Low PAPR Waveform Design for OFDM Systems Based on Convolutional Autoencoder

Yara Huleihel
Ben Gurion University
Beer Sheva, Israel
halihal@post.bgu.ac.il

Eilam Ben-Dror
Tel-Aviv Research Center
Huawei Technologies Co. Ltd.
eilam.ben.dror@huawei.com

Haim H. Permuter
Ben Gurion University
Beer Sheva, Israel
haimp@bgu.ac.il

***Abstract*—This paper introduces the architecture of a convolutional autoencoder (CAE) for the task of peak-to-average power ratio (PAPR) reduction and waveform design, for orthogonal frequency division multiplexing (OFDM) systems. The proposed architecture integrates a PAPR reduction block and a non-linear high power amplifier (HPA) model. We apply gradual loss learning for multi-objective optimization. We analyse the model's performance by examining the bit error rate (BER), the PAPR and the spectral response, and comparing them with common PAPR reduction algorithms.**

***Index Terms*—Autoencoder, convolutional neural network, deep learning, OFDM, PAPR, wireless signal processing.**

I. Introduction

Orthogonal frequency division multiplexing (OFDM) has been adopted as a standard technology in various wireless communication systems, such as WiFi, 4G and 5G standards for wireless communications. Nonetheless, a major drawback of the OFDM multi-carrier system is its tendency to produce signals with high peak-to-average power ratio (PAPR) in the time-domain, since many subcarrier components are added via a fast Fourier transform (FFT) operation. The contribution of each subcarrier to the total power is dynamic, which makes the total power highly variant. These high PAPR signals pass through a non-linear high power amplifier (HPA), which is a major power-consuming analog component, resulting in severe nonlinear signal distortions. Consequently, the resulting signal exhibits spectral regrowth in the form of in-band signal distortions and out-of-band radiation [1], and the bit error rate (BER) increases.

The design of OFDM signals aims to simultaneously achieve high data rate, high spectral efficiency (measured by the adjacent channel power ratio - ACPR) and low computational complexity [2]–[5]. This design is highly affected by the nonlinear effects of the HPA. While keeping the PAPR level low is favorable, there are other design criteria that should be taken into account. Specifically, it is of particular importance to have acceptable signal spectral behavior and BER, which are often referred to as *waveform design*. These criteria often collide, and some trade-offs appear, thus lower PAPR levels are achieved with higher BER.

Various OFDM PAPR reduction techniques have been proposed in the literature [6], [7]. Generally, these techniques can be categorized into *model-driven* and *data-driven* techniques. The first category refers to standard approaches in classical communications theory, while the second relies on recent approaches based on machine learning techniques.

A. Classical Approaches (Model Driven)

PAPR reduction schemes are roughly classified into three categories: The signal distortion category consists of techniques such as clipping and filtering (CF) [8]–[11], which limits the peak envelope of the input signal in the time domain to a predetermined value. The multiple signaling probabilistic category includes methods such as selective mapping (SLM) [11], [12], partial transmit sequence (PTS) [6], [12], ton reservation and ton injection [6], [7], [13], and constellation shaping [14]–[16]. The main principle of SLM is to generate several different candidates for each OFDM block by multiplying the symbols vector with a set of different pseudo-random sequences, and to choose the candidate with the lowest PAPR. In the PTS scheme, input data are divided into smaller disjoint sub-blocks, which are multiplied by rotating phase factors. The sub-blocks are then added to form the OFDM symbol for transmission. The objective of PTS is to design an optimal phase factor for a sub-block set that minimizes the PAPR. The coding technique category is presented in [6], [7], [17].

B. Deep Learning Based Schemes (Data Driven)

In recent years much research has been dedicated to applying deep learning (DL) techniques for designing and optimizing wireless communication networks, e.g. [2], [18]–[20]. Several papers propose DL methods to handle PAPR reduction. For example, the authors of [21], [22], added a neural network (NN) to reduce the complexity of the active constellation scheme, followed by CF. In [16], [23] the authors present an autoencoder (AE) solution for PAPR reduction, while minimizing the BER degradation. The authors in [24], [25] proposed a deep NN combined with SLM to mitigate the high PAPR issue of OFDM signal types. They use an AE structure to represent the constellation mapping and demapping of the transmitted symbols.

C. Main Contributions

Some of the aforementioned approaches suffer from in-band interference, out-of-band distortions and high computational complexity. In this paper we aim to handle the PAPR problem

as an integral part of a waveform design objective. In particular, we design a communication system, which simultaneously achieves PAPR reduction, acceptable spectral behavior of the PA's output and good BER performance. Novelties we introduce include using a CAE combined with a gradual loss learning technique to handle the multi-objective optimization of the network, and adding the HPA effect to an integrated end-to-end communication system. We demonstrate our algorithm's results on simulated data, and we compare them with classical methods for PAPR reduction and waveform design, showing competitive results for all three aforementioned objectives. The proposed algorithm allows performance improvement of future wireless communication systems.

II. Problem Definition

In an OFDM system with N complex orthogonal subcarriers, the discrete-time transmitted OFDM signal is given by the inverse discrete Fourier transform (IDFT),

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{j\frac{2\pi k}{LN}kn}, \;\; 0 \le n \le LN-1, \qquad (1)$$

where $\{X_k\}_{k=0}^{N-1}$ are random input symbols modulated by a finite constellation, and $L \ge 1$ is the over-sampling factor ($L = 1$ is the Nyquist sampling rate). As shown in [6], [7], oversampling by a factor of four results in a good approximation of the continuous-time PAPR of complex OFDM signals. The PAPR of the transmitted signal in (1) is defined as the ratio between the maximum peak power and the average power of the OFDM signal, i.e.,

$$\mathsf{PAPR} \triangleq \frac{\max_{0\le n\le LN-1} |x_n|^2}{\mathbb{E}|x_n|^2}, \qquad (2)$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator.

As HPA non-linearity causes spectral regrowth, an important assessment for the spectral purity of the system is the ACPR criterion, which is the ratio between the power of the adjacent channel and the power of the main channel, defined as [26]:

$$\mathsf{ACPR} \triangleq \frac{\max\left(\int_{\mathsf{BW}/2}^{3\mathsf{BW}/2} \mathsf{P_{ss}}(f)\,\mathrm{d}f, \int_{-3\mathsf{BW}/2}^{\mathsf{BW}/2} \mathsf{P_{ss}}(f)\,\mathrm{d}f\right)}{\int_{-\mathsf{BW}/2}^{\mathsf{BW}/2} \mathsf{P_{ss}}(f)\,\mathrm{d}f}, \qquad (3)$$

where $\mathsf{P_{ss}}(\cdot)$ is the power spectral density (PSD) of the signal at the HPA's output, and BW is the main channel bandwidth, which is assumed to be equal to the data signal bandwidth.

A block diagram of the communication system model is shown in Fig. 1. Specifically, the encoder and filter blocks mitigate the PAPR effect and design the waveform to comply with predefined spectral mask requirements. For example, the encoder block can model a clipping operation, while the filter can be a standard band-pass filter (BPF). The filtered signal x_n^{F} is amplified by a non-linear PA. The amplified signal, $x_n^{\mathsf{P}} = \mathrm{G}(x_n^{\mathsf{F}})$, is transmitted through an additive white Gaussian noise (AWGN) channel. The channel decoder receives the noisy signal and tries to reconstruct the transmitted signal. Finally, maximum likelihood (ML) is applied for detecting the estimated symbol denoted by $\hat{X}_k$.

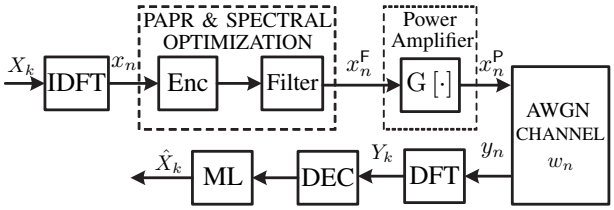

Fig. 1: System model diagram.

The role of the HPA is to convert the low-level transmission signal to a high power signal, capable of driving the antenna at the desired power level. For achieving maximal power efficiency, the HPA has to operate close to its saturation region. If the HPA exceeds the saturation point and enters the non-linear region of operation, the output signal becomes non-linear. Accordingly, in order to operate the amplifier only in the linear region, we need to make sure that the amplifier operates at a power level that is lower than the saturation point, so that even if the amplifier's input signal increases, the HPA will not enter the non-linear region. This is achieved by down-scaling the input signal by an input back-off (IBO) factor. The drawback of adding the IBO attenuation is that the output power decreases, which makes the HPA power-inefficient.

There are several commonly used models for the non-linearity of an HPA. Here, we will focus on the RAPP model [27], which is very accurate for solid-state-power-amplifiers (SSPA), and where only the amplitude is affected (AM/AM conversion!!!). The model's AM/AM conversion is given by

$$\mathsf{G}(A_{\mathsf{in}}) = v \cdot A_{\mathsf{in}} \cdot \left(1 + \left(\frac{vA_{\mathsf{in}}}{A_0}\right)^{2p}\right)^{-\frac{1}{2p}}, \qquad (4)$$

where A_{in} is the input amplitude, A_0 is the limiting output amplitude, v is the small signal gain, p is a smoothness parameter controlling the transition from the linear region to the saturation region, and $\mathsf{G}(A)$ is the output amplitude. Fig. 2 shows RAPP HPA outputs versus input for several smoothing factor values.

III. Proposed Waveform Design Structure

In this section, we first briefly discuss the CAE general concept. The proposed architecture in Fig. 3 is then elaborated, including the Bussgang's nonlinearity compensation, followed by a description of the gradual learning process.

A. Convolutional Autoencoder (CAE)

The proposed implementation uses an AE learning system based on a convolutional neural network (CNN). The general structure of an AE consists of two main blocks: the encoder $f(\boldsymbol{x})$ and the decoder $g(\boldsymbol{x})$, where $\boldsymbol{x}$ is the input data. The AE is trained to minimize a certain joint loss function, which we denote by $\mathcal{L}(\boldsymbol{x}, g(f(\boldsymbol{x})))$. An end-to-end communication system can be interpreted as an AE in which the encoder and the decoder are part of the transmitter and the receiver

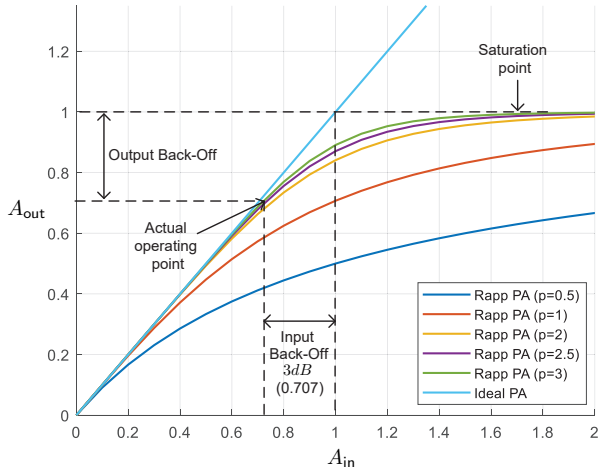

Fig. 2: RAPP HPA output versus input signal for different smoothness p values.

respectively, and they can be jointly optimized through an end-to-end learning procedure. AEs have been applied in recent years to various wireless communication tasks, such as MIMO detection [28], channel coding [29] and blind channel equalization [30].

CNNs are widely used for feature extraction and pattern recognition in ML models. Compared with a fully connected (FC) network, a CNN has significantly less connections between adjacent layers, thus less parameters and fewer weights to train, resulting in lower complexity and much faster training.

B. Proposed CAE Architecture

We propose a CAE learning system, as depicted in Fig. 3, where the input data are the N subcarries $\{X_k\}_{k=0}^{N-1}$ in the frequency domain. Then the signal is zero-padded and converted to the time domain by FFT, outputting $\{x_n\}_{n=0}^{LN-1}$. These symbols serve as input to the encoder, which acts as a PAPR reduction block, followed by a filter for optimizing the spectral behavior. Both the encoder and the decoder are composed of two convolutional layers, each followed by a nonlinear activation function and batch normalization [31], and then a fully connected layer. We have tested several activation functions, including sigmoid, rectified linear unit (RELU), Gaussian error linear unit (GELU), and scaled exponential linear unit (SELU) [32]. Empirically, SELU activation provides the best results for our CAE scheme. In addition, the encoder has a power normalization layer [33], which insures that the transmitted signal meets the power constraints.

In the transmitter, we use a BPF, whose frequency response is a rectangular window with the same bandwidth as that of X_k, for reducing the out-of-band radiation. Then, a predefined IBO is applied, and the signal is amplified by the HPA. The signal is then transmitted through an AWGN channel.

To overcome the nonlinearity of the HPA we compensate the receiver input signal by applying an attenuation factor represented by α. The Bussgang's theorem [34] states that if a zero-mean Gaussian signal passes through a memory-less nonlinear device, then the output-input cross correlation function is proportional to the input autocovariance. Accordingly, the value of α is chosen to minimize the variance of the nonlinear signal distortions. It can be shown that

$$\alpha = \frac{\mathbb{E}\left(x_n \overline{x}_n^{\mathsf{P}}\right)}{\mathbb{E}\left(|x_n|^2\right)}, \tag{5}$$

where x_n^{P} is the complex output signal of the PA, and $\overline{x}_n^{\mathsf{P}}$ is its complex conjugate.

On the receiver side, FFT converts the signal to the frequency domain, and the signal is divided by α to compensate for the nonlinear distortions. The zero-unpadding block removes the out of band samples, and finally the decoder of the proposed CAE reconstructs the estimated signal.

C. Training of the CAE Network

We train a single CAE model for all tested SNR values. We use AdamW optimizer [35] that runs back-propagation to optimize the model during training. AdamW is designed such that it improves gradients when using L_2 regularization.

Our loss function is set to optimize three objectives: accurate signal reconstruction (minimal BER), minimal PAPR and minimal ACPR. These objectives are represented by three loss components $\mathcal{L}_1$, $\mathcal{L}_2$ and $\mathcal{L}_3$, respectively. That is,

$$\mathcal{L}(x, \hat{x}) = \mathcal{L}_1(x, \hat{x}) + \lambda_2 \mathcal{L}_2(x) + \lambda_3 \mathcal{L}_3(x), \tag{6}$$

where λ_2 and λ_3 are hyper-parameters, which balance the contribution of each loss component to the joint loss function.

The loss functions we use for optimizing signal reconstruction is the minimum square error (MSE) function with L_2 regularization to reduce over-fitting. Denoting by x the input sample (which is also the output target), $\hat{x}$ as the estimated signal, Θ as the model's weights, and λ_1 as a hyper-parameter for tuning the L_2 regularization, the loss function is given by

$$\mathcal{L}_1(x, \hat{x}) = \|x - \hat{x}\|_2^2 + \lambda_1 \|\Theta\|_2^2 . \tag{7}$$

For minimizing the PAPR, we calculate it according to the BPF output, x_n^{F} (cf. Fig. 3), so that

$$\mathcal{L}_2(x) = \mathsf{PAPR}\{x_n^{\mathsf{F}}\}. \tag{8}$$

The ACPR loss component is given by

$$\mathcal{L}_3(x) = \mathsf{ACPR}\{x_n^{\mathsf{P}}\} - \mathsf{ACPR}_{\mathsf{req}}, \tag{9}$$

where, x_n^{P} is the PA's output, and $\mathsf{ACPR}_{\mathsf{req}}$ is the required ACPR value, which is usually dictated by a standard. $\mathsf{ACPR}_{\mathsf{req}}$ was set according to LTE standard requirements for high spectral purity: $\mathsf{ACPR}_{\mathsf{req}} \leq -45\mathrm{dB}$ [26].

We have applied a gradual loss learning technique: In the first stage, the loss functions consisted only of $\mathcal{L}_1$, so that only the reconstruction loss was optimized. Then, after a predetermined number of epochs, the loss function defined in (6) was used to reduce the PAPR and improve the spectral behavior.

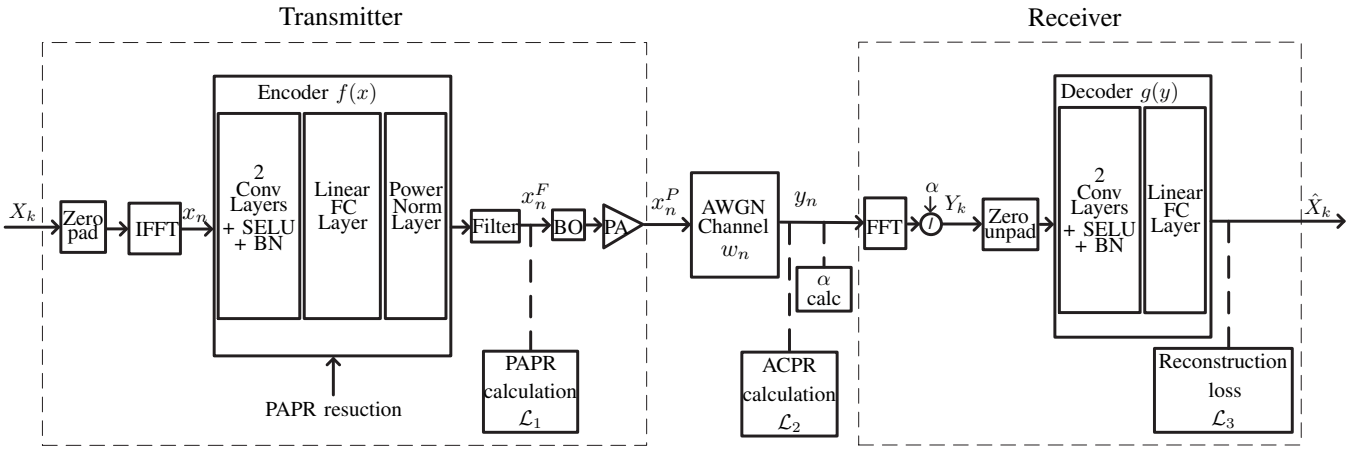

Fig. 3: Structure of the proposed conv-AE scheme.

IV. Results and Insights

We consider a SISO OFDM system with 72 subcarriers and 4-QAM modulation. 4375 batches of 32 samples each were used for training. An oversampling factor $L = 4$, and smoothness factor $p = 2$ were used. The structure of the proposed CAE for the above system is described in Table I. We compare our CAE model to a CF algorithm with clipping ratio of 1.58 dB, and to SLM with $U = 128$ phase sequences.

TABLE I: CAE Proposed Structure

Parameter	value	Kernel	In-channel	Out-channel
Transmitter				
Input size	360			
Conv (SELU)	-	3	1	13
Conv (SELU)	-	3	13	11
FC (Linear) size	360			
Receiver				
Input size	72			
Conv (SELU)	-	3	1	11
Conv (SELU)	-	3	11	13
FC (Linear) size	72			
General definitions				
Conv padding	2			
Learning rate	0.001			
epochs num	160			
Subcarriers number	72			
λ_2	0.004			
λ_3	0.001			

A. BER Analysis

Peak Signal to Noise Ratio (P_SNR) is defined as the ratio between the maximal HPA power A_0 and the noise power σ_w^2, such that

$$\mathsf{P\_SNR} = \frac{A_0^2}{\sigma_w^2}. \tag{10}$$

As shown in Fig. 4, the CAE has better BER vs. P_SNR performance compared to the other examined methods.

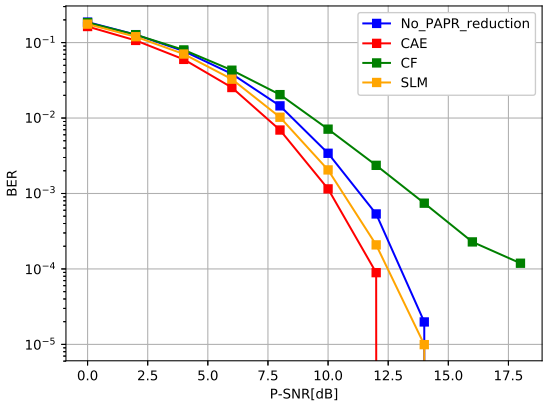

Fig. 4: BER vs. P_SNR for the considered methods

B. CCDF for PAPR Comparison

In order to evaluate the PAPR performance of different methods, a complementary cumulative distribution function (CCDF) curve is presented in Fig. 5. The CCDF of the PAPR denotes the probability that the PAPR exceeds a certain threshold, i.e. $\mathbb{P}(\mathsf{PAPR} > \mathsf{PAPR}_0)$. As can be seen in Fig. 5, the proposed CAE achieves better PAPR reduction compared to the CF and SLM methods.

C. Spectrum Analysis

Fig. 6 compares the spectral performance in terms of PSD of the transmitted signals for all examined methods. The dashed rectangle shows perfect spectral behavior for a linear HPA with no non-linear components. The proposed CAE decreases the out-of-band distortions at the expense of lower transmitted power efficiency. The transmitter's output back-off (OBO), which evaluates the power efficiency of the system, is defined as the ratio between the maximal HPA output power A_0, and the mean transmitted power:

$$\mathsf{OBO} = \frac{A_0^2}{\mathbb{E}\left(|x_n^{\mathsf{F}}|^2\right)}. \tag{11}$$

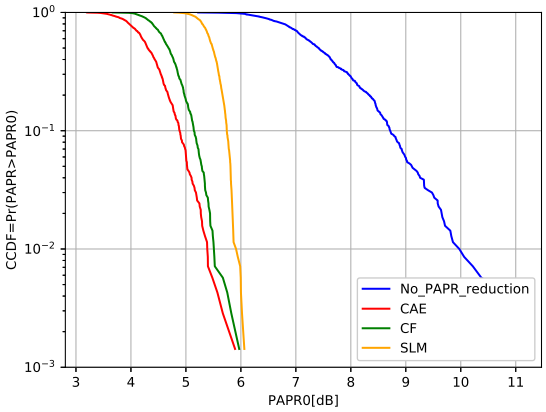

Fig. 5: CCDF of PAPR for the considered methods

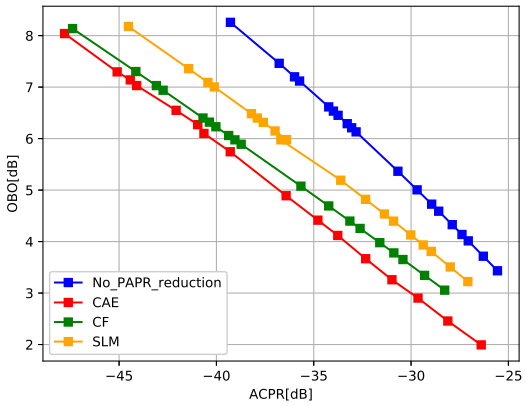

Fig. 7: OBO vs. ACPR for the considered methods

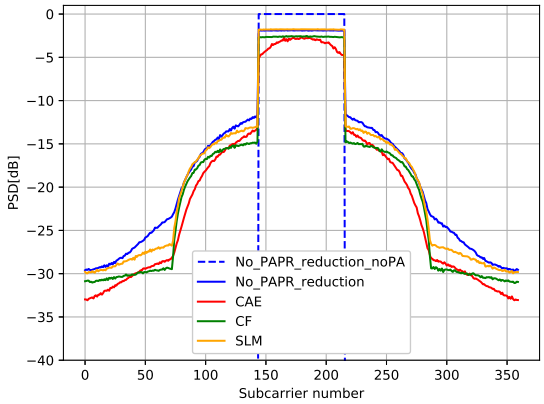

Fig. 6: PSD for the considered methods

Table II compares the ACPR and the OBO of the proposed CAE to the other methods. As can be seen the ACPR value of the CAE is comparable with the considered methods. It should be noted that for 4-QAM and 72 subcarriers, adding the ACPR constraint showed only a little improvement. We expect it to have a stronger influence for higher constellations and number of subcarriers.

TABLE II: ACPR and OBO

Parameter	CAE	FC-AE	CF	SLM	No-reduction
ACPR[dB]	-28.24	-25.54	-29.3	-27.9	-26.28
OBO[dB]	2.5	1.58	3.34	3.5	3.7

In Fig. 7 we further compare the OBO performance for different ACPR values. It can be seen that the CAE system has better power efficiency, while maintaining better BER compared to the other methods.

D. Autoencoder - FC vs. CNN

We have investigated various NN types for the AE, in particular, FC and CNN. Fig. 8 compares the BER performance of two AE architectures: the proposed CAE, which contains convolutional layers, and a fully connected autoencoder (FC-AE), which contains only FC layers. It can be observed that the CAE network has better BER performance compared to the FC-AE. As shown in Table II, the ACPR of the CAE is better than that of the FC-AE. Moreover, the CAE has lower complexity and thus faster training: The two convolutional layers have a total of 468 parameters, while for FC layers of sizes 2500 and 3500, as were used for the FC-AE in Fig. 8 and Table II, the number of parameters is around 10^7.

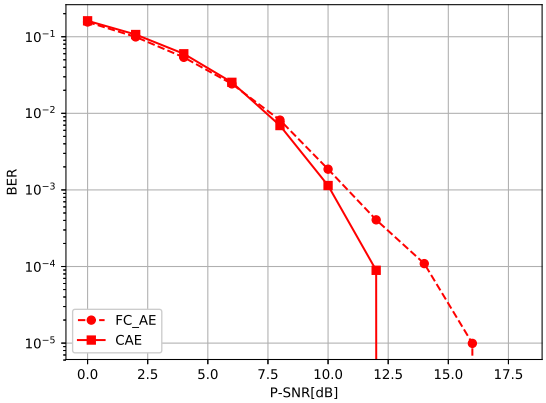

Fig. 8: BER vs. P_SNR of FC-AE and CAE

E. Fixed vs. Gradual Loss Learning

For showing the benefits of using a gradual loss learning procedure, Fig. 9 compares its BER performance to that of a fixed-loss training procedure, where the loss function's weights are fixed for the entire training. It can be observed that the gradual loss learning procedure significantly improves the BER. In addition, improving the BER while keeping the PAPR and spectral performance at the desired levels is easier to control when applying the gradual loss learning method than manipulating loss function weights in fixed-loss training.

V. Conclusions and Future Work

In this study we have presented a CAE model for PAPR reduction and waveform design in an OFDM system. We have applied a gradual loss learning method to optimize the performance of three objectives: low BER, low PAPR and adherence

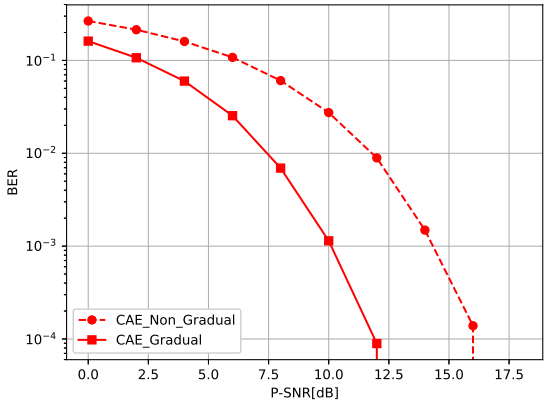

Fig. 9: BER vs. P_SNR of fixed and gradual loss learning

to ACPR spectral requirements. The proposed CAE was shown to outperform the CF and the SLM algorithms. In future work we plan to extend the model to a multiple-input-multiple-output (MIMO) scenario with higher modulation schemes and more complex channel models, aiming to achieve a functional utility for future wireless communication networks.

ACKNOWLEDGMENT

We would like to express our appreciation and gratitude to Prof. Dov Wulich, and Dr. Ilia Yoffe for their professional guidance throughout the work on this paper.

REFERENCES

[1] A. Bo, Y. Z. xing, P. C. yong, Z. T. tao, and G. J. hua, "Effects of PAPR reduction on HPA predistortion," *IEEE Transactions on Consumer Electronics*, vol. 51, no. 4, pp. 1143–1147, 2005.

[2] C. Wang, M. D. Renzo, S. Stanczak, S. Wang, and E. G. Larsson, "Artificial intelligence enabled wireless networking for 5G and beyond: Recent advances and future challenges," *IEEE Wireless Communications*, vol. 27, no. 1, pp. 16–23, 2020.

[3] Y. S.Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM wireless communications with MATLAB*. John Wiley & Sons, 2010.

[4] G. T. 38.104, "Base station (BS) radio transmission and reception (release 15), 3 GPP technical specification group radio access network; nr, v16.2.0," Dec. 2019.

[5] Y. Levinbook, D. Ezri, and E. Melzer, "Low-PAPR OFDM-based waveform for fifth-generation cellular communications," in *2017 IEEE International Conference on Microwaves, Antennas, Communications and Electronic Systems (COMCAS)*, 2017, pp. 1–6.

[6] K. D. Wong, M. Pun, and H. V. Poor, "The continuous-time peak-to-average power ratio of OFDM signals using complex modulation schemes," *IEEE Transactions on Communications*, vol. 56, no. 9, pp. 1390–1393, 2008.

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

[8] I. Gutman, I. Iofedov, and D. Wulich, "Iterative decoding of iterative clipped and filtered OFDM signal," *IEEE Transactions on Communications*, vol. 61, no. 10, pp. 4284–4293, 2013.

[9] S. P. Yadav and S. C. Bera, "PAPR reduction using clipping and filtering technique for nonlinear communication systems," in *International Conference on Computing, Communication Automation*, 2015, pp. 1220–1225.

[10] K. Anoh, C. Tanriover, B. Adebisi, and M. Hammoudeh, "A new approach to iterative clipping and filtering PAPR reduction scheme for OFDM systems," *IEEE Access*, vol. 6, pp. 17 533–17 544, 2018.

[11] A. P. More and S. B. Somani, "The reduction of PAPR in OFDM systems using clipping and SLM method," in *2013 International Conference on Information Communication and Embedded Systems (ICICES)*, 2013, pp. 593–597.

[12] Di-xiao Wu, "Selected mapping and partial transmit sequence schemes to reduce PAPR in OFDM systems," in *2011 International Conference on Image Analysis and Signal Processing*, 2011, pp. 1–5.

[13] J. Hou, C. Tellambura, and J. Ge, "Clipping noise-based tone injection for PAPR reduction in OFDM systems," in *2013 IEEE International Conference on Communications (ICC)*, 2013, pp. 5759–5763.

[14] A. Kakkavas, W. Xu, J. Luo, M. Castañeda, and J. A. Nossek, "On PAPR characteristics of DFT-s-OFDM with geometric and probabilistic constellation shaping," in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2017, pp. 1–5.

[15] A. Mobasher and A. K. Khandani, "Integer-based constellation-shaping method for PAPR reduction in OFDM systems," *IEEE Transactions on Communications*, vol. 54, no. 1, pp. 119–127, 2006.

[16] M. Kim, W. Lee, and D. Cho, "A novel papr reduction scheme for OFDM system based on deep learning," *IEEE Communications Letters*, vol. 22, no. 3, pp. 510–513, 2018.

[17] Z. Q. Taha and X. Liu, "An adaptive coding technique for PAPR reduction," in *IEEE GLOBECOM 2007 - IEEE Global Telecommunications Conference*, 2007, pp. 376–380.

[18] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?" *IEEE Transactions on Communications*, vol. 67, no. 10, pp. 7331–7376, 2019.

[19] Y. Yang, F. Gao, X. Ma, and S. Zhang, "Deep learning-based channel estimation for doubly selective fading channels," *IEEE Access*, vol. 7, pp. 36 579–36 589, 2019.

[20] O. Sholev, H. H. Permuter, E. Ben-Dror, and W. Liang, "Neural network MIMO detection for coded wireless communication with impairments," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020, pp. 1–8.

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

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

[23] L. Shi, X. Zhang, W. Wang, Z. Wang, A. Vladimirescu, Y. Zhang, and J. Wang, "PAPR reduction based on deep autoencoder for VLC DCO-OFDM system," in *2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2019, pp. 1–4.

[24] L. Hao, D. Wang, Y. Tao, W. Cheng, J. Li, and Z. Liu, "The extended SLM combined autoencoder of the PAPR reduction scheme in DCO-OFDM systems," *Applied Sciences*, vol. 9, no. 5, p. 852, 2019.

[25] L. Hao, D. Wang, W. Cheng, J. Li, and A. Ma, "Performance enhancement of ACO-OFDM-based VLC systems using a hybrid autoencoder scheme," *Optics Communications*, vol. 442, pp. 110–116, 2019.

[26] E. U. T. R. Access, "LTE: Base station (BS) conformance testing (3gpp ts 137.145 version 15.4.0)," *European Telecommunications Standards Institute*, 2019.

[27] P. Kenington, "Methods linearize RF transmitters and power amps," *Microwaves & RF*, vol. 37, no. 13, pp. 102–116, 1998.

[28] T. J. O'Shea, T. Erpek, and T. C. Clancy, "Deep learning based MIMO communications," *arXiv preprint arXiv:1707.07980*, 2017.

[29] H. A. S. K. S. O. Y. Jiang, H. Kim and P. Viswanath, "Turbo autoencoder: Deep learning based channel codes for point-to-point communication channels," in *Advances in Neural Information Processing Systems*, 2019, pp. 2758–2768.

[30] A. Caciularu and D. Burshtein, "Blind channel equalization using variational autoencoders," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2018, pp. 1–6.

[31] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" in *Advances in Neural Information Processing Systems*, 2018, pp. 2483–2493.

[32] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks," in *Advances in neural information processing systems*, 2017, pp. 971–980.

[33] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *arXiv preprint arXiv:1607.06450*, 2016.

[34] H. E. Rowe, "Memoryless nonlinearities with gaussian inputs: Elementary results," *The BELL system technical Journal*, vol. 61, no. 7, pp. 1519–1525, 1982.

[35] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.