



HAL
open science

CSI-aided Robust Neural-based Decoders

Meryem Benammar, Eduardo Dadalto-Camara-Gomes, Pablo Piantanida

► **To cite this version:**

Meryem Benammar, Eduardo Dadalto-Camara-Gomes, Pablo Piantanida. CSI-aided Robust Neural-based Decoders. 11th International Symposium on Topics in Coding, Aug 2021, Montréal, Canada. pp.0. hal-03522702

HAL Id: hal-03522702

<https://hal.science/hal-03522702>

Submitted on 12 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open Archive Toulouse Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of some Toulouse researchers and makes it freely available over the web where possible.

This is an author's version published in: <https://oatao.univ-toulouse.fr/28710>

Official URL : <https://doi.org/10.1109/BRAINS52497.2021.9569808>

To cite this version :

Benammar, Meryem and Dadalto-Camara-Gomes, Eduardo and Piantanida, Pablo CSI-aided Robust Neural-based Decoders. (2021) In: 11th International Symposium on Topics in Coding, 30 August 2021 - 3 September 2021 (Montréal, Canada).

Any correspondence concerning this service should be sent to the repository administrator:

tech-oatao@listes-diff.inp-toulouse.fr

CSI-aided Robust Neural-based Decoders

Meryem Benammar
ISAE-Supaéro, Université de Toulouse
meryem.benammar@isae-supaero.fr

Eduardo Dadalto Câmara Gomes and Pablo Piantanida
Laboratoire des Signaux et Systèmes (L2S)
CentraleSupélec CNRS Université Paris Saclay
{eduardo.dadalto-camara-gomes,pablo.piantanida}@centralesupelec.fr

Abstract—In this work ¹, we investigate the design of neural based channel decoders for the Binary Asymmetric Channel (BAC), which exhibits robustness issues related to training/testing channel parameters mismatch. Rather than enforcing the independence of the trained model to the channel parameter as in our previous work, we show that providing even a coarse (possibly imperfect) quantized CSI to the decoder, allows to build a *single* robust neural decoder for all values of channel parameters.

I. INTRODUCTION

Authors in [1], and references therein, showed that near-to-optimal and low complexity channel decoders can be built using neural networks. Following these results, extensive research was undertaken in order to generalize the design of neural based channel decoders and encoders to a variety of communication channel models. However, while neural based decoders allow to trade the online complexity with an offline complexity pertaining to the training phase, the performances of such neural based decoders depend strongly on the design metaparameters (structure of the network, loss functions, optimizers, training channel parameter, ...) , and can exhibit severe robustness limitations especially to possible channel model mismatches between training and testing. Indeed, when for most classical channels, alike the Binary Input Additive White Gaussian (BI-AWGN) channel, the Binary Symmetric Channel (BSC) and the Binary Erasure Channel (BEC), the neural based channel decoders developed in literature exhibit good generalization capability to channel parameters unseen during training (SNR, crossover probability, erasure probability), there exist a family of channels for which such robust neural based decoders cannot be built using the existing design methodology, namely, the Binary Asymmetric Channel (BAC) [2]. The BAC serves as an equivalent channel model for density based communication channels, alike optical channels with ON-OFF Keying (OOK) and direct detection, or molecular communication, whose probability of misdetection depends on the transmitted signal.

A first approach, which we implemented in [3], was to implement Domain Adaptation (DA) techniques to the design of neural based decoders. DA techniques are mostly inspired from the supervised learning community [4], with particular applications to computer vision, and consist in improving the robustness of a given neural network classifier (decoder) to mismatches between training (source) and testing (target)

datasets. Most DA techniques aim at enforcing some independence between the neural classifier (decoder) and environment variables (Channel State Information (CSI)), which improves slightly the generalization property of the neural based decoders, but does not allow to perfectly adapt to all channel parameters since enforcing a trade-off in the training phase.

Contribution: In this work, rather than enforcing the invariance of the neural decoder to the channel parameters (CSI), we will rather enforce a full dependence of the decoder on the channel parameters, by means of a carefully designed quantized CSI. To this end, we start by showing, through a simple information theoretic analysis, that a quantization of the CSI can be as informative as a perfect CSI, from which we infer an approximation of the number of quantization levels. Then we show that we can build a *single* neural decoder which remains optimal for all parameters, provided that a carefully quantized CSI is provided to the neural decoder during training. To this end, we compare two sources of quantized CSI, namely a genie-aided CSI quantizer, and a blind quantized CSI estimator, and show that in both cases, optimal neural based decoding can be built for the BAC.

The remainder of the work is as follows. We introduce the definitions and previous results in Sections II and III. In Section IV, we introduce the suggested method and list possible constructions of the CSI quantizer.

Notations: Random variables (resp. their realizations) are denoted by capital case letters (resp. lower case letters), e.g. X (resp. x). In this paper, pmf stands for probability mass function. The pmf of a random variable X is denoted as $P_X(\cdot)$, while the conditional pmf of a random variable X knowing Y , is denoted as $P_{X|Y}(\cdot|\cdot)$. The operators $\mathbb{E}(\cdot)$ and $\mathbb{1}[\cdot]$ stand respectively for the expectation operation and the indicator function. The notations $H(X)$, $H(X|Y)$ and $I(X;Y)$ refer respectively to the entropy of X and the conditional entropy of X knowing Y and the mutual information between X and Y . The conditional (termed also ergodic) mutual information between two variables X and Y given a variable P writes as

$$I(X;Y|P) \triangleq \mathbb{E}_P I(X;Y|P=p). \quad (1)$$

The binary entropy is defined on $[0 : 1]$ by

$$h_2(x) \triangleq -x \log_2(x) - (1-x) \log_2(1-x). \quad (2)$$

An n -dimensional sequence is denoted as $x^n = (x_1, \dots, x_n)$ where x_i is the i -th component of the sequence.

¹This work was carried out when Mr Dadalto was a MSc student at ISAE-Supaéro, he is currently with L2S, CentraleSupélec Université Paris Saclay.

II. DEFINITIONS AND PREVIOUS RESULTS

A. The channel coding problem

In this work, we investigate the channel coding problem depicted in Figure 1, and which consists in:



Fig. 1. The channel coding problem

- A k -bit input message u^k with i.i.d Bern(0.5) entries.
- An encoder $f_e(\cdot)$ which maps the k -bit input message u^k into an n -symbol sequence x^n from an alphabet $\mathcal{X} = \{0, 1\}$. Throughout this work, we will assume that the encoder $f_e(\cdot)$ is an (n, k) Polar code with generator matrix G , i.e., $x^n = G.u^k$ where G is a submatrix of $T \triangleq T_2^{\otimes \log_2(n)}$ and $T_2 \triangleq \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ is Arikan's binary kernel. The generator matrix G corresponds to the sub-space generated by the non-frozen positions of the matrix T .
- A memoryless noisy channel with output alphabet $\mathcal{Y} = \{0, 1\}$ and conditional pmf $P_{Y^n|X^n}$ given by, for all $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$,

$$P_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^n P_{Y|X}(y_i|x_i). \quad (3)$$

- A decoder $g_d(\cdot)$ which maps each received sequence y^n into an estimate message sequence \hat{u}^k .

The channel coding problem consists in devising, given a fixed encoding rule $f_e(\cdot)$, decoding rules $g_d(\cdot)$ which minimize the probability of error given by:

$$P_e \triangleq \mathbb{P}(\hat{U}^k \neq U^k). \quad (4)$$

In this work, we will be interested in two families of decoding rules: the theoretically optimal Maximum A Posteriori (MAP) decoding rule $g_{MAP}(\cdot)$, and neural based decoding rules. To this end, let us first introduce in more details the channel model of interest, and characterize the MAP decoder for this channel.

B. The Binary Asymmetric Channel

In this work, we investigate the channel coding problem for the BAC. In this binary-input binary-output channel, depicted in figure 2, the crossover probability takes two possible values p or q , depending on the value of the input bit, $X = 0$ or 1 .

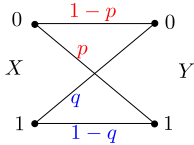


Fig. 2. The BAC (p, q)

The channel transition probability of the BAC is given by

$$P_{Y|X}(y|x) = \begin{cases} p^{x \oplus y} (1-p)^{1-x \oplus y} & \text{if } x = 0, \\ q^{x \oplus y} (1-q)^{1-x \oplus y} & \text{if } x = 1. \end{cases}$$

The BAC generalizes three main classes of binary-input binary-output channels, namely the BSC($p = q$), the Z-channel ($p = 0$) and the S-channel ($q = 0$). In the following, we characterize the MAP decoder for the BAC(p, q).

Lemma 1 (MAP decoding for the BAC).

The MAP decision rule of the BAC(p, q) is given by

$$g_{MAP}(y^n) = \underset{u^k \in \{0:1\}^k}{\operatorname{argmax}} \left[\log(1-p) |I_0(x^n(u^k))| + \log(1-q) |I_1(x^n(u^k))| + \log\left(\frac{p}{1-p}\right) (x_0^n(u^k) \oplus y_0^n) + \log\left(\frac{q}{1-q}\right) (x_1^n(u^k) \oplus y_1^n) \right], \quad (5)$$

where, for any codeword x^n , the quantities $I_0(x^n)$, $I_1(x^n)$, x_0^n , x_1^n , y_0^n , and y_1^n are defined as follows, for all $b \in \{0, 1\}$:

$$I_b(x^n) \triangleq \{i \in [1:n], x_i = b\}, \quad (6)$$

$$x_b^n \triangleq (x_i)_{i \in I_b(x^n)} \text{ and } y_b^n \triangleq (y_i)_{i \in I_b(x^n)}, \quad (7)$$

and where $|I_b(x^n)|$ refers to the cardinality of the set $I_b(x^n)$.

Proof. The proof of this lemma follows from [2, 3.3] \square

As can be seen from (5), the MAP decoding rule complexity is exponential in the message length k , which renders its implementation impractical even for moderate message lengths. Besides, it relies on prior knowledge of the specific parameters (p, q) . Following the general understanding that quasi-optimal low complexity decoders can be built from neural networks, we investigate in the following the design and performances of neural based decoders as an alternative to the MAP.

III. NEURAL BASED DECODING FOR THE BAC

The idea behind neural based decoding is to replace the decoding rule $g_d(\cdot)$ with a pretrained neural network which maps every received sequence y^n to a message estimate \hat{u}^k . To this end, the design of the neural based decoder calls for a careful choice of the training meta-parameters, and more specifically, of the training crossover probabilities of p and q . In the following, we summarize our previous findings [3] on neural based decoding for the BAC. Throughout this work, we will fix a value for $q = 0.07$, and assess the performances of neural based decoders for all possible values of p .

A. Robustness issues for neural based decoders

In [3], we built neural based decoders for the $(16, 4)$ polar code, trained under different values of the parameter p : $p = 0.01$, $p = 0.1$ and $p = 0.4$. Table I summarizes the parameters of the neural based decoder.

TABLE I
PARAMETERS OF CSI-AIDED NEURAL DECODER

Parameter	Value
(n, k) code	$(16, 4)$ Polar
Network structure	$[128, 16]$
Activation	[Relu, Softmax]
Epochs	2^{12}
# Messages	1000
Batch size	256

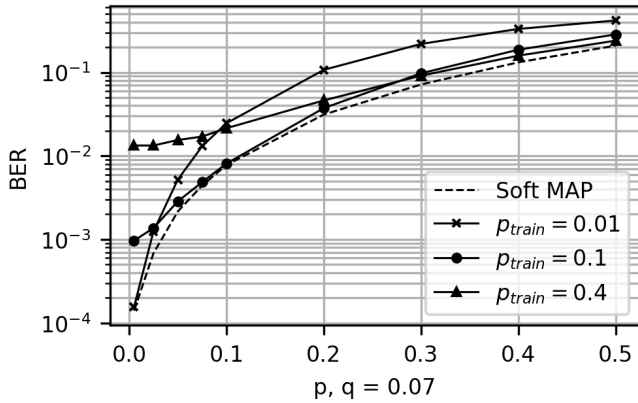


Fig. 3. Neural network decoder BER performance trained with 3 different training channel parameter p .

We show in Figure 3 the performances of the different neural networks obtained at these different training p . The results highlight clearly a robustness issue when generalizing to other values of p . This can be intuitively understood from the fact that even the optimal decision rule, i.e., the MAP, requires the knowledge of the channel parameters (p, q). This robustness impairment was scarcely noticed for most usual channels, namely, the BSC, BEC and BI-AWGN channels which have been investigated in literature, partly due to the fact that their MAP decision rule is invariant to the channel parameter (crossover probability, erasure probability or SNR).

B. Domain adaptation techniques

Our first attempt to tackle this robustness impairment of neural based decoding for the BAC was to apply Domain Adaptation (DA) techniques to the channel coding problem under investigation. DA techniques, mostly used in computer vision [4], consist in tackling the mismatch between the source dataset (training) and the target dataset (testing) which might exhibit distinct statistics. In the following, we summarize our previous contributions on how to build robust neural based decoders by enforcing a channel-oblivious neural decoder.

- Loss mixtures: we train the neural network using samples drawn from multiple training values of p . This amounts to training the neural decoder on a mixture of losses, with mixing proportions corresponding to the relative proportions of the corresponding training p samples in the training dataset.
- Models mixtures: we train distinct neural decoders, each for a distinct parameter p , and then train a meta model which combines the outputs of the different models to yield a decision without retraining the individual models.
- Domain invariant training: we train the neural decoder on two different training values of p (namely $p = .01$ and $p = 0.1$) by passing each transmitted sequence u^k through the two channels in parallel, and regularizing the loss function to enforce also that the decoded message u^k from the received signals on both channels be the same.
- Domain independent training: we train the neural decoder in an adversarial manner to enforce that the decoder output

be independent statistically from the channel parameter. To this end, we regularize the loss function so as to impose the maximization of the conditional entropy $H(\hat{U}^K|P)$ while minimizing the original loss function.

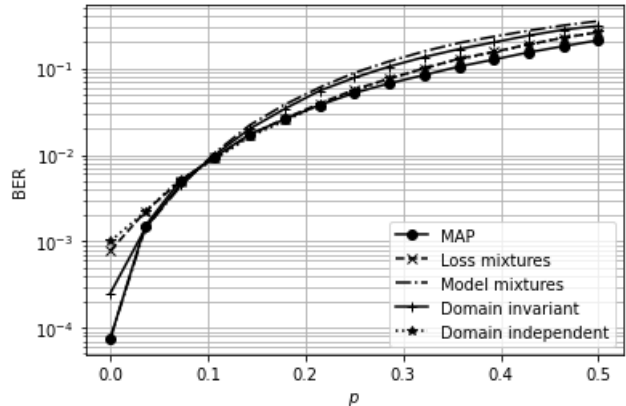


Fig. 4. Comparison of different DA techniques and MAP decoding

Figure 4 shows the results obtained by each of these DA techniques, as compared to the MAP decoding. It can be seen that none of these methods, even with extensive fine tuning of the metaparameters (mixture parameters, networks design, optimizers) allows to approach the MAP decoder everywhere.

IV. CSI-AIDED NEURAL BASED CHANNEL DECODING (CSI-ND)

Throughout the remainder of this work, we will denote by CSI any form of knowledge the decoder may have on the channel parameter p . In this section, we first show that even a coarse quantized CSI can be informative enough, and then assess the performances of two possible quantized CSI aided neural based decoder: Genie-Aided quantized CSI and a blind quantized CSI estimator.

A. Quantized CSI is sufficient

Unlike classical decoders (MAP, Successive Cancellation decoders (SC), BCJR, ...), to which we can feed continuous valued CSI, a neural based decoder requires a discrete representation of the CSI, which we term hereafter *quantized CSI*. In this work, we suggest to obtain the quantized CSI by first applying a (possibly non-uniform) quantization to the channel parameter p , and then encoding the obtained classes in a one-hot vector which is then passed on as an additional input to the neural decoder. We show through an information theoretic analysis that a coarsely quantized CSI should be informative enough for the decoding task.

To this end, let us assume that the channel parameter P is a uniform random variable over the interval $[0 : p_m]$, where p_m is the maximum value of crossover probabilities under investigation. Assume as well that C_P is a uniformly quantized version of P over N levels in the interval $[0 : p_m]$. The quantized CSI variable, which can be interpreted as well as the class to which P belongs, can be written as

$$C_P = \left\lfloor \frac{P}{\Delta} \right\rfloor \Delta, \quad (8)$$

where $\lfloor x \rfloor$ stands for the integer floor of x , and Δ is the quantization step. Note that, since we apply uniform quantization over the interval $[0 : p_m]$, $\Delta = \frac{p_m}{N}$, and thus, the quantized CSI is a function of both p_m and N and is given by

$$C_P = \left\lfloor \frac{P}{p_m} N \right\rfloor \frac{p_m}{N}. \quad (9)$$

In the following, we will define the set \mathcal{C} as

$$\mathcal{C} \triangleq \left\{ 0, \frac{p_m}{N}, \dots, \frac{N-1}{N} p_m \right\}. \quad (10)$$

In order to design the number of levels N , we compare two conditional mutual information, respectively, $I(X; Y|P)$ and $I(X; Y|C_P)$, which capture the information content shared between the input and output of the channel given, respectively, perfect or quantized CSI.

Lemma 2 (Conditional MI under perfect CSI). *The Conditional mutual information between the channel input X and the channel output Y , given the perfect knowledge of P , is*

$$I(X; Y|P) = \mathbb{E}_P \left[h_2 \left(\frac{1-P+q}{2} \right) - \frac{1}{2} h_2(P) - \frac{1}{2} h_2(q) \right].$$

Proof. The proof follows from the definition of conditional mutual information in (1) assuming $P_X \sim \text{Bern}(0.5)$. \square

Lemma 3 (Conditional MI under quantized CSI). *The conditional mutual information under quantized CSI writes as*

$$I(X; Y|C_P) = \mathbb{E}_{C_P} \left[h_2 \left(\frac{1 - c_p - \frac{p_m}{2N} + q}{2} \right) - \frac{1}{2} h_2 \left(c_p + \frac{p_m}{2N} \right) - \frac{1}{2} h_2(q) \right] \quad (11)$$

where C_P is uniformly distributed over the set \mathcal{C} (9).

Proof. The proof of this lemma is given in Appendix A.

In Figure 5, we plot the conditional mutual information under perfect and quantized CSI, as a function of the number of quantization levels N and considering $p_m = 0.5$. It can be seen that $N = 4$ yields an already satisfying result. Having proved that a quantized estimate is enough to capture all the information necessary for the channel knowledge, we will hereafter assume that the parameter of interest C_P lies in a discrete set, and consists in the class to which p belongs, with a cardinality of at least 4 classes. In the following, we investigate two possible constructions of CSI-aided neural decoders based on either a Genie-Aided (GA) quantized CSI or a blind quantized CSI.

B. Genie-aided CSI quantizer and neural decoding

Let us assume that the true value of the channel parameter is given by a genie. Then, Figure 7 shows the block diagram for the GA neural based decoder. The genie encodes the true channel parameter p into a one hot vector as follows:

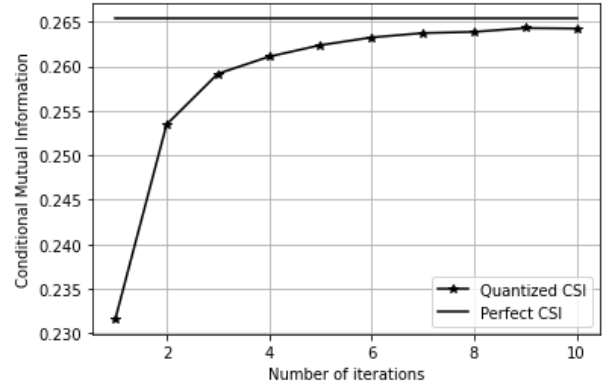


Fig. 5. Comparison of information rates for perfect and quantized CSI

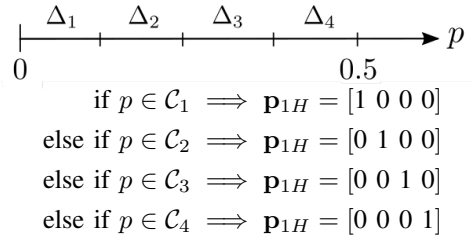


Fig. 6. One hot encoding of p into discrete intervals for model prediction. The intervals of choice were $\{(0, 0.01), (0.01, 0.1), (0.1, 0.3), (0.3, 0.4)\}$.

This CSI is then concatenated with the observed sequence y^n and used as the input for the neural based decoder in the training phase as shown in Figure 7 (training parameters are given in I). Figure 8 shows the resulting BER as compared to that of the MAP. It can be clearly seen that, although provided with only coarsely quantized CSI, the neural based decoder can actually encode the information of the state in its decoding structure, and hence, adapt to varying channel parameters p .

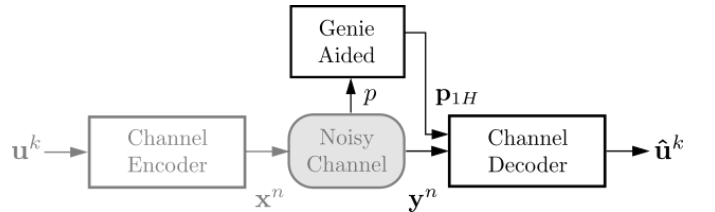


Fig. 7. Genie aided decoder illustrated.

C. Blind CSI quantizer and neural decoding

In this section, we replace the genie with a blind (non pilot-aided) quantized CSI estimator based on an iterative MAP estimator of the class C_p , which alleviated the short blocklength limitation. To this end, let us denote $P_{C_p}^t$ the prior at iteration t , and let $P_{C_p}^0 = [0.25, 0.25, 0.25, 0.25]$ be a totally uninformative initial prior. We choose to approximate the prior on C_p at iteration $t + 1$ by the following

$$P_{C_p}^{t+1}(c_p) \leftarrow P_{C_p}^t(c_p|y^n) \triangleq \frac{P_{Y^n|C_p}(y^n|c_p)P_{C_p}^t(c_p)}{\sum_{c'_p} P_{Y^n|C_p}(y^n|c'_p)P_{C_p}^t(c'_p)}.$$

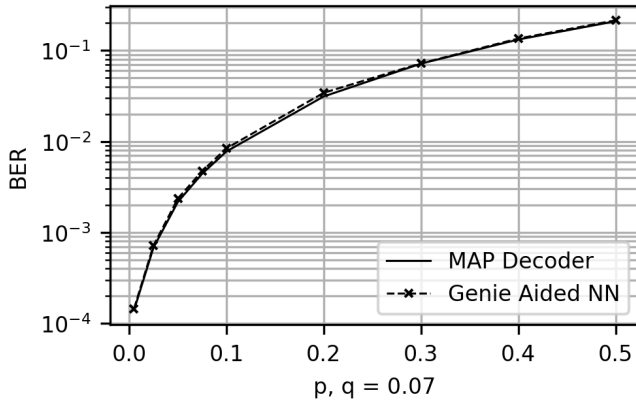


Fig. 8. Genie aided decoder and MAP decoder BER performance.

In order to reduce the complexity of this algorithm, we train a neural network to generate from both a prior $P_{C_p}^t$ and a received sequence y^n , the prior of the following iteration $P_{C_p}^{t+1}$. Alike the iterative MAP, we will use this blind neural CSI estimator in an iterative manner as well as shown in Figure 9 and concatenate it with the neural based decoder.

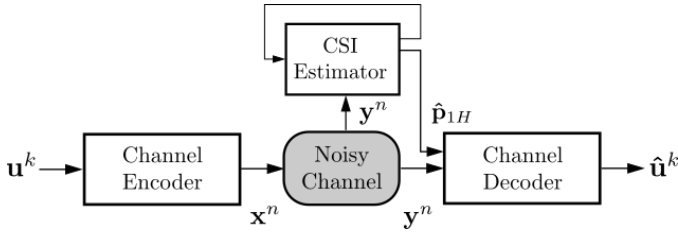


Fig. 9. Joint CSI estimation and decoding illustrated diagram.

Figure 10 shows that the performance of the obtained neural based decoder meets that of the MAP decoder although a blind quantized CSI estimator is used.

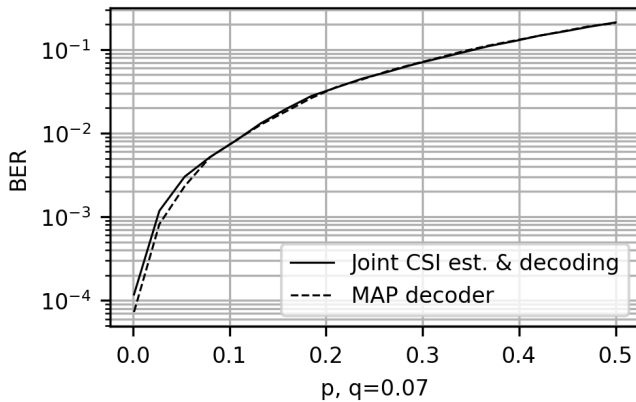


Fig. 10. Joint CSI estimation and decoding BER performance.

V. CONCLUSION

In this work, we showed that a *unique, robust and quasi-optimal* neural based decoder can be built for the BAC, by feeding a quantized (even coarse, and possibly blind) CSI to the neural decoder in the training phase. Further work could

be done on the design of more evolved channel decoders based on RNN or CNN which could encompass the CSI estimation step, whilst remaining low complexity and low latency.

APPENDIX A PROOF OF LEMMA 3

In order to express this information rate, we will first show that under quantized CSI-R, the channel $P_{Y|X,C_P}$ is equivalent to a BAC with a specific pair of parameters which depend on P , N (number of quantization levels) and q . Then the conditional mutual information will follow similarly to a BAC with perfect CSI. We can write that

$$\begin{aligned} P_{Y|X,C_P}(y|x, c_p) &= \int_p f_{P|C_P}(p|c_p) P_{Y|X,C_P,P}(y|x, c_p, p) dp \\ &= \int_{p \in \mathcal{P}(c_p)} \frac{N}{p_m} P_{Y|X,P}(y|x, p) dp \\ &= \frac{N}{p_m} \int_{c_p}^{c_p+\Delta} P_{Y|X,P}(y|x, p) dp \end{aligned}$$

where we used the fact that

$$\begin{aligned} f_{P|C_P}(p|c_p) &= \frac{f_P(p)}{P_{C_P}(c_p)} P_{C_P|P}(c_p|p) \\ &= \frac{1}{p_m} \frac{1}{\int_{p \in \mathcal{P}(c_p)} f_P(p) dp} \mathbb{1}(p \in \mathcal{P}(c_p)) \\ &= \frac{N}{p_m} \mathbb{1}(p \in \mathcal{P}(c_p)). \end{aligned}$$

Next, by simplifying the integral, we can show that:

$$P_{Y|X,P}(y|x=1, c_p) = q^{y \oplus x} (1-q)^{1-y \oplus x} \quad (12)$$

$$P_{Y|X,C_P}(1|x=0, c_p) = c_p + \frac{p_m}{2N} \quad (13)$$

$$P_{Y|X,C_P}(0|x=0, c_p) = 1 - \left(c_p + \frac{p_m}{2N} \right), \quad (14)$$

where we have used the fact that:

$$P_{Y|X,C_P}(1|x=0, c_p) = \frac{N}{p_m} \int_{c_p}^{c_p+\Delta} p dp = c_p + \frac{p_m}{2N}. \quad (15)$$

Thus, the channel under quantized CSI c_p , can be considered to be a BAC with parameters $\left(c_p + \frac{p_m}{2N}, q \right)$. It then follows that the mutual information writes as

$$\begin{aligned} I(X; Y|C_P = c_p) &= h_2 \left(\frac{1 - c_p - \frac{p_m}{2N} + q}{2} \right) - \frac{1}{2} h_2 \left(c_p + \frac{p_m}{2N} \right) - \frac{1}{2} h_2(q), \end{aligned}$$

where C_P is chosen uniformly in its support set \mathcal{C} in (9).

REFERENCES

- [1] T. Gruber, S. Cammerer, J. Hoydis, and S. ten Brink, "On deep learning-based channel decoding," in *2017 51st Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2017, pp. 1–6.
- [2] S. M. Moser, "Error probability analysis of binary asymmetric channels."
- [3] M. Benammar and P. Piantanida, "On robust deep neural decoders," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, 2019, pp. 527–531.
- [4] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, pp. 135–153, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231218306684>