Low Complexity Channel Prediction Using TFOS-ELM Method for Massive MIMO Systems

TONGTONG CHENG¹, YIGANG HE²,³, WEI HE¹, LUQIANG SHI¹, YONGBO SUI¹, AND YUAN HUANG¹

¹Electrical Engineering Department, Hefei University of Technology, Hefei 232000, China
²School of Electrical Engineering, Wuhan University, Wuhan 430072, China
Corresponding author: Yigang He (18655136887@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51977153 and Grant 51577046, in part by the State Key Program of National Natural Science Foundation of China under Grant 51637004, in part by the National Key Research and Development Plan “Important Scientific Instruments and Equipment Development” under Grant 2016YFF0102200, and in part by the Equipment Research Project in advance Grant 41402040301.

ABSTRACT Multiple-input multiple-output (MIMO) technology can potentially help to achieve high data rates for multiuser communication. To achieve better performance, the channel state information (CSI) is estimated by the pilot. However, the estimated CSI cannot be used in downlinks when the mobile speed is very high, since it becomes outdated due to the rapid channel variation. In a massive MIMO system, the issue of outdated CSI is serious when using traditional techniques. Therefore, in order to obtain accurate CSI, the prediction of future CSI is required. In this paper, a low complexity online extreme learning machine (ELM) is proposed for the online prediction of the fast fading channel. First, we introduce the structure of the online sequential extreme learning machine (OS-ELM) and combine the training process of the OS-ELM with a forgetting mechanism (FM) to predict fast changing channels. Second, we use the truncated polynomial expansion (TPE) to reduce the computational complexity of the OS-ELM with the FM (FOS-ELM). In addition, the performance of the proposed algorithm is verified through simulation results, and we apply channel prediction in the precoding process. It is shown that the communication quality is improved by our channel prediction algorithm.

INDEX TERMS Massive MIMO, OS-ELM, TFOS-ELM, channel prediction, precoding, low complexity, MSE and BER.

I. INTRODUCTION
Multiple-input multiple-output (MIMO) wireless communication improves the transmission rate by increasing the number of transmitter and receiver antennas [1], [2]. However, those results rely on perfect knowledge of the channel state information (CSI) at the transmitter and receiver. But in the fast time-varying channel, the CSI can be outdated due to the delay of computations at the base station (BS), which might cause significant system performance degradations. There has been considerable work conducted on this issue, and the channel prediction has received significant critical attention in wireless communication research.

Channel prediction is beneficial for communication performance, and topics include adaptive coding, modulation, decoding processes, and channel equalization [3]–[9]. In wireless communication links, channel prediction helps the communication system to deploy physical-layer parameters and schemes [3]. The improper deployment of modulation schemes causes higher bit-error-rate (BER) or the lower data throughput. In addition, channel prediction is favorable for radio resource allocation and interference management. Therefore, channel prediction is essential for a wide range of wireless communication technologies [9].

In the past few years, several researches were conducted on channel prediction. The existing researches on channel prediction fall into three main categories: 1) the auto-regressive (AR) model [3], [10], [11], 2) the sum-of-sinusoids (SOS) [12]–[15], [17], 3) the neural network model [18]–[20]. The AR model, which adopts the AR algorithm, uses the outdated CSI to predict the future CSI. The greatest advantage is that the AR model does not need to know

The associate editor coordinating the review of this manuscript and approving it for publication was Usama Mir.
the information of the physical parameters. The AR model only uses the outdated CSI to calculate the AR coefficients. Duel-Hallen [3], [10] employed an AR model to trace the channel changes and computes the MMSE estimates of the future CSI based on the outdated CSI. Since the AR model is a linear prediction model [16], it has wonderful effect and low computational complexity in slow changing channel, but the performance of the AR model is limited by fast changing channels.

The SOS model begins with an approximation of the physical propagation process. The wireless channel is modeled as a sum of complex sinusoids. Unlike the AR model, the outdated CSI is not directly used by the SOS model. The parameters of the SOS model are estimated using outdated CSI. J. K. Hwang proposed MUSIC to estimate the physical parameters [17]. M. Chen investigated ESPRIT to predict future CSI [15]. If the parameters of the wireless channel do not vary with time, the SOS model has excellent predictive performance. However, the channel parameters change in practical communications, which causes performance degradation as time passes. The feasible measure is frequently used to estimate the parameters of the SOS model, but it increases the computational complexity.

Recently, neural networks become major channel prediction methods. Navabi et al. [18] assessed the viability of using a backpropagation (BP) neural network technique for estimating the user-channel feature. Potter et al. [19] proposed a recursive neural network (RNN) method for channel prediction. Jhihoon Joo offered a deep learning-based prediction model for predicting future channel information [20]. Neural networks provide excellent prediction performance for fast changing channels. However, the traditional BP neural network, RNN and deep learning require a large number of computational complexity to calculate the parameters of the fading channel. Therefore those neural networks are not applicable to online channel prediction.

With the coming of fifth generation mobile communication systems (5G), massive MIMO is utilized to promote the spectral efficiency and data throughput of communication systems by deploying large scale arrays with multiple antennas at the BS [1], [21], [22]. However, the traditional channel prediction methods have a large number of matrix operations with high computational complexity, which impose formidable pressure on the user due to the limited computational resources.

In this paper, the online sequential extreme learning machine with the forgetting mechanism and truncated polynomial expansion (TFOS-ELM) is proposed. We first apply a novel prediction algorithm developed from [23], [24], [26], which takes both neural network structure and computational speed into account. Our algorithm is based on the online sequential extreme learning machine (OS-ELM) with the forgetting mechanism (FOS-ELM). The TFOS-ELM algorithm applies in the online channel prediction with a low computational burden. In addition, the TFOS-ELM algorithm has a wonderful channel prediction effect for fast changing channels. We give more detailed analysis in the simulation results. Moreover, our algorithm is applied in the precoding stage under, and the results show that channel prediction can improve the precoding performance in a massive MIMO system.

The main contribution of our work is first to utilize OS-ELM for online channel prediction in a Massive MIMO system. The online channel prediction has more accuracy than off-line channel prediction in real communication environment. The FOS-ELM can avoid the influence of obsolete data but it causes huge complexity. Therefore, we exploit the TPE method to avoid the complicated matrix inversion. The TFOS-ELM model utilizes lower computational complexity to realize online channel prediction. Due to low complexity, the TFOS-ELM is easy to apply in online channel prediction for fast changing channels.

The rest of this paper is organized as follows. Section II briefly describes the massive MIMO systems, the prediction procedure and channel datasets. In Section III, the OS-ELM and FOS-ELM are introduced simply. In Section IV, we propose a channel prediction framework for the TFOS-ELM model. In Section V, the performance analysis and simulation results are provided. Finally, the paper is concluded in Section VI.

II. SYSTEM MODEL

A. CHANNEL MODEL

Multipath delay can be expressed as some channel impulse responses (CIRs) between BS antennas and user antennas in the massive MIMO system [1], [21], [22]. The multipath delay fading channel can be modeled as [3]:

\[ g_{m,n_r}(t) = \sum_{p=1}^{P} A_p \exp(j2\pi f_p t + \phi_p) \]  

(1)

where \( P \) is the number of reflection paths or obstacles between the \((m, n_r)\)-antenna pair; \( N_t \) and \( N_r \) are the transmitter and receiver antenna, respectively; and \((1 \leq m \leq N_t, 1 \leq n_r \leq N_r)\). For the \(p\)-th path, \(A_p, f_p, \) and \( \phi_p \) are the amplitude, Doppler frequency, and initial phase, respectively. The Doppler frequency shift is given as follows:

\[ f_p = f_c \frac{v}{c} \cos \theta_p \]

(2)

where \( f_c \) is the carrier frequency, and \( v \) is the speed of the vehicle. \( c \) is the speed of light. \( \theta_p \) is the angle between the radio wave and the direction of the vehicle. The fast changing channel has a large \( v \). The parameters \(A_p, f_p, \) and \( \phi_p \) are very slowly time-varying, and all of them are viewed as relevant in a coherent time period. Thus, the signal \(g_{m,n_r}(t)\) in Eq.(1) is converted into a superposition of complex sinusoids as

\[ g_{m,n_r}(t) = \sum_{p=1}^{P} A_p \exp(j2\pi f_p t + \phi_p) \]

\[ = \sum_{p=1}^{P} A_p \cos(2\pi f_p t + \phi_p) + j \sum_{p=1}^{P} A_p \sin(2\pi f_p t + \phi_p) \]

(3)
Therefore \( g_{m,n}(t) = g_{m,n,1}(t) + jg_{m,n,Q}(t) \)

where

\[
g_{m,n,1}(t) = \sum_{p=1}^{P} A_p \cos(2\pi f_p t + \phi_p) 
\]

\[
g_{m,n,Q}(t) = \sum_{p=1}^{P} A_p \sin(2\pi f_p t + \phi_p) 
\]

The discrete-time channel is formed by sampling the wireless channel over a fixed time as

\[
g_{m,n,k} = g_{m,n,1}(k) + jg_{m,n,Q}(k) 
\]

where \( g_{m,n,k} \) is the discrete-time complex channel value. \( g_{m,n,k} \) is expressed as the real part \( g_{m,n,1}(k) \) and the imaginary part \( g_{m,n,Q}(k) \), respectively. \( g_{m,n,k} \) is the complex-value of the channel information of the \((m_1, n_1)\)-antenna pair at time \( k \).

In an OFDM (Orthogonal Frequency Division Multiplexing) system, the pilot structure has three different types, including the block type, the comb type, and the lattice type. In our paper, the block-type pilots are arranged to estimate the CSI. \( S_t \) is the period of the pilot symbols with respect to time. The pilot symbol period satisfies the following inequality

\[
S_t \leq 1/f_d.
\]

Then, the channel estimation matrix can be depicted as follows.

\[
g(k) = \begin{bmatrix}
g(1.1)f(k) & \cdots & g(1.N_1)f(k) \\
\vdots & \ddots & \vdots \\
g(N_1,1)f(k) & \cdots & g(N_1,N_1)f(k) \\
\end{bmatrix} 
\]

\[
+ j \begin{bmatrix}
g(1.1)Q(k) & \cdots & g(1.N_1)Q(k) \\
\vdots & \ddots & \vdots \\
g(N_1,1)Q(k) & \cdots & g(N_1,N_1)Q(k) \\
\end{bmatrix}
\]

\[
(7)
\]

B. DESCRIPTION OF THE SLFN STRUCTURE

The traditional ELM is a simple structure of SLFNs [23] and its network structure is shown in Figure 1. However, the ELM requires that all channel datasets of the SLFN are ready for learning.

The SLFNs can be described as follows [23]:

\[
f(x) = \sum_{i=1}^{\tilde{N}} \beta_i F(a_i, b_i, x), \quad x \in \mathbb{R}^n
\]

where \( a_i \in (0, 1) \) is the weight vector between the input layer and the \( i \) th hidden node. \( \tilde{N} \) is the number of SLFN hidden nodes. \( b_i \in (-1, 1) \) is the bias of the \( i \) th hidden node. \( \beta_i \) is the weight vector, which connects hidden layer and output layer; and \( F(a_i, b_i, x) \) is the output of the \( i \) th hidden node with respect to the input. \( F(a_i, b_i, x) \) is denoted by \( F(a_i x + b_i) \), where \( a_i \cdot x \) denotes the inner product of \( a_i \) and \( x \); and \( F(a_i x + b_i) \) indicates the activation function including the “Sigmoid”, “Sine”, “RBF”, “Hardlim” and so on.

C. STRUCTURE OF THE CHANNEL DATASETS

Before channel prediction model based on SLFN, we need consist channel datasets which give us a way to easily predict channel in the OS-ELM, FOS-ELM and TFOS-ELM model. The sliding window is applied to consist channel dataset and training model in channel prediction. In online channel prediction, Channel datasets improve the accuracy of prediction and suit the structure of prediction model.

As shown in Figure 2.(a), the length of sliding window is \((k - i + 1)\), \( M \) is the step of sliding window. The blue dotted box is sliding window at time \( T_k \), the gray dotted box is sliding window at time \( T_{(k-1)} \). It means that the parameters of prediction model are updated when every \( M \) CSI are received.

As shown in Figure 2.(b), the structure of channel datasets are consisted in time \( T_k \). The \((v + 1)\) represents the size of channel dataset \((x, t)\) and \( v \) is the number of SLFN input nodes. The input of channel dataset \( x_i \) and output of channel dataset \( t_i \) are represented as follows.

\[
x_i = \{g_i, g_{(i+1)}, \ldots, g_{(i+v-1)}\}
\]

\[
t_i = \{g_{(i+v)}\}
\]

Channel datasets is not only suit to the structure of SLFN, but also exploit the information of every CSI. In order to predict every future CSI, the step of sliding window is \( 1 \) in channel dataset. In Figure 2.(a), the sliding window contains \((k - i - v + 1)\) channel datasets.

In real channel prediction scenario, the step of sliding window \( M \) is equal to the number of CSI in coherent time. In other words, we need to update the parameters of SLFN in every coherent time.

III. STRUCTURE OF THE ONLINE PREDICTION MODEL BASED ON ELM

In practical applications, the ELM algorithm needs to update its own parameters for every online sequence in the prediction, which causes large computational complexity. The OS-ELM algorithm [24] can avoid the repeated training of past samples, and the extreme machine learning algorithm is realized with online learning. The AR model is a linear model, the predictive performance cannot be satisfied with a
large Doppler frequency shift. The activation function of the OS-ELM is non-linear and can be well adapted to fast fading situations.

The OS-ELM is implemented with a single hidden layer feed-forward neural network (SLFN) which has a fast learning speed and can also achieve wonderful generalization performance compared to traditional algorithms.

A. THE DESCRIPTION OF THE OS-ELM

Assume the step of sliding windows is a fixed size $M$. Therefore each chunk of the channel datasets is $(x_i, t_i, i=1, 2, ..., kM)$ has arrived. We fix the number of hidden nodes $N$ and randomly assign the node parameters $[(a_i, b_i), (i = 1, 2, ..., N)]$. Let $H_i$ denote the $i$-th hidden-layer output matrix [24]:

$$H_i = \begin{bmatrix} F(a_1, b_1, x_{1M}) & \cdots & F(a_N, b_N, x_{1M}) \\ \vdots & \ddots & \vdots \\ F(a_1, b_1, x_{(i+1)M-1}) & \cdots & F(a_N, b_N, x_{(i+1)M-1}) \end{bmatrix}$$  \hspace{1cm} (11)

where $T_i = [t_{iM} \cdots t_{(i+1)M-1}]^T$.

The output weight $\beta(k)$ can be expressed as follows [24]:

$$\beta(k) = P_k \begin{bmatrix} H_{(0)} \\ \vdots \\ H_{k} \end{bmatrix}^T \begin{bmatrix} T_{(0)} \\ \vdots \\ T_{k} \end{bmatrix}$$ \hspace{1cm} (12)

where

$$P_k = \left( \sum_{q=0}^{k} H_q^T H_q \right)^{-1}$$ \hspace{1cm} (13)

Second, when $M$ new channel datasets arrive, $H_{k+1}$ is calculated by Eq. (11), and the $\beta(k+1)$ can be expressed as follows [24]:

$$P_{k+1} = P_k - P_k H_{k+1}^T (I + H_{k+1} P_k H_{k+1}^T)^{-1} \times H_{k+1} P_k$$ \hspace{1cm} (14)

$$\beta(k+1) = \beta(k) + P_{k+1} H_{k+1}^T (T_{k+1} - H_{k+1} \beta(k))$$ \hspace{1cm} (15)

where $m = 1, 2, ..., M$.

Although the OS-ELM algorithm can avoid the repeated training of historical data, channel datasets are often timely with respect to some real applications. In network training research [25], [26], forgetting mechanism can reduce the influence of useless information.

B. THE DESCRIPTION OF THE FOS-ELM

When the forgetting mechanism is introduced into OS-ELM. When $M$ new channel datasets arrive, the earliest channel datasets are deleted. The channel datasets are composed by Eq. (9) and Eq. (10). Therefore, the formula of $\beta(k+1)$ is given as [26]

$$P_{k+1} = P_k - P_k \left[ -H_{k+1}^{-1} \begin{bmatrix} -H_{k+1}^{-1} \end{bmatrix}^T (I + \bar{P}_k^{-1})^{-1} \times \begin{bmatrix} -H_{k+1}^{-1} \end{bmatrix} P_k \right]$$ \hspace{1cm} (16)
\[P_k = \begin{bmatrix}
  -H_{k-s+1}^T P_k H_{k-s+1}^T & H_{k-s+1}^T P_k H_{k-s+1}^T \\
  -H_{k+1} P_k H_{k+1}^T & H_{k+1} P_k H_{k+1}^T
\end{bmatrix} \]  

Eq. (17)

\[\beta(k + 1) = \beta(k) + P_{k+1} \left[ -H_{k-s+1} \right]^T \times \left[ \begin{bmatrix}
  T_{k-s+1}^T \\
  T_{k+1}^T
\end{bmatrix} - \begin{bmatrix}
  H_{k-s+1} \\
  H_{k+1}
\end{bmatrix} \beta(k) \right] \]  

Eq. (18)

However, the complexity of the FOS-ELM resists the channel prediction performance in fast changing channels. The complexity makes it difficult to apply the FOS-ELM algorithm in online predictions. Therefore, we proposed a novel prediction model, which reduced the computations of the training process and the influence of invalid data. It is easy to conduct online channel predictions in fast changing channel.

IV. THE TFOS-ELM CHANNEL PREDICTION ALGORITHM

A. THE DESCRIPTION OF THE TFOS-ELM ALGORITHM

**Lemma 1:** For any positive definite Hermitian matrix \(X\),

\[X^{-1} = \kappa (I - (I - \kappa X))^{-1} = \kappa \sum_{\ell=0}^{\infty} (I - \kappa X)^\ell\]  

(19)

\[0 < \kappa < \frac{2}{\max \lambda(X)}\]  

(20)

Therefore, the Hermitian matrix is converted into a matrix polynomial expansion, when all the maximum of eigenvalues \(\max \lambda(X)\) are smaller than unity, hence any eigenvalues of the matrix power \(\lambda(X)\) converge to 0 for \(\ell\) close to infinite. The matrix disconverges when one eigenvalue is larger than unity, because the matrix power \(\lambda(X)\) can be infinity when \(\ell\) approaches infinity. So, the eigenvalues use the normalization factor \(\kappa\). If a \(\kappa\) is chosen that satisfies Eq.(21), \(X^{-1}\) can be approximated by the TPE algorithm.

B. THE DESCRIPTION OF THE TFOS-ELM ALGORITHM

First, suppose the chunk of the channel datasets \((\mathbf{x}_t, \mathbf{t}_t)\) is available. Similar to the FOS-ELM, the channel datasets have been assembled. The \(l\)-th hidden-layer output matrix \(H_l\) is calculated using Eq.(11), and the \(\beta(k)\) of the TFOS-ELM was calculated using Eq.(12).

When the chunk of \(M\) new CSI data has arrived \((g(1M + 1), \ldots, g(1M + M))\), \(M\) new channel datasets \((\mathbf{x}_t, \mathbf{t}_t)\) is constructed, and \(H_{k+1}\) can be calculated using Eq.(11). The advantage of the TFOS-ELM algorithm is that it conducts matrix inversion using the TPE algorithm.

Because \((I + \bar{P}_k)\) is a Hermitian matrix in Eq.(17), the inversion of matrix \((I + \bar{P}_k)\) can be converted into a special form. The TPE algorithm conducts the matrix inversion through polynomial expansion, and then it intercepts the first few terms, which reduces calculation complexity. According to Lemma 1, the inverse operation of matrix \(P_k\) is found using Eq.(20). It is denoted by

\[P_{k+1} = \kappa \sum_{n=0}^{\infty} \left( \frac{\kappa}{n} \right) (1 - \kappa)^n (1 - \kappa)^n (\bar{P}_k)\]  

(21)

\[P_{k+1}^T = W_k (\bar{P}_k)\]  

(22)

where \(W_k = \sum_{n=0}^{\infty} \left( \frac{\kappa}{n} \right) (1 - \kappa)^n (1 - \kappa)^n (\bar{P}_k)\).

Therefore, the \(\beta(1)\) of the TFOS-ELM algorithm was calculated using Eq.(24)

\[\hat{g}(k+1) = \hat{g}(k+1) + \beta(k) + \bar{P}_k^T \left[ -H_{k-s+1} \right]^T \times \left[ \begin{bmatrix}
  T_{k-s+1}^T \\
  T_{k+1}^T
\end{bmatrix} - \begin{bmatrix}
  H_{k-s+1} \\
  H_{k+1}
\end{bmatrix} \beta(k) \right].\]  

(23)

The channel prediction value can be expressed as Eq.(24).

\[\hat{g}(k+1)M + m = \hat{g}(k+1)F(\mathbf{a}_i, b_i, \mathbf{x}_{(k+1)M + m})\]  

(24)

Thus \(\hat{g}(k+1)M + m\) can be predicted using \(\beta(k+1)\) when \(g(k+1)M + m\) is arrived. The \(\beta(k+1)\) can be updated when \(M\) new channel datasets arrive.

C. THE PROCESS OF THE TFOS-ELM ALGORITHM

In Figure 4, the general scheme of sub-channel prediction with the TFOS-ELM algorithm is depicted, and the steps of MIMO channel prediction is concluded in the following Algorithm 1 and Figure 3.

Since the CSI is a complex value, the channel prediction is divided into two parts: the real-part and the imaginary-part. The TFOS-ELM calculates value of the \(\hat{g}_{mn}(k+1)\) depending on the real-part and the imaginary-part. In our paper, all \(N_rN_c\) channels are simultaneously predicted. Then, the channel matrix \(\hat{g}(k+1)\) is composed of every \(\hat{g}_{mn}(k+1)\).

**Algorithm 1** The Channel Prediction Procedure of MIMO

**Input:** \(N_r, N_c, f_d, j_d\)

**Output:** \(\hat{g}(k+1)\)

1. **for** \(m_r = 1; m_r \leq N_r; m_r++\):
2. **for** \(n_r = 1; n_r \leq N_c; n_r++\):
3. Generate the sub-channels \(\hat{g}_{mn}\) in Eq.(4-6).
4. **for** the channel datasets using Eq.(9-10).
5. **for** the TFOS-ELM to predict future \(\hat{g}_{mn}(k+1)\).
6. **end for**
7. **end for**
8. Combine the prediction \(\hat{g}(k+1)\).
9. **Return** \(\hat{g}(k+1)\).

Next, the specific process of our algorithm is shown in algorithm 2:

D. COMPLEXITY ANALYSIS OF THE TFOS-ELM

The channel prediction complexities of the FOS-ELM and the TFOS-ELM are compared. The complexity of prediction
**FIGURE 3.** The procedure of the precoding with channel prediction.

### Algorithm 2 TFOS-ELM

**Input:** datasets \((x_l, t_l)| l = 0, 1, 2, \ldots, kM\), \(M\), \(\tilde{N}\), \(v\).

**Output:** \(\hat{g}(k+1)M + m\)

**Step I: initial step**
1. Randomly assign the hidden parameters: \((a_i, b_i), i = 1, \ldots, \tilde{N}\).
2. Calculate the \(l\)-th partial hidden-layer output matrix: \(\mathbf{H}_l^T\) from Eq. (11) and \(\mathbf{T}_l = [t_{lM} \ldots t_{(l+1)M-1}]^T\).
3. Calculate \(P_T^{k+1}\) using Eq. (14).
4. Estimate the output weight \(\beta(k)\) in Eq. (12).
5. Calculate \(\hat{g}(kM + m) = \beta(k)F(a_i, b_i, x((k+1)M + m - v), m)\)

**Step II: The TFOS-ELM Algorithm.**

**While** new \(M\) CSI arrives, **do**:
1. Compose new \((x, t)\) using Eq. (9-10).
2. Calculate the partial hidden-layer output matrix \(\mathbf{H}_{k+1}\).
3. Calculate \(W_{k+1}\) using Eq. (21) and \(P_{T_{k+1}}^T\) using Eq. (22).
4. Calculate the output weight \(\beta(k+1)\) using Eq. (23).
5. Calculate the prediction value \(\hat{g}(k+1)M + m\) based on the input dataset \(x((k+1)M + m - v), m \in (0, M)\).

The algorithms includes complex addition and multiplication in every update. Assume that all new chunks of CSI data are the same size (\(M\)) and \(\tilde{N}\) is fixed. The matrix inversion method of OS-ELM and FOS-ELM is Cholesky factorization [28]. The cost of matrix inversion is \((1/3)n^3 + 2n^2\) (where \(n\) is size of matrix). The complexity of the OS-ELM, FOS-ELM algorithm and the TFOS-ELM algorithm are shown as follows.

There are standard operations for the matrix. We get the computational complexity of \(\beta(k + 1)\) for the OS-ELM and FOS-ELM algorithm in TABLE 1. The complexity of OS-ELM and FOS-ELM are expressed as two part. One is the complexity of matrix inversion, other is process of \(\beta(k + 1)\). The specific data shown in TABLE 1.

### TABLE 1. The complexity of OS-ELM, FOS-ELM and TFOS-ELM.

| Prediction model | Complexity [28] |
|------------------|-----------------|
|                  | Total complexity | standard operations |
| OS-ELM           | \(\frac{M^3}{3} + 4M^2 + 2MN\) | \(\frac{M^3}{3} + 2M^2\) |
|                  | \(2M^2 + 2MN\) |
| FOS-ELM          | \(\frac{8M^3}{3} + 16M^2 + 16MN\) | \(\frac{8M^3}{3} + 8M^2\) |
|                  | \(8M^2 + 16MN\) |
| TFOS-ELM         | \(8JM^2 - 2JM\) | \(8JM^2 - 2JM\) |

The TFOS-ELM algorithm avoiding the matrix inversion, the computational complexity of \(\beta(k + 1)\) is converted into terms of matrix multiplication.

\[
\begin{align*}
C_{\text{TFOS-ELM}} & = 8JM^2 - 2JM. \\
C_{\text{FOS-ELM}} & \approx \frac{M + 6}{3J} \\
C_{\text{OS-ELM}} & \approx \frac{M + 12}{24J} \\
C_{\text{TFOS-ELM}} & \approx \frac{M + 12}{24J} \\
\end{align*}
\]

From Eq. (26~27), the complexity of OS-ELM has lower complexity than others when \(M + 12 < 24J\). As the time goes by, the performance of OS-ELM model could degradation due to obsolescence of earliest channel datasets. Therefore, the OS-ELM is not suit in online channel prediction. Comparing FOS-ELM with TFOS-ELM, \(M\) is more than \(\tilde{N}\) in the
Due to $M \geq \tilde{N}$ and $M \gg J$, the computational complexity of the TFOS-ELM is less than that of the FOS-ELM. In addition, the complexity of $\beta(k+1)$ is only one sub-channel in the massive MIMO, thus, as the number of BS antennas increases, the TFOS-ELM algorithm is better than the FOS-ELM algorithm with respect to the overall complexity of all channels. Overall, when the step of sliding windows

V. SIMULATION

The Massive MIMO-OFDM simulation parameters are listed in Table 2. The numbers of receiving terminals, BS antennas and receiving terminal antennas are 16, 100 and 1, respectively. All of our simulations are completed in MATLAB 2014.a environment on a personal computer including a quad-processor with 3.3 GHz and 16 GB of RAM.

In our simulations, the pilot symbol period ($S_t$) is set to 0.5ms in the block-type pilot arrangement when the Doppler frequency is 100Hz. The size of channel dataset is 5, the step of sliding window $M$ is 30. When the predictors collect $M$ channel datasets, the predictors update the OS-ELM, FOS-ELM and TFOS-ELM once. The AR model is different from the OS-ELM, FOS-ELM and TFOS-ELM; therefore it should update its parameters using the MMSE when 30 CSI arrive.

In order to evaluate performance of four prediction model, root mean square error (RMSE), correlation coefficients (R) and R-squared ($R^2$) analyze the performance of prediction between the actual and obtained CSI. Due to the online channel prediction, the testing accuracy play a key role in performance criteria. In this paper, we define a model, which reaches a best validation [29] performance in terms of RMSE, $R$ and $R^2$ for the testing accuracy. The results are presented by using the OS-ELM, FOS-ELM, TFOS-ELM ($J=4, 6, 8$) and AR (MMSE) in Table 3.

In Table 3, the testing results are presented by using the AR, OS-ELM, FOS-ELM and TFOS-ELM. In this process, 2000 CSI data are utilized. Compare with AR model, the OS-ELM, FOS-ELM and TFOS-ELM have higher RMSE and R. In addition, the TFOS-ELM approaches to the performance of FOS-ELM in RMSE and R criteria. But the TFOS-ELM has lower computational complexity according to Eq.(27).

Figures 6-9 can explain the high-performance of the four prediction algorithms.

Figure 5 depicts the amplitude of the (1,1)-antenna pair channel prediction with the OS-ELM, FOS-ELM and TFOS-ELM algorithm.

| TABLE 2. Simulation parameters. |
|---------------------------------|
| **parameter type**   | **Value** |
| Carrier frequency        | 1.8GHz    |
| Sampling Rate            | 5 MHz     |
| No. of antennas(bs,ms)   | (100, 16) |
| size of channel dataset $\nu$ | 5     |
| bandwidth                | 5 MHZ     |
| subcarrier               | 256       |
| No. of Nodes             | 10        |
| step of sliding window $M$ | 30       |
| Doppler frequency        | 100Hz     |
| Activation function      | sigmoid   |

| TABLE 3. Comparison of the OS-ELM, FOS-ELM, TFOS-ELM($J=4,6,8$) and AR (MMSE) for testing process (SNR=0 dB). |
|--------------------------------------------------|
| **Prediction type** | **RMSE** | **R** | **$R^2$** |
| OS-ELM                             | 0.156    | 0.952 | 0.902 |
| FOS-ELM                             | 0.062    | 0.994 | 0.982 |
| TFOS-ELM($J=4$)                    | 0.096    | 0.972 | 0.938 |
| TFOS-ELM($J=6$)                    | 0.072    | 0.982 | 0.954 |
| TFOS-ELM($J=8$)                    | 0.065    | 0.988 | 0.960 |
| AR(MMSE)                            | 0.198    | 0.929 | 0.854 |
TFOS-ELM algorithms in a scenario in which the Maximum Doppler shift is 100 Hz and the delay is 0.5ms. We can find that the channel coefficients are correlated in the coherent time.

Figure 6 shows a situation where the maximum Doppler shift is set to 100 Hz with a 0.5ms delay. As the SNR increases, the OS-ELM, the FOS-ELM, the TFOS-ELM with $J = (4, 6, 8)$ and the AR algorithms have different MSEs. The results show that all algorithms have better performance than the AR algorithm, since the OS-ELM, the FOS-ELM and the TFOS-ELM have higher non-linear prediction performance. In addition, at a low SNR, the OS-ELM and the FOS-ELM perform worse than the TFOS-ELM algorithm, since our channel prediction model will become inaccurate when the noise is large. As $J$ increases, the influence of the noise on the TFOS-ELM algorithm will also increase. The TFOS-ELM prediction algorithm is not sensitive to the effects of noise when $J$ is small.

In Figure 7, the average performance rate of massive MIMO precoding is investigated. The OS-ELM, the FOS-ELM, the TFOS-ELM, and the AR algorithm are adopted before the Zero-Forcing (ZF) precoding in massive MIMO system. The channel delay is set to 0.5ms, the Doppler frequency is 100Hz, and $J = (4, 6, 8)$. We observe that the average performance rate of the simulation with the channel prediction is better than the simulation without the prediction when the SNR is more than $-7$dB. This because the channel prediction can compensate for the effects of the delay. In addition, when the SNR is less than $-7$dB, the accuracy of the channel prediction may be influenced by noise. The AR algorithm has lower average rates than the other prediction algorithms in our simulation. In addition, the average rates of the OS-ELM, the FOS-ELM, and the TFOS-ELM are similar, but the TFOS-ELM runs faster than the other algorithms when the number of BS antennas is large.

In Figure 8, the simulation results demonstrate that the BER performance of a system using channel prediction is better than that with no prediction. The BER of the AR algorithm is higher than those of other prediction algorithms. The TFOS-ELM ($J = 4, 6, 8$) has best performance when low SNR exists. This is because the ZF precoding enhances the influence of noise and the TPE algorithm decreases the impact of the noise. The simulation results in Figure 8 show that the FOS-ELM algorithm has higher performance than the TFOS-ELM algorithm when the SNR more than $-7$dB, but it cause high computational complexity (the computational complexity is shown in Table 1). A suitable value of $J$ can allow the TFOS-ELM to attain high communication performance. Comparing the OS-ELM and the TFOS-ELM prediction algorithms, the TFOS-ELM algorithm ($J \geq 6$) has similar performance to the FOS-ELM algorithm.
The computational complexities of the three prediction algorithms with steps of sliding windows \( M \) (the number of CSI in one coherent time) are presented in Figure 9. As the step of sliding window \( M \) increases the computational complexities of the FOS-ELM are more than those of the TFOS-ELM and OS-ELM in the massive MIMO system. When \( M \geq N \) and \( N = 10 \), Figure 9 depicts that the TFOS-ELM algorithm with \( J = (4, 6, 8) \) has lower computational complexity than the FOS-ELM algorithm. In Figures 6-9, where \( J = (4, 6, 8), v = 5 \), the number of nodes \( N \) is 10. The OS–ELM and TFOS-ELM have the similar computational complexity, but the prediction accuracy of the TFOS-ELM is better than the OS-ELM. The computational complexity of the FOS-ELM is larger than the OS-ELM and the TFOS-ELM. Therefore, as the step of sliding window \( M \) and the number of BS antennas increase in the massive MIMO system, our proposed prediction algorithm has a substantial advantage with respect to the computational complexity.

VI. CONCLUSION

In this paper, we proposed the TFOS-ELM algorithm for online channel prediction. We first introduced the FOSELM to predict channel, and utilized TPE to reduce the complexity of FOSELM algorithm. The TFOS-ELM algorithm was simulated using a fast changing channel. The simulation results show that the TFOS-ELM algorithm has superior performance compared to the OS-ELM, FOS-ELM and AR algorithms for channel prediction using fast changing channels. Furthermore, we applied the channel prediction to precoding. The results show that our algorithm improves the communication quality. Our future work includes improving the performance of modulation, power control and beamforming performances based on channel prediction.

REFERENCES

[1] R. Zi, X. Ge, J. Thompson, C.-X. Wang, H. Wang, and T. Han, “Energy efficiency optimization of 5G radio frequency chain systems,” IEEE J. Sel. Areas Commun., vol. 34, no. 4, pp. 758–771, Apr. 2016.
[2] K. J. Kim, M. Pun, and R. A. Illis, “Channel prediction for limited feedback precoded MIMO-OFDM systems over time-varying fading channels,” in Proc. 42nd Annu. Conf. Inf. Sci. Syst., Mar. 2008, pp. 972–977.
[3] A. Duel-Hallen, “Fading channel prediction for mobile radio adaptive transmission systems,” Proc. IEEE, vol. 95, no. 12, pp. 2299–2313, Dec. 2007.
[4] H. Shirani-Mehr, D. N. Liu, and G. Caire, “Channel state prediction, feedback and scheduling for a multiuser MIMO-OFDM downlink,” in Proc. 42nd Asilomar Conf. Signals, Syst. Comput., Oct. 2008, pp. 136–140.
[5] M. Sernad and S. Falahati, “Maximizing throughput with adaptive M-QAM based on imperfect channel predictions,” in Proc. IEEE PIMRC, Sep. 2004, pp. 2289–2293.
[6] A. Alsawah, I. Fijalkow, and C. R. Johnson, “Improving minimum rate maximization link adaptation strategy using channel prediction,” in Proc. IEEE ICASSP, vol. 4, May 2006, pp. 265–268.
[7] S. Hu and A. Duel-Hallen, “Combined adaptive modulation and transmitter diversity using long range prediction for flat fading mobile radio channels,” in Proc. GLOBECOM. IEEE Global Telecommun. Conf., vol. 2, Nov. 2001, pp. 1256–1261.
[8] Q. Liu, “Channel prediction for multiuser downlink LTE-TDD systems,” in Proc. IEEE 26th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC), Aug. 2015, pp. 160–165.
[9] A. Papazafeiropoulos and T. Ratnarajah, “Linear precoding for downlink massive MIMO with delayed CSIT and channel prediction,” in Proc. IEEE Wireless Commun. Netw. Conf. (WCNC), Apr. 2014, pp. 809–914.
[10] A. Duel-Hallen, S. Hu, and H. Hallen, “Long-range prediction of fading signals,” IEEE Signal Process. Mag., vol. 17, no. 3, pp. 62–75, May 2000.
[11] H. Hallen, A. Duel-Hallen, S. Hu, T.-S. Yang, and M. Lei, “A physical model for wireless channels to provide insights for long range prediction,” in Proc. MILCOM, vol. 1, Oct. 2002, pp. 627–631.
[12] I. C. Wong and B. L. Evans, “Joint channel estimation and prediction for OFDM systems,” in Proc. GLOBECOM . IEEE Global Telecommun. Conf., vol. 4, Dec. 2005, pp. 2255–2259.
[13] I. C. Wong and B. L. Evans, “WLC43-5: low-complexity adaptive high-resolution channel prediction for OFDM systems,” in Proc. IEEE Globalcom, vol. 27, Nov. 2006, pp. 1–5.
[14] P. D. Teal and R. G. Vaughan, “Simulation and performance bounds for real-time prediction of the mobile multipath channel,” in Proc. 11th IEEE Signal Process. Workshop Stat. Signal Process., Aug. 2001, pp. 548–551.
[15] M. Chen, T. Ekman, and M. Viberg, “New approaches for channel prediction based on sinusoidal modeling,” EURASIP J. Adv. Signal Process., vol. 2007, no. 1, Sep. 2006, Art. no. 049393.
[16] K. E. Baddour and N. C. Beaulieu, “Autoregressive modeling for fading channel simulation,” IEEE Trans. Wireless Commun., vol. 4, no. 4, pp. 1650–1662, Jul. 2005.
[17] J. Hwang and J. Winters, “Sinusoidal modeling and prediction of fast fading processes,” Proc. GLOBECOM, vol. 2, Nov. 1998, pp. 892–897.
[18] S. Navabi, C. Wang, O. Y. Bursalioglu, and H. Papadopoulos, “Predicting wireless channel features using neural networks,” 2018, arXiv:1802.00107. [Online]. Available: http://arxiv.org/abs/1802.00107
[19] C. Potter, G. K. Venayagamoorthy, and K. Khorash, “RNN based MIMO channel prediction,” Signal Process., vol. 90, no. 2, pp. 440–450, Feb. 2010.
[20] J. Joo, M. C. Park, D. S. Han, and V. Pejovic, “Deep learning-based channel prediction in realistic vehicular communications,” IEEE Access, vol. 7, pp. 27846–27858, 2019.
[21] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, “Channel state information prediction for 5G wireless communications: A deep learning approach,” IEEE Trans. Netw. Sci. Eng., to be published.
[22] A. Muller, A. Kammoun, E. Bjornson, and M. Debbah, “Efficient linear precoding for massive MIMO systems using truncated polynomial expansion,” in Proc. IEEE 8th Sensor Array Multichannel Signal Process. Workshop (SAM), Jun. 2014, pp. 273–276.
[23] G. B. Huang, Q. Y. Zhu, and C. K. Siew, “Extreme learning machine: A new learning scheme of feedforward neural networks,” in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2004, pp. 985–990.
[24] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, “A fast and accurate online sequential learning algorithm for feedforward networks,” IEEE Trans. Neural Netw., vol. 17, no. 6, pp. 1411–1423, Nov. 2006.
[25] P. Zhang, C. Li, T. Huang, L. Chen, and Y. Chen, “Forgetting memristor based neuromorphic system for pattern training and recognition,” Neurocomputing, vol. 222, pp. 47–53, Jan. 2017.
[26] J. Zhao, Z. Wang, and D. S. Park, “Online sequential extreme learning machine with forgetting mechanism,” Neurocomputing, vol. 87, pp. 79–89, Jun. 2012.
[27] A. Mueller, A. Kammoun, E. Bjornson, and M. Debbah, “Linear precoding based on polynomial expansion: Reducing complexity in massive MIMO,” EURASIP J. Wireless Commun. Netw., vol. 2016, no. 1, Feb. 2016.
[28] S. Boyd and L. Vandenberghe. Numerical Linear Algebra Background. Accessed: Feb. 12, 2016. [Online]. Available: http://www.seas.ucla.edu/~vandenbe/ee236b/lectures/num-lin-alg.pdf
[29] T. Gurgenc, F. Ucar, C. Ozel, Y. Ortac, and D. Korkmaz, “A study on the extreme learning machine based prediction of machining times of the cycloidal gears in CNC milling machines,” Prod. Eng., vol. 13, pp. 635–647, Oct. 2019, doi: 10.1007/s11740-019-00923-1.
YIGANG HE received the M.Sc. degree in electrical engineering from Hunan University, Changsha, China, in 1992, and the Ph.D. degree in electrical engineering from Xi’an Jiaotong University, Xi’an, China, in 1996. His teaching and research interests are in the areas of power electronic circuit theory and its applications, testing and fault diagnosis of analog and mixed-signal circuits, electrical signal detection, smart grid, satellite communication monitoring, and intelligent signal processing. He has published some 300 journal and conference papers, which was included more than 1000 times in Science Citation Index of American Institute for Scientific Information in the aforementioned areas and several chapters in edited books. He has been on the Technical Program Committee of a number of international conferences. He was the recipient of a number of national and international awards, prizes, and honors, i.e., the winner of the National Outstanding Youth Science Fund, China National Excellent Science and Technology Worker.

WEI HE received the bachelor’s degree in automation from the Hefei University of Technology, China, in 2009, where he is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Automation. He is also a joint Ph.D. student with the University of Duisburg-Essen, Germany. His research interests include condition monitoring, fault diagnosis, machine learning, and signal processing.

LUQIANG SHI received the bachelor’s degree in electrical engineering from the Hefei University of Technology, Hefei, Anhui, in 2015, where he is currently pursuing the Ph.D. degree. His research interest is wireless communication.

YONGBO SUI received the B.E. degree in electrical engineering from Xiangtan University, Xiangtan, Hunan, in 2014. He is currently pursuing the Ph.D. degree with the Hefei University of Technology, Hefei, Anhui. His research interests include intelligent algorithms and wireless communication channels.

YUAN HUANG received the B.S. degree from the Hefei University of Technology, Anhui, China, in 2016, where he is currently pursuing the Ph.D. degree in electrical engineering. His current research interests include the research of digital signal processing, wireless communication, deep learning, and compressed sensing.