Robust Communication Systems in Unknown Environments

Yasser Mestrah

Robust Communication Systems in Unknown Environments

Yasser Mestrah

UNIVERSITÉ DE REIMS CHAMPAGNE-ARDENNE

ÉCOLE DOCTORALE SCIENCES DU NUMÉRIQUE ET DE L'INGÉNIEUR

THÈSE

Pour obtenir le grade de

Docteur de l'Université de Reims Champagne-Ardenne

Discipline: Génie informatique, automatique et traitement du signal

---

Yasser MESTRAH

16 Decembre 2019

Robust Communication Systems in Unknown Environments

JURY

Valeriu Vrabie	Professeur, Université de Reims	Président
Charly Poulliat	Professeur, Université de Toulouse	Rapporteur
Philippe Mary	MdC (HDR), INSA Rennes, CNRS	Rapporteur
Emmanuel Boutillon	Professeur, Université de Bretagne Sud	Examinateur
Malcolm Egan	Chargé de recherche, INRIA	Examinateur
Anne Savard	MdC, IMT Lille Douai	Examinateur
Alban Goupil	MdC, Université de Reims	Examinateur
Laurent Clavier	Professeur, IMT Lille Douai	Directeur
Guillaume Gellé	Professeur, Université de Reims	Directeur

Acknowledgments

Living these last few years has been as intense as a roller-coaster ride. I have had my share of ups and downs; late-night studying marathons, the stresses involved in meeting multiple deadlines, the euphoria that comes with a sudden research breakthrough. Some say that graduating with a PhD degree is a self-accomplishment - a direct consequence of the sheer number of arduous hours one has clocked. In my case, I was very fortunate to be associated with amazing people throughout my life. I am a firm believer, that their combined effort and support has propelled me and my career to this point in time. To them, I am forever grateful.

Life has a weird way of unraveling itself. The decisions one takes may inadvertently present surprising opportunities in time. My arrival in France was one such decision. I am indebted to my supervisors Laurent CLAVIER, Alban GOUPIL and Guillaume GELLE. They have my utmost respect and gratitude. Without their support, encouragement and expertise, this work would never have achieved completion. It was a true privilege to be under their tutelage. Over the course of these few years, I am glad to have developed a strong friendship with them. In retrospect, if I had the chance of starting my PhD afresh, I would have definitely chosen the same path all over again.

My special thanks go to Charly POULLIAT and Philippe MARY for their acceptance as my PhD reviewers. Thank you so much for your valuable comments. I also would like to thank Emmanuel BOUTILLON, Valeriu VRABIE, Malcolm EGAN, and Anne SAVARD for their acceptance as my PhD committee members and for the fruitful discussion we had through my PhD defense. In addition, I am equally indebted to my colleagues who were always ready to be either technical or caring depending on the circumstances. Many thanks and appreciation goes out to all my teachers and friends I have had over the years. Thanks for being there!

Where would one be without family? Foremost, I am grateful to God, Ahl al-Bayt and my parents; my mother whose unconditional love, prayers and wishes have been with me since day one. I want to thank her from the bottom of my heart for all the sacrifices she has made for me. To my father, your support and advice were a paramount factor in my decisions. Your influence in my life will always be remembered and cherished. Special thanks to my sisters and brothers, the biological and the in-laws, and of course my sweetheart. They all were my source of constant endorsement and motivation, which kept me focused, and without which I could not have achieved any of the serious goals. Most importantly, I would like to thank my leader "Al Mahdi", who without any doubt would have been so proud to witness this moment. He has been with me, side by side, throughout all the ups and downs; his unique ultimate love and goal has been reflected in all sides of my life.

Abstract

Future networks will become more dense and heterogeneous due to the inevitable increase in the number of communicated devices and the coexistence of numerous independent networks. One of the consequences is the significant increase in interference. Many studies have shown the impulsive nature of such an interference that is characterized by the presence of high amplitudes during short time durations. In fact, this undesirable phenomenon cannot be captured by the Gaussian model but more properly by heavy-tailed distributions. Beyond networks, impulsive noises are also found in other contexts. They can be generated naturally or be man-made. Systems lose their robustness when the environment changes, as the design takes too much into account the specificities of the model. The problem is that most of the communication systems implemented are based on the Gaussian assumption.

Several techniques have been developed to limit the impact of interference, such as interference alignment at the physical layer or simultaneous transmission avoidance techniques like CSMA at the MAC layer. Finally, other methods try to suppress them effectively at the receiver as the successive interference cancellation (SIC). However, all these techniques cannot completely cancel interference. This is all the more true since we are heading towards dense networks such as LoRa, Sigfox, 5G or in general the internet of things (IoT) networks without centralized control or access to the radio resources or emission powers. Therefore, taking into account the presence of interference at the receiver level becomes a necessity, or even an obligation.

Robust communication is necessary and making a decision at the receiver requires an evaluation of the log-likelihood ratio (LLR), whose derivation depends on the noise distribution. In the presence of additive white Gaussian noise (AWGN) the performance of digital communication schemes has been widely studied, optimized and simply implemented thanks to the linear-based receiver. In impulsive noise, the LLR is not linear anymore and it is computationally prohibitive or even impossible when the noise distribution is not known. Besides, the traditional linear behaviour of the optimal receiver exhibits a significant performance loss. In this study, we focus on designing a simple, adaptive and robust receiver that exhibits a near-optimal performance over Gaussian and non-Gaussian environments. The receiver must strive for universality by adapting automatically and without assistance in real conditions.

We prove in this thesis that a simple module between the channel output and the

decoder input allows effectively to combat the noise and interference that disrupt point-to-point (P2P) communications in a network. This module can be used as a front end of any LLR-based decoder and it does not require the knowledge of the noise distribution including both thermal noise and interference. This module consists of a LLR approximation selected in a parametric family of functions, flexible enough to be able to represent many communication contexts (Gaussian or non-Gaussian). Then, the judicious use of an information theory criterion allows to search effectively for the LLR approximation function that matches the channel state. Two different methods are proposed and investigated for this search, either using supervised learning or with an unsupervised approach. We show that it is even possible to use such a scheme for short packet communications with a performance close to the true LLR, which is computationally prohibitive. Overall, we believe that our findings can significantly contribute to many communication scenarios and will be desired in different networks wireless or wired, point to point or dense networks.

Index terms— Impulsive noise, coding theory, information theory, statistical learning and inference, receiver design, detection, soft channel decoding, Log-likelihood ratio (LLR) estimation, supervised learning, unsupervised learning.

Résumé

Le nombre croissant des appareils communicants et la coexistence de réseaux indépendants toujours plus abondants en augmenteront dans le futur la densité et l'hétérogénéité avec pour conséquence une accentuation des interférences. De nombreuses études ont montré leur nature impulsive qui se caractérise par des événements de fortes intensités sur de courtes périodes. Toutefois, ces phénomènes ne sont pas correctement capturés par un modèle gaussien et nécessitent plutôt le recours à des distributions à queues lourdes. Ces bruits impulsifs ne sont pas l'apanage des réseaux et se retrouvent aussi dans d'autres contextes d'origines naturelles ou humaines. Les systèmes perdent leur robustesse lorsque leur environnement se modifie et lorsqu'ils reposent trop fortement sur les spécificités de leur modèle. La plupart des systèmes de communications étant basés sur le modèle gaussien souffrent de tels problèmes en milieu impulsif.

Plusieurs techniques ont été développées pour limiter l'impact des interférences comme l'alignement d'interférences au niveau de la couche physique ou par des techniques d'évitement de transmissions simultanées comme le CSMA au niveau de la couche MAC. Enfin, d'autres méthodes essaient de les supprimer efficacement au niveau du récepteur à l'instar de l'annulation successives d'interférences. Toutes ces techniques ne peuvent parfaitement annuler toutes les interférences ; d'autant plus que nous nous dirigeons vers des réseaux denses comme LoRa, Sigfox, la 5G ou en général l'Internet des objets sans contrôle centralisé ni d'accès à la ressource radio ni aux puissances des émissions. Par conséquent, prendre en compte la présence des interférences au niveau du récepteur devient une nécessité, voire une obligation.

La robustesse des communications est nécessaire et prendre de bonnes décisions au niveau du récepteur requiert l'évaluation du log rapport de vraisemblance (LLR) qui dépend de la distribution du bruit. Le cas du bruit blanc gaussien additif est bien connu avec son récepteur linéaire et ses performances bien étudiées. Les non-linéarités apparaissent avec le bruit impulsif et le LLR devient alors difficilement calculable lorsque la distribution de bruit n'est pas parfaitement connue. Malheureusement, dans cette situation, les récepteurs classiques montrent des pertes de performances significatives. Nous nous concentrons ici sur la conception d'un récepteur adaptatif simple et robuste qui affiche des performances proches de l'optimum sous bruit gaussien ou non. Ce récepteur aspire à être suffisamment générique pour s'adapter automatiquement en situation réel.

Nous montrons par nos travaux qu'un simple module entre la sortie du canal et

le décodeur de canal permet de combattre efficacement le bruit impulsif et améliore grandement les performances globales du système. Ce module approche le LLR par une fonction adéquate sélectionnée parmi une famille paramétrée qui reflète suffisamment de conditions réelles du canal allant du cas gaussien au cas sévèrement impulsif. Deux méthodes de sélection sont proposées et étudiées : la première utilise une séquence d'apprentissage, la seconde consiste en un apprentissage non supervisé. Nous montrons que notre solution reste viable même pour des communications en paquets courts tout en restant très efficace en terme de coût de calcul. Nos contributions peuvent être amenées à être appliquées à d'autres domaine que les communications numériques.

Index terms— Bruit impulsif, détection, rapport de vraisemblance, apprentissage supervisé, apprentissage non-supervisé, décodage de canal souple.

Contents

Chapter 1

Introduction

1.1 Introduction

DIGITAL communication systems are one of the major technology revolutions of the last few decades. In our current societies these systems have to be more and more efficient in different terms: maximizing the spectrum efficiency, quality of service (QoS), mobility, energy consumption... in order to address the needs of many applications that facilitate our livings. These modern applications invaded almost the vast majority of disciplines, the list is too long to go through for instance: medical and healthcare, agriculture, environmental monitoring, security, smart cities, transportation, energy management, commercial usage, city infrastructure, public utilities, oil and gas extraction, artificial intelligence, etc. In short, mobile data communications will include any device, not only those carried by humans. This opens the appetite for many prestigious companies that realize this potential and start rolling to invest in this field, for instance, AT&T, Verizon, Huawei, Google, Apple, Orange, IBM, etc. But also creates many opportunities for startups to engage.

Hence, in the coming few years an extraordinary number of connected devices is expected. The number of these devices will be multiplied by more than 100 compared to nowadays reaching 50 billion by 2020 according to Ericsson's former CEO Hans Vestburg [Ahm19]. Due to the proliferation of these devices, it gives rise in future scenarios to situations where a large number of devices are located in physical proximity, creating hybrid products with large independent traffic quantities and different requirements that share the same radio resources simultaneously. In such situations, the number of competitors for radio resources can be much higher than those manageable by conventional wireless architectures, protocols, and procedures. Thus, radio access networks must evolve towards new paradigms in order to adapt to such contemporary scenarios.

The architecture of radio small cells has gained attention due to the high data capacity that can be realized and the ability to meet users QoS requirements while keeping the end devices costs low [AWM14]. To increase the cell coverage and capacity, small cells include femtocells, picocells, metrocells, and microcells [HM12], that is different area coverage provided by different Base Stations (BS). These areas are becoming more dense and heterogeneous due to the continuous increase of users.

Under the constraint of limited resources at each BS, the interference will increase putting a limit on the development of such techniques.

To capitalize on the knowledge acquired by small cells, Device to Device (D2D) communications need to be within small ranges where they communicate directly without traversing the BS or core network [AWM14]. Differently from the traditional systems where all communications must go through the BS. D2D communications can be classified into two categories inband and outband, where the former occurs on the cellular frequencies and the latter uses an unlicensed spectrum. However, a major issue in the inband D2D communication is the power control and interference management between D2D and cellular users [AWM14]. In outband D2D the interference level is uncontrollable, therefore, QoS cannot be guaranteed in such highly saturated wireless areas making it a challenging task.

In addition to cellular networks, IoT devices are employed in the unlicensed industrial, scientific and medical (ISM) bands. The most two common connectivity technologies for IoT devices being deployed in these bands are Sigfox [NGK16] and long range (LoRa) [SYH17]. In fact, these bands suffer from the spectrum scarcity problem and are shared by many technologies that may overlap over frequency spectrum. As an example, at 2.4 GHz, we can mention different standards: IEEE 802.15.1 (Bluetooth), IEEE 802.11b (Wi-Fi), 6LoWPAN which is the acronym of IPv6 over low-power wireless personal area networks, IEEE 802.15.4 (e.g. ZigBee), etc, leading to a contracted band. As a consequence, this will load the networks with different data traffic patterns and leads to a significant increase in the impact of dynamic interference. This poses a challenge for the way that interference can be handled, either by considering it as noise or by the enhancement of interference mitigation techniques.

Nevertheless, interference will become a fundamental limit in many communication systems due to the densification of communicating devices and the scarcity of available spectrum resources. If we consider the interference as noise, this poses a question about the interference characteristics and the models that can represent it. Is the traditional Gaussian model suitable to represent the statistical nature of this interference? It was shown that the interference observed in these networks is not Gaussian but have an impulsive characteristics [PW10a, ZEC+19]. That means high amplitude interference is significantly more likely than in Gaussian models. The problem is that most of the conventional communication systems implemented nowadays are based on the Gaussian assumption which will be undermined by the impulsive interference phenomena.

Beyond networks, impulsive noises are also found in other contexts. Impulsive noise can be generated naturally or by man-made noise, which is observed in many other different scenarios including: underwater acoustic noise [CPH04], indoor wireless communication systems [BRB93], the background noise of power line communications (PLC) [MGC05], multiple access interference (MAI) in ultra-wideband (UWB) systems [BY09], digital subsrciber line (DSL) transmission [KG95, NMLD02], *ad hoc* networks [HEg07, Car10], in multi-user systems [ZB02], molecular communications [FGCE15], radio frequency interference (RFI) for radars [GSY09], atmospheric noise [HH-56], electromagnetic interference (EMI) [Mid72b, Mid77], multi-path noise in satellite transmission [Nah09], etc.

We focus on this thesis on impulsive noise which is a fundamental limit in many communication systems as aforementioned. The impulsive nature must be taken into account if reliable communications are to be established. While this observation is widely shared, little work has been done on "robust" coded systems in this type of environment. This is the subject of this work which studies solutions to make communications more reliable when the channel state (here the interference distribution) is not precisely known. What strategies, both in terms of coding and decoding, can ensure the reliability of the communications regardless of the level of impulsivity encountered?

1.2 Motivation and challenges

A major consequence of the massive increase in wireless transmissions is heterogeneous systems and the wide variability of environments is the impact of interference. Future networks will face two challenges: robustness and adaptability, with high energy and lifetime constraints.

- **Interference:** increasing the number of objects without more available frequency bands inevitably implies stricter spatial reuse of radio resources. If we try to limit interference (interference alignment [JG12]) or to consider them as a signal (Network Coding [BMR+13]), high complexity arises because of the timing and knowledge of the channel which is needed. Moreover, trying to create systems without interference is a sub-optimal strategy [Cos83]. Alternatively, the optimal approach is to consider the interference as noise and create codes that take advantage of it [Cos83]. However, if they are considered as noise, their statistical nature depends strongly on the environment and they are often not Gaussian but have impulsive characteristics [PW10a, GCA+10]. The problem is that most of the communication systems implemented are based on Gaussian assumption: the capacity is well studied with additive Gaussian noise, but less with impulsive interference; the conventional linear receiver under the Gaussian noise assumption is not suited anymore and new strategies have to be implemented; even the SNR (Signal to Noise Ratio) is not sufficient to represent the link quality and another criterion must be defined.

- **Adaptability:** communicating systems need to be robust against the change in a wide variety of environments (dense or not, static or mobile ...), supporting heterogeneous radio interfaces and different quality of service. This implies, in particular, adapting to different types of noise, more or less impulsive. Systems lose their robustness when the environment changes, as the design takes too much into account the specificities of the model. As a motivation let's take the example of a receiver that receives two versions r_1 and r_2 of the same sample x that takes either $+1$ or -1 values. The usual receiver, based on the Gaussian noise model assumption is resulting in the decision regions shown in case (a) in Figure 1.1 where a linear boundary differentiates a $+1$ from a -1. However, the performance of such a receiver will significantly degrade in the impulsive case where non-linear separations of the decision regions will arise (case (b)).

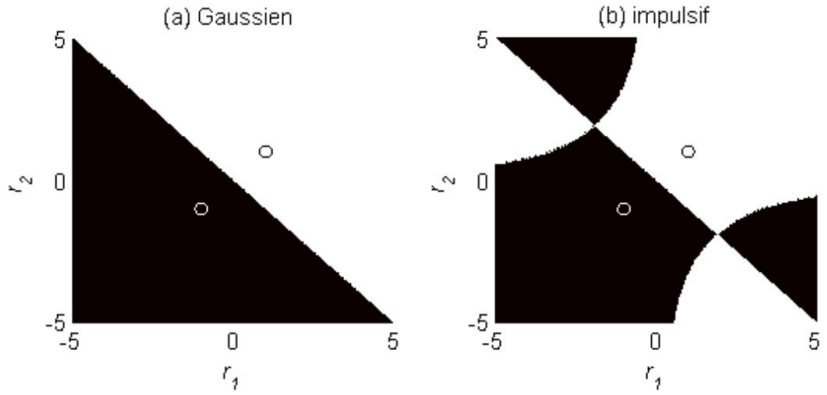


Fig. 1.1. Decision regions when the transmitted bit is repeated twice; in black areas, the best decision is -1 while in white areas, it is $+1$. Case (a) corresponds to additive white Gaussian additive noise and we can see a linear separation of the regions whereas in (b) the noise is impulsive and non-linear region separation will appear.

This visual analysis is confirmed in many papers [GCA+10, BY09] and this is also verified when error-correcting codes are used but not adapted to the noise distribution [GC12, MGCG10a]. If the observation of non-Gaussian interference has been made in many papers, fewer studies are available on error-correcting codes in impulsive environments that are of interest in our study.

The impulsive behavior study was founded in Middleton's work [SM77] and more recently in the results of stochastic geometry and the α-stable distributions [GCA+10]. These works often lead to interference distributions that are difficult to handle in receivers. Indeed the probability density function is sometimes expressed as an infinite series (Middleton) or has no closed-form expression (α-stable). Receivers frequently need the evaluation of the likelihood of the received sequence, which cannot be evaluated simply with such distributions. In literature, different approaches have been considered to overcome these issues. Nevertheless, the approaches are designed for a specific noise model and their robustness against a model mismatch is not ensured [GC12, Chu05, FC09]. The choice of a more universal solution that can be used for various impulsive noise is thus salutary.

Eventually, we have to take into account that in IoT networks, simple end devices have only a limited amount of data to transmit. Due to their limited energy resources, it is important to avoid adding data to be transferred that are not information. This leads to short packets, short training sequence or even, if possible, no training sequence at all.

1.3 Aims, objectives and contributions

In this thesis we aim to design a robust receiver that exhibits a near-optimal performance over Gaussian and non-Gaussian environments without relying on the interference plus noise statistical properties knowledge. We propose to select the LLR in a parametric family of functions, flexible enough to be able to represent many

different communication contexts and to estimate it directly. The great advantage of this approach is that it does not require the knowledge of the noise (including both thermal noise and interference) distribution nor the knowledge of its parameters but only the estimation of the LLR approximation parameters.

We aim to construct an adaptive robust receiver in a way that can be used at the front end of any LLR-based decoder which will be of great advantage as it can be migrated to any LLR-based decoder like LDPC or turbo, thus capitalizing on the knowledge acquired by the development of such types of decoders. The receiver must strive for universality by adapting automatically and without assistance in real conditions. We aim to select efficiently the LLR parameters' with low computational complexity in order to be implemented online. We propose to perform the LLR parameter estimation under an information theory criterion based on the capacity of the Memoryless Binary Input Symmetric Output (MBISO) channel, which can be expressed as a function of the LLR [RU08]. The LLR parameter estimation method will be performed in both supervised and unsupervised manner under long and short block length regimes. This study is attractive for many communication networks either wireless networks or power line communications, point to point or dense networks. For instance, having an unsupervised receiver where the parameters can be estimated efficiently under the short block length regime will be highly desired in the case of IoT networks with short length packets where the receiver in this context must cope with the impact of dynamic interference.

1.3.1 Main contributions of this thesis

The contributions of this thesis brought into focus on how to design a robust receiver that strives for universality without relying on the noise plus interference knowledge. The novelty of this thesis is described as follows

In Chapter 3, we proposed a new framework for robust receiver design directed towards a realization of universal receiver promised by information theory.

- To avoid relying on a noise distribution assumption, we propose to directly estimate a LLR approximation. In doing so, our solution can be used with any LLR-based decoder (LDPC, turbo, convolutional, etc).

- To adapt to many different types of noises, we approximate the LLR by a function chosen in a parametric family. Besides, we consider a family defined by a limited number of parameters and easy to implement, thus, both the estimation and implementation complexities are reduced.

- To adapt to interference that may change in time and space, we propose an online real-time parameter estimation method, which is based on maximizing the mutual information between the channel input and channel output. Furthermore, we demonstrate the rationale behind this approach to obtain the LLR approximation and why it is expected to be efficient and robust.

In Chapter 4, we introduce a supervised online estimation approach where a learning sequence is used to maximize the mutual information.

- To assess the optimal performance that can be achieved using the supervised framework we studied it first under the long blocklength regime, we show an indirect relation between the optimization problem and minimizing the bit error rate.
- We study the robustness of the supervised approach under different noise models. It is shown to be efficient for a large variety of noise distribution. Furthermore, there is no need for a detection step to distinguish between Gaussian and impulsive situations.
- We propose a new LLR approximation that requires the estimation of three parameters. Numerical simulations show that the performance achieved matches the one obtained with the true LLR and outperforms the one obtained with previously proposed solutions.
- We propose to evaluate the quality of the approximation by a mean squared error and ascertain that this criterion is sufficient to identify the approximations that will be efficient instead of using the extensive Monte Carlo simulations.
- We study the effect of shortening the learning sequence and introduce the mismatch risk, meaning that the learning sequence may not match the real condition of the message.

In Chapter 5 we introduce an unsupervised online real-time estimation. It avoids the need of training data that reduces the useful information rate. It also allows taking benefit from the whole data sequence to improve the accuracy of the estimation.

- We study the robustness of the unsupervised approach under a long blocklength regime for different noise models, it exhibits a near-optimal performance in a large variety of noises and it is even better than the supervised approach if the training sequence is not sufficiently long.
- We investigate the impact of reducing the length of the packet on the Bit Error Rate (BER) when the LLR approximation parameters' estimation is unsupervised.
- We analyze the reasons for the degradation that we observed compared to a longer packet case and derive an analytical tool to assess the risk of estimation failure.
- We propose solutions to keep a robust scheme with shorter packets by increasing the diversity in the noise sequence extracted from the received packet and adding a regularization term.

Chapter 2

Theoretical background

With the denser deployment of wireless networks, the induced interference becomes the main system performance limitation, due to the collection of undesired signals broadcasted by other transmitters. If the thermal noise caused by receiver equipment is well modeled by Gaussian distribution, it has been shown in many works that interference exhibits an impulsive behavior [PW10a, GCA+10, SAC04, ECdF+17] as it does not represent the impulsive nature of the interference. In this context, other models must therefore be considered. The study of such models is important for two major reasons. First, from a fundamental point of view, it brings a better understanding of the phenomena. Then, from a practical point of view, it guides the design of communication systems and the choice of communication strategies. This chapter presents noise/interference models often seen in the literature and introduces the dynamic interference characterization. In particular, α-stable model and its properties are studied as well as their interests to model impulsive processes. This is followed by LDPC codes review which will be considered in the rest of this manuscript and a brief of some useful information theory elements.

2.1 Impulsive interference

With the denser deployment of wireless networks, the induced noise and interference becomes the main system performance's limitation on capacity, coverage, etc. By definition, noise is an unsought signal that implicates *unpredictable* perturbations that degrade the desired information. Its source can be separated in many categories [Vas00], for instance, acoustic noise, electronic noise, electrostatic noise, quantization noise, and communication channel, etc. Interference represents the collection of undesired signals broadcasted by other transmitters and added to the desired one. It is created when multiple uncoordinated links share a common communication medium, hence, it is featured as being a special case of *artificial* noise generated by other signals. In the cellular communication context, the frequency planning is used in the sake of improving the spectral efficiency and the system capacity, however, it will induce inevitable interference.

Limiting the impact of interference either by avoiding it or by mitigating it [CAG08, Gol05, CLCC11] is tackled in several works, in the sake of improving the

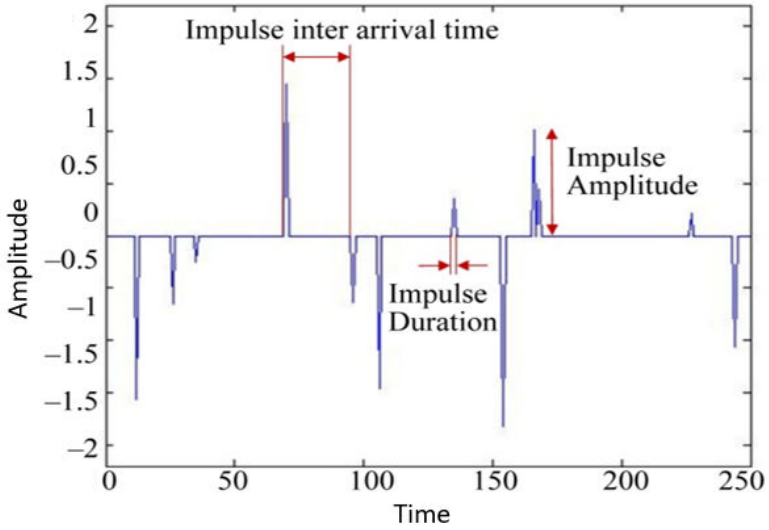


Fig. 2.1. Character of Impulse noise.

system capacity, the node coverage in heterogeneous networks, or any other features in the interest of system designer. Several techniques are proposed to reduce the effect of interference: interference alignment at the physical layer [EPH13], successive interference cancellation [WAYd07, And05], carrier sensing and many other useful techniques at the MAC layer [YV05, JHMB05] that try to avoid simultaneous transmissions, attributing orthogonal resource blocks to different users. However, orthogonal schemes imply coordination in the network, which is not always feasible. For this reason, several techniques have been proposed as utilizing the same time-frequency resource or spatial separation [CBV+09] or the Non-Orthogonal Multiple Access (NOMA) technique.

Creating systems with no interference is an optimal strategy, but in fact, the indispensable increase of transmitting devices may burden such an approach. Another approach can be achieved by designing codes that benefit from the interference. This motivates the researchers to understand the fundamental characteristics in transmissions containing interference.

Understanding the main characteristics of the noise and interference is essential to evaluate its effect on transmission systems. A Gaussian random variable (r.v.) choice is a fundamental model justified by the Central Limit Theorem (CLT). This model is attractive due to the stability property and its simplicity and its analytical tractability. Furthermore, the optimal receiver is linear and easy to implement. Nevertheless, it has been shown in many works that interference exhibits an impulsive behavior [PW10a, GCA+10, SAC04]. The receiver based on the Gaussian assumption exhibit a dramatic performance degradation [PW10a] in such cases. During the last years, many new models have been investigated in various scenarios to handle the interference phenomenon.

Impulsive noise consists of irregular pulses of short duration and high amplitude and may occur in bursts or discrete impulses, as shown in Figure 2.1. Impulsive noise is measured by the number of impulses above a certain threshold over a time interval [KTH95]. It is worth mentioning that each spike has a broad spectral

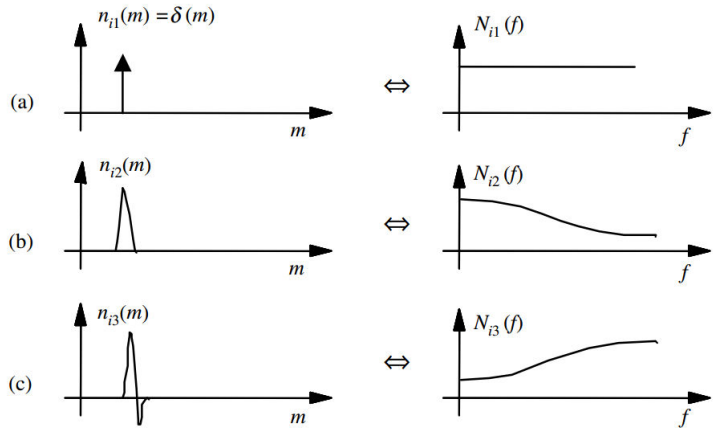


Fig. 2.2. Time and frequency sketches of (a) ideal impulse, (b) and (c) two different short duration pulses.

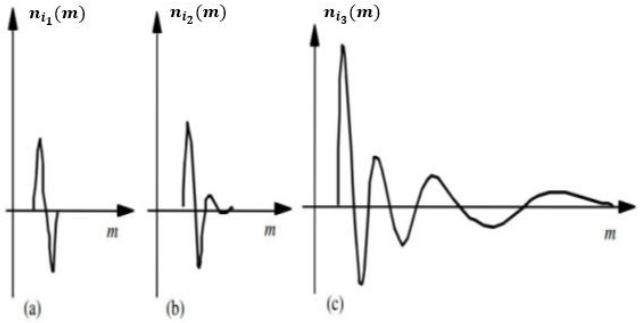


Fig. 2.3. Illustration of variation of the impulse response of a non-linear system with increasing amplitude of the impulse.

content. Figure 2.2 shows two examples of short-duration pulses with their respective spectra, where m indicates the discrete-time index and f indicates the frequency. In a communication system, at a certain point in space and time the impulsive noise will be generated, and then propagates to the receiver through the channel.

The channel will shape the received noise which can be seen as the channel impulse response. Generally, the communication channel characteristics may be variant or invariant, linear or non-linear. Moreover, the response to a large-amplitude impulse in many communication systems will exhibit a nonlinear characteristic as shown in Figure 2.3. This gives a glance behind the effect of large amplitude impulses from a signal processing aspects and allows better understanding. One can define the impulsive behavior as a random variable having a heavy-tailed probability density function (PDF). From a probability theory aspect, heavy-tailed distributions are probability distributions whose tails are not exponentially bounded, in other words, they have heavier tails compared to exponential distributions.

In Figure 2.4 we show the ambient noise detected by an underwater acoustic receiver operating in shallow waters of the coast of Singapore. The collected data was taken from the National University of Singapore [KTMP03] during sea-trials by the Acoustic Research Laboratory. Obviously, the ambient noise exhibits impulsive

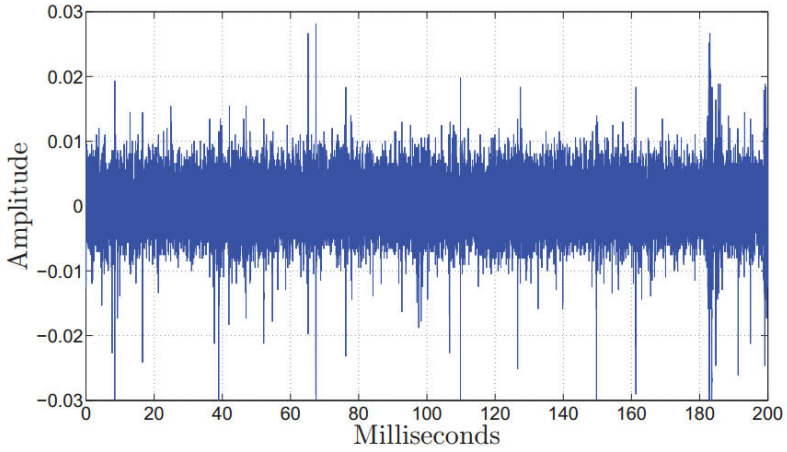


Fig. 2.4. Realization of the ambient noise in Singapore shallow coastal waters.

behavior. As such, in order to establish reliable and efficient communications, one needs to take into account this impulsive nature while designing the receivers. Various models have been proposed for such noises as will be detailed in the following section.

2.2 Impulsive interference models

Some classical noise models encountered in the literature are reviewed in this section. The Gaussian noise model represents accurately the thermal noise in the receiver. However, dealing with dense and heterogeneous networks, i.e. 5G networks, the interference may exhibit impulsive behavior [PW10b, Car10, ECdF+17], and the Gaussian assumption is no longer suited. Thus, several models have been proposed to better take into account the impulsive behavior; we can identify some approaches as:

- **Theoretical approaches:** several works try to derive the interference distribution which can be found mainly in Middleton's work [SM77] and more recently in the results of stochastic geometry [WA12, WPS09]. This last approach, under some assumptions, can lead to the important α-stable distributions [GCA+10, PW10a] that will be considered mainly in this thesis. These works allow a better understanding of the physical underlying phenomena and its link to the model parameters. However, these works often lead to interference distributions that might be difficult to handle in receivers. Indeed the probability density function is sometimes expressed as an infinite series (Middleton) or has no closed-form expression (α-stable).
- **Mixture model approaches:** Due to the difficulties raised by the theoretical approaches, other solutions are proposed based on some mixtures of distributions like the Gaussian-mixture model, generalized Gaussian model, ϵ-contaminated model, etc. Most of these models can be seen as limiting the series in Middleton's work at its main terms (ϵ-contaminated, Gaussian-mixture). The main idea is to have some components of the mixture that will

increase the heaviness of the tails. Mixture models may have good attributes, mainly the closed-form PDF and the existence of the 1st and 2nd order moments. However, they may not truly portray the noise characteristic at the tails [Leg09] and lack the stability property, thus, restricting arguments must be considered in order to characterize the resulting distribution.

- **Empirical approaches:** they are often based on practical choices of a distribution that allow a good fit with generally simulated data, for instance, Pareto model [K60], T-student model [Hal66], etc.

One can note that the aforementioned models cover the main solutions and are not an extensive list of the different impulsive models. In the following of this section, we will detail some of the most commonly used models.

2.2.1 Gaussian

The Gaussian distribution is the most common noise model used in wireless systems [Vas06, BSE10], it is characterized by two parameters: the mean μ and the variance σ^2. Essentially, it appears from external environment sources and the vibration of atoms in conductors, known as the thermal noise. An important theorem, often justifying the good adequacy of this model, is the CLT [Fel70, Dur10]: the noise is obtained by the superposition of a large number of independent contributions. And by the CLT given below tends to be Gaussian.

Theorem 1. ***(Classical CLT)*** *Let $X_1, X_2, ..., X_N$ be a sequence of random variables independently and identically distributed (i.i.d) and let the mean $\mu = \mathbb{E}[X_1]$ and finite variance $\sigma^2 = \mathbb{E}[(X_1 - \mu)^2] < \infty$, then*

$$\frac{1}{\sigma\sqrt{n}}\left(\sum_{j=1}^{n} X_j - n\mu\right) \xrightarrow[n\to\infty]{d} Z \sim \mathcal{N}(0,1), \tag{2.1}$$

To match the following notation, this can be rewritten as

$$a_n(X_1 + \ldots + X_n) - b_n \xrightarrow[n\to\infty]{d} Z \sim \mathcal{N}(0,1), \tag{2.2}$$

where $a_n = 1/(\sigma\sqrt{n})$ and $b_n = \sqrt{n}\mu/\sigma$.

Conceptually, the CLT explains the Gaussian nature of the processes that are generated by the superposition of a very large number of small independent random causes and which present identical probability distributions. For instance, this is the case of the thermal noise, which is generated by the superposition of a large number of independent random interactions at the molecular level. In summary, the CLT imposes that regardless of the X_j distribution, as $n \to \infty$ the sum tends to a Gaussian random variable if X_j are i.i.d. and have a finite variance.

Formally, for a continuous Gaussian random variable X the Gaussian noise PDF is given by

$$f_G(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{2.3}$$

where μ is the mean and σ is the standard deviation. On the other hand, the characteristic function is represented by

$$\phi_G(\theta) = e^{i\mu\theta} e^{-\frac{1}{2}(\sigma\theta)^2}. \tag{2.4}$$

The analytical and tractable forms of the Gaussian model make it a convenient model. Indeed, this model appears as a universal model because of adapting to many situations. Most of the conventional signal processing researches use this model. Among these situations, we find the field of communication theory where many algorithms assume that the studied signals obey a Gaussian law or add to Gaussian noise. This hypothesis generally makes it possible to obtain compact and fast analytical solutions, but it is restrictive.

One limitation is that it is difficult to take into account a large variability in the data. This is due to the fast tail decay of the PDF that can be quantified as

Proposition 1. *Let* $X \sim \mathcal{N}(\mu, \sigma^2)$ *then*

$$\Pr(|X - \mu| > t) \leq \sqrt{\frac{2}{\pi}} \frac{\sigma}{t} e^{-\frac{t^2}{2\sigma^2}}, t > 0, \tag{2.5}$$

meaning that the probability of large samples decays exponentially. As such, the probability of having large values that appear in the interference depicted by samples far from the mean μ are not predicted by the Gaussian model. However, in the field of telecommunications, many phenomena encountered have this important variability impulsive by nature. Examples include noises encountered during transmission on the power grid [ZD02], digital subscriber lines [Coo93], UWB systems [PCG+06], interference in *ad hoc* networks [IH98a], interference in wireless multi-user systems [Sou92a], etc.

In the following, we illustrate main approaches that have lead to impulsive models: analytical approaches, for instance, Middleton class A, B, C models or under some assumptions α-stable model and other empirical approaches based on Gaussian mixtures or mixture of distributions.

2.2.2 Middleton model

One of the first significant models in communication can be seen in Middleton's work who described the phenomenon of impulsive noise [Mid96], where he gave a model for impulsive noise in communications systems.

The nature, the origins, the measurement, and the prediction of the general electromagnetic (EM) interference environment are a major concern of any adequate spectral management program. Middleton divided the origin of impulse noise in two main categories: (1) man-made, which is caused by other devices connected in a communications network and (2) naturally occurring, for instance, due to thunderstorms, and atmospheric phenomena, etc. Nevertheless, most man-made and natural electromagnetic interference are highly non-Gaussian random processes.

General EM noise environments can be conveniently classified into three broad categories vis-à-vis any narrow-band receiver [Mid77]: Middleton class A, class B, and class C.

- **Class A interference** (represents narrowband noise:) originally defined so that the bandwidth of the noise is comparable to, or less than, the bandwidth of the receiving system [Mid72b], now modified to include all noise pulses that do not produce transients in the front end of the receiver [Mid72a]. In particular, class A noise describes the type of electromagnetic interference (EMI) often encountered in telecommunication applications, where this ambient noise is largely due to other, "intelligent"[1] telecommunication operations.

- **Class B interference** (represents broadband noise:) the bandwidth of the noise is greater than the bandwidth of the receiving system, i.e., the noise pulses produce transients in the receiver. So that transient effects, both in the build-up and decay, occur, with the latter predominating. Usually, ambient Class B noise represents man-made or natural "non intelligent"[2] and is highly impulsive.

- **Class C interference:** a combination of Class A and Class B.

Spaulding and Middleton have studied optimum reception of signals in Class A and B noise [SM77, Spa81].

Middleton Class A noise model is one of the most famous models which has been extensively studied and utilized in the literature to become widely accepted to model the effects of impulse noise in communications systems. Moreover, it has been shown that Class A noise accurately model electromagnetic interference (EMI) and background noise, for instance, in Power line Communications (PLC) [AP10], Orthogonal Frequency-Division Multiplexing (OFDM) [III07] and Multiple Input Multiple Output (MIMO) [CGE+09].

Due to its success, we will use this model to evaluate our proposals so that we dedicate space in the following to describe the Class A noise model. The PDF of Middleton Class A model is a Gaussian Mixture Model (GMM) with an infinite number of components

$$f_M(n_k) = \sum_{m=0}^{\infty} P_m \mathcal{N}(n_k; 0; \sigma_m^2), \tag{2.6}$$

where $\mathcal{N}(x_k; \mu; \sigma_m^2)$ represents a Gaussian PDF with mean μ and variance σ^2, and

$$P_m = \frac{A^m e^{-A}}{m!}; \tag{2.7}$$

$$\sigma_m^2 = \sigma^2 \frac{\frac{m}{A} + \Gamma}{1 + \Gamma} = \sigma_I^2 \frac{m}{A} + \sigma_G^2 = \sigma_G^2 \left(\frac{m}{A\Gamma} + 1 \right). \tag{2.8}$$

The noise variance is σ^2 which can be decomposed into two parts, $\sigma^2 = \sigma_G^2 + \sigma_I^2$, where σ_G^2 is the thermal noise power and σ_I^2 is the impulsive noise power. The

[1] Intelligent noise or interference [Mid72b]: is man-made and intended to convey a message or information of some sort;

[2] Non-Intelligent noise or interference [Mid72b]: may be attributable to natural phenomena, e.g., receiver noise or atmospheric noise, for example, or may be man-made, but conveys no intended communication, such as automobile ignition, or radiation from power lines, etc.

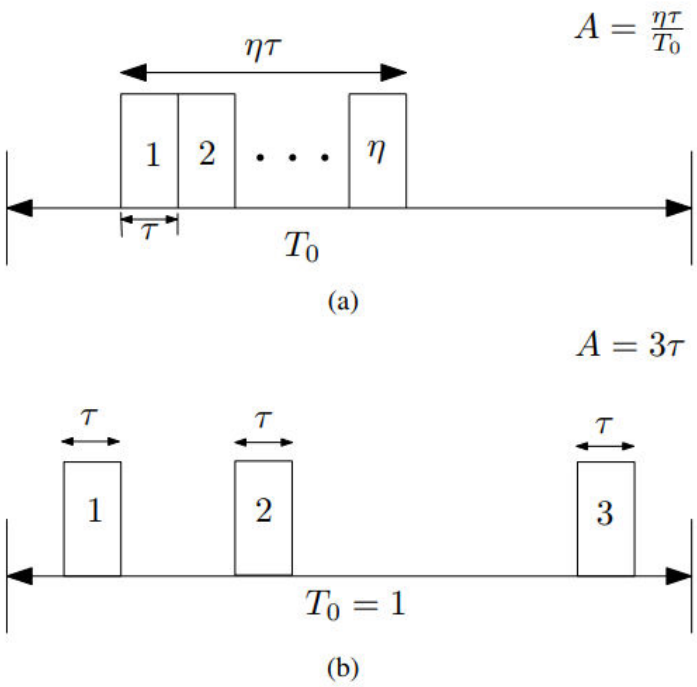


Fig. 2.5. Example of Impulsive index: (a) Impulsive index (density) of η impulses, each with width (duration) τ, occupying a given time period T_0 and (b) impulsive index (density) of 3 impulses, each with width (duration) τ, occupying a given time period $T_0 = 1$.

parameter $\Gamma = \sigma_G^2/\sigma_I^2$ is the ratio between them, which gives the Gaussian to impulse noise power. The parameter A denotes the impulsive index or more recently it is denoted as the overlap index and used to control the impulsiveness as $A > 0$. It is related to the average number of emission "events" that collide at the receiver times the mean duration of a typical interfering source emission [Mid77]. In other words, it represents the density of impulses (of a certain width) in an observation period. Therefore, $A = \eta\tau/T_0$, where η is the average number of impulses per second and $T_0 = 1$, which is unit time. The parameter τ, is the average duration of each impulse, where all impulses are taken to have the same duration. Therefore, instead of talking about the number of impulses we use the density of impulses as shown in (2.7) where we have the density of impulse noise occurring according to a Poisson distribution.

The impulsive index A is rarely well explained in the literature, so we will give some details to enhance its understanding. First note that $A \leq 1$, this follows from the definition of the impulsive index being a fraction of impulses in a given period T_0. Therefore, for $\eta\tau > T_0$, no matter how large $\eta\tau$ is in the observation period T_0, the impulsive index is capped at 1. Moreover, no matter if the impulses occur whether in bursts (next to each other) or not, the calculation of the impulsive index follows the same procedure. To clarify that we show in Figure 2.5 both scenarios. In Figure 2.5 (a) we present η impulses each of duration τ that occur next to each other which is defined as bursts. On the other hand, in Figure 2.5 (b) we present $\eta = 3$ impulses that do not necessarily occur in bursts each of duration τ, we show the

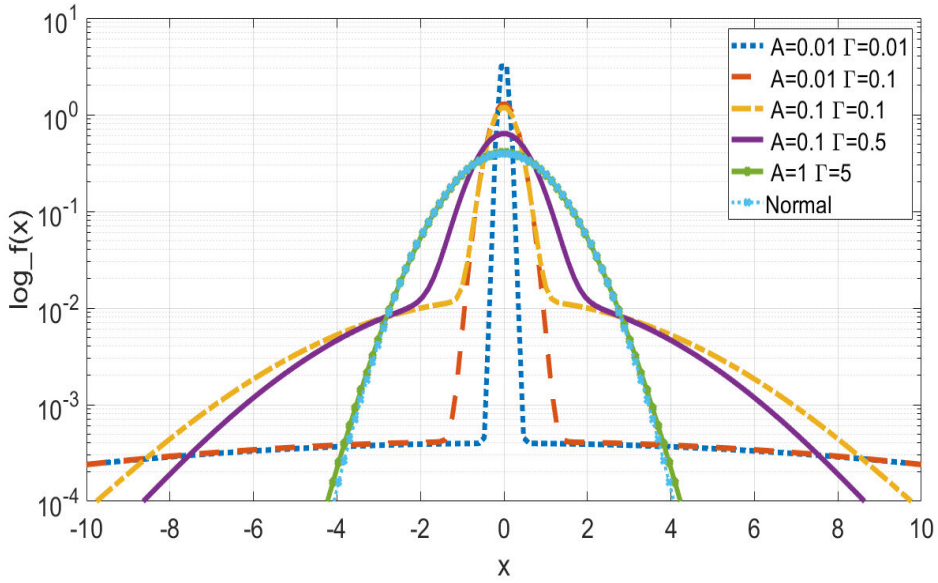


Fig. 2.6. PDF comparison of Middleton Class A noise with different impulsive index A, different Gaussian to impulsive noise power ratio Γ and compared to the normal PDF with $\sigma_G^2 = 1$.

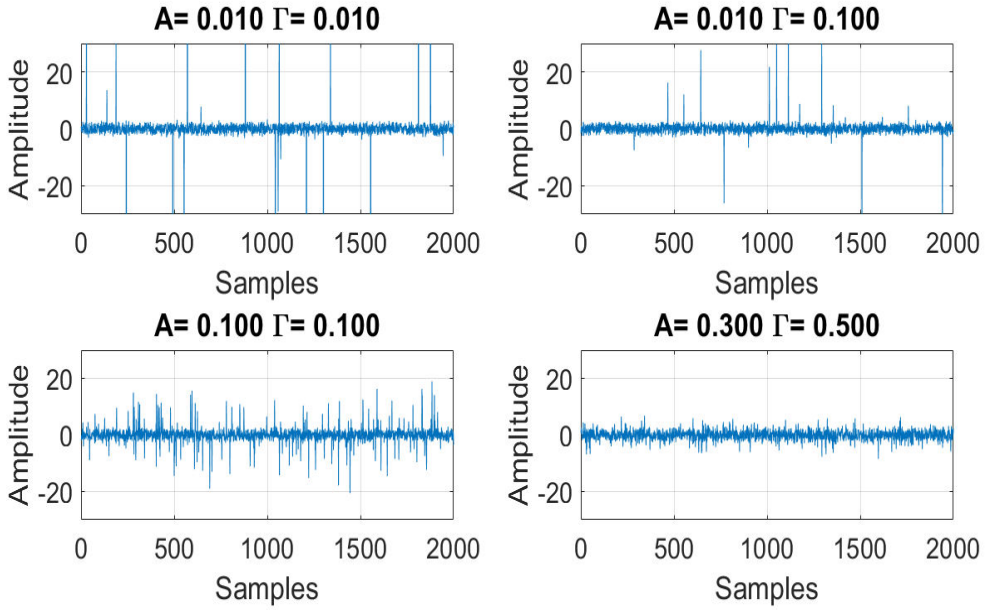


Fig. 2.7. Samples generated from a Middleton class A noise with different A, Γ and $\sigma_G^2 = 1$.

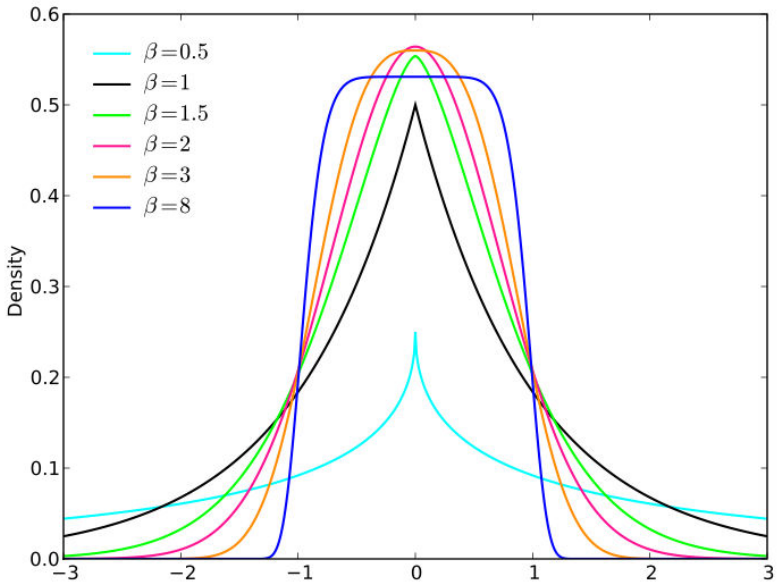


Fig. 2.8. Probability density of a Generalized Gaussian distribution with different β values.

usual case, in the sense that impulsive noise is sporadic by nature, the calculation of A where spread over an observation period $T_0 = 1$. Eventually, one can note from Figure 2.5 that whether the impulses occur in bursts or not, the impulsive index calculation follows the same procedure.

Figure 2.6 compares the PDF of the Middleton Class A with different parameters' to the Normal PDF. The y-axis is given in logarithmic scale to highlight the heaviness of the tail distribution as a function of parameters. The smaller values of A produce impulses with high amplitude samples noise as the tail of the distribution becomes heavier, thus, the probability to have severe impulses increases. Remark that despite the fact that $A = 0$ degenerates into purely Gaussian (see (2.6)), conversely, as A increases the noise tends towards the Gaussian noise.

In Figure 2.7 we generate 2000 samples from a Middleton Class A noise with different values of A and Γ. Figure 2.7 show that as A decreases (except for the special case $A = 0$) the probability to receive extremely large values will increase, however, as Γ increases the Gaussian noise will become dominant compared to the impulsive one.

2.2.3 Generalized Gaussian distribution

The generalized Gaussian distribution (GGD) distribution has been used in several areas such as multiple access interference in UWB systems [BY09], *ad hoc* networks [HEg07], in multi-user systems [ZB02], etc. This distribution is described in several

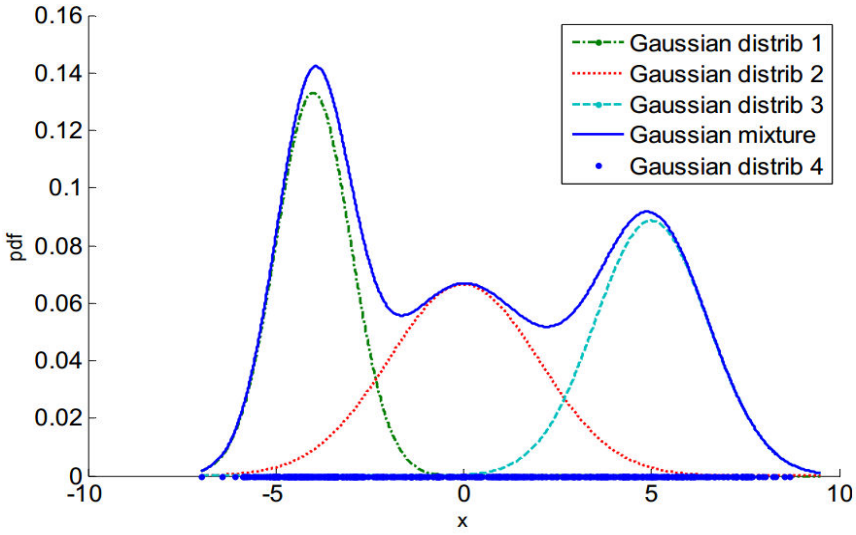


Fig. 2.9. Probability density of a Gaussian mixture distribution

forms [VA89], but we will use the following PDF form

$$f(x) = \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left(-\left|\frac{x-\mu}{\alpha}\right|^{\beta}\right) \tag{2.9}$$

where $\mu \in \mathbb{R}$ represents the location parameter, $\beta \in \mathbb{R}^+$ is the shape parameter or characteristic exponent, $\alpha \in \mathbb{R}^+$ is a scaling parameter and $\Gamma(\cdot)$ is the Gamma function. This family includes different types of distributions: Gaussian ($\beta = 2$), Laplace ($\beta = 1$) and uniform law ($\beta \to \infty$) as shown in Figure 2.8.

One can note that the nature of this distribution depends essentially on the value of the characteristic exponent β. They allow either heavier ($\beta < 2$) or lighter ($\beta > 2$) tails compared to the Gaussian case. Due to this flexibility, they become very useful to model the impulsive noise, however, they remain empirical without any theoretical justifications. In addition, when $\beta < 1$, this distribution inherits the pointed shape or peaky shape of the Laplace distribution. Furthermore, when $\beta > 1$ the tails of the generalized Gaussian distributions are characterized by exponential decay, in contrast to the heavy tails that appear in practice [NS95, WTS88].

2.2.4 Gaussian mixture model

The Gaussian mixture [MT76, Kas88] is a statistical model represented by a density of probability seen as a weighted combination of several weighted Gaussian. We can write it as follows:

$$f(x) = \sum_{k=1}^{K} \lambda_k \mathcal{N}(x|\mu_k, \sigma_k), \tag{2.10}$$

where $\mathcal{N}(x|\mu_k, \sigma_k)$ is a Gaussian density described by a vector of averages μ_k and a variance σ_k and the weighting is given by the coefficient λ_k that satisfies $\sum_{k=1}^{K} \lambda_k =$

1. Figure 2.9 shows the construction of this probability density from four components.

This distribution was used to model the impulsive processes that appear in the transmission systems such as noise due to multi-path in satellite transmission [Nah09], multi-user interference that occurs in UWB systems [ECD08]. The Gaussian mixture is a very flexible distribution. It can represent different situations, symmetric or not, and if a large number of components is accepted the tail can be very precisely controlled. However, this model has some drawbacks, in particular, it requires the estimation of the parameters μ_k, σ_k and λ_k, which can be complex [APDR77].

2.2.5 ϵ-contaminated model

Middleton models have been proven to be difficult to work with in many practical scenarios, due to the infinite series nature as shown in (2.6). For that reason, many approximation models have been proposed such as ϵ-contaminated noise [AB07, DYZB03, AALM17a], Gaussian mixture [GDK06]. Mainly, such approximations consider the most significant terms in (2.6). To obtain a good approximation of Middleton model it is claimed in [Vas84] that two or three terms are sufficient. As a consequence, by taking two terms we can obtain the ϵ-contaminated noise. The main idea of such approximation is to mix two Gaussian models with different weight and standard deviation but the same mean. An ϵ-contaminated mixture PDF is given as:

$$f(x) = (1-\epsilon)\mathcal{N}(x; 0, \sigma^2) + \epsilon\mathcal{N}(x; 0, k\sigma^2), \tag{2.11}$$

where ϵ represents the probability of impulsive occurrence or the contamination level and k represents the impulsive strength. The first term of (2.11) represents the PDF of Gaussian thermal noise with zero mean and variance σ^2. The second term represents the impulsive part with zero mean and larger variance $k\sigma^2$. Hence, we can adjust the impulsiveness of the model by adjusting the weight of the part with largest standard deviation. For instance, as ϵ increases more often large values will appear.

Figure 2.10 compares the density function of ϵ-contaminated distribution with different parameters to the Gaussian distribution. We notice that the y-axis is given in logarithmic scale in order to highlight the effect of the parameters in the tail of the distribution. Figure 2.10 show that the ϵ-contaminated distribution has a heavier tail compared to the Gaussian distribution. Moreover, the tail becomes much heavier as ϵ and k increases which means that it is more likely to receive extremely large values. Figure 2.11 present different sample realizations generated from ϵ-contaminated noise, we can see that as ϵ and k increases the probability to have impulsive samples will increase.

In the next section, we present the α-stable distribution that becomes in the literature a cornerstone to model the impulsive noise.

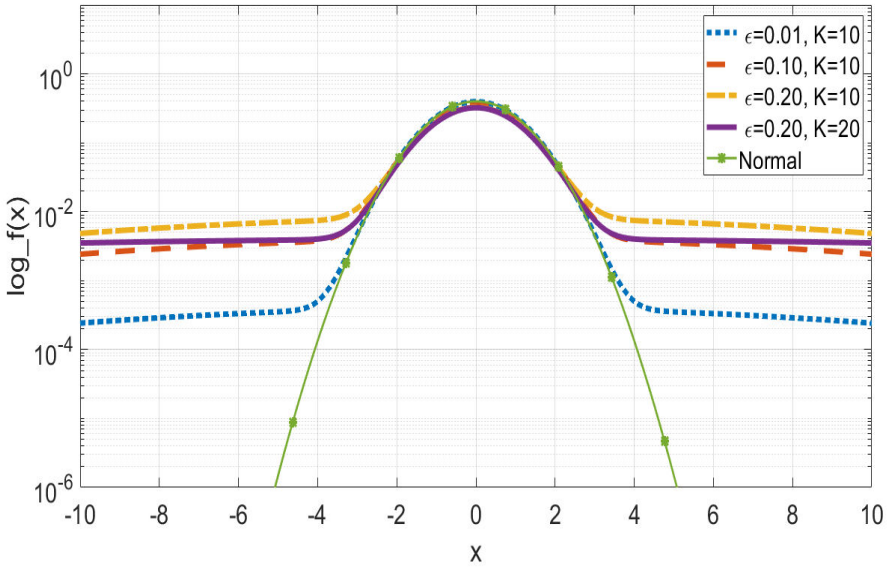


Fig. 2.10. Comparison of the density function of ϵ-contaminated distribution with different ϵ, K and $\sigma^2 = 1$ to Gaussian distribution ($\mu = 0, \sigma^2 = 1$).

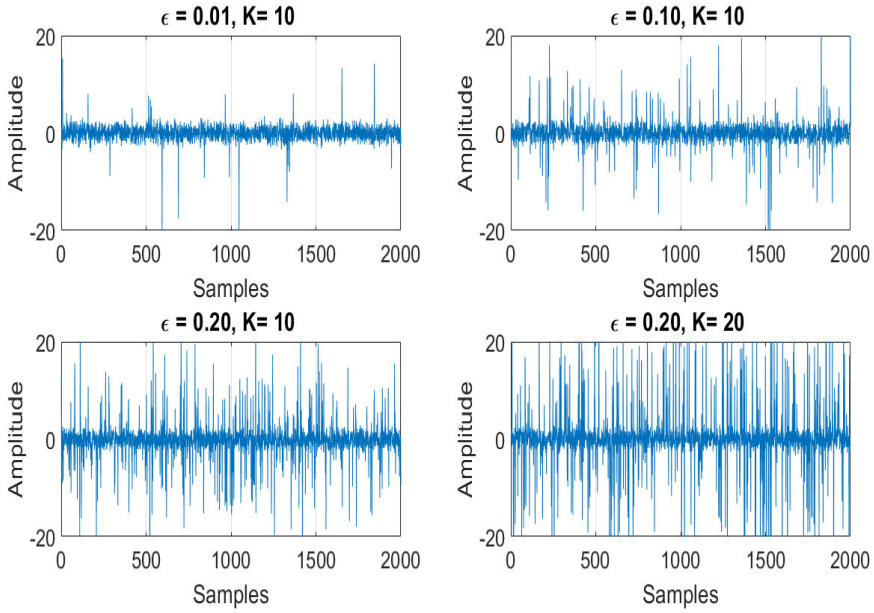


Fig. 2.11. Samples generated from an ϵ-contaminated noise with different values of ϵ and k with $\sigma^2 = 1$.

2.2.6 α-stable distributions

Stable distributions also called α-stable distributions, were introduced by Paul Lévy in the 1920's. They form a very rich class of probability distributions capable of representing skewness, heavy tails and have many fascinating mathematical properties. As a consequence, they are considered to be an important class that models impulsive noise.

This family distribution is used in many areas such as telecommunications [CTB98], finance [BEK98], computer science [KH01], biomedical... Several books are available devoted to them: Zolotarev [Zol86] studied the α-stable laws in the uni-varied context; Samorodnitsky and Taqqu [ST94] have studied in depth several properties of these laws in the univariate case as well as the multivariate case. Nikias and Shao [NS95] applied these laws in the signal processing, such as signal detection and classification, development of optimal and sub-optimal receivers in the presence of impulsive signals and the estimation of the parameters of an α-stable distribution.

One may adopt the α-stable distributions to describe a system essentially for three reasons:

- The existence of theoretical justifications for the adequacy of this model. The recent work by Pinto and Win [PW10a, PW10b] are a good illustration of these justifications, moreover, there are so many in the literature Nolan [Nol97, Sou92b, NS95, ST94]. One can note that, when the radius of the network is large with no guard zones and the active interferer set changes rapidly, the induced true interference can efficiently be approximated by stable distributions [GEAT10, IH98b].

- The Generalized Central limit Theorem (GCLT) which states that the only possible non-trivial limit of normalized sums of i.i.d terms (with or without finite variance) is stable. It is argued that a stable model should be used to describe systems from which the observed quantities are results from the sum of many small terms (i.e. the noise in a communication system, the price of stock, ..., etc).

- The third argument to adopt the stable models is empirical: many large data sets exhibit skewness and heavy tails, combining these features with the GCLT is used by many in the literature to justify the use of stable models. Examples in communication systems which are our concern can be seen by Stuck and Kleiner [SK74], Zolotarev [Zol86], Nikias and Shao [NS95]. Using a Gaussian model, such data sets are poorly described, however, using a stable distribution they can be well described.

The major drawback and challenges to the use of stable distributions is the lack of closed formulas for the densities and distribution functions except for few stable distributions such as Gaussian, Cauchy, Lévy. However, reliable computer programs nowadays allow us to compute densities numerically, distribution functions and quantiles. With such programs, a variety of practical problems can be solved using stable models.

α-stable definitions and properties

One main property of a normal random variable is that the sum of two of them will lead to a normal random variable. Consequently, if X is normal, then for X_1 and X_2 independent copies of it and any $a, b \in \mathbb{R}_{>0}$,

$$aX_1 + bX_2 \stackrel{d}{=} cX + d, \tag{2.12}$$

holds for some positive $c \in \mathbb{R}_{>0}$ and a real number $d \in \mathbb{R}$. The symbol $\stackrel{d}{=}$ means that both right and left expressions have the same probability law (equality in distribution).

Similarly stable random variables are defined

Definition 1. *A stable random variable X is said to have a stable distribution or it follows a stable law if for X_1 and X_2 independent copies of X all positive real numbers $a, b \in \mathbb{R}_{>0}$, (2.12) holds for some positive $c \in \mathbb{R}_{>0}$ and a real number $d \in \mathbb{R}$. Particularly, if (2.12) holds with $d = 0$ for all choices of a and b then X is strictly stable. If X is stable and symmetrically distributed around 0, e.g. $X \stackrel{d}{=} -X$ then it is symmetric stable.*

One can note that the term stable is used because the shape is stable or preserved up to potential shift and scaling under sums of the type (2.12). In fact, there are other equivalent definitions of stable random variables. We will state two of them in the following.

Definition 1 can be extended to a sum of n random variables.

Definition 2. *A non-degenerate Z is stable iff $X_1, \ldots, X_n$ are i.i.d copies of Z and there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that*

$$X_1 + \ldots + X_n \stackrel{d}{=} c_n Z + d_n. \tag{2.13}$$

Lemma 1. *In [P.N18, Section 3.1] it is shown that Definition 2 holds only if the scaling constants is $c_n = n^{1/\alpha}$ for some $\alpha \in (0, 2]$.*

Putting it differently will state what is called the Generalized Central Limit Theorem.

Theorem 2. ***Generalized Central Limit Theorem (GCLT)*** *for some $0 < \alpha \leq 2$, a nondegenerate random variable Z is stable iff there is an i.i.d sequence $\{X_i\}_{i \in \mathbb{N}}$, and constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that*

$$\frac{1}{c_n}\left(\sum_{i=1}^{n} X_i - d_n\right) \xrightarrow[n\to\infty]{d} Z \tag{2.14}$$

or, equivalently,

$$\lim_{n\to\infty} \Pr\left\{\frac{1}{c_n}\left(\sum_{i=1}^{n} X_i - d_n\right) < x\right\} = G(x) \tag{2.15}$$

where for all continuity points x of G, $G(x)$ denotes a non-degenerate random variable Z, in other words, a limiting distribution [Nol97].

Theorem 2 says that the only possible non-degenerate distributions with a domain of attraction are stable. In other words, this theorem states that if the normalized sum of i.i.d. random variables with or without finite variance converge to a distribution by increasing the number of variables, the limit distribution must belong to the family of stable laws. In particular, having a finite variance gives the CLT and the Gaussian limit distribution.

Both Definition 1 and Definition 2 use distributional properties of X or distributional characterization of GCLT. While useful, these conditions do not give a concrete way of parameterizing stable distributions. However, another way to describe stable distributions is done with the characteristic function (CF) [CL97] $\phi_X(t) = \mathbb{E}[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} dF(x)$, where X is a random variable with a distribution $F(X)$.

Definition 3. *A random variable X is stable iff $X \stackrel{d}{=} \gamma Z + \delta$, where $\gamma \neq 0$, $\delta \in \mathbb{R}$ and Z is a random variable with CF*

$$\mathbb{E}[e^{itX}] = \begin{cases} \exp(-|\gamma t|^{\alpha}[1 - i\beta \tan \frac{\pi\alpha}{2}(\text{sign}(t))] + i\delta t) & \text{if} \quad \alpha \neq 1 \\ \exp(-|\gamma t|[1 + i\beta \frac{2}{\pi}(\text{sign}(t)) \log |t|] + i\delta t) & \text{if} \quad \alpha = 1. \end{cases} \tag{2.16}$$

where $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$ and $\text{sign}(t)$ defined as

$$\text{sign}(t) = \begin{cases} 1 & if\, t > 0 \\ 0 & if\, t = 0 \\ -1 & if\, t < 0. \end{cases} \tag{2.17}$$

When $\beta = 0$ and $\delta = 0$, these distributions are symmetric around zero and denoted by SαS for which $X \stackrel{d}{=} -X$. In that case, the CF reduced to a simpler form

$$\phi(t) = e^{-|\gamma t|^{\alpha}}, \qquad t \in \mathbb{R}. \tag{2.18}$$

Except for three different cases (Normal, Cauchy, and Lévy distributions), the density cannot be written in closed form. In practice, this appears to doom the use of stable models.

Parameterizations of stable laws

Definition 3 introduces the four parameters α, β, γ, and δ that characterize a general stable distribution:

- The characteristic exponent or index of stability α: it characterizes the thickness of the distribution tail. For instance, as α increases the probability of observing values far from the central position is interpreted as rare events or impulses decreases. In the wireless context, α is directly associated with the path loss exponent of the radio channel [RSU01a]. By letting $\alpha = 0.5, 1$ and 2, we obtain three special cases: Lévy, Cauchy and Gaussian distributions, respectively.

Parameter	Name	Range
α	characteristic exponent	(0,2]
β	skew parameter	[-1, +1]
γ	scale parameter	$(0, +\infty)$
δ	location parameter	$(-\infty, +\infty)$

Table 2.1 – Parameter description for Stable Distributions

- Skewness parameter β: it measures the skewness of the probability density. If $\beta = -1$, the distribution is totally skewed to the left. If $\beta = 1$, the distribution is totally skewed to the right. For $\beta = 0$, the distribution is symmetric. One can note that when $\alpha = 2$, β has no effect on the distribution (the Gaussian distribution cannot be skewed).

- Dispersion or scale parameter γ: it measures the spread of the noise and is considered as a scale parameter, similarly to the variance for Gaussian distribution, which is a special case with $\alpha = 2$ and $\gamma = \sigma/\sqrt{2}$.

- Location parameter δ: determines the shift of the distribution. Thus, for a given distribution most of the samples are concentrated around this value. For symmetric distributions when $1 < \alpha \leq 2$ it is equal to the mean and to the median when $0 < \alpha < 1$.

A summary of these parameters and their restricted ranges is listed in Table 2.1.

In literature, the notation $X \sim \mathcal{S}_\alpha(\gamma, \beta, \delta)$ is used to represent an α-stable random variable. Particularly, when dealing with symmetric α-stable ($\beta = 0$) the notation is $S\alpha S(\gamma, \delta)$.

Figure 2.12 illustrates the effect of the characteristic exponent parameter α on the α-stable PDF by varying $\alpha = 0.5, 1.0, 1.4, 2$ and for $\delta = 0, \gamma = 1$ and $\beta = 0$. The y-axis is given in a logarithmic scale to highlight the heaviness of the tails for each α. Obviously, as α decreases the tail becomes heavier which delineates a higher probability to receive samples far from the origin. Moreover, for the special case $\alpha = 2$ the tail decreases exponentially. Furthermore, in Figure 2.13, 1000 samples are generated from a standard SαS distribution with different values of α, showing the impulsiveness behavior as α decreases and the special case of when $\alpha = 2$ representing the Gaussian noise.

Figure 2.14 highlights the effect of the skewness parameter β. A standard α-stable PDF is presented, with $\alpha = 1.4, \delta = 0, \gamma = 1$ and the three main skewness values. First, when $\beta = -1$, it depicts a totally left-skewed distribution and consequently, increases the heaviness of the tail for the negative values. Second, when $\beta = 0$ a symmetric distribution is observed. Third when $\beta = 1$, a totally right-skewed distribution increases the tail heaviness for the positive values.

As such, in Figure 2.15 we highlight the effect of γ. The increase in the value of γ increases the variability around the central value.

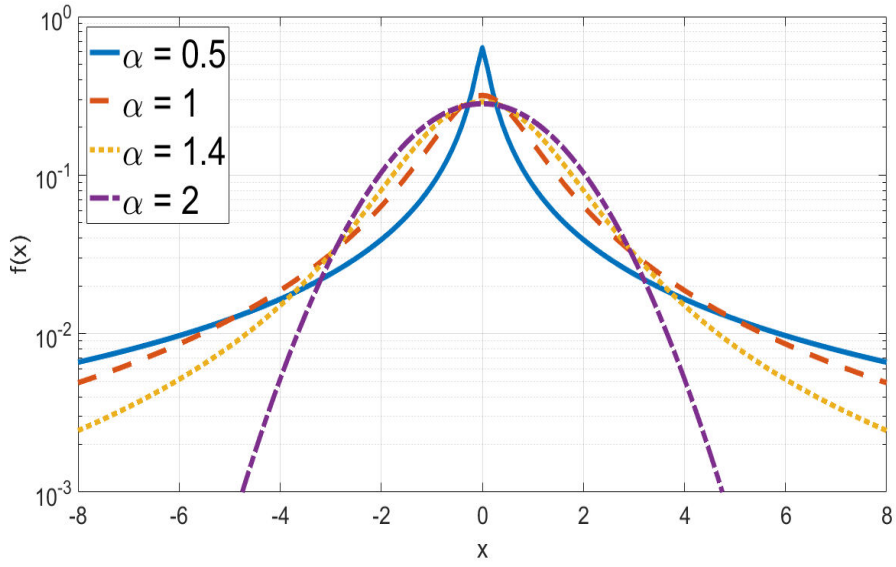


Fig. 2.12. Effect of the characteristic exponent parameter α on the α-stable PDF with $\delta = 0, \gamma = 1$ and $\beta = 0$ and varying $\alpha = 0.5, 1.0, 1.4, 2$.

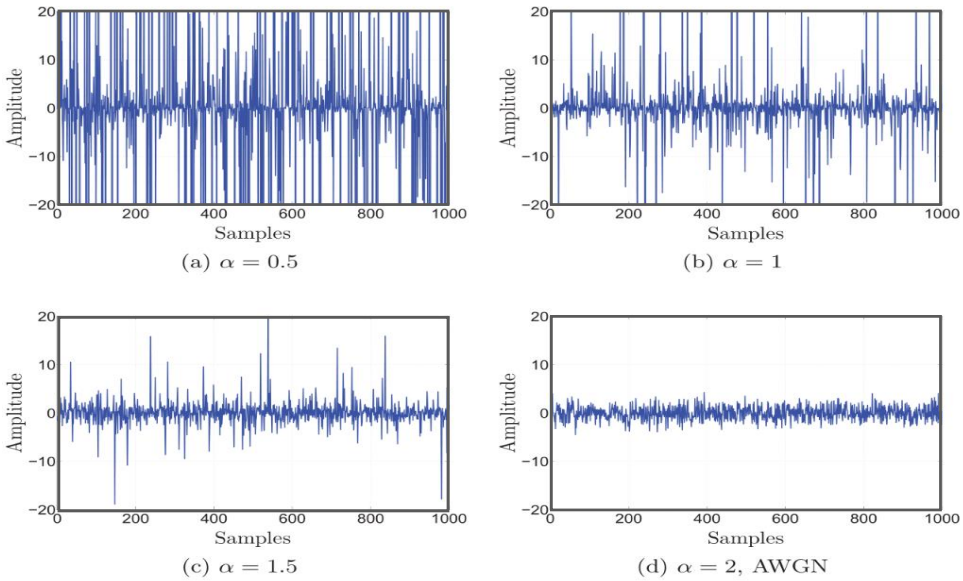


Fig. 2.13. Effect of the characteristic exponent parameter α on the noise Samples generated by an α-stable noise with $\delta = 0, \gamma = 1$ and $\beta = 0$ and varying $\alpha = 0.5, 1.0, 1.5, 2$.

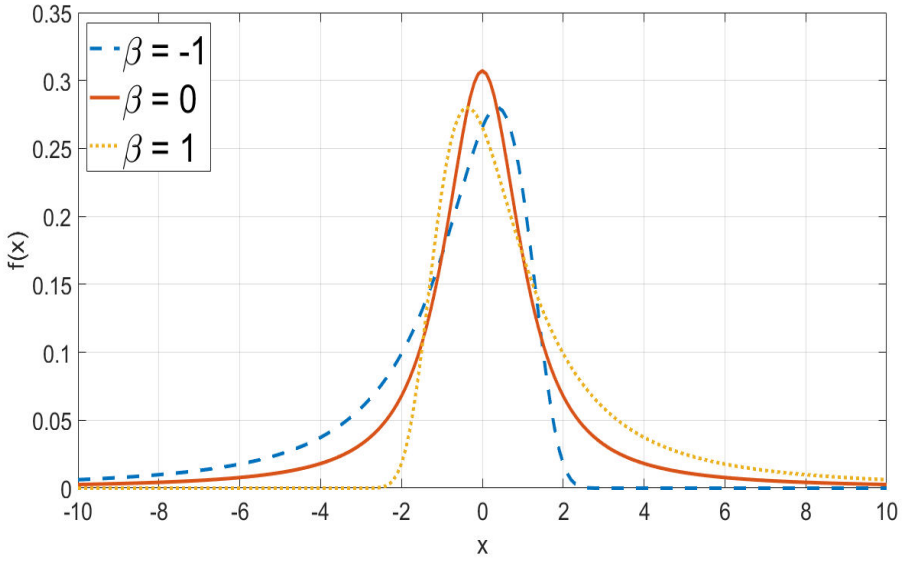


Fig. 2.14. Effect of the symmetry parameter β on the α-stable PDF for $\alpha = 1.1, \gamma = 1, \delta = 0$ and varying the skewness $\beta = -1, 0, 1$.

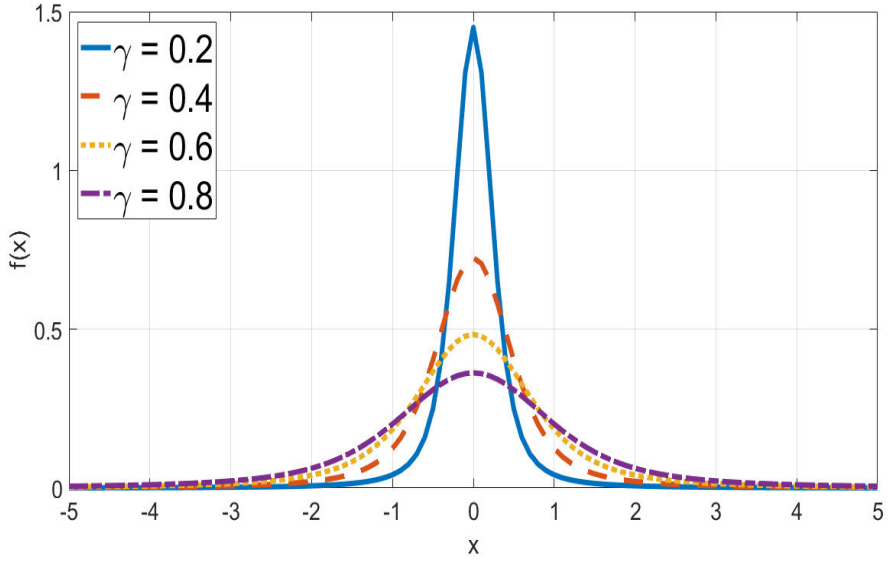


Fig. 2.15. Effect of the dispersion parameter γ on the α-stable PDF for $\alpha = 1.4, \delta = 0, \beta = 0$ and varying the dispersion $\beta = 0.2, 0.4, 0.6, 0.8$.

c	$P(X > c)$		
	Normal	Cauchy	Lévy
0	0.5	0.5	1.0
1	0.1587	0.25	0.6827
2	0.228	0.1476	0.5205
3	0.001347	0.1024	0.4363
4	0.00003467	0.0780	0.3829
5	0.0000002866	0.0628	0.3453

Table 2.2 – Comparison for the tail probabilities of Gaussian, Cauchy and Lévy distributions [P.N18].

Probability density function

As mentioned before that there are three special cases where one can write a closed-form expressions for the PDF:

Example 1. *Normal or Gaussian distribution.* $X \sim S_2(\gamma, 0, \delta) \stackrel{d}{=} S_2S(\gamma, \delta)$ *if it has density*

$$f(x) = \frac{1}{\sqrt{2\pi}\gamma} \exp\left(-\frac{(x-\delta)^2}{4\gamma^2}\right), \qquad -\infty < x < \infty. \tag{2.19}$$

Therefore the normal distribution $\mathcal{N}(\mu, \sigma^2)$ *can be obtained by adjusting the dispersion as* $X \sim S_2S(\sigma/\sqrt{2}, \mu)$*, thus the density will become*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad -\infty < x < \infty. \tag{2.20}$$

and CF becomes

$$\phi(t) = \exp\left(-\frac{\sigma^2}{2}t^2 + i\mu t\right) \tag{2.21}$$

Example 2. *Cauchy distribution.* $X \sim S_1(\gamma, 0, \delta) \stackrel{d}{=} S_1S(\gamma, \delta)$ *if it has density*

$$f(x) = \frac{1}{\pi}\frac{\gamma}{\gamma^2 + (x-\delta)^2}, \qquad -\infty < x < \infty. \tag{2.22}$$

with CF as

$$\phi(t) = \exp\left(-\gamma|t| + i\delta t\right) \tag{2.23}$$

Example 3. *Lévy distribution.* $X \sim Lévy(\gamma, \delta) \stackrel{d}{=} S_{1/2}(\gamma, 1, \delta)$ *if it has density*

$$f(x) = \sqrt{\frac{\gamma}{2\pi}}\frac{1}{(x-\delta)^{3/2}} \exp\left(-\frac{\gamma}{2(x-\delta)}\right), \qquad \delta < x < \infty. \tag{2.24}$$

Table 2.2 compares the tail probabilities for the three aforementioned examples. One can note that the probability is very small above 3 for the normal distribution, but significant for the other two distributions. In particular, there will be approximately on average 100 times more values above 3 in the Cauchy case compared to

the Gaussian. Which illustrates the reason behind calling the stable distributions as heavy-tailed. Moreover, both normal and Cauchy distributions are symmetric, while the Lévy distribution is totally skewed to the right with all the probability concentrated on $x > 0$.

For α-stable random variables, there is no explicit expression of the probability density in the general case. However, we can obtain an expression in the form of the inverse Fourier transform of the characteristic function as

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx)\phi_X(t)dt. \tag{2.25}$$

If X is symmetric,

$$\begin{aligned} f_X(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} e^{-\gamma^\alpha |t|^\alpha} dt. \\ &= \frac{1}{\pi} \int_{0}^{\infty} e^{-\gamma^\alpha |t|^\alpha} \cos(xt)dt. \end{aligned} \tag{2.26}$$

Nikias and Shao presented a simple asymptotic expansion that may well approximate the probability density function f_X of stable variables with unitary scale parameter [NS95].

Proposition 2. ***(Asymptotic Expansion).*** *Let $X \sim S_\alpha(1, 0, 0)$, the corresponding probability density function f_X can be approximated as*

$$f_X(x) = \sum_{k=1}^{n} \frac{b_k}{|x|^{\alpha k+1}} + O\left(|x|^{-\alpha(n+1)-1}\right), \tag{2.27}$$

Corollary 1. *Suppose that $X \sim S_\alpha(1, 0, 0)$ with a probability density function $f_X(x)$, for x large enough, $f_X(x)$ can be written as*

$$f_X(x) \propto |x|^{-(\alpha+1)} \tag{2.28}$$

where (2.28) is based on the first term of (2.27) (see [NS95] for more details).

Tails and moments

The asymptotic tail properties for the normal distribution ($\alpha = 2$) is well understood. A brief discussion is given in the following about the tails of non-Gaussian ($\alpha < 2$) stable laws. When $\alpha < 1$ and $\beta = \pm 1$, stable distributions will have one tail, while for all other cases it will have both tails, these tails are asymptotically power laws with heavy tails.

Proposition 3. *Let $X \sim S_\alpha(\gamma, \beta, \delta)$ with $0 < \alpha < 2$, then*

$$\begin{cases} \lim_{x\to\infty} x^\alpha \Pr(X > x) & = \gamma^\alpha C_\alpha \frac{1+\beta}{2} \quad \textit{right tail,} \\ \lim_{x\to\infty} x^\alpha \Pr(X < -x) & = \gamma^\alpha C_\alpha \frac{1-\beta}{2} \quad \textit{left tail.} \end{cases} \tag{2.29}$$

where the constant C_α depends on α and is given as

$$C_\alpha = \left(\int_0^\infty x^{-\alpha} \sin x dx \right)^{-1} = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha)cos(\pi\alpha/2)} & if\, \alpha \neq 1, \\ \frac{2}{\pi} & if\, \alpha = 1. \end{cases} \tag{2.30}$$

and $\Gamma(\cdot)$ *is the Gamma function defined as*

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-x} dx, \qquad \Re(z) > 0. \tag{2.31}$$

We refer the reader to [ST94][Property 1.2.15] for more details. Obviously, (2.29) show that α-stable densities have heavy tails or inverse power (i.e. algebraic) tails, in contrast, the Gaussian distribution has exponential tails. Polynomial expressions have a slower decay than exponential ones. This proves that the tail of stable laws for $\alpha < 2$ are heavier than those of the Gaussian distribution, moreover, as α decreases the tails become thicker. Formally, heavy tails can be defined as

Definition 4. ***(Heavy Tail)****. Let's consider a distribution with a right heavy tail, then* X *will follow this distribution iff the tail probability* $\Pr(X > x)$ *decay slower than any exponential distribution, such that*

$$\lim_{x\to\infty} e^{\lambda x} \Pr(X > x) = \infty, \qquad \forall \lambda > 0. \tag{2.32}$$

One consequence of heavy tails is that not all moments exist. In stable case, second order moment is infinite (except for $\alpha = 2$). In most statistical problems, integer moments (IM) like the first and second moments are used to describe a distribution. Instead, sometimes it is useful to use fractional lower-order moments (FLOMs), where they represent all moments of order less than α and do exist: $\mathbb{E}|X|^p = \int_{-\infty}^{\infty} |x|^p f(x) dx$, where p is any real number.

Proposition 4. *Let* X *be an* α*-stable random variable. If* $0 < \alpha < 2$*, then*

$$\mathbb{E}|X|^p = \infty, \quad if\, p \geq \alpha \tag{2.33}$$

and

$$\mathbb{E}|X|^p < \infty, \quad if\, 0 \leq p < \alpha, \tag{2.34}$$

however, if $\alpha = 2$*, then*

$$\mathbb{E}|X|^p < \infty, \quad \forall\, p \geq 0. \tag{2.35}$$

Hence, the first and higher-order moments do not exist for $0 < \alpha \leq 1$; on the other hand, $1 < \alpha < 2$, the first-order moment exists, moreover, all the fractional moments of order p with $p < \alpha$ exist as well. Besides, all the non-Gaussian stable distributions have infinite variance. From α and γ, one can find the FLOM's of a SαS random variable as follows.

Theorem 3. *Let* X *be a* $S\alpha S(\gamma, 0)$ *then the FLOMs will have the form*

$$\mathbb{E}|X|^p = C(p,\alpha)\gamma^{\frac{p}{\alpha}} \quad for\, 0 < p < \alpha, \tag{2.36}$$

where

$$C(p,\alpha) = \frac{2^{p+1}\Gamma(\frac{p+1}{2})\Gamma(-p/\alpha)}{\alpha\sqrt{\pi}\Gamma(-p/2)} \tag{2.37}$$

depends only on α *and* p *not on* X*, and* $\Gamma(\cdot)$ *function is given in* (2.31)*.*

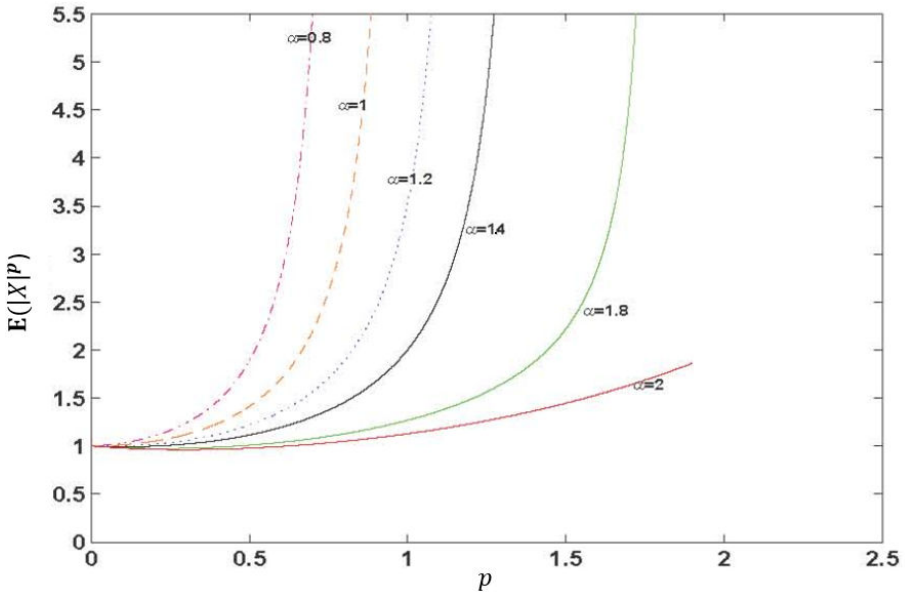


Fig. 2.16. FLOM's for a SαS random variable with different values of α.

Proof. We refer the reader to [SN93] for details. ■

Figure 2.16 shows the moments of order p for different α values corresponding to the (2.36) and (2.37) in Theorem 3.

Many author's [NS95, ST94, Zol86], have investigated the fractional moments to represent in different contexts the density of univariate and bivariate random variables. Indeed, they are important when the density of the random variable has inverse power-law tails and, consequently, it lacks integer order moments. Moreover, FLOMs can be used as a measure of the dispersion of a stable distribution, and are used in some estimation schemes.

Nevertheless, FLOMs has some limitations as they do not provide a universal framework for the characterization of algebraic tail processes: when $p \geq \alpha$, the moments become infinite and the statistical characterization is then no longer possible; In order to overcome this difficulty Gonzalez proposed in [GPA06, Gon97] another approach based on zero order statistics (ZOS) which use logarithmic moments. The latter makes it possible to introduce new parameters that allow characterizing the impulsive processes with an infinite variance which is based on so-called logarithmic process classes [Zol86].

Remark 1. *In the case of an impulsive environment with $\alpha < 2$, the second-order moment of a stable variable is infinite as shown before, making the conventional noise power measurement infinite. However, this does not mean that there are no other adequate measures of variability of stable random variables. Accordingly, in the rest of this thesis, we present our simulation results as a function of the dispersion of a stable random variable γ, which is used as a measurement of the strength of the α-stable noise and plays an analogous role of the variance [BSF08, Theorem 3].*

Parameter estimation

The knowledge of the parameters plays a significant role in the performance of the receiver. In the literature, several approaches exist to estimate them. Due to the lack of analytical expressions most of the methods are based on the law of large numbers and rely on the observed samples at the receiver side. Classical parameter estimation techniques for SαS random variables can be categorized into six main categories [WLGM15]; 1- Fractional lower order moment (FLOM) method [BMA10], 2- Method of log-cumulant (MoLC) [HSSL11], 3- Extreme value method (EVM), 4- Maximum likelihood (ML) [DuM73], 5- Quantile method (QM), 6- Empirical characteristics function (ECF). We discuss the latter three methods.

- **The maximum likelihood (ML) method:** It is known that the ML approach is widely favored due to its generality and asymptotic efficiency [Yu04]. DuMouchel [DuM73] proposed this method and under the assumption of $\delta = 0, \beta = 0$ he obtained a ML estimation of α and γ. He showed that the estimation of the parameters is asymptotically normal and consistent after approximating the likelihood function by multinomial.

- **Quantiles method:** Fama and Roll originate the quantile method in [FR71]. They gave an approximation of the dispersion γ using the sample quantiles properties of the SαS variable. Thereafter, they used the tail property of the α-stable distributions to estimate the parameter α. Later, the work of McCulloch [McC86a] makes the quantile method much more desired after its extension to include asymmetric distributions and for cases where $\alpha \in [0.6, 2]$, unlike the former approach that restricts it to $\alpha \geq 1$.

- **Empirical characteristic function (ECF):** This method was discussed in [Yu04]. Although the likelihood function can be unbounded or not be in a closed-form, its Fourier transform, which gives the CF hence the name of this method, is always bounded and has a closed form expression. The ECF real and imaginary parts will provide estimates of the parameters through given expressions. One can note that such a method provides an alternative way to envision the regression estimation method.

Nolan [Nol97] provides a useful software package that can be used to estimate the parameters of the stable distributions. Extensively, Zolotarev [Zol86] discussed a more theoretical approach to the statistical estimation of the parameters of stable laws. Readers interested in how to simulate stable process can refer to two excellent works of literature, Weron [WW95] and Zolotarev [Zol86].

Among the known techniques for parameter estimation, only QM, FLOM and MoLC yield to a closed form expression [WLGM15, WLF14]. However, it is worth mentioning that they are computationally expensive as they require either tables obtained through linear interpolation as for the QM, or for ECF method one needs Monte-Carlo approach to approximate the CF. Eventually, the burden of these methods is too expensive in terms of time and number of computations or under short sequences the variances of their estimates are high. As a result, they are inconvenient for online estimation, in other words, for applications that require low latency.

In this thesis, we propose a solution to design a robust receiver that does not rely on noise knowledge. Hence, it allows us to overpass the need of the noise distribution parameters estimation which is a considerable advantage. Furthermore, it shows a good performance even with a few number of samples.

Properties and further definitions

Property 1. ***(Stability property).*** *let X_1, X_2 be independent stable r.v. with the same characteristic exponent α as $X_1 \sim S_\alpha(\gamma_1, \beta_1, \delta_1)$ and $X_2 \sim S_\alpha(\gamma_2, \beta_2, \delta_2)$. Then, by adding the two r.v's $X_3 = X_1 + X_2$ we get a new stable r.v. $X_3 \sim S_\alpha(\gamma_3, \beta_3, \delta_3)$ with*

$$\left(\gamma_3 = (\gamma_1^\alpha + \gamma_2^\alpha)^{\frac{1}{\alpha}}\,; \quad \beta_3 = \frac{\beta_1\gamma_1^\alpha + \beta_1\gamma_1^\alpha}{\gamma_1^\alpha + \gamma_2^\alpha}; \quad \delta_3 = \delta_1 + \delta_2\right). \tag{2.38}$$

Corollary 2. *Let $X_i \sim S_\alpha(\gamma_i, 0, 0)$, $i = 1, 2, \ldots, K$, then $\sum_{i=1}^K X_i \sim S_\alpha(\gamma, 0, 0)$ where $\gamma = (\sum_{i=1}^K \gamma_i^\alpha)^{\frac{1}{\alpha}}$.*

Property 2. ***(Scaling property).*** *Suppose $X \sim S_\alpha(\gamma, \beta, \delta)$, then scaling X by a non null constant $b \in \mathbb{R}_{\neq 0}$ will give*

$$\begin{aligned} bX &\sim S_\alpha\left(|b|\gamma,\ \mathrm{sign}(b)\beta,\ b\delta\right), && \text{if} \quad \alpha \neq 1 \\ bX &\sim S_1\left(|b|\gamma,\ \mathrm{sign}(b)\beta,\ b\delta - 2/\pi\gamma\beta b(\ln|b|)\right), && \text{if} \quad \alpha = 1 \end{aligned} \tag{2.39}$$

Corollary 3. *Suppose $X \sim S_\alpha(\gamma, \beta, \delta)$ where $1 < \alpha < 2$, then multiplying X by a real positive constant $b \in \mathbb{R}^+$ gives*

$$bX \sim S_\alpha(b\gamma, \beta, b\delta). \tag{2.40}$$

The following property concerns the asymptotic behavior of the probability density function of a SαS distributions as presented by Samorodnitsky and Taqqu [ST94].

Property 3. ***(Asymptotic behavior).*** *Let $X \sim S_\alpha(\gamma, 0, \delta)\}$ then the probability density function $f_X(x)$ satisfies*

$$f_X(x) = \frac{\alpha(1-\alpha)\gamma^\alpha}{\Gamma(2-\alpha)\cos(\frac{\alpha\pi}{2})}|x|^{-\alpha-1} \quad \text{as } |x| \to \infty. \tag{2.41}$$

2.3 Physical mechanism that leads to α-stable model

We claimed before that one of the main reasons to adopt the α-stable model is the theoretical justifications behind it. In this section, we will demonstrate that an α-stable model may arise from a realistic physical mechanism under some reasonable assumptions and conditions. The transmission in a common band between networks with different protocols, multiple applications, etc, will end up with different data type and different symbol durations, these types of networks are known as heterogeneous networks [AGM+15]. In particular, the network heterogeneity comes out in *ad-hoc* networks and small-cells due to the variations in transmit power constraints and the varied placement of the base stations.

Two key physical mechanisms can create dynamic interference:

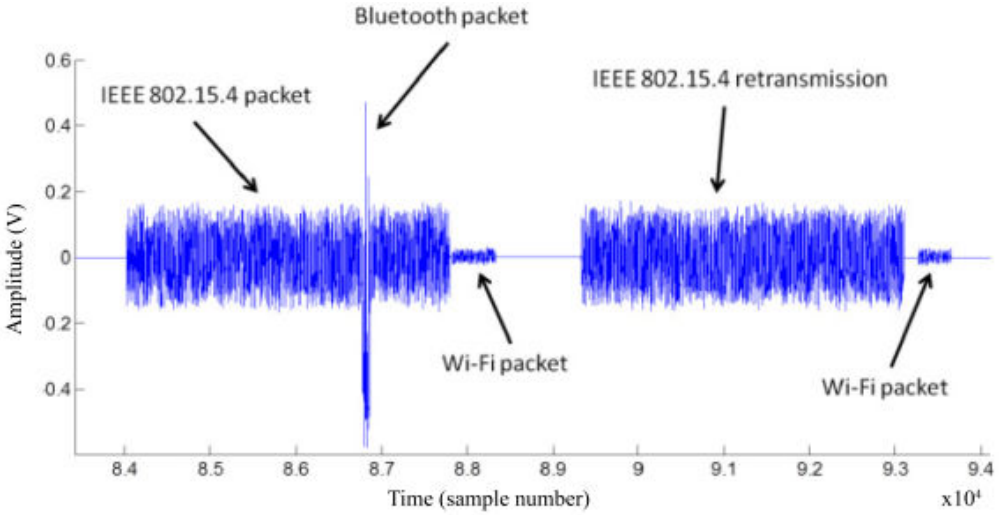


Fig. 2.17. Coexistence of technologies in the 2.4-GHz ban. Measurements made by the National Instruments USRP [Per06].

- Dynamic interference can be a consequence of the rapid change in time of the active set of transmitting devices. This rapid change depicts protocols with non-contiguous transmitted data blocks, meaning that interferers do not transmit data continuously. A good example is a machine to machine (M2M) communications which transmit rare and short packets [Dig09]. In addition, the coexisting of different PHY and MAC layer protocols will arise heterogeneous networks, for instance, several protocols rely on the new ALOHA approach as Sigfox [VLN+17]. Besides, the non-orthogonal multiple access (NOMA) technique which is the most promising radio access technique in the next-generation of wireless communications [SKB+13]. Therefore, when considering the change at the symbol level in these networks a non-Gaussian interference will be induced [ECZ+18].

- Another mechanism that creates dynamic interference arises when multiple coexisting communication systems share the same band, for instance, the ISM band. The uncoordinated access between the different networks and even in some networks like Sigfox and LoRa (using ALOHA) arises a collision between the frames. For example, the Wi-Fi frame on-air time is 42.07 μs (fixed 34 bytes header; 250 bytes of payload; data rate of 54 Mbps); while the Zigbee frame on-air time is 1.28 ms (40 bytes with a data rate 250 kbps). Furthermore, the Bluetooth utilizes frequency-hopping and for a very short time, it appears in the 802.15 bands. Figure 2.17 illustrates the second mechanism where an experimental study with the coexisting Wi-fi, Bluetooth and Zigbee transmissions is presented. Due to the low data rate of the Zigbee technology, a considerable on-air time for the Zigbee packet is observed followed by a small Wi-fi packet. A Bluetooth packet interferes with the Zigbee packet and corrupt it even if it is much smaller. Thus, a re-transmission is required for the corrupted packet we refer the interested reader to [Per06] where a detailed explanation can be found. The short time presence of the Wi-Fi and

Bluetooth interferers compared to Zigbee transmissions will result in dynamic interference which means independent in time.

By considering a large scale wireless communication networks composed of K access points and K devices, where each device transmit data to a unique access point. Several works assert that the channel is characterized by a memoryless additive isotropic α-stable noise (AIαSN) [PW10a, PW10b, ECdF+17].

2.4 Low-Density Parity-Check codes

The aim of channel coding is to make the noisy channel behave like a noiseless channel. The main idea behind forward error correction (FEC) or channel coding is to increase the message length with deliberately introduced repetition, by adding *extra check* bits to produce a *codeword* for the message. In other words, the check bits are added on purpose of having distinct codewords, thus, any corrupted bits within a codeword induced by noisy channel can be inferred correctly at the receiver side.

2.4.1 History and definitions

Low-Density Parity-Check (LDPC) codes first proposed in the PhD dissertation of Gallager in 1960 and then published in [Gal63] are FEC codes. LDPC codes were mostly ignored for a long time over 35 years, due to the computational limitation of the hardware. More precisely, the computational demands of simulations in an era when vacuum tubes were only just being replaced by the first transistors. Meanwhile, the development of Reed-Solomon codes in addition to highly structured algebraic block and convolutional codes dominate the field of forward error correction. Despite the huge practical success of these codes, their performance fell way far from the theoretical achievable limits set down by Shannon in his seminal 1948 paper [Sha48]. One can note that by the late 1980s, most of the researchers resigned to this seemingly insurmountable theory, despite decades of attempts. In 1981, Tanner [Tan81] proposed a graphical representation of the LDPC codes known as Tanner graphs.

Berrou, Glavieux, and Thitimajshima transformed thoroughly the relative quiescence of the coding field by introducing "turbo codes" in 1993. Turbo-codes are based on two concepts: the concatenation of two sub-codes and decoding using feedback. Thereafter, all the main components of successful error-correcting codes were replaced: turbo codes focus on average rather than worst-case performance, employ iterative, distributed algorithms, rely on soft information extracted from the channel. Using such decoders with manageable complexity, the gap to the Shannon limit was all but eliminated. Eventually, turbo-codes paved the way for corrective codes to approach the Shannon limit with manageable length codes and decoding that can be implemented effectively. As researchers struggled just to understand the successfulness of turbo codes through the 1990s, MacKay and Neal, introduced a new class of block codes in 1996 [MN96] where most of the turbo code features are possessed. In fact, it was recognized that such block codes are just a rediscovery

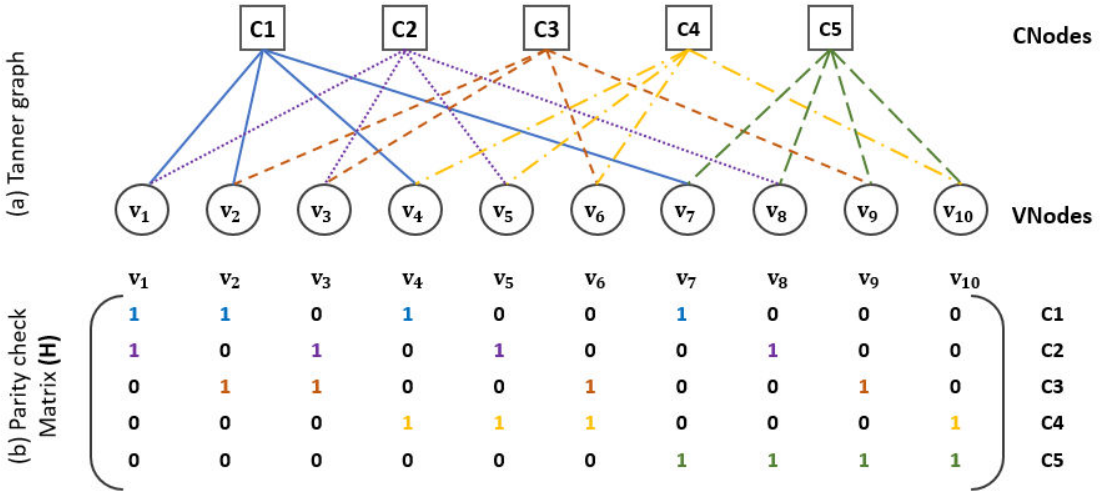


Fig. 2.18. Illustration of a parity check matrix and tanner graph representation.

of the LDPC codes developed by Gallager [Gal63]. Indeed, the decoding algorithm for turbo codes was shown to be a special case of the decoding algorithm for LDPC codes.

Subsequently, irregular LDPC codes were introduced which easily outperform the best turbo codes and they can be seen as a generalization of the regular LDPC codes. Moreover, they offer some practical advantages and a disputable cleaner setup for theoretical results. MacKay proposed Non-Binary (NB) LDPC codes over finite fields [DM98]. It was shown that NB-LDPC codes outperform the binary LDPC codes, but with a higher decoding complexity. Consequently, many works are proposed to reduce such complexity [GCGD07, PFD08, BD03, WSM04, DF07, Sav08]. Nowadays, several design techniques for LDPC codes exist which gives the power to construct codes that approach within hundredths of a decibel to Shannon's capacity [SFRU01].

In addition to the strong theoretical interest in LDPC codes, such codes are widely used in various applications and standards due to: their superior decoding performance [SFRU01, RSU01b]; power efficiency [MS02, PCZ+11]; high-throughput capabilities [KC04, ZHW11]; their advantages related to hardware implementations, such as low-cost [BALE+07, DHB06]. Consequently, these codes are gradually replacing other well-established FEC schemes [RCS+11]. One can mention Satellite-based digital video broadcasting 2nd generation (DVB-S2/T2/C2), ITU-T standard for networking over power lines, phone lines and coaxial cable (G.hn/ G.9960), in IEEE wireless local area network standards like Wi-Fi Standard (802.11n-2009), WiMAX (802.16e), etc. Moreover, LDPC codes have been proposed as a potential candidate for 5G cellular system [Mau16].

An LDPC code is a linear block code defined by a (M, N) *sparse* parity-check matrix denoted by $\mathbf{H}$. The M rows and N columns corresponds to the parity check equations and the coded bits, respectively. A vector $\mathbf{v} = (v_1, \ldots, v_N) \in \{0, 1\}^N$ is a codeword iff:

$$\mathbf{H}\mathbf{v}^T = \mathbf{0} \mod 2. \tag{2.42}$$

By assuming a full rank matrix **H** and systematic code, any codeword of length N is comprised of $K = N - M$ information bits and M parity bits. The coding rate of the code is given by $R = K/N$ which characterizes the amount of redundancy. Figure 2.18 depicts a miniature example of **H** with an illustration of its Tanner graph representation also known as a bipartite graph. It consists of two sets of nodes: variable-nodes represent **H** columns (coded bits); check-nodes represent rows of **H** (parity-check equations). The edges connecting both sets correspond to the 1's entries of **H**.

2.4.2 Construction of LDPC codes

Normally, two ways can be used to construct a LDPC code: algebraic construction and random construction [Joh08]. The latter gives a good performance when large code lengths LDPC codes are considered since they avoid short cycles in the Tanner graph. However such cycles are highly probable when constructing short length LDPC codes and for an irregular LDPC ensemble, the minimum distance becomes a critical issue as it is related to the level of error floor phenomenon. On the other hand, a famous construction method called the Progressive Edge-Growth (PEG) algorithm was proposed in [XEA05] which can effectively construct LDPC codes with medium or short length. PEG algorithm allows constructing the tanner graph with a large girth by progressively establishing edges between variable and check nodes. Two main advantages result by using the PEG algorithm: first, it is very flexible in the sense that it can be used to construct codes for any code length and code rate, furthermore, linear time encoding is feasible under certain modifications. Second, the simplicity to construct LDPC codes with a good girth property.

2.4.3 Decoding

The majority of LDPC codes cannot be decoded algebraically due to their length and the limited structure of their parity matrix. Thus the decoding algorithms are based, like turbo codes, on an iterative process. These algorithms are not optimal for MAP decoding but allow to approach the Shannon boundary on simple channels. The non-optimality is explained by the fact that the graph representing LDPC codes has cycles. Authors in [ETV99] show that cycle-free Tanner graphs cannot support good codes, moreover, the 4-cycle are inevitable for many classical binary linear block codes as shown in [HGC06]. The minimum distance of LDPC codes that have a representation without cycles is low regardless of the performance. The use of iterative decoding in Turbo and LDPC codes is an important element to achieve a good performance. For LDPC codes, the iterative message-passing algorithm has shown a close performance compared to the maximum likelihood method with very low complexity.

The advantage of LDPC codes over other codes is that they have relevant decoding whose complexity only varies linearly with the length of the code. This important aspect is mainly due to the low density of the parity matrix. To decode LDPC codes, a class of algorithms based on message passing along the edges of the tanner graph which is so-called message passing algorithms. Two types can

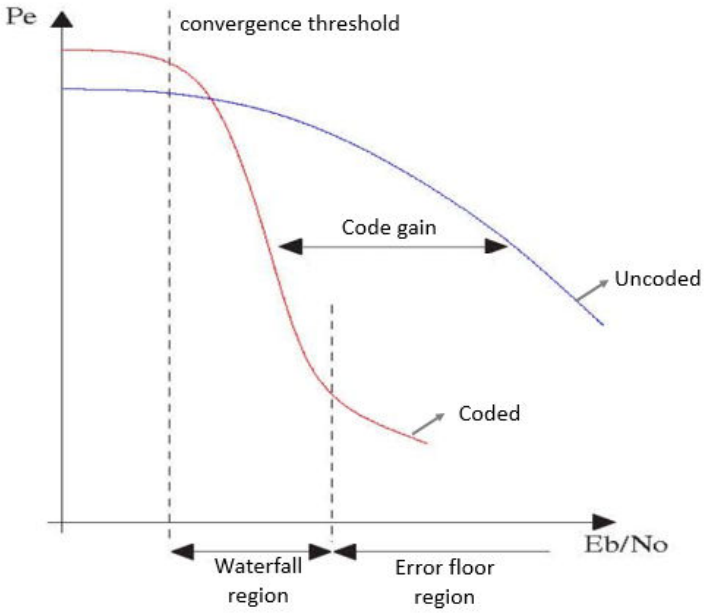


Fig. 2.19. Illustration of three regions characterizing the performance of an LDPC code.

be identified by observing the messages passing through the graph: hard decision and soft decision. The former corresponds to binary messages such as bit flipping algorithms [Gal63] (Gallager A and B) and the latter corresponds to probabilities or log-likelihood ratios (LLRs) such as sum-product algorithm (SPA).

The Belief Propagation (BP) and Min-Sum algorithms are the most important decoding algorithms. One can note that implementing the BP algorithm is complex due to the calculation products, and evaluating hyperbolic trigonometric functions at the check node. Even if these functions are tabulated, the number of evaluations remains important. At high signal-to-noise ratio, few messages are erroneous, in this case, the magnitudes of the messages from check node to variable node are large and their hyperbolic tangent will be close to $+1$ or -1. As a result, during the Check-Pass only the message of *lower* magnitude will be taken into account in the product and that's where the Min-sum algorithm name came from. Hence these remarks make it possible to greatly simplify the complexity of decoding.

Using advanced coding techniques, the evolution of the code performance in function of the signal-to-noise ratio can be divided into three regions as shown in Figure 2.19. The first corresponds to a behaviour where the decoding does not converge (non-convergence region), the decoding then tends to degrade performance compared to an uncoded system. From a certain signal-to-noise ratio, called the convergence threshold or decoding threshold, the decoding enters a phase known as the waterfall region where the probability of error decreases very quickly with the signal to noise ratio. Finally, the error floor region where the probability of error decreases less rapidly than the waterfall region. The coding gain corresponds to the difference between two systems in terms of the required energy per useful bit to achieve a given error rate.

2.4.4 Analyzing tools

One can mention three different techniques to analyze the iterative decoding, in particular, the asymptotic performance of LDPC codes: Density evolution, Gaussian Approximation, and EXIT charts.

- **Density Evolution (DE):** such a method is used to analyze the asymptotic performance of iterative decoders [RU01]. A main objective of using DE for LDPC codes with Message-Passing decoding is to find the SNR decoding threshold (convergence threshold) of a specific LDPC ensemble whose length tends to infinity, by tracking the messages transmitting information on emitted codeword back and forth on the bipartite graph. Extending the DE to irregular LDPC ensembles will allow designing it by enabling optimization of the degree distribution [RSU01b]. Nevertheless, the large amounts of numerical calculations required by the DE will arise with instability and complexity, in particular, when tracking the messages from check to variable nodes. For such a reason, authors in [SFRU01] proposed the quantized density evolution (QDE) method and showed that with only 11-bit quantization one can avoid the instability and reduce the complexity.

- **Gaussian Approximation (GA):** another solution to reduce the complexity and implement the DE was proposed in [SUR00] denoted as Gaussian approximation (GA). This method was proposed for binary-input additive white Gaussian noise (BI-AWGN) channels. The GA method assumes that the message passed through iterative decoders is Gaussian or Gaussian mixtures. The algorithm tracks the parameters of the message distributions instead of tracking whole densities in the decoding process. Thus, a huge reduction in the complexity will be gained as one needs to track only the messages' density mean and variance.

- **EXIT charts:** the extrinsic information transfer (EXIT) chart was proposed as an alternative method compared to DE and GA. It was proposed in first to track the exchange of intrinsic information between component decoders [ten01, Hag04]. Then it was extended to LDPC codes [tKA04] as a tool to design good codes and to find the decoding threshold SNR by reducing it to a curve-fitting problem.

Nevertheless, the DE, GA and EXIT chart assumes that the LDPC codes are infinite codeword length and cycle-free. Moreover, both GA and EXIT charts rely on the assumption that the channel is Gaussian, thus, they cannot be applied directly when a non-Gaussian noise has to be considered. However, DE can be applied to any binary memory-less symmetric output (MBISO) channels including Gaussian and non-Gaussian i.e SαS noise.

LDPC code design is out of the scope of this thesis. Because we focus on designing an estimation of the LLR without relying on the interference plus noise knowledge such that it can adapt for Gaussian or impulsive noise. Thus, one can use any type of LLR based decoders (i.e LDPC, turbo codes, ..., etc) over any Gaussian or non-Gaussian channel type and still observe an improvement compared to the linear

receiver designed only for Gaussian noise. Furthermore, having such a solution is of great advantage and it is considered inevitable when it comes to designing codes for a specific channel because it solves the problem of mismatch decoding that has a significant influence in the receiver performance.

2.5 Elements of Information theory

In this section, we outline the machinery to measure the information that passes through the channel. We introduce some notions and definitions that will be useful in our study.

Definition 5. ***(Discrete channel.)*** *A discrete channel is a statistical model with an input X and an output Y. During each signaling interval (symbol period), the channel accepts an input symbol from X, and in response, it generates an output symbol from Y, generally a noisy version of X. The channel is discrete when the alphabets of X and Y are both finite.*

Definition 6. ***(Memoryless channel.)*** *A channel, characterized by its transition probability $p(y|x)$, is said to be memoryless[3] if when we transmit a sequence $x = \{x_1, x_2, \ldots, x_N\}$ and observe a sequence of output symbols $y = \{y_1, y_2, \ldots, y_N\}$,*

$$p(y|x) = \prod_{n=1}^{N} p(y_n|x_n), \tag{2.43}$$

in other words, the current output symbol depends only on the current input symbol and not on any of the previous input symbols which means that there is no Inter-Symbol Interference.

Definition 7. ***(Symmetric channel.)*** *Assume a random variable Y that takes on a possible value in $\mathbb{R}$. We say that a binary memoryless is symmetric[4] (more precisely, output-symmetric) if*

$$p(y|x = +1) = p(-y|x = -1), \tag{2.44}$$

or as defined by Gallager (1968), if the set of outputs can be partitioned into subsets in such a way that for each subset the matrix of transition probabilities has the property that each row (if more than 1) is a permutation of each other row and each column is a permutation of each other column.

Definition 8. ***(Ensemble X).*** *An ensemble X is a triple $(x, \mathcal{A}_X, \mathcal{P}_X)$, where the outcome x is the random variable value, which takes on one of a set possible values, $\mathcal{A}_X = \{a_1, a_2, \ldots, a_I\}$, having probabilities $\mathcal{P}_X = \{p_1, p_2, \ldots, p_I\}$, with $P(x = a_i) = p_i, p_i \geq 0$ and $\sum_{a_i \in \mathcal{A}_X} P(x = a_i) = 1$.*

[3] The above definition is adequate if we restrict ourselves to channels without feedback.

[4] It is a consequence of the definition of the symmetric channel that $H(Y|X)$ is independent of the input distribution $p(x)$, and depends only on the channel probabilities $p(y_j|x_i)$.

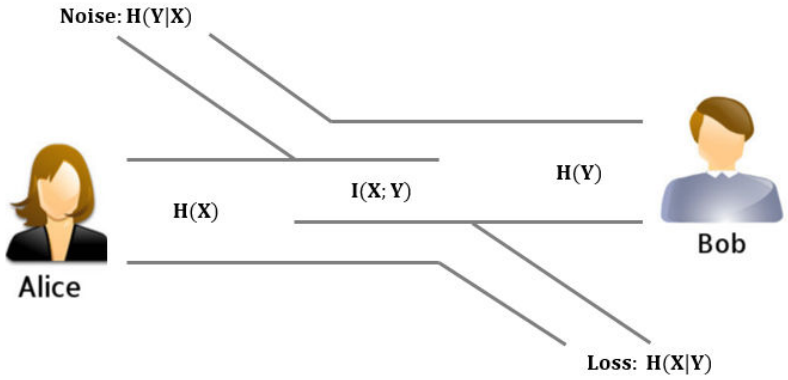


Fig. 2.20. Information distribution over a simple communication scheme.

Definition 9. ***(Joint Ensemble** XY**).*** *A joint ensemble XY is an ensemble in which each outcome is an ordered pair x, y with $x \in \mathcal{A}_X = \{a_1, a_2, \ldots, a_I\}$ and $y \in \mathcal{A}_Y = \{b_1, b_2, \ldots, b_J\}$.*

We denote $p(x, y)$ as the joint probability of x and y which can be defined as $p(x, y) = p(y|x)p(x)$, and represents the joint probability of transmitting x_i and receiving y_j; $p(y|x)$ is the transition probabilities that describe a channel.

The question that we raise now is how to measure the information content of a random variable?

One can classify the measures into three different cases:

- discrete which contains countable values,
- continuous, when $p(X = x) = 0$,
- a mixed version of both discrete and continuous.

In the following, we start by considering the information content of discrete probability distributions over finite sets $\mathcal{A}_X, \mathcal{A}_Y$.

Discrete Information

In the context of communicating over a noisy channel, one has to consider a joint ensemble in which the random variables are dependent. It would be impossible to communicate over the channel if they were independent as will be shown shortly. Figure 2.20 depicts a simple point to point (P2P) communication scheme where: Alice wants to transmit a message x over a noisy channel; the output of the channel y is received by Bob; one can define the following

Definition 10. ***(Entropy**[5]**).*** *The average Shannon information content of an outcome is defined as the entropy of the ensemble X*

$$H(X) = \sum_{x \in \mathcal{A}_X} p(x) \log_2 \frac{1}{p(x)}. \tag{2.45}$$

[5] Also referred to as marginal entropy, it is used to contrast it with the conditional entropy.

It's a sensible measure of the ensemble's average information content, in other words, it is a measure of Bob's uncertainty about the channel input X before observing the channel output Y and measured in bits.

Definition 11. ***(Joint entropy).*** *The joint entropy of X, Y is:*

$$H(X,Y) = \sum_{xy \in \mathcal{A}_X \mathcal{A}_Y} p(x,y) \log_2 \frac{1}{p(x,y)}. \tag{2.46}$$

where entropy is additive for independent random variables $H(X,Y) = H(X) + H(Y)$ if and only if $p(x,y) = p(x)p(y)$.

Definition 12. ***(Conditional entropy.)*** *The conditional entropy of X given Y, is the average, over y, of the conditional of X given y,*

$$H(X|Y) = \sum_{y \in \mathcal{A}_Y} p(y) \left(\sum_{x \in \mathcal{A}_X} p(x|y) \log_2 \frac{1}{p(x|y)} \right), \tag{2.47}$$

this measures the average residual uncertainty about x when y is known; representing the information that goes into the channel by Alice but does not received by Bob (information loss).

The conditional entropy of Y given X is the average, over x, of the conditional of Y given x,

$$H(Y|X) = \sum_{x \in \mathcal{A}_X} p(x) \left(\sum_{y \in \mathcal{A}_Y} p(y|x) \log_2 \frac{1}{p(y|x)} \right), \tag{2.48}$$

this measures the average residual uncertainty about y when x is known; representing the information that came out of the channel (by a noise source) but was not sent by Alice.

Definition 13. ***(Mutual information).*** *The mutual information between X and Y is*

$$I(X;Y) = H(X) - H(X|Y), \tag{2.49}$$

Mutual information is symmetric and non-negative, i.e., $I(X;Y) = I(Y;X) \geq 0$. It measures the average reduction in uncertainty about x that is resolved by observing the value of y; or vice versa, the average amount of information that x conveys about y. Thus, the mutual information is a good measure of the amount of information that goes through the channel between Alice and Bob.

In Figure 2.21 illustrates the following useful relationship between joint information, marginal entropy, conditional entropy and mutual entropy

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y). \tag{2.50}$$

After presenting the discrete definitions of entropy with discrete measures, the extension continuous information measurement will be tackled in the following.

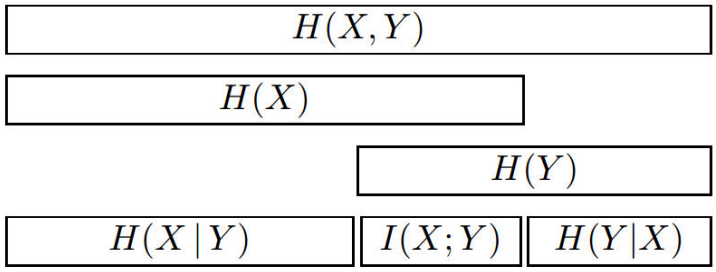


Fig. 2.21. The relationship between joint information, marginal entropy, conditional entropy and mutual entropy.

Continuous Information

One may ask, what if the signals or messages we wish to transfer are a continuous variable. Indeed, many practical channels have real, rather than discrete, inputs and outputs;

From the discrete case we can obtain the results for the continuous case as a limiting process. Consider a random variable X with an outcome $x_k = k\delta x$ for $\{k = 0, \pm 1, \pm 2, \ldots\}$ having probability $p(x_k)\delta x$, where $p(x_k)$ represents the probability density function that satisfy $\sum_k p(x_k)\delta x$=1.

By taking the limit of the discrete entropy formula of $H(X)$ given in (2.45) we obtain

$$\begin{aligned} H(X) &= \lim_{\delta x \to 0} \sum_k p(x_k)\delta x \log_2 \left(\frac{1}{p(x_k)\delta x} \right) \\ &= \lim_{\delta x \to 0} \left[\sum_k p(x_k) \log_2 \left(\frac{1}{p(x_k)} \right) \delta x - \log_2(\delta x) \sum_k p(x_k)\delta x \right] \\ &= \int_{-\infty}^{\infty} p(x_k) \log_2 \left(\frac{1}{p(x)} \right) dx - \left(\lim_{\delta x \to 0} \log_2(\delta x) \right) \times \int_{-\infty}^{\infty} p(x) dx \\ &= h(X) - \lim_{\delta x \to 0} \log_2(\delta x) \end{aligned} \tag{2.51}$$

This is rather worrying as the second term does not exist. However, as we are often interested in the differences between entropies in the consideration of mutual entropy or capacity we define the problem by using the first term only as a measure of differential entropy[6].

Hence, we can now define also the joint and conditional entropies for continuous

[6]Differential entropy is not a measure of the average amount of information contained in a continuous random variable (in general, it contains an infinite amount of information).

distributions:

$$h(X,Y) = \int_Y \int_X p(x,y) \log_2 \left(\frac{1}{p(x,y)} \right) dxdy \tag{2.52}$$

$$h(X|Y) = \int_Y \int_X p(x,y) \log_2 \left(\frac{p(y)}{p(x,y)} \right) dxdy \tag{2.53}$$

$$h(Y|X) = \int_Y \int_X p(x,y) \log_2 \left(\frac{p(x)}{p(x,y)} \right) dxdy \tag{2.54}$$

with:

$$p(x) = \int_Y p(x,y)dy \tag{2.55}$$

$$p(y) = \int_X p(x,y)dx \tag{2.56}$$

Eventually, the mutual information between two continuous random variable X and Y is defined as

$$I(X,Y) = \int_Y \int_X p(x,y) \log_2 \left(\frac{p(x|y)}{p(x)} \right) dxdy \tag{2.57}$$

Channel capacity

The mutual information relies on both the channel characteristics (represented by the conditional probability distribution) and the input probability distribution, which is clearly independent of the channel[7]. Therefore, by changing the input probability distribution, the average mutual information will change. Let us assume that we want to maximize the mutual information that goes through the channel by choosing the best possible input ensemble, one can note that it is possible to find multiple optimal input distributions achieving the same value of $I(X;Y)$. We can define the channel capacity to be its maximum mutual information

$$C = \sup_{\mathcal{P}_X} I(X;Y). \tag{2.58}$$

Under the input average power constraint the r.v. X satisfies $\mathbb{E}[X_i^2] \leq P$. The capacity is determined by the intrinsic properties of the channel and is independent of the content of the transmitted information and the way it is encoded.

For any binary memory less symmetric channels one can compute the capacity as a function of LLR's [RU08].

Lemma 2. ***(Capacity MBISO channels.)*** *Let f be the density associated with a MBISO channel. Then the capacity of this channel in bits per channel use is*

$$C = 1 - \mathbb{E}\left[\log_2\left(1 + e^{-X\Lambda(Y)}\right)\right] \tag{2.59}$$

[7] let $p(y|x)$ be the conditional probability distribution of y given x, which is an inherent fixed property of the communication channel. Then the choice of the marginal distribution $p(x)$ completely determines the joint distribution $p(x,y)$ due to the identity $p(x,y) = p(y|x)p(x)$ which, in turn induces a mutual information.

Proof.

$$\begin{aligned} C &= H(Y) - H(Y|X) \\ &= \int_Y P(y) \log_2 \left(\frac{1}{P(y)} \right) dy - \sum_{x=\pm 1} P(x) \left(\int_Y P(y|x) \log_2 \left(\frac{1}{P(y|x)} \right) dy \right) \\ &= \int_Y \frac{1}{2} \sum_{x=\pm 1} P(y|x) \log_2 \left(\frac{2P(y|x)}{P(y|1) + P(y|-1)} \right) dy \\ &= \int_Y P(y|1) \log_2 \left(\frac{2P(y|1)}{P(y|1) + P(y|-1)} \right) dy \end{aligned} \tag{2.60}$$

where for the transition to the fourth line we have used Definition 7 (page 38) for a symmetric channel $P(y|1) = P(-y|-1)$.

Let $f(y)$ be the probability density at the channel output y when $x = 1$ is transmitted. Then, $P(y|1) = f(y)$ and $P(y|-1) = f(-y)$. The logarithm of the likelihood ratio LLR[8] $\Lambda(y)$ then takes the following particular expression:

$$\Lambda(y) = \log \frac{P(y|+1)}{P(y|-1)} = \log \frac{f(y)}{f(-y)} \tag{2.61}$$

from (2.61) we can obtain $f(-y) = f(y)e^{-\Lambda(y)}$, replacing it in (2.60) gives

$$\begin{aligned} C &= \int_Y f(y) \log_2 \left(\frac{2f(y)}{f(y) + f(-y)} \right) dy \\ &= \int_Y f(y) \log_2 \left(\frac{2}{1 + e^{-\Lambda(y)}} \right) dy \\ &= \int_Y f(y) dy - \int_Y f(y) \log_2 \left(1 + e^{-\Lambda(y)} \right) dy \\ &= 1 - \mathbb{E}_Y \left[\log_2 \left(1 + e^{-\Lambda(Y)} \right) \right]. \end{aligned} \tag{2.62}$$

By considering the two possible input values $+1$ or -1, (2.62) becomes

$$C = 1 - \mathbb{E}_Y \left[\log_2 \left(1 + e^{-X\Lambda(Y)} \right) \right]. \tag{2.63}$$

■

2.6 Capacity for MBISO channels

The channel capacity is an extremely important quantity since it is possible to transmit information through a channel at any rate less than the channel capacity with an arbitrarily small probability of error; completely reliable transmission is not

[8] The LLR is a prime tool in information theory as it constitutes a sufficient statistic relative to the channel input [RU08]; in other words, knowing $\Lambda(Y)$ or Y is equivalent for the decoding process.

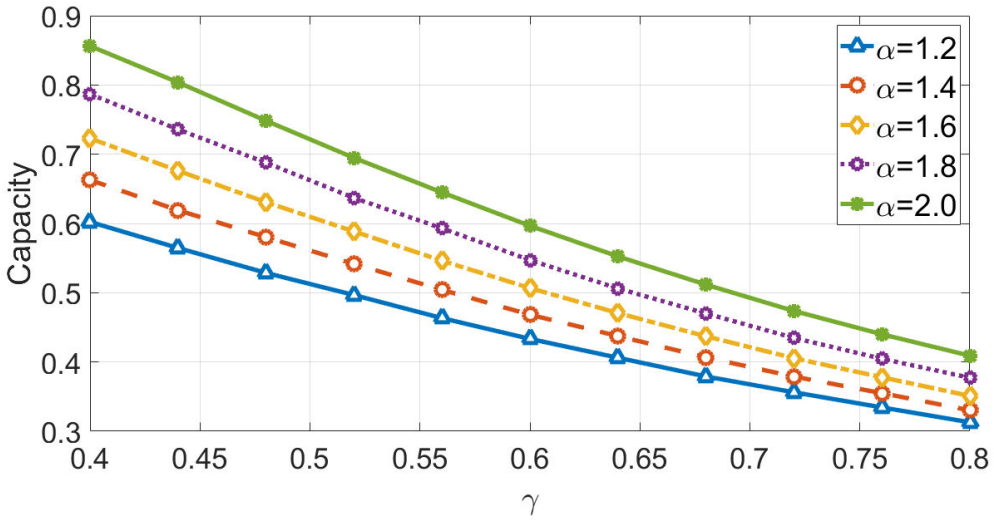


Fig. 2.22. Numerical computation of the α-stable noise capacity with different values of α as a function of the dispersion γ.

possible if the information processed is greater than the channel capacity. However, in general, the calculation of the channel capacity is a difficult problem.

For the class of discrete memoryless channels, the capacity is now well understood [Sha48]. However, generalizing to continuous channels has proven to be non-trivial, with the important exception of the linear additive white Gaussian noise (AWGN) channel subject to a power constraint [Sha48, Theorem 18]. Due to the difficulty in deriving closed-form expressions for the capacity and the optimal input distribution of continuous channels, the focus has shifted to determining structural properties of the optimal input distribution, as well as bounds and numerical methods to compute the capacity. Hence, we will adopt such an approach to compute the capacity numerically in this manuscript for different noise models which will be used as a benchmark in order to assess the impact of dynamic interference on long block length codes.

We rely on Monte Carlo calculations to estimate the capacity by replacing the expectation of (2.63) by an empirical average, hence, the capacity can be obtained as

$$C = 1 - \lim_{K\to\infty} \left(\frac{1}{K} \sum_{k=1}^{K} \log_2 \left(1 + e^{-x_k \Lambda(y_k)}\right) \right) \tag{2.64}$$

Figure 2.22, Figure 2.23, Figure 2.24 will be used to assess the impact of dynamic interference on long block length codes in α-stable, Middleton class A and ϵ-contaminated noises, respectively.

2.7 Conclusion

To have a better understanding of the interference or impulsive noise that appears in many modern communication systems, different models were depicted. We started

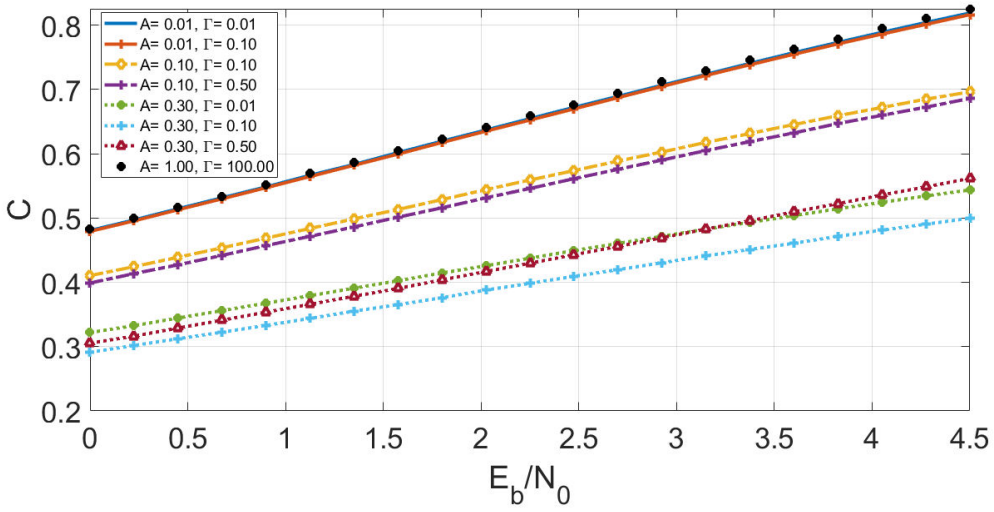


Fig. 2.23. Numerical computation of the Middleton class A noise with different values of A and Γ as a function of E_b/N_0.

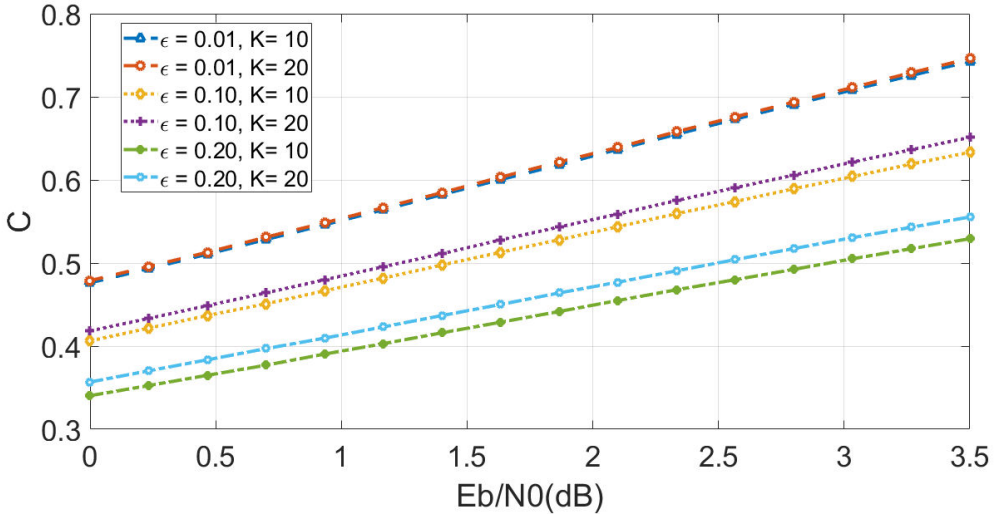


Fig. 2.24. Numerical computation of the ϵ-contaminated noise with different values of ϵ and K as a function of E_b/N_0.

by the Gaussian distribution, however, this model is unable to represent the large variability in the data that arises by an impulsive environment due to the exponential decay tail behavior. Next, we introduce non-Gaussian models and classify them into three different categories: theoretical approaches; mixture model approaches; empirical approaches. In the first step, we present the different Middleton classes and show their physical perspectives and the difficulty to use in practical systems. Then, we present several mixture models, mainly Gaussian mixtures, which can be seen as a reduced version of Middleton classes in order to reduce its complexity. However, they may not truly portray the noise characteristic at the tails and lack the stability property. Thereafter we focus on the α-stable family as it is suggested as an accurate solution in the class of heavy-tailed distributions to model impulsive noise. Moreover, we presented the motivation behind using the α- family to model the dynamic impulsive interference phenomena. Then, we introduced LDPC codes and we present some elements of information theory that are needed in the rest of this thesis. Finally, we computed numerically the capacity of different noise models and different channels, in order to be used as a benchmark in the long blocklength regime.

Chapter 3

Robust Receiver Design

Establishing reliable and efficient communications require to take into account the impulsive nature while designing the receivers. The interference modeling question was answered in the aforementioned part; it exhibits in many situations an impulsive behavior. Without being exhaustive we presented several approaches adapted to specific situations. However, designing a specific receiver for each situation is not efficient as the interference characteristics can highly vary in time and space. Thus, having a receiver able to cope with a large set of different interference models (impulsive or not) and with different degrees of impulsiveness is highly desired. This is the main challenge we address in this manuscript. For this purpose, we start by clarifying the impact of impulsive interference on the optimal decision regions. Then, we classify the different receiver design approaches seen in the literature. Thereafter, we introduce the system scenario and propose a new framework inspired by information theory for the sake of designing a robust receiver.

3.1 Impact of impulsive interference on optimal decision regions

An efficient way to characterize and understand the influence of impulsive noise is to visualize the noise impact on the optimal receiver by representing the decision regions. This was proposed by Saaifan and Henkel [SH13] for the Middleton class A case and by Shehat *et al.* [SME10] and by Saleh *et al.* [SME12] for the α-stable case.

We represent in Figure 3.1 six different examples of noise realizations from impulsive noise models. We follow the framework proposed in [SME12] and use the same parameters defined in Figure 3.1: the received vector Y is composed of two received samples (two dimensions, $Y = [y_1, y_2]$), and we consider two possible transmitted values $x \in \Omega = \{-1, 1\}$. Then we show in Figure 3.2 the decision regions r that the optimal receiver must produce in a binary case under different interference models, i.e., the regions that maximize the probability of having transmitted x when $Y = (y_1, y_2) = (x + n_1, x + n_2)$ is received with two i.i.d. impulsive noise realizations

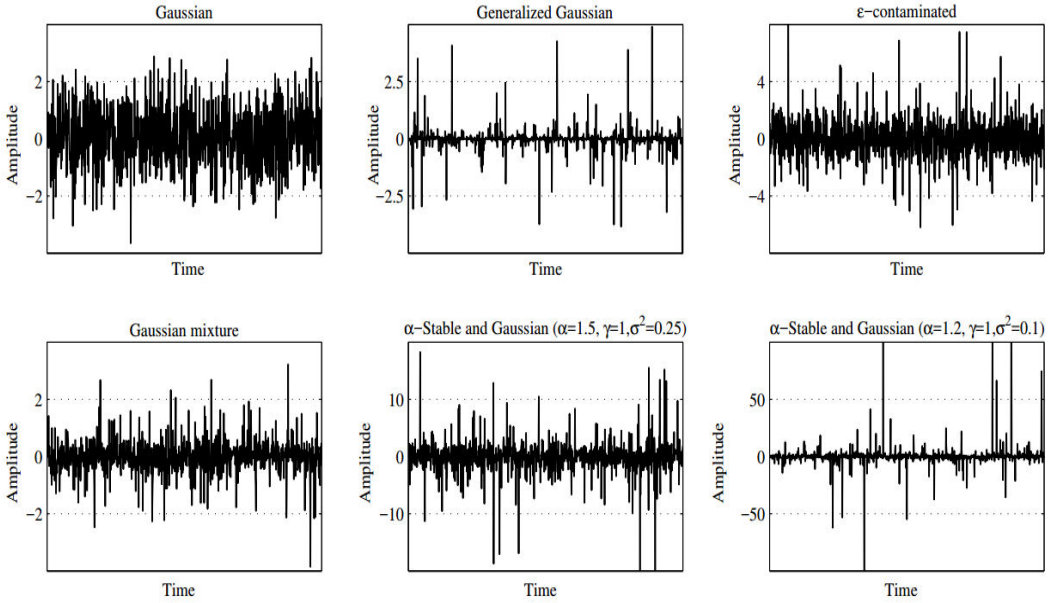


Fig. 3.1. Example realizations for each different sub-exponential impulsive noise processes. The following parameters were used in each case: Gaussian case ($\mu = 0$ and $\sigma^2 = 1$); generalized Gaussian case ($\alpha = 1$ and $\beta = 0.2$); ϵ-contaminated case ($\epsilon = 0.1, k = 10, \sigma^2 = 1$); Gaussian mixture case ($P = 3; \lambda_1 = 0.1, \mu_1 = -0.1, \sigma_1^2 = 1; \lambda_2 = 0.8, \mu_2 = 0, \sigma_2^2 = 0.1; \lambda_3 = 0.1, \mu_3 = 0.1, \sigma_3^2 = 1$; sum of Gaussian and α-stable in a moderately impulsive case ($\alpha = 1.5, \gamma = 1$ and $\sigma^2 = 0.25$); sum of Gaussian and α-stable in a highly impulsive case ($\alpha = 1.2, \gamma = 1$ and $\sigma^2 = 0.1$)

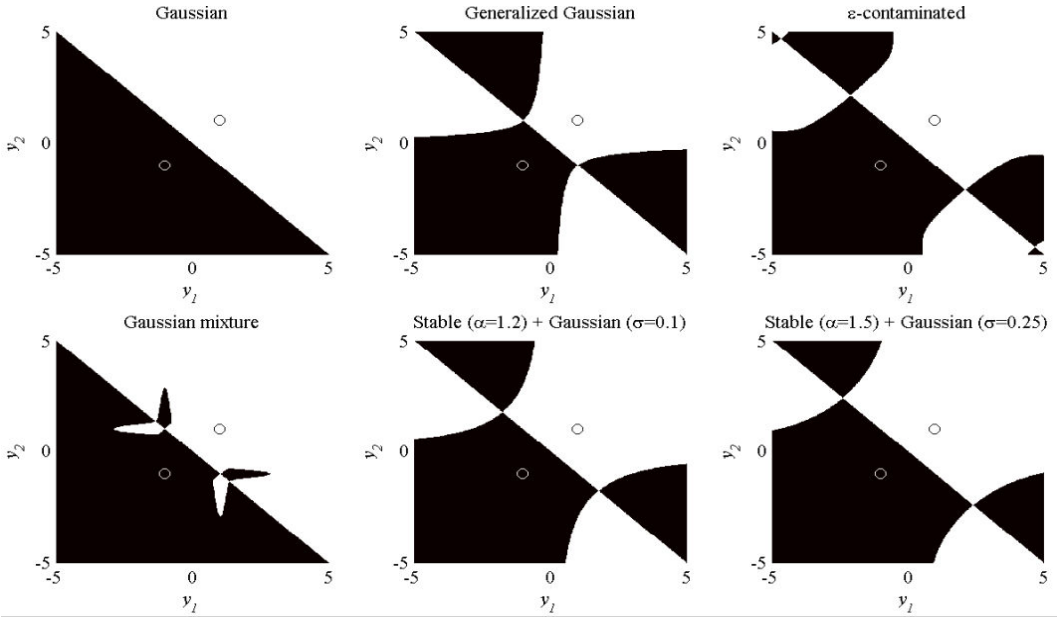


Fig. 3.2. Optimal decision regions for the different noise processes.

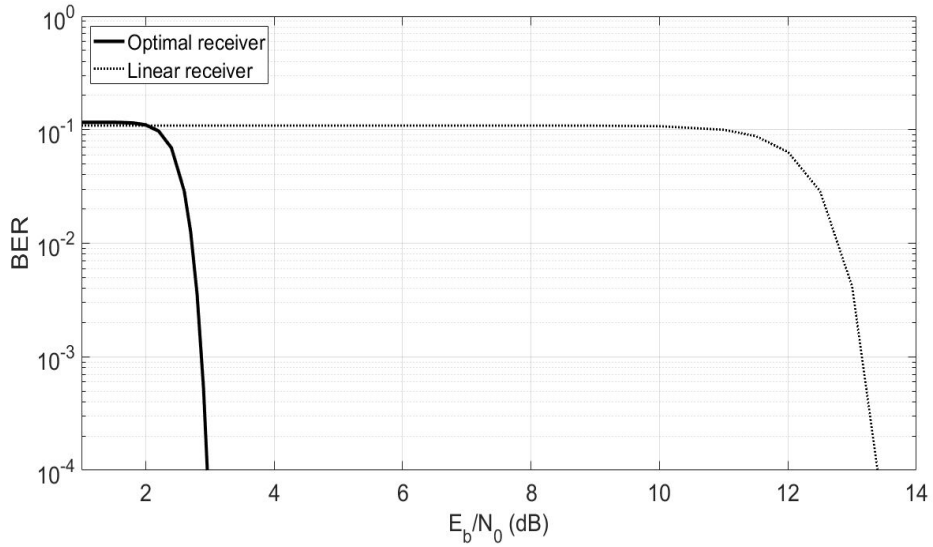


Fig. 3.3. Effect of mismatch decoding, where the linear receiver is used in an impulsive environment (Middleton Class A).

$n_1, n_2,$

$$r = \underset{x \in \Omega}{\operatorname{argmax}}\, p(y_1, y_2|x). \tag{3.1}$$

The black and white regions shown in Figure 3.2 represent the -1 and $+1$ symbol decision, respectively. The Generalized Gaussian distribution, when the shape parameter is less than 1, is sub-exponential in nature. The Mixture of Gaussians, including the ϵ-contaminated, are not strictly sub-exponential. However, as noted in [KFR98] one can approximate for instance an α-stable sub-exponential interference model to an arbitrary accuracy over any tail probabilities with enough Gaussian mixture components.

Notably, under interference with exponential tail decay, i.e. the Gaussian case shown is in Figure 3.2, the optimal decision regions are linearly separated. However, the optimal decision regions under heavy-tailed sub-exponential interference produce non-linear frontiers and disjoint regions. In each impulsive case, two operating regions can be identified: for small received values y_1, y_2, the boundaries are linear. However, when at least one of the values becomes larger, the linear boundaries completely fail to recover the most likely transmitted symbol. It is worth mentioning that the point at which this non-linearity appears in the decision regions depicts the beginning of the heavy-tailed distribution dominance over the Gaussian noise.

3.2 Different approaches for receiver design

Regarding the receiver design in the impulsive case, several observations can be noted. Firstly, the significant performance degradation obtained by the linear receiver (optimal for Gaussian noise and simple to implement) is due to the model mismatch as shown in Figure 3.3. Second, the construction of an optimal receiver, which assumes to have knowledge of the channel is complicated for several reasons:

1- the various interference models proposed in the literature will raise several questions when it comes to design a receiver for instance, which model must be selected and is it robust against environment changes? 2- when theoretical impulsive models are considered, it is difficult to implement the receiver. Moreover, if empirical models are chosen to offer analytical solutions, their ability to adapt to different contexts is to be proven. In the following, we do not try to be exhaustive about the existing receiver strategies but we propose to classify the different receiver design approaches into four categories.

3.2.1 Optimal approach

To implement the soft iterative codes we first encode the channel outputs into log-likelihood ratios. The channels we are interested in are memoryless, binary input, and symmetric output. It means that the symbols of the input X are $+1$ and -1. Due to the symmetry of the output Y, we have $p(y|x = +1) = p(-y|x = -1)$. Under MBISO channel, the capacity achieving distribution of the input is uniform (that is $p(+1) = p(-1) = \frac{1}{2}$), because of the symmetry of the output. In this case, the LLR is given equivalently by

$$\Lambda(y) = \log \frac{p(y|x = +1)}{p(y|x = -1)} = \log \frac{p(x = +1|y)}{p(x = -1|y)}. \tag{3.2}$$

Notably because of uniform prior, 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 [RU08]; 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 BP algorithm.

This is very attractive when noise is Gaussian because it leads to a linear receiver, straightforward to implement. However, with impulsive noise, the LLR becomes a non-linear function. Its implementation is complex and highly depends on the noise distribution either because of the lack of a closed-form expression such as for α-stable noise, or because it needs high computational burden such as for Middleton noise. Consequently, the extraction of a simple metric based on the noise PDF in the decoding algorithm is prevented. It is worth mentioning that under the α-stable assumption the LLR can still be computed numerically, for instance, by numerical integration of the inverse Fourier transform of the characteristic function as shown in (2.25). However, the integral in (2.25), induces a prohibitive computation and the evaluation of the LLR requires the knowledge of the noise parameters. Hence, a sub-optimal receiver is then salutary to reduce complexity.

3.2.2 Noise distribution approximation

The main idea behind this approach is to find a distribution that well approximates the true noise plus interference PDF, with an analytical expression and parameters that can be estimated in a simple manner. For UWB communications and in this context one can find a review in [BY09]. For instance, Erseghe *et al.* used a

Gaussian mixture [ECD08]; a receiver adapted to Multi-User interference based on a generalized Gaussian distribution approximation was proposed by Fiorina in [Fio06]. Nammi *et al.* [NBST06] studied the impact of impulsive noise on Parity Check Codes by using ϵ-contaminated distribution. Beaulieu and Niranjayan [BN10] considered a mixture of Laplacian and Gaussian noise. El-Ghannudi *et al.* [GCA+10] proposed the Cauchy receiver based on Cauchy distributions, which are a special case of SαS distributions with $\alpha = 1$. An improved version of the Cauchy receiver called Myriad receiver was proposed in [NB08], which is based on the Cauchy distribution but with a modified dispersion parameter to improve the adaptability to different impulsive degrees. The Normal Inverse Gaussian (NIG) distribution was proposed in [GPC+12] to study cooperative communications, it is a flexible family of distributions that includes Gaussian and Cauchy distributions.

A significant performance improvement has been shown for each solution in their specific context compared to the linear approach. However, in the case of a model mismatch, we can wonder how robust they will be. Moreover, the estimation of the parameters can become in these situations significantly important. For instance, the Cauchy receiver is suitable solely for highly impulsive noise, but the performance will be limited when the impulsiveness decreases ranging from moderately impulsive to purely Gaussian noise. Furthermore, the Cauchy receiver is highly influenced by the model mismatch i.e. Middleton Class A. The Myriad and NIG receiver will show better immunity against the model mismatch but they show poor performance under less impulsive and purely Gaussian noises. Furthermore, the computation of the LLR is complicated when the NIG distribution is considered as it requires more complex function, including Bessel function.

3.2.3 Different metric measures

An alternative way to interpret detection is to consider that the likelihood measures the distance between all the received signals and the possible transmitted signals. For the Gaussian case, the optimal distance is the Euclidean one which is not suitable for the impulsive case. For this purpose and to be less specific on a noise model, several papers proposed different metrics, e.g. a robust metric mixing euclidean distance and erasure [FC09], Hubber metric [Chu05], p-norm [GC12], ..., etc. The p-norm is a distance measurement in the α-stable case with $p < \alpha$, see [GC12], as the α-norm can be estimated via:

$$\|X - Y\|_\alpha = \begin{cases} \left[\mathbb{E}|X - Y|^p / C(\alpha, p)\right]^{1/p} & , 1 \leq \alpha \leq 2 \\ \left[\mathbb{E}|X - Y|^p / C(\alpha, p)\right]^{\alpha/p} & , 0 \leq \alpha < 1 \end{cases} \tag{3.3}$$

where

$$C(\alpha, p) = \frac{2^{p+1}\Gamma((p+1)/2)\Gamma(-p/\alpha)}{\alpha\sqrt{\pi}\,\Gamma(-p/2)} \tag{3.4}$$

and $\Gamma(\cdot)$ is the Gamma function given in (2.31).

In [SD99], the L_p-norm was proposed instead of the standard Least Mean square as an interference suppression scheme for DS/CDMA in the presence of SαS interfer-

ence. The decision statistic for the p-norm metric is given without the normalization constant by

$$L(y) = |y-1|^p - |y+1|^p. \tag{3.5}$$

Attractively, only a rough knowledge of α is needed and there is no need for any distribution parameter estimation. However, the estimation of the p value will play a significant role in terms of performance.

3.2.4 Direct LLR approximation

Linear approximation: evaluating the LLR with a Gaussian noise assumption results in a linear operation for detection and is denoted as the linear receiver. We primarily consider this choice for its simple implementation structure and also as a reference to show the gain achieved by using other approaches, though it is known to perform poorly in impulsive situations. The LLR for the Gaussian noise is expressed by:

$$\begin{aligned}\Lambda_G(y) &= \log \frac{f_N(y-1)}{f_N(y+1)}, \\ &= \log \frac{e^{-(y-1)^2/2\sigma^2}}{e^{-(y+1)^2/2\sigma^2}} = \frac{2y}{\sigma^2}.\end{aligned} \tag{3.6}$$

where $f_N(\cdot)$ denotes the PDF (2.3) of the noise N and σ is the standard deviation.

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 (meaning a less reliable sample) so that the LLR should be small.

Non-linear LLR approximation: under impulsive noise, the LLR is computationally prohibitive as discussed in Section 3.2.1, consequently, we consider parametric approximation L_θ of the LLR $\Lambda(y)$. The family of functions L_θ is chosen for its simplicity and for its flexibility to represent the LLR of different channel types. Besides, if we consider a family defined by a limited number of parameters and easy to be implemented, both the estimation and implementation complexities are reduced. Hence, to narrow down the search space and to have an easy to implement approximation, we assume that the estimated LLR L_θ is an odd piece-wise function of the optimization parameter θ. More precisely, we are interested in functions that can be represented as

$$L_\theta(y) = \operatorname{sign}(y) \min\left\{\theta_1\phi_1(|y|), \theta_2\phi_2(|y|), \ldots, \theta_n\phi_n(|y|)\right\}, \tag{3.7}$$

where $\operatorname{sign}(x)$ is defined as $\operatorname{sign}(x) = x/|x|$; $\phi_i(|y|)$ are functions depending solely on the channel output $|y|$ but not necessarily linear in $|y|$; the parameter θ, given as $\theta = [\theta_1, \theta_2, \ldots, \theta_n] \in \mathbb{R}_+^n$, is the the optimization parameter that needs to be estimated. This formulation includes the linear case. The family of function is then

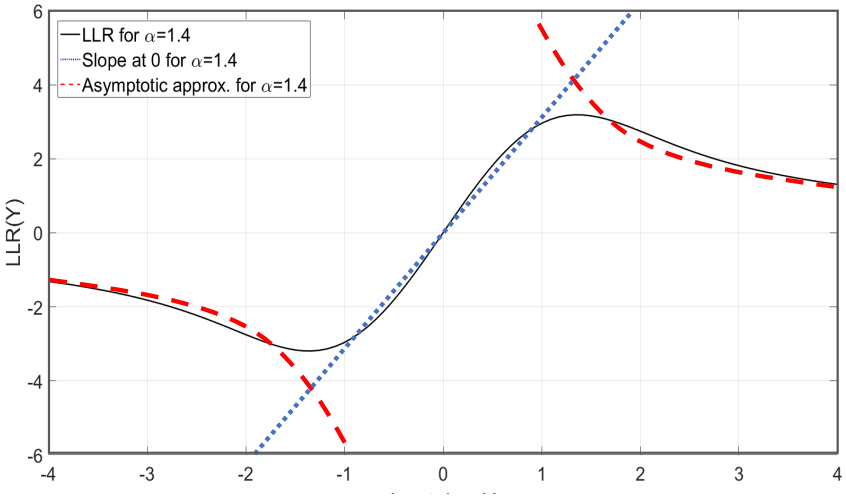


Fig. 3.4. LLR demapper for $\alpha = 1.4$, $\gamma = 0.5$, and its approximation.

simply $L_\theta(y) = ay$, this is called thereafter the L_a receiver. The formulation (3.7) allows us to take into account asymptotical approximations of the LLR.

Figure 3.4 lightens the non-linearity of the LLR function for the channel output Y when the noise is α-stable. 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. When the noise is Gaussian, the linear region spreads with the decrease of the noise impulsiveness until reaching the limit (only the linear part exists).

Remark 2. *Even if Figure 3.4 delineates a specific noise model (α-stable), the overall appearance of the LLR exhibits a similar behavior for other impulsive models.*

The presence of these two parts has been used in the literature to propose several LLR approximations [DGCG14, AIH94, MGCG13, SMET12, MSG+18] and justifies the proposed piece-wise affine set for the LLRs approximation. Firstly, several approximations cope with impulses by limiting the received values such as soft limiter and hole-puncher [AIH94]. In [MGCG13], a variant of the soft limiter was proposed called the clipper, where the LLR equation is:

$$L_{clipp}(y) = \begin{cases} py & \text{if } -h/p < y < h/p \\ h\,\mathrm{sign}(y) & \text{otherwise,} \end{cases} \tag{3.8}$$

where p represents the signal amplitude and the impulse clipping level is represented by h, using the density evolution they obtain both parameters. The clipper demapper also falls into the formulation (3.7). In order to match the parametric notation defined in (3.7), the clipper demapper can be rewritten as:

$$L_{\text{clip}}(y) = \mathrm{sgn}(y)\ \min\big(a|y|, \sqrt{ab}\big) = \begin{cases} ay & \text{if } |y| < \sqrt{b/a}, \\ \sqrt{ab} & \text{otherwise,} \end{cases} \tag{3.9}$$

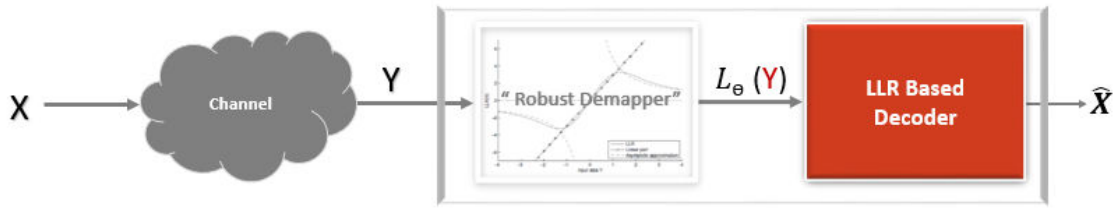


Fig. 3.5. The demapper scheme.

where a represent the slope of the linear part and $\sqrt{ab}$ represent the clipping threshold.

Despite the performance improvement seen in the clipping receiver compared to the linear one, we expect that approximating the optimal LLR might give a near-optimal performance. Indeed, recently the approximated LLR receiver [DGCG14] which will be called thereafter L_{ab} was proposed. The L_{ab} demapper decomposes the LLR into two parts: a linear part and an asymptotic part. The linear part is proportional to the received signal and is related to the noise dispersion γ and the asymptotic part is derived from the asymptotic expansion approximation of the density function of stable variable proposition 2 (page 27) and it reflects the heaviness of the tail. Hence, the L_{ab} demapper is given by

$$L_{ab}(y) = \operatorname{sgn}(y)\ \min\big(a|y|, b/|y|\big) = \begin{cases} ay & \text{if } |y| < \sqrt{b/a}, \\ b/y & \text{otherwise}, \end{cases} \tag{3.10}$$

where $a = \sqrt{2}/\gamma$ and $b = 2(\alpha + 1)$. Remark that, the demapper may be directly parameterized by a and b instead of α and γ.

Recently, we proposed [MSG+19c] a new approximation based on three parameters which will be called thereafter L_{abc}, expecting to have an approximation that fits the true LLR to a high degree without adding much complexity. Consequently, near-optimum performance can be achieved. Hence, the L_{abc} demapper is given by

$$L_{abc}(y) = \operatorname{sgn}(y)\ \min\big(a|y|, b/|y|, c\big) = \begin{cases} ay & \text{if } |y| < c/a, \\ c & \text{if } c/a < |y| > bc, \\ b/y & \text{if } |y| > bc, \end{cases} \tag{3.11}$$

where a represents the slope of the linear part (for small values of y), b represents the degree of decay of the asymptotic part (for large values of y) and c represents a saturation level. The L_{abc} demapper will be further detailed in Section 4.2.3.

This thesis mainly deals with this transformation between channel output and decoder input. This transformation is called thereafter demapping as shown in Figure 3.5. Attractively, constructing an adaptive robust receiver in a way that can be used at the front end of any LLR-based decoder will be of great advantage as it can be migrated to any LLR-based decoder like LDPC, turbo, etc.

The main question arising after having chosen the parametric function L_θ, is how to estimate θ in order to minimize the BER or the Frame Error Rate (FER)? For this purpose, we study in the following different methods.

3.3 Different parameter optimization methods

The approximated LLRs require the knowledge of the parameter θ in order to be tuned and match the channel state. The first idea to estimate these parameters is to compute them directly from noise samples—which may be obtained by a training sequence. We can mention three methods to estimate these parameters:

- direct estimation of the noise model parameters α and γ and then taking $a = \sqrt{2}/\gamma$ and $b = 2(\alpha + 1)$. This is equivalent to considering that the noise distribution follows an α-stable law. Two different methods are used to estimate the parameters of the stable distribution: the first is based on quantiles proposed by McCulloch [McC86b] and the second use regression type estimation method based on the characteristic function proposed by Koutrouvelis in [Kou81]. Apparently, the performance in terms of BER of such methods will be degraded as the environment changes due to the noise model mismatch effect.

- direct noise distribution estimation: in this case, a kernel density estimation is used followed by fitting the estimated LLR ($\widehat{L}(x)$) with the approximation L_{ab}: first the noise PDF will be estimated using some classical kernel-based approaches. Thereafter, by using the density estimation, the LLR is estimated directly by computing (3.2). Unfortunately, $\widehat{L}(x)$ is difficult to use as an online real-time likelihood approximation. This is essentially due to missing samples which results in the appearance of zeros in the density estimation. Consequently, the two parameters a and b will be estimated from $\widehat{L}(x)$ by considering its maximum (x_m): $a = \widehat{L}(x_m)/x_m$ and $b = ax_m^2$. It is worth mentioning that such a method will be degenerated under some extreme situations, for instance, when the noise magnitude is too small or when the noise samples are too large.

- maximization of the mutual information between the source and the output of the approximated LLR: this approach is attractive since it does not assume a particular noise-plus-interference model. It is inspired by the work of Yazdani and Ardakani in [YA09]. In order to find the best slope of the linear receiver, they used the maximization of mutual information between the source and the receiver output. Nevertheless, this approach can be adjusted easily in order to fit our estimation problem, in general, to find the optimal parameter θ^* that will tune the demapper L_{θ^*} and thus match the channel state. The receiver parameter θ^* is now given by

$$\theta^* = \underset{\theta}{\operatorname{argmax}}\, I(X; L_\theta(Y)). \tag{3.12}$$

  This method will be studied in more detail in the rest of the manuscript.

3.4 Proposed framework and system scenario

Source: we use a LDPC code associated with a BP-decoding algorithm. This case is well-suited to our proposal because the LLR has to be estimated and fed to the

BP algorithm. Nevertheless, 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.

In the following, we study the proposed framework with two different block lengths, long and short. We assume that the binary message X is encoded using a regular (3,6) LDPC code of length 20000 and 408, respectively. We did the same study over different LDPC codes, but the conclusions are the same.

Channel: the channels we are interested in are memoryless, binary input, and symmetric output (MBISO). A large number of channels fall into this class such as the Binary Symmetric Channel (BSC), the Binary Erasure Channel (BEC) or the Additive White Gaussian Noise (AWGN) Channel. A wide range of situations can be dealt with rather straightforwardly once we have mastered MBISO channels [RU08]. More generally, channels defined by $Y = X + N$, whose inputs X are perturbed by an additive noise N belong to the MBISO family as long as the additive perturbations are symmetric and independent from the input. 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 [Mid77] or a symmetric α-stable distribution [PW10a], etc. In particular, we will focus on Symmetric α stable (SαS) distributions. Nevertheless, we also want to check the robustness against different noise models. Consequently, we will study the behavior of our approach with other classical noise models: Gaussian, Middleton class A and $\epsilon-$contaminated.

Receiver: thereafter, we will adopt the L_{ab} demapper for further study and analysis since it outperforms the most used LLR approximations [HLZ14, MJLC15]. This work can be extended with more or less efforts to other approximations or to higher modulation schemes.

Parameter optimization method: we aim at the smallest BER and FER; but this criterion is not within reach in practice. Moreover, we constraint ourselves to have a criterion that can be used directly at the receiver (to achieve online real-time parameter estimation) which requires: simple implementation and fast learning (on the fly) in order to match practical receiver implementation. We propose to estimate the LLR approximation parameters by maximizing the mutual information between the source and the channel output (3.14). This does not rely on the knowledge of the noise distribution. We consequently expect robustness of this approach in different interference contexts. The justification of such an approach will be tackled in the following.

3.5 Online real-time parameter estimation method

The LLR approximations depend on several parameters, grouped here under the variable θ, which must be optimized to make the approximation as close as possible to the LLR.

3.5.1 LLR approximation as an optimization problem

We propose to perform the estimation under an information theory criterion based on the capacity of MBISO channel, which can be expressed as a function of the

LLR [RU08]. Recall that the capacity of such a channel is given as

$$C_\Lambda = H(X) - H(X|Y) = 1 - \mathbb{E}\left[\log_2\left(1+e^{-X\Lambda(Y)}\right)\right], \tag{3.13}$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator over the pair of random variables X, Y. Under the approximated LLR L_θ, one does not have access to (3.13) but to the lower bound

$$\widehat{C}_{L_\theta} = H(X) - \widehat{H}(X|Y) = 1 - \mathbb{E}\left[\log_2\left(1+e^{-XL_\theta(Y)}\right)\right]. \tag{3.14}$$

In order to evaluate the accuracy of this approximation, we follow the work of [YA09]. Approximating the LLR is equivalent to approximate $p(x|y) = 1/(1+e^{-x\Lambda(y)})$ by $q(x|y) = 1/(1+e^{-xL_\theta(y)})$ and thus $H(X|Y)$ by $\widehat{H}(X|Y) = \mathbb{E}\left[\log_2\left(1+e^{-XL_\theta(Y)}\right)\right]$.

Whereas (3.14) is only a heuristic, it appears to be a good criterion. Indeed the difference between the true (3.13) and the approximated (3.14) capacities is directly related to the Kullback-Leibler divergence between the densities $p(X|Y)$ and $q(X|Y)$ as

$$\begin{aligned} C_\Lambda - \widehat{C}_{L_\theta} &= H(X|Y) - \widehat{H}(X|Y) \\ &= \mathbb{E}\left[\log_2 p(X|Y)\right] - \mathbb{E}\left[\log_2 q(X|Y)\right] \\ &= \mathbb{E}\left[\log_2 \frac{p(X|Y)}{q(X|Y)}\right] \\ &= D_{KL}\left(p \,||\, q\right) \end{aligned} \tag{3.15}$$

where $D_{KL}(p\|q)$ is the Kullback-Leibler divergence between densities p and q [CT06]. We draw several facts from (3.15): the non-negativity of the divergence implies that our criterion is lower bounded, and the bound is reached when $q(x|y) = p(x|y)$. In other words, $\widehat{C}_{L_\theta} = C_\Lambda$ for $L_\theta = \Lambda$ if the LLR Λ belongs to the parametric family L_θ. Consequently, our objective is to minimize the distance between both conditional entropies by bringing both densities closer, and thus, maximizing (3.14). Equivalently, our optimization problem writes as:

$$\theta^* = \arg\min_\theta \widehat{H}(X|Y) = \arg\min_\theta \mathbb{E}\left[\log_2\left(1+e^{-XL_\theta(Y)}\right)\right] \tag{3.16}$$

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 use a Monte Carlo approach to estimate it, replacing the expectation by an empirical average $\widehat{H}_K(X|Y)$. Hence $\widehat{H}(X|Y)$ can be obtained as

$$\widehat{H}(X|Y) \approx \widehat{H}_K(X|Y) = \frac{1}{K}\sum_{k=1}^{K} \log_2\left(1+e^{-x_k\, L_\theta(y_k)}\right), \tag{3.17}$$

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 (3.17) over the possible choices of θ. 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

$$\begin{aligned}\theta^* &= \arg\min_{\theta} \widehat{H}_K(X|Y) \\ &= \arg\min_{\theta} \frac{1}{K}\sum_{k=1}^{K} \log_2\left(1+e^{-x_k L_\theta(y_k)}\right) \\ &= \arg\min_{\theta} \frac{1}{K}\sum_{n=1}^{K} \log_2\left(1+e^{-L_\theta(x_k y_k)}\right),\end{aligned} \tag{3.18}$$

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

Finally, one can rewrite the objective function as

$$\begin{aligned}\widehat{H}_K(X|Y) = \frac{1}{K} &\sum_{\substack{k=1 \\ x_k y_k \geq 0}}^{K} \log_2\left(1+e^{-L_\theta(x_k y_k)}\right) \\ &+ \frac{1}{K}\sum_{\substack{k=1 \\ x_k y_k < 0}}^{K} \log_2\left(1+e^{-L_\theta(x_k y_k)}\right).\end{aligned} \tag{3.19}$$

The proposed learning criterion is attractive because the optimum is achieved if the *a posteriori* is well approximated. Even without knowing the LLR, it is possible to approach it by learning. The obtained performance will show that such a model is efficient. The complexity study of the estimation step is out of the scope of this thesis. Nevertheless, the used simplex method based on Nelder Mead-algorithm [NM65] converges within 10 iterations, which is suitable for our application. Other optimization methods could be used such as Newton's descent, but as we will see shortly, the obtained value with the simplex method falls within the region of small BER.

Note that various LLR approximations can fit into this affine framework as we proposed some in Section 3.2.4. Nevertheless, other approximations that do not belong to our considered piece-wise affine function can be found in the literature, as for instance the Hole puncher demapper [SN93] or non-linear approximation like in [HLZ14]. However, the proposed framework remains valid but attention has to be paid to the optimization algorithm that will be used.

3.5.2 Analysis of the optimization problem

The authors in [YA11], proved the convexity of (3.18) by considering piecewise linear (in y) LLR approximations with fixed boundary regions assumption under Gaussian noise. However, our proposal does not assume linearity in y (only in θ) and boundaries are not fixed.

For clarity reasons, let us rewrite the odd piece-wise function given in (3.7) by taking into account the sign of the product $x_k y_k$, which will be

$$L_\theta(x_k y_k) = \begin{cases} \min\left\{\theta_1\phi_1(x_k y_k), \theta_2\phi_2(x_k y_k), \ldots, \theta_n\phi_n(x_k y_k)\right\} & \text{if } x_k y_k > 0 \\ \max\left\{\theta_1\phi_1(x_k y_k), \theta_2\phi_2(x_k y_k), \ldots, \theta_n\phi_n(x_k y_k)\right\} & \text{if } x_k y_k < 0 \end{cases} \tag{3.20}$$

In order to minimize (3.19), one needs to minimize a sum of terms, which can be treated separately according to the sign of $x_k y_k$. Let $f_\theta = \log_2(1 + e^{-L_\theta(x_k y_k)})$, hence, on one hand, if $x_k y_k > 0$ then

$$\begin{aligned} f(\theta) &= \log_2\left(1 + e^{-\min\{\theta_1\phi_1(x_k y_k),\ldots,\theta_n\phi_n(x_k y_k)\}}\right) \\ &= \log_2\left(1 + \max\left\{e^{-\theta_1\phi_1(x_k y_k)}, \ldots, e^{-\theta_n\phi_n(x_k y_k)}\right\}\right) \\ &= \max_i \left\{\log_2(1 + e^{-\theta_i\phi_i(x_k y_k)})\right\}. \end{aligned} \tag{3.21}$$

Consequently, in order to minimize (3.21), one needs to increase the parameters θ_i, because the term inside $\max\{\cdot\}$ decrease with θ_i. On the other hand, if $x_k y_k < 0$, then

$$\begin{aligned} f(\theta) &= \log_2\left(1 + e^{-\max\{\theta_1\phi_1(x_k y_k),\ldots,\theta_n\phi_n(x_k y_k)\}}\right) \\ &= \log_2\left(1 + \min\left\{e^{-\theta_1\phi_1(x_k y_k)}, \ldots, e^{-\theta_n\phi_n(x_k y_k)}\right\}\right) \\ &= \min_i \left\{\log_2(1 + e^{-\theta_i\phi_i(x_k y_k)})\right\}. \end{aligned} \tag{3.22}$$

In order to minimize (3.22), one thus needs to decrease the parameters θ_i. Thus, minimizing $\widehat{H}_K(X|Y)$ results in a compromise between minimizing each of the two parts, 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 θ, 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 [NM65] to obtain at least a local minimum. The use of an algorithm adapted to non-convex methods does not result in any significant gain. It could, however, be different for other approximation families, for instance, the non-linear approximation [HLZ14]. However, the best optimization method, as well as its complexity study, remains out of the scope of this thesis.

In the following chapters, we are going to investigate and study the behavior of our framework. Under long and short block length regimes the parameter optimization will be tackled in two different ways: using supervised learning and unsupervised learning as will be handled in Chapter 4 and Chapter 5, respectively.

3.6 Conclusion

Establishing reliable and efficient communications require to take into account the impulsive nature of interference while designing the receivers. To understand the impact of impulsive noise we started by visualizing the noise impact on the optimal receiver by representing the decision regions. Under the Gaussian case, the optimal decision regions are linearly separated, however, in the presence of impulsive noise these regions are non-linearly separated. Then, we presented different approaches to design a robust receiver: optimal approach; noise distribution approximation approach; different metric measures approach and the LLR inspired approach. We show that the optimal receiver cannot be achieved as the LLR is

computationally prohibitive and it relies on the noise-plus-interference knowledge, thus, sub-optimal receivers are salutary. We show that the aforementioned second and third approaches may be well suited to a specific noise context, but, their performance is subject to channel estimation accuracy. Designing a specific receiver for each situation is not as efficient as the interference characteristics can highly vary due to many situations. Thus, having a receiver able to cope with a large set of different interference models (impulsive or not) and with different degrees of impulsiveness is highly desired. For this reason, we proposed to directly estimate the LLR, without relying on the interference plus noise statistics knowledge. We chose a LLR approximation function f_θ in a parametric family, flexible enough to represent many different communication contexts. Furthermore, we proposed a new method to estimate θ. This method is inspired by information theory where the parameter θ is estimated by maximizing the mutual information between the channel input and the channel output. The justifications behind this method are illustrated as well.

Chapter 4

A supervised LLR estimation with unknown noise distribution

In this chapter, we will study in detail the novel supervised-learning framework to perform estimation of the LLR approximation parameters, without relying on the interference plus noise knowledge, that we introduced in the previous chapter. We demonstrate the supervised learning in long and short block length regime. Moreover, we propose a near-optimal demapper which features an easy implementation. The estimation method is shown to be efficient in a large variety of noises and the receiver exhibits a near-optimal performance. In addition, we propose a simple metric that is based on a simple distance calculation to evaluate the performance of an approximation instead of using intensive Monte Carlo simulations to evaluate error rates. Eventually, we show that supervised learning suffers from significant performance degradation when the training sequence is shortened and present the reason behind this loss.

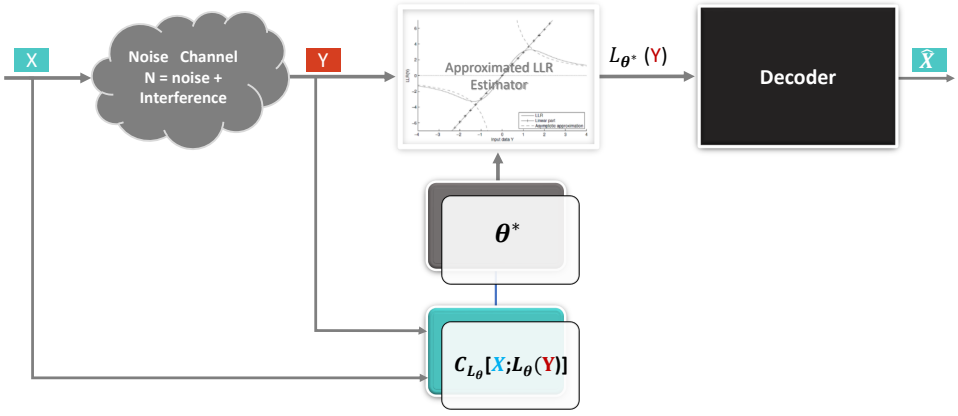


Fig. 4.1. Supervised LLR scheme.

4.1 Supervised scheme

If the optimization of the criterion $\widehat{H}_K(X|Y)$ (3.18) is performed in a supervised manner, the value of the parameter θ^* is triggered after receiving the channel output Y as shown in Figure 4.1. As we rely on Monte Carlo calculation to estimate (3.18), we need a known sequence of channel inputs, thanks to the learning sequence $x_1, \ldots, x_K$ and the corresponding outputs $y_1, \ldots, y_K$. Once the optimal parameter θ^* is obtained over the designed channel, it is used to build the approximated demapper for the rest of the block as $L_{\theta^*}(Y)$.

We investigate the accuracy of the supervised process with long block length performance analysis in Section 4.2, thereafter, we study the effect of shortening the training sequence in Section 4.3. For clarity reasons we first focus on SαS noise distributions, while we extend to other noise models with high and low impulsive cases, in particular, Middleton, $\epsilon-$contaminated and Gaussian to investigate the robustness and the ability to adapt to different contexts. In order to assess the optimal performance of the supervised approach, we use a large learning sequence of 20000 samples to estimate a and b. This allows having a good grasp of the results with high estimation accuracy. To evaluate the performance of the proposed scheme we will use L_{ab} (3.10) and present the BER curves using LDPC codes.

4.2 Supervised learning with long block length regime

In this section, we investigate the accuracy of the supervised parameters estimation.

4.2.1 Parameter estimation performance analysis

The performance of the estimation process is evaluated over different noise models. To perform so, we compare the LLR shape obtained by the tuned θ^* with the true LLR. In order to be efficient, we can expect that the approximated demapper L_{ab} is close to the true one.

Estimation over impulsive SαS additive noise

Figure 4.2 and Figure 4.3 illustrates the LLR behavior of a SαS channel output for fixed value of α and by varying γ. First note, we can see that in all cases linear and asymptotic regions appear, as described in Section 3.2.4. Second, for each channel state realization depicted by the couple α, γ, different LLR shapes are observed, thus, to match the channel state it is important to estimate correctly the LLR parameters. Third, by comparing the asymptotic part behavior between Figure 4.3 and Figure 4.2 we can see that as the environment becomes more impulsive (by decreasing α) the LLR asymptotic part will be almost the same in spite of varying γ. Consequently, we can expect that the LLR approximation in high impulsive environments will be less sensitive to parameters' variations.

Figure 4.4 compares the evolution of the mean of the estimated parameters a and b, as a function of the dispersion γ of a SαS noise with $\alpha = 1.4$ for the supervised

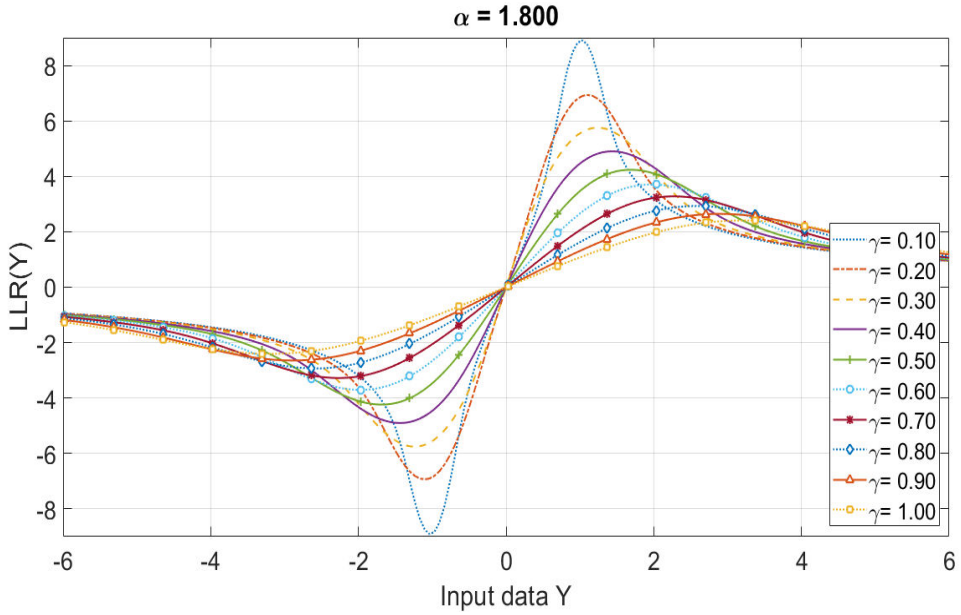


Fig. 4.2. Behavior of the LLR according to the output of the SαS channel for a fixed $\alpha = 1.8$ and different values of γ.

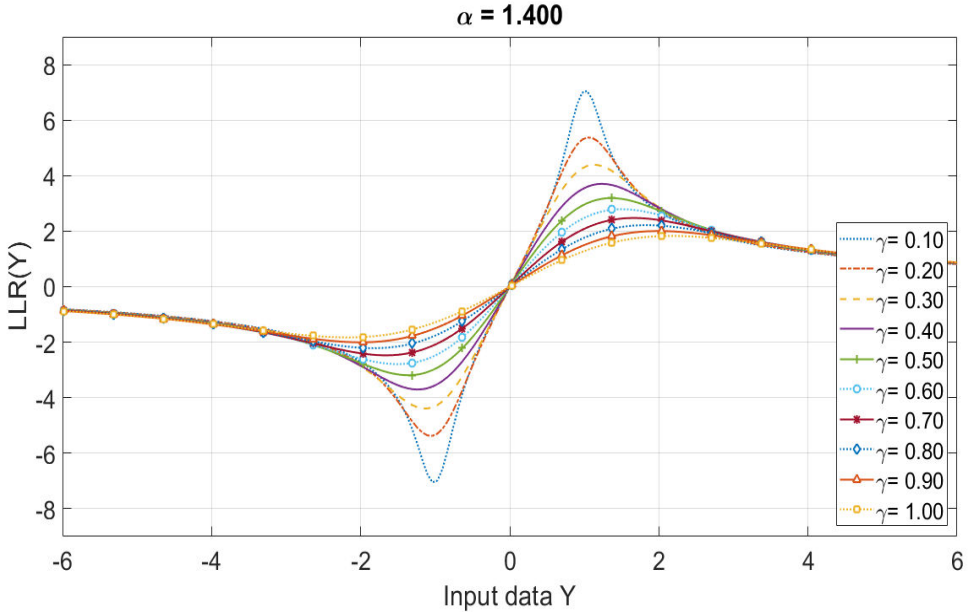


Fig. 4.3. Behavior of the LLR according to the output of the SαS channel for a fixed $\alpha = 1.4$ and different values of γ.

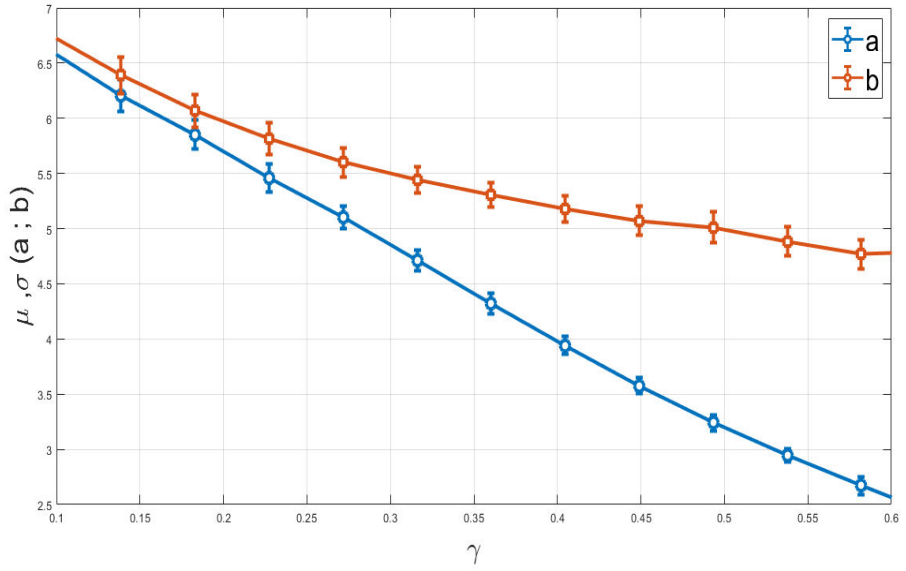


Fig. 4.4. Comparison of the mean and standard deviation evolution for parameter a and b as a function of the dispersion γ of a SαS noise with $\alpha = 1.4$ for the supervised optimization.

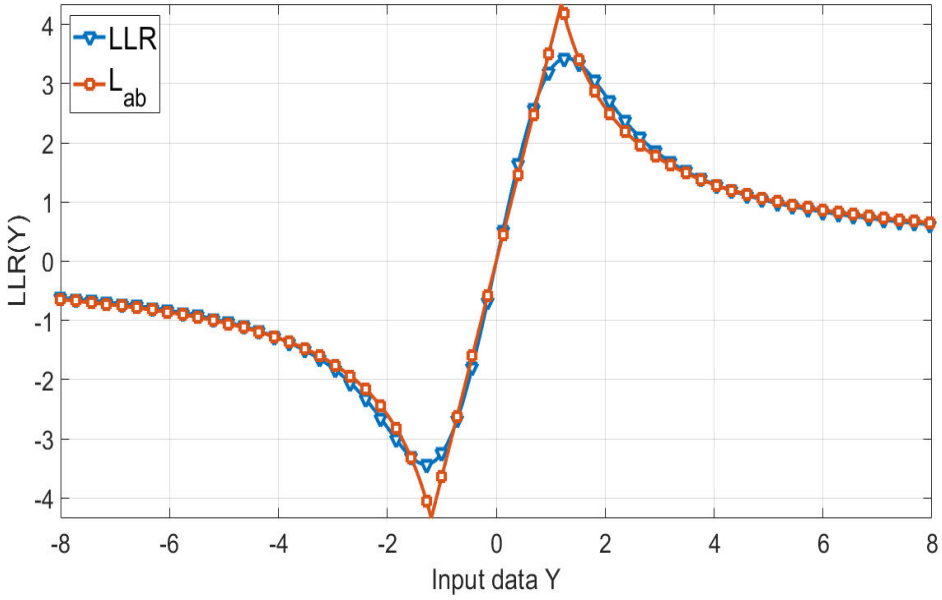


Fig. 4.5. Comparison of the LLR shapes L_{ab} approximation with the LLR obtained by numerical integration. Under the effect of the estimated a and b parameters with $\gamma = 0.45$ and $\alpha = 1.4$.

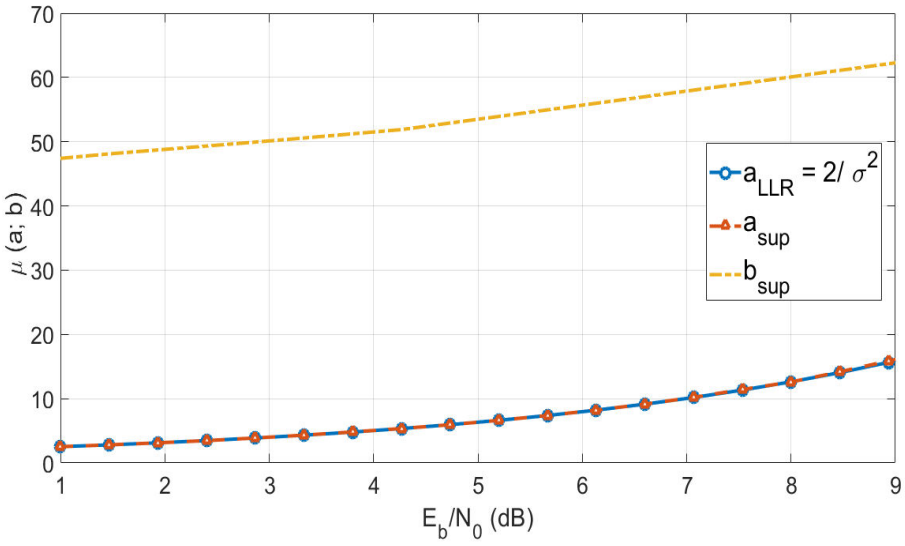


Fig. 4.6. Comparison of the mean evolution for parameter a as a function of E_b/N_0, under the Gaussian noise, the supervised optimization and the optimal receiver.

optimization. Remark that the same behaviors can be obtained for other values of α between 0 and 2. For each channel state, a and b are resultant of 5000 experiments. We notice that the variability in the estimated values is very small.

Figure 4.5 compares the LLR shapes of L_{ab} obtained under supervised optimization with the true LLR obtained by numerical integration. For this comparison we select from Figure 4.4 $\gamma = 0.45$ (depicts the waterfall region) the estimated parameters $a = 3.6$ and $b = 5.1$ that will be used to tune the L_{ab} demapper. This comparison shows the good fit between the approximated demapper L_{ab} and the true LLR. Accordingly, we can expect a good BER performance.

Estimation over Gaussian noise

It is an essential feature that our proposal does not degrade the performance of communication when no impulsiveness is present. Figure 4.6 compares the evolution of the mean of the optimized parameter a and b obtained under the supervised approach to the optimal slope in the linear receiver a_{LLR}. For each noise variance, we ran 5000 experiments. Recall that with our proposed demapper, the decreasing part starts at $\pm\sqrt{b/a}$. If this value becomes large enough, the power-law part will not impact the receiver. For Gaussian noise, large received samples are very rare, so that all samples fall into the linear part of the demapper. Numerically, when simulating AWGN channel, we obtained values bigger than 49 for b as shown in Figure 4.6 which are larger than the one obtained for a, showing that only the linear part impacts the receiver[1]. The decreasing part has consequently a negligible impact. The linear part is given by the optimal slope as $a = 2/\sigma^2$. This comparison shows

[1] As E_b/N_0 increases the noise standard deviation σ will decrease, making thus all the received samples well-framed around the mean, that is the amplitude are mostly within 3σ of the mean. Consequently, all the received samples fall between $\pm\sqrt{b/a}$ values and thus effected solely by the linear part.

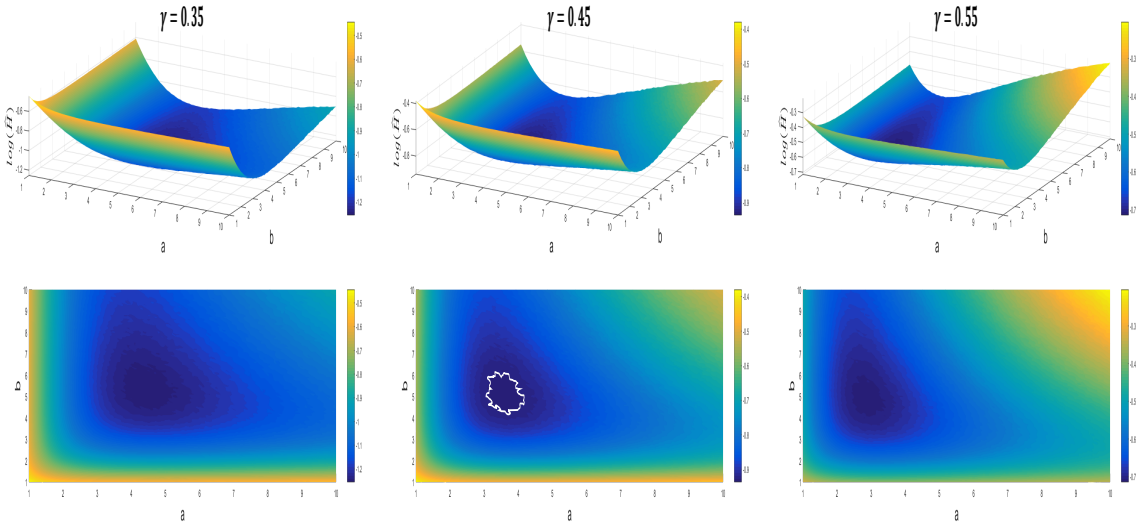


Fig. 4.7. Behavior study of the $\widehat{H}_K$ function, as a function of parameters a and b for different values of γ, under highly impulsive $S\alpha S$ noise with $\alpha = 1.4$.

clearly the good fit between the slopes, proving thus the strength of the supervised optimization of the demapping function L_{ab} over the AWGN channel.

4.2.2 Performance evaluation

Indirect link between the optimization method and minimizing BER

In a first step, we investigate the shape of the function $\widehat{H}_K$ (3.18) according to $\theta = (a, b)$. 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 Figure 4.7, we represent a 3D plot of the function $\widehat{H}_K$ for three values of γ, namely $\gamma = 0.35$, $\gamma = 0.45$ and $\gamma = 0.55$, representing the levels of $\widehat{H}_K$ under the supervised criterion using a learning sequence of length 20000. The white contour in the middle figure delineates the area where the $\widehat{H}_K$ is less than 0.3 and this value decreases as γ decrease. The values of γ are selected in a way to represent the shape of the function $\widehat{H}_K$ before, within and after the waterfall region, respectively, as will be seen later in Section 4.2.1. First note that $\widehat{H}_K$ is quite flat around its minimum value. As a consequence, it may be quite sensitive to noise and thus to the length of the training sequence.

In Figure 4.8, we illustrate the link between the function $\widehat{H}_K$ and the obtained BER. The contour plot delineates different BER values, ranging from 10^{-5} to 10^{-1}, whereas the white contour delineates the set of a and b values yielding the smallest values of $\widehat{H}_K$ as shown in the middle figure of Figure 4.7. Note that the regions match in the sense that the set of optimal values for a and b allows the decoder to achieve a BER below 10^{-5}.

Indeed, as shown in Figure 4.8 when $\gamma = 0.45$, both a and b estimated mean values under the supervised optimization shown in Figure 4.4 fall in the small BER region. Besides, the small variance of the estimated θ^* ensures that the estimated

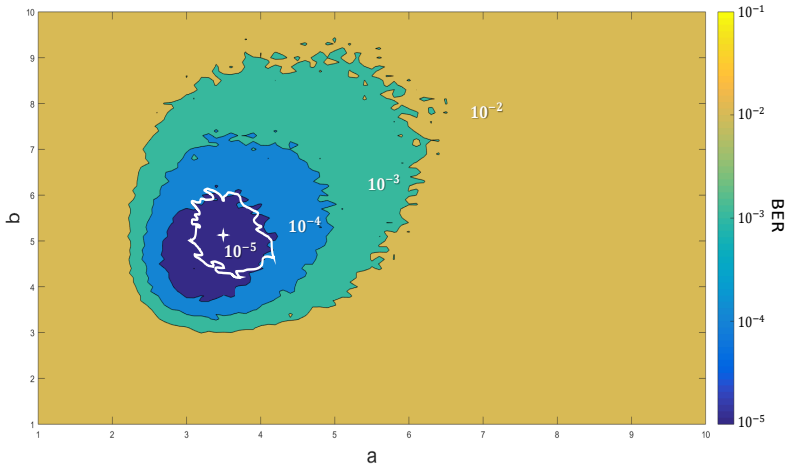


Fig. 4.8. BER comparison as a function of a and b parameters with $\gamma = 0.45$ and $\alpha = 1.4$ under the supervised approximation.

values will fall in the region yielding the smallest BER.

The sensitivity to errors due to the flatness of the landscape of $\widehat{H}_K$ is thus lessened by the flatness of the BER region.

BER performance under SαS noise

Once our demapper is tuned with the estimated value θ^*, it is used as a front-end to the 20000 bits long regular (3,6) LDPC decoder using the BP algorithm. We test this scheme first over an additive impulsive $S\alpha S$ noise.

In this case, we study a highly impulsive situation with $\alpha = 1.4$ and a less impulsive case with $\alpha = 1.8$. Figure 4.9 and Figure 4.10 present the BER for $\alpha = 1.4$ and $\alpha = 1.8$ respectively, as a function of the dispersion γ of the α-stable noise[2]. In both cases, we compare the BER obtained via the demapping function in a supervised manner, to the BER obtained with the true LLR computed via numerical integration.

First, we note that in both cases, the estimation with such a long training sequence gives performance close to the optimal LLR which shows the good behavior of our solution. Secondly, we can figure out the ability to adapt to different impulsive states from low impulsive up to highly impulsive.

In order to stress the need of good LLR shape design, we also optimized with our solution a linear approximation L_a. This receiver shows very poor performance behavior in all impulsive cases. Thus, the impulsive nature must be tackled with a decreasing part of the LLR approximation such as with L_{ab}.

[2]Recall that in case of an impulsive environment with $\alpha < 2$, the second-order moment of a stable variable is infinite [BSF08, Theorem 3], making the conventional noise power measurement infinite. Accordingly, we present our simulation results as a function of the dispersion parameter γ, which is used as a measurement of the strength of the α-stable noise.

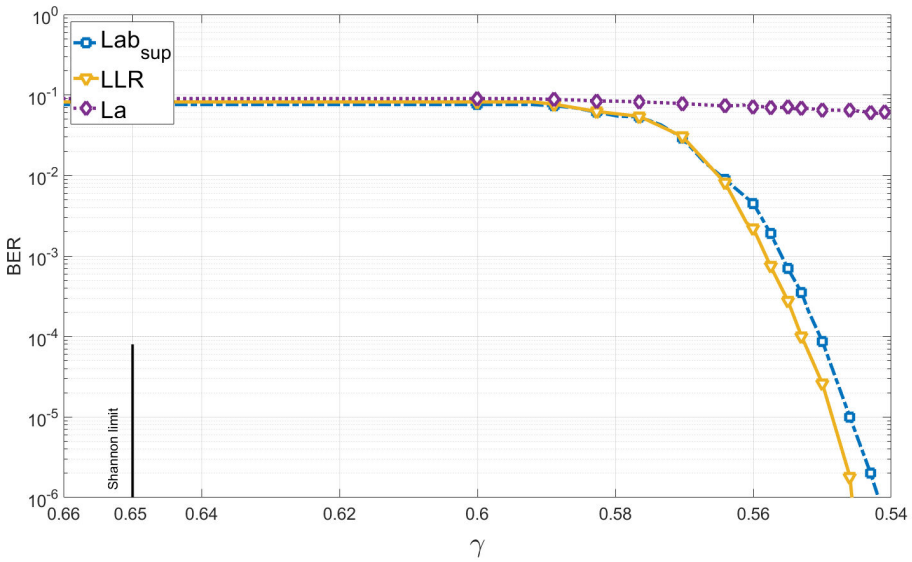


Fig. 4.9. Evolution comparison of the BER as a function of the dispersion γ of a SαS noise in low impulsive environment with $\alpha = 1.8$, between the supervised, linear LLR approximations and the LLR obtained by numerical integration.

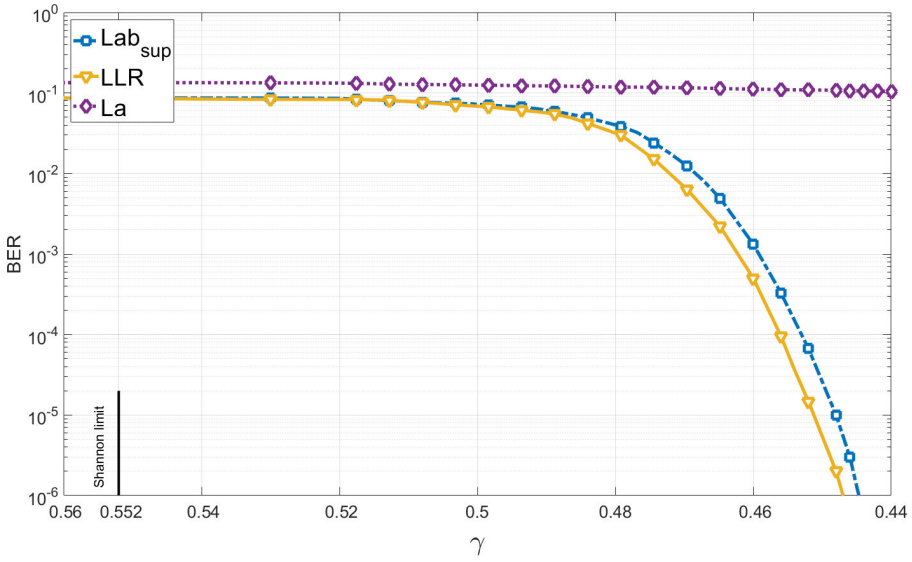


Fig. 4.10. Evolution comparison of the BER as a function of the dispersion γ of a SαS noise in highly impulsive environment with $\alpha = 1.4$, between the supervised, linear LLR approximations and the LLR obtained by numerical integration.

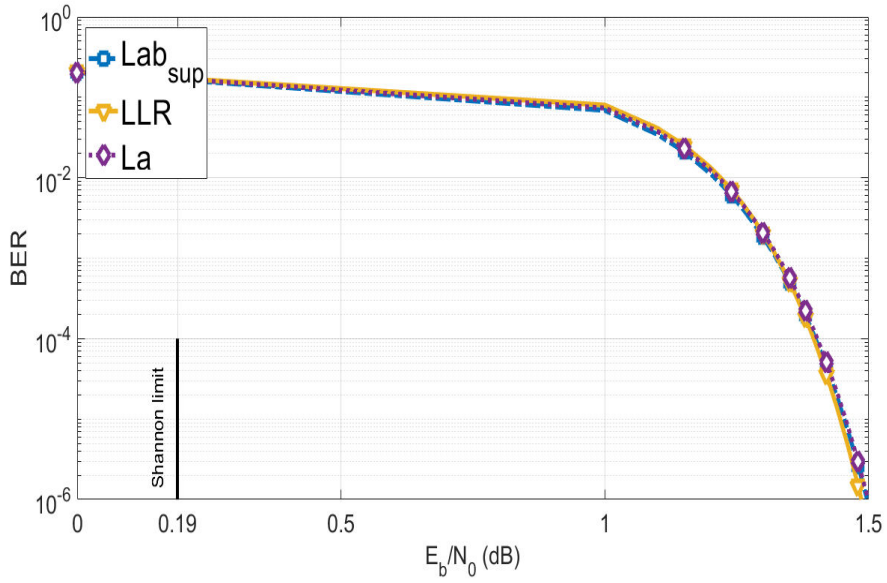


Fig. 4.11. BER comparison as a function of E_b/N_0 between the supervised approximation and the optimal LLR in AWGN channel.

Investigation of the robustness and adaptability of the proposed framework

In order to further show the robustness of our proposed demapper L_θ, we investigate in the following its performance when instead of suffering from a $S\alpha S$ noise, the channel exhibits an impulsive nature modeled by Middleton Class A [SM77] or ε-contaminated noises [AALM17b] or non-impulsive like the Gaussian case. We keep the linear approximation L_a and the proposed LLR approximation L_{ab} and test them under different configurations:

- additive Gaussian noise, in Figure 4.11;
- low impulsive ε-contaminated ($\varepsilon = 0.01, K = 10$), in Figure 4.13;
- high impulsive ε-contaminated ($\varepsilon = 0.1, K = 10$), in Figure 4.12;
- high impulsive Middleton class A ($A = 0.01, \Gamma = 0.01$), in Figure 4.14;
- moderate impulsive Middleton class A ($A = 0.1, \Gamma = 0.1$), in Figure 4.15.

Note that in these cases, one can compute the noise variance, thus the error rates can be given as a function of E_b/N_0. For each scenario, we compare our proposal based on L_{ab} with the true LLR, obtained via numerical integration, and the linear demapper L_a. For each channel set, in the supervised case, a learning sequence of length 20000 is used to optimize θ. For each case, we studied the evolution of the BER as a function of the E_b/N_0.

The high robustness and the adaptability of our demapper can be seen through the close performance obtained between the supervised case and the true LLR. Moreover, the ability to adapt to different noise type scenarios in spite of the change of the impulsiveness degree.

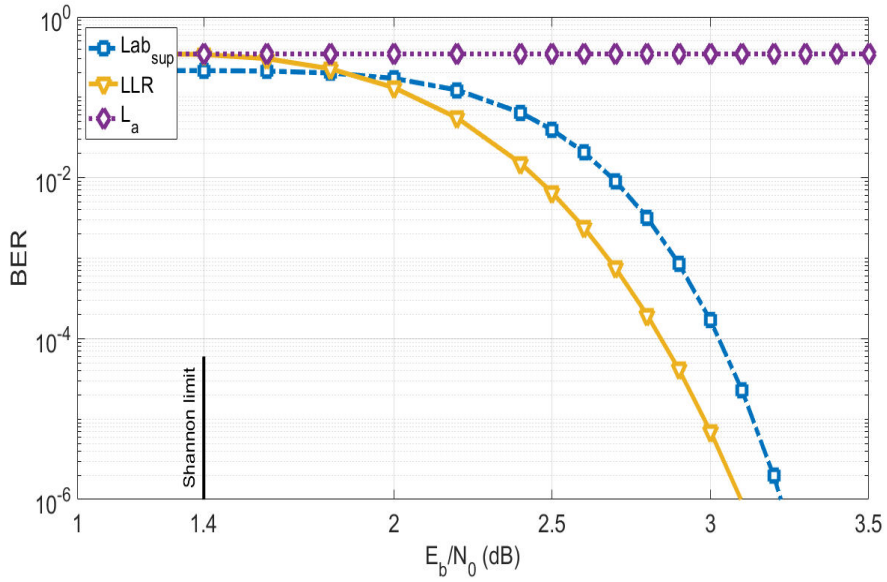


Fig. 4.12. BER comparison as a function of E_b/N_0 between the supervised, linear LLR approximations and the LLR obtained by numerical integration, in high ε-contaminated with $\varepsilon = 0.1$ and $K = 10$.

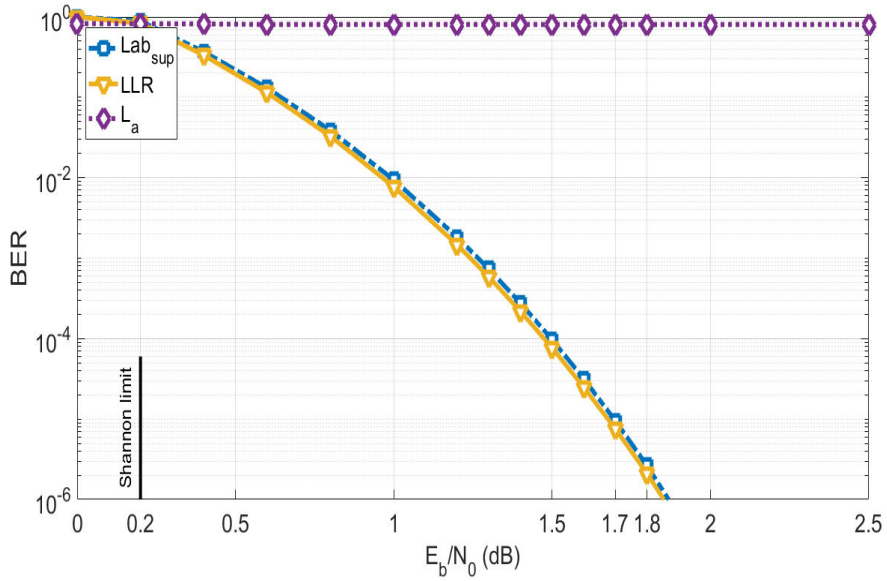


Fig. 4.13. BER comparison as a function of E_b/N_0 between the supervised, linear LLR approximations and the LLR obtained by numerical integration, in low ε-contaminated with $\varepsilon = 0.01$ and $K = 10$.

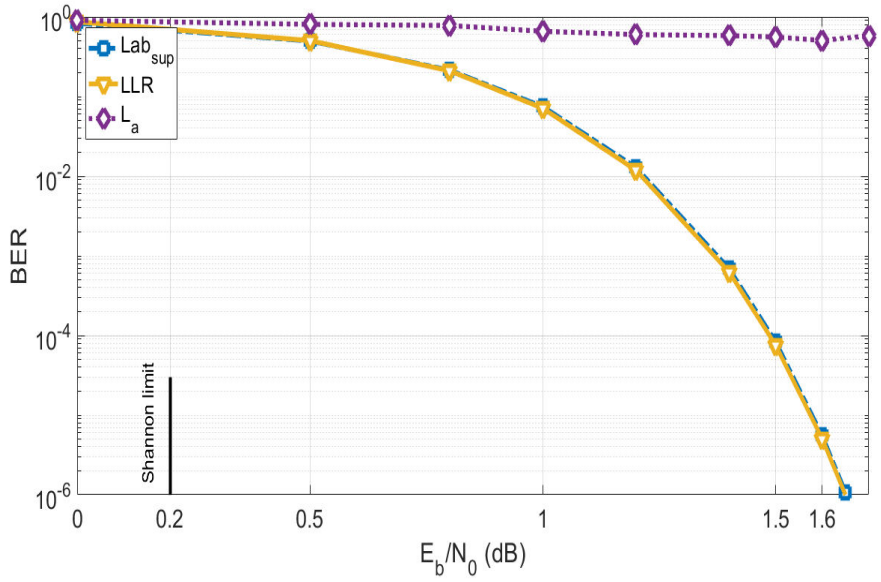


Fig. 4.14. BER comparison as a function of E_b/N_0 between the supervised, linear LLR approximations and the LLR obtained by numerical integration, in highly impulsive Middleton Class A noise with $A = 0.01$ and $\Gamma = 0.01$.

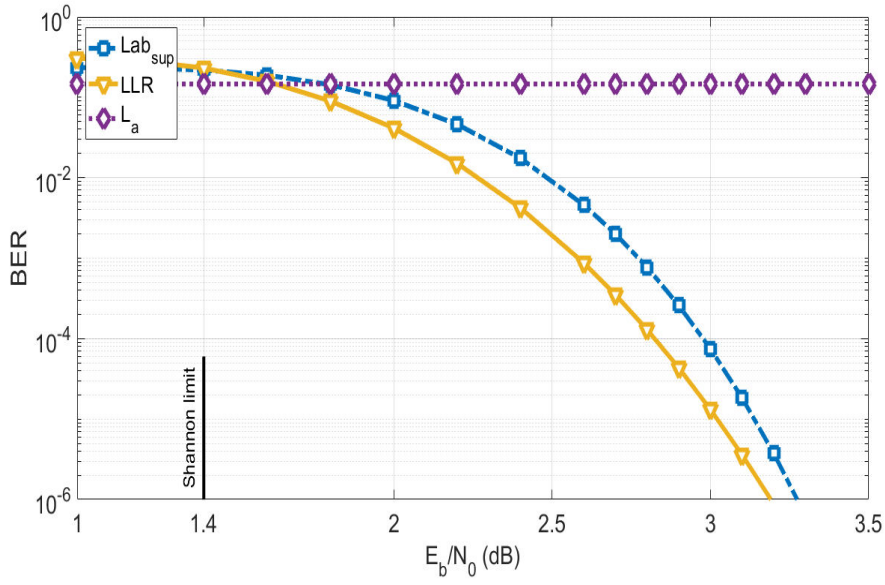


Fig. 4.15. BER comparison as a function of E_b/N_0 between the supervised, linear LLR approximations and the LLR obtained by numerical integration, in moderately impulsive Middleton Class A noise with $A = 0.1$ and $\Gamma = 0.1$.

Analysis and discussion: these numerical simulations illustrate the universality of the supervised approach framework. The LLR family has to be wide enough to encompass 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 that does not depend on any noise assumption. Consequently, this allows our supervised framework to adapt to different noise models with different impulsive degrees.

The gap between the supervised optimization and the true LLR is small in all the studied examples. We extended this study to other noises parameters or a mixture of distributions, for instance, the Gaussian plus stable noise and obtained similar conclusions. In impulsive situations, the gap between the non-linear LLR approximation L_θ 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 impact the performance when noise is not impulsive so that we do not need a detection step to distinguish between Gaussian and impulsive situations.

In the aforementioned study, we used a long learning sequence of 20000 samples to estimate a and b. Consequently, the performance of the decoder depends only on the quality of the LLR approximation and not on estimation errors. Thus, the existing gap compared to the true LLR in the BER figures especially when the impulsiveness increases is relevant to the approximated LLR by itself. In the sake of completeness, we deal with this loss by proposing in the next section a new demapper with an added parameter which allows to significantly reduce the gap.

4.2.3 A robust and simple LLR approximation for receiver design

Note that the approximated LLR receiver L_{ab} will remain one reference as it gives the best performance compared to other proposed approximations [HLZ14], [MJLC15].

The main problem behind L_{ab} is the rough transition between the linear part and the asymptotic part, where one can note, see Figure 4.16, a large gap between the true LLR and the approximated one. The purple curve (or dots) denoted by the PDF of the received samples, represents the conditional probabilities $\Pr(Y|X=+1)$, respectively $\Pr(Y|X=-1)$. One can note that a large proportion of the received samples fall in the transition part, where the gap to the true LLR is the largest. Consequently, we propose a new approximation as represented in Figure 4.16. The third parameter corresponds to a saturation of the approximated LLR in the transition part, allowing to decrease the gap to the true LLR. This new approximation is detailed in the following.

Proposed Approximation

We call the new demapper L_{abc}. It decomposes the LLR into three regions:

- First, when the channel output y is small the linear approximation of the LLR

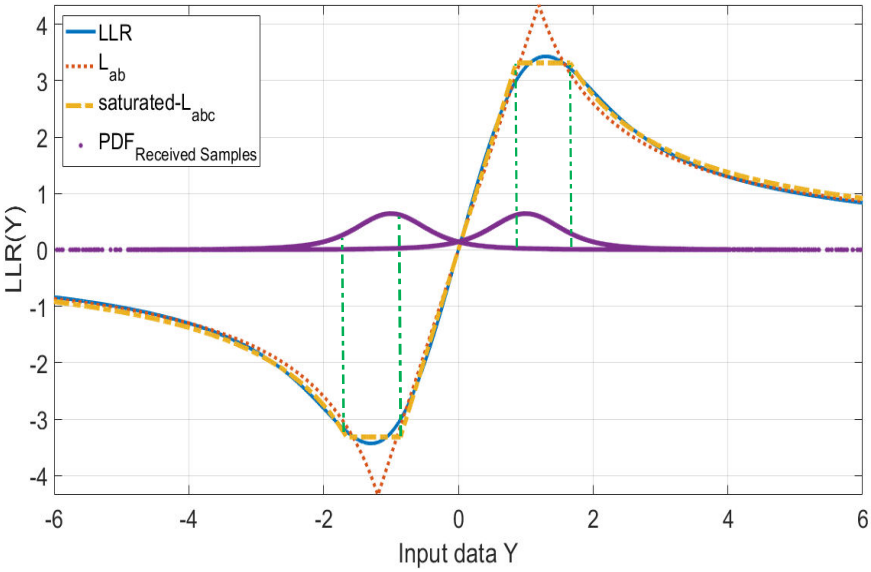


Fig. 4.16. Comparison of the LLR shapes under the effect of the estimated a and b parameters with $\gamma = 0.45$ and $\alpha = 1.4$, in the L_{ab} and L_{abc} approximations with the LLR obtained by numerical integration.

around zero is given by:

$$\begin{aligned} \mathrm{LLR}(y) &= \log \frac{f(y-1)}{f(y+1)} \\ &= \log \frac{f(-1) + f'(-1)y + O(y^2)}{f(1) + f'(1)y + O(y^2)} \\ &= -2\frac{f'(1)}{f(1)} y + O(y^3) \approx ay. \end{aligned} \tag{4.1}$$

- Second, using Proposition 2, a simple asymptotic expansion may well approximate the probability density function f_α of stable variables with unitary scale parameter [NS95],

$$f_\alpha(y) = \sum_{k=1}^{n} \frac{b_k}{|y|^{\alpha k+1}} + O\left(|y|^{-\alpha(n+1)-1}\right). \tag{4.2}$$

Using such asymptotic expansion of α-stable distribution provides a very simple LLR approximation based on the first term $f_\alpha(y) \propto |y|^{-(\alpha+1)}$ for large values of the channel output y,

$$\begin{aligned}\mathrm{LLR}(y) &= \log\frac{f(\frac{y-1}{\gamma})}{f(\frac{y+1}{\gamma})} \\ &\approx \log\frac{(y-1)^{-(\alpha+1)}}{(y+1)^{-(\alpha+1)}}, \qquad \text{for large values of } y, \\ &\approx -(\alpha+1)\log\frac{y-1}{y+1} \approx 2\frac{\alpha+1}{y}.\end{aligned} \tag{4.3}$$

- The third part makes the transition between the linear and the asymptotic parts. As shown in Figure 4.16, the true LLR behaves smoothly in this region, whereas with L_{ab} demapper, the transition is rather sharp. Moreover, the green shaded part illustrates the advantage of introducing a constant term to improve the approximation. In order to keep the simplicity and easy implementation of our demapper, we propose to introduce a new parameter c that saturate the LLR at this part.

The three aforementioned points lead to the demapping function

$$L_{abc}(y) = \mathrm{sgn}(y)\ \min\big(a|y|, b/|y|, c\big) = \begin{cases} ay & \text{if } |y| < c/a, \\ c & \text{if } c/a < |y| > bc, \\ b/y & \text{if } |y| > bc. \end{cases} \tag{4.4}$$

Discussion and Analysis

To match the channel situation, the receiver must be tuned by the optimized parameters $\theta^* = (a; b; c)$ using the same framework as for L_{ab} in previous sections.

Figure 4.16 compares the LLR shapes for L_{ab} $(a = 3.64; b = 5.19)$ and L_{abc} $(a = 3.86; b = 5.5; c = 3.31)$ to the true LLR obtained via numerical integration for a SαS noise of parameters $\alpha = 1.4$ and $\gamma = 0.45$, where this specific γ represents the waterfall region for such α and code. This comparison shows the convergence between the L_{abc} and the true LLR, and clearly shows the improvement in terms of LLR shapes compared to the demapper L_{ab}. Moreover, Figure 4.16 shows that the majority of samples will fall in the transition part as can be seen from the PDF of the received samples, for instance, for $\alpha = 1.4$ and $\gamma = 0.45$ the percentage error is (44 ± 4), we mean by percentage error the percentage of samples that fall in the transition phase where the gap to the true LLR is the largest. This percentage error is able to increase for different γ. In Figure 4.17, we present the frequency of receiving a sample within each region S_a (linear part), S_b (asymptotic part), or S_c (saturated part) obtained numerically using 1000 frames. For our simulation, we use low impulsive SαS noise where $\alpha = 1.8$. Figure 4.17 shows that receiving a sample that falls within the transition phase depicted by the c parameter is highly probable for a large range of noise scale γ. Furthermore, those samples highly influence the selection of the optimized parameters as well as the decoder performance in terms of BER.

Figure 4.18, respectively Figure 4.19, compares the evolution of the mean and variance of the optimized parameters $\theta^* = (a; b; c)$ as a function of the dispersion γ, for less or more impulsive SαS noise, respectively. For each channel state, we ran

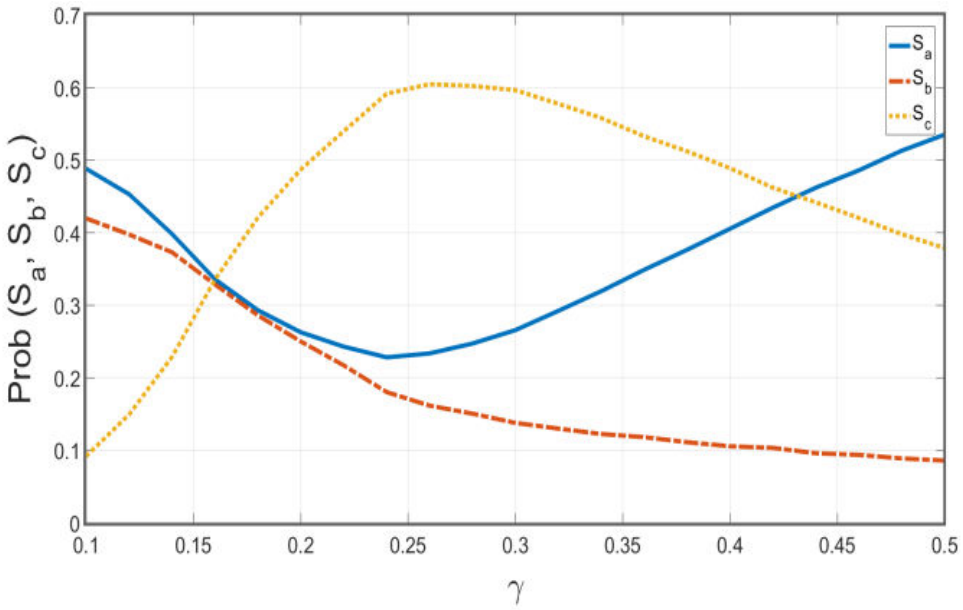


Fig. 4.17. Comparison of the mean evolution of the probability to fall within the a; b and c regions, as a function of the dispersion γ of a SαS noise with $\alpha = 1.8$ for the L_{abc}.

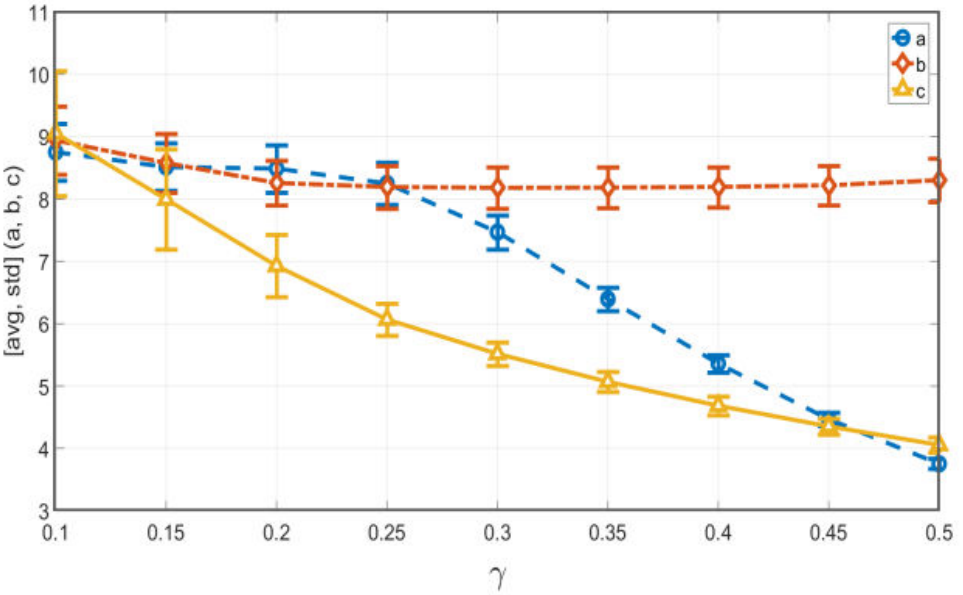


Fig. 4.18. Comparison of the mean and standard deviation evolution for parameter a; b and c as a function of the dispersion γ of a SαS noise with $\alpha = 1.8$ for the demapper L_{abc}.

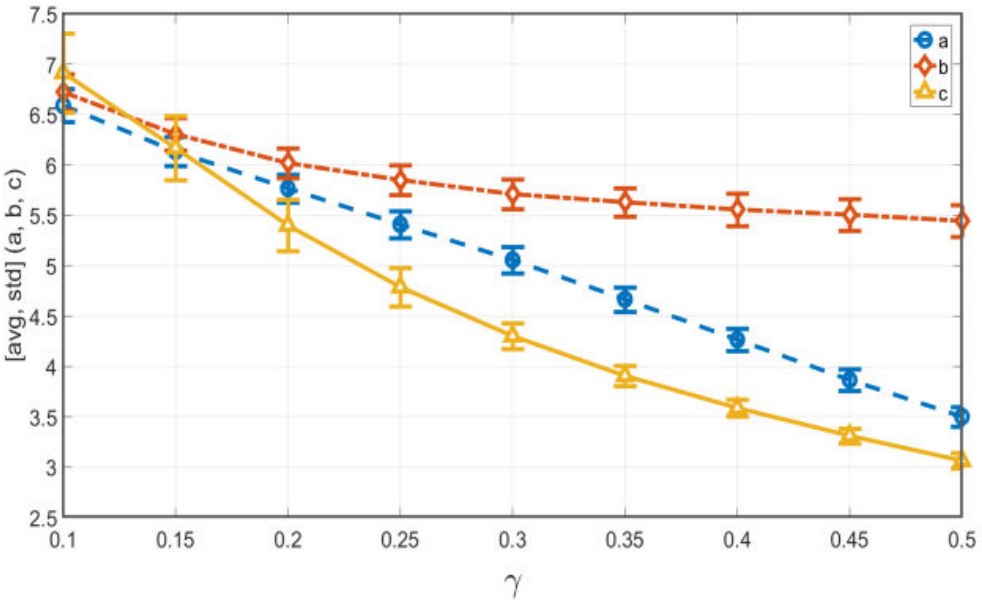


Fig. 4.19. Comparison of the mean and standard deviation evolution for parameter a; b and c as a function of the dispersion γ of a SαS noise with $\alpha = 1.4$ for the demapper L_{abc}.

10000 experiments. The error bars indicate the small impact of different realizations in which we can infer the robustness of the parameter estimation method.

Figure 4.20 compares the shape of the true LLR, and the one obtained with the demapper L_{ab} and L_{abc}. The third parameter c allows a better fit to the true LLR. Moreover, even in the linear and asymptotic parts, the fitting to the true LLR is better when three optimization parameters are used instead of only two. This can be explained by the fact that the new approximation suppresses the dependence between a and b which implies a compromise. These two parameters can be optimally chosen for their own region, without really impacting each other.

Performance Investigation

In Figure 4.21 and Figure 4.22 we present the BER and FER performance of different receivers for $\alpha = 1.8$ and $\alpha = 1.4$ which represent less and more impulsive channels, respectively. For each γ, the demapping functions L_{ab} and L_{abc} are compared with the optimal receiver. Our proposed solution improves the performance over the L_{ab} demapping function and matches the performance of the optimum receiver in both less and more impulsive channels, as it was expected, since the shape of the demapping function L_{abc} is really close to the shape of the true LLR. It should be mentioned that similar conclusions are observed for other values of α or other impulsive environment types, as for instance, Middleton Class A, ϵ-contaminated or Gaussian Noise.

This approximation is designed for additive impulsive noise channels, nevertheless, it is not computationally demanding and it remains easy to implement. It requires the estimation of three parameters and our proposed scheme is shown to be efficient. Moreover, in terms of performance, our solution is barely discernible from

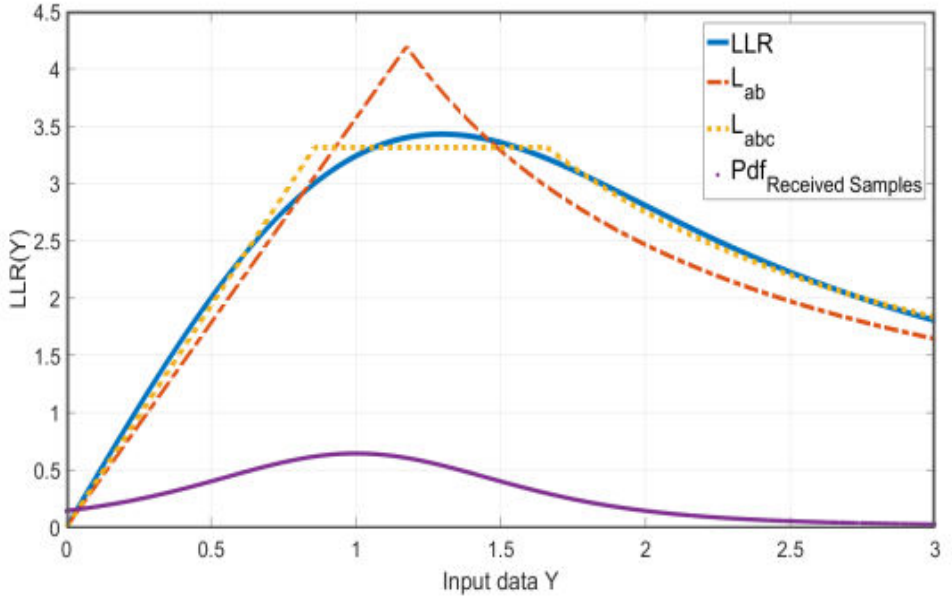


Fig. 4.20. Comparison of the LLR shapes obtained by numerical integration (true LLR) or with the approximation L_{ab} and L_{abc} under an SαS noise with $\gamma = 0.4$, $\alpha = 1.4$.

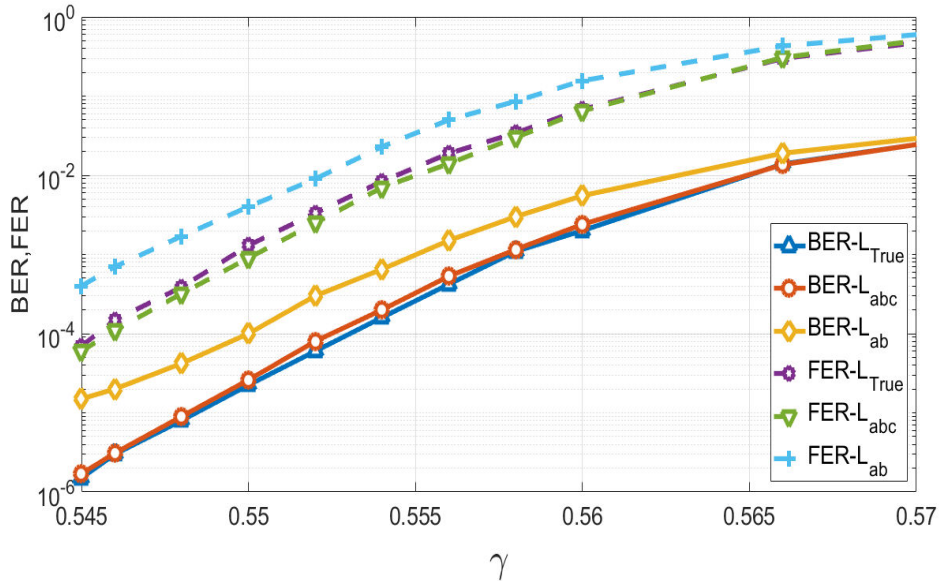


Fig. 4.21. Comparison of the BER and FER as a function of the dispersion γ of a SαS noise in less impulsive environment with $\alpha = 1.8$, between L_{ab}, L_{abc} and the LLR obtained by numerical integration.

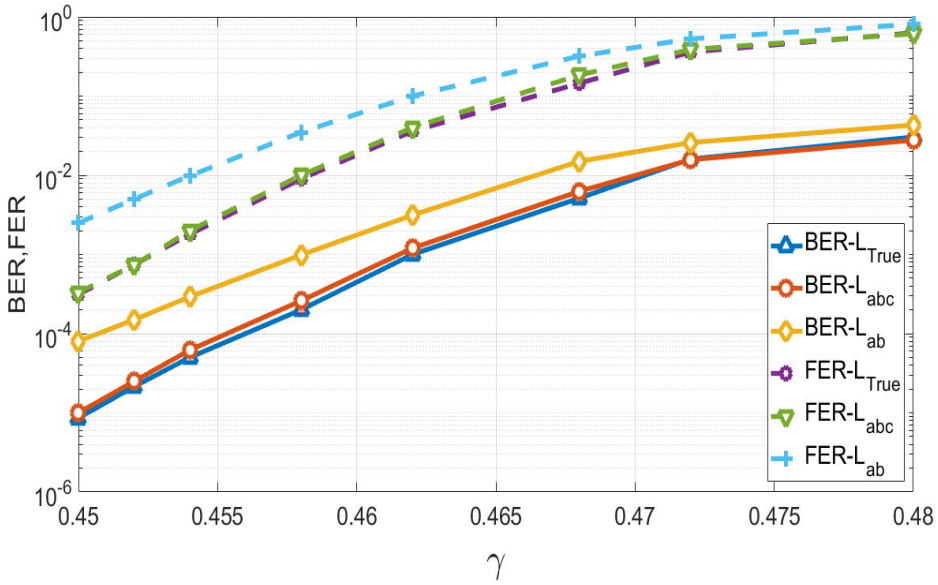


Fig. 4.22. Comparison of the BER and FER as a function of the dispersion γ of a SαS noise in highly impulsive environment with $\alpha = 1.4$, between L_{ab}, L_{abc} and the LLR obtained by numerical integration.

the optimal receiver which is computationally prohibitive.

4.3 Shortening the training sequence.

So far we considered a learning sequence of the size of the LDPC code, which allows to evaluate the LLR approximation without considering any performance impact induced by the estimation step. However, 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.

In this section we will study the effect of the estimation when shortening the learning sequence size.

We present in Table 4.1 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 training sequences under moderate and less impulsive SαS noise with $\alpha = 1.4$ and $\alpha = 1.8$. 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.

Note that one can expect a performance degradation in terms of BER by projecting the estimated values from Table 4.1 in Figure 4.7. Furthermore, in Figure 4.23 we compare the LLR shapes of the L_{ab} approximations obtained in the case of an α-stable noise with parameters $\alpha = 1.8$ and $\gamma = 0.55$ for $LS = 20000$ and $LS = 900$

			μ_a	σ_a	μ_b	σ_b
$\alpha = 1.4$	$\gamma = 0.43$	$\text{Sup}_{\text{LS=20000}}$	3.73	0.07	5.10	0.12
		$\text{Sup}_{\text{LS=1200}}$	3.77	0.35	5.18	0.58
		$\text{Sup}_{\text{LS=900}}$	3.77	0.40	5.28	0.71
	$\gamma = 0.45$	$\text{Sup}_{\text{LS=20000}}$	3.57	0.07	5.06	0.13
		$\text{Sup}_{\text{LS=1200}}$	3.60	0.32	5.15	0.58
		$\text{Sup}_{\text{LS=900}}$	3.61	0.37	5.16	0.66
$\alpha = 1.8$	$\gamma = 0.53$	$\text{Sup}_{\text{LS=20000}}$	3.25	0.05	7.59	0.28
		$\text{Sup}_{\text{LS=1200}}$	3.27	0.24	8.50	14.48
		$\text{Sup}_{\text{LS=900}}$	3.27	0.27	11.72	46.15
	$\gamma = 0.55$	$\text{Sup}_{\text{LS=20000}}$	3.05	0.05	7.62	0.28
		$\text{Sup}_{\text{LS=1200}}$	3.07	0.22	7.97	1.54
		$\text{Sup}_{\text{LS=900}}$	3.07	0.26	10.73	30.41

Table 4.1 – Comparison of the mean and standard deviation evolution for the parameters (a, b) as a function of the dispersion γ of a SαS noise with $\alpha = 1.4$ and $\alpha = 1.8$ for the supervised with different learning sequence sizes.

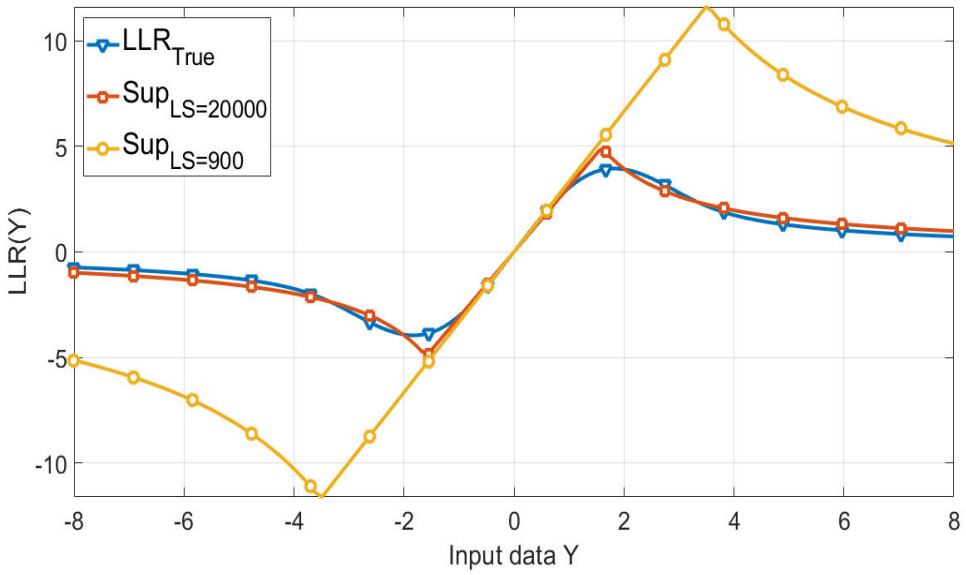


Fig. 4.23. Comparison of the L_{ab} approximations obtained in the case α-stable noise with parameters $\alpha = 1.8$ and $\gamma = 0.55$ for different LS sizes with the True LLR obtained using numerical calculations.

with the true LLR using the mean of the estimated parameters in Table 4.1. Obviously, the LLR shape obtained by the estimated parameters of $LS = 900$ is far from the LLR shape. While the LLR shape obtained by the estimated parameters of $LS = 20000$ is close to the true LLR shape. This degradation in the LLR shape will result in a bad matching of the actual channel output Y and eventually a degraded

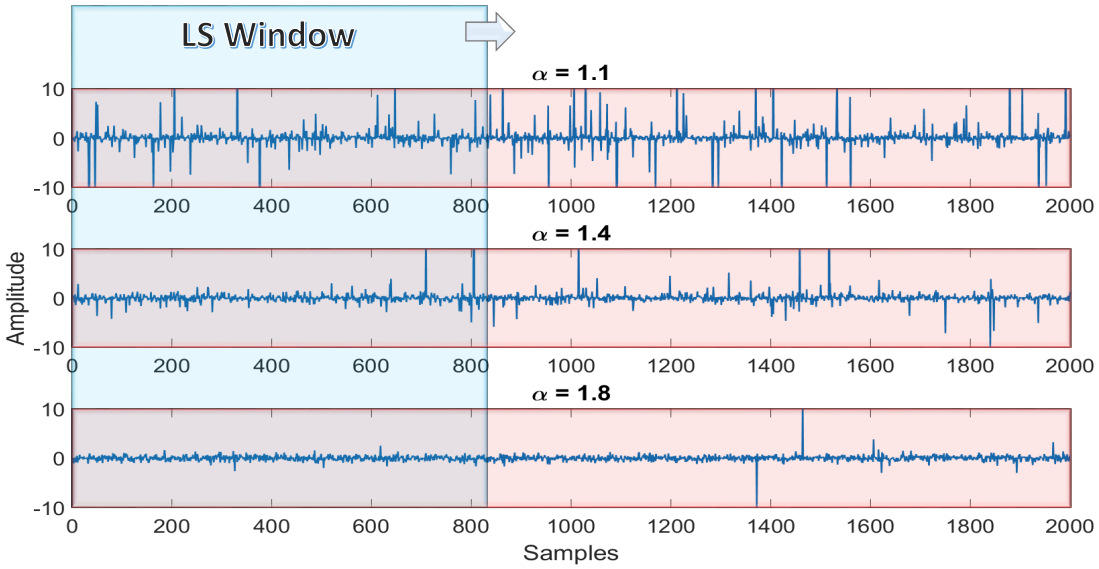


Fig. 4.24. Showing the mismatch.

performance in terms of BER and FER.

We also (note even if it is counter-intuitive) that the estimated parameters are degraded as α increases and γ decreases, in other words, when the channel becomes less impulsive. However, the intuitive way of thinking is that the more the channel becomes impulsive the most difficult the estimation is, which is shown to be wrong.

In fact, when the channel becomes less impulsive the probability to observe rare events (impulses) in the short training sequence (for instance LS=900) becomes smaller. However, this probability remains important in the whole message (20000). Thus, one can define a mismatch between what is learned from the LS and the real condition of the whole message. To have a better understanding we demonstrate the existence of such a phenomenon in the following.

Mismatch between training phase and decoding phase: the estimation step can be tricky with impulsive noises because the most significant events are rare. The risk is that what we learn from the learning sequence may not match the condition of the message, leading to severe performance degradation.

Figure 4.24 shows three different impulsive states with SαS noise: less (α = 1.8), medium (α = 1.4) and highly (α = 1.1) impulsive. The learning sequence (highlighted in the blue box) is chosen to be small (816 in our illustrative example). In each case, we compare the LS window to the up next 2000 received samples.

First, in Figure 4.24, α = 1.8 case, bottom plot, no impulse is observed in the training sequence and noise looks Gaussian. Consequently, the estimation will result in a linear demapper which is perfect in such a case. However, some impulses with high amplitudes do appear outside the LS window. Due to the linear effect, they will have a high likelihood and fool the receiver.

Second, in Figure 4.24, α = 1.4 case, few impulses fall in the LS window when

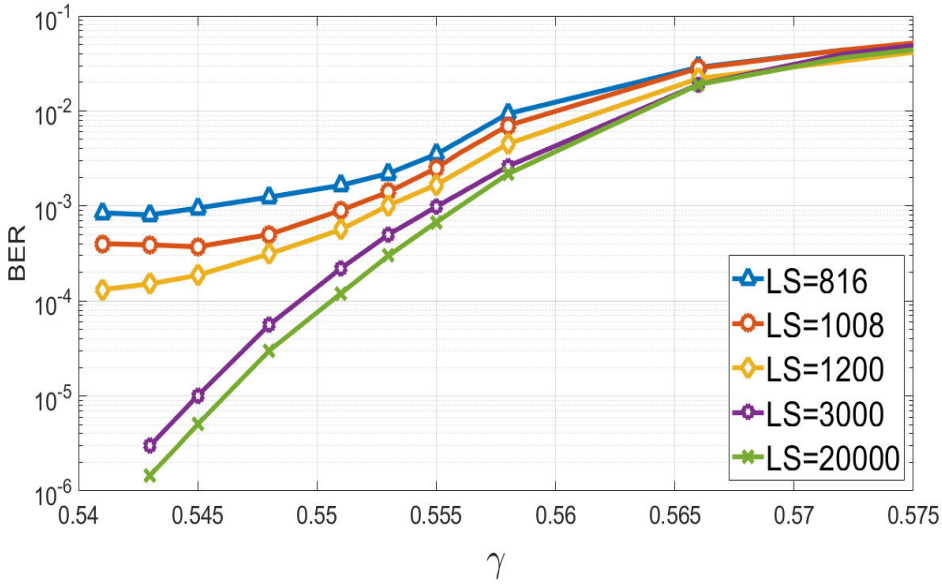


Fig. 4.25. Comparison of the BER in a less impulsive environment ($\alpha = 1.8$) obtained for the approximation L_{ab} where the estimated parameters are effected by the supervised design using different LS sizes.

compared to the data samples. It is not clear how the estimator can deal with that but the risk of mismatch is certainly present.

Third, for ($\alpha = 1.1$), the noise statistics in the LS window looks to a certain extent similar to that in the data sequence and consequently, the mismatch risk should decrease.

BER simulation results and analysis: for our simulation results, a SαS noise is considered of parameter $\alpha = 1.1$, $\alpha = 1.4$ and $\alpha = 1.8$. For each channel state, we used different LS sizes (816, 1008, 1200, 3000, 20000) to optimize the parameters of the L_{ab} approximation. The long training sequence (20000) allows us to assess the best performance of the supervised estimation, and as the LS is shortened it allows us to evaluate the loss due to estimation errors.

The obtained BER curves are presented as a function of γ,

- For low impulsive environment $\alpha = 1.8$ in Figure 4.25.
- For high impulsive environment $\alpha = 1.4$ in Figure 4.26.
- For extremely high impulsive environment $\alpha = 1.1$ in Figure 4.27.

We can confirm the previous analysis and the counter-intuitive results. As obviously, for $\alpha = 1.8$ and LS=(816, 1008 and 1200) the performance is significantly degraded in terms of BER as it is highly influenced by the mismatch. As α decreases the probability of mismatch decreases as can be seen via Figure 4.25 to Figure 4.27, and the performance shows a reduced loss when the LS length is shortened.

As a conclusion, in order to be adaptive enough, the length of the training sequence should be defined in low impulsiveness. The question, however, remains

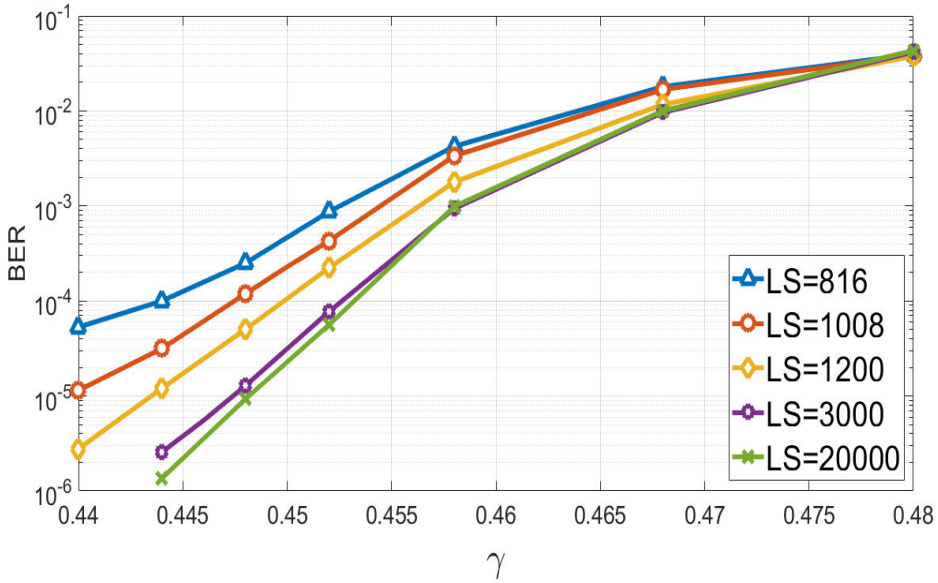


Fig. 4.26. Comparison of the BER in a moderately impulsive environment ($\alpha = 1.4$) obtained for the approximation L_{ab} where the estimated parameters are effected by the supervised design using different LS sizes.

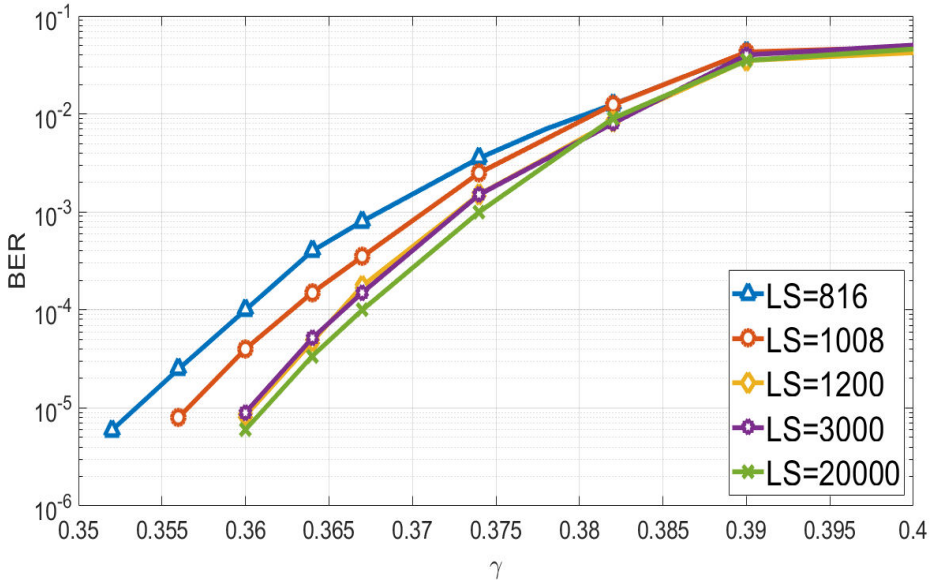


Fig. 4.27. Comparison of the BER in a highly impulsive environment ($\alpha = 1.1$) obtained for the approximation L_{ab} where the estimated parameters are effected by the supervised design using different LS sizes.

complex and has also to account for the packet length. Quantifying the mismatch risk in a simple way could be a solution to adjust the TS length.

But up to now we mainly rely on BER or FER to assess the performance of the system. Because we encompass a large number of noise models, analytical tractability is complex to implement. We will in the next section introduce a metric, easier to handle than BER, to evaluate the performance of a LLR approximation function. This could be used in future work to design receivers and assess their performance in a more efficient way.

4.4 Indirect performance measurement of the LLR approximations

In literature, different methods can be applied to evaluate the quality of different approximations $(L_{clip}, L_{ab}, ...)$ in a given channel. The most trivial method is the direct exhaustive search using Monte Carlo simulation which can be done for various choices of LLR approximations for a given code. Then, the LLR approximation which gives rise to the best BER curve is chosen. However, this approach of finding the best LLR approximation becomes rapidly too complex and long. In the previous section and for estimation purposes, we used an indirect way of measuring the LLR fit with the capacity criterion (3.13). A *capacity loss* could be introduced to evaluate a LLR approximation.

Another way to measure the accuracy is the minimum mean square error (MMSE) between the true LLR and its approximation L_θ. In our study we consider three different approximations, however, any approximation can fit the proposed framework. The three different approximations are: the clipping demapper $L_{\text{clip}}(y) = \text{sgn}(y)\ \min\big(a|y|, \sqrt{ab}\big)$; the approximated LLR $L_{ab}(y) = \text{sgn}(y)\ \min\big(a|y|, b/|y|\big)$; our previously proposed solution $L_{abc}(y) = \text{sgn}(y)\ \min\big(a|y|, b/|y|, c\big)$.

Figure 4.28 compares these different approximations and the true LLR obtained numerically for an α-stable noise with $\alpha = 1.4$ and a scale factor $\gamma = 0.4$. If the parameters are selected correctly, the simple functions L_{clip}, L_{ab} and L_{abc} are rather close to the LLR, at least in the linear region. In superimposition, the figure also displays the density of the output Y knowing the input X; clearly, all regions of the various approximations are likely to be used.

The LLR approximation should be directly comparable to each other in order to choose the best compromise between simplicity and performance. We propose to use the following criterion to rank the approximations

$$\text{MSE} = \int_{-\infty}^{\infty} \left[\Lambda(y) - L_\theta(y)\right]^2 p(y)\, dy = \mathbb{E}_y \left[\Lambda(y) - L_\theta(y)\right]^2 . \tag{4.5}$$

This metric is justified by the need for an approximation to get close to the true likelihood of $\Lambda(y)$. However, all regions must be weighted differently because this approximation must be better for the most likely y values, hence the $p(y)$ factor.

More formally, it is possible to find two constants K and K' to frame the integral

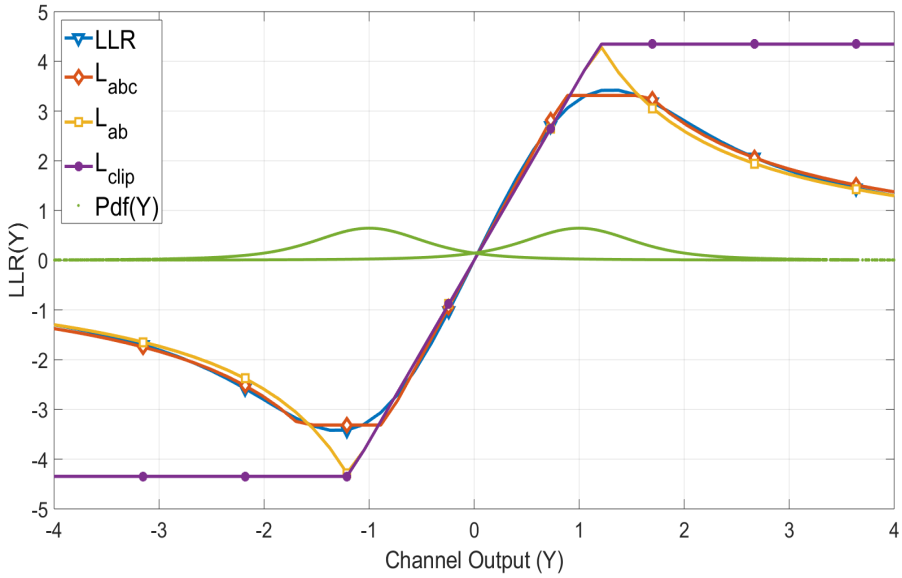


Fig. 4.28. Comparison of the LLR and different approximations in the case α-stable noise with parameters $\alpha = 1.4$ and $\gamma = 0.4$. For L_{ab}, the parameters used are $(a = 3.64; b = 5.19)$ and for L_{abc} $(a = 3.86; b = 5.5; c = 3.31)$

on x of (3.15)

$$K\mathbb{E}_y\big[\Lambda(y) - L_\theta(y)\big]^2 \leq \int \log \frac{q(x\,|\,y)}{p(x\,|\,y)} p(x\,|\,y)\,dxdy \leq K'\mathbb{E}_y\big[\Lambda(y) - L_\theta(y)\big]^2, \tag{4.6}$$

which shows a strong link between the Kullback-Leibler distance and the MSE criterion,

$$K\ \mathrm{MSE} \leq D\big(q(x\mid y)\|p(x\mid y)\big) \leq K'\ \mathrm{MSE}. \tag{4.7}$$

Thus, comparing receivers according to the MSE criterion is a first coherent approach and makes it possible to quickly select the best proposals.

Performance investigation and discussion: To study the performance of our approach, we use regular LDPC codes $(3, 6)$ with a length of 20000 bits. The results will be presented according to the dispersion parameter γ.

First we calculate the distance (4.5) for different approximations according to the dispersion value γ. We see in Figure 4.29, for $\alpha = 1.8$, the significant distance reduction from clipping L_{clip} to L_{ab} then to L_{abc}. The results are similar in a more impulsive environment with $\alpha = 1.4$ as shown in the Figure 4.30. However, the difference between clipping and other approximations is more significant.

To confirm that the chosen metric is relevant, we then study the performance of the three approximations in terms of BER. Figure 4.31 and Figure 4.32 present the results, respectively for $\alpha = 1.8$ and $\alpha = 1.4$.

As predicted by MSE calculations, clipping is significantly less efficient than the other two LLR approximations and the gap is much more significant when α decreases. The L_{abc} function also offers a gain over $L_{a,b}$ as predicted by the MSE

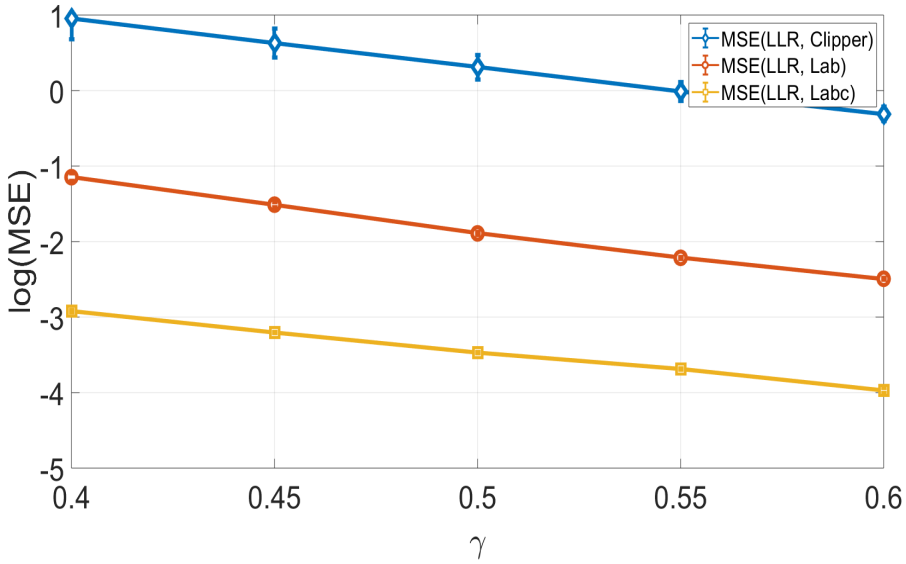


Fig. 4.29. MSE between L_{clip}, L_{ab}, L_{abc} and LLR depending on the dispersion γ in a moderately impulsive environment ($\alpha = 1.8$).

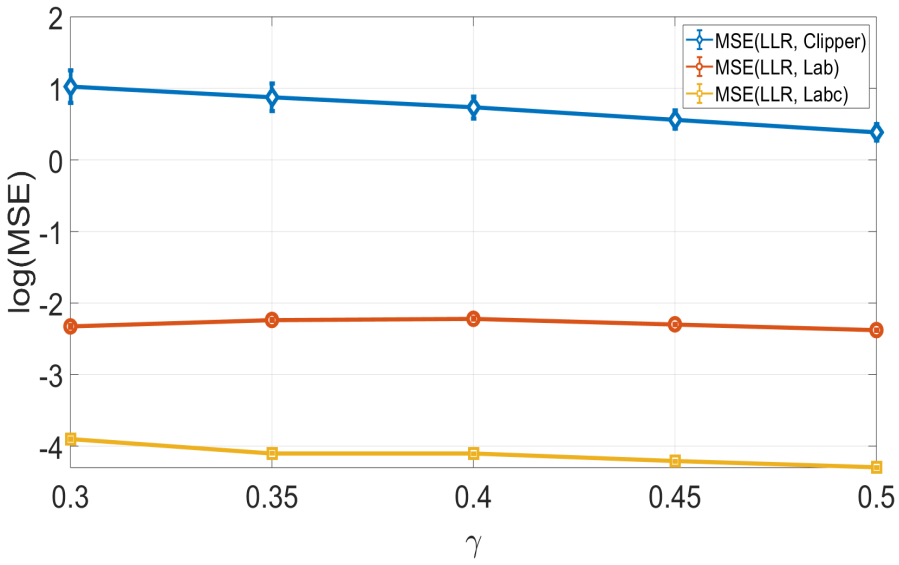


Fig. 4.30. MSE between L_{clip}, L_{ab}, L_{abc} and LLR depending on the dispersion γ in a highly impulsive environment ($\alpha = 1.4$).

criterion. Classically in model selection, weighted by the number of parameters could be introduced to find a compromise between complexity and accuracy.

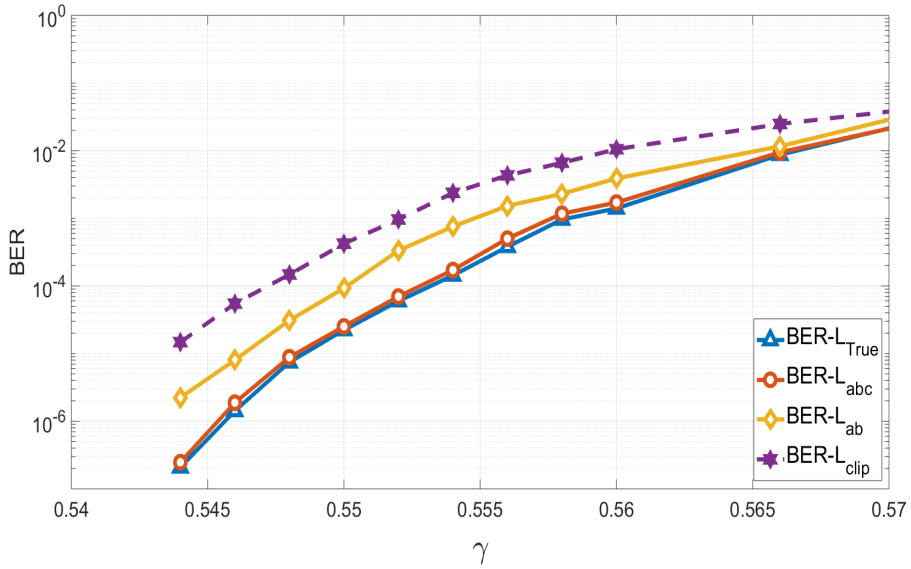


Fig. 4.31. Comparison of the BER for a moderately impulsive environment ($\alpha = 1.8$) obtained for the approximations L_{clip}, L_{ab}, L_{abc} and LLR.

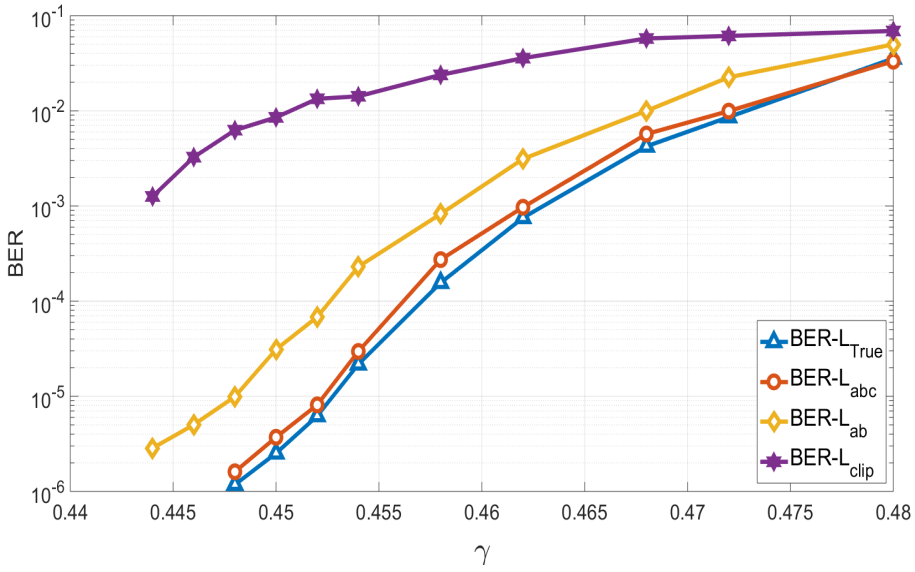


Fig. 4.32. Comparison of the BER for a highly impulsive environment ($\alpha = 1.4$) obtained for the approximations L_{clip}, L_{ab}, L_{abc} and LLR.

4.5 conclusion

The presence of interference or impulsive noise makes it difficult to calculate the LLRs required for decoding. We proposed in this chapter a flexible receiver design based on a LLR approximation function f_θ in a parametric family. The parameter θ is estimated through the maximization of the mutual information between the channel input and the demapping of the channel output. A supervised estimation is proposed where first we test our framework using long sequence size. Our results show that the supervised receiver design is efficient and robust in a large variety of noises and allows to reach performance close to the optimal.

In addition, we proposed a new LLR demapper based on a parameterized approximation function with three optimization parameters. Numerical simulations show that the performance achieved match the one obtained with the true LLR.

Furthermore, to evaluate the performance of an approximation, we show that it is not necessary to use intensive Monte Carlo simulations but that a simple distance calculation gives a precise idea of its performance. This makes it possible to work on the quality of the approximations and their adaptability to various contexts. Consequently, we evaluate the quality of the approximation by a mean squared error and ascertain that this criterion is sufficient to identify the approximations that will be efficient.

All the obtained performances are investigated in a first approach by the use of a training sequence of the size of the codeword, in other words, assuming a very accurate estimation of the channel state. This is not realistic in real settings, for such, we evaluated the impact of shortening the training sequence. The task is especially difficult with impulsive noise that makes the estimation step more complex and requires longer training sequence to observe enough rare events. In such a case, a risk of mismatch arises: the risk that what we learned from the learning sequence may not match the condition of the message, leading to severe performance degradation. Indeed, the learning sequence has to be long enough to guarantee the presence of the expected rare events. If this is not ensured, the estimation of θ won't lead to a good LLR estimation.

A major drawback of the supervised approach relies on the necessity of the learning sequence. As a consequence, it induces an increase in signaling and a decrease in the useful rate. Moreover, we show the LS has to be quite long to avoid the mismatch effect. For this reason, we propose in the following chapter an unsupervised estimation of θ, which focuses directly on the output of the channel, without any prior on the input.

Chapter 5

An unsupervised LLR estimation with unknown noise distribution

In this chapter, we propose an unsupervised estimation of θ^*, avoiding the need of a training sequence. The performance of our estimation method is first analyzed with long codewords. It is shown to be efficient in a large variety of noises and the receiver exhibits a near-optimal performance. Next, this estimation is performed with short codewords, rending the estimation very error-sensitive. We analyze what conditions lead to the estimation failure, derive an analytical tool to assess the probability of failure and then propose to jointly use two mechanisms that prevent the aforementioned failures. Our estimation is shown to be efficient and the receiver exhibits a near-optimal performance under various noisy environment types, ranging from very impulsive one to Gaussian.

5.1 Online parameter estimation and unsupervised optimization

To ease the reading we recall that our optimization problem writes as:

$$\theta^* = \arg\min_{\theta} \widehat{H}(X|Y) = \arg\min_{\theta} \mathbb{E}\left[\log_2\left(1 + e^{-XL_\theta(Y)}\right)\right] \tag{5.1}$$

and can be solved using data samples as

$$\theta^* = \arg\min_{\theta} \frac{1}{K} \sum_{k=1}^{K} \log_2\left(1 + e^{-x_k L_\theta(y_k)}\right), \tag{5.2}$$

where K is the number of samples, x_k and y_k represent the input and the output of the channel, respectively. For more details see Section 3.5.1.

5.2 Unsupervised optimization

To solve (5.2), one needs a received sequence y_k as well as the corresponding transmitted one x_k. This is usually obtained thanks to the supervised approach where

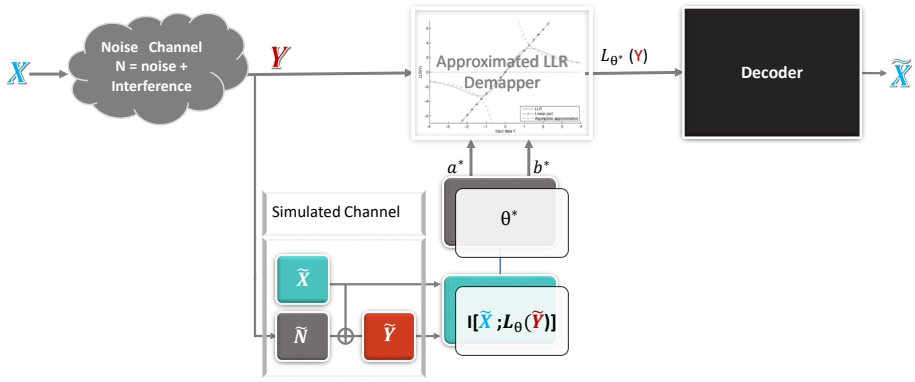


Fig. 5.1. Unsupervised LLR demapper scheme.

a training sequence is used as discusses in the previous chapter. However, this induces an increase in 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 function 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 and the mismatch discussed in Section 4.3 becomes irrelevant.

Since one needs the sent sequence X as well as the corresponding channel output Y, we propose [MSG+19d, MSG+18] to extract a noise sequence $\widetilde{N}$ directly from the received channel output Y and to simulate at the receiver side the transmission of a known sequence $\widetilde{X}$. The corresponding channel output is build as $\widetilde{Y} = \widetilde{X} + \widetilde{N}$, as depicted in Figure 5.1. To do so, we propose to use a sign-detector yielding

$$\widetilde{N} = Y - \text{sign}(Y). \tag{5.3}$$

The simulated channel input is an i.i.d. BPSK random variable, independent of $\widetilde{N}$. Due to this independence, the input sequence can be chosen as a sequence of $+1$'s, so that the new channel output is $\widetilde{Y} = +1 + \widetilde{N}$. The optimization parameter θ can be estimated based on (5.2) but with the newly generated input and output as

$$\begin{aligned} \theta^* &= \arg\min_\theta \widehat{H}(\widetilde{X}|\widetilde{Y}) \\ &= \arg\min_\theta \frac{1}{K}\sum_{k=1}^{K} \log_2\left(1 + e^{-\widetilde{x}_k L_\theta(\widetilde{y}_k)}\right) \\ &= \arg\min_\theta \frac{1}{K}\sum_{k=1}^{K} \log_2\left(1 + e^{-L_\theta(\widetilde{y}_k)}\right), \end{aligned} \tag{5.4}$$

where the last transition comes from the fact that we use $\widetilde{X} = +1$. Once θ^* is obtained, the LLR is approximated by $L_{\theta^*}(Y)$, where Y is the true received sequence over the MBISO channel.

In order to investigate the performance of the unsupervised approach, we are going to compare it in the following to the best achieved performance of the supervised approach (LS=20000). In the next section, we propose to apply our blind

LLR approximation optimization to LDPC coding where the noise exhibits either impulsive or a Gaussian nature.

For our simulations, once our demapper is tuned with the estimated value θ^*, it is used as a front-end to the 20000 bits long regular (3,6) LDPC decoder when long block length regime analysis is considered and the 408 bits long regular (3,6) LDPC decoder when short block length regime analysis is considered using the BP algorithm.

5.3 Unsupervised learning with long block length regime

5.3.1 Parameter estimation

Estimation over impulsive SαS additive noise

In a first step, we compare the obtained θ^* under blind optimization with the one obtained under a supervised approach. In this section, we only present the obtained results for a highly impulsive noise when $\alpha = 1.4$, but similar observations and conclusions can be made for other choices.

Figure 5.2, respectively 5.3, compares the evolution of the mean and variance of the estimated parameter a, respectively b, as a function of the dispersion γ. For each noise dispersion, we ran 5000 experiments.

We can see from Figure 5.2 that the gap between the obtained values for parameter a under supervised and unsupervised optimization keeps very small. But unfortunately, as shown in Figure 5.3, 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 more difficult.

In Figure 4.7, we represent a 3D plot of the function $\widehat{H}_K$ (5.2). We show that $\widehat{H}_K$ 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. We illustrate the link between the function $\widehat{H}_K$ and the obtained BER. The contour plot delineates different BER values, ranging from 10^{-5} to 10^{-1}.

In Figure 5.4, we project the two white crosses correspond to the mean value of the optimization parameters a and b obtained under supervised and unsupervised optimization, as provided on Figure 5.2 and Figure 5.3, respectively. Furthermore, the white contour delineates the set of a and b values yielding the smallest values of $\widehat{H}_K$ 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 small variance of the estimated θ^* under both cases ensures that most of the estimated values will fall in the region yielding the smallest BER. Thus, we can expect that the error on b will have a limited impact in terms of BER performance. Through intensive simulations, we noticed that the connection between $\widehat{H}_K$ and the BER is always assessed, irrespective of the noise model and noise parameters value.

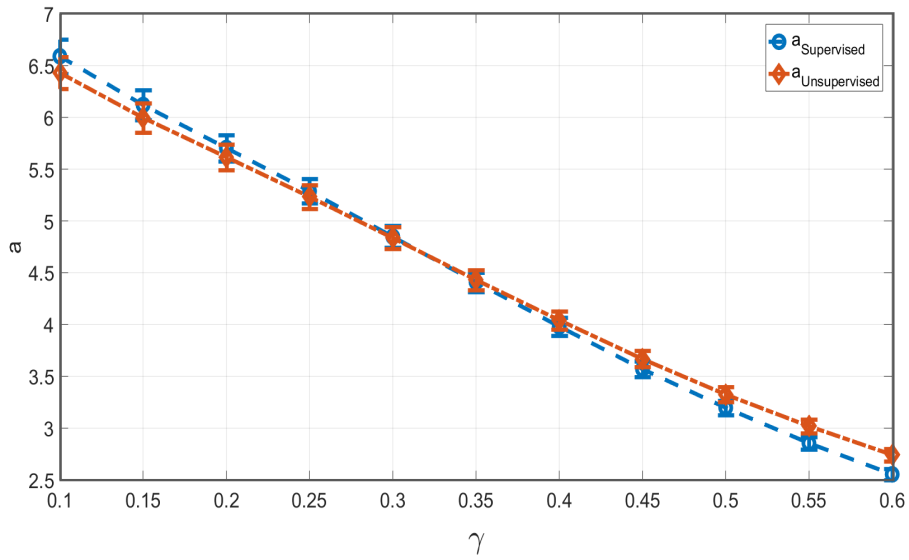


Fig. 5.2. Comparison of the mean and standard deviation evolution for parameter a as a function of the dispersion γ of a SαS noise with $\alpha = 1.4$ for the supervised and unsupervised optimization.

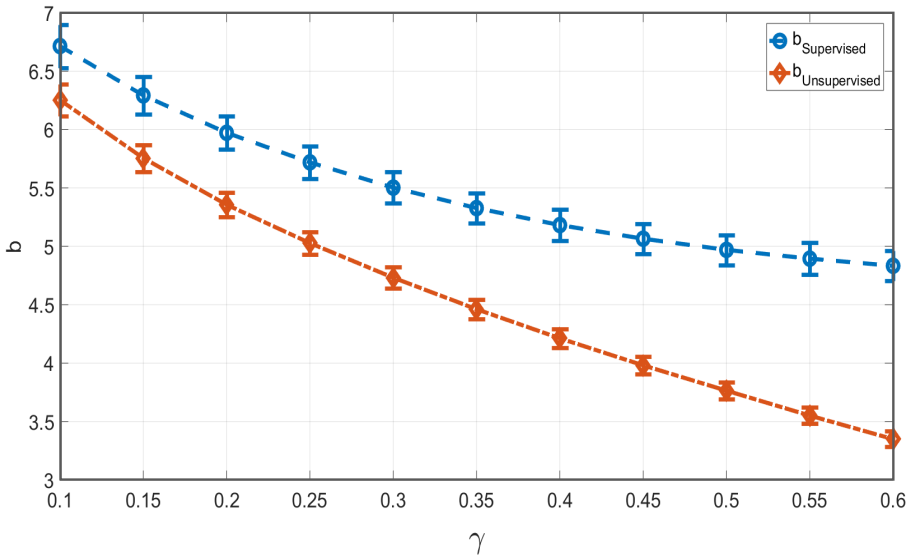


Fig. 5.3. Comparison of the mean and standard deviation evolution for the parameter b as a function of the dispersion γ of a SαS noise with $\alpha = 1.4$ for the supervised and unsupervised optimization.

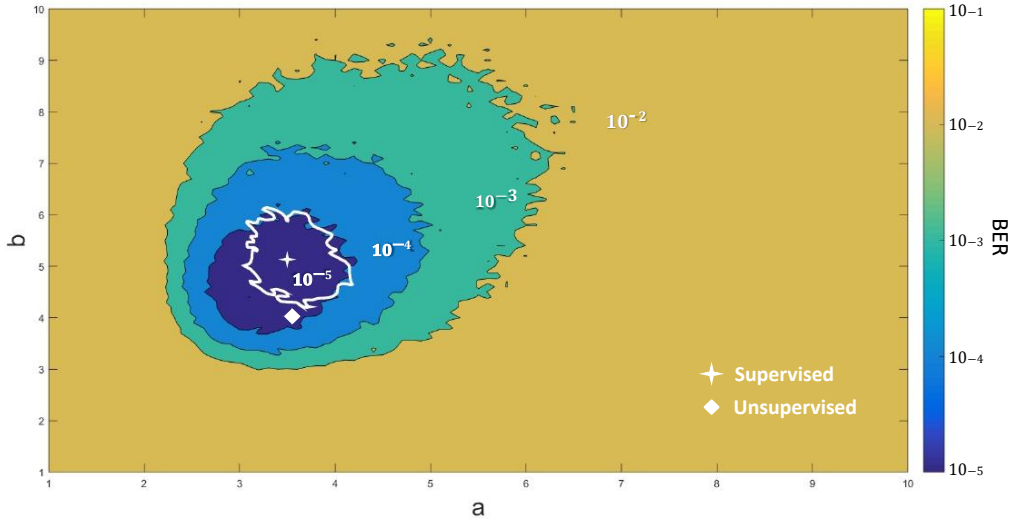


Fig. 5.4. BER comparison as a function of a and b parameters with $\gamma = 0.45$ and $\alpha = 1.4$ under the supervised approximation.

			μ_a	σ_a	μ_b	σ_b
$\alpha = 1.8$	$\gamma = 0.53$	Unsupervised	3.43	0.06	5.73	0.15
		$\text{Sup}_{\text{LS=20000}}$	3.25	0.05	7.59	0.28
		$\text{Sup}_{\text{LS=1200}}$	3.27	0.24	8.50	14.48
		$\text{Sup}_{\text{LS=900}}$	3.27	0.27	11.72	46.15
	$\gamma = 0.55$	Unsupervised	3.23	0.05	5.61	0.14
		$\text{Sup}_{\text{LS=20000}}$	3.05	0.05	7.62	0.28
		$\text{Sup}_{\text{LS=1200}}$	3.07	0.22	7.97	1.54
		$\text{Sup}_{\text{LS=900}}$	3.07	0.26	10.73	30.41

Table 5.1 – Parameter Estimation. Comparison of the mean and standard deviation evolution for the parameters (a, b) as a function of the dispersion γ of a SαS noise with $\alpha = 1.8$ for the supervised with different learning sequences and unsupervised optimization.

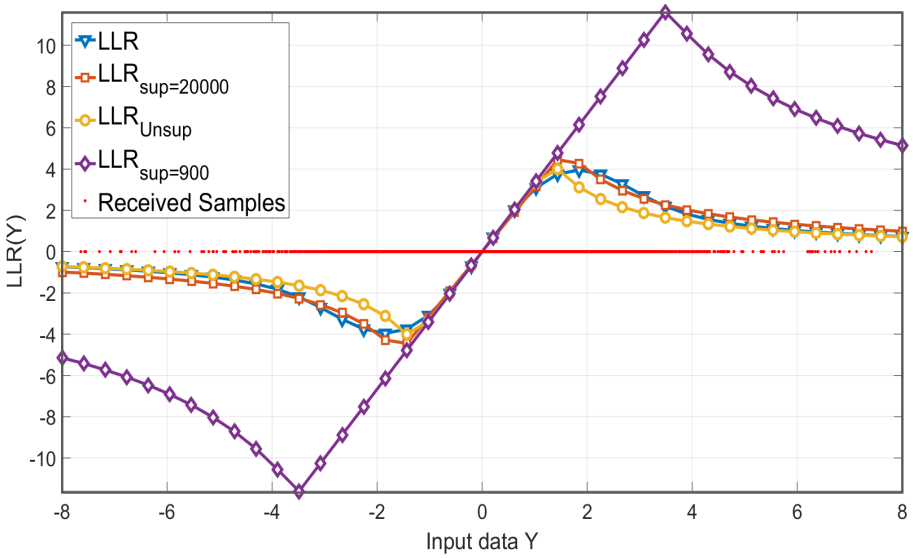


Fig. 5.5. Comparison of the LLR shapes under the effect of the estimated a and b parameters with $\gamma = 0.55$ and $\alpha = 1.8$, in the supervised LLR approximation different learning sequences, the unsupervised LLR approximation and with LLR obtained by numerical integration.

To complete the study we present in Table 5.1 the estimated θ^* obtained under both approaches unsupervised and supervised with different LS sizes. Table 5.1 highlights the performed gain when the whole data set under the unsupervised approach is used and compared to shortening the LS used under the supervised approach. 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. 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 shows a higher error. On another hand, the estimated b parameter slightly changes under the unsupervised approach while shortening the learning sequence under the supervised approach will induce large error values, which indicate the significant impact of different realizations being more likely to fall outside the optimal BER region.

Nevertheless, the induced large error values will effect the shape of the demapper. To show that, we compare in Figure 5.5 the LLR shapes obtained under supervised optimization with learning sequence (900, 20000) and unsupervised optimization to the true LLR obtained via numerical integration for a $S\alpha S$ noise of parameters $\alpha = 1.8$ and $\gamma = 0.55$. Figure 5.5 compares the LLR shapes under the worst case estimation by taking the sum of the mean and standard deviation of both parameters a and b. This comparison shows clearly the fitting between the LLR shapes, except for the supervised with short training sequence, which induces a significant number of received samples treated incorrectly. Consequently, we can expect that the approximated demapper under unsupervised optimization performs close to the true LLR and even better than supervised when the learning sequence is shortened.

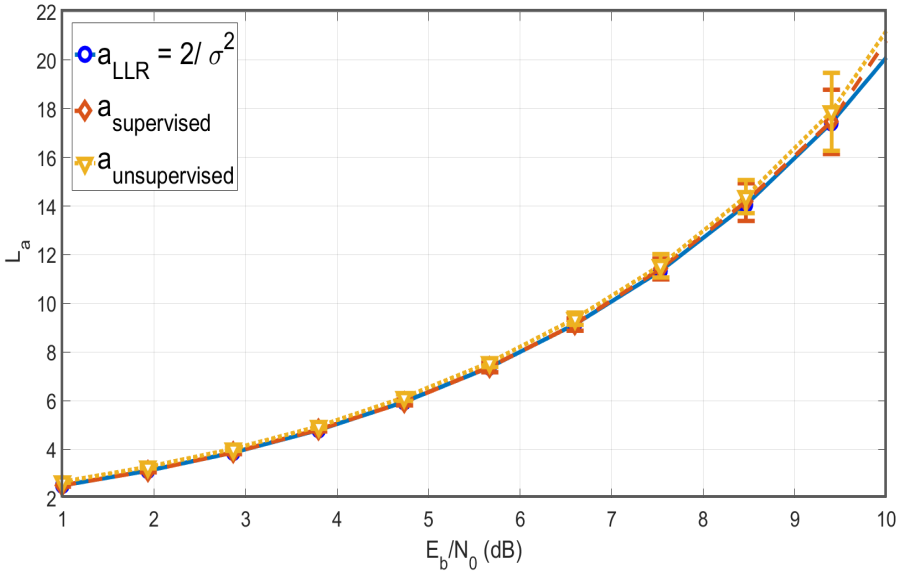


Fig. 5.6. Comparison of the mean and standard deviation evolution for parameter a as a function of E_b/N_0, under the Gaussian noise, for the supervised and unsupervised optimization to the optimal receiver.

Estimation over Gaussian noise

Recall that with our proposed demapper, the decreasing part starts at $\pm\sqrt{b/a}$. If this value becomes large enough, the power-law part will not impact the receiver. For Gaussian noise, large received samples are very rare, so that all samples fall into the linear part of the demapper. Numerically, when simulating AWGN channel, we obtained values bigger than 49 for b see Section 4.2.1 which are significantly larger than the one obtained for a, showing that the non-linear part of the demapper will rarely be used and will have a very small impact on the performance. The linear part is given by the optimal slope as $a = 2/\sigma^2$.

Figure 5.6 compares the evolution of the mean and variance of the optimized parameter a, under the supervised (LS=20000) and unsupervised approaches to the optimal slope. For each noise variance, we ran 5000 experiments.

This comparison shows clearly the convergence between the slopes, proving thus the strength of both the supervised and unsupervised optimization of the demapping function L_{ab} over the AWGN channel. The error bars indicate the small impact of different realizations. In the previous chapter, we demonstrated the robust performance of the supervised approach in terms of BER and thus we can expect the same behavior for the unsupervised approach.

5.3.2 Performance evaluation

BER performance under SαS additive noise

Figure 5.7 and Figure 5.8 present the obtained BER for $\alpha = 1.4$ and $\alpha = 1.8$, respectively, as a function of the dispersion parameter γ. which is used as a measurement

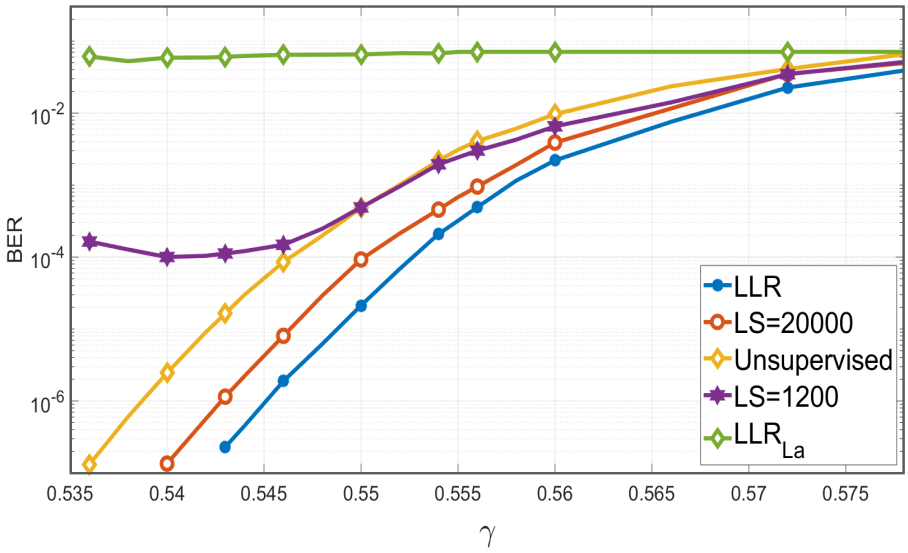


Fig. 5.7. Evolution comparison of the BER as a function of the dispersion γ of a SαS noise in poorly impulsive environment with $\alpha = 1.8$, between the supervised with different learning sequence sizes, unsupervised, Gaussian designed LLR approximations and the LLR obtained by numerical integration.

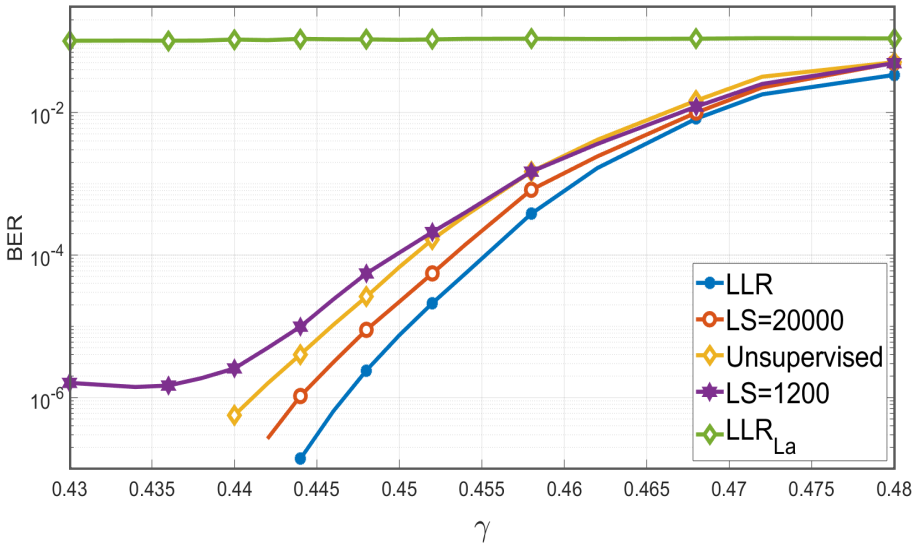


Fig. 5.8. Evolution comparison of the BER as a function of the dispersion γ of a SαS noise in highly impulsive environment with $\alpha = 1.4$, between the supervised with different learning sequence sizes, unsupervised, Gaussian designed LLR approximations and the LLR obtained by numerical integration.

of the strength of the α-stable noise. In both cases, we compare the BER obtained via the demapping function L_{ab}, 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 samples to optimize θ in the supervised case; the 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. The unsupervised approach does not perform as well as the supervised one with a long training sequence but the gap is restrained. However, in comparison to a linear receiver gain is huge. Moreover, 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 unsupervised framework, we investigate in the following its performance when the channel exhibits either an impulsive nature modeled by: Middleton Class A [SM77] noise, ε-contaminated [AALM17b] noise or simply a Gaussian noise.

BER over Gaussian and other impulsive noises

The different configurations

- Gaussian noise in Figure 5.9
- high impulsive Middleton environment with $A = 0.01$ and $\Gamma = 0.01$ in Figure 5.10
- moderate impulsive Middleton environment with $A = 0.1$ and $\Gamma = 0.1$ in Figure 5.11
- low impulsive $\epsilon-$contaminated environment with $\varepsilon = 0.01$ and $K = 10$ in Figure 5.12.
- high impulsive $\epsilon-$contaminated environment with $\varepsilon = 0.1$ and $K = 10$ in Figure 5.13.

For each scenario, we compare the true LLR, obtained via numerical integration to the LLR approximations under supervised with 20000 long learning sequence, unsupervised parameter estimation and Gaussian designed demapper L_a.

In Figure 5.9 (Gaussian noise) all curves are almost superposed, with a small performance loss under the unsupervised optimization. Our proposed approach L_θ does not degrade the decoding performance in a purely Gaussian case under the unsupervised estimation.

In Figure 5.10 to Figure 5.13, the high robustness and adaptability of our receiver 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 in all scenarios.

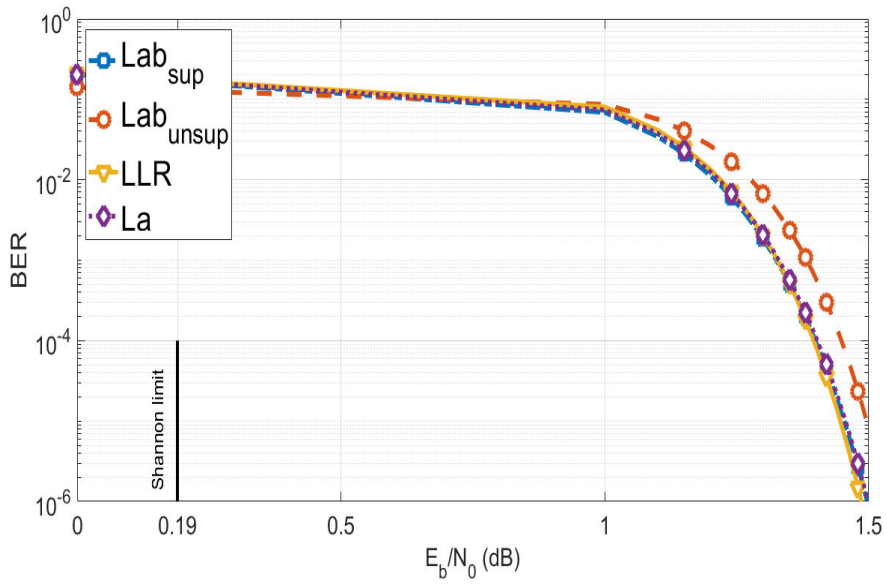


Fig. 5.9. BER comparison as a function of E_b/N_0 between the supervised, unsupervised LLR approximations and the optimal LLR in Additive Gaussian noise channel.

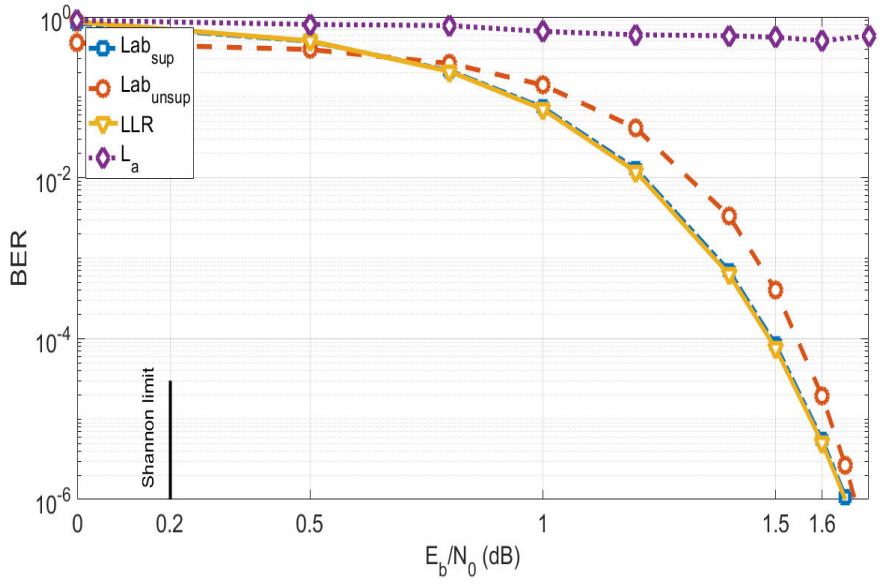


Fig. 5.10. BER comparison as a function of E_b/N_0 between the supervised, unsupervised, Gaussian designed LLR approximations and the LLR obtained by numerical integration, in high impulsive Middleton Class A noise with ($A = 0.01$ and $\Gamma = 0.01$).

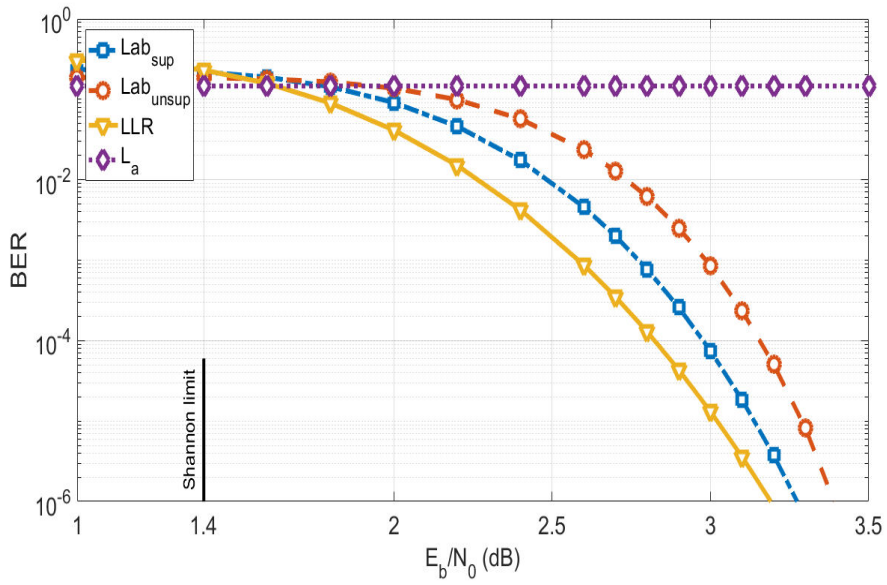


Fig. 5.11. BER comparison as a function of E_b/N_0 between the supervised, unsupervised, Gaussian designed LLR approximations and the LLR obtained by numerical integration, in moderately impulsive Middleton Class A noise with $A = 0.1$ and $\Gamma = 0.1$.

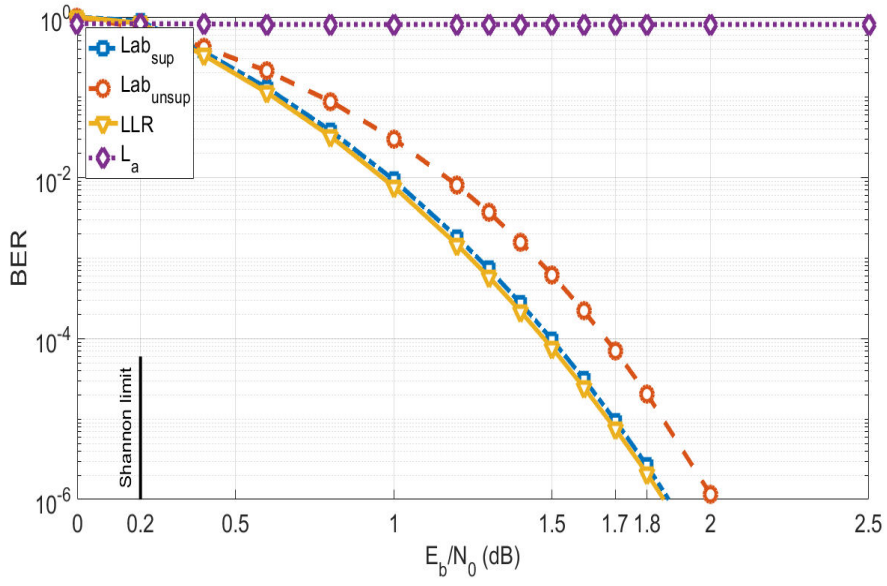


Fig. 5.12. BER comparison as a function of E_b/N_0 between the supervised, unsupervised, Gaussian designed LLR approximations and the LLR obtained by numerical integration, in low impulsive $\epsilon-$contaminated noise with $\varepsilon = 0.01$ and $K = 10$.

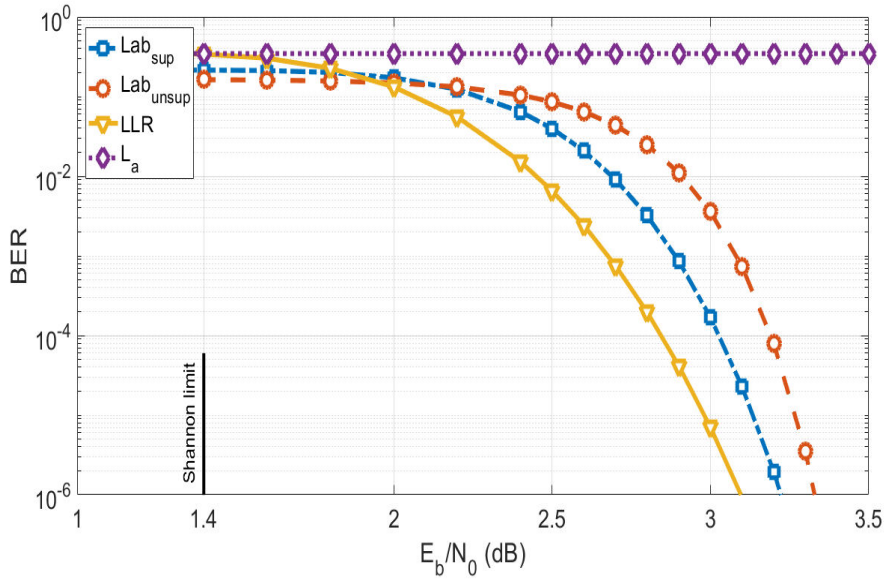


Fig. 5.13. BER comparison as a function of E_b/N_0 between the supervised, unsupervised, Gaussian designed LLR approximations and the LLR obtained by numerical integration, in highly impulsive $\epsilon-$contaminated noise with $\varepsilon = 0.10$ and $K = 10$.

Analysis and discussion: these numerical simulations illustrate the universality of the unsupervised framework which is proposed to benefit from the whole received sequence to avoid the mismatch problem raised by the supervised case, but also to increase the useful data rate. The approximation family has to be wide enough to encompass 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 that does not depend on any noise assumption. Our results show that the receiver design is efficient in a large variety of noises and that the blind estimation allows to reach performance close to the optimal and even better than the supervised approach if the training sequence is not sufficiently long.

In the following section, we are going to investigate the performance of the unsupervised framework influenced by short packets.

5.4 Unsupervised learning with short block length regime

In this section, we consider the case where the transmitted packets contain a limited number of information bits sent over an impulsive channel, in which it simulates for instance the IoT networks [MSG+19b]. The use of a training sequence, in that case, will significantly impact the useful throughput of the communication, and it increases the mismatch risk probability between the training sequence and the payload 4.3. Besides, impulsive noises are used to model rare events and estimation

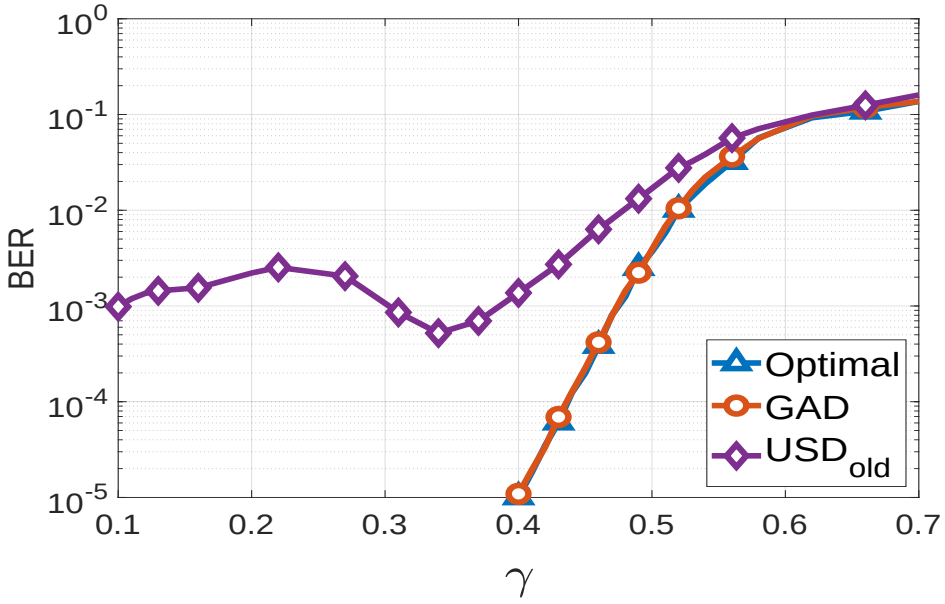


Fig. 5.14. BER Comparison as a function of the dispersion γ of SαS noise in low impulsive environment with $\alpha = 1.8$, between GAD, our previously proposed USD and the optimal decoder where the LLR obtained by numerical integration when $K = 408$.

becomes difficult so that more samples are needed in the training set. The question we raise is: can we use an unsupervised approach to evaluate the LLR approximation parameters that will remain robust with short packets?

The main contributions of this section are the following:

1. first, we investigate the impact of reducing the length of the packet on the bit error rate when the LLR approximation parameters' estimation is unsupervised.

2. we then analyze the reasons of the significant degradation that we observed compared to a longer packet case.

3. we derive an analytical tool to assess the probability of failure.

4. finally, we propose solutions to keep a robust scheme with shorter packets by increasing the diversity in the noise sequence extracted from the received packet and adding a regularization term.

In order to distinguish the unsupervised learning to the LLR approximation, we introduce the genie aided decoder (GAD), which uses the true input x_k to infer the approximation parameter θ^* from our criterion $\widehat{H}_K$.

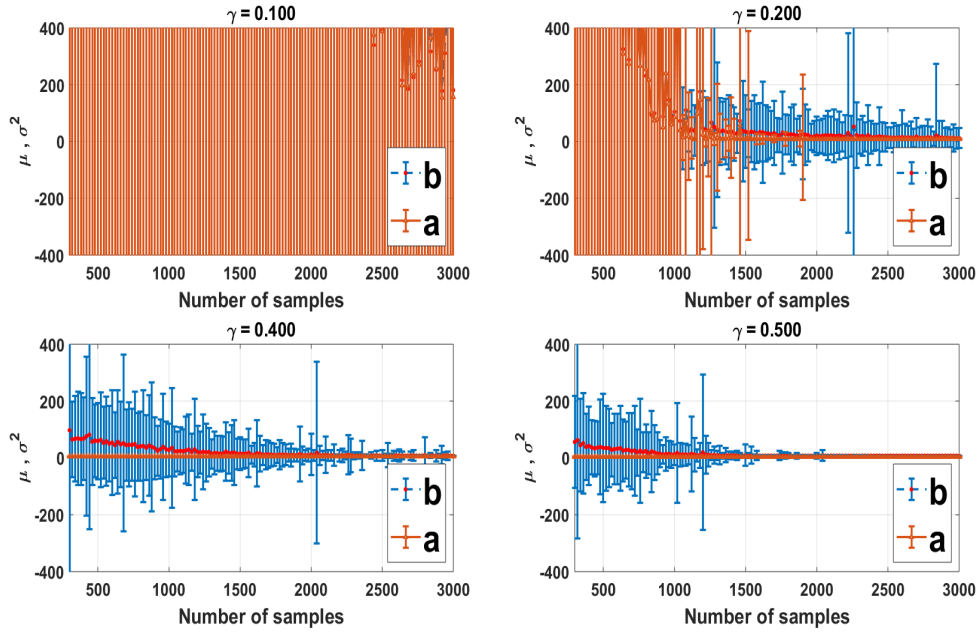


Fig. 5.15. Evolution study of a and b parameters as a function of the number of samples for $S\alpha S$ noises with $\alpha = 1.8$.

5.4.1 Estimation with short sequences

Problem Statement

In the aforementioned results, we performed the optimization of θ in both a supervised and unsupervised manner under a large number of training samples $K = 20000$. The resulting BER curves under both the supervised and unsupervised optimization were close to each other and moreover close the one obtained with numerically computed true LLR, showing thus the efficiency of our proposed LLR approximation under large codewords. However, reducing the available number of samples available to solve (5.2) has a large impact in terms of BER as can be seen in Figure 5.14, where we shorten n to 408 and used the whole received sequence for the unsupervised estimation, meaning $K = n$. One can indeed notice a large degradation of the BER obtained under the unsupervised optimization (USD), compared to the one obtained either with the GAD or with the true LLR (optimal). A bump can even be noticed and the BER is not a monotonic function of the noise dispersion anymore. Nevertheless, this performance loss is not due to the form of the LLR approximation L_θ, since the BER curves obtained under the GAD and with the true LLR still match, but solely to the estimation step yielding wrong values of θ^*.

To investigate the causes of the performance loss, we compare in Figure 5.15 the evolution of the mean and variance of the optimized parameters a and b as a function of the number of samples for an $S\alpha S$ noise of parameter $\alpha = 1.8$ and $\gamma \in \{0.1, 0.2, 0.4, 0.5\}$. We see that the variance of the estimated parameters are extremely large when packets are short, especially for low values of γ where the BER degradation is the strongest in Figure 5.14. In fact, the optimization step sometimes outputs extremely large parameter values. Consequently, the BP algorithm input

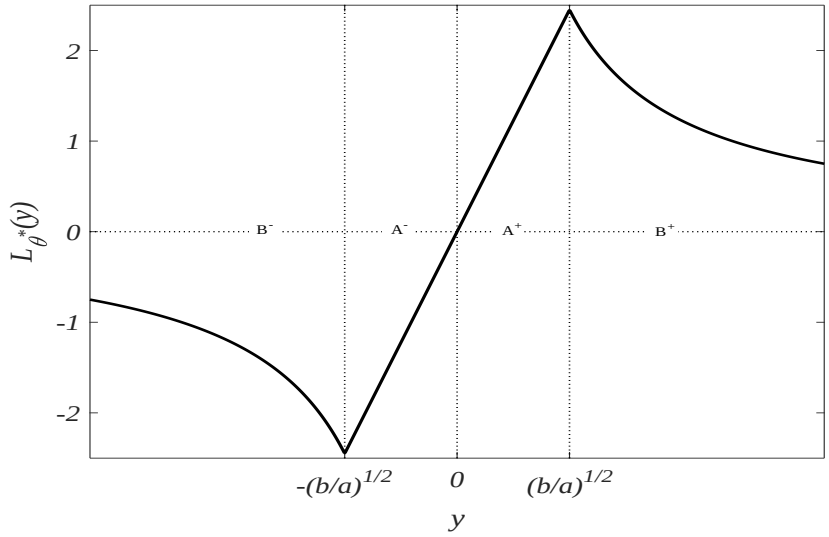


Fig. 5.16. Illustration of the LLR approximation with the four regions influencing the shape of L_θ.

does not match the LLR, leading to poor BER. The same conclusion can be made for other values of α.

In the rest of this section, we are going to (a) analyze in depth the reasons why the estimation fails and (b) propose solutions to improve the robustness of this step. Thereafter we will call *degeneration* the fact that the optimization step outputs a very large value for parameters a or b, resulting in a bad tuning of L_{θ^*} that does not match the real state.

Problem exploration

Let Z denote $Z = XY$. Z is an interesting way to present the following derivation because $z > 0$ means that x and y have the same sign (this bit detection would not be in error) when $z < 0$ means that the noise modifies the sign of the transmitted bit. Since LLRs are odd functions, one can rewrite $x\Lambda(y)$, respectively $xL_\theta(y)$, as $\Lambda(z)$ and $L_\theta(z)$. Further, let f_{ab} denote $f_{ab}(z) = \log_2(1 + e^{-L_{ab}(z)})$. Based on the two optimization parameters a and b and on the L_{ab} approximation, one can define four regions that influence the shape of L_θ as depicted in Figure 5.16:

$$B^- = \left[-\infty, -\sqrt{b/a}\right]; A^- = \left[-\sqrt{b/a}, 0\right]; A^+ = \left[0, \sqrt{b/a}\right]; B^+ = \left[\sqrt{b/a}, +\infty\right]. \tag{5.5}$$

Using these four regions, one can rewrite (3.19) as

$$\widehat{H}(X|Y) \propto \sum_{z_i \in B^- \cup B^+} \log_2(1+e^{-\frac{b}{z_i}}) + \sum_{z_i \in A^- \cup A^+} \log_2(1+e^{-az_i}). \tag{5.6}$$

In each of these regions, only one of the two optimization parameters influences the shape of L_θ: samples falling in B^- and B^+ are only impacting the optimization parameter b, while samples falling in A^- and A^+ are only impacting the optimization

parameter a. However, due to different exponent signs, minimizing $\widehat{H}(X|Y)$ requires reverse optimization tendencies in the regions B^+ and B^-, respectively, A^+ and A^-. Indeed, samples z_i in B^- tend to decrease the value of b in order to minimize $f_{ab}(z_i)$ since $b > 0$ and $z_i < 0$, whereas samples in B^+ tend to increase the value of b, since $z_i > 0$. The optimal value of b results thus from a compromise between the two sets B^+ and B^-. The same analysis holds for the parameter a and one can conclude that if one of these regions is empty, the minimization problem will converge towards an approximated LLR far away from the true one.

Having empty A^+ or empty B^+ region is very unlikely since it would require all the training samples to be decoded in error, or in other words to only have large noise samples of the opposite sign than the input sequence. Hence only two cases remain that trigger the optimization failure:

1. empty B^- region and populated B^+ region: the samples in B^+ will make b parameter goes to infinity in order to minimize f_{ab}. As a consequence, the threshold $\sqrt{|b/a|}$ will increase and tend to infinity and thus the B^- region will still remain empty. This case, however, is not necessarily an optimization failure since under Gaussian noise, the optimal receiver is a linear one with $b^* \to \infty$. Under the GAD approach, such a case can also occur when no negative z is received with a large amplitude.

2. empty A^- region and populated A^+ region: the terms in A^+ will make a parameter goes to infinity in order to minimize f_{ab}. As a consequence, the threshold $\sqrt{|b/a|}$ will decrease and tend to zero and thus the A^- region will remain empty. Under weak noise, the sign detector only extracts small noise samples and thus the probability of having zero samples in A^- becomes non negligible, leading thus to a very large value of a. The fact that it happens more when the dispersion of the noise decreases explains the bump shape observed in Figure 5.14.

To conclude, the optimization step will fail when at least one of the aforementioned regions is empty, which is more likely with short training sequences since the extracted noise does not represent its whole statistics. Indeed, impulsive noises are characterized by large, but rare, events and their probability to appear several times during the training sequence decreases when its length decreases. This can also happen in the unsupervised case, especially when the noise strength is low or/and when the impulsiveness is low. In fact, the sign detector will tend to generate noise samples with values less than one.

Quantifying the risk of bad estimation

We will quantify the risk of optimization failure by assessing the probability of having zero samples falling in the regions A^- and B^-. However, these two regions being delimited by a non-fixed boundary, we will relax the studied problem and consider the case $A^- \cup B^- = \varnothing$, i.e. no negative samples z are obtained, which will be called *degeneration*.

The degeneration risk η is the probability that the sequence $Z_1, \ldots, Z_K$ has only positive elements:

$$\eta_k = P\left(Z > 0,\ \forall k \in \{1, ..., K\}\right). \tag{5.7}$$

Let Q_α denote the right tail of the probability function for a $S\alpha S$ random variable as $Q_\alpha(x) = \int_x^{+\infty} f_\alpha(u)du$, where f_α is the probability density function of the $S\alpha S$ distribution with dispersion parameter $\gamma = 1$.

Assuming K identically distributed and independent noise realizations, we have

$$\eta = \prod_{k=1}^{K}(1 - P\left(Z \leq 0\right)) = (1 - P\left(Z \leq 0\right))^K. \tag{5.8}$$

In the next Proposition, we will provide the degeneration risk under both the GAD and the unsupervised approach.

Proposition 5. *Under the Genie-aided decoder, the degeneration risk is given as*

$$\eta_{K,GAD} = \left[1 - Q_\alpha\left(\frac{1}{\gamma}\right)\right]^K = (\eta_{1,GAD})^K, \tag{5.9}$$

whereas under the unsupervised optimization, it is given as

$$\eta_{K,USD} = \left[1 - \frac{1}{2}Q_\alpha\left(\frac{3}{\gamma}\right) - \frac{1}{2}Q_\alpha\left(\frac{1}{\gamma}\right)\right]^K = (\eta_{1,USD})^K. \tag{5.10}$$

Proof. Let us consider the two cases separately.

Under the GAD, $X \in \{-1, 1\}$ and $Y = X + N$, so that $Z = XY$ is negative if $X = +1$ and $N < -1$ or if $X = -1$ and $N > 1$. Thus,

$$\begin{aligned} P\left(Z < 0\right) &= \frac{1}{2}P\left(N < -1 | x = +1\right) + \frac{1}{2}P\left(N > 1 | x = -1\right) \\ &= \frac{1}{2}\left(\int_{-\infty}^{-1} f_\alpha(u)du + \int_{1}^{+\infty} f_\alpha(u)du\right) \\ &= Q_\alpha\left(\frac{1}{\gamma}\right), \end{aligned} \tag{5.11}$$

and $\eta_{K,GAD} = \left[1 - Q_\alpha\left(\frac{1}{\gamma}\right)\right]^K = (\eta_{1,GAD})^K$.

It should be emphasized that, due to the full access to a genie-aided provided bits X, the estimation failure will be triggered only when no negative sample exists which translates in this case as an error-free received codeword. Thus, the BER performance will not be influenced in spite of a badly tuned LLR approximation.

Under unsupervised optimization, the training sequence only consists of $\widetilde{X} = +1$s and $\widetilde{Y} = +1 + \widetilde{N}$. As a consequence, the probability to have a negative $\widetilde{Z} = \widetilde{X}\widetilde{Y}$ is

$$P\left(\widetilde{Z} < 0\right) = P\left(\widetilde{N} < -1\right) = P\left(Y - \text{sign}\left(Y\right) < -1\right).$$

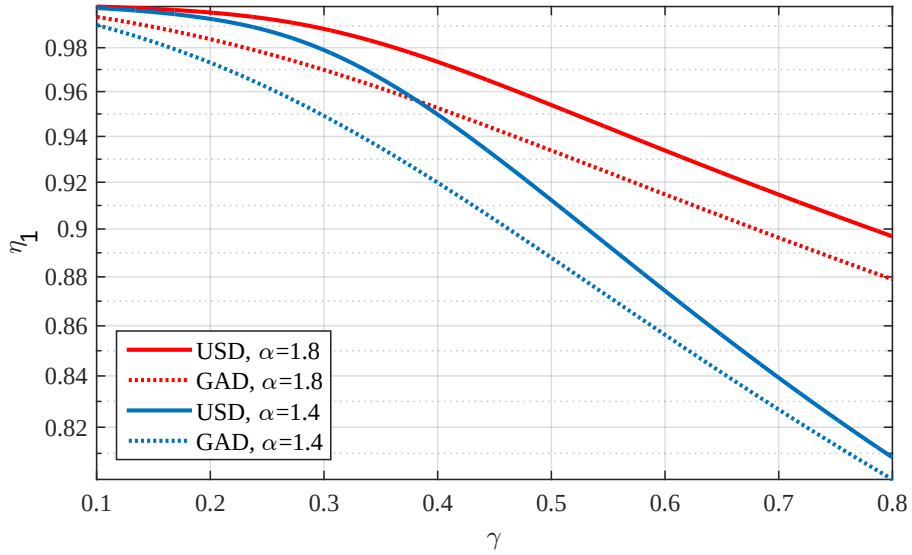


Fig. 5.17. Degeneration risk for $K = 1$ and $\alpha = 1.4$ or 1.8 for the USD and the GAD receivers. It is important to notice the gap between the two approaches, explaining the gap between the BER obtained with USD and the one obtained with GAD.

Necessarily, $\text{sign}(Y) = -1$ since $\text{sign}(Y) = 1$ would lead to $Y < 0$ which is in contradiction. Thus,

$$\begin{aligned} P\left(\widetilde{Z} < 0\right) &= P(X + N < -2) \\ &= \frac{1}{2}P(N < -3|X = 1) + \frac{1}{2}P(N < -1|X = -1) \\ &= \frac{1}{2}\left[Q_\alpha\left(\frac{3}{\gamma}\right) + Q_\alpha\left(\frac{1}{\gamma}\right)\right]. \end{aligned} \tag{5.12}$$

Finally, we have $\eta_{K,USD} = \left[1 - \frac{1}{2}Q_\alpha\left(\frac{3}{\gamma}\right) - \frac{1}{2}Q_\alpha\left(\frac{1}{\gamma}\right)\right]^K = (\eta_{1,USD})^K$. ■

Discussion

In Figure 5.17, we represent the degeneration risk for $K = 1$ for the genie-aided decoder as well as for the unsupervised decoder (USD) as given in Proposition 5. Note that the degeneration risk is only a qualitative measure which gives an idea of the probability of badly estimating a and b: when the value of this risk decreases, it indicates that region $A^- \cup B^-$ has a higher probability to contain samples and thus the estimation to converge.

Based on this numerical simulation, three conclusions can be made:

1. The degeneration risk under USD is significantly higher than under GAD, which shows that the estimation will be more difficult in the unsupervised setting and explains the BER degradation observed in Figure 5.14.

2. The degeneration risk increases when γ decreases. The dispersion γ is a scale parameter that measures the spread of the samples around the mean, similarly to the variance in the Gaussian case. Figure 5.17 shows that an estimation failure (η_K tends to 1) is inevitable when the noise dispersion is small, making all samples falling close to the mean (in our case $\widetilde{X} = 1$). Thus, it is less likely to have samples in $A^- \cup B^-$. The consequence is most significant in the medium and low γ regions. That consequence explains the bump observed in Figure 5.14.

3. The degeneration risk decreases as α becomes smaller (more impulsive). Indeed, the smaller the value of α, the heavier the tail of the probability density function, which increases the likelihood of having impulses with large amplitudes and far from the central location. Thus, it is more likely to have samples in $A^- \cup B^-$.

These aspects are not intuitive because they suggest that the reduction of the packet length will be more detrimental when noise is less impulsive and with a smaller strength. However, If the noise strength is low, its impact will be very low and even if the estimation is not accurate, the decoding should succeed. Nevertheless, the most impacting regime is the medium noise strength regime, in which the bump can be observed on the BER curves.

In the following, we propose a new training sequence $\widetilde{Y}$ design to reduce the degeneration risk under the unsupervised optimization.

5.4.2 Proposed solution for the estimation with short block length

We now provide two mechanisms in order to decrease the degeneration risk obtained under unsupervised optimization: one regularization term in order to prevent the parameter a to tend to infinity and one algorithm to increase the diversity in the extracted noise sequence, such that the convergence of the parameter b is improved.

- **Parameter a:** the value of a defines the slope of the linear part for low values of the received signal y. Thus, if this parameter tends toward infinity, the approximated LLR would tend towards b/y which gives the highest likelihood in samples around 0, which obviously is not efficient. To overcome this issue, we propose to add a regularization term to (5.6) such that reasonable values for a are obtained even if A^- is empty. The new objective function writes as

$$\widehat{H}(X|Y)_{\text{New}} = \widehat{H}(X|Y) + \log_2(1 + e^{a\epsilon}), \tag{5.13}$$

  where ϵ is a small value. The regularization term corresponds to an artificial small negative sample of amplitude ϵ in the A^- region. It prevents an empty A^- region that causes degeneration. With small ϵ, this term could be approximated by a linear regularization $\log_2(1 + e^{a\epsilon}) \approx a\epsilon/(2\log 2)$.

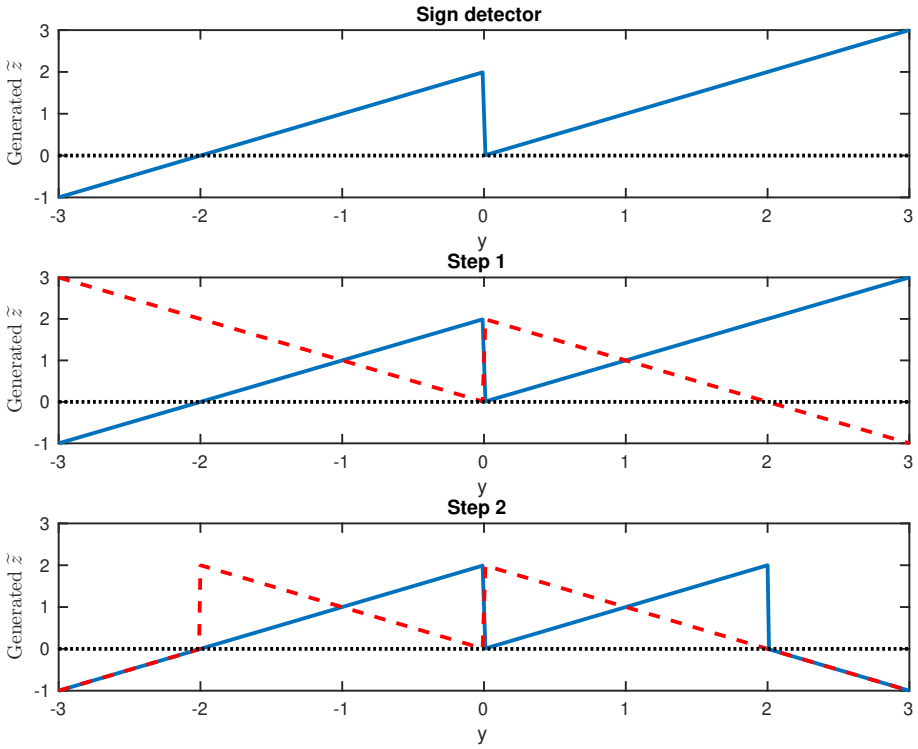


Fig. 5.18. Modification of the training sequence generated in the unsupervised way: the objective is to increase the probability of having Y values that result in a negative $\widetilde{Z}$. In the original sign detector, we see that only the Y values smaller than -2 result in negative $\widetilde{Z}$. In the concatenation sequence (step 1) the red curve corresponds to the opposite versions of $\widetilde{N}$. Y samples smaller than -2 but also larger than 2 result in negative $\widetilde{Z}$. In step 2, We force all Y values larger than 2 and smaller than -2 to give only negative $\widetilde{Z}$.

- **Parameter b:** the value of b defines the presence of the large samples. One important fact is that the optimization parameter b can go to infinity when the noise is Gaussian in order to have a fully linear receiver, which is known to be optimal. For this reason, we do not want to include a regularization term on b. Our objective is thus to increase the diversity in the extracted noise sequence and in turn to tune $\widetilde{Y}$ in order to reduce the degeneration risk, which is performed in Algorithm 1. As a result of this algorithm, the received samples $\widetilde{Z}$ are obtained. The principle of the algorithm is illustrated in Figure 5.18. In the first step of the proposed algorithm, we concatenate the extracted noise $\widetilde{N}$ with its opposite version $(-\widetilde{N})$, yielding a new sequence $\widetilde{N}_c$. Figure 5.18, middle plot, exemplifies the effect of this first step, where one can see that $\widetilde{N}_c$ is symmetric and that the number of negative samples $\widetilde{Z}$ has increased.

  In the second step of the algorithm, we consider the generation of the $\widetilde{Z}$ sequence. In order to have a sample that falls in the B^+ region, the noise must be of the same sign than $\widetilde{X}$ and such that $|\widetilde{N}_c| \geq \sqrt{b/a} - 1$, whereas to fall in B^- the noise sample must be of opposite sign than $\widetilde{X}$ and such that

Data: Channel output Y
Result: $\widetilde{Z}$
`/* Step-1 */`
$\widetilde{N} = Y - \text{sign}(Y)$
$\widetilde{N}_c = -\widetilde{N}||\widetilde{N}$ *% Concatenate N and $-N$ into $\widetilde{N}_c$;* `/* Step-2 */`
for *k=1 to length($\widetilde{N}_c$)* **do**
  if $\widetilde{N}_c(k) \geq 1$ **then**
    $\widetilde{Z}(k) = +1 - \widetilde{N}_c(k)$
  else
    $\widetilde{Z}(k) = +1 + \widetilde{N}_c(k)$
  end
end

Algorithm 1: Generating the training sequence $\widetilde{Z} = +1.\widetilde{Y}$

$|\widetilde{N}_c| \geq \sqrt{b/a}+1$. The probability of this second event is much smaller than the probability of the first: having samples in B^+ is highly probable but having an empty B^- has also a high probability. To reduce this last probability, we force a negative $\widetilde{Z}$ when $|\widetilde{N}_c| > 1$.

The next proposition states the degeneration risk of our new training design, allowing thus to evaluate its benefits.

Proposition 6. *Under the new training sequence design with unsupervised optimization, the degeneration risk is given as* $\eta_{K,NUSD} = \left[1 - Q_\alpha\left(\frac{3}{\gamma}\right) - Q_\alpha\left(\frac{1}{\gamma}\right)\right]^K = (\eta_{1,NUSD})^K$.

Proof. under this new design, $\widetilde{Z}$ is given either as $\widetilde{Z} = 1 \pm \widetilde{N}_c$, depending on the value of $\widetilde{N}_c$, thus

$$\begin{aligned} P(\widetilde{Z} < 0) &= P(\widetilde{Z} < 0, \widetilde{N} \geq 1) + P(\widetilde{Z} < 0, \widetilde{N} \leq -1) \\ &\quad + P(\widetilde{Z} < 0, -1 \leq \widetilde{N} \leq 1) \\ &= P(\widetilde{Z} < 0, \widetilde{N} \geq 1) + P(\widetilde{Z} < 0, \widetilde{N} \leq -1) \\ &= P(\widetilde{N} \geq 1) + P(\widetilde{N} \leq -1) \\ &= 2\left(P(\widetilde{N} < -1)\right) \\ &= Q_\alpha\left(\frac{3}{\gamma}\right) + Q_\alpha\left(\frac{1}{\gamma}\right) \end{aligned} \tag{5.14}$$

and $\eta_{K,NUSD} = \left[1 - Q_\alpha\left(\frac{3}{\gamma}\right) - Q_\alpha\left(\frac{1}{\gamma}\right)\right]^K = (\eta_{1,NUSD})^K$. ■

Discussion

Figure 5.19 compares the degeneration risk η_K of the GAD and USD as a function of γ under an additive impulsive SαS noise with $\alpha \in \{1.8; 1.4\}$ representing low

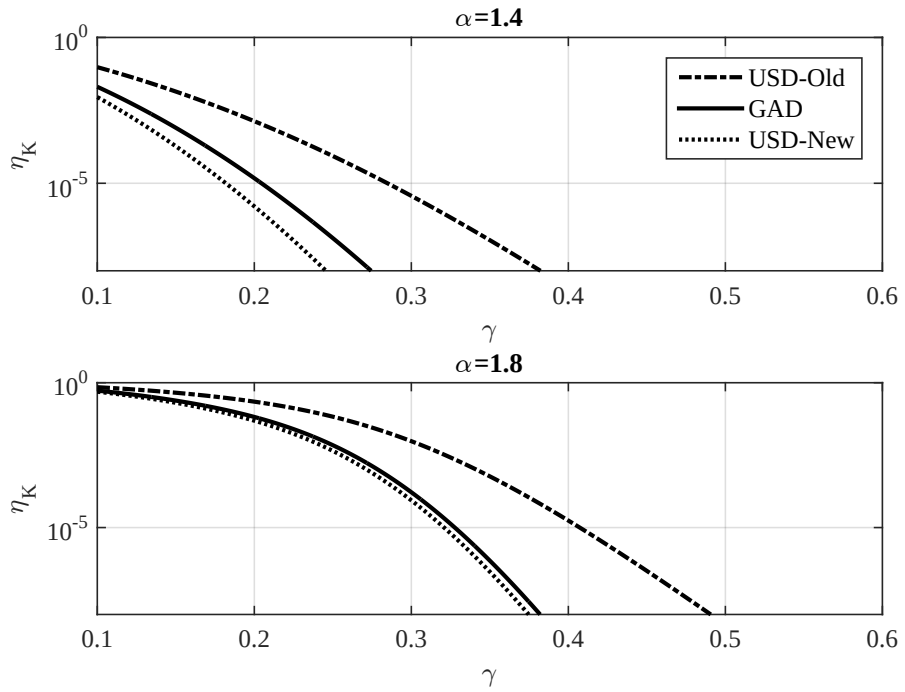


Fig. 5.19. Comparison of the degeneration risk η_K for GAD, the simple sign detector USD or the new unsupervised case NUSD as a function of γ, under SαS noise with $\alpha = 1.8$ and $\alpha = 1.4$ representing less and high impulsive environments, respectively and with $K = 408$.

and high impulsive environments respectively. The block size considered is 408 which depicts short codewords. The new unsupervised approach, which improves the diversity in the generated $\widetilde{Z}$ sequence, allows to significantly decrease η_K, which gets even smaller than the one under the GAD. Nevertheless, this does not guarantee better BER performance but only that the estimation should be more robust from one packet to another.

In Figure 5.20, we provide the evolution of the mean and the variance of the optimization parameters a and b under our proposed approach, which can be compared to the one obtained in Figure 5.15 without the two mechanisms allowing the degeneration risk decrease. One can note that under our proposed method, the optimization parameters do converge for all values of the noise strength. Even if the parameter b still exhibits some variability, it is significantly reduced: the variations around the mean values are much more confined compared to the previous method and will not result in catastrophic values. The threshold $\sqrt{|b/a|}$ remains stable so that the shape of the LLR remains similar whatever the training set. It should be mentioned that similar conclusions are observed for other values of α or other impulsive environment types, as for instance, Middleton Class A, ϵ-contaminated or Gaussian noise.

In the following, we present some numerical simulations showing the performance of our new designed training sequence under blind optimization.

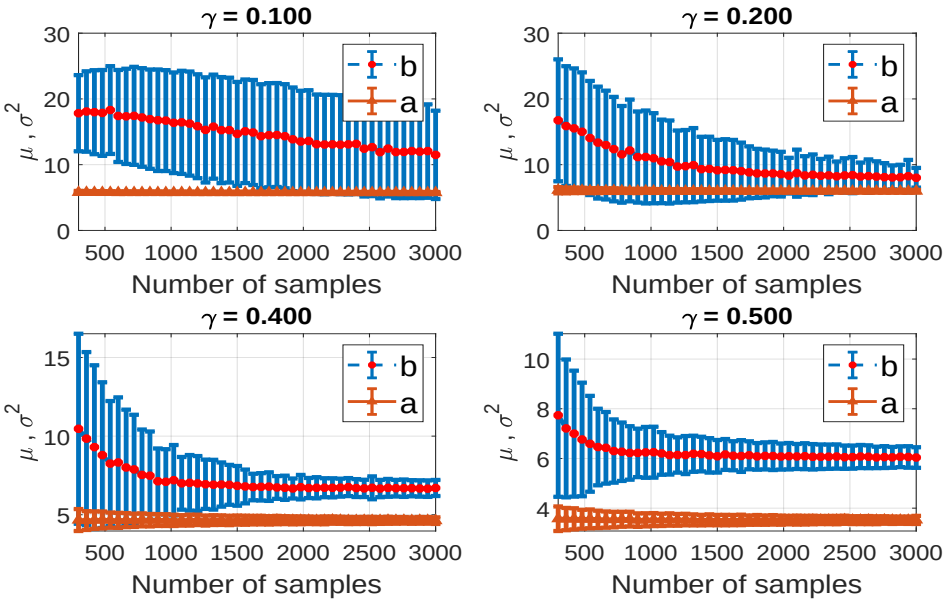


Fig. 5.20. Evolution study of a and b parameters using the proposed solution as a function of the number of samples for $S\alpha S$ noise with $\alpha = 1.8$.

5.4.3 Numerical results using LDPC codes

Simulation Setup

We assume that the input sequence is encoded using a regular (3,6) LDPC code of size $n = 408$. Consequently, the available number of samples to solve (3.14) is set to $K = 408$, allowing to assess the performance gain of the new training sequence design in the short codeword regime. In order to assess the robustness of our method, we will consider the five following noisy environments:

- low impulsive $S\alpha S$ environment with $\alpha = 1.8$ in Figure 5.21;
- high impulsive $S\alpha S$ environment with $\alpha = 1.4$ in Figure 5.22;
- Gaussian noise in Figure 5.23;
- moderate impulsive Middleton environment with $A = 0.1$ and $\Gamma = 0.1$ in Figure 5.25.
- high impulsive Middleton environment with $A = 0.01$ and $\Gamma = 0.01$ in Figure 5.24.
- low impulsive $\epsilon-$contaminated environment with $\epsilon = 0.01$ and $K = 10$ in Figure 5.26.
- high impulsive $\epsilon-$contaminated environment with $\epsilon = 0.1$ and $K = 10$ in Figure 5.27.

Under all noise types, we provide the BER curves obtained with the GAD, with the unsupervised optimization with our previously proposed training sequence

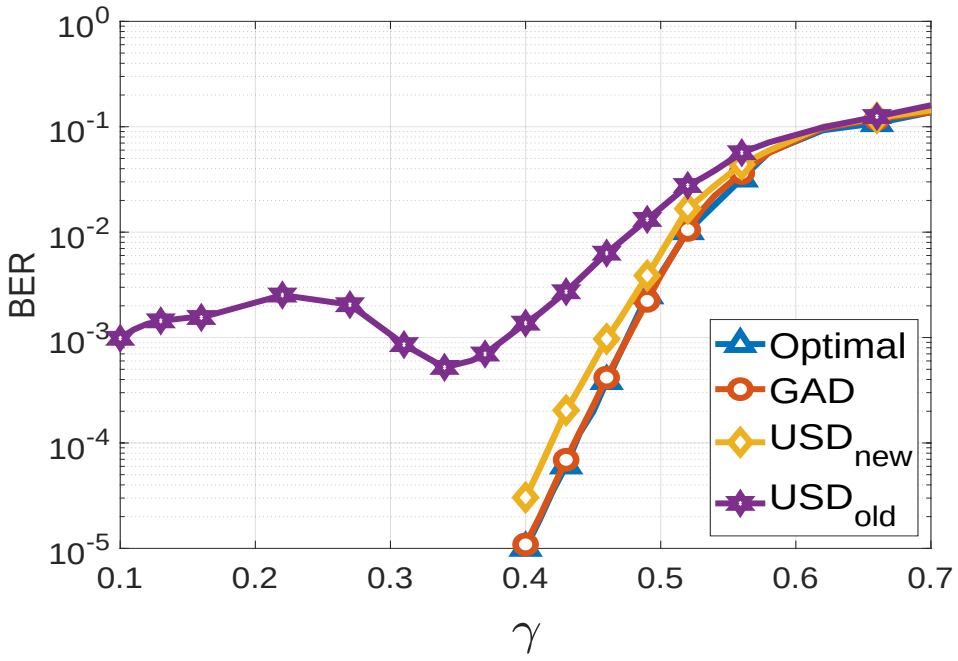


Fig. 5.21. BER evolution as a function of γ under a low impulsive $S\alpha S$ noise with $\alpha = 1.8$

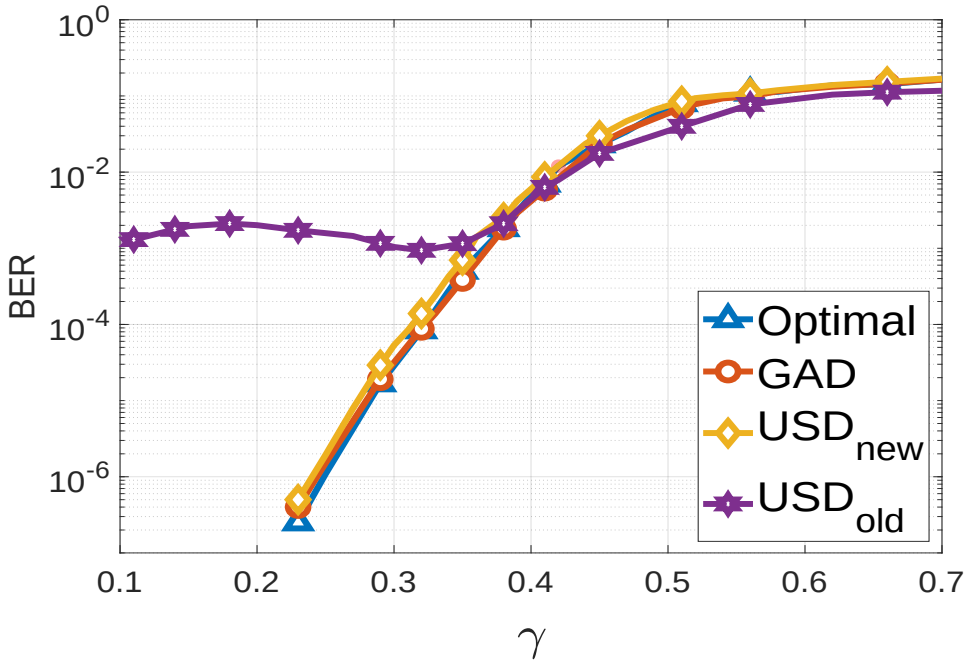


Fig. 5.22. BER evolution as a function of γ under a high impulsive $S\alpha S$ noise with $\alpha = 1.4$

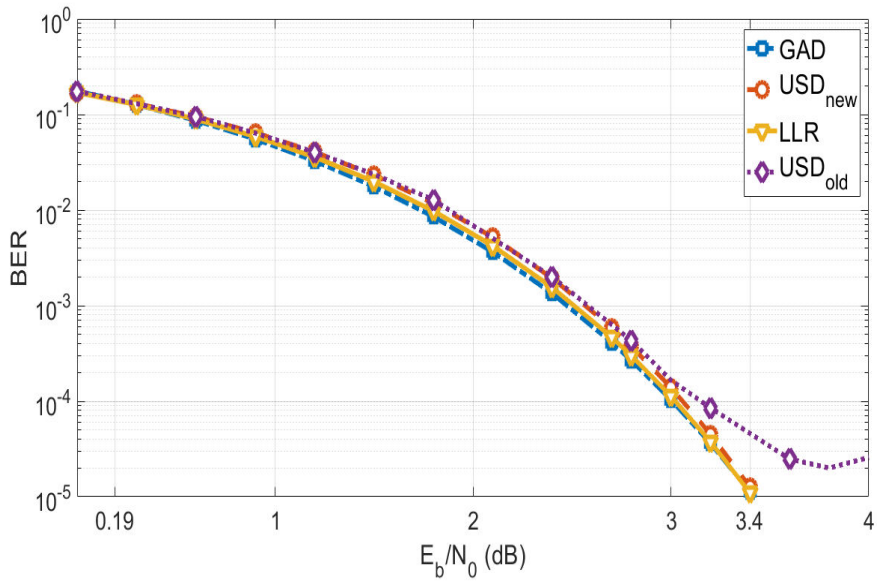


Fig. 5.23. BER evolution as a function of E_b/N_0 under a Gaussian noise

design [MSG+18, MSG+19d] (USD old) as well as with the newly proposed training sequence (USD new). We furthermore provide the BER curves obtained with the numerically obtained true LLR to represent the best possible case with the given code and BP. Under $S\alpha S$ noises, we present the BER curves as a function of the dispersion γ; whereas under the Gaussian, Middleton and ϵ-contaminated cases, we present the BER curves a function of the signal-to-noise ration E_b/N_0.

Discussion

Under all types of noises, impulsive or not, the GAD reaches the optimum, proving the validity of the LLR approximation choice L_θ under the short packet assumption. It seems unnecessary in that case to use a more accurate approximation, as L_{abc} that was proposed in Section 4.2.3. Moreover, under the old training sequence design, the BER curve is not monotonic and exhibits a bump due to the degeneration explained in Section 5.4.1, whereas the BER curve obtained under the new training sequence design is close to the one obtained with GAD. The new training sequence improves the performance of the optimization step under short codewords for all types of additive noise, impulsive or not, without requiring the noise model knowledge at the receiver. Furthermore, our method does not decrease the performance under Gaussian noise, where the optimal b parameter should be infinite. Thus, our proposed method allows to achieve almost optimal BER performance for noises ranging from highly impulsive, moderately impulsive to Gaussian, with a low implementation complexity without the knowledge of the noise model, even with a small available number of data samples to perform the LLR approximation optimization.

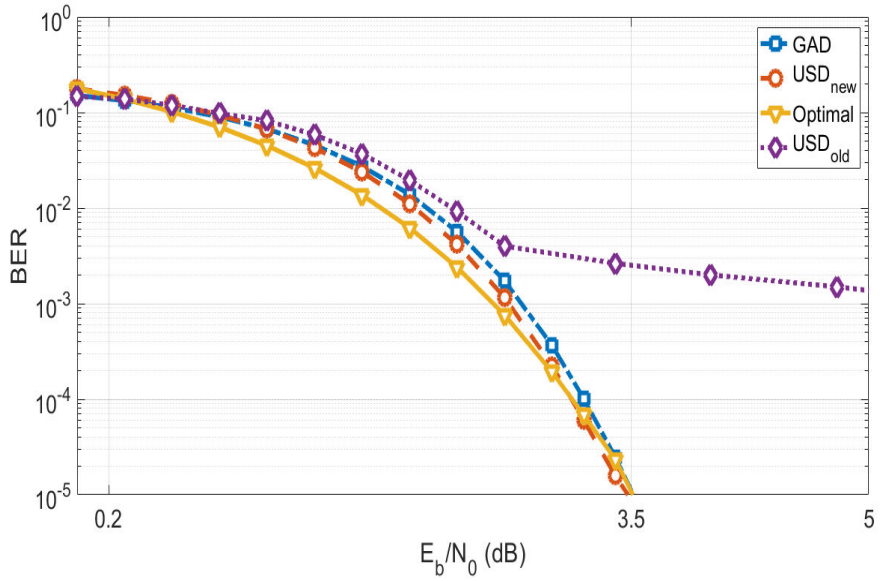


Fig. 5.24. BER evolution as a function of E_b/N_0 under a high impulsive Middleton noise with $A = 0.01$ and $\Gamma = 0.01$.

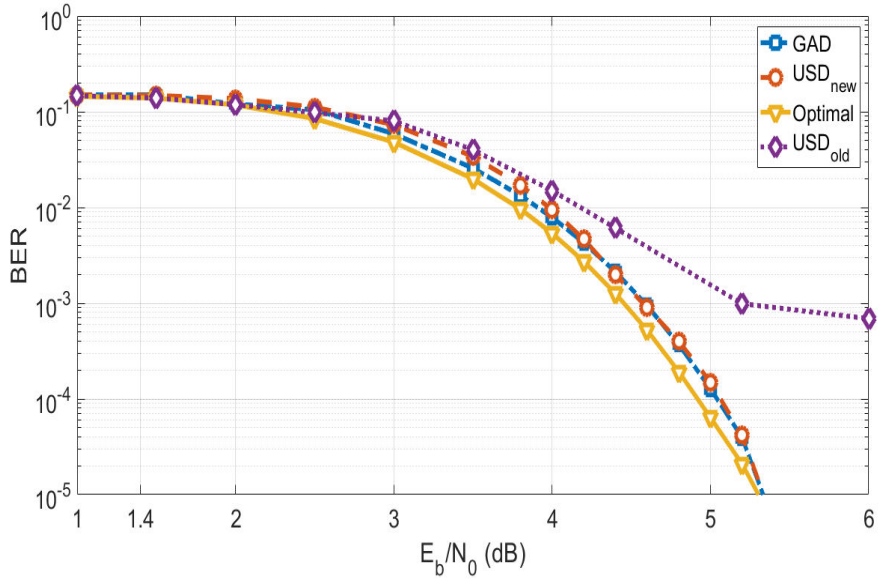


Fig. 5.25. BER evolution as a function of E_b/N_0 under a moderate impulsive Middleton noise with $A = 0.1$ and $\Gamma = 0.1$.

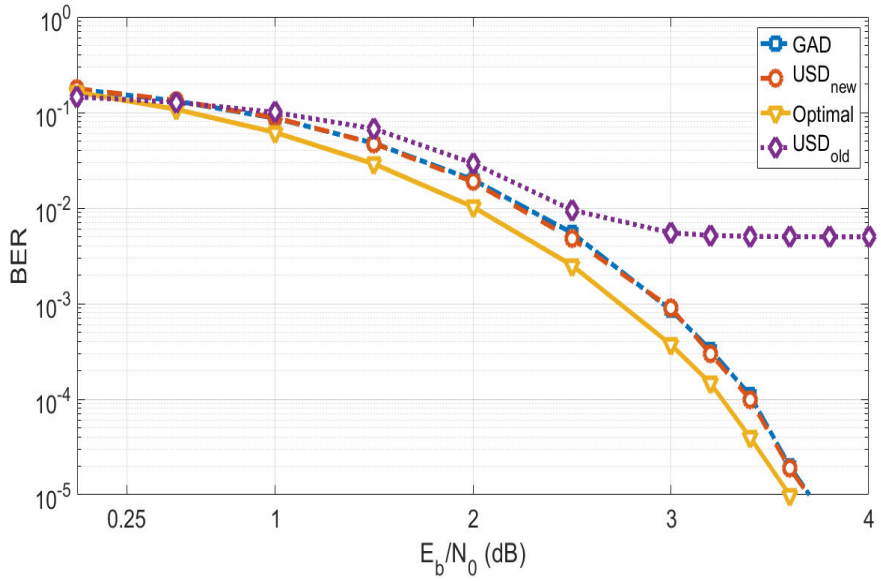


Fig. 5.26. BER evolution as a function of E_b/N_0 under an ϵ-contaminated noise with $\varepsilon = 0.01$ and $K = 10$.

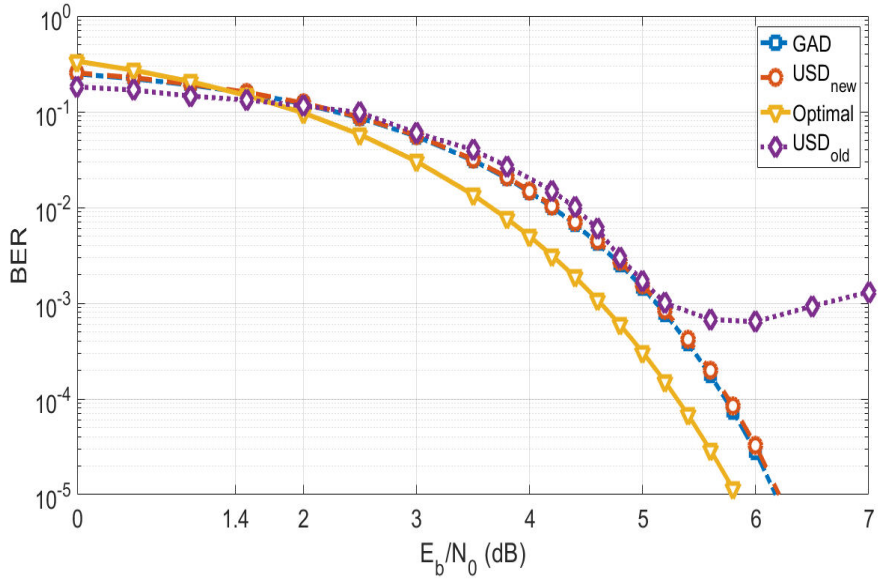


Fig. 5.27. BER evolution as a function of E_b/N_0 under an ϵ-contaminated noise with $\varepsilon = 0.1$ and $K = 10$.

5.5 Conclusion

Many decoding schemes rely on the log-likelihood ratio, whose derivation depends on the noise distribution. Unfortunately, in various dense and heterogeneous networks, this knowledge is missing and the noise plus interference term exhibits an impulsive nature, under which the LLR is not linear anymore and computationally prohibitive. We proposed an unsupervised universal receiver design. We choose a LLR approximation function f_θ in a parametric family. The parameters θ are estimated through the maximization of mutual information. An unsupervised solution is proposed in order to benefit from the whole received sequence but also to increase the useful data rate. We studied the performance of the unsupervised estimation in the asymptotic and finite block length regime.

However, in the short block length regime, we show that if the simulated transmission is designed by simply adding the extracted noise to the all one-codeword, the BER is non-monotonic and exhibits a severe performance loss compared to a genie-aided approach. We first analytically derived a metric to measure the probability of this performance loss to occur. Based on this analysis, we proposed a new design for the simulated transmission as well as adding a regularization term in the optimization problem so that the unsupervised LLR approximation achieves almost the same performance as the genie-aided decoder even under short codewords for a wide range of noise types while exhibiting a limited implementation complexity.

Chapter 6

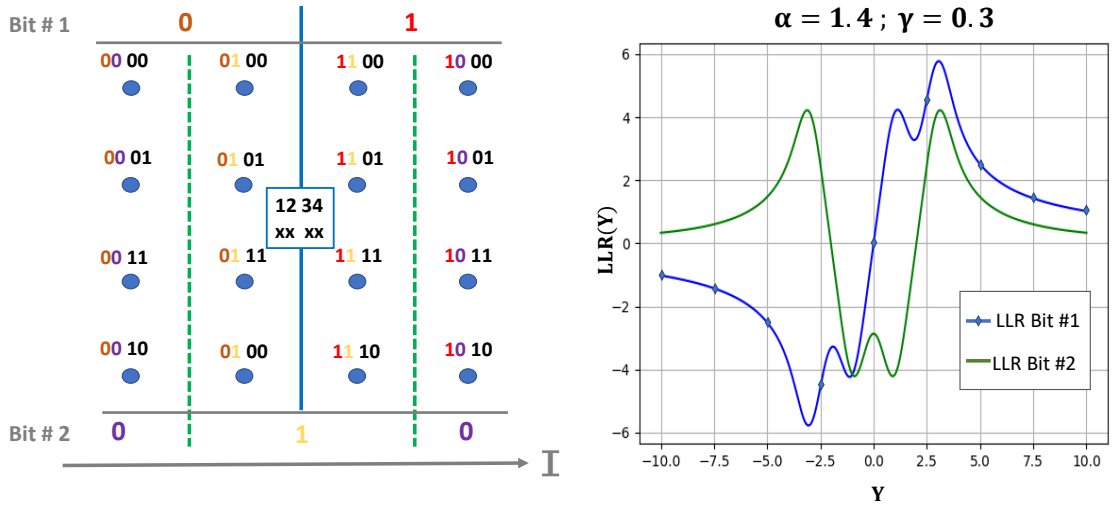
Conclusion and perspectives

This thesis focused on designing a robust receiver that exhibits a near-optimal performance over Gaussian and non-Gaussian environments without relying on the knowledge of interference plus noise statistical properties. This receiver strives for universality by adapting automatically to a large variety of different conditions. Such receivers can play an important role in future networks that will face two crucial challenges, robustness and adaptability, in a highly interfering environment and with strict energy and lifetime constraints.

We have shown in our work that a simple module between the channel output and the decoder input allows to combat effectively the noise and interference that disrupt point-to-point communications in a network. This module can be used as a front end of any LLR-based decoder. It consists of a function approximating the calculation of the likelihood ratio, which is rarely accessible and that requires the knowledge of the channel status. However, thanks to a judicious criterion, it is possible to search for this approximation function effectively either by supervised learning or by an unsupervised one. We show that it is even possible to use such a scheme for short packet communications without performance losses.

The perspectives of our work are many. To start with we can give some issues that have not been really discussed in the thesis but would deserve some attention. For instance, the complexity issues are not properly addressed and if we think that our demapper would not significantly increase the decoding complexity, it should be carefully analyzed. In the short term, it is relevant to extend our results for higher order modulations such as quadrature amplitude modulations (QAM). In this case, the likelihood ratios depend on the position of the bits in the binary coding of the symbols. It is, therefore, necessary to propose families of functions that take into account this position. Our preliminary works show that it is possible to return to a mixture of our functions L_{ab} to limit the size of the parameters or even to use the L_{abc}. Figure 6.1 shows an example of a gray coded 16-QAM constellation as shown in plot (a) and the corresponding LLR for the first and second most significant bits shown in plot (b), where the received samples Y is influenced by a SαS noise. For instance, the LLR of the first bit L_{b1} can be approximated as

$$L_{b1} = \max\left\{L_{ab}(y), L_{ab}(y-2)\right\},$$



(a) Gray coded 16-QAM constellation.

(b) LLR computation for the first and second most significant bits.

Fig. 6.1. Example of a LLR-computation for higher-order constellations (16-QAM)

or perhaps simply by using L_{abc}.

Another promising perspective is to integrate the LLR approximation to the code design. The LLR may uniformize the distribution of the messages coming from the channel outputs for several impulsive noises. Looking for a good code candidate will reduce the gap between the performances and channel capacity, which, as we have shown, is sometimes rather large. Of course, the asymptotic limit should be replaced by the short block length regime bound proposed by Polyanskiy [PPV10] to have a better insight on the gain that can be reached (section 5.4.3), this is an on-going work.

In the longer term, since the proposed criterion is close to a variational approach to learning, it may be appropriate to integrate it directly into channel decoding so that there is an exchange of information between the code and the LLR calculation. This is all the more relevant as in the unsupervised case, limiting the number of decision errors improves the quality of the noise samples obtained and therefore the estimation of the parameters. A simple joint coupling between channel decoding and LLR approximation has already been tested but we did not observe any significant improvement. However, it deserves more detailed analysis, for example by using a coupling of factor graphs between that of the LDPC code and that of the LLR approximation.

Besides, we have always considered approximations that belong to a certain family built *a priori*. Since the criterion comes from information theory, it would be beneficial to make the approach even more generic by choosing a function according to Jaynes [Jay03] maximum entropy approach and combining it with channel decoding directly to make a decoder based on the principle of maximum entropy.

Also, it has been shown that interference can exhibit a non-trivial dependence structure, especially some tail dependence [ZEC+19]. How to account for this dependence is an interesting open question that could significantly impact the robustness

of the communication.

Annexe A

Résumé détaillé en français

A.1 Introduction

A.1.1 Contexte

Lors de communications filaires, l'agitation thermique au niveau du récepteur explique en grande partie les bruits perturbant les signaux porteurs d'information. Le théorème central limite sert alors de justification à leur modélisation par undistribution gaussienne centrée. En effet, l'agitation thermique consiste en un grand nombre de fluctuations centrées indépendantes et de puissance finie qui correspond aux prémisses du théorème.

Les communications sans fils et en réseau subissent en plus du bruit thermique des interférences avec les communications voisines qui ne peuvent pas être considérées comme du bruit indépendant et gaussien. Plusieurs techniques ont été développées pour les limiter comme l'alignement d'interférences au niveau de la couche physique [EPH13] ou des techniques d'évitement de transmissions simultanées comme le CSMA au niveau de la couche MAC [YV05, JHMB05]. Enfin, d'autres méthodes essaient de les supprimer efficacement au niveau du récepteur comme l'annulation successives d'interférences (SIC) [WAYd07, And05].

Toutes ces techniques ne peuvent parfaitement annuler les et divers. Cela se vérifie d'autant plus que nous nous dirigeons vers des réseaux denses comme LoRa, Sigfox, la 5G ou en général l'IoT sans contrôle centralisé ni d'accès à la ressource radio ni aux puissances des émissions.

Par conséquent, prendre en compte la présence des interférences au niveau du récepteur devient une nécessité, voire une obligation. Une modélisation fine et pratique des interférences s'avère alors un prérequis pour concevoir la chaîne de réception. Plusieurs approches ont été envisagées dans la littérature pour construire ces modèles. La première, tournée vers la théorie, consiste à utiliser les outils mathématiques comme la géométrie aléatoire [WA12, WPS09] avec des hypothèses sur la conception et l'utilisation du réseau pour en déduire des modèles pertinents. Ces travaux mènent par exemple aux modèles alpha-stables [PW10a, GCA+10] ou aux différentes classes des modèles de Middleton [Mid77]. Le second type d'approche est plus pragmatique : il s'agit de trouver des familles de lois de probabilité qui soient simples d'utilisation et qui correspondent aux aspects impulsifs des interférences. Ces lois doivent avoir

une queue lourde pour modéliser les événements impulsifs qui se caractérisent par des valeurs de forte amplitude. Parmi ces modèles nous retrouvons les gaussiennes généralisées ou les mélanges de gaussiennes.

Nous avons concentré particulièrement nos travaux sur les distributions alpha-stables symétriques et centrées comme modèle des interférences même si d'autres modèles ont été utilisés pour tester nos solutions.

A.1.2 Objectifs et contributions

L'objectif principal de nos travaux consiste à développer un récepteur simple, adaptatif et performant qui prenne en compte le caractère impulsif des interférences vues comme un bruit additif. Simplicité signifie ici une mise en œuvre peu coûteuse en temps de calcul et en mémoire. Le caractère adaptatif nécessite de prendre en compte l'état réel du canal sans le connaître par ailleurs ; pour cela un apprentissage devient nécessaire. Enfin, les performances ne doivent pas s'écarter de l'optimum fourni par le log rapport de vraisemblance.

Nous avons répondu à ces objectifs par nos contributions qui sont

- l'ajout d'une approximation directe du rapport de vraisemblance entre la sortie du canal et le décodeur de canal ;
- la proposition d'un critère pertinent pour l'adaptation de cette approximation selon le type de bruit et d'interférence ;
- l'utilisation de ce critère pour le développement d'un algorithme d'adaptation supervisé, c'est-à-dire basé sur une séquence d'apprentissage ;
- l'analyse de cet algorithme sur les critères de performances et de robustesse selon la longueur de la séquence d'apprentissage ;
- l'extension de cet algorithme à un apprentissage non supervisé et son étude ;
- l'amélioration de l'apprentissage non supervisé dans le cadre des communications à paquets courts.

À cela s'ajoute la proposition d'une nouvelle famille de fonctions d'approximation à trois paramètres plus performantes mais restant toujours simples et la proposition d'un critère plus simple reliant la divergence de Kullback-Leibler avec la norme L_2 pour comparer les récepteurs.

A.2 Description du système

Nos travaux reposent sur un bruit modélisé par les distributions alpha-stables et sur un système qui utilise les codes LDPC. Ces deux points sont discutés dans cette section pour finir par la spécification du modèle de canal considéré.

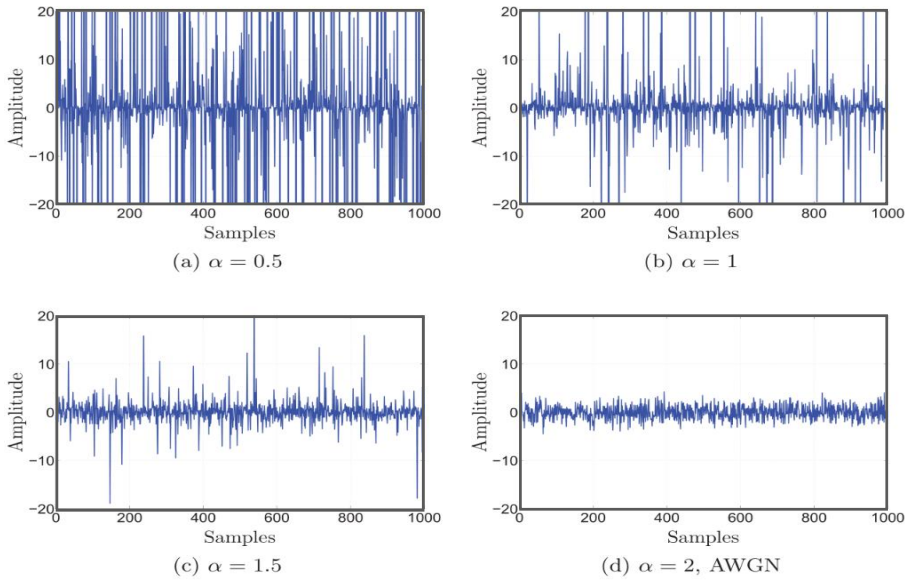


Fig. A.1. Exemple de réalisations de bruit alpha-stable.

A.2.1 Distributions alpha-stables

Les distributions alpha-stables sont les distributions limites des sommes de variables aléatoires normalisées et centrées lorsqu'elles existent [GH97]. Nous nous limiterons par la suite aux versions symétriques, et, dans cette situation, ces distributions limites se décrivent par leur fonction caractéristique

$$\phi(t) = e^{-|\gamma\, t|^{\alpha}}, \tag{A.1}$$

paramétrée par leur exposant caractéristique α et leur facteur d'échelle γ. Parmi ces distributions se trouvent les lois normales avec $\alpha = 2$ et $\gamma = \sigma/\sqrt{2}$ et les distributions de Cauchy avec $\alpha = 1$ et $\gamma = 1/a$.

Quelques réalisations de bruits selon les distributions alpha-stables sont représentées sur la figure A.1. Il est visible que le paramètre α caractérise l'impulsivité du bruit : plus α est faible plus la probabilité d'avoir des échantillons de forte amplitude augmente ainsi que la valeur de ces amplitudes.

Des densités de probabilité de lois alpha-stables pour différentes valeurs de α sont tracées sur la figure A.2. Elle illustre bien le caractère "queues lourdes" de ces modèles qui permet de prendre en compte le caractère impulsif des interférences. Le cas gaussien $\alpha = 2$ se particularise vraiment car la probabilité des grandes amplitudes décroît exponentiellement et non plus selon une loi inverse comme pour les autres α.

Bien qu'attrayantes les distributions stables souffrent de deux défauts importants. Le premier concerne leur variance qui est infinie lorsque $\alpha < 2$. Par conséquent, la notion de rapport sur bruit devient inopérante et c'est pourquoi, pour indiquer l'intensité des interférences nous utiliserons plutôt le facteur d'échelle γ. Le second défaut est l'absence d'expression simple de la densité de probabilité en général ; pour

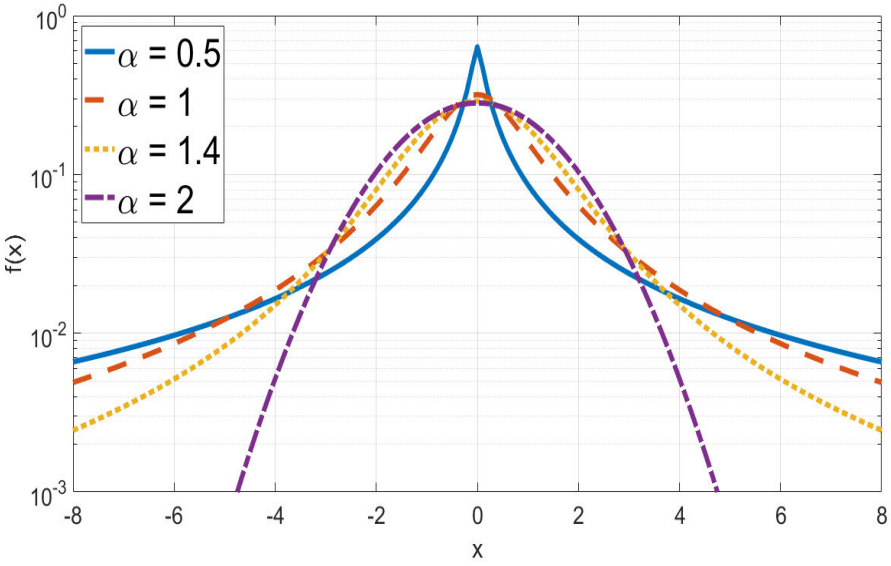


Fig. A.2. Densité de probabilité de lois alpha-stables.

la calculer il faut avoir recours à des intégrations numériques approchées calculant la transformée de Fourier inverse de la fonction caractéristique (A.1).

A.2.2 Codes LDPC

Nos travaux permettent de transformer la sortie du canal perturbée par des interférents et du bruit pour qu'elle soit utilisable pour la suite de la réception. Pour pouvoir mesurer les performances obtenues, il est nécessaire de recourir à un code correcteur d'erreurs. Le choix s'est naturellement porté vers les codes LDPC pour plusieurs raisons : ils forment une famille simple à décrire et performante pour une large gamme de longueurs de mots de code. L'algorithme de propagation de croyance appelé Sum Product Algorithm permet une mise en œuvre générique qui utilise directement en entrée les log rapports de vraisemblance (LLR) de la sortie du canal. Ainsi, nos travaux s'implémentent comme un élément de la chaîne entre la sortie du canal sous la forme du signal et du bruit et le décodeur LDPC sous la forme d'un LLR.

Les performances d'une chaîne avec un code LDPC suivent généralement le tracé de la figure A.3. La région de décroissance rapide du taux d'erreur s'appelle la région du waterfall qui est suivie par une stagnation de cette décroissance, la région du plancher d'erreur (error floor).

La suite du manuscrit ne concerne pas spécifiquement les codes LDPC et nous ne nous étendrons pas sur ce point. Toutefois, nous utiliserons dans nos simulations les codes régulier $(3, 6)$ de MacKay de longueur 20000 et de longueur 408 construit par l'algorithme du PEG [XEA05].

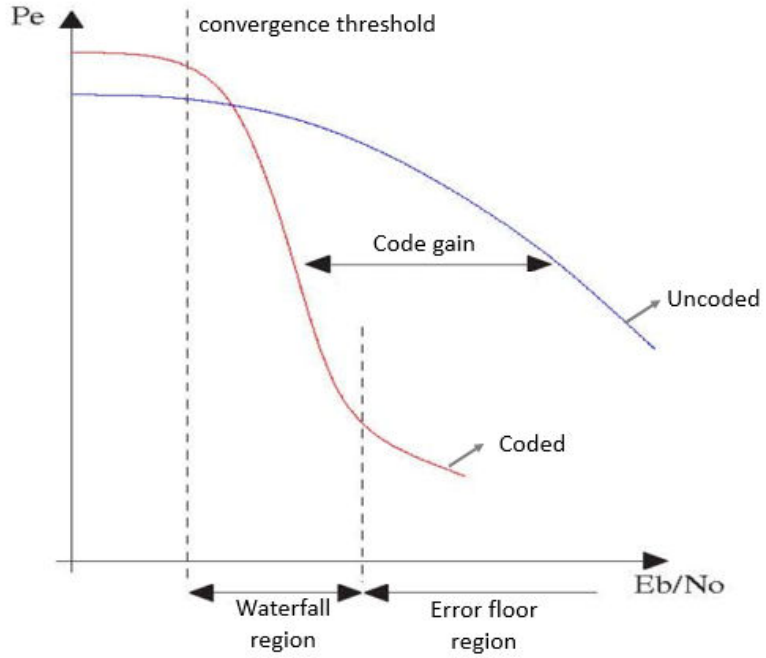


Fig. A.3. Regions caractéristiques des performances d'un code LDPC.

A.2.3 Modèle du système considéré

Le canal considéré ici est un canal binaire sans mémoire à sortie réelle symétrique — MBISO en anglais pour Memoryless Binary Input Symmetric Output [RU08]. Le canal additif sera le plus utilisé,

$$Y = X + N, \tag{A.2}$$

où l'entrée X est une variable aléatoire valant $+1$ ou -1 à l'instar d'une modulation bipolaire BPSK et N le bruit et l'interférence modélisés par une distribution symétrique, principalement de type alpha-stable. Ces deux variables sont indépendantes l'une de l'autre.

La sortie du canal est ensuite convertie en un rapport de vraisemblance ou une approximation de ce dernier

$$\Lambda(y) = \log \frac{\Pr[Y = y|X = +1]}{\Pr[Y = y|X = -1]}, \tag{A.3}$$

pour ensuite être décodée par le code LDPC grâce à l'algorithme de propagation de croyance.

A.3 Conception de récepteurs robustes

Le récepteur est un organe essentiel dans un environnement impulsif car il permet ensuite de prendre une décision sur des mesures pertinentes. La difficulté de sa conception vient de l'impulsivité qui détruit l'isotropie d'un vecteur de bruit.

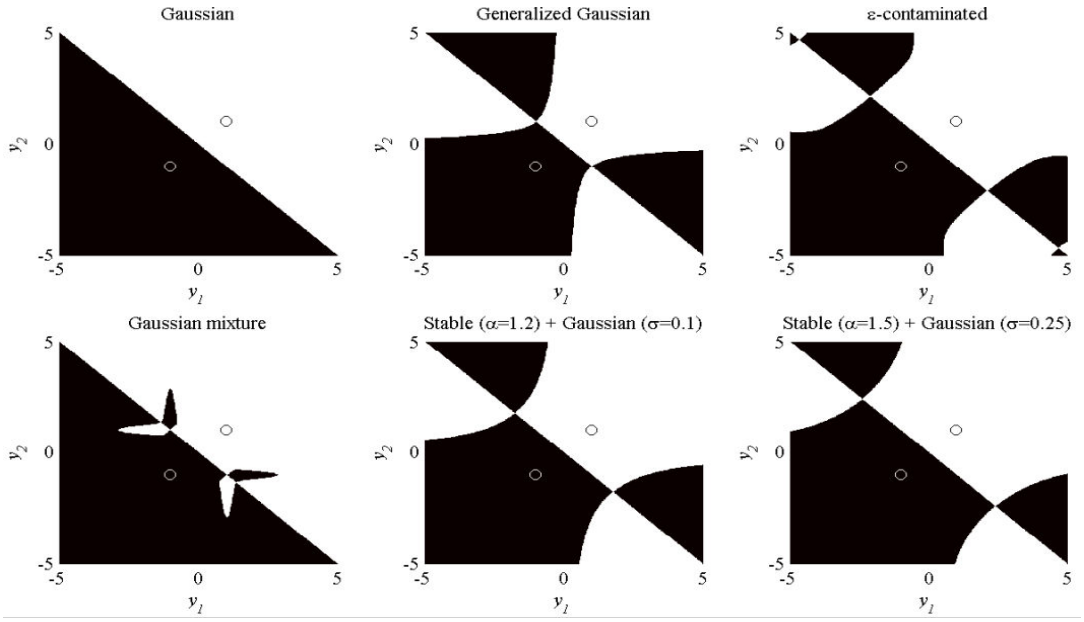


Fig. A.4. Régions de décision optimale pour différents type de bruit.

A.3.1 Régions de décision et impulsivité

Considérons la réception d'un signal valant $+1$ ou -1 et dont on récupère la valeur deux fois y_1 et y_2 mais bruitée par deux réalisations indépendantes de bruit impulsif.

La figure A.4 montre les régions de décision en fonction des valeurs mesurées y_1 et y_2 pour différents types de bruit. La région blanche indique une décision pour l'envoi de $+1$ et la région noire pour -1. Lorsque le bruit est gaussien, vignette en haut à gauche, les régions sont particulièrement simples car séparées par la médiatrice des points $(+1,+1)$ et $(-1,-1)$. Par contre, en présence d'impulsivité, les régions de décision ne sont plus nécessairement connexe comme en témoignent les autres vignettes où le bruit est soit un mélange de gaussiennes, soit la somme d'un bruit alpha-stable et d'un bruit gaussien. Dans ces cas, contrairement au cas gaussien, la distance aux points $(+1,+1)$ et $(-1,-1)$ n'est plus suffisante pour décider, il faut ajouter une information sur la direction des valeurs reçues. En ce sens, la brisure d'isotropie du vecteur bruit déconnecte les régions de décision.

A.3.2 Conception de récepteur

Pour répondre aux difficultés de détection en environnement impulsif, il est nécessaire de revenir au rapport de vraisemblance qui est une statistique suffisante [CT06]. Autrement dit, l'objectif du récepteur, encore appelé parfois demappeur, est de calculer le LLR $\Lambda(y)$ donné par (A.3).

Dans le cas d'un canal MBISO, ce LLR représente à la fois le rapport de vraisemblance mais aussi le rapport des probabilités *a posteriori* car l'*a priori* est uniforme,

$$\Lambda(y) = \log \frac{\Pr[Y=y|X=+1]}{\Pr[Y=y|X=-1]} = \log \frac{\Pr[X=+1|Y=y]}{\Pr[X=-1|Y=y]}, \tag{A.4}$$

et comme X est binaire et uniforme, la donnée du LLR est équivalente à celle de l'*a*

posteriori

$$\Pr[X = x|Y = y] = \frac{1}{1 + e^{-x\,\Lambda(y)}}. \tag{A.5}$$

Malheureusement, pour évaluer $\Lambda(y)$, il est nécessaire de connaître la distribution du bruit affectant le canal. En effet, pour un bruit additif, le LLR s'écrit $\Lambda(y) = \log\big[p(y-1)/p(y+1)\big]$ où $p(\cdot)$ est la densité de probabilité du bruit. Or cette densité est inconnue du récepteur.

Plusieurs stratégies sont envisageables pour calculer le LLR. La première consiste à présupposer un modèle de bruit, par exemple gaussien ou alpha-stable, paramétré par des coefficients qui seront estimés à partir de la sortie du canal. Par exemple, dans le cas gaussien, il est nécessaire d'estimer la puissance du bruit et dans le cas alpha-stable l'exposant caractéristique et le facteur d'échelle doivent être obtenus par ailleurs. Cette stratégie est limitée d'une part par la difficulté d'obtenir des estimations des paramètres suffisamment fiables notamment dans un environnement impulsif et d'autre part par une mauvaise modélisation du bruit du canal. Par exemple, utiliser un modèle gaussien dans un environnement impulsif s'avère désastreux en terme de performances [MGCG10a, MGCG13]. Enfin, lorsque l'expression de la densité de probabilité du bruit n'est pas une expression simple, le calcul du LLR s'avère soit coûteuse en temps soit en espace pour mettre en place des un système pré-calcul par table de correspondance. Cette situation malheureuse se présente par exemple lorsque le bruit se modélise par une distribution alpha-stable ou une distributions de Middleton qui nécessite alors le calcul d'une série.

Une seconde stratégie est de travailler plus près du décodage en modifiant directement la métrique utilisée, par exemple les métriques de branches dans l'algorithme de Viterbi pour les codes convolutifs. Citons par exemple les normes p [GC12] ou la métrique de Huber [Chu05]. Il reste alors à déterminer la bonne valeur des paramètres selon l'état du canal. Cette approche est simple avec une mise en œuvre peu coûteuse mais ne permet pas de dissocier clairement le décodage du récepteur.

Enfin, l'approximation directe de la fonction $\Lambda(y)$ du LLR fournit la dernière stratégie. Ces approximations sont bien sûr paramétrées mais ces coefficients sont directement estimés sans recourir à un modèle particulier même si la famille des fonctions approchantes sous-entend un modèle.

Plusieurs familles ont été proposées dans la littérature pour limiter les impulsivités du bruit comme le récepteur linéaire avec saturation [MGCG13], les récepteurs de Cauchy ou logistique [MGCG10a] ou encore le Hole-puncher [MGCG10b].

Dans la suite, nous utiliserons surtout l'approximation $L_{ab}(y)$ proposée dans [DGCG14] et définie par

$$L_{ab}(y) = \begin{cases} ay & \text{si } -\sqrt{b/a} < y < \sqrt{b/a}, \\ b/y & \text{sinon.} \end{cases} \tag{A.6}$$

Initialement, cette approximation est la conséquence de l'étude du LLR pour les bruits de type alpha-stable pour les valeurs de y proche de 0 et pour celles tendant vers l'infini. Les graphes du LLR pour un canal à bruit alpha-stable pour $\alpha = 1.4$ et des approximations linéaires et asymptotiques sont tracés sur la figure A.5. La correspondance est visuellement pertinente et prend bien en compte l'impulsivité en

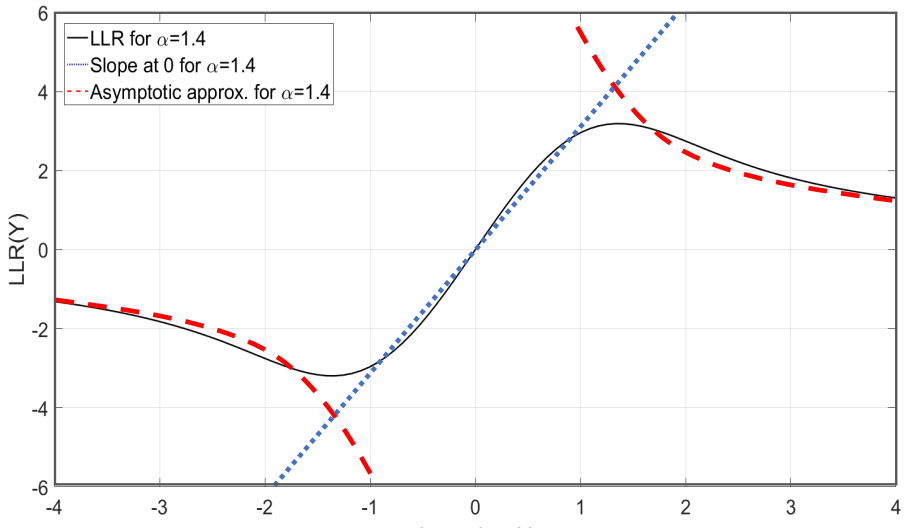


Fig. A.5. LLR pour un bruit alpha-stable $\alpha = 1.4$ et approximation.

faisant décroître le LLR pour de grande valeur de y. De plus seulement 2 paramètres a et b sont nécessaires pour caractériser une approximation.

Nos travaux ne présupposent pas une approximation de la forme $L_{ab}(y)$. Certaine partie de ce manuscrit étend les résultats à des formes plus génériques ; mais dans la plupart des cas, nos résultats se généralisent sans difficulté.

A.3.3 Critères de sélection

Une fois la famille de fonctions approchant le LLR $\Lambda(y)$ choisie, reste-t-il encore à sélectionner la meilleure fonction en fonction de la sortie du canal ou du bruit. Naturellement, l'introduction du terme "meilleure" nécessite un critère permettant d'ordonner les différentes fonctions candidates.

Le premier critère est naturellement celui de minimisation du taux d'erreur de la chaîne de communication, par exemple à la sortie du décodeur du code LDPC. Mais celui-ci demeure hors de portée en pratique car il nécessite la connaissance parfaite de la distribution du bruit et demande une puissance de calcul importante. Toutefois ce critère servira à comparer par simulation les différentes solutions et les différentes familles d'approximations.

Autre critère, une distance entre le LLR et l'approximation permettrait de mesurer l'adéquation entre les deux courbes avec pour attente des performances équivalentes en terme de taux d'erreur. La norme L_2 entre le LLR $\Lambda(y)$ et une fonction $L_\theta(y)$ dépendante des paramètres θ pourrait servir pour sélectionner la meilleure approximation par l'optimisation

$$\theta^* = \arg\min_\theta \int_{\mathbb{R}} \left|\Lambda(y) - L_\theta(y)\right|^2 dy. \tag{A.7}$$

Une approche plus intéressante serait de pondérer cette norme L_2 par la densité de probabilité pour que l'approximation $L_\theta(y)$ soit d'autant plus proche du LLR sur

les domaines fréquents de la sortie du canal, c'est-à-dire,

$$\theta^* = \arg\min_{\theta} \int_{\mathbb{R}} \left|\Lambda(y) - L_\theta(y)\right|^2 p(y)\, dy. \tag{A.8}$$

Nous avons montré dans [MSG+19a] que ce critère permet de borner la divergence de Kullback-Leibler entre l'*a posteriori* $\Pr[X = x|Y = y]$ et son approximation $\Pr_\theta[X = x|Y = y]$ obtenue en remplaçant $\Lambda(y)$ par $L_\theta(y)$ dans (A.5),

$$K\text{EQM} \leq D\Big(\Pr_\theta[X = x|Y = y]\Big\|\Pr[X = x|Y = y]\Big) \leq K'\text{EQM}, \tag{A.9}$$

où sont EQM, pour erreur quadratique moyenne, est la norme pondérée ou (A.8) et où K et K' deux constantes. Ainsi la sélection par le critère EQM est proche en un certain sens d'une sélection par une recherche de probabilité la plus proche de l'*a posteriori.* selon une approche variationnelle.

Le dernier critère qui est à la base de l'approche que nous développerons par la suite repose sur la théorie de l'information et notamment sur l'idée présentée dans [YA09]. La capacité d'un canal MBISO s'écrit comme $C = H(X) - H(X|Y)$. Or, d'une part, par équiprobabilité de l'entrée binaire, $H(X) = 1$ et comme d'autre part, la loi de $X|Y$ est donnée par (A.5),

$$C = 1 - \mathbb{E}\,\log_2\Big(1 + e^{-X\Lambda(Y)}\Big). \tag{A.10}$$

Il est possible de borner inférieurement la capacité C par une valeur approchée C_θ obtenue en remplaçant comme ci-dessus le LLR $\Lambda(y)$ par une fonction approchée $L_\theta(y)$,

$$C \geq C_\theta = 1 - \mathbb{E}\,\log_2\Big(1 + e^{-XL_\theta(Y)}\Big). \tag{A.11}$$

Le critère de sélection devient ici

$$\theta^* = \arg\max_{\theta} 1 - \mathbb{E}\,\log_2\Big(1 + e^{-XL_\theta(Y)}\Big), \tag{A.12}$$

ou, de façon équivalente,

$$\theta^* = \arg\min_{\theta} H_\theta(X|Y) \quad \text{avec} \quad H_\theta(X|Y) = \mathbb{E}\,\log_2\Big(1 + e^{-XL_\theta(Y)}\Big). \tag{A.13}$$

Le critère (A.13) est pertinent car l'écart entre C et C_θ vaut la divergence de Kullback-Leibler entre l'*a posteriori* $\Pr\theta[X = x|Y = y]$ et son approximation $\Pr_\theta[X = x|Y = y]$.

A.4 Apprentissage supervisé du LLR

Le critère de sélection d'une fonction approchant le LLR dans une famille paramétrée est maintenant fourni par (A.13). Il n'est pas utilisable tel quel car il fait intervenir l'espérance mathématique sur X et sur Y. Nous proposons et étudions par la suite une méthode d'optimisation de ce critère utilisable dans une chaîne de communications.

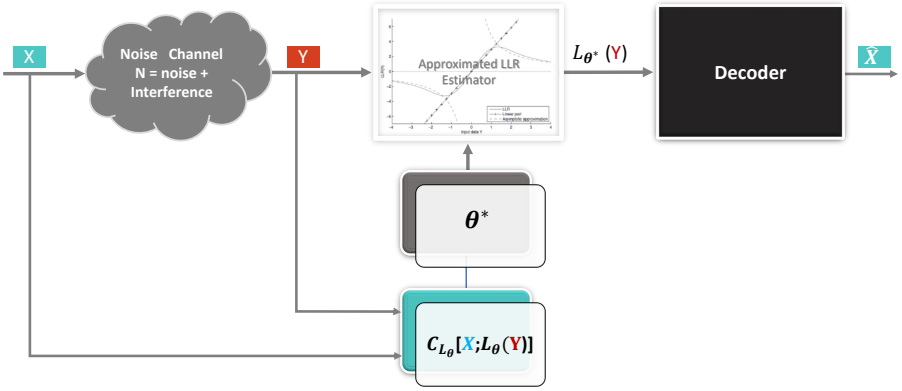


Fig. A.6. Synoptique de l'apprentissage supervisé.

A.4.1 Schéma de l'apprentissage supervisé

Dans le cadre de l'apprentissage supervisé, l'émetteur transmet au récepteur une séquence binaire connue des deux extrémités. Grâce à cette séquence dite d'apprentissage, le récepteur adapte l'approximation du LLR en optimisant le critère précédent. Le synopsis de cette construction est représenté sur la figure A.6. L'apprentissage terminé, le système utilise alors l'approximation apprise pour décoder les autres symboles binaires émis.

Naturellement, l'application du critère (A.13) nécessite d'être remaniée en utilisant la loi des grands nombres qui permet de remplacer l'espérance mathématique par une moyenne empirique

$$\theta^* = \arg\min_{\theta} \widehat{H}_\theta(x_{1:K}|y_{1:K}), \tag{A.14}$$

où

$$\widehat{H}_\theta(x_{1:K}|y_{1:K}) = \frac{1}{K}\sum_{i=1}^{K} \log_2\left(1 + e^{-x_k L_\theta(y_k)}\right), \tag{A.15}$$

en notant $x_{1:K}$ les K symboles de la séquence d'apprentissage et $y_{1:K}$ les K sorties de canal correspondantes.

La taille de la séquence d'apprentissage devient alors un paramètre important pour la bonne estimation des paramètres de la fonction d'approximation du LLR. La prochaine section s'intéresse au cas ou cette longueur est suffisante pour analyser la pertinence du critère et de l'approche sans pour autant s'encombrer des difficultés de l'estimation par la loi des grands nombres. La section suivante sera dédiée à la situation lorsque la taille de la séquence d'apprentissage diminue.

A.4.2 Cas de séquences d'apprentissage longues

La figure A.7 montre le résultat de l'optimisation de (A.14) lorsque la famille d'approximations L_{ab} donnée par (A.6) est utilisée sur un canal à bruit additif alpha-stable de paramètres $\alpha = 1.4$ et $\gamma = 0.45$. La séquence d'apprentissage est de 20000 symboles binaires. Sur cette même figure apparaît le graphe du LLR obtenu par intégration numérique et en utilisant les paramètres du canal.

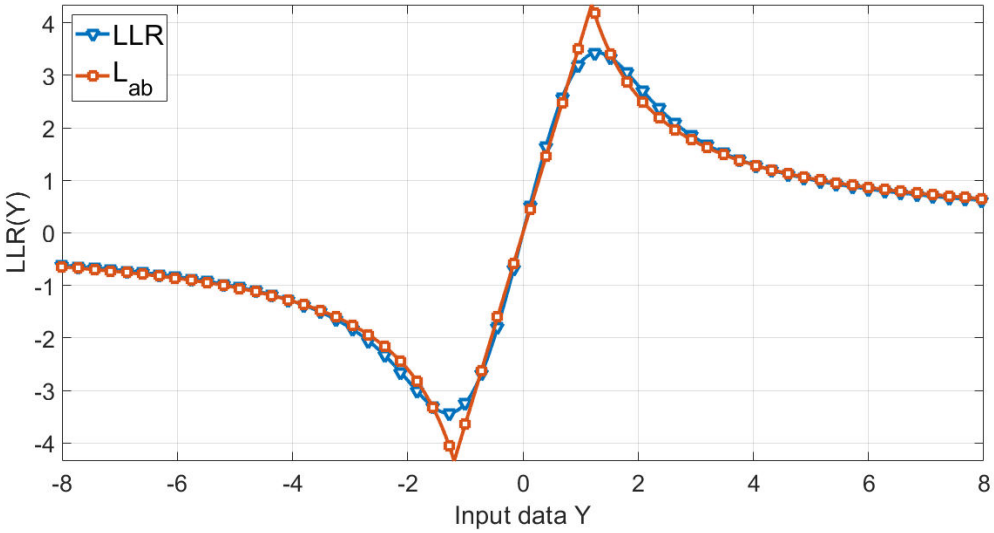


Fig. A.7. LLR et approximation par apprentissage supervisé avec une longue séquence pour un canal alpha-stable de paramètres $\alpha = 1.4$ et $\gamma = 0.45$.

Il est remarquable de constater à quel point les deux courbes sont similaires. Cette adéquation conforte la pertinence de l'approche. Des résultats similaires sont obtenus pour d'autres valeurs de α et γ.

Lorsque les conditions du canal sont apprises, la fonction d'approximation trouvée sert à faire le décodage. Sur la figure A.8, le taux d'erreur binaire finalement obtenu est tracé en fonction du facteur d'échelle γ. Le paramètre α du canal reste fixé à 1.4. Pour comparaison, les performances obtenues en utilisant le LLR calculé par intégration numérique avec les valeurs exactes des paramètres du canal sont aussi tracées en ligne pleine. L'écart entre l'approximation et le LLR est particulièrement faible. La solution proposée est attractive du fait de sa faible complexité de mise en œuvre et de ses performances. Enfin, la prise en compte de l'impulsivité du canal est nécessaire car si un récepteur linéaire était utilisé avec la méthode d'apprentissage supervisé présentée précédemment, les performances seraient dégradées. En effet dans ce cas, le taux d'erreur binaire ne descend pas en dessous de 10 % comme l'illustre la courbe associé de la figure A.8 en trait pointillé.

Des conclusions similaires peuvent être tirées de simulations effectuées sur d'autres types de canaux, pour lesquels les écarts entre l'approximation et le LLR diffèrent mais reste limités.

Pour analyser plus précisément le lien entre le taux d'erreur binaire et le critère d'apprentissage, nous avons superposé sur la figure A.9 les contours du taux d'erreur pour le canal alpha-stable avec $\alpha = 1.4$ et $\gamma = 0.45$ en fonction des paramètres a et b de l'approximation L_{ab} et l'optimum du critère $\widehat{H}_\theta$ signalé par la croix blanche. Comme le confirme cette figure, l'optimum est bien dans la région de faible taux d'erreur. De plus en traçant en blanc le bord de la région pour laquelle le critère $\widehat{H}_\theta$ reste proche de son optimum, nous constatons une certaine robustesse sur la recherche du minimum. En effet cette région de faible critère reste dans la région de faible taux d'erreur.

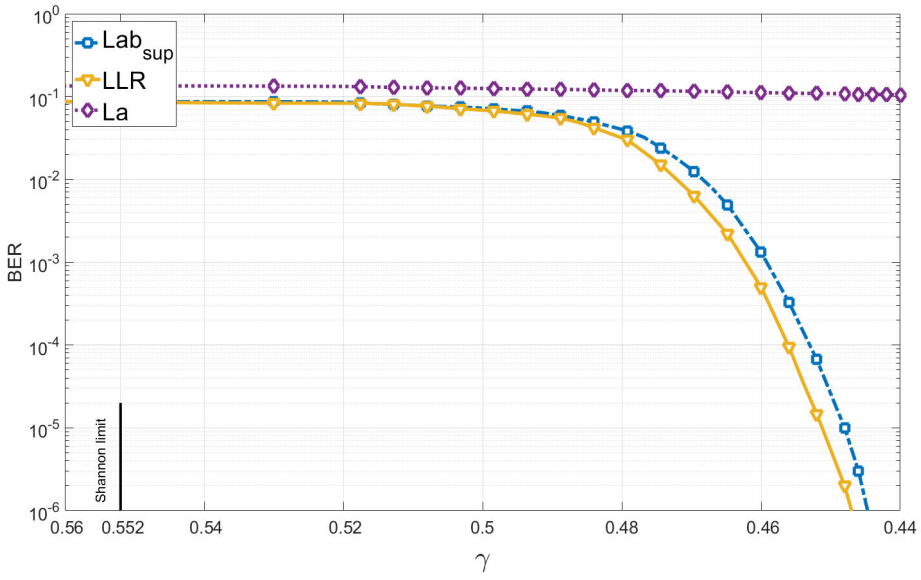


Fig. A.8. Taux d'erreur après décodage utilisant l'approximation du LLR, le LLR ou une approximation linéaire avec un canal alpha-stable d'exposant caractéristique $\alpha = 1.4$.

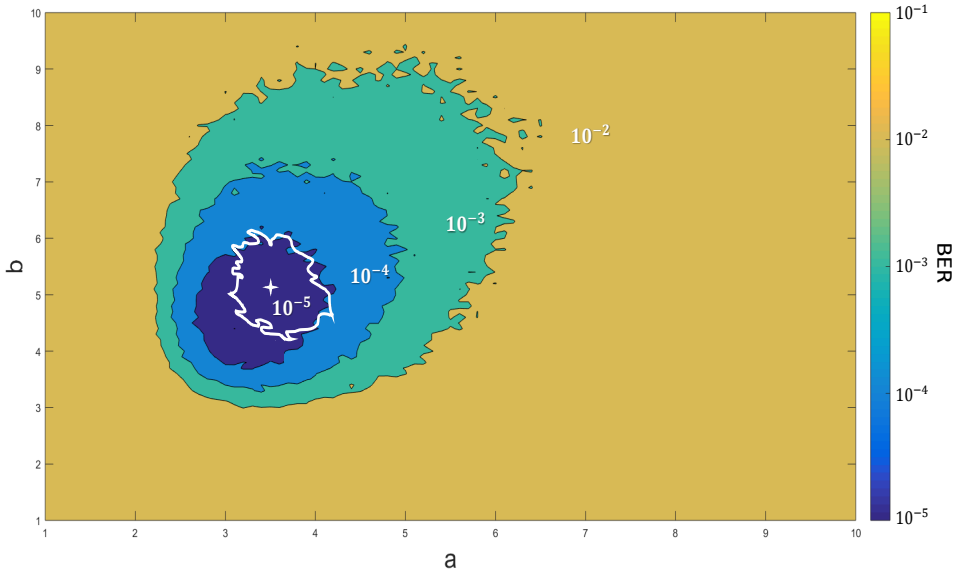


Fig. A.9. Taux d'erreur binaire et critère $\widehat{H}_\theta$.

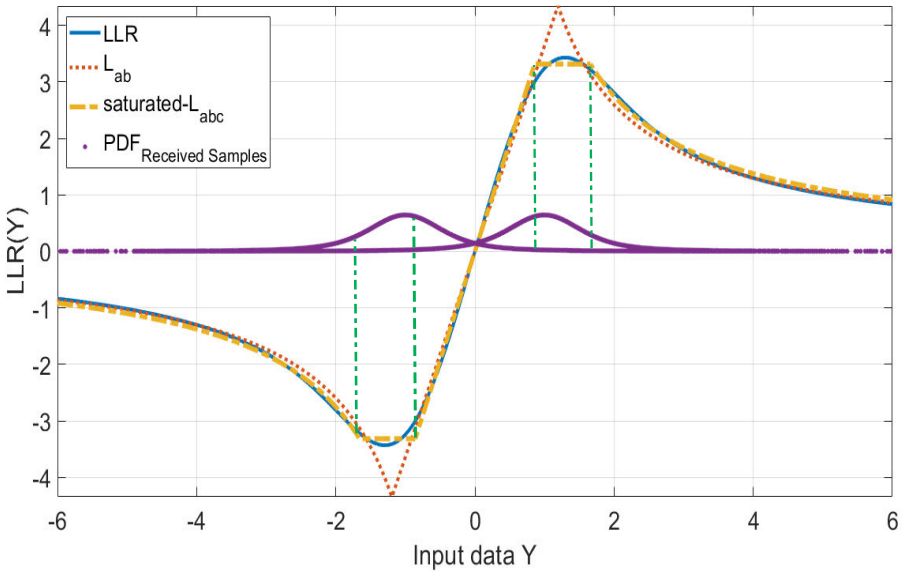


Fig. A.10. Comparaison des familles L_{ab} et L_{abc}.

A.4.3 Une nouvelle famille d'approximations

Une analyse détaillée de l'approximation L_{ab} nous a amené à ajouter un nouveau paramètre c pour modéliser le passage de la partie linéaire à la partie asymptotique. La nouvelle famille L_{abc} d'approximations est définie par

$$L_{abc}(y) = \begin{cases} ay & \text{si } -c/a < y < c/a, \\ c & \text{si } c/a < y < bc, \\ b/y & \text{sinon.} \end{cases} \tag{A.16}$$

La figure A.10 permet de comparer visuellement l'adéquation entre le LLR, la meilleure fonction de type L_{ab} et la meilleure de type L_{abc}. L'ajout de ce paramètre c permet de mieux prendre en compte la transition entre les deux régimes linéaire et asymptotique. Cela est particulièrement important car cette zone représentent environ 40 % des sorties du canal qui sont par conséquent mal pris en compte par l'approximation — la densité de la sortie du canal est aussi représentée sur cette figure.

Comme le montre la figure A.11, l'écart de performance entre le LLR et l'optimum des familles L_{ab} est réduit par l'introduction de ce paramètre c aussi bien pour le taux d'erreur binaire que pour le taux d'erreur paquet. Les conditions sont identiques aux simulations de la figure A.8.

Nous resterons malgré cela sur la famille L_{ab} par la suite pour rester sur une optimisation sur deux paramètres.

A.4.4 Cas de séquence d'apprentissage courtes

Lorsque la séquence d'apprentissage devient plus courte, deux phénomènes apparaissent. Le premier concerne l'estimation de H_θ par $\widehat{H}_\theta$ qui devient moins précise

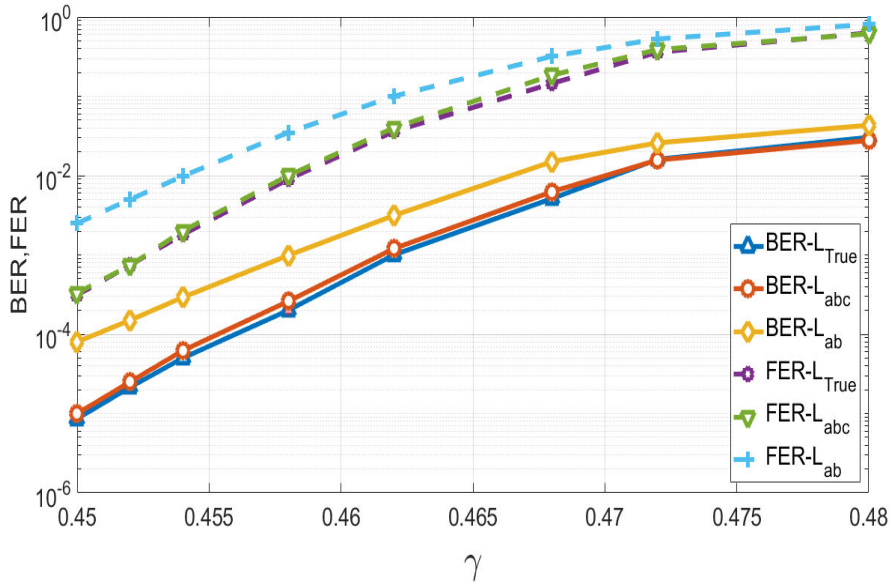


Fig. A.11. Taux d'erreur pour les familles d'approximations L_{ab} et L_{abc}.

car reposant sur une moyenne sur un plus petit nombre de termes. Il y a donc une certaine volatilité des paramètres θ obtenus par optimisation comme l'indique le tableau A.1. On remarque que la moyenne de a sur l'ensemble des tests est peu affectée mais qu'elle subit une grande variabilité lorsque la séquence d'apprentissage est courte. Le cas du paramètre b est pire : et la moyenne et la variabilité sont sujetes à la taille de la séquence.

Le deuxième phénomène est lié à l'apprentissage supervisé dans un environnement impulsif. Il est possible que la séquence d'apprentissage ne soit pas affectée par des impulsions de bruit alors que le paquet suivant le soit ou inversement. Il y a donc une mauvaise correspondance entre l'état du canal appris par le récepteur et l'état du canal lors de la transmission du mot de code. Ce phénomène se produit d'autant plus souvent que la séquence d'apprentissage est courte. De plus, et cela peut paraître contre-intuitif, ce phénomène apparaît aussi plus souvent lorsque le bruit est peu impulsif, c'est-à-dire dans le cas alpha-stable lorsque α est plus proche de 2 que de 1.

En prenant en compte ces deux phénomènes, les performances du système pour les courtes séquences d'apprentissage sont fortement dégradées comme le montre la figure A.12. Les simulations se font dans les même conditions que précédemment mais la longueur est réduite. Nous remarquons qu'ici, la longueur 1200 semble être un seuil, en dessus de celui-ci, le décodage subit plus tôt un plancher d'erreur.

A.5 Apprentissage non-supervisé du LLR

L'apprentissage non supervisé permet de répondre au problème de l'inadéquation entre le canal appris et le canal réel qui a été soulevé précédemment. De plus, en supprimant la nécessité d'une séquence d'apprentissage, le débit s'en trouve augmenté. Nous proposons un tel schéma dans cette partie.

			μ_a	σ_a	μ_b	σ_b
$\alpha = 1.4$	$\gamma = 0.43$	$\text{Sup}_{\text{LS}=20000}$	3.73	0.07	5.10	0.12
		$\text{Sup}_{\text{LS}=1200}$	3.77	0.35	5.18	0.58
		$\text{Sup}_{\text{LS}=900}$	3.77	0.40	5.28	0.71
	$\gamma = 0.45$	$\text{Sup}_{\text{LS}=20000}$	3.57	0.07	5.06	0.13
		$\text{Sup}_{\text{LS}=1200}$	3.60	0.32	5.15	0.58
		$\text{Sup}_{\text{LS}=900}$	3.61	0.37	5.16	0.66
$\alpha = 1.8$	$\gamma = 0.53$	$\text{Sup}_{\text{LS}=20000}$	3.25	0.05	7.59	0.28
		$\text{Sup}_{\text{LS}=1200}$	3.27	0.24	8.50	14.48
		$\text{Sup}_{\text{LS}=900}$	3.27	0.27	11.72	46.15
	$\gamma = 0.55$	$\text{Sup}_{\text{LS}=20000}$	3.05	0.05	7.62	0.28
		$\text{Sup}_{\text{LS}=1200}$	3.07	0.22	7.97	1.54
		$\text{Sup}_{\text{LS}=900}$	3.07	0.26	10.73	30.41

TABLE A.1 – Variabilité des paramètres a et b en fonction de la longueur de la séquence d'apprentissage.

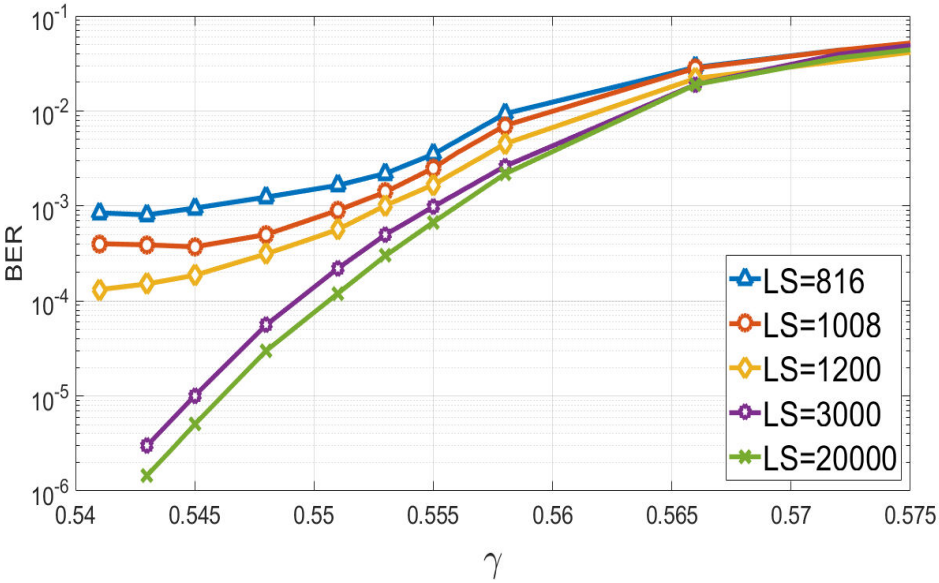


Fig. A.12. Performances de l'apprentissage supervisé pour différentes longueurs de séquences d'apprentissage.

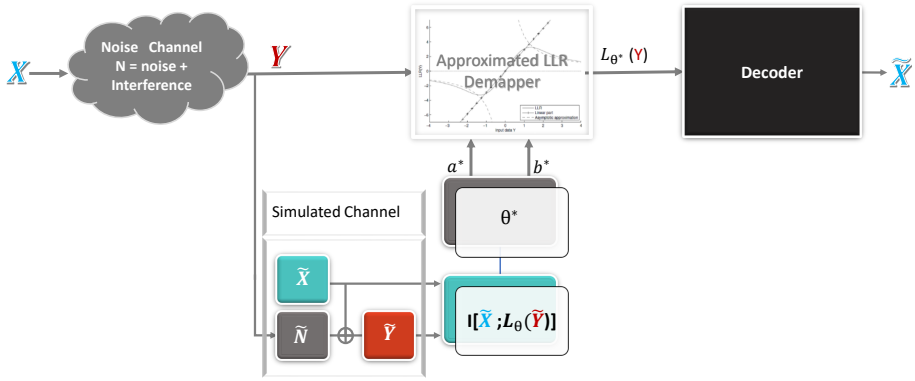


Fig. A.13. Synopsis de l'apprentissage non-supervisé.

A.5.1 Schéma de l'apprentissage non-supervisé

Notre méthode cherche à reconstruire un canal équivalent avec une source connue du récepteur. Pour cela, le schéma de la figure A.13 est utilisé. Des échantillons de bruit sont d'abord extraits grâce à une détection par le signe. Ensuite, ces échantillons de bruit sont injectés dans un canal simulé dont la pseudo-source est générée par le récepteur. Grâce à cela, la méthode précédente dite supervisée est utilisée pour estimer les paramètres θ de l'approximation du LLR.

Même si certaines décisions sont erronées, l'estimation obtenue pour a et b peuvent être suffisamment proche de l'optimum pour permettre un décodage performant. Deux cas sont à considérer : soit les codes sont longs soit ils sont courts. Les parties suivantes développent ces cas.

A.5.2 Cas des paquets longs

La figure A.14 compare les paramètres a et b dans le cadre d'un apprentissage supervisé et d'un apprentissage non-supervisé avec une taille de séquence ou de mot de code de 20000. Le paramètre a est peu affecté par les erreurs de décision du schéma non-supervisé contrairement au second paramètre b. Toutefois les ordres de grandeurs de ce paramètre ne sont pas aberrants ce qui présage d'une dégradation limitée des performances.

Comme le montre la figure A.15, en reprenant les mêmes paramètres de simulations que pour le cas supervisé, les performances de notre système non supervisé restent proches du système supervisé sur de longues séquences de taille 20000. On remarque que ces performances sont même supérieures à un système supervisé avec une séquence d'apprentissage de longueur 1200. Naturellement, l'écart de performance entre le supervisé et le non-supervisé dépend du canal mais reste cependant limité.

A.5.3 Cas des paquets courts

Le cas des paquets courts en non-supervisé est plus complexe à étudier car plusieurs phénomènes interviennent. Comme le montre la courbe légendée " USD " de la figure A.16, l'apprentissage ne parvient pas à approcher correctement le LLR

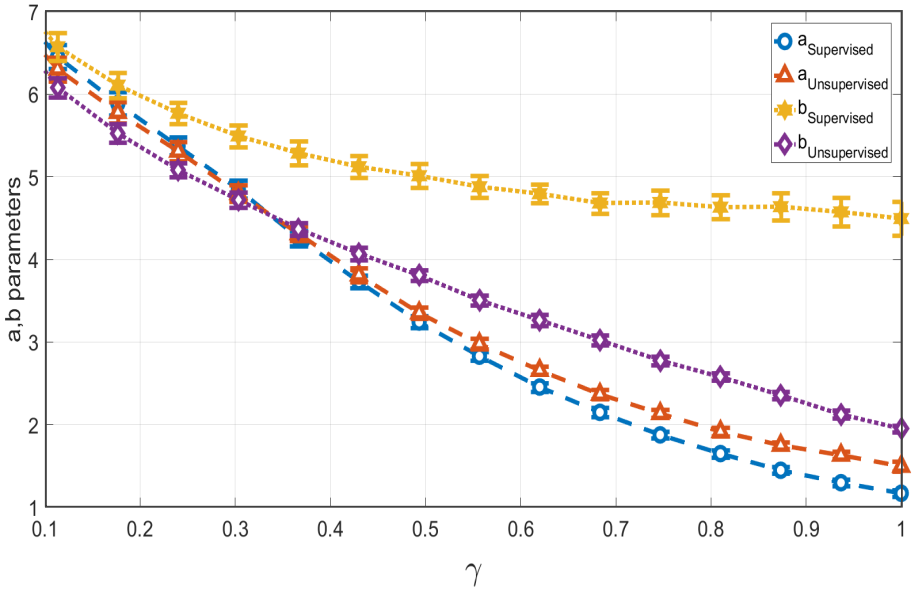


Fig. A.14. Estimation des paramètres en supervisé et non-supervisé.

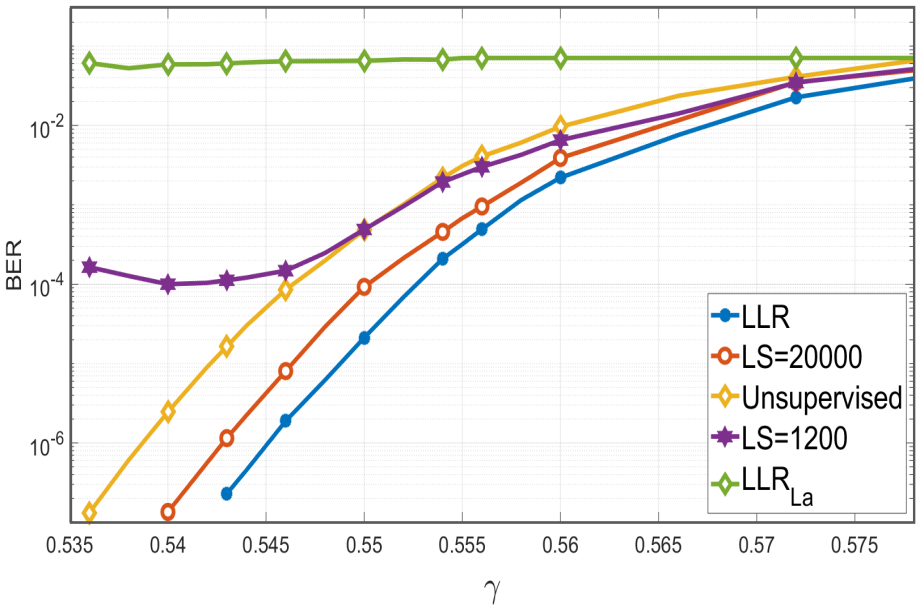


Fig. A.15. Performances du schéma non-supervisé avec des paquets longs de 20000 bits pour le canal alpha-stable avec $\alpha = 1.8$.

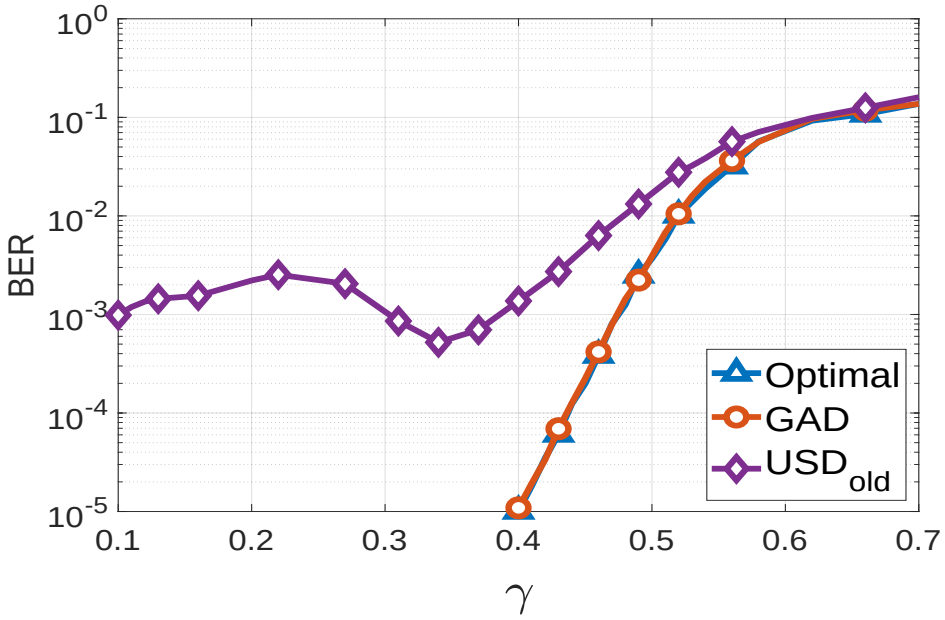


Fig. A.16. Performances du schéma non-supervisé avec des paquets courts de 408 bits sur un canal alpha-stable de paramètre $\alpha = 1.8$.

représentée par la courbe légendée " Optimal". De plus, un sursaut d'erreurs apparaît quand bien même le bruit est plus faible.

Pour comprendre les raisons de ce comportement, un schéma qualifié de GAD pour Genie Aided Decoder est utilisé. Il utilise le même schéma que l'apprentissage non-supervisé mais où l'entrée estimée $\widetilde{X}$ est remplacée par les bits réellement émis X. Les performances du schéma GAD sont quasiment indiscernables de l'optimal.

Une étude fine permet de constater que les paramètres a et b deviennent hautement variables en fonction des conditions du canal. Par conséquent, le LLR est rarement correctement approché par la fonction L_{ab}. Cette variabilité s'explique par une dégénérescence du processus optimisation. En effet, le critère (A.15) exprime un compromis entre quatre termes : l'un tend à faire augmenter le paramètre a tandis que l'autre tend à l'annuler, les deux autres termes jouent des rôles identiques avec le paramètre b. Il y a dégénérescence si l'un de ces quatre termes disparaît faute d'échantillon de bruit associé. Dans ce cas, l'un des paramètres a ou b devient soit infini soit nul ce que rend l'approximation du LLR inopérante et explique la dégradation des performances.

Nous avons quantifié théoriquement le risque d'apparition d'une dégénérescence. La figure A.17 en trace les valeurs. L'apprentissage non-supervisé présente un risque accru de dégénérescence par rapport au GAD. Et cette différence est la plus grande justement dans la zone de bruit où un sursaut d'erreur est observée dans la figure A.16.

Pour corriger ces problèmes de l'apprentissage non-supervisé pour les paquets courts, nous avons alors proposé une solution simple. Le résultat se retrouve sur la figure A.18 sous la légende USD$_{\text{new}}$. La solution proposée permet de retrouver des performances proche de l'optimum du LLR.

Cette correction consiste à symétriser la distribution des échantillons de bruit

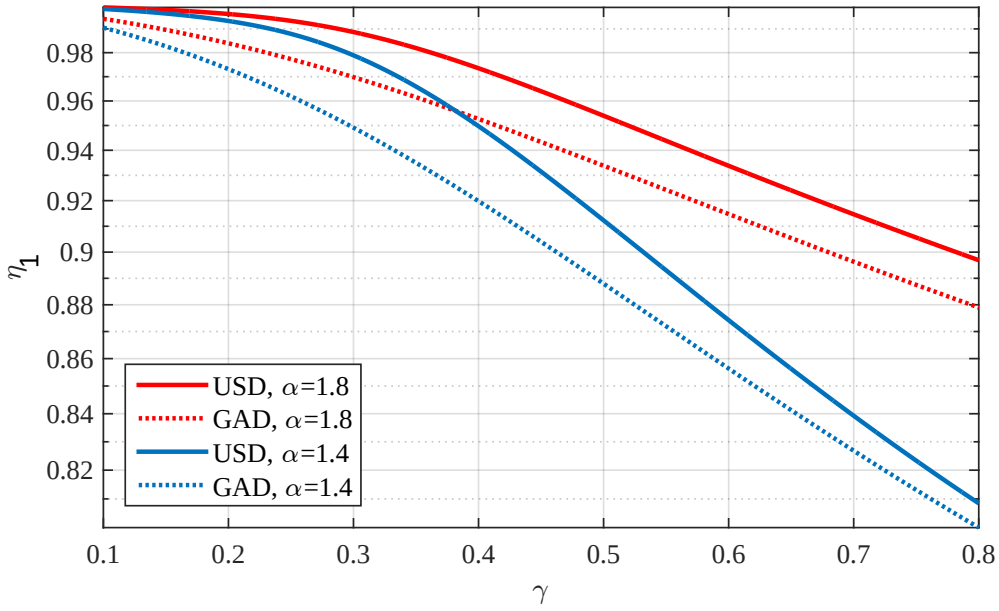


Fig. A.17. Risque de dégénérescence en fonction de γ pour un canal alpha-stable.

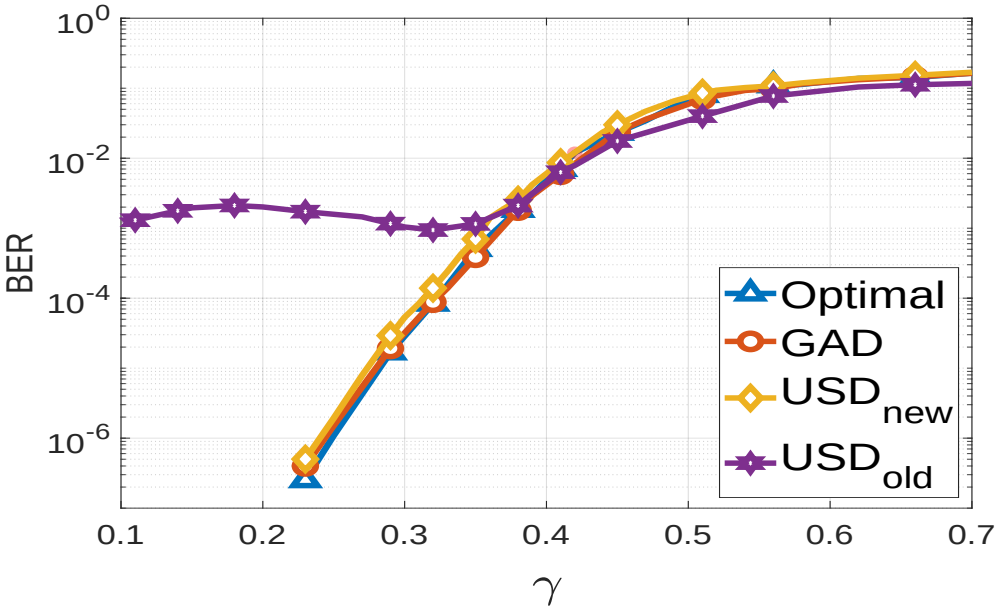


Fig. A.18. Performances de l'apprentissage supervisée corrigé.

reconstitués puis à introduire un terme de régulation sur le paramètre a critère $\widehat{H}_\theta$ qui est équivalent à ajouter un échantillon de bruit faible est négatif. Cette solution a quasiment aucun coût de calcul supplémentaire par rapport à la version originale et pourtant ses performances sont satisfaisantes.

A.5.4 Conclusion et perspectives

Nous avons montré dans nos travaux qu'un simple module entre la sortie du canal et l'entrée du décodeur de canal permet de combattre efficacement le bruit et les interférences qui perturbent les communications point à point dans un réseau. Ce module consiste en un fonction approchant le calcul du rapport de vraisemblance qui est rarement accessible et qui nécessite la connaissance de l'état du canal. Or, grâce à un critère judicieux, il est possible de chercher cette fonction d'approximation efficacement soit par un apprentissage supervisé soit par une approche non-supervisé. Nous avons montré qu'il devient même possible d'utiliser un tel schéma pour des communications par paquets courts sans pertes rédhibitoires de performances.

Les perspectives de nos travaux sont nombreux. La première, à court terme, consiste à étendre nos résultats pour des modulations d'ordre supérieur comme les modulations d'amplitude en quadrature. Dans ce cas les rapports de vraisemblances dépendent de la position des bits dans le codage binaire des symboles. Il faut donc proposer des familles de fonctions approchantes qui prennent en compte cette position. Toutefois, nos travaux préliminaires montrent qu'il est possible de revenir à un mélange de nos fonctions L_{ab} pour limiter la taille des paramètres.

À plus long terme, comme le critère proposé est très proche d'une approche variationnelle de l'apprentissage, il peut être judicieux de l'intégrer directement dans le décodage canal pour qu'il y ait un échange d'information entre le code et le calcul du LLR. Cela est d'autant plus pertinent que dans le cas non-supervisé, la limitation du nombre d'erreurs de décision améliore la qualité des échantillons de bruit obtenus et donc de l'estimation des paramètres. Un couplage simple entre le décodage de canal et l'approximation a déjà été testé sans résultat significatif ce qui incite à rechercher une approche plus fine, par exemple par l'utilisation d'un couplage de graphes factoriels entre celui du code LDPC et celui de l'approximation du LLR.

Enfin, nous avons toujours considéré des approximations qui appartiennent à une certaine famille construite *a priori*. Le critère étant issu de la théorie de l'information, il serait profitable de rendre l'approche encore plus générique en choisissant une fonction selon une approche par maximum d'entropie à la Jaynes [Jay03] et en la combinant avec le décodage de canal directement pour faire un décodeur basé sur le principe du maximum d'entropie.

Bibliographie

[AALM17a] O. Alhussein, I. Ahmed, J. Liang, and S. Muhaidat. Unified analysis of diversity reception in the presence of impulsive noise. *IEEE Transactions on Vehicular Technology*, 66(2) :1408–1417, Feb 2017.

[AALM17b] O. Alhussein, I. Ahmed, J. Liang, and S. Muhaidat. Unified analysis of diversity reception in the presence of impulsive noise. *IEEE Transactions on Vehicular Technology*, 66(2) :1408–1417, February 2017.

[AB07] T. Aysal and K. Barner. Generalized mean-median filtering for robust frequency-selective applications. *IEEE Transactions on Signal Processing*, 2007.

[AGM+15] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash. Internet of things : A survey on enabling technologies, protocols, and applications. *IEEE Communications Surveys Tutorials*, 17(4) :2347–2376, Fourthquarter 2015.

[Ahm19] Asaad Ahmed Gad-Elrab Ahmed. Benefits and challenges of internet of things for telecommunication networks. In Mohammad Abdul Matin, editor, *Telecommunication Networks*, chapter 6. IntechOpen, Rijeka, 2019.

[AIH94] S. Ambike, J. Ilow, and D. Hatzinakos. Detection for binary transmission in a mixture of gaussian noise and impulsive noise modeled as an alpha-stable process. *IEEE Signal Processing Letters*, 1(3) :55–57, March 1994.

[And05] J. G. Andrews. Interference cancellation for cellular systems : a contemporary overview. *IEEE Wireless Communications*, 12(2) :19–29, April 2005.

[AP10] N. Andreadou and F. Pavlidou. Modeling the noise on the ofdm powerline communications system. *IEEE Transactions on Power Delivery*, 25(1) :150–157, Jan 2010.

[APDR77] N. M. Laird A. P. Dempster and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 1977.

[AWM14] A. Asadi, Q. Wang, and V. Mancuso. A survey on device-to-device communication in cellular networks. *IEEE Communications Surveys Tutorials*, 16(4) :1801–1819, Fourthquarter 2014.

[BALE+07] Torben Brack, M. Alles, Timo Lehnigk-Emden, F. Kienle, Norbert Wehn, N.E. L'Insalata, F. Rossi, Massimo Rovini, and Luca Fanucci. Low complexity ldpc code decoders for next generation standards. *2008 Design, Automation and Test in Europe*, 0 :69, 04 2007.

[BD03] L. Barnault and D. Declercq. Fast decoding algorithm for ldpc over gf(2/sup q/). In *Proceedings 2003 IEEE Information Theory Workshop (Cat. No.03EX674)*, pages 70–73, March 2003.

[BEK98] Franco Bassi, Paul Embrechts, and Maria Kafetzaki. *A Practical Guide to Heavy Tails.* Birkhauser Boston Inc., Cambridge, MA, USA, 1998.

[BMA10] H. Bhaskar, L. Mihaylova, and A. Achim. Video foreground detection based on symmetric alpha-stable mixture models. *IEEE Trans. Cir. and Sys. for Video Technol.*, 20(8) :1133–1138, August 2010.

[BMR+13] R. Bassoli, H. Marques, J. Rodriguez, K. W. Shum, and R. Tafazolli. Network coding theory : A survey. *IEEE Communications Surveys Tutorials*, 15(4) :1950–1978, Fourth 2013.

[BN10] N. C. Beaulieu and S. Niranjayan. Uwb receiver designs based on a gaussian-laplacian noise-plus-mai model. *IEEE Transactions on Communications*, 58(3) :997–1006, March 2010.

[BRB93] K. L. Blackard, T. S. Rappaport, and C. W. Bostian. Measurements and models of radio frequency impulsive noise for indoor wireless communications. *IEEE Journal on Selected Areas in Communications*, 11(7) :991–1001, Sep. 1993.

[BSE10] A. Skrondal B. S. Everitt. *The Cambridge dictionary of Statistics.* Fourth Edition Cambridge university press, 2010.

[BSF08] N. C. Beaulieu, Hua Shao, and J. Fiorina. P-order metric UWB receiver structures with superior performance. *IEEE Transactions on Communications*, 56(10) :1666–1676, October 2008.

[BY09] N. C. Beaulieu and D. J. Young. Designing time-hopping ultrawide bandwidth receivers for multiuser interference environments. *Proceedings of the IEEE*, 97(2) :255–284, Feb 2009.

[CAG08] V. Chandrasekhar, J. G. Andrews, and A. Gatherer. Femtocell networks : a survey. *IEEE Communications Magazine*, 46(9) :59–67, Sep. 2008.

[Car10] P. Cardieri. Modeling interference in wireless Ad Hoc networks. *IEEE Communications Surveys Tutorials*, 12(4) :551–572, Fourth 2010.

[CBV+09] N. Czink, B. Bandemer, G. Vazquez-Vilar, L. Jalloul, C. Oestges, and A. Paulraj. Spatial separation of multi-user mimo channels. In *2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1059–1063, Sep. 2009.

[CGE+09] A. Chopra, K. Gulati, B. L. Evans, K. R. Tinsley, and C. Sreerama. Performance bounds of mimo receivers in the presence of radio frequency interference. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2817–2820, April 2009.

[Chu05] T. C. Chuah. Robust iterative decoding of turbo codes in heavy-tailed noise. *IEE Proceedings - Communications*, 152(1) :29–38, Feb 2005.

[CL97] R. C. Cheng and W. Liu. A continuous representation of the family of stable law distributions. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, pages 137–145, 1997.

[CLCC11] S. Cheng, S. Lien, F. Chu, and K. Chen. On exploiting cognitive radio to mitigate interference in macro/femto heterogeneous networks. *IEEE Wireless Communications*, 18(3) :40–47, June 2011.

[Coo93] John W. Cook. Wideband impulsive noise survey of the access network. 1993.

[Cos83] M. Costa. Writing on dirty paper (corresp.). *IEEE Transactions on Information Theory*, 29(3) :439–441, May 1983.

[CPH04] M. Chitre, J. Potter, and O. S. Heng. Underwater acoustic channel characterisation for medium-range shallow water communications. In *Oceans '04 MTS/IEEE Techno-Ocean '04 (IEEE Cat. No.04CH37600)*, volume 1, pages 40–45 Vol.1, Nov 2004.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, 2 edition, 2006.

[CTB98] Mark E. Crovella, Murad S. Taqqu, and Azer Bestavros. *A Practical Guide to Heavy Tails*. Birkhauser Boston Inc., Cambridge, MA, USA, 1998.

[DF07] D. Declercq and M. Fossorier. Decoding algorithms for nonbinary ldpc codes over gf(q). *IEEE Transactions on Communications*, 55(4) :633–643, April 2007.

[DGCG14] V. Dimanche, A. Goupil, L. Clavier, and G. Gelle. On detection method for soft iterative decoding in the presence of impulsive interference. *IEEE Communications Letters*, 18(6) :945–948, June 2014.

[DHB06] J. Dielissen, A. Hekstra, and V. Berg. Low cost ldpc decoder for dvb-s2. In *Proceedings of the Design Automation Test in Europe Conference*, volume 2, pages 1–6, March 2006.

[Dig09] Digi. *Efficient data transfer over cellular networks*. White Paper, 2009.

[DM98] M. C. Davey and D. J. C. MacKay. Low density parity check codes over gf(q). In *1998 Information Theory Workshop (Cat. No.98EX131)*, pages 70–71, June 1998.

[DuM73] William H. DuMouchel. On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *Ann. Statist.*, 1(5) :948–957, 09 1973.

[Dur10] R. Durrett. *Probability : Theory and Examples*. 4th edition Cambridge Series in Statistical and Probabilistic Mathematics, 2010.

[DYZB03] Duc Son Pham, Yee Hong Leung, A. Zoubir, and R. Brcic. On the computational aspect of robust multiuser detection. In *Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology (IEEE Cat. No.03EX795)*, pages 22–25, Dec 2003.

[ECD08] T. Erseghe, V. Cellini, and G. Dona. On uwb impulse radio receivers derived by modeling mai as a gaussian mixture process. *IEEE Transactions on Wireless Communications*, 7(6) :2388–2396, June 2008.

[ECdF+17] M. Egan, L. Clavier, M. de Freitas, L. Dorville, J.M. Gorce, and A. Savard. Wireless communication in dynamic interference. In *IEEE GLOBECOM*, Singapore, December 2017.

[ECZ+18] Malcolm Egan, Laurent Clavier, Ce Zheng, Mauro de Freitas, and Jean-Marie Gorce. Dynamic interference for uplink scma in large-scale wireless networks without coordination. *EURASIP Journal on Wireless Communications and Networking*, 2018(1) :213, Aug 2018.

[EPH13] O. El Ayach, S. W. Peters, and R. W. Heath. The practical challenges of interference alignment. *IEEE Wireless Communications*, 20(1) :35–42, February 2013.

[ETV99] T. Etzion, A. Trachtenberg, and A. Vardy. Which codes have cycle-free tanner graphs ? *IEEE Transactions on Information Theory*, 45(6) :2173–2181, Sep. 1999.

[FC09] D. Fertonani and G. Colavolpe. A robust metric for soft-output detection in the presence of class-A noise. *IEEE Transactions on Communications*, 57(1) :36–40, January 2009.

[Fel70] W. Feller. *An Introduction to Probability Theory and its Applications, volume I of Wiley Series in Probability and Mathematical Statistics*. John Wiley & Sons, New York, 1970.

[FGCE15] N. Farsad, W. Guo, C. Chae, and A. Eckford. Stable distributions as noise models for molecular communication. In *2015 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, Dec 2015.

[Fio06] J. Fiorina. Wlc28-2 : A simple ir-uwb receiver adapted to multi-user interferences. In *IEEE Globecom 2006*, pages 1–4, Nov 2006.

[FR71] Eugene F. Fama and Richard Roll. Parameter estimates for symmetric stable distributions. *Journal of the American Statistical Association*, 66(334) :331–338, 1971.

[Gal63] Robert G. Gallager. *Low density parity check codes*. M.I.T press, 1963.

[GC12] W. Gu and L. Clavier. Decoding metric study for turbo codes in very impulsive environment. *IEEE Communications Letters*, 16(2) :256–258, February 2012.

[GCA+10] H. E. Ghannudi, L. Clavier, N. Azzaoui, F. Septier, and P. a. Rolland. α-stable interference modeling and cauchy receiver for an IR-UWB Ad Hoc network. *IEEE Transactions on Communications*, 58(6) :1748–1757, June 2010.

[GCGD07] A. Goupil, M. Colas, G. Gelle, and D. Declercq. Fft-based bp decoding of general ldpc codes over abelian groups. *IEEE Transactions on Communications*, 55(4) :644–649, April 2007.

[GDK06] N. Guney, H. Delic, and M. Koca. Robust detection of ultra-wideband signals in non-gaussian noise. *IEEE Transactions on Microwave Theory and Techniques*, 54(4) :1724–1730, June 2006.

[GEAT10] K. Gulati, B. L. Evans, J. G. Andrews, and K. R. Tinsley. Statistics of co-channel interference in a field of poisson and poisson-poisson clustered interferers. *IEEE Transactions on Signal Processing*, 58(12) :6207–6222, Dec 2010.

[GH97] Jaap Geluk and Laurens Haan. Stable probability distributions and their domains of attraction. *Tinbergen Institute, Tinbergen Institute Discussion Papers*, 20, 01 1997.

[Gol05] A. Goldsmith. *Wireless communications*. Cambridge university press, 2005.

[Gon97] J. G. Gonzalez. *Robust techniques for wireless communications in non-Gaussian environments*. PhD thesis, Dept. Elec. Eng., Univ. of Delaware, Newark, DE,, 1997.

[GPA06] J. González, J. L. Paredes, and G. R. Arce. Zero-order statistics : A mathematical framework for the processing and characterization of very impulsive signals. *IEEE Transactions on Signal Processing*, 54 :3839–3851, 2006.

[GPC+12] W. Gu, G. Peters, L. Clavier, F. Septier, and I. Nevat. Receiver study for cooperative communications in convolved additive alpha-stable interference plus gaussian thermal noise. In *2012 International Symposium on Wireless Communication Systems (ISWCS)*, pages 451–455, Aug 2012.

[GSY09] X. Guo, H. Sun, and Yilong Lu. Joint suppression of radio frequency interference and lightning impulsive noise in hfswr. In *2009 International Radar Conference "Surveillance for a Safer World" (RADAR 2009)*, pages 1–6, Oct 2009.

[Hag04] J. Hagenauer. The exit chart - introduction to extrinsic information transfer in iterative processing. In *2004 12th European Signal Processing Conference*, pages 1541–1548, Sep. 2004.

[Hal66] Harry Hall. A new model for impulsive phenomena : Application to atmospheric noise communication channels. page 179, 08 1966.

[HEg07] P.A. Rolland H. El ghannudi, L. Clavier. "modélisation du mai d'un système ir-uwb transposè à 60 ghz pour des réseaux ad hoc" modélisation du mai d'un système ir-uwb transposè à 60 ghz pour des réseaux ad hoc. *GRETSI*, 2007.

[HGC06] T. R. Halford, A. J. Grant, and K. M. Chugg. Which codes have4-cycle-free tanner graphs ? *IEEE Transactions on Information Theory*, 52(9) :4219–4223, Sep. 2006.

[HH-56] *An investigation of atmospheric radio noise at very low frequencies*, volume 103. Proceedings of the IET, 1956.

[HLZ14] Y. Hou, R. Liu, and L. Zhao. A non-linear LLR approximation for LDPC decoding over impulsive noise channels. In *2014 IEEE/CIC International Conference on Communications in China (ICCC)*, pages 86–90, Oct 2014.

[HM12] J. Hoadley and P. Maveddat. Enabling small cell deployment with hetnet. *IEEE Wireless Communications*, 19(2) :4–5, April 2012.

[HSSL11] Y. Hao, Z. Shan, F. Shen, and D. Lv. Parameter estimation of alpha-stable distributions based on mcmc. In *2011 3rd International Conference on Advanced Computer Control*, pages 325–327, Jan 2011.

[IH98a] J. Ilow and D. Hatzinakos. Analytic alpha-stable noise modeling in a poisson field of interferers or scatterers. *IEEE Transactions on Signal Processing*, 46(6) :1601–1611, June 1998.

[IH98b] J. Ilow and D. Hatzinakos. Analytic alpha-stable noise modeling in a poisson field of interferers or scatterers. *IEEE Transactions on Signal Processing*, 46(6) :1601–1611, June 1998.

[III07] H. Ishikawa, M. Itami, and K. Itoh. A study on adaptive modulation of ofdm under middleton's class-a impulsive noise model. In *2007 Digest of Technical Papers International Conference on Consumer Electronics*, pages 1–2, Jan 2007.

[Jay03] E. T. Jaynes. *Probability Theory : The Logic of Science*. Cambridge University Press, 2003.

[JG12] Sang-Woon Jeon and Michael Gastpar. A survey on interference networks : Interference alignment and neutralization. *Entropy*, 14(10) :22. 1842–1863, 2012.

[JHMB05] Kyle Jamieson, Bret Hull, Allen Miu, and Hari Balakrishnan. Understanding the real-world performance of carrier sense. In *Proceedings of the 2005 ACM SIGCOMM Workshop on Experimental Approaches to Wireless Network Design and Analysis*, E-WIND '05, pages 52–57, New York, NY, USA, 2005. ACM.

[Joh08] S. J. Johnson. *Iterative error correction : turbo, low-density parity-check and repeat-accumulate codes*. Cambridge University Press, 2008.

[K60] G. Hagstroem K. Remarks on pareto distributions. *Scandinavian Actuarial Journal*, 1960(1-2) :59–71, 1960.

[Kas88] S. A. Kassam. *Signal Detection in Non-Gaussian Noise*. Springer-Verlag, 1988.

[KC04] M. Karkooti and J. R. Cavallaro. Semi-parallel reconfigurable architectures for real-time ldpc decoding. In *International Conference on Information Technology : Coding and Computing, 2004. Proceedings. ITCC 2004.*, volume 1, pages 579–585 Vol.1, April 2004.

[KFR98] E. E. Kuruoglu, W. J. Fitzgerald, and P. J. W. Rayner. Near optimal detection of signals in impulsive noise modeled with a symmetric /spl alpha/-stable distribution. *IEEE Communications Letters*, 2(10) :282–284, Oct 1998.

[KG95] K. J. Kerpez and A. M. Gottlieb. The error performance of digital subscriber lines in the presence of impulse noise. *IEEE Transactions on Communications*, 43(5) :1902–1905, May 1995.

[KH01] A. Karasaridis and D. Hatzinakos. Network heavy traffic modeling using alpha-stable self-similar processes. *IEEE Transaction on Communication*, 2001.

[Kou81] Ioannis A. Koutrouvelis. An iterative procedure for the estimation of the parameters of stable laws. *Communications in Statistics - Simulation and Computation*, 10(1) :17–28, 1981.

[KTH95] S. Kuramoto, F. Takemoto, and M. Hattori. A study of impulse noise caused by connecting or disconnecting coaxial cable between equipment. In *Proceedings of International Symposium on Electromagnetic Compatibility*, pages 501–506, Aug 1995.

[KTMP03] Koay Teong Beng, Tan Eng Teck, Mandar Chitre, and J. R. Potter. Estimating the spatial and temporal distribution of snapping shrimp using a portable, broadband 3-dimensional acoustic array. In *Oceans 2003. Celebrating the Past ... Teaming Toward the Future (IEEE Cat. No.03CH37492)*, volume 5, pages 2706–2713 Vol.5, Sep. 2003.

[Leg09] M. W. Legg. *Non-Gaussian and non-homogeneous Poisson models of snapping shrimp noise*. PhD thesis, Curtin University of Technology, 2009.

[Mau16] Robert G Maunder. A vision for 5g channel coding. *AccelerComm White Paper*, October 2016.

[McC86a] J. H. McCulloch. Simple consistent estimators of stable distribution parameters. *Communications on Satatistical Simulations*, pages 1109–1136, 1986.

[McC86b] J. Huston McCulloch. Simple consistent estimators of stable distribution parameters. *Communications in Statistics - Simulation and Computation*, 15(4) :1109–1136, 1986.

[MGC05] H. Meng, Y. L. Guan, and S. Chen. Modeling and analysis of noise effects on broadband power-line communications. *IEEE Transactions on Power Delivery*, 20(2) :630–637, April 2005.

[MGCG10a] H. Ben Mâad, A. Goupil, L. Clavier, and G. Gelle. Asymptotic performance of ldpc codes in impulsive non-gaussian channel. In *Signal Processing Advances in Wireless Communications (SPAWC), 2010 IEEE Eleventh International Workshop on*, pages 1–5, June 2010.

[MGCG10b] H. Ben Mâad, A. Goupil, L. Clavier, and G. Gelle. Robust clipping demapper for LDPC decoding in impulsive channel. In *2010 6th International Symposium on Turbo Codes Iterative Information Processing*, pages 231–235, Sept 2010.

[MGCG13] H. B. Maad, A. Goupil, L. Clavier, and G. Gelle. Clipping demapper for LDPC decoding in impulsive channel. *IEEE Communications Letters*, 17(5) :968–971, May 2013.

[Mid72a] D. Middleton. Performance of telecommunications systems in the spectraluse environment : Vii. interference scenarios and the canonical and quasi-canonical (first-order) probability models of class a interference. 1972.

[Mid72b] David Middleton. Statistical-physical models of urban radio-noise environments - part i : Foundations. 1972.

[Mid77] D. Middleton. Statistical-physical models of electromagnetic interference. *IEEE Transactions on Electromagnetic Compatibility*, EMC-19(3) :106–127, August 1977.

[Mid96] D. Middleton. *An Introduction to Statistical Communication Theory*, chapter 11. IEEE, 1996.

[MJLC15] Z. Mei, M. Johnston, S. Le, and L. Chen. Density evolution analysis of LDPC codes with different receivers on impulsive noise channels.

In *2015 IEEE/CIC International Conference on Communications in China (ICCC)*, pages 1–6, Nov 2015.

[MN96] D. J. C. MacKay and R. M. Neal. Near shannon limit performance of low density parity check codes. *Electronics Letters*, 32(18) :1645–, Aug 1996.

[MS02] M. M. Mansour and N. R. Shanbhag. Low-power vlsi decoder architectures for ldpc codes. In *Proceedings of the International Symposium on Low Power Electronics and Design*, pages 284–289, Aug 2002.

[MSG+18] Yasser Mestrah, Anne Savard, Alban Goupil, Laurent Clavier, and Guillaume Gellé. Blind estimation of an approximated likelihood ratio in impulsive environment. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 2018.

[MSG+19a] Yasser Mestrah, Anne Savard, Alban Goupil, Guillaume Gellé, and Laurent Clavier. Mesure indirecte des performances de llr approché. In *GRETSI*, 2019.

[MSG+19b] Yasser Mestrah, Anne Savard, Alban Goupil, Guillaume Gellé, and Laurent Clavier. On unsupervised llr estimation under finite block-length with unknown noise distribution. *To be published*, 2019.

[MSG+19c] Yasser Mestrah, Anne Savard, Alban Goupil, Guillaume Gellé, and Laurent Clavier. Robust and simple log-likelihood approximation for receiver design. In *IEEE Wireless Communications and Networking Conference (IEEE WCNC)*, 2019.

[MSG+19d] Yasser Mestrah, Anne Savard, Alban Goupil, Guillaume Gellé, and Laurent Clavier. An unsupervised llr estimation with unknown noise distribution. *EURASIP Journal on Wireless Communications and Networking*, 2019.

[MT76] J. H. Miller and J. B. Thomas. The detection of signals in impulsive noise modeled as a mixture process. *IEEE Trans. on Comm.*, 1976.

[Nah09] D. F. Nahimana. *Impact des multi trajets sur les performances des systèmes de navigation par satellite : Contribution à l'amélioration de la précision de localisation par modélisation bayèsienne.* PhD thesis, Ecole centrale Lille, 2009.

[NB08] S. Niranjayan and N. C. Beaulieu. A myriad filter detector for uwb multiuser communication. In *2008 IEEE International Conference on Communications*, pages 3918–3922, May 2008.

[NBST06] S. Nammi, D. K. Borah, C. Schneider, and R. Thoma. Effects of impulse noise on the performance of multidimensional parity check codes. In *IEEE Wireless Communications and Networking Conference, 2006. WCNC 2006.*, volume 4, pages 1966–1971, April 2006.

[NGK16] K. E. Nolan, W. Guibene, and M. Y. Kelly. An evaluation of low power wide area network technologies for the internet of things. In *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 439–444, Sep. 2016.

[NM65] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(1) :308–313, January 1965.

[NMLD02] N. Nedev, S. McLaughlin, D. Laurenson, and R. Daley. Data errors in adsl and shdsl systems due to impulse noise. In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages IV–4048–IV–4051, May 2002.

[Nol97] John P. Nolan. Numerical calculation of stable densities and distribution functions. communications in statistics. *Stochastic Models*, 1997.

[NS95] Chrysostomos L. Nikias and Min Shao. *Signal processing with alpha-stable distributions and applications.* Wiley-Interscience, New York, NY, USA, 1995.

[PCG+06] P. C. Pinto, C. Chong, A. Giorgetti, M. Chiani, and M. Z. Win. Narrowband communication in a poisson field of ultrawideband interferers. In *2006 IEEE International Conference on Ultra-Wideband*, pages 387–392, Sep. 2006.

[PCZ+11] X. Peng, Z. Chen, X. Zhao, D. Zhou, and S. Goto. A 115mw 1gbps qc-ldpc decoder asic for wimax in 65nm cmos. In *IEEE Asian Solid-State Circuits Conference 2011*, pages 317–320, Nov 2011.

[Per06] R. J. Igual Perez. *Platform Hardware/software for the energy optimization in a node of wireless sensor networks.* PhD thesis, Lille 1, 2006.

[PFD08] Charly Poulliat, Marc Fossorier, and David Declercq. Design of regular (2,dc)-ldpc codes over gf(q) using their binary images. *Communications, IEEE Transactions on*, 56 :1626 – 1635, 11 2008.

[P.N18] John P.Nolan. *Stable Distributions models for heavy Tailed Data.* American University, 2018.

[PPV10] Y. Polyanskiy, H. V. Poor, and S. Verdu. Channel coding rate in the finite blocklength regime. *IEEE Transactions on Information Theory*, 56(5) :2307–2359, May 2010.

[PW10a] P. C. Pinto and M. Z. Win. Communication in a poisson field of interferers–part I : Interference distribution and error probability. *IEEE Transactions on Wireless Communications*, 9(7) :2176–2186, July 2010.

[PW10b] P. C. Pinto and M. Z. Win. Communication in a poisson field of interferers-part II : Channel capacity and interference spectrum. *IEEE Transactions on Wireless Communications*, 9(7) :2187–2195, July 2010.

[RCS+11] C. Roth, A. Cevrero, C. Studer, Y. Leblebici, and A. Burg. Area, throughput, and energy-efficiency trade-offs in the vlsi implementation of ldpc decoders. In *2011 IEEE International Symposium of Circuits and Systems (ISCAS)*, pages 1772–1775, May 2011.

[RSU01a] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke. Design of capacity-approaching irregular low-density parity-check codes. *IEEE Transactions on Information Theory*, 47(2) :619–637, Feb 2001.

[RSU01b] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke. Design of capacity-approaching irregular low-density parity-check codes. *IEEE Transactions on Information Theory*, 47(2) :619–637, Feb 2001.

[RU01] T. J. Richardson and R. L. Urbanke. The capacity of low-density parity-check codes under message-passing decoding. *IEEE Transactions on Information Theory*, 47(2) :599–618, Feb 2001.

[RU08] T. J. Richardson and R. Urbkane. *Modern Coding Theory*. Cambridge University Press, 2008.

[SAC04] M. G. Sanchez, A. V. Alejos, and I. Cuinas. Urban wide-band measurement of the umts electromagnetic environment. *IEEE Transactions on Vehicular Technology*, 53(4) :1014–1022, July 2004.

[Sav08] V. Savin. Min-max decoding for non binary ldpc codes. In *2008 IEEE International Symposium on Information Theory*, pages 960–964, July 2008.

[SD99] Seoyoung Lee and J. Dickerson. Least l/sub p/-norm interference suppression for ds/cdma systems in non-gaussian impulsive channels. In *1999 IEEE International Conference on Communications (Cat. No. 99CH36311)*, volume 2, pages 907–911 vol.2, June 1999.

[SFRU01] Sae-Young Chung, G. D. Forney, T. J. Richardson, and R. Urbanke. On the design of low-density parity-check codes within 0.0045 db of the shannon limit. *IEEE Communications Letters*, 5(2) :58–60, Feb 2001.

[SH13] K. A. Saaifan and W. Henkel. Decision boundary evaluation of optimum and suboptimum detectors in class-a interference. *IEEE Transactions on Communications*, 61(1) :197–205, January 2013.

[Sha48] Claude Elwood Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3) :379–423, 1948.

[SK74] B. W. Stuck and B. Kleiner. A statistical analysis of telephone noise. *The Bell System Technical Journal*, 53(7) :1263–1320, Sep. 1974.

[SKB+13] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi. Non-orthogonal multiple access (noma) for cellular future radio access. In *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, pages 1–5, June 2013.

[SM77] A. Spaulding and D. Middleton. Optimum reception in an impulsive interference environment - part I : Coherent detection. *IEEE Transactions on Communications*, 25(9) :910–923, September 1977.

[SME10] T. S. Shehata, I. Marsland, and M. El-Tanany. A novel framework for signal detection in alpha-stable interference. In *2010 IEEE 71st Vehicular Technology Conference*, pages 1–5, May 2010.

[SME12] T. S. Saleh, I. Marsland, and M. El-Tanany. Suboptimal detectors for alpha-stable noise : Simplifying design and improving performance. *IEEE Transactions on Communications*, 60(10) :2982–2989, October 2012.

[SMET12] T. S. Saleh, I. Marsland, and M. El-Tanany. Simplified LLR-based viterbi decoder for convolutional codes in symmetric alpha-stable noise. In *2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 1–4, April 2012.

[SN93] Min Shao and C. L. Nikias. Signal detection in impulsive noise based on stable distributions. In *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, pages 218–222 vol.1, Nov 1993.

[Sou92a] E. S. Sousa. Performance of a spread spectrum packet radio network link in a poisson field of interferers. *IEEE Transactions on Information Theory*, 38(6) :1743–1754, Nov 1992.

[Sou92b] E. S. Sousa. Performance of a spread spectrum packet radio network link in a poisson field of interferers. *IEEE Transactions on Information Theory*, 38(6) :1743–1754, Nov 1992.

[Spa81] A. D. Spaulding. Optimum threshold signal detection in broadband impulsive noise employing both time and spatial sampling. *IEEE trans. Communication*, COM-29, 1981.

[ST94] G. Samorodnitsky and M. Taqqu. *Stable Non-Gaussian Random Processes : Stochastic Models with Infinite Variance*. Chapman & Hall, New York, 1994.

[SUR00] Sae-Young Chung, R. Urbanke, and T. J. Richardson. Gaussian approximation for sum-product decoding of low-density parity-check codes. In *2000 IEEE International Symposium on Information Theory (Cat. No.00CH37060)*, pages 318–, June 2000.

[SYH17] Rashmi Sinha, Wei Yiqiao, and Seung-Hoon Hwang. A survey on lpwa technology : Lora and nb-iot. *ICT Express*, 3, 03 2017.

[Tan81] R. Tanner. A recursive approach to low complexity codes. *IEEE Transactions on Information Theory*, 27(5) :533–547, Sep. 1981.

[ten01] S. ten Brink. Convergence behavior of iteratively decoded parallel concatenated codes. *IEEE Transactions on Communications*, 49(10) :1727–1737, Oct 2001.

[tKA04] S. ten Brink, G. Kramer, and A. Ashikhmin. Design of low-density parity-check codes for modulation and detection. *IEEE Transactions on Communications*, 52(4) :670–678, April 2004.

[VA89] Mahesh K. Varanasi and Behnaam Aazhang. Parametric generalized gaussian density estimation. In *J. Acoust. Soc. Am.*, 1989.

[Vas84] K. Vastola. Threshold detection in narrow-band non-gaussian noise. *IEEE Transactions on Communications*, 32(2) :134–139, February 1984.

[Vas00] S. V. Vaseghi. *Advanced Digital Signal Processing and Noise Reduction.* John Wiley & Sons ltd, 2000.

[Vas06] S. V. Vaseghi. *Advanced Digital Signal Processing and Noise Reduction.* Third Edition John Wiley & Sons Ltd, 2006.

[VLN+17] B. Vejlgaard, M. Lauridsen, H. Nguyen, I. Z. Kovacs, P. Mogensen, and M. Sorensen. Coverage and capacity analysis of sigfox, lora, gprs, and nb-iot. In *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, pages 1–5, June 2017.

[WA12] Steven Weber and Jeffrey G. Andrews. Transmission capacity of wireless networks. *CoRR*, abs/1201.0662, 2012.

[WAYd07] S. P. Weber, J. G. Andrews, X. Yang, and G. de Veciana. Transmission capacity of wireless ad hoc networks with successive interference cancellation. *IEEE Transactions on Information Theory*, 53(8) :2799–2814, Aug 2007.

[WLF14] C. Wang, X. Liu, and B. Fan. Estimation method for weak sinusoidal amplitude in alpha noise. In *2014 12th International Conference on Signal Processing (ICSP)*, pages 46–51, Oct 2014.

[WLGM15] Xueyun Wang, Kui Li, Pengyu Gao, and Suxia Meng. Research on parameter estimation methods for alpha stable noise in a laser gyroscope's random error. *Sensors (Basel, Switzerland)*, 15 :18550–64, 08 2015.

[WPS09] M. Z. Win, P. C. Pinto, and L. A. Shepp. A mathematical theory of network interference and its applications. *Proceedings of the IEEE*, 97(2) :205–230, Feb 2009.

[WSM04] H. Wymeersch, H. Steendam, and M. Moeneclaey. Log-domain decoding of ldpc codes over gf(q). In *2004 IEEE International Conference on Communications (IEEE Cat. No.04CH37577)*, volume 2, pages 772–776 Vol.2, June 2004.

[WTS88] Edward J. Wegman, J. B. Thomas, and S. C. Schwartz, editors. *Topics in Non-Gaussian Signal Processing.* Springer-Verlag, Berlin, Heidelberg, 1988.

[WW95] Aleksander Weron and Rafal Weron. Computer simulation of lévy α-stable variables and processes. In Piotr Garbaczewski, Marek Wolf, and Aleksander Weron, editors, *Chaos — The Interplay Between Stochastic and Deterministic Behaviour*, pages 379–392, Berlin, Heidelberg, 1995. Springer Berlin Heidelberg.

[XEA05] Xiao-Yu Hu, E. Eleftheriou, and D. M. Arnold. Regular and irregular progressive edge-growth tanner graphs. *IEEE Transactions on Information Theory*, 51(1) :386–398, Jan 2005.

[YA09] R. Yazdani and M. Ardakani. Linear LLR approximation for iterative decoding on wireless channels. *IEEE Transactions on Communications*, 57(11) :3278–3287, Nov 2009.

[YA11] R. Yazdani and M. Ardakani. Efficient LLR calculation for non-binary modulations over fading channels. *IEEE Transactions on Communications*, 59(5) :1236–1241, May 2011.

[Yu04] Jun Yu. Empirical Characteristic Function Estimation and Its Applications. *Econometric Reviews*, 23(2) :93–123, 2004.

[YV05] X. Yang and N. Vaidya. On physical carrier sensing in wireless ad hoc networks. In *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, volume 4, pages 2525–2535 vol. 4, March 2005.

[ZB02] A. Zoubir and R. Brcich. Multiuser detection in non-gaussian channels. *Digital Sig. Proces.*, 12 :262–273, 2002.

[ZD02] M. Zimmermann and K. Dostert. Analysis and modeling of impulsive noise in broad-band powerline communications. *IEEE Transactions on Electromagnetic Compatibility*, 44(1) :249–258, Feb 2002.

[ZEC+19] C. Zheng, M. Egan, L. Clavier, G. W. Peters, and J. Gorce. Copula-based interference models for iot wireless networks. In *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, pages 1–6, May 2019.

[ZHW11] K. Zhang, X. Huang, and Z. Wang. A high-throughput ldpc decoder architecture with rate compatibility. *IEEE Transactions on Circuits and Systems I : Regular Papers*, 58(4) :839–847, April 2011.

[Zol86] V. M. Zolotarez. *One dimensional Stable Distribution*. American Mathematical society, 1986.