Deep Learning Constellation Design for the AWGN Channel with Additive Radar Interference

Florence Alberge

Deep Learning Constellation Design for the AWGN Channel with Additive Radar Interference

Florence Alberge
Laboratoire des Signaux et Systèmes (L2S), Université Paris-Sud-CNRS-CentraleSupélec, Université Paris-Saclay,
3, rue Joliot Curie, 91192, Gif-sur-Yvette, France
e-mail: florence.alberge@u-psud.fr

***Abstract*—Radar and wireless communication coexistence is considered in this paper as a possible solution to face the exploding demand and rising congestion in wireless networks. The transmission medium is modeled as an AWGN channel with additive radar interference. Standard constellations are not optimal in this context and an auto-encoder (AE) is used to design proper constellations and corresponding receiver devices. AE is a powerful tool in neural networks that shares strong similarities with communication systems. This technique is particularly relevant in the lack of an analytical expression of the loss function. In the asymptotic region (high interference regime), the optimal constellation shape is known and the AE always converges towards this optimal solution. In the other regions, the AE is able to yield solutions that outperform the standard configurations. Several demapping alternatives are also considered leading to the conclusion that it is possible to maintain the communication link in the presence of radar interference independently of the interference power. This is a step further compared to previous works in which solutions were limited to low or high interference regimes.**

***Index Terms*—Machine learning, optimization methods, digital modulation, radar interference.**

I. Introduction

Increasing data traffic demand in cellular and wireless networks leads to a rethinking of the spectrum allocation policy. The actual spectrum regulation with bands allocated to specific users or services is suboptimal and a significant part of the spectrum is seldom used. Indeed radars and radio navigation infrastructure occupy an important portion of the available spectrum with low usage efficiency. Several regulators across the world have started considerations for release portions of governmental radar bands to be shared with commercial wireless services. Consequently, the concept of spectrum sharing in radar has gained lots of interest [1], [2], [3]. Literature on radar and wireless communication coexistence is vast. It is possible however to design the radar and/or the communication system to alleviate its impact on the other device by using different techniques such as waveform design [4], [5], interference mitigation [6] or spatial separation [7]. In [8], a joint design of the radar transmission waveform's power spectrum and the power spectral density of a multi-carrier communication system is proposed. As a complementary approach, it is also interesting to consider an unaltered radar signal and a communication system and in particular it is useful to have an idea of the attainable performance when both systems coexist. The performance of standard communication systems in the presence of radar interference has been studied through numerical experiments in several publications. For example, WiMax and ground based radar are considered in [9] and an OFDM system subject to ultra-wide band radar interference is considered in [10]. A theoretical performance analysis is given in [11] where bounds on performance of the joint system are measured in terms of data information rate for communications and radar estimation information rate for radar. The capacity of the communication channels with both additive white Gaussian noise and radar interference is investigated in [12]. It is proved that communication is possible even in the presence of high radar interference. This latter case results in a loss of half the degrees of freedom compared to a channel without interference. The capacity achieving input distribution under average power constraint is characterized in [12]. The optimal input distribution has independent modulus and phase. The phase is uniformly distributed in $[0, 2\pi]$ whereas the modulus is discrete with countably infinite many mass points but only finitely many in any bounded interval. Finite inputs are used in practice whereas the optimal input is continuous (Gaussian distributed). The objective of this paper is to design finite constellations for the additive radar interference channel. To the best of our knowledge, this problem has not been widely investigated in the past. Very recently, this question was addressed in [13]. The objective function considered for the minimization process is the symbol error rate which is based on the analytical expressions given in [14] for the weak and for the very strong interference regime leading to very interesting results for these two situations. The method in [13] cannot be extended to other regimes since closed form expressions are not known in the general case. Instead, it is proposed here to use deep learning and AEs for constellation design. With deep learning, the network is able to learn directly from the observations and closed-form expressions are no longer required.

Deep learning is a branch of machine learning research which has been introduced to allow machines and computer systems to improve with experience and data accumulation. An auto-encoder (AE) is a type of artificial neural network used to learn efficient representations of the data in an unsupervised manner. The analogy between AEs and communication systems has been used in recent publications [15]. More generally, deep learning techniques have already been used for solving communication problems. The vast majority is dedicated to the design of the receiver. For example, multi-user detection in

code-division multiple access (CDMA) has been investigated in [16], [17]. The ability of a neural network to decode polar codes with MAP performance is analyzed in [18]. Neural network techniques demonstrate convincing results for short block length codes but suffer from high decoding complexity. Indeed, the training complexity of deep learning-based channel decoders scales exponentially with the number of information bits and cannot be reasonably considered for standard codes used in practice. In [19], the emitter and decoder are considered jointly. Here again the complexity limits the usefulness of the method since the proposed method can only be applied to short block length.

Taking into consideration those works, we believe that AEs are well-suited for constellation design. Indeed, the dimensionality of this problem is modest and the encoder of the AE can be trained to optimize the constellation whereas the decoder is trained to recover the binary message from the received data. Our objective is not end-to-end learning of communications systems through deep neural network like in [19]. Here, the neural network is considered only for modulation and demodulation tasks. The proposed solution is compatible with other elements of the communication chain such as error-correcting codes. The strengths of the proposed methodology are listed below. The objective function to be optimized is a bound of the mutual information [20] which is an usual criterion in communication and in particular in constellation design. Neither an analytical expression of the mutual information (or of the symbol error rate) nor an *a priori* on the geometry of the constellation are needed for the optimization process. This is in contrast with previous works [21], [22], [23], [13]. The constellation geometry and the labeling are optimized jointly. The proposed mapping/demapping solution can be used in combination with powerful channel decoders such as turbo-codes or Low Density Parity-Check codes (LDPC) and is fully compatible with wireless communications schemes already deployed.

The paper is organized as follows. The system model is depicted in Section II. Past work related to this model is recalled and our contribution detailed. The strong interference regime is considered in section III. It is proved that, in this specific regime, an M-PAM like constellation should be used. AEs are discussed in Section IV-A and the proposed structure is presented. Section V is devoted to constellation design. Numerical results are given in Section VI where the demapping operation is discussed in order to exhibit the best association for each interference regime.

II. System model

The system under consideration is composed of a narrow-band communication channel suffering from the interference of a high power, short duty cycle radar pulse. An analytical model for the discrete-time complex-valued signal at the communication receiver is given in [24]. The channel model is an AWGN with additive interference, with known constant amplitude $\sqrt{I}$ and unknown random phase θ_I uniformly distributed in $[0, 2\pi]$. A justification for this model can be found in [25] and references therein. From the communication perspective, a single carrier system is considered with narrow-band frequency-flat slowly-varying fading model with known channel gain at the receiver. Extension to time-varying channels is obtained by averaging the error rates over the channel statistics. The amplitude of the radar interference can be accurately estimated since the radar system has slow-varying parameters [25]. At the opposite, the phase suffers rapid variations and is difficult to track. Let's denote the discrete-time signal at the communication receiver as Y then

$$Y = X + \sqrt{I}e^{j\theta_I} + Z \tag{1}$$

where X is the channel input chosen from the constellation $\mathcal{X} = \{X_i = \sqrt{S_i}e^{j\theta_i}, i = 0, ..., M-1\}$ with cardinality M and Z is a complex Gaussian noise with zero-mean and unit-variance. The average power of the constellation is S meaning that $\mathbb{E}[|X|^2] = S$. The random variables (X, θ_I, Z) are mutually independent and the channel is memoryless. From these notations, the average Signal to Noise Ratio (SNR) is S and the average Interference to Noise Ratio (INR) is I. The objective of this paper is to optimize the constellation alphabet $\mathcal{X}$ for reliable transmission through the AWGN channel with additive interference given in (1). Both S and I are fixed and known at the receiver. Three different regimes will be considered here. For each one, the state of the art is first recalled and emphasis is given on our contribution.

1) $I \gg S$. In the high interference regime, it is optimal at the receiver to estimate the phase of the radar interference as the phase of the received signal and to use this estimation to subtract its contribution from the received signal [24]. In this interference cancellation process, a portion of the useful signal is also subtracted and the capacity is only half the capacity of the interference-free system. More specifically, it is proved in [12] that
$$\lim_{I\to\infty} C(S, I) = \frac{1}{2}\log(1 + S) \tag{2}$$
where $C(S, I)$ denotes the capacity of the channel in (1). The Gaussian input is optimal in the high interference regime. However, in practice, the channel input is constrained to a finite size alphabet.
2) $I \ll S$. The Gaussian input performs also very well (but is not optimal) in the weak interference regime with achievable rates close to the upper bound in [26]. The capacity achieving distribution under average power constraint has independent modulus and phase [12]. The phase is uniformly distributed in $[0, 2\pi]$ and the modulus is discrete with countably infinitely many mass points, but only finitely many in any bounded interval. It can also be seen in [14] that, in this regime, the optimal MAP decoder is the classical minimal distance decoder applied to (1) in which the interference is considered as Gaussian noise.
3) The intermediate regime is the most challenging case. The interference is neither small enough to be neglected nor large enough to be accurately estimated. The expression of the channel capacity does not admit a simple expression like in the other two cases.

Our contributions are given below. The main contribution is the design of finite size constellations with enhanced perfor-

mance compared to standard constellations and the design of the corresponding soft demapping rules. The constellation as well as the soft demapping are optimized with an AE. Thanks to this optimization framework, the constellation shape, the constellation labeling[1] and the soft demapping device are optimized jointly. This in contrast with [13] in which the constellation labeling is not addressed. Indeed, in [13] the symbol error rate (SER) is considered in the optimization problem whereas here the loss function is based on bit probabilities. This is a key difference that will be discussed in section V-C. Note also that [13] focuses on the low and high INR regime whereas here solutions are proposed for the three regimes. The resulting constellations are evaluated in terms of achievable rates and/or error probability and compared to theoretical bounds and also to standard constellations. The second contribution is to show that the designed schemes are efficient both with uncoded and coded data. Indeed the AE outputs soft information which can be easily connected to a turbo or LDPC decoder. To the best of our knowledge, previous works considered uncoded data only. However, the gain from the constellation design may be absorbed by the channel coding. It can be seen here that the proposed schemes outperform the standard ones even when coded data are considered. Finally, a new demapping rule (called DMHighSNR in the following) well adapted to the additive radar interference channel is introduced here. This new rule outperforms the existing ones in the three regimes.

An analysis of the interference channel in the high INR regime is provided in the next section. It is proved that the constellation points should be collinear to avoid equivocal areas. This is consistent with the expression of the capacity of the real AWGN channel with power constraint given in (2).

III. Interference cancellation in the high INR regime

In the high INR regime ($I \gg S$), $Y \approx \sqrt{I}e^{j\theta_I}$ since the other terms in (1) become negligible when I goes to infinity. The phase of the radar interference can thus be estimated as $\hat{\theta}_I = arg(Y)$ and $\sqrt{I}e^{j\hat{\theta}_I} = \sqrt{I}\frac{Y}{|Y|}$ is an estimation of the radar interference. The accuracy of this estimation depends on the SNR value. Let's define Y_{IC} as

$$Y_{IC} = Y - \sqrt{I}e^{j\hat{\theta}_I} = Y\left(1 - \frac{\sqrt{I}}{|Y|}\right) \tag{3}$$

where $|Y|$ stands for the modulus of the complex variable Y. In the whole paper, X stands for the channel input, Y is the channel output in (1) and Y_{IC} is the channel output after interference estimation and cancellation given in (3).

Result 1: Without loss of generality, $X = X_0 = \sqrt{S_0}e^{j\theta_0}$ where X_0 is a given point of the constellation $\mathcal{X}$. If $SNR \to \infty$ and $I \gg S$ then

$$Y_{IC} = \frac{\sqrt{S_0}}{2}\left(e^{j\theta_0} + e^{j(2\theta_I - \theta_0)}\right) + \sqrt{S_0}o(1) \tag{4}$$

where $\sqrt{S_0}o(1)$ is the remainder term of the Taylor expansion and $o(1) \to 0$ when $\sqrt{\frac{S_0}{I}} \to 0$.

[1]A constellation labeling is the assignment of a bit pattern to each symbol in a signal-set constellation.

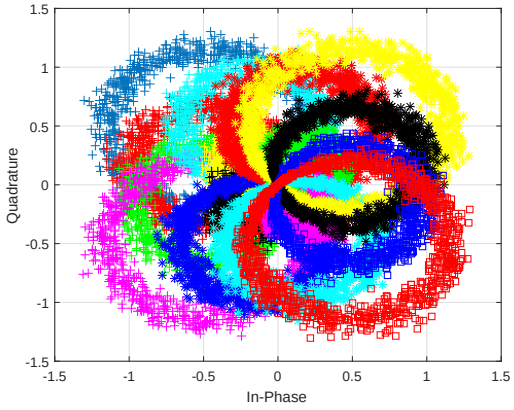

Figure 1. Locus of Y_{IC}, 16QAM constellation, $I = S^2$, $SNR = 20$ dB.

Proof: The absolute square of Y reads

$$|Y|^2 = I\left(1 + \sqrt{\frac{S_0}{I}}\left(e^{j(\theta_0 - \theta_I)} + e^{j(\theta_I - \theta_0)}\right) + \frac{S_0}{I}\right) \tag{5}$$

Since S is the average energy of the constellation and since the constellation is bounded and of finite size, $I \gg S$ leads to $I \gg S_0$ Thus, $\frac{\sqrt{I}}{|Y|}$ admits the first-order asymptotic expansion

$$\frac{\sqrt{I}}{|Y|} = 1 - \frac{1}{2}\sqrt{\frac{S_0}{I}}\left(e^{j(\theta_0 - \theta_I)} + e^{j(\theta_I - \theta_0)}\right) + \sqrt{\frac{S_0}{I}}o(1) \tag{6}$$

where $\sqrt{\frac{S_0}{I}}o(1)$ is the remainder term of the Taylor expansion and $o(1) \to 0$ when $\sqrt{\frac{S_0}{I}} \to 0$. From $Y = \sqrt{I}\left(e^{j\theta_I} + \sqrt{\frac{S_0}{I}}e^{j\theta_0}\right)$ and from (3), we obtain the first-order expansion of $\frac{Y_{IC}}{\sqrt{I}}$ as

$$\frac{Y_{IC}}{\sqrt{I}} = \frac{1}{2}\sqrt{\frac{S_0}{I}}\left(e^{j\theta_0} + e^{j(2\theta_I - \theta_0)}\right) + \sqrt{\frac{S_0}{I}}o(1) \tag{7}$$

which concludes the proof. ■

From result 1, the locus of Y_{IC} for a given transmitted symbol $X = X_0 = \sqrt{S_0}e^{j\theta_0}$ is a circle with center $\frac{\sqrt{S_0}}{2}e^{j\theta_0}$ and radius $\frac{\sqrt{S_0}}{2}$. Let's denote the point with Cartesian coordinates $(0;0)$ as O. The circles corresponding to two different constellation points overlap at least in O since each circle passes through the origin. The loci for a 16QAM constellation are depicted in Figs. 1 and 2 for $I = S^2$ and $SNR = 20$ dB and $SNR = 40$ dB respectively. We can observe several overlapping points which may complicate the de-mapping task. We prove in the next result that there exists a constellation geometry with one and only one overlapping point, namely O.

Result 2: Let's assume that $\mathcal{C}_1$ and $\mathcal{C}_2$ are two circles defined as

$$\mathcal{C}_1 = \left\{z_1 = \frac{\sqrt{S_1}}{2}\left(e^{j\theta_1} + e^{j\theta}\right), \theta \in [0; 2\pi)\right\}$$

$$\mathcal{C}_2 = \left\{z_2 = \frac{\sqrt{S_2}}{2}\left(e^{j\theta_2} + e^{j\theta}\right), \theta \in [0; 2\pi)\right\}$$

Let's define X_1 and X_2 as two distinct points of $\mathcal{X}$ with respective polar coordinates $\frac{\sqrt{S_1}}{2}e^{j\theta_1}$ and $\frac{\sqrt{S_2}}{2}e^{j\theta_2}$. Then,

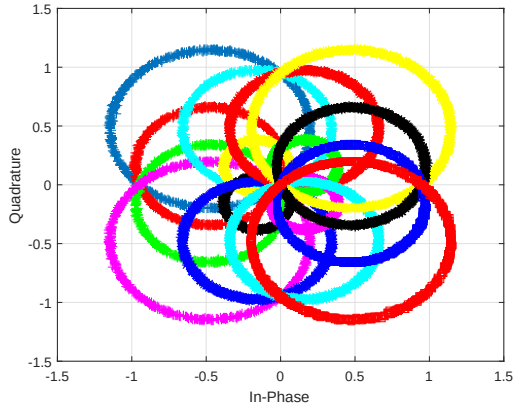

Figure 2. Locus of Y_{IC}, 16QAM constellation, $I = S^2$, $SNR = 40$ dB.

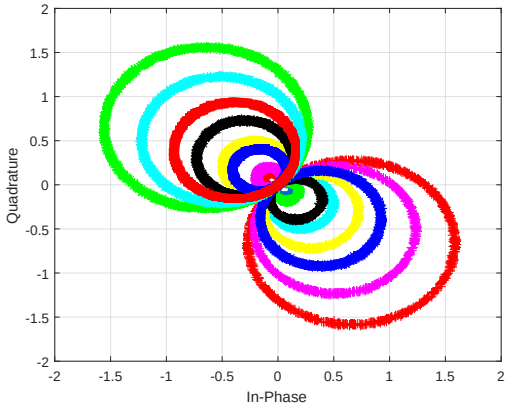

Figure 3. Locus of Y_{IC} - $|\mathcal{X}| = 16$ -$I = S^2$ - $SNR = 40$ dB

$\mathcal{C}_1 \cap \mathcal{C}_2 = \{O\}$ if and only if the three points X_1, X_2 and O are collinear.

Proof: For compactness, $(x_1; y_1)$ denote the Cartesian coordinates of X_1 and $(x_2; y_2)$ denote the Cartesian coordinates of X_2. With these notations and for $i \in \{1, 2\}$, $\mathcal{C}_i$ is a circle with center coordinates (x_i, y_i) and radius $R_i = \sqrt{x_i^2 + y_i^2}$. If M with Cartesian coordinate $(x; y)$ is a point of $\mathcal{C}_1 \cap \mathcal{C}_2$, then x and y are such that

$$x^2 - 2x_1x + y^2 - 2y_1y = 0 \quad (8)$$
$$x^2 - 2x_2x + y^2 - 2y_2y = 0 \quad (9)$$

We distinguish here the two following cases:

1) $x_1 = x_2$. Then, by subtraction of (8) and (9), either $y = 0$ or $y_1 = y_2$. Since it is assumed that $X_1 \neq X_2$, the case $\{x_1 = x_2$ and $y_1 = y_2\}$ does not require to be considered. If $y = 0$ then $x = 0$ or $x = 2x_1$ hence $|\mathcal{C}_1 \cap \mathcal{C}_2| = 2$.
2) $x_1 \neq x_2$. By subtracting (8) and (9), we have

$$x = \frac{y_2 - y_1}{x_1 - x_2} y \quad (10)$$

which can be plugged into (8) leading to

$$\left(\left(\frac{y_1 - y_2}{x_2 - x_1}\right)^2 + 1\right) y^2 = 2\left(\frac{x_2y_1 - x_1y_2}{x_2 - x_1}\right) y \quad (11)$$

As a consequence, $|\mathcal{C}_1 \cap \mathcal{C}_2| = 1$ if and only if $x_2y_1 = x_1y_2$. This condition is equivalent to

$$sin(\theta_1 - \theta_2) = 0 \quad (12)$$

Hence X_1, X_2 and O are collinear.

■

An illustration of this result is given in Fig. 3 where the constellation is composed of 16 collinear points and is symmetric with respect to O. As expected, the overlapping is limited to O. In the following, a constellation with cardinality M and such that all points are on a straight line passing through the origin is termed an M-RPAM by analogy with a rotated M-PAM. Note however that regular spacing between the constellation points is not required. It is proved in the next result that perfect decoding is possible even when $Y_{IC} = 0$.

Result 3: Let's consider the high INR regime ($I \gg S$) and the noiseless case. Let's assume that $\mathcal{X}$ is an M-RPAM constellation and that $\{\theta_{\mathcal{X}}; \theta_{\mathcal{X}} + \pi\}$ denote the possible phases of the constellation points. The channel input is $X = X_0 = \sqrt{S_0}e^{j\theta_0}$ and the corresponding channel output is $Y = \sqrt{S_0}e^{j\theta_0} + \sqrt{I}e^{j\theta_I}$. If $Y_{IC} = 0$ then X_0 is recovered from the observation Y as

$$X_0 = 2\mathcal{R}e(Ye^{-j\theta_{\mathcal{X}}})e^{j\theta_{\mathcal{X}}} \quad (13)$$

Proof: From (3), $Y_{IC} = 0$ is equivalent to $Y = 0$ or $|Y| = \sqrt{I}$. In the high INR regime, $I \neq S_0$ thus $Y = \sqrt{S_0}e^{j\theta_0} + \sqrt{I}e^{j\theta_I} \neq 0$. The other alternative $|Y| = \sqrt{I}$ is equivalent to $|Y|^2 = I$. Since $|Y|^2 = S_0 + I + 2\sqrt{S_0I}cos(\theta_0 - \theta_I)$, $|Y| = \sqrt{I}$ is equivalent to $cos(\theta_0 - \theta_I) = -\frac{1}{2}\sqrt{\frac{S_0}{I}}$. The expression of $sin(\theta_0 - \theta_I)$ is inferred from $cos(\theta_0 - \theta_I)^2 + sin(\theta_0 - \theta_I)^2 = 1$ leading to $sin(\theta_0 - \theta_I) = \epsilon\sqrt{1 - \frac{S_0}{4I}}$ with $\epsilon \in \{-1; 1\}$. Replacing these two equations into the expression of the noiseless channel output leads to

$$Y = \sqrt{S_0}e^{j\theta_0}\left(\frac{1}{2} - j\epsilon\sqrt{\frac{I}{S_0}}\sqrt{1 - \frac{S_0}{4I}}\right) \quad (14)$$

Since $\theta_0 - \theta_{\mathcal{X}} = 0 \mod \pi$, we have $e^{j(\theta_0 - \theta_{\mathcal{X}})} \in \{+1, -1\}$ and we can conclude that

$$\mathcal{R}e(Ye^{-j\theta_{\mathcal{X}}}) = \frac{1}{2}\sqrt{S_0}e^{j(\theta_0 - \theta_{\mathcal{X}})} \quad (15)$$

The channel input is thus recovered from the observation Y by computing $2\mathcal{R}e(Ye^{-j\theta_{\mathcal{X}}})e^{j\theta_{\mathcal{X}}}$. ■

Hence, at high SNR, a M-RPAM constellation yields perfect decoding. Other geometric configurations may share this property. However, since collinear points decrease the number of equivocal area, it is likely to observe, with noisy data, an improvement in terms of BER with M-RPAM compared to other constellations. This point will be confirmed in the simulation part.

IV. Auto-encoders and communication systems

A. Deep learning

Deep learning is a specific subfield of machine learning. The central problem in machine learning and deep learning is to

learn useful representations of the input data from exposure to known examples of inputs and outputs. Deep learning involves several successive layers of representation. A deep neural network is obtained by stacking layers one on top of another. The neural layers considered in this article transform an input data ℓ_{in} into an output ℓ_{out} as follows

$$\ell_{out} = g(\mathbf{w}\ell_{in} + \mathbf{b}) \tag{16}$$

where $\mathbf{w}$ and $\mathbf{b}$ are called weights or trainable parameters and where $g(.)$ is either a non-linear function or the identity mapping [27]. The weights contain the information learned by the network from exposure to training data. The weights of the whole layers are optimized jointly, each layer being updated to follow both the representational needs of the layer above and the needs of the layer below [28]. The optimizer implements the back-propagation algorithm. The loss function takes the prediction of the output of the network and the true target and computes a distance score that is used as a feedback signal to adjust the weights. In this representation, ℓ_{in} and ℓ_{out} are vectors with respective length d_{in} and d_{out}, $\mathbf{w}$ is a matrix with size $d_{out} \times d_{in}$ and $\mathbf{b}$ is a vector with length d_{out}. In the following, we use the shorthand notation

$$\ell_{out} = f_{\mathcal{P}}^{d_{in}\to d_{out},g}(\ell_{in}) \tag{17}$$

where $\mathcal{P} = \{\mathbf{w}, \mathbf{b}\}$ is the set of trainable parameters.

B. Auto-encoders and proposed structure

AEs are a specific type of feedforward neural networks where the input is the same as the output. The Denoising AE is of particular interest here. The idea behind Denoising AEs is simple. In order to force the neural network to discover more robust features and prevent it from simply learning the identity, we train the AE to reconstruct the input from a corrupted version of it. The network may be viewed as consisting of two parts termed as encoder and decoder. Usually, the corruption is introduced at the input. To mimic a communication system, the perturbation is introduced here at the output of the encoder and before entering the decoder [15]. From the network perspective, the perturbation is modeled by a layer (Channel layer) that does not involve any trainable parameters. A schematic representation of the proposed structure is given in Fig. 4.

Let's denote the length of the binary message C to be transmitted as k. C is the input of the AE. The input space is of size $M = 2^k$. X is the complex symbol at the output of the encoder and is thus a vector of length 2 (real and imaginary part of the symbol). X is a representation of C that will be optimized to be robust to the perturbation introduced by the communication channel. Unfortunately, there are no formula to determine the right number of layers and the right size for each layer. The general workflow to find the appropriate network's size is to start with relative few layers and parameters and increase the size of the layers or the number of layers until reaching a minimum or a threshold in the validation loss [28]. With this methodology, the encoder (modulator) is defined as

$$X = f_{\mathcal{P}_1}^{k\to 16,t} \circ f_{\mathcal{P}_2}^{16\to 8,t} \circ f_{\mathcal{P}_3}^{8\to 4,t} \circ f_{\mathcal{P}_4}^{4\to 2,l}(C) \tag{18}$$

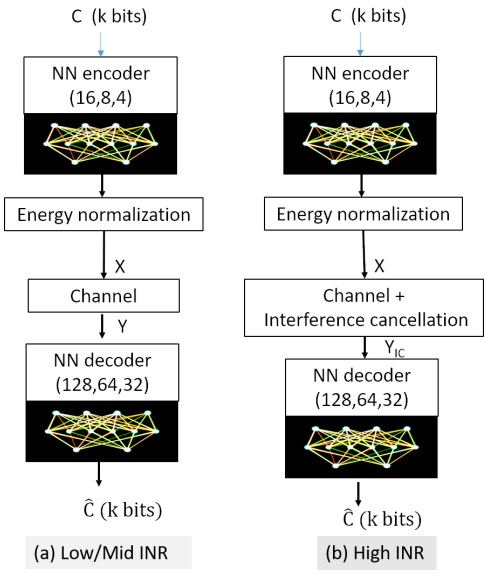

Figure 4. Schematic representation of the AE - Black pictograms indicate layers with trainable parameters.

where t and l stand respectively for the hyperbolic tangent and linear activation functions. This structure involves a moderate number of layers with small size but appears to be sufficient for performing the encoding task. The activation functions are chosen symmetric with respect to 0 since X should not be restricted to the quadrant of the complex plane with positive real part and positive imaginary part. Before entering the channel, the block of data is normalized such that the constellation has an average energy equal to S. The decoder (demodulator) is defined as

$$\hat{C} = f_{\mathcal{P}'_1}^{2\to 128,r} \circ f_{\mathcal{P}'_2}^{128\to 64,r} \circ f_{\mathcal{P}'_3}^{64\to 32,r} \circ f_{\mathcal{P}'_4}^{32\to M,s}(Y) \tag{19}$$

where r and s stand respectively for the rectified linear unit (relu) and sigmoid activation functions. The rectified linear unit is recognized as one of the most efficient activation function. The sigmoid activation function should be used in the output layer for a binary classification problem. Since $\hat{C} \in \{0,1\}^k$, the problem considered here can be seen as a binary classification problem. This decoder structure has also been used in [18] for decoding polar codes.

C. Training the network

For a given training session, the triplet (k, S, α) is fixed. For all the simulations in this paper, the number of layers and the size of each layer is also fixed to the values given in (18-19) and in Fig. 4. The training set contains all the possible values of C (binary words with length k). In an AE during the training phase, the targets values are equal to the inputs *ie* $\hat{C}_i = C_i$ where C_i is a realization of C. The output of the channel layer could be either Y (see Fig. 4-a, low/mid INR) or Y_{IC} (see Fig. 4-b, high INR). A third alternative will be discussed later on. Like in [18], the gradient of the loss function is calculated over the entire input set during each epoch. The batch size is chosen equal to 2^k, the number of epochs is 2^{19} which is a good compromise between performance and complexity. The

stochastic optimization is performed by a stochastic gradient with fixed learning rate $l_r = 0.01$. In our architecture, the channel layer generates a new realization of the noise Z and of the phase θ_I each time it is used. The network is then trained in order to optimize the reconstruction error

$$L(C, \hat{C}) = -\log p(C|\hat{C})$$

The problem under consideration here is a binary classification problem. In that case, the reconstruction error is the cross-entropy loss [20], namely

$$L(C, \hat{C}) = -\sum_j \Big(C(j) \log \hat{C}(j) + (1 - C(j)) \log (1 - \hat{C}(j)) \Big) \tag{20}$$

and $\hat{C}(j)$ is such that $\hat{C}(j) = P(C(j) = 1|\hat{C})$ where $C(j)$ stands for bit j of C and $\hat{C}(j)$ stands for bit j of $\hat{C}$. Since $\hat{C}(j) = P(C(j) = 1|\hat{C})$, connection with soft decoders such as LDPC or turbo-codes is straightforward. The training of the network is performed by solving the following optimization problem

$$\arg \min_{\mathcal{P}, \mathcal{P}'} \mathbb{E}_{C,Z,\theta_I}[L(C, \hat{C})] \tag{21}$$

When the distribution of the input is unknown, the expectation is replaced by the empirical average over the training samples [27]. It is proven in [20] that the optimization problem in (21) is strongly connected to the maximization of the mutual information $I(X; C)$ between X and C which makes sense in the present context. The optimization problem is solved with a unit-norm constraint on the columns of $\mathbf{w}$ to avoid scaling effects. The goal of training is to find the optimal set of parameters $\{\mathcal{P}_1, ...\mathcal{P}_4, \mathcal{P}'_1, ..., \mathcal{P}'_4\}$. The training of the neural network is implemented with Keras[2].

D. Using a trained network

Once trained, the optimized network can be introduced into the communication chain. Let's define the binary representation of i for $0 \leq i \leq 2^k$ as C_i. Then, $X_i = f_{\mathcal{P}_1}^{k\to 16,t} \circ f_{\mathcal{P}_2}^{16\to 8,t} \circ f_{\mathcal{P}_3}^{8\to 4,t} \circ f_{\mathcal{P}_4}^{4\to 2,l}(C_i)$. The association $C_i \leftrightarrow X_i$ can be also stored in a table. This association between a constellation point and a particular input is usually called labeling. At the receiver and when the network is the one in Fig. 4-a, the demapping task can be handled as $\hat{C}_i = f_{\mathcal{P}'_1}^{2\to 128,r} \circ f_{\mathcal{P}'_2}^{128\to 64,r} \circ f_{\mathcal{P}'_3}^{64\to 32,r} \circ f_{\mathcal{P}'_4}^{32\to M,s}(Y(X_i))$ where $Y(X_i)$ is the channel output corresponding to the input $X = X_i$. The demapping is called DMAE(Y) in that case. When the network used during the training is the one in Fig. 4-b, the demapping task can be handled as $\hat{C}_i = f_{\mathcal{P}'_1}^{2\to 128,r} \circ f_{\mathcal{P}'_2}^{128\to 64,r} \circ f_{\mathcal{P}'_3}^{64\to 32,r} \circ f_{\mathcal{P}'_4}^{32\to M,s}(Y_{IC}(X_i))$ where $Y_{IC}(X_i)$ is the channel output after interference cancellation. This demapping will be referred as DMAE(Y_{IC}). With this framework, the constellation geometry, the labeling and the demapping device are optimized jointly. This in contrast with previous publications. The proposed structure is deployed, in the next section, in various channel conditions. The AE is conceived here as an off-line optimization tool with the objective of finding the appropriate constellation and demapping device for each scenario.

[2]https://keras.io/

V. Constellation design

We are now ready to use the AE defined above. Our goal here is to evaluate the capability of the AE to find the most appropriate constellation for both the low and high INR regime in which communication should be possible without experiencing a severe degradation [12]. The capacity of the AE to cope with the severe channel impairments inherent to the intermediate regime will also be tested.

A. Constellation design at high INR

As a first step, the performance of several standard constellations are compared in terms of average symbol error rate (SER) for $SNRs$ ranging from 10 to 30 dB. The results are displayed in Fig. 5 where we used the analytical expression of the SER given in [14] for an arbitrary constellation at high INR and with an interference cancellation decoder. The numerical results in Fig. 5 illustrate perfectly the theoretical analysis of section III. For any constellation size, there exists a threshold on the SNR such that beyond this threshold the M-PAM constellation outperforms the other ones. The ability of the AE to find this optimal constellation is now under consideration. For the high INR regime, the receiver should estimate and subtract the interference. Hence, the AE is trained with Y_{IC} as channel output instead of Y. Training is performed for different values of the triplet (M, SNR, α) where α is such that $I = S^\alpha$. The values of α are chosen between 1.8 and 2.5. The corresponding SER can be found in Fig. 5. We can observe that for each triplet the AE converges towards a constellation which has either the same SER as the M-PAM (best standard constellation) or slightly better. For example, the constellation obtained with ($M = 16$, $SNR = 20$ dB, $\alpha = 2$) is shown in Fig. 8. This is a 16-RPAM[3] with non-equally spaced symbols. The SER with this constellation is 0.3499 whereas the SER of the 16-PAM at 20 dB is 0.3547. As a conclusion, the AE has been able to find the appropriate constellation for each triplet. The efficiency of the corresponding decoder will be evaluated in Section VI with LDPC-coded data.

B. Constellation design at low to intermediate INR

The AE is trained here with Y as a channel output. Several pairs (M, α) are considered with $SNR = 20$ dB. The achievable rates for the constellations designed by the AE are given in Fig. 6. The results are compared with the upper bound from Ihara's work in [26] which coincides, for $\alpha \in [0\ 1]$, with the rate achieved with a Gaussian input. The capacity of the Gaussian channel $\log_2(1 + \frac{S}{1+I})$ where the interference is replaced by a Gaussian noise with same power is also plotted and serves also to benchmark the achievable performance. Finally, our results are compared to the achievable rates of an input distribution with uniform phase and one mass-point at $\sqrt{S}$ for the modulus. This constellation corresponds to the class of optimal input distributions for the AWGN channel with additive radar interference ([12], theorem 1). The one

[3]The constellations obtained with the AE at the high INR regime are all RPAM. The AE always converges towards the optimal constellation shape.

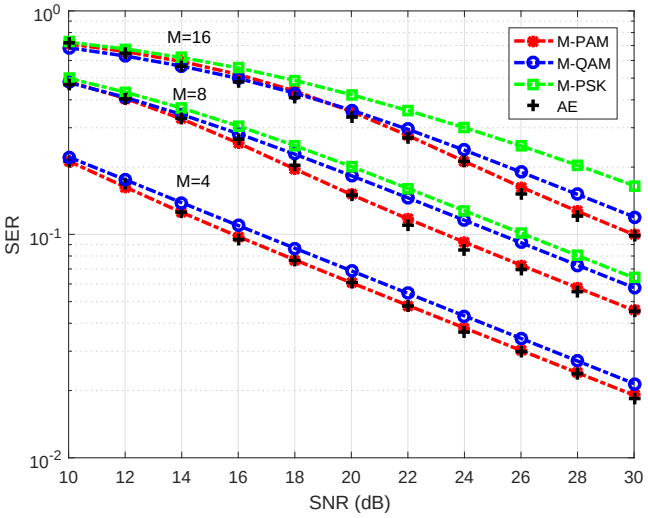

Figure 5. SER vs SNR at high INR ($I = S^{\alpha}$, $\alpha \approx 2$) and with interference cancellation.

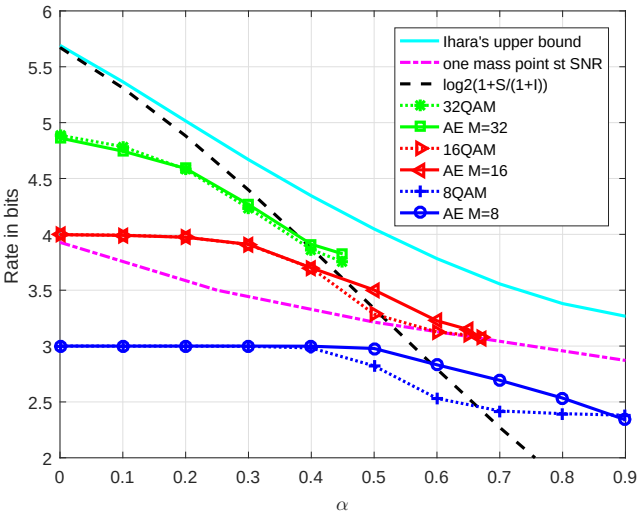

Figure 6. Achievable rates vs α at $SNR = 20$ dB and $I = S^{\alpha}$.

mass-point constellation is obviously easy to design since the modulus is necessarily equal to $\sqrt{S}$. The optimization remains challenging in the general case for which the optimal number of mass-points is unknown. This is here that the AE is expected to bring new solutions. We can observe in Fig. 6, that when α is small ($\alpha < 0.4$), the standard constellation cannot be improved with the AE. Indeed, the constellations obtained in this area are in general rotated standard QAM[4]. In contrast, when $0.4 \leq \alpha \leq 0.8$ the constellations obtained with the AE outperform significantly the standard constellations in particular when $M = 8$ or $M = 16$. Significant improvements can also be obtained when $M = 32$ but at higher SNR. When the channel impairment is too important, the AE converges automatically towards a degenerated constellation. For example, at $SNR = 20$ dB and at $\alpha = 0.7$, even if M is chosen equal to 16, a constellation with cardinality 8 will be obtained. In that case, the optimization process should be run again with $M = 8$ instead of $M = 16$ to obtain a constellation with M distinct points. This is a nice property of the AE making the choice of M much easier. The constellations obtained at ($SNR = 20$ dB, $M = 8$, $\alpha = 0.6$) and ($SNR = 20$ dB, $M = 16$, $\alpha = 0.5$) are given in Fig. 7 and are discussed below.

C. Analysis of constellation shapes

The constellations obtained with the AE and for various triplets (SNR, α, M) are given in Figs. 7 and 8. It is noticeable that the constellations in Fig. 7 can be seen as optimal input distributions ([12], theorem 1) after uniform phase sampling. Except for the point at the origin, the constellation with cardinality equal to 8 is a sampled version of the "one mass point at SNR". To take the analysis one step further, the constellation corresponding to ($SNR = 20$ dB, $M = 16$, $\alpha = 0.5$) is again plotted in Fig. 9. We can see that this constellation has four different values of the modulus and thus 4 orbits with 4 points on each $((0, 2, 4, 6); (1, 3, 5, 7); (8, 10, 12, 14); (9, 11, 13, 15))$ and 16 evenly distributed values for the phase. However on a specific orbit, the phase is not evenly distributed. From the labels given in the scatter plot, decision regions for each bit can be inferred. Let's denote the binary decomposition of C as $(C(3), C(2), C(1), C(0))$. For clarity, the constellation is again plotted in Fig. 9 in which red markers indicate a constellation point such that $C(i) = 0$ whereas blue stars indicate a constellation point such that $C(i) = 1$. We can observe a clear dividing line between the two categories ($C(i) = 0$ / $C(i) = 1$) whatever the value of i is. The demapping task can thus be handled with binary classification and this is exactly the function of the AE decoder. Constellation design for the radar interference channel is also addressed in [25]-[13]. In those publications, the loss function is the symbol error rate (SER) with power constraint on the constellation. The labelling is not addressed. The constellation given in Fig. 9 (right) is obtained at ($SNR = 20$ dB, $M = 16$, $\alpha = 0.5$) with the AE and with the SER as loss function. It can be compared with the constellation in Fig. 9 (left) obtained with our method and also at ($SNR = 20$ dB, $M = 16$, $\alpha = 0.5$). When the optimization is based on the SER, the constellation tends to shape like a concentric hexagon as already mentioned in [25]-[13]. We can observe in this particular example that the cardinality of the constellation on the right is 13 instead of 16 (several constellation points coincide). The irregular spacing between the constellation points is optimized by taking into account the binary representation of the input which is not possible when the optimization is based on the SER. From this example, we can conclude that a joint optimization (labelling + constellation shape) is mandatory in the intermediate regime. We then turn our attention to Fig. 8. We can observe that when the INR increases, degrees of freedom are being lost on the

[4]It should be noticed however that $\alpha = 0.4$ is the threshold obtained when $SNR = 20$ dB. We will see in the next section, at $\alpha = 0.2$ for example, that improvements are possible at lower SNR (compared to standard constellations).

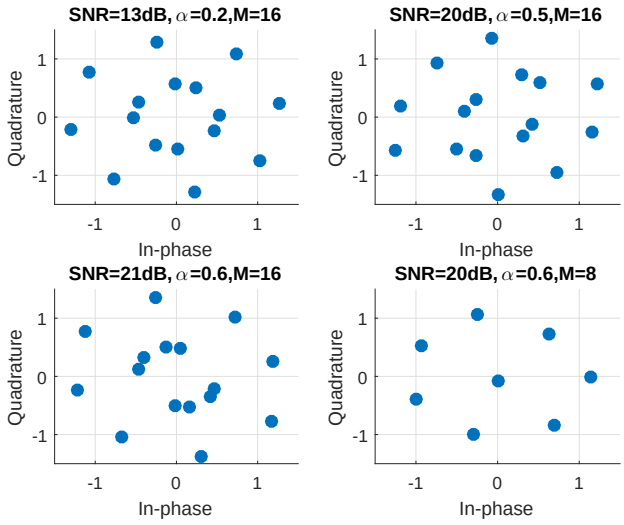

Figure 7. Constellation designed with AE. Low to intermediate INR regime.

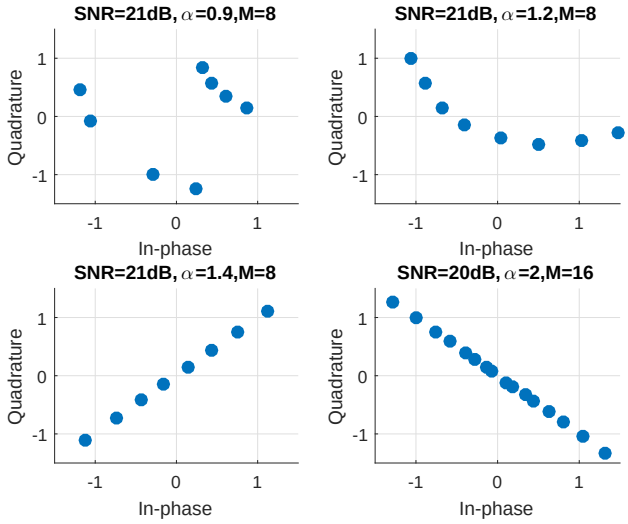

Figure 8. Constellation designed with AE. Intermediate to high INR regime.

constellation geometry. Indeed, the phase of the constellation points is allowed to span $[0; 2\pi]$ in the constellations of Fig. 7 whereas the phase is limited to smaller intervals when the INR increases. In the extreme case (high INR regime) the phase belongs to a set with cardinality 2. Loosing degrees of freedom is equivalent to increasing the prior knowledge on the transmitted symbol. This prior knowledge is a key point for recovering the transmitted symbol from the observation for large INR values (see the proof of Result 3). The constellation shape at high INR is consistent with the results in section III and with [13]. The AE is therefore able to find the most suitable constellation for given channel parameters. At this point, the efficiency of the demapping has not been evaluated yet. In the next section, the joint performance of the mapping/demapping optimized with the AE are compared to standard settings in the context of LDPC-encoded data.

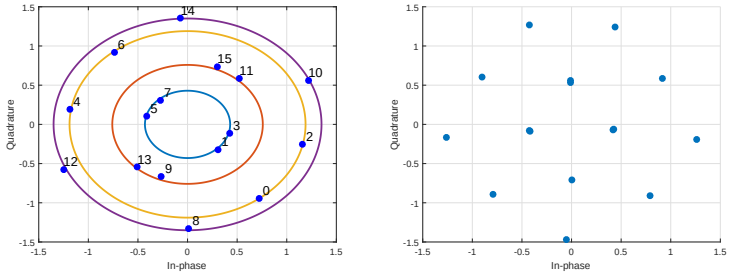

Figure 9. Constellation designed at ($SNR = 20$ dB, $\alpha = 0.5$, $M = 16$). Left: loss function in (20). Right: loss function = SER [25].

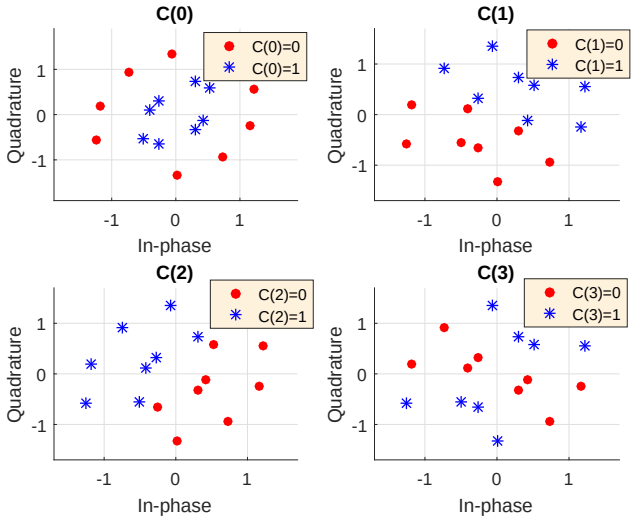

Figure 10. Constellation designed with AE at ($SNR = 20$ dB, $\alpha = 0.5$, $M = 16$). Decision region for the individual bits.

VI. Numerical results

Let's denote the information message as B and the corresponding LDPC-encoded message as C. The data-block C is split into binary words of length $log_2(M)$ which are mapped to $X \in \mathcal{X}$ and sent through the channel in (1). At the receiver side a soft demapper is used. The output of the demapper is converted into LLR and serves as an input to the LDPC decoder which is based on the sum-product algorithm. The parity-check matrix of the LDPC code proceeds from the DVB-S.2 standard. The block length of the code is fixed and equal to 64800, the code rate R is chosen according to the achievable rates of the constellation. The constellation is either a standard one (QAM, PAM, PSK) or one of the constellations in Figs. 7 - 8. The demapping is characterized by the conditional probability $p_{Y|X}(y, x)$ from which the LLR of bit j is obtained as

$$LLR(C_j) = \log\left(\frac{\sum_{x\in\mathcal{X}:x_j=0} p_{Y|X}(y,x)}{\sum_{x\in\mathcal{X}:x_j=1} p_{Y|X}(y,x)}\right) \qquad (22)$$

and serves as an input to the LDPC decoder. Depending on the values of the channel parameters, several expressions can be obtained for $p_{Y|X}(y, x)$:

- At low INR ($I \ll S$), the channel is expected to behave as an AWGN channel [14]. In that case, the channel conditional distribution reads

$$p_{Y|X}(y,x) \propto e^{-\frac{|y-x|^2}{1+I}} \tag{23}$$

We shall refer to this demapping as DMLowINR.

- At high INR ($I \gg S$), it is possible to obtain an accurate estimate $\widehat{\theta}_I = arg(Y)$ leading to Y_{IC} after interference cancellation. It can be proven with the same line of arguments as in the proof of result 1 (see also [14] for an alternative proof) that

$$\mathcal{R}e(y_{IC}e^{-j\widehat{\theta}_I}) = \sqrt{s_0}cos(\theta_0 - \widehat{\theta}_I) + V \tag{24}$$

where $V \sim \mathcal{N}(0, \frac{1}{2})$ and where the polar form of x is $\sqrt{s_0}e^{j\theta_0}$. Therefore, the conditional probability to be used at high INR reads

$$\frac{1}{\sqrt{\pi}}e^{-(\mathcal{R}e(y_{IC}e^{-j\widehat{\theta}_I}) - \sqrt{s_0}cos(\theta_0 - \widehat{\theta}_I))^2} \tag{25}$$

We shall refer to the soft demapping based on (25) as DMHighINR.

- When $I \approx S$, $arg(Y)$ is not anymore an accurate estimation of θ_I. Instead and since we want to compute a conditional probability, an estimation of θ_I conditioned on the knowledge of the transmitted symbol X is obtained as $\widehat{\theta}_I(X) = arg(Y - X)$. The accuracy of the estimation depends on the noise level. In that case

$$p_{Y|X}(y,x) \propto e^{-|y-x-\sqrt{I}e^{j\widehat{\theta}_I(x)}|^2} \tag{26}$$

The corresponding soft demapping shall be referred as DMHighSNR.

- Once trained, the AE provides a constellation alphabet associated with a valid soft demapping. The output $\hat{C}$ of the demapping block is given in (19) where the input should be Y at low and intermediate INR and Y_{IC} at high INR. The LLR of bit j is computed as $LLR(C_j) = \log\left(\frac{1-\hat{C}_j}{\hat{C}_j}\right)$. This soft-demapping is termed DMAE. As before, three cases are considered. In the high INR regime, the input of the AE decoder should be Y_{IC}. The corresponding soft-demapping is denoted DMAE(Y_{IC}). Outside of the high INR regime, the input of the AE decoder is Y and the resulting demapping is called DMAE(Y). We can observe that DMAE(Y) is based on the same observations (received samples) as DMLowINR whereas DMAE(Y_{IC}) is based on the same observations as DMHighINR. The last one, DMHighSNR, is a bit more intricate since the observation required is $Y - X$. It is possible to mimic DMHighSNR by giving $\{Y - X_i\}_{0 \leq i \leq M-1}$ as an input to the AE decoder where X_i belongs to a given constellation $\mathcal{X}$. In that case, the training process is divided into two steps: 1) constellation design with one of the schemes in Fig. 4, 2) update of the decoder weights with one of the schemes in Fig. 4 for which the weights $\mathcal{P}$ of the encoder are frozen to output the constellation $\mathcal{X}$ given by step one. The corresponding soft-demapping is denoted DMAE($Y - \{X_i\}$).

NAME	ASSUMPTION	METRIC
DMLowINR	$I \ll S$ Gaussian noise + Gaussian interference	eq. (23)
DMHighINR	$I \gg S$ Gaussian noise	eq. (25)
DMHighSNR	$S \gg 1$ Gaussian noise	eq. (26)
NAME	ASSUMPTION	INPUT OF AE DECODER
DMAE(Y)	$I \ll S$ or $I \approx S$	$\mathcal{R}e(Y)$, $\mathcal{I}m(Y)$
DMAE(Y_{IC})	$I \gg S$	$\mathcal{R}e(Y_{IC})$, $\mathcal{I}m(Y_{IC})$
DMAE($Y - \{X_i\}$)	$S \gg 1$	$\mathcal{R}e(Y - X_i)$, $\mathcal{I}m(Y - X_i)$ $X_i \in \mathcal{X}$, $i = 0, 1, ...M-1$

Table I
LIST OF DEMAPPING DEVICES.

The full list of soft-demapping devices under consideration is given in table I. The numerical results are depicted in Figs. 11 to 14 in terms of bit error rate (BER) for different INR and SNR values. We can draw some conclusions. Whatever the value of α, the best configuration always involves $\mathcal{X}_{AE}$. The AE is therefore able to yield appropriate constellations in the low and high INR regime but also in the intermediate one. This is a substantial advantage compared to the state of the art. Even in the high INR regime where we already know that a PAM-like constellation should be used, the AE is able to bring an improvement thanks to an optimized spacing of the constellation points. The highest gains are obtained in the intermediate regime and it is thus mandatory to design a proper communication chain in this region. These gains could be very important. For example, at $\alpha = 0.5$, a gap of more than 2 dB is observed between the best association ($\mathcal{X}_{AE}$+DMHighSNR) and the standard one (16QAM+DLowINR). Not surprisingly DMLowINR and DMHighSNR lead to reasonable performance in their target area but are dismissed outside. For all configurations under consideration, DMHighSNR outperforms the other alternatives even when $SNR \approx 14$ dB but may fail at lower SNR. We turn to the discussion on the performance of DMAE. In Fig. 11, DMAE(Y) exhibits close performance to DMLowSNR. Indeed, when $\alpha = 0.2$, the assumption $\sqrt{I}e^{j\theta_I} + Z \sim \mathcal{N}(0, 1 + I)$ is true. Then eq. (23) is the correct metric and we cannot expect the training process to perform better. The gap of 0.1 dB between DMLowSNR and DMAE(Y) should then be considered as a good result for DMAE(Y). When $\alpha = 0.5$, the radar interference can no longer be considered as a Gaussian variable. As a consequence, DMLowINR exhibits poor results as can be seen in Fig. 12. DMAE(Y) does not rely on a Gaussian assumption (on the radar interference) and is still a good solution at $\alpha = 0.5$. However, when the INR level increases, DMAE(Y) is not the best option as can been seen in Fig. 13. At high INR, we can observe that DMAE(Y_{IC}) and DMHighINR are within a gap of 0.2 dB meaning that the AE has been able to "learn" the metric. Finally, DMAE((Y-$\{X_i\}$) is considered in Figs. 13-14 as an alternative to DMHighSNR. We can see that, in both cases, the gap between DMHighSNR and DMAE((Y-$\{X_i\}$) is less than 0.5 dB which means that DMAE((Y-$\{X_i\}$) is an appropriate alternative to DMHighSNR in particular when the noise has unknown or intractable closed-form expression.

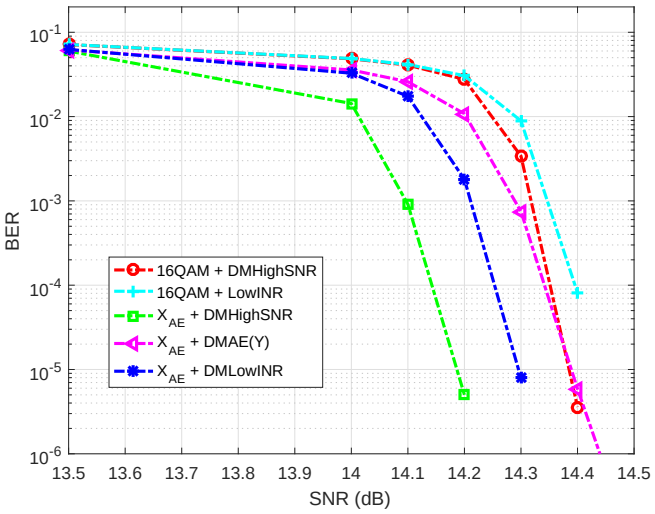

Figure 11. BER vs SNR at $\alpha = 0.2$ and with LDPC rate R = 3/4. $\mathcal{X}_{AE}$ optimized with ($M = 16$, $\alpha = 0.2$, $SNR = 13$ dB).

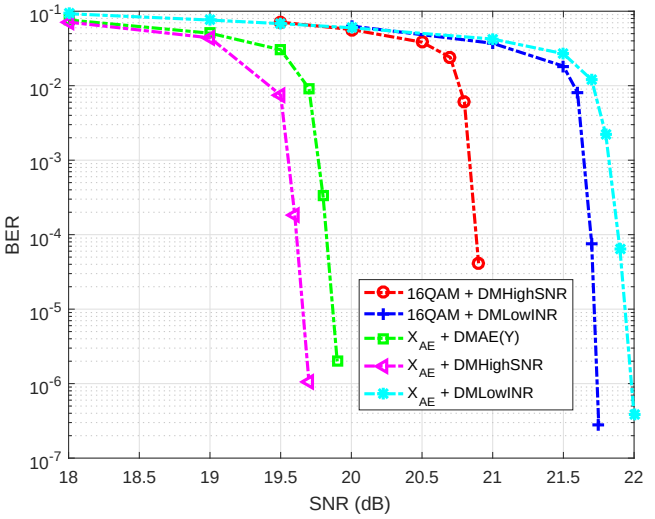

Figure 12. BER vs SNR at $\alpha = 0.5$ and LDPC rate R = 5/6. $\mathcal{X}_{AE}$ optimized with ($M = 16$, $\alpha = 0.5$, $SNR = 20$ dB).

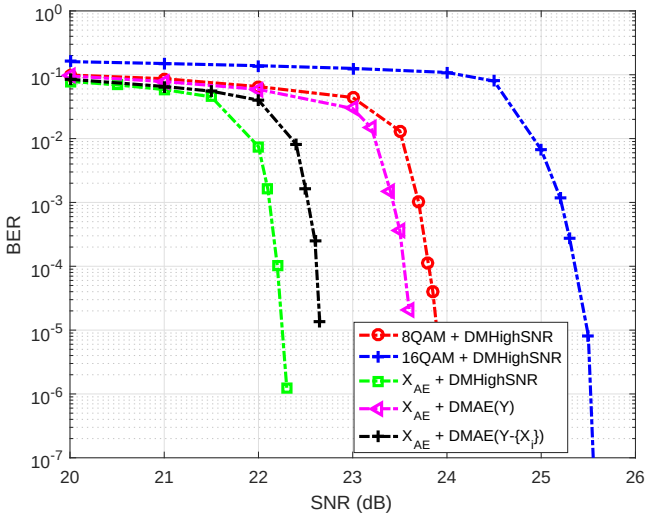

Figure 13. BER vs SNR at $\alpha = 0.9$ and LDPC rate R = 8/10 ($M = 8$) or R = 3/5 ($M = 16$). $\mathcal{X}_{AE}$ optimized with ($M = 8$, $\alpha = 0.9$, $SNR = 21$ dB).

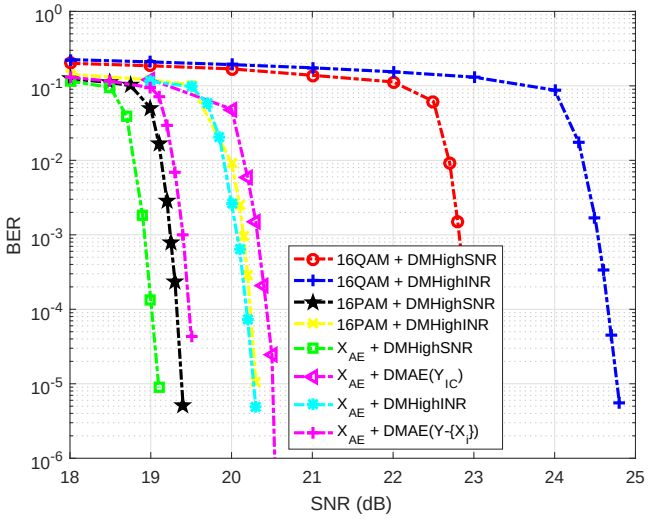

Figure 14. BER vs SNR at $\alpha = 2$ and LDPC rate R = 3/5. $\mathcal{X}_{AE}$ optimized with ($M = 16$, $\alpha = 2$, $SNR = 20$ dB).

VII. Conclusions

The coexistence of a short duty cycle, wide-band radar signaling and of a narrow-band communication signal was discussed in this paper from the communication system perspective. Thanks to deep learning techniques, it was proven that it is possible to design an appropriate constellation and to maintain an acceptable communication link whatever the interference power is. In this paper, AEs were envisioned as an off-line optimization tool. The self adaptation of a given node to the transmission medium is the next step forward.

References

[1] G. Locke, "An Assessment of the Near-Term Viability of Accommodating Wireless Broadband Systems in the 16751710 MHz, 17551780 MHz, 35003650 MHz, and 42004220 MHz, 43804400 MHz Bands," NTIA, Tech. Rep., Nov. 2010.

[2] R. S. P. G. . (EU), "Report on Collective Use of Spectrum (CUS) and other spectrum sharing approaches," Tech. Rep., 2011.

[3] Z. Khan, J. J. Lehtomaki, R. Vuohtoniemi, E. Hossain, and L. A. Dasilva, "On opportunistic spectrum access in radar bands: Lessons learned from measurement of weather radar signals," *IEEE Wireless Communications*, vol. 23, no. 3, pp. 40–48, June 2016.

[4] A. Aubry, A. D. Maio, M. Piezzo, and A. Farina, "Radar waveform design in a spectrally crowded environment via nonconvex quadratic optimization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 2, pp. 1138–1152, April 2014.

[5] K.-W. Huang, M. Bic, U. Mitra, and V. Koivunen, "Radar waveform design in spectrum sharing environment: Coexistence and cognition," in *2015 IEEE Radar Conference (RadarCon)*, May 2015, pp. 1698–1703.

[6] H. Deng and B. Himed, "Interference Mitigation Processing for Spectrum-Sharing Between Radar and Wireless Communications Systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 49, no. 3, pp. 1911–1919, July 2013.

[7] A. Khawar, A. Abdel-Hadi, and T. C. Clancy, "Spectrum sharing between S-band radar and LTE cellular system: A spatial approach,"

in *2014 IEEE International Symposium on Dynamic Spectrum Access Networks (DYSPAN)*, April 2014, pp. 7–14.

[8] A. Turlapaty and Y. Jin, "A joint design of transmit waveforms for radar and communications systems in coexistence," in *2014 IEEE Radar Conference*, May 2014, pp. 315–319.

[9] A. Lackpour, M. Luddy, and J. Winters, "Overview of interference mitigation techniques between WiMAX networks and ground based radar," in *2011 20th Annual Wireless and Optical Communications Conference (WOCC)*, April 2011, pp. 1–5.

[10] I. Pasya, A. Mahyuni, S. F. S. Adnan, and Z. Awang, "Analysis of interference from UWB radar signals on a digital OFDM transmission system," in *2011 IEEE International Conference on System Engineering and Technology*, June 2011, pp. 91–95.

[11] A. R. Chiriyath, B. Paul, G. M. Jacyna, and D. W. Bliss, "Inner Bounds on Performance of Radar and Communications Co-Existence," *IEEE Trans. on Signal Processing*, vol. 64, no. 2, pp. 464–474, Jan 2016.

[12] S. Shahi, D. Tuninetti, and N. Devroye, "On the Capacity of the AWGN Channel With Additive Radar Interference," *IEEE Transactions on Communications*, vol. 66, no. 2, pp. 629–643, Feb 2018.

[13] N. Nartasilpa, D. Tuninetti, and N. Devroye, "Signal constellation design in the presence of radar interference and Gaussian noise," in *MILCOM 2017 - 2017 IEEE Military Communications Conference (MILCOM)*, Oct 2017, pp. 719–724.

[14] N. Nartasilpa, D. Tuninetti, N. Devroye, and D. Erricolo, "On the Error Rate of a Communication System Suffering from Additive Radar Interference," in *2016 IEEE Global Communications Conference (GLOBECOM)*, Dec 2016, pp. 1–6.

[15] 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.

[16] U. Mitra and H. V. Poor, "Neural network techniques for adaptive multiuser demodulation," *IEEE Journal on Selected Areas in Commun.*, vol. 12, no. 9, pp. 1460–1470, Dec 1994.

[17] B. Aazhang, B. . Paris, and G. C. Orsak, "Neural networks for multiuser detection in code-division multiple-access communications," *IEEE Transactions on Commun.*, vol. 40, no. 7, pp. 1212–1222, July 1992.

[18] T. Gruber, S. Cammerer, J. Hoydis, and S. ten Brink, "On deep learning-based channel decoding," in *CISS'17*, 2017.

[19] T. J. O'Shea, K. Karra, and T. C. Clancy, "Learning to communicate: Channel auto-encoders, domain specific regularizers, and attention," in *2016 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, Dec 2016, pp. 223–228.

[20] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising auto-encoders : learning useful representation in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, no. 11, pp. 3371–3408, 2010.

[21] L. Yang, S. Xu, and H. Yang, "Design of Circular Signal Constellations in the Presence of Phase Noise," in *2008 4th International Conference on Wireless Commun., Networking and Mobile Computing*, Oct 2008.

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

[23] B. J. Kwak, N. O. Song, B. Park, and D. S. Kwon, "Spiral QAM: A Novel Modulation Scheme Robust in the Presence of Phase Noise," in *2008 IEEE 68th Vehicular Technology Conference*, Sept 2008, pp. 1–5.

[24] N. Nartasilpa, D. Tuninetti, N. Devroye, and D. Erricolo, "Let's share CommRad: Effect of radar interference on an uncoded data communication system," in *2016 IEEE Radar Conference (RadarConf)*, May 2016.

[25] N. Nartasilpa, A. Salim, D. Tuninetti, and N. Devroye, "Communications System Performance and Design in the Presence of Radar Interference," *IEEE Transactions on Communications*, vol. 66, no. 9, pp. 4170–4185, Sept 2018.

[26] S. Ihara, "On the capacity of channels with additive non-Gaussian noise," *Information and Control*, vol. 37, no. 1, pp. 34 – 39, 1978.

[27] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[28] F. Chollet, *Deep Learning with Python*. Manning, 2018.