

Massive MIMO as an Extreme Learning Machine

Dawei Gao ¹, Qinghua Guo ¹, *Senior Member, IEEE*,
and Yonina C. Eldar ², *Fellow, IEEE*

Abstract—This work shows that a massive multiple-input multiple-output (MIMO) system with low-resolution analog-to-digital converters (ADCs) forms a natural extreme learning machine (ELM). The receive antennas at the base station serve as the hidden nodes of the ELM, and the low-resolution ADCs act as the ELM activation function. By adding random biases to the received signals and optimizing the ELM output weights, the system can effectively tackle hardware impairments, such as the nonlinearity of power amplifiers and the low-resolution ADCs. Moreover, the fast adaptive capability of ELM allows the design of an adaptive receiver to address time-varying effects of MIMO channels. Simulations demonstrate the promising performance of the ELM-based receiver compared to conventional receivers in dealing with hardware impairments.

Index Terms—Massive MIMO, ELM, signal detection, nonlinear distortion, low-resolution ADC, hardware impairments.

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) systems, where the base station (BS) is equipped with a large number of antennas, is a promising technology for 5G and future generation wireless communications [1]. However, the requirement for a large number of radio frequency (RF) chains leads to high power consumption. To address this challenge, various techniques have been proposed to reduce the number of RF chains [2] and low-resolution analog-to-digital converters (ADCs) have been considered [3]–[5]. In addition to low-resolution ADCs at the BS, hardware impairments of the user equipment may also need to be addressed. For example, the use of cheap power amplifiers (PAs) can lead to severe nonlinear distortions to transmitted signals [6], [7]. Previous work has shown that hardware impairments have to be handled properly to avoid system performance degradation [3]–[5], [7]. Existing results consider impairments on either the BS or the user side, however they do not consider impairments on both sides.

An extreme learning machine (ELM) is a single-hidden layer feed-forward neural network (NN). In an ELM, the input weights and biases are randomly initialized and fixed, so that the only parameters to be learned are its output weights. Learning reduces to solving a least squares (LS) problem, making ELM fast in learning [8]. ELM receivers have been designed in our previous works [9], [10] to handle LED nonlinearity and/or cross-LED interference. These works show

Manuscript received July 1, 2020; revised November 27, 2020; accepted December 22, 2020. Date of publication December 29, 2020; date of current version February 12, 2021. This work was supported by the Air Force Office of Scientific Research under Grant FA9550-18-1-0208. The work of Yonina C. Eldar was also supported by the Futurewei and Huawei Technologies. The review of this article was coordinated by Prof. Jingon Joung. (*Corresponding author: Qinghua Guo.*)

Dawei Gao and Qinghua Guo are with the School of Electrical, Computer, and Telecommunications Engineering, University of Wollongong, Wollongong, NSW 2522, Australia (e-mail: dg687@uowmail.edu.au; qguo@uow.edu.au).

Yonina C. Eldar is with the Faculty of Math and CS, Weizmann Institute of Science, 7610001 Rehovot, Israel (e-mail: yonina.eldar@weizmann.ac.il).

Digital Object Identifier 10.1109/TVT.2020.3047865

that ELM is very effective in dealing with nonlinear distortions and delivers much better performance than conventional polynomial based techniques [9], [10]. ELM has also been used for channel estimation and data detection in OFDM systems [11], [12].

In this work, we address the hardware impairments at both the BS and the user side in a massive MIMO system using the concept of an ELM. Although deep NNs have attracted much attention for MIMO receiver design (such as [4], [13]), we focus on ELM (with a single-hidden layer) due to its fast learning and adaptive capabilities. In particular, no back propagation is needed and the size of training samples required is relatively small. We consider the uplink of a massive MIMO system with low-resolution ADCs, where the transmitted signals of users suffer from PA nonlinear distortions. We show that the massive MIMO system forms a *natural* ELM. Specifically, the transmit antennas of users are regarded as the input nodes of the ELM, the massive number of antennas at the BS serve as the hidden nodes of the ELM so that the channel matrix functions as the ELM input weight matrix, and the low-resolution ADCs act as the activation function. We add random biases to the received signals before the ADC and learn the output weights of the ELM from training signals. As the ELM output weight optimization is simply an LS problem, an adaptive receiver is designed to deal with time-varying channels. Simulation results demonstrate the promising performance of the ELM based system in handling hardware impairments compared to conventional techniques.

Notations: Boldface lower-case and upper-case letters denote vectors and matrices, respectively. The superscripts $(\cdot)^T$ and $(\cdot)^H$ represent the transpose and conjugate transpose operations. We use $|x|$ and $\|\mathbf{x}\|$ to denote the amplitude of x and the norm of \mathbf{x} , and $\Re\{\cdot\}$, $\Im\{\cdot\}$ represent the real and imaginary parts of a complex number, respectively.

II. MASSIVE MIMO WITH HARDWARE IMPAIRMENTS

Consider the uplink transmission of a massive MIMO system with K active users. Each user has a single antenna and the BS is equipped with N antennas, where N can be much larger than K . We consider two hardware impairments: PA nonlinear distortion at the transmitter (user) side and low-resolution ADCs at the receiver (BS) side.

The PA nonlinear distortion can be characterized by the amplitude to amplitude conversion $A(|x|)$ and amplitude to phase conversion $\Phi(|x|)$ [14]:

$$A(|x|) = \frac{\alpha_a |x|}{1 + \epsilon_a |x|^2}, \quad \Phi(|x|) = \frac{\alpha_\phi |x|^2}{1 + \epsilon_\phi |x|^2}, \quad (1)$$

where x is the signal input to the PA, and α_a , ϵ_a , ϵ_ϕ and β_ϕ are parameters. The distorted signal due to the PA nonlinearity is expressed as

$$s = f(x) = A(|x|)e^{j(\text{angle}(x) + \Phi(|x|))}, \quad (2)$$

where $\text{angle}(x)$ denotes the angle of the complex signal x .

The received signal at time instant m is modeled as

$$\mathbf{y}(m) = \mathbf{H}\mathbf{s}(m) + \mathbf{n}(m), \quad (3)$$

where \mathbf{H} is an $N \times K$ MIMO channel matrix, $\mathbf{s}(m) = f(\mathbf{x}(m))$ with $\mathbf{x}(m) = [x_1(m), x_2(m), \dots, x_K(m)]^T$ being the transmitted signals of the K users at time instant m , and $\mathbf{n}(m)$ denotes an additive white Gaussian noise vector. Conventionally the receiver signal is quantized directly, however in this work, a random bias is added to the received signal, which is then quantized, i.e.,

$$\mathbf{r}(m) = Q(\Re\{\mathbf{y}(m)\} + \mathbf{b}^{\text{Re}}) + jQ(\Im\{\mathbf{y}(m) + \mathbf{b}^{\text{Im}}\}), \quad (4)$$

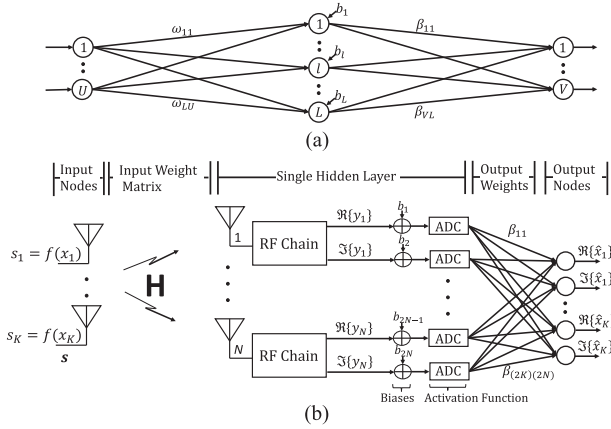


Fig. 1. Massive MIMO system forms an ELM, where the received signals are biased before quantization. (a) ELM. (b) Massive MIMO.

where \mathbf{b}^{Re} and \mathbf{b}^{Im} represent the biases for the real part and imaginary part of $\mathbf{y}(m)$, respectively, and $Q(\cdot)$ denotes an element-wise quantization operation. For a uniform mid-rise quantizer, we have

$$Q(c) = \Delta (\lfloor c/\Delta \rfloor + 0.5), \quad (5)$$

where c denotes the signal to be quantized, $\lfloor \cdot \rfloor$ is the floor function and Δ is the quantization step size. Due to the limited number of bits used for the ADCs, the input signal c is clipped when its amplitude exceeds a threshold. The aim of the receiver is to recover $\mathbf{x}(m)$ based on $\mathbf{r}(m)$.

III. FROM ELM TO ELM BASED MASSIVE MIMO SYSTEM

A. ELM Structure and Training

The structure of an ELM with U input nodes, L hidden nodes and V output nodes is shown in Fig. 1(a). The input weights $\{\omega_{lu}\}$ and biases $\{b_l\}$ are randomly initialized and fixed [8]. The parameters to be learned are the output weights $\{\beta_{vl}\}$.

Suppose we have an input $\mathbf{p}(m) = [p_1(m), p_2(m), \dots, p_U(m)]^T$. The output of the hidden nodes of the ELM shown in Fig. 1(a) can be expressed as

$$\mathbf{z}(m) = g(\mathbf{W}\mathbf{p}(m) + \mathbf{b}), \quad (6)$$

where \mathbf{W} is the input weight matrix, whose (l, u) th element ω_{lu} associates the u th input node and the l th hidden node, $g(\cdot)$ is the activation function, and $\mathbf{b} = [b_1, b_2, \dots, b_L]^T$ is the bias vector. The v th ELM output is then

$$\psi_v(m) = \mathbf{z}(m)^T \boldsymbol{\beta}_v, \quad (7)$$

where $\boldsymbol{\beta}_v = [\beta_{v1}, \beta_{v2}, \dots, \beta_{vL}]^T$, and β_{vl} denotes the output weight associating the l th hidden node with the v th output node.

Suppose that there are M training samples $\{\mathbf{p}(m), \mathbf{t}(m)\}, m = 1, \dots, M$, where $\mathbf{t}(m) = [t_1(m), t_2(m), \dots, t_V(m)]^T$ denotes the expected output. We can write (7) in matrix form as

$$\boldsymbol{\psi}_v = \mathbf{Z}\boldsymbol{\beta}_v, \quad (8)$$

with $\boldsymbol{\psi}_v = [\psi_v(1), \psi_v(2), \dots, \psi_v(M)]^T$, where \mathbf{Z} is an $M \times L$ matrix given by

$$\mathbf{Z} = [\mathbf{z}(1), \mathbf{z}(2), \dots, \mathbf{z}(M)]^T. \quad (9)$$

The output weight vector $\boldsymbol{\beta}_v$ is trained by minimizing the cost function $\|\boldsymbol{\psi}_v - \mathbf{t}_v\|^2$ based on the linear model (8), where $\mathbf{t}_v =$

$[t_v(1), t_v(2), \dots, t_v(M)]^T$. The L_2 regularized LS solution is given by [15]

$$\boldsymbol{\beta}_v = (\mathbf{Z}^T \mathbf{Z} + \gamma \mathbf{I})^{-1} \mathbf{Z}^T \mathbf{t}_v, \quad v = 1, \dots, V \quad (10)$$

where \mathbf{I} is an identity matrix and γ is a regularization parameter.

B. Massive MIMO as an ELM

Our design is inspired by the fact that the BS is equipped with a large number of antennas N (normally $N > K$), and the quantization is nonlinear. As illustrated in Fig. 1, by biasing the received signals, a massive MIMO system naturally forms an ELM.

Comparing Fig. 1(b) to Fig. 1(a), we consider the K transmit antennas as the input nodes of the ELM. The distorted signal $\mathbf{s}(m) = f(\mathbf{x}(m))$ is transmitted over the air, and detected by the receive antennas at the BS. We treat the receive antennas as the hidden nodes of the ELM, so that the channel matrix \mathbf{H} resembles the input weight matrix of the ELM. The received signals are biased before quantization, as shown in Fig. 1(b). As the signals are complex-valued, we separate their real parts and imaginary parts. The quantized signals are then represented in vector form as

$$\begin{aligned} \mathbf{r}'(m) &= Q \left(\begin{bmatrix} \Re\{\mathbf{H}\mathbf{s}(m)\} \\ \Im\{\mathbf{H}\mathbf{s}(m)\} \end{bmatrix} + \mathbf{b} + \begin{bmatrix} \Re\{\mathbf{n}(m)\} \\ \Im\{\mathbf{n}(m)\} \end{bmatrix} \right) \\ &= Q(\mathbf{H}'\mathbf{s}'(m) + \mathbf{b} + \mathbf{n}'(m)), \end{aligned} \quad (11)$$

where the length- $2N$ bias vector \mathbf{b} is randomly generated and fixed,

$$\mathbf{H}' = \begin{bmatrix} \Re\{\mathbf{H}\} & -\Im\{\mathbf{H}\} \\ \Im\{\mathbf{H}\} & \Re\{\mathbf{H}\} \end{bmatrix}, \quad \mathbf{s}'(m) = \begin{bmatrix} \Re\{\mathbf{s}(m)\} \\ \Im\{\mathbf{s}(m)\} \end{bmatrix}, \quad (12)$$

and $\mathbf{n}'(m) = [\Re\{\mathbf{n}(m)^T\}, \Im\{\mathbf{n}(m)^T\}]^T$. Treating $Q(\cdot)$ as the activation function, the only difference between (11) and (6) is the extra noise term $\mathbf{n}'(m)$.

If we ignore the noise term, then $\mathbf{r}'[m]$ resembles the hidden layer output vector of the ELM. By mimicking ELM, the outputs of the ADCs are weighted to recover the signal $\mathbf{x}(m)$ at the output nodes in Fig. 1 (b). The transmitters, the massive MIMO channel and the BS receiver form a complete ELM, where the input is the distorted signal $\mathbf{s}(m)$ and the output is an estimate of $\mathbf{x}(m)$ (or more precisely estimates of $\Re\{\mathbf{x}(m)\}, \Im\{\mathbf{x}(m)\}$).

1) Receiver Training: The ELM output weight vectors $\{\beta_k^{\text{Re}}, \beta_k^{\text{Im}}, k = 1, 2, \dots, K\}$, each pair corresponding to a user, can be learned using training signals $\{\mathbf{s}^{\text{Train}}(m), \mathbf{x}^{\text{Train}}(m), m = 1, \dots, M\}$. In the resulting ELM, the output vector of the hidden nodes $\mathbf{r}'(m)$ in (11) is formed by signal transmission and low resolution ADC quantization without any additional computations, enabling a low complexity receiver. Collecting the outputs of the ADCs, we get the hidden node output matrix

$$\mathbf{R}' = [\mathbf{r}'(1), \mathbf{r}'(2), \dots, \mathbf{r}'(M)]^T. \quad (13)$$

The values of β_k^{Re} and β_k^{Im} can then be obtained by solving two regularized LS problems, i.e.,

$$\beta_k^{\text{Re}} = (\mathbf{R}'^T \mathbf{R}' + \gamma \mathbf{I})^{-1} \mathbf{R}'^T \mathbf{t}_k^{\text{Re}}, \quad (14)$$

$$\beta_k^{\text{Im}} = (\mathbf{R}'^T \mathbf{R}' + \gamma \mathbf{I})^{-1} \mathbf{R}'^T \mathbf{t}_k^{\text{Im}}, \quad (15)$$

where

$$\mathbf{t}_k^{\text{Re}} = \Re\{[x_k^{\text{Train}}(1), \dots, x_k^{\text{Train}}(M)]^T\}, \quad (16)$$

$$\mathbf{t}_k^{\text{Im}} = \Im\{[x_k^{\text{Train}}(1), \dots, x_k^{\text{Train}}(M)]^T\}. \quad (17)$$

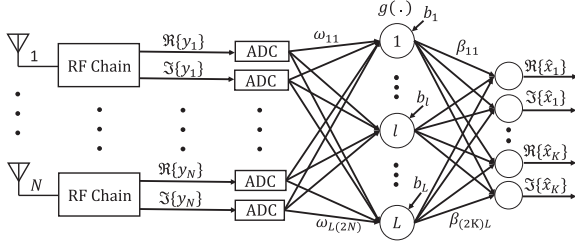


Fig. 2. ELM receiver borrowed from [9].

2) *Data Detection*: Once the output weights for the users are learned, they are applied to the received signals to estimate the transmitted data symbols $\{x_k(m)\}$ via

$$\tilde{x}_k(m) = (\beta_k^{\text{Re}})^T \mathbf{r}'(m) + j(\beta_k^{\text{Im}})^T \mathbf{r}'(m), \quad (18)$$

where $\mathbf{r}'(m)$ is the output of the ADCs at time instant m . A decision based on $\tilde{x}_k(m)$ is then made, i.e.,

$$\hat{x}_k(m) = \underset{c}{\text{argmin}} | \tilde{x}_k(m) - c |^2, \quad (19)$$

where c belongs to the symbol alphabet.

C. Comparisons With Other Receivers

1) *Conventional ZF and MMSE Receiver*: Given the channel matrix \mathbf{H} , the weight for user k using the zero-forcing (ZF) receiver is represented as

$$(\mathbf{w}_k^{\text{ZF}})^T = [(\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H]_k, \quad (20)$$

where $[\cdot]_k$ denotes the k th row of the matrix. If the noise power or the signal to noise power ratio (SNR) is known, we can use the minimum mean squared error (MMSE) receiver whose weight for user k is

$$(\mathbf{w}_k^{\text{MMSE}})^T = \left[\left(\mathbf{H}^H \mathbf{H} + \frac{1}{\text{SNR}} \mathbf{I} \right)^{-1} \mathbf{H}^H \right]_k. \quad (21)$$

Both receivers ignore the nonlinear distortion at the transmitter side and the impact of the low-resolution ADCs at the receiver side, resulting in very poor performance as shown in Section V. It is noted that both receivers require the channel matrix, which needs to first be estimated with training signals.

2) *ZF Receiver With Training*: The ZF receiver can also directly be trained using training signals. In this case the weight of the receiver for user k can be expressed as

$$\mathbf{w}_{k(\text{Train})}^{\text{ZF}} = (\mathbf{R}^H \mathbf{R} + \gamma \mathbf{I})^{-1} \mathbf{R}^H \mathbf{x}_k^{\text{Train}}, \quad (22)$$

where $\mathbf{R} = [\mathbf{r}(1), \mathbf{r}(2), \dots, \mathbf{r}(M)]^T$ is the matrix of the quantized signals (no biasing before quantization), and γ is a regularization parameter. We can also separate the real and imaginary parts, so that the weights are obtained similarly to (14) and (15). The difference between the proposed receiver and the trained ZF receiver is that the received signals are biased before quantization in the proposed receiver. It is interesting that the directly trained detector performs better than the detectors with perfect \mathbf{H} (see Section V) because the hardware impairments are considered in training.

3) *ELM Receiver Borrowed From [9]*: In [9], we proposed an ELM receiver to handle both the LED nonlinearity and cross-LED interference in MIMO LED communications. The receiver can be readily extended to massive MIMO, as shown in Fig. 2. The input to the ELM is the quantized signal (the output of the ADCs), the number of hidden nodes is L , and the activation function is the sigmoid function.

With the training signals, the output weights of the ELM can also be learned, and then applied for data symbol detection, similar to the proposed receiver. The proposed ELM based receiver in Fig. 1(b) is very different, where the multiplication of the input weight matrix with the input vector is naturally accomplished by signal transmission over the air, which leads to much lower complexity of the new ELM based receiver in both training and detection. In data detection, the proposed ELM based receiver only needs to carry out (18) and (19). However, the ELM receiver borrowed from [9] needs to implement matrix-vector multiplications. In addition, as shown in Section V, the proposed ELM based receiver results in considerably better performance.

IV. ELM BASED ADAPTIVE RECEIVER DESIGN

ELM is attractive in that it allows fast adaptive learning as its output weights can be easily updated. This endows ELM with the unique capability of dealing with time-varying massive MIMO channels. By leveraging the online sequential-ELM (OSELM) [16], the ELM based receiver is readily extended to an adaptive setting.

The adaptive receiver consists of two learning phases, an initialization phase and a sequential learning phase. Suppose that we have M_0 training samples for initialization. The output weights for the users can be computed using (14) and (15), which are denoted as $\beta_k^{\text{Re}}(0)$ and $\beta_k^{\text{Im}}(0)$ for user k . Define

$$\mathbf{P}(0) = (\mathbf{R}'^T \mathbf{R}' + \gamma \mathbf{I})^{-1}. \quad (23)$$

In the subsequent learning phase with new training samples $\{(\mathbf{s}^{\text{Train}}(m), \mathbf{x}^{\text{Train}}(m))\}$ arriving one by one, the adaptive receiver recursively updates its weights. Taking β_k^{Re} as an example, we have

$$\mathbf{q}(m) = \frac{\mathbf{P}(m-1) \mathbf{r}'(m)}{\lambda + (\mathbf{r}'(m))^T \mathbf{P}(m-1) \mathbf{r}'(m)}, \quad (24)$$

$$e_k(m) = \Re\{x_k^{\text{Train}}(m)\} - (\beta_k^{\text{Re}}(m-1))^T \mathbf{r}'(m), \quad (25)$$

$$\beta_k^{\text{Re}}(m) = \beta_k^{\text{Re}}(m-1) + \mathbf{q}(m) e_k(m), \quad (26)$$

$$\mathbf{P}(m) = \frac{1}{\lambda} [\mathbf{P}(m-1) - \mathbf{q}(m) (\mathbf{r}'(m))^T \mathbf{P}(m-1)], \quad (27)$$

where $\lambda \in [0, 1]$ is a forgetting factor to adjust the tracking capability and convergence rate. The above sample-by-sample update can also be extended to a chunk-by-chunk update [17].

Although the adaptive receiver may require a relatively large number of training samples (still smaller compared to deep NNs) for initialization, a small number of training samples are sufficient for tracking. This makes the adaptive receiver very attractive in handling time-varying massive MIMO channels, as demonstrated in Section V.

V. SIMULATION RESULTS

Assume that the BS is equipped with a uniform linear array (ULA) of $N = 256$ antennas, number of users $K = 10$, and 16-QAM is used. As in [14], the parameter setting for the PA nonlinearity is $\alpha_a = 1.96$, $\epsilon_a = 0.99$, $\alpha_\phi = 2.53$ and $\epsilon_\phi = 2.82$. Assume 6-bit ADCs. The SNR is defined as $\text{SNR} = P_s / \sigma_n^2$, where P_s is the power of the signal (per transmit antenna), and σ_n^2 is the power of the noise (per receive antenna). We employ the massive channel model in [18], [19] and [20], which considers both spatial and temporal correlations, and assume that the ULA has a half-wavelength spacing. Table I lists the parameters used for massive MIMO channel generation. Small scale fading is considered and the multiple rays of each user are normalized. The power angular spectrum is modelled using a truncated Laplacian distribution. We consider the channel under a typical mobility scenario, i.e., urban

TABLE I
 PARAMETERS SETTING FOR MASSIVE MIMO CHANNEL

Parameters of the massive MIMO channel	Value
Number of antennas N at BS	256
Number of users K	10
Carrier frequency	2GHz
Symbol duration T_s	$1\mu s$
Mean value of AOA	$[-0.5\pi, 0.5\pi]$
Angular spread (AS)	10°
Number of multiple rays for each user	5

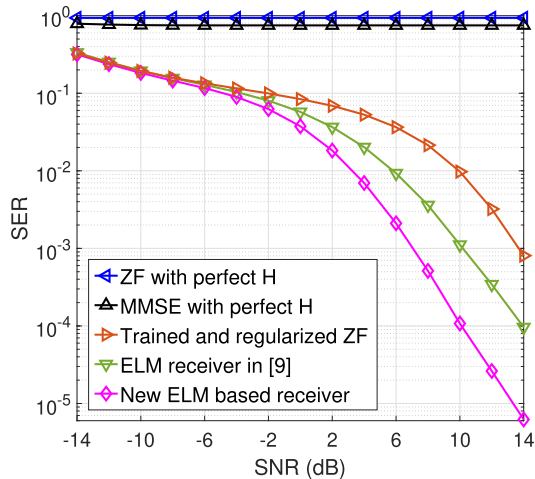


Fig. 3. SER performance of various receivers.

macro [20]. For the ZF receiver with training, the weights are computed with real and imaginary parts separated. For the ELM receiver in [9], 512 hidden nodes are used, and the input weights and the biases (also the biases of the new ELM based receiver) are drawn independently from a uniform distribution $[-0.1, 0.1]$.

Figure 3 shows the symbol error rate (SER) of the new ELM based receiver, the ELM receiver borrowed from [9] and the ZF and MMSE receivers with perfect channel state information and noise power (but note that it is difficult to acquire them when low resolution ADCs are used). Quasi-static channels are assumed without considering the mobility of users, and the training length is 3000. It can be seen from Fig. 3 that the conventional ZF and MMSE receivers deliver poor performance due to the inability to mitigate the PA nonlinearity. The trained and regularized ZF receiver performs better than them as the impact of nonlinearity is considered in training. In contrast, the ELM receiver in [9] and the new ELM based receiver can effectively handle the hardware impairments, and the new ELM based receiver delivers the best performance, with significantly lower complexity compared to the ELM receiver in [9].

The difference between the trained and regularized ZF receiver and the new ELM receiver is that the received signals are biased before quantization in the new receiver. Their huge performance difference indicates the impact of signal biasing. To further examine the impact of signal biasing and quantization, we assume two trained and regularized ZF receivers without quantization: one performs signal biasing but the other one does not. The results are shown in Fig. 4. It is interesting to see that the trained ZF with biasing performs better than the trained ZF without biasing. This indicates that, for a linear receiver, adding biases to the received signals is helpful to deal with the nonlinear distortions. However, the trained ZF receiver with biasing still performs much worse than the new ELM based receiver. This indicates that the low resolution ADCs are helpful in dealing with the nonlinear distortion when they are

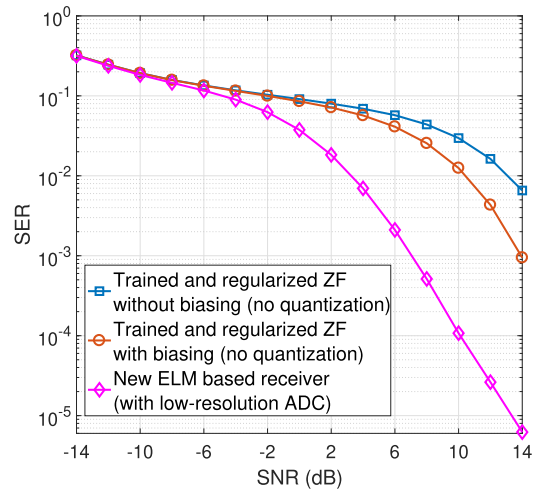


Fig. 4. Impacts of biasing and quantization.

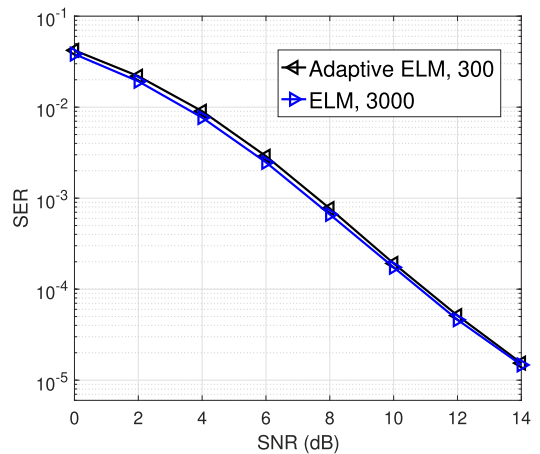


Fig. 5. Adaptive receiver with time-varying channels.

exploited as the ELM activation function. The ADCs mimic a scaled version of typically used activation functions, e.g., \tanh , due to the clipping affect.

Figure 5 shows the SER performance of the adaptive receiver with time-varying channels, where a mobile velocity of 100 km/h is assumed. After the initialization with 3000 training symbols, only 300 training symbols in each subsequent frame are used for receiver update. The forgetting factor λ is 0.98, and chunk-by-chunk update is used. As a benchmark, we also show the performance of the receiver, where we assume that 3000 training symbols are available for training in each frame. We can see that the adaptive receiver with training length 300 can achieve almost the same performance as that with training length 3000, indicating that training length 300 is sufficient for the receiver to track the channel.

Note that no explicit channel estimation is needed for the new receiver and the ZF receiver with training, which is in contrast to the conventional ZF and MMSE receivers. Batch training of the new ELM receiver and the ZF receiver requires the same complexity of $\mathcal{O}(N^3 + MN^2)$. As for detection, the new ELM receiver has the same complexity as the ZF and MMSE receivers, i.e., $\mathcal{O}(N)$ per user.

VI. CONCLUSION

We have shown that massive MIMO with low resolution ADCs can be treated as a natural ELM where the massive number of antennas serve as the hidden nodes and the ADCs act as the activation function of the ELM. By adding biases to the received signals and optimizing the ELM output weights, the receiver can effectively handle hardware impairments. An adaptive receiver is also designed and its capability of tracking time-varying channels is demonstrated.

REFERENCES

- [1] J. G. Andrews *et al.*, "What will 5G be?," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [2] S. S. Ioushua and Y. C. Eldar, "A family of hybrid analog-digital beamforming methods for massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 12, pp. 3243–3257, Jun. 2019.
- [3] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 4038–4051, Apr. 2017.
- [4] N. Shlezinger and Y. C. Eldar, "Task-based quantization with application to MIMO receivers," *Commun. Inf. Syst.*, vol. 20, pp. 131–162, 2020.
- [5] J. Singh, O. Dabeer, and U. Madhow, "On the limits of communication with low-precision analog-to-digital conversion at the receiver," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3629–3639, Dec. 2009.
- [6] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.
- [7] L. Ding *et al.*, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 159–165, Mar. 2004.
- [8] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, Dec. 2006.
- [9] D. Gao and Q. Guo, "Extreme learning machine-based receiver for MIMO LED communications," *Digit. Signal Process.*, vol. 95, Dec. 2019, Art. no. 102594.
- [10] D. Gao, Q. Guo, J. Tong, N. Wu, J. Xi, and Y. Yu, "Extreme-learning-machine-based noniterative and iterative nonlinearity mitigation for LED communication systems," *IEEE Syst. J.*, vol. 14, no. 4, pp. 4674–4683, Dec. 2020.
- [11] 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.
- [12] L. Yang, Q. Zhao, and Y. Jing, "Channel equalization and detection with ELM-based regressors for OFDM systems," *IEEE Commun. Lett.*, vol. 24, no. 1, pp. 86–89, Nov. 2020.
- [13] H. He, C. Wen, S. Jin, and G. Y. Li, "Model-driven deep learning for MIMO detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 1702–1715, Feb. 2020, doi: [10.1109/TSP.2020.2976585](https://doi.org/10.1109/TSP.2020.2976585).
- [14] A. A. M. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers," *IEEE Trans. Commun.*, vol. 29, no. 11, pp. 1715–1720, Nov. 1981.
- [15] W. Deng, Q. Zheng, and L. Chen, "Regularized extreme learning machine," in *Proc. IEEE Symp. CIDM*, Mar. 2009, pp. 389–395.
- [16] N. Liang, G. 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.
- [17] J.-S. Lim, S. Lee, and H.-S. Pang, "Low complexity adaptive forgetting factor for online sequential extreme learning machine (OS-ELM) for application to nonstationary system estimations," *Neural Comput. Appl.*, vol. 22, nos. 3/4, pp. 569–576, 2013.
- [18] J. Ma, S. Zhang, H. Li, F. Gao, and S. Jin, "Sparse Bayesian learning for the time-varying massive MIMO channels: Acquisition and tracking," *IEEE Trans. Wireless Commun.*, vol. 67, no. 3, pp. 1925–1938, Mar. 2019.
- [19] K. Liu, V. Raghavan, and A. M. Sayeed, "Capacity scaling and spectral efficiency in wide-band correlated MIMO channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2504–2526, Oct. 2003.
- [20] L. You, X. Gao, A. L. Swindlehurst, and W. Zhong, "Channel acquisition for massive MIMO-OFDM with adjustable phase shift pilots," *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1461–1476, Mar. 2016.