Higher Order Constellations for Channels with Residual Phase Noise and Nonlinear Power Amplifiers

David Kopyto, Amar Gopi Nadh and Gerhard Bauch
Institute of Communications
Hamburg University of Technology
Hamburg, Germany
{david.kopyto, amar.nadh, bauch}@tuhh.de

***Abstract*—In satellite transponders, hardware impairments play a significant role when transmission at very high data rates is desired. Two components in satellite transponders causing such impairments are the power amplifier and the oscillator. Particularly for high data rates, the power amplifier needs to be driven as close to saturation as possible, as doing so maximizes transmit power. Operation close to saturation, however, causes clipping effects. The local oscillator, on the other hand, poses challenges to synchronization at the receiver particularly at high data rates. Hence, imperfect synchronizers can result in high residual phase noise which needs to be taken into account as additional hardware impairment. When it comes to designing optimal transmit constellations, past research has treated these problems individually, resulting in spiral constellations for channels with high phase noise and amplitude and phase-shift keying (APSK) constellations for channels with amplifiers operating close to saturation. In this work, we optimize high order constellation, i.e., with order 256, for channels with joint impairment of power amplifier and residual phase noise. We optimize our constellations using neural networks, and propose an extension of spiral constellation optimization using a feed-forward network. We compare our proposed constellations to APSK constellations from DVB-S2X and spiral constellations and provide information rates for different severities of each impairment.**

***Index Terms*—autoencoders, phase noise, power amplifiers, constellation shaping, satellite communications**

I. Introduction

Local oscillators produce phase noise channels, when the hardware operates unstably [1]. Phase noise phenomena affect wireless transmissions, particularly satellites [2]. Furthermore, satellites include nonlinear power amplifiers, which also change the channel conditions compared to additive white Gaussian noise (AWGN) channels [3]. Hence, each phenomenon results in different optimal transmit constellations than in the AWGN case, whose optimal geometries have been displayed, e.g., in [4], [5].

For power amplifiers, APSK constellations work well, as their peak-to average power ratio is rather low, which makes their shape suitable for channels with peak power constraints, which are induced by power amplifiers [5], [6].

On the other hand, suitable constellations for memoryless phase noise channels are, e.g., spiral constellations [7] or similar shapes [8]–[10]. Furthermore, the construction of spiral constellations can be adapted to work well on channels with nonlinear impairments [11].

Both spiral and APSK constellations are usually optimized with respect to a parametric channel model, e.g., for spirals, the distance between consecutive constellation points is modelled as a second order polynomial in [11] to model nonlinear effects of the power amplifier and in [7] the constellation points sampled from a continuous Archimedian spiral are spaced in an appropriate distance to be robust to strong phase noise, whose angular and radial noise components differ in amplitude.

In contrast, neural networks offer a model-agnostic training paradigm. By means of backpropagation, the constellation geometry and bit labeling can be directly optimized with respect to the channel mutual information, which has been shown to be related to the cross-entropy loss [4]. For higher order modulations, the authors in [12] introduced regularizers for neural constellation shaping to symmetrize constellations on the AWGN channel and optimized quadrant-symmetric constellations, which are also known as non-uniform constellation (NUC) in the ATSC 3.0 standard [13].

In this work, we construct higher order constellations of order $M = 256$ for channels which are jointly impaired by power amplifiers and residual phase noise. To the best of our knowledge, this work is the first, which presents trained constellations for channels affected by this joint impairment. We compare spiral constellations and neurally-shaped constellations resulting from unconstrained and quadrant-symmetric optimization. Furthermore, we propose a novel extension of the nonlinear spiral from [11] which was optimized using a neural network. For comparison, a non-regular APSK constellation from DVB-S2X [14] is considered. Optimization is done for a hard limiter nonlinearity with pulse shaping and matched filter and residual phase noise following a Gaussian distribution [7]. We provide information rates for all setups for different phase noise levels and input back-off (IBO) of the power amplifier.

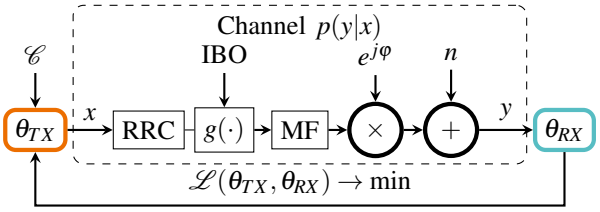


Fig. 1: Neural transmitter θ_{TX} and receiver θ_{RX} together with the channel model with pulse shaping, ideal limiter, matched filter, residual phase noise and AWGN. The loss function $\mathscr{L}(\theta_{TX},\theta_{RX})$ is minimized during training to adapt transmitter and receiver to the channel conditions.

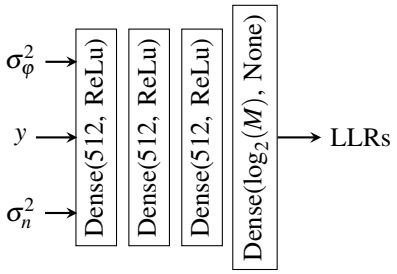


Fig. 2: Neural demapper with variance of both noise processes as input.

II. System Model

Within this work, we consider a channel with pulse shaping using a root-raised cosine (RRC) filter with roll-off factor α, a memoryless power amplifier $g(\cdot)$, matched filtering followed by memoryless residual phase noise, which reflects imperfect synchronization and receiver oscillator noise [7], [8], [15], and AWGN. A pulse shaping filter followed by nonlinear device and matched filter (MF) constitutes a Wiener-Hammerstein model [16]. This model impacts the transmit symbols with two effects: warping and clustering. The warping effect is a memoryless nonlinear effect which changes the decision regions of the constellation at the receiver [17]. Furthermore, the so-called clustering effect introduces intersymbol interference (ISI), which is summarized as additional noise d_{ISI}, as explained in [17]. In total, the discrete representation of the channel with input x and output y is

$$y = g(x)e^{j\varphi} + n + d_{\text{ISI}}. \tag{1}$$

Here, φ denotes a phase rotation and n is AWGN. We model both noise processes as Gaussian, i.e. $\varphi \sim \mathscr{N}(0,\sigma_\varphi^2)$ and $n \sim \mathscr{N}(0,\sigma_n^2)$. The transmitter and receiver are implemented as neural networks with weights θ_{TX} and θ_{RX} representing trainable components (e.g., constellation $\mathscr{C}$ and demapper from Fig. 2). Transmitter and receiver network comprise a communication autoencoder [4], which is trained using a loss $\mathscr{L}(\theta_{TX},\theta_{RX})$ as explained in Sec. III. In our work, the distortion term d_{ISI} is not modeled by a closed-form expression, but results from simulating the channel with pulse shaping and matched filtering, and is measured by the neural network. As we do not consider predistortion in this work, the ISI will be treated as additional noise by the receiver θ_{RX}.

The power amplifier is modeled as a normalized ideal limiter, meaning that its input-output characteristic can be written as:

$$|g(x)| = \min(|x|, A_{\text{sat}}), \tag{2}$$

$$\arg[g(x)] = \arg(x). \tag{3}$$

Here, A_{sat} denotes the maximum amplitude of the power amplifier and is set to $A_{\text{sat}} = 1$ during our simulations.

Commonly, the IBO and output back-off (OBO) are used to express how close the measured signal power is to amplifier saturation. The back-off is a ratio of average signal power and the power at amplifier saturation. The average signal power at amplifier input is denoted by $P_{\text{av}}^{\text{IN}}$, the input power at saturation is $P_{\text{sat}}^{\text{IN}}$, the average signal power at the output by $P_{\text{av}}^{\text{OUT}}$ and the output power at saturation is $P_{\text{sat}}^{\text{OUT}}$. To quantify the distance to amplifier saturation, the IBO and OBO are used within this work:

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

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

For channels, which are constrained by a maximum power, as the hard limiter in (2), the peak signal-to-noise ratio (PSNR) is preferred to quantify information rates or bit error rates. The PSNR is defined by

$$\text{PSNR} = \frac{P_{\text{sat}}^{\text{OUT}}}{P_N}, \tag{6}$$

where P_N denotes the noise power of AWGN.

III. Autoencoder Model

A. Mapper and Demapper Architecture

As introduced in [4], the constellation $\mathscr{C}$ can be jointly optimized with the demapper to maximize the mutual information using neural networks. Here, the mapper is treated as a neural network with trainable parameters θ_{TX}. The constellation $\mathscr{C}$ can be optimized with different constraints as summarized in Sec. IV.

The demapper is then jointly optimized with the constellation and is implemented as another neural network θ_{RX}. The demapper architecture for this work is displayed in Fig. 2. As the channel in Fig. 1 is memoryless, the demapper is implemented as a feedforward network consisting of four dense layers. The output layer then computes log-likelihood ratio (LLR) per bit, i.e., a bit-level soft demapper is trained by θ_{RX}. To train the parameters θ_{TX} and θ_{RX} a loss function needs to be chosen, such that mutual information is maximized. This loss function is introduced in Sec. III-B.

B. Loss Function for Bit-Wise Training

The authors in [4] have shown that neural networks can be used to train transmitter and receiver components to maximize the bit-wise mutual information (BMI) by leveraging the binary cross-entropy (BCE) loss between transmitted bits and estimated soft bits at the receiver. The BMI is an achievable

rate for bit-interleaved coded modulation (BICM) systems. Let $\mathbf{B}$ be the vector binary random variable with entries B_j with the input bit vector $\mathbf{b}$ of length N with entries b_j as concrete realization. The BMI is then:

$$R_{\mathrm{BMI}} = H(\mathbf{B}) - \sum_{j=1}^{N} H(B_j|Y) \leq I(X;Y), \tag{7}$$

where X and Y are the channel input and output random variables. As illustrated in Fig. 2, the demapper constructs the LLR per bit. The goal of every transmission in BICM systems is to maximize BMI. The BCE loss function is closely related to the BMI, i.e., $\mathscr{L}(\theta_{TX}, \theta_{RX}) =$

$$H(\mathbf{B}) - R_{\mathrm{BMI}} + \sum_{j=1}^{N} \mathbb{E}_y[\mathrm{D_{KL}}(p_{\theta_{TX}}(b_j|y)||\tilde{p}_{\theta_{RX}}(b_j|y))], \tag{8}$$

where $\tilde{p}_{\theta_{RX}}(b_j|y)$ are the probabilities over the N bits. The last term in (8) vanishes, when the network components are well-designed for the transmission channel [4].

Training of neural networks is done in batches, where the loss $\mathscr{L}$ is estimated using Monte Carlo iterations. For (8) this is done as follows. Let N be the block size of the l-th bit vector $\mathbf{b}_{(l)} \in \{0,1\}^N$ in a batch of size $\mathscr{B}$. Furthermore, let $\hat{\mathbf{b}}_{(l)} \in [0,1]^N$ denote the l-th soft bit vector in a batch of size $\mathscr{B}$ estimated by the demapper. The BCE loss between $\mathbf{b}_{(l)}$ and $\hat{\mathbf{b}}_{(l)}$ is then defined by $\mathrm{BCE}(\mathbf{b}_{(l)}, \hat{\mathbf{b}}_{(l)}) =$

$$\frac{1}{N}[-\mathbf{b}_{(l)}^T \cdot \log_2 \hat{\mathbf{b}}_{(l)} - (\mathbf{1} - \mathbf{b}_{(l)}^T) \cdot \log_2(\mathbf{1} - \hat{\mathbf{b}}_{(l)})]. \tag{9}$$

Please note, that the $\log_2(\cdot)$ operation is performed element-wise. By sending batches of bit vectors through the model in Fig. 1, the transmitter and receiver weights θ_{TX}, θ_{RX} are optimized by minimizing

$$\mathscr{L}(\theta_{TX}, \theta_{RX}) \approx \frac{1}{\mathscr{B}} \sum_{l=1}^{\mathscr{B}} \mathrm{BCE}(\mathbf{b}_{(l)}, \hat{\mathbf{b}}_{(l)}) \tag{10}$$

using backpropagation. Please note, that the loss also depends on the channel characteristics. The channel is, however, not trainable, and remains fixed during training. Hence, we highlight only the trainable variables θ_{TX}, θ_{RX} in $\mathscr{L}$.

IV. Geometric Shaping

A. Unconstrained Shaping and Quadrant Symmetry Constraints

In most autoencoder setups, geometric shaping has been applied without constraints [4], i.e., each of the M constellation points is optimized in real and imaginary part and the points can be placed at any position in the complex plane. Hence, $\mathscr{C} \in \mathbb{C}^M$ leading to $2M$ parameters to be optimized with respect to the channel conditions. The constellation $\mathscr{C}$ can be, for example, initialized by an M-QAM constellation. The autoencoder then adjusts the $2M$ values using backpropagation to yield the best geometry with respect to the channel mutual information. However, as no constraints are posed to the constellation, convergence to a good solution may take longer, or may be even infeasible to achieve.

A simplification can be made by introducing quadrant symmetry, as done in [12], [13]. This way, the number of trainable parameters reduces by a factor of four, resulting in $\frac{M}{2}$ parameters. In [13], such constellations are called NUC.

B. Spiral Constellations

1) Spiral Constellations for Phase Noise Channels: In [7], spiral constellations are proposed to be used for transmission under severe phase noise. The authors argue, that phase noise affects symbols with higher magnitude more severely than those with lower magnitude. More specifically, the angular noise component increases with the symbol magnitude. Appropriate constellations for strong phase noise should, hence, increase the angular distance between points when the magnitude increases. To do so, the authors propose to sample the points of an Archimedian spiral in a non-equidistant fashion. Consider the Archimedian spiral:

$$s(t) = te^{jt}. \tag{11}$$

The m-th constellation point $c_m \in \mathscr{C}_{\mathrm{spiral}}$ of a spiral constellation is found by sampling from the Archimedian spiral, i.e.,

$$s(t_m) = c_m = t_m e^{jt_m}. \tag{12}$$

To take the effects of phase noise into account, the authors then propose to sample the parameters t_m in a non-equidistant fashion:

$$t_m = \sqrt{\frac{(4\pi m)^2 f_s}{2} + \sqrt{\frac{(4\pi m)^4 f_s^2}{4} + (4\pi m)^2}}. \tag{13}$$

Notice that this construction contains only one free scalar parameter f_s, which makes the construction of spiral constellations for phase noise channels simple, as only the parameter f_s need to be adjusted to the channel conditions.

This approach yields excellent symbol-wise performance for residual phase noise as illustrated in [7], where the single parameter f_s can be well optimized to maximize the symbol-wise information rate. However, for maximization of bit-wise information rate, the bit labels need to be taken into account as well. As the authors point out, bit labeling is not straight forward for the proposed spiral construction for arbitrary f_s and remains an open research question.

However, for some f_s the number of branches in the spiral can become $\sqrt{M}$, i.e., a Gray labeling can be achieved in each branch. One value, where the number of branches equals $\sqrt{M}$ for $M = 256$ is $f_s = 0.00112$, where the corresponding spiral is shown in Fig. 4a. In this work, we use this spiral with a per-branch Gray labeling, as it was done in [7] for the comparison of bit-wise information rates.

2) Spiral Constellations for Nonlinear Channels: In [11] the authors propose to use a modified spiral:

$$s(t) = f(t)e^{jt}, \tag{14}$$

i.e., the m-th constellation point can be obtained by

$$s(t_m) = c_m = f(t_m)e^{jt_m}, \tag{15}$$

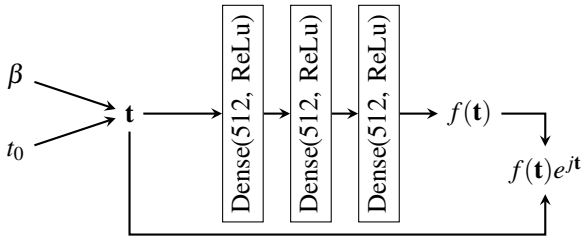


Fig. 3: Construction of the nonlinear spiral $f(\mathbf{t})$ using a feedforward network.

where $t_1 < t_2 < \cdots < t_M$ and $\mathscr{C}_{\text{spiral}} = \{c_m | m \in 1, \ldots, M\}$. This nonlinear extension can be used to design the spiral according to the amplifier model in the channel. The authors in [11] assume a third order nonlinearity, and derive the function $f(\cdot)$ to be a second order polynomial. To adapt to arbitrary memoryless $g(\cdot)$, we propose to learn the function $f(\cdot)$ by neural network training, using the architecture in Fig. 3.

In contrast to [11], we choose a simpler approach and construct the parameters t_m equidistantly:

$$t_{m+1} = t_m + \beta. \tag{16}$$

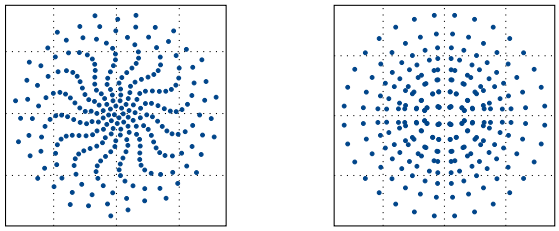
The distance between consecutive parameters β and the first entry t_0 are set as trainable parameters. The proposed construction of the nonlinear spiral (14) differs from [11], where the authors construct both the function $f(\cdot)$ and the parameters t_m as a second order polynomial. Our approach, however, allows arbitrary functions for $f(\cdot)$ as the neural network is also capable of learning more complex mappings. The restriction to equidistant t_m also gets rid of the recursive formulation proposed in [11] and a single vector $\mathbf{t} = \{t_m | m \in 0, \ldots, M-1\}$ can be constructed after defining t_0 and β which is then transformed by $f(\cdot)$.

As far as labeling is concerned, the authors in [11] proposed to use Gray labels along the spiral, meaning that labels differ only by one bit when looking at the previous and next point on the spiral. This labeling strategy performed well in their work, and hence we adapt it within our simulations and extend it to our proposed neural optimization of Fig. 3.

V. Results

To evaluate the performance of the different constellation shaping strategies on the channel with phase noise and nonlinearity, we train the autoencoder of Fig. 1 on different channel setups for all shaping strategies. We then evaluate the estimated BMI for each trained model and plot the result for different signal-to-noise ratio (SNR).

The optimization strategies, where $\mathscr{C}$ is trained with respect to the channel condition include unconstrained optimization, NUC optimization (both Sec. IV-A) and the nonlinear spiral from Fig 3. For the nonlinear spiral, we set the initial parameters as $t_0 = 2\pi$ and $\beta = 0.2$, which are optimized during training together with the mapping function $f(\cdot)$. We use a roll-off factor $\alpha = 0.1$, a batch size of $\mathscr{B} = 3000$ with bit vectors of size $N = 2000$.



(a) Spiral with $\sqrt{M} = 16$ branches. (b) APSK of DVB-S2X $R = 20/30$.

Fig. 4: Spiral constellation from [7] with $f_s = 0.00112$ for convenient Gray labeling on $\sqrt{M}$ branches. In Fig. 4b the APSK constellation from DVB-S2X [18] for rate $R = 20/30$ is displayed.

We compare these trainable setups with the spiral in Fig. 4a, which results from setting $f_s = 0.00112$. As another baseline, we consider the APSK constellation optimized for $R = 20/30$ in DVB-S2X [18], which is displayed in Fig. 4b. The constellations from Fig. 4 are kept fixed during training and only the demapper is optimized for transmission.

A. Channels with Phase Noise and AWGN

At first, the shaped constellations under severe phase noise are studied. As found in [7], spirals with many branches are suitable by construction (13) for such transmissions. The spirals from [11] and our proposed neural optimization in Fig. 3, however, are primarily constructed for nonlinear channels.

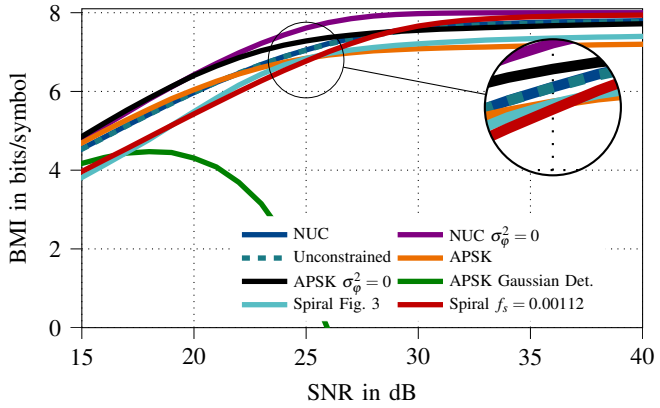
We train the constellations on two phase noise channels. For $\sigma_\varphi^2 = 0.01$, we trained the models on a range of SNR $\in [10, 50]$ dB, that is, we randomly sample noise levels for the AWGN channel n in the aforementioned interval. For $\sigma_\varphi^2 = 0.1$, we used SNR $\in [10, 70]$ dB to illustrate the effects at higher SNR. This way, the constellation and demapper can be made robust to different SNR levels while we keep the phase noise level fixed.

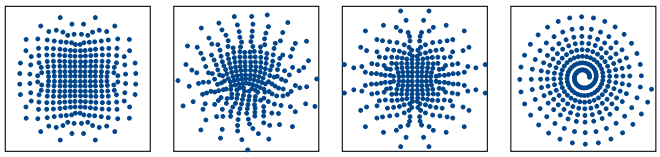
First, a phase noise impairment of $\sigma_\varphi^2 = 0.01$ is considered. We compare unconstrained training, NUC, the spiral from Fig. 3 with the spiral from Fig. 4a and the APSK in Fig. 4b.

Fig. 5 shows the BMI for all setups for $\sigma_\varphi^2 = 0.01$. For comparison, we train a NUC on a pure AWGN channel (Fig. 5a), i.e., $\sigma_\varphi^2 = 0$ which serves as an upper bound. A lower bound is given by the performance of the APSK constellation with a conventional AWGN detector on the phase noise channel.

It can be observed, that quadrant symmetric and unconstrained constellation achieve similar performance. However, the constellation resulting from unconstrained training in Fig. 5b has a skewed symmetry similar to the constructed spiral in Fig. 4a. Both unconstrained and quadrant symmetric optimization have a spiral-like geometry, similar to Fig. 4a, particularly for high amplitude, branches are built. However, for low amplitudes, the constellations deviate from the spiral scheme, and resemble a quadrature amplitude modulation (QAM)-like structure.

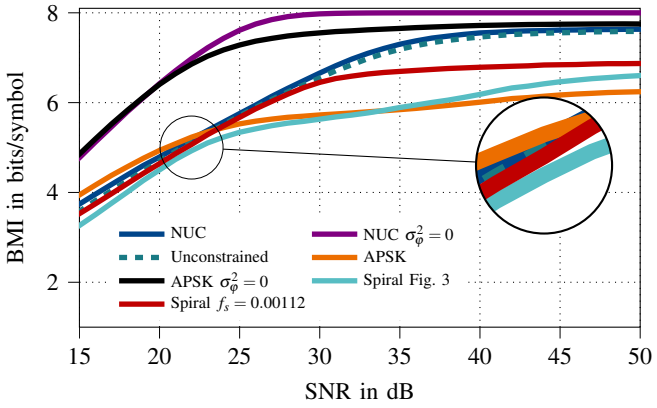
Unconstrained and NUC outperform the other setups. However, for lower SNR, the APSK constellation achieves similar performance and for higher SNR, the spiral with $f_s = 0.00112$

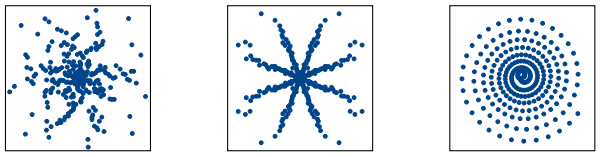




(a) NUC $\sigma_\varphi^2 = 0$. (b) Unconstrained. (c) NUC. (d) Spiral Fig. 3.

Fig. 5: Analysis of transmission over a high phase noise channel with $\sigma_\varphi^2 = 0.01$. The learned constellations are displayed underneath the information rate plot.





(a) Unconstrained. (b) NUC. (c) Spiral Fig. 3.

Fig. 6: Analysis of transmission over a high phase noise channel with $\sigma_\varphi^2 = 0.1$. The learned constellations are displayed underneath the information rate plot.

is competitive. The spiral from Fig. 3 performs slightly worse, as this construction is primarily intended for nonlinear channels.

For very high phase noise of $\sigma_\varphi^2 = 0.1$, the performance is given in Fig. 6. Naturally, the information rates decrease with higher phase noise and are now even farther away from the AWGN upper bounds. However, unconstrained and quadrant symmetric optimization can almost perform equally to the AWGN bounds for very high SNR. The constellations are now mostly 'star-like', as the high noise in the phase components degrades detection performance, forcing the demapper to take decisions mostly in amplitude direction.

B. Channels with Low IBO and Phase Noise

To illustrate how power amplifiers with low IBO impact performance when impaired by phase noise, we train different constellations on a setup with IBO $= 4$ dB, i.e., operation close to amplifier saturation. First, we show the constellations and information rates for AWGN and $\sigma_\varphi^2 = 0$ in Fig. 7. In Fig. 8, the performance on the joint channel with IBO $= 4$ dB and $\sigma_\varphi^2 = 0.01$ is displayed. For the two following setups, we train on a range of SNR $\in [10, 50]$ dB.

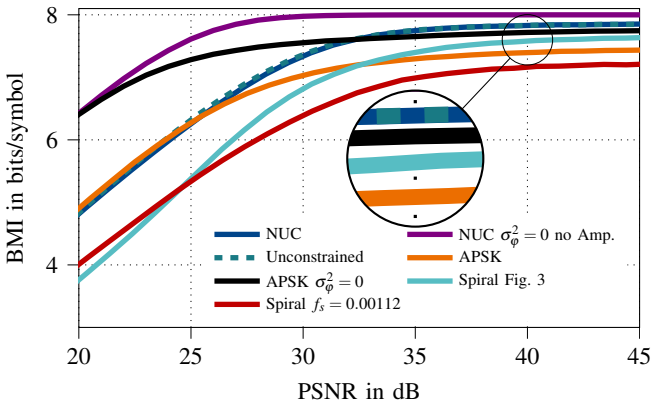
For $\sigma_\varphi^2 = 0$, the constellations show a similar shape, as explained in [3]. That is, the unconstrained and NUC optimization yields an inner constellation and an outer ring. The inner constellation looks like a 'warped' QAM. Furthermore, the spiral in Fig. 7c has the outermost lap distanced farther away from the inner laps, which mimics the same geometry of an outer ring with inner constellation. In terms of BMI, NUC and unconstrained optimization yield the best performance, while APSK is competitive for lower SNR and the spiral in Fig. 7c yields almost equal performance as NUC and unconstrained for higher SNR. This finding backs the results found in [11], where the spiral construction (15) was motivated for nonlinear channels, particularly for high SNR.

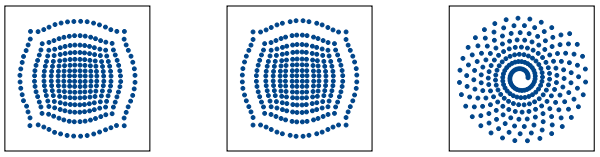
For $\sigma_\varphi^2 = 0.01$ and nonlinearity, the constellations from unconstrained and NUC optimization are also structured in conformity with [3], i.e., an outer ring and an inner constellation are visible. However, the inner constellation shows a similar shape, as for the channel without power amplifier (Fig. 5c). In contrast to Fig. 7 the spiral constructed as in Fig. 3 can not outperform the spiral with $f_s = 0.00112$ in Fig. 8. The APSK can, however, perform similarly to the unconstrained and NUC constellations for lower SNR.

From these results, one may infer, that the nonlinear spiral from Fig. 3 is optimized for channels with nonlinearity, but without residual phase noise. In contrast, the spiral from Fig. 4a is constructed for channels with high phase noise as found in Fig. 5. Hence, both spiral construction methods are designed for channels with either strong phase noise or strong nonlinearity, at least for the analyzed setups of IBO $= 4$ dB and $\sigma_\varphi^2 \in \{0.01, 0.1\}$. When both impairments occur at the same time, constellations need to be defined differently, e.g., by neural constellation shaping using NUC or unconstrained optimization.

VI. Conclusion

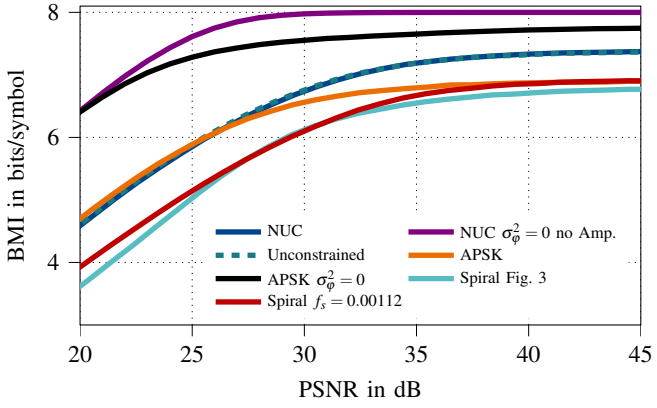
In this paper, we analyzed neural constellation shaping for channels with joint impairment of residual phase noise and low IBO of the power amplifier. We introduced a neural construction of spiral constellations, which extends the idea of [11]. For channels with strong phase noise impairment,

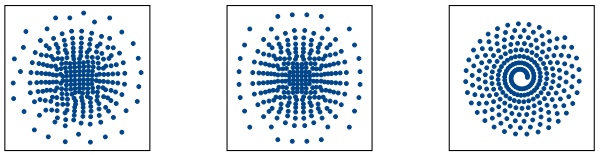




(a) Unconstrained. (b) NUC. (c) Spiral Fig. 3.

Fig. 7: Information rate and trained constellation for IBO = 4 dB and no phase noise.





(a) Unconstrained. (b) NUC. (c) Spiral Fig. 3.

Fig. 8: Information rate and trained constellation for IBO = 4 dB and $\sigma_\varphi^2 = 0.01$.

we have shown that unconstrained optimization leads to constellations with similar shape as the spirals in [7], i.e., constellations with branches. The shape is particularly similar for high amplitudes, whereas for low amplitudes, unconstrained optimization yields different shapes, e.g., QAM-like appearances.

The neural spiral construction of Fig. 3 performs well for channels with nonlinearity, but does not yield equal performance on a channel with both nonlinearity and residual phase noise. In contrast, the constellation constructed by the method in [7] yields good performance on channels with phase noise and no nonlinearity. For channels, where both impairments occur, constellation shaping using quadrant-symmetry constraints or unconstrained optimization yields superior performance for the considered setups.

References

[1] A. Piemontese, G. Colavolpe, and T. Eriksson, "A New Analytical Model of Phase Noise in Communication Systems," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2022, pp. 926–931.

[2] T. Colin, T. Delamotte, and A. Knopp, "Phase Noise Characterization for Ultra High-Throughput Satellite Systems," in *ICC 2022 - IEEE International Conference on Communications*, May 2022, pp. 2369–2374.

[3] D. Kopyto and G. Bauch, "Neural Constellation Shaping and Back-Off Training for Memoryless Power Amplifiers," in *2024 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2024, pp. 1–6.

[4] 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.

[5] F. Kayhan and G. Montorsi, "Constellation design for transmission over nonlinear satellite channels," in *2012 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2012, pp. 3401–3406.

[6] 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.

[7] A. Ugolini, A. Piemontese, and T. Eriksson, "Spiral Constellations for Phase Noise Channels," *IEEE Transactions on Communications*, vol. 67, no. 11, pp. 7799–7810, Nov. 2019.

[8] R. Krishnan, A. Graell i Amat, T. Eriksson, and G. Colavolpe, "Constellation Optimization in the Presence of Strong Phase Noise," *IEEE Transactions on Communications*, vol. 61, no. 12, pp. 5056–5066, Dec. 2013.

[9] F. Kayhan and G. Montorsi, "Constellation Design for Memoryless Phase Noise Channels," *IEEE Transactions on Wireless Communications*, vol. 13, no. 5, pp. 2874–2883, May 2014.

[10] D. Marasinghe, L. H. Nguyen, J. Mohammadi, Y. Chen, T. Wild, and N. Rajatheva, "Constellation Shaping Under Phase Noise Impairment for Sub-THz Communications," in *ICC 2024 - IEEE International Conference on Communications*, Jun. 2024, pp. 3833–3838.

[11] A. Piemontese, A. Ugolini, M. Morini, G. Colavolpe, and T. Eriksson, "Spiral Constellations for Nonlinear Channels," *IEEE Communications Letters*, pp. 1–1, 2024.

[12] P. Madadi, J. Cho, C. J. Zhang, and D. Burghal, "AI/ML Optimized High-Order Modulations," in *ICC 2023 - IEEE International Conference on Communications*, May 2023, pp. 6373–6378.

[13] N. S. Loghin, J. Zöllner, B. Mouhouche, D. Ansorregui, J. Kim, and S.-I. Park, "Non-Uniform Constellations for ATSC 3.0," *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 197–203, Mar. 2016.

[14] "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)," http://www.etsi.org.

[15] R. Krishnan, M. R. Khanzadi, T. Eriksson, and T. Svensson, "Soft metrics and their Performance Analysis for Optimal Data Detection in the Presence of Strong Oscillator Phase Noise," *IEEE Transactions on Communications*, vol. 61, no. 6, pp. 2385–2395, Jun. 2013.

[16] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. USA: Krieger Publishing Co., Inc., Mar. 2006.

[17] M. H. Moghaddam, S. R. Aghdam, N. Mazzali, and T. Eriksson, "Statistical Modeling and Analysis of Power Amplifier Nonlinearities in Communication Systems," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 822–835, Feb. 2022.

[18] "Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications; Part 1 (DVB-S2)," http://www.etsi.org.