



**HAL**  
open science

# An Unsupervised LLR Estimation with unknown Noise Distribution

Yasser Mestrah, Anne Savard, Alban Goupil, Guillaume Gelle, Laurent Clavier

► **To cite this version:**

Yasser Mestrah, Anne Savard, Alban Goupil, Guillaume Gelle, Laurent Clavier. An Unsupervised LLR Estimation with unknown Noise Distribution. EURASIP Journal on Wireless Communications and Networking, 2020, 2020 (1), pp.26. 10.1186/s13638-019-1608-9 . hal-02503050

**HAL Id: hal-02503050**

**<https://hal.science/hal-02503050v1>**

Submitted on 20 May 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.



Distributed under a Creative Commons Attribution 4.0 International License

REVIEW

Open Access



# An Unsupervised LLR Estimation with unknown Noise Distribution

Yasser Mestrah<sup>1\*</sup> , Anne Savard<sup>2</sup>, Alban Goupil<sup>3</sup>, Guillaume Gellé<sup>3</sup> and Laurent Clavier<sup>2</sup>

## Abstract

Many decoding schemes rely on the log-likelihood ratio (LLR) whose derivation depends on the knowledge of the noise distribution. In dense and heterogeneous network settings, this knowledge can be difficult to obtain from channel outputs. Besides, when interference exhibits an impulsive behavior, the LLR becomes highly non-linear and, consequently, computationally prohibitive. In this paper, we directly estimate the LLR, without relying on the interference plus noise knowledge. We propose to select the LLR in a parametric family of functions, flexible enough to be able to represent many different communication contexts. It allows limiting the number of parameters to be estimated. Furthermore, we propose an unsupervised estimation approach, avoiding the need of a training sequence. Our estimation method is shown to be efficient in large variety of noises and the receiver exhibits a near-optimal performance.

**Keywords:** Receiver design, Log-likelihood ratio (LLR) estimation, Impulsive noise, Unsupervised learning

## 1 Introduction

5G will have to deal with dense and heterogeneous networks. In such situations, interference may exhibit an impulsive behavior [1–3] and the Gaussian assumption is no longer suited. In order to establish reliable and efficient communications, one needs to take into account this impulsive nature while designing the receivers. Indeed, traditional linear receivers exhibit a dramatic performance degradation [4] under impulsive noise. Several papers proposed ways to overcome this issue by using different metrics to make the decision, e.g., a robust metric mixing euclidean distance and erasure [5],  $p$ -norm [6], Hubber metric [7], ... Nevertheless, the approaches are designed for a specific noise model and their robustness against a model mismatch is not ensured. The choice of a more universal solution that can be used for various impulsive noise is thus salutary.

Many receivers rely on the likelihood of the channel outputs according to its hypothetical input. In the binary case, this can be captured through the LLR. This is very attractive when noise is Gaussian because it leads to a linear receiver, straightforward to implement. However,

when rare events with large amplitudes (impulsive noise) arise, the LLR becomes a non-linear function. Its implementation is complex and highly depends on the noise distribution. Consequently, to obtain the optimal receiver, the noise distribution has to be known and if it falls in a parametric family, the parameters have to be estimated.

In this paper, we propose to approximate the LLR by a parametric function. If the family is large enough, it allows to adapt to many different types of noises without requiring any noise distribution assumption. Besides, if we consider a family defined by a limited number of parameters and easy to implement, both the estimation and implementation complexities are reduced. This work can thus be seen as a generalization of some previous works that dealt with an approximation of the LLR function [8–10]: if the soft limiter and the hole puncher [11] are probably the best-known solutions, we previously proposed the approximation function  $f(y) = \text{sgn}(y) \min(a|y|, b/|y|)$  [12], where  $\text{sgn}(y)$  extracts the sign of  $y$ . Nevertheless, we will not focus in this paper on the best family choice but on a generic way to estimate the approximation parameters in an unsupervised manner.

Our main contributions are the following:

- First, we propose to approximate the LLR by a function chosen in a parametric family large enough to adapt to various noise types but characterized by a

\*Correspondence: [yasser.mestrah@gmail.com](mailto:yasser.mestrah@gmail.com)

<sup>1</sup>IMT Lille Douai, IRCICA - USR 3380, University of Reims Champagne-Ardenne, CReSTIC - EA 3804, 50 Avenue Halley, 59650 Lille, France  
Full list of author information is available at the end of the article

enough limited number of parameters to obtain an efficient implementation.

- Second, we introduce an unsupervised estimation of the approximated LLR. Using an unsupervised method avoids the need of training data that reduces the useful information rate and allows to take benefit from the whole data sequence to improve the accuracy of the estimation.

The remaining of the paper is organized as follows. Once the system model and some background material are given in Section 2.2, Section 2.3 presents the optimization problem and Section 2.4 discusses the parameters' optimization in an unsupervised manner. Section 3 presents the application of our proposed scheme to low-density parity check (LDPC) coding; our simulation setup is explained and results are presented for different noise conditions. Finally, Section 4 concludes the paper.

## 2 Methods

### 2.1 Notations

We use notations like  $p(x)$  for probability mass function (pmf) or probability density function (pdf) according to the random variables involved. We keep uppercase letters for random variables and their corresponding lowercase for their realizations. We will denote  $+1$  by  $+$  and  $-1$  by  $-$  for short. In the remaining,  $\mathbb{H}(X)$  denotes the entropy of the source  $X$ ,  $\mathbb{H}(X|Y)$  the conditional entropy of  $X$  given  $Y$ , and  $\mathbb{I}(X; Y)$  the mutual information of the random variables  $X$  and  $Y$ .

### 2.2 System model and channel capacity

In information theory, channels are completely described by the conditional probability  $p(y|x)$  of the output  $Y$  given the input  $X$ . The channels we are considering in this paper are memory less, binary input, and symmetric output (MBISO) channels defined by  $Y = X + N$ , whose inputs  $X$  are perturbed by an additive noise  $N$ . A channel belongs to the MBISO family as long as the additive perturbations are symmetric, i.e., such that  $p(y|+) = p(-y|-)$ , and independent from the input. Furthermore, we assume uniform input distribution. The noise may represent thermal noise but also impulsive interference. In the latter,  $N$  can be modeled, for instance, by a Middleton class A distribution [13] or a symmetric  $\alpha$ -stable distribution [4]. Further details are given in Section 3.1.2.

In our considered case, the log-likelihood ratio (LLR) is given by

$$\Lambda(y) = \log \frac{p(y|+)}{p(y|-)} = \log \frac{p(+|y)}{p(-|y)}. \quad (1)$$

Notably, the LLR  $\Lambda(y)$  is a rewrite of the posterior probability  $p(x|y) = 1/(1 + e^{-x\Lambda(y)})$ . The LLR is a prime tool in information theory as it constitutes a sufficient statistic

relative to the channel input [14]; in other words, knowing  $\Lambda(Y)$  or  $Y$  is equivalent for the decoding process. Moreover, in practice, the LLR provides also a lingua franca to represent the input of most soft decoder algorithms such as the belief propagation (BP) algorithm.

The capacity of such a MBISO channel is given as

$$C = 1 - \mathbb{E} \left[ \log_2 \left( 1 + e^{-X\Lambda(Y)} \right) \right], \quad (2)$$

where  $\mathbb{E}[\cdot]$  denotes the expectation operator. This expression comes from the decomposition  $I(X, Y) = H(X) - H(X|Y)$  and the expression given above of the conditional density  $p(x|y)$  as a function of the likelihood ratio,  $H(X|Y) = \mathbb{E}[-\log_2 p(X|Y)] = \mathbb{E}[\log_2(1 + e^{-X\Lambda(Y)})]$ .

### 2.3 LLR approximation and optimization problem

To decode the received packets, computing the LLR is unfortunately often too demanding either because of the lack of a closed-form expression, such as for  $\alpha$ -stable noise, or because it needs high computational burden such as for Middleton noise. Furthermore, the LLR depends on the channel state information which should be estimated by the receiver: for example, the LLR is linear in AWGN channel but the slope depends on the signal to noise ratio.

In this paper, we consider a parametric approximation  $L_\theta$  of the LLR  $\Lambda(y)$ . The family of functions  $L_\theta$  is chosen for its simplicity and for its flexibility to match the LLR of different channel types. In order to narrow down the search space and to have an easy to implement approximation, we assume that the estimated LLR  $L_\theta$  is an odd piece-wise affine function of the optimization parameter  $\theta$ . More precisely, we are interested in functions that can be represented as

$$L_\theta(y) = \text{sgn}(y) \min \{ \theta_1 \phi_1(y), \theta_2 \phi_2(y), \dots, \theta_n \phi_n(y) \}, \quad (3)$$

where  $\phi_i(y)$  are functions depending solely on  $y$  but not necessarily linear in  $y$ , and  $\theta$ , given as  $\theta = [\theta_1, \theta_2, \dots, \theta_n]$ , is the optimization parameter that needs to be estimated. This approach was already investigated in previous works [8, 12]. In [12], we considered the case of two parameters, which showed good results according to bit error rate (BER).

A suitable criterion is thus needed to select the best parameter  $\theta$ . Of course, we aim at the smallest BER, but this criterion is not within reach in practice. We thus follow the idea from [15] that proposed to evaluate the accuracy of the approximation using

$$\widehat{C}_{L_\theta} = 1 - \mathbb{E} \left[ \log_2 \left( 1 + e^{-XL_\theta(Y)} \right) \right]. \quad (4)$$

In fact, using the approximated LLR is equivalent to approximate  $p(x|y)$  by the density  $q(x|y) = 1/(1 + e^{-xL_\theta(y)})$ , and thus to approximate  $H(X|Y)$  by  $\widehat{H}(X|Y) = \mathbb{E}[\log_2(1 + e^{-XL_\theta(Y)})]$ .

Whereas the previous derivation is only an heuristic, it appears to be a good criterion. Indeed, the difference between the approximated and the true conditional entropies is

$$\begin{aligned} H(X|Y) - \widehat{H}(X|Y) &= \mathbb{E} [\log_2 p(X|Y)] - \mathbb{E} [\log_2 q(X|Y)] \\ &= \mathbb{E} \left[ \log_2 \frac{p(X|Y)}{q(X|Y)} \right] \\ &= D(p(x|y) \| q(x|y)), \end{aligned} \quad (5)$$

where  $D(p\|q)$  is the Kullback-Leibler divergence between densities  $p$  and  $q$  [16]. We draw several facts from (5): the non-negativity of the divergence implies that our criterion is lower bounded,  $H(X|Y) \leq \widehat{H}(X|Y)$ , and the bound is reached when  $q(x|y) = p(x|y)$ . In other words,  $\widehat{C}_{L_\theta} = C$  for  $L_\theta = \Lambda$  if the LLR  $\Lambda$  belongs to the parametric family  $L_\theta$ .

However, in our setting, the approximated conditional entropy  $\widehat{H}(X|Y)$  is not available directly, since the expectation operator depends on the noise distribution that we assume unknown. We thus rely on the law of large numbers to estimate it, replacing the expectation by an empirical average  $\widehat{H}_K(X|Y)$ . Hence,  $\widehat{H}(X|Y)$  can be obtained as

$$\widehat{H}(X|Y) \approx \widehat{H}_K(X|Y) = \frac{1}{K} \sum_{k=1}^K \log_2 \left( 1 + e^{-x_k L_\theta(y_k)} \right), \quad (6)$$

where  $x_k$  and  $y_k$  are samples that represent the input and output of the channel respectively.

Our objective is to minimize  $\widehat{H}_K$  in (6) over the possible choices of  $\theta$ . This will allow us to find an approximation of the LLR in the considered family that should be a good choice for our decoding algorithm. Our optimization problem is therefore given as

$$\begin{aligned} \theta^* &= \arg \min_{\theta} \widehat{H}_K(X|Y) \\ &= \arg \min_{\theta} \frac{1}{K} \sum_{k=1}^K \log_2 \left( 1 + e^{-x_k L_\theta(y_k)} \right) \\ &= \arg \min_{\theta} \frac{1}{K} \sum_{k=1}^K \log_2 \left( 1 + e^{-L_\theta(x_k y_k)} \right), \end{aligned} \quad (7)$$

where the last equality holds since  $L_\theta(\cdot)$  is an odd function and since  $x_k$  belongs to  $\pm 1$ .

Finally, one can rewrite the objective function as

$$\begin{aligned} \widehat{H}_K(X|Y) &= \frac{1}{K} \sum_{\substack{k=1 \\ x_k y_k \geq 0}}^K \log_2 \left( 1 + e^{-L_\theta(x_k y_k)} \right) \\ &+ \frac{1}{K} \sum_{\substack{k=1 \\ x_k y_k < 0}}^K \log_2 \left( 1 + e^{-L_\theta(x_k y_k)} \right). \end{aligned} \quad (8)$$

In order to minimize (8), one needs to minimize the two sums, which can be treated separately according to the sign of  $x_k y_k$ . On the one hand, if  $x_k y_k > 0$  then

$$\begin{aligned} &\log_2 \left( 1 + e^{-\min\{\theta_1 \phi_1(x_k y_k), \dots, \theta_n \phi_n(x_k y_k)\}} \right) \\ &= \max_i \left\{ \log_2 \left( 1 + e^{-\theta_i \phi_i(x_k y_k)} \right) \right\}. \end{aligned} \quad (9)$$

Consequently, in order to minimize (9), one needs to increase the parameters  $\theta_i$ . On the other hand, if  $x_k y_k < 0$ , then

$$\begin{aligned} &\log_2 \left( 1 + e^{-\max\{\theta_1 \phi_1(x_k y_k), \dots, \theta_n \phi_n(x_k y_k)\}} \right) \\ &= \min_i \left\{ \log_2 \left( 1 + e^{-\theta_i \phi_i(x_k y_k)} \right) \right\}. \end{aligned} \quad (10)$$

In order to minimize (10), one thus needs to decrease the parameters  $\theta_i$ . Thus, minimizing  $\widehat{H}_K(X|Y)$  results in a compromise between minimizing each of the two sums in (8), one of it tends to increase the value of the parameters while the other tends to decrease it.

Unfortunately, based on this study, the optimization problem we are considering is not convex: indeed, if  $xy > 0$ , one can show that the objective function is convex in  $\theta$ , but this does not hold in the case  $xy < 0$ . Despite the non-convexity of the problem, we will use a simplex method based algorithm [17] to obtain at least a local minimum. This method converges within ten iterations, which is suited for our application, and the use of an algorithm adapted to non-convex methods could not result in any significant gain. It could however be different for other approximation families but the best optimization method as well as its complexity study remain out of the scope of this paper.

**Remark 1** Note that various LLR approximations can fit into the proposed affine framework, for instance, in [12], the approximation  $L_\theta(x) = \text{sgn}(x) \min(\theta_1 |x|, \theta_2 / |x|)$ . Other examples of piece-wise affine LLR approximations can be found in the literature. A classical solution is the clipping demapper [18] defined as  $L_\theta(x) = \text{sgn}(x) \min(\theta_1 |x|, \theta_2)$ . Nevertheless, other approximations that do not belong to our considered piece-wise affine function can also be found, as for instance the Hole puncher demapper [19] or non-linear approximation like in [8]. The proposed framework remains valid but attention has to be paid to the optimization algorithm that will be used.

## 2.4 Unsupervised optimization

To solve (7), one needs a received sequence  $Y$  as well as the corresponding transmitted one  $X$ . This is usually obtained thanks to the use of a training sequence [20]. However, this induces an increase in the signaling and a decrease in the

useful data rate. Unsupervised optimization is thus attractive since it does not imply any overload. Besides, a good aspect of having such an unsupervised approach is that we optimize the approximation directly from the sequence that we are going to decode. In other words, the noise impacting the training phase and the decoding phase will be the same ensuring the best knowledge of the actual channel state.

Since one needs the sent sequence  $X$  as well as the corresponding channel output  $Y$ , we propose thus to extract a noise sequence  $\tilde{N}$  directly from the received channel output  $Y$  and to simulate at the receiver side the transmission of a known sequence  $\tilde{X}$ . The corresponding channel output is build as  $\tilde{Y} = \tilde{X} + \tilde{N}$ , as depicted in Fig. 1. To do so, we propose to use a sign-detector yielding  $\tilde{N} = Y - \text{sgn}(Y)$ . The simulated channel input is an i.i.d. BPSK random variable that is independent of  $\tilde{N}$ . The optimization parameter  $\theta$  can thus be estimated based on (7) but with the newly generated input and output as

$$\theta^* = \arg \max_{\theta} \hat{C}_{L_{\theta}}(\tilde{X}, \tilde{Y}). \quad (11)$$

Once adapted parameters  $\theta^*$  are obtained, the LLR will be approximated by  $L_{\theta^*}(y)$ , where  $Y$  is the true received sequence over the MBISO channel.

In the next section, we propose to apply our unsupervised LLR approximation optimization to LDPC coding where the noise exhibits an impulsive nature. However, our solution is not limited to these codes, but could be applied to any code families whose decoding relies on the LLR, as for instance convolutional codes or turbo-codes.

### 3 Results and discussion

In this section, we first present our simulation setup. Then, we investigate the accuracy of the unsupervised optimization process and evaluate the BER performance of our proposed demapper. For clarity reasons, we first focus on  $S\alpha S$  noise distributions, while we will extend to Middleton for the robustness study.

### 3.1 Simulation setup

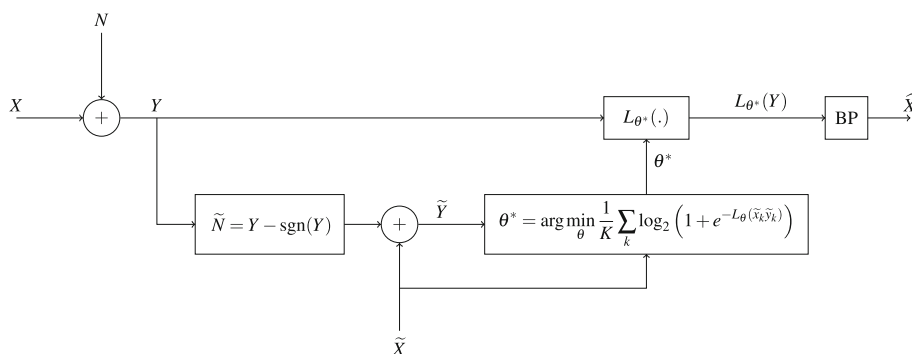
#### 3.1.1 Source

We use an LDPC code associated with a BP-decoding algorithm. This case is well-suited to our proposal because the LLR have to be estimated and fed to the BP algorithm. We refer the interested reader to [21] for a detailed treatment of LDPC codes.

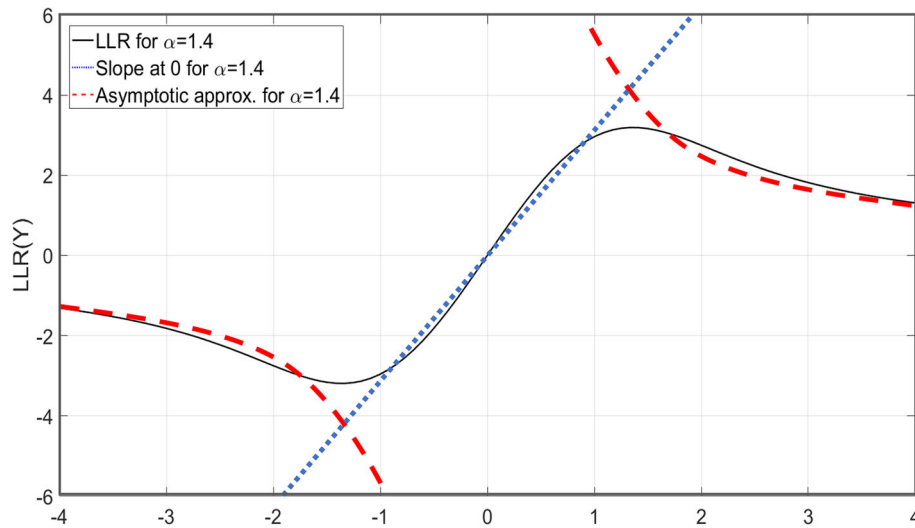
Throughout this paper, we assume that the binary message  $X$  is encoded using a regular (3,6) LDPC code of length 20000. We performed the same study over different LDPC codes and the conclusions are the same.

#### 3.1.2 Non-Gaussian noise

In the following, we assume that the additive noise impacting the transmission exhibits an impulsive nature. In a first step, we will use symmetric  $\alpha$ -stable ( $S\alpha S$ ) distributions to model this impulsive interference, since the heavy tail property of their pdf has been shown to coincide with the impulsive nature of network interference in various environment types [22–25]. One way to define stable distributions is as follows: if for any  $n \geq 2$  there exist a strictly positive constant  $C_n$  and a real  $D_n$  such that the sum of  $n$  independent copies of  $X$ ,  $X_1 + X_2 + \dots + X_n$ , and  $C_n X + D_n$  have the same distribution, then  $X$  is an  $\alpha$ -stable distribution. The finite variance case leads to the central limit theorem and  $X$  being Gaussian ( $\alpha = 2$ ), whereas the non finite variance case leads to the generalized central limit theorem and  $X$  being  $\alpha$ -stable ( $0 < \alpha < 2$ ). The parameter  $\alpha$  ( $0 < \alpha \leq 2$ ) is the characteristic exponent indicating the heaviness of the tail: the smaller  $\alpha$ , the more frequent rare events are, which we call impulsive. If  $X$  is in addition symmetric, only a second parameter is necessary to characterize the distribution: the dispersion  $\gamma$  that plays a similar role to the variance in the Gaussian case. Further details on these distributions can be found in [26] and their interest for network interference can be found for instance in [4]. Unfortunately, in general for  $S\alpha S$  distribution, no closed-form expression of the pdf exists, which prevents the extraction of a simple metric based on the



**Fig. 1** Unsupervised LLR demapper



**Fig. 2** LLR demapper approximations. The following parameters were used in our case:  $\alpha = 1.4$  and  $\gamma = 0.5$

noise pdf in the decoding algorithm. Transmission over an additive  $S\alpha S$  noise channel is thus a perfect example where one can use our proposed LLR approximation.

Since we do not want to enforce restriction on the noise distribution, we will study the behavior of our approach with two other classical noise models: Gaussian and Middleton class A. The latter was proposed by Middleton [13] to model thermal noise plus impulsive interference. This distribution is a mixture of centered Gaussian distributions of increasing variances, whose weights follow a Poisson's law of parameter  $A$  called the impulsive index. The remaining parameters are the total noise power  $\sigma^2$  and the thermal to interference power ratio  $\Gamma$ .

### 3.1.3 LLR approximation under impulsive noise

Note that when the noise exhibits a Gaussian nature, the LLR is given as  $L_{\alpha^*}(y) = \frac{2}{\sigma_N^2}y$ , which belongs to our proposed parametric function. Nevertheless, using only a linear scaling whose slope depends on the additive noise variance leads to severe performance loss as soon as noise is impulsive. This performance loss occurs because with this linear scaling, large values in  $Y$  result into large LLR. However, under impulsive noise, large values in  $Y$  are more likely due to an impulsive event so that the LLR should be small, meaning a less reliable sample due to the presence of a large noise sample.

Figure 2 lightens the non-linearity of the LLR function for the channel output  $Y$  when the noise is  $\alpha$ -stable. Even if Fig. 2 delineates a specific noise model, the overall appearance of the LLR exhibits a similar behavior when noise is impulsive.

At a first look, two different parts in the LLR can be observed: a first one when  $y$  is close to zero and another

one when  $y$  becomes large enough. When  $y$  is close to zero, the LLR is almost linear, whereas when  $y$  is large enough, the LLR presents a power-law decrease. The presence of these two parts has been used in the literature to propose several LLRs [12, 18, 27–29] and justifies the proposed piece-wise affine set for the LLRs approximation.

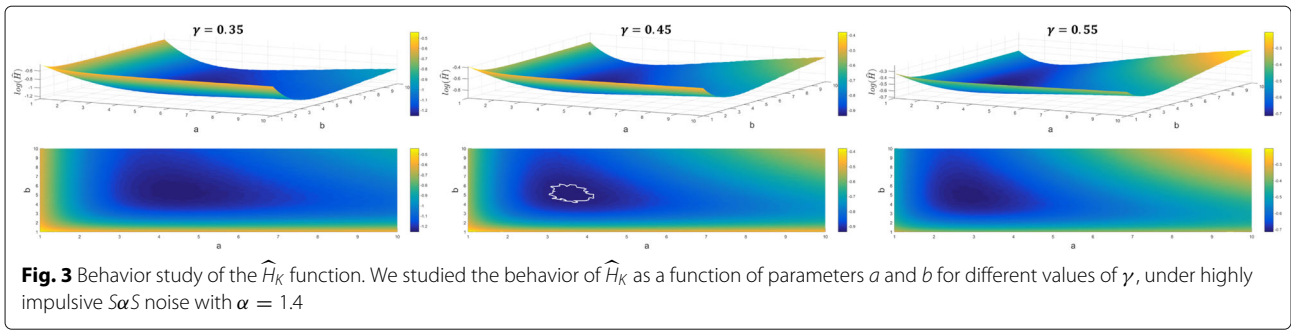
In the remainder of the paper and without loss of generality, we focus on a LLR approximation based on two parameters  $\theta = \{a, b\}$  and  $L_{\theta}(x) = \text{sgn}(x) \min(a|x|, b/|x|)$ , which exhibits performance close to the true LLR [8, 30].

### 3.2 Estimation in additive $s\alpha S$ noise

In a first step, we investigate the shape of the function  $\hat{H}_K$  given in (7). In this paper, we present the obtained results for a highly impulsive noise when  $\alpha = 1.4$ , but similar observations and conclusions would be made for other choices.

In Fig. 3, we represent a 3D plot of the function  $\hat{H}_K$  for three values of  $\gamma$ , namely  $\gamma = 0.35$ ,  $\gamma = 0.45$ , and  $\gamma = 0.55$ , as well as a contour plot representing the levels of  $\hat{H}_K$  under the supervised criterion using a learning sequence of length 20000. The values of  $\gamma$  are selected in a way to represent the shape of the function  $\hat{H}_K$  before, within, and after the waterfall region<sup>1</sup> respectively. First, note that  $\hat{H}_K$  is quite flat around its minimum value. As a consequence, it may be quite sensitive to errors and thus to the length of the training sequence. Using the whole data set in an unsupervised approach can then be a source of robustness.

<sup>1</sup>In iterated sparse graph-based error correcting codes, as for instance LDPC codes, the BER decreases quickly with the increase of the SNR up to a given point, where the performance starts to flatten. The waterfall region corresponds the set of SNRs before the change of speed in the BER decrease.



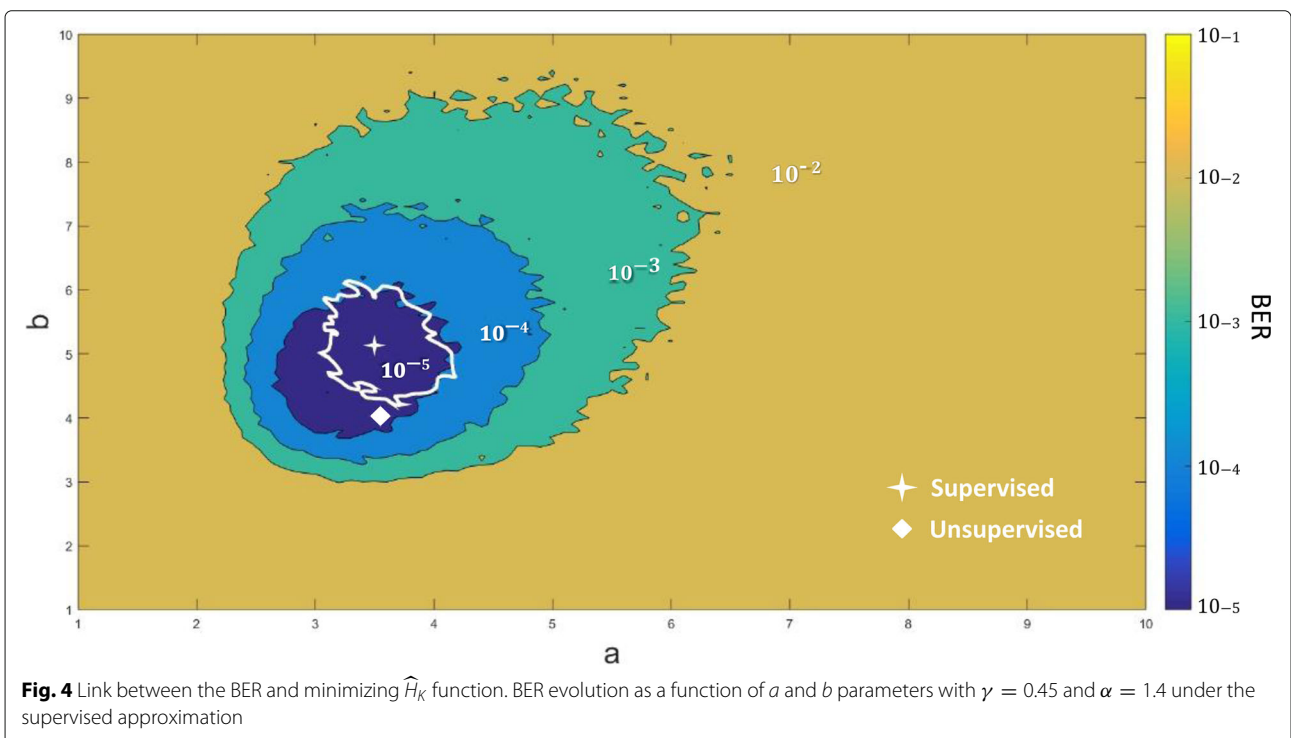
In Fig. 4, we illustrate the link between the function  $\hat{H}_K$  and the obtained BER. The contour plot delineates different BER values, ranging from  $10^{-5}$  to  $10^{-1}$ .

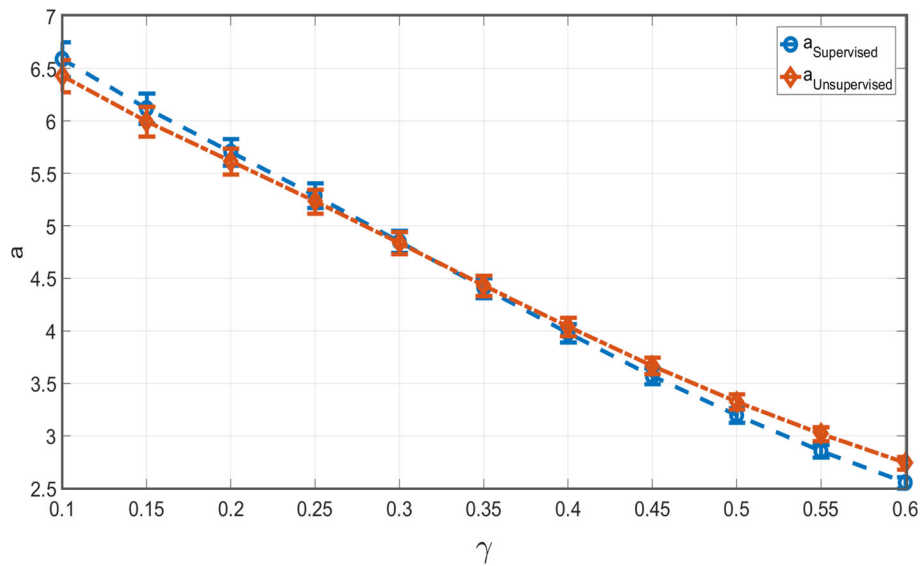
The two white symbols correspond to the mean value of the optimization parameters  $a$  and  $b$  obtained under supervised and unsupervised optimization, respectively, as provided on Figs. 5 and 6. Furthermore, the white contour delineates the set of  $a$  and  $b$  values yielding the smallest values of  $\hat{H}_K$  within a small precision error. First note that the obtained mean values of  $a$  and  $b$  under both types of optimization fall within the set of points achieving a BER less than  $10^{-5}$ . The sensitivity to errors due to the flatness of the landscape of  $\hat{H}_K$  is thus lessened by the flatness of the BER region. Moreover, the set of points minimizing the optimization function  $\hat{H}_K$  belongs to the set of points achieving a BER less than  $10^{-5}$ , which means that using  $\hat{H}_K$  as a criterion turns out to be

a relevant choice robust to lightly non-perfect optimization. Through intensive simulations, we noticed that the connection between  $\hat{H}_K$  and the BER is always assessed, irrespective from the noise model and noise parameters value.

We next evaluate the performance of the estimation process. To perform so, we compare the obtained  $\theta^*$  under unsupervised optimization with the one obtained under a supervised approach. In the latter, instead of building a training sequence  $\tilde{X}$  at the decoder, we directly use the learning sequence to estimate the optimal  $\theta^*$ . More details on the supervised optimization can be found in our previous work [29].

Figures 5, respectively 6, compares the evolution of the mean and variance of the estimated parameter  $a$ , respectively  $b$ , as a function of the dispersion  $\gamma$ , of a  $S\alpha S$  noise with  $\alpha = 1.4$  under supervised and unsupervised opti-



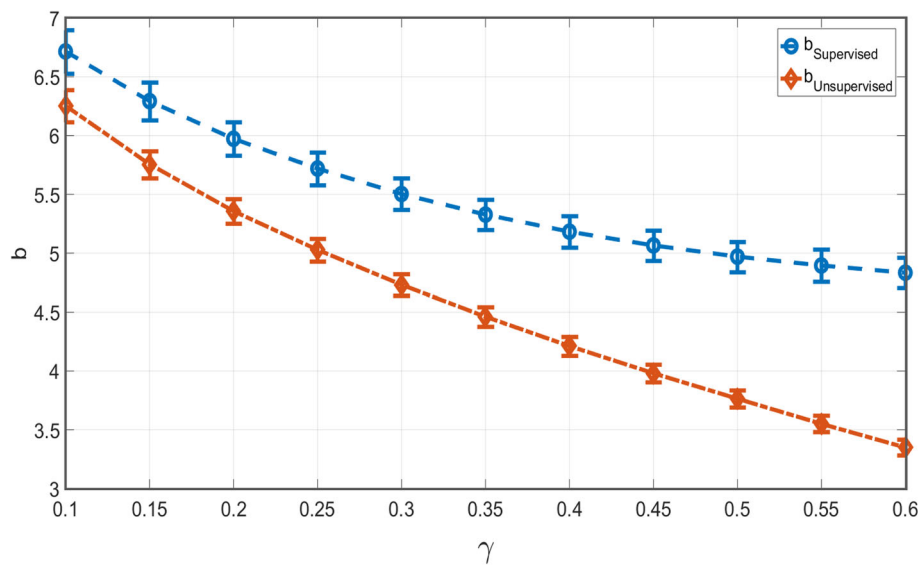


**Fig. 5** Estimation of the  $a$  parameter. Comparison of the mean and standard deviation evolution for parameter  $a$  as a function of the dispersion  $\gamma$  of a  $S\alpha S$  noise with  $\alpha = 1.4$  for the supervised and unsupervised optimization

mization. However, the same behaviors can be obtained for other values of  $\alpha$  between 0 and 2. For each noise dispersion, we ran 5000 experiments. For the supervised case, we use a learning sequence of 20000 samples to estimate  $a$  and  $b$ . This allows to have a good idea of the results with a very small estimation error. In a practical setting, such a long training sequence is not reasonable and additional errors can be expected as the length of the learning sequence decreases, which would benefit our proposal.

We can see from Fig. 5 that the gap between the obtained values for parameter  $a$  under supervised and unsupervised optimization is small. Unfortunately, as shown in Fig. 6, the one obtained for  $b$  is significantly larger. This difference can be explained since  $b$  mainly depends on large noise samples which are rare events; consequently, its estimation is less accurate.

We can however expect that the error on  $b$  will have a limited impact in terms of BER performance. Indeed, as shown in Fig. 4 when  $\gamma = 0.45$ , the estimated mean values



**Fig. 6** Estimation of the  $b$  parameter. Comparison of the mean and standard deviation evolution for the parameter  $b$  as a function of the dispersion  $\gamma$  of a  $S\alpha S$  noise with  $\alpha = 1.4$  for the supervised and unsupervised optimization



**Table 1** Parameter estimation

			$\mu_a$	$\sigma_a$	$\mu_b$	$\sigma_b$
$\alpha = 1.8$	$\gamma = 0.53$	Unsupervised	3.43	0.06	5.73	0.15
		Sup <sub>LS=20000</sub>	3.25	0.05	7.59	0.28
		Sup <sub>LS=1200</sub>	3.27	0.24	8.50	14.48
		Sup <sub>LS=900</sub>	3.27	0.27	11.72	46.15
	$\gamma = 0.55$	Unsupervised	3.23	0.05	5.61	0.14
		Sup <sub>LS=20000</sub>	3.05	0.05	7.62	0.28
		Sup <sub>LS=1200</sub>	3.07	0.22	7.97	1.54
		Sup <sub>LS=900</sub>	3.07	0.26	10.73	30.41

Comparison of the mean and standard deviation evolution for the parameters ( $a, b$ ) as a function of the dispersion  $\gamma$  of a  $S\alpha S$  noise with  $\alpha = 1.8$  for the supervised with different learning sequence sizes and unsupervised optimization

under both the supervised and the unsupervised approach fall in the small BER region. Besides, the small variance of the estimated  $\theta^*$  ensures that most of the estimated values will fall in the region yielding the smallest BER.

To complete the study, we present in Table 1 the influence of the training sequence length. In this table, the mean and the variance of the estimation of the parameters  $a$  and  $b$  are collected for 20000, 1200, and 900 bits long learning sequences and also compared to the unsupervised algorithm. The mean value of  $a$  is only slightly affected by the learning sequence length even for short or moderate length sequences. However, the standard deviation of the estimation significantly increases. On another hand, parameter  $b$  is more volatile and the mean of its estimation varies significantly with the training sequence length. Such variability will affect the performance of the system and degrade the BER, asking for a trade-off between the targeted BER and the sequence length.

### 3.3 BER performance under $S\alpha S$ additive noise

Once our demapper is tuned with the estimated value  $\theta$ , it is used as a front-end to the 20000 bits long regular (3,6) LDPC decoder using the BP algorithm. We study a highly impulsive situation with  $\alpha = 1.4$  and a more moderate case with  $\alpha = 1.8$ .

Figures 7 and 8 present the obtained BER for  $\alpha = 1.4$  and  $\alpha = 1.8$  respectively, as a function of the dispersion  $\gamma$  of the  $\alpha$ -stable noise<sup>2</sup>. In both cases, we compare the BER obtained via the demapping function, either in an unsupervised or supervised manner, to the BER obtained with the true LLR computed via numerical integration. For each channel set, we use a learning sequence of length (1200 or 20000) to optimize  $\theta$  in the supervised case; the

<sup>2</sup>In case of an impulsive environment with  $\alpha < 2$ , the second-order moment of a stable variable is infinite ([22], Theorem 3), making the conventional noise power measurement infinite. Accordingly, we present our simulation results as a function of the dispersion parameter  $\gamma$ , which is used as a measurement of the strength of the  $\alpha$ -stable noise.

long training sequence (20000) allows to assess the optimal performance of the supervised estimation, the shorter one (1200) allows to evaluate the loss due to estimation with more realistic training sequences.

First, we note that the estimation with a long training sequence gives performance close to the optimal LLR which shows the good behavior of our demapping function. Moreover, the unsupervised approach does not perform as well as the supervised one with long training sequence but the gap is not so large and the gain in comparison to a linear receiver is enormous. However, when the training sequence is shortened, the supervised estimation degrades and the performance of the unsupervised approach is then much better.

In order to show the robustness of our proposed demapper  $L_\theta$ , we investigate in the next subsection its performance when the channel exhibits different statistics.

### 3.4 Robustness study of the proposed LLR approximation

We use a linear approximation  $L_a(y) = ay$  and the proposed LLR approximation with  $\theta = \{a, b\}$  and test them under two different configurations:

- Highly impulsive Middleton class A with  $A = 0.1$  and  $\Gamma = 0.1$  taken from [31], where one varies the total noise variance  $\sigma^2$ ,
- Gaussian noise.

Note that in these cases, one can compute the noise variance, thus the numerical simulations can be given as a function of the normalized signal-to-noise ratio  $E_b/N_0$ . For each scenario, we compare the BER performance using the true LLR, obtained via numerical integration to the one using LLR approximations under supervised and unsupervised parameter estimation and Gaussian designed demapper  $L_a$ . For each channel set, in the supervised way, a learning sequence of length 20000 is used to optimize  $\theta$ .

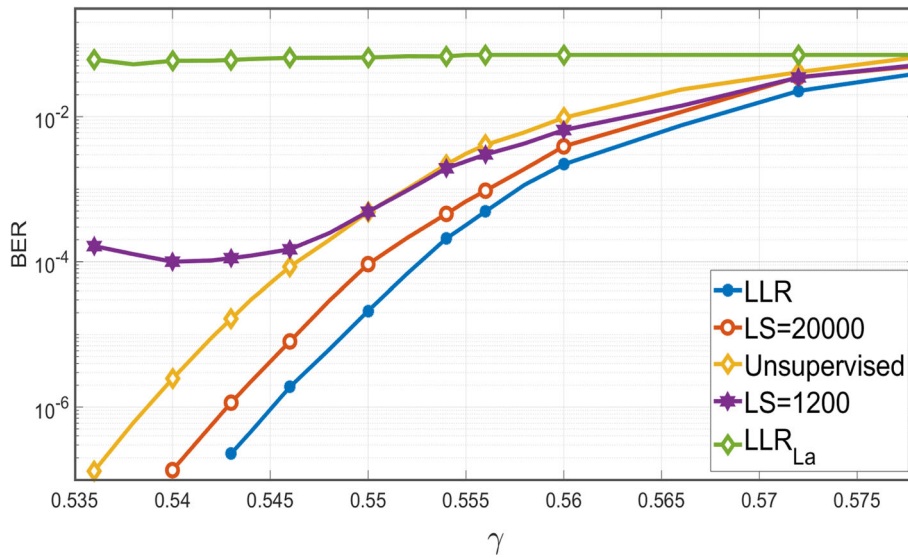
#### 3.4.1 Other additive impulsive noise channels

In the impulsive context, we choose a highly impulsive Middleton class A model. Figure 9 shows the evolution of the BER as a function of the  $E_b/N_0$ .

The high robustness of our demapper can be seen through the close performance obtained between the unsupervised and supervised case from one side and between the approximations and the true LLR from the other side in spite of the change of the type of noise and the degree of impulsiveness.

#### 3.4.2 Additive Gaussian noise channel

It is important that our receiver behaves well even if the noise does not present impulsiveness characteristics. The approximated LLR  $L_\theta$  is thus tested in the presence of a Gaussian noise and the obtained BER is shown in Fig. 10.

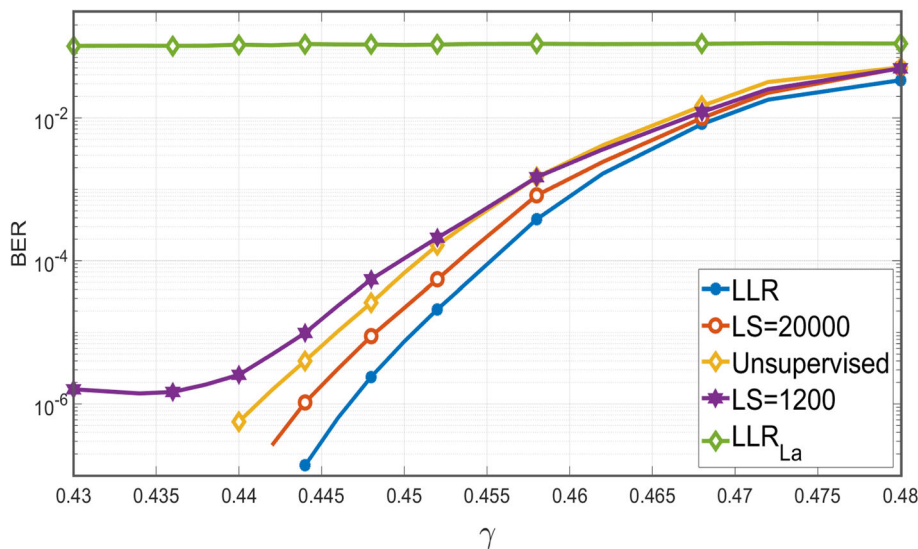


**Fig. 7** BER comparison in low impulsive S $\alpha$ S noise. Evolution comparison of the BER as a function of the dispersion  $\gamma$  of a S $\alpha$ S noise in low impulsive environment with  $\alpha = 1.8$ , between the supervised with different learning sequence sizes, unsupervised, Gaussian designed LLR approximations, and the LLR obtained by numerical integration

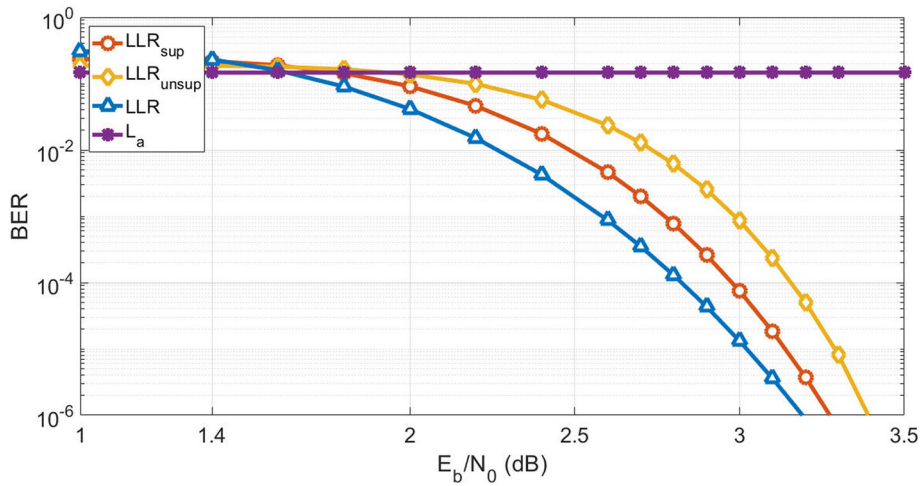
We note that all curves, i.e., the BER obtained under both the supervised and unsupervised optimization, the linear demapper  $L_{a^*}$ , designed for Gaussian noise and the optimal receiver obtained with the true LLR are almost superposed, with a negligible performance loss under the unsupervised optimization. We can conclude that our proposed approach  $L_\theta$  does not degrade the decoding performance in a purely Gaussian case.

**3.4.3 Analysis**

These numerical simulations illustrate the universality of the approach. The LLR family has to be wide enough to be able to represent the linear behavior of exponential-tail noises like the Gaussian and the non-linear behavior of sub-exponential distributions of the impulsive noises. The estimation of the LLR approximation parameter relies on an information theory criteria which does not depend on any noise assumption.



**Fig. 8** BER comparison in highly impulsive S $\alpha$ S noise. Evolution comparison of the BER as a function of the dispersion  $\gamma$  of a S $\alpha$ S noise in highly impulsive environment with  $\alpha = 1.4$ , between the supervised with different learning sequence sizes, unsupervised, Gaussian designed LLR approximations and the LLR obtained by numerical integration



**Fig. 9** BER comparison in highly impulsive Middleton class A noise. BER comparison as a function of  $E_b/N_0$  between the supervised, unsupervised, Gaussian designed LLR approximations and the LLR obtained by numerical integration, in highly impulsive Middleton class A noise with  $A = 0.1$  and  $\Gamma = 0.1$

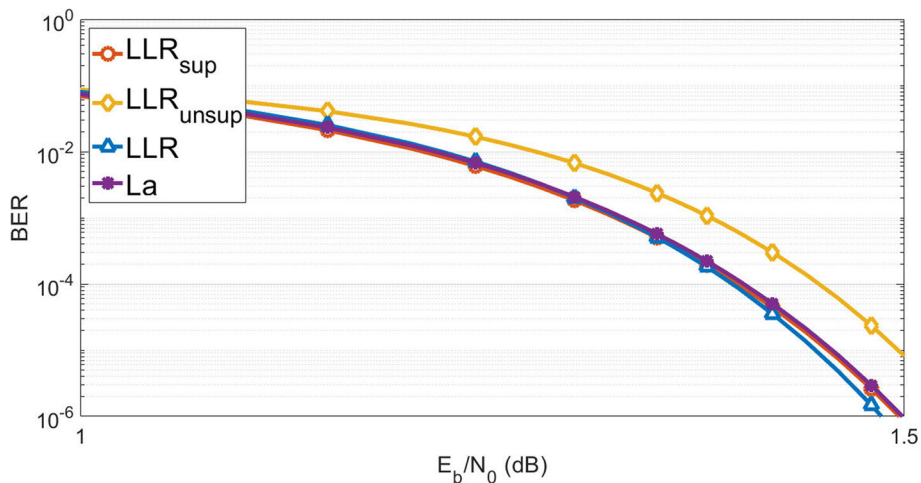
The gap between the unsupervised optimization and the true LLR is small in all the studied examples. We extended this study to other noises parameters or distribution like the  $\epsilon$ -contaminated [32] and obtained similar conclusions. Besides, this gap is partly due to the choice of an approximation function described by only two parameters.

In impulsive situations, the gap between the non linear (with respect to  $y$ ) LLR approximation  $L_\theta$  and the linear receiver  $L_a$  is huge. It proves the influence of handling correctly the impulses that arise due to the presence of interference. Moreover, our demapper function does not significantly impact the performance when noise is not

impulsive, so that we do not need a detection step to distinguish between Gaussian and impulsive situations.

#### 4 Conclusion

We proposed in this paper a receiver design that can adapt to various noise models ranging from very impulsive to non-impulsive by approximating the LLR fed to the iterative decoder. We choose an LLR approximation function  $L_\theta$  in a parametric family. The parameters  $\theta$  are estimated through the maximization of the mutual information. An unsupervised solution is proposed in order to benefit from the whole received sequence and to increase the useful



**Fig. 10** BER comparison in Gaussian noise. BER comparison as a function of  $E_b/N_0$  between the supervised and unsupervised LLR approximations and the optimal LLR in additive Gaussian noise channel

data rate. Our results show that the receiver design is efficient in a large variety of noises and that the unsupervised estimation allows to reach performance close to the optimal and even better than the supervised approach if the training sequence is not sufficiently long.

#### Acknowledgment

This work has been partially supported by COST ACTION CA 15104 - IRACON, IRCICA USR CNRS 3380, and the French National Agency for Research (ANR) under grant ANR-16-CE25-0001 - ARBURST.

#### Abbreviations

AWGNC: Additive white Gaussian noise channel; BER: Bit error rate; BP: Belief propagation; BSC: Binary symmetric channel; LDPC: Low-density parity check; LLR: Log-likelihood ratio; MBISO: Memoryless binary input symmetric output; pdf: Probability density function; pmf: Probability mass function;  $\alpha$ S: Symmetric alpha-stable

#### Authors' contributions

All the authors fully contributed to this paper through many discussions and collaborative works. They all participated to the writing. All authors read and approved the final manuscript.

#### Funding

This work is funded by IMT Lille Douai & University of Reims Champagne Ardenne - France.

#### Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study. All results are from simulations that can be easily reproduced.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>IMT Lille Douai, IRCICA - USR 3380, University of Reims Champagne-Ardenne, CReSTIC - EA 3804, 50 Avenue Halley, 59650 Lille, France. <sup>2</sup>IMT Lille Douai, Univ. Lille, CNRS, UMR 8520 - IEMN, F-59000, 50 Avenue Halley, 59650 Lille, France. <sup>3</sup>CReSTIC - EA 3804, University of Reims Champagne-Ardenne, Moulin de la Housse, 51100 Reims, France.

Received: 13 May 2019 Accepted: 28 November 2019

Published online: 29 January 2020

#### References

- P. C. Pinto, M. Z. Win, Communication in a poisson field of interferers-part II: channel capacity and interference spectrum. *IEEE Trans. Wirel. Commun.* **9**(7), 2187–2195 (2010)
- P. Cardieri, Modeling interference in wireless ad hoc networks. *IEEE Commun. Surv. Tutor.* **12**(4), 551–572 (2010)
- M. Egan, L. Clavier, M. de Freitas, L. Dorville, J. Gorce, A. Savard, in *IEEE GLOBECOM*. Wireless communication in dynamic interference, (Singapore, 2017)
- P. C. Pinto, M. Z. Win, Communication in a poisson field of interferers—part I: interference distribution and error probability. *IEEE Trans. Wirel. Commun.* **9**(7), 2176–2186 (2010)
- D. Fertoni, G. Colavolpe, A robust metric for soft-output detection in the presence of class-A noise. *IEEE Trans. Commun.* **57**(1), 36–40 (2009)
- W. Gu, L. Clavier, Decoding metric study for turbo codes in very impulsive environment. *IEEE Commun. Lett.* **16**(2), 256–258 (2012)
- T. C. Chuah, Robust iterative decoding of turbo codes in heavy-tailed noise. *IEE Proc. Commun.* **152**(1), 29–38 (2005)
- Y. Hou, R. Liu, L. Zhao, in *2014 IEEE/CIC International Conference on Communications in China (ICCC)*. A non-linear LLR approximation for LDPC decoding over impulsive noise channels, (2014), pp. 86–90. <https://doi.org/10.1109/icchina.2014.7008248>
- C. Gao, R. Liu, B. Dai, in *2017 IEEE/CIC International Conference on Communications in China (ICCC)*. Gradient descent bit-flipping based on penalty factor for decoding LDPC codes over symmetric alpha-stable noise channels, (2017), pp. 1–4. <https://doi.org/10.1109/icchina.2017.8330404>
- Y. Mestrah, A. Savard, A. Goupil, Gellé G., L. Clavier, *Robust and simple log-likelihood approximation for receiver Design*, (Morocco, 2019). <https://doi.org/10.1109/wcnc.2019.8886114>
- H. B. Mâad, A. Goupil, L. Clavier, G. Gelle, in *2010 6th International Symposium on Turbo Codes Iterative Information Processing*. Robust clipping demapper for LDPC decoding in impulsive channel, (2010), pp. 231–235. <https://doi.org/10.1109/istc.2010.5613845>
- V. Dimanche, A. Goupil, L. Clavier, G. Gelle, On detection method for soft iterative decoding in the presence of impulsive interference. *IEEE Commun. Lett.* **18**(6), 945–948 (2014)
- D. Middleton, Statistical-physical models of electromagnetic interference. *IEEE Trans Electromagn Compat.* **EMC-19**(3), 106–127 (1977)
- T. J. Richardson, R. Urbkane, *Modern Coding Theory*. (Cambridge University Press, 2008). <https://doi.org/10.1017/cbo9780511791338>
- R. Yazdani, M. Ardakani, Linear LLR approximation for iterative decoding on wireless channels. *IEEE Trans. Commun.* **57**(11), 3278–3287 (2009)
- T. M. Cover, J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley-Interscience, Wiley, 2006)
- J. A. Nelder, R. Mead, A simplex method for function minimization. *Comput. J.* **7**(1), 308–313 (1965)
- H. B. Maad, A. Goupil, L. Clavier, G. Gelle, Clipping demapper for LDPC decoding in impulsive channel. *IEEE Commun. Lett.* **17**(5), 968–971 (2013)
- M. Shao, C. L. Nikias, in *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers vol. 1*. Signal detection in impulsive noise based on stable distributions, (1993), pp. 218–222
- V. Dimanche, A. Goupil, L. Clavier, Gellé G., in *2016 IEEE Wireless Communications and Networking Conference*. Estimation of an approximated likelihood ratio for iterative decoding in impulsive environment, (2016), pp. 1–6
- T. J. Richardson, R. L. Urbanke. The capacity of low-density parity-check codes under message-passing decoding, vol. 47, (2001), pp. 599–618
- N. C. Beaulieu, H. Shao, J. Fiorina, P-order metric UWB receiver structures with superior performance. *IEEE Trans. Commun.* **56**(10), 1666–1676 (2008)
- M. Z. Win, P. C. Pinto, L. A. Shepp, A mathematical theory of network interference and its applications. *Proc. IEEE.* **97**(2), 205–230 (2009)
- H. E. Ghannudi, L. Clavier, N. Azzouzi, F. Septier, P. A. Rolland,  $\alpha$ -stable interference modeling and cauchy receiver for an IR-UWB ad hoc network. *IEEE Trans. Commun.* **58**(6), 1748–1757 (2010)
- E. S. Sousa, Performance of a spread spectrum packet radio network link in a poisson field of interferers. *IEEE Trans. Inf. Theory.* **38**(6), 1743–1754 (1992)
- G. Samorodnitsky, M. S. Taqqu, *Stable non-Gaussian Random processes: stochastic models with infinite variance*. (Chapman & Hall, 1994)
- S. Ambike, J. Ilow, D. Hatzinakos, Detection for binary transmission in a mixture of gaussian noise and impulsive noise modeled as an alpha-stable process. *IEEE Sig. Proc. Lett.* **1**(3), 55–57 (1994)
- T. S. Saleh, I. Marsland, M. El-Tanany, in *2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. Simplified LLR-based viterbi decoder for convolutional codes in symmetric alpha-stable noise, (2012), pp. 1–4. <https://doi.org/10.1109/ccece.2012.6334964>
- Y. Mestrah, A. Savard, A. Goupil, L. Clavier, Gellé G., in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. Blind estimation of an approximated likelihood ratio in impulsive environment, (2018). <https://doi.org/10.1109/pimrc.2018.8580788>
- Z. Mei, M. Johnston, S. Le, L. Chen, in *2015 IEEE/CIC International Conference on Communications in China (ICCC)*. Density evolution analysis of LDPC codes with different receivers on impulsive noise channels, (2015), pp. 1–6
- Y. Nakano, D. Umehara, M. Kawai, Y. Morihoro, Viterbi decoding for convolutional code over class A noise channel. *IEEE ISPLC*, 10 (2017)
- O. Alhussein, I. Ahmed, J. Liang, S. Muhaidat, Unified analysis of diversity reception in the presence of impulsive noise. *IEEE Trans. Veh. Technol.* **66**(2), 1408–1417 (2017)

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.