Safi *et al. J Wireless Com Network* (2021) 2021:47
https://doi.org/10.1186/s13638-021-01929-z

EURASIP Journal on Wireless Communications and Networking

RESEARCH Open Access

Autoencoder-bank based design for adaptive channel-blind robust transmission

Hossein Safi[1], Mohammad Akbari[2*], Elaheh Vaezpour[2], Saeedeh Parsaeefard[3] and Raed M Shubair[4,5]

*Correspondence:
m.akbari@itrc.ac.ir
[2] Department of Communication Technology, ICT Research Institute (ITRC), North Karegar, Tehran, Iran
Full list of author information is available at the end of the article

Abstract

The idea of employing deep autoencoders (AEs) has been recently proposed to capture the end-to-end performance in the physical layer of communication systems. However, most of the current methods for applying AEs are developed based on the assumption that there exists an explicit channel model for training that matches the actual channel model in the online transmission. The variation of the actual channel indeed imposes a major limitation on employing AE-based systems. In this paper, without relying on an explicit channel model, we propose an adaptive scheme to increase the reliability of an AE-based communication system over different channel conditions. Specifically, we partition channel coefficient values into sub-intervals, train an AE for each partition in the offline phase, and constitute a bank of AEs. Then, based on the actual channel condition in the online phase and the average block error rate (BLER), the optimal pair of encoder and decoder is selected for data transmission. To gain knowledge about the actual channel conditions, we assume a realistic scenario in which the instantaneous channel is not known, and propose to blindly estimate it at the Rx, i.e., without any pilot symbols. Our simulation results confirm the superiority of the proposed adaptive scheme over existing methods in terms of the average power consumption. For instance, when the target average BLER is equal to 10^{-4}, our proposed algorithm with 5 pairs of AE can achieve a performance gain over 1.2 dB compared with a non-adaptive scheme.

Keywords: Autoencoder, Blind estimation, Deep learning

1 Introduction

To reliably transmit data from a source to a destination, conventional communication systems employ multiple independent blocks which are separately optimized to perform isolated functions (e.g., source/channel coding, modulation, channel estimation, equalization) [1]. However, such a divided architecture is known to be sub-optimal [2], and thus, achieving optimal performance through the end-to-end optimization of a communication system retains appealing for carrying out further investigations [3].

Recently, considerable advances in Deep Learning (DL) have empowered researchers to efficiently perform end-to-end learning of communication systems [3]. This way, transmitter (Tx) and receiver (Rx) can be trained in an end-to-end fashion under a specific performance metric and channel model [4]. Accordingly, via modeling the Tx and Rx as neural networks (NNs), autoencoders (AEs) have emerged as a useful tool for

end-to-end modeling of the physical layer of communication systems [5]. In particular, such a setup is enabled to optimize Tx and Rx without being limited to conventional component-wise optimization methods, and hence, moving away from carefully optimized sub-blocks to adaptive and flexible NNs [6]. Using offline data-set, AEs can be trained and optimized for a practical communication system, and this architecture can outperform the conventional separable design of the physical layer of such systems.

Owing to these benefits, a number of studies on employing AEs in the physical layer of communication systems has been reported [6–10]. Particularly, the idea of end-to-end learning of communication systems through deep NN-based AEs has been applied to an orthogonal frequency division multiplexing (OFDM)-based communication system in [6]. Moreover, the authors in [7] investigated the problem of joint source and channel coding of structured data (i.e, natural language) over a noisy channel and attained lower word error rates by developing an AE-based system. A new AE-based peak-to-average power ratio reduction scheme has been proposed in [8]. The authors in [9] developed an AE-based deep learning architecture to model a multiuser single-input multiple-output communication system. The work in [10] employs an AE to find proper constellations and corresponding receiver when a radar system coexists with the other interfering wireless systems.

Furthermore, there exist some studies on DL-based wireless transmission systems in which different functionalities of physical layer have been modeled and investigated as a deep NN (DNN). In this regard, ref [11] presents an overview of physical layer DL and the state of the art for fifth-generation of wireless and beyond systems. Meanwhile, the potential of DL approaches to address problems in the physical layer has been shown in several recent studies. More precisely, a novel method for synthesizing new physical layer modulation and coding schemes for communications systems using a learning-based approach is proposed in [12]. The dynamic interference channel in a communication system has been investigated in [13], modulation recognition has been studied in [14], radio fingerprinting has been evaluated in [15], and medium access control mechanisms have been studied in [16]. Furthermore, the authors in [17] investigated mobile edge computing networks for intelligent internet of things (IoT), where multiple users have some computational tasks assisted by multiple computational access points. Accordingly, a system is devised by proposing an intelligent off-loading strategy in which the deep reinforcement learning algorithm is used to automatically learn the optimal offloading strategy. An NN is also trained to predict the offloading action, where the training data is provided by the environment. Also in [18], a DL-based ultra reliable multi-user multiple-input multiple-output (MIMO) detector for 5G enabled IoT is proposed, where the system is operating in interfering environments correlated over the time or frequency domain. To this end, an iterative detection framework including a conventional symbol-by-symbol detector and a deep convolutional neural network (DCNN) is utilized, where the DCNN is used to suppress the interfering signals by capturing their characteristics through deep learning.

Nevertheless, most of these prior works assume an exact mathematical channel model to perform training. More precisely, in an end-to-end communication system, the channel is considered as a layer in the NN. Thus, to backpropagate error during the training phase, the AE needs to know the gradient of the channel transfer function. Moreover, to

capture the maximum end-to-end performance, considered channel model in the training phase must match the actual channel model in the online transmission phase. This imposes a major limitation on employing AE-based approaches to achieve maximum end-to-end performance of a communication system when the actual channel varies over the time. To cope with this problem, prior works have proposed different online training methods based on measured data during online transmission. For instance, the work in [2] considers training the DL-based system using a channel model, and then fine-tuning the Rx with measured data. However, fully capturing of end-to-end performance is not possible since no fine-tuning is done at the Tx side in this approach. Moreover, the authors in [19] approximate the loss function gradient with respect to the Tx parameters and develop an alternating algorithm for end-to-end training without channel model knowledge. This algorithm iterates between two phases: (i) training of the Rx using the true gradient of the loss, and (ii) training of the Tx based on an approximation of the loss function gradient. However, this method takes more samples to converge and is relied on a two phase time-consuming training paradigm over online transmission, thus decreasing the link availability.

To mitigate the need for undergoing complex online training over actual channels as well as to obtain the maximum end-to-end performance of a communication system, we propose a robust adaptive scheme for data transmission over a random channel with no specific mathematical model using an AE bank. For online transmission, we assume a realistic scenario where the instantaneous channel gain is not known to Tx/Rx. Thus, we need to estimate the channel gain at the RX and feed it back to the Tx. Then, based on the actual channel conditions in the online transmission phase, the pair of encoder and decoder that satisfies the system average block error rate (BLER) constraint is selected for data transmission. To increase bandwidth efficiency, as well as to avoid data framing at the Tx, we propose a method to estimate the channel blindly, i.e., without using any pilot symbols. We then compare the proposed blind method for channel estimation with existing methods in terms of average BLER of the system. It is worth mentioning that, in this paper, the term "robust" indicates that, by using the proposed adaptive scheme, the performance of the system will not be affected by channel variations over the time, and hence, the communication system can deliver solid performance during transmission.

Our major contributions can be summarized as follows:

- A robust AE-based transmission scheme consists of n pairs of AE, each of which corresponds to a specific sub-interval of possible values of channel coefficients, is proposed. Considering the instantaneous channel state, the AE that satisfies the best BLER is selected for data transmission in the online transmission phase.
- In a realistic scenario where the instantaneous channel gain is not known to Tx/Rx, a bandwidth-efficient blind channel estimation is proposed which avoids any pilot transmission. Therefore, the proposed scheme does not impose additional overhead that arises in prevalent pilot-based schemes.
- We evaluate performance of the proposed adaptive scheme in terms of the required number of encoders and decoders and also the average power consumption to satisfy a BLER constraint for data transmission. In this regard, we seek to balance an inherent tradeoff between the deployment cost (represented by the required number

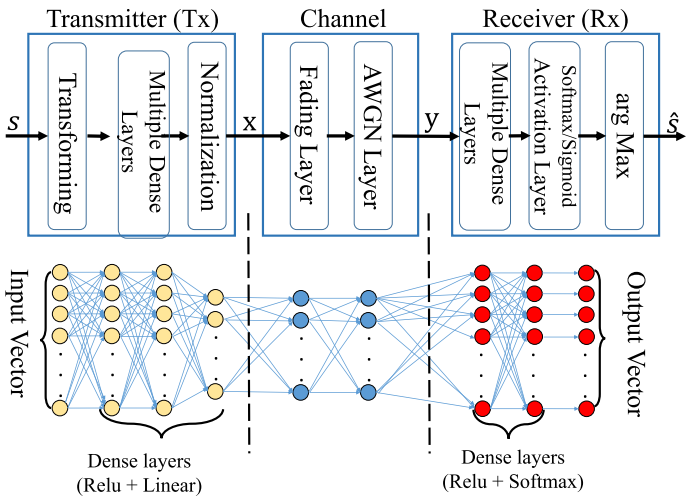

Fig. 1 The layered structure of an end-to-end deep learning-based communications system modeled as an autoencoder

of encoders and decoders), and the system performance (represented by the average power consumption and target BLER).

The rest of this paper is organized as follows. Section 2 briefly presents our method for modeling and evaluating the considered DL-based system. Section 3 describes our system model including the end-to-end communication system and the channel estimation methods. In Sect. 4, we present our adaptive transmission scheme. In Sect. 5, we present the numerical results of the proposed adaptive scheme and system performances. Finally, we conclude the paper in Sect. 6.

2 Methods

The inspiration for an end-to-end deep learning model, also known as, AE , is rooted in the functioning of the proposed method in this study. More precisely, Fig. 1 represents the layered structure of the end-to-end deep learning-based communication system modeled as an AE. Accordingly, without relying on an explicit channel model, we propose an adaptive scheme to increase the reliability of an AE-based communication system over different channel conditions. During the training process, the input symbols, modeled as one-hot vectors, go through the AE where the weights of the neural nodes are initialized with random values. After that, the weight vectors of the nodes will be tuned. The main training goal is to minimize the loss function and maximize the accuracy of the whole process. Using TensorFlow framework [20], the performance of the proposed DL-based method is numerically evaluated. TensorFlow is an end-to-end open source platform for machine learning. It has a comprehensive and flexible ecosystem of tools, libraries, and community resources that lets researchers push the state-of-the-art in learning and provides some facilities for the developers to easily build and deploy ML powered applications as well [20]. Accordingly, one can build and train machine learning models easily using intuitive high-level APIs like Keras with eager execution, which makes for immediate model iteration and easy debugging. As a result, to model and

simulate the proposed design as well as to build and train the deep neural network, we use Keras [21] with TensorFlow in its back-end. Moreover, the set of parameters for simulation is provided in Tables 1 and 2.

3 System model

3.1 End-to-end communication system (autoencoder)

As thoroughly expressed in [2] and depicted in Fig. 1, an AE describes a NN that is trained to reconstruct the input at the output. As the information must pass each layer, the AE will have to find a robust representation of the input message at every layer. Here, we assume a DL-based communication system modeled as an AE. Particularly, the AE includes three main blocks, namely, the Tx, the Rx, and channel, as shown in Fig. 1. The input message $s \in \mathcal{M} = \{1, 2, ..., M\}$ has been received by the Tx. Here, we have $M = 2^b$ where b is the number of bits per message. Before entering to the dense layers, the input message is transformed into a one-hot vector $\mathbf{s}_m$ of dimension M which consists of a single element equal to "1" in position m, whereas all other elements are equal to "0". The one-hot vector is then sent to the hidden dense layers, where at each layer an activation function is applied individually to each element of the input vector. Table 3 shows some commonly used activation functions in the dense layers of an AE. After passing the one-hot vector through the multiple dense layers at the Tx, the transmitted signal $\mathbf{x} = [x[1], \ldots, x[L]]$ is formed for L discrete channel uses. Accordingly, for the considered setup, the data rate is defined as $r = \frac{b}{L}$ (bit/channel use). It is worth noting that, given a certain power constraint and for a feasible rate r, the AE-based system that is trained to minimize a loss function, can automatically build zero-error codes (which is the case in our considered model). However, due to the problems such as vanishing and exploding gradients [22], there may be no guarantee for finding an optimal capacity-achieving code especially in deep NNs. Therefore, the problem of designing capacity-approaching codes is an interesting future research topic in this domain and is beyond the scope of this work.

Furthermore, the Tx last layer normalizes the transmit vectors to guarantee that the average energy per symbol is equal to a predefined value $E_s = \left[\frac{1}{L}\|\mathbf{x}\|_2^2\right]$, where $\|\cdot\|_2$ is the Euclidean norm.

The channel is implemented by including both fading and noise layers whose output $\mathbf{y} = [y[1], \ldots, y[L]]$, i.e., a noisy and distorted version of $\mathbf{x}$, is given by

$$\mathbf{y} = h\mathbf{x} + \mathbf{w}, \tag{1}$$

where h is the channel gain[1] , and $\mathbf{w} \sim \mathcal{N}\left(\mathbf{0}, \sigma_L^2 \mathbf{I}_L\right)$ is a zero-mean additive white Gaussian noise (AWGN) vector with noise variance σ_L^2.

At the Rx side, the received signal $\mathbf{y}$ is passed through multiple dense layers to reach the last layer. Accordingly, at this layer, a softmax activation function is utilized where its output consists of an estimate of the corresponding posterior probability vector $\mathbf{p} \in \mathbb{R}^M$ over all possible messages. The index of the element of $\mathbf{p}$ with the largest value is returned to estimate the transmitted message $\hat{s}$ based on the maximum a posteriori

[1] It is worth mentioning that, the NN architecture is not able to perform complex operation and complex numbers are represented by two real numbers. As a result, all channel gains are real-valued.

(MAP) criterion. Moreover, we utilize the mean square error (MSE) as the loss function of the training process. This way, the loss function is obtained as

$$\mathcal{L}(\mathbf{s}_m, \mathbf{p}) = \|(\mathbf{s}_m - \mathbf{p})\|_2^2. \tag{2}$$

Also, the BLER of considered setup is obtained as

$$P_e = \sum_{s \in \mathcal{M}} \Pr(s \neq \hat{s}) p_s, \tag{3}$$

where p_s is the *a priori* probability of transmission message s. Since the message probability distribution is commonly assumed to be uniform, we have $p_s = \frac{1}{M}$.

3.2 Challenge in training over physical channel

To fully exploit the end-to-end performance of an AE-based communication system, channel model in the training phase must match the actual channel over which the system is supposed to communicate. Nevertheless, for an actual system, the channel is not perfectly known and varies over the time. Therefore, an AE that is trained over a specific realization of the channel may not deliver the expected performance with the change in channel conditions. Thus, AE should be re-trained from scratch for each new channel condition in order to minimize the loss function and BLER which is a time-consuming process that greatly restricts the link availability and reliability. In this paper, we propose to obviate this practical limitation by employing an AE bank consisting of multiple pairs of trained encoder and decoder for different channel conditions. Subsequently, regarding the actual channel state in the online transmission phase, one pair of trained encoder and decoder is selected for data transmission. To this aim, we estimate the channel gain at the Rx and feed it back to the Tx. The details of the proposed adaptive scheme are provided in Sect. 4. To monitor actual channel conditions for adopting the adaptive scheme, and depending on whether pilot symbols are used or not, two methods for estimating the channel gain of the considered AE-based communication system are presented in the sequel.

3.2.1 Channel estimation using pilot symbols

We assume that the pilot symbol s' is transmitted. Therefore, the channel output $\mathbf{y}'$ corresponding to the encoded signal $\mathbf{x}'$ is given by

$$\mathbf{y}' = h\mathbf{x}' + \mathbf{w}. \tag{4}$$

Given $\mathbf{x}'$, one can obtain an estimate of the channel gain, h, by applying the maximum-likelihood (ML) criterion as

$$\begin{aligned} \hat{h}_{\mathrm{ML}} &= \arg\max_h P(\mathbf{y}'|\mathbf{x}'; h) \\ &= \arg\min_h \left|\mathbf{y}' - h\mathbf{x}'\right|^2 = \frac{\langle \mathbf{x}', \mathbf{y}' \rangle}{\langle \mathbf{x}', \mathbf{x}' \rangle}, \end{aligned} \tag{5}$$

where $\langle \cdot, \cdot \rangle$ denotes the dot product. Although estimating the communication channel via pilot symbols results in an accurate estimation, adopting this approach requires data

framing at the Tx side which leads to the loss of data rate and increases the required channel bandwidth. In the sequel, we propose a near optimal method for channel estimation without using pilot symbols.

3.2.2 Blind channel estimation

To increase the bandwidth efficiency by avoiding pilot symbols and, at the same time, to decrease the system complexity by avoiding data framing at the Tx, here, we tend to estimate the channel in a blind way, i.e., without using any pilot symbols. For this purpose, we assume that the encoded symbol is accumulated during an observation window composed of K intervals. As a result, at each interval k, $k \in \{1, \ldots, K\}$, the received signal (or equally the channel output) is obtained as

$$y[k] = hx[k] + w[k]. \tag{6}$$

Squaring both sides of (6) implies that

$$\begin{aligned} y[k]^2 &= (hx[k] + w[k])^2 \\ &= h^2 x[k]^2 + 2hx[k]w[k] + w[k]^2. \end{aligned} \tag{7}$$

Since the term $w[k]^2$ is much smaller than $w[k]$ i.e., $w[k]^2 \ll w[k]$, we can reasonably neglect $w[k]^2$ and approximate (7) as

$$y[k]^2 \simeq h^2 x[k]^2 + 2hx[k]w[k]. \tag{8}$$

By performing summation over k, (8) can be reformulated as

$$\tau_y \simeq h^2 + w'', \tag{9}$$

where $\tau_y = \frac{1}{KE_s} \sum_k y[k]^2$, and w'' is a zero-mean white Gaussian noise with variance $\sigma^2 = \frac{4h^2\sigma_N^2}{KE_s}$. Therefore, τ_y has a Gaussian distribution with a mean equal to h^2 and variance σ^2. By using the ML criterion, a blind estimate of the channel is attained as

$$\begin{aligned} \hat{h}_{BL} &= \arg\min_h \left| \tau_y - h^2 \right|^2 \\ &= \sqrt{\tau_y}. \end{aligned} \tag{10}$$

Moreover, by employing a buffer at the Rx, the communication system can support a real-time decision-feedback channel estimation process.[2]

4 Adaptive transmission scheme

In this section, an adaptive transmission scheme is proposed for the AE-based communication system to increase the system reliability over different channel conditions. Meanwhile, we aim to minimize the number of trained encoder-decoder pairs required for the communication link with the average BLER and transmit power constraints

[2] It is worth noting that, in the online phase, the communication system only needs to estimate the channel blindly and feeds it back to the Tx and Rx. Here, the delay occurs only for accumulating the first K bits to perform the first channel estimation. After that, by using the buffer at the Rx, the estimation updates symbol by symbol and there will be no delay for performing channel estimation.

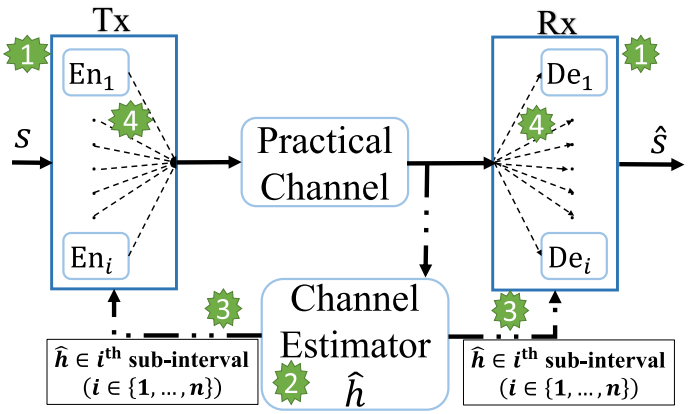

Fig. 2 Four-step structure of the proposed adaptive transmission scheme for the DL-based communication system

under different channel conditions. Figure 2 depicts the four-step structure of our proposed adaptive DL-based transmission scheme. Each step is described as follows.

The first step is offline training. Different pairs of the encoder and decoder in the AE bank are trained offline over different channel conditions. Since the AE-represented communication system should be applicable for any type of channel without a tractable mathematical model, in this paper, we assume there is no channel model information. As a result, the channel fading block acts as a random gain block where its input is multiplied by a random number (i.e., the instantaneous channel gain) to produce the output[3]. We divide the interval of channel gains, $\mathbf{h} = [h_{\min}, h_{\max}]$, into n sub-intervals. Therefore, the channel gain interval is obtained as

$$\mathbf{h} = \underbrace{[[h_{\min}, h_1], (h_1, h_2], \ldots, (h_{\text{n-1}}, h_{\max}]]}_{n \text{ sub-intervals}}. \tag{11}$$

Accordingly, in the training phase, we have n fading blocks for which the gain of each block is randomly selected from its associated sub-interval. Then, AE_i (the ith pair of encoder and decoder for each sub-interval, $i \in \{1, \ldots, n\}$), is trained under the assumption that the channel gain in the channel layer of AE_i lies in $(h_i - 1, h_i]$. After training, the trained encoders and decoders are employed with fixed parameters (i.e, the input and output weights and bias of the neurons remain constant during online transmission). As a result, each trained pair is optimized for a specific channel condition and is used when the practical channel conditions are within the same interval as the one used in the training. Note that, by training the system for each sub-interval, the associated encoder carefully learns a robust representations $\mathbf{x}$ of the different symbol $\mathbf{s}$ regarding to

[3] Although this is indeed an appealing theoretical idea to model the end-to-end system as a whole learning system, its biggest drawback impeding practical implementation is that the gradient of the instantaneous channel transfer function should be known [2]. Generally, there is no tractable mathematical model in a real-world communication system. More precisely, in the context of AE, the actual channel is generally considered as a black box for which only the inputs and outputs can be observed. Here, we just need to perform some simple measurements at different time intervals to determine the range of the channel gain variations.

the possible distortions created by the channel at that interval. Therefore, the whole system is expected to deliver a robust performance over a wide range of channel conditions.

The second step is channel estimation during online transmission. As we mentioned in Sect. 3.2, the Rx can perform channel estimation by employing two methods, i.e., using pilot symbols and blind estimation. For estimating the channel via pilot symbol, the Tx should enclose transmit data in different frames and insert pilot symbols into the frames which results in less bandwidth efficiency. To avoid this, the Rx can also perform blind channel estimation over an observation window as thoroughly discussed in Sect. 3.2.2.

In the third and fourth steps, first, the estimated channel gain at the Rx, $\hat{h}$, is fed back to the Tx. Next, if $\hat{h}$ lies in the ith sub-interval, the ith pair of encoder and decoder is selected for sending and receiving data. We summarize the main steps of the proposed adaptive transmission scheme in Algorithm 1.

Algorithm 1 Robust adaptive DL-based end-to-end transmission using real-time channel estimation.

Input: n sub-intervals of the channel gain, n pairs of encoder and decoder;
Output: An estimate value for the actual communication channel $\hat{h}$, i^{th} pair of encoder and decoder for sending and receiving data;
Step 1 (offline): Train n pairs of encoder and decoder over n channel sub-intervals;
Step 2 (online): Estimate the instantaneous channel gain, $\hat{h}$, at the Rx using (10), and feed it back to the Tx;
Step 3 (online): if $\hat{h}$ lies in the i^{th} sub-interval, choose the i^{th} pair of encoder and decoder for sending and receiving data;
Step 4 (online): Monitor the values of real-time channel estimation and switch to the appropriate pair of encoder and decoder whenever the estimated value has changed from the current sub-interval to another one.

It should be noted that the robust performance of the proposed adaptive scheme comes at the expense of using multiple pairs of encoder and decoder instead of using one pair which increases the deployment cost. Clearly, this price should be paid for being agnostic to the actual channel during training phase. This gives rise to a natural question: what is the minimum number of pairs of encoder and decoder which satisfies a target level of performance for the considered system. In the sequel, we propose an answer to this question by evaluating the performance of the proposed adaptive scheme in terms of average power consumption and deployment cost to fulfill a predetermined average BLER. More precisely, we impose a target average BLER, P_e^t, as a constraint for the system, and minimize the number of encoder and decoder pairs (or equally the number of sub-intervals n) to satisfy the BLER constraint under different channel conditions. Hence, the optimization problem can be formulated as

$$\min \ n \tag{12a}$$

$$\text{s.t.} \quad \bar{P}_e \leq P_e^t. \tag{12b}$$

Note that, in the optimization problem (12), the minimum number of encoder and decoder pairs should be found under different channel conditions and average transmit power. From the system performance's perspective, the more encoders and decoders, the

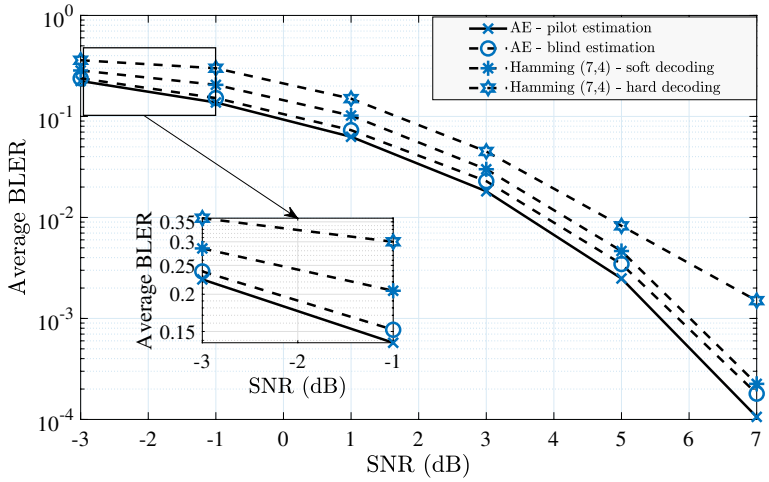

Fig. 3 Average BLER versus SNR when the instantaneous channel gain, $\hat{h}$, is estimated by using two methods, i.e., pilot symbol method, and blind estimation method. The parameters for the AE are set as follows: $M = 16$, $L = 7$, and $n = 5$

Table 1 Layout of autoencoder

	Parameter #	Output shape
Tx		
Input	0	$\{0, 1\}^M$
Dense (Relu)	72	$\mathbb{R}^M$
Dense (Relu)	72	$\mathbb{R}^M$
Dense (linear)	63	$\mathbb{R}^L$
Normalization	14	$\mathbb{R}^L$
Channel		
Random gain	14 (non-trainable)	$\mathbb{R}^L$
AWGN	14 (non-trainable)	$\mathbb{R}^L$
Rx		
Dense (Relu)	64	$\mathbb{R}^M$
Dense (Softmax)	72	$\mathbb{R}^M$

Table 2 Parameters for the autoencoder setup

Parameter	Value
Optimizer	Adam
Loss function	MSE
Epoch	150
Batch size	45
Trained samples	10^5
Test samples	10^6

more robust is the system which is designed for different channel conditions. Therefore,

Table 3 List of activation functions used in the AE layers

Activation functions		
Name	Function	Output range
Linear	s_m	$(-\infty, +\infty)$
Relu	$\max(0, s_m)$	$[0, +\infty)$
Softmax	$\frac{e^{s_m}}{\sum_m s_m}$	(0, 1)
Sigmoid	$\frac{1}{1+e^{-s_m}}$	(0, 1)
tanh	$\tanh(s_m)$	$(-1, 1)$

finding the optimum number of encoders and decoders requires balancing an inherent tradeoff between the cost of system deployment and desirable performance. This tradeoff is thoroughly studied in Sect. 5.

5 Numerical results and discussion

In this section, we evaluate the numerical results of the proposed adaptive transmission scheme in the DL-based communication system. We use Keras [21] with TensorFlow [20] in its back-end in order to build and train our deep NN. For training, we use a variant of stochastic gradient descent (SGD) known as Adam with widely accepted thumb rule for the parameter values as follows, learning rate $\eta = 0.001$, $\beta_1 = 0.9$, and $\beta_2 = 0.99$ [2]. Also, the set of parameters for each pair of encoder and decoder (the AE) is provided in Table 1.

We first compare the proposed blind method for channel estimation with pilot estimation in terms of average BLER. To this aim, we plot the average BLER of the considered system when different methods for channel estimation are employed in Fig. 3. The results of this figure are obtained by assuming $M = 16$ and $L = 7$, thus $r = \frac{4}{7}$. Moreover, in Fig. 3, we assume a communication system employing binary phase-shift keying (BPSK) modulation and a Hamming (7, 4) code with either binary hard-decision decoding or soft decoding against the BLER achieved by the trained AEs as a baseline system for comparison. From the results of this figure, first, we can observe that given the same information transmission rate, $r = \frac{4}{7}$, the performance of the DL-based system is better than that of communication system employing Hamming code. It is worth mentioning that, the DL-based system does not employ any error control coding approach for the noisy channel, and it still outperforms a classical communication system that utilizes error control schemes. Second, for the considered rate in our setup, the proposed blind method for channel estimation achieves an acceptable level of accuracy (SNR gap less than 0.3 dB for $L = 7$ in our setup) compared with the pilot estimation. Thus, proposed blind estimation method can be applied in our considered adaptive scheme.

To investigate the tradeoff of finding the optimum number of encoder and decoder over different channel conditions, we have presented curves of average BLER as a function of transmit power for different number of sub-intervals, n, and for the case of an AE with $M = 16$, and $L = 7$ in Fig. 4. Firstly, as expected, our considered system gives its worst performance in terms of average BLER when one pair of encoder and decoder is used for all channel conditions, i.e., when $n = 1$ (non-adaptive scheme). Indeed, this worst performance highlights the necessity of employing a robust transmission scheme

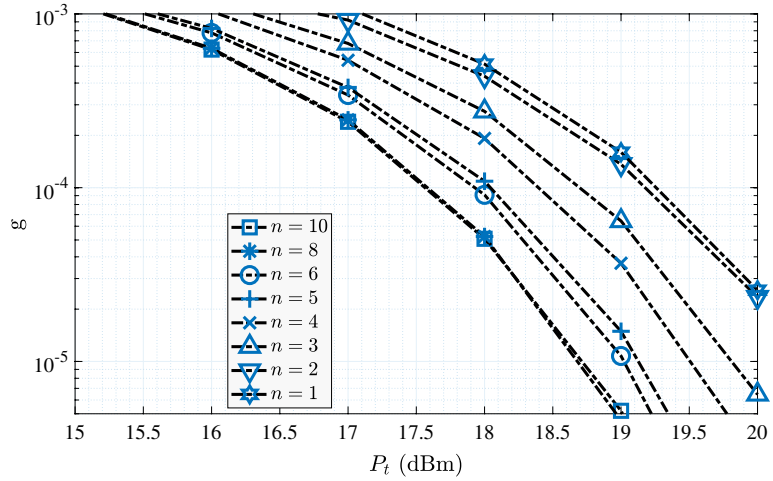

Fig. 4 Average BLER versus p_t for different numbers of sub-intervals n. The parameters for the AE are set as follows: $M = 16, L = 7$

over different channel conditions. Subsequently, by applying our adaptive scheme, one can observe that the performance of the considered system improves by increasing n. On the other hand, from a specific number onwards (e.g., $n = 8$ in our setup), increasing the number of sub-intervals (or equally the number of pairs of encoders and decoders), does not necessarily improve the system performance. This can be justified by the fact that, in this case, the n pairs of encoder and decoder are trained over closed sub-intervals (i.e., nearly the same channel conditions) as the number of these sub-intervals increases. Hence, when the sub-interval is short, those pairs of encoder and decoder trained over close sub-intervals deliver the same performance when they are used in actual channel conditions. Hence, the improvement of performance is negligible in this situation. Finally, as it is depicted in Fig. 4, the minimum value of average transmit power, as well as the minimum number of pairs of encoder and decoder can be obtained to satisfy the average BLER constraint in Eq. (12). For instance, as we can observe from Fig. 4, for a target average BLER equal to 10^{-4}, our proposed algorithm with $n = 5$ can achieve a performance gain over 1.2 dB in terms of average power consumption compared with a non-adaptive scheme.

Finally, to better evaluate the performance and effectiveness of the proposed scheme, we have carried out another evaluation by contrasting the performance of our proposed scheme with the proposed scheme in [2], referred as fine tuning,[4] under a practical optical wireless scenario. More precisely, we consider a free space optical channel under moderate atmospheric turbulence regime, which has been experimentally proven to have a log-normal distribution [23], and compare the performance of our proposed scheme with the fine tuning method. We note that, since optical wireless channel is quasi-static [24], the channel can be estimated with good accuracy and fed back to the transmitter, thus making it an appealing practical case for employing our proposed AE-based communication systems. From the result of Fig. 5, one can readily observe

[4] In this approach the encoder is trained again in the online phase.

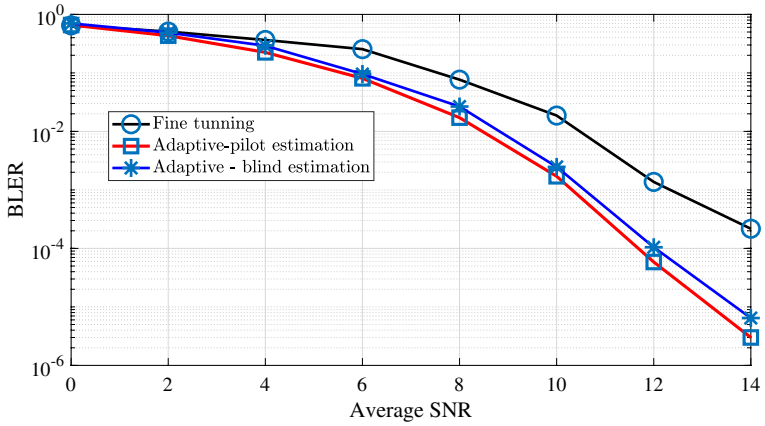

Fig. 5 Performance of the proposed AE-based systems versus fine tuning approach proposed in [2] over the lognormal fading channel

that our proposed scheme outperforms the fine tuning approach. The reason for the enhancement is that, compared to [2] which uses an unchangeable encoding strategy over different channel conditions (or equally, it just fine tunes the receiver, resulting in sub-optimal performance), our proposed scheme is able to use different encoder to generate a robust representation of transmitted massages for different channel conditions (messages that will be least affected by random channel variations). Hence, the effect of channel uncertainties on the performance of the learning-based systems is decreased, and the proposed scheme outperforms the fine tuning method with more than 2dB margin in the power consumption.

6 Conclusion

In this paper, we proposed an adaptive scheme to increase the reliability of an AE-based communication system over different channel conditions. Accordingly, we divided the interval of random channel gains into n sub-intervals and assigned n pairs of encoder and decoder to each interval. The encoders and decoders are trained offline, and, regarding the actual channel state in the online transmission phase, one pair of trained encoder and decoder is selected for data transmission. To this aim, we estimated the channel gain at the Rx and fed it back to the Tx without using any pilot symbols. We showed that, compared with a non-adaptive scheme, by using the proposed adaptive method, the DL-based system can deliver a robust performance in terms of average BLER over different channel conditions. Also, it is observed that our proposed adaptive scheme can achieve a performance gain in terms of average power consumption to achieve the same average BLER as existing methods.

Abbreviations

AE: Autoencoder; AWGN: Additive white Gaussian noise; BLER: Block error rate; BPSK: Binary phase shift keying; DL: Deep learning; MAP: Maximum a posteriori; ML: Machine learning; NN: Neural networks; Rx: Receiver; SGD: Stochastic gradient descent; Tx: Transmitter.

Author's contributions
Hossein Safi, and Mohammad Akbari are the main authors of the current paper. They contributed to the development of the ideas, design of the study, theory, result analysis, and paper writing. Elaheh Vaezpour, Saeedeh Parsaeefard, and Raed M Shubair contributed to the paper revision. All authors read and approved the final manuscript.

Funding
Not applicable.

Availability of data and materials
Not applicable.

Declaration

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Department of Electrical Engineering, Shahid Beheshti University, Velenjak, Tehran, Iran. [2] Department of Communication Technology, ICT Research Institute (ITRC), North Karegar, Tehran, Iran. [3] Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada. [4] Massachusetts Institute of Technology (MIT), Massachusetts, CA 02139, USA. [5] New York University (NYU) Abu Dhabi, Abu Dhabi 129188, UAE.

Received: 10 September 2020 Accepted: 12 February 2021
Published online: 06 March 2021

References

1. A. Goldsmith, *Wireless Communications* (Cambridge University Press, Cambridge, 2005)
2. S. Dörner, S. Cammerer, J. Hoydis, S. ten Brink, Deep learning based communication over the air. IEEE J. Sel. Topics Signal Process. **12**(1), 132–143 (2017)
3. T. O'Shea, J. Hoydis, An introduction to deep learning for the physical layer. IEEE Trans. Cognit. Commun. Netw. **3**(4), 563–575 (2017)
4. Z. Qin, H. Ye, G.Y. Li, B.-H.F. Juang, Deep learning in physical layer communications. IEEE Wirel. Commun. **26**(2), 93–99 (2019)
5. T.J. O'Shea, K. Karra, T.C. Clancy, Learning to communicate: channel auto-encoders, domain specific regularizers, and attention, in *Proceedings of 16th IEEE ISSPIT*, Limassol, Cyprus (2016)
6. A. Felix, S. Cammerer, S. Dörner, J. Hoydis, S. Ten Brink, OFDM-autoencoder for end-to-end learning of communications systems, in *Proceedings of 19th IEEE SPAWC*, Kalamata, Greece (2018)
7. N. Farsad, M. Rao, A. Goldsmith, Deep learning for joint source-channel coding of text, in *Proceedings of IEEE ICASSP*, Calgary, Alberta, Canada (2018)
8. M. Kim, W. Lee, D.-H. Cho, A novel PAPR reduction scheme for OFDM system based on deep learning. IEEE Commun. Lett. **22**(3), 510–513 (2017)
9. S. Xue, Y. Ma, N. Yi, R. Tafazolli, Unsupervised deep learning for MU-SIMO joint transmitter and noncoherent receiver design. IEEE Wirel. Commun. Lett. **8**(1), 177–180 (2018)
10. F. Alberge, Deep learning constellation design for the AWGN channel with additive radar interference. IEEE Trans. Commun. **67**(2), 1413–1423 (2018)
11. F. Restuccia, T. Melodia, Deep learning at the physical layer: system challenges and applications to 5G and beyond. IEEE Commun. Mag. **58**(10), 58–64 (2020)
12. T.J. O'Shea, T. Roy, N. West, B.C. Hilburn, Physical layer communications system design over-the-air using adversarial networks, in *Proceedings of 26th EUSIPCO Conference* (2018)
13. F. Qamar, M.N. Hindia, K. Dimyati, K.A. Noordin, I.S. Amiri, Interference management issues for the future 5g network: a review. Telecommun. Syst. **71**(4), 627–643 (2019)
14. T.J. O'Shea, T. Roy, T.C. Clancy, Over-the-air deep learning based radio signal classification. IEEE J. Sel. Top. Signal Process. **12**(1), 168–179 (2018)
15. F. Restuccia, et al.: Deepradioid: real-time channel-resilient optimization of deep learning-based radio fingerprinting algorithms, in *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 51–60 (2019)
16. O. Naparstek, K. Cohen, Deep multi-user reinforcement learning for distributed dynamic spectrum access. IEEE Trans. Wirel. Commun. **18**(1), 310–323 (2019)
17. R. Zhao, X. Wang, J. Xia, L. Fan, Deep reinforcement learning based mobile edge computing for intelligent internet of things. Phys. Commun. **43**(1), 101148–101156 (2020)
18. K. He, Z. Wang, D. Li, F. Zhu, L. Fan, Ultra-reliable MU-MIMO detector based on deep learning for 5G/B5G-enabled IoT. Phys. Commun. **43**(1), 101181–101190 (2020)
19. F.A. Aoudia, J. Hoydis, End-to-end learning of communications systems without a channel model, in *Proceedings of 52nd IEEE ACSSC*, Pacific Grove, CA, USA (2018)
20. M. Abadi, et al., TensorFlow: large-scale machine learning on heterogeneous systems. Software available from tensorflow.org (2015)
21. F. Chollet, et al., Keras. https://keras.io (2015)
22. Y. Bengio, P. Simard, P. Frasconi, Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Netw. **5**(2), 157–166 (1994)
23. L.C. Andrews, R.L. Phillips, *Laser Beam Propagation Through Random Media*, SPIE (2005)

24. H. Safi, A.A. Sharifi, M.T. Dabiri, I.S. Ansari, J. Cheng, Adaptive channel coding and power control for practical FSO communication systems under channel estimation error. IEEE Trans. Veh. Technol. **68**(8), 7566–7577 (2019)

Publisher's note