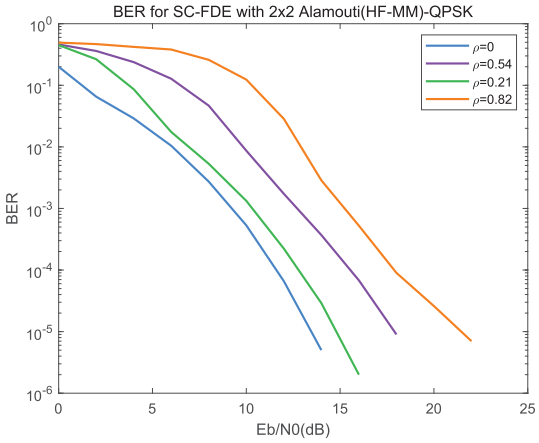


FIGURE 24. BER Comparison of Different correlation for online LSTM Estimator.

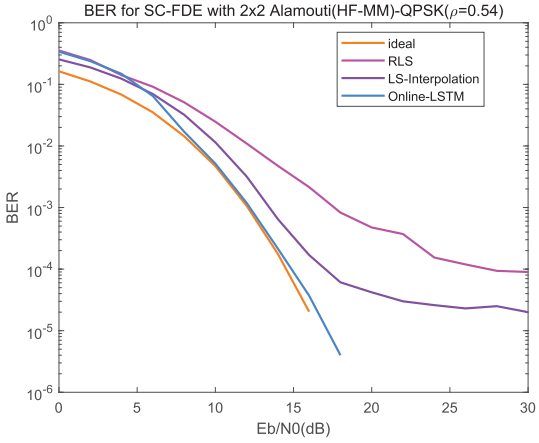


FIGURE 25. BER Comparison of Different Estimator for Correlated MIMO Channel.

of the online LSTM does not exceed that of the LS and RLS estimators. When the SNR is higher than 10, the BER of the online LSTM outperforms that of the LS and RLS estimators, and it is close to that of the real channel response.

The simulation results demonstrate that our method can be used in physical HF systems. To establish a HF MIMO-SC-FDE system, some function modules, such as IQ imbalance and synchronization, should be added besides the modules shown in Fig. 1. We will build an experimental system in the near future.

Although our research focuses on the HF communication, the presented method can be used in other bands. To adapt to other bands, the only change is made on the parameters of Table 1 and Table 2. If our method is applied to the lower frequency, the online LSTM estimator may achieve better performance due to the longer wavelength and more stationary channel.

V. CONCLUSIONS

Long distance HF communication suffers from time varying channel. MIMO and SC-FDE have been introduced to HF communication to combat the multipath, avoid wave distortion at transmitters and improve the data rate. We designed the architecture of HF MIMO SC-FDE and the signal flow chart. Channel estimation is essential in the HF MIMO SC-FDE system. Restoring the transmitted data at the receivers depends on the accuracy of channel estimation.

We present an online LSTM channel estimator for HF MIMO SC-FDE system. Different from the training and prediction loop of the LSTM estimator that is built on the sliding window of channel response series, the online LSTM channel estimator uses the received data symbols and the restored transmitting data symbols to re-compute the channel responses in terms of the communication process of HF MIMO SC-FDE. The corrected channel responses are added to the channel response series for next LSTM network training and next channel response prediction.

The online LSTM network is firstly trained by the pilots and their channel responses. The channel responses of pilots are estimated by the circular orthogonal sequences based channel estimation. The pilots are constructed referring to the Zadoff-chu sequence. The configured pilots between different transmitters are orthogonal, making the pilots from different transmitters can be recognized. After training on the pilot blocks, the online LSTM estimator is applied to the channel estimation of data blocks.

We established simulation system. The introduced online LSTM estimator was compared with LS, RLS and ideal channel estimation methods. The results of simulation show that BER and MSE of the online LSTM estimator are lower than other methods. The online LSTM estimator outperforms LS and RLS methods. It has potentiality in long distance HF MIMO communication.

REFERENCES

[1] G. R. Kadambi, K. Krishnan, and B. R. Karthikeyan, "HF channel estimation for MIMO systems based on particle filter technique," *J. Commun.*, vol. 5, no. 9, pp. 674–683, Oct. 2010.

[2] Y. Erhel, D. Lemur, M. Oger, J. Le Masson, and F. Marie, "Evaluation of ionospheric HF MIMO channels: Two complementary circular polarizations reduce correlation.," *IEEE Antennas Propag. Mag.*, vol. 58, no. 6, pp. 38–48, Dec. 2016.

[3] S. D. Gunashekar, E. M. Warrington, S. Salous, S. M. Feeney, N. M. Abbasi, L. Bertel, D. Lemur, and M. Oger, "Investigations into the feasibility of multiple input multiple output techniques within the HF band: Preliminary results," *Radio Sci.*, vol. 44, no. 1, pp. 1–5, Feb. 2009.

[4] M. P. Scheible, L. J. Teig, J. D. Fite, K. M. Cuomo, J. L. Werth, G. W. Meurer, N. C. Ferreira, and C. R. Franzini, "High data rate, reliable wideband HF communications demonstration," MITRE Corp., Tech. Paper, 2014.

[5] S. B. Sahay, K. C. Bhagwat, and P. R. J. Mohan, "Exploitation of MIMO techniques for reliable HF communication," in *Proc. Int. Conf. Signal Process. Commun. (SPCOM)*, Jul. 2012, pp. 1–4.

[6] R. C. Daniels and S. W. Peters, "A new MIMO HF data link: Designing for high data rates and backwards compatibility," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Nov. 2013, pp. 1256–1261.

[7] F. Pancaldi, G. Vitetta, R. Kalbasi, N. Aldhahir, M. Uysal, and H. Mheidat, *Single-Carrier Frequency Domain Equalization*. Hoboken, NJ, USA: Wiley, 2008.

[8] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Commun. Mag.*, vol. 40, no. 4, pp. 58–66, Apr. 2002.

[9] P. M. Ndao, Y. Erhel, D. Lemur, and J. Masson, "Design of a high-frequency (3–30 MHz) multiple-input multiple-output system resorting to polarisation diversity," *IET Microw., Antennas Propag.*, vol. 5, no. 11, pp. 1310–1318, Aug. 2011.

[10] D. Lemur, Y. Erhel, M. Oger, and J. Le Masson, "MIMO channel measurements on a ionospheric HF (3–30 MHz) radio link," in *Proc. IWSSIP*, May 2014, pp. 191–194.

[11] Y. Nechaev, I. Dvorakova, and A. Malyutin, "Fast channel estimation method for HF serial-tone MIMO system," in *Proc. 38th Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2015, pp. 1–5.

[12] D. Li, Y. Liu, and H. Zhang, "Circular orthogonal sequences based channel estimation for MIMO SC-FDE systems," in *Proc. Int. Conf. Wireless Commun. Signal Process.*, Nov. 2009, pp. 1–4.

[13] F. Arikan, "A brief review of HF channel response estimation," *J. Electromagn. Waves Appl.*, vol. 18, no. 6, pp. 837–851, Jan. 2004.

[14] J. Wang, G. Ding, and H. Wang, "HF communications: Past, present, and future," *China Commun.*, vol. 15, no. 9, pp. 9–17, 2018.

[15] H. Jie and Y. Ling, "Semi-blind channel estimation of MIMO-OFDM systems based on RBF network," in *Proc. IET Int. Commun. Conf. Wireless Mobile Comput. (CCWMC)*, 2011, pp. 187–191.

[16] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 852–855, Oct. 2018.

[17] J. Zhang, C.-K. Wen, S. Jin, and G. Y. Li, "Artificial intelligence-aided receiver for a CP-free OFDM system: Design, simulation, and experimental test," *IEEE Access*, vol. 7, pp. 58901–58914, 2019.

[18] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, "Deep learning-based channel estimation," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 652–655, Apr. 2019.

[19] H. Wang and J. Li, "Development of intelligent receiver for MIMO-OFDM system," in *Proc. Int. Conf. Comput., Commun. Signal Process. (ICCCSP)*, Jan. 2017, pp. 1–5.

[20] P. Gogoi and K. K. Sarma, "Recurrent neural network based channel estimation technique for STBC coded MIMO system over Rayleigh fading channel," in *Proc. CUBE Int. Inf. Technol. Conf. (CUBE)*. New York, NY, USA: Association for Computing Machinery, 2012, pp. 294–298, doi: 10.1145/2381716.2381771.

[21] M. N. Seyman and N. Taşpınar, "Channel estimation based on neural network in space time block coded MIMO–OFDM system," *Digit. Signal Process.*, vol. 23, no. 1, pp. 275–280, Jan. 2013.

[22] M. Jiang, C. Li, H. Li, and D. Yuan, "Channel tracking based on neural network and particle filter in MIMO-OFDM system," in *Proc. Int. Conf. Natural Comput. (ICNC)*, vol. 5, Oct. 2008, pp. 192–196.

[23] J. Liu, K. Mei, X. Zhang, D. Ma, and J. Wei, "Online extreme learning machine-based channel estimation and equalization for OFDM systems," *IEEE Commun. Lett.*, vol. 23, no. 7, pp. 1276–1279, Jul. 2019.

[24] H. He, S. Jin, C.-K. Wen, F. Gao, G. Y. Li, and Z. Xu, "Model-driven deep learning for physical layer communications," *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 77–83, Oct. 2019.

[25] F. Liang, W. G. Hatcher, G. Xu, J. Nguyen, W. Liao, and W. Yu, "Towards online deep learning-based energy forecasting," in *Proc. 28th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Jul. 2019, pp. 1–9.

[26] T. Ergen and S. S. Kozat, "Online training of LSTM networks in distributed systems for variable length data sequences," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 5159–5165, Oct. 2018.

[27] T. Ergen and S. Serdar Kozat, "Efficient online learning algorithms based on LSTM neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3772–3783, Aug. 2018.

[28] D. Chu, "Polyphase codes with good periodic correlation properties (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 4, pp. 531–532, Jul. 1972.

[29] N. Al-Dhahir, "Single-carrier frequency-domain equalization for space-time-coded transmissions over broadband wireless channels," in *Proc. 12th IEEE Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, vol. 1, Oct./Sep. 2001, p. B.

[30] C. Watterson, J. Juroshek, and W. Bensema, "Experimental confirmation of an HF channel model," *IEEE Trans. Commun.*, vol. TCOMM-18, no. 6, pp. 792–803, Dec. 1970.
[31] J. A. Hoffmeyer, L. E. Vogler, J. F. Mastrangelo, L. E. Pratt, and C. J. Behm, "A new HF channel model and its implementation in a real-time simulator," in *Proc. Int. Conf. Hf Radio Syst. Techn.*, 2002, pp. 173–177.
[32] *Testing of HF Modems With Bandwidths of up to About 12 kHz, Using Ionospheric Channel Simulators*, document Recommendation ITU-R F.1487(05/2000), IRF Recommendation, 2000.
[33] X. Liu and M. E. Bialkowski, "Optimization of training signal transmission for estimating MIMO channel under antenna mutual coupling conditions," *Int. J. Antennas Propag.*, vol. 2010, pp. 1–10, 2010.
[34] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[35] G. Bontempi, S. B. Taieb, and Y. A. L. Borgne, *Machine Learning Strategies for Time Series Forecasting*. Berlin, Germany: Springer, 2013.
[36] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *Proc. ICML*, 2003, pp. 928–935.
[37] W. M. Younis, A. H. Sayed, and N. Al-Dhahir, "Efficient adaptive receivers for joint equalization and interference cancellation in multiuser space-time block-coded systems," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2849–2862, Nov. 2003.
[38] L. Wang, J. Li, and K. Huang, "Improved RLS algorithm in use of MIMO-OFDM system," in *Proc. Int. Conf. Meas., Inf. Control*, vol. 2, May 2012, pp. 879–882.
[39] P. Jiang, T. Wang, B. Han, X. Gao, J. Zhang, C.-K. Wen, S. Jin, and G. Y. Li, "Artificial intelligence-aided OFDM receiver: Design and experimental results," 2018, *arXiv:1812.06638*. [Online]. Available: https://arxiv.org/abs/1812.06638

ZHIYONG WANG received the B.E. degree in communication engineering from Wuhan University, Wuhan, China, in 2018, where he is currently pursuing the M.S. degree in information and communication engineering. His research interests include machine learning and MIMO communications.

FANGLING PU (Member, IEEE) received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, in 2005. She was a Research Assistant with the Institute for Theoretical Information Technology, RWTH Aachen University, in 2007. Since 2000, she has been an Associate Professor with the School of Electronic Information, Wuhan University. Her research interests include system and networking, signal and information processing, robotics, and artificial intelligence.

XIAOSHI YANG received the B.E. degree in electronic and information engineering from Wuhan University, Wuhan, China, in 2018, where he is currently pursuing the M.S. degree in information and communication engineering. His research interests include machine learning and wireless communications.

NING CHEN received the B.E. degree in communication engineering from Wuhan University, Wuhan, China, in 2019, where she is currently pursuing the M.S. degree in electronic and communication engineering. Her research interests include MIMO communication systems and software defined networks.

YONGMIN SHUAI received the Ph.D. degree in communication and information system from Wuhan University, Wuhan, China, in 2011. He is currently a Senior Engineer with the Wuhan Maritime Communication Research Institute, Wuhan. His research interest includes signal processing and its applications in digital communication systems.

RUI YANG received the M.S. degree in signal and information processing from Xidian University, Xi'an, China, in 2012. She is currently a Senior Engineer with the Wuhan Maritime Communication Research Institute, Wuhan. Her research interest includes signal processing and its applications in digital communication systems.