

Probabilistic Shaping for Multidimensional Signals with Autoencoder-based End-to-end Learning

Xinyue Liu and Izzat Darwazeh

Dept. of Electronic and Electrical Engineering
University College London
London WC1E 7JE, UK
{x.liu.17, i.darwazeh}@ucl.ac.uk

Nader Zein

NEC Laboratories Europe
NEC Europe
Middlesex HA4 6QE, UK
Nader.Zein@emea.nec.com

Eisaku Sasaki

1st Wireless Access Solutions Division
NEC Corporation
211-8666, Japan
sasaki-e@nec.com

Abstract—This work proposes a system that optimises multidimensional signal transmission, utilising signals with probabilistic shaping designed with the aid of end-to-end learning of an autoencoder-based architecture. For the first time, this work reports bit mapping optimisation for multidimensional signals and applied the newly derived optimised signals to the probabilistic shaping system. The autoencoder employs two neural networks for the transceiver, separated by the embedded channel. The optimisation of the autoencoder configuration is implemented for probabilistic shaping for n -dimensional signals. Specifically, We investigate a 4-dimensional (4D) signal employing 2 successive time slots that has better noise immunity relative to regular 2-dimensional quadrature amplitude modulation (QAM) signals. We propose a new application of autoencoders in communication systems based on 4D signals and apply machine learning to optimise the 4D probabilistic shaping on the basis of receiver signal-to-noise-ratio (SNR). The performance of the optimised probabilistically shaped 4D signals is evaluated in terms of the bit error rate (BER) and mutual information. Simulation results show that the proposed probabilistically shaped 4D signal achieves better BER performance relative to the unshaped 4D and regular 2D QAM. We demonstrate the mutual information of the proposed signal with varying SNR, showing its improved capacity in comparison with other constellations.

Index Terms—probabilistic shaping, end-to-end learning, multidimensional modulation, autoencoder

I. INTRODUCTION

Probabilistic shaping, a non-uniform signalling technique, has attracted wide research interests, because of its advantages of higher capacity, lower power requirements and flexible data rate adaption. The basic idea of probabilistic shaping is to encode the data stream appropriately so that symbols are transmitted with unequal probabilities. Therefore, the achievable data rate can be maximised by optimising the occurrence probabilities of such symbols. To this end, conventional probabilistic shaping is implemented by coding schemes [1] such that the input bit stream can be encoded to symbols with a target probability distribution. This procedure was later on termed distribution matching (DM). In 2015, Böcherer and Steiner in [2] reviewed the classic shaping schemes, such as trellis shaping and shell mapping, and introduced a reverse concatenation architecture of probabilistic amplitude shaping (PAS). In [3], the same authors proposed the use of constant composition distribution matcher (CCDM) with PAS for efficient implementation of probabilistic shaping. Since

then PAS has been widely used employing different DM algorithms [3]–[6], with the focus on improving the robustness of probabilistically shaped signals against different channel impairments, such as fading in wireless channels [7] and nonlinearities in optical fibre systems [8].

The DM-based probabilistic shaping has been intensively investigated with many applications emerging for various communication systems. Typical designs of probabilistic shaping scheme are mainly 1-dimensional (1D), which can be directly applied to singular dimensional modulation formats such as pulse amplitude modulation (PAM) [9]. To construct n -dimensional (nD , $n \geq 2$) modulations with probabilistic shaping, the n -fold Cartesian product of the 1D probability distribution is used [10]. Particularly, probabilistically shaped 2D quadrature amplitude modulation (QAM) can be achieved by applying 1D probabilistic shaping for the in-phase (**I**) and quadrature (**Q**) components. However, the Cartesian product is not feasible to shape non-regular 2D constellations such as 32-QAM and 128-QAM. To tackle this problem, a universal probabilistic shaping scheme has been proposed in [11] based on a 2D distribution matcher, which exploits shaping to 2D symbols straightforwardly. For multidimensional signals, additional dimensions for nD symbols ($n > 2$), apart from the **I** and **Q** dimensions, can be achieved by employing successive time slots [12] or a pair of polarisation of optical signals [13]. While there have been numerous studies on geometrical shaping for multidimensional modulations [14] [15], only a few papers exist on multidimensional probabilistic shaping, with such shaping itself only being applied to 2D signals in all existing literature [16]–[18].

Deep learning technology has been applied to communication systems by approximating and optimising the transmitter and receiver functions using neural networks [19]. In particular, A very recent work in [20] proposes a special end-to-end learning approach employing the autoencoder to perform the shaping and optimise the probability distribution for 2D constellation symbols. In the autoencoder, two neural networks, i.e., the encoder and decoder, are jointly trained to minimise the error rate of the reconstructed symbols at the receiver. By adding the channel effects (i.e., channel embedded autoencoder), the learned constellation shapes will be robust against channel impairments. More precisely, the

encoder finds optimal representation of the signal for a given channel and then the decoder recovers the input from the contaminated data. The learning procedure of autoencoder for 2D probabilistic shaping is leveraged on the Gumbel-Softmax trick [21], which provides a continuous and therefore differentiable approximation of the originally discrete probability distribution for the training [20].

Inspired by other researchers' recent successes in 2D probabilistic shaping with autoencoders [11] [22], in this paper we study the probabilistic shaping for multidimensional modulations using autoencoder and end-to-end learning techniques. Herein, for the learning procedure, the sampling method for the discrete distribution approximation [21] is utilised for the nD constellation. We employ the multidimensional signals proposed in [12], where the constellation points in the nD space are selected from a lattice. The nD signaling is implemented by transmitting 2D signals in $n/2$ successive signaling intervals. Such nD signal designs have improved Gaussian noise immunity and are advantageous in their power efficiency relative to regular 2D QAM that is of the same bit-per-symbol. This work, for the first time, optimises the bit-to-symbol mapping for the aforementioned 4D constellation by using the reactive tabu search (RTS) algorithm [23]. The optimised bit mapping is expected to have better bit error rate (BER) performance relative to the original mapping. For space limitation reasons optimisation details will not be discussed in this paper.

In this work, we study an nD probabilistic shaping by implementing the autoencoder-based system. End-to-end learning is used to optimise the occurrence probability distribution of the nD signals, for different signal-to-noise ratio (SNR) values, ranging from 5 dB to 20 dB. Specifically, we simulate the proposed scheme for probabilistically shaped 4D-256-QAM (PS-4D-256-QAM) with an additive white Gaussian noise (AWGN) channel. The optimised autoencoder-based probabilistic shaping scheme was then tested. The system performance is evaluated in terms of BER and mutual information. We demonstrate simulated results showing an improved error performance and enhanced mutual information relative to the 4D-256-QAM and 2D-16-QAM.

The rest of the paper is organised as follows: section II describes the basics of multidimensional modulation scheme and the 4D constellations used in this work; section III details the proposed autoencoder architecture design for nD probabilistic shaping and addresses the loss function and the key merits of the end-to-end learning process; then in section IV, simulation results are discussed to show the system performance; finally, section V concludes the paper.

II. MULTIDIMENSIONAL MODULATION BASICS

The multidimensional constellation in this work is based on a subset of an infinite nD lattice Λ . In particular, we adopt a 4D constellation proposed in [12] in 1984, which selects a sublattice of 256 points in 4D space. This scheme proposed by Gersho and Lawrence provides a more densely packed constellation in multidimensional space for a given

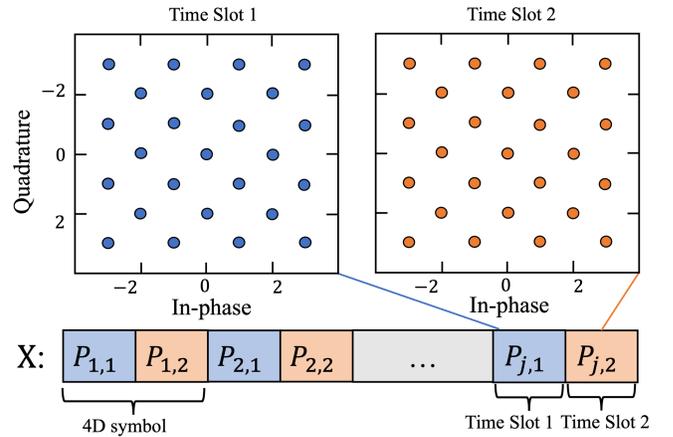


Fig. 1. Constellation diagram of 4D-256-QAM projected on 2D complex plane. A 4D symbol consists of two successive 2D complex symbols, for example, $P_{j,1} = x_{j,1} + jx_{j,2}$ and $P_{j,2} = x_{j,3} + jx_{j,4}$

TABLE I
SEVEN LEAD POINTS FOR 4D-256-QAM CONSTELLATIONS (FROM [12])

Coordinates	Quantity	Power	No. of Neighbours
1 1 1 1	16	4	23
2 0 0 0	8	4	22
2 2 0 0	24	8	22
2 2 2 0	32	12	19
2 2 2 2	16	16	15
3 1 1 1	64	12	15
3 3 1 1	96	20	8

minimum Euclidean distance. Equivalently, the minimum Euclidean distance of the 4D constellation can be increased with respect to regular 2D QAM when the same number of bits per symbol is achieved. Advantageously, the 4D signal requires less power to reach a certain BER when compared to the 2D QAM signal. The 4D QAM modulation scheme employed here is formed by using a pair of 2D complex symbols, which are transmitted in two successive signalling intervals, as shown in Fig.1. In a 4D-256-QAM scheme, each 4D symbol encodes 8 bits and hence it achieves 8 bits per 4D symbol, which is equivalent to 4 bits per 2D symbol of 2D-16-QAM. The I-Q coordinates of the 4D signal can be given by a 4D vector $\mathbf{x}_j = [x_{j,1}, x_{j,2}, x_{j,3}, x_{j,4}] \in \mathbb{Z}^4$, wherein the coordinates are either all even or all odd integers. Fig.1 shows that the 2D projection diagram of 4D-256-QAM signals is identical for the two signalling time slots, consisting of 25 constellation points each.

The total 256 4D symbol points are generated by a process of permutation and manipulation of a set of coordinates, termed lead points, following the system of [12]. Table I lists the coordinates of the 7 lead points and the cardinality of their associated subsets. The 4D symbols drawn from the sublattice have 8 to 23 nearest neighbours, rendering an average number of neighbours $\bar{N}_n = 14.25$, which is smaller than the 24

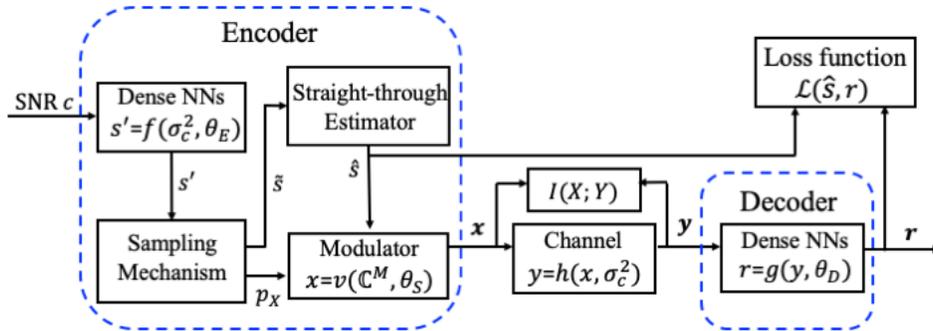


Fig. 2. Block diagram of autoencoder based probabilistic shaping system

possible nearest neighbours in this sublattice. The reduced number of neighbours of 4D-256-QAM leads to a reduced upper bound of the symbol error rate (SER).

The 4D-256-QAM design in [12] is based on a special bit-to-symbol mapping rule as well as an associated detection method. In this work, we first reproduce the 4D modulation scheme, with its original mapping as a base line. We then aim to optimise the bit-to-symbol mapping to minimise the BER of the 4D-256-QAM constellation. This work, for the first time, optimises the aforementioned 4D constellation by a metaheuristic optimisation method based on the RTS algorithm [23]. Specifically, the optimisation objective involves the Euclidean distance of the multidimensional symbols and the Hamming distance between the corresponding bits to minimise the upper bound of the achievable BER. The optimum bit mapping found by the RTS algorithm is then used in the proposed probabilistic shaping scheme. Extensive simulations and numerical results confirm the efficacy of the optimised mapping in improving the error performance. To improve further the achievable data rate and the BER performance by optimising the 4D signal design, we apply multidimensional probabilistic shaping to the 4D constellation using a specially designed autoencoder, which is detailed in the next section.

III. AUTOENCODER DESIGN FOR 4D PROBABILISTIC SHAPING OPTIMISATION

A typical communication system can be interpreted as an autoencoder, of which the encoder and the decoder are jointly optimised to perform the same tasks as traditional transceivers. In general, the goal of an autoencoder is to generate optimal representations of the input data, pass these through the channel and reconstruct the initial data from the contaminated channel outputs. In the case of nD probabilistic shaping, the autoencoder-based system is designed to optimise the probability distribution of the nD constellation symbols by training in an end-to-end manner.

As shown in Fig.2, the encoder outputs the nD symbols $\mathbf{x} \in X = \mathbb{C}^M$ with an occurrence probability $p_X(x)$. In the encoder, an implicit selection procedure is conducted such that the symbol probability distribution is applied to the finite set X of M constellation points in nD space. The

Gumbel-Softmax trick [21] is used as the sampling mechanism to approximate the discrete probability distribution $p_X(x)$ using the continuous, differentiable softmax function. An n -dimensional approximation vector is employed in this work.

Our optimisation method is set to find the optimum probability distribution of the transmitted symbols. This is based on training the autoencoder at different received SNR values and then choosing the optimum probability distribution with the best error performance. To generate the discrete probability distribution $p_X(x)$, a softmax activation function is applied to the encoder neural network outputs (i.e., the unscaled logits s'). The aforementioned Gumbel-Softmax trick is applied to obtain the approximation continuum of $p_X(x)$, which is used for the end-to-end learning. While the true one-hot representation $\hat{\mathbf{s}} = [0, 0, \dots, 1, \dots, 0]$ of the discrete distribution is used to perform the selection among the nD constellation points. More specifically, the normalised constellation \mathbf{C}^M with an unity energy expectation is multiplied by the one-hot representation $\hat{\mathbf{s}}$ to give the output nD symbol, which is passed to the channel.

The channel can be described by a set of layers embedded in an autoencoder structure. In this work, AWGN is considered and the Gaussian noise layer simply performs $\mathbf{y} = \mathbf{x} + \mathbf{w}$, wherein for $w_i \in \mathbf{w}, w_i \sim \mathcal{N}(0, \sigma_c^2)$, $i = 1, \dots, n$ and σ_c^2 denotes the noise variance determined by the given SNR c for the training. Subsequently, the noisy symbols \mathbf{y} are sent to the decoder of the autoencoder, which carries out classification to reconstruct the input message. The decoder neural network process the impaired data with several dense layers followed by a softmax layer. The softmax function outputs a probability vector that essentially indicates the predicted probability of each element that has been sent as in the input message.

The end-to-end training of the autoencoder based system employs stochastic gradient descent (SGD). To simplify the learning of the autoencoder, a set of transformations can be used to describe the general processes:

$$\begin{aligned}
 \mathbf{p}_X, \hat{\mathbf{s}} &= f(\sigma_c^2, \theta_E), \\
 \mathbf{x} &= v(\mathbf{p}_X, \hat{\mathbf{s}}, \theta_S), \\
 \mathbf{y} &= h(\mathbf{x}, \sigma_c^2), \\
 \mathbf{r} &= g(\mathbf{y}, \theta_D),
 \end{aligned} \tag{1}$$

where $f(\cdot)$ and $v(\cdot)$ construct the encoder neural network, $h(\cdot)$ denotes the channel layer and $g(\cdot)$ denotes the decoder neural network. θ_E and θ_S are the trainable parameters of the neural network for the encoder and θ_D is for the decoder. In particular, θ_S is the parameter for the sampling mechanism. The loss function \mathcal{L} measures how correct the decoder reproduces the input. The categorical cross entropy is used to construct the loss function, as discussed in the section below.

A. Categorical Cross Entropy

Neural networks that perform classification tasks usually adopt one-hot vectors for labeling and probability distribution vectors as their classification outcomes. Categorical cross entropy is commonly used as the loss function for classifications, which assesses the difference between two probability distributions over the same set of multiple variables, namely the target labels and the output categories. It is defined as the average number of bits required to encode data with probability p to the codewords with probability q . Based on the concept, the categorical cross entropy between p and q is given by

$$\mathbb{H}(p, q) = \mathbb{E}_p[-\log(q)] = - \sum_{x \in \mathcal{X}} p(x) \log(q(x)), \quad (2)$$

where $\mathbb{E}[\cdot]$ denotes the expectation operation. The cross entropy can also be derived by a summation of the entropy of p and the Kullback-Leibler (KL) divergence D_{KL} as:

$$\mathbb{H}(p, q) = \mathbb{H}(p) + D_{\text{KL}}(p||q), \quad (3)$$

where the entropy $\mathbb{H}(p) = \mathbb{E}_p[-\log(p)]$ denotes the average amount of bits that are needed to represent the symbols with probability distribution p and D_{KL} denotes the difference between the two probability distributions.

Recall the objective of the end-to-end learning with autoencoder for probabilistic shaping in this work, the encoder learns the latent representations that are robust to the channel impairments and the decoder learns to reconstruct the input symbols. Herein, the representations are the probability distribution of the transmitted symbols and the decoder performs classification task to recover the original constellation symbols from the noisy data. The two neural networks in the encoder and the decoder blocks are trained jointly to optimise the reconstruction. Therefore, the loss function can be derived from the difference between the transmitted symbol \mathbf{x} and the recovered symbol \mathbf{r} . To this end, we develop the loss function based on the derivation in [20] as:

$$\begin{aligned} \mathcal{L}_{\mathbf{x}, \mathbf{r}}(\theta_E, \theta_D) &= \mathbb{E}_{\mathbf{x}, \mathbf{y}}[-\log(P(\mathbf{r}|\mathbf{y}, \theta_D))] - \mathbb{H}(\mathbf{X}) \\ &= \mathbb{E}_{\mathbf{y}}[D_{\text{KL}}(P(\mathbf{r}|\mathbf{y}, \theta_D)||P(\mathbf{x}|\mathbf{y}))] - I(\mathbf{X}, \mathbf{Y}), \end{aligned} \quad (4)$$

where the $p(\mathbf{r}|\mathbf{y}, \theta_D)$ is the empirical probability density function (PDF) of \mathbf{r} given \mathbf{y} , which estimates the true posterior PDF $p(\mathbf{x}|\mathbf{y})$, $\mathbf{x}, \mathbf{y} \sim P_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y})$ and $\mathbb{H}(\mathbf{X})$ is the entropy of the transmitted symbol. This can be further derive to the expectation of Kullback-Leibler (KL) divergence of the true posterior PDF and the empirical PDF, subtracting the mutual

information $I(\mathbf{X}; \mathbf{Y})$ between the transmitted and received symbols, as discussed in the next section. This means that the mutual information is maximised while minimising the loss function \mathcal{L} in the training.

B. Mutual Information

Achievable information rates (AIR) is an asymptotic metric that measures the number of bits per symbol that can be reliably transmitted through a channel. A typical AIR, the mutual information $I(\mathbf{X}; \mathbf{Y})$, is commonly used to provide predictions of maximum throughput for fair comparisons among different modulation schemes for a given channel. More specifically, the channel capacity can be obtained by maximising the the mutual information over the PDF of the transmitted symbol \mathbf{X} (i.e. $P_{\mathbf{X}}(x)$). Such maximisation of the channel capacity can be expressed as:

$$C = \max_{P_{\mathbf{X}}(x)} I(\mathbf{X}; \mathbf{Y}). \quad (5)$$

The mutual information is defined as:

$$I(\mathbf{X}; \mathbf{Y}) = \mathbb{E}[\log \frac{P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})}{P_{\mathbf{Y}}(\mathbf{y})}], \quad (6)$$

where $P_{\mathbf{Y}}(\mathbf{y})$ is the marginal entropy of \mathbf{Y} and $P_{\mathbf{Y}|\mathbf{X}}$ is a conditional PDF that is used to characterise the channel. The channel transition function of an n D AWGN channel is given by an n -dimensional Gaussian distribution

$$P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp(-\frac{1}{2}(\mathbf{y} - \mathbf{x})^T \Sigma^{-1}(\mathbf{y} - \mathbf{x})), \quad (7)$$

where Σ denotes the covariance matrix. For i.i.d Gaussian distribution, Σ is equivalent to the product of an $n \times n$ identity matrix and the 1D noise variance σ_{1d}^2 . Hence the determinant of such covariance matrix $|\Sigma| = \sigma_{1d}^{2n}$. Therefore, the transition function for the n -dimensional AWGN channel can be rewritten as:

$$P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi\sigma_{1d}^2)^{n/2}} \exp(-\frac{\|\mathbf{y} - \mathbf{x}\|^2}{2\sigma_{1d}^2}). \quad (8)$$

Applying equation (8) to equation (6) derives the mutual information $I(\mathbf{X}; \mathbf{Y})$. This is used to evaluate the mutual information of the 4D signal with probabilistic shaping achieved by the proposed autoencoder-based system.

IV. SYSTEM PERFORMANCE AND EVALUATION

This section presents the performance of the proposed autoencoder-based system for multidimensional probabilistic shaping for the 4D-256-QAM signal with end-to-end learning. Tensorflow framework [24] is used to implement the training. In terms of the neural networks in the autoencoder, we employ one dense layer of 128 nodes with ReLU activation function followed by a linear layer in the encoder. For the decoder, three dense layers are used; the first two layers consist of 128 nodes with ReLU activation function; the last employs softmax function to output the probability vector. We utilise the Adam optimiser [25] for the learning of the autoencoder. The i.i.d.

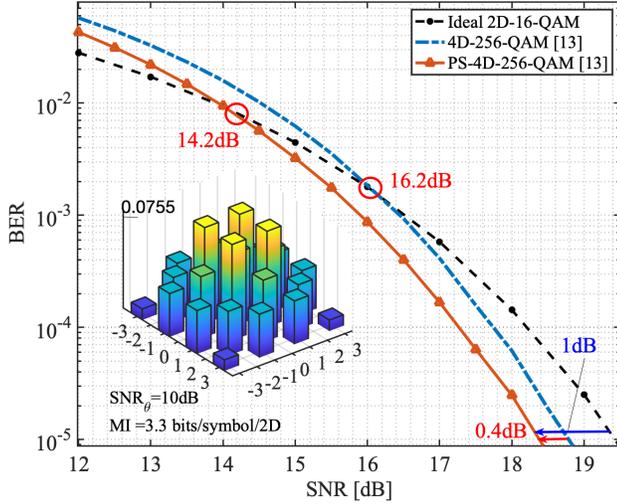


Fig. 3. BER performance comparison for the 4D signals with/without probabilistic shaping using bit mapping from [12]

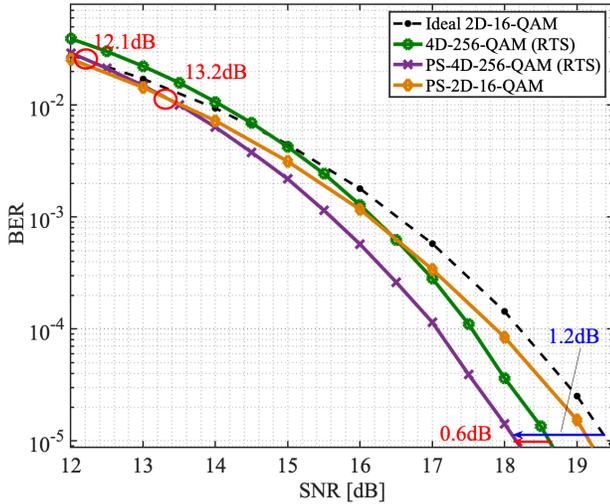


Fig. 4. BER performance comparison for the 4D and 2D signals with/without probabilistic shaping. The bit labelling of the 4D QAM is obtained by optimisation using the RTS algorithm

n -dimensional Gaussian noise channel is employed for the embedded channel in the autoencoder architecture as discussed in the section.III B. The results of the PS-4D-256-QAM are presented in this section including the BER performance and the mutual information.

A. BER Performance

To verify numerically the designs, the BER of 2D-16-QAM and 4D-256-QAM are used for comparison. It is worth noting that both the bit-to-symbol mapping from [12] and the optimised mapping, found by RTS algorithm in this work, are tested for the proposed scheme for the 4D-256-QAM. The probability distribution of the PS-4D-256-QAM constellation

was optimised by the end-to-end learning procedure with the autoencoder structure using the original and the optimised bit mapping are given by Fig.3 and Fig.4, respectively. Such optimisation requires certain SNR value to be fed into the model for each training. Therefore, varying values of SNR have been tested and SNR = 10 dB was chosen for its optimality with the corresponding mutual information of 3.3 bits per 2D symbol.

Figure 3 depicts the comparison of the BER performance of the PS-4D-256-QAM with respect to that of unshaped 2D-16-QAM and 4D-256-QAM. The 4D-256-QAM was proved in [12] to have better BER performance in high-SNR regime relative to the 2D-16-QAM, which has the same mutual information of 4 bits/symbol/1D. Fig.3 shows that the PS-4D-256-QAM given by autoencoder outperforms the 2D-16-QAM when the SNR exceeds 14.2 dB. An extra shaping gain of 0.4 dB can be achieved for the probabilistically shaped 4D QAM relative to the unshaped 4D QAM, leading to an overall power reduction of 1 dB at BER = 10^{-5} .

Figure 4 shows the power advantage of PS-4D-256-QAM, resulting from the combined effects of optimisation of bit mapping and probabilistic shaping, over 4D and 2D signals, including probabilistically shaped 2D signal similar to that discussed recently in [20] for fair comparison. The results show that the use of optimised bit mapping by RTS leads to a better error performance for 4D-256-QAM such that 1.2 dB power advantage can be achieved for PS-4D-256-QAM when compared to the 2D-16-QAM. While even compared to the PS-2D-16-QAM signal optimised by autoencoder, there is still approximately 1 dB advantage achieved by the PS-4D-256-QAM. Besides, the crossing points of the BER performance shows that the optimal the PS-4D-256-QAM with both optimised bit mapping and optimised probabilistic shaping outperforms the PS-2D-16-QAM and the unshaped 2D-16-QAM when the SNR exceeds 13.2 dB and 12.1 dB as the corresponding BER is above 10^{-2} .

B. Mutual Information Performance

We assess the mutual information $I(X;Y)$ of the 4D QAM with optimised probabilistic shaping from the proposed scheme regarding to other schemes. Since the bit-to-symbol mapping does not affect the mutual information of a modulation based on its definition in equation (6). The mutual information of probabilistically shaped QAM signals are analysed while the SNR varies for a range from 5 dB to 20 dB. Fig.5 shows significant gain achieved by the probabilistically shaped 4D QAM signal when compared to the 4D-256-QAM and the 2D-16-QAM. The unshaped 4D is shown to achieve higher mutual information relative to the PS-2D-16-QAM generated by autoencoder in high SNR regime. Both probabilistic shaping and 4D constellation design are shown to achieve enhanced achievable capacity when compared to the regular 2D QAM. It can be observed that the mutual information of the PS-4D-256-QAM approaches Shannon's limit. In particular, the shaped 4D signal is shown to have enhanced capacity in the low SNR regime where the

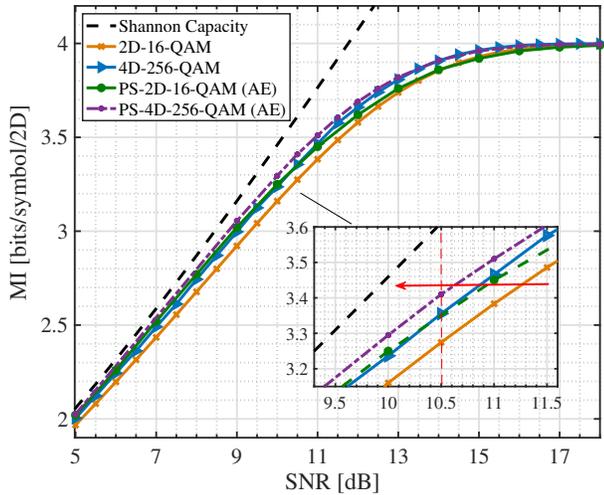


Fig. 5. Mutual information versus SNR, with/without probabilistic shaping with autoencoder (AE)

mutual information is below the benchmark of the maximum achievable data rate for the modulation scheme concerned.

V. CONCLUSIONS

This work proposes, for the first time, the use of autoencoder-based end-to-end learning to optimise the probabilistic shaping for multidimensional signals. The autoencoder-based system is implemented to optimise the probabilistic shaping for a 4D QAM signal, thereby concurrently minimising the transmission error rate and maximising the mutual information. The work also reports the use of a specially designed optimisation algorithm, based on reactive tabu search, to optimise the bit-to-symbol mapping for 4D QAM signals. Simulation results show that the probabilistically shaped 4D signal, optimised by the proposed scheme, achieves better BER performance when compared to the unshaped 4D and 2D signals, and to the probabilistically shaped 2D signal. The mutual information of the 4D signal demonstrates its enhanced capacity relative to the 2D-16-QAM by applying the multidimensional probabilistic shaping.

ACKNOWLEDGEMENT

Xinyue Liu is grateful to NEC Laboratories Europe GmbH for part funding of her PhD studies through the NEC Laboratories Europe GmbH internship funding.

REFERENCES

- [1] G. Forney, R. Gallager, G. Lang, F. Longstaff, and S. Qureshi, "Efficient modulation for band-limited channels," *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 5, pp. 632–647, 1984.
- [2] G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 4651–4665, 2015.
- [3] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 430–434, 2016.

- [4] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons, "Multiset-partition distribution matching," *IEEE Transactions on Communications*, vol. 67, no. 3, pp. 1885–1893, 2019.
- [5] T. Yoshida, M. Karlsson, and E. Agrell, "Hierarchical distribution matching for probabilistically shaped coded modulation," *Journal of Lightwave Technology*, vol. 37, no. 6, pp. 1579–1589, 2019.
- [6] Y. C. Gültekin, W. J. van Houtum, A. G. C. Koppelaar, F. M. J. Willems, and W. J. van Houtum, "Enumerative sphere shaping for wireless communications with short packets," *IEEE Transactions on Wireless Communications*, vol. 19, no. 2, pp. 1098–1112, 2020.
- [7] O. İşcan, R. Böhnke, and W. Xu, "Probabilistically shaped multi-level coding with polar codes for fading channels," in *2018 IEEE Globecom Workshops (GC Wkshps)*, 2018, pp. 1–5.
- [8] J. Renner, T. Fehenberger, M. P. Yankov, F. Da Ros, S. Forchhammer, G. Böcherer, and N. Hanik, "Experimental comparison of probabilistic shaping methods for unrepeated fiber transmission," *Journal of Lightwave Technology*, vol. 35, no. 22, pp. 4871–4879, 2017.
- [9] G. Böcherer, P. Schulte, and F. Steiner, "Probabilistic shaping and forward error correction for fiber-optic communication systems," *Journal of Lightwave Technology*, vol. 37, no. 2, pp. 230–244, 2019.
- [10] J. Cho and P. J. Winzer, "Probabilistic constellation shaping for optical fiber communications," *Journal of Lightwave Technology*, vol. 37, no. 6, pp. 1590–1607, 2019.
- [11] Z. Qu, S. Zhang, and I. B. Djordjevic, "Universal hybrid probabilistic-geometric shaping based on two-dimensional distribution matchers," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, 2018, pp. 1–3.
- [12] A. Gersho and V. Lawrence, "Multidimensional signal constellations for voiceband data transmission," *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 5, pp. 687–702, 1984.
- [13] E. Agrell and M. Karlsson, "Power-efficient modulation formats in coherent transmission systems," *Journal of Lightwave Technology*, vol. 27, no. 22, pp. 5115–5126, 2009.
- [14] R. Laroia, N. Farvardin, and S. Tretter, "On optimal shaping of multidimensional constellations," *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1044–1056, 1994.
- [15] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "On shaping gain in the nonlinear fiber-optic channel," in *2014 IEEE International Symposium on Information Theory*, 2014, pp. 2794–2798.
- [16] P. Schulte, F. Steiner, and G. Böcherer, "Four dimensional probabilistic shaping for fiber-optic communication," in *Signal Processing in Photonic Communications*. Optical Society of America, 2017.
- [17] Q. Guo, W.-R. Peng, Y. Cui, and Y. Bai, "Multi-dimensional probabilistic shaping for higher fibre nonlinearity tolerance," in *45th European Conference on Optical Communication (ECOC 2019)*, 2019, pp. 1–3.
- [18] A. Amari and A. Richter, "Multi-dimensional short blocklength probabilistic shaping for digital subcarrier multiplexing systems," *arXiv preprint arXiv:2105.04956*, 2021.
- [19] 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, 2017.
- [20] M. Stark, F. Ait Aoudia, and J. Hoydis, "Joint learning of geometric and probabilistic constellation shaping," in *2019 IEEE Globecom Workshops (GC Wkshps)*, 2019, pp. 1–6.
- [21] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with gumbel-softmax," *arXiv preprint arXiv:1611.01144*, 2016.
- [22] 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*. IEEE, 2020, pp. 1–6.
- [23] R. Battiti and G. Tecchiolli, "The reactive tabu search," *ORSA journal on computing*, vol. 6, no. 2, pp. 126–140, 1994.
- [24] "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. [Online]. Available: <https://www.tensorflow.org/>
- [25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.