A Novel PAPR Reduction Scheme for OFDM System Based on Deep Learning

Minhoe Kim, Woongsup Lee, *Member, IEEE*, and Dong-Ho Cho, *Senior member, IEEE*

***Abstract*—High peak-to-average power ratio (PAPR) has been one of the major drawbacks of orthogonal frequency division multiplexing (OFDM) systems. In this letter, we propose a novel PAPR reduction scheme, known as PAPR reducing network (PRNet), based on the autoencoder architecture of deep learning. In the PRNet, the constellation mapping and demapping of symbols on each subcarrier is determined adaptively through a deep learning technique, such that both the bit error rate (BER) and the PAPR of the OFDM system are jointly minimized. We used simulations to show that the proposed scheme outperforms conventional schemes in terms of BER and PAPR.**

***Index Terms*—Orthogonal frequency division multiplexing, autoencoder, deep learning, peak-to-average power ratio.**

I. Introduction

ORTHOGONAL frequency division multiplexing (OFDM) has been used in many wireless communication systems including the IEEE 802.11, Long-Term Evolution (LTE) and LTE Advanced systems, due to its robustness in multipath propagation environments and the relatively low complexity transceiver design. However, it suffers from high peak-to-average power ratio (PAPR) which possibly results in the distortion of transmitted signal due to the nonlinearity of power amplifiers. Consequently, the reduction of PAPR of OFDM system has been the subject of extensive investigation by many researchers [1]–[4].

One of most well known schemes for reducing the PAPR in an OFDM system is the clipping scheme [1] in which the original signal is distorted deliberately. Although the clipping scheme can reduce the PAPR of an OFDM system, the bit error rate (BER) of the system is degraded due to the distortion of the signal, which is the drawback of the clipping scheme. Other notable schemes for PAPR reduction are signal scrambling techniques, such as partial transmission sequence (PTS) and selective mapping (SLM) [2], [3]. However, in the signal scrambling schemes, the phase factor to generate a scrambled signal is needed at the receiver in order to reconstruct the transmitted information properly. If this side information is inaccurate, the performance is greatly degraded. Moreover, the signal scrambling scheme requires huge computational complexity to find the phase factor that generates an optimal scrambled sequence. In addition, a nonlinear companding transform scheme for reducing PAPR was proposed in [4]; however, the BER of this system was degraded due to the signal distortion, just as it is in the clipping scheme.

In recent years, deep learning which is based on an architecture of multiple layers of perceptron (MLP) connected between input and output layers in a way that mimics the neurons in a human brain, has provided a significant improvement in performance compared with conventional hand-crafted schemes in many fields. Far from being merely a tool for cognitive tasks such as image recognition, deep learning shows potential for applications in communication systems, offering an improvement over traditional schemes. For instance, various deep learning techniques were shown to have better performances than conventional schemes in the tasks of channel encoding and decoding [5]. Moreover, radio modulation recognition using convolution neural networks was investigated in [6]. Accordingly, the deep learning technique can be adopted to solve the problem of high PAPR in OFDM systems. Sohn [7] and Sohn and Kim [8], considered an artificial neural network (ANN) to reduce the complexity in solving PAPR reduction problem; however, the main focus of these studies was to reduce the complexity of active constellation extension (ACE) scheme which is based on clipping, thus they cannot outperform conventional schemes.

Herein, we investigate the reduction of PAPR by utilizing the specific type of DNN, i.e., an autoencoder, which is usually used for denoising corrupted data [9]. In our proposed scheme, constellation mapping and demapping of symbols[1] on each subcarrier in an OFDM system are learned through the training of DNN such that the transmit signal sequence has a low PAPR, while minimizing the degradation of the BER. Moreover, once trained, our proposed scheme can be performed with negligible overhead such that the real-time operation is possible. The main contributions of this paper are summarized in two folds as follows.

- We propose a novel PAPR reduction scheme, namely PRNet, in an OFDM system based on deep autoencoder architecture. In our proposed scheme, constellation mapping and demapping of symbols on each subcarrier in the OFDM system are found through a deep learning technique. To the best of our knowledge, this is the first attempt to apply a deep autoencoder for the reduction of PAPR in an OFDM system.
- We evaluate the performance of the proposed scheme using computer simulations. We confirm that PRNet significantly reduces the PAPR of an OFDM system while retaining the BER.

[1]Note that the constellation shaping problem to minimize PAPR is a multivariate non-convex problem that is NP-hard [10].

Manuscript received November 22, 2017; revised December 19, 2017; accepted December 20, 2017. Date of publication December 27, 2017; date of current version March 8, 2018. This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No. 2014-0-00282, Development of 5G Mobile Communication Technologies for Hyper-connected smart services). The associate editor coordinating the review of this paper and approving it for publication was B. Smida. *(Corresponding author: Woongsup Lee.)*

M. Kim and D.-H. Co are with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea (e-mail: kimminhoe@kaist.ac.kr; dhcho@kaist.ac.kr).

W. Lee is with the Department of Information and Communication Engineering, Institute of Marine Industry, Gyeongsang National University, Tongyeong 53064, South Korea (e-mail: seotaijiya@gmail.com).

Digital Object Identifier 10.1109/LCOMM.2017.2787646

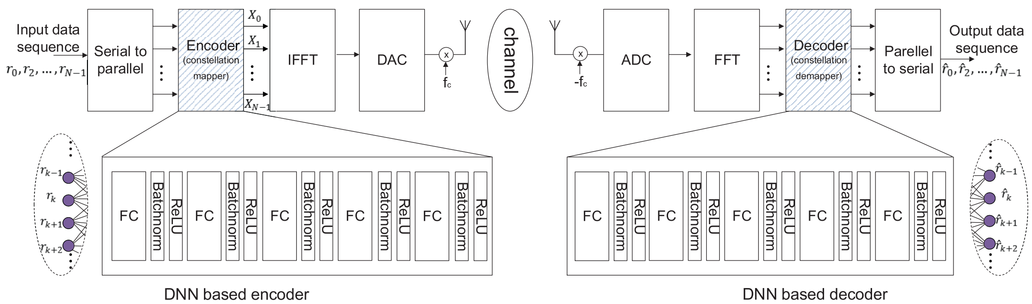

Fig. 1. Structure of proposed PRNet with DNN encoder and DNN decoder.

The remainder of this paper is organized as follows. In Section II, we describe the system model and the principles behind the autoencoder. Our proposed scheme is then introduced in Section III. Its performance is assessed in Section IV, and in Section V we provide our conclusions.

II. System Model and Autoencoder

A. System Model

We consider an OFDM system in which bandwidth is divided into N orthogonal subcarriers where an encoded symbol is transmitted on each subcarrier. We assume an independent and identically distributed (i.i.d.) input data sequence which consists of symbol r_k that refers to M-bit data, where k is the index of the symbol and $\boldsymbol{r}$ is the vector of r_k. In our system model, the input data sequence is mapped into I-Q constellations, using a constellation mapper that we denote as an encoder, $f(\boldsymbol{r})$. The encoded symbol of the k-th subcarrier, X_k, can be described as a symbol mapped from the k-th information symbol, i.e., $f(r_k) = X_k$. Then, $x[n]$, which is time domain signal that becomes the time domain input for the signal amplifier at the RF chain, can be obtained by feeding X_k into an inverse discrete Fourier transform as follows:

$$x[n] = \sum_{k=0}^{N-1} X_k e^{j2\pi nk/N}, \tag{1}$$

where $0 \le n \le N-1$.

Then, the PAPR of the OFDM sequence $x[n]$, which we denote as $PAPR\{x[n]\}$, can be defined as follows:

$$PAPR\{x[n]\} = \frac{\max\limits_{0\le n\le N-1} x[n]^2}{\mathbb{E}\left[|x[n]|^2\right]}. \tag{2}$$

Here, the denominator is the average power of $x[n]$ over $0 \le n \le N-1$.

It is worth noting that in an OFDM system without PAPR reduction, the distribution of the real part of the time domain signal, $\mathbb{R}\{x[n]\}$, and that of the imaginary part, $\mathbb{I}\{x[n]\}$, become Gaussian distribution as N is sufficiently large according to the central limit theorem, which results in a high PAPR [11].

B. Autoencoder

In this paper, we have used an autoencoder of deep learning, which is one of the most well-known generative models in deep learning [12]. Let $\boldsymbol{x}$, $f(\boldsymbol{x})$, and $g(\boldsymbol{x})$ be the input, encoder, and decoder of the autoencoder, respectively. Then, the autoencoder is trained to generate output $g(f(\boldsymbol{x}))$ such that the loss function, $\mathcal{L}(\boldsymbol{x}, g(f(\boldsymbol{x})))$ is minimized. For example, in a denoising autoencoder (DAE), the reconstruction of the original input, $\boldsymbol{x}$, from a corrupted copy of input, $\tilde{\boldsymbol{x}} = \boldsymbol{x} + \epsilon$, where ϵ is random noise, is the main objective [12]. Then, the loss function becomes $||\boldsymbol{x} - g(f(\tilde{\boldsymbol{x}}))||_2$ where $||\cdot||_2$ denotes the L2 norm. Note that in this case, the autoencoder is trained to have the capability of reconstructing the original signal $\boldsymbol{x}$ from the corrupted one. For the training of the autoencoder, the stochastic gradient descent (SGD) optimization method can be used in which the weights and biases of the autoencoder are updated as follows:

$$\theta^{+} := \theta - \alpha \nabla \mathcal{L}(\boldsymbol{x}, \hat{\boldsymbol{x}}; \theta). \tag{3}$$

Here, α is the learning rate that determines the step size of the update value and θ denotes parameters involved in building the DNN layers, which will be explained in the next section.

III. Proposed Scheme

In this section, we first describe the structure and operation of the proposed PRNet. We then discuss how PRNet can be trained to achieve our goal, i.e., reduction of PAPR without sacrificing BER. The considered PRNet is depicted in Fig. 1.

A. Structure and Operation of Proposed PRNet

A simple encoder and decoder structure is considered for PRNet. More specifically, both the encoder and the decoder are composed of the identical sub-blocks (hidden layers) which are connected in tandem, cf. Fig. 1. Each sub-block is composed of fully connected layer (FC), batch normalization (Batchnorm), and rectifier linear unit (ReLU), which are also connected together.

In the FC, the matrix multiplication of the weights is applied, as the addition of biases is applied. Let x_{FC} be the input of the m-th FC, then the output of the FC, $\boldsymbol{y}_{FC}$, becomes $\boldsymbol{y}_{FC} = \boldsymbol{W}_m x_{\text{FC}} + \boldsymbol{b}_m$, where $\boldsymbol{W}_m$ and $\boldsymbol{b}_m$ are the weights and bias for the m-th FC, respectively. In addition, we assume that each FC contains 2048 hidden nodes.

The output of FC is fed into the Batchnorm unit, which normalizes the input of the ReLU unit such that the PRNet can be trained more efficiently [13]. The Batchnorm can be mathematically expressed as $|\boldsymbol{y}_{FC}|_{norm} = \gamma \frac{\boldsymbol{y}_{FC} - \mathbb{E}[\boldsymbol{y}_{FC}]}{\sqrt{Var[\boldsymbol{y}_{FC}] + \nu}} + \beta$, where γ and β are a scaling and shift factor, respectively. Moreover, $\nu = 0.001$ is a constant which prevents the division by zero. It should be noted that both γ and β are also learnable parameters whose values are found through training.

Finally, the normalized value is fed into the activation function $\phi(\cdot)$, which provides nonlinearity to the PRNet [14]. In our proposed scheme, the ReLU is used as the activation function for PRNet such that the output becomes the same as the input if the input is greater than 0, and the output becomes 0, otherwise, i.e., the output of m-th ReLU becomes $\max(|\boldsymbol{y}_{FC}|_{\text{norm}},\ 0)$.

In the PRNet, the input data is encoded to constellation plane using the encoder of PRNet, which is composed of $L_f = 5$ sub-blocks. Accordingly, the output of encoder can be expressed mathematically as $f(\boldsymbol{r}) = \phi_{L_f}(|\boldsymbol{W}^f_{L_f}\phi_{L_f-1}(\cdots\phi_1(|\boldsymbol{W}^f_1\boldsymbol{r}+\boldsymbol{b}^f_1|_{norm})\cdots)+\boldsymbol{b}^f_{L_f}|_{norm})$, where $\boldsymbol{W}^f_{l_f}$ and $\boldsymbol{b}^f_{l_f}$ are the weights and bias for the l_f-th FC of the encoder, respectively. Then the encoded symbols pass through an IFFT operation, which generates the transmit signal. After the IFFT, the signal is transmitted through a wireless channel before arriving at the receiver. Finally, the received signal passes through an FFT operation and is decoded using the decoder of the PRNet. We assume that the decoder is composed of $L_g = 5$ sub-blocks like the encoder such that the output of the decoder, $g(\boldsymbol{y})$, can be expressed as $g(\boldsymbol{y}) = \phi_{L_g}(|\boldsymbol{W}^g_{L_g}\phi_{L_g-1}(\cdots\phi_1(|\boldsymbol{W}^g_1\boldsymbol{y}+\boldsymbol{b}^g_1|_{norm})\cdots)+\boldsymbol{b}^g_{L_g}|_{norm})$ where $\boldsymbol{y}$ denotes the input of decoder and $\boldsymbol{W}^g_{l_g}$ and $\boldsymbol{b}^g_{l_g}$ are the weights and bias for the l_g-th FC of the decoder, respectively.

Accordingly, the reconstructed symbol at the receiver, $\hat{\boldsymbol{r}}$, can be written as follows:

$$\hat{\boldsymbol{r}} = g \circ \text{FFT} \circ \mathbb{H} \circ \text{IFFT} \circ f(\boldsymbol{r}), \tag{4}$$

where $\mathbb{H}$ denotes the effects of the wireless channel such as multipath fading and thermal noise.

B. Training of PRNet

In our proposed scheme, the network is trained to reduce PAPR while preventing the deterioration in the BER. Accordingly, two distinct objectives must be taken into account. First, the PRNet needs to be able to reconstruct the transmitted signals from the received signals such that the BER of the system does not deteriorate. Second, the PRNet must also generate a transmission signal that shows a low PAPR. In the following, we explain the appropriate loss function of the PRNet to achieve the former and latter goal and also provide a solution for achieving both goals jointly.

In order to achieve the former objective, the encoder of the PRNet is trained to find the proper constellation mapping from the input data, r_k, to the output, X_k, and the decoder of the PRNet must be able to decode the received signal. Then, the proper loss function to achieve this objective can be written as follows[2]:

$$\mathcal{L}_1(\boldsymbol{r},\hat{\boldsymbol{r}}) = ||\boldsymbol{r} - g(FFT(\mathbb{H} \circ IFFT(f(\boldsymbol{r};\theta_f)) + \epsilon);\theta_g)||_2, \tag{5}$$

where $f(\cdot;\theta_f)$ and $g(\cdot;\theta_g)$ are the parametric representation of the encoder and decoder, respectively, and ϵ is the noise at the receiver. We denote this as a parametric representation because the weight matrix, biases, and activation layers can be represented by simple matrix operations of hidden node parameters as explained in the previous subsection, i.e. $\theta = \{\boldsymbol{W}, \boldsymbol{b}\}$. Through the training, θ_f and θ_g, i.e., the weight and bias of the autoencoder, which minimize the loss function, are found such that the constellation that is robust to a random channel $\mathbb{H}$ is attained by the encoder, and an efficient way of decoding this constellation mapping is achieved by the decoder. On the other hand, the following loss function, $\mathcal{L}_2(\boldsymbol{r})$, can be used to achieve the latter objective, i.e., decrease of PAPR,

$$\mathcal{L}_2(\boldsymbol{r}) = PAPR\{x[n]\} = PAPR\{IFFT(f(\boldsymbol{r};\theta_f))\}. \tag{6}$$

In our proposed scheme, the training is divided into two stages. In the first stage of training, the proper corruption level, $\eta = \frac{\mathbb{E}[|\epsilon|^2]}{\mathbb{E}[|r|^2]}$ which denotes the ratio of noise power and signal power, is determined by using the loss function $\mathcal{L}_1$. Then, in the second stage of training, the weights and biases of the autoencoder (θ_f and θ_g) are learned by considering the joint loss function, $\mathcal{L}(\boldsymbol{r},\hat{\boldsymbol{r}})$, which combines $\mathcal{L}_1$ and $\mathcal{L}_2$ such that both PAPR and BER, can be minimized. $\mathcal{L}(\boldsymbol{r},\hat{\boldsymbol{r}})$ can be expressed as follows:

$$\mathcal{L}(\boldsymbol{r},\hat{\boldsymbol{r}}) = \mathcal{L}_1(\boldsymbol{r},\hat{\boldsymbol{r}}) + \lambda\mathcal{L}_2(\boldsymbol{r}). \tag{7}$$

Here, λ denotes the weight parameter that determines which loss, i.e., $\mathcal{L}_1$ or $\mathcal{L}_2$, is dominant. For example, when the value of λ is small, the autoencoder is trained to improve the BER but to put less effort into reducing the PAPR.

IV. Performance Evaluation

In this section, we evaluate the performance of our proposed PRNet scheme. In particular, the BER and complementary cumulative distribution function (CCDF) of the PAPR, defined as $\text{Prob}\{PAPR > PAPR_0\}$, are examined. Moreover, the characteristics of the PRNet are also investigated. Our simulation codes can be found at [15].

We consider an OFDM system with 64 subcarriers and 4-QAM modulation.[3] We also assume the Rayleigh fading channel for a wireless channel, i.e., $\mathbb{H}$. In the training of PRNet, total 500,000 batches are used for training and the batch size that denotes the number of training samples used for updating a single step in the SGD algorithm, is set to 400, and the weight parameter, λ, is set to 0.01. For comparison, we consider two well known conventional PAPR reduction schemes, which are the PTS algorithm and the clipping algorithm, where 4 sub-blocks with 4 possible phase factors are taken into account for the PTS algorithm. It should be noted that although the PRNet is based on deep learning which is known to have high computational burden, the computational overhead of the PRNet is rather small with complexity of $O(L \cdot K \cdot M)$ where L is the number of hidden layers; it has an execution time of 480μs with parallel computation and 2304μs without parallel computation, which is lower or similar compared to that of the PTS scheme or clipping schemes, whose execution times are 2890μs, and 1415μs, respectively. For the training

[2] In our proposed scheme, we have used mean squared error (MSE) for the loss function, which is widely used for the autoencoder structure because our goal is to regenerate $\boldsymbol{r}$ from the noisy sample. The cross entropy which is also widely taken into account in DNN design is inappropriate in our scheme because the considered problem is not classification problem and the label data is unavailable.

[3] Although we have considered an OFDM system with a small number of subcarriers due to the problem of number of required training samples, the considered OFDM system is plausible because some wireless systems, e.g., the IEEE 802.11 a-g systems, use the small OFDM size.

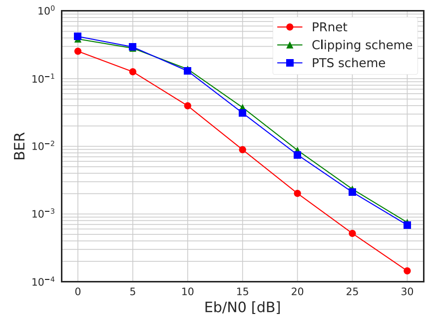

Fig. 2. BER vs. SNR of conventional and proposed schemes.

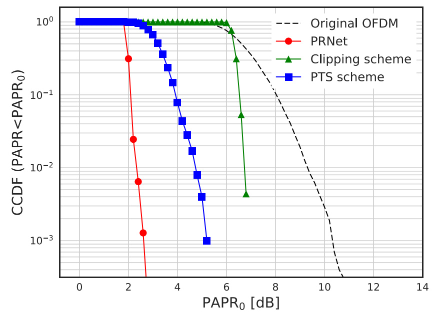

Fig. 3. CCDF of PAPR for conventional and proposed schemes.

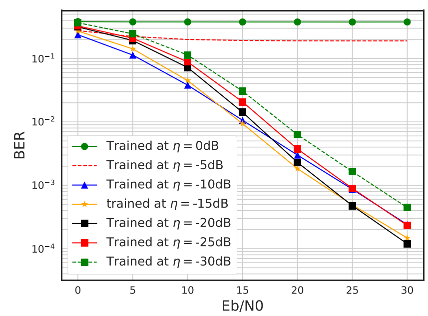

Fig. 4. BER of PRNet at different corruption levels.

stage, the average power of the noise, $\mathbb{E}[|\epsilon|^2]$, is fixed.[4] Finally, we used a PRNet trained at $\eta = -15$ dB for the performance evaluation.

In Fig. 2, the BER of the PRNet is compared with that of conventional schemes. The results indicate that our proposed scheme has a lower BER than conventional PAPR reduction schemes over the whole range of values shown.

The CCDF of the PAPR of conventional schemes and of the PRNet are shown in Fig. 3 which are evaluated with 50,000 random OFDM input data sequences. From these results, we find that the PAPR of the PRNet is much lower than that of conventional schemes. We also conclude that our proposed scheme outperforms conventional PAPR reduction schemes in terms of both BER and PAPR.

In Fig. 4, the BER of the PRNet at different corruption levels, η, is shown. As seen from these results, training the PRNet at $\eta = -15$ dB provides lower BER than other cases

[4] As can be seen from the loss function used in our PRNet, the performance of our proposed scheme depends on the amount of noise applied to corrupt the original signal in training. If the amount of noise is too small, the decoder learns only the small region near the original symbol location in the constellation plane, whereas if the amount of noise is too large, the decoder is confused by the overlapping region of the received signal distribution in the constellation plane. In fact, the appropriate level of noise depends on the modulation level, because the average distance between symbols in a constellation plane depends on the modulation level.

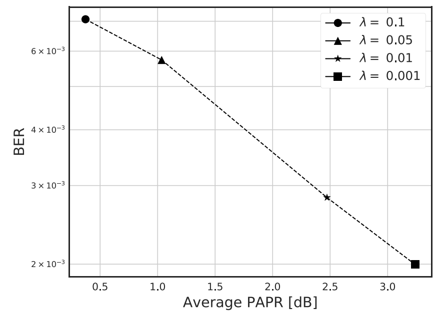

Fig. 5. BER vs. PAPR of PRNet by varying λ when SNR is 20dB.

over the whole range of SNR, which justifies our choice of corruption level.

Finally, the average BER and PAPR of the PRNet obtained by varying λ when the SNR is 20 dB, are shown in Fig. 5. As expected, the values of BER and PAPR vary according to λ, such that the performance of the PRNet can be adjusted according to the importance of the two metrics.

V. Conclusions and Future Works

In this paper, a novel deep learning-based PAPR reduction scheme for an OFDM system, namely PRNet, was proposed. Using the autoencoder of deep learning, our proposed scheme is capable of reducing the PAPR while maintaining BER. We showed that our proposed scheme significantly outperforms conventional schemes in view of both PAPR and BER.

References

[1] H. Ochiai and H. Imai, "Performance analysis of deliberately clipped OFDM signals," *IEEE Trans. Commun.*, vol. 50, no. 1, pp. 89–101, Jan. 2002.

[2] S. H. Müller and J. B. Huber, "OFDM with reduced peak-to-average power ratio by optimum combination of partial transmit sequences," *Electron. Lett.*, vol. 33, no. 5, pp. 368–369, Feb. 1997.

[3] R. J. Baxley and G. T. Zhou, "Comparing selected mapping and partial transmit sequence for PAR reduction," *IEEE Trans. Broadcast.*, vol. 53, no. 4, pp. 797–803, Dec. 2007.

[4] T. Jiang *et al.*, "Exponential companding technique for PAPR reduction in OFDM systems," *IEEE Trans. Broadcast.*, vol. 51, no. 2, pp. 244–248, Jun. 2005.

[5] T. J. O'Shea *et al.*, "Learning to communicate: Channel auto-encoders, domain specific regularizers, and attention," in *Proc. ISSPIT*, Limassol, Cyprus, Dec. 2016, pp. 223–228.

[6] T. J. O'Shea *et al.*, "Convolutional radio modulation recognition networks," in *Proc. EANN*, Aberdeen, U.K., Sep. 2016, pp. 213–226.

[7] I. Sohn, "A low complexity PAPR reduction scheme for OFDM systems via neural networks," *IEEE Commun. Lett.*, vol. 18, no. 2, pp. 225–228, Feb. 2014.

[8] I. Sohn and S. C. Kim, "Neural network based simplified clipping and filtering technique for PAPR reduction of OFDM signals," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1438–1441, Aug. 2015.

[9] P. Vincent *et al.*, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, Dec. 2010.

[10] M. C. P. Paredes *et al.*, "PAPR reduction via constellation extension in OFDM systems using generalized benders decomposition and branch-and-bound techniques," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5133–5145, Jul. 2016.

[11] T. Jiang and Y. Wu, "An overview: Peak-to-average power ratio reduction techniques for OFDM signals," *IEEE Trans. Broadcast.*, vol. 54, no. 2, pp. 257–268, Jun. 2008.

[12] Y. Bengio *et al.*, "Generalized denoising auto-encoders as generative models," in *Proc. NIPS*, Lake Tahoe, NV, USA, Dec. 2013, pp. 899–907.

[13] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. ICML*, Lille, France, Jul. 2015, pp. 448–456.

[14] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. ICML*, Haifa, Israel, Jun. 2010, pp. 807–814.

[15] *PRnet Github Code*. Accessed: Dec. 2017. [Online]. Available: http://github.com/gomman/ofdm