Online Extreme Learning Machine-Based Channel Estimation and Equalization for OFDM Systems

Jun Liu, Kai Mei, Xiaochen Zhang, Dongtang Ma, *Senior Member, IEEE*, and Jibo Wei, *Member, IEEE*

***Abstract*— Machine learning-based channel estimation and equalization methods may improve the robustness and bit error rate (BER) performance of communication systems. However, the implementation of these methods has been blocked by some limitations, mainly including channel model-based offline training and high-computational complexity for training deep neural network (DNN). To overcome those limitations, we propose an online fully complex extreme learning machine (C-ELM)-based channel estimation and equalization scheme with a single hidden layer feedforward network (SLFN) for orthogonal frequency-division multiplexing (OFDM) systems against fading channels and the nonlinear distortion resulting from an high-power amplifier (HPA). Computer simulations show that the proposed scheme can acquire the information of channels accurately and has the ability to resist nonlinear distortion and fading without pre-training and feedback link between receiver and transmitter. Furthermore, the robustness of the proposed scheme is well investigated by extensive simulations in various fading channels, and its excellent generalization ability is also discussed and compared with the DNN.**

***Index Terms*— Fully complex extreme learning machine (C-ELM), channel estimation and equalization, single hidden layer feedforward network (SLFN), OFDM, fading channels.**

I. Introduction

THE depiction of channel capacity in communication systems has been exhaustively investigated by Shannon. Nevertheless, whether communication systems can approach the channel capacity heavily depends on acquisition of accurate channel state information (CSI). Actually, precisely estimating CSI is extremely difficult. Traditionally, least squares (LS) channel estimator is implemented in practical communication systems. If the correlation matrices of channels are known, minimum mean-square error (MMSE) channel estimator can achieve the best performance in linear communication systems [1].

Lately, a variety of machine learning (ML)-based channel estimation and CSI feedback schemes have emerged. In [2], a deep neural network (DNN)-based orthogonal frequency division multiplexing (OFDM) channel estimation and signal detection approach has been introduced and its performance is better than LS and MMSE estimators when considering adversities. Moreover, through computer vision (CV) or natural language processing (NLP) techniques, exploitation of the correlations among time, frequency and space of channels also have been considered. In [3], image super-resolution (SR) and image restoration (IR) algorithms are used for channel interpolation and noise suppression, respectively. Similarly, [4] proposes two high-resolution schemes to realize DOA estimation and channel estimation in sparse case. A real-time CSI feedback architecture based on convolution neural network (CNN) and long short-term memory (LSTM) is proposed in [5], in order to achieve channel compressed sensing (CS). Besides, LSTM is also used to detect sequences in [6].

However, those approaches have some common limitations. One of the major limitations is that those methods require channel model-based pre-training before they can be used for channel estimation, CSI feedback and so on. Nevertheless, the real channel scenarios are significantly different from the channel models used for pre-training. Therefore, learning-based algorithms may converge in optimum based on pre-training instead of optimum in terms of real channel scenarios. Another limitation lies in the training process of ML-based methods. For the mobile wireless communication systems, gradient propagation between transmitter and receiver is difficult to achieve. Although [7] presents a model using conditional generative adversarial network (GAN) to tackle the difficulty of gradient propagation, channel models and pre-training are still indispensable. The last but not least is that most of the previous works are based on DNN, which requires prohibitive computational complexity.

In order to overcome those limitations and achieve better performance on channel estimation and equalization, four main challenges should be tackled on. First, communication systems should be trained online instead of pre-training based on channel models. Second, online training should be done at receiver solely. Otherwise, gradient propagation between transmitter and receiver requires lots of communication resources and thus this deviates from the objective of communications. Third, systems should have the flexibility to adapt to any channels. Forth, the performance of channel estimation and equalization can achieve similar performance with MMSE estimator by consuming low computational complexity and even superior performance when signals suffer from nonlinear distortion.

In this article, we propose an online fully complex extreme learning machine (C-ELM)-based communication system in which the receiver can learn the information about channels through C-ELM with single hidden layer feedforward network (SLFN) and the system can achieve optimal performance. The main contributions of this letter are summarized as follows.

- We propose an online learning-based scheme to realize channel estimation and equalization, where the C-ELM can learn the statistical information of channels and restore the nonlinear distortion.

Manuscript received April 5, 2019; accepted April 30, 2019. Date of publication May 15, 2019; date of current version July 10, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61601480 and Grant 61372099 and in part by the Research Foundation of National University of Defense Technology under Grant ZK-17-03-13. The associate editor coordinating the review of this letter and approving it for publication was M. Z. Z. Shakir. *(Corresponding author: Jun Liu.)*

The authors are with the College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China (e-mail: liujun15@nudt.edu.cn).

Digital Object Identifier 10.1109/LCOMM.2019.2916797

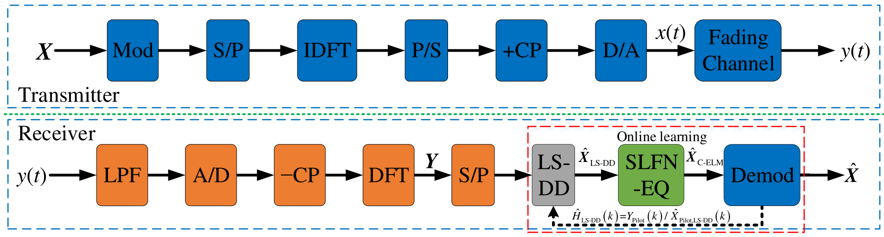

Fig. 1. OFDM system model with C-ELM-based SLFN equalizer. The block of Demod only feeds back the decision-directed (DD) estimation of pilot symbols $\hat{\boldsymbol{X}}_{\mathrm{Pilot,LS-DD}}$. Then, DD channel estimation $\hat{\boldsymbol{H}}_{\mathrm{LS-DD}}$ and DD pre-equalization of data symbols $\hat{\boldsymbol{X}}_{\mathrm{Data,LS-DD}}$ can be acquired, successively.

- We provide performance analysis of the proposed learning scheme of OFDM in different cases. Specifically, we simulate bit error ratio (BER) for assessing and comparing the performance of the proposed scheme. Also, extensive simulation results and comparisons have demonstrated the robustness and generalization ability of the proposed scheme.

The remainder of this letter is organized as follows. In Section II, we propose a scheme that incorporates C-ELM into the OFDM system, in which SLFN is used to filter the signal being equalized by the estimated CSI from LS estimator. Then, Numerical results and analysis for evaluating the performance of the proposed scheme are provided in Section III, which is followed by conclusions in Section IV.

Notations: The notations adopted in the letter are as follows. We use boldface small letters and capital letters to denote vectors and matrices, respectively. Specifically, let $\mathbf{X}^T$, $\mathbf{X}^H$, $\mathbf{X}^\dagger$ denote the transpose, conjugate transpose and the Moore-Penrose generalized inverse of a matrix $\mathbf{X}$ respectively. Exceptionally, italic boldface capital letters such as $\boldsymbol{X}$ and $\boldsymbol{Y}$ denote frequency-domain vectors. $\circledast$, $\mathrm{E}\{\cdot\}$ and j denote cyclic convolution, the expectation and the imaginary unit.

II. System Model

A. Traditional Channel Estimation for OFDM

We consider an OFDM system with N_c subcarriers and cyclic prefix (CP) duration T_G. The received signal $\boldsymbol{Y}$ is then modeled using the N_c-point discrete-time Fourier transform (DFT) as

$$\boldsymbol{Y} = \mathrm{DFT}\left(\mathrm{IDFT}\left(\boldsymbol{X}\right) \circledast \frac{\mathbf{g}}{\sqrt{N_c}} + \tilde{\mathbf{n}}\right), \tag{1}$$

where transmitted signal $\boldsymbol{X} = [X_0\ X_1 \cdots X_{N_c-1}]^T$, $\boldsymbol{Y} = [Y_0\ Y_1 \cdots Y_{N_c-1}]^T$. $\tilde{\mathbf{n}} = [\tilde{n}_0\ \tilde{n}_1 \cdots \tilde{n}_{N_c-1}]^T$, is a noise vector of i.i.d. complex Gaussian variables, and channel impulse response (CIR) $\mathbf{g} = [g_0\ g_1 \cdots g_{N_c-1}]^T$, is determined by the cyclic equivalent of sinc-functions. Specifically, The vector $\mathbf{g}/\sqrt{N_c}$ is the observed CIR after sampling the impulse response of $g(t)$. The corresponding continuous CIR $g(t)$ can be denoted as

$$g(t) = \sum_m \alpha_m \delta\left(t - \tau_m T_s\right) \tag{2}$$

using tapped-delay line model, where the amplitudes α_m are complex-valued, T_s is symbol period and path delays $\tau_m \geq 0$. The system described by (1) can be written in matrix notation

$$\boldsymbol{Y} = \mathbf{X}\mathbf{F}\mathbf{g} + \boldsymbol{N}, \tag{3}$$

where $\mathbf{X}$ is a matrix with the elements of $\boldsymbol{X}$ on its diagonal, $\boldsymbol{N} = [n_0\ n_1 \cdots n_{N_c-1}]^T = \mathrm{DFT}_{N_c}(\tilde{\mathbf{n}})$ is an i.i.d. complex zero-mean Gaussian noise vector and $\mathbf{F}$ is the DFT-matrix with $W_{N_c}^{nk} = \frac{1}{\sqrt{N_c}} e^{-j2\pi\frac{nk}{N_c}}$. CSI can be represented in the form of frequency response $\boldsymbol{H} = [H_0\ H_1 \cdots H_{N_c-1}]^T = \mathrm{DFT}_{N_c}(\mathbf{g})$.

To estimate CSI, LS estimator has been used in practical systems extensively and it can be written as

$$\hat{\boldsymbol{H}}_{\mathrm{LS}} = \mathbf{X}^{-1}\boldsymbol{Y}. \tag{4}$$

Different from the LS, MMSE estimator is more complicated. If the channel vector $\mathbf{g}$ is Gaussian and uncorrelated with the channel noise $\boldsymbol{N}$, the frequency-domain MMSE estimate $\hat{\boldsymbol{H}}_{\mathrm{MMSE}}$ becomes [1]

$$\hat{\boldsymbol{H}}_{\mathrm{MMSE}} = \mathbf{F}\mathbf{Q}_{\mathrm{MMSE}}\mathbf{F}^H\mathbf{X}^H\boldsymbol{Y} \tag{5}$$

where $\mathbf{Q}_{\mathrm{MMSE}}$ can be shown to be

$$\mathbf{Q}_{\mathrm{MMSE}} = \mathbf{R}_{\mathrm{gg}}\left[\left(\mathbf{F}^H\mathbf{X}^H\mathbf{X}\mathbf{F}\right)^{-1}\sigma_n^2 + \mathbf{R}_{\mathrm{gg}}\right]^{-1} \times \left(\mathbf{F}^H\mathbf{X}^H\mathbf{X}\mathbf{F}\right)^{-1}. \tag{6}$$

$\mathbf{R}_{\mathrm{gg}}$ is the auto-covariance matrix of $\mathbf{g}$ and σ_n^2 denotes the noise variance $\mathrm{E}\left\{\left|n_k^2\right|\right\}$. These two parameters of channel are hard to acquire for practical OFDM systems. In this respect, we propose a learning-based scheme which can achieve similar performance with that of MMSE.

B. Online C-ELM-Based Channel Estimation and Equalization

Neural network has some outstanding merits such as nonlinear mapping and trainable weights, which means that it can be used to restore nonlinear distortion and learn the information about channels. In the following, we outline the proposed method.

Fig. 1 illustrates the architecture of OFDM system with C-ELM-based channel equalizer. Apart from the SLFN channel equalizer, all other blocks are same to traditional OFDM systems. There are shaping filter and high-power amplifier (HPA) in the D/A block on the side of transmitter. Correspondingly, matched filter and automatic gain control (AGC) are in the A/D block on the side of receiver. All these processes might distort signals nonlinearly and we mainly consider the effect resulting from HPA in this article. The block of LS-DD on the receiver side represents LS-based decision-directed (DD) channel estimation and the LS channel estimation has been described in the previous. There are two reasons why LS-DD is chosen instead of MMSE. The first reason is that receiver does not have $\mathbf{R}_{\mathrm{gg}}$ and σ_n^2. Therefore, we hope SLFN

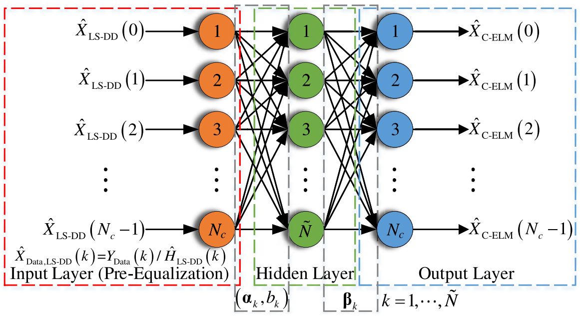

Fig. 2. The structure of C-ELM with SLFN. $\hat{\boldsymbol{X}}_{\text{Pilot,LS-DD}}$ and $\hat{\boldsymbol{X}}_{\text{Data,LS-DD}}$ are the input in the training and data stages, respectively.

could acquire both of them by learning from received data. The other reason is that received pilots equalized by LS-DD channel estimation values can be used to train the SLFN by using C-ELM algorithm without any redundant pilot symbols.

The most significant difference between C-ELM and deep neural network is that C-ELM can be trained by matrix calculus instead of backpropagation (BP)-based gradient descent algorithm. As Fig. 2 shows, the input, output and weights of C-ELM are fully complex. Samples $\left(\hat{\boldsymbol{X}}_{\text{Pilot,LS-DD}}, \boldsymbol{X}_{\text{Pilot}}\right)$ are used for training C-ELM and then received signal $\boldsymbol{Y}$ is equalized by $\hat{\boldsymbol{H}}_{\text{LS-DD}}$ preliminarily. In the following, for simplicity, the subscript "Pilot" of $\hat{\boldsymbol{X}}_{\text{Pilot,LS-DD}}$ are omitted. C-ELM algorithm can be summarized as follows [8]:

Algorithm 1 The Training Algorithm for C-ELM

Given a training set: $\mathbf{N} = \left\{\left(\hat{\boldsymbol{X}}_{\text{LS-DD},i}, \boldsymbol{X}_{\text{Pilot},i}\right) | i = 1, \cdots, N\right\}$, complex activation function $g_c(\cdot)$, and hidden neuron number $\tilde{N}$. $\hat{\boldsymbol{X}}_{\text{LS-DD},i}, \boldsymbol{X}_{\text{Pilot},i} \in \mathbb{C}^{N_c}$ and these two are corresponding to the input and desired output of the C-ELM, respectively.

step1: Randomly choose the values of complex input weight $\boldsymbol{\alpha}_k$ and the complex bias b_k, $k = 1, \cdots, \tilde{N}$.

step2: Calculate the complex hidden layer output matrix $\mathbf{H}$.

step3: Calculate the complex output weight $\boldsymbol{\beta}$ using $\hat{\boldsymbol{\beta}} = \mathbf{H}^{\dagger}\mathbf{X}_{\text{Pilot}}$.

Specifically, the actual outputs of the C-ELM are given by

$$\sum_{k=1}^{\tilde{N}} \boldsymbol{\beta}_k g_c\left(\boldsymbol{\alpha}_k \cdot \hat{\boldsymbol{X}}_{\text{LS-DD},i} + b_k\right) = \hat{\boldsymbol{X}}_{\text{C-ELM},i}, \quad i = 1, \cdots, N, \tag{7}$$

where column vector $\boldsymbol{\alpha}_k \in \mathbb{C}^{N_c}$ is complex input vector connecting the input layer neurons to the kth hidden neuron, $\boldsymbol{\beta}_k = [\beta_{k1}, \beta_{k2}, \cdots, \beta_{kN_c}]^T \in \mathbb{C}^{N_c}$ denotes the complex output weight vector connecting the kth hidden neuron and the output neurons, and $b_k \in \mathbb{C}$ means the complex bias of the kth hidden neuron. $\boldsymbol{\alpha}_k \cdot \hat{\boldsymbol{X}}_{\text{LS-DD},i}$ denotes the inner product of column vectors $\boldsymbol{\alpha}_k$ and $\hat{\boldsymbol{X}}_{\text{LS-DD},i}$. The above N equations can be written as

$$\mathbf{H}\boldsymbol{\beta} = \hat{\mathbf{X}}_{\text{C-ELM}} \tag{8}$$

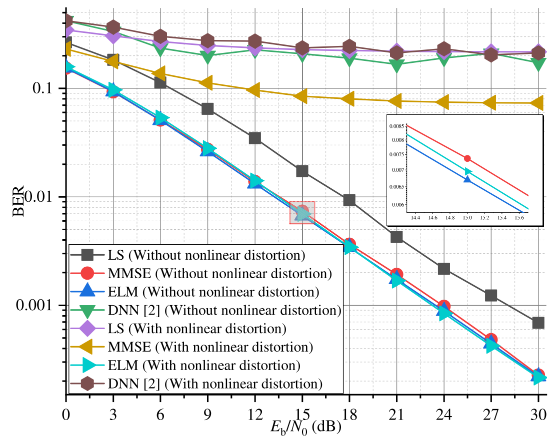

Fig. 3. Comparisons of the BER performance of the proposed C-ELM-based scheme, traditional schemes and DNN-based scheme, where with/without nonlinear distortion cases are considered.

where $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_{\tilde{N}}]^T_{\tilde{N}\times N_c}$, $\hat{\mathbf{X}}_{\text{C-ELM}} = \left[\hat{\mathbf{X}}_{\text{C-ELM},1}, \cdots, \hat{\mathbf{X}}_{\text{C-ELM},N}\right]^T_{N\times N_c}$ and

$$\mathbf{H} = \begin{bmatrix} g_c\left(\boldsymbol{\alpha}_1 \cdot \hat{\boldsymbol{X}}_{\text{LS-DD},1} + b_1\right) & \cdots & g_c\left(\boldsymbol{\alpha}_{\tilde{N}} \cdot \hat{\boldsymbol{X}}_{\text{LS-DD},1} + b_{\tilde{N}}\right) \\ \vdots & \ddots & \vdots \\ g_c\left(\boldsymbol{\alpha}_1 \cdot \hat{\boldsymbol{X}}_{\text{LS-DD},N} + b_1\right) & \cdots & g_c\left(\boldsymbol{\alpha}_{\tilde{N}} \cdot \hat{\boldsymbol{X}}_{\text{LS-DD},N} + b_{\tilde{N}}\right) \end{bmatrix}. \tag{9}$$

Different from deep learning, ELM can be trained in one shot according to minimizing training error instead of iterative training. Analytically, the LS solution $\hat{\boldsymbol{\beta}}$ of $\mathbf{H}\boldsymbol{\beta} = \mathbf{X}_{\text{Pilot}}$ with minimum norm of output weights $\boldsymbol{\beta}$ is given by

$$\left\|\mathbf{H}\hat{\boldsymbol{\beta}} - \mathbf{X}_{\text{Pilot}}\right\| = \min_{\boldsymbol{\beta}} \left\|\mathbf{H}\boldsymbol{\beta} - \mathbf{X}_{\text{Pilot}}\right\| \rightarrow \hat{\boldsymbol{\beta}} = \mathbf{H}^{\dagger}\mathbf{X}_{\text{Pilot}}. \tag{10}$$

$\tilde{N}$ should be larger than N_c and g_c is $\text{arcsinh}(z) = \int_0^z dt / \left[\left(1+t^2\right)^{1/2}\right]$ [9], where $z \in \mathbb{C}$. Once enough pilots are collected, SLFN will be trained and operate. Then, in the data stage, the input of the C-ELM is $\hat{\boldsymbol{X}}_{\text{Data,LS-DD}}$ and its corresponding output $\hat{\boldsymbol{X}}_{\text{C-ELM}}$ can be demodulated directly.

III. Simulation Results and Analysis

Theoretically, MMSE with accurate channel parameters $\mathbf{R}_{\text{gg}}$ and σ_n^2 can achieve the best performance. Therefore, it is necessary to verify that whether the proposed method can learn the information about channels and achieve performance near to MMSE. In the following simulations, the performance and generalization of the proposed method compared with LS, MMSE and DNN under various disadvantages are given.

A. Performance Under Nonlinear Distortion

The transmitted signals are affected by HPA which is characterized by nonlinear amplitude conversion (AM/AM) and phase conversion (AM/PM) as following formulas shown:

$$A(x) = \frac{\alpha_a x}{1+\beta_a x^2}, \quad \Phi(x) = \frac{\alpha_\phi x^2}{1+\beta_\phi x^2}, \tag{11}$$

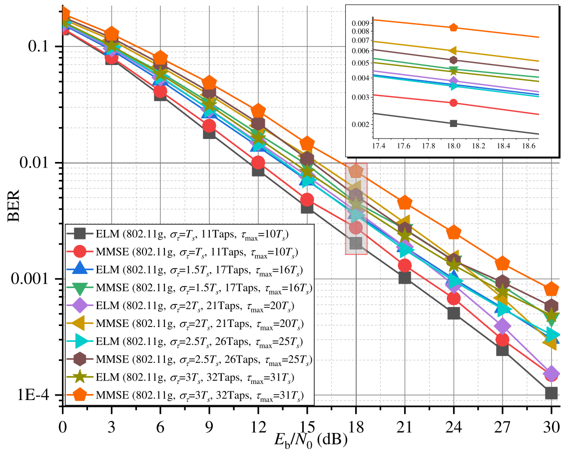

Fig. 4. BER versus the E_b/N_0 in the C-ELM-based scheme and MMSE, where five channel conditions are considered. σ_τ and $\tau_{\max}$ represent root mean square (RMS) delay spread and maximum delay, respectively.

where x is the amplifier input signal amplitude and $A(x)$ and $\Phi(x)$ represent AM/AM and AM/PM nonlinear distortion of amplifier respectively. We set $\alpha_a = 1.96$, $\beta_a = 0.99$, $\alpha_\phi = 2.53$ and $\beta_\phi = 2.82$ according to [10]. $N_c = 64$, $\tilde{N} = 512$, $N = 10^4$ and $T_G = N_c/4$. Modulation mode is 4-QAM. Channel is modeled as exponential model with 8 stochastic fading delay paths and perfect frame synchronization is guaranteed. Assuming that the channel is quasi-static during each channel realization and block pilots are used for channel estimation and training for C-ELM. Fig. 3 shows the BER performance versus E_b/N_0 of the proposed scheme, MMSE, LS and DNN [2], in which the cases with/without nonlinear distortion are investigated. It can be seen that the proposed scheme significantly outperforms DNN and slightly outperforms MMSE when there is not nonlinear distortion caused by HPA. In other words, the C-ELM has learnt the information about the channel. When signals are distorted by HPA, the performance of proposed scheme is significantly better than the MMSE, LS and DNN.

B. Analysis of Robustness Under Various Channels

Practically, the parameters of channels vary with scenarios and these are unknown to receivers. Therefore, if a system can perform well in various channels, its robustness can be proved. Fig. 4 illustrates the BER performance versus the E_b/N_0 of the proposed scheme and MMSE under five channel conditions referring 802.11g channel model. It can be seen that the proposed method shows stable performance and achieves more gain comparing with MMSE as $\tau_{\max}$ increases.

C. Analysis of Generalization Ability

Fig. 5 shows the BER performance versus the E_b/N_0 of the proposed scheme, MMSE and DNN with different waveform parameters. From this figure, the performance of ELM-based scheme is same with MMSE when $N_c = 16$ or 64 with modulation order $M \leq 2$, which outperforms the offline DNN scheme [2] with three hidden layers (the numbers of neurons in each hidden layer are 500, 250, 120, respectively). However its performance degrades when $N_c > 64$ or $M > 2$. It means that if the number of input pattern (e.g., the number

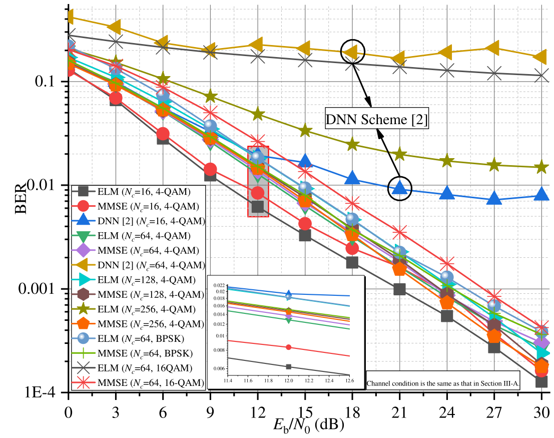

Fig. 5. The BER of the C-ELM-based scheme, MMSE and DNN in the case of $N_c = 16,\ 64,\ 128,\ 256$ with $M = 2$ and $N_c = 64$ with $M = 1,\ 4$.

of subcarriers and modulation order) of neural networks is large, the performance of networks will be affected. From this aspect, C-ELM only has single hidden layer and thus it has lower computational complexity, but it still has excellent performance and outperforms offline deeper networks.

IV. Conclusions

In this letter, we proposed an online C-ELM-based channel estimation and equalization scheme for OFDM systems. Simulation results show that the proposed scheme has the best performance comparing with MMSE, LS and DNN-based scheme and it has excellent robustness under different multi-path fading channels. We also find that the generalization ability of C-ELM scheme is better than offline DNN-based scheme but still limited. Therefore, enhancing the generalization ability of C-ELM with limited pilots will be considered in the future.

References

[1] J.-J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in OFDM systems," in *Proc. IEEE 45th Veh. Technol. Conf.*, vol. 2, Jul. 1995, pp. 815–819.

[2] H. Ye, G. Y. Li, and B. H. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018.

[3] 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.

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

[5] N. Farsad and A. J. Goldsmith. (2017). "Detection algorithms for communications systems using deep learning." [Online]. Available: https://arxiv.org/abs/1705.08044

[6] T. Wang, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based CSI feedback approach for time-varying massive MIMO channels," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 416–419, Apr. 2019.

[7] H. Ye, G. Y. Li, B.-H. F. Juang, and K. Sivanesan, "Channel agnostic end-to-end learning based communication systems with conditional GAN," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–5.

[8] M.-B. Li, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "Fully complex extreme learning machine," *Neurocomputing*, vol. 68, nos. 1–4, pp. 306–314, Oct. 2005.

[9] T. Kim and T. Adali, "Approximation by fully complex multilayer perceptrons," *Neural Comput.*, vol. 15, no. 7, pp. 1641–1666, Jul. 2003.

[10] A. A. M. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers," *IEEE Trans. Commun.*, vol. COM-29, no. 11, pp. 1715–1720, Nov. 1981.