# Neural Constellation Shaping and Back-Off Training for Memoryless Power Amplifiers

David Kopyto, Gerhard Bauch
*Institute of Communications*
*Hamburg University of Technology*
Hamburg, Germany
{david.kopyto, bauch}@tuhh.de

***Abstract*—Traveling-wave tube amplifiers (TWTAs) are common power amplifiers in satellite communications. Saturation and warping effects caused by the nonlinearity make constellation shaping a non-trivial task, when nonlinear amplifiers are involved in the channel. Constellation shaping optimizes geometry and probability of occurrence of constellation points to maximize the mutual information. This paper newly introduces constellation shaping using neural networks to satellite communications by showing its benefits on channels with TWTAs. We study how the TWTA nonlinearity impacts learned constellations both in terms of geometry and point probability of occurrence. The peak-power constraint introduced by TWTA saturation leads to a decreased impact of probabilistic shaping compared to geometric shaping which is illustrated using mutual information curves and displayed constellations. The tuning of the input back-off (IBO) of the TWTA causes an additional trade-off between transmit power at the cost of more severe nonlinear effects. This paper makes a new contribution by training IBO and transmit constellation jointly for TWTAs. By displaying constellations at different IBOs the interdependency of constellation shaping and IBO optimization is motivated. Total degradation analysis on mutual information curves resulting from the cross-entropy loss of the autoencoder illustrates the impact of joint IBO training and constellation shaping. A study of coded bit error rate performance and comparison to conventional constellations with memoryless predistortion emphasizes the effectiveness of training IBO and constellation jointly using autoencoders.**

***Index Terms*—Constellation Shaping, Autoencoders, Nonlinear Channels, Satellite Communications**

## I. Introduction

Satellite communication links typically include a traveling-wave tube amplifier (TWTA). Their nonlinear characteristic makes the design of transmit constellations non-trivial, as in contrast to a linear additive white Gaussian noise (AWGN) channel, the optimal input distribution is not known. Classical optimization of constellations for the nonlinear satellite channel lead to amplitude and phase-shift keying (APSK) constellations such as the ones found in [1]. In comparison to quadrature amplitude modulation (QAM) constellations, APSK constellations have a lower peak-to-average power ratio, therefore warping effects of the nonlinear TWTA are less pronounced. Many past works dealt with the geometric optimization of APSK constellations [2]–[4]. However, in these works the nonlinear effect is not modelled, apart from taking a peak-power constraint into account in [3].

Optimal transmit constellations may be different when directly optimizing on a joint channel of TWTA nonlinearity and AWGN.

End-to-end learning using autoencoders [5] offers the possibility to directly estimate the channel characteristics of unknown channels and optimizing the constellation with respect to the learned channel. Past work studied the joint optimization of transmit constellation, i.e. joint probabilistic and geometric shaping [6], the bit mappings [7] together with a neural demapper on an AWGN channel, as well as a Rayleigh fading channel [6], [8].

In this work, we use end-to-end learning to *jointly* optimize transmitter and receiver parameters for channels with TWTA nonlinearity. A peculiarity of power amplifiers is the influence of the input back-off (IBO) on the effective transmit power. Low IBO results in high transmit power, but an operation point close to saturation introduces a high amount of nonlinear effects. This trade-off is typically tackled by extensive bit-error rate simulations for different IBO settings, resulting in a so-called total degradation curve [4].

Traditionally, the optimization of transmission chains involving TWTAs is done separately, i.e. first, a modulation and coding scheme is optimized, which is afterwards tested on different IBOs, resulting in a total degradation curve, from which the best IBO can be computed.

In this work, we use end-to-end learning to optimize transmit constellation *jointly* with the IBO while adapting both components to the nonlinear channel. By using the cross-entropy loss, the network implicitly estimates the mutual information of the channel, enabling automated optimization of transmitter and receiver components. The communication chain is interpreted as one autoencoder, where transmitter and receiver are trainable and the channel is fixed.

End-to-end learning has been applied to various problems in communications, including OFDM [9], optical communications [10] and waveform learning [11] including an investigation for reduced out-of-band emissions in power amplifiers [12].

This paper newly proposes the joint optimization of IBO, transmit constellation and demapper to improve transmission over channels with TWTAs. We demonstrate that IBO optimization can be done together with constellation shaping. The joint optimization of IBO and constellation is relevant as both

parameters influence the mutual information of a channel with TWTA. We show that end-to-end learning for joint IBO and constellation optimization is possible by comparing the idea to a separate optimization scheme, where autoencoders are trained on various TWTA channels with fixed IBOs. Inspired by [4] we compare the total degradation for the autoencoders trained on fixed IBOs to joint training of constellation and IBO and show that the joint approach offers better performance both in terms of mutual information and bit error rate while offering a fast search procedure for the optimal IBO. This finding is non-trivial, as the introduction of further trainable parameters makes the global optimization of a transmission scheme highly non-convex as noted in [10].

To the best of our knowledge, this work is the first, which applies end-to-end learning to TWTAs for joint optimization of transmit constellation and IBO.

This paper is organized as follows: Sec. II provides an overview of common models for TWTAs, power back-off and evaluation with total degradation. In Sec. III symbol-wise and bit-wise end-to-end learning for constellation shaping are contrasted. Sec. IV provides simulation results for constellation shaping with symbol-wise and bit-wise optimization with trainable IBO as well as total degradation analysis and a comparison to a conventional scheme with memoryless predistortion according to [13]. For geometric shaping, bit error rate results are presented in Sec. IV-F and compared to conventional transmission schemes.

## II. Power Amplifiers: Models and Relations

### A. The Saleh Model for Traveling-Wave Tube Amplifiers

We model the nonlinear behavior of the TWTA by the commonly used Saleh model [14], which lets amplitude variations (AM) of the input signal cause nonlinear amplitude variations (AM/AM-conversion) and phase variations (AM/PM-conversion) of the output signal:

$$r(t) = g(|x(t)|)e^{j[\phi(t)+\vartheta(|x(t)|)]} \tag{1}$$

where $\phi(t)$ is the initial phase of the input signal. In the Saleh model, the AM/AM-conversion $g(a)$ and the AM/PM-conversion $\vartheta(a)$ are modeled by:

$$g(a) = \frac{\alpha_g a}{\beta_g a^2 + 1}, \tag{2}$$

$$\vartheta(a) = \frac{\alpha_\vartheta a^2}{\beta_\vartheta a^2 + 1}. \tag{3}$$

Throughout our simulations, we fix the parameters as follows: $\alpha_g = 2$, $\beta_g = 1$, $\alpha_\vartheta = \pi/3$ and $\beta_\vartheta = 2$, which forces the input and output saturation power to be normalized to one. Constellation points at the output of the TWTA are warped by the nonlinearity [15]. If the TWTA operates with pulse shaping filter at the transmitter and matched filter at the receiver, clustering of the received constellation points due to intersymbol interference occurs as well. Furthermore, TWTAs themselves can also have memory, which is not captured in the Saleh model. However, in this work, we restrict ourselves to warping effects and model the direct impact of a memoryless TWTA on constellation points by considering only a discrete-time model without pulse shaping and receive filter.

### B. Back-Off and Signal-to-Noise Ratio of Power Amplifiers

The average input power $P_{\text{av}}^{\text{IN}}$ and average output power $P_{\text{av}}^{\text{OUT}}$ relative to input amplifier saturation power $P_{\text{sat}}^{\text{IN}}$ and output amplifier saturation power $P_{\text{sat}}^{\text{OUT}}$ are defined by the input back-off (IBO) and output back-off (OBO), respectively:

$$\text{IBO} = \frac{P_{\text{sat}}^{\text{IN}}}{P_{\text{av}}^{\text{IN}}}, \tag{4}$$

$$\text{OBO} = \frac{P_{\text{sat}}^{\text{OUT}}}{P_{\text{av}}^{\text{OUT}}}. \tag{5}$$

By using the OBO, we can define the signal-to-noise ratio (SNR) of a channel with nonlinear power amplifier denoted as $\text{SNR}^{\text{nlin}}$ as:

$$\left.\frac{P_{\text{sat}}^{\text{OUT}}}{P_N}\right|_{\text{dB}} = \left.\frac{E_b}{N_0}\right|_{\text{dB}} + \text{OBO}\big|_{\text{dB}} + (R\cdot M)\big|_{\text{dB}} = \text{SNR}^{\text{nlin}}\big|_{\text{dB}} \tag{6}$$

where $R$ denotes the code rate and $M$ the number of bits per symbol. The noise power is $P_N$. The modulation order is $2^M$.

### C. Evaluation Measures for Power Amplifiers

The total degradation is an evaluation measure for power amplifiers, which allows to determine the optimal IBO. Usually, the total degradation is computed by bit error simulations on channels with TWTAs at different IBOs. By fixing a target bit error probability $P_b = P_b^c$, the SNRs at which $P_b^c$ is reached for each IBO are compared to the SNR for $P_b = P_b^c$ on an AWGN channel:

$$\text{TD}_{\text{BER}} = \left(\left.\frac{E_b}{N_0}\right|_{\text{dB}} + \text{OBO}\big|_{\text{dB}}\right)_{[P_b^c]} - \left(\left.\frac{E_b}{N_0}\right|_{\text{dB}}\right)_{[P_b^c]}^{\text{AWGN}}. \tag{7}$$

In contrast, we adopt the total degradation computed on mutual information curves, as explained in [4]. For each IBO configuration of the TWTA, the mutual information is computed for different SNRs. When fixing a certain target mutual information $I_c$, we can define the total degradation as:

$$\text{TD}_{\text{IT}} = \text{SNR}_{[I_c]}^{\text{nlin}}\big|_{\text{dB}} - \text{SNR}_{[I_c]}^{\text{AWGN}}\big|_{\text{dB}}. \tag{8}$$

As explained in [6], the cross-entropy loss of an autoencoder yields an estimate for the maximum mutual information of the channel, which is further discussed in Sec. III. By training and evaluating autoencoders on different SNRs we can use the loss function to plot a mutual information curve.

## III. End-to-End Learning for Constellation Shaping

To jointly optimize IBO and transmit constellation for a discrete channel with TWTA nonlinearity and AWGN, we compare autoencoder structures for symbol-wise, as well as bit-wise optimization. The end-to-end model during training is depicted in Fig. 1. As explained in [6], the autoencoder consists of $\theta = [\theta_{TX}, \theta_{RX}]$, which can be adapted to the

channel to minimize the loss function. The transmitter parameters $\theta_{TX}$ include constellation geometry $\mathscr{C}$ and their distribution $p_{\mathscr{C}^{(j)}}(x)$ according to [6], [8]. As our channel also includes a power amplifier, the IBO is added to $\theta_{TX}$, which is a novelty introduced in this work. The trainable parameters of the mapper together with the channel result in a posterior distribution $p_{\theta_{TX}}$, which is approximated by the neural demapper to obtain $\tilde{p}_{\theta_{RX}}$. The neural demapper produces log-likelihood ratios (LLRs), i.e. soft estimates $\hat{\mathbf{s}}$ or $\hat{b}_i$ of the true labels (symbols $\mathbf{s}$ or bits $b_i$). The weights of the neural demapper $\theta_{RX}$ are jointly optimized with the trainable parameters of the transmitter $\theta_{TX}$ using either the categorical cross-entropy or the binary cross-entropy for symbol-wise or bit-wise training.

For probabilistic shaping, the constellation is divided into normalized sub-constellations, where each sub-constellation $\mathscr{C}^{(j)}$ is distributed according to $p_{\mathscr{C}^{(j)}}(x)$ to enable the integration of distribution matching [8].

### A. Symbol-Wise Autoencoder

For symbol-wise optimization we use the categorical cross-entropy loss function, which is connected to mutual information, as derived in [6]. The mutual information is defined as:

$$I_{\theta_{TX}}(X;Y) = H_{\theta_{TX}}(X) - H_{\theta_{TX}}(X|Y). \tag{9}$$

where the entropy of the source is denoted by $H_{\theta_{TX}}(X)$. Let us denote the ground truth label vector as $\mathbf{s} = \mathbb{1}_{\mathscr{C}}(i)$, where $\mathbb{1}_{\mathscr{C}}$ is a one-hot vector of dimension $|\mathscr{C}| = 2^M$, which is one at position $i$ representing the transmit symbol $x$ of the constellation $\mathscr{C}$. The soft estimates generated by the neural demapper are denoted by $\hat{\mathbf{s}} \in [0,1]^{|\mathscr{C}|}$. Then the categorical cross-entropy CCE is defined as:

$$\mathrm{CCE}(\mathbf{s},\hat{\mathbf{s}}) = -\sum_{i=1}^{|\mathscr{C}|} s_i \cdot \log_2 \hat{s}_i. \tag{10}$$

The loss function, which we denote as $\mathscr{L}_\theta^{sw}$, is the negative mutual information plus a correction term:

$$\mathscr{L}_\theta^{sw} = -I_{\theta_{TX}}(X;Y) + \mathbb{E}\{\mathrm{D}_{\mathrm{KL}}[p_{\theta_{TX}}(x|y)||\tilde{p}_{\theta_{RX}}(x|y)]\}. \tag{11}$$

As shown in [6], [8], one can rewrite Eq. 11 by using the categorical cross-entropy.

We construct batches of transmit constellations, which are divided into sub-constellations according to the code rate, as explained in [8]. For each constellation point the one-hot vector $\mathbf{s}$ is added as ground truth label to each batch. The categorical cross-entropy is computed for each label and its estimate, weighted by the symbol distribution $p_{\mathscr{C}^{(j)}}(x)$ and the result is averaged over the transmit constellation points and batch. The loss function in Eq. 11 is hence effectively computed by

$$\mathscr{L}_\theta^{sw} \approx -H_{\theta_{TX}}(X) - \frac{1}{B}\sum_{l=1}^{B}\left(\sum_{j=1}^{2^{M-K}}\sum_{x\in\mathscr{C}^{(j)}} p_{\mathscr{C}^{(j)}}(x)\mathrm{CCE}(\mathbf{s},\hat{\mathbf{s}})\right)^{(l)}. \tag{12}$$

where $B$ is the batch size. The inner sums over the constellation are computed $B$ times with different channel realizations to perform Monte Carlo sampling. The entropy can be changed during training, when probabilistic shaping shall be learned, hence depends on the trainable parameters $\theta_{TX}$. The symbol distribution $p_{\mathscr{C}^{(j)}}(x)$ is trainable when probabilistic shaping shall be learned, and is a fixed uniform distribution, if only geometric shaping shall be performed.

### B. Bit-Wise Autoencoder

Bit-wise optimization enables joint learning of bit labeling of each constellation point, as well as constellation shaping [7]. The concept has been extended to joint probabilistic and geometric shaping [8]. In bit-wise optimization we aim at maximizing the bit-wise mutual information $\mathscr{R}$, which is an achievable rate for bit-interleaved coded modulation (BICM) systems [16]:

$$\mathscr{R}_{\theta_{TX}} = \left[H_{\theta_{TX}}(X) - \sum_{i=1}^{M} H_{\theta_{TX}}(B_i|Y)\right]^{+} \leq I(X;Y). \tag{13}$$

Similar to Eq. 11, the bit-wise loss function is the negative $\mathscr{R}$ plus a correction term, which is determined by the approximation quality of the autoencoder:

$$\mathscr{L}_\theta^{bw} = -\mathscr{R}_{\theta_{TX}} + \sum_{i=1}^{M}\mathbb{E}\{\mathrm{D}_{\mathrm{KL}}[p_{\theta_{TX}}(b_i|y)||\tilde{p}_{\theta_{RX}}(b_i|y)]\}. \tag{14}$$

We can approximate the loss function similarly to symbol-wise training by using the cross-entropy loss. As bit-wise optimization is a binary classification problem, we can use the binary cross-entropy BCE, which is defined as:

$$\mathrm{BCE}(b_i,\hat{b}_i) = -b_i \cdot \log_2(\hat{b}_i) - (1-b_i)\cdot\log_2(1-\hat{b}_i). \tag{15}$$

Here, $b_i \in \{0,1\}$ is a ground truth bit and $\hat{b}_i \in [0,1]$ its soft estimate generated by the demapper. The loss function can be approximated [8] by

$$\mathscr{L}_\theta^{bw} \approx -H_{\theta_{TX}}(X) - \frac{1}{B}\sum_{l=1}^{B}\left(\sum_{j=1}^{2^{M-K}}\sum_{x\in\mathscr{C}^{(j)}} p_{\mathscr{C}^{(j)}}(x)\sum_{i=1}^{M}\mathrm{BCE}(b_i,\hat{b}_i)\right)^{(l)}. \tag{16}$$

Here, $B$ denotes the batch size.

## IV. Simulation Results

In the following, we present results of constellation shaping using end-to-end learning on a TWTA. In Sec. IV-A, results for symbol-wise joint probabilistic and geometric shaping are shown. Sec. IV-B shows results for bit-wise joint probabilistic and geometric shaping, as well as a comparison to pure geometric shaping and conventional transmission schemes. We compare joint shaping to pure geometric shaping in Sec. IV-C, discuss the performance compared to conventional constellations in Sec. IV-D and provide a total degradation analysis in Sec. IV-E. Sec. IV-F provides bit error rate results for pure geometric shaping. For all autoencoder systems, we use a neural demapper, which consists of two dense layers

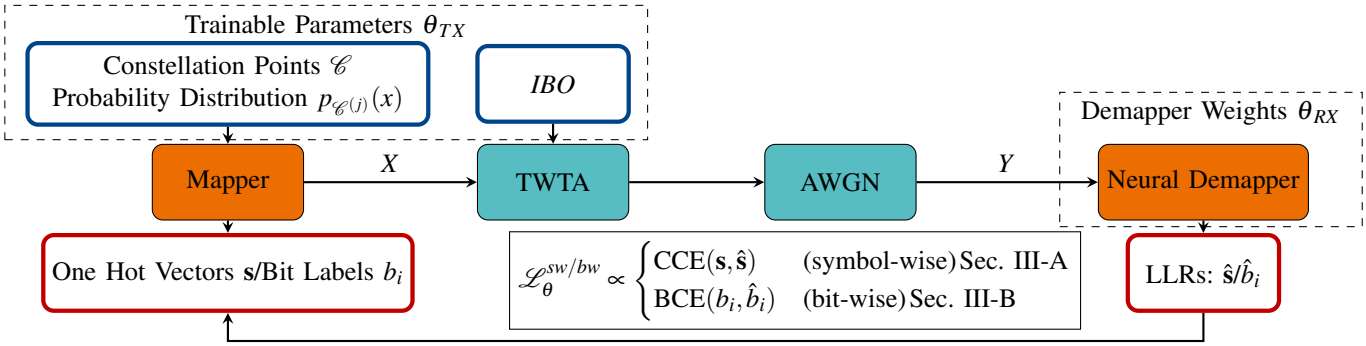



Fig. 1: Block diagram of end-to-end learning scheme used for simulations. Orange: trainable components of transmitter and receiver. Blue: trainable parameters of mapper and TWTA. The *IBO* extends the trainable transmitter parameters $\theta_{TX}$ compared to [6], [8]. Cyan: channel. Red: ground truth/predicted labels.

with 64 hidden units per layer, and a tanh activation function. The output layer dimension depends on whether symbol-wise ($2^M$ output LLRs) or bit-wise optimization ($M$ output LLRs) is used. We fixed the constellation size to $2^M = 64$, leading to $M = 6\frac{\text{bits}}{\text{symbol}}$. For training all neural systems we used a batch size of $B = 20000$ constellations with 5000 iterations per epoch.

### A. Symbol-Wise Optimization

To see the effect of nonlinear compression by the TWTA on joint probabilistic and geometric shaping, we train the symbol-wise autoencoder with the loss in Eq. 12.

Exemplary constellations for different IBOs are provided in Fig. 2. Constellation points are displayed, where the size of points represents the learned probabilistic shaping. Interestingly, constellations with low IBOs (2.36 dB and 2 dB) consist of an "outer" ring, which contains more points for lower IBOs, and an "inner" constellation, where points are shaped according to a lattice-like geometry. For low IBOs, symbol-wise optimization yields a geometry similar to APSK constellations, with points closer to the origin. The inner constellation is only slightly shaped probabilistically for IBOs 2.36 dB and 2 dB and probabilities mostly differ between "outer" ring and "inner" constellation. Only for a higher IBO of 10 dB, which is very close to an AWGN channel, probabilistic shaping is dominant in the constellation. Overall, the following statement for symbol-wise joint probabilistic and geometric shaping can be made: For low IBO probabilistic shaping has less impact on the optimized constellation, as fewer amplitude deviations are allowed due to the compression.

In Fig. 3 we provide mutual information curves for joint optimization of probabilistic and geometric shaping for the AWGN channel, as well as different values of the IBO. Results show that the trainable IBO* outperforms transmission schemes with fixed IBO.

### B. Bit-Wise Optimization

As stated in Sec. III-B, constellation shaping can also be performed by means of the bit-wise mutual information, which enables joint geometric and probabilistic shaping with bit labels. In this section we compare the results of bit-wise optimization.

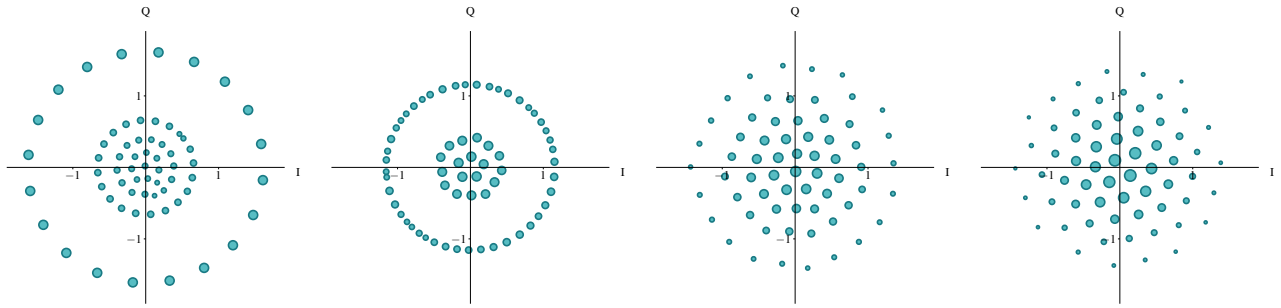
Fig. 4 shows the bit-wise mutual information for different SNRs for IBOs ranging from 0-10 dB and the AWGN performance for computing the total degradation. Results show, that bit-wise optimization yields similar behavior in terms of mutual information curves and total degradation, where the trained IBO* performs better than all other considered IBO configurations.

### C. Comparison of Joint Shaping to Pure Geometric Shaping

In Fig. 5 we compare the mutual information of joint shaping with pure geometric shaping for an AWGN channel and the TWTA channel where for each approach the IBO is jointly trained with the constellation. For the AWGN channel, the difference between the mutual information is higher than for channels with TWTA with trained IBO*. As already observed in Fig. 2, the TWTA saturation results in less impact of probabilistic shaping.

### D. Comparison to Conventional Constellations with Memoryless Predistortion

As noted in Sec. IV-A, the network learns an APSK-like constellation with unevenly spread points. This indicates that the autoencoder learned to invert the TWTA by implicitly learning a predistorted constellation. In Fig. 6, the learned constellation for IBO* for geometric shaping is compared before and after the TWTA. The constellation points after the TWTA are distributed more evenly and are closer to an APSK constellation than before warping. To support this finding, we compare the autoencoder to conventional transmission consisting of a memoryless predistorter, as proposed in [13] using a QAM and an APSK constellation (64 APSK 11/15 from [1]) with a conventional demapper. The memoryless predistorter with the TWTA generates a hard limiter with the same saturation as the TWTA. To avoid clipping on the hard limiter, we transmit QAM and APSK with an appropriate IBO, i.e. $\mathrm{IBO}_{\mathrm{QAM}} = 3.68\,\mathrm{dB}$ for QAM and $\mathrm{IBO}_{\mathrm{APSK}} = 2.38\,\mathrm{dB}$ for APSK. Figs. 4 and 5 show the bit-wise mutual information for QAM and APSK on the hard limiter. As the bit-wise mutual information is higher for IBO* compared to APSK and QAM,




(a) Trained IBO* = 2.36 dB (b) IBO = 2 dB (c) IBO = 10 dB (d) AWGN

Fig. 2: Exemplary constellations of symbol-wise joint probabilistic and geometric shaping for different IBOs and AWGN. The constellations obtained the same $I(X;Y)$ of $I_c = 4.5 \frac{\text{bit}}{\text{channel use}}$. The size of the constellation points is proportional to their probability of occurrence. The constellations are normalized to have unit energy.

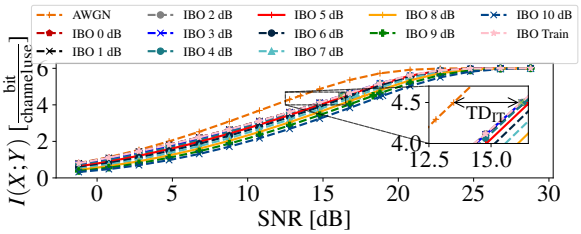



Fig. 3: Mutual information curves for AWGN channel and different IBOs with symbol-wise training for joint shaping.

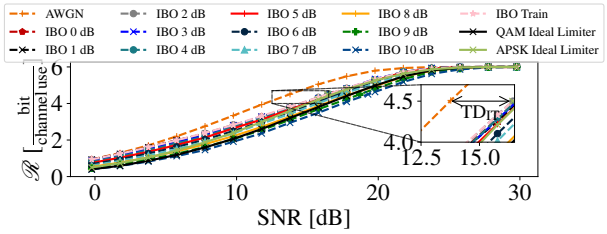



Fig. 4: Mutual information curves for AWGN channel and different IBOs with bit-wise training for joint shaping and conventional transmission with QAM and APSK.

although we assume perfect predistortion, this implies that the autoencoder learned to predistort the transmit constellation.

### *E. Total Degradation Analysis*

Fig. 7 shows the total degradation for a (bit-wise) mutual information of $I_c = 4.5 \frac{\text{bit}}{\text{channel use}}$. The trained IBO* for symbol-wise optimization (green star) is found at $\text{IBO}^* = 2.36\,\text{dB}$. For bit-wise joint shaping, the trained IBO* is at $\text{IBO}^* = 2.38\,\text{dB}$ (orange star). For bit-wise pure geometric shaping, the trained IBO* is at $\text{IBO}^* = 2.9\,\text{dB}$ (red star). All IBO* have a lower total degradation than the minima of the total degradation curves, which shows the effectiveness of incorporating the IBO to the training parameters. Furthermore, total degradation analysis supports the decreasing effect of probabilistic shaping due to TWTA saturation, as the total degradation for geometric shaping is lower than for joint shaping, meaning that probabilistic shaping is more beneficial on an AWGN channel than on a channel with TWTA.

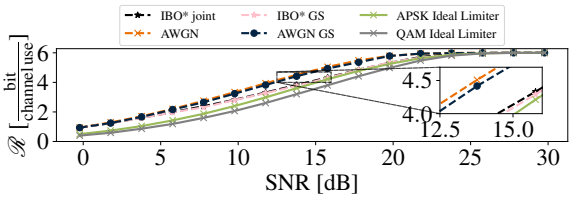
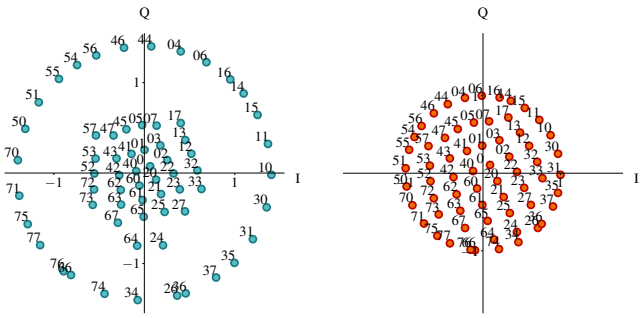



Fig. 5: Geometric (GS) vs. joint shaping for AWGN channel and trained IBO.




(a) IBO* = 2.9 dB unwarped (b) IBO* = 2.9 dB warped by TWTA

Fig. 6: Constellations for pure geometric shaping with trained IBO* before (green) and after the TWTA (red) for $I_c = 4.5 \frac{\text{bit}}{\text{channel use}}$. Constellation points are distributed more evenly after the TWTA.

### *F. BER curves for Geometric Shaping*

We present bit error rates for geometric shaping on the discrete memoryless TWTA model.

We evaluate our findings using a 5G LDPC code implemented using the open source library Sionna [17] with a block

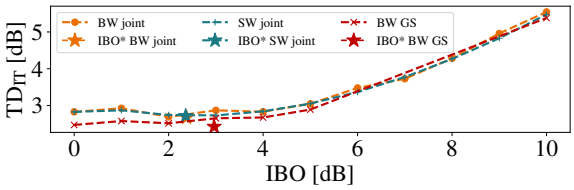



Fig. 7: Total degradation for $I_c = 4.5 \frac{\text{bit}}{\text{channel use}}$ over IBO. Orange: Bit-wise (BW) joint geometric and probabilistic shaping. Green: Symbol-wise (SW) joint geometric and probabilistic shaping. Red: Bit-wise geometric shaping (GS). The stars denote IBO* for each optimization strategy.

length of $N = 1500$ and a code rate of $R = 0.5$. To increase robustness we train the constellations on an $\frac{E_b}{N_0}$ range of 4 dB to 11 dB. Training is performed on channels with fixed IBOs ranging from from 0 dB to 3 dB, as well as the trained IBO*. As a baseline, the bit error rate of QAM and APSK from Sec. IV-B on the hard limiter are given together with their AWGN performance. The bit error rates of each considered transmission are displayed in Fig. 8.

Results show that a trainable IBO yields the lowest total degradation also in terms of bit error rates. Furthermore, joint constellation and IBO training outperforms the baseline when transmitted using memoryless predistortion, allowing transmission close to TWTA saturation.

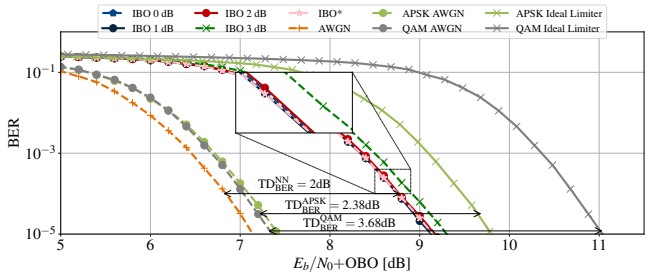



Fig. 8: BER curves for bit-wise geometric shaping for different IBO values vs. AWGN channel. For comparison, the performance of QAM and APSK on AWGN and ideal limiter channel are provided.

## V. Conclusion

We have studied the behavior of end-to-end learning under TWTA nonlinearities. As the IBO of TWTAs introduces a trade-off between transmit power and nonlinear effects, we jointly optimized the IBO with constellation and demapper. A comparison with total degradation curves showed the effectiveness of including IBO to the training parameters of the autoencoder.

By comparing joint shaping to pure geometric shaping, we have shown that the autoencoder predominantly learns geometric shaping for low IBOs.

Furthermore, joint IBO training outperforms conventional APSK and QAM constellations on a memoryless discrete TWTA channel even if perfect memoryless predistortion is used for APSK and QAM. This finding indicates that neural constellation shaping implicitly learns a predistorted constellation which is transmitted with a trained IBO making joint IBO training and constellation shaping a promising candidate for future predistortion algorithms, including those incorporating TWTA memory effects.

## References


[1] "Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications; Part 2: DVB-S2 Extensions (DVB-S2X)," https://dvb.org/?standard=second-generation-framing-structure-channel-coding-and-modulation-systems-for-broadcasting-interactive-services-news-gathering-and-other-broadband-satellite-applications-part-2-dvb-s2-extensions.

[2] R. De Gaudenzi, A. Guillen i Fabregas, and A. Martinez, "Performance analysis of turbo-coded APSK modulations over nonlinear satellite channels," *IEEE Transactions on Wireless Communications*, vol. 5, no. 9, pp. 2396–2407, Sep. 2006.

[3] F. Kayhan and G. Montorsi, "Joint signal-labeling optimization under peak power constraint," *International Journal of Satellite Communications and Networking*, vol. 30, no. 6, pp. 251–263, 2012.

[4] A. Modenini, A. Ugolini, and G. Colavolpe, "A Fast Method for optimizing the Amplifier Output Back-off by Means of the Total Degradation," in *2019 8th International Workshop on Tracking, Telemetry and Command Systems for Space Applications (TTC)*, Sep. 2019, pp. 1–6.

[5] T. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[6] M. Stark, F. Ait Aoudia, and J. Hoydis, "Joint Learning of Geometric and Probabilistic Constellation Shaping," in *2019 IEEE Globecom Workshops (GC Wkshps)*, Dec. 2019, pp. 1–6.

[7] S. Cammerer, F. A. Aoudia, S. Dörner, M. Stark, J. Hoydis, and S. ten Brink, "Trainable Communication Systems: Concepts and Prototype," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5489–5503, Sep. 2020.

[8] F. A. Aoudia and J. Hoydis, "Joint Learning of Probabilistic and Geometric Shaping for Coded Modulation Systems," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, Dec. 2020, pp. 1–6.

[9] A. Felix, S. Cammerer, S. Dörner, J. Hoydis, and S. Ten Brink, "OFDM-Autoencoder for End-to-End Learning of Communications Systems," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2018, pp. 1–5.

[10] M. Srinivasan, J. Song, A. Grabowski, K. Szczerba, H. K. Iversen, M. N. Schmidt, D. Zibar, J. Schröder, A. Larsson, C. Häger, and H. Wymeersch, "End-to-End Learning for VCSEL-Based Optical Interconnects: State-of-the-Art, Challenges, and Opportunities," *Journal of Lightwave Technology*, vol. 41, no. 11, pp. 3261–3277, Jun. 2023.

[11] F. A. Aoudia and J. Hoydis, "End-to-end Waveform Learning Through Joint Optimization of Pulse and Constellation Shaping," in *2021 IEEE Globecom Workshops (GC Wkshps)*, Dec. 2021, pp. 1–6.

[12] D. Korpi, M. Honkala, J. M. J. Huttunen, F. A. Aoudia, and J. Hoydis, "Waveform Learning for Reduced Out-of-Band Emissions Under a Nonlinear Power Amplifier," *arXiv:2201.05524 [cs, eess]*, Jan. 2022.

[13] V. Dantona, T. Delamotte, G. Bauch, and B. Lankl, "Impact of nonlinear power amplifiers on the performance of precoded MIMO satellite systems," in *2012 IEEE First AESS European Conference on Satellite Telecommunications (ESTEL)*, Oct. 2012, pp. 1–7.

[14] A. Saleh, "Frequency-Independent and Frequency-Dependent Nonlinear Models of TWT Amplifiers," *IEEE Transactions on Communications*, vol. 29, no. 11, pp. 1715–1720, Nov. 1981.

[15] E. Casini, R. D. Gaudenzi, and A. Ginesi, "DVB-S2 modem algorithms design and performance over typical satellite channels," *International Journal of Satellite Communications and Networking*, vol. 22, no. 3, pp. 281–318, 2004.

[16] G. Böcherer, "Achievable Rates for Shaped Bit-Metric Decoding," *CoRR*, vol. abs/1410.8075, 2014.

[17] J. Hoydis, S. Cammerer, F. A. Aoudia, A. Vem, N. Binder, G. Marcus, and A. Keller, "Sionna: An Open-Source Library for Next-Generation Physical Layer Research," Mar. 2022.