



N° d'ordre : 2012-10-TH

THÈSE DE DOCTORAT

DOMAINE : STIC

SPECIALITE : Télécommunications

Ecole Doctorale « Mathématiques, Télécommunications, Informatique, Signal, Systèmes Electroniques »

Présentée par :

Wassim JOUINI

Sujet :

Contribution to learning and decision making under uncertainty for Cognitive Radio

Soutenue le 15 juin 2012

devant les membres du jury :

Prof. Damien ERNST	University of Liège	Prof.-Chair EDF-Luminus Leader
Prof. Friedrich K. JONDRAL*	Karlsruhe Institute of Technology	Prof.-Research Director (CEL)
Dr. Apostolos KOUNTOURIS	Orange Lab, Grenoble	Expert Engineer
Prof. Christophe MOY ***	Supélec, Rennes	Professor (Supervisor)
Prof. Rémi MUNOS **	INRIA, Lille	Senior Researcher (Sequel Team)
Prof. Jacques PALICOT	Supélec, Rennes	Prof.-Research Director (SCEE)
Prof. Jordi PEREZ-ROMERO*	UPC, Barcelona	Associate Professor

* Rapporteurs (External Referees)

** Président du jury (Chair of PhD Defense Committee)

*** Directeur de thèse (Supervisor)

Abstract

During the last century, most of the meaningful frequency bands were licensed to emerging wireless applications. Because of the static model of frequency allocation, the growing number of spectrum demanding services led to a spectrum scarcity. However, recently, series of measurements on the spectrum utilization showed that the different frequency bands were underutilized (sometimes even unoccupied) and thus that the scarcity of the spectrum resource is virtual and only due to the static allocation of the different bands to specific wireless services. Moreover, the underutilization of the spectrum resource varies on different scales in time and space offering many opportunities to an unlicensed user or network to access the spectrum. Cognitive Radio (CR) and Opportunistic Spectrum Access (OSA) were introduced as possible solutions to alleviate the spectrum scarcity issue.

In this dissertation, we aim at enabling CR equipments to exploit autonomously communication opportunities found in their vicinity. For that purpose, we suggest decision making mechanisms designed and/or adapted to answer CR related problems in general, and more specifically, OSA related scenarios. Thus, we argue that OSA scenarios can be modeled as Multi-Armed Bandit (MAB) problems. As a matter of fact, within OSA contexts, CR equipments are assumed to have no prior knowledge on their environment. Acquiring the necessary information relies on a sequential interaction between the CR equipment and its environment. Finally, the CR equipment is modeled as a cognitive agent whose purpose is to learn while providing an improving service to its user.

During a preliminary phase, we discuss different solutions borrowed from the Machine Learning literature. We chose in the dissertation to focus on a simple yet efficient learning algorithm known as UCB_1 algorithm. The rest of the analysis aims at exploring the performance of UCB_1 in more complex and realistic scenarios. Namely, we consider one secondary user (SU) willing to exploit communication opportunities left vacant by their incumbent users. The SU is allowed to access a frequency band if he senses it free. Consequently, he needs to learn the availability of the different bands in order to select the most available one (i.e., the optimal band). The sensing process is unfortunately prone to errors.

Thus, firstly we analyze the performance of UCB_1 algorithm when dealing with OSA problems with imperfect sensing. More specifically, we show that UCB_1 can efficiently cope with sensing errors. We prove its convergence to the optimal channel and quantify its loss of performance compared to the case with perfect sensing. Secondly, we combine UCB_1 algorithm with collaborative and coordination mechanism to model a secondary network (i.e. several SUs). We show that within this complex scenario, a coordinated learning mechanism can lead to efficient secondary networks. These scenarios assume that a SU can efficiently detect incumbent users' activity while having no prior knowledge on their characteristics. Usually, energy detection is suggested as a possible approach

to handle such task. Unfortunately, energy detection is known to perform poorly when dealing with uncertainty. Consequently, we ventured in this Ph.D. to revisit the problem of energy detection limits under uncertainty. We present new results on its performances as well as its limits when the noise level is uncertain and the uncertainty is modeled by a log-normal distribution (as suggested by Alexander Sonnenschein and Philip M. Fishman in 1992).

Within OSA contexts, we address a final problem where a sensor aims at quantifying the quality of a channel in fading environments. In such contexts, UCB_1 algorithms seem to fail. Consequently, we designed a new algorithm called Multiplicative UCB (MUCB) and prove its convergence. Moreover, we prove that MUCB algorithms are order optimal (i.e., the order of their learning rate is optimal). This last work provides a contribution that goes beyond CR and OSA. As a matter of fact, MUCB algorithms are introduced and solved within a general MAB framework.

Résumé Général Français

L'allocation des ressources spectrales à des services de communications sans fil, sans cesse plus nombreux et plus gourmands, a récemment mené la communauté radio à vouloir remettre en question la stratégie de répartition des bandes de fréquences imposée depuis plus d'un siècle. En effet une étude rendue publique en 2002 par la commission fédérale des communications aux Etats-Unis (Federal Communications Commission - FCC) mit en évidence une pénurie des ressources spectrales dans une large bande de fréquences comprise entre quelques mégahertz à plusieurs gigahertz. Cependant, cette même étude expliqua cette pénurie par une allocation statique des ressources aux différents services demandeurs plutôt que par une saturation des bandes de fréquences. Cette explication fut par la suite corroborée par de nombreuses mesures d'occupation spectrale, réalisées dans plusieurs pays, qui montrèrent une forte sous-utilisation des bandes de fréquences en fonction du temps et de l'espace, représentant par conséquent autant d'opportunité spectrale inexploitée. Ces constatations donnèrent naissance à un domaine en pleine effervescence connu sous le nom d'Accès Opportuniste au Spectre (Opportunistic Spectrum Access).

Nos travaux suggèrent l'étude de mécanismes d'apprentissage pour la radio intelligente (Cognitive Radio) dans le cadre de l'Accès Opportuniste au Spectre (AOS) afin de permettre à des équipements radio d'exploiter ces opportunités de manière autonome. Pour cela, nous montrons que les problématiques d'AOS peuvent être fidèlement représentées par des modèles d'apprentissage par renforcement. Ainsi, l'équipement radio est modélisé par un agent intelligent capable d'interagir avec son environnement afin d'en collecter des informations. Ces dernières servent à reconnaître, au fur et à mesure des expériences, les meilleurs choix (bandes de fréquences, configurations, etc.) qui s'offrent au système de communication. Nous nous intéressons au modèle particulier des bandits manchots (Multi-Armed Bandit appliqué à l'AOS).

Nous discutons, lors d'une phase préliminaire, différentes solutions empruntées au domaine de l'apprentissage machine (Machine Learning). Ensuite, nous élargissons ces résultats à des cadres adaptés à la radio intelligente. Notamment, nous évaluons les performances de ces algorithmes dans le cas de réseaux d'équipements qui collaborent en prenant en compte, dans le modèle suggéré, les erreurs d'observations. On montre de plus que ces algorithmes n'ont pas besoin de connaître la fréquence des erreurs d'observation afin de converger. La vitesse de convergence dépend néanmoins de ces fréquences. Dans un second temps nous concevons un nouvel algorithme d'apprentissage destiné à répondre à des problèmes d'exploitation des ressources spectrales dans des conditions dites de fading.

Tous ces travaux présupposent néanmoins la capacité de l'équipement intelligent à détecter efficacement l'activité d'autres utilisateurs sur la bande (utilisateurs prioritaires dits utilisateurs primaires). La principale difficulté réside dans le fait que l'équipement intelligent ne suppose aucune connaissance a priori sur son environnement (niveau du

bruit notamment) ou sur les utilisateurs primaires. Afin de lever le doute sur l'efficacité de l'approche suggérée, nous analysons l'impact de ces incertitudes sur le détecteur d'énergie. Ce dernier prend donc le rôle d'observateur et envoie ses observations aux algorithmes d'apprentissage. Nous montrons ainsi qu'il est possible de quantifier les performances de ce détecteur dans des conditions d'incertitude sur le niveau du bruit ce qui le rend utilisable dans le contexte de la radio intelligente. Par conséquent, les algorithmes d'apprentissage utilisés pourront exploiter les résultats du détecteur malgré l'incertitude inhérente liée à l'environnement considéré et aux hypothèses (sévères) d'incertitude liées au problème analysé.

Acknowledgment

First of all, I would like to thank Christophe Moy, my supervisor, and Jacques Palicot, head of the SCEE department, for offering me the opportunity to pursue this research. Their helpful comments and critical minds kept challenging my work enabling it to tremendously improve. Their enthusiasm, their patient and friendship made this experience a success from both scientific and human perspective.

I would like to extend my gratitude to the SCEE members who are always available and supporting: Amor Nafkha, Yves Louët, Daniel Le Guennec, Pierre Leray, Renaud Séguier, Gilles Tourneur and Jacques Weiss.

My gratitude goes then to Damien Ernst whose advices and guidance at the early stage of this Ph.D. helped identifying promising approaches to further investigate. I would like to thank him very warmly for his patient, his rigorousness and his high spirit. He is a very capable, highly motivated and inspiring researcher as well as a very friendly person. He proved to be a very valuable collaborator and friend.

Moreover, along these years, I had the pleasure to work with very stimulating and friendly teams. I would like to thank Damien Ernst, Agustí Ramon and Jordi Pérez-Romero as well as Mérouane Debbah for inviting me into their respective laboratories: the University of Liège, the Universitat Politècnica de Catalunya and at the Chair-Alcatel at Supélec.

I would like to acknowledge the support and friendship of those, within the SCEE team and beyond, who helped me through this journey : Ziad Khalaf, Hongzhi Wang, Salma Bourbia, Abel Gouba, Patricia Kaiser, Stéphane Lecompte, Hanan Salam, Catherine Soladié, Samba Traoré, Nicolas Stoiber, Noël Tchidjio, Sosthène Yamego, Biyi Lin, Miguel Lopez Benitez, Marco Di-Felice, Luciano Bononi, Zayen Bassem, Leonardo Cardoso, Samir Perlaza, Tembine Hamidou, Najett Neji, Ana Galindo-Serrano, Ismael Gómez Miguelez, Raphael Fonteneau, Florence Fonteneau-Belmudes, Francis Maes, Boris Defourny, Emmanuel Rachelson and many others friends and colleagues that I forgot to mention here.

I would like to express my gratitude to the jury members: Friedrich K. Jondral and Jordi Pérez-Romero for carefully reading this document and for their insightful comments, as well as Apostolos Kountouris, Rémi Munos and Damien Ernst, for evaluating this work and helping improving its clarity and quality.

I would like thank Marc Cuggia and Olivier Dameron as well as Joël Boissolle and Marine Olivo, respectively from the medical school of Rennes and from the University of Rennes 1, for allowing me to join their teaching teams during the years 2008 to 2011. A

very special thanks to Marc Cuggia, on the one hand for including me into his research team, and on the other hand for the numerous very interesting general purpose debates as well as technical discussions about medical care issues and high technology innovations.

Last but not least, my deepest personal gratitude goes to my family, my dear friends and my beloved one for their unconditional support and encouragement.

To all of you and for those I omitted, I dedicate this work,

Wassim JOUINI

Acronyms

GSM Global System for Mobile Communications	6
SMS Short Message Service	6
GPRS General Packet Radio Service	6
EDGE Enhanced Data rate for GSM Evolution.....	6
UMTS Universal Mobile Telecommunication System.....	6
LTE Long Term Evolution.....	6
WLAN Wireless Local Area Network.....	6
ISM Industrial, Scientific and Medical	6
Wi-Fi Wireless Fidelity, IEEE 802.11	6
ADSL Asymmetric Digital Subscriber Line	7
VoIP Voice over IP	7
ITU International Telecommunication Union.....	6
ICT Information and Communications Technology	6
SDR Software Defined Radio	13

PDA Personal Digital Assistant	14
CR Cognitive Radio	14
QoS Quality of Service	14
CC Cognitive Cycle	14
SRB The Sensorial Radio bubble	16
HDCRAM Hierarchical and Distributed Cognitive Architecture Management	16
USRP Universal Software Radio Peripheral	17
FCC Federal Communications Commission	24
ETSI European Telecommunications Standards Institute	24
DCA Dynamic Configuration Adaptation	22
DSA Dynamic Spectrum Access	22
OSA Opportunistic Spectrum Access	22
CA Cognitive Agent	25
CE Cognitive Engine	25
DEUM Dynamic Exclusive Use Model	31
OSM Open Sharing Model	31
SCM Spectrum Commons Model	31
HAM Hierarchical Access Model	31

RKRL Radio Knowledge Representation Language	34
XML eXtensible Markup Language	34
RDF Resource Description Framework	34
OWL Web Ontology Language	34
XG neXt Generation	34
W3C World Wide Web Consortium	34
SiA Simulated Annealing	35
GA Genetic Algorithm	35
SwA Swarm Algorithm	35
ICA Insect Colony Algorithms	36
ANN Artificial Neuronal Networks	37
ECS Evolving Connectionist Systems	37
SNR Signal-to-Noise Ratio	40
NP-ED Neyman-Pearson Energy Detector	41
i.i.d. independent and identically distributed	43
AWGN Additive White Gaussian Noise	43
PDF Probability Density Function	46
MAB Multi-Armed Bandit	59

PU Primary User 60

SU Secondary User 60

PN Primary Network 64

UCB Upper Confidence Bound 72

SN Secondary Network 95

MUCB Multiplicative Upper Confidence Bound 121

Mathematical Notations

\mathbf{y}_t Vector of independent and identically distributed samples gathered by the receiver at the slot $t \in \mathbb{N}$.

\mathbf{n}_t Vector of independent and identically distributed noise samples at the slot $t \in \mathbb{N}$.

\mathbf{x}_t Vector of independent and identically distributed signal samples at the slot $t \in \mathbb{N}$.

$\sigma_{x,t}^2$ Signal's power level.

$\sigma_{n,t}^2$ Noise's power level.

$\mathbb{P}_{fa,t}$ The probability of false alarm.

$\mathbb{P}_{d,t}$ The probability of detection.

π A decision making policy.

\mathcal{T}_t Energy Statistic.

$F_{\chi_M^2}(\cdot)$ Cumulative distribution function of a χ^2 -distribution with M degrees of freedom.

$Q(\cdot)$ Complementary cumulative distribution function of Gaussian random variable (also known as *Marcum function*).

$f_{\chi_M^2}(\cdot)$ The probability density function of a χ^2 distribution with M degrees of freedom.

$f_{\text{LogN}(\mu_{\mathcal{L}}, \sigma_{\mathcal{L}}^2)}(\cdot)$ The probability density function of a Log-Normal distribution with parameters $\{\mu_{\mathcal{L}}, \sigma_{\mathcal{L}}^2\}$.

$f_{\mathcal{N}(\mu_{\mathcal{N}}, \sigma_{\mathcal{N}}^2)}(\cdot)$ The probability density function of a normal distribution with parameters $\{\mu_{\mathcal{N}}, \sigma_{\mathcal{N}}^2\}$.

$\xi_t(\alpha_{fa})$ Detection threshold to guaranty $\mathbb{P}_{fa} \leq \alpha_{fa}$.

$$\Delta_1(x) = f_{\mathcal{L}og\mathcal{N}(\mu_{\mathcal{L}}, \sigma_{\mathcal{L}}^2)}(x) - f_{\chi_M^2}(x)$$

$$\Delta_2(x) = f_{\chi_M^2}(x) - f_{\mathcal{N}(\mu_{\mathcal{N}}, \sigma_{\mathcal{N}}^2)}(x)$$

$\tilde{\Delta}_1(x)$ Taylor Approximation of $\Delta_1(x)$.

$\tilde{\Delta}_2(x)$ Taylor Approximation of $\Delta_2(x)$.

\mathcal{W}_t Normalized Log-Energy-Ratio statistic, $\mathcal{W}_t = \log\left(\frac{\mathcal{T}_t}{M\hat{\sigma}_n^2}\right)$

γ_t SNR, $\gamma_t = \sigma_{x,t}^2/\sigma_n^2$

$\mathbb{E}(\mathcal{W}_t|\mathbb{H}_0)$ Expected value of \mathcal{W}_t under hypothesis \mathbb{H}_0

$\mathbb{E}(\mathcal{W}_t|\mathbb{H}_1)$ Expected value of \mathcal{W}_t under hypothesis \mathbb{H}_1

$\mathbb{V}(\mathcal{W}_t)$ Variance of the random variable \mathcal{W}_t .

N Number of resources in a MAB problem and number of channels in OSA.

Θ MAB machines' distributions. In OSA it refers to the channels' occupation pattern.

θ_n The distribution of a specific MAB machine n . The occupation pattern of a specific channel n .

\mathbf{S}_t Channels' state at the slot number t : $\mathbf{S}_t = \{S_{1,t}, \dots, S_{N,t}\} \in \{0, 1\}^N$

μ_n Expected income of a machine n . Availability of a channel n for all t , $\mu_n \triangleq \mathbb{E}_{\theta_n} [S_{n,t}]$.

r_t Reward received at the slot t .

W_t^π Cumulated reward at the slot t .

$\overline{W}_{T_n(t)}$ Averaged cumulated reward at the slot t .

R_t^π Cumulated expected regret at the slot t .

Contents

Acknowledgment	v
Contents	xv
I Ph.D. Dissertation	1
1 Twenty Years of Wireless Communication Innovations: Towards Cognitive Radio	5
1.1 Twenty years of Wireless Communications	6
1.1.1 The emergence of licensed cellular networks	6
1.1.2 WLAN and unlicensed standards: the success of the WiFi standard	6
1.1.3 Cooper's law and the physical layer's limits	11
1.2 Towards Cognitive Radio	13
1.2.1 Software Defined Radio	13
1.2.2 The rise of Cognitive Radio	14
1.3 Ph.D. motivations	16
1.3.1 SDR and CR related topics	16
1.3.2 Research Objectives in this Ph.D.	17
1.4 Presentation and results	19
2 Decision Making and Learning for Cognitive Radio	21
2.1 Introduction	22
2.2 Cognitive Radio	22
2.2.1 Definitions	23
2.2.2 Basic Cognitive Cycle	25
2.3 Decision Making Problems for Cognitive Radio	27
2.3.1 Design Space and Dynamic Configuration Adaptation Problem	27
2.3.2 Spectrum Scarcity and Dynamic Spectrum Access	30
2.4 Decision Making Tools for Dynamic Configuration Adaptation	32
2.4.1 Expert approach	34
2.4.2 Exploration based decision making	35
2.4.3 Learning approaches: exploration and exploitation	37
2.5 Conclusions	38

3	Energy detection limits under noise uncertainty and Log-Normal approximation of Chi-square distributions	39
3.1	Introduction	41
3.2	System model	43
3.2.1	Network assumption	43
3.2.2	Performance evaluation of a detection policy π	43
3.2.3	Neyman-Pearson Energy Detector	45
3.2.4	Energy detection with noise uncertainty	46
3.3	Log-Normal Approximation of χ^2 distributions	46
3.3.1	Mathematical Model	46
3.3.2	Main Results	48
3.3.3	Simulations and Empirical Evaluation of Log-Normal based Approximations	50
3.4	Energy Detector under Log-Normal noise uncertainty	52
3.4.1	Noise Uncertainty and Energy Statistic's Approximation	52
3.4.2	Energy Detector's Performances and Limits	53
3.5	Conclusion	56
4	Learning for Opportunistic Spectrum Access: A multi-Armed Bandit Framework	57
4.1	Introduction	59
4.1.1	General Context and Challenges	59
4.1.2	Classic Illustration: Opportunistic Spectrum Access	59
4.1.3	Outline and contributions	61
4.2	Learning for Opportunistic Spectrum Access: Multi-Armed Bandit Paradigm and Motivations	61
4.2.1	Multi-Armed Bandit Paradigm: Conceptual Problem Statement	61
4.2.2	Multi-Armed Bandit Paradigm: Stochastic Environment Vs Adversarial Environment	62
4.2.3	Multi-Armed Bandit Paradigm: Index based Policies	63
4.2.4	Opportunistic Spectrum Access Modeled as a Multi-Armed Bandit Problem	64
4.3	Opportunistic Spectrum Access Mathematic Model : a Multi-Armed Bandit problem	68
4.3.1	Basic Opportunistic Spectrum Access Model : Multi-Armed Bandit Notations	68
4.3.2	General Performance Evaluation of a Learning Policy and Optimality	70
4.3.3	Sample Mean Based Upper Confidence Bound Algorithms and Theoretical Performance Results	72
4.3.3.1	UCB_1	73
4.3.3.2	UCB_V	74
4.3.4	Complexity	75
4.4	Algorithm Illustration, Limits and Discussion	75
4.4.1	Configuration Adaptation Problem	75
4.4.2	OSA under perfect Channel State Information	79
4.4.3	Limits of the Theory and Discussion	82
4.5	Opportunistic Spectrum Access with Imperfect Sensing	84

4.5.1	UCB_1 Performance Analysis	84
4.5.2	Simulation Results	86
4.6	Simulink based Reinforcement Learning Scenario	87
4.6.1	Primary Network: OFDM	88
4.6.2	Sensing: Energy Detector	88
4.6.3	Simulation results	90
4.7	Conclusion	90
5	Collaboration and Coordination in Secondary Networks for Opportunistic Spectrum Access	93
5.1	Introduction	96
5.2	Related Work	97
5.3	Network model	98
5.3.1	Primary Network	98
5.3.2	Secondary Users model	99
5.4	Learning Mechanism	100
5.5	General Resource Allocation Problem	101
5.5.1	Coordination and Job Assignment Problems	101
5.5.2	Coordination Mechanisms based on The Hungarian Algorithm	103
5.5.3	Coordination Mechanisms based on Round Robin Algorithm	103
5.6	Theoretical Analysis	104
5.6.1	Definitions of the Reward and the Expected Cumulated Regret	104
5.6.2	Theoretical Results: Symmetric Network	105
5.6.3	Non-Symmetric Network, the Heterogeneous case	106
5.7	Information Sharing: Discussion	107
5.7.0.1	Configuration adaptation	107
5.7.0.2	Acknowledgment	108
5.7.0.3	Information sharing	108
5.8	Empirical Evaluation: Simulation Results	108
5.8.1	Scenario and experimental protocol for the regret analysis	108
5.8.2	Simulation results: Regret Analysis	109
5.8.3	Simulation results: Network Performance Analysis	110
5.9	Conclusion	112
5.10	Appendix	113
6	Fading Environments, Exponential Reward Distributions and MUCB Algorithms	119
6.1	Introduction	121
6.2	Multi-Armed Bandits	122
6.3	Multiplicative upper confidence bound algorithms	123
6.4	Analysis of $MUCB(\alpha)$ policies	124
6.4.1	Consistency and order optimality of MUCB indexes	125
6.4.2	Learning Anomalies and Consistency of MUCB policies	126
6.5	Simulation Results	128
6.6	Conclusion	128
6.7	Appendix	131
6.7.1	Large Deviations Inequalities	131

6.7.2	Proof of Lemma 1	132
6.7.3	Proof of Lemma 2	133
6.7.4	Proof of Lemma 3	135
7	Overview, General Conclusions and Future Work	137
7.1	Conclusion and Overview	138
7.2	Perspectives and Future Work	138
	Appendix	141
A	Three years of Ph.D. research, teachings and talks	143
B	Publications	145
B.1	Journal Papers	145
B.2	Book Chapter(s)	145
B.3	Invited Papers	145
B.4	International Conference papers (with peer-reviews)	146
B.5	National Conference paper(s) (with peer-reviews)	147
B.6	Technical Reports	147
II	Résumé en Français	149
C	Vingt années de communications sans-fil: vers la radio intelligente	151
C.1	Vingt années de communication sans-fil	152
C.1.1	Les limites de la couche physique et la loi de Cooper	152
C.2	Vers la radio intelligente	154
C.2.1	La radio logicielle	154
C.2.2	Radio intelligente	155
C.3	Travaux de Thèse	156
D	Limites de la détection d'énergie dues à une connaissance incertaine du niveau du bruit	159
D.1	Introduction	160
D.2	Détection d'Energie	161
D.2.1	Modèle, Hypothèses et Notations	161
D.2.2	Evaluation d'un détecteur	161
D.2.3	Détecteur d'Energie: modèle de Neyman-Pearson	161
D.2.4	Détection d'énergie avec un niveau de bruit incertain	162
D.3	Approximation Log-Normal d'une distributions χ^2	163
D.4	Détecteur d'Energie et le Modèle Log-Normal de l'Incertitude	165
D.4.1	Approximation de la Statistique d'Energie et Incertitude sur le Niveau du Bruit	165
D.4.2	Performances et limites du détecteur d'énergie	165
D.5	Conclusion	168

E Apprentissage pour l'Accès Opportuniste au Spectre : Prise en Compte des Erreurs d'Observation	169
E.1 Introduction	169
E.1.1 Accès Opportuniste au Spectre	169
E.1.2 Le Paradigme des Bandits Manchots	170
E.2 Modélisation de la problématique	171
E.2.1 Le réseau primaire	171
E.2.2 L'utilisateur secondaire	172
E.2.3 Stratégie d'accès au canal	172
E.3 Algorithmes et performances	173
E.3.1 Algorithme de selection du canal	173
E.3.2 Performances	174
E.4 Conclusion	175
List of Figures	177
List of Tables	183

Part I

Ph.D. Dissertation

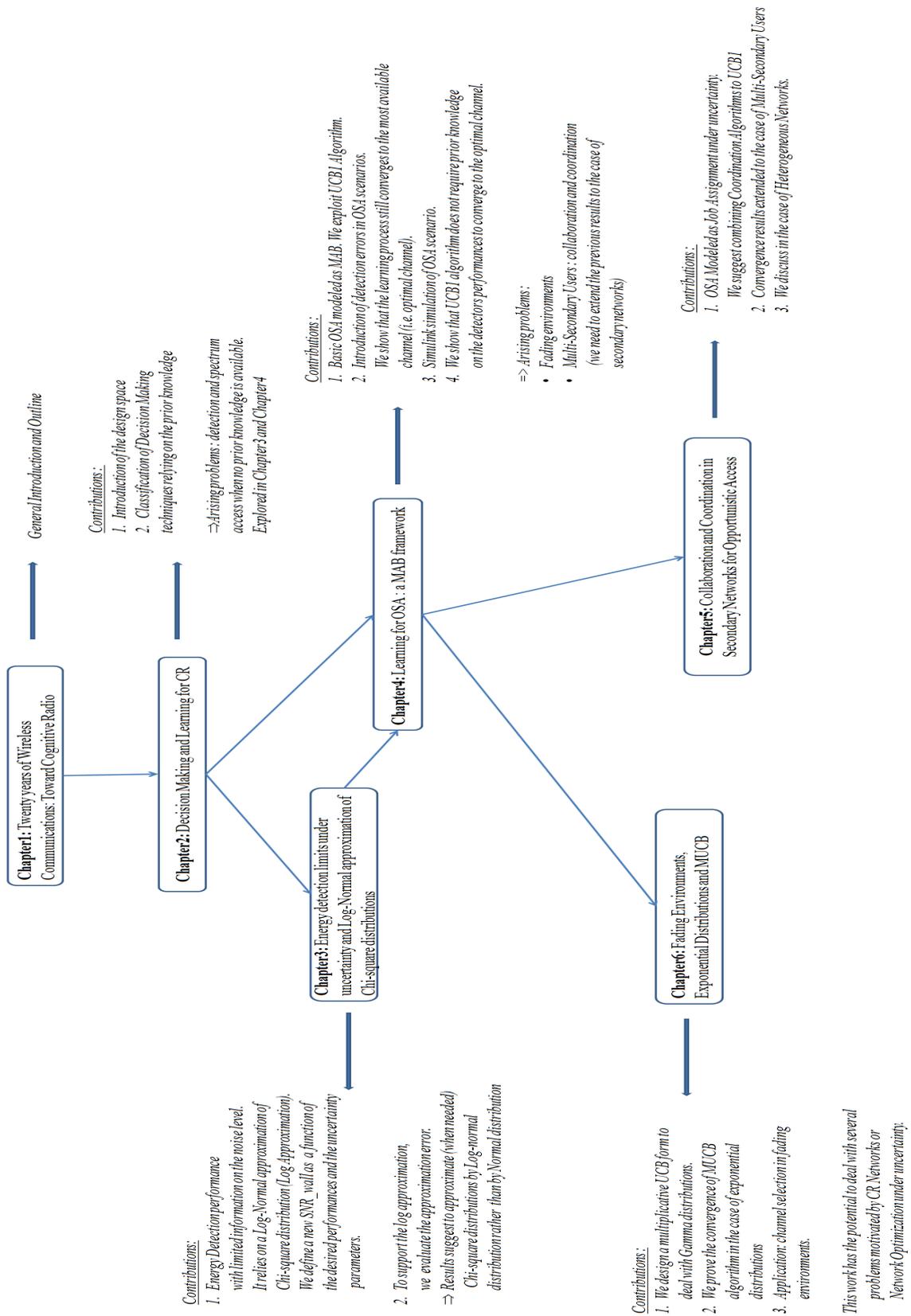


Figure 1: General Synopsis

Chapter 1

Twenty Years of Wireless Communication Innovations: Towards Cognitive Radio

1.1 Twenty years of Wireless Communications

1.1.1 The emergence of licensed cellular networks

In April 1973, in New York City, the first handset based wireless communication was performed by Martin Cooper (former Motorola vice president and division manager). At that time, the experiment was held to convince the Federal Communication Commission to support wireless communication innovations (more specifically cellular networks) by allocating and licensing specific frequency bands to private companies⁽¹⁾. Since then, Radio technologies went a long way.

As a matter of fact, the advent of the digital wireless standard, Global System for Mobile Communications (GSM), demonstrated in 1991 in Finland, allowed a global democratization of cellular phones, and opened the way to both voice and data transmissions⁽²⁾. The GSM standard, also known as 2nd generation wireless telephony technology, saw many improvements and upgrades since its first design. On the one hand, General Packet Radio Service (GPRS) then Enhanced Data rate for GSM Evolution (EDGE) aimed at improving packet management and at raising the transmission speed. On the other hand, the Universal Mobile Telecommunication System (UMTS), also referred to as 3rd generation wireless telephony technology, 3G for short, in Europe and Japan, aimed at providing video based wireless -cellular- communications. Although, its design was mainly concerned with mobile TV and video calls, it is however most of the time used for mobile Internet access due to its satisfactory speed to fulfill the basic bandwidth requirements of Internet communications.

The success of cellular technology is unarguable: today, according to the statistics provided by International Telecommunication Union (ITU), more than 86% of the worldwide community has a mobile cellular subscription [1]. For illustration purpose, the global Information and Communications Technology (ICT) developments during the period 2001-2011 are drawn in Figure 1.1

1.1.2 WLAN and unlicensed standards: the success of the WiFi standard

Complementary to licensed cellular networks, GSM, EDGE, UMTS and the future Long Term Evolution (LTE) standard, but on a much smaller cellular scale, Wireless Local Area Network (WLAN) were designed for short range and high speed data transmissions. No specific license is required to deploy a fully operating network or to extend an existing one in the Industrial, Scientific and Medical (ISM) band. Thus, it enables to create an efficient personal or professional network at a low cost compared to a hard wired network. Moreover, it enables all compatible machines to communicate on a relatively large cell (around 50-300 meters depending on the environment).

One of the most famous unlicensed standards, Wireless Fidelity, IEEE 802.11 (Wi-Fi) specifications first commercialized in 1999, quickly spread over the world. Its wide success is probably due to the fact that Wi-Fi provides a mean of high rate communication, at a low cost, with an application range only limited by the designer's imagination. Wi-Fi is

⁽¹⁾Amusing anecdote found at: <http://www.cellular.co.za/cellphoneinventor.htm>

⁽²⁾Data transmissions started with short and instantaneous messages usually referred to as Short Message Service (SMS)

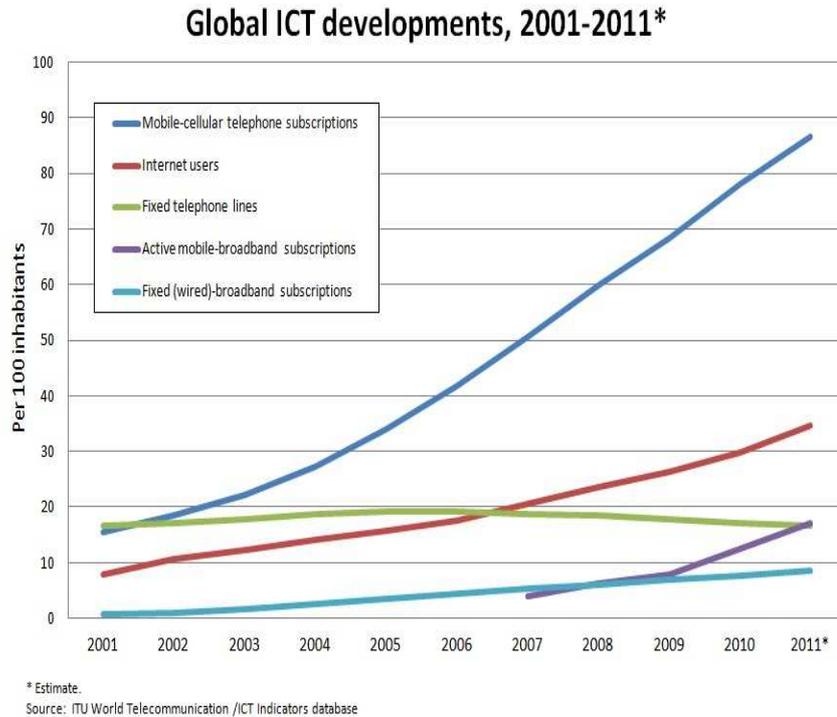


Figure 1.1: the global ICT developments during the period 2001-2011 [1].

today generally used for Internet access and machine to machine connections (e.g., with a printer, a camera or other computers). However, Wi-Fi networks that are used to extend Internet networks are also increasingly exploited to provide wireless voice services, e.g., Voice over IP (VoIP).

This latter application can also be useful to unload overcrowded licensed cellular networks and is currently exploited in France by the major wireless communication operators and providers (*SFR, Bouygues Telecom* and *Free*). As a matter of fact, the democratization and large deployment of the so called ‘ADSL⁽³⁾ Boxes’ provided by the main operators to private clients, led to quasi-ubiquitous wireless access points. As one can observe in Figure 1.2, all providers share the same space and obviously possess a dense Wi-Fi network. The density of the network is however mainly due to private users’ Wi-Fi Access points. These access points are shared by the providers in a transparent way, and the private user (usually at home) is not necessarily aware that his ADSL Internet access might be shared by external users through a virtual secondary Wi-Fi access managed by his Internet provider. It is important to understand at this level, that the secondary access is virtual and both access points share the same wireless network card. This matter will be discussed and further detailed later as it provides an interesting introduction and illustration to Opportunistic Spectrum Access related concepts.

Wi-Fi networks also appear to become victims of their own success. As a matter of fact, unlike cellular networks, no specific channel access coordination is planned among different

⁽³⁾ Asymmetric Digital Subscriber Line (ADSL)

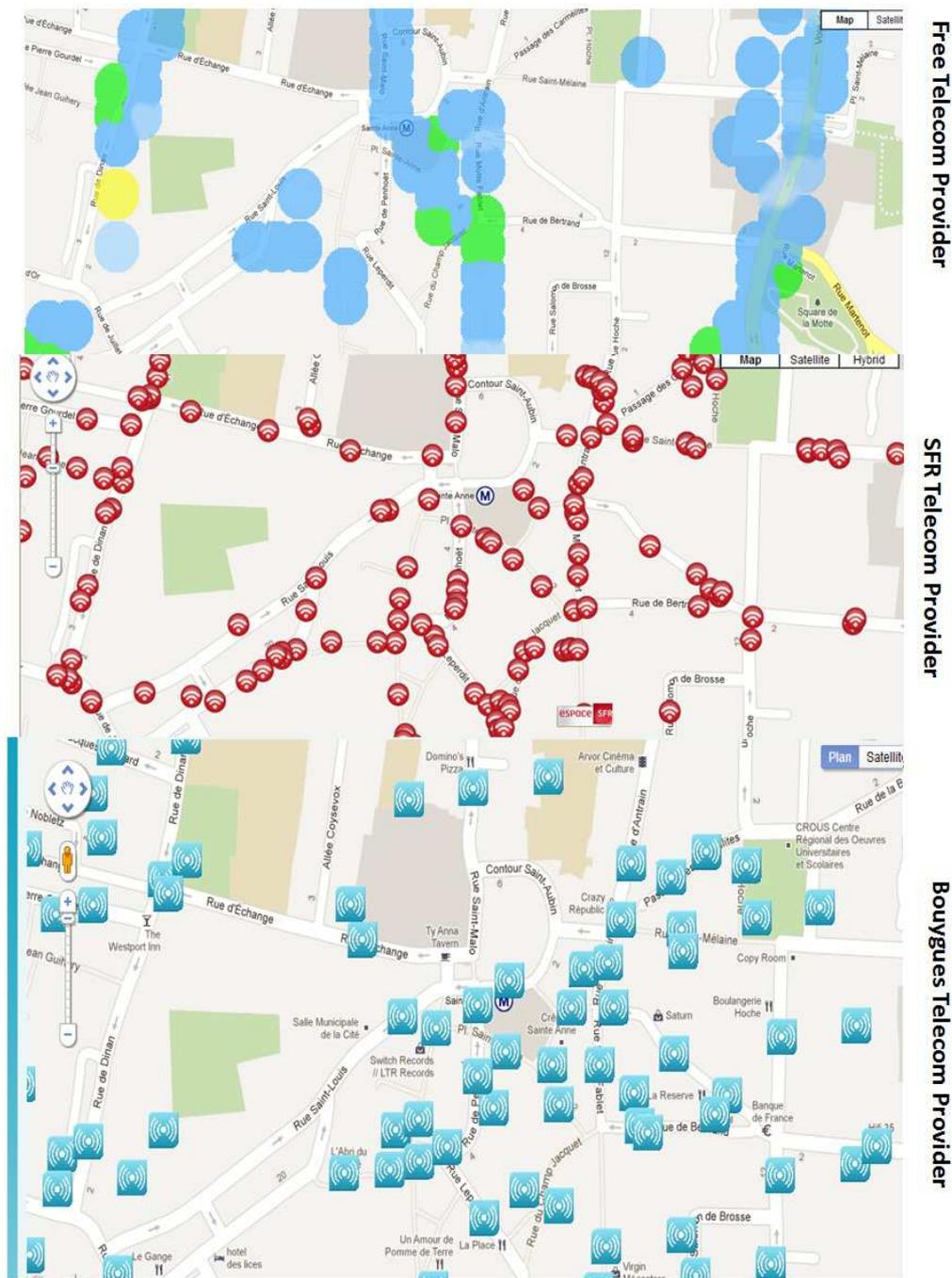


Figure 1.2: Major France Telecommunication providers' Wi-Fi access points in Rennes around Saint-Anne metro station. To quickly densify their Wi-Fi network, main operators exploit their subscribers' ADSL box. Thus, usually the Wi-Fi connexion of an ADSL box runs two or three virtual networks that share the same connexion. Usually we find two virtual networks: the first wireless network is dedicated to the subscriber, whereas the second wireless network is managed by the operator. This latter network is shared with other mobile subscribers in the vicinity of the box. Both networks share the same wireless physical card and thus share the same frequency band. Finally, we also noticed, in the case of the operator *Free*, the existence of a third virtual network dedicated to VoIP.

Wi-Fi networks. Consequently they interfere and their performances can rapidly degrade when other similar networks are in their vicinity operating in the same band. Although the case of Steve Jobs in June 2010 at Apple's Worldwide Developers Conference might seem extreme -where no less than 500 separate base stations (or Hot-Spots) were detected for 5000 persons in the conference room - it yet offers an illustration full of teachings on the limits of non-coordinated access among different users. Thus, due to the extreme traffic congestion (and probably a very high collision rate) in the room, Steve Jobs was not able to maintain a stable Wi-Fi connection on his new and fancy iPhone 4 , which led him to fail the performed demo.

In general, nowadays, depending on the location, a Wi-Fi device can usually detect at least a few Wi-Fi Hot-Spots. In residential and commercial zones, such a survey usually exhibits no less than 10 or 20 -real and virtual- access points and up to a few hundreds (between several commercial and residential buildings where all surrounding access points are visible)⁽⁴⁾. Unfortunately, from the user's perspective, only few access points are open and free. Moreover, since no high level coordination is available among the different access points, to share time/frequency resources, harmful collisions might frequently occur in crowded areas leading to very poor network performances.

Thus, the technological success of Wi-Fi network, approved by a large majority of system designers and consumers, led to a major economical success. In general, and, as suggested in Figure 1.3 presented by Dave Cleevly at the annual conference of Spectrum management IEEE DySPAN2011 (which took place in Aachen, Germany), unlicensed use of spectrum bands opens the way to technologies that seem to be profitable in terms of revenue. Hence, he suggested to provide new bands to allow the development of such technologies. The interpretation of these results as well as the credibility of the sources supporting the presented analysis have naturally been questioned. However, as asked and answered by Linda Doyle on her web blog in an article entitled 'To License or Not to License':

I am not sure I completely buy into the message or interpretation of the data but I find it an interesting suggestion. I wonder if it is possible to create an equivalent graph showing the opposite? Having said that, David has gathered his facts and figures from multiple reliable sources and did say that even if the calculations are off, the big difference in magnitude between the value of the licensed and unlicensed remains.

In other words, the sources and the results seem to be solid enough to be accepted by the community and to suggest opening new spectrum bands to allow new innovative communication services based on unlicensed spectrum use.

As a first conclusion, due to their astonishing success, both Cellular and Wireless Local Area Networks grew to become crowded. However, WLANs are yet to be expanded, optimized and coordinated in order to fully exploit the communication opportunities offered by these technologies.

⁽⁴⁾The reported estimations were observed on my personal, not so fancy, *smartphone* relying a free Wi-Fi analyzer application. A more rigorous survey is provided at the following web link: <http://www.silicon.com/technology/mobile/2007/01/31/peter-cochranes-blog-wi-fi-london-39165548/>

Although the analysis was performance 5 years ago, in London, it appears to remain relevant, from my point of view, to describe the situation in current middle sized towns.

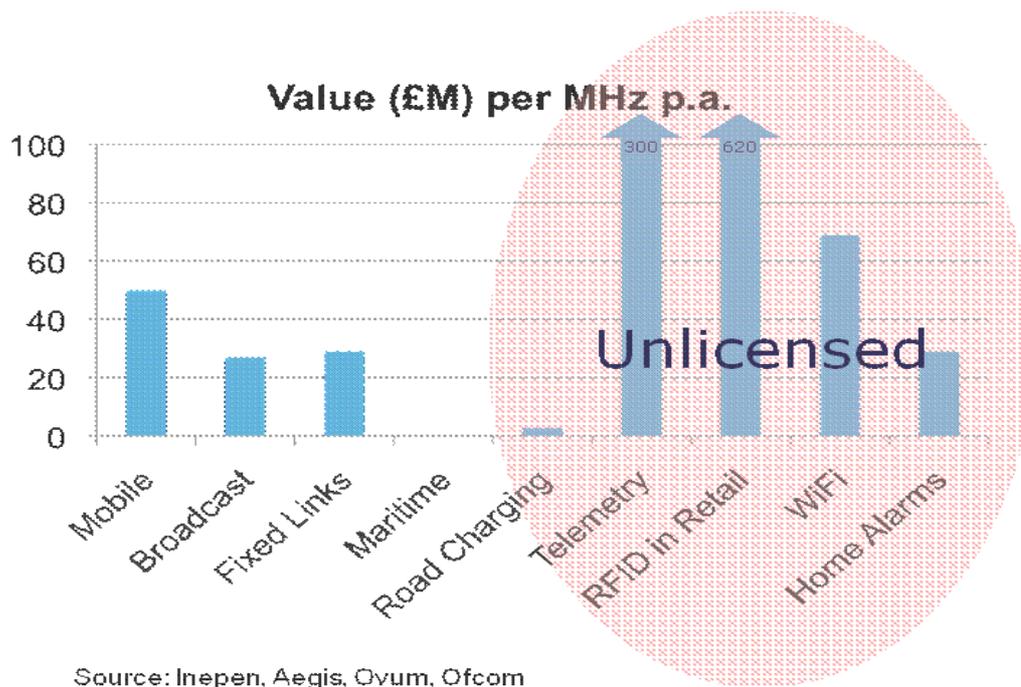


Figure 1.3: Estimated economical impact of both licensed and unlicensed radio technologies: results presented by Dave Cleevly during the annual conference of Spectrum management IEEE DySPAN2011. This figure support the idea that more bandwidth should be opened to unlicensed radio exploitation.

At this level a natural question arises: considering the increasing need for bandwidths and higher data rates, which degrees of freedom are still to be exploited to introduce new wireless communication opportunities?

To provide a piece of answer, we revisit the evolution of the overall effective simultaneously achievable throughput by all operating wireless systems⁽⁵⁾.

1.1.3 Cooper's law and the physical layer's limits

1991-2011: Twenty years of digital wireless communication innovations. Twenty years during which the increase of communication data rate provided by both licensed and unlicensed networks, as illustrated in Figure 1.4, marched along with the improvements of computing abilities: a higher electronic density and higher computing frequencies increasing at a regular rate (as extrapolated by Gordon E. Moore in 1965). Indeed, with more efficient computing tools, the wireless communication community was able to integrate and compute more sophisticated and complex physical layer related algorithms (e.g., higher sampling rate to probe larger bands, more efficient source coding as well as channel coding and waveforms, adaptive equalization and channel estimation to name a few).

As illustrated in Figure 1.4, the increase of achievable data rates by wireless communication devices, due to the physical layer's breakthroughs, is indeed substantial. Yet, surprisingly enough, it only represents a small fraction of the capacity increase observed during the last century of wireless communication! As a matter of fact, Martin Cooper recently claimed that the 'wireless capacity⁽⁶⁾ has doubled every 30 months over the last 104 years' [4]. However, the overall increase of wireless capacity can be fractioned into three main technological contributors:

- Better use of the spectrum through advanced signal processing tools (coding and modulation) and spectrum access management enabled a 25 fold increase.
- Managing wider bands allowed a 25 fold increase.
- Reducing cellular cells' scales led to a 1600 fold improvement⁽⁷⁾!

Many straightforward conclusions can be drawn from this statement and several questions arise. On the one hand, both physical layer related contributions only provided a 625 fold improvement in the overall wireless capacity. Whereas, the reduction of cellular networks' scale improves frequency bands reuse [5] and thus enables an easy improvement of the overall achievable throughput in a given region. Future wireless networks based on Femto-cells and Heterogeneous networks tend to exploit this degree of freedom [4]. As a matter of fact, as already noticed, unlicensed WLANs for instance enabled to intensify drastically the density of the network and thus to significantly increase, at a low cost, existing networks.

⁽⁵⁾Thus, anticipating the results, it naturally highly depends on the number of possible simultaneous wireless connections.

⁽⁶⁾Personal note: the term *capacity* used in this statement seems to be different from the one introduced in Information Theory. We thus prefer the following equivalent statement: 'The number of simultaneous voice and data connections has doubled every 2.5 years since wireless began (1900)'[3]

⁽⁷⁾It is not clear whether WLAN based networks are accounted in the final result proposed by Martin Cooper or not.

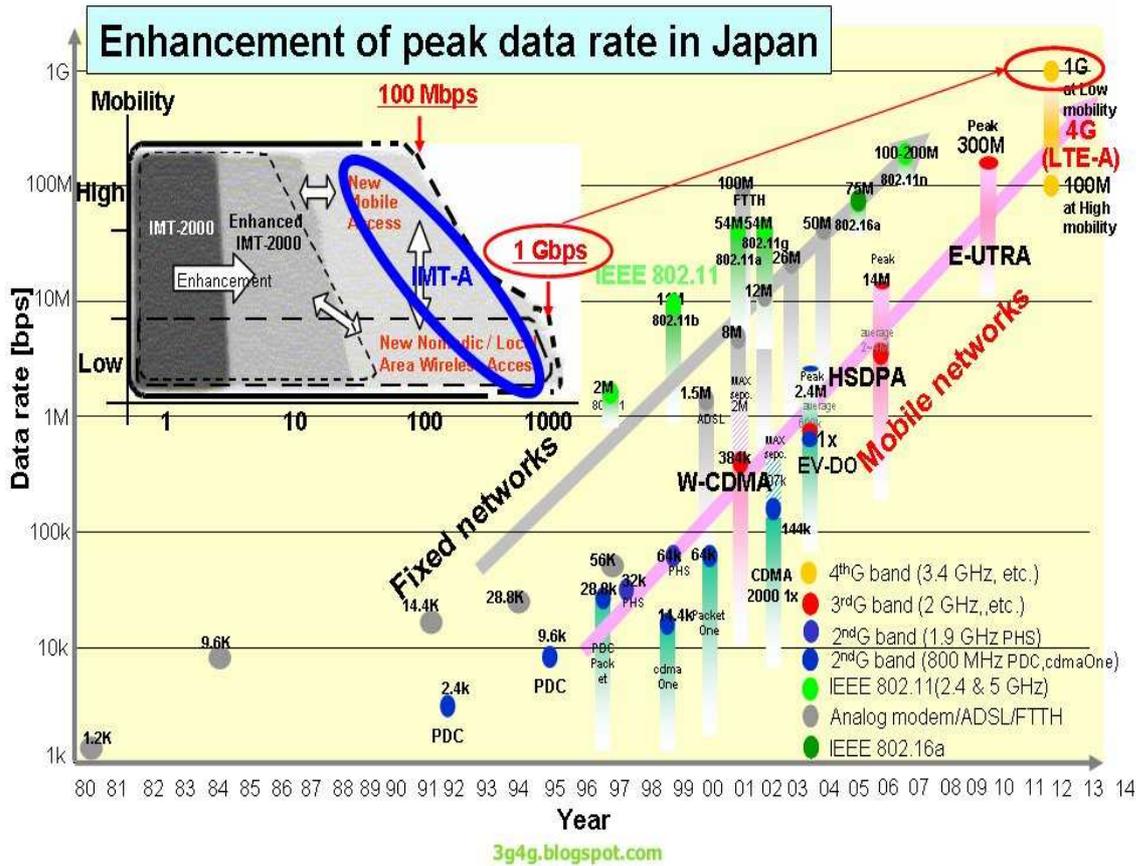


Figure 1.4: Evolution of the bandwidth of the main Radio communications standards (found online: http://3g4g.blogspot.fr/2008_04_01_archive.html) . Similar graphs are provided in [2].

On the other hand, physical layer based improvements do not seem to be able to sustain the increasing need for higher data rates. Considering that only minor improvement can be expected from the physical layer to reach the fundamental Shannon limit, is it still worth investigating [6]? The answer is of course yes: new opportunities appeared due to techniques such as beam-forming and smart-antennas techniques offering new spatial degrees of freedom allowing simultaneous communications in a same location.

However, these new degrees of freedom usually assume a high adaptability of radio devices to their environment. Thus, it presupposes that future wireless communication systems have the ability to probe their environment and to reconfigure on the fly their architecture relying on appropriate decisions made depending on the context (user's expectation and environments constraints). Consequently, to maintain the growth of wireless *capacity* at the pace estimated by Cooper's law, hardware flexibility and basic cognition abilities (sensing and decision making) need to be combined into both radio equipments and networks.

1.2 Towards Cognitive Radio

1.2.1 Software Defined Radio

The increase of computational capacity associated with (rather) cheap flexible hardware technologies (such as Programmable Logic Devices, Digital Signal Processors and Central Processing Units) offer us a glimpse into new ways to designing and managing future non military communication systems⁽⁸⁾. As a matter of fact in 1991, Joseph Mitola III argued that in a few years, at least in theory, software design of communication systems should be possible. The term coined by Joseph Mitola to present such technologies is Software Defined Radio (SDR).

For illustration purposes, today's radio devices need a specific dedicated electronic chain for each standard, switching from one standard to another when needed (known as the *Velcro* approach). With the growth of the number of these standards (GSM, EDGE, Wi-Fi, Bluetooth, LTE, etc.) in one equipment, the design and development of these radio devices has become a real challenge and the practical need for more flexibility became urgent. Recent hardware advances have offered the possibility to design, at least partially, software solutions to problems which were requiring in the past hardware signal processing devices: a step closer to SDR systems.

More specifically speaking, several possible definitions exist -and are still a matter of debate in the community- to define SDR systems. For consistency reasons, we briefly describe software related radio concepts as agreed on by the SDR Forum [7]. This matter is further discussed in [8].

The SDR Forum defines Software Defined Radio as *radio in which some or all of the physical layer functions are software defined* where *physical layer* and *software defined* terms are respectively described as:

- Physical layer: *The layer within the wireless protocol in which processing of Radio Frequency, Intermediate Frequency, or baseband signals including channel coding*

⁽⁸⁾Both US and European military have been working on such flexible and inter-operable defense systems since the late 70's.

occurs. It is the lowest layer of the ISO 7-layer model as adapted for wireless transmission and reception.

- Software Defined: *Software defined refers to the use of software processing within the radio system or device to implement operating (but not control) functions.*

Thus, SDR systems are defined only from the design and the implementation perspectives. Consequently it appears as a simple evolution from the usual hardwired radio systems. However, with the added software layer, it is technically possible with current technology to control a large set of parameters in order to adapt radio equipment to their communication environment (e.g., bandwidth, modulation, protocol, power level adaptation to name a few). However the control and optimization of reconfigurable radio devices need the definition of optimization criteria related to the equipment hardware capabilities, the users' needs as well as the regulators' rules. Introducing autonomous optimization capabilities in radio terminals and networks is the basis of Cognitive Radio, term also suggested and coined by Joseph Mitola III [9, 10].

1.2.2 The rise of Cognitive Radio

J. Mitola defined Cognitive Radio (CR), in his Ph.D. dissertation as follows [10]:

The term cognitive radio identifies the point at which wireless Personal Digital Assistant (PDA) and the related networks are sufficiently computationally intelligent about radio resources and related computer to computer communication to:

1. *Detect user communication needs as a function of use context, and*
2. *Provide radio resources and wireless services most appropriate to these needs.*

Thus, the purpose of this new concept is to autonomously meet the user's expectations, i.e., maximizing his 'profit', in terms of Quality of Service (QoS), throughput or power efficiency to name a few, without compromising the efficiency of the network. Hence, the needed intelligence to operate efficiently must be distributed in both the network and the radio device.

To fulfill these requirements, J. Mitola and J.Q. Maguire introduced the notion of *Cognitive Cycle (CC)* as described in Figure 1.5 [9, 10], where the *Cognitive Cycle* presupposes the capacity to collect information from the surrounding environment (perception), to digest it (i.e., learning, decision making and predicting tools) and to act in the best possible way by considering several constraints and the available information. The reconfiguration of radio equipment is not discussed in depth, however, it is generally accepted that SDR technology is needed to support Cognitive Radio [8].

As illustrated in Figure 1.5, a full Cognitive Cycle⁽⁹⁾ demands at every iteration five steps: *Observe, Orient, Plan, Decide* and *Act*. The *Observe* step deals with internal as well as external metrics. It aims at capturing the characteristics of the environment of the communication device (e.g., channel state, interference level or battery level to name a few.). This information is then processed by the three following steps : *Orient, Plan* and *Decide* steps, where priorities are set, schedules are planed according to the systems constraints, and decisions are made. Finally an appropriate action is taken during the

⁽⁹⁾It is called full cognitive radio to oppose it to other simplified versions suggested in the literature.

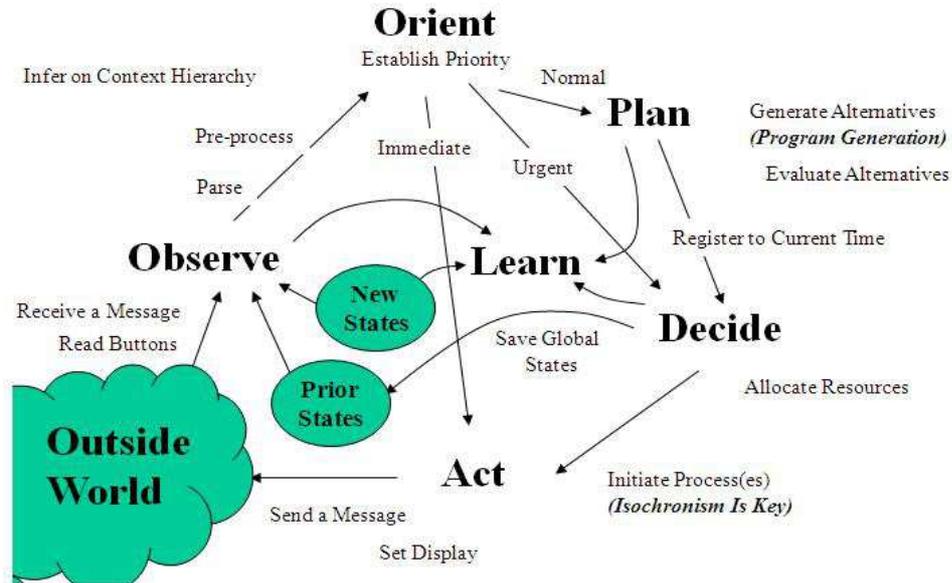


Figure 1.5: Cognitive radio decision making context: the CR cycle as introduced by Joseph Mitola III [10].

Act step (such as *send a message*, *reconfigure*, *modify power level* to name a few). In order to complete the Cognitive Cycle, a last and final step is needed to enhance the decision making engine of the communication device: the *Learn* step. As a matter of fact, learning abilities enable communication equipment to evaluate the quality of their past actions. Thus, the decision making engine learns from its past *successes* and *failures* to tune its parameters and adapt its decision rules to its specific environment. Learning can consequently help the decision making engine to improve the quality of future decisions.

As far as we can track the emergence of a Cognitive Radio literature and to the best of our knowledge, the today's plethora of publications started with three major contributions: On the one hand, the Federal Communication Commission (FCC) pointed out in 2002 the inefficiency of static frequency bands' allocation to specific wireless applications, and suggested Cognitive Radio as a possible paradigm to alleviate the resulting spectrum scarcity [11]. Then, S. Haykin in Paper [12] in 2005, suggested a simplified Cognitive Cycle to represent Cognitive Radio decision making engines as illustrated in Figure 1.6. Haykin's model tackled the particular dynamic spectrum management problem and discussed different possible models to design future Cognitive Radio Networks. Paper [12] inspired many studies on Cognitive Radio application fields such as Spectrum Hole Detection and Game Theory Based Cognitive Networks. Eventually, these two subjects led to two very active research fields as illustrated in these recent surveys [13, 14, 15].

On the other hand, while the two contributions [11] and [12] focus on spectral efficiency, C.J. Rieser suggested, through various publications, synthesized in his Ph.D. dissertation, [16] in 2004, a biologically inspired cognitive radio engine that relies on Genetic Algo-

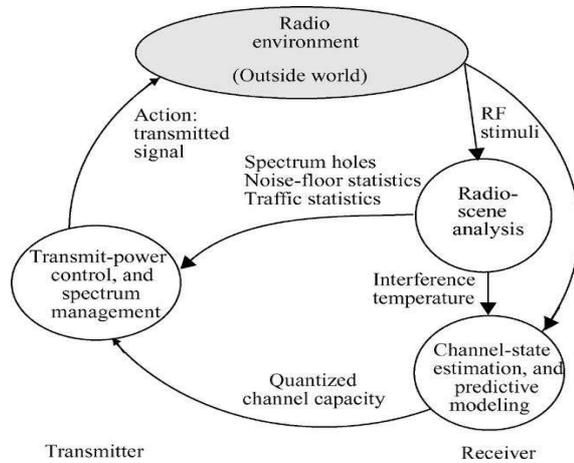


Figure 1.6: Cognitive radio decision making context [12]. The Cognitive Cycle as introduced by S. Haykin to answer Dynamic Spectrum Access related problems.

rithms. To the best of our knowledge, it was the first suggested and partially implemented Cognitive Radio Engine presented to the community.

1.3 Ph.D. motivations

1.3.1 SDR and CR related topics

Within SDR and CR contexts, The SCEE lab at Supélec defined and explored, among others, two main concepts related to this Ph.D.:

1. The Sensorial Radio bubble (SRB) [17, 18, 19].
2. Hierarchical and Distributed Cognitive Architecture Management (HDCRAM) [20].

On the one hand, the SRB models virtual bubbles through which radio equipment observe various metrics related to their surrounding environment. As described in [19], it *relies on a plurality of sensors based on several signal processing elements. It gives to communication systems the ability to explore the radio environment in order to provide knowledge of the spatial and spectrum environment, and some context awareness. Such a CR system knows all about the signals coming inside and going outside its bubble, as well as the state of many parameters inside the bubble.* Thus, it describes a concept that aims at exploring and extracting useful information regarding the radio device. It can be seen as a sensorial agent that gathers information on the environment and sends them to an adequate analysis center for further processing.

On the other hand, the HDCRAM concept tackles both reconfigurations management and distributed decision making related issues. Both topics arise when dealing with SDR equipment and/or cognitive radios. As a matter of fact, the SDR concept allows the reconfiguration of a large set of parameters. In theory, it enables the reconfiguration of a whole communication chain on the fly. The reconfiguration process needs however to be transparent from the user's point of view. To answer this challenge a reconfiguration and decision

making architecture named HDCRAM has been suggested as a possible solution to allow distributed decision making and reconfiguration operations. The HDCRAM architecture relies on three reconfiguration and decision making levels denoted (from the highest level): $L1$, $L2$, $L3$ as illustrated in Figure 1.7. Figure 1.7 shows first the basic operators composing any HDCRAM system. Generically, we find two classes: a reconfigurable operator and an operator referred to as sensor. Note that on the one hand a reconfigurable operator receives orders from its dedicated $L3_ReMU$. On the other hand the defined *sensor* can be any part of the system as long as it provides a given level $L3_CRMU$ with a metric. All blocks $(\cdot)_CRMU$ deal with decision making and send their decisions to the adequate $(\cdot)_ReMU$ on their same level. This latter is the only one dealing with reconfiguration matters as explicitly illustrated at the center of the figure. Of course a certain operator can have both $L3_ReMU$ and $L3_CRMU$. Finally, when a decision made at a low level, i.e., at a local level, involves other parts of the system, the metrics are send to a cognitive unit at a higher level. For further details on the HDCRAM concept, we would suggest the following papers: [20, 21, 22, 23, 24].

During this Ph.D., I have been involved with both topics. However, the work presented in this dissertation can be seen as a part of the HDCRAM architecture, where we deal with local decision making. In other words, all suggested approaches in this dissertation are to be implemented into appropriate CRMU units.

The work realized during my thesis on the SRB can be found online. The four first papers concern the implementation of detection solutions on Universal Software Radio Peripheral (USRP) cards [25, 26, 27], and blind energy detection relying on the Expectation-Maximization algorithm applied on real measurements using the USRP cards [28] (An empirical evaluation). The last paper investigates the concept of hot-spot migration on flexible hardwares in the context of Green Cognitive Radio [29]. The results related to the aforementioned work are not reported in this dissertation. As a matter of fact, this report focuses on fundamental theoretical results on learning and decision making under uncertainty for CR, in general, and Opportunistic Spectrum Access in particular.

1.3.2 Research Objectives in this Ph.D.

At first, the goals of the thesis were threefold:

1. Exploration of the main decision making algorithms designed in the literature to tackle recent cognitive radio related decision making and learning problems.
2. Focus on low complexity learning algorithms capable of learning with (almost) no prior information.
3. Adaptation of the selected machine learning techniques to CR learning problems (e.g., understanding the impact of sensing errors on the performance of the algorithms applied to Opportunistic Spectrum Access).

Then, while exploring these matters new problems needed further investigation:

4. Quantifying energy detection uncertainty. Is it possible? If yes under what assumptions? These questions are very important to evaluate the impact of imperfect sensing on decision making and leaning.

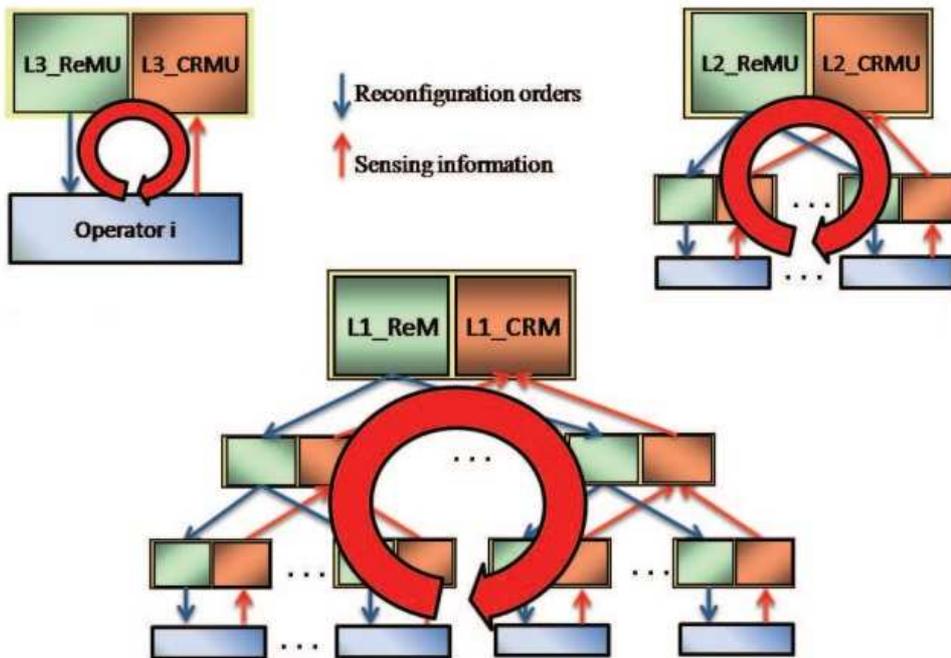
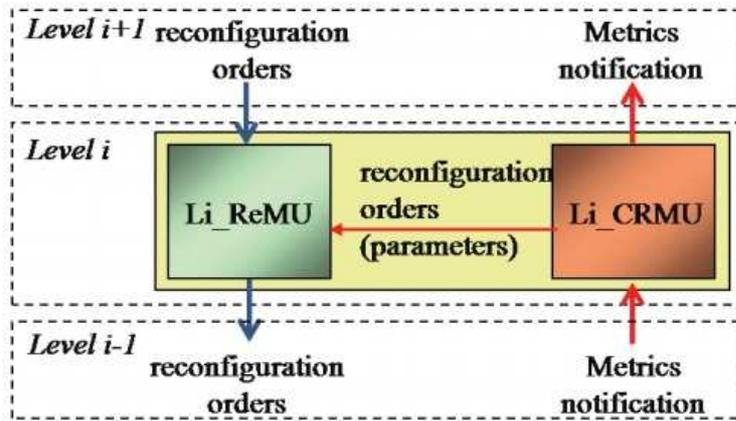
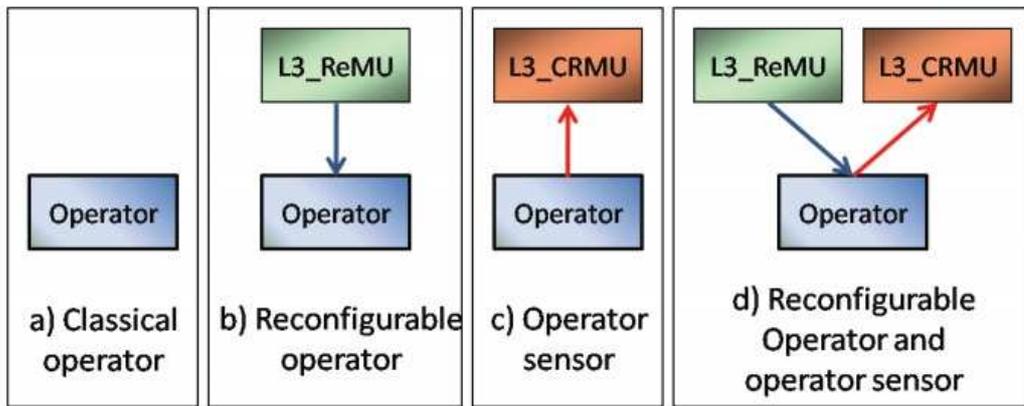


Figure 1.7: Hierarchical and Distributed Cognitive Radio Architecture Management (HDCRAM) [20, 21, 22, 23, 24]. A cognitive cycle management architecture is required to efficiently cycle through the observe, decide and adapt steps. The specificity of the CR context requires the management architecture to be hierarchical and distributed over several processing units. This figure illustrates the basic components and behavior of HDCRAM systems.

5. Collaboration and Coordination among different CR users.
6. Adapting or designing learning algorithms to different contexts such detection in fading environment.

We present in the next section the main contributions introduced in this dissertations and detail the organization of this document.

1.4 Presentation and results

The contributions of the Ph.D., presented in this report, answer the previously asked questions.

Chapter 2, explores CR related literature. It focuses on decision making and learning for CR. We synthesize all decision making problems in CR through the Dynamic Configuration Adaptation (DCA) Problem. Within this framework, Dynamic Spectrum Access (DSA) appears as a specific instantiation of DCA. While we briefly discuss DSA⁽¹⁰⁾, we focus on the literature that aims at defining Cognitive Decision Making Engines. Then we introduce the conceptual notion of *design space*. This latter constrains the set of possible decision making problems CR needs or might face. We argue that CR decision making problem share the same *design space*. Finally, relying on the notion of design space and the notion of *prior knowledge*, we suggest an original classification of decision making techniques for CR engines.

Thus, The main contributions of Chapter 2:

- Definition of the DCA problem.
- Introduction of the notion of *Design Space*.
- General classification of decision making techniques for DCA relying on the *design space* and *prior* knowledge.

The chapter 3 revisits the impact of noise uncertainty on the performance of the well known energy detector and present new results on its limits.

Thus, The main contributions of Chapter 3

- Introduction of a new measure of uncertainty related to the variance of noise estimation. The distribution of noise uncertainty is assumed to follow Log-Normal distributions.
- Analysis of Energy Detection characteristics with limited information on the noise level. It relies on a Log-Normal approximation of Chi-square distributions (Log Approximation). We define a new SNR_{wall} as a function of the desired performances and the uncertainty parameters.
- To support the Log Approximation, we evaluate the approximation error. Results suggest to approximate (when needed) Chi-square distributions by Log-normal distributions rather than by Normal distributions.

⁽¹⁰⁾It contains various very actives sub-topics that would be impossible to extensively report in this Ph.D.

This work highlights interesting information on the quality of the observations provided to the decision maker. One key element can be stated as follows: the designed decision making or/and learning algorithm must be able to operate without knowing the quality of the observations. As a matter of fact, such information is not always available as discussed in the case of Opportunistic Spectrum Access (OSA) in Chapters 4 and 5

The chapter 4 introduces, analyzes and discusses sequential learning applied to OSA. More specifically, we suggest modeling OSA problems relying on Multi-Armed Bandit models. Then we illustrate the performance of a low complexity algorithm known as UCB_1 to answer the designed academic OSA scenarios. We extend the theoretical results of UCB_1 in more realistic scenarios where the observation are soiled with errors due to imperfect sensing. We show that UCB_1 remains efficient and does not need prior knowledge on the sensor's performance.

Thus, The main contributions of Chapter 4

- Basic OSA scenarios modeled as MAB. We exploit UCB1 Algorithm to tackle them.
- Introduction of detection errors in OSA scenarios. We show that the learning process still converges to the most available channel (i.e. optimal channel).
- We show that UCB1 algorithm does not require prior knowledge on the detectors performances to converge to the optimal channel.
- Simulink based illustrations of our considered OSA scenario.

Several question arise from this chapter. More specifically:

- How to learn in Secondary Networks? Dealing with coordination and collaboration?
- How to deal with more complex scenario involving channel fading?

These questions are answered, respectively in Chapters 5 and 6. As a matter of fact, their contributions are respectively:

Chapter 5:

- OSA problems are Modeled as Job Assignment under uncertainty. We suggest combining Coordination Algorithms to UCB_1 learning mechanism.
- Convergence results extended to the case of Multi-Secondary Users.
- We discuss the case of Heterogeneous Networks.

Chapter 6:

- We designed a multiplicative UCB form to deal with Gamma distributions.
- We prove the convergence of MUCB algorithm in the case of exponential distributions.
- Application: channel selection in fading environments.

This work has the potential to deal with several problems motivated by CR Networks or Network Optimization under uncertainty.

Chapter 2

Decision Making and Learning for Cognitive Radio

Contents

1.1	Twenty years of Wireless Communications	6
1.1.1	The emergence of licensed cellular networks	6
1.1.2	WLAN and unlicensed standards: the success of the WiFi standard	6
1.1.3	Cooper's law and the physical layer's limits	11
1.2	Towards Cognitive Radio	13
1.2.1	Software Defined Radio	13
1.2.2	The rise of Cognitive Radio	14
1.3	Ph.D. motivations	16
1.3.1	SDR and CR related topics	16
1.3.2	Research Objectives in this Ph.D.	17
1.4	Presentation and results	19

2.1 Introduction

Due to the plethora, and constantly increasing, number of publications related to decision making applied to CR, we do not intend to present an exhaustive state-of-the-art on this topic. However, for consistency reasons, we suggest an analysis of the main decision making problems tackled by the radio community during the first decade of CR research. We propose to refer to the general problem as Dynamic Configuration Adaptation (DCA). We define the problem and show that practically speaking it can be declined into two main topics. On the one hand, the first topic aims at finding an adequate configuration (code or modulation adaptation for instance) adapted to the channel used for the transmission. In this topic various decision making tools were suggested. Mainly, these techniques were borrowed from the Artificial Intelligence community. The first topic is usually terminal centric. Namely, the interaction of several users is not considered or is assumed implicit. The second Topic on the other hand, specifically tackles the urgent problem of efficient spectrum allocation. This topic is usually referred to as Dynamic Spectrum Access (DSA) problem. Once again a brief state-of-the-art is provided in this specific case. For the occasion, we describe the main possible axis proposed by the community and look deeper into Opportunistic Spectrum Access (OSA) problems, a sub-topic of DSA. Note that in this case both terminal and network centric problems are considered. OSA problems are further discussed in Chapters 4, 5 and 6.

The main contributions of this chapter are three fold: On the one hand we introduce the concept of *design space*. It is presented as a conceptual object that defines a set of cognitive decision making problems by their constraints rather than by their degrees of freedom. On the other hand, relying on the notion of *design space* as well as prior knowledge on the environment, we suggest a qualitative classification of decision making techniques for Dynamic Configuration Adaptation problems. As a matter of fact we argue that all CR related decision making problems can be embedded, as a first approximation, in the same DCA framework. Finally, we briefly describe the main approaches presented in the classification: Expert approaches, Exploration based approaches and finally learning and partial monitoring based approaches.

The outline of the rest of this chapter is the following: Section 2.2 revisits extensively the definitions and the main concepts related to CR and introduces the the basic cognitive cycle. Then Section 2.3, discusses decision making for CR. More specifically, it introduces the notion of design space as well as the general DCA Framework. Moreover, Dynamic Spectrum Access, is discussed as a specific instantiation of DCA problems. Section 2.4 discusses the importance of prior knowledge to determine the decision making tool to tackle a specific CR application. Then, it briefly reviews the main approaches suggested in the literature to design CR decision making engines. Finally Section 2.5

2.2 Cognitive Radio

In this section, we remind the readers of various definitions for Cognitive Radio. These definitions usually depend on the context of application the authors intend to tackle. Then we describe the basic cognitive cycle well known in the AI community [30] and discuss it in the case of CR contexts.

2.2.1 Definitions

Since the original definition suggested by Joseph Mitola III, several other definitions were proposed to define the edges of Cognitive Radio. We remind the reader in the next few paragraphs of the main definitions found in the literature [8]:

Definition 1 (Cognitive Radio, by J. Mitola). *The term cognitive radio identifies the point at which wireless personal digital assistant (PDAs) and the related networks are sufficiently computationally intelligent about radio resources and related computer to computer communication to:*

1. *Detect user communication needs as a function of use context, and*
2. *Provide radio resources and wireless services most appropriate to these needs.*

In 2005, F. K. Jondral [31] suggested a definition that insists on the one hand, on a tight relationship between SDR technologies and CR paradigm and on the other hand, on the importance of information exchange among different CRs. This definition however keeps its generality and do not seem to tackle a particular application:

Definition 2 (Cognitive Radio, by F. K. Jondral [31]). *A CR is an SDR that additionally senses its environment, tracks changes, and reacts upon its findings. A CR is an autonomous unit in a communication environment that frequently exchanges information with the networks it is able to access as well as with other CRs.*

Whether CR is necessarily based on SDR devices is still a matter of debate. The evolution from SDR to CR seems relevant; however, CR can be seen as a paradigm that allows the design of a general purpose decision making engine. This latter adapts then its strategies to the flexibility of the equipment it is running on. Such vision is presented in the work of C.J. Rieser [16] and T.W. Rondeau [32, 33]⁽¹⁾. This approach has its pros and cons. On the one hand it provides a general optimizer that can transform any adaptable radio into a decent cognitive radio. On the other hand, for the same reasons it must be recurrently updated to face new radio designs and capabilities. Eventually this approach could lead to very complex and heavy systems that would probably be underutilized by their host radio equipment. Thus dimensioning the decision making capabilities is an important task. We discuss this problem later in this chapter while introducing the notion of *Design Space*.

F. K. Jondral's definition further stresses the importance of communication and information exchange between a CR and its surrounding environment, viz., *the networks it is able to access as well as with other CRs*. Information exchange is usually synonym of communication overhead and loss of throughput; however if the information exchange enables interference mitigation and avoids conflicts, it is worth the time and energy spent on it.

Anticipating later chapters, we shall take this point of view into account when dealing with multi-users CR networks. In general we address scenarios where collaboration is

⁽¹⁾From Virginia Tech under the supervision of J. Reed. Their approach based on Genetic Algorithms is further discussed later in this Chapter.

considered among the CR users (e.g., interference avoidance policies for instance when accessing frequency bands resources). However it does not necessarily imply information exchange.

During the same year 2005, the Federal Communications Commission (FCC) [34]⁽²⁾, in the United-States, and S. Haykin [12], respectively, suggested more pragmatic definitions that aim at defining cognitive radio as a possible mean to enable better spectrum use:

Definition 3 (Cognitive Radio, FCC 2005 [34]). *A Cognitive Radio is a radio that can change its transmitter parameters based on interaction with the environment in which it operates.*

Definition 4 (Cognitive Radio, S. Haykin 2005 [12]). *Cognitive radio is an intelligent wireless communication system that is aware of its surrounding environment (i.e. its outside world), and uses the methodology of understanding-by-building to learn from the environment and adapt its internal states to statistical variations in the incoming RF stimuli by making corresponding changes in certain operating parameters (e.g. transmit power, carrier-frequency and modulation strategy) in real-time, with two primary objectives in mind: highly reliable communications whenever and wherever needed and efficient utilization of the radio spectrum.*

Generally, the parameters considered in this definitions refer to, the transmission frequency, the modulation scheme, the bandwidth or/and the power allocated to each user for instance.

More recently, in 2009, the ITU [35] also suggested a general definition, that appears to synthesize both the definitions proposed by the normalization task force P1900.1 and the European Telecommunications Standards Institute (ETSI) [8]:

Definition 5 (Cognitive Radio, ITU). *Cognitive Radio System (CRS) is a radio system employing technology that allows the system to obtain knowledge of its operational and geographical environment, established policies and its internal state ; to dynamically and autonomously adjust its operational parameters and protocols according to its obtained knowledge in order to achieve predefined objectives ; and to learn from the result obtained.*

The definition proposed by the ITU is the closest to the one considered in our work and introduced in [36]⁽³⁾:

Definition 6 (Cognitive Radio, W. Jouini). *Cognitive Radio presents itself as a set of concepts and technologies that enable radio equipments to have the **autonomy and the***

⁽²⁾(<http://transition.fcc.gov/aboutus.html>About) About the FCC: The Federal Communications Commission (FCC) is an independent United States government agency. The FCC was established by the Communications Act of 1934 and is charged with regulating interstate and international communications by radio, television, wire, satellite and cable. The FCC's jurisdiction covers the 50 states, the District of Columbia, and U.S. possessions

⁽³⁾It is worth mentioning that the submission of our considered paper is prior to the publication of the ITU's CR definitions

cognitive abilities to become aware of their environment as well as of their own operational abilities.

The purpose of this new concept is to **meet the user's expectations** and to maximize operators' resources usage (e.g. spectral resource allocation) without compromising the efficiency of the network.

Thus, it presupposes the capacity to **collect information** from its surrounding environment (perception), to digest it (**learning and decision making problems**) and **to act** in the best possible way by considering several constraints (equipment parameters, regulations, enforcement policies and so on).

This definition centers on three axis:

- First, it introduces the new characteristics of a CR device: autonomy and cognitive abilities. These characteristics are carried out by a decision making engine that can be seen as the brain of the equipment. Since it is commonly accepted that such engine -also referred to as Cognitive Agent (CA) or Cognitive Engine (CE)- is a software program, the device can be seen as a software agent.
- Then, it defines the purpose of such agents: optimization of the radio device depending on the objectives defined (explicitly or not) by the user while constrained by the network. Consequently, the agent is an intelligent agent.
- Finally, the agent's abilities are emphasized through a cognitive cycle: perception - analysis and decision making - action. We refer to this cycle as basic cognitive cycle.

2.2.2 Basic Cognitive Cycle

Defining *cognition* is, in general, a harsh task. In the context of CR, basic cognitive abilities are considered:

- *environment perception* (or *Observation*)
- and *reasoning* (or *Analysis/Decision*).

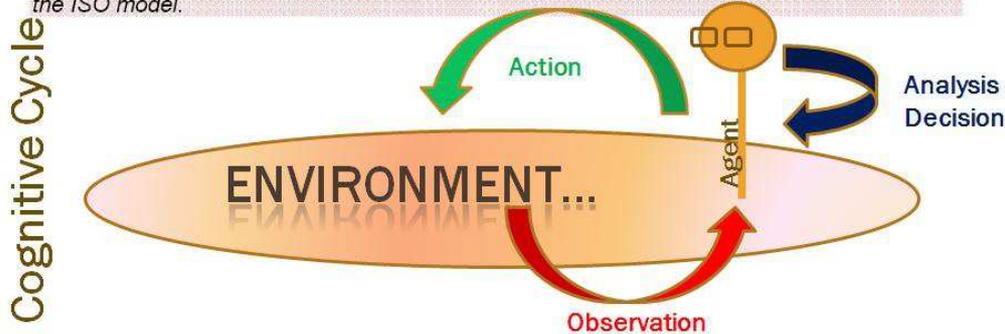
Based on these cognitive abilities, a CR needs to take appropriate *actions* to adapt itself to its surrounding *environment*.

Once again these notions know several possible definitions that we do not explicit in this report. However, the basic cognitive cycle considers three macro-steps as illustrated in Figure 2.1 and that we can define as follows:

1. **Observation:** through its sensors the agent gathers information on its environment. Raw data and preprocessed information helps the agent to build a knowledge base. In this context, the term environment is used in a broad sense referring to any source of information that could improve the CR's behavior (internal state, interference level, regulators' rules and enforcement policies, to name a few).
2. **Analysis/Decision:** This macro-step, presented as a black box in this case, includes all needed operations before given specific orders to the actuators (i.e., before reconfiguration in CR contexts). Depending on the level of sophistication, this step can deal with metric analysis, performance optimization, scheduling and learning.

Cognitive Radio :

Summarising Mitola, a full CR can be defined as "...a radio that is aware of its surroundings and adapts intelligently". This may require adaptation and intelligence at all the 7 layers of the ISO model.



A full CR will also use long-term analysis to learn about its environment and its own behaviour.

Figure 2.1: Illustration of the basic Cognitive Cycle [38, 39]. As illustrated, an agent, usually referred to as Cognitive Agent (CA) faces an Environment in a broad sense. The CA repeats the Cognitive Cycle where he *Observes* the environment, *Analyzes* the collected information and *Decides* the next action to take. Note that the arrow **Action** could suggest always an action on the environment. This is possible to evaluate the reaction on the environment to given stimuli. However, the arrow also suggests an action on the CR in order to adapt to the environment

3. **Action:** Mainly parameter reconfiguration and waveform transmission. A reconfiguration management architecture needs to be implemented to ensure efficient and quick reconfigurations [37].

This definition is quite general. It can incorporate simple designs as well as complex ones. Most of the published papers deal however with a restricted problem : spectrum management. In such context, the term environment finds more specific definitions such as the followings to name a few.

An environment can refer to:

- Geolocation [40], [41], [42], [43].
- Spectrum Occupation [44], [45], [46], [47], [48].
- Interference level (or Interference Temperature [12]).
- Noise level uncertainty [49], [50], [51].
- Regulatory rules (that define the open opportunities [13] for instance).

Thus, depending on the considered environment, specific sensors are to be designed [8, 19, 17]. The captured -and/or computed- metrics by the sensors are then processed by the decision making engine. The kind of process highly depends on the quality of

the metrics (level of uncertainty on the captured numerical value for instance) as well as the global information held by the CR. Finally, the made decisions are translated to appropriate bandwidth occupation and power allocation actions.

2.3 Decision Making Problems for Cognitive Radio

Within the basic cognitive cycle, we focus in this section on the *analysis* step, and more specifically on learning and decision making. We mainly find, in the literature two approaches.

On the one hand, some of the papers focus on implementing smart behavior into radio devices to enable more adequate configurations, adapted to their environment, than those imposed by radio standards. As a matter of fact, standard configurations are usually over dimensioned to meet the requirements of various critical communication scenarios. This approach mainly focuses on a single equipment, ignoring the rest of the network. We refer to the problem related to the first approach as *Dynamic Configuration Adaptation Problem*.

On the other hand due to a more pressing matter, most of CR related papers focus on spectrum management. These latter papers aim at enabling a more efficient use of the frequency resources to alleviate its scarcity. This second problem is usually referred as *Dynamic Spectrum Access Problem*

2.3.1 Design Space and Dynamic Configuration Adaptation Problem

In this subsection, we discuss some of the limits related to the idealized CR concept before introducing the so called DCA problem. Several questions arise when designing a Cognitive Radio Engine. We summarize our conceptual approach, presented in paper [52], to dimension the decision making and learning abilities of a Cognitive Engine. Thus, we introduce the notion of *design space* as a conceptual object that defines a set of cognitive radio decision making problems by their constraints rather than by their degrees of freedom. We identified, in our analysis work, three dimensions of constraints: the environment's, the equipment's and the user's related constraints.

Ideally speaking, CR concept -supported by an ideal SDR platform- opens the way to *infinite* possibilities. Autonomous and aware of its surrounding environment as well as of its own behavior (and thus of its own abilities), any part of the radio chain could be probed and tested to evaluate its impact on the device's performance. This however implies that the equipment is also able, in its reasoning process, to validate its own choices. Namely, it must self-reference its cognition components [53]. Unfortunately, *this class of reasoning is well known in the theory of computing to be a potential black hole for computational resources. Specifically, any Turing-capable (TC) computational entity that reasons about itself can enter a Gödel-Turing⁽⁴⁾ loop from which it cannot recover* [53].

To alleviate this paradox, time limited reasoning has been suggested by Mitola. As a matter of fact, radio systems need to observe, decide and act within a limited amount of time: *The timer and related computationally indivisible control construct is equivalent to*

⁽⁴⁾A specific example of such paradox can be illustrated by the following sentence: 'This sentence is false!' [54] as suggested by Mitola during a recent seminar at Supélec.

the computer-theoretic construct of a step-counting function over ‘finite minimalization’. It has been proved that computations that are limited with reliable watchdog timers can avoid the Gödel-Turing paradox to the reliability of the timer. This proof is a fundamental theorem for practical self-modifying systems [53].

Realistic CR frameworks need to take into account a large set of possible configurations, however, as mentioned hereabove through the Gödel-Turing paradox, the decision making engine also needs to be constrained in order to avoid the system to crash. Thus, we argue in the rest of this paragraph that, in general, cognitive radio decision making problems are better defined by their constraints rather than by their degrees of freedom.

When designing such CR equipments the main challenge is to find an appropriate way to correctly dimension its cognitive abilities according to its environment as well as to its purpose (i.e., providing a certain service to the user). Several papers in the literature have already been concerned by this matter however their description of the problem usually remained fuzzy (e.g., [10, 16, 33, 39, 55]). We summarize their analysis by defining three “constraints” on which the design of a CR equipment depends: First, the constraints imposed by the surrounding environment, then the constraints related to the user’s expectations and finally, the constraints inherent to the equipment. We argue that these constraints help dimensioning the CR decision making engine. Consequently, a *prior* formulation of these elements helps the designer to implement the right tools in order to obtain a flexible and adequate cognitive radio.

- **The environment constraints:** since a cognitive radio is a wireless device that operates in a surrounding communicating environment, it shall respect its rules: those imposed by regulation for instance (e.g., allocated frequency bands, tolerated interference, etc.) as well as its physical reality (propagation, multi-path and fading to name a few) and network conditions (channel load or surrounding users’ activities for instance). Thus the behavior of cognitive radio equipments is highly coordinated by the constraints imposed by the environment. As a matter of fact, if the environment allows no degrees of freedom to the equipments, this latter has no choice but to obey and thus loses all cognitive behavior. On the other side, if no constraints are imposed by the environment, the cognitive radio will still be constrained by its own operational abilities and the expectations of the user.
- **User’s expectations:** when using his wireless device for a particular application (voice communication, data, streaming and so on), the user is expecting a certain quality of service. Depending on the awaited quality of service, the cognitive radio can identify several criteria to optimize, such as, minimizing the bit error rate, minimizing energy consumption, maximizing spectral efficiency, etc. If the user is too greedy and imposes too many objectives, the designing problem to solve might become intractable because of the constraints imposed by the surrounding environment and the platform of the cognitive radio. However if the user is expecting nothing, then again there is no need for a flexible cognitive radio. Usually it is assumed that the user is reasonable in a sense that he accepts the best he could get with a minimum cost as long as the quality of service provided is above a certain level⁽⁵⁾.

⁽⁵⁾Note that this assumption introduces the notion of satisfactory behavior that we also refer to, in this report, as pragmatic behavior. We oppose it to rational thinking where the decision making engine always

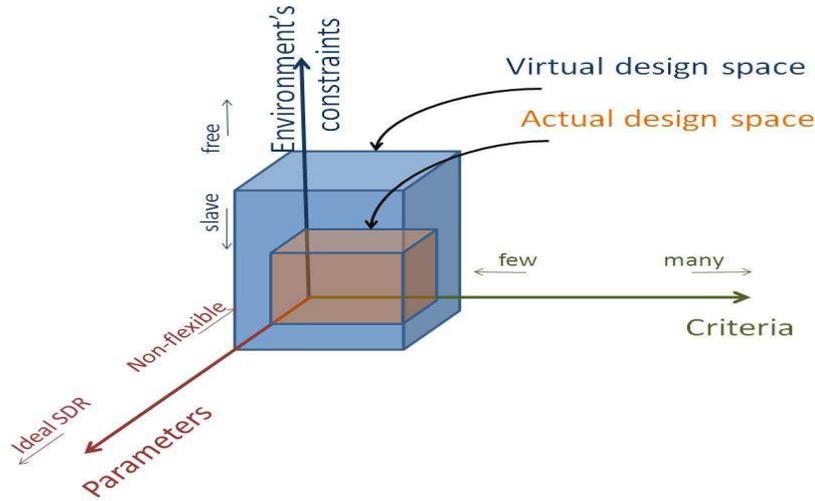


Figure 2.2: Cognitive radio decision making design space.

- Equipment's operational abilities:** These limitations are perhaps the most obvious since one cannot ask the cognitive radio equipment to adapt itself more than what it can perform (sense and/or act). It is usually assumed in the cognitive radio literature that the equipment is an ideal software defined radio, and thus, that it has all the needed flexibility for the designed framework. On a real application the efficiency of cognitive radio equipments depends of course on the degrees of freedom (or equivalently the constraints) inherent to the wireless platform used to communicate. As examples of commonly analyzed degrees of freedom one can find: modulation, pulse shape, symbol rate, transmit power, equalization to name a few. In all cases, a CR is designed to target and support given scenarios. We do not consider that CR can be designed to answer all scenarios or concepts [37].

The interaction between all three constraints is further emphasized through the notion of *design space*. We denote by *cognitive radio design space* an abstract three dimensional space that characterizes the CR decision making engine as shown in Figure 2.2. It is indeed abstract since it does not have any rigorous mathematical meaning but it is only used to visually and conceptually illustrate the dependencies of the CR decision making engine to the 'design dimensions': environment, parameters (usually referred to as knobs) and objectives (or criteria defined from the user's expectations).

In Figure 2.2, we represent two sub-spaces referred to as *actual design space* and *virtual design space*. On the one hand, the virtual design space refers to the upper bound support of the design space where all three dimensions are considered independently from each others. Its volume can be interpreted as the largest space of decision problems one could define from the three dimensions. On the other hand, the actual design space is included

aims at the most rewarding option. Thus when the decision making engine needs to learn in an uncertain environment, satisfaction based reasoning can be introduced to accelerate the convergence rate of learning algorithms for instance.

in the virtual design space. It results from the reduction of the design space when taking into account the correlation between the different constraints imposed by every dimension of the design space. For instance, some constraints on the environment such as, “imposed fixed waveform” might limit some objectives such as “find a waveform that maximizes the spectral efficiency”.

To define a specific decision making problem, one needs to introduce a last -possibly implicit- function. This latter represents a functional relationship between all three dimensions, more specifically the correlation between the different constraints as illustrated by the design space. Thus, it models the interdependence of all three constraints. A simple representation of this interdependence can be expressed through an explicit objective function which numerical value is computed as a function of the equipment parameters, the environment’s conditions as well as the values of other objective functions. Unfortunately such functions are not always available and might remain implicit. In such scenarios, optimization might prove problematic without using appropriate learning tools.

Finally, based on the hereabove presented analysis, all configuration adaptation problems seem to have the same roots. However, to define a specific problem among the set of possibilities in the design space, prior knowledge is important. This latter notion is further detailed in Section 2.4, where a classification of decision making tools as a function prior knowledge is suggested. Nevertheless, the general DCA problem can be described as the most general decision making design space that we can state as follows [36, 52]:

Within this framework, we assume that the environment constrains the cognitive radio by allowing only N possible configurations to use. This condition characterizes the environment and the equipment. Moreover we assume that there exist one or several objectives that evaluates how well the equipment performs to meet the users expectations.

To conclude, we usually observe in the literature that these constrained based characterizations are implicitly made, then final assumptions are done to define the decision making framework. These assumptions concern what we refer to as the ‘*a priori* model knowledge’. In Section 2.4, we introduce and explain the notion of *a priori* knowledge and we present a brief state of the art on decision making for cognitive radio configuration adaptation using the DCA design space. We show that although the design space is the same, depending on the *a priori* model knowledge, different approaches are suggested by the community to tackle the defined decision making problems.

The next section describes an important case of DCA know as Dynamic Spectrum Access.

2.3.2 Spectrum Scarcity and Dynamic Spectrum Access

Since the early 90’s, the radio community captured the potential industrial and economic opportunities that could emerge from a better frequency resource usage as noticed in 2004 in Paper [56]: *A trend that has the potential to change the current industrial structure is the emergence of alternative spectrum management regimes, such as the introduction of so called ‘unlicensed bands’, where new technologies can be introduced if they fulfill some very simple and relaxed ‘spectrum etiquette’ rules to avoid excessive interference on existing systems. The most notable initiative in this area is the one of the FCC (Federal Commu-*

nications Commission, the regulator in USA) in the early 90's driving the development of short range wireless communication systems and WLANs (Wireless Local Area Networks).

Exploiting portions of the spectrum to unlicensed usage was a first step to introducing alternative frequency management schemes. Rethinking the main regulatory frameworks imposed for decades is the next step. As a matter of fact, during the last century, most of the meaningful spectrum resources were licensed to emerging wireless applications, where the static frequency allocation policy combined with a growing number of spectrum demanding services led to a spectrum scarcity. However, several measurements conducted in the United-States first, and then in numerous other countries [11], [44], [45], [46], [47], [48], showed a chronic underutilization of the frequency band resources, revealing substantial communication opportunities.

With the advent of SDR technology, it became, at least theoretically, possible to design agile systems capable of switching from one frequency band to another depending on given communication constraints. Thus, during the years 2002 and 2003 several task forces and researches suggested new frequency management policies and regulatory frameworks to enable efficient use of the spectrum resource [11], [57], [58], [59], [58], [60], [61], [62]. *The consequences of this new framework are that the spectrum management model of today is abolished for large parts of the spectrum. Instead, 'free'⁽⁶⁾ spectrum trading becomes the preferred mechanism and technical systems that allow for the dynamic use and reuse of spectrum becomes a necessity* [56].

The DSA encompasses all suggested approaches that emerged from the early definitions of efficient and 'free' spectrum access or trading. In 2007, Paper [63] suggested one possible and simple taxonomy⁽⁷⁾ to classify the different suggested spectrum management approaches as illustrated in Figure 2.3. Three main approaches can be discriminated: Dynamic Exclusive Use Model, Open Sharing Model (Spectrum Commons Model) and Hierarchical Access Model:

- Dynamic Exclusive Use Model (DEUM): the spectrum basically is allocated exclusively to specific services or operators. However, the Spectrum Property Rights framework allows opening a secondary market where the licensed users can sell and trade portion of their spectrum, whereas the Dynamic Spectrum Allocation framework aims at providing a better allocation of the spectrum, to exclusive services, by adapting the spectrum allocation to space and time network load information.
- Open Sharing Model (OSM) or Spectrum Commons Model (SCM): Aims at generalizing the success encountered by WLAN technologies within the ISM band. In other words, it mainly suggests opening new portions of the spectrum to unlicensed users.
- Hierarchical Access Model (HAM): this framework introduced a secondary network that aims at exploiting resources left vacant by the incumbent users (usually referred to as primary users). Secondary users are able to communicate as long as they do not cause harmful interference to primary users. In this report, we do not further subdivide this framework. As a matter of fact, there are as many subsets

⁽⁶⁾[...] 'trade, lease and rent of licenses were possible without incurring excessive administrative procedures and overhead costs' [56].

⁽⁷⁾A different, more detailed and more exhaustive, DSA taxonomy can be found in Paper [64].

as the possible communication opportunities to exploit: power control, ultra-wide band communication beneath primary users' noise level, spectrum hole detection and exploitation, and directional communications to name a few [13]. In general, it is referred to as OSA.

Since the seminal paper of S. Haykin [12] in 2005, OSA research community has been, to the best of our knowledge, the most active in the field of DSA. With several network models based on Game Theory [15], Markov Chains optimization or Multi-Armed Bandit (and machine learning in general) [65, 63, 66, 36, 67, 68, 69, 70, 71], to reference a few, and relying on the concept of cognitive radio, the community tackled several challenges encountered when dealing with OSA such as (non exhaustive): dynamic power allocation, optimal band selection (with or without prior knowledge on the occupancy pattern of the spectrum bands by primary users), as well as cooperation among the different secondary users [14] centralized or decentralized, with or without observation errors.

In the next section, we introduce *prior knowledge* as a classification criteria among the main (non exhaustive) learning and decision making tools suggested in CR papers.

2.4 Decision Making Tools for Dynamic Configuration Adaptation

The *a priori* knowledge is a set of assumptions made by the designer on the amount and representation of the available information to the decision making engine when it first deals with the environment. As a matter of fact, "knowledge" is defined by the Oxford English Dictionary as: (i) *expertise, and skills acquired by a person through experience or education; the theoretical or practical understanding of a subject*, (ii) *what is known in a particular field or in total; facts and information or (iii) awareness or familiarity gained by experience of a fact or situation*. Consequently, within the cognitive radio framework, we can define the *a priori* knowledge as the set of *theoretical or practical* assumptions provided by the designer to the CR decision making engine. These assumptions, if they are accurate, provide the CR with valuable information on the problem to deal with. These remarks lead us to suggest that the decision making problems the cognitive radio has to deal with are defined by the set {design space, *a priori* knowledge}. In other words, depending on the *a priori* knowledge on the environment, some decision making approaches offer a better fit to the decision making framework than others. Moreover, we assert that a few, if not many, different cognitive engines could cohabit in a single cognitive radio equipment and will have to coordinate their actions [24]. Thus, recently (2011), a cognitive radio decision making engine based on prior knowledge has been suggested in [72], which supports our analysis.

In the next subsections we briefly describe the different approaches provided by the community depending on the *a priori* knowledge assumed relevant to tackle the environment the CR might face during its life time. In Figure 2.4 we suggest [52, 73] to classify these techniques depending on the *a priori* knowledge provided to the cognitive decision making engine.

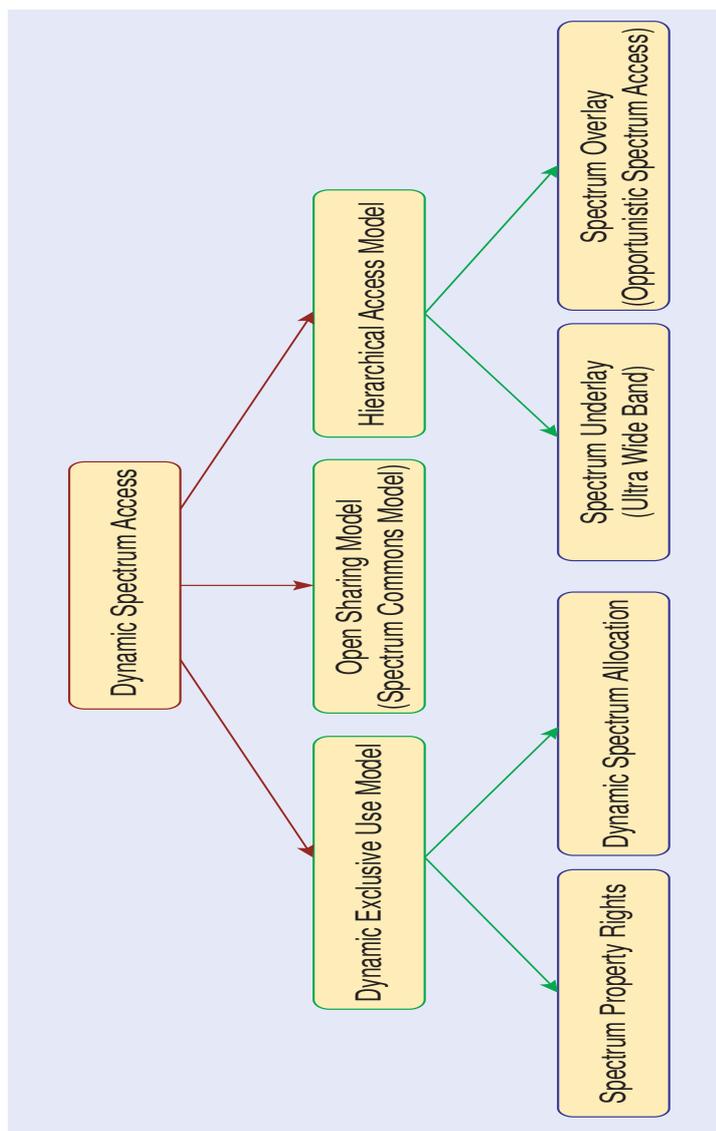


Figure 2.3: Classification of several Dynamic Spectrum Access approaches as suggested in Paper [63]. Three main approaches can be discriminated : Dynamic Exclusive Use Model, Open Sharing Model (Spectrum Commons Model) and Hierarchical Access Model.

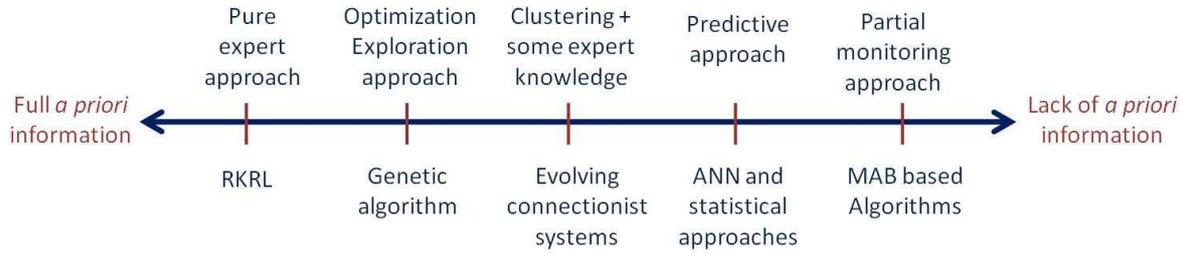


Figure 2.4: Suggested decision making techniques depending on the assumed *a priori* knowledge.

2.4.1 Expert approach

The expert approach relies on the important amount of knowledge collected by telecommunication engineers and researchers. This knowledge is based, on the one hand, on theoretical consideration and practical measures on the environment and radio communication parameters, and on the other hand, on structured set of wireless communication related concepts.

It was first suggested by Mitola in his Ph.D. dissertation on cognitive radio [10]. Through intensive off-line simulations, expert systems are provided with a set of inference rules. These rules are then used on-line to adapt the equipment depending on the context faced by cognitive radio equipments. Thus, the more available knowledge, the better the equipment can adapt itself to its surrounding dynamic environment. However, this knowledge is usefully as long as if the cognitive radio can represent its knowledge in a way that enables to exploit it and to react to the environment by adequate adaptations of its operating configuration.

For that purpose, Mitola suggested representing the knowledge of cognitive radio equipments using a new dedicated language radio communication: Radio Knowledge Representation Language (RKRL) [10, 53]. This representation of knowledge relies on web semantic related tools such as eXtensible Markup Language (XML), Resource Description Framework (RDF) and Web Ontology Language (OWL)⁽⁸⁾. The expert knowledge based approach had a large success especially due to the neXt Generation (XG) project supported by the DARPA (e.g. [75] and for spectrum sharing: [76]). As a matter of fact, if the knowledge is well represented and provided to the equipment as a set of rules, the decision making process becomes very simple. However this approach has a few drawbacks:

- The behavior of the designed system is not tuned to a particular user but to all users and to a set of probable environments. Moreover in order to acquaint the CR decision making engine with valuable and large knowledge, an important amount of effort is needed from the designer.
- Expert knowledge is mainly based on models. Thus the system might behave in a poor way when it is facing unexpected dynamics in the environment.

⁽⁸⁾Note that Mitola's work on OWL was published during the period 1999-2001. OWL specifications knew several improvements since then and it became recommended in 2004 by the World Wide Web Consortium (W3C) [74].

The techniques based on expert systems can, however be supported by several other tools (some are discussed later) to help them acquire new knowledge on the environment or help them avoid conflicts between different configuration adaptation rules.

A similar approach, based on an ontology to model the knowledge of the decision making engine was recently suggested [77, 78, 79, 80]. Where a common language to radio devices is suggested based on an ontology, expressed in OWL and implemented on the USRP card [81] using GNU radio [82].

2.4.2 Exploration based decision making

In some contexts, one can consider that there is *a priori* knowledge available on the complex relationships existing between, the metrics observed, the parameters to adapt and the criteria to satisfy as described in Figure 2.5. In this case the problem appears to be a multi-criteria optimization problem. Within this framework, the CR decision making engine aims at finding the best parameters to meet the users expectations by solving a set of equations as shown in Table II of paper [32] from which is extracted Figure 2.5). This problem is known to be complex for several reasons:

- there exists no universal definition of optimality in this case. Thus the solution of this problem are satisfactory (or not) with respect to a certain function, usually named *fitness* that evaluates how well the criteria were satisfied.
- Thus usually a large space of possible “satisfactory” configurations can be available.
- The criteria are correlated and can be in conflict (e.g., Figure 2.5).

If we assume that the previously mentioned off-line expert rule extraction phase has not been (or partially) accomplished, an exploration of the space of possible configurations is needed.

There exists various possible algorithm to explore a large set of potential candidates. The most obvious one is probably ‘exhaustive search’, where all possible candidates are computed and evaluated in order to find the best solution. However, when the number of candidates grows large, such approaches can become computationally burdensome and miss the imposed decision making deadlines. Usually in such contexts, heuristics are preferred.

In the context of Cognitive Radio, finding the best solution might not be necessary. Instead, the Cognitive Engine would rather find, within the imposed limited amount of time, a satisfactory solution.

Consequently, if the following criteria are met:

- Available *a priori* knowledge on the complex relationships existing between, the metrics observed, the parameters to adapt and the criteria to satisfy.
- Possible heavy parallel computing.

Then a large set of decision making tools are possible such as: Simulated Annealing (SiA), Genetic Algorithm (GA) and Swarm Algorithm (SwA) to name a few [30]. Note that such approaches did not wait for Cognitive Radio to be used on radio technologies. In

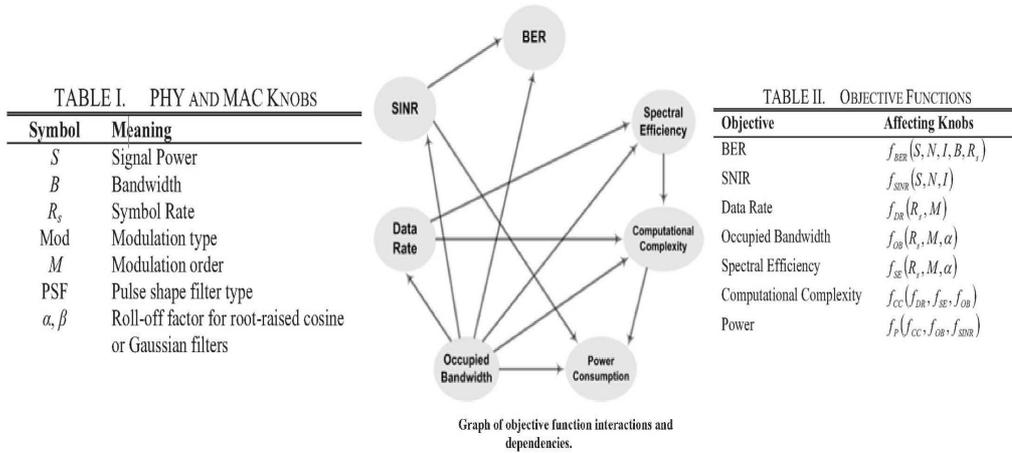


Figure 2.5: Multi-criteria optimization problem [32].

1993, Paper [83] already suggested simulated annealing as a possible solution to deal with channel assignment for cellular networks⁽⁹⁾.

GAs [16, 32, 33], SwAs [84, 85] and Insect Colony Algorithms (ICA) [86]⁽¹⁰⁾ techniques are usually referred to as bio-inspired or evolutionary techniques.

Bio-inspired cognitive radio decision making framework was first analyzed by C. J. Rieser and T. W. Rondeau. They suggested the use of Genetic Algorithms (GA) to tackle this framework [16], [32], [33]. Genetic algorithms were first designed to mimic Darwin's evolutionary theory and are well known for their capacity to adapt themselves to a changing environment. Without using our formalism, their work showed that under what we define as *design space* and with the described *a priori* knowledge, the genetic algorithms provide cognitive radios with an efficient and flexible decision making engine. But we can not consider their model as a generality for all cognitive radio use cases, so that other solutions have to be considered additionally. Further details on the different versions suggested and implemented by Virginia Tech can be found in the following recent survey [87]⁽¹¹⁾.

Note that once again prior knowledge can substantially enhance the behavior of these algorithms. An interesting illustration can be found in paper [72] in the case of genetic algorithms based decision making engines, where the authors showed how prior knowledge can improve the performance of GAs.

⁽⁹⁾It is indeed a very restrictive case of DCA and DSA where a centralized entity, seen as the Cognitive Agent assigns frequency channels to its users depending on the channel conditions.

⁽¹⁰⁾To the best of our knowledge Swarm Algorithms have only been exploited in case of resource allocation. No complex configuration adaptation decision making engine was found in the literature based on such techniques.

⁽¹¹⁾This document is presented as a survey of the various suggested decision making architectures for CR. We notice however, that except the one designed by J. Mitola, during the DARPA XG Program, and those designed and implemented by Virginia Tech, the community around this topic seems thin and advances slowly toward efficient architectures. Other suggested architectures relying mostly on bio-inspired techniques tackle spectrum resource allocation related problems.

2.4.3 Learning approaches: exploration and exploitation

As we argued in the previous subsections and as several other authors [88, 55] noticed, "Many CR proposals, such as [32, 89, 90], rely on a priori characterization of these performance metrics which are often derived from analytical models. Unfortunately, [...], this approach is not always practical due to e.g., limiting modeling assumption, non-ideal behaviors in real-life scenarios, and poor scalability" [88]. To avoid these limitations and in order to tackle more realistic scenarios, many methods based on learning techniques were suggested: Artificial Neuronal Networks (ANN), Evolving Connectionist Systems (ECS) [91, 92], statistical learning [93], regression models and so on. All of these approaches have their cons and pros, however they all have in common that they mainly rely on trials conducted within a real environment to try and infer from it decision making rules for CR equipments. Since this learning tools aim at representing the functional relationship between the environment (through the sensed metrics), the systems parameters and the criteria to satisfy, they need a direct interaction with the environment in order to build a *posteriori* knowledge on their environment.

In this work we sub-classify these methods depending on the way they learn and exploit their rules. On the one hand (i), we find a set of techniques that separates *exploration* and *exploitation* phases. On the other hand (ii), we find other techniques more flexible that combine both processes.

In the first mentioned case (i) we find several tools such as ANNs or statistical learning already used and exploited in other domain requiring some cognitive abilities (robotics, video games, etc.). These methods have two phases: a phase of pure "exploration" where the CR decision making engine learns and infers to find (explicitly or implicitly) decision making rules, then uses in a second phase this *a posteriori* knowledge to make decision. Since these learning techniques rely on a first learning phase, a large amount of data and computational power is needed in order to extract reliable knowledge. This difficulty is already known concerning ANNs for instance. It is still true for statistical learning. As noticed by Weingart in paper [93], the provided techniques are still computationally prohibitive, and not ready yet to be used in a real equipment. However if the first phase is well achieved the second phase is usually very simple and does not require much time or energy [89].

In the second case (ii), we find promising techniques recently introduced to the community and still need to be further investigated [55, 36]. These techniques try to provide the CR with a flexible and incremental learning decision making engine. In the case of ECS based decision making engine, Colson suggested the use of an evolving neural network [91, 92]. Unlike the usual ANN, the ECS-NN can change its structure without "forgetting" already learned knowledge. Thus new rules can be learned by adding new neurons to the neural structure. In order to be efficient the architecture proposed in [55] needs some expert advice (*a priori* knowledge) on the several available configurations. These added information ranks the different configurations based on some criteria (robustness, spectral efficiency, etc.) but without knowing *a priori* which one is more adequate when facing a certain environment.

More recently, we suggested in 2009 an approach to solving the problem without prior knowledge [36]. Thus, the performance of the equipment can only be estimated when trying a specific configuration. The associated tools are based on the so-called Multi-Armed

Bandit (MAB) framework. Such approaches provide learning solutions while operating, even if the CE is facing a completely new environment. Of course, performance increases while the learning process progresses. Note that this approach is also proving its accuracy in the opportunistic spectrum access (OSA) context [67, 68].

To conclude on this brief overview on decision making tools for CR, we would like to emphasize the fact that the proposed classification in this thesis shows that a CR equipment cannot depend on only one core decision making tool but on a pool of techniques. Every time it faces an environment, the equipment needs to have an estimation of its *a priori* knowledge and on its reliability. To tackle a particular context, the general process can be summarized through three questions: What can't I do (design space)? What do I already know (*a priori* knowledge)? And what technique should I select to solve the decision making problem?

Such a mixed approach can be efficiently handled by the HDCRAM architecture for instance.

Since the main goals of this Ph.D. aims at analyzing decision making and learning scenarios with minimum prior knowledge, we chose naturally to investigate an algorithm from the class of partial monitoring algorithms. The chosen algorithm is known as UCB_1 and is further described and analyzed in Chapters 4 and 5.

2.5 Conclusions

We analyzed in this Chapter decision making aspects related to CR. We showed that it is possible to model CR decision making problems as one general statement. We refer to this problem as DCA problem. Then, we ventured a classification of the main suggested tools to tackle CR decision making problems. Thus, we showed, through the notion of design space that all studies seem to tackle the same design space; however, the prior knowledge assumed available differs. As a consequence, the chosen decision making techniques also differ. Our classification offers a qualitative insight on the choices made by the community to tackle decision making problem within the context CR.

As illustrated through the notion of basic cognitive cycle, decision making and learning rely on prior observations of the environment. Consequently, the performance of the implemented decision making tools highly depends on the quality of the observations. Unfortunately, we could not find substantial quantitative material evaluating the impact of sensing errors on decision making and learning tools.

In the next chapters we deal with sensing and decision making under uncertainty. Thus, Chapter 3 analyses the performances of a very popular detector, known as energy detector, under uncertain noise level. Then Chapter 4 analyses the impact of imperfect sensing on a popular decision making algorithm known as UCB_1 algorithm.

Chapter 3

Energy detection limits under noise uncertainty and Log-Normal approximation of Chi-square distributions

Contents

2.1	Introduction	22
2.2	Cognitive Radio	22
2.2.1	Definitions	23
2.2.2	Basic Cognitive Cycle	25
2.3	Decision Making Problems for Cognitive Radio	27
2.3.1	Design Space and Dynamic Configuration Adaptation Problem	27
2.3.2	Spectrum Scarcity and Dynamic Spectrum Access	30
2.4	Decision Making Tools for Dynamic Configuration Adaptation	32
2.4.1	Expert approach	34
2.4.2	Exploration based decision making	35
2.4.3	Learning approaches: exploration and exploitation	37
2.5	Conclusions	38

Usually, learning algorithms assume perfect observations or labels. Thus, the designed learning mechanisms rely on relevant information to tackle their assigned tasks. Unfortunately, such assumptions no longer hold in engineering fields in general and within Cognitive Radio contexts in particular. Observation limits and uncertainties are major issues to which learning is confronted. Such matter is discussed in this chapter.

Mainly, this chapter is divided into two parts. On the one hand, we remind the reader that radio equipment sensors can lead to fundamental limits when it comes to signal detection. These limits, due to the lack of knowledge of the decision maker regarding the level of the noise (known as noise uncertainty), are particularly burdensome when dealing with Opportunistic Spectrum Access for instance. This chapter only presents such limits in the case of Energy Detectors. On the other hand, we revisit the impact of noise uncertainty on the performance of the well known energy detector and present new results on its limits. Mainly, we reconsider the case of a Log-Normal approximated noise uncertainty suggested in the work of Alexander Sonnenschein and Philip M. Fishman in 1992. We show that under a Log-Normal noise uncertainty, closed form expressions of the detector's performances and limits can be provided. Thus we show that, relying on mild approximations, we can design a detector with a fixed probability of false alarm function of the uncertainty, and present a new expression of the *SNR-wall*⁽¹⁾ that depends on the desired performances of the detector as well as the introduced uncertainty parameter.

⁽¹⁾Signal-to-Noise Ratio (SNR)

3.1 Introduction

The Neyman-Pearson Energy Detector (NP-ED) - also known as *Energy Detector* or *radiometric detector* - is a commonly used spectrum sensor. It has been extensively analyzed [94, 95] for its properties as a semi-blind low complexity spectrum sensor, since it ignores the characteristics of the received signals and only relies on the perceived energy of the signal. The main detection process relies on the comparison of the perceived energy to a fixed threshold that depends on the desired performances of the detector as well as the noise power level. However, despite its general assets, the NP-ED's performances decrease quickly in case of imperfect knowledge on the noise level [49, 50].

Recently, Energy Detectors have been the center of a lot of attention. They are mainly explored as possible low complexity alternatives [13, 96] to tackle Cognitive Radio detection problems. Usually, no *a priori* information is assumed available on the radio activity in the vicinity of the detector (e.g., in Opportunistic Spectrum Access[63, 67, 68] or spectrum measurement campaigns [46, 47]). Consequently, energy detection seems to offer simple and low complexity tools to probe CR equipment electromagnetic environment.

However, despite its general assets, the NP-ED's performances decrease quickly in case of imperfect knowledge on the noise power level [49, 50]. Moreover, even if an energy detector is combined with sophisticated signal processing techniques to extract information on the noise, it only results on stochastic estimations of the noise level leading to noise uncertainty. This led the radio communication community to address the problem of energy detection with noise uncertainty [49, 50, 97].

Thus, in their seminal paper, Alexander Sonnenschein and Philip M. Fishman [49] performed a worst case analysis on the performances of the energy detector in the case of imperfect knowledge on the noise level, referred to as *noise uncertainty*. As a matter of fact, depending on the information held by the decision maker on the noise level and its uncertainty, the analysis suggested in [49] relies on an upper-bound of the probability of false alarm and on a lower bound of the probability of detection. The main results showed that if the noise power level is only known through a confidence bound, there exists an *SNR-wall*, value of the Signal-to-Noise Ratio (SNR) beyond which detection is theoretically impossible.

The first contribution of this chapter is to reconsider the case of a Log-Normal approximated noise uncertainty as suggested in [49]. However, rather than reducing the analysis of the noise uncertainty to a bounded distribution, we suggest to redefine the uncertainty based on the estimated noise distribution's variance. This analysis, however, involves the knowledge of the probability density function of a ratio of χ^2 and Log-Normal random variables. Unfortunately, the considered ratio distribution does not seem to have a simple explicit form. To alleviate this mathematical problem, we suggest, as the second contribution of this chapter to evaluate a Log-Normal approximation of χ^2 distributions. This approximation is then used to develop our analysis on energy detection limits under Log-Normal uncertainty. Although, from this application's point of view, the Log-Normal approximation can seem as a convenient and opportunistic trick to by-pass the initial problem ; we show, relying on some mild calculus considerations, that from a mathematical point of view, the Log-Normal approximation of χ^2 distributions offers a better fit than the usually used Normal approximation of χ^2 distributions. Consequently, one should in general prefer a Log-Normal approximation rather than a Normal approximation to a χ^2

distributions (if needed). Finally, relying on the introduced Log-Normal approximation of χ^2 distributions, we tackle the initial problem and propose a new expression of the *SNR-wall* function of the uncertainty parameter. Moreover, unlike the previously introduced *SNR-wall* expression [49], the herein introduced *SNR-wall* formula takes into account the desired performances of the detector. Consequently, it enables an accurate evaluation of the detection limits.

In order to make this chapter as self-content as possible, we made the choice to simultaneously introduce both mentioned contributions due to their connexity: On the one hand, energy detection limits analysis under Log-Normal uncertainty, and on the other hand, a Log-Normal approximation of χ^2 distributions. As a matter of fact, without the analysis of Log-Normal approximations for χ^2 distributions, the first topic remains incomplete. While, the considered framework for energy detection limits under Log-Normal uncertainty offers a natural application for the second topic. However, the chapter is structured in a way that enables the reader to focus on one topic or the other of these contributions. Of course, we encourage the reader to follow the thread of the chapter.

The rest of this chapter is organized as follows: first, we start by presenting the general system model related to energy detection in Section 3.2. Then, we introduce and evaluate a Log-Normal approximation of χ^2 distributions in Section 3.3. The validation of the Log-Normal model for energy detection is then exploited in Section 3.4 where we analyze the performances of the ED in case of Log-Normal noise uncertainty. Several simulations illustrate and support the theoretical results of Section 3.3 and 3.4. These illustrations are presented and commented in their related sections. Finally Section 3.5 concludes this chapter.

3.2 System model

In this section, we introduce the detection characteristics usually considered when dealing with an Energy Detector, as well as the detection limits of the Energy Detector.

3.2.1 Network assumption

Let $\mathbf{y}_t = [y_{t,0}, y_{t,1}, \dots, y_{t,M-1}]$ be M independent and identically distributed (i.i.d.) samples gathered by the receiver at the current slot $t \in \mathbb{N}$. The outcome of the sensing process can be modeled as a binary hypothesis⁽²⁾ test described as follows:

$$\mathbf{y}_t = \begin{cases} \mathbf{n}_t, & \mathbb{H}_0 \\ \mathbf{x}_t + \mathbf{n}_t, & \mathbb{H}_1 \end{cases}$$

where hypotheses \mathbb{H}_0 and \mathbb{H}_1 refer respectively to the case of an absent or a present signal on the current slot. On the one hand, $\mathbf{x}_t = [x_{t,0}, x_{t,1}, \dots, x_{t,M-1}]$ refers to the source signal where every sample $x_{t,k}$ is perceived as an i.i.d. realization of a Gaussian stochastic distribution $\mathcal{N}(0, \sigma_{x,t}^2)$. On the other hand, $\mathbf{n}_t = [n_{t,0}, n_{t,1}, \dots, n_{t,M-1}]$ refers to i.i.d. Additive White Gaussian Noise (AWGN) samples $\mathcal{N}(0, \sigma_{n,t}^2)$. Moreover, \mathbf{x}_t and \mathbf{n}_t are assumed to be independent. Thus, we consider the following Gaussian received signals under either hypothesis $\forall y_{t,i} \ i \in \{0, \dots, M-1\}$:

$$\begin{cases} \mathbb{H}_0 : y_{t,i} \sim \mathcal{N}(0, \sigma_{n,t}^2) \\ \mathbb{H}_1 : y_{t,i} \sim \mathcal{N}(0, \sigma_{x,t}^2 + \sigma_{n,t}^2) \end{cases}$$

Within this context, the detection outcome can be modeled as the output of a decision making policy π that maps the current samples \mathbf{y}_t into a binary value $d_t = \pi(\mathbf{y}_t)$, $d_t \in \{0, 1\}$, where 0 refers to the possible absence of signal and reciprocally 1 indicates the detection of a signal.

It is worth mentioning that this work is motivated by the conclusions introduced in Papers [49] and [50]. For the sake of consistency, we chose to work with the same network assumptions, i.e., hypothesizes \mathbb{H}_0 and \mathbb{H}_1 follow both a central Gaussian distribution. Strictly speaking however, from Wireless Communications' perspective, samples under \mathbb{H}_1 should follow a non-central Normal distribution. Consequently, the work presented hereafter can be seen as a preliminary work that needs further investigations to answer the more general mathematical framework that models energy detection under uncertainty. This latter mathematical problem is a perspective of this thesis work and currently under investigation.

In the next subsection, we summarize the usually used criteria to evaluate the performance of a signal detection policy.

3.2.2 Performance evaluation of a detection policy π

Under the previously considered binary hypothesis test, one can define two probabilities that characterize the performance of the detection policy π at the slot number t : The probability of false alarm ($\mathbb{P}_{fa,t}$) and the probability of detection ($\mathbb{P}_{d,t}$):

$$\begin{cases} \mathbb{P}_{fa,t} = \mathbb{P}(d_t = 1 | \mathbb{H}_0) \\ \mathbb{P}_{d,t} = \mathbb{P}(d_t = 1 | \mathbb{H}_1) \end{cases}$$

⁽²⁾In this chapter, we associate the numerical value 1 to the existence of a signal to detect and 0 otherwise.

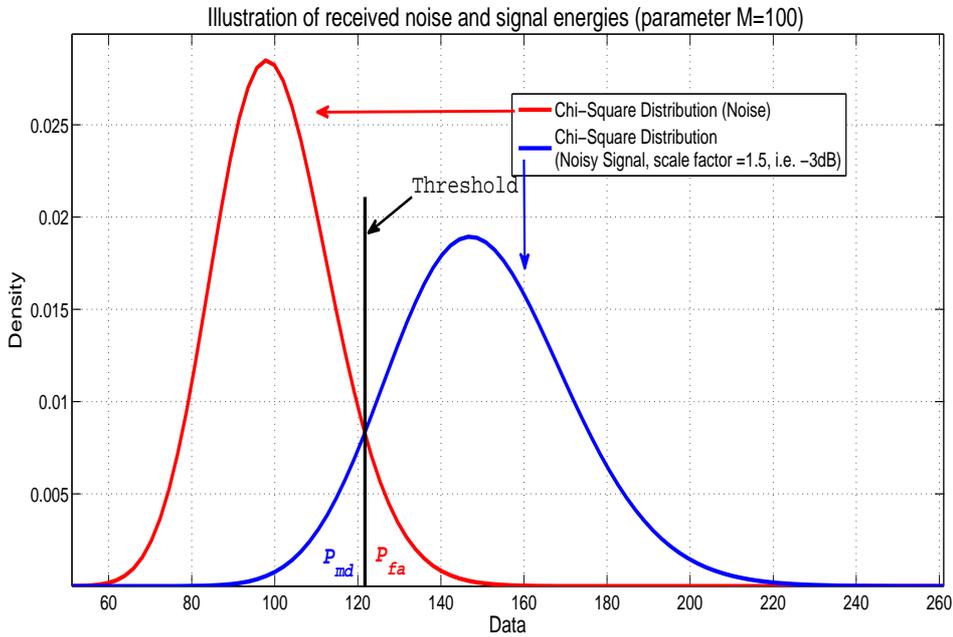


Figure 3.1: Illustration of the energy detection threshold based policy with two outcome classes \mathbb{H}_0 and \mathbb{H}_1 (in this case ‘data’ refers to the estimated power statistic). On the one hand, the first curve on the left refers to the probability density function of the observed energy of noise. On the other hand, the second curve on the right shows the probability density function of the observed energy of the signal.

Thus the performance of a detector highly depends on the distributions of the received samples in both environments, viz. pure noise and noisy signal. This last remark is illustrated in Figure 3.1. Indeed, it suggests an illustration of an energy detection threshold based policy with two outcome classes (\mathbb{H}_0 and \mathbb{H}_1). On the one hand, the first curve on the left refers to the probability density function of the observed energy of noise (for illustration purpose $\sigma_{n,t}^2 = 1$). On the other hand, the second curve on the right shows the probability density function of the observed energy of the signal (for illustration purpose $\sigma_{x,t}^2 = 0.5$, i.e., a signal-to-noise ratio equal to $-3dB$). Thus the probabilities of false alarm and miss detection, respectively $\mathbb{P}_{fa,t}$ and $\mathbb{P}_{md,t} = 1 - \mathbb{P}_{d,t}$ are equal to the surface integral under the density functions and limited by the threshold vertical line.

Usually, constraints impose to fix the $\mathbb{P}_{fa,t}$ under a given level α_{fa} , such that $\mathbb{P}_{fa,t} \leq \alpha_{fa}$. The most powerful decision policy is then defined as the one having the largest $\mathbb{P}_{d,t}$ value for a given $\mathbb{P}_{fa,t} = \alpha_{fa}$. Note that Figure 3.1 shows a particular scenario where the threshold is set at the intersection of both density functions such that the sum of the error probabilities, $\mathbb{P}_{fa,t} + 1 - \mathbb{P}_{d,t}$ is minimized in the case of equi-probable Hypotheses (i.e., in the case of Maximum Likelihood detection).

3.2.3 Neyman-Pearson Energy Detector

NP-ED assumes known the noise level $\sigma_{n,t}^2$ at every slot number t . For the sake of simplicity and without loss of generality, we consider in the rest of the chapter a constant noise level for all t , $\sigma_{n,t}^2 = \sigma_n^2$. Under these assumptions, the NP-ED is proven to be the most powerful test.

To make a decision on the presence or absence of a signal, NP-ED relies on the computation of the received energy statistic \mathcal{T}_t at the slot number t defined such as:

$$\mathcal{T}_t = \sum_{i=0}^{M-1} |y_{t,i}|^2$$

The decision policy π_{NP-ED} is a simple Heaviside function $\mathcal{H}(\cdot)$ that only depends on the evaluation of the statistic \mathcal{T}_t at the current slot t :

$$d_t = \pi_{NP-ED}(y_t) \iff d_t = \mathcal{H}(\mathcal{T}_t - \xi_t(\alpha_{fa}))$$

where $\xi_t(\alpha_{fa})$ is the selected threshold to guaranty $\mathbb{P}_{fa} \leq \alpha_{fa}$. Such policies are usually described using the following notation:

$$\mathcal{T}_t \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\leq}} \xi_t(\alpha_{fa})$$

The following equations remind us of the expressions of $\mathbb{P}_{fa,t}$ and $\mathbb{P}_{d,t}$ (where $\mathcal{T}_t \sim \chi_M^2$) as well as their approximations for large M (where \mathcal{T}_t is assumed to follow a Gaussian distribution):

$$\begin{cases} \mathbb{P}_{fa,t} = 1 - F_{\chi_M^2} \left(\frac{\xi_t(\alpha_{fa})}{\sigma_n^2} \right) \\ \mathbb{P}_{d,t} = 1 - F_{\chi_M^2} \left(\frac{\xi_t(\alpha_{fa})}{\sigma_n^2 + \sigma_{x,t}^2} \right) \end{cases}$$

where $F_{\chi_M^2}(\cdot)$ refers to the cumulative distribution function of a χ^2 -distribution with M degrees of freedom.

When the number of gathered samples is large enough ($M \geq 200$) Normal approximation of Chi-Square distributions is generally considered as satisfactory [94, 49] (note however that in general such approximations are not necessary):

$$\begin{cases} \mathbb{P}_{fa,t} \approx Q \left(\sqrt{\frac{M}{2}} \left(\frac{\xi_t(\alpha_{fa})/M}{\sigma_n^2} - 1 \right) \right) \\ \mathbb{P}_{d,t} \approx Q \left(\sqrt{\frac{M}{2}} \left(\frac{\xi_t(\alpha_{fa})/M}{\sigma_n^2 + \sigma_{x,t}^2} - 1 \right) \right) \end{cases}$$

where $Q(\cdot)$ is the complementary cumulative distribution function of Gaussian random variable (also known as *Marcum function*) [98]:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{y^2}{2}} dy$$

NP-ED provides satisfactory behavior when σ_n^2 is known. Unfortunately, when such knowledge is unavailable, its performances, through a worst-case analysis, is shown to significantly degrade [49, 50].

3.2.4 Energy detection with noise uncertainty

The authors of Paper [49] suggested to analyze the impact of noise uncertainty on the performances of an energy detector. Two models were discussed at different levels : on the one hand, they introduced a bounded model of the noise estimation. Within this framework, they performed a worst case analysis on the performances of the energy detector. Thus, they proved the existence of an *SNR-wall*, value of the SNR beyond which detection is theoretically impossible. On the other hand, they suggested a more realistic description of the noise power estimation as a Log-Normal distribution. However, its analysis would involve the knowledge of the Probability Density Function (PDF) of a ratio statistic composed of a χ^2 and a Log-Normal distribution. Since its analytical expression has no known simple form, the problem was reduced to a bounded distribution leading to the same hereabove stated results.

Mainly, to summarize, the bounded model for the noise uncertainty impose an upper bound on the probability of false alarm, while it analyzes the lower bound of the probability of miss detection. Which leads to an underestimation of the system capabilities, resulting in the assumed “fragile” behavior of energy detectors toward noise uncertainty.

In our work, we are motivated by the case of a Log-Normal approximated noise uncertainty, as suggested in [49]. As a matter of fact, in order to provide an accurate estimation of the behavior of an energy detector with imperfect knowledge, it is important to analytically approximate the hereabove described ratio statistic.

To that purpose, we suggest in the next Section to approximate the considered χ^2 distribution by an adequate Log-Normal distribution. Thus, this approximation reduces the problem to the analysis of a ratio of Log-Normal distributions.

3.3 Log-Normal Approximation of χ^2 distributions

In this section, we introduce and analyze a Log-Normal approximation of χ^2 distributions. The purposes of this approximation are twofold : on the one hand, it offers a better fit, than the usually used Normal approximation. While on the other hand, it offers a convenient mathematical solution to tackle ratios of χ^2 and Log-Normal random variables as it is further detailed in the next paragraphs.

3.3.1 Mathematical Model

Definition 7 (Distributions). *Let $f_{\chi_M^2}(\cdot)$, $f_{\mathcal{L}og\mathcal{N}(\mu_{\mathcal{L}},\sigma_{\mathcal{L}}^2)}(\cdot)$ and $f_{\mathcal{N}(\mu_{\mathcal{N}},\sigma_{\mathcal{N}}^2)}(\cdot)$ denote, respectively, the Probability Density Function (PDF) of a χ^2 distribution with M degrees of freedom, a Log-Normal distribution with parameters $\{\mu_{\mathcal{L}},\sigma_{\mathcal{L}}^2\}$ and a normal distribution with parameters $\{\mu_{\mathcal{N}},\sigma_{\mathcal{N}}^2\}$, such that:*

$$\begin{cases} f_{\chi_M^2}(x) = \frac{1}{2^{M/2}\Gamma(M/2)}x^{M/2-1}e^{-x/2}, x \in \mathbb{R}_+, 0 \text{ otherwise} \\ f_{\mathcal{L}og\mathcal{N}(\mu_{\mathcal{L}},\sigma_{\mathcal{L}}^2)}(x) = \frac{1}{x\sqrt{2\pi\sigma_{\mathcal{L}}^2}}e^{-\frac{(\log(x)-\mu_{\mathcal{L}})^2}{2\sigma_{\mathcal{L}}^2}}, x \in \mathbb{R}_+, 0 \text{ otherwise} \\ f_{\mathcal{N}(\mu_{\mathcal{N}},\sigma_{\mathcal{N}}^2)}(x) = \frac{1}{\sqrt{2\pi\sigma_{\mathcal{N}}^2}}e^{-\frac{(x-\mu_{\mathcal{N}})^2}{2\sigma_{\mathcal{N}}^2}}, x \in \mathbb{R} \end{cases} \quad (3.1)$$

Anticipating the analysis of energy detection limits under Log-Normal approximation, we consider in the rest of this chapter the following specific parameters such that:

$$\begin{cases} \{\mu_{\mathcal{L}}, \sigma_{\mathcal{L}}^2\} = \{\log(M) - \sigma_{\mathcal{L}}^2/2, \log(1 + 2/M)\} \\ \{\mu_{\mathcal{N}}, \sigma_{\mathcal{N}}^2\} = \{M, 2M\} \end{cases} \quad (3.2)$$

These parameters were chosen such that all three distributions have the same mean and variance.

Fact 1. ⁽³⁾[Convergence to Gaussian Distributions] Let $f_{\chi_M^2}(\cdot)$, $f_{\mathcal{L}og\mathcal{N}(\mu_{\mathcal{L}}, \sigma_{\mathcal{L}}^2)}(\cdot)$ and $f_{\mathcal{N}(\mu_{\mathcal{N}}, \sigma_{\mathcal{N}}^2)}(\cdot)$ be the distribution described in Definition 7, then for $k \rightarrow \infty$:

$$f_{\chi_M^2} \rightarrow f_{\mathcal{N}(\mu_{\mathcal{N}}, \sigma_{\mathcal{N}}^2)} \quad (3.3)$$

Moreover, since for M large enough, the expectation of the the Log-Normal distribution (equal to M by definition) is much larger than its standard deviation (equal to $\sqrt{2M}$ by definition), for $M \rightarrow \infty$, it converges to a Gaussian distribution with the same mean and variance, i.e., in this case and with respect to the previously introduced notations:

$$f_{\mathcal{L}og\mathcal{N}(\mu_{\mathcal{L}}, \sigma_{\mathcal{L}}^2)} \rightarrow f_{\mathcal{N}(\mu_{\mathcal{N}}, \sigma_{\mathcal{N}}^2)} \quad (3.4)$$

Thus we aims at evaluating the converging rates of the following error functions:

$$\begin{cases} \Delta_1(x) = f_{\mathcal{L}og\mathcal{N}(\mu_{\mathcal{L}}, \sigma_{\mathcal{L}}^2)}(x) - f_{\chi_M^2}(x) \\ \Delta_2(x) = f_{\chi_M^2}(x) - f_{\mathcal{N}(\mu_{\mathcal{N}}, \sigma_{\mathcal{N}}^2)}(x) \end{cases} \quad (3.5)$$

For that purpose, we propose to develop and approximate both functions $\Delta_1(x)$ and $\Delta_2(x)$ as polynomial series. This approach is further detailed in the rest of this subsection.

Definition 8 (Partial Taylor polynomial). Let $f_{\mathcal{D}}(\cdot)$ the PDF of a distribution $\mathcal{D} \in \{\chi^2, \mathcal{L}og\mathcal{N}\}$. We denote by $T_{\mathcal{D}, x_0}(\cdot)$ the following polynomial evaluated at the finite real point x_0 :

$$T_{x_0, \mathcal{D}}^{(n)}(x) = \frac{f_{\mathcal{D}}(x)}{f_{\mathcal{N}(\mu_{\mathcal{N}}, \sigma_{\mathcal{N}}^2)}(x)} = 1 + \sum_{j=0}^n C_{j, \mathcal{D}}(x - x_0)^j + \epsilon_{\mathcal{D}}^{(n)}(x) \quad (3.6)$$

where n is the approximation order, $\{C_{j, \mathcal{D}}\}_{j=0, \dots, n}$ are the polynomial components of the power series and $\epsilon_{\mathcal{D}}^{(n)}(\cdot)$ is an implicit function that contains the missing terms to respect the equality. $\epsilon_{\mathcal{D}}^{(n)}(\cdot)$ is very small compared to the other terms and converges to 0 as x tends to x_0 .

In the case of our analysis, the polynomial components of the considered series : $T_{M, \chi_M^2}^{(n)}(\cdot)$ and $T_{M, \mathcal{L}og\mathcal{N}(\mu_{\mathcal{L}}, \sigma_{\mathcal{L}}^2)}^{(n)}(\cdot)$ are regular functions of the parameter M . Their expression is usually very complex. However, since we are only interested in the asymptotic behavior of these functions, we simplify the general expression of these parameters using the first existing order of their polynomial expression evaluated as M tends to infinity such that for all $j = \{1, \dots, n\}$:

$$C_{j, \mathcal{D}}(M) = \sum_{i=0}^{\infty} \frac{c_{i, j, \mathcal{D}}}{M^i} \approx \frac{c_{i_0, j, \mathcal{D}}}{M^{i_0}} \quad (3.7)$$

where $i_0(j)$ is the first index in \mathbb{N} such that $c_{i_0(j), j, \mathcal{D}} \neq 0$.

⁽³⁾Although these well known properties are presented as facts, no clear references of anteriority could be found.

Definition 9 (Approximation and evaluated functions). Let $\tilde{C}_{j,\mathcal{D}}(M) = \frac{c_{i_0(j),j,\mathcal{D}}}{M^{i_0}}$ be the asymptotically approximated polynomial component, and $\tilde{T}_{x_0,\mathcal{D}}^{(n)}(\cdot)$ the partial approximation of Taylor series as defined in Definition 8:

$$\tilde{T}_{x_0,\mathcal{D}}^{(n)}(x) = 1 + \sum_{j=0}^n \tilde{C}_{j,\mathcal{D}}(x - x_0)^j \quad (3.8)$$

We evaluate in this chapter the asymptotic behavior of the following error functions for large M :

$$\begin{cases} \Delta_1(x) = f_{\mathcal{L}og\mathcal{N}(\mu_{\mathcal{L}},\sigma_{\mathcal{L}}^2)}(x) - f_{\chi_M^2}(x) \approx \left(\tilde{T}_{M,\mathcal{L}og\mathcal{N}(\mu_{\mathcal{L}},\sigma_{\mathcal{L}}^2)}^{(n)}(x) - \tilde{T}_{M,\chi_M^2}^{(n)}(x) \right) f_{\mathcal{N}(\mu_{\mathcal{N}},\sigma_{\mathcal{N}}^2)}(x) \\ \Delta_2(x) = f_{\chi_M^2}(x) - f_{\mathcal{N}(\mu_{\mathcal{N}},\sigma_{\mathcal{N}}^2)}(x) \approx \left(\tilde{T}_{M,\chi_M^2}^{(n)}(x) - 1 \right) f_{\mathcal{N}(\mu_{\mathcal{N}},\sigma_{\mathcal{N}}^2)}(x) \end{cases} \quad (3.9)$$

In the rest of this section, we focus on the analysis of the approximations :

$$\begin{cases} \tilde{\Delta}_1(x) = \left(\tilde{T}_{M,\mathcal{L}og\mathcal{N}(\mu_{\mathcal{L}},\sigma_{\mathcal{L}}^2)}^{(n)}(x) - \tilde{T}_{M,\chi_M^2}^{(n)}(x) \right) f_{\mathcal{N}(\mu_{\mathcal{N}},\sigma_{\mathcal{N}}^2)}(x) \\ \tilde{\Delta}_2(x) = \left(\tilde{T}_{M,\chi_M^2}^{(n)}(x) - 1 \right) f_{\mathcal{N}(\mu_{\mathcal{N}},\sigma_{\mathcal{N}}^2)}(x) \end{cases} \quad (3.10)$$

Relying on the previously introduced definitions, we present hereafter the main contributions related to the Log-Normal approximation of χ^2 distributions. Since, the analytical results provided rely on asymptotic approximations for large M , they cannot be generalized to small values of M . Some of the stated properties, are however empirically generalized and illustrated in the last part of this section.

Note that we do not present the following results as ‘‘Lemmas’’ or ‘‘Theorems’’: as a matter of fact, they are mainly based on heavy yet straightforward calculus⁽⁴⁾. Moreover, such calculus were possible because of some justified approximations that need, nonetheless, to be further investigated to validate them from a rigorous mathematical perspective. We, however, detail the mathematical protocols and approximations that led to the stated results when needed.

3.3.2 Main Results

We focus on an approximation of the third order $n = 3$ and analyze the extrema of the functions $\tilde{\Delta}_1(\cdot)$ and $\tilde{\Delta}_2(\cdot)$. The choice of a third order approximation was motivated by the necessity of obtaining analytical solutions for the extrema.

Property 1 (Approximated error functions). Let the function $\tilde{\Delta}_1(\cdot)$ and $\tilde{\Delta}_2(\cdot)$ two approximation errors function as defined in Equation 3.10, then we can show that:

$$\begin{cases} \tilde{\Delta}_1(x) = \left(-\frac{1}{6M} - \frac{x-M}{M} + \frac{(x-M)^2}{M^2} + \frac{(x-M)^3}{6M^2} \right) f_{\mathcal{N}(\mu_{\mathcal{N}},\sigma_{\mathcal{N}}^2)}(x) \\ \tilde{\Delta}_2(x) = \left(-\frac{5}{12M} - \frac{x-M}{2M} + \frac{5(x-M)^2}{8M^2} + \frac{(x-M)^3}{12M^2} \right) f_{\mathcal{N}(\mu_{\mathcal{N}},\sigma_{\mathcal{N}}^2)}(x) \end{cases} \quad (3.11)$$

⁽⁴⁾The results introduced hereafter were obtained relying on *Mathematica*. Consequently, they do not constitute proof. They are however introduced to support the suggested approximations of Chi-Square distributions by Log-Normal distributions. Nevertheless, further investigations are needed to rigorously confirm and proof these results.

Property 2 (Extrema : position and amplitude). *Let the real values $\{y_{1,i}(M)\}_{i=1}^4$ and $\{y_{2,i}(M)\}_{i=1}^4$ denote the approximated extrema amplitudes, of, respectively, $\tilde{\Delta}_1(\cdot)$ and $\tilde{\Delta}_2(\cdot)$ at the positions $\{x_1(M) < x_2(M) < x_3(M) < x_4(M)\}$. Then there exist two real constants $\{a, b\}$ such that:*

$\{a, b\} = \{4 - 2^{\frac{1}{3}} - 2^{\frac{2}{3}}, 2 + 2^{\frac{1}{3}} + 2^{\frac{2}{3}}\}$ such that for large M :

$$\begin{cases} x_1(M) \approx M + \left(-\sqrt{a} - \sqrt{b}\right) \sqrt{M} \\ x_2(M) \approx M + \left(\sqrt{a} - \sqrt{b}\right) \sqrt{M} \\ x_3(M) \approx M + \left(-\sqrt{a} + \sqrt{b}\right) \sqrt{M} \\ x_4(M) \approx M + \left(\sqrt{a} + \sqrt{b}\right) \sqrt{M} \end{cases} \quad (3.12)$$

Which leads to the following expressions for the approximated extrema of $\tilde{\Delta}_1(\cdot)$ for large M :

$$\begin{cases} \tilde{\Delta}_1(x_1(M)) \approx y_{1,1}(M) \approx \frac{e^{-\frac{1}{4}(\sqrt{a}+\sqrt{b})^2}}{24\sqrt{\pi}} \left(-\frac{2}{M}\sqrt{ab}(\sqrt{a}+\sqrt{b}) + \frac{5}{2M^{3/2}}(2+3(\sqrt{a}+\sqrt{b})^2)\right) \\ \tilde{\Delta}_1(x_2(M)) \approx y_{1,2}(M) \approx \frac{e^{-\frac{1}{4}(12-(\sqrt{a}+\sqrt{b})^2)}}{24\sqrt{\pi}} \left(-\frac{2}{M}\sqrt{ab}(\sqrt{a}-\sqrt{b}) + \frac{5}{2M^{3/2}}(38-3(\sqrt{a}+\sqrt{b})^2)\right) \\ \tilde{\Delta}_1(x_3(M)) \approx y_{1,3}(M) \approx \frac{e^{-\frac{1}{4}(12-(\sqrt{a}+\sqrt{b})^2)}}{24\sqrt{\pi}} \left(\frac{2}{M}\sqrt{ab}(\sqrt{a}-\sqrt{b}) + \frac{5}{2M^{3/2}}(38-3(\sqrt{a}+\sqrt{b})^2)\right) \\ \tilde{\Delta}_1(x_4(M)) \approx y_{1,4}(M) \approx \frac{e^{-\frac{1}{4}(\sqrt{a}+\sqrt{b})^2}}{24\sqrt{\pi}} \left(\frac{2}{M}\sqrt{ab}(\sqrt{a}+\sqrt{b}) + \frac{5}{2M^{3/2}}(2+3(\sqrt{a}+\sqrt{b})^2)\right) \end{cases} \quad (3.13)$$

As well as the following expressions for the approximated extrema of $\tilde{\Delta}_2(\cdot)$ for large M :

$$\begin{cases} \tilde{\Delta}_2(x_1(M)) \approx y_{2,1}(M) \approx \frac{e^{-\frac{1}{4}(\sqrt{a}+\sqrt{b})^2}}{12\sqrt{\pi}} \left(-\frac{2}{M}\sqrt{ab}(\sqrt{a}+\sqrt{b}) + \frac{1}{M^{3/2}}(-1+6(\sqrt{a}+\sqrt{b})^2)\right) \\ \tilde{\Delta}_2(x_2(M)) \approx y_{2,2}(M) \approx \frac{e^{-\frac{1}{4}(12-(\sqrt{a}+\sqrt{b})^2)}}{12\sqrt{\pi}} \left(-\frac{2}{M}\sqrt{ab}(\sqrt{a}-\sqrt{b}) + \frac{1}{M^{3/2}}(71-6(\sqrt{a}+\sqrt{b})^2)\right) \\ \tilde{\Delta}_2(x_3(M)) \approx y_{2,3}(M) \approx \frac{e^{-\frac{1}{4}(12-(\sqrt{a}+\sqrt{b})^2)}}{12\sqrt{\pi}} \left(\frac{2}{M}\sqrt{ab}(\sqrt{a}-\sqrt{b}) + \frac{1}{M^{3/2}}(71-6(\sqrt{a}+\sqrt{b})^2)\right) \\ \tilde{\Delta}_2(x_4(M)) \approx y_{2,4}(M) \approx \frac{e^{-\frac{1}{4}(\sqrt{a}+\sqrt{b})^2}}{12\sqrt{\pi}} \left(\frac{2}{M}\sqrt{ab}(\sqrt{a}+\sqrt{b}) + \frac{1}{M^{3/2}}(-1+6(\sqrt{a}+\sqrt{b})^2)\right) \end{cases} \quad (3.14)$$

Sketch of the proof: Let us consider the approximation functions $\tilde{\Delta}_1(\cdot)$ and $\tilde{\Delta}_2(\cdot)$ as defined in Equation 3.11. The analysis of their derivative functions is equivalent to a root analysis of a fourth degree polynomial, which can be solved using the well known Ferrari approach. This latter provides us with values which leading terms (for large M) are equal to $\{x_1(M) < x_2(M) < x_3(M) < x_4(M)\}$. These solutions appear to be the same for both error functions. Equations 3.13 and 3.14 are, then, computed as the evaluation of $\tilde{\Delta}_1(\cdot)$ and $\tilde{\Delta}_2(\cdot)$ at the positions $\{x_1(M) < x_2(M) < x_3(M) < x_4(M)\}$.

One can notice from the previous property, that for large M the function $\tilde{\Delta}_1(\cdot)$ is approximately two times smaller than $\tilde{\Delta}_2(\cdot)$. Moreover the ratio of their respective amplitudes tends to 2 as M tends to infinity, which suggests preferring Log-Normal approximations of χ^2 distributions rather than the usually used Normal approximation.

This last result is a corollary of Property 2

Property 3 (Maximum absolute error). *Let, $\{y_{1,i}(M)\}_{i=1}^4$ and $\{y_{2,i}(M)\}_{i=1}^4$ the real values defined in Property 2, then we can show that,*

$$\begin{cases} \max\{|y_{1,i}(M)|\}_{i=1}^4 = |y_{1,2}(M)| \\ \max\{|y_{2,i}(M)|\}_{i=1}^4 = |y_{2,2}(M)| \end{cases} \quad (3.15)$$

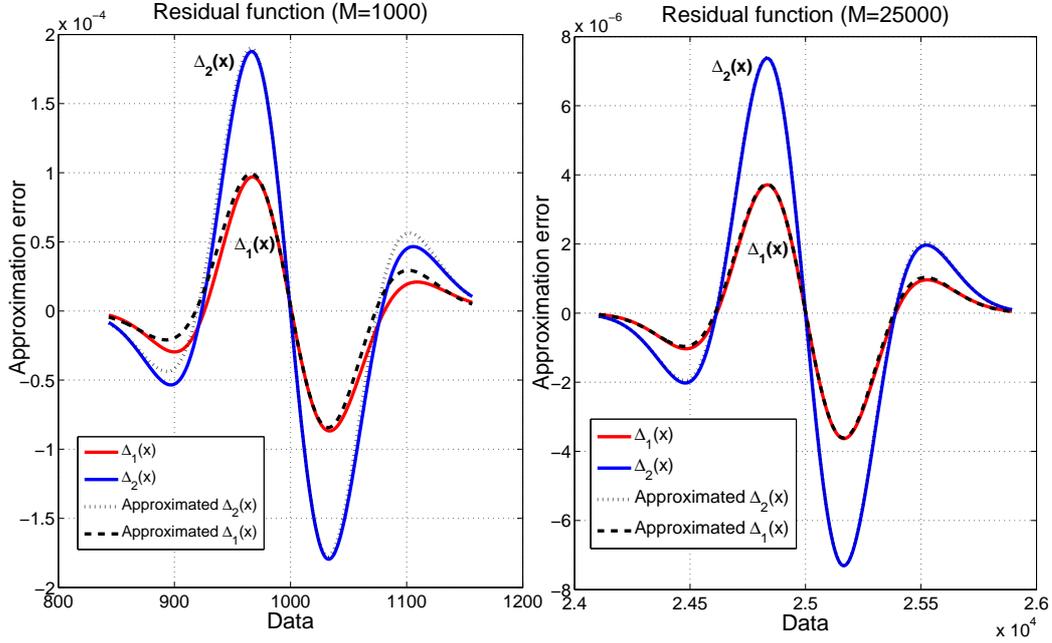


Figure 3.2: Approximation Error functions. Both left and right figures plot the functions $\Delta_1(\cdot)$, $\Delta_2(\cdot)$, $\tilde{\Delta}_1(\cdot)$ and $\tilde{\Delta}_2(\cdot)$. However, the left figure shows these functions for a parameter $M = 1000$, while the right figure shows them for $M = 25000$. We observe in this figure that the theoretical approximation introduced in Property 1 seems to converge to the real values as M grows large.

Moreover:

$$\begin{cases} |y_{1,2}(M)| < |y_{2,2}(M)| \\ |y_{1,2}(M)| \approx |y_{2,2}(M)|/2, \text{ as } M \rightarrow \infty \end{cases} \quad (3.16)$$

This last property investigates the maximum absolute error due to the approximation of χ^2 distribution by Log-Normal or Normal distributions. It shows once again that the bias due to a Log-Normal approximation is smaller than the bias due to a Normal distribution.

3.3.3 Simulations and Empirical Evaluation of Log-Normal based Approximations

Finally, we illustrate in this part the previously introduced results.

Figure 3.2 plots the errors functions $\Delta_1(\cdot)$ and $\Delta_2(\cdot)$ as well as their evaluated approximations, $\tilde{\Delta}_1(\cdot)$ and $\tilde{\Delta}_2(\cdot)$, as computed in Equation 3.11. The left subplot shows the approximation errors for a parameter $M = 1000$ while the right subplot considers a parameter $M = 25000$. The purpose of these figures is twofold: on the one hand, they aim at illustrating the error of fit due to Log-Normal or Normal approximations of χ^2 distributions. This aspect is highlighted by the curves $\Delta_1(\cdot)$ and $\Delta_2(\cdot)$ on the figures, appearing respectively, in red and blue solid lines. On the other hand, this figure shows the accuracy of the introduced approximations of this error functions $\tilde{\Delta}_1(\cdot)$ and $\tilde{\Delta}_2(\cdot)$ as the parameter M grows large.

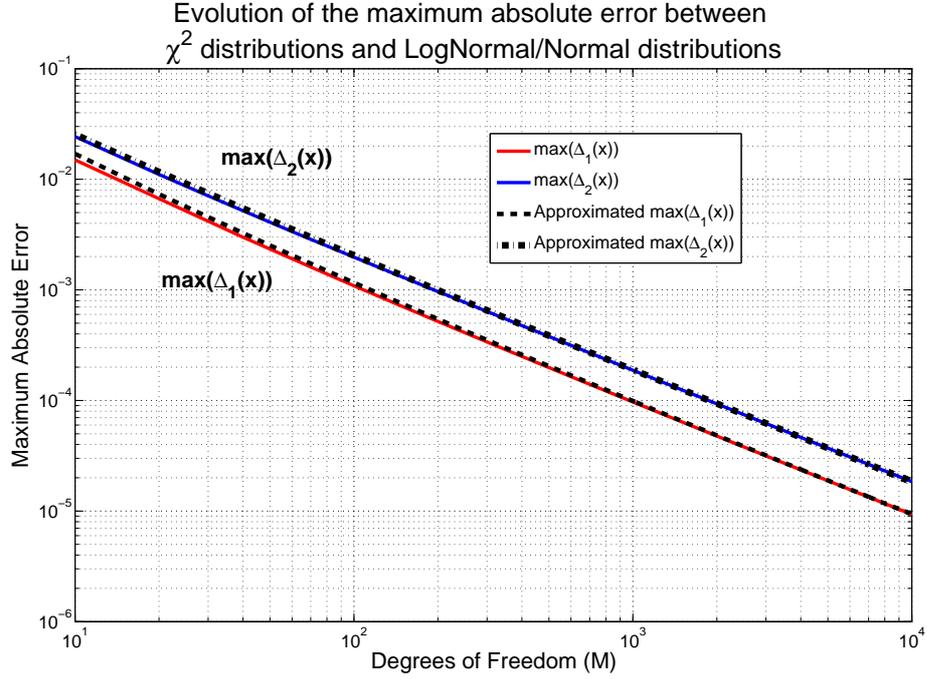


Figure 3.3: Maximum Absolute Error. In this figure, four curves are represented: two of them, in solid line, illustrate the decreasing rate of the global maximum of the error functions $\Delta_1(\cdot)$ and $\Delta_2(\cdot)$. Whereas, the two other curves, plot the theoretical maximum of the approximations $\tilde{\Delta}_1(\cdot)$ and $\tilde{\Delta}_2(\cdot)$. As we can notice, the theoretical approximations developed in Equations 3.13 and 3.14 describe well the reality in both cases.

Firstly, we observe that the errors functions, $\Delta_1(\cdot)$ and $\Delta_2(\cdot)$, have similar forms however, the error due to a Normal approximation of χ^2 distributions seems to have a higher amplitude than a Log-Normal approximation. This indicates, as already suggested in Property 3, that a Log-Normal approximation should be preferred in general.

Secondly, we observe that the approximations $\tilde{\Delta}_1(\cdot)$ and $\tilde{\Delta}_2(\cdot)$ seem to be reasonably accurate especially when it comes to evaluating the maximum of the functions. Thus, this observation validates the properties introduced in this section. As one can observe, the higher the value of the parameter M are, the closer the results get to the real values.

Figure 3.3 emphasizes this last remark. As a matter of fact, it describes the evolution of the absolute value of the maximum of the functions $\Delta_1(\cdot)$ and $\Delta_2(\cdot)$, as well as their theoretical approximations: $\tilde{\Delta}_1(\cdot)$ and $\tilde{\Delta}_2(\cdot)$. We observe, on the one hand, that the amplitude of the maximum absolute error of both $\Delta_1(\cdot)$ and $\Delta_2(\cdot)$ decrease at the same rate, as illustrated by the red and blue curves. On the other hand, this figure plots the theoretically computed functions of the maximums $y_{1,2}(k)$ and $y_{2,2}(k)$ as shown in Property 3. We observe that the reported results in Equations 3.13 and 3.14, fit well the reality even for small values of M in spite of the approximations.

3.4 Energy Detector under Log-Normal noise uncertainty

In this section, We investigate the design of an energy detector that takes into account Log-Normal noise uncertainty. We show that we can accurately control the false alarm, and deduce explicit formulas for the probability of detection as well as the *SNR-wall*. To that purpose, we exploit the Log-Normal approximation of χ^2 distributions, introduced and analyzed in Section 3.3, and apply it within the considered framework detailed in Section 3.2.

3.4.1 Noise Uncertainty and Energy Statistic's Approximation

Let $\hat{\sigma}_n^2$ denote the estimated noise power level. Assuming that $\hat{\sigma}_n^2$ follows a unbiased Log-Normal distribution [49], such that the expectation and the variance of the estimated noise level, respectively, verify $\mathbb{E}[\hat{\sigma}_n^2] = \sigma_n^2$ and $\mathbb{V}[\hat{\sigma}_n^2] = u \cdot \sigma_n^4$, with u the defined uncertainty parameter, also defined through a (non conventional) decibel value $u = 10^{U_{dB}/10} - 1$, then:

$$\hat{\sigma}_n^2 \sim \text{LogN}(\mu_u, \mathbb{V}_u), \text{ s.t. : } \begin{cases} \mathbb{V}_u = \log(1 + u) \\ \mu_u = 2 \log(\sigma_n) - \frac{\mathbb{V}_u}{2} \end{cases} \quad (3.17)$$

where μ_u and \mathbb{V}_u respectively refer to the mean and variance parameters of the uncertainty noise distribution.

Moreover, for mathematical reasons, we consider a Log-Normal approximation of the χ^2 distribution as described in Section 3.3. Approximation analytically and empirically validated in Figure 3.3 for M usually considered large enough, i.e., $M > 50$. As a matter of fact, we can observe in Figure 3.2, that the Log-Normal approximation induces an estimation error on the probability density function. However, the amplitude of this error function, called residual function, is uniformly bounded by a function that mainly decreases as $1/M$. This results are analytically corroborated in Equations 3.14 and 3.13 where we can observe that asymptotically, the amplitudes of the extrema of the error functions mainly decrease as $1/M$. Thus for the rest of this chapter we assume that Log-Normal approximations of Chi-Square distributions is valid. We exploit this approximation to analyze energy detection limits under Log-Normal uncertainty as suggested in Paper [49] by Alexander Sonnenschein and Philip M. Fishman.

Thus, let us assume that the power statistic \mathcal{T}_t/M can be accurately approximated by a Log-Normal distribution such that $\mathbb{E}[\mathcal{T}_t/M] = \sigma_T^2$ and $\mathbb{V}[\mathcal{T}_t/M] = 2\sigma_T^4/M$ then:

$$\frac{\mathcal{T}_t}{M} \sim \text{LogN}(\mu_T, \mathbb{V}_T), \text{ s.t. : } \begin{cases} \mathbb{V}_T = \log\left(1 + \frac{2}{M}\right) \\ \mu_T = 2 \log(\sigma_T) - \frac{\mathbb{V}_T}{2} \end{cases} \quad (3.18)$$

where σ_T^2 is the value of the power level of the collected samples depending on the current state of the channel at the slot t such that:

$$\begin{cases} \mathbb{H}_0 : \sigma_T^2 = \sigma_n^2 \\ \mathbb{H}_1 : \sigma_T^2 = \sigma_n^2 + \sigma_{x,t}^2 \end{cases} \quad (3.19)$$

Finally, we introduce the following statistic \mathcal{W}_t defined as:

$$\mathcal{W}_t = \log\left(\frac{\mathcal{T}_t}{M\hat{\sigma}_n^2}\right) \quad (3.20)$$

In the next subsection, we analyze the performance of the Energy Detector based on the statistic \mathcal{W}_t .

3.4.2 Energy Detector's Performances and Limits

Under the previously introduced assumptions, we present the main results of the chapter. For that purpose, anticipating the next results, we use the following notations: the Signal-to-Noise Ratio $\gamma_t = \sigma_{x,t}^2/\sigma_n^2$ and:

$$\begin{cases} \mathbb{E}(\mathcal{W}_t|\mathbb{H}_0) = \frac{1}{2}(\mathbb{V}_u - \mathbb{V}_T) = \frac{1}{2} \log\left(\frac{1+u}{1+2/M}\right) \\ \mathbb{E}(\mathcal{W}_t|\mathbb{H}_1) = \log(1 + \gamma_t) + \mathbb{E}[\mathcal{W}_t|\mathbb{H}_0] \\ \mathbb{V}(\mathcal{W}_t) = \mathbb{V}_u + \mathbb{V}_T = \log\left((1 + \frac{2}{M})(1 + u)\right) \end{cases} \quad (3.21)$$

Lemma 1 (Distribution of \mathcal{W}_t). *Let \mathcal{W}_t be the random variable defined in Equation 3.20. We assume that the previously introduced assumptions hold, then:*

$$\frac{\mathcal{W}_t - \mathbb{E}[\mathcal{W}_t|\mathbb{H}_0]}{\sqrt{\mathbb{V}[\mathcal{W}_t]}} \sim \begin{cases} \mathbb{H}_0 : \mathcal{N}(0, 1) \\ \mathbb{H}_1 : \mathcal{N}\left(\frac{\log(1+\gamma_t)}{\sqrt{\mathbb{V}[\mathcal{W}_t]}}, 1\right) \end{cases} \quad (3.22)$$

Sketch of the proof: Note that we can write :

$$\mathcal{W}_t = \log\left(\frac{\mathcal{T}_t}{M}\right) - \log(\hat{\sigma}_n^2) \quad (3.23)$$

Thus, \mathcal{W}_t is a linear combination of two independent Gaussian random variables with, respectively, parameters: $\{\mu_T, \mathbb{V}_T\}$ and $\{\mu_u, \mathbb{V}_u\}$. Consequently,

$$\mathcal{W}_t \sim \mathcal{N}(\mu_T - \mu_u, \mathbb{V}_T + \mathbb{V}_u) \quad (3.24)$$

which can be written as:

$$\mathcal{W}_t \sim \mathcal{N}\left(\log\left(\frac{\sigma_T^2}{\sigma_n^2}\right) + \frac{1}{2}(\mathbb{V}_u - \mathbb{V}_T), \mathbb{V}_T + \mathbb{V}_u\right) \quad (3.25)$$

We can notice that if the channel is idle, $\log\left(\frac{\sigma_T^2}{\sigma_n^2}\right) = 0$, otherwise $\log\left(\frac{\sigma_T^2}{\sigma_n^2}\right) = \log(1 + \gamma_t)$. Finally, using the previously introduced notations, we can write:

$$\mathcal{W}_t \sim \mathcal{N}\left(\log\left(\frac{\sigma_T^2}{\sigma_n^2}\right) + \mathbb{E}[\mathcal{W}_t|\mathbb{H}_0], \mathbb{V}[\mathcal{W}_t]\right) \quad (3.26)$$

which concludes this proof.

It is interesting to notice that the variance of \mathcal{W}_t does not depend on the hypotheses; it only depends on the uncertainties. Thus, it illustrates the uncertainty of the energy detector.

More specifically, when the noise level is perfectly known, $u = 0$, we can reach any desired performance by simply increasing, if possible, the sample's size M . This is no more the case, when there exist an uncertainty on the level of the noise.

As a matter of fact, when an uncertainty $u \neq 0$ exists, one can notice that even if the number of samples M tends to infinity, the distribution of \mathcal{W}_t , with either hypothesis, still has a positive variance, leading to unavoidable detection errors.

We next present the performances of the herein analyzed detector.

Theorem 1 (Detector's performances). *Let $\xi_t(\alpha_{fa})$ be a real value such that the threshold based policy is: $\mathcal{W}_t \underset{\mathcal{H}_0}{\leq} \underset{\mathcal{H}_1}{\geq} \log(\xi_t(\alpha_{fa}))$, then the probabilities of false alarm and detection have the following forms:*

$$\begin{cases} \mathbb{P}_{fa,t} = Q\left(\frac{\log\left(\xi_t(\alpha_{fa})\sqrt{\frac{1+2/M}{1+u}}\right)}{\sqrt{\log((1+2/M)(1+u))}}\right) \\ \mathbb{P}_{d,t} = Q\left(\frac{\log\left(\frac{\xi_t(\alpha_{fa})}{1+\gamma_t}\sqrt{\frac{1+2/M}{1+u}}\right)}{\sqrt{\log((1+2/M)(1+u))}}\right) \end{cases} \quad (3.27)$$

Sketch of the proof:

Since \mathcal{W}_t follows a Gaussian distribution under either hypotheses, we can write the probabilities of false alarm $\mathbb{P}_{fa,t}$ and detection $\mathbb{P}_{d,t}$ of this energy detector as follows:

$$\begin{cases} \mathbb{P}_{fa,t} = Q\left(\frac{\log(\xi_t(\alpha_{fa})) - \mathbb{E}[\mathcal{W}_t|\mathbb{H}_0]}{\sqrt{\mathbb{V}[\mathcal{W}_t]}}\right) \\ \mathbb{P}_{d,t} = Q\left(\frac{\log(\xi_t(\alpha_{fa})) - \mathbb{E}[\mathcal{W}_t|\mathbb{H}_1]}{\sqrt{\mathbb{V}[\mathcal{W}_t]}}\right) \end{cases} \quad (3.28)$$

Using the previously introduced notations, we obtain the stated results.

The contributions of Theorem 1 are twofold: On the one hand, Theorem 1 provides closed form expressions of $\mathbb{P}_{fa,t}$ and $\mathbb{P}_{d,t}$. On the other hand, as a corollary, it provides an explicit expression of the threshold for a given probability of false alarm:

$$\log(\xi_t(\alpha_{fa})) = Q^{-1}(\mathbb{P}_{fa})\sqrt{\mathbb{V}[\mathcal{W}_t]} + \mathbb{E}[\mathcal{W}_t|\mathbb{H}_0]$$

Thus, it shows one asset of the Log-Normal model for the noise uncertainty compared to the usually considered bounded uncertainty. As a matter of fact, it enables to define *a priori*, the values of the threshold depending on the false alarm and on the uncertainties. This enables objective and theoretical evaluations of the loss of performance due to noise uncertainty, in terms of probability of detection, for a given false alarm. This comparison was not possible in the case of bounded noise uncertainty, since, by definition, the chosen false alarm only guaranties an upper bound on the desired false alarm.

The following result provides a general form of the *SNR-wall* as a function of the desired performances of the detector and the uncertainties $\{2/M; u\}$.

Theorem 2 (SNR-wall). *Let \mathcal{W}_t be the random variable defined in Equation 3.20 and let $\Delta = Q^{-1}(\mathbb{P}_{fa}) - Q^{-1}(\mathbb{P}_d)$, then the SNR-wall of the ED under a Log-Normal approximated noise level is equal to:*

$$\gamma_{wall,t} = e^{\Delta\sqrt{\mathbb{V}[\mathcal{W}_t]}} - 1 \quad (3.29)$$

Sketch of the proof:

Note that by inverting the equations of the probabilities of false alarm and detection, we can write:

$$\log(\xi_t(\alpha_{fa})) = \begin{cases} Q^{-1}(\mathbb{P}_{fa})\sqrt{\mathbb{V}[\mathcal{W}_t]} - \log\left(\sqrt{\frac{1+2/M}{1+u}}\right) \\ Q^{-1}(\mathbb{P}_{d,t})\sqrt{\mathbb{V}[\mathcal{W}_t]} - \log\left(\sqrt{\frac{1+2/M}{(1+u)(1+\gamma_t)^2}}\right) \end{cases} \quad (3.30)$$

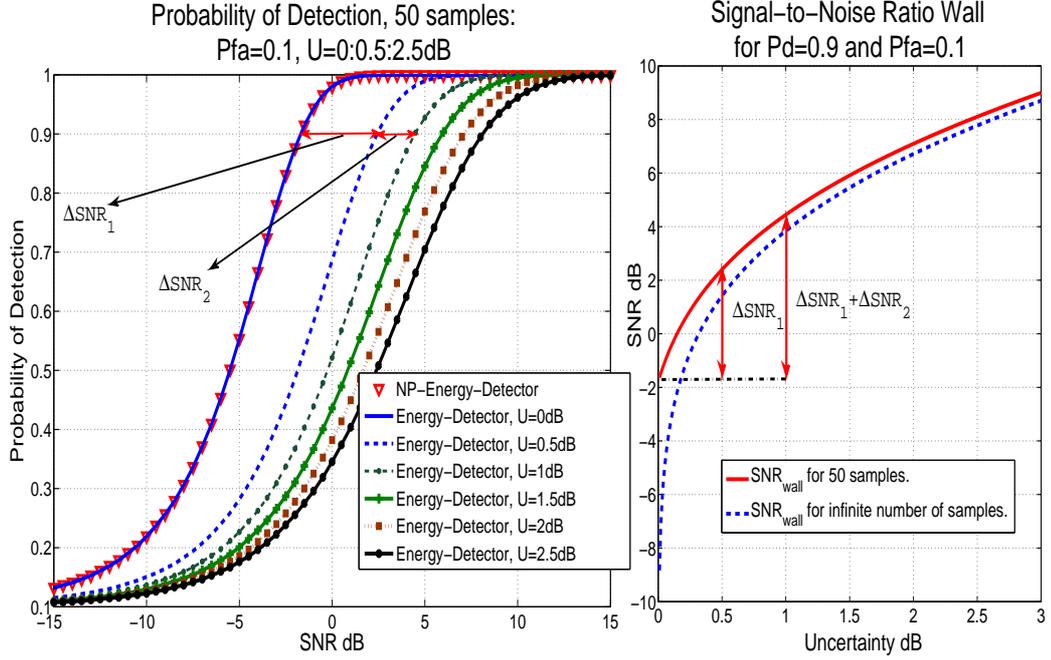


Figure 3.4: Probability of detection (left) and SNR -wall (right). On the left figure, we observe the impact of noise uncertainty, as defined in this chapter, on the detection performances. The right curves shows that the lost of performances can be predicted using the new formula of the SNR -wall. Note that the impact of the “classic” uncertainty[49, 50] does not appear in these curves since it is, generally, impossible to impose a given false alarm, which would make any comparison biased.

which lead to the following expression:

$$\log(1 + \gamma_t) = \Delta \sqrt{\mathbb{V}[\mathcal{W}_t]} \quad (3.31)$$

which concludes the proof.

As already mentioned, the expression of the SNR -wall, within a Log-Normal model for noise uncertainty, depends on the desired performances, in terms of probability of false alarm and probability of detection, as well as on the values on the uncertainties. Thus, rather than an “ SNR -wall”, this results offers a function that accurately predicts the needed SNR to reach given desired performances depending on the uncertainties. In the case of the bounded model for the uncertainty, the SNR -wall is due to an underestimation of the system capabilities, resulting from a worst case analysis.

The impact of noise uncertainty on the detection performances, as described in the equations of Theorem 1, is illustrated in Figure 3.4 (left figure).

On the left figure seven curves of probability of detection are drawn for a false alarm equal to 0.1. The first curve entitled “NP-Energy Detector” represents the results obtained by the NP-ED with no approximation and with no uncertainty on the noise level as described in Section 3.2. The second curve, entitled ‘Energy Detector $U = 0$ dB’, that matches the first curve, illustrates the results as described in Section 3.4 with $U_{dB} = 0$.

The other five curves show the impact of the uncertainty on the herein introduced detector under a Log-Normal model for the noise uncertainty. As expected, we can notice that the detection abilities of the suggested detector degrade as the uncertainty increases. However, the loss of performances can be accurately predicted. As a matter of fact, the right figure of Figure 3.4 illustrates the result of Theorem 2, where we can observe the evolution of the *SNR-wall* as a function of the uncertainty (in this case to achieve a probability of false alarm equal to 0.1 and a probability of detection equal to 0.9). We can notice that the existing gap between the second and third curves (left figure), i.e. the case where the uncertainties are respectively equal to $0dB$ (i.e., with no uncertainty) and $0.5dB$, and referred to as ΔSNR_1 in both figures, matches the result of Theorem 2 (right figure). Similar comments apply for the gap ΔSNR_2 . Finally, this results show that, indeed, Theorem 1 and 2 enable to evaluate the detection limits of the system and conclude this chapter.

3.5 Conclusion

This chapter tackled simultaneously two challenges: on the one hand, we investigated the goodness of a Log-Normal approximation for χ^2 distributions. We showed that not only does the Log-Normal model provide a satisfactory approximation for χ^2 distributions, it, in fact, offers a better fit than the usually suggested Normal approximation. On the one hand, exploiting this result, we revisited the impact of Log-Normal noise uncertainty on energy detection. We showed that the problem involved the analysis of a ratio statistic composed of χ^2 and Log-Normal distributions. The exact analysis of the PDF of the ratio statistic being complex, we suggested to simplify the analysis relying on a Log-Normal approximation of χ^2 distributions. Thus, the considered new model, as well as the mathematical approximation, enabled us to design a detector with a fixed false alarm in spite of the uncertainty. Moreover relying on those results we were able to present a new expression of the *SNR-wall* that depends on the desired performances of the detector as well as the noise uncertainty parameter.

The theoretical analysis of Log-Normal approximations of χ^2 distributions can however be improved. Moreover, the relationship between uncertainty and *SNR-wall* could further be exploited to estimate the uncertainty. These topics are currently under investigation and should lead to new advances in low complexity signal detection for cognitive radio.

Chapter 4

Learning for Opportunistic Spectrum Access: A multi-Armed Bandit Framework

Contents

3.1	Introduction	41
3.2	System model	43
3.2.1	Network assumption	43
3.2.2	Performance evaluation of a detection policy π	43
3.2.3	Neyman-Pearson Energy Detector	45
3.2.4	Energy detection with noise uncertainty	46
3.3	Log-Normal Approximation of χ^2 distributions	46
3.3.1	Mathematical Model	46
3.3.2	Main Results	48
3.3.3	Simulations and Empirical Evaluation of Log-Normal based Approximations	50
3.4	Energy Detector under Log-Normal noise uncertainty	52
3.4.1	Noise Uncertainty and Energy Statistic's Approximation	52
3.4.2	Energy Detector's Performances and Limits	53
3.5	Conclusion	56

Introducing decision making algorithms to tackle wireless communication related issues is not new. However usually, either a substantial amount of information is available to answer the decision making problems in an efficient way, or the system is over-dimensioned to provide satisfactory behavior within a large set of communication scenarios. In other words, one can consider that the basic cognitive cycle has already been implemented, however only in the case where few sensors and possible parameters were operational. As a matter of fact, the sensors are carefully implemented for very specific purposes and parameters already tuned to meet communications standards' requirements. Consequently, the decision making *design space*, discussed in Section 2.3 is shrunk to its minimal volume, leaving few degrees of freedom to radio equipment to optimize their behavior regarding their environment and the user's expectations.

We focus in the sequel on OSA problems. As a matter of fact, these problems appear as a particular instance of the general DCA problems (cf. Section 2.3 and Papers [52, 73]). However due to the scarcity of the spectrum and the need to quickly provide smart allocation techniques, we have decided to tackle this problem. It can be formulated as follows: one (or several) CR user(s) seek to exploit spectrum opportunities left vacant by incumbent users. How should the CA learn to access the most profitable resource while providing a service and without excessively interfering with primary users?

As further detailed in this sequel, the general framework selected for this matter is the so called Multi-Armed Bandit paradigm. It is worth noting that when we started working on the DCA problems in 2008, only few papers were dealing with this matter. Today, we count a large number of papers dealing with MAB models for CR. Moreover, the period 2008-2011 proved to be a prolific period for the MAB community as shown in our brief state of the art in this chapter, Section 4.2.

Our main contributions, as depicted in the next chapters, can be summarized as follows:

- We modeled CR online learning problems as MAB problems. Then we applied the existing results on a basic OSA framework to emphasize its potentials.
- We introduced sensing errors into the OSA model and proved the convergence of any consistent MAB algorithm (under mild assumptions).
- We analyzed the case of multi-secondary users and showed their consistency when they collaborate. We moreover discuss the limits of theoretical guaranties versus empirical 'risky' yet satisfactory behavior of the algorithms.
- We discuss the limits of our model in the case of heterogeneous networks and show the conditions for consistency. This problem combines both resources allocation techniques and machine learning techniques.
- We designed a new form of algorithms to deal with gamma distributions in a simple way. This new algorithm opens the way to dealing with more complex environments such as Rayleigh channels or multi-paths networks.

In this specific chapter, we introduce the considered MAB framework, the considered algorithm, the OSA model as well as the first results, namely investigating the impact of detection errors on the considered MAB based OSA problem.

4.1 Introduction

4.1.1 General Context and Challenges

There are many cognitive radio related problems that can be formalized as follows: to a (or many) radio device(s) is (are) associated a performance criterion which is a (finite) sum of terms, named rewards, observed sequentially. Every reward of the sum is the realization of a random function called reward function which is influenced by the value of the parameters of the devices and the environment. The objective is to determine the sequence of values for the parameters to maximize the expected value of the performance criterion, often while having only limited information about the reward function itself. For example, there is poor knowledge on the reward function when one seeks to operate a radio device in a minimum energy consumption mode (under various operational constraints). Indeed, in such a case, it is very difficult to find the ‘right’ analytical or even algorithmic expression that could model the power consumption. Intuitively, the appropriate way for solving these problems would be to try to overcome this lack of information on the reward function by exploiting past information on the rewards obtained, the environment and the values of the parameters. Moreover, the parameters should be modified to address at best the trade-off between the exploitation of existing past information to generate immediately as high as possible rewards and the generation of new information which could lead to strategies to get perhaps even better rewards, but probably in a more distant future.

From the CR perspective, and as already depicted in Chapter 2, several challenges arise. As a matter of fact designing a CA for CR equipments is challenging for the following reasons:

1. The environment in which the CA operates is stochastic and unknown.
2. The CR device has multiple objectives that involve trade-offs. For instance a CR should minimize its power consumption, while maximizing its transmission range.

To the environment model we add user oriented communication constraints that should influence the cognitive agent’s design:

1. A CR equipment must behave at least as well as current non-CR equipments.
2. The CA should lead to a service that improves over time.
3. The algorithms used by the CA should be able to operate with limited memory and computational resources in order to be embedded in a CR equipment.

These constraints led us to consider sequential decision making problems under uncertainty.

In our research work, solutions for these cognitive radio problems have been built based on research results related to the Multi-Armed Bandit (MAB) [99, 100, 101, 102, 103].

4.1.2 Classic Illustration: Opportunistic Spectrum Access

We discussed in Chapter 2 the concept of Dynamic Spectrum Access (DSA). More specifically, we noted that it has been suggested as a promising approach to exploiting frequency

band resources efficiently, taking advantage of the various available communication opportunities. As a matter of fact, during the last century, most of the meaningful spectrum resources were licensed to emerging wireless applications, where the static frequency allocation policy combined with a growing number of spectrum demanding services led to a spectrum scarcity. However, several measurements conducted in the United-States [11], first, and then in numerous other countries, showed a chronic underutilization of the frequency band resources, revealing substantial communication opportunities.

Thus, DSA was introduced as a possible concept that could alleviate spectrum scarcity [11, 12, 63]. This concept however embed several different approach to manage and allocate spectrum resources among users. We invite the reader to refer to the chapter 2 for further details.

Among the suggested approaches to alleviate spectrum scarcity, Opportunistic Spectrum Access (OSA) has been the center of a lot of attention [12, 63]. The general concept of OSA, as considered in this chapter, defines two types of users: a Primary User (PU) or primary users (PUs) and a Secondary User (SU) or secondary users (SUs). PUs access spectrum resources dedicated to the services provided to them, while SUs refer to a pool of users willing to exploit the spectrum resources unoccupied by PUs at a particular time in a particular geographical area. Since SUs need to access the spectrum while ensuring minimum interference with PUs and without *a priori* knowledge on the behavior of PUs, cognitive abilities (sensing its environment, processing the gathered information, and finally adapting its behavior depending on the environment constraints and users' expectations) are required to enable the coexistence of SUs and PUs. To fulfill these requirements, Cognitive Radio (CR) has been suggested as a promising technology to enable the OSA concept [9, 11, 34, 12, 63].

Although, to the best of our knowledge, there are still no commercialized OSA services, an interesting illustration of this concept can be provided. This example was introduced in Chapter 1. Briefly, to quickly densify their (Wi-Fi) network, main operators in France exploit their subscribers ADSL box. Usually the Wi-Fi connexion of an ADSL box runs two or three virtual networks that share the same connexion. Measurements showed, in most cases, two virtual networks: the first wireless network is dedicated to the subscriber, whereas the second wireless network is managed by the operator. This latter network is shared with other mobile subscribers in the vicinity of the box. Thus, both networks share the same wireless physical card, therefore share the same frequency band. Finally, we also noticed, in the case of the operator *Free*, the existence of a third virtual network dedicated to VoIP. Consequently, in this example, the primary users, refer naturally the incumbent ADSL users. While the secondary users refer to the other subscribers of the same provider in the vicinity of the ADSL box.

This example shows the necessity to quickly provide means to exploit communication opportunities. Although this illustration provides a convenient introduction to OSA, it does not face the same challenges we suggest to tackle. As a matter of fact, in this case, the operator has a full knowledge and control regarding the allocated frequencies. Thus, it may face resource allocation problems yet no specific learning challenges arise. For instance, and for illustration purpose, the models we deal with might design a SU willing to exploit the most interesting ADSL box among those surrounding him, with no prior knowledge on their availability or QoS. Such essential knowledge must be acquired through a learning process.

4.1.3 Outline and contributions

The rest of the chapter is organized as follows. First, we describe the Multi-Armed Bandit paradigm and related to OSA problems in Section 4.2. Moreover, the algorithms designed to tackle MAB problems are briefly discussed depending on the nature of the MAB problem (e.g., stochastic or adversarial). Then, Section 4.3 introduces our first main contribution: a MAB mathematical model for OSA. The model considers detection errors and thus provides a first basic yet realistic model. Section 4.5 provides our second main contribution of this chapter: the evaluation of the performance of the UCB_1 algorithm under the model of Section 4.3. Finally, Section 4.7 concludes on this work and suggests perspectives discussed in Chapter 5 and Chapter 6.

4.2 Learning for Opportunistic Spectrum Access: Multi-Armed Bandit Paradigm and Motivations

4.2.1 Multi-Armed Bandit Paradigm: Conceptual Problem Statement

As discussed in Chapter 2 and, stated hereabove, decision making for CR appears in many scenarios as a sequential decision making problem. In general, sequential decision making problems face a dilemma between the exploration of a space of choices, or solutions, and the exploitation of the information available to the decision maker. The problem described herein, within OSA contexts, is known as sequential decision making under uncertainty.

In this thesis we focus on a sub-class of this problem, where the decision maker has a discrete set of stateless choices and the added information is a real valued sequence (of feedbacks, or rewards) that quantifies how well the decision maker behaved in the previous time steps. This particular instance of sequential decision making problems is generally known as the multi-armed bandit (MAB) problem. Throughout this thesis we focus on a stochastic formulation of the MAB problem (cf., next subsection 4.2.2 for a discussion on the different models the MAB paradigm).

A traditional analogy to this problem is a slot machine (one-armed bandit) with more than one arm. The decision maker (the gambler in the analogy) has to make a choice between several levers⁽¹⁾ to pull. If the gambler had all the information about the expected rewards of the different levers, he would always pull the one maximizing his expected reward. However, since he lacks that essential information, he has no choice but to try all levers to earn an estimation of their performances. This example illustrates the exploration versus exploitation dilemma: the gambler has to find a policy that balances the exploration of the different levers, in order to collect information, and the exploitation of the already gathered estimations by selecting the arm that seems to provide the highest expected reward. Consequently, solving a MAB problem consists in finding a good arm selection strategy. Such a strategy maps the current information about the different arms to the next decision. Playing a given policy results in a cumulated reward over the execution. The largest expected cumulated reward is obtained by the policy which always pulls the arm providing the maximum expected one-step reward. The difference between

⁽¹⁾We use indifferently the words “lever”, “arm”, or “machine” in the remainder of this document.

this maximum cumulated reward and the expected reward of a given policy is called the expected *cumulated loss* or expected (cumulated) *regret* for this policy.

We refer to the sequence of three steps: arm selection, action and reward computation as the MAB cycle. We argue that the MAB cycle can accurately model the basic cognitive cycle, as illustrated in Figure 4.1. This matter is further detailed and illustrated through OSA related scenarios. These scenarios are detailed when needed throughout the sequel of this report.

4.2.2 Multi-Armed Bandit Paradigm: Stochastic Environment Vs Adversarial Environment

When modeling the rewards observed from the environment there exist mainly two approaches: stochastic or adversarial (also referred to as non-stochastic). The stochastic model, assumes that the observed rewards are drawn from a stochastic distribution. Depending on the working assumptions, the distributions can be stationary i.e., with the same statistic parameters throughout the learning process, or non stationary. In the latter, the distribution's parameters can evolve depending on either a deterministic function or a stochastic function. In either one of these cases, the rewards are drawn randomly from the defined distribution at the considered iteration. The adversarial model, however, considers that an opponent chooses the rewards to be provided on the set of arms available. The gambler aims at maximizing the cumulated sum of the collected rewards.

Both topics are still under investigation. Recently, extensive work was published on both stochastic or non-stochastic models [102, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128] as well as many possible related topics (extension to trees, convex functions, infinite number of arms to name a few). These results open the way to many applications, in general, and in particular to CR related problems: spectrum allocation problems, power allocation problems, resource allocation problems under uncertainty, network optimization problems, finding the shortest path in a network, cooperation and/or collaboration in OSA networks, smart wireless communication jamming, and so on. In a more general scope, this work could also target, clinical trials, strategy games (such as Go), advertisement, Smart-Grid applications, to name a few.

The algorithms suggested by the machine learning community to tackle the main MAB problems can be divided into three sets that target respectively (without been exhaustive):

- stochastic-stationary MAB such as Bayesian [129] approach (e.g., Gittings index), or non Bayesian techniques such as: Epsilon-greedy [102] and UCB algorithms [100, 101, 102, 103] to name a few.
- stochastic-non-stationary MAB, e.g., with the discounted UCB algorithms (D-UCB) and sliding-window UCB (SW-UCB) [113].
- Adversarial e.g., Exp3 algorithm [104, 107] (the most popular algorithm which belongs to the class of Softmax algorithms).

For specific problems, this method can be combined with other optimization techniques in order to obtain efficient optimization solutions to problems under uncertainty.

Although, the designed algorithms are getting closer to optimal performances, the machine learning community faces however a remaining challenge: the algorithm that prove to efficiently deal with stochastic MAB behave very poorly with adversarial MAB problems and *vice-versa*. Finding an algorithm that presents satisfactory performances in either cases is still an open challenge.

Thus, the MAB literature offers solutions to problems currently encountered within the CR community. In our work, we identified the simple stochastic MAB model as a promising framework for our OSA related applications. Among this prolific MAB contributions, available in 2008, we aimed at finding, investigating and exploiting an algorithm that offers a satisfactory compromise between optimality and complexity. The general class of algorithm of interest is the so-called index policies. This class of algorithms is further detailed in the next subsection.

4.2.3 Multi-Armed Bandit Paradigm: Index based Policies

A common approach to solving the exploration versus exploitation dilemma consists in assigning an utility value to every arm. An arm's utility aggregates all the past information about the lever and quantifies the gambler's interest in pulling it. Such utilities are called *indexes*. Paper [101], in 1995, emphasized the family of indexes minimizing the expected cumulated loss and called them Upper Confidence Bound (UCB) indexes. UCB indexes provide an optimistic estimation of the arms' performances while ensuring a rapidly decreasing probability of selecting a suboptimal arm. The decision maker builds its policy by greedily selecting the largest index. In his work, Agrawal [101] provided explicit formulae for such indexes and analyzed their performances asymptotically for different reward distributions. Unfortunately, although the complexity of the indexes suggested therein was smaller than those previously analyzed in the work of [100], evaluating these indexes remained computationally costly. Paper [102], in 2002, focused on the case of bounded distributions (with an extension to the case of Gaussian distributions). They showed that a simple index form, named UCB_1 , induces an expected cumulated loss which is upper bounded by a logarithmic function of the total number of pulls. Thus, UCB_1 algorithms are said to be order optimal ⁽²⁾.

Since the work of Auer et al. in 2002 [102], several studies were presented. The suggested studies aim: on the one hand at improving the theoretical guaranties of the algorithm UCB_1 [103, 130], and on the other hand at introducing alternative UCB forms [103, 119, 120, 126, 128]⁽³⁾ to reach the optimal performance predicted by Robbins in 1985 [100].

In this thesis, we mainly consider the algorithm UCB_1 . As a matter of fact, the state of the art conducted in 2008 showed that it presented an interesting *complexity-efficiency* compromise compared to Robbins' indexes [100], Agrawal's indexes [101] and even other algorithms suggested by Auer [102]. Note that Auer et al, in 2002, also suggested in [102] a new Epsilon-greedy algorithm, as well as two other algorithms UCB_2 and UCB_{tuned} .

The Epsilon-greedy algorithm presented constraining conditions to converge. As a matter of fact, the decision making engine needs to know *a priori* the lower bound on

⁽²⁾The notion of optimality and order optimality are thoroughly explained in Section 4.3

⁽³⁾Since 2008 and with major contributions in 2011.

the difference between the expected income of the best arm and the second best arm. Condition that is usually unknown prior to the learning process. UCB_2 presents a more complex form than UCB_1 and proves to behave well. Convergence guaranties for UCB_2 are also provided. However, the paper [103], in 2008, proves that the parameter used in UCB_1 can still be tuned. Consequently, UCB_2 provides a more complex form without providing a substantial improvement in the performances compared to UCB_1 . We believe today that with the new guaranties provided on UCB_1 , relying on a tuned parameter, UCB_1 competes fairly with, UCB_2 [130]. Finally, Auer et al. also suggested a UCB form called UCB_{tuned} . It was shown empirically to outperform the other suggested algorithms. However, we discarded its use in this thesis as we seek strong mathematical guaranties. As a matter of fact, since our scenarios are mainly based on simulations, we need to be able to guaranty the soundness of the algorithm for a large set of problems, not only on a set of (perhaps well chosen) scenarios. It is worth mentioning that with the new theoretical guaranties provided recently on UCB_1 algorithms [130], this latter shows, in practice, a regret twice as large as UCB_{tuned} as well. However, the convergence of UCB_1 to the optimal choice is proven; which is not the case for UCB_{tuned} . To improve even more the performance of UCB_1 , beyond the theoretical bounds, its parameter can be tuned empirically. In this case, UCB_1 and UCB_{tuned} have fairly similar results. This latter matter is briefly discussed in Section 4.4.

Consequently, during this thesis we solely focus on the UCB_1 algorithm⁽⁴⁾.

Thus we ventured the analysis of its potential to answer CR related problems. Moreover, we investigated the possibility to generalize its theoretical results to meet wireless communication intrinsic constraints: for illustration purposes we can note that detection errors lead to deceitful rewards, Rayleigh channels lead to different reward distributions to consider, many gamblers competing on the same machines can lead to interference and reward loss, and so on. These generalizations are discussed in this chapter as well as the next chapters.

We end this section by depicting, in the next subsection, a simple yet important OSA scenario where we insist on the MAB-OSA equivalence.

4.2.4 Opportunistic Spectrum Access Modeled as a Multi-Armed Bandit Problem

As a first approximation, we consider the simplest OSA framework within which a single CR device aims at exploring and exploiting vacant communication opportunities. Thus, the gambler is the SU. More specifically, the CA represents the gambler in this example.

The communication opportunities can be of different nature as illustrated in Figure 4.2 found in Paper [13]. However, for the sake of simplicity and without loss of generality we consider, in this illustration, frequency bands as the sole exploitable opportunities if they appear unused. Consequently, the arms of the slot machine, that the CR can play are the frequency bands.

It is usually assumed that there exist a network of users, referred to as primary users, that have the priority on the resources. The secondary user however, does not assume prior knowledge on the behavior of the Primary Network (PN). In other words, the CA

⁽⁴⁾We however quickly describe a promising algorithm known as UCB_V [103, 130].

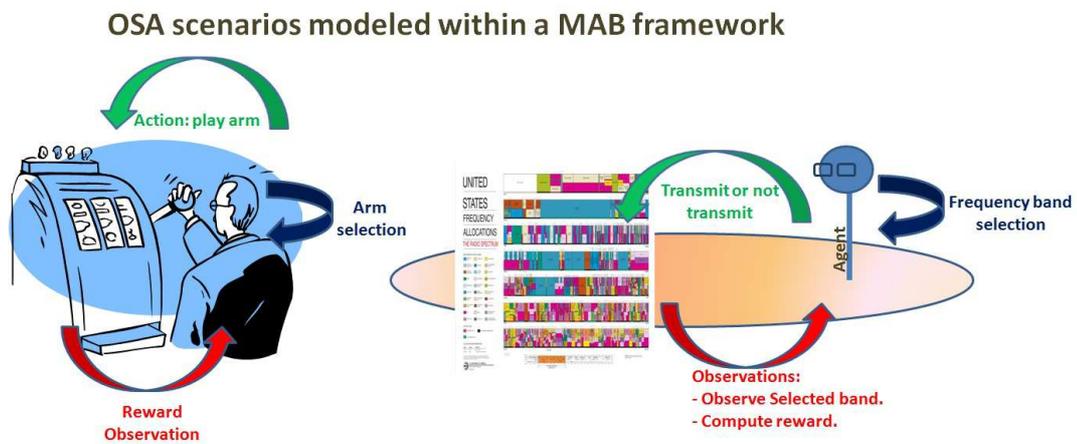


Figure 4.1: Opportunistic Spectrum Access Modeled as a Multi-Armed Bandit Problem. MAB problems represent simple machine learning problems where a gambler have the choice between several machines (or equivalently different arms on a single machine). With no prior knowledge on the machine, the gambler sequentially plays the machines in order to earn information on their expected rewards. Meanwhile, the gambler aims at maximizing his cumulated gains. This last conflict between the time spent on testing the machines and the time spent at playing the machines that seem to be the most profitable is known as Exploration Vs Exploitation dilemma. This figure illustrates side by side the MAB cycle and the CR cycle in OSA contexts. The mean objective is to show that both problems have similar models. In an OSA problem Exploration Vs Exploitation dilemma appears when a secondary user aims at maximizing his cumulated transmitted data parquets while earning more information on the availability of the frequency bands.

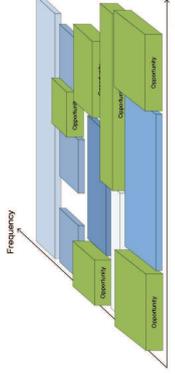
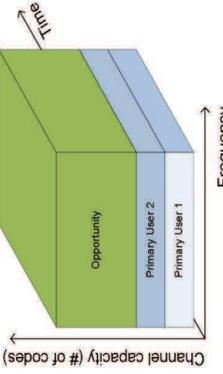
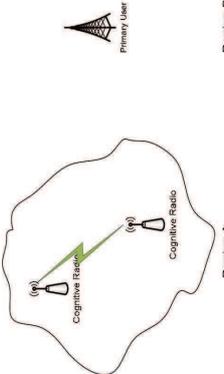
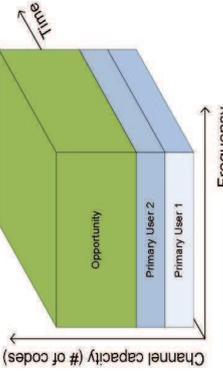
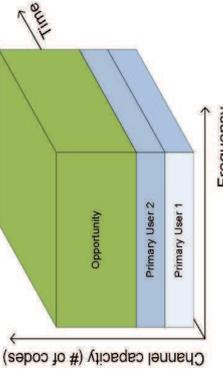
Dimension	What needs to be sensed?	Comments	Illustrations
Frequency	Opportunity in the frequency domain.	Availability in part of the frequency spectrum. The available spectrum is divided into narrower chunks of bands. Spectrum opportunity in this dimension means that all the bands are not used simultaneously at the same time, <i>i.e.</i> , some bands might be available for opportunistic usage.	
Time	Opportunity of a specific band in time.	This involves the availability of a specific part of the spectrum in time. In other words, the band is not continuously used. There will be times where it will be available for opportunistic usage.	
Geographical space	Location (latitude, longitude, and elevation) and distance of primary users.	The spectrum can be available in some parts of the geographical area while it is occupied in some other parts at a given time. This takes advantage of the propagation loss (path loss) in space. These measurements can be avoided by simply looking at the interference level. No interference means no primary user transmission in a local area. However, one needs to be careful because of hidden terminal problem.	
Code	The spreading code, time hopping (TH), or frequency hopping (FH) sequences used by the primary users. Also, the timing information is needed so that secondary users can synchronize their transmissions with primary users. The synchronization estimation can be avoided with long and random code usage. However, partial interference in this case is unavoidable.	The spectrum over a wideband might be used at a given time through spread spectrum or frequency hopping. This does not mean that there is no availability over this band. Simultaneous transmission without interfering with primary users would be possible in code domain with an orthogonal code with respect to codes that primary users are using. This requires the opportunity in code domain, <i>i.e.</i> , not only detecting the usage of the spectrum, but also determining the used codes, and possibly multipath parameters as well.	
Angle	Directions of primary users' beam (azimuth and elevation angle) and locations of primary users.	Along with the knowledge of the location/position or direction of primary users, spectrum opportunities in angle dimension can be created. For example, if a primary user is transmitting in a specific direction, the secondary user can transmit in other directions without creating interference on the primary user.	

Figure 4.2: Communication Opportunities in Wireless Communication for Cognitive Radio. This figure borrowed from the survey [13] illustrates several possible communication dimensions that could be exploited by a SU within the OSA context. Whether SUs can exploit or not these opportunities depends on their abilities to detect them and to use the available resources while ensuring minimum interference with primary users.

Multi-Armed Bandit Context	Opportunistic Spectrum Access Context
Gambler	Secondary User (Equivalently: CR or CA)
1-Armed bandit	Frequency channel to exploit
Observe payoff at every iteration	Instantaneous availability or number of transmitted packets.

Table 4.1: Summary of the equivalence between MAB paradigm and OSA related problems.

cannot deterministically predict the behavior of primary users, viz., the occupancy or non occupancy of the resources by the primary network. Consequently, every time the CA explores a frequency band, the observed state appears as a random variable. In this case the random variable can be labeled as either *idle* or *busy*. Other scenarios where the reward is related to the Signal-to-Noise Ratio (SNR) in fading environments are also considered in Chapter 6. The reward obtained by the CA is thus a function of the numerical values affected to these labels at every iteration.

The OSA scenarios introduced in this thesis rely on a discrete and sequential decision making process. In this case ‘iteration’ refers to the set of computations needed to complete a MAB cycle. This cycle is similar, as summarized in both Figure 4.1 and Table 4.1, to the basic cognitive cycle. The exact structure of the slot depends on the amount of computations needed. Consequently, we propose to detail the slot structure every time needed, with the mathematical models, depending on the considered scenarios. Such explicit mathematical models are proposed, first in the the next sections, and then in the next chapters depending on the scenarios.

4.3 Opportunistic Spectrum Access Mathematic Model : a Multi-Armed Bandit problem

In this section, we introduce the mathematical notations related to MAB problems as well as those related to UCB algorithms.

4.3.1 Basic Opportunistic Spectrum Access Model : Multi-Armed Bandit Notations

OSA related problems introduce several constraints that are not, to the best of our knowledge, considered within the MAB framework. Thus, for clarity reasons, in order to emphasize the contributions of this thesis, we start by modeling a basic OSA problem as a MAB problem. In this case we consider, as will be detailed in the next paragraphs, that there is one SU and that the sensing and detections abilities of the SU are flawless. In other words, no observation errors are considered during the evaluation of the state of a probed channel. However in order to avoid unnecessary redundancies, we introduce a complete model, that includes a model of the sensing abilities of the CR⁽⁵⁾.

We consider the case of one secondary user willing to opportunistically exploit the available spectrum in its vicinity. The spectrum of interest is licensed to a primary network providing N independent but non-identical channels. We denote by $n \in \{1, \dots, N\}$ the n^{th} most available channel. Every channel n can appear, when observed, in one of these two possible states {idle, busy}. In the rest of the dissertation, we associate the numerical value 0 to a busy channel and 1 to an idle channel. The temporal occupancy pattern of every channel n is thus supposed to follow an unknown Bernoulli distribution θ_n . Moreover, the distributions $\Theta = \{\theta_1, \theta_2, \dots, \theta_n, \dots, \theta_N\}$ are assumed to be stationary.

In this thesis, we tackle the particular case where PUs are assumed to be synchronous and the time $t = 0, 1, 2 \dots$, is divided into slots. We denote by \mathbf{S}_t the channels' state at the slot number t : $\mathbf{S}_t = \{S_{1,t}, \dots, S_{N,t}\} \in \{0, 1\}^N$. For all $t \in \mathbb{N}$, the numerical value $S_{n,t}$ is assumed to be an independent random realization of the stationary distributions $\theta_n \in \Theta$. Moreover, the realizations $\{S_{n,t}\}_{t \in \mathbb{N}}$ drawn from a given distribution θ_n are assumed to be independent and identically distributed. The expected availability of a channel is characterized by its probability of being idle. Thus, we define the availability μ_n of a channel n , for all t as:

$$\mu_n \triangleq \mathbb{E}[\theta_n] = \mathbb{P}(\text{channel } n \text{ is free}) = \mathbb{P}(S_{n,t} = 1) \quad (4.1)$$

where $\mu_1 > \mu_2 \geq \dots \geq \mu_n \geq \dots \geq \mu_N$ without loss of generality.

At every slot number t , the SU has to choose a channel to sense. To do so, the cognitive agent relies on the outcome of past trials. We denote by i_t the gathered information until the slot t . We assume, for simplicity reasons, that the SU can only sense one channel per slot. Thus selecting a channel can be seen as an action $a_t \in \mathcal{A}$ where the set of possible actions $\mathcal{A} = \{1, 2, \dots, N\}$ refers to the set of channels available.

⁽⁵⁾We would like to stress that although there exists a paper dealing with MAB modeling for OSA [131] (found online in 2008 on Arxiv), the found paper did not consider imperfect sensing. Thus, one of our contributions is to suggest such a model. Of course, anticipating next sections, the analysis of the impact of sensing errors on the behavior of the chosen algorithms is also part of our contributions.

Thus, we can model the CA as a policy π that maps for all $t \in \mathbb{N}$, the information i_t to an action a_t :

$$a_t = \pi(i_t) \quad (4.2)$$

The outcome of the sensing process is denoted by the binary random variable $X_t \in \{0, 1\}$. In the case of perfect sensing, $X_t = S_{a_t,t}$, where a_t refers to the channel selected at the slot number t . However since we assumed that sensing errors can occur, the value of X_t depends on the receiver operating characteristic (ROC). The ROC defines the accuracy and the reliability of a sensor through the measure of two types of errors: on the one hand, detecting a PU on the channel when it is free usually referred to as *false alarm*. On the other hand, assuming the channel free when a PU is occupying it usually referred to as *miss detection*. Let us denote by ϵ and δ , respectively the probability of false alarm, and the probability of miss-detection characterizing the CR equipment:

$$\begin{cases} \epsilon = \mathbb{P}_{fa} = \mathbb{P}(X_t = 0 | S_{a_t,t} = 1) \\ \delta = \mathbb{P}_{md} = \mathbb{P}(X_t = 1 | S_{a_t,t} = 0) \end{cases} \quad (4.3)$$

Finally, the outcome of the sensing process can be seen as the output of a random policy $\pi_s(\epsilon, \delta, S_{a_t,t})$ such that:

$$X_t = \pi_s(\epsilon, \delta, S_{a_t,t}) \quad (4.4)$$

The design of such policies is however out of the scope of this chapter. Thus we invite the reader to refer to the survey [13] for a general overview on sensing techniques. In the context of uncertainty, we invite the reader to refer to Chapter 3.

Depending on the sensing outcome $X_t \in \{0, 1\}$, the CA can choose to access the channel or not. We denote by $\pi_a(X_t) \in \{0, 1\}$ the access decision, where 0 refers to *access denied* and 1 refers to *access granted*. The access policy π_a chosen in this chapter can be described as: “*access the channel if sensed available*”, i.e. $\pi_a(X_t) = \mathbf{1}_{\{X_t=1\}}$ ⁽⁶⁾.

Note that we assume the ROC to be designed such that the probability of miss detection δ is smaller or equal to a given interference level allowed by the primary network, although $\{\epsilon, \delta\}$ are not necessarily known⁽⁷⁾.

Moreover, we assume that if interference occurs, it is detected and the transmission of the secondary user fails. When channel access is granted, the CA receives a numerical acknowledgment. This feedback informs the CA of the state of the transmission {succeeded, failed}. Finally, we assume that for every transmission attempt, a packet D_t is sent. At the end of every slot t , the CA can use the different information available to compute a numerical value, usually referred to as reward r_t in the MAB literature. This reward informs the CA of its current performance. The form of the reward as well as the evaluation of the selection policy π are described and discussed in the next subsection.

Finally, the sequential steps described hereabove formalize the OSA framework we are dealing with as a MAB problem with sensing errors. A schematic representation of a CA observing and accessing an RF environment is illustrated in Figure 4.3.

⁽⁷⁾As discussed in Chapter 3, in case of uncertainty, it might be difficult to know the exact values of $\{\epsilon, \delta\}$. We show in Section 4.5 that the *UCB1* algorithm does not require that knowledge to converge.

⁽⁷⁾Indicator function: $\mathbf{1}_{\{\text{logical_expression}\}} = \{1 \text{ if logical_expression=true ; } 0 \text{ if logical_expression=false}\}$.

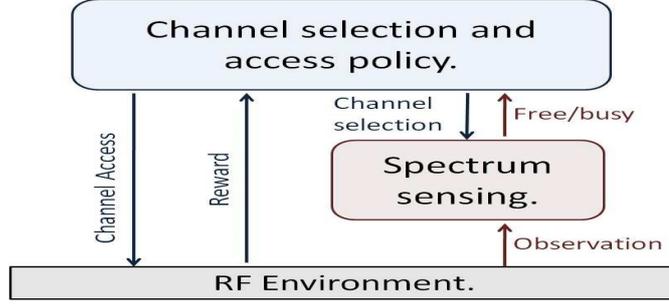


Figure 4.3: Representation of a CA observing and accessing an RF environment.

4.3.2 General Performance Evaluation of a Learning Policy and Optimality

The purpose of this subsection are two-fold. On the one hand, we explicit the notions of reward and regret with the mathematical model introduced in the previous subsection as first described in our paper [68]. On the other hand, we summarize the main criteria and results that enable a fair analysis of learning algorithms. These latter results are the fruit of an early study conducted by T.L. Lai . and H. Robbins in Paper [100] in 1985.

Thus, at the end of every slot t , the CA can compute a numerical value that evaluates its performance. In the case of OSA, we focus on the transmitted throughput. Relying on the previously introduced notations, the throughput achieved by the SU at the slot number t can be defined as:

$$r_t \triangleq D_t S_{a_t,t} \pi_a(X_t) \quad (4.5)$$

which is the reward considered in this particular framework. For the sake of simplicity we assume a normalized transmitted packet for all channels and all t , $D_t = 1$ bit. We can notice that the choices made on the access policy π_a and D_t , simplify the expression of the reward such that:

$$r_t = S_{a_t,t} X_t \quad (4.6)$$

where r_t equals 1 only if the channel is free and the CA senses it free. Consequently, the expected reward achievable using a channel $a_t \in \mathcal{A}$ can be easily computed:

$$\mathbb{E}[r_t] = \mathbb{P}(X_t = 1 | S_{a_t,t} = 1) \mathbb{P}(S_{a_t,t} = 1) = (1 - \epsilon) \mu_{a_t}^{(8)} \quad (4.7)$$

We refer to the channel $\mu_1 = \max_n \mu_n$, that maximizes the reward, as optimal whereas the other channels are said to be suboptimal. We usually evaluate the performance of a policy by its expected cumulated throughput after t slots defined as:

$$W_t^\pi = \mathbb{E} \left[\sum_{m=0}^{t-1} r_m \right] \quad (4.8)$$

An efficient policy π is assumed to maximize the quantity W_t^π .

An alternative representation of the expected performance of a policy π until the slot number t is described through the notion of *regret* R_t^π (or expected regret). The regret

is defined as the gap between the maximum achievable performance in expectation, if the most available channel is chosen, and the expected cumulated throughput achieved by the policy π :

$$R_t^\pi = \sum_{m=0}^{t-1} \max_{a_t \in \mathcal{A}} \mathbb{E}[r_t] - W_t^\pi \quad (4.9)$$

Hence, we define the regret of a channel selection policy π when sensing errors can occur as:

$$R_t^\pi = \sum_{m=0}^{t-1} (1 - \epsilon) \mu_1 - W_t^\pi \quad (4.10)$$

The general idea behind the notion of *regret* can be explained as follows: if the CA knew *a priori* the values of $\{\mu_n\}_{n \in \mathcal{A}}$, the best choice would be to always select the optimal channel μ_1 . Unfortunately, since usually the CA lacks that information, it has to learn it. For that purpose, the CA explores the different channels to acquire better estimations of their expected availability. While exploring it should also exploit the already collected information to minimize the regret during the learning process. This leads to an exploration-exploitation trade-off. Thus, the *regret* represents the loss due to suboptimal channel selections during the learning process.

Maximizing the expected throughput is equivalent to minimizing the cumulated expected regret. In the rest of the chapter, we use the following equivalent formula of the regret:

$$R_t^\pi = (1 - \epsilon) \sum_{n=1}^N \Delta_n \cdot \mathbb{E}[T_n(t)] \quad (4.11)$$

where $\Delta_n = \mu_1 - \mu_n$ and $T_n(t)$ refers to the number of times the channel n has been selected from instant 0 to instant $t - 1$.

Finally we introduce a loss function $\mathcal{L}^\pi(t)$ that evaluates the loss of performance due to sensing errors compared to the perfect sensing framework.

$$\mathcal{L}^\pi(t) = t \max_{a_t \in \mathbb{A}} \mu_{a_t} - W_t^\pi \quad (4.12)$$

An efficient policy is one that maximizes the cumulated expected reward of the gambler. In order to provide a general and fair formalization of the the notion of efficiency, Lai and Robbins introduced the notions of regret as well as beta-consistency [100]. Using our own words we can define both notions as:

Definition 10 (β -Consistent Strategy). *A policy π is said to be β -consistent, $0 < \beta \leq 1$, if it satisfies:*

$$\lim_{t \rightarrow \infty} \frac{E[R_t^\pi]}{t^\beta} = 0 \quad (4.13)$$

This notion gives information on the growth rate of the regret. We usually expect a good policy to be at least *1-consistent*, i.e.:

$$\lim_{t \rightarrow \infty} \frac{\sum_{m=0}^{t-1} r_m}{t} = \max_{a_t \in \mathcal{A}} \mathbb{E}[r_t] \quad (4.14)$$

As a matter of fact, this property ensures that asymptotically the mean expected reward is optimal.

In their seminal paper, Lai and Robbins [100] defined a uniformly good policy as a β -consistent policy for all positive β . A fundamental result shows that any uniformly good policy suffers, in expectation, a regret that grows asymptotically at least as a logarithmic function of the number of iterations. In other words, there exists a fundamental real positive constant C_θ that depends on the reward distributions involved, such that for any uniformly good policy π , we can write:

$$\lim_{t \rightarrow \infty} \frac{R_t^\pi}{\ln(t)} \geq C_\theta \quad (4.15)$$

Moreover Lai and Robbins [100] provided explicit formulas of leaning policies, for several distributions (Bernoulli and Normal distributions for instance) able to achieve such bounds. However the policies introduced in [100] were based on complex indexes whose computation is burdensome. The machine learning community dealing with MAB issues aim at finding simple policies π verifying at least:

$$\lim_{t \rightarrow \infty} \frac{R_t^\pi}{\ln(t)} \geq C_\theta^\pi \quad (4.16)$$

where C_θ^π is real positive constant that depends on both the reward distributions involved as well as the learning policy. Note that we always have $C_\theta^\pi \geq C_\theta$. Since the growth rate of the regret still follows a logarithmic function of the number of trials, such policies are said to be *order optimal*.

Considering the context of OSA and its related constraints as described, in Subsection 4.1.1, we focus on a particular class of order optimal policies. As explained in Section 4.2, these algorithms are particularly simple to compute, have strong mathematical guaranties uniformly over time, rather than only asymptotically, and show satisfactory behavior since they are order optimal. Thus we chose to illustrate them in this Chapter. Then, we shall focus on the performances of the UCB_1 on several scenarios.

4.3.3 Sample Mean Based Upper Confidence Bound Algorithms and Theoretical Performance Results

Building a Cognitive Agent for a CR device requires to find a policy π for this agent that offers satisfactory performances. In this section, we describe an approach for designing well-performing policies π when the CA faces a decision problem that can be formalized as a Multi-Armed Bandit problem, as we believe it is often the case for cognitive radio's decision making problems. The approach is based on the computation of Upper Confidence Bound (UCB) indexes introduced first in [100, 101]. From the UCB indexes computed from the information vector i_t at time t , the action a_t can be inferred in a straightforward way.

Parameters: N , exploration coefficient α

Input: $i_t = [I_0, r_0, I_1, r_1, \dots, I_{t-1}, r_{t-1}]$

Output: a_t

Algorithm:

If: $t \leq N$ return $a_t = t + 1$

Else:

- $T_n(t) \leftarrow \sum_{m=0}^{t-1} \mathbf{1}_{\{I_m=n\}}, \forall n$
 - $A_{T_n(t)} \leftarrow \sqrt{\frac{\alpha \cdot \ln(t)}{T_n(t)}}, \forall n$
 - $B_{T_n(t)} \leftarrow \bar{W}_{T_n(t)} + A_{T_n(t)}, \forall n$
 - return $a_t = \arg \max_k (B_{T_n(t)})$
-

Figure 4.4: A tabular version of a $\pi(i_t)$ policy using a UCB_1 algorithm for computing actions a_t .

As we will see later in this section, these UCB based policies offer good performance guarantees and lend themselves to software implementations compliant with the limited computational resources of a CA embedded in a CR device. The empirical performances of these policies will be evaluated on academic CR and OSA problems in Section 4.4.

At every instant t , an upper confidence bound index is computed for every machine n . This upper confidence bound index, denoted by $B_{T_n(t)}$, is computed from i_t and gives an optimistic estimation of the expected reward of machine n .

Let $B_{T_n(t)}$ denote the index of the policies we are dealing with:

$$B_{T_n(t)} = \bar{W}_{T_n(t)} + A_{T_n(t)} \quad (4.17)$$

where $\bar{W}_{T_n(t)}$ is the sampled mean of the machine n after been played $T_n(t)$ times at the step t , and $A_{T_n(t)}$ is an upper confidence bias added to the sampled mean.

A policy π computes from i_t these indexes from which it deduces an action a_t as follows:

$$a_t = \pi(i_t) = \arg \max_n (B_{T_n(t)}) \quad (4.18)$$

In the rest of this section, we describe two specific upper confidence biases $A_{T_n(t)}$ that are illustrated in our simulations in Section 4.4 and discuss the theoretical properties of the policies associated to these indexes.

4.3.3.1 UCB_1

When using the following upper confidence bias:

$$A_{T_n(t)} = \sqrt{\frac{\alpha \cdot \ln(t)}{T_n(t)}} \quad (4.19)$$

with $\alpha > 1$, we obtain an upper confidence index referred to as UCB_1 in the literature. Under some mild assumptions, given in the following theorem, a UCB policy (cf. tabular version Figure 4.4) using this index is order optimal uniformly over time.

Theorem 1. (cf. [103] for proofs) For all $N \geq 2$, if policy $UCB_1(\alpha > 1)$ is run on N machines/arms having arbitrary reward distributions $\theta_1, \dots, \theta_N$ with support in $[0, 1]$, then:

$$E[R_t^{\pi=UCB_1}] \leq \sum_{n:\Delta_n>0} \frac{4.\alpha}{\Delta_n} . \ln(t) \quad (4.20)$$

Note that a similar theorem could be written if the reward distributions had a bounded support rather than a support in $[0, 1]$. Moreover, in 2009, Paper [130] showed that for $\alpha > 0.5$, UCB_1 algorithm remains order-optimal, whereas for $\alpha < 0.5$ the algorithm has a polynomial regret. Consequently $\alpha = 0.5$ appears a fundamental limit on the mathematical guaranties provided by the machine learning community.

4.3.3.2 UCB_V

A UCB_V policy refers to a policy which uses as upper confidence bias:

$$A_{T_n(t)} = \sqrt{\frac{2\xi.V_n(t). \ln(t)}{T_n}} + \frac{3.c.\xi. \ln(t)}{T_n(t)} \quad (4.21)$$

The UCB_V upper confidence index was first introduced in [103]. In the same research paper, the authors have also proven the theorem given hereafter which shows that UCB_V policies are also order optimal uniformly over time.

Theorem 2. (cf. [103] for proofs) For all $N \geq 2$, if policy $UCB_V(\xi \geq 1, c = 1)$ is run on N machines/arms having arbitrary reward distributions $\theta_1, \dots, \theta_N$ with support in $[0, 1]$, then $\exists C_\xi$ s.t.

$$E[R_t^{\pi=UCB_V}] \leq C_\xi \sum_{n:\Delta_n>0} \left(\frac{\sigma_n^2}{\Delta_n} + 2 \right) . \ln(t) \quad (4.22)$$

Actually a similar result would still hold if $c \neq 1$ but satisfies nonetheless $3.\xi.c > 1$.

By anticipating on the simulation results reported in the next subsection to illustrate the performance of these algorithms, due to the chosen parameters for the algorithms, the UCB_V index performs better on our test problem than the UCB_1 index. This is due to the fact that by adapting its behavior according to the empirical variance of every arm (see the term $\sqrt{\frac{2\xi.V_n(t). \ln(t)}{T_n(t)}}$ in Equality 4.21), a UCB_V based policy seems to be able to better address the exploration-exploitation tradeoff. Other authors have also noticed that by using upper confidence bound indexes based on the empirical variance, better performances could be obtained (see, e.g., [102]).

However, when tuning the parameter of UCB_1 , $\alpha = 0.6$, UCB_1 performs much better than the UCB_V on the horizon that we could simulate (around 10^7 slots). As a matter of fact, with a smaller parameter, UCB_1 algorithms explore less suboptimal channels and focuses on the optimal one.

4.3.4 Complexity

A Cognitive Agent which exploits the tabular version of the UCB_1 algorithm given in Figure 4.4 or its UCB_V counterpart will have at every instant t to carry out a number of operations which is proportional to t and store an information vector whose length grows linearly with t . Therefore, after a certain time of interaction with its environment, a Cognitive Agent having limited computational and memory resources won't be unable to store the information vector i_t and process it fast enough. To overcome this problem, one can program the UCB_1 and UCB_V policy in such a way that part of the solution computed at time t can be used at time $t + 1$, and so that the time required to compute a new solution can be bounded and the memory requirements independent from t .

This can be achieved by noticing that the upper confidence bounds from which the action a_t is computed at time t (see Equations 4.19 and 4.21) are functions of the arguments $\bar{W}_{T_n(t)}$, $T_n(t)$ and also $V_n(t)$ for UCB_V and that these arguments can be computed from the only knowledge of their values at time $t - 1$, a_{t-1} and r_{t-1} . Indeed if $n = a_{t-1}$ we have:

$$T_n(t) = T_n(t - 1) + 1 \quad (4.23)$$

$$D \triangleq r_t - \bar{W}_{T_n(t)} \quad (4.24)$$

$$\bar{W}_{T_n(t)} = \bar{W}_{T_n(t-1)} + \frac{D}{T_n(t)} \quad (4.25)$$

$$V_n(t) = \frac{V_n(t-1) + D \cdot (r_t - \bar{W}_{T_n(t)})}{T_n(t-1)} \quad (4.26)$$

and if $k \neq I_{t-1}$, $T_k(t)$, $\bar{X}_{k, T_k(t)}$, $V_k(t)$ are equal $T_k(t-1)$, $\bar{X}_{k, T_k(t-1)}$, $V_k(t-1)$, respectively.

4.4 Algorithm Illustration, Limits and Discussion

In papers [36, 67], we suggested the use of the algorithms UCB_1 and UCB_V to tackle CR related learning problems in general and OSA access related problems in particular. Moreover we illustrated their performances through several simulations. The paper considered flawless detectors. In other words the detectors parameters $\{\epsilon, \delta\}$ were both chosen equal to 0.

Note that for the sake of clarity, notation and definition redundancies might occur in the next subsections.

4.4.1 Configuration Adaptation Problem

This illustration is extracted from our early work published in 2009 in [36]. We consider Configuration Adaptation Problem, where the performance of several configuration is perceived as a Gaussian or truncated Gaussian distribution. The illustration presented hereafter is reported as described in our paper [36], in 2009.

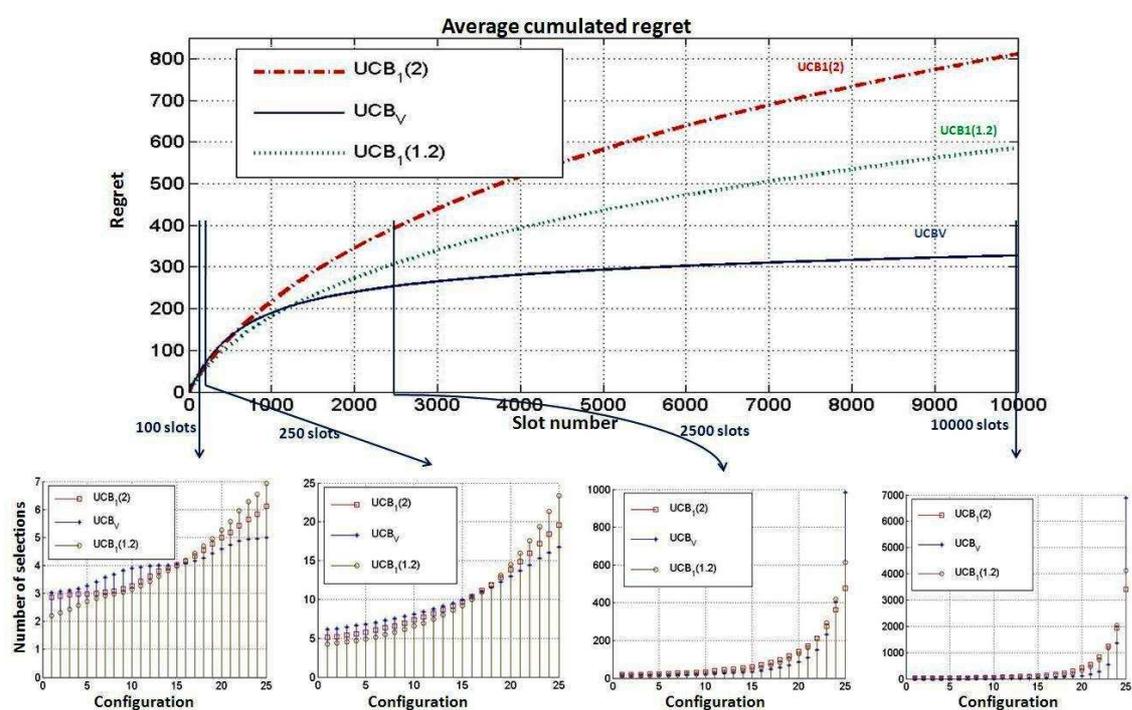


Figure 4.5: UCB based policies and dynamic configuration problem: simulation results. Figure on top plots the average cumulated reward as a function of the number of slots for the different UCB based policies. The figures on the bottom represent the number of times every configuration has been selected after a specific number of time slots. From the left to the right, 100 slots, 250 slots, 2500 slots and 5000 slots.

1-CR equipment:

- N possible configurations C_n , $n \in \{n = 1, \dots, N\}$, verifying the operational constraints but with unknown performances.
- A Cognitive Agent: can learn and make decisions to help the CR equipment to improve its behavior.

2-Time representation:

- Time divided into slots $t = 0, 1, 2, \dots$ (Figure 4.7)
- At the beginning of every slot t , the CA decides to reconfigure or not the CR equipment.

3-Environment and performance evaluation:

- Typical observations: SNR, BER, network load, throughput, spectrum bands, etc.
- A numerical signal is computed at the end of every slot t and informs the CA of the performance of the CR equipment. The numerical signal obtained when using configuration C_n is a function of the observations and the configurations.
- The numerical results computed with a configuration C_n are assumed to be *i.i.d.* and drawn from an unknown stochastic distribution θ_n .

Figure 4.6: Description of the Dynamic Configuration Adaptation problem.

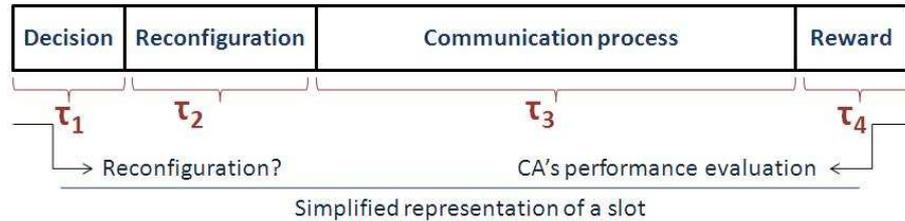


Figure 4.7: Slot representation for a radio equipment controlled by a CA. A slot is divided into 4 periods. During the first period, the CA chooses the next configuration. If the new configuration is different from the current one, a reconfiguration is carried out during the second period before communicating. If a reconfiguration is not needed, the CR equipment keeps the current configuration to communicate. At the end of every slot, the CA computes a reward that evaluates its performance during the communication process. It is assumed here that $\tau_1 + \tau_2 + \tau_4$ are small with respect τ_3 .

The followed experimental protocol is the following. The CA can choose between 25 configurations. To every of these configurations is associated a reward distribution which is Gaussian distribution truncated to the interval $[0,1]$. The pdf of such a distribution is given by

$$\frac{\text{Gauss}_{(\mu,\sigma^2)}(x)}{E_{X \sim \text{Gauss}_{(\mu,\sigma^2)}}[\mathbf{1}_{\{X \in [0,1]\}}]} \quad (4.27)$$

where $\text{Gauss}_{(\mu,\sigma^2)}(\cdot)$ refers to the pdf of a non-truncated Gaussian distribution having a mean μ and a standard deviation σ . The parameter σ has always been chosen equal to 0.1 in our simulations. The parameter μ differs from one distribution to another and has been selected by drawing a number at random and with uniform probability in $[0,1]$ for every configuration.

Every numerical result reported hereafter is the average of the values obtained over 100 experiments. For each experiment, new reward distributions are first generated. To ease the presentation of the results, we will in each experiment refer by n the configuration to which is associated the reward distribution having the n th smallest mean.

In this section, the parameter α of the UCB_1 algorithm is chosen either equal to 1.2 (in which case the algorithm is referred to as $UCB_1(1.2)$) or to 2 (referred to as $UCB_1(2)$). The parameters ξ and c of the UCB_V algorithm are equal to 1 and 0.4, respectively.⁽⁹⁾

The simulation results are reported in the figures of this subsection. Figure 4.5-top shows the evolution of the average cumulated regret for the different UCB policies. For all three policies, the cumulated regret first increases rather rapidly with the slot number and then more and more slowly. This shows that UCB policies are able to process the past information in an appropriate way such that configurations leading to high rewards are favored with time. This is further illustrated by the four graphics on the bottom of Figure 4.5. These graphics show the number of times every individual configuration has been selected after a specific slot number. As we observe, the UCB policies indeed select more often the best configurations when the slot number increases. The coefficient α seems to affect significantly the performance of the UCB_1 policies (see Figure 4.5). This in turn suggests that tuning well the parameters of UCB based policies is important. With respect to this particular parameter α , theory suggests to take α as close as possible to 1 to have the smallest possible upper bound on the expected cumulated regret. As we can observe, we have indeed obtained better results with the smallest value of α considered in our simulations.

In this academic problem, the number of possible configurations was relatively small (25). One may wonder how the UCB policies would scale up to larger sets of configurations. In an attempt to answer this question, we have run also simulations by using 50 configurations. The results are reported on Figure 4.8. The bold curve represents for different number of slots the percentage of times the optimal configuration was selected when using 50 configurations. The dashed curve shows the results obtained with 25 configurations. As we observe, the dashed curve stands well-above the plain one when the number of slots is small. When the numbers of slots starts growing, the distance between both curves decreases and almost vanishes after a large number of slots.

⁽⁹⁾With such values for c and ξ , the condition $3.\xi.c > 1$ is satisfied and the bound of Equation 4.22 of the expected cumulated regret still holds.

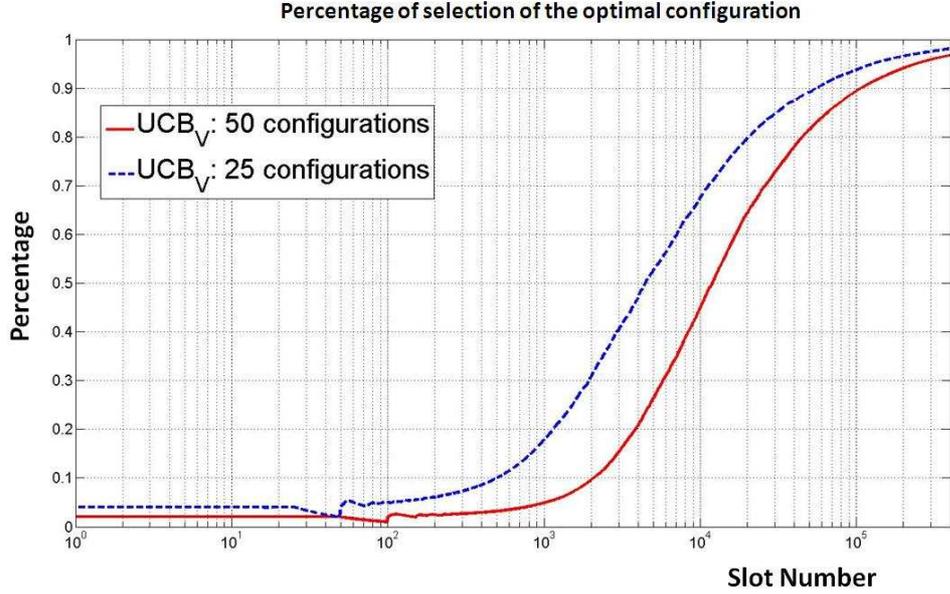


Figure 4.8: Percentage of times a UCB-based policy selects the optimal configuration.

These results suggest that when dealing with larger set of configurations, UCB based policies still lead to acceptable performances if the number of slots is large enough.

4.4.2 OSA under perfect Channel State Information

In Chapter 2, we defined OSA as a particular instance of DCA problems. In this subsection, we provide results similar to the previous subsection. However since spectrum availability is the interesting resources, the distribution of the rewards is modeled through a Bernoulli distribution as formalized in Section 4.3.

Figure 4.9-top shows similar results as Figure 4.5. This is further illustrated by the 3 graphics on the bottom of Figure 4.9. These graphics show the average throughput achieved by the UCB policies. As we observe, the throughput increases with time. Actually, one has the theoretical guarantee that it will converge to 0.9, which is the largest probability of availability of a channel. Figure 7 shows the percentage p of times a UCB policy selects the optimal channel until the slot number t ($p = 100 \cdot \frac{\sum_{m=0}^{t-1} \mathbf{1}_{\{a_m=K\}}}{t}$). As one can observe, this percentage tends to get closer and closer to 100 as the slot number increases.

Although, both UCB policies converge to the optimal channel (i.e. the most available channel), their learning process is different. As a matter of fact, as illustrated in Figure 4.10, while UCB_1 seems to have a balanced behavior between exploration and exploitation, UCB_V tends to spend more time collecting information on the different channels in a first phase (around 300 slots in our case), then exploiting them efficiently taking advantage of the empirical variances. Consequently even though, the behavior of UCB_1 seems better in the first slots, a UCB_V based agent is actually having a more satisfying behavior from the user point of view. As a matter of fact, one might note that

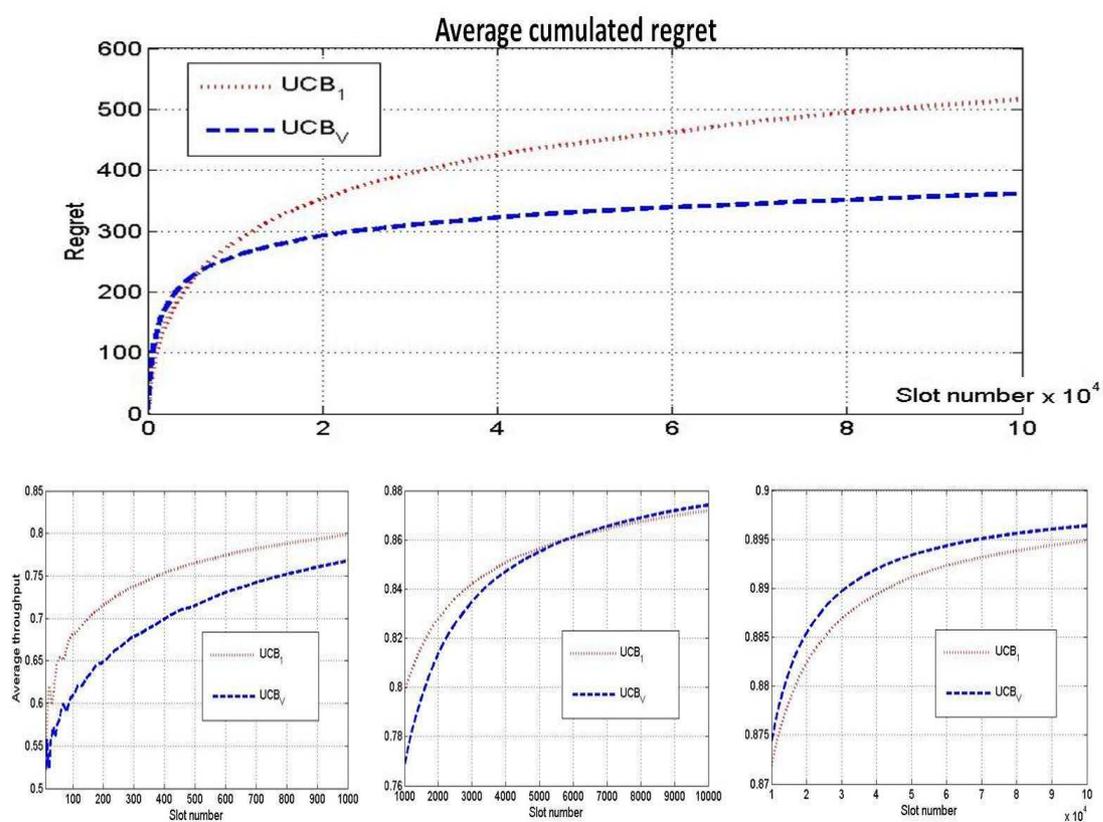


Figure 4.9: UCB based policies and opportunistic spectrum access problem: simulation results. Figure on top plots the average cumulated reward as a function of the number of slots for the different UCB based policies. The figures on the bottom represent the evolution of the normalized average throughput achieved by these policies.

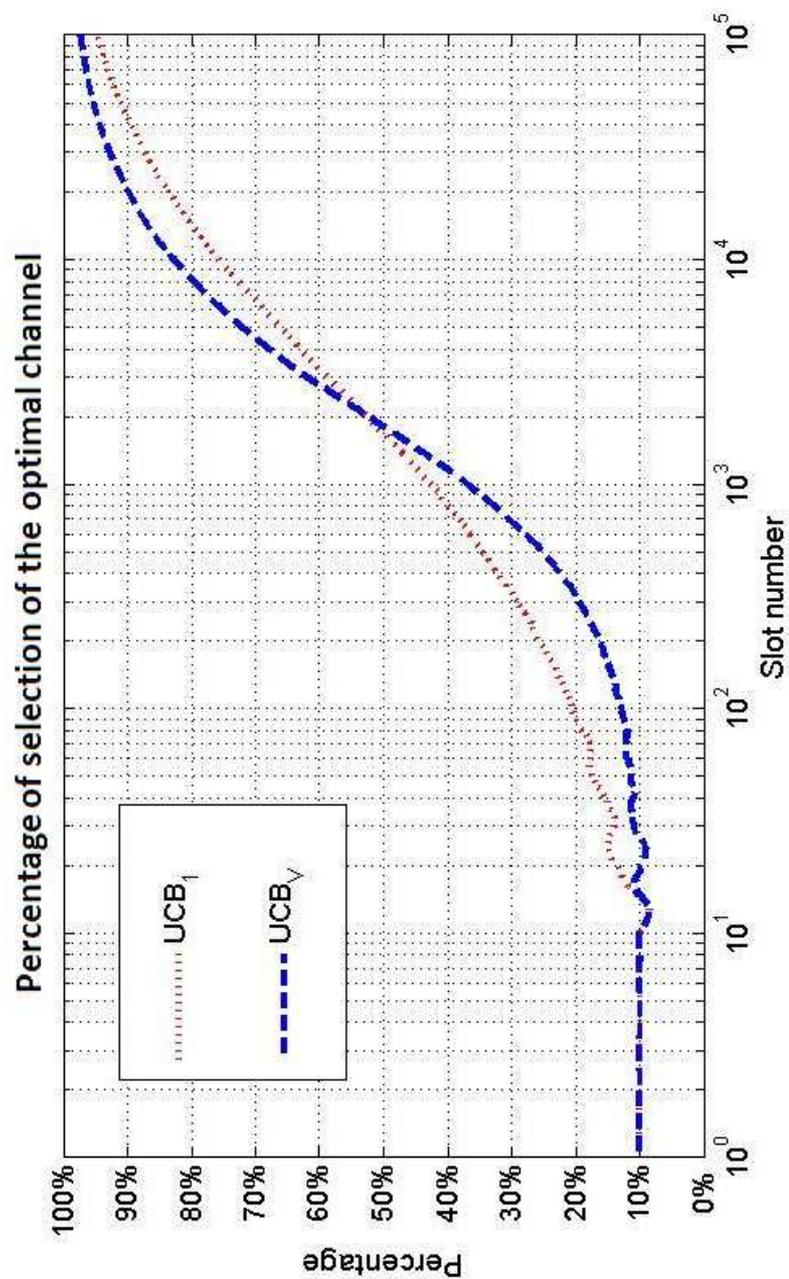


Figure 4.10: Percentage of time a UCB-based policy selects the optimal channel in case of perfect channel state information at every iteration t .

starting from a slot number T_0 where the distance between the two curves is maximal (around 300 slots), the growth rate of UCB_V 's curve increases faster than UCB_1 's curve. This means that the UCB_V based agent selects more often the best channel than UCB_1 which also leads to good performance in terms of regret and throughput.

In our simulations results, we found out that $UCB_1(\alpha > 1)$ seems to outperform UCB_V at the beginning of the learning process and that, afterwards, UCB_V outperforms UCB_1 . This may be explained by the fact at the beginning of the learning UCB_V spends more time collecting information on the different channels than UCB_1 since it also depends on the variances of the different channels and not only on their empirical mean. During this phase, it mainly has a pure exploration strategy while UCB_1 starts already exploiting the information that has been gathered

However, once it starts having good estimates of these variances, it addresses the exploration-exploitation tradeoff, in this example, in a more efficient way than UCB_1 .

The performance of UCB_1 algorithm highly depends on its exploration parameter α . Relying on recent results in [130], when the exploration parameter of UCB_1 is smaller than 1, viz., $1 > \alpha > 0.5$, UCB_1 is still order optimal while it seems to always outperform UCB_V algorithm on the time horizon that we were able to simulate (around 10^7 slots).

Finally, it is worth noticing that although some algorithms still have no convergence proofs, they seem to outperform those that do. We can refer to UCB_{tuned} for instance. This matter led the machine learning community to investigate the performance of UCB algorithms beyond the limits and constraints fixed by the theory [132].

4.4.3 Limits of the Theory and Discussion

In the case of UCB_1 for instance, we usually consider that it is risky to choose an exploration parameter α smaller than 0.5. However, at the ICML2011 conference, J.-Y. Audibert and R. Munos presented an empirical evaluation suggesting that the lowest regret in average is observed for a parameter $\alpha \approx 0.2$ as illustrated in Figure 4.11. The results suggest that for non infinite horizons, small values of α , that seem to contradict the limits imposed by the theory, can lead to better results than those predicted by the theory.

The fact is that it does not contradict theory. As a matter of fact, in the case of UCB_1 algorithms for instance, for $\alpha > 0.5$, the algorithm is order optimal; however for $\alpha < 0.5$ theory shows that the expected cumulated regret increases as a sub-linear function. Yet, there exists a finite interval where the sub-linear function is smaller than the logarithmic function predicted by the theory. This explains why for values of α close to 0.2, and for relatively short time horizons, the average computed regret is smaller than the regret computed with parameters larger than 0.5.

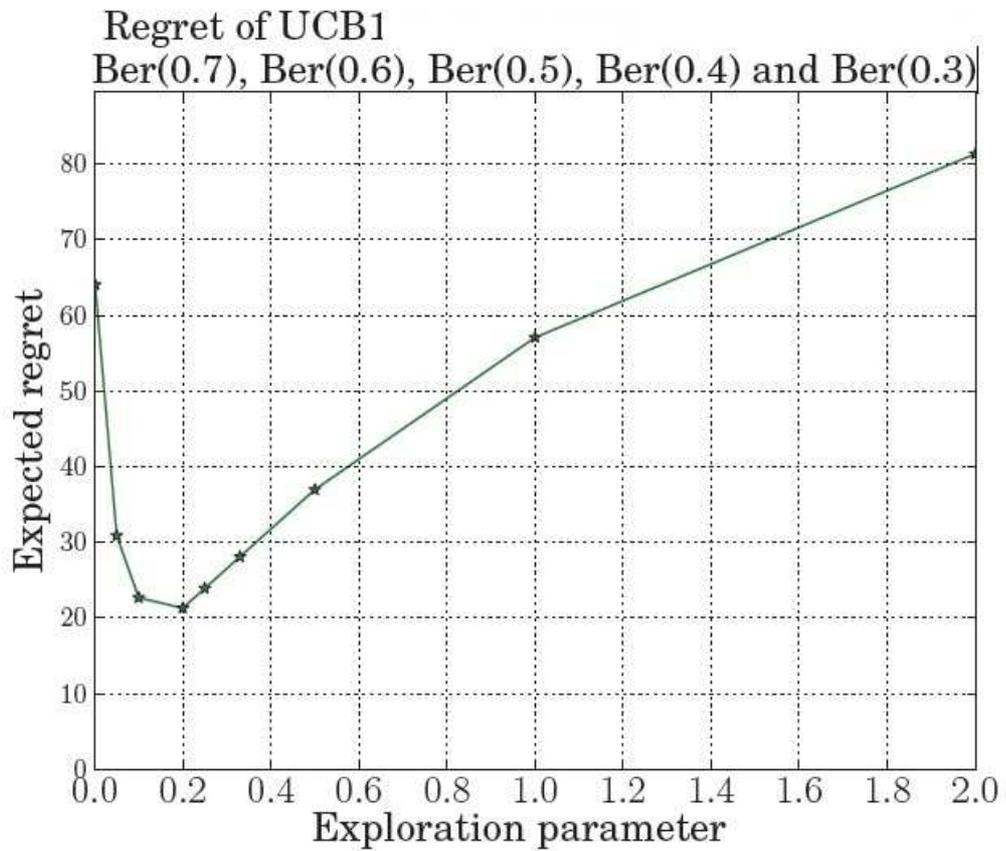


Figure 4.11: Presented by J.-Y. Audibert and R. Munos at ICML 2011. Averaged empirical Regret plotted as a function of the exploration parameter α . It considers $N = 5$ machines and a time horizon of 1000 slots. The results suggest that for non infinite horizons, small value of α , that seem to contradict the limits imposed by the theory, can lead to better results than those predicted by the theory.

4.5 Opportunistic Spectrum Access with Imperfect Sensing

In this Section, we analyze the impact of sensing errors on the performances of UCB_1 algorithm.

4.5.1 UCB_1 Performance Analysis

Relying on the general model provided in Section 4.3, we provide in this section the proof of the convergence of UCB_1 algorithms in the case of OSA scenarios with sensing errors.

The following theorem shows that although the CA suffers imperfect sensing, it still can converge quickly to the most available channel.

Theorem 3 (Logarithmic suboptimal channel selection). *Let us consider a receiver with sensing characteristics $\{\epsilon, \delta\}$, and an “access the channel if sensed available” policy. We consider the instantaneous normalized throughput as the CA’s reward.*

Then for all $N \geq 2$, if the receiver runs the $UCB_1(\alpha > 1)$ policy on N channels having Bernoulli occupation pattern distributions $\theta_1, \dots, \theta_N$ with support in $[0, 1]$, the expected number of selections $\mathbb{E}[T_n(t)]$ for all suboptimal channels $n \in \{2, \dots, N\}$ after t slots is upper bounded by a logarithmic function such that:

$$\mathbb{E}[T_n(t)] \leq \frac{4\alpha \ln(t)}{((1 - \epsilon)\Delta_n)^2} \quad (4.28)$$

Proof. we provide an intuitive proof (a full proof is provided in Chapter 5 as a particular case of the Multi-secondary user scenario):

Let us consider Bernoulli occupation pattern distributions $\Theta = \{\theta_1, \dots, \theta_N\}$ with support in $[0, 1]$. As noticed through Equation 4.4, SUs’ sensors can be seen as functions $\pi_s(\epsilon, \delta, \cdot)$ with parameters $\{\epsilon, \delta\}$ that map a random realization $S_{n,t}$ drawn from the distribution θ_n , at the slot number $t \in \mathbb{N}$, into a binary value $X_t \in \{0, 1\}$ such that:

$$X_t = \pi_s(\epsilon, \delta, S_{n,t}) \quad (4.29)$$

Let us define the set of reward distributions $\tilde{\Theta} = \{\tilde{\theta}_1, \dots, \tilde{\theta}_N\}$ such that: $\forall t \in \mathbb{N}$, the reward $r_t = S_{n,t}X_t$ computed when the channel n is selected follows the distribution $\tilde{\theta}_n$. Then the distributions $\tilde{\Theta} = \{\tilde{\theta}_1, \dots, \tilde{\theta}_N\}$ are bounded distributions with support in $[0, 1]$.

Moreover let us define:

$$\forall n \in \{1, 2, \dots, N\}, \tilde{\mu}_n \triangleq \mathbb{E}[\tilde{\theta}_n] \quad (4.30)$$

Under the assumptions of this theorem, we can write for all $n \in \{1, 2, \dots, N\}$:

$$\begin{cases} \tilde{\mu}_n = (1 - \epsilon)\mu_n \\ \tilde{\Delta}_n = (1 - \epsilon)\Delta_n \end{cases} \quad (4.31)$$

Consequently we can apply the following theorem (Cf. [103] for proof or in Chapter 5 as a particular case of the Multi-secondary user scenario):

For all $N \geq 2$, if policy $UCB_1(\alpha > 1)$ is run on N channels having arbitrary reward distributions $\theta_1, \dots, \theta_N$ with support in $[0, 1]$, then:

$$\mathbb{E}[T_n(t)] \leq \frac{4\alpha}{\tilde{\Delta}_n^2} \ln(t) \quad (4.32)$$

Finally, by substituting: $\mu_n \Leftarrow (1 - \epsilon)\mu_n$ and $\Delta_n \Leftarrow (1 - \epsilon)\Delta_n$ we obtain the stated result:

$$\mathbb{E}[T_n(t)] \leq \frac{4\alpha \ln(t)}{((1 - \epsilon)\Delta_n)^2} \quad (4.33)$$

□

The consequences of Theorem 3 are twofold: on the one hand as for the case of perfect sensing UCB_1 policies, applied on OSA scenarios with sensing errors, spend exponentially more time probing the optimal channel than suboptimal channels⁽¹⁰⁾. On the other hand, we note that the exploration phase, characterized by the time spent on suboptimal channels increases with a scale $\frac{1}{(1-\epsilon)^2}$ compared to the perfect sensing framework. Thus, as expected the accuracy of the sensor is crucial in order to maximize SUs' profit. This last statement partially motivated our work provided in Chapter 3. As a matter of fact it aims at allowing higher observation accuracy - better control- in the case of noise uncertainty.

The consistency of UCB_1 algorithms, when their learning process is disturbed with sensing errors, can be generalized to other learning algorithms. As a matter of fact, the provided proof relies on a key analysis: the detector modifies the observed channels' respective qualities homogeneously. Their observed quality is scaled by a factor $(1 - \epsilon)$. consequently, the observation process does not modify the order of the channels: the most available channel is still observed as such. Therefore, we can state that the learning algorithm still converges to the optimal channel in spite of the errors.

Corollary 1 (Regret and Loss function). *Assuming that we verify the assumptions and conditions of Theorem 3, the regret and the loss function can be upper bounded as follows:*

$$\begin{cases} R_t^\pi \leq \sum_{n=1}^N \frac{4\alpha \ln(t)}{((1-\epsilon)\Delta_n)} \\ \mathcal{L}^\pi(t) \leq \epsilon t + \sum_{n=1}^N \frac{4\alpha \ln(t)}{((1-\epsilon)\Delta_n)} \end{cases} \quad (4.34)$$

Proof. First, we can note that:

$$\mathcal{L}^\pi(t) = \epsilon t + R_t^\pi \quad (4.35)$$

The rest of the proof is an immediate application of the result of Equation 4.28 of Theorem 3, to Equation 4.11 and Equation 4.12. □

The first result of the corollary shows that the regret, as defined in machine learning, is still upper bounded by a logarithmic function of the slot number t . However, as for $\mathbb{E}[T_n(t)]$, due to sensing errors, the regret increases by a scaling factor equal to $1/(1 - \epsilon)$. The second result shows that compared to the perfect sensing framework, the SU suffers unavoidable linear expected loss due to sensing errors.

⁽¹⁰⁾Note that $\mathbb{E}[T_n(t)]$ only depends explicitly on ϵ because of the feedback. This latter avoids considering failed transmissions as rewards (Equation 4.7)!

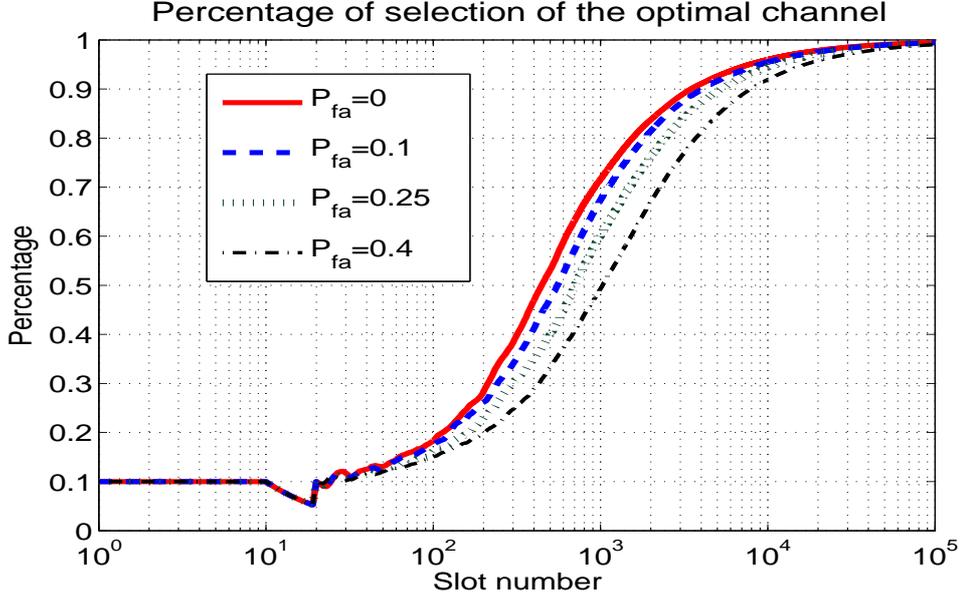


Figure 4.12: Percentage of time the UCB_1 -based CA selects the optimal channel under various sensing errors frameworks (over 10 available channels).

4.5.2 Simulation Results

In this section we present and comment simulation curves focusing on the regret and on the optimal channel selection. The curves compare the behavior of the UCB_1 algorithm under various sensing characteristics.

We consider, in our simulations, one SU willing to exploit a pool of 10 channels. The parameters of the Bernoulli distributions are $[\mu_1, \mu_9, \dots, \mu_{10}] = [0.9, 0.8, 0.8 : -0.1 : 0.1]$. These distribution characterize the temporal occupancy of these channels. To avoid causing interference to PU's, we assume that an adequate δ is guaranteed. Since, ϵ and δ are related to one another through their ROC, the values of ϵ are imposed depending on the channels' conditions.⁽¹¹⁾ In order to evaluate the impact of these parameters on the CA's behavior, we chose to simulate the UCB_1 algorithm with four different sensors: $\epsilon = [0, 0.1, 0.25, 0.4]$. Moreover, in order to respect the conditions stated in Theorem 3, UCB_1 was run with the parameter $\alpha = 1.2$. Every numerical result reported hereafter is the average of the values obtained over 100 experiments.

Figure 4.13 shows the evolution of the average regret achieved by the UCB_1 policy under various sensing characteristics. As expected (Cf. Corollory 1), we observe that the regret first increases rather rapidly with the slot number and then more and more slowly. We remind that the smaller the regret is, the better is the algorithm behaving. This shows that the UCB policy is able to process the past information in an appropriate way even if there are sensing errors such that most available resources are favored with time. Actually, one has the theoretical guarantee that it will converge to $(1 - \epsilon)\mu_1$, which is the largest

⁽¹¹⁾The sensing capabilities of the SU are fixed by the parameter ϵ . Since we only evaluate, in this section, the regret and the optimal channel selection, we purposely ignore the parameter δ since it does not appear in the theoretical results. Note however that in real scenarios, ϵ and δ are related to one another.

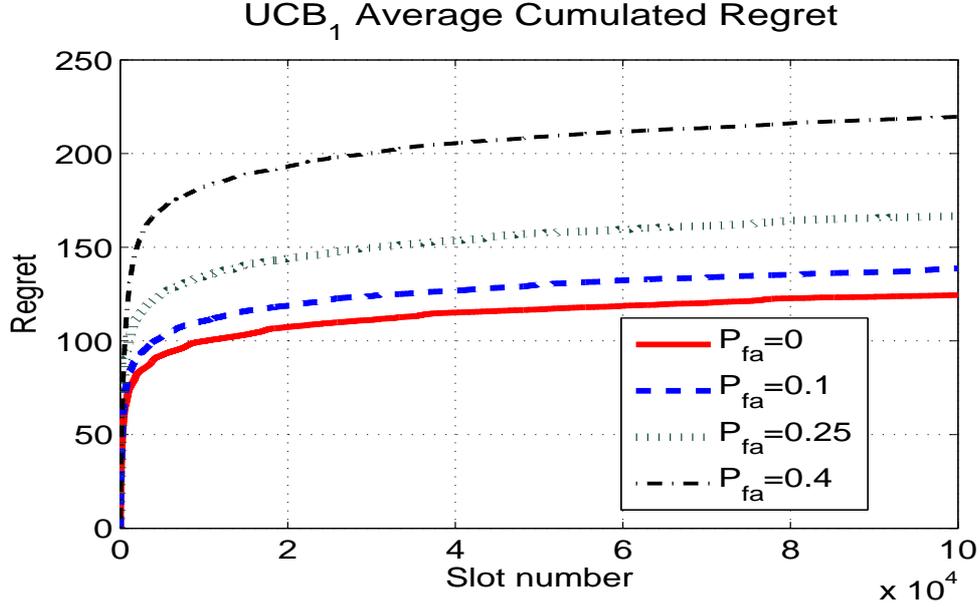


Figure 4.13: UCB_1 algorithm and Opportunistic Spectrum Access problem with sensing errors: regret simulation results.

probability of availability of the optimal channel within the herein modeled imperfect sensing framework. We however note that the sensing errors increase the cumulated regret. The smallest regret is achieved as expected in the case of perfect sensing ($\epsilon = \mathbb{P}_{fa} = 0$). Moreover, we can notice that the ratio of the regret in the case of perfect sensing and in the case of sensing errors characterized by $\epsilon \neq 0$ is approximately equal to $1/(1 - \epsilon)$ which supports the theoretical results.

The optimal channel selection percentage p achieved by the UCB_1 algorithm until the slot number t is illustrated in Figure 4.12, where $p = 100 \cdot \frac{\sum_{m=0}^{t-1} \mathbf{1}_{\{a_m=1\}}}{t}$. As one can observe the percentage of optimal channel selection increases progressively and tends to get closer and closer to 100% as the slot number increases.

As for the regret analysis, we observe that the performance of the UCB_1 algorithm decreases when the \mathbb{P}_{fa} increase. Thus, the UCB_1 with perfect sensing performs best. The increasing rate of the other curves is slower depending on their sensing capabilities. As proven in the theoretical analysis provided hereabove, all UCB_1 algorithms converge to the best channel, however the less accurate is their sensing outcome, the slower becomes their convergence rate.

4.6 Simulink based Reinforcement Learning Scenario

This scenario is described in our Paper [133] and aims at illustrating a more realistic framework. For that purpose, primary users are modeled through a OFDMA network, while the secondary user relies on an energy detector and UCB_1 algorithms in order to exploit found communication opportunities. The results of this scenario support the theoretical results provided in Section 4.5.

4.6.1 Primary Network: OFDM

In our simulation model an OFDM transmitter is used to represent the bands' occupancy by primary users. As a matter of fact, OFDM enables to simulate a large spectrum which is assimilated to the spectrum band that PUs and SUs can use and share. Consequently, it is convenient to model independent primary transmitters by adjusting the OFDM sub-bands occupancy relying on appropriate stochastic distributions $\{\theta_n\}_{n \in \mathcal{A}}$ and add SUs to fill the spectrum holes left vacant by PUs.

As a matter of fact, the outputs $\{x_{m,t}\}_{m=0}^{m=N_{sub}-1}$ of the OFDM modulator with N_{sub} sub-carriers affected with the weights $\{c_{m,t}\}_{m=0}^{m=N_{sub}-1}$ can be written as:

$$x_{m,t} = \sum_{l=0}^{N_{sub}-1} c_{m,t} e^{2\pi j \frac{l}{N_{sub}} m} \quad (4.36)$$

Within our cognitive radio model, the N_{sub} sub-carriers are divided among the PUs to design N channels accessible to the SUs. We assume that the value N_{sub}/N is an integer. In order to simulate the extinguishing and the turning on of each virtual transmitter, a part of the model consist in setting to zero subsets of the sub-carriers. Thus, we can write Equation 4.36 as follows:

$$x_{m,t} = \sum_{n=0}^{N-1} S_{n,t} \left(\sum_{l=0}^{N_{sub}/N-1} e^{2\pi j (\frac{n}{N} N_{sub} + l) \frac{m}{N_{sub}}} \right) \quad (4.37)$$

In this section, we consider an OFDM spectrum divided into 16 channels such that their non-occupancy probabilities verify: $[\mu_1, \dots, \mu_{16}] = [0.1, 0.1 : 0.5 : 0.9]^{(12)}$.

4.6.2 Sensing: Energy Detector

In our work, we chose to implement the Neyman-Pearson Energy Detector (NP-ED) as described in Chapter 3. It has been extensively analyzed [94] for its properties as a semi-blind low complexity spectrum sensor, since it ignores the characteristics of the received signals and only relies on the perceived energy of the signal (in a given band). The main detection process relies on the comparison of the perceived energy, \mathcal{T}_t at the slot t , to a fixed threshold that depends on the desired performances of the detector as well as the noise power level. The following equations remind us of the expressions of \mathbb{P}_{fa} (also referred to by the letter ϵ) and $\mathbb{P}_{md,t}$:

$$\begin{cases} \mathbb{P}_{fa} = 1 - F_{\chi_M^2} \left(\frac{\xi(\alpha_{fa})}{\sigma_n^2} \right) \\ \mathbb{P}_{md,t} = F_{\chi_M^2} \left(\frac{\xi(\alpha_{fa})}{\sigma_n^2 + \sigma_{x,t}^2} \right) \end{cases}$$

where $F_{\chi_M^2}(\cdot)$ refers to the cumulative distribution function of a χ^2 -distribution with M degrees of freedom (i.e., the number of sensed samples in a slot), $\xi(\alpha_{fa})$ is the chosen threshold to guaranty a false alarm equal to α_{fa} and $\sigma_{x,t}^2$ refers to the power lever of the received signal at the slot t . Finally for the illustration's sake, we assume a perfect knowledge of the noise level.

The next subsection, illustrates this study by presenting a display of the considered Simulink model, a throughput curve as well as a channel selection curve.

⁽¹²⁾Namely $[0.1, 0.1, 0.2, 0.25, 0.3, \dots, 0.8, 0.85, 0.9]$

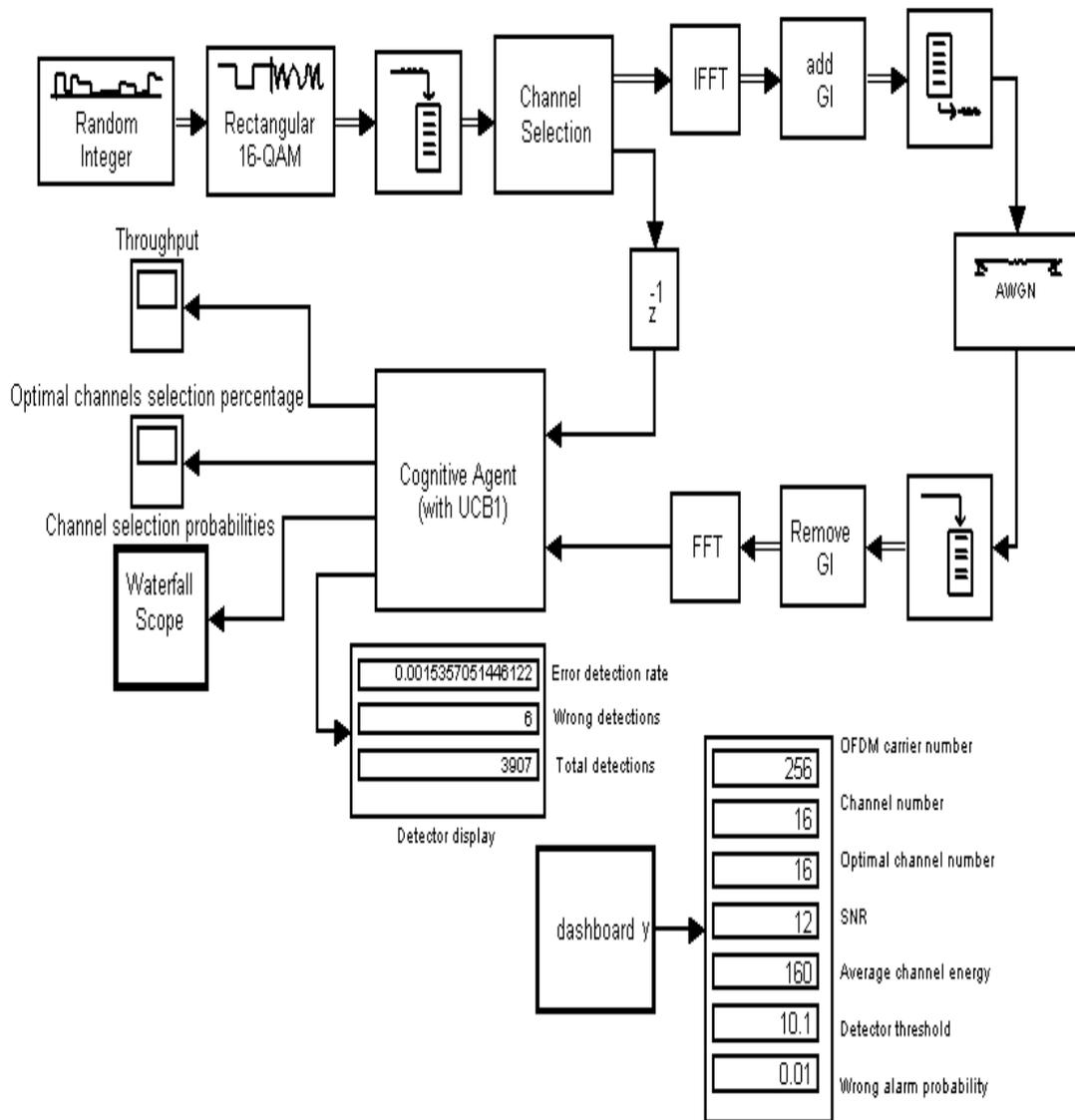


Figure 4.14: Simulink model of the OSA framework. At the top, the OFDM transmitter simulates the primary network (many independent transmitters). The cognitive agent refers to the SU's decision making engine. The dashboard presents the main parameters regarding the current simulation.

4.6.3 Simulation results

The simulations were conducted with an exploration parameter $\alpha = 1.2$. This is considered as very cautious. As a matter of fact, any $\alpha > 0.5$ leads to a guaranteed theoretical convergence. On practical scenarios, $\alpha = 0.5$ would lead to the best results with a guaranteed convergence to the optimal channels. Consequently, our results underestimate (on purpose) the rate of convergence of UCB_1 . Thus, in these simulations, we aim at supporting the previously introduced theoretical results rather than tuning UCB_1 algorithm.

The complete model is summarized in Figure 4.14. In this snapshot, we can observe at the top chain of the figure the primary network, while the SU is represented by its cognitive agent (CA, i.e., the decision making engine of the CR). Note that at the reception chain, the operations referred to as ‘Remove GI’ and ‘FFT’ are not needed. They are used to validate the OFDM chain. Both signal detection and channel selection policies are implemented in the block ‘Cognitive Agent’. Several display screens are added to monitor the behavior and performance of the SU.

Figure 4.15 plots the average throughput of the secondary user. we can see, that it grows quickly to the optimal expected performances. This implies that the CA converges to the optimal channel. This is confirmed in Figure 4.16. As a matter of fact, we observe the evolution of the channel selection process conducted by the SU. In this case, we consider 16 channels and 4000 slots. As expected, during the first iterations, the channel explores. This appears in the quasi-uniform selection of the channels. Then, as the number of trials grows, we observe that the proportion of selections of the channel 16, viz. the optimal channel in this case, grows significantly. This can be understood as an often selection of that channel during the remaining iterations, which means that the SU is exploiting (while exploring) this resources. In this specific scenario, we can see that after 4000 iterations, the CA selected the optimal channel almost 65 % of the time. Moreover the 3 best channels are selected 80 % of the time. Finally, the UCB_1 channel selection policy combined with the energy detector seems to be efficient in this scenario, which concludes this section.

4.7 Conclusion

In this chapter, we briefly described the general concept of Multi-Armed Bandit and applied it to Dynamic Configuration Adaptation as well as Opportunistic Spectrum Access problems. We discussed different possible approaches to deal with sequential decision making under uncertainty within the MAB framework. Thus, we chose to focus on the performances of the algorithm UCB_1 as it provides a satisfactory compromise between complexity and mathematical guarantees. Several illustrations were provided to explore and understand the behavior of the UCB_1 algorithm.

Moreover, we introduced a general OSA model that takes into account sensing errors. Then we proved that the UCB_1 conserves its order-optimality; the convergence rate however slows as the sensing errors’ frequencies grow large. Once again, we illustrated the theoretical results through several simulations and scenarios. The illustration aimed at supporting the theoretical results while describing the empirical behavior of the learning algorithm.

This chapter opens the ways to many possible research paths. As a matter of fact, one cannot consider OSA in a CR context without taking into account CR networks. Moreover,

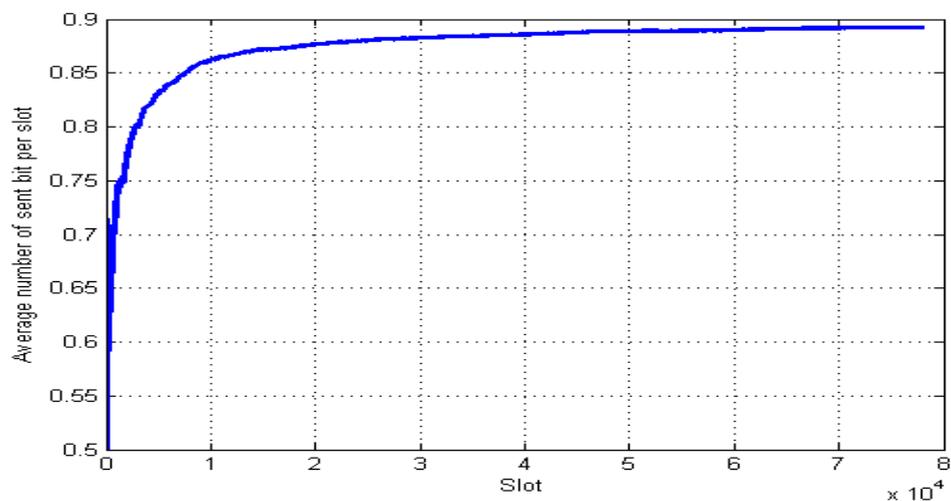


Figure 4.15: Validation of the expected theoretical convergence to the optimal throughput (i.e., $(1 - \epsilon)(1 - \mu_{16}) = 0.891$ in this case).

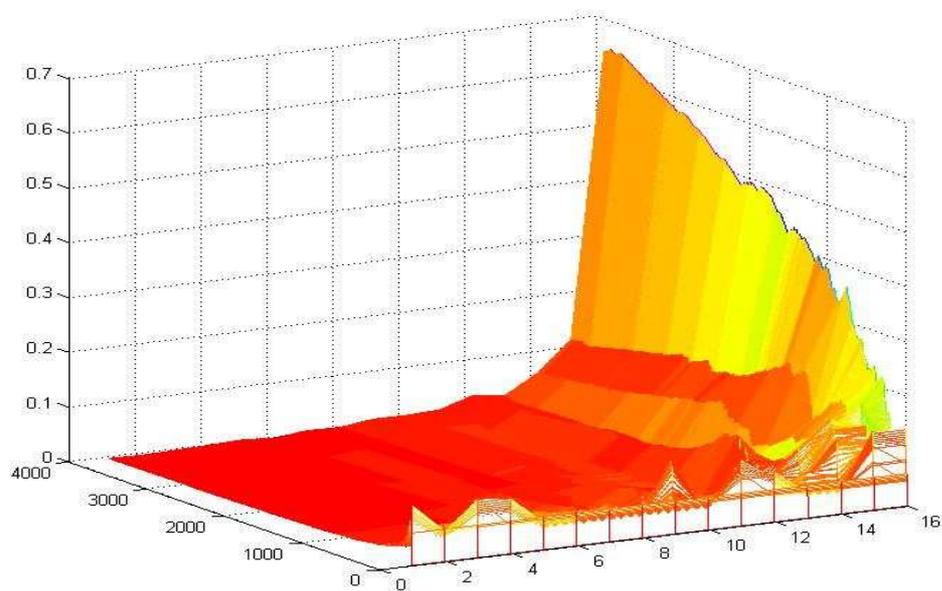


Figure 4.16: Channel selection proportions: We observe the evolution of the channel selection process conducted by the SU. In this case, we consider 16 channels and 4000 slots. In this specific scenario, we can see that after 4000 iterations, the CA selected the optimal channel almost 65 % of the time. Moreover the 3 best channels are selected 80 % of the time.

in the case of fading, for instance, the reward distributions might not comply well with the chosen algorithm. In order to answer these questions, we investigate in the next chapters both issues. On the one hand, we show that in a collaborative network, UCB_1 behaves well even in heterogeneous networks. On the other hand, we design and introduce a new algorithm referred to as *Multiplicative UCB* (i.e., MUCB) to tackle exponential distributions.

Chapter 5

Collaboration and Coordination in Secondary Networks for Opportunistic Spectrum Access

Contents

4.1	Introduction	59
4.1.1	General Context and Challenges	59
4.1.2	Classic Illustration: Opportunistic Spectrum Access	59
4.1.3	Outline and contributions	61
4.2	Learning for Opportunistic Spectrum Access: Multi-Armed Bandit Paradigm and Motivations	61
4.2.1	Multi-Armed Bandit Paradigm: Conceptual Problem Statement	61
4.2.2	Multi-Armed Bandit Paradigm: Stochastic Environment Vs Adversarial Environment	62
4.2.3	Multi-Armed Bandit Paradigm: Index based Policies	63
4.2.4	Opportunistic Spectrum Access Modeled as a Multi-Armed Bandit Problem	64
4.3	Opportunistic Spectrum Access Mathematic Model : a Multi-Armed Bandit problem	68
4.3.1	Basic Opportunistic Spectrum Access Model : Multi-Armed Bandit Notations	68
4.3.2	General Performance Evaluation of a Learning Policy and Optimality	70
4.3.3	Sample Mean Based Upper Confidence Bound Algorithms and Theoretical Performance Results	72
4.3.4	Complexity	75
4.4	Algorithm Illustration, Limits and Discussion	75
4.4.1	Configuration Adaptation Problem	75
4.4.2	OSA under perfect Channel State Information	79
4.4.3	Limits of the Theory and Discussion	82
4.5	Opportunistic Spectrum Access with Imperfect Sensing	84

4.5.1	<i>UCB</i> ₁ Performance Analysis	84
4.5.2	Simulation Results	86
4.6	Simulink based Reinforcement Learning Scenario	87
4.6.1	Primary Network: OFDM	88
4.6.2	Sensing: Energy Detector	88
4.6.3	Simulation results	90
4.7	Conclusion	90

In this chapter, we address the general case of a coordinated secondary network willing to exploit communication opportunities left vacant by a licensed primary network. Since secondary users (SU) usually have no prior knowledge on the environment, they need to learn the availability of each channel through sensing techniques, which however can be prone to detection errors. We argue that cooperation among secondary users can enable efficient learning and coordination mechanisms in order to maximize the spectrum exploitation by SUs, while minimizing the impact on the primary network. To this goal, we provide three novel contributions. First, we formulate the spectrum allocation problem through a general learning model that takes into account the observation limits of the SUs. Second, we derive fundamental limits on the optimality of the Upper Confidence Bound algorithm, in case secondary users can share the rewards and have symmetric goals and communication capabilities. Third, we introduce a general coordination mechanism under uncertainty based on the Hungarian algorithm and we show, through several simulations, that it converges to the optimal channels that maximize the overall performance of the Secondary Network (SN) without prior knowledge on its nature (symmetric or not).

5.1 Introduction

The detection of opportunities and their exploitation in secondary networks can be challenging. On the one hand, the secondary users can have different perceptions of a same opportunity depending on their observation abilities. Thus, a channel available with high probability -offering substantial communication opportunities- could be discarded by a SU unable to properly detect PUs' activity. On the other hand, several SUs can be competing for the same resources. Consequently, high interference can occur among them degrading the observed quality of the resources and the realized performance of the secondary network.

This chapter addresses the spectrum allocation problem in secondary networks, through the key concepts of *learning*, *collaboration* and *coordination*. In order to implement the OSA paradigm in an efficient way, the SUs must be able to detect the communications opportunities left vacant by incumbent users. Since usually no prior knowledge is available on the occupancy pattern of the channels, *learning* abilities are needed. Between 2008 and 2011, several machine learning-based techniques have been proposed for spectrum allocation in secondary networks. Among these, the MAB techniques [100, 101, 102] introduced discussed in Chapter 4 have gained an increasing interest, due to the possibility to derive theoretical bounds on the performance of learning algorithms. However, the impact of individual sensing error on the convergence of the learning algorithm is far to be completely explored. For this reason, in this chapter we consider a collaborative network environment, where the secondary users can collaborate and share the information learnt on the occupancy pattern of the channels. *Collaboration* is a key element in Cognitive Radio (CR) networks [134, 135]. Here, we investigate if and how the utilization of collaborative techniques can enhance the performance of the learning schemes, in order to enable secondary users to fully and quickly exploit vacant resources. At the same time, while collaborative learning is fundamental to mitigate the impact of PU interference, *coordination* among SUs is required to guarantee optimal sharing of spectrum resources and to avoid inter-SU interference. The coordinator entity can be either real or virtual, but it should guarantee that -in the optimal configuration - a single SU is allocated per-channel.

In this chapter, we introduce and analyze a joint coordination-learning mechanism. We state that the suggested mechanism enables secondary networks to deal with dynamic and uncertain environment in spite of sensing errors. We propose three novel contributions in this chapter. First, we formulate the spectrum allocation problem in secondary networks as a special instance of the Multi-Armed Bandit (MAB) problem. and we propose to solve it through the UCB_1 algorithm discussed in Chapter 4. Compared to previous applications of MAB techniques on OSA issues, we address the case of cooperative learning, i.e. SUs share the rewards in order to speedup the convergence of the learning algorithm to the optimal solution. Second, while learning PUs' occupation patterns of each spectrum band, we consider general coordination algorithm whose purpose is to allocate at every iteration a unique SU per channel, in order to nullify the interference among SUs. The coordination algorithm relies on a modified Hungarian algorithm [136], and our modification aims at providing a fair allocation of the resources. Third, we derive some fundamental results on the performance of collaborative learning schemes among SUs using Round Robin based time division access scheme. More specifically, we demonstrate that -in a symmetric

scenario where all SU have the same perception of the quality of the resources (yet with sensing errors) ⁽¹⁾- the UCB_1 algorithm can efficiently learn accessing optimal solutions even without prior knowledge on the sensors performances. Both results, in the case of homogeneous and heterogeneous environments are illustrated through several simulations.

The rest of this chapter is organized as follows:

Section 5.2 discusses the works related to this chapter and found in the open literature. Section 5.3 details the considered OSA framework. To deal with uncertainty, a collaborative learning mechanism is proposed in Section 5.4. The considered coordination mechanisms are modeled as instances of Job Assignment problems, and are detailed in Section 5.5. The theoretical analysis of the joint learning-coordination framework is discussed in Section 5.6. Section 5.7 describes the collaboration mechanisms implicated in this OSA context. Finally, Section 5.8 empirically evaluates the introduced coordination and learning mechanisms, while Section 5.9 concludes the chapter.

5.2 Related Work

Several authors have already proposed to borrow algorithms from the machine learning community to design strategies for SUs that can successfully exploit available resources. We focus this brief overview on MAB related models applied to OSA problems.

To the best of our knowledge, the first extensive work that tackles spectrum band allocation under uncertainty applied to OSA, was presented in [131]. The paper presented various models where a single or multiple secondary user(s) aim(s) at opportunistically exploiting available frequency bands. Among other models, a MAB model was suggested in the case of perfect sensing (i.e., the state of a sensed channel is acquired without errors). The authors of [131] suggested the use of the algorithm UCB_1 and extended its results to the case of multi-channel selection by a single user. The case of multi-secondary users was also discussed. However a game theory based approach was suggested to tackle the problem. Such approaches lead to asymptotic Nash equilibrium, that is known to be difficult to compute in practice.

Since then, several papers suggested MAB modeling to tackle OSA related problems. In [36, 67], we compared UCB_1 and UCB_V algorithm [103, 130] in the context of OSA problems, while [70, 137] suggested to tackle multi-secondary users OSA problems modeled within a MAB framework. The algorithm analyzed in [70, 137] was borrowed from [100]. This algorithm is designed for observations drawn from Bernoulli distributions and known to be asymptotically optimal in the case of one single user. Thus, to adapt to OSA contexts, they extended the results to multi-users first. Then proved that mild modification of the algorithm, that take into account the frequencies of the errors (i.e., false alarms and miss detection), maintain the order optimality of their approach. Finally, they also considered the case of decentralized secondary networks and proved their convergence asymptotically.

Taking the sensing errors into account is a fundamental step to achieving realistic OSA models. However considering that the error frequencies are perfectly known can be limiting in some scenarios [49, 50, 51] (as discussed in Chapter 3). In the papers [68, 133]⁽²⁾,

⁽¹⁾We refer to this scenario as *symmetric* or homogeneous scenario in the following

⁽²⁾These contributions of these papers are summarized in Sections 4.5 and 4.6.

we showed that UCB_1 does not require prior knowledge on the sensors' performance to converge. However, we showed that the loss of performance is twofold. On the one hand, false alarms (i.e., detection of a signal while the band is free) lead to missing communication opportunities. On the other hand, they also lead to slower convergence rates to the optimal channel. Relying on these results, [138] provided complex empirical evaluations to estimate the benefit of UCB_1 combined with various multi-user learning and coordination mechanisms (such as softmax-UCB approach for instance).

Within a similar context, an interesting contribution can be found in [69]. They analyzed, in the case of errorless sensing, the performance of UCB_1 algorithms in the context of several secondary users competing to access the primary channel. No explicit communication or collaboration is considered in this scenario, yet, once again, UCB algorithms are proven to be efficient to handle this scenario and to have an order optimal behavior.

All hereabove mentioned papers, consider homogeneous environment (or sensing). Namely, the frequency errors for all users and through all channels are the same. An exception can be found in [71, 139]. As a matter of fact, they provided a general heterogeneous framework. It is worth mentioning that these papers do not consider a specific OSA framework. They rather consider that the observed expected *quality* of a resource can be different. Consequently, the suggested model tackles multi-users in a general MAB framework rather than a specific application. The model, referred to as combinatorial MAB framework, is solved relying on a modified version of UCB_1 algorithms and the Hungarian algorithm.

The work [139] is the closest to the one provided within this chapter⁽³⁾. Unfortunately, since their model presents a general framework, it does not explicitly take into account the impact of sensing errors, nor does it show how would perform the algorithm in the case of collaborative homogeneous networks. Moreover, the Hungarian algorithm was only introduced as a possible optimization tool to solve their mathematical problem, but it was not considered from a network coordination perspective. The latter problems is addressed by this chapter.

5.3 Network model

In this section we detail the considered OSA framework. It generalizes the model presented in Chapter 4.

5.3.1 Primary Network

The spectrum of interest is licensed to a primary network providing N independent but non-identical channels. We denote by $n \in \mathcal{D} = \{1, \dots, N\}$ the n^{th} channel. Every channel n can appear, when observed, in one of these two possible states {idle, busy}. In the rest of the chapter, we associate the numerical value 0 to a busy channel and 1 to an idle channel. The temporal occupancy pattern of every channel $n \in \mathcal{D}$ is thus supposed to follow an unknown Bernoulli distribution θ_n . Moreover, the distributions $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$ are assumed to be stationary.

⁽³⁾The work presented in this chapter was conducted independently from Papers [71, 139]. Unfortunately, due to their anteriority, we had to renounce claiming some of our findings, and to adapt our presentation to highlight our approach and the contributions yet unpublished in the literature.

As Chapter 4, we tackle hereafter the particular case where PUs are assumed to be synchronous and the time $t = 0, 1, 2, \dots$, is divided into slots. We denote by \mathbf{S}_t the channels' state at the slot number t : $\mathbf{S}_t = \{S_{1,t}, \dots, S_{N,t}\} \in \{0, 1\}^N$. For all $t \in \mathbb{N}$, the numerical value $S_{n,t}$ is assumed to be an independent random realization of the stationary distributions $\theta_n \in \Theta$. Moreover, the realizations $\{S_{n,t}\}_{t \in \mathbb{N}}$ drawn from a given distribution θ_n are assumed to be independent and identically distributed. The expected availability of a channel is characterized by its probability of being idle. Thus, we define the availability μ_n of a channel n , for all t as:

$$\mu_n \triangleq \mathbb{E}_{\theta_n} [S_{n,t}] = \mathbb{P}(\text{channel } n \text{ is free}) = \mathbb{P}(S_{n,t} = 1)$$

5.3.2 Secondary Users model

We detail in this subsection the generic characteristics of all considered SUs.

At every slot number t , the SU has to choose a channel to sense. To do so, the SU relies on the outcome of past trials. We denote by $i_t^{(k)}$ the gathered information until the slot t by the k^{th} SU. We assume that all SUs can only sense and access one channel per slot. Thus selecting a channel by a SU k can be seen as an action $a_t^{(k)} \in \mathcal{A}$ where the set of possible actions $\mathcal{A} \subseteq \mathcal{D} = \{1, 2, \dots, N\}$ refers to the set of channels available. In this chapter, all SUs collaborate through a coordination mechanism as described in the previous section. This latter, through either a centralized or decentralized approach allocates at every iteration t a different channel to each SU.

The outcome of the detection phase is denoted by the binary random variable $X_t^{(k)} \in \{0, 1\}$, where $X_t^{(k)} = 0$ denotes the detection of a signal by the k^{th} SU and $X_t^{(k)} = 1$ the absence of a signal, respectively. In the case of perfect sensing, $X_t^{(k)} = S_{a_t^{(k)}, t}$ for all SUs, where $a_t^{(k)}$ refers to the channel selected at the slot number t . However since we assumed that sensing errors can occur, the value of $X_t^{(k)}$ depends on accuracy of the detector characterized through the measure of two types of errors: on the one hand, detecting a PU on the channel when it is free usually referred to as *false alarm*. On the other hand, assuming the channel free when a PU is occupying it usually referred to as *miss detection*. Let us denote by $\epsilon_n^{(k)}$ and $\delta_n^{(k)}$, respectively the probability of false alarm, and the probability of miss detection characterizing the observation of a channel $n \in \mathcal{D}$ by the k^{th} SU:

$$\begin{cases} \epsilon_n^{(k)} = \mathbb{P}\left(X_t^{(k)} = 0 \mid S_{a_t^{(k)}, t} = 1\right) \\ \delta_n^{(k)} = \mathbb{P}\left(X_t^{(k)} = 1 \mid S_{a_t^{(k)}, t} = 0\right) \end{cases}$$

Finally, the outcome of the sensing process can be seen as the output of a random policy $\pi_s^{(k)}(\epsilon_n^{(k)}, \delta_n^{(k)}, S_{a_t, t})$ such that: $X_t^{(k)} = \pi_s^{(k)}(\epsilon_n^{(k)}, \delta_n^{(k)}, S_{a_t, t})$. The design of such policies [13] is however out of the scope of this chapter.

Depending on the sensing outcome $X_t^{(k)} \in \{0, 1\}$, the SU k can choose to access the channel or not. The access policy chosen in this chapter can be described as: “*access the channel if sensed available*”, i.e. if $X_t^{(k)} = 1$.

Notice that we assume the SUs' detectors to be designed such that for all $k \in \mathcal{K}$ and $n \in \mathcal{D}$, $\delta_n^{(k)}$ (respectively $\epsilon_n^{(k)}$) is smaller or equal to a given interference level allowed by the

primary network (respectively, smaller or equal to a given level desired by the SU), although $\{\epsilon_n^{(k)}, \delta_n^{(k)}\}$ are not necessarily known. Moreover, we assume that if interference occurs, it is detected and the transmission of the secondary user fails. Regardless of the channel access policy, when the channel access is granted, the SU receives a numerical acknowledgment. This feedback informs the SU of the state of the transmission {succeeded, failed}. Finally, we assume for simplicity reasons that for every transmission attempt, a packet $D_t = 1$ is sent.

At the end of every slot t , the SUs use the different information available to compute a numerical value, usually referred to as reward $r_t^{(k)}$ in the Machine Learning literature. This reward informs the SU of its current performance. For cooperation purposes, every secondary user shares its reward with the other SUs. All shared information as well as the used communication interface are further discussed in Section 5.7.

5.4 Learning Mechanism

The learning mechanism aims at exploiting all gathered information to evaluate the most promising resources. Thus, the performance of a learning mechanism highly depends on the sampling model of the rewards (deterministic, stochastic or adversarial for instance). In the case of stochastic sampling as defined in Section 5.3, we exploit UCB_1 learning mechanisms. As emphasized in the introduction and in Chapter 4, they have proven to be efficient while having a very low complexity. Our approach remains however consistent for a different reward sampling (cf. Paragraph 4.2.2).

The estimation of the performance of a resource $n \in \mathcal{D}$ considered by UCB_1 indexes relies on the computation of the average reward provided by that resource until the iteration t to which a positive bias is added. We remind the reader of the general form of UCB_1 indexes:

$$B_{T_n(t)} = \bar{W}_{T_n(t)} + A_{T_n(t)} \quad (5.1)$$

where $A_{T_n(t)}$ is an upper confidence bias added to the sample mean $\bar{W}_{T_n(t)}$ of the resource n after being selected $T_n(t)$ times at the step t :

$$\begin{cases} A_{T_n(t)} = \sqrt{\frac{\alpha \cdot \ln(t)}{T_n(t)}} \\ \bar{W}_{T_n(t)} = \frac{\sum_{m=0}^{t-1} r_m \cdot \mathbf{1}_{\{a_m=n\}}}{T_n(t)} \end{cases} \quad (5.2)$$

For that purpose, we define $B_{T_n^{(k)}(t)}^{(k)}$ as the computed index associated to a resource n observed $T_n^{(k)}$ times by the k^{th} decision maker until the iteration t , and $A_{T_n^{(k)}(t)}^{(k)}$ its associated bias.

Let $B(t)$ refer to a K by N matrix such that component of $\{B(t)\}_{\{k,n\}} = B_{T_n^{(k)}(t)}^{(k)}$, where $k \in \mathcal{K}$, $n \in \mathcal{D}$ and:

$$\tilde{k} = (k - 1 + t) \oslash K + 1$$

The form of $B(t)$ is explicitly designed to ensure fairness among Secondary Users. As a matter of fact, the rows of $B(t)$ switch at every iteration in a Round Robin way. It is important as some algorithm, such as the Hungarian algorithm, allocates the resources

considering the first lines first. Thus if there are several optimal solutions they would always pick the same. The design of $B(t)$ alleviates this problem.

For the rest of this chapter, $B(t)$ is the considered estimated weight matrix for coordination algorithms.

Channel Selection Policy 1 ($CC-UCB_1(R, \alpha)$). *The overall algorithm can be described as follows. Let R be a positive integer, $R = 1$ if heterogeneous network and $R = K$ if homogeneous network.*

Every R rounds: computation and coordination.

- *Step 1: Compute $B(t)$ using $UCB_1(\alpha)$ algorithm.*
- *Step 2: Compute the output of the coordination mechanism $a_t^{(k)}$ for all users k .*

$$\max_{\{a_t^1, \dots, a_t^K\}} \sum_{k=1}^K \{B(t)\}_{\{k, a_t^k\}} \quad (5.3)$$

Thus, every SU is allocated R channels to access in a Round Robin fashion for the next R iteration.

At every iteration during R rounds: sense and access the channels:

- *Step 3 (for R iterations): Sense the channels and Access them if sensed free.*

At the end of R rounds: collaboration-information sharing

- *Step 4: Share the sensing-access outcomes of the last R rounds.*

As shown by the Channel Selection Policy 1, the second step relies on a coordination mechanism to perform channel allocation among the SUs. These mechanisms are usually equivalent to *Job Assignment* problems. In the following, we introduce two coordination algorithms in order to allow fair resource allocation among SUs: (i) the Hungarian algorithm based coordination and (ii) the Round Robin based coordination.

5.5 General Resource Allocation Problem

The first contribution of this chapter, is to introduce a general resource allocation framework to discuss OSA scenarios. We show that OSA related problems are simple applications of this general framework.

5.5.1 Coordination and Job Assignment Problems

We argue in this subsection that the coordination of multi-secondary users can be formulated as a job assignment problem. We first introduce the general notations related to the

job assignment framework. Then we present this latter as an adequate tool to model OSA related coordination problems.

Let us consider a set \mathcal{K} of K *workers* or *decision makers* and a set \mathcal{D} of N *jobs* or *resources*. Let us denote by λ the K by N weight (or cost) matrix where $\{\lambda\}_{\{k,n\}} = \lambda_n^{(k)}$ refers to a weight associate to the *decision maker* $k \in \mathcal{K}$ assigned to the *job* or *resource* $n \in \mathcal{D}$ such that:

$$\lambda = \begin{bmatrix} \lambda_1^{(1)} & \cdots & \lambda_N^{(1)} \\ \lambda_1^{(2)} & \cdots & \lambda_N^{(2)} \\ \vdots & \vdots & \vdots \\ \lambda_1^{(K)} & \cdots & \lambda_N^{(K)} \end{bmatrix}$$

We assume that every *decision maker* can be assigned on a unique *resource*. Moreover, every *resource* can be handled by only one *decision maker*. Let $a_n \in \mathcal{K}$ refer to the assigned *decision maker* to the *resource* $n \in \mathcal{D}$. The resource allocation problem can be formalized as follows. Find an optimal set of assignments such that the total weight is maximized (or equivalently, the total cost minimized):

$$\max_{\{a_1, \dots, a_N\}} \sum_{n=1}^N \lambda_n^{(a_n)} \mathbf{1}_{\{\exists a_n\}} \quad (5.4)$$

where the logic expression⁽⁴⁾ $\{\exists a_n\}$ refers to the existence of a *decision maker* assigned to the resource n .

In the case of OSA a coordinator generally aims at canceling harmful interference between the different users. To that purpose, a coordinator usually allocates different resources to different users, or uses advanced signal processing techniques to alleviate interference effects on the users' performances (e.g., Time Division Multiple Access, Frequency Division Multiple Access or Code Division Multiple Access to name a few):

Definition 11 (Coordinator or Facilitator). *Let \mathcal{K} refer to a set of decision making agents. We refer to as Coordinator or Facilitator any real or virtual entity that enables the different decision makers to jointly plan their decisions at every iteration.*

For the sake of coherence in speech, let us consider a set \mathcal{K} of K SUs (viz., the workers or decision makers) willing to exploit a set \mathcal{D} of N primary channels (viz., the resources). Moreover let $\{\mu_n\}_{\{n \in \mathcal{D}\}}$ and $\pi_s^{(k)}$ denote, respectively, a characteristic measure that quantifies the quality of the primary channels (e.g., their expected availability or Signal to Noise Ratio for instance) and a sensing policy that characterizes the observation abilities of the k^{th} SU. Then $\lambda_n^{(k)} = f_{\pi_s^{(k)}}(\mu_n)$ represents the quality of a primary resource observed by the k^{th} SU, where $f(\cdot)$ represents a (possibly implicit) functional relationship that relates primary resources' quality to SUs observations.

Consequently, the stated problem in Equation 5.4 is equivalent, when allocating primary resources among secondary users, to maximize the secondary network's observed performance.

⁽⁴⁾Indicator function: $\mathbf{1}_{\{\text{logical_expression}\}} = \{1 \text{ if logical_expression=true ; } 0 \text{ if logical_expression=false}\}$.

5.5.2 Coordination Mechanisms based on The Hungarian Algorithm

Suggested in 1955 by H. W. Kuhn [136], the Hungarian method is a matching algorithm that solves the job assignment problem in polynomial time. It mainly takes as an input the matrix λ (or its opposite, depending on whether it is a maximization or minimization approach) and provides as an output a binary matrix that contains a unique 1 per row and per column. This output indicates the resource allocation to the workers.

Many assignment combinations can verify the stated problem in Equation 5.4. The Hungarian algorithm provides one solution among the set of optimal matching solutions. This solution mainly depends on the matrix λ . Inverting two columns can lead to a different optimal solution if such solution exists. It is thus necessary to consider, for fairness reasons among secondary users, a permutation mechanism that changes the order of the rows of the weight matrix at every new iteration t . To this goal, we introduce the following coordination algorithm:

Coordination 1 (Hungarian Algorithm based Coordination). *Let $t = 0, 1, 2, \dots$ refers to a discrete sampling time and let $\{\lambda_n^{(k)}(t)\}_{n \in \mathcal{D}}$ refers to weights associated to the decision maker $k \in \mathcal{K}$ at the iteration t . Let $\lambda(t)$ refer to a K by N matrix such that $\{\lambda(t)\}_{\{k,n\}} = \lambda_n^{(\tilde{k})}(t)$, $k \in \mathcal{K}$, $n \in \mathcal{D}$ and:*

$$\tilde{k} = (k - 1 + t) \oslash K + 1$$

where $a \oslash b$ refers to the modulo operator that returns the remainder of the division of a by b .

Let $H(t)$ refer to the output of the Hungarian algorithm with input $\lambda(t)$.

Then the k^{th} decision maker is assigned the resource $a_t^{(k)}$ verifying:

$$a_t^{(k)} = n \text{ s.t. } H(t)_{\{\tilde{k},n\}} = 1; \quad (5.5)$$

5.5.3 Coordination Mechanisms based on Round Robin Algorithm

We consider in this subsection, a particular case of the introduced job assignment problem: Symmetric workers.

Definition 12 (Symmetric Behavior). *Let \mathcal{K} refer to a set of decision making agents. These agents are said to have a Symmetric Behavior if their optimization criteria, their communication abilities as well as their decision making policies are the same. In OSA contexts, a network with Symmetric Behavior transceivers is thus referred to as a Symmetric Network.*

This can be formalized as particular weight matrix with the same rows for all $k \in \mathcal{K}$, i.e., let n be a resource, $n \in \mathcal{D}$ then:

$$\forall k \in \mathcal{K} \ \{\lambda\}_{\{k,n\}} = \lambda_n$$

In this context a very simple coordination algorithm can ensure fairness among workers⁽⁵⁾:

⁽⁵⁾ Although the suggested form is original, the coordination algorithm is a simple Round Robin allocation scheme. It has been already suggested in [137] in an OSA context with no observation errors and no collaboration.

Coordination 2 (Circular Coordination (Round Robin)). Let $t = 0, 1, 2, \dots$ refers to a discrete sampling time. We define $t' = \{0, 1, \dots, \lfloor t/K \rfloor\}$ as a sequel of integers updated every K iterations. Let $\{\lambda(t)\}$ refer to the weight matrix computed at the iteration t , and let $\sigma_n(t)$ be the permutation function used at the iteration t to order the rows of the weight matrix values. We assume that $\{\lambda(t)\}$ and $\sigma_n(t)$ are computed every K iterations such that for all $t \in [Kt', K(t' + 1) - 1]$, $\{\lambda(t)\} = \lambda(Kt')$ and $\sigma_n(t) = \sigma_n(Kt')$.

Then the k^{th} decision maker selects the channel n verifying:

$$\begin{cases} \sigma_n(t) = a_t^{(k)} \\ a_t^{(k)} = (k - 1 + t) \oslash K + 1 \end{cases} \quad (5.6)$$

Coordination algorithm 2 needs to know that the network is perfectly symmetric. In case this knowledge is unavailable or the network is non-symmetric, this coordination scheme could fail. Moreover, in real scenarios, the weight matrix is usually unknown and every worker can solely access one row of the matrix: the one related to his own perception of the environment (usually prone to detection errors). Consequently, OSA related problems appears as Job Assignment problems under uncertainty. Thus, we suggest in this chapter to introduce collaboration and coordination based learning mechanisms among secondary users to alleviate the lack of information so as to converge to optimal resource allocation.

In order to compute an estimation of the weight matrix, we assume a collaboration behavior among workers to share information (introduced in Section 5.3 and discussed in Section 5.7). The shared information enables a learning mechanism to compute the estimated quality matrix as described in Section 5.4. The performance of Coordination algorithm 1 is empirically analyzed in Section 5.8, while, in the case of symmetric networks, the performance of Coordination algorithm 2 is theoretically analyzed in Section 5.6.

5.6 Theoretical Analysis

the main results of this section are both the general multi-user reward model presented in Subsection 5.6.1 and Theorem 4 in Subsection 5.6.2.

5.6.1 Definitions of the Reward and the Expected Cumulated Regret

Since we consider a coordinated network, it reasonable to assume that interference among SUs is null. Thus, relying on the previously introduced notations and assumptions, the throughput achieved by a SU_k , $k \in \mathcal{K}$, at the slot number t can be defined as:

$$r_t^{(k)} = S_{a_t^{(k)}, t} X_t^{(k)} \quad (5.7)$$

which is the reward considered in this particular framework, where $r_t^{(k)}$ equals 1 only if the channel is free and the SU observes it as free. Consequently, the expected reward achievable by a given secondary user SU_k using a channel $a_t^{(k)} \in \mathcal{D}$ can be easily computed:

$$\mathbb{E} \left[r_t^{(k)} \right] = \mathbb{P} \left(X_t^{(k)} = 1 | S_{a_t^{(k)}, t} = 1 \right) \mathbb{P} \left(S_{a_t^{(k)}, t} = 1 \right) \quad (5.8)$$

which equals to, in this case:

$$\mathbb{E} \left[r_t^{(k)} \right] = \left(1 - \epsilon_{a_t^{(k)}}^{(k)} \right) \mu_{a_t^{(k)}} \quad (5.9)$$

To relate with the job assignment problem described in Section 5.5, the weight matrix λ is equal, in this context, to the matrix $\{\mathbb{E} [r_t^{(k)}]\}_{\{k \in \mathcal{K}, n \in \mathcal{D}\}}$.

We usually evaluate the performance of a user k by its expected cumulated throughput after t slots defined as:

$$\mathbb{E} \left[W_t^{(k)} \right] = \mathbb{E} \left[\sum_{m=0}^{t-1} r_m^{(k)} \right] \quad (5.10)$$

Note that in this case, $r_t^{(k)}$ follows a Bernoulli distribution. Several other algorithms are possible to answer this distribution within MAB problems such as Robbins or Agrawal's index policies [100, 101]. They would however fail in more complex scenarios where $r_t^{(k)}$ follows an unknown distribution in $[0, 1]$ for instance. As a matter of fact, they need the knowledge of the exact reward distribution, whereas UCB_1 guarantees order optimality for any bounded reward distribution (even unknown). It briefly justifies the choice of UCB_1 in this chapter. For further details, please refer to Chapter 4.

An alternative representation of the expected performance of the learning mechanism until the slot number t is described through the notion of *regret* $R_t^{(k)}$ (or expected regret of the SU_k). The regret is defined as the gap between the maximum achievable performance in expectation and the expected cumulated throughput achieved by the implemented policy.

$$R_t^{(k)} = \sum_{m=0}^{t-1} \max_{a_t^{(k)} \in \mathcal{A}_t^{(k)}} \mathbb{E}[r_t^{(k)}] - \mathbb{E} \left[W_t^{(k)} \right] \quad (5.11)$$

where $\mathcal{A}_t^{(k)}$ denotes the subspace of channels that a given SU_k can access at the slot time t , $\mathcal{A}_t^{(k)} \subseteq \mathcal{D}$.

5.6.2 Theoretical Results: Symmetric Network

In Symmetric Networks, the expected quality of a channel n observed by all SUs is the same: $\forall k \in \mathcal{K} \lambda_n^{(k)} = \lambda_n$. If the symmetry property is known to SU, all collected information on the probed channels at the slot number t is relevant to every SU. Thus, it can be used to improve their overall learning rate. As matter of fact, in this context, the SUs combine at every iteration all gathered rewards into one common information vector i_t such that $i_t = \{i_{t-1}, \{a_t^{(k)}, r_t^{(k)}\}_{k \in \mathcal{K}}\}$. Hence, the UCB indexes computed by the SUs at every slot number t are also the same, i.e., for all users $k \in \mathcal{K}$, $B_{T_n^{(k)}}^{(k)}(t) = B_{T_n}(t)$. Notice that in Symmetric Networks the optimal set of channels \mathcal{D}^* is composed of the K channels with the highest expected reward. Consequently a simple Round Robin based coordination algorithm, as described in Coordination 2 is optimal (avoids harmful interference and is fair).

In the next Theorem, we show that the regret of the k^{th} SU in a Coordinated and Collaborative Symmetric Network is upper bounded by a logarithmic function of the number of iterations t .

Theorem 4 (Upper Bound of the Regret). *Let us consider $K \geq 1$ Symmetric Secondary Users and $N \geq K$ Primary channels. The SUs are assumed to have limited observation abilities defined by their parameters $\{\epsilon_n, \delta_n\}$ for every channel n . Assuming that the Secondary Network follows the Coordination Policy 2 to select and access the primary channels, relying on UCB₁ algorithm with parameter $\alpha > 1$, then every SU suffers an expected cumulated regret $R_t^{(k)}$, after t slots, upper bounded by a logarithmic function of the iteration t :*

$$R_t^{(k)} \leq \sum_{n \notin \mathcal{D}^*} \frac{4\alpha (\bar{\lambda}^* - \lambda_n)}{K \Delta_n^2} \ln(t + K - 1) + o(\ln(t)) \quad (5.12)$$

where the following notations were introduced:

$$\begin{cases} \lambda_n = (1 - \epsilon_n) \mu_n \\ \bar{\lambda}^* = \frac{\sum_{n \in \mathcal{D}^*} \lambda_n}{K} \\ \Delta_n = \min_{n \in \mathcal{D}^*} \{\lambda_n\} - \lambda_n \end{cases}$$

Proof. This proof relies on two main results stated and proven in Lemma 2 and Lemme 3 (C.f. Appendix). As a matter of fact, Lemma 2 shows that the regret can be upper bounded by a function of the expected number of pulls of sub-optimal channels:

$$R_t^{(k)} \leq \sum_{n \notin \mathcal{D}^*} \frac{(\bar{\lambda}^* - \lambda_n) \mathbb{E} [T_n (\lfloor \frac{t}{K} \rfloor K + K - 1)]}{K} \quad (5.13)$$

Then Lemma 3 upper bounds $\mathbb{E} [T_n (\lfloor \frac{t}{K} \rfloor K + K - 1)]$ by a logarithmic function of number of iterations t :

$$\mathbb{E} \left[T_n \left(\left\lfloor \frac{t}{K} \right\rfloor K + K - 1 \right) \right] \leq \frac{4\alpha}{\Delta_n^2} \ln(t + K - 1) + o(\ln(t)) \quad (5.14)$$

□

For the case $K = 1$, $\epsilon_n = \epsilon$ and $\delta_n = \delta$, we find the classic result stated in our paper [68] (cf. Section 4.5):

$$R_t \leq \sum_{n \neq 1} \frac{4\alpha}{((1 - \epsilon)(\max_{n \in \mathcal{D}} \{\mu_n\} - \mu_n))} \ln(t) + o(\ln(t)) \quad (5.15)$$

5.6.3 Non-Symmetric Network, the Heterogeneous case

In the case of Non-Symmetric Networks, we can apply the upper bound provided in Paper [139]. As a matter of fact, our approach that decomposes, on the one hand the learning step and on the other hand the coordinating step, is equivalent to the algorithm referred to as *Learning with Linear Regret* (LLR) in [139]. More specifically, the authors of [139]

prove that if the exploration parameter of the UCB_1 algorithm, i.e. the α factor, verifies this condition: $\alpha \geq L$ where $L = N \wedge K = K$ and \wedge refers to the minimum operator, then the LLR algorithm has an order optimal behavior (i.e., expected cumulated regret upper bounded by a logarithmic function of the time). In our case, the logarithmic regret scales linearly with the value: $(N \wedge K)^3 NK$ as reported in [139].

However fairness is not considered in Paper [139]. Our suggested joint coordination-learning mechanism alleviates these problem. It is easy to verify that the same results discussed in [139] hold also when when the Coordination algorithm 1 is used for spectrum selection. Consequently, a joined coordination-learning mechanism in Non-Symmetric environments is order optimal.

Although this result is fundamental to many resource allocation problems under uncertainty, two questions remain unanswered in [139]:

- Although the theory, in [139], constrains α to values larger than K (in our case), does it mean that the algorithm fails for smaller values? Note that the larger α is, the longer it takes to converge.
- With the result provided for Non-Symmetric Environment, it is obvious that the same mechanisms would also work for Coordinated, non-collaborative, Symmetric Environments. Is it possible to provide tighter bounds for the regret and to use smaller value for the exploration parameter α ?

Both questions are tackled in this chapter. On the one hand, the previous subsection tackled the first question. We see from the results of Theorem 4 that the logarithmic function scales as $1/K$, improving tremendously the scale found in the case of heterogeneous environments. On the other hand, the simulations discussed in Section 5.8 suggest a piece of answer to the second question.

5.7 Information Sharing: Discussion

An efficient communication process relies on reliable information exchange. Thus, we assume in this chapter that the communication interface used by Cognitive Radio (CR) SUs to share information is a Common Control Channel⁽⁶⁾ (CCC). The CCCs are used, on the one hand, between a transmitter and a receiver (which can be a secondary base station or another SU), and on the other hand, among all transmitters and receivers for cooperation purposes. The information transmitted through this vessel is furthermore assumed to be received without errors.

Thus from a *Transmitter-Receiver's* perspective, CCCs' purposes are twofold: configuration adaptation and acknowledgment messages transmission.

5.7.0.1 Configuration adaptation

To initiate a transmission, both the transmitter and the receiver have to agree on a particular frequency band and on a communication configuration (e.g., modulation). In this

⁽⁶⁾Whether to use or not CCCs for cognitive radio networks is still a matter of debate in the CR community. This debate is however out of the scope of this chapter. Notice that the conclusions of this study would still apply if we assumed any other kind of reliable information exchange interface among secondary users.

particular case, *configuration* refers, solely, to *frequency band*. Thus we assume that at every slot t the transmitter informs the receiver of the channel selection outcome before transmitting.

5.7.0.2 Acknowledgment

At the end of every transmission attempt the receiver has to confirm the reception of the transmitted packet. In case of a successful transmission, the transmitter receives an *ACK* message from the receiver. Otherwise, in case of PU interference, it receives a *NACK* message.

5.7.0.3 Information sharing

As mentioned in Section 5.3.2, at the end of every slot t , and for cooperation purposes, a communication period is dedicated to share feedback information among SUs.

Thus, from the secondary users' network perspective, CCCs are used, in general, to share SUs computed rewards. As a consequence, a given SU can coordinate its behavior according to other SUs. Moreover, in the case of Symmetric Networks, he can learn faster by relying on the outcomes of the other SUs' attempts, gathered on bands it did not address, at the slot number t .

5.8 Empirical Evaluation: Simulation Results

In this section, we describe and show the simulation results aimed at illustrating the herein suggested resource selection mechanisms. We first describe the general experimental protocol and the considered scenarios in Subsection 5.8.1. Subsection 5.8.2 presents and discusses the simulation results pertaining to the regret analysis. Subsection 5.8.3 show the results pertaining to the secondary network performance analysis.

5.8.1 Scenario and experimental protocol for the regret analysis

We consider 3 secondary users willing to exploit 10 primary channels with unknown expected occupancy patterns $\mu = \{\mu_n\}_{\{1, \dots, 10\}}$. For the sake of generality, we do not provide explicit numerical values to PUs' channel occupancy and to the probability of false alarms. The impact of sensing errors has been analyzed and illustrated in a previous work [68].

We denote by $\lambda_n^{(k)}$ the expected reward of a resource n observed by a user k . We, however consider that the occupation state n observed by a user k at the slot t follows a Bernoulli distribution with parameter $\lambda_n^{(k)}$. Thus, the application to OSA related scenarios is straightforward as: $\lambda_n^{(k)} = (1 - \epsilon_n^{(k)}) \mu_n$ in this context.

For illustration purposes we tackle two scenarios. On the one hand we consider 3 symmetric users. While on the other hand, we consider that the 3 secondary users are divided into 2 sets: two symmetric users sharing the spectrum with a last secondary user whose optimal channel do not belong to the set of optimal channels of the other secondary users, such that:

Scenario 1 (Symmetric network). *We consider a quality matrix λ defined as:*

$$\lambda = \begin{bmatrix} 0.1 & 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & 0.7 & 0.8 & 0.9 \\ 0.1 & 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & 0.7 & 0.8 & 0.9 \\ 0.1 & 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & 0.7 & 0.8 & 0.9 \end{bmatrix}$$

Scenario 2 (Non-symmetric network). *We consider a quality matrix λ defined as:*

$$\lambda = \begin{bmatrix} 0.1 & 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & 0.7 & 0.8 & 0.9 \\ 0.1 & 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & 0.7 & 0.8 & 0.9 \\ 0.1 & 0.1 & 0.2 & 0.3 & 0.4 & 0.7 & 0.9 & 0.7 & 0.7 & 0.6 \end{bmatrix}$$

These scenarios aim at illustrating both Hungarian and Round Robin based coordination algorithms. We expect the channel selection algorithm, relying on both learning and coordinations mechanisms to be able to converge to the set of optimal channels in Scenario 1. However, in Scenario 2 only the Hungarian algorithm based coordinator is illustrated as a Round Robin approach would be inefficient.

During all experiments, the learning parameter α is selected such that $\alpha = 1.1$ (to respect the conditions of Theorem 4). Notice that these simulations were conducted so as their respective results and conclusion could be generalized to more complex scenarios.

Finally, the presented results are averaged over 30 experiments with a final horizon equal to 1 000 000 slots to obtain reliable results.

5.8.2 Simulation results: Regret Analysis

The averaged regret -over the number of SUs- of four algorithms are illustrated in Figures 5.1(a) and 5.1(b) in the context of Scenario 1: Figure 5.1(a) shows the regrets of the Hungarian algorithm, respectively, with or without common information vector (i.e. with or without collaborative learning), while Figure 5.1(b) correspond to Round Robin based coordination algorithms with common information vector. In this latter case, one algorithm updates its information vector every 3 iterations (i.e., every K iterations as considered in Theorem 4), while the second one updates its information vector every slot.

On the one hand, Figure 5.1(b) illustrates Theorem 4. As a matter of fact, we observe that the regrets of Round Robin based algorithms are similar and have indeed a logarithmic like behavior as a function of the slot number. This behavior is observed for all four simulated algorithms. Secondly, as expected, the Hungarian based coordinator with collaborative learning performs as well as Round Robin based coordinators.

On the other hand, Figure 5.1(a) shows the impact of coordination with individual learning (the shared information is only used for coordination purpose). In this case the regret grows, as expected, larger by a factor approximatively equal to K . In this case where the users are symmetric but unaware of that fact, they do not exploit other users' information to increase their respective learning rate. The collected information from their neighbors is solely used to compute the quality matrix λ to enable coordination. Thus, we observe in Figure 5.1(a) that the Hungarian algorithm is still able to handle it however, as already noticed, with a loss of performance.

In the case of Scenario 2, Round Robin based coordination algorithms are in general not efficient. Consequently, we do not illustrate them in this context. Figure 5.2 shows the proportion of time the Hungarian algorithm based coordinator allocates the different

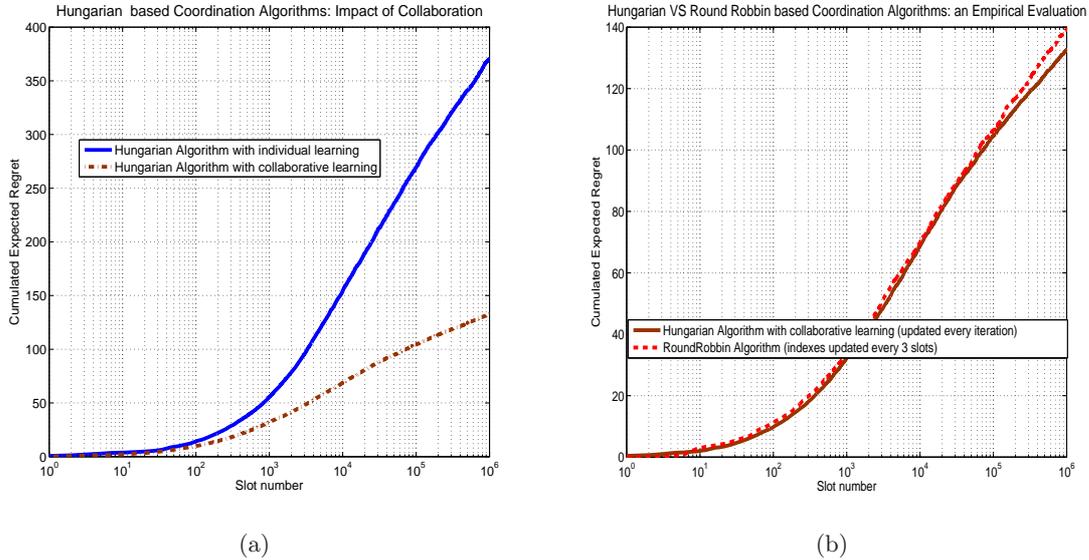


Figure 5.1: Collaboration, Learning and Coordination in the case of Symmetric Networks: averaged regret. The simulation results show that both Hungarian algorithm and Round Robin based coordinators can efficiently learn to allocate the resources among the SUs. All curves are computed with $\alpha = 1.1$. Left Figure shows the impact of collaboration on the learning process in symmetric networks. Right curves compares learning mechanisms with both Hungarian coordination or Round Robin coordination. We notice that their performance is quite similar.

secondary users to their respective optimal sets. We can observe that the curves increase rather quickly which indicates that the algorithm allocates the SUs to their respective optimal sets most of the time after a first learning phase. Theoretical analysis as well as testbed-based experiments are currently under investigation to confirm these results.

5.8.3 Simulation results: Network Performance Analysis

In this subsection, we evaluate the performance of joint collaboration-cooperative learning scheme from the point of view of secondary network performance. To this aim, we model a primary network with $N=10$ channels, and a secondary network composed of $K=4$ transmitter nodes. The temporal occupation pattern of the N channels is defined by this vector θ of Bernoulli distributions: $\{0.1, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. All the SU have a fixed probability of sensing miss-detection and sensing false-alarm, i.e. a Symmetric Network scenario is considered. Unless specified otherwise, we set $\epsilon_n^{(k)}=0.2$, for all SU k and channel n . At each slot, each SU k decides a channel to sense, and transmits a packet of 1000 bytes if the channel is found idle. No transmission attempt is performed in case the channel is sensed occupied by a PU. Both interferences among SUs and between a SU and a PU are taken into account in the model. If no interference occurs during the SU transmission, then an ACK message is sent back to the SU transmitter. Otherwise, the data packet is discarded by the SU receiver node. Thus, at each slot t , each SU k can experience a local throughput $TP^k(t)$ equal to 0 or 1000 bytes, based on interference and sensing conditions. The average network throughput $NTP(t)$ is defined

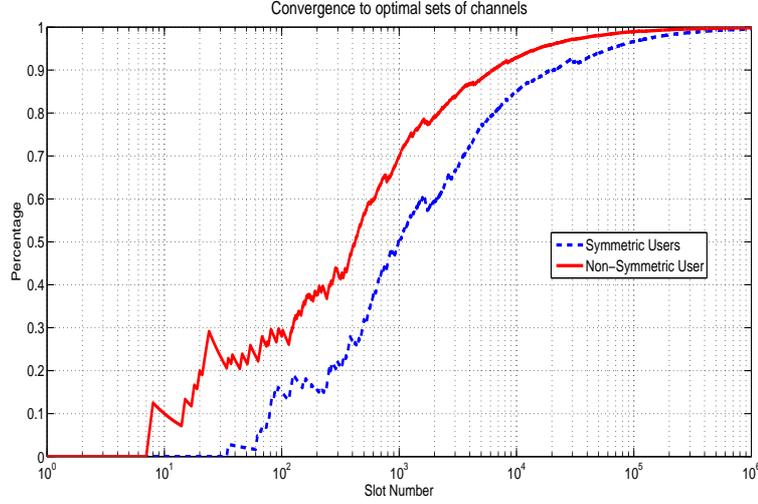


Figure 5.2: Percentage of time the Hungarian algorithm based coordinator allocates the different secondary users to their respective optimal sets. The exploration parameter α is chosen equal to 1.1. This value is smaller than the minimum value suggested by the theory. We observe however that the algorithm remains consistent.

as the average amount of byte successfully transmitted in the secondary network at each slot t , i.e.: $NTP(t) = E[\sum_{k=1}^K TP^k(t)]$.

We consider four different configurations of learning, cooperation and coordination schemes in our analysis:

- C1 (*Random, No Learning*): no learning is employed by SUs. At each slot, each SU chooses randomly the channel to sense among the available N channels.
- C2 (*Individual Learning, No Coordination*): each SU employs the UCB_1 algorithm to learn the temporal channel usage. No coordination and collaboration mechanisms are used. At each slot t , each SU k chooses randomly based on the local UCB_1 -index associated to each channel. More specifically, the probability to select channel n is computed proportional to $1 - B_{T_n^{(k)}(t)}^{(k)}$. The probabilities are normalized so that their value is between 0 and 1, and their sum equals one.
- C3 (*Cooperative Learning, No Coordination*): as before, each SU employs the UCB_1 algorithm to learn the temporal channel usage, and shares the rewards received at each slot t . However, no collaboration mechanism is used. The channel selection is performed as the previous case.
- C4 (*Cooperative Learning, Cooperation*): the complete Channel Selection Policy 1 described in Section 5.4 is evaluated. The Round Robin algorithm is considered for channel access coordination.

Figure 5.3(a) shows the network throughput as a function of the time slot t , averaged over 1000 simulation runs. As expected, the S1 scheme experiences the lowest throughput, since it does not take into account any mechanism to prevent SU and PU interference.

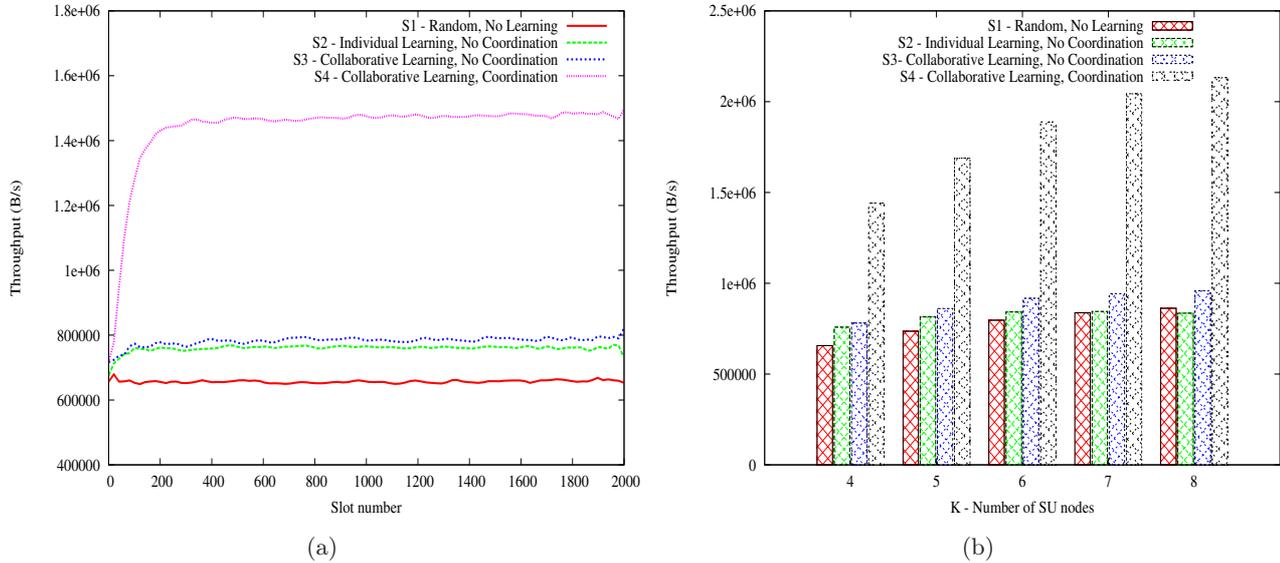


Figure 5.3: The network throughput over simulation time in a scenario with $K=4$ is shown in Figure 5.3(a). The network throughput as a function of the number of SUs (i.e. K) is shown in Figure 5.3(b).

On the other hand, both S2 and S3 schemes employ learning mechanisms to derive the PU occupation patterns of each channel, and thus are able to mitigate the interference caused by incumbent PU transmissions. Moreover, Figure 5.3(a) shows that the S3 scheme slightly enhances the S2 scheme since the usage of collaborative mechanism with reward sharing reduces the occurrence of wrong channel selection events due to local sensing errors. However, both S2 and S3 schemes do not include coordination mechanisms, and thus suffer of packet losses caused by SU interference i.e. by the fact that multiple SU transmitters are allocated on the same channel. The S4 scheme nullifies the harmful interference among SUs through Round Robin coordination, and thus provides the highest performance. Figure 5.3(b) shows the average network throughput as a function of the number of *SU* transmitters in the network. Again, Figure 5.3(b) shows that the joint cooperative learning and cooperative scheme provides the highest performance over all the scenarios considered.

5.9 Conclusion

In this chapter, we have addressed the problem of Opportunistic Spectrum Access (OSA) in coordinated secondary networks. We have formulated the problem as a cooperative learning task where SUs can share their information about spectrum availability. We have analyzed the case of symmetric secondary networks, and we have provided some fundamental results on the performance of cooperative learning schemes. Moreover, we have proposed a general coordination mechanism based on the Hungarian algorithm to address the general case (i.e. both symmetric and asymmetric networks). We are planning to validate our approach on cooperative learning schemes through further theoretical analysis and Cognitive Radio testbed-based implementations.

5.10 Appendix

We introduce and prove in this section technical results used to justify the important results stated in this chapter.

Lemma 2 (Regret, general upper bound). *Let us consider $K \geq 1$ Symmetric Secondary Users and $N \geq K$ Primary channels. The SUs are assumed to have limited observation abilities defined by their parameters $\{\epsilon_n, \delta_n\}$ for every channel n . Assuming that the Secondary Network follows the Coordination Policy 2 to select and access the primary channels, relying on UCB₁ algorithm with parameter $\alpha > 1$, then every SU suffers, after t slots, an expected cumulated regret $R_t^{(k)}$ upper bounded such that:*

$$R_t^{(k)} \leq \sum_{n \notin \mathcal{D}^*} \frac{(\bar{\lambda}^* - \lambda_n) \mathbb{E} [T_n (\lfloor \frac{t}{K} \rfloor + K - 1)]}{K} \quad (5.16)$$

where $\mathbb{E} [T_n(t)]$ refers to the expected number of pulls of a given channel n (by all SUs), and where the following notations were introduced:

$$\begin{cases} \lambda_n = (1 - \epsilon_n) \mu_n \\ \bar{\lambda}^* = \frac{\sum_{n \in \mathcal{D}^*} \lambda_n}{K} \end{cases}$$

Proof. We can upper bound the regret of a user k as defined in Equation 5.11 by the regret that he suffers at the end of the considered round of K plays, i.e.,

$$R_t^{(k)} \leq R_{\lfloor t/K \rfloor K + K - 1}^{(k)} \leq \sum_{m=0}^{\lfloor t/K \rfloor} \sum_{p=0}^{K-1} \left(\bar{\lambda}^* - \mathbb{E} [r_{Km+p}^{(k)}] \right)$$

where the sum $\sum_{p=0}^{K-1} \left(\bar{\lambda}^* - \mathbb{E} [r_{Km+p}^{(k)}] \right)$ which refers to the cumulated loss during the round of K plays indexed by the round number m , can also be written as:

$$\sum_{p=0}^{K-1} \left(\bar{\lambda}^* - \mathbb{E} [r_{Km+p}^{(k)}] \right) = \sum_{n \in \mathcal{D}^*} \lambda_n - \sum_{p=0}^{K-1} \mathbb{E} [r_{Km+p}^{(k)}]$$

which justifies the second inequality. Notice that this sum is positive if and only if at least one sub-optimal channel, $n \notin \mathcal{D}^*$, is selected among the best K channels to be played during the round m .

Thus we can further upper bound the regret as follows:

$$R_t^{(k)} \leq \sum_{m=0}^{\lfloor t/K \rfloor} \sum_{n \notin \mathcal{D}^*} (\bar{\lambda}^* - \lambda_n) \mathbb{P} \left(n \in \mathcal{A}_{Km}^{(k)} \right)$$

where $\mathcal{A}_{Km}^{(k)}$ refers to the K channels with the highest indexes evaluated at the round number m evaluated by the k^{th} SU. An inversion of the two sum leads to the following expression inequality:

$$R_t^{(k)} \leq \sum_{n \notin \mathcal{D}^*} (\bar{\lambda}^* - \lambda_n) \sum_{m=0}^{\lfloor t/K \rfloor} \mathbb{P} \left(n \in \mathcal{A}_{Km}^{(k)} \right) \quad (5.17)$$

Finally, we notice that the three following equalities are verified:

$$\begin{cases} \sum_{m=0}^{\lfloor t/K \rfloor} \mathbb{P} \left(n \in \mathcal{A}_{Km}^{(k)} \right) = \mathbb{E} \left[\sum_{m=0}^{\lfloor t/K \rfloor} \mathbf{1}_{\{n \in \mathcal{A}_{Km}^{(k)}\}} \right] \\ \sum_{m=0}^{\lfloor t/K \rfloor} \mathbf{1}_{\{n \in \mathcal{A}_m^{(k)}\}} = T_n^{(k)} (\lfloor t/K \rfloor K + K - 1) \\ T_n^{(k)} (\lfloor t/K \rfloor K + K - 1) = T_n (\lfloor t/K \rfloor K + K - 1) / K \end{cases}$$

where the second equality can be read as: the number of time a channel n is selected by a user k , until the slot number $\lfloor t/K \rfloor K + K - 1$, is equal to the number of rounds the event $\{n \in \mathcal{A}_{Km}^{(k)}\}$ is verified. The third equality on the other hand, reminds us that in the context of symmetric users, all SUs share the same information vector and obtain the same index values. Consequently, if a channel n is selected at a given round, it is played exactly once by every SU. In other words, the channel is selected K times during a round of K plays.

Thus substituting and combining the three previous equalities with Equation 5.17 leads to the stated result and ends this proof. \square

Lemma 3. *Let us consider $K \geq 1$ Symmetric Secondary Users and $N \geq K$ Primary channels. The SUs are assumed to have limited observation abilities defined by their parameters $\{\epsilon_n, \delta_n\}$ for every channel n . Assuming that the Secondary Network follows the Coordination Policy 2 to select and access the primary channels, relying on UCB₁ algorithm with parameter $\alpha > 1$, then every suboptimal channel n , after t slots, has an expected number of pulls upper bounded by a logarithmic function of the number of iterations that:*

$$\mathbb{E} \left[T_n \left(\left\lfloor \frac{t}{K} \right\rfloor K + K - 1 \right) \right] \leq \frac{4\alpha}{\Delta_n^2} \ln(t + K - 1) + o(\ln(t)) \quad (5.18)$$

Proof. We start by a first coarse upper bound verified for all $u_n \in \mathbb{N}$: since every channel is to be sensed at least K times, we can write:

$$\begin{aligned} \mathbb{E} \left[T_n \left(\left\lfloor \frac{t}{K} \right\rfloor K + K - 1 \right) \right] &\leq K + u_n \\ + K \sum_{m=u_n+1}^{\lfloor \frac{t}{K} \rfloor} \mathbb{P} (n \in \mathcal{A}_{Km}; T_n(Km) > u_n + 1) &\end{aligned} \quad (5.19)$$

Since we have the following event inclusion:

$$\{n \in \mathcal{A}_{mK}\} \subseteq \left\{ B_{T_n}(m) \geq \min_{n \in \mathcal{D}^*} \{B_{T_n}(m)\} \right\}$$

We can write:

$$\begin{aligned} \mathbb{E} \left[T_n \left(\left\lfloor \frac{t}{K} \right\rfloor K + K - 1 \right) \right] &\leq K + u_n \\ + K \sum_{m=u_n+K}^{\lfloor \frac{t}{K} \rfloor} \mathbb{P} (B_{T_n}(m) \geq \min_{n \in \mathcal{D}^*} \{B_{T_n}(m)\}) &\end{aligned} \quad (5.20)$$

In this last inequality, the joint event $\{T_n(Km) > u_n + 1\}$ is left implicit to ease the notations. This will be the case in the next assertion. Moreover notice that: $\forall n \in \mathcal{D}^*$, $K \leq T_n(Km) \leq m$. Since for all $\tau \in \mathbb{R}^+$ we have the following event inclusion:

$$\begin{aligned} \{B_{T_n}(Km) \geq \min_{n \in \mathcal{D}^*} \{B_{T_n}(Km)\}\} \\ \subseteq \{B_{T_n}(Km) \geq \tau\} \cup \{\min_{n \in \mathcal{D}^*} \{B_{T_n}(Km)\} < \tau\} \end{aligned} \quad (5.21)$$

We can write:

$$\begin{aligned} \mathbb{E} [T_n (\lfloor \frac{t}{K} \rfloor K + K - 1)] &\leq K + u_n \\ + K \sum_{m=u_n+K}^{\lfloor \frac{t}{K} \rfloor} \mathbb{P} (B_{T_n}(Km) \geq \tau) & \\ + K \sum_{m=u_n+K}^{\lfloor \frac{t}{K} \rfloor} \mathbb{P} (\min_{n \in \mathcal{D}^*} \{B_{T_n}(Km)\} < \tau) & \end{aligned} \quad (5.22)$$

For the rest of the proof we assume that:

$$\begin{cases} u_n = u_n(t) = \frac{4\alpha \ln(\lfloor \frac{t}{K} \rfloor K + K - 1)}{\Delta_n^2} \\ \tau = \min_{n \in \mathcal{D}^*} \{\lambda_n\} \end{cases} \quad (5.23)$$

then we prove that:

$$\begin{cases} \sum_{m=u_n+K}^{\lfloor \frac{t}{K} \rfloor} \mathbb{P} (B_{T_n}(Km) \geq \tau) = o(\ln(t)) \\ \sum_{m=u_n+K}^{\lfloor \frac{t}{K} \rfloor} \mathbb{P} (\min_{n \in \mathcal{D}^*} \{B_{T_n}(Km)\} < \tau) = o(\ln(t)) \end{cases} \quad (5.24)$$

First, we start by the following term: $\mathbb{P} (B_{T_n}(Km) \geq \tau)$. Notice that if the event (including its implicit event) $\{B_{T_n}(Km) \geq \tau; T_n(Km) > u_n + 1\}$ is verified then there exists an integer $s : u_n + 1 \leq s \leq m$ such that the real value verifies $B_s(Km) \geq \tau$. Consequently, we can write:

$$\mathbb{P} (B_{T_n}(Km) \geq \tau; T_n > u_n + 1) \leq \sum_{s=u_n+1}^m \mathbb{P} (B_s(Km) \geq \tau) \quad (5.25)$$

Considering an index value computed as detailed in Equations 5.1 and 5.2, we can write:

$$\begin{aligned} \mathbb{P} (B_s(Km) \geq \tau) &= \mathbb{P} (\bar{W}_s(Km) \geq \tau - A_s(Km)) \\ &= \mathbb{P} (\bar{W}_s(Km) - \lambda_n \geq \tau - \lambda_n - A_s(Km)) \end{aligned} \quad (5.26)$$

Since $s > u_n + 1$, then:

$$\tau - \lambda_n - A_s(Km) \geq \Delta_n - \sqrt{\frac{\alpha \ln(Km)}{u_n}} \geq \frac{\Delta_n}{2}$$

Consequently, we can write:

$$\mathbb{P} (B_s(Km) \geq \tau) \leq \mathbb{P} \left(W_s(Km) - \lambda_n \geq \frac{\Delta_n}{2} \right) \quad (5.27)$$

$$\leq e^{-2(\frac{\Delta_n^2}{4})s} \quad (5.28)$$

$$\leq e^{-2\alpha \ln(mK+K-1)} \leq \frac{1}{(mK)^{2\alpha}} \quad (5.29)$$

where the second inequality is a concentration inequality known as Hoeffding's inequality [140]. The third inequality is once again due to the inequality $s > u_n + 1$. Finally assuming that $\alpha > 1$,

$$\sum_{m=u_n+K}^{\lfloor \frac{t}{K} \rfloor} \mathbb{P} (B_{T_n}(Km) \geq \tau) \leq \sum_{m=u_n+K}^{\lfloor \frac{t}{K} \rfloor} \sum_{s=u_n+1}^m \frac{1}{(mK)^{2\alpha}} \quad (5.30)$$

$$\leq \sum_{m=u_n+K}^{\infty} \frac{1}{(m)^{2\alpha-1}(K)^{2\alpha}} \quad (5.31)$$

$$= C_{n,\alpha} = o(\ln(t)) \quad (5.32)$$

where $C_{k,\alpha}$ exist for $\alpha > 1$, is finite and is defined as the limit of Reimann's serie:

$$\sum_{m=(u_n+K)}^{\infty} \frac{1}{(m)^{2\alpha-1}(K)^{2\alpha}}$$

We deal know with the following term: $\mathbb{P}(\min_{n \in \mathcal{D}^*} \{B_{T_n}(Km)\} < \tau)$ (including the implicit event). In order to avoid confusing optimal channels and sub-optimal channels, for the rest of this proof, we denote by n^* a channel that belongs to the optimal set \mathcal{D}^* . As for the previous proof, and since for any $\{T_{n^*}\}_{\min_{n^* \in \mathcal{D}^*}$, $K \leq T_{n^*} \leq m$, if the event $\{\min_{n^* \in \mathcal{D}^*} \{B_{T_{n^*}}(Km)\} < \tau; K \leq T_{n^*} \leq m\}$ is verified then there exists a channel $n^* \in \mathcal{D}^*$ and an integer $s_{n^*} : K \leq s_{n^*} \leq m$ such that the real value verifies $B_{s_{n^*}}(Km) \leq \tau$. To ease notations we introduce \mathbb{P}_{n^*} the considered event:

$$\mathbb{P}_{n^*} = \mathbb{P}\left(\min_{n \in \mathcal{D}^*} \{B_{T_n}(Km)\} < \tau; T_{n^*} > K\right)$$

Consequently we can write:

$$\mathbb{P}_{n^*} \leq \sum_{n^* \in \mathcal{D}^*} \sum_{s_n=K+1}^m \mathbb{P}(B_{s_{n^*}}(Km) < \tau) \quad (5.33)$$

Notice that for any $n^* \in \mathcal{D}^*$:

$$\min_{n^* \in \mathcal{D}^*} \{\lambda_{n^*}\} - \lambda_{n^*} \leq 0$$

Consequently, as for the previous proof, relying on Hoeffding's inequality, we can write:

$$\mathbb{P}_{n^*} \leq \sum_{n^* \in \mathcal{D}^*} \sum_{s_{n^*}=K+1}^m \mathbb{P}(\bar{W}_{s_{n^*}}(m) - \lambda_{n^*} < -A_{n^*}) \quad (5.34)$$

$$\leq \sum_{s_{n^*}=K+1}^m K e^{-2A_{n^*}^2 s_{n^*}} \quad (5.35)$$

$$\leq \sum_{s_{n^*}=K+1}^m K e^{-2\alpha \ln(Km)} \quad (5.36)$$

$$\leq \frac{1}{(Km)^{2\alpha-1}} \quad (5.37)$$

Finally, we can write:

$$\sum_{m=u_n+1}^{\lfloor \frac{t}{K} \rfloor} \mathbb{P}_{n^*} \leq \sum_{m=u_n+1}^{\lfloor \frac{t}{K} \rfloor} \frac{1}{(Km)^{2\alpha-1}} \quad (5.38)$$

$$\leq \sum_{m=u_n+1}^{\infty} \frac{1}{(Km)^{2\alpha-1}} \quad (5.39)$$

$$= C_{n^*,\alpha} = o(\ln(t)) \quad (5.40)$$

where $C_{n^*,\alpha}$ exist for $\alpha > 1$, is finite and is defined as the limit of Reimann's serie:

$$\sum_{m=u_n+1}^{\infty} \frac{1}{(Km)^{2\alpha-1}}$$

Finally, since: $\lfloor \frac{t}{K} \rfloor K + K - 1 \leq t + K - 1$, combining Inequalities 5.20, 5.32 and 5.40, we can finally write:

$$\mathbb{E} [T_n (\lfloor \frac{t}{K} \rfloor K + K - 1)] \leq \frac{4\alpha}{\Delta_n^2} \ln(t + K - 1) + o(\ln(t)) \quad (5.41)$$

Which ends the proof. \square

Chapter 6

Fading Environments, Exponential Reward Distributions and MUCB Algorithms

Contents

5.1	Introduction	96
5.2	Related Work	97
5.3	Network model	98
5.3.1	Primary Network	98
5.3.2	Secondary Users model	99
5.4	Learning Mechanism	100
5.5	General Resource Allocation Problem	101
5.5.1	Coordination and Job Assignment Problems	101
5.5.2	Coordination Mechanisms based on The Hungarian Algorithm	103
5.5.3	Coordination Mechanisms based on Round Robin Algorithm	103
5.6	Theoretical Analysis	104
5.6.1	Definitions of the Reward and the Expected Cumulated Regret	104
5.6.2	Theoretical Results: Symmetric Network	105
5.6.3	Non-Symmetric Network, the Heterogeneous case	106
5.7	Information Sharing: Discussion	107
5.8	Empirical Evaluation: Simulation Results	108
5.8.1	Scenario and experimental protocol for the regret analysis	108
5.8.2	Simulation results: Regret Analysis	109
5.8.3	Simulation results: Network Performance Analysis	110
5.9	Conclusion	112
5.10	Appendix	113

This chapter contrasts with Chapters 4 and 5 as it ventures the analysis of MAB environments with unbounded reward distributions. More specifically, we explore the case of exponentially distributed rewards in the sequel. Such distributions occur in many scenarios involving *channel fading* or *network services* to name a few. Thus, to answer this challenging matter we designed a new UCB algorithm. However rather than an additive form, we designed a multiplicative form that seems to provide a simple yet efficient behavior. We called this new algorithm *Multiplicative UCB* (MUCB). We prove that the suggested algorithm is order optimal in the case of Exponential distributions and we conjecture that it remains order optimal for a larger class of distributions known as Gamma distributions.

Although, this chapter shares many similarities with Chapters 4 and 5, we made the choice to allow redundancy. As matter of fact, many notions related to MAB were introduced in the context of specific OSA scenarios. Thus, to avoid ambiguity in the speech and to ensure that this chapter is self-content, necessary redundancies with previous chapters might occur in the description of the mathematical model and notations. Note that the notations regarding the MUCB index slightly differ from those used in Chapters 4 and 5. It aims at emphasizing the fact that we are dealing with two different approaches. The notations remain however coherent with previous chapters.

6.1 Introduction

Quickly evaluating the quality of a resource is a challenging matter in OSA related contexts. Since usually no prior knowledge is available on the quality of the channels, learning abilities are needed. We discussed several solutions in Chapter 4 and 5 where resources are modeled by the availability or the throughput of a pool of channels. The general approach relies on MAB models where the rewards are assumed to be drawn from bounded distributions. The UCB_1 analyzed in this thesis proved to be efficient while maintaining a low computational complexity.

Channel selection in fading environment remains however a challenging issue. As a matter of fact, it involves reward drawn from unbounded distributions such as exponential distributions. Unfortunately, optimal learning algorithms to tackle this matter prove to be complex to implement [100, 101]⁽¹⁾.

The general problem that motivates this work can be summarized as follows. We consider in this chapter the case of one SU willing to exploit a set of primary channels. Due to fading conditions, the sensed Signal-to-Noise Ratio (SNR) in every channel is assumed to follow an exponential distribution.

The main contributions of this chapter are twofold. On the one hand, we model channel selection in fading environments as a MAB problem in Section 6.2. On the other hand, we design and analyze a simple, deterministic, multiplicative index-based policy we refer to as Multiplicative Upper Confidence Bound (MUCB) index. This form is inspired from the additive form usually suggested in the Machine Learning community. Yet the form of MUCB policies and the large set of possible applications it could target, provide a contribution to machine learning that goes beyond CR's scope.

In a nutshell, the decision making strategy computes an index associated to every available arm, and then selects the arm with the highest index. Every index associated to an arm is equal to the product of the sample mean of the reward collected by this arm and a scaling factor. The scaling factor is chosen so as to provide an optimistic estimation of the considered arm's performance. The general expression of the MUCB algorithm is introduced in Section 6.3.

As for Chapter 4, it is important to clarify the MAB vocabulary used in this chapter. First of all, an 'arm' and a 'lever' refer to the same element that a gambler pulls to play a bandit-machine. Then, the ambiguity in Multi-armed bandit terminology is the following: should we consider several machines, with one arm per machine? Or should we consider one machine with several arms. Both concepts are equivalent in machine learning. For the sake of simplicity, we usually consider one arm per machine. In such case, an arm and its machine refer to the same reward source. Consequently, we can use one term or the other to designate the reward source. For the sake of clarity we shall solely use the term arm in the sequel.

Section 6.4 detail our main theoretical results showing that the MUCB policy leads to a logarithmic loss over time under some non-restrictive conditions.

⁽¹⁾Very recently (December 2011), new algorithms that tackle various heavy tailed or light tailed distributions were found on the Open Literature. They are inspired from the UCB_1 algorithm. Since, these algorithms are not able to deal with exponentially distributed rewards with a prior knowledge on their parameters (that we seek to learn), we ignore them on purpose in this Chapter.

Section 6.5 discusses some simulation results. Finally, Section 6.6 concludes.

6.2 Multi-Armed Bandits

N -Channel selection in fading environments can be modeled as a N -armed bandit. In the sequel, the decision maker -referred to as gambler- represents the SU; while the N -armed bandit refers to the probed resources.

Such a problem is defined by the N -tuple $(\theta_1, \theta_2, \dots, \theta_N) \in \Theta^N$, Θ being the set of all positive reward distributions. When pulled at a time $t \in \mathbb{N}$, each lever $n \in \llbracket 1, N \rrbracket$ (where $\llbracket 1, N \rrbracket = \{1, \dots, N\}$) provides a reward r_t drawn from a distribution θ_n associated to that specific lever. As already discussed in previous chapters, we assume that the different payoffs drawn from an arm are independent and identically distributed (i.i.d.) and that the independence of the rewards holds between the arms. However the different arms' reward distributions $(\theta_1, \theta_2, \dots, \theta_N)$ are not supposed to be the same.

In OSA problems, the reward usually quantifies the instantaneous performance of the probed channel. Namely, the availability of the channel or its throughput. In this case, we consider the measured SNR as the reward. As a matter of fact, this quantity is closely related to the capacity of the channel. In Rayleigh fading channels, the SNR follows an exponential distribution. Consequently the probability density function, $f_n(\cdot)$, of a distribution $n \in \llbracket 1, N \rrbracket$ is equal for x real positive:

$$f_n(x) = \lambda_n e^{-\lambda_n x}$$

Let $a_t \in \llbracket 1, N \rrbracket$ denote the arm selected at a time t , and let i_t be the history vector available to the gambler at instant t , i.e.

$$i_t = [a_0, r_0, a_1, r_1, \dots, a_{t-1}, r_{t-1}].$$

We assume that the SU uses a policy π to select a arm a_t at the instant t , such that $a_t = \pi(i_t)$. We shall also write $\forall n \in \llbracket 1, N \rrbracket$, $\mu_n \triangleq \frac{1}{\lambda_n} \triangleq \mathbb{E}[\theta_n]$. Moreover we assume that $\mu_n > 0$ for all n .

We briefly remind the reader of the expressions of the cumulated regret as well as the expected cumulated regret considered in this work.

The (cumulated) regret of a policy π at time t (after t pulls) is defined as follows:

$$R_t = t\mu^* - \sum_{m=0}^{t-1} r_m,$$

where $\mu^* = \max_{n \in \llbracket 1, N \rrbracket} \{\mu_n\}$ refers to the expected reward of the optimal arm.

$$\mathbb{E}[R_t] = \sum_{n \neq n^*} \Delta_n \mathbb{E}[T_{n,t}], \quad (6.1)$$

where $\Delta_n = \mu^* - \mu_n$ is the expected loss of playing arm n , and $T_{n,t}$ refers to the number of times the arm n has been played from instant 0 to instant $t - 1$.

The general idea behind the regret can be summarized as follows: if the gambler knew *a priori* which arm was the best one, he would only pull that one, and hence maximize his expected collected rewards. However, since he lacks that essential information he will suffer unavoidable loss due to exploration of suboptimal pulls.

6.3 Multiplicative upper confidence bound algorithms

This section presents our main contribution: the introduction of a new multiplicative index. Let $B_{n,t}(T_{n,t})$ denote the index of arm n at time t after being pulled $T_{n,t}$. We refer to as Multiplicative Upper Confidence Bound algorithms (MUCB) the family of indexes that can be written in the form

$$B_{n,t}(T_{n,t}) = \bar{W}_{n,t}(T_{n,t})M_{n,t}(T_{n,t}),$$

where $\bar{W}_{n,t}(T_{n,t})$ is the sample mean of arm n at step t after $T_{n,t}$ pulls, i.e.,

$$\bar{W}_{n,t}(T_{n,t}) = \frac{1}{T_{n,t}} \sum_{i=0}^{t-1} \mathbb{1}_{\{a_i=n\}} r_i,$$

and $M_{n,t}(\cdot)$ is an upper confidence scaling factor chosen to insure that the index $B_{n,t}(T_{n,t})$ is an increasing function of the number of rounds t . This last property insures that the index of an arm that has not been pulled for a long time will increase, thus eventually leading to the sampling of this arm. We introduce a particular parametric class of MUCB indexes, which we call $MUCB(\alpha)$, given as follows:

$$\forall \alpha \geq 0, M_{n,t}(T_{n,t}) = \frac{1}{\max\left\{0; \left(1 - \sqrt{\frac{\alpha \ln(t)}{T_{n,t}}}\right)\right\}} \quad (6.2)$$

We adopt the convention that $\frac{1}{0} = +\infty$.

This form offers a compact mathematical formula. However practically speaking, an arm n is played when $T_{n,t} \leq \alpha \ln(t)$. Otherwise the arm with largest finite index is played. As a matter of fact, given a history i_t , one can compute the values of $T_{n,t}$ and $M_{n,t}$ and derive an index-based policy π as follows:

$$a_t = \pi(i_t) \in \arg \max_{n \in \llbracket 1, N \rrbracket} \{B_{n,t}(T_{n,t})\}. \quad (6.3)$$

The intuition behind this multiplicative index is comparable to the one underlying the additive bounds introduced in the literature. An arm n with a high expected reward μ_n (estimated by $\bar{W}_{n,t}(T_{n,t})$) will be more likely to have a high index, and thus will be pulled more often than another arm n' having a lower $\mu_{n'}$. The multiplicative term accounts for the uncertainty in μ_n 's estimation via $\bar{W}_{n,t}(T_{n,t})$, which quickly allows to identify sub-optimal arms⁽²⁾. On the other hand, the $\ln(t)$ term insures that an arm that not been pulled for a long time will see its index increase slowly, despite the sub-optimal $\bar{W}_{n,t}(T_{n,t})$, and will eventually be pulled, thus insuring statistical consistency in the limit. Nevertheless, the $\frac{\ln(t)}{T_{n,t}}$ term drives the process towards pulling suboptimal arms less and less often as t grows, and the α parameter allows to control this aspect. Note also that, with the convention that the initial value of $M_{n,t}(T_{n,t})$ is $+\infty$, and with small enough values of α , every arm will be played exactly once in the N first pulls, since $M_{n,t}(T_{n,t})$ becomes finite as soon as arm n has been pulled once (with α small enough⁽³⁾).

⁽²⁾All UCB approaches are related to the principle of *optimism in the face of uncertainty* introduced for instance in the work of [141].

⁽³⁾For larger values of α , the number of times an arm has to be pulled before its index becomes finite is the same for each arm (it does not depend on $\bar{W}_{n,t}(T_{n,t})$).

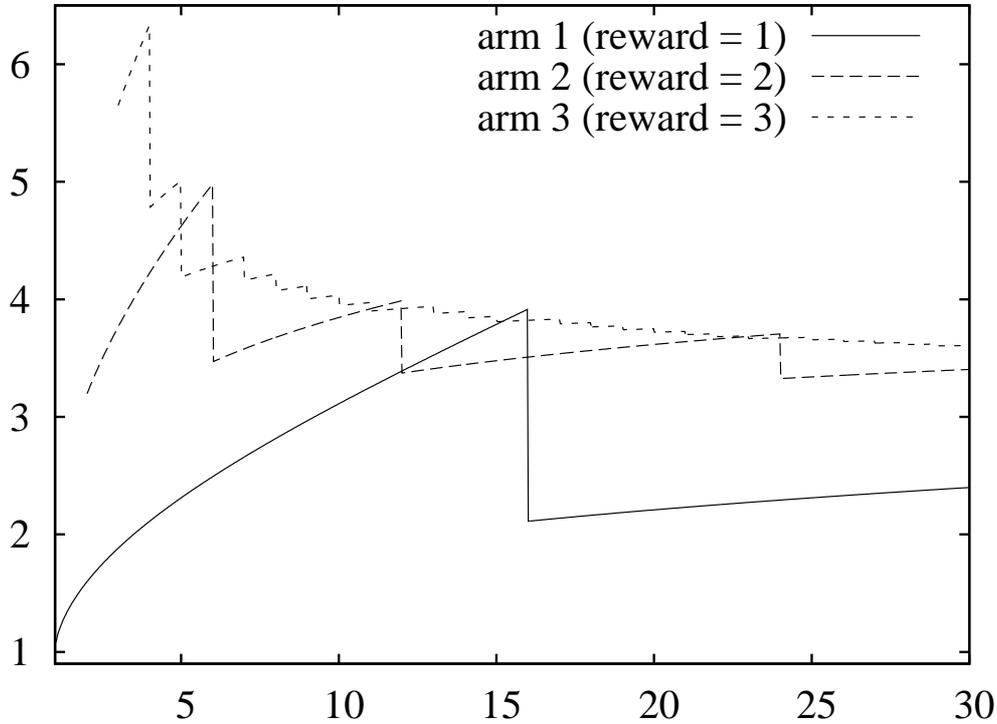


Figure 6.1: A deterministic 3-armed bandit $MUCB(0.2)$ scenario

To illustrate how this multiplicative index behaves, from a mathematical perspective, we plotted a simple scenario with three arms. All three arms are deterministic. The first one always provides a unit reward, the second one has a reward of two and the third provides a reward of three. Suppose $\alpha = 0.2$. Initially, all arms have an index of $+\infty$. At the first time step, arm 1 is pulled and its index becomes finite. Similarly for arms 2 and 3 at steps 2 and 3. Figure 6.1 presents the evolution of each arm's index as the game unfolds for 30 time steps. Note that pulling an arm decreases its index; while the indexes of arms that are not being pulled steadily increase.

In the end, the $\frac{\ln(t)}{T_{n,t}}$ -based evolution of the indexes draws the optimal arm (arm 3) to be pulled much more often than the other ones. However, it also leads to pulling the suboptimal arms once in a while, but with a decreasing frequency — hence a low probability, in a stochastic environment — to insure statistical consistency. Finally, it seems all indexes asymptotically tend to the μ^* value, and the frequency of suboptimal pulls tends to zero. The next section focuses on formalizing and proving the properties intuitively underlined by Figure 6.1.

6.4 Analysis of $MUCB(\alpha)$ policies

This section analyses the theoretical properties of $MUCB(\alpha)$ algorithms. More specifically, it shows that the MUCB policy is order optimal when dealing with exponential

distributions. Thus, it focuses on determining how fast is the optimal arm identified and what are the probabilities of *anomalies*, that is sub-optimal pulls.

6.4.1 Consistency and order optimality of $MUCB$ indexes

This subsection's aim is twofold. First we briefly recall the general notions of consistency and order optimality which are used by the algorithms in the literature to characterize the asymptotic behavior of a policy. Then, we provide a general result concerning the expected cumulated regret for $MUCB$ policies.

Definition 13 (β -consistency). *Consider the set Θ^N of N -armed bandit problems. A policy π is said to be β -consistent, $0 < \beta \leq 1$, with respect to Θ^N , if and only if:*

$$\forall(\theta_1, \dots, \theta_N) \in \Theta^N, \lim_{t \rightarrow \infty} \frac{\mathbb{E}[R_t]}{t^\beta} = 0 \quad (6.4)$$

Definition 14 (Order optimality). *A policy is referred to as order optimal if:*

$$\exists C > 0 : \lim_{t \rightarrow \infty} \frac{\mathbb{E}[R_t]}{\ln(t)} \leq C$$

The constant C introduced in the previous definition depends on the arm's distributions parameters (e.g., expected mean, variances, etc.). Some policies may even be *order optimal over time*:

Definition 15 (Order optimality over time). *A policy is said to be order optimal over time if:*

$$\exists C > 0 : \forall t \in \llbracket 2, \infty \llbracket, \mathbb{E}[R_t] \leq C \ln(t)$$

In the sequel, we introduce the main result of this chapter: we prove the order optimality over time of $MUCB$ policies.

From the expression of Equation 6.1 one can remark that it is sufficient to upper bound the expected number of times $\mathbb{E}[T_{n,t}]$ one plays a suboptimal arm n after t rounds, to obtain an upper bound on the expected cumulated regret. This leads to the following theorem.

Theorem 5 (Order optimality of $MUCB(\alpha)$ policies). *Let $\rho_n = \mu_n/\mu^*$, $n \in \llbracket 1, N \rrbracket \setminus \{n^*\}$. For all $N \geq 2$, if policy $MUCB(\alpha > 4)$ is run on N arms having rewards drawn from exponential distributions $\theta_1, \dots, \theta_N$ then:*

$$\mathbb{E}[R_t] \leq \sum_{n: \Delta_n > 0} \frac{4\mu^*\alpha}{1 - \rho_n} \ln(t) + o(\ln(t)) \quad (6.5)$$

Note that Theorem 5 upper bounds the regret of MUCB for $\alpha > 4$. As illustrated in Section 6.5, there might exist smaller values for α such that this upper bound still holds. Further investigations on this matter are needed to improve the provided upper bound.

Proving Theorem 5 relies on three lemmas that we analyze and prove in the next subsection. The lemma 1 provides a general bound for the regret regardless of the policy considered. The expression is function of two probabilities related to learning anomalies. These anomalies depend on the learning algorithm. They are introduced and analyzed. Then lemma 2 and 3 upper bound them.

6.4.2 Learning Anomalies and Consistency of MUCB policies

Let us introduce the set $\mathbb{S} = \mathbb{N} \times \mathbb{R}$; then, one can write $S_{n,t} = (T_{n,t}, B_{n,t}) \in \mathbb{S}$ the decision state of arm n at time t . We associate the product order to the set \mathbb{S} : for a pair of states $S = (T, B) \in \mathbb{S}$ and $S' = (T', B') \in \mathbb{S}$, we write $S \geq S'$ if and only if $T \geq T'$ and $B \geq B'$.

In order to analyze the behavior of these indexes within different problems, we focus on two types of *anomalies* where the indexes associated to the suboptimal arms are ‘too large’ or the index associated to the optimal arm is ‘too small’. In a nutshell, anomalies describe situations where the current indexes of the arms differ from their asymptotic values and lead to suboptimal arm pulls. We will show that the behavior of an index in terms of regret highly depends on the decreasing rate of these anomalies’ probabilities.

Definition 16 (Anomaly of type 1). *We assume that there exists at least one suboptimal arm, i.e., $\llbracket 1, N \rrbracket \setminus \{n^*\} \neq \emptyset$. We call anomaly of type 1, denoted by $\{\phi_1(u_n)\}_{n,t}^\pi$, for a suboptimal arm $n \in \llbracket 1, N \rrbracket \setminus \{n^*\}$, and with parameter $u_n \in \mathbb{N}$, the following event:*

$$\{\phi_1(u_n)\}_{n,t}^\pi = \{S_{n,t} \geq (u_n, \mu^*)\} .$$

This anomaly can be explained as follows. Assume that arm n has been already played at least u_n times ($T_{n,t} \geq u_n$) at round t . Then $\{\phi_1(u_n)\}_{n,t}^\pi$ describes the situation where the computed index $B_{n,t}(T_{n,t})$ is larger than the desired asymptotic value of the optimal arm’s index, μ^* . We will show that for a policy π , there is a specific lower bound on the integer u_n associated to each arm n that guaranties a seldom occurrence of the type 1 anomaly and, thus, will ensure a sound behavior of the policy in terms of regret. In this case, u_n can be interpreted as an upper bound on the minimum number of times that a arm n should be played using a policy π to ensure that the event $\{\phi_1(u_n)\}_{n,t}^\pi$ is rare.

Definition 17 (Anomaly of type 2). *We refer to as anomaly of type 2, denoted by $\{\phi_2\}_t^\pi$, associated to the optimal arm n^* , the following event:*

$$\{\phi_2\}_t^\pi = \{S_{n^*,t} < (\infty, \mu^*) \cap T_{n^*,t} \geq 1\} .$$

Unlike the anomaly of type 1, we are concerned in this case with the underestimation of the optimal arm’s index for all $T_{n^*,t} \geq 1$. We show, in the rest of this subsection, that the occurrence of this event can be made as rare as wished by an appropriate choice of the policy’s parameters values.

First we introduce the following lemma that provides a general bound of the regret as a function of the anomalies.

Lemma 1 (Expected cumulated regret. Proof in 6.7.2). *Given a policy π and a MAB problem, let $\mathbf{u} = [u_1, \dots, u_N]$ represent a set of integers, then the expected cumulated regret is upper bounded by:*

$$\mathbb{E}[R_t] \leq \sum_{n \neq n^*} \Delta_n u_n + \sum_{n \neq n^*} \Delta_n \mathbb{P}_t(u_n)$$

with,

$$\mathbb{P}_t(u_n) = \sum_{m=u_n+1}^t (\mathbb{P}(\{\phi_2\}_m^\pi) + \mathbb{P}(\{\phi_1(u_n)\}_{n,m}^\pi))$$

In the sequel, we consider the following values for the set \mathbf{u} , for all suboptimal arms n ,

$$u_n(t) = \left\lceil \frac{4\alpha}{(1-\rho_n)^2} \ln(t) \right\rceil$$

where for $n \neq n^*$, $\rho_n = \mu_n/\mu^*$.

We show in the two following lemmas that for the defined set \mathbf{u} the anomalies are upper bounded by exponentially decreasing functions of the number of iterations.

Lemma 2 (Upper bound of Anomaly 1. Proof in 6.7.3). *For all $N \geq 2$, if policy $MUCB(\alpha)$ is run on N arms having rewards drawn from exponential distributions $\theta_2, \dots, \theta_N$ then $\forall n \in \llbracket 1, N \rrbracket \setminus \{n^*\}$:*

$$\mathbb{P}(\{\phi_1(u_n)\}_{n,t}^\pi) \leq t^{-\alpha/2+1} \quad (6.6)$$

Lemma 3 (Upper bound of Anomaly 2. Proof in 6.7.4). *For all $N \geq 2$, if policy $MUCB(\alpha)$ is run on N arms having rewards drawn from exponential distributions $\theta_1, \dots, \theta_N$ then:*

$$\mathbb{P}(\{\phi_2\}_t^\pi) \leq t^{-\alpha/2+1} \quad (6.7)$$

We end this section proving Theorem 5.

Proof of Theorem 5. For $\alpha > 4$, relying on Lemmas 1, 2 and 3 we can write:

$$\mathbb{E}[R_t] \leq \sum_{n \neq n^*} \Delta_n \left\lceil \frac{4\alpha}{(1-\rho_n)^2} \ln(t) \right\rceil + o(\ln(t))$$

where Lemma 1, is simplified using the following equality for $\alpha > 4$

$$\sum_{n \neq n^*} \Delta_n \mathbb{P}_t(u_n) = o(\ln(t))$$

As a matter of fact, for $\alpha > 4$:

$$\sum_{u=u_n k}^t u^{-\alpha/2+1} \leq \sum_{u=u_n}^t \frac{1}{u^{1+\delta}} = o(\ln(t))$$

for any δ such that $\alpha = 4 + 2\delta$. Finally, since $\Delta_n = \mu^*(1-\rho_n)$ and $u_n(t) = \frac{4\alpha}{(1-\rho_n)^2} \ln(t) + o(\ln(t))$, we find the stated result in Theorem 5. \square

6.5 Simulation Results

This section presents and discusses simulation results that aim at illustrating and validating the stated upper bound provided in Theorem 5. We detail first the simulation protocol, then we discuss the simulation results.

For illustration purpose we consider a SU willing to evaluate the quality of $N = 10$ channels. The SU relies on the measure of the channels' SNR to evaluate the best channel. We assume that the SU suffers Rayleigh fading. Consequently, for every channel, the measured SNR follows an exponential distribution. The presented simulation consider the following parameters $\mu = \{\mu_1, \dots, \mu_{10}\}$ for the channels, where $\mu_1 \leq \dots \leq \mu_{10}$ without loss of generality: and $\mu = \{0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9; 1\}$.

The simulations compare three MUCB policies for α equal to respectively, $\{1; 2; 4.01\}$. These algorithms are referred to as $MUCB(1)$, $MUCB(2)$ and $MUCB(4)$ respectively. Notice that $MUCB(4)$ is chosen so as to respect the condition imposed in Theorem 5, i.e., $\alpha > 4$. $MUCB(1)$ and $MUCB(2)$ on the contrary are considered as possibly risky by Theorem 5 as already discussed in Section 6.4. The simulations consider a time horizon of 10^6 iterations.

Figure 6.3 plots the cumulated averaged regret of MUCB policies. In order to obtain relevant results, the curves were averaged over 100 experiments. All curves show a similar behavior: first an exploration phase where the regret grows quickly. Then the curves tend to confirm that the regret of MUCB policies grow as a logarithmic function of the number of iterations. As matter of fact, we notice that after the first exploration phase, on a logarithmic scale, the regret grows as a linear function. Moreover, since $MUCB(1)$ and $MUCB(2)$ seem to respect this trend, these curves suggest that the imposed condition in Theorem 5, $\alpha > 4$, might be improvable.

Figure 6.2 further illustrates the behavior of MUCB policies. A typical channel selection figure is plotted. Thus we can see for different exploration coefficients α the selected channels. As predicted by Theorem 5, suboptimal channels are selected regularly on a logarithmic scale depending on their quality.

6.6 Conclusion

A new low complexity algorithm for MAB problems is suggested and analyzed in this chapter: MUCB. The analysis of its regret proves that the algorithm is order optimality over time. In order to quantify its performance compared to optimal algorithms, further empirical evaluations are needed and are currently under investigation.

Moreover, combined with coordination mechanisms similar to those suggested in Chapter 5, MUCB algorithms could solve various problems: such as finding the shortest path in a network or selecting optimal sensors in a wireless networks. Thus, this contributions open the way to many new applications that go beyond the scope of this thesis.

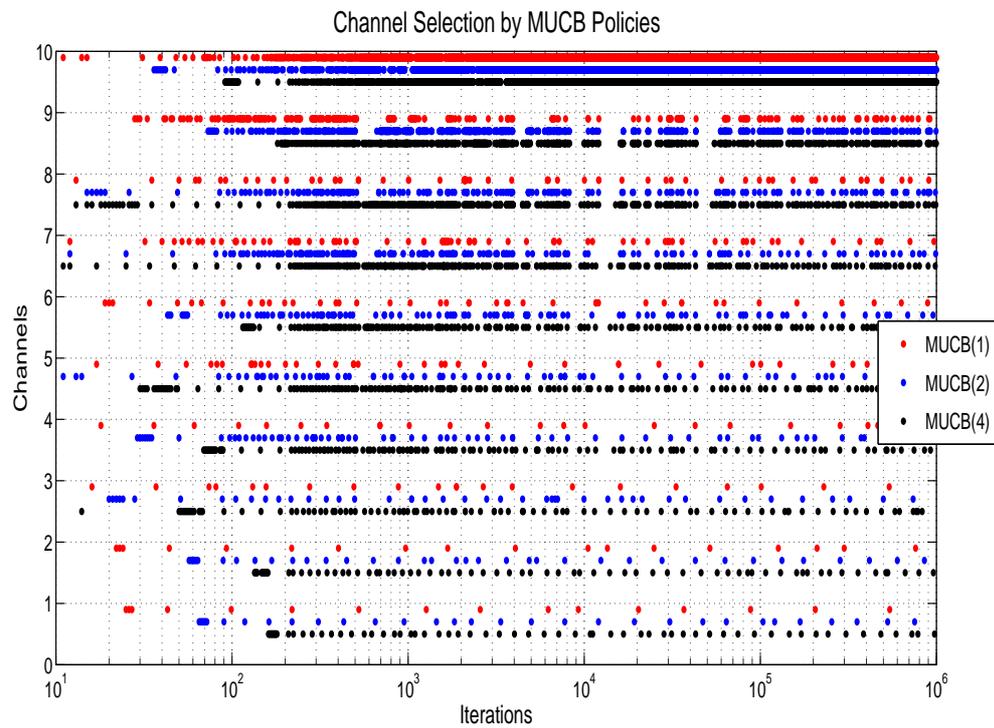


Figure 6.2: Channel selection process over time: a typical run. The dots between the values $[n; n + 1]$ (Y – axis) represents the time instants $t \in \mathbb{N}$ where channel n is selected by MUCB policies. This curve illustrates a typical run of MUCB policies. Thus we can see for different exploration coefficients α the selected channels. As predicted by Theorem 5, suboptimal channels are selected regularly on a logarithmic scale depending on their quality.

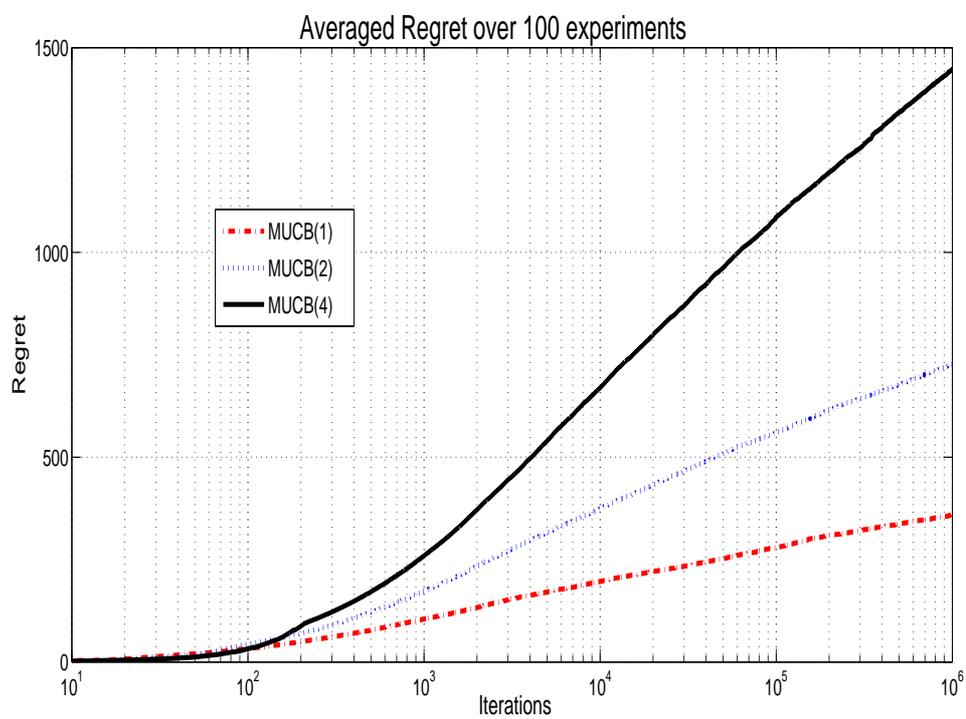


Figure 6.3: Average Regret Over 100 experiments: Illustration of Theorem 5. These curves confirm that the regret of MUCB policies grow as a logarithmic function of the number of iterations. Moreover, since $MUCB(1)$ and $MUCB(2)$ seem to respect this trend, these curves suggest that the imposed condition in Theorem 5, $\alpha > 4$, might be improvable.

6.7 Appendix

6.7.1 Large Deviations Inequalities

We introduce hereafter the *Cramer condition* for a random variable. Then, the *Cramer-Chernoff Theorem* allows to compute bounds on the probability that the sample mean of a random variable satisfying the Cramer condition deviates from its expected value.

Assumption 1 (Cramer condition). *Let X be a real random variable. X satisfies the Cramer condition if and only if:*

$$\exists \gamma > 0 : \forall \eta \in (0, \gamma), \mathbb{E} [e^{\eta X}] < \infty .$$

Lemma 4 (Cramer-Chernoff Lemma for the sample mean). *Let X_1, \dots, X_m ($m \in \mathbb{N}$) be a sequence of i.i.d. real random variables satisfying the Cramer condition with expected value $\mathbb{E}[X]$. We denote by \bar{X}_m the sample mean $\bar{X}_m = \frac{1}{m} \sum_{i=1}^m X_i$. Then, there exist two functions $l_1(\cdot)$ and $l_2(\cdot)$ such that:*

$$\forall \beta_1 > \mathbb{E}[X], \mathbb{P}(\bar{X}_m \geq \beta_1) \leq e^{-l_1(\beta_1)m} ,$$

$$\forall \beta_2 < \mathbb{E}[X], \mathbb{P}(\bar{X}_m \leq \beta_2) \leq e^{-l_2(\beta_2)m} .$$

Functions $l_1(\cdot)$ and $l_2(\cdot)$ do not depend on the sample size m and satisfy the following properties:

- $l_1(\beta_1)$ is a continuous non-negative, strictly increasing, non-constant function for all $\beta_1 > \mathbb{E}(X)$ and $l_1(\mathbb{E}(X)) = 0$.
- $l_2(\beta_2)$ is a continuous non-negative, strictly-decreasing, non-constant function for all $\beta_2 < \mathbb{E}(X)$ and $l_2(\mathbb{E}(X)) = 0$.

This result was initially proposed and proved in [142]. The bounds provided by this lemma are called *Large Deviations Inequalities* (LDIs) in this chapter.

In the case of exponential distributions this theorem can be applied and LDI functions have the following expressions:

$$l_1(\beta) = l_2(\beta) = \frac{\beta}{\mathbb{E}[X]} - 1 - \ln \left(\frac{\beta}{\mathbb{E}[X]} \right)$$

This function as well as the following well known inequality are very useful to prove the results of this chapter.

$$\frac{\beta}{\mathbb{E}[X]} - 1 - \ln \left(\frac{\beta}{\mathbb{E}[X]} \right) \geq \frac{3 \left(1 - \frac{\beta}{\mathbb{E}[X]} \right)^2}{2 \left(1 + 2 \frac{\beta}{\mathbb{E}[X]} \right)}$$

6.7.2 Proof of Lemma 1

First, according to Equation 6.1, recall that

$$\mathbb{E}[R_t^\pi] = \sum_{n \neq n^*} \Delta_n \mathbb{E}[T_{n,t}]$$

Since $\Delta_{n^*} = 0$ for the optimal arm n^* , we shall only consider suboptimal arms $n \in \llbracket 1, N \rrbracket \setminus \{n^*\}$ in the remainder of the proof. $T_{n,t}$ is the number of times one pulled arm n before time t , so $T_{n,t} = \sum_{m=0}^{t-1} \mathbb{1}_{a_m=n}$. Then,

$$\mathbb{E}[T_{n,t}] = \sum_{m=0}^{t-1} \mathbb{E}[\mathbb{1}_{a_m=n}]$$

After playing an arm u_n times, bounding the first u_n terms by 1 yields:

$$\mathbb{E}[T_{n,t}] \leq u_n + \sum_{m=u_n+1}^{t-1} \mathbb{P}(\{a_m = n\} \cap \{T_{n,m} > u_n\}) \quad (6.8)$$

Then we can notice that the following events are equivalent:

$$\{a_m = n\} = \left\{ B_{n,m} > \max_{n' \neq n} B_{n',m} \right\}$$

Moreover we can notice that:

$$\left\{ B_{n,m} > \max_{n' \neq n} B_{n',m} \right\} \subset \{B_{n,m} > B_{n^*,m}\}$$

Which can be further included in the following union of events:

$$\{B_{n,m} > B_{n^*,m}\} \subset \{B_{n,m} > \mu^*\} \cup \{\mu^* > B_{n^*,m}\}$$

Consequently we can write:

$$\{a_m = n\} \cap \{T_{n,m} > u_n\} \subset \{\Phi_1(u_n)\}_{n,m}^\pi \cup \{\Phi_2\}_m^\pi \quad (6.9)$$

Finally, we apply the probability operator:

$$\mathbb{E}[T_{n,t}] \leq u_n + \sum_{m=u_n+1}^{t-1} \mathbb{P}(\{\Phi_1(u_n)\}_{n,m}^\pi) + \mathbb{P}(\{\Phi_2\}_m^\pi). \quad (6.10)$$

The combination of Equation 6.1 - given at the beginning of this proof - and Equation 6.10 concludes this proof.

6.7.3 Proof of Lemma 2

From the definition of $\{\phi_1(u_n)\}_{n,t}^\pi$ we can write that :

$$\begin{aligned} \mathbb{P}(\{\phi_1(u_n)\}_{n,t}^\pi) &= \sum_{S_{n,t} \in \mathbb{S}} \mathbb{P}(S_{n,t} \geq (u_n, \mu^*)), \\ &= \sum_{u=u_n}^{t-1} \mathbb{P}(T_{n,t} = u) \mathbb{P}(B_{n,t}(T_{n,t}) \geq \mu^* | T_{n,t} = u), \\ &\leq \sum_{u=u_n}^{t-1} \mathbb{P}(B_{n,t}(u) \geq \mu^*). \end{aligned}$$

In the case of MUCB policies, we have:

$$\forall u \leq t, \mathbb{P}(B_{n,t}(u) \geq \mu^*) = \mathbb{P}\left(\overline{W}_{n,t}(u) \geq \frac{\mu^*}{M_{n,t}(u)}\right)$$

Consequently, we can upper bound the probability of occurrence of type 1 anomalies by:

$$\mathbb{P}(\{\phi_1(u_n)\}_{n,t}^\pi) \leq \sum_{u=u_n}^{t-1} \mathbb{P}\left(\overline{W}_{n,t}(u) \geq \frac{\mu^*}{M_{n,t}(u)}\right).$$

Let us define $\beta_{n,t}(T_{n,t}) = \frac{\mu^*}{M_{n,t}(T_{n,t})}$.

Since we are dealing with exponential distributions, the rewards provided by the arm n satisfy the Cramer condition. As a matter of fact, since $u \geq u_n \geq \alpha \frac{\ln(t)}{(1-\rho_n)^2}$ then:

$$\beta_{n,t}(u)\lambda_n = \rho_n^{-1} \left(1 - \sqrt{\alpha \frac{\ln(t)}{u}}\right) \geq 1$$

So, according to the large deviation inequality for $\overline{W}_{n,t}(T_{n,t})$ given by Lemma 4 (with $T_{n,t} \geq u_n$ and u_n large enough), there exists a continuous, non-decreasing, non-negative function $l_{1,k}$ such that:

$$\mathbb{P}(\overline{W}_{n,t}(T_{n,t}) \geq \beta_{n,t}(T_{n,t}) | T_{n,t} = u) \leq e^{-l_{1,n}(\beta_{n,t}(u))u}.$$

Finally:

$$\mathbb{P}(\{\phi_1(u_n)\}_{n,t}^\pi) \leq \sum_{u=u_n}^{t-1} e^{-l_{1,n}(\beta_{n,t}(u))u}. \quad (6.11)$$

The end of this proof aims at proving that for $u \geq u_n$:

$$l_{1,n}(\beta_{n,t}(u)) \geq \alpha \frac{\ln(t)}{2u}$$

Note that since we are dealing with exponential distributions we can write (C.f. Subsection 6.7.1):

$$l_{1,n}(\beta_{n,t}(u)) \geq \frac{3(1 - \beta_{n,t}(u)\lambda_n)^2}{2(1 + 2\beta_{n,t}(u)\lambda_n)}$$

Moreover since $u \geq u_n \geq \alpha \frac{\ln(t)}{(1-\rho_n)^2}$ then:

$$0 \leq 1 - \sqrt{\alpha \frac{\ln(t)}{u}} \leq 1$$

Thus,

$$\beta_{n,t}(u)\lambda_n = \rho_n^{-1} \left(1 - \sqrt{\alpha \frac{\ln(t)}{u}} \right) \leq \rho_n^{-1}$$

Consequently it is sufficient to prove that:

$$\frac{3(1 - \beta_{n,t}(u)\lambda_n)^2}{2(1 + 2\rho_n^{-1})} \geq \alpha \frac{\ln(t)}{2u}$$

Let us define $h(t)$ as a function of time: $h(t) = \sqrt{\alpha \frac{\ln(t)}{u}} \in [0, 1]$. We analyze the sign of the function:

$$g(t) = (\rho_n^{-1}h(t) - (\rho_n^{-1} - 1))^2 - \frac{(1 + 2\rho_n^{-1})}{3}h(t)^2 \quad (6.12)$$

Consequently we need to prove that for $u \geq u_n$, $g(\cdot)$ has positive values.

Factorizing last equation leads to the following to terms:

$$\left\{ \begin{array}{l} \left(\rho_n^{-1} - \sqrt{\frac{(1+2\rho_n^{-1})}{3}} \right) h(t) - (\rho_n^{-1} - 1) \\ \left(\rho_n^{-1} + \sqrt{\frac{(1+2\rho_n^{-1})}{3}} \right) h(t) - (\rho_n^{-1} - 1) \end{array} \right. \quad (6.13)$$

Since per definition:

$$\left\{ \begin{array}{l} h(t) \in [0, 1] \\ \rho_n^{-1} \geq 1 \end{array} \right. \quad (6.14)$$

Then,

$$\left(\rho_n^{-1} - \sqrt{\frac{(1 + 2\rho_n^{-1})}{3}} \right) h(t) - (\rho_n^{-1} - 1) \leq 0$$

Consequently, $g(\cdot)$ is positive only if the second term of Equation 6.13 is negative, i.e.,

$$\sqrt{\alpha \frac{\ln(t)}{u}} \leq \frac{(\rho_n^{-1} - 1)}{\left(\rho_n^{-1} + \sqrt{\frac{(1+2\rho_n^{-1})}{3}} \right)}$$

Since $u \geq u_n$, the last inequation is verified.

We conclude this proof by upper bounding Equation 6.11 for $u \geq u_n$:

$$\mathbb{P}(\{\phi_1(u_n)\}_{n,t}^\pi) \leq \sum_{u=u_n}^{t-1} e^{-\alpha \ln(u)/2} \leq \sum_{u=u_n}^{t-1} \frac{1}{u^{\alpha/2}} \leq \frac{1}{t^{\alpha/2-1}}$$

6.7.4 Proof of Lemma 3

This proof follows the same steps as the the proof in Subsection 6.7.3. From the definition of $\{\phi_1(u)\}_{n,t}^\pi$ we can write that :

$$\mathbb{P}(\{\phi_2\}_t^\pi) \leq \sum_{u=1}^{t-1} \mathbb{P}(B_{n^*,t}(u) \leq \mu^*).$$

In the case of MUCB policies, we have:

$$\forall u \leq t, \mathbb{P}(B_{n^*,t}(u) \leq \mu^*) = \mathbb{P}\left(\overline{W}_{n^*,t}(u) \leq \frac{\mu^*}{M_{n^*,t}(u)}\right)$$

Consequently, we can upper bound the probability of occurrence of type 2 anomalies by:

$$\mathbb{P}(\{\phi_2\}_t^\pi) \leq \sum_{u=1}^{t-1} \mathbb{P}\left(\frac{\overline{W}_{n^*,t}(u)}{\mu^*} \leq \max\left\{0; \left(1 - \sqrt{\frac{\alpha \ln(t)}{T_{n,t}}}\right)\right\}\right)$$

Since $\mu^* \max\left\{0; \left(1 - \sqrt{\frac{\alpha \ln(t)}{T_{n,t}}}\right)\right\} \leq \mu^*$ Cramer's condition is verified. Moreover since the arm is played when the maximal of the previous term is equal to 0, we can consider that $u \geq \alpha \ln(t)$ and that:

$$\mu^* \max\left\{0; \left(1 - \sqrt{\frac{\alpha \ln(t)}{T_{n,t}}}\right)\right\} = \mu^* \left(1 - \sqrt{\frac{\alpha \ln(t)}{T_{n,t}}}\right)$$

Consequently, we can upper-bound the occurrence of Anomaly 2:

$$\mathbb{P}(\{\phi_2\}_t^\pi) \leq \sum_{u=\alpha \ln(t)}^{t-1} e^{-l_2(\beta_{n^*,t}(u))u} \tag{6.15}$$

Where, $l_2(\beta_{n^*,t}(u))$ verifies the LDI as defined in Appendix 6.7.1. Thus, after mild simplifications we can write,

$$l_2(\beta_{n^*,t}(u)) \geq \frac{\frac{3\alpha \ln(t)}{u}}{2\left(1 + 2\left(1 - \sqrt{\frac{\alpha \ln(t)}{u}}\right)\right)} \geq \frac{\alpha \ln(t)}{2u}$$

Consequently, including this last inequality into Equation 6.15 ends the proof.

Chapter 7

Overview, General Conclusions and Future Work

Contents

6.1	Introduction	121
6.2	Multi-Armed Bandits	122
6.3	Multiplicative upper confidence bound algorithms	123
6.4	Analysis of $MUCB(\alpha)$ policies	124
6.4.1	Consistency and order optimality of MUCB indexes	125
6.4.2	Learning Anomalies and Consistency of MUCB policies	126
6.5	Simulation Results	128
6.6	Conclusion	128
6.7	Appendix	131
6.7.1	Large Deviations Inequalities	131
6.7.2	Proof of Lemma 1	132
6.7.3	Proof of Lemma 2	133
6.7.4	Proof of Lemma 3	135

7.1 Conclusion and Overview

In this report we tackled several fundamental decision making and learning problems related to CR, in general, and to OSA in particular. These problems appeared naturally with the advent of CR.

Thus, after a first general introduction in Chapter 1, we positioned the general class of problems we are dealing with in Chapter 2. Moreover, this latter relates our approach, based on partial monitoring under uncertainty, to the CR literature. Assuming minimum information on CR equipment's environments, we ventured the analysis of sequential decision making techniques to enable the design of efficient and reliable cognitive agents.

Two main topics were tackled: signal detection under uncertainty and resource exploitation under uncertainty:

- On the one hand, introduced in Chapter 3, we revisit the performance of energy detection under noise uncertainty. Relying on the seminal work of Alexander Sennschein and Philip M. Fishman in 1992 [49], we proved new results on the limits of detection under noise level uncertainty. More specifically, we showed that under a log-normal distributed noise uncertainty, close-formed expression of the ED's performances can be written. Such performances appear to depend on the length of the sampling window as well as the noise uncertainty parameter introduced in the chapter.
- On the other hand, in Chapters 4, 5 and 6, we venture the analysis of OSA related models. We modeled OSA scenarios as Multi-Armed Bandit problems. Through these chapters we complexified the scenarios taking into account OSA specificities: detection errors, sensing uncertainties, coordination and collaboration among players and fading channels. To solve the previously mentioned problems, we focus on the UCB_1 proving its efficiency at every step. In the case of fading channels, we introduced a different form of algorithms to handle exponentially distributed observations.

Note that, the results of Chapter 6 offer a new contribution to the MAB literature and open the way to a large set of machine learning applications.

7.2 Perspectives and Future Work

The presented work in this dissertation answers many questions asked in Chapter 1, Section 1.4. The provided answers open the way to many new studies and potential applications.

- Chapter 3 shows that it possible to quantify, avoiding worst case analyzes, the uncertainty on the noise level. The results assume that the uncertainty level is known. We can see from the equations that if the window size is large enough, the distribution of log-energy ratio depends only on the uncertainty parameter. This open the way to estimating, perhaps even alleviating, energy detection limits for CR applications.
- The relationship between the uncertainty model suggested by Tandra et al. in [50] and our uncertainty model is not quite clear yet. It is currently under investigation. The aim is to provide a unified uncertainty framework on energy detection limits under noise level uncertainty.

- Explaining the benefit of coordination and collaboration in Secondary Networks, Chapter 5 open the way the real applications. The protocol involved are yet to be clearly investigated. Hopefully, this work sill help designing real and efficient secondary networks in the next few years.
- Combining MUCB algorithms introduced in Chapter 6 and coordination and collaboration mechanisms discussed in Chapter 5, it becomes possible to tackle complex sensor networks or network applications. For instance, we can think of problems involving: shortest paths, quality chains and sensing networks in fading environments to name a few.

These topics are currently under investigation at Supélec in the team SCEE.

Appendix

Appendix A

Three years of Ph.D. research, teachings and talks

This Ph.D. took place in Supélec on the Campus of Rennes. Within the team SCÉE, we dealt with signal detection under uncertainty and decision making for CR. During the period between October 2008 and December 2011, my time was split between my Ph.D. research, computer science teachings at the University of Rennes1 and the Medical School of Rennes1, and finally undergraduates' project supervision. In a nutshell:

- My teaching activities occupied 224 hours during 3 years. The count does not include the preparation time.
- I followed 99 hours of various courses (scientific and management related courses).
- I supervised 9 undergraduate projects related to signal processing, game theory and machine learning.
- I published 2 journal papers and co-authored 15 conference papers (10 papers as first author, 5 as second author). Among these papers, 3 were 'invited'. I participated to most of the related conventions where I presented my work. Finally, I participated to 2 technical reports within the project NEWCOM++.
- I held 4 main Seminar Talks: At the university of Liège (Belgium) invited by Damien ERNST. At INRIA-Lille Nord (France) with the team of Rémi MUNOS. At the UPC (Barcelona, Spain) within the department of Signal Processing and Communication with Miguel LOPEZ-BENITEZ and Jordi PEREZ-ROMERO. Finally at Supélec (to greet Prof. MITOLA visiting the campus of Rennes).
- I wrote a chapter in the book on SDR and CR entitled 'De la Radio Logicielle à la Radio Intelligente' supervised by J. PALICOT. Moreover I supervised its English translation: 'from Software to Cognitive Radio'.

During my Ph.D., I had the pleasure to spend some time with many different research teams in various European universities:

- Several weeks at the University of Liège in Belgium, in the Machine Learning Department (with Damien ERNST).

- Almost three months at the Universitat Politècnica de Catalunya (UPC), in the department of Signal Processing and Communication (with Miguel LOPEZ-BENITEZ and Jordi PEREZ-ROMERO). This exchange was organized within the European project NEWCOM++.
- I discussed network related issues at the Chair Alcatel in Supélec, held by Mérouane DEBBAH, (on the Campus of Gif) where I spent several weeks near Paris.

Finally, we worked on various research and implementation topics that are not presented in this report. The work involved the design of algorithms and their implementations on the USRP platforms [25, 26], interfacing HDCRAM with Simulink and USRP cards [27], blind energy detection relying on Expectation-Maximization algorithms [28], as well as the exploration of hot-spot migration techniques on FPGA platforms in the context of Green CR [29].

Appendix B

Publications

B.1 Journal Papers

1. W. Jouini.
Energy Detection Limits under Log-Normal Approximated Noise Uncertainty,
Signal Processing Letters, Volume 18, Issue 7, Pages: 423-426, July 2011.
2. W. Jouini, C. Moy, J. Palicot.
Decision making for cognitive radio equipment: analysis of the first 10 years of exploration,
EURASIP Journal on Wireless Communications and Networking 2012, 2012:26.

B.2 Book Chapter(s)

1. J. Palicot (Surprised by)
De la radio logicielle à la radio intelligente, Collection Télécom,
Lavoisier Librairie, juin 2010; ISBN : 978-2-7462-2598-5

B.3 Invited Papers

1. W. Jouini, D. Ernst, C. Moy et J. Palicot.
Multi-armed bandit based decision making for cognitive radio,
28th Benelux Meeting on Systems and Control, Spa, Belgium, March 2009. (Extended Abstract)
2. C. Moy, W. Jouini et N. Michael.
Cognitive Radio Equipments Supporting Spectrum Agility,
3rd International Workshop on Cognitive Radio and Advanced Spectrum Management, Rome, Italy, November 2010.
3. W. Jouini, R. Bollenbach, M. Guillet, C. Moy et A. Nafkha.
Reinforcement learning application scenario for Opportunistic Spectrum Access,
54th IEEE International Midwest Symposium on Circuits & Systems (MWSCAS 2011), August 7-10 2011, Seoul, South Korea

B.4 International Conference papers (with peer-reviews)

1. W. Jouini, D. Ernst, C. Moy et J. Palicot.
Multi-Armed Bandit Based Policies for Cognitive Radio's Decision Making Issues,
3rd conference on Signal Circuits and Systems, Jerba, Tunisia, November 2009.
2. W. Jouini, D. Ernst, C. Moy et J. Palicot.
Upper confidence bound based decision making strategies and dynamic spectrum access,
International Conference on Communications (ICC), Cape Town, South Africa, May 2010.
3. W. Jouini, C. Moy et J. Palicot.
On decision making for dynamic configuration adaptation problem in cognitive radio equipments: a multi-armed bandit based approach,
6th Karlsruhe Workshop on Software Radios, WSR'10, Karlsruhe, Germany, March 2010.
4. H. Wang, W. Jouini , L. S. Cardoso, R Hachemani, J. Palicot, M. Debbah.
Blind bandwidth shape recognition for standard identification using USRP platforms and SDR4all tools,
Sixth Advanced International Conference on Telecommunications (AICT2010), may 9 2010 Barcelona, Spain
5. H. Wang, W. Jouini , L. S. Cardoso, R. Hachemani, J. Palicot, M. Debbah.
Blind Standard Identification with Bandwidth Shape and GI Recognition using USRP platforms and SDR4all Tools,
5th International Conference on Cognitive Radio Oriented Wireless Networks and Communications, CrownCom'10, Cannes, France, June 9-11, 2010.
6. W. Jouini, A. Nafkha, M. Lopez-Benitez et J. Pérez-Romero.
Joint Learning-Detection Framework: an Empirical Analysis,
Proceedings of the Joint COST2100 & IC0902 Workshop on Cognitive Radio and Networking, Bologna, Italy, November 23, 2010, Pages: 1-6.
7. X. Zhang, W. Jouini, P. Leray and J. Palicot.
Heat Emission Issues in Power Efficient Design for FPGA based Cognitive Radio,
International Conference on Green Computing and Communications (GreenCom2010), Hangzhou, China, December 2010.
8. S. Lecomte, W. Jouini, C. Moy, P. Leray.
A SystemC Radio-in-the-Loop Modeling for Cognitive Radio Equipments,
Software Defined Radio Forum 2010, 30 Nov-3 Dec 2010, Washington DC, USA.
9. W. Jouini, C. Moy et J Palicot.
Upper Confidence Bound Algorithm for Opportunistic Spectrum Access with sensing errors,
6th International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications, Osaka, Japan, 2011.

10. W. Jouini, D. Le Guennec, C. Moy and J. Palicot.
Log-Normal Approximation of Chi-square Distributions for Signal Processing,
XXX URSI General Assembly and Scientific Symposium of International Union of Radio Science, URSI 2011, August 2011.
11. W. Jouini and C Moy.
Channel Selection with Rayleigh Fading: A multi-Armed Bandit Framework,
The 13th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC, 2012)

B.5 National Conference paper(s) (with peer-reviews)

1. W. Jouini, C. Moy and J. Palicot.
Apprentissage pour l'Accès Opportuniste au Spectre : Prise en Compte des Erreurs d'Observation,
XXIII Colloque GRETSI (Groupement de Recherche et d'Etude du Traitement du Signal et de l'Image, Bordeaux, Septembre 2011.

B.6 Technical Reports

1. V. Corvino, M. Moretti, S. M. Perlaza, M. Debbah, S. Lasaulce, W. Jouini, J. Palicot, C. Moy, A. Serrador, H. Bogucka, P. Sroka, E. B. Rodrigues, M. López-Benítez, A. Umbert, F. Casadevall, J. Pérez-Romero.
Definitions and evaluation of JRRM and ASM algorithms,
NEWCOM++ Network of Excellence, Deliverable DR9.2, January 12, 2010
2. A. Serrador, L. Caeiro, L. M. Correia, M. Moretti, E. Bezerra, P. Sroka, H. Bogucka, M. López-Benítez, A. Umbert, F. Casadevall, W. Jouini, J. Palicot, C. Moy, A. Kliks, M. Debbah.
Final report of the JRRM and ASM activities,
NEWCOM++ Network of Excellence, Deliverable DR9.3, February 1, 2011

Part II

Résumé en Français

Appendix C

Vingt années de communications sans-fil: vers la radio intelligente

Contents

B.1	Journal Papers	145
B.2	Book Chapter(s)	145
B.3	Invited Papers	145
B.4	International Conference papers (with peer-reviews)	146
B.5	National Conference paper(s) (with peer-reviews)	147
B.6	Technical Reports	147

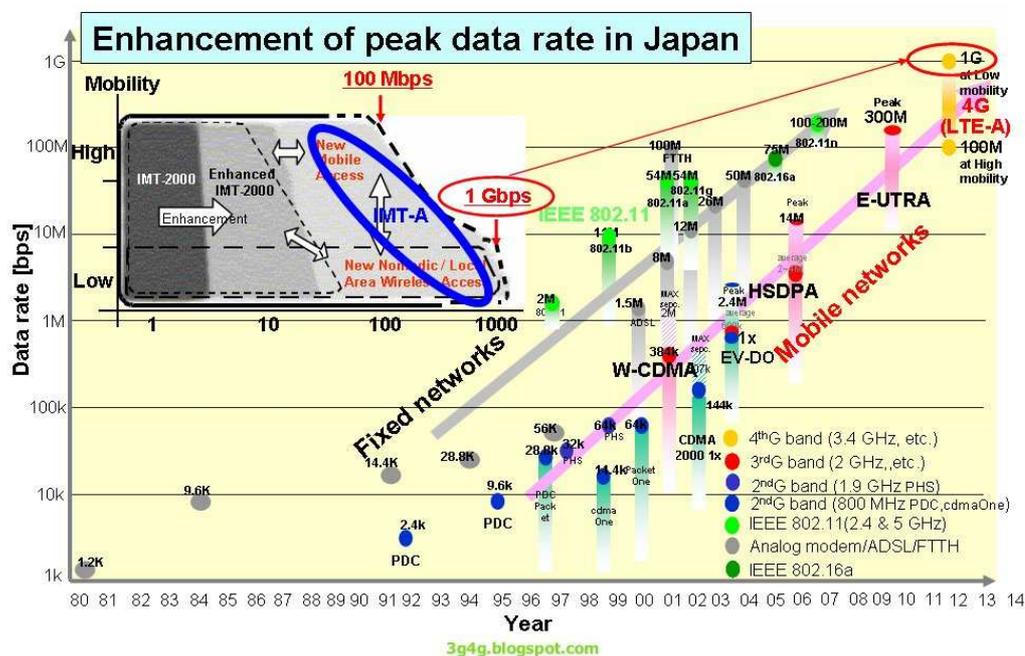


Figure C.1: Evolution des débits des principaux standards de communication sans-fil (Figure trouvé en ligne : http://3g4g.blogspot.fr/2008_04_01_archive.html). Des courbes similaires peuvent être trouvées dans le papier [2].

C.1 Vingt années de communication sans-fil

C.1.1 Les limites de la couche physique et la loi de Cooper

1991-2011 : vingt années d'innovation dans le domaine des communications numériques. Vingt années durant lesquelles les débits proposés par les technologies cellulaires et WLAN ne cessèrent de croître soutenus par une augmentation régulière de la capacité de calcul des processeurs (vérifiant les extrapolations de Gordon E. Moore en 1965). En effet, avec des outils de calcul plus rapides et performants, la communauté radio fut capable de concevoir et implanter des moyens de traitement du signal plus complexes au niveau de la couche physique.

Ainsi illustré par la figure C.1 les débits atteints par les technologies sans-fil, dus notamment aux innovations au niveau de la couche physique, sont en effet substantiels. Cela ne représente néanmoins qu'une petite fraction de l'augmentation de la capacité cumulée effective de communication ! En effet, Martin Cooper déclara récemment que la capacité de communication sans-fil avait doublé tous les 30 mois depuis 104 ans [4]. Cette progression régulière et exponentielle repose sur de diverses améliorations technologiques. Nous pouvons compter parmi celles-ci trois citées par Martin Cooper pour soutenir son discours :

- L'amélioration de la gestion du spectre (e.g., codage et modulation), ainsi que la gestion de l'accès à cette ressource, permet l'augmentation de la capacité cumulée effective de 25 fois.
- La gestion de bandes de fréquences plus large permet aussi la multiplication de la capacité cumulée effective d'un facteur 25 fois.
- enfin, la réduction de la taille des cellules (ou rayon) de transmission, permettant d'améliorer la réutilisation des fréquences, mène à une amélioration d'un facteur 1600.

Ainsi nous observons que toutes contributions confondues associées à la couche physique ne permirent d'obtenir 'qu'une amélioration d'un facteur 625'. Alors que la réduction de la taille des cellules permet, au dépend d'un coût important en infrastructure, d'apporter un gain substantiel à la capacité cumulée effective de communication. Ce gain augmente aujourd'hui rapidement grâce à l'implantation massive de terminaux Wi-Fi dans tous les bâtiments et dans les espaces urbains en général. Notons que la prise en compte des réseaux Wi-Fi -ou leurs contributions le cas échéant- dans le calcul effectué par Martin Cooper n'est pas explicite. Ces conclusions restent donc sujettes à des précisions supplémentaires.

De plus, la recherche dans le domaine de la couche physique semble aujourd'hui s'essouffler. Ainsi, en considérant une simple communication entre deux utilisateurs, aucun codage ni aucune forme de modulation ne permettra d'apporter un réel gain. En effet, les moyens de communication actuels permettent déjà d'atteindre des débits proches de ceux annoncés par Shannon. La question se pose alors : est-il encore intéressant de poursuivre les recherches sur la couche physique ? La réponse est bien entendu oui : les nouvelles opportunités de communication ne viendront probablement pas en ne se focalisant que sur la capacité de Shannon mais plutôt sur la capacité équivalente du réseau. Dans ce contexte apparaissent alors de nouvelles pistes prometteuses qui ont pour objectif d'augmenter les opportunités de communications des utilisateurs. Pour cela chaque utilisateur aura pour mission de répondre à ses besoins tout en évitant de 'polluer' son environnement. C'est dans cet esprit qu'apparaissent des techniques telles que le 'beam-forming' ou encore les 'smart-antennas'.

Enfin, optimiser les performances d'un utilisateur, en prenant en compte le fonctionnement du réseau, peut bien entendu se faire à travers les différentes couches du système : de la couche physique à la couche applicative. Néanmoins, augmenter les degrés de liberté d'un équipement afin d'optimiser son comportement présuppose d'une part une grande flexibilité, du point de vue de l'électronique et du traitement du signal, et d'autre part une grande autonomie dans la prise de décision. En d'autres termes, le système ainsi conçu observe son environnement, prend une décision appropriée vis-à-vis des changements de l'environnement et/ou les objectifs de l'utilisateur, et enfin, reconfigurer en temps réel l'architecture de l'équipement afin de s'adapter. Par conséquent, une approche prometteuse, dont l'objectif est de soutenir la loi de Cooper, consiste à combiner flexibilité électronique et intelligence computationnelle aussi bien dans les futurs équipements radio que dans les réseaux. Ces besoins, anticipés depuis une vingtaine d'années ont donné lieu à deux domaines très actifs : la radio logicielle (Software Radio) et à la radio intelligente (Cognitive Radio).

C.2 Vers la radio intelligente

C.2.1 La radio logicielle

Les nombreux progrès réalisés dans les domaines de l'électronique ont permis de concevoir des plateformes à la fois plus rapides et plus flexibles offrant, de nouvelles perspectives en termes de nouveaux moyens de communications pour des applications non militaires⁽¹⁾. En effet, en 1991, Joseph Mitola III présentait déjà la possibilité de concevoir, au moins en théorie, des systèmes de communication reposant sur une exécution de logiciels. Cela représente donc un changement de paradigme important, remplaçant ainsi l'exécution matérielle habituelle par des traitements logiciels autant que possible. Mitola nomma ce paradigme 'Software Defined Radio', en d'autres termes radio logicielle.

A titre illustratif, les systèmes radio conçus aujourd'hui nécessitent une chaîne électronique dédiée pour tout standard inclus dans l'équipement. Utiliser un standard plutôt qu'un autre revient à éteindre la première chaîne pour activer la seconde. Avec l'augmentation du nombre de standards intégrés (GSM, EDGE, UMTS, Wi-Fi, Bluetooth, etc.) dans un même équipement, la conception de tels systèmes devient particulièrement complexe. Le besoin pour une plus grande flexibilité dans l'équipement s'impose petit à petit en tant que nécessité. Il deviendrait alors possible, au moins en théorie, de reconfigurer l'équipement à la volée en n'implantant que les chaînes utiles à chaque instant, et par extension, uniquement les opérateurs de traitement du signal nécessaires à chaque instant (en fonction des conditions de transmission). Dans la pratique, le problème est encore un sujet de recherche en cours de résolution.

Proposer une définition de la radio logicielle qui soit sans ambiguïté et acceptée de tous semble aujourd'hui compliqué. Plusieurs définitions ont été proposées et restent aujourd'hui encore sujets à de nombreuses discussions au sein de la communauté radio. Pour des raisons de clarté, nous décrivons brièvement quelques notions liées à la radio logicielle tel que définies par le SDR Forum [7] (d'autres définitions alternatives existent [8]). Les définitions exactes traduites par l'auteur ici peuvent être retrouvées dans le chapitre 1.

Ainsi, le SDR Forum définit la radio logicielle en tant qu'*équipement radio dans lequel, au moins quelques fonctions de la couche physique sont traitées de manière logicielle*. Par conséquent, au moins du point de vue du SDR Forum, les radios dites logicielles ne sont définies que par la manière dont sont implantés les blocs de traitement au niveau de la couche physique. La radio logicielle apparaît donc comme une simple évolution des systèmes paramétrables déjà existants. Or, avec cette nouvelle couche logicielle, il devient possible de contrôler un large jeu de paramètres afin que l'équipement puisse s'adapter à son environnement (largeur de bande, modulation, codage, niveau de puissance, ainsi de suite). Cela n'est néanmoins possible que si des critères relatifs aux objectifs à atteindre sont définis. Ces critères prennent en compte les besoins de l'utilisateur, les degrés de libertés de la plateforme radio ainsi que les lois de régulations en vigueur. Introduire les moyens d'optimiser, de manière autonome, un équipement et/ou un réseau radio est la base de la radio intelligente, terme suggéré aussi par Joseph Mitola III [9, 10].

⁽¹⁾Ces recherches avaient commencé dans les années 70 pour des applications militaires aux Etats-Unis

C.2.2 Radio intelligente

Mitola définit la radio intelligente dans son manuscrit de thèse [10] de la manière suivante :

The term cognitive radio identifies the point at which wireless PDA and the related networks are sufficiently computationally intelligent about radio resources and related computer to computer communication to:

1. *Detect user communication needs as a function of use context, and*
2. *Provide radio resources and wireless services most appropriate to these needs.*

En d'autres termes, le concept de la radio intelligente présuppose que les équipements radios ainsi que les réseaux auxquels ils sont reliés ont une connaissance suffisante de leurs propres ressources et les moyens de communication inter-machines afin de :

- détecter les besoins des utilisateurs en fonction du contexte et,
- fournir les ressources radio et les services sans fil les plus appropriés pour répondre à ces besoins.

Ainsi, ce concept vise à répondre de manière autonome aux besoins des utilisateurs, i.e. maximiser leurs 'gains', en termes de qualité de service, de débit ou d'efficacité énergétique par exemple, sans corrompre le bon fonctionnement du réseau. Par conséquent, l'intelligence nécessaire doit être distribuée aussi bien au niveau des équipements que dans le réseau. Afin de répondre à ces exigences, J. Mitola et J.Q. Maguire introduisirent la notion de 'cycle cognitif'. Le cycle est décrit dans la figure C.2 [9, 10]. Il suppose l'aptitude à collecter de l'information sur son environnement (perception/observation), à digérer ces informations (apprentissage et prise de décision) et à agir de la manière la plus appropriée en considérant les diverses contraintes imposées à l'équipement ainsi que les informations disponibles. La reconfiguration de l'équipement n'est pas détaillée, néanmoins il est généralement admis que la technologie radio logicielle est nécessaire au moins partiellement pour atteindre les objectifs de la radio intelligente [8].

Illustré par la figure C.2, un cycle cognitif complet ⁽²⁾ fait appel à chaque itération à cinq étapes : Observer, Orienter, Planifier, décider et Agir.

L'observation s'intéresse aussi bien aux métriques internes à l'équipement qu'aux métriques externes. Elle vise à s'informer sur certaines métriques nécessaires au raisonnement de l'équipement. On peut citer, par exemple, l'état du canal de transmission, le niveau d'interférence, ou encore le niveau de la batterie. Ces informations sont ensuite envoyées au centre décisionnel qui poursuit en orientant la stratégie de l'équipement, en planifiant les différentes tâches à réaliser en décidant de la démarche à suivre afin d'atteindre les objectifs fixés. Enfin, l'équipement prend en considération les décisions prises en se reconfigurant. Cette dernière étape agit sur l'équipement (paramètres internes liées aux chaînes de transmissions par exemple) ainsi que sur l'environnement externe. Une dernière étape est nécessaire afin de compléter le cycle cognitif : l'apprentissage. Cette étape prend en considération les résultats des décisions passées afin d'améliorer le comportement futur de l'équipement.

⁽²⁾On l'appelle cycle cognitif complet afin de le différencier d'autres cycles proposées dans la littérature et qui apparaissent comme des formes simplifiées de celui-ci.

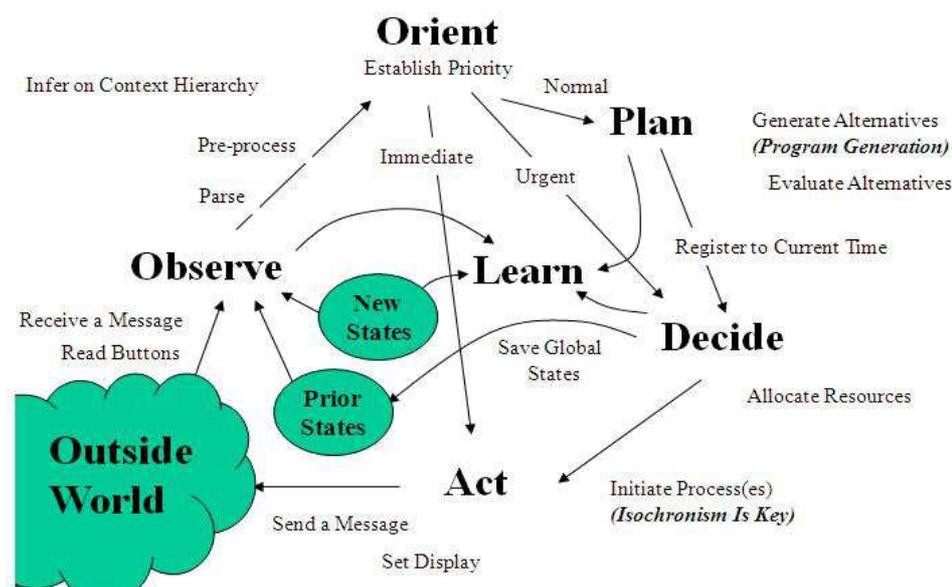


Figure C.2: Cycle cognitif présenter par J.Mitola III [10].

Depuis les travaux de J.Mitola, il y a une dizaine d'années, une véritable communauté de recherche sur les questions liées à la radio logicielle et la radio intelligente s'est créée. Les travaux de cette thèse s'inscrivent dans cette dynamique en focalisant sur les aspects décisions et apprentissage.

C.3 Travaux de Thèse

Les objectifs de la thèse ont été définis en deux temps. Dans un premier temps, trois objectifs ont été fixés :

1. Exploration des outils d'apprentissage et de prise de décisions exploités dans le cadre de la radio intelligente.
2. Identification, dans la littérature *Machine Learning*, des techniques de faibles complexités capables d'opérer sans information *a priori* sur leurs environnements.
3. Adapter les algorithmes ainsi identifiés au contexte de la radio intelligente. Notamment comprendre l'impact des erreurs d'observation sur les performances des algorithmes d'apprentissage.

Répondre à ces problématiques permet de relever de nouvelles interrogations :

1. Quantifier l'incertitude liée au détecteur d'énergie. Est-ce Possible ? Si oui, sous quelles hypothèses ? Cette question est cruciale afin de dimensionner correctement l'algorithme d'apprentissage. En effet ce dernier reposera ces décisions sur les observations du détecteur de signaux.

2. Intégrer les notions de coordination et de collaboration parmi plusieurs moteurs de prise de décisions. Notamment lorsqu'on analyse un réseau de communication secondaire.
3. Adapter, ou développer des algorithmes d'apprentissage, afin de répondre à des contextes d'exploitation de ressources spectrales lorsque l'observation de ces dernières est sujette à du *fading*.

Ce rapport de thèse a pour objectif d'apporter des éléments de réponses à toutes ces questions qui combinent des problèmes mathématiques nouveaux appliqués à des problématiques de télécommunications. Ce résumé en Français n'apporte qu'une introduction rapide aux notions fondamentales impliquées dans ces travaux. Nous invitons le lecteur intéressé à lire la Section 1.4 en anglais, au début de ce manuscrit, pour de plus amples détails. Le reste de ce résumé, à travers les Sections D et E, introduit d'une part les notions importantes relatives à la détection d'énergie dans des contextes d'incertitude, et d'autre part, les notions fondamentales liées à l'accès opportuniste au spectre combinée à la théorie des bandits manchots (Multi-Armed Bandit Paradigm).

Appendix D

Limites de la détection d'énergie dues à une connaissance incertaine du niveau du bruit

Contents

C.1	Vingt années de communication sans-fil	152
C.1.1	Les limites de la couche physique et la loi de Cooper	152
C.2	Vers la radio intelligente	154
C.2.1	La radio logicielle	154
C.2.2	Radio intelligente	155
C.3	Travaux de Thèse	156

D.1 Introduction

En général, les modèles d'apprentissage supposent que les mesures traitées sont exactes. Ces hypothèses ne correspondent souvent pas à la réalité. En effet, dans le cas de l'accès au spectre, l'issue de la phase d'observation (i.e., de détection) est souvent entachée d'erreurs. Ainsi le détecteur peut observer un canal libre alors que celui-ci est occupé et vice-versa. Connaître les limites des systèmes d'observation est donc une étape cruciale afin de dimensionner de manière pertinente les algorithmes d'apprentissage adjacents. Nous nous proposons donc dans le contexte de la radio intelligente (Cognitive Radio en anglais, CR) d'évaluer les limites d'un des algorithmes les plus populaires du domaine : le détecteur d'énergie. En effet, le détecteur d'énergie est un élément crucial en CR puisque ce dernier a une très faible complexité et permet, sans connaître la structure des signaux analysés, de détecter les bandes libres dans le spectre, ce qui représente l'un des scénarios CR les plus courants. Les contributions de ce chapitre sont les suivantes :

- **Approximation d'une distribution χ^2 par une loi Log-Normal.** Le résultat montre que l'approximation par une loi Log-Normal mène à des erreurs d'approximations plus faible que dans le cas d'une approximation d'une distribution χ^2 par une loi normale.
- En général, le niveau du bruit est connu via une estimation. Cette estimation semble être bien approximée en pratique par des lois Log-Normal modélisant l'incertitude du bruit [49]. Le détecteur se ramène donc à la comparaison de l'énergie du paquet analysé à l'estimation courante du niveau du bruit. En pratique cela se caractérise par l'analyse du ratio entre l'énergie du paquet courant et l'estimation du bruit. Malheureusement la statistique ainsi définie ne connaît pas de loi simple et explicite. **Nous proposons alors d'exploiter l'approximation de la loi χ^2 par une loi Log-Normal afin d'approcher le problème initial par un problème plus simple.**
- Les études proposées dans la littérature ne permettent pas d'évaluer les performances d'un détecteur d'énergie dans le cas où une incertitude existe sur la valeur du niveau du bruit. En effet elles considèrent que le bruit est connu de manière certaine dans un certain intervalle d'incertitude. Cela mène alors à des études du pire cas, i.e., évaluer les performances du détecteur en fixant une borne supérieure de la probabilité de fausse alarme. Cela mène naturellement à définir une borne inférieure de la probabilité de détection. Les études ainsi menées [49, 50] aboutissent à la définition d'une valeur limite de rapport signal à bruit en deçà duquel les résultats du détecteur ne sont plus pertinents. Ce résultat est souvent interprété comme la définition d'une limite en deçà de laquelle il est impossible de détecter la présence ou l'absence d'un signal.

Nous offrons ainsi une alternative aux études 'pire cas' généralement proposées dans la littérature [50]. **Nos résultats donnent une forme explicite de l'incertitude du bruit sur la dégradation des capacités de détection du détecteur d'énergie.**

Seuls les résultats principaux sont présentés dans ce résumé. Pour plus de détails, les lecteurs sont invités à lire les chapitres correspondant en Anglais.

Ces travaux ont été publiés dans une revue et une conférence.

D.2 Détection d'Énergie

D.2.1 Modèle, Hypothèses et Notations

Soit $\mathbf{y}_t = [y_{t,0}, y_{t,1}, \dots, y_{t,M-1}]$, un vecteur de M échantillons indépendants et identiquement distribués (i.i.d.) obtenu pendant la durée du paquet temporel $t \in \mathbb{N}$. L'issue du processus de détection peut être modélisé par le test d'hypothèse binaire suivant: ⁽¹⁾

$$\mathbf{y}_t = \begin{cases} \mathbf{n}_t, & \mathbb{H}_0 \\ \mathbf{x}_t + \mathbf{n}_t, & \mathbb{H}_1 \end{cases}$$

où les hypothèses \mathbb{H}_0 et \mathbb{H}_1 désignent respectivement les cas d'absence ou présence de signal dans le paquet analysé. D'une part, $\mathbf{x}_t = [x_{t,0}, x_{t,1}, \dots, x_{t,M-1}]$ désigne le signal où chaque échantillon $x_{t,k}$ est une réalisation i.i.d. d'une distribution normale $\mathcal{N}(0, \sigma_{x,t}^2)$. D'autre part, $\mathbf{n}_t = [n_{t,0}, n_{t,1}, \dots, n_{t,M-1}]$ désigne les échantillons i.i.d. extraits d'un bruit additif Gaussien $\mathcal{N}(0, \sigma_{n,t}^2)$. De plus, \mathbf{x}_t et \mathbf{n}_t sont supposés indépendants. Ainsi, nous considérons les vecteurs d'échantillons reçus en fonction de l'hypothèse $\forall y_{t,i} \ i \in \{0, \dots, M-1\}$:

$$\begin{cases} \mathbb{H}_0 : y_{t,i} \sim \mathcal{N}(0, \sigma_{n,t}^2) \\ \mathbb{H}_1 : y_{t,i} \sim \mathcal{N}(0, \sigma_{x,t}^2 + \sigma_{n,t}^2) \end{cases}$$

Le paragraphe suivant rappelle les critères habituellement utilisés afin d'évaluer la qualité d'un détecteur.

D.2.2 Évaluation d'un détecteur

La qualité d'un détecteur est habituellement évaluée via le calcul de la probabilité de fausse alarme $\mathbb{P}_{fa,t}$ et de la probabilité de détection $\mathbb{P}_{d,t}$. Soit d_t la décision prise relative à la présence ou l'absence du signal au test t :

$$\begin{cases} \mathbb{P}_{fa,t} = \mathbb{P}(d_t = 1 | \mathbb{H}_0) \\ \mathbb{P}_{d,t} = \mathbb{P}(d_t = 1 | \mathbb{H}_1) \end{cases}$$

En règle générale, les détecteurs sont paramétrés afin de satisfaire une contrainte relative à la probabilité de fausse alarme, $\mathbb{P}_{fa,t} \leq \alpha$ pour un certain $\alpha \in [0, 1]$. Le détecteur le plus puissant est celui qui maximise $\mathbb{P}_{d,t}$ sous la contrainte $\mathbb{P}_{fa,t} = \alpha$.

D.2.3 Détecteur d'Énergie: modèle de Neyman-Pearson

Le détecteur d'énergie dit de Neyman-Person (ED-NP) suppose la connaissance du niveau du bruit à chaque instant $\sigma_{n,t}^2$. Par conséquent, pour des raisons de clarté et sans perte de généralité, nous considérons que le niveau du bruit est constant : $\sigma_{n,t}^2 = \sigma_n^2$. Il est alors connu que sous ces conditions, l'ED-NP est le détecteur le plus puissant.

Prendre une décision revient à comparer l'énergie \mathcal{T}_t du paquet de M échantillons à un seuil :

$$\begin{aligned} \mathcal{T}_t &= \sum_{i=0}^{M-1} |y_{t,i}|^2 \\ \mathcal{T}_t &\underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\leq}} \xi_t(\alpha) \end{aligned}$$

⁽¹⁾Dans ce chapitre, nous associons la valeur numérique 1 à l'existence d'un signal à détecter, 0 sinon.

où $\xi_t(\alpha)$ représente le seuil choisi afin de garantir $\mathbb{P}_{f_a,t} \leq \alpha$.

Nous rappelons les expressions de $\mathbb{P}_{f_a,t}$ et $\mathbb{P}_{d,t}$ (où $\mathcal{T}_t \sim \chi_M^2$ suit une distribution 'Chi-deux' avec M degrés de liberté, ainsi que son approximation usuelle lorsque M devient suffisamment grand, où la statistique \mathcal{T}_t est approximée par une distribution normale) :

$$\begin{cases} \mathbb{P}_{f_a,t} = 1 - F_{\chi_M^2} \left(\frac{\xi_t(\alpha)}{\sigma_n^2} \right) \\ \mathbb{P}_{d,t} = 1 - F_{\chi_M^2} \left(\frac{\xi_t(\alpha)}{\sigma_n^2 + \sigma_{x,t}^2} \right) \end{cases}$$

$F_{\chi_M^2}(\cdot)$ désigne la fonction de répartition de la distribution χ^2 avec M degrés de liberté.

Lorsque le nombre d'échantillons considérés est suffisamment grand ($M \geq 200$), il est souvent admis qu'une approximation normale est satisfaisante [94, 49]. Notons, néanmoins, qu'en général cette approximation est inutile puisque la forme exacte de la distributions χ_M^2 est connue.

$$\begin{cases} \mathbb{P}_{f_a,t} \approx Q \left(\sqrt{\frac{M}{2}} \left(\frac{\xi_t(\alpha)/M}{\sigma_n^2} - 1 \right) \right) \\ \mathbb{P}_{d,t} \approx Q \left(\sqrt{\frac{M}{2}} \left(\frac{\xi_t(\alpha)/M}{\sigma_n^2 + \sigma_{x,t}^2} - 1 \right) \right) \end{cases}$$

où $Q(\cdot)$ désigne le complémentaire de la fonction de répartition d'une variable aléatoire gaussienne (connue aussi sous le nom de *fonction de Marcum*) [98]:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{y^2}{2}} dy$$

Lorsque la puissance du bruit σ_n^2 est connue, le détecteur ED-NP est satisfaisant. Lorsque cette information n'est pas disponible, les performances du détecteur se dégradent rapidement. Nous proposons de quantifier de manière explicite la dégradation des performances du détecteur.

D.2.4 Détection d'énergie avec un niveau de bruit incertain

Les auteurs du papier [49] proposent l'analyse de l'impact d'une information incertaine relative au niveau du bruit sur les performances du détecteur d'énergie. Pour cela, Ils définissent deux modèles. Le premier considère que la puissance du bruit est uniquement connue dans un certain intervalle. Dans ce contexte ils montrent, à travers une étude du pire cas, qu'il existe une limite du rapport signal à bruit en-deçà duquel il est impossible de détecter un signal. La difficulté en pratique d'aboutir à un intervalle certain (avec probabilité 1) a mené les auteurs à considérer un modèle plus réaliste mais plus complexe à résoudre. Il repose sur des mesures empiriques. Ces dernières montrent que dans de nombreux scénarios, l'estimation de la puissance du bruit semble suivre une loi Log-Normale. Cela ramène l'étude de leur détecteur à un ratio entre une variable de loi χ^2 et une variable de loi Log-Normale. Un tel ratio n'a malheureusement pas de forme explicite simple. Ils ramenèrent donc le second problème au premier modèle à travers un intervalle de confiance.

Afin de répondre au second modèle, nous proposons dans cette thèse une approximation de la loi χ^2 par une loi Log-Normale. Cette approximation est évaluée puis exploitée afin de résoudre le problème resté non-résolu dans le papier [49].

• **Approximation errors (3rd order approximation)**

$$\begin{cases} \tilde{\Delta}_1(x) = \left(-\frac{1}{6M} - \frac{x-M}{M} + \frac{(x-M)^2}{M^2} + \frac{(x-M)^3}{6M^2} \right) f_{\mathcal{N}(\mu_N, \sigma_N^2)}(x) \\ \tilde{\Delta}_2(x) = \left(-\frac{5}{12M} - \frac{x-M}{2M} + \frac{5(x-M)^2}{8M^2} + \frac{(x-M)^3}{12M^2} \right) f_{\mathcal{N}(\mu_N, \sigma_N^2)}(x) \end{cases}$$

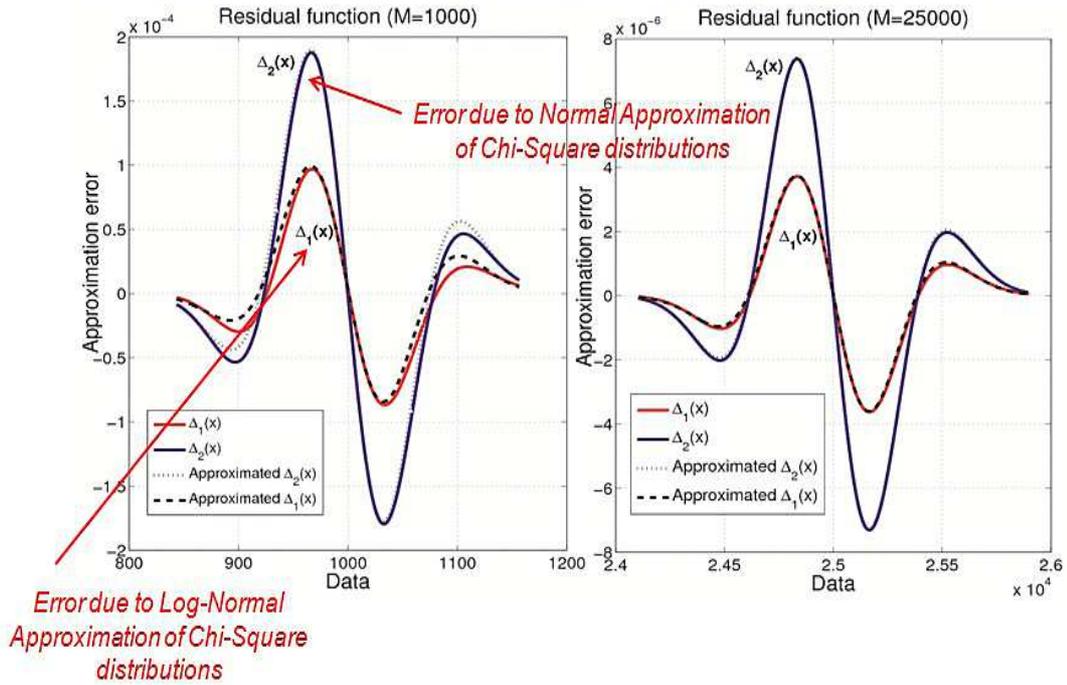


Figure D.1: Approximation de la fonction erreur (Résultats et figures issues du papier [143]). Les deux figures de droite et de gauche présentent les fonctions $\Delta_1(\cdot)$, $\Delta_2(\cdot)$, $\tilde{\Delta}_1(\cdot)$ et $\tilde{\Delta}_2(\cdot)$ (C.f. paragraphe 3.3.2). La figure de gauche montre ces fonctions pour un paramètre $M = 1000$, alors que la figure de droite montre pour un paramètre $M = 25000$. Nous observons que les approximations théoriques introduites dans la propriété 1, paragraphe 3.3.2- et les équations en haut de la figure ci-dessus- convergent bien vers les vraies valeurs lorsque M grandit.

D.3 Approximation Log-Normal d'une distributions χ^2

Dans ce résumé nous ne présenterons pas les détails de l'étude. Afin de présenter la contribution liée à cette section, nous allons néanmoins brièvement la décrire à travers la simulation des résultats. Pour les détails de l'étude, nous invitons le lecteur à se référer au Chapitre 3.

La figure D.1 montre les fonctions erreurs $\Delta_1(\cdot)$ et $\Delta_2(\cdot)$ qui représentent la différence entre la fonction de densité de probabilité de la loi χ^2 et celles des lois Log-Normale et loi normale respectivement, ainsi que leurs approximations respectives, $\tilde{\Delta}_1(\cdot)$ et $\tilde{\Delta}_2(\cdot)$, introduites via l'équation 3.11 au paragraphe 3.3.2. La figure de gauche montre les résultats pour une valeur du paramètre $M = 1000$ tandis que la figure de droite considère la valeur $M = 25000$. L'objectif de ces deux figures s'articule en deux temps : d'une part,

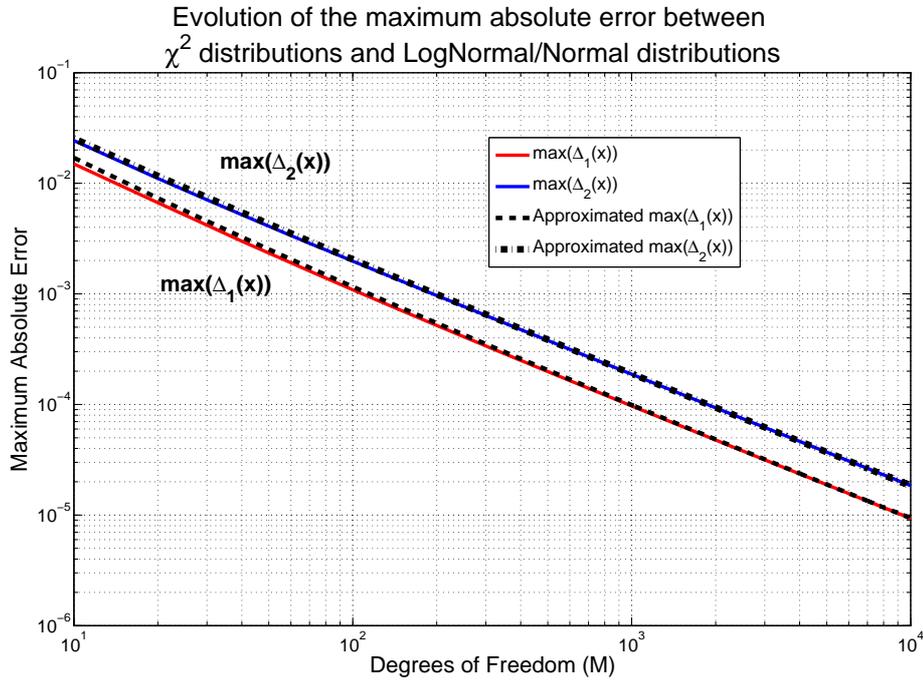


Figure D.2: Erreur Maximale en valeur absolue (Maximum Absolute Error). Sur cette figure, quatre courbes sont affichées: deux en pointillé et deux en trait plein. Les courbes en pointillé illustrent le taux de décroissance de l'amplitude maximale des fonctions $\Delta_1(\cdot)$ et $\Delta_2(\cdot)$. Tandis que, les deux autres courbes, affichent les valeurs théoriques approximées à l'aide des fonctions $\tilde{\Delta}_1(\cdot)$ et $\tilde{\Delta}_2(\cdot)$. Nous pouvons remarquer que les résultats théoriques développés à travers les équations 3.13 et 3.14 décrivent bien la réalité dans les deux cas.

elles cherchent à illustrer les erreurs dues à l'approximation de la distribution χ^2 par une distribution normale ou une distribution Log-Normale. Cet aspect est mis en valeur par les courbes $\Delta_1(\cdot)$ et $\Delta_2(\cdot)$ sur les figures (rouge et bleu en pointillé). D'autre part, nous cherchons à évaluer la précision de l'approximation d'ordre fini de ces deux fonctions $\tilde{\Delta}_1(\cdot)$ et $\tilde{\Delta}_2(\cdot)$ en fonction du paramètre M .

Nous observons tout d'abord que les fonctions $\Delta_1(\cdot)$ and $\Delta_2(\cdot)$, présentent une courbe similaire. Néanmoins, une approximation gaussienne de la distribution χ^2 semble mener à des erreurs d'amplitude plus larges qu'une approximation par une loi Log-Normale. Cela suggère qu'une approximation Log-Normal devrait être en général préférée (si une approximation est nécessaire).

Deuxièmement, nous observons que l'approche théorique à travers les fonctions $\tilde{\Delta}_1(\cdot)$ et $\tilde{\Delta}_2(\cdot)$ offre des approximations précises des fonctions d'erreur, surtout lorsqu'il s'agit d'évaluer le maximum global de la fonction. La figure D.2, insiste sur cette dernière remarque.

D.4 Détecteur d'Énergie et le Modèle Log-Normal de l'Incertitude

D.4.1 Approximation de la Statistique d'Énergie et Incertitude sur le Niveau du Bruit

Soit $\hat{\sigma}_n^2$ la puissance estimée du bruit. En supposant que $\hat{\sigma}_n^2$ suit un loi Log-Normale non biaisée [49], telle que l'espérance et la variance de cette distribution vérifient respectivement, $\mathbb{E} [\hat{\sigma}_n^2] = \sigma_n^2$ et $\mathbb{V} [\hat{\sigma}_n^2] = u \cdot \sigma_n^4$, avec u le paramètre d'incertitude, introduit aussi par la définition (non conventionnelle) en décibel $u = 10^{U_{dB}/10} - 1$, alors :

$$\hat{\sigma}_n^2 \sim \text{LogN}(\mu_u, \mathbb{V}_u), \text{ s.t. : } \begin{cases} \mathbb{V}_u = \log(1+u) \\ \mu_u = 2 \log(\sigma_n) - \frac{\mathbb{V}_u}{2} \end{cases} \quad (\text{D.1})$$

tel que μ_u et \mathbb{V}_u désignent respectivement l'espérance et la variance de la distribution d'incertitude du bruit.

Ainsi supposons que la statistique de puissance \mathcal{T}_t/M peut être approximée par une loi Log-Normale telle que $\mathbb{E} [\mathcal{T}_t/M] = \sigma_T^2$ et $\mathbb{V} [\mathcal{T}_t/M] = 2\sigma_T^4/M$ alors :

$$\frac{\mathcal{T}_t}{M} \sim \text{LogN}(\mu_T, \mathbb{V}_T), \text{ s.t. : } \begin{cases} \mathbb{V}_T = \log\left(1 + \frac{2}{M}\right) \\ \mu_T = 2 \log(\sigma_T) - \frac{\mathbb{V}_T}{2} \end{cases} \quad (\text{D.2})$$

avec σ_T^2 le niveau de puissance moyen des échantillons obtenus en fonction de la présence ou l'absence de signal au slot t :

$$\begin{cases} \mathbb{H}_0 : \sigma_T^2 = \sigma_n^2 \\ \mathbb{H}_1 : \sigma_T^2 = \sigma_n^2 + \sigma_{x,t}^2 \end{cases} \quad (\text{D.3})$$

Finalement nous introduisons la variable aléatoire suivante \mathcal{W}_t telle que :

$$\mathcal{W}_t = \log\left(\frac{\mathcal{T}_t}{M\hat{\sigma}_n^2}\right) \quad (\text{D.4})$$

Nous présentons dans la prochaine sous-section les performances du détecteur d'énergie à l'aide de la variable aléatoire \mathcal{W}_t .

D.4.2 Performances et limites du détecteur d'énergie

En supposant les hypothèses précédentes valables, nous présentons dans la suite de ce chapitre les résultats principaux de notre étude. Afin d'anticiper les résultats des prochains paragraphes, nous introduisons les notations suivante, rapport signal-à-bruit (Signal-to-Noise Ratio ou SNR) $\gamma_t = \sigma_{x,t}^2/\sigma_n^2$ et :

$$\begin{cases} \mathbb{E} [\mathcal{W}_t | \mathbb{H}_0] = \frac{1}{2} (\mathbb{V}_u - \mathbb{V}_T) = \frac{1}{2} \log\left(\frac{1+u}{1+2/M}\right) \\ \mathbb{E} [\mathcal{W}_t | \mathbb{H}_1] = \log(1 + \gamma_t) + \mathbb{E} [\mathcal{W}_t | \mathbb{H}_0] \\ \mathbb{V} [\mathcal{W}_t] = \mathbb{V}_u + \mathbb{V}_T = \log\left((1 + \frac{2}{M})(1+u)\right) \end{cases} \quad (\text{D.5})$$

Lemma 4 (Distribution de \mathcal{W}_t). *Soit \mathcal{W}_t la variable aléatoire introduite dans l'équation D.4. Nous supposons que les hypothèses introduites précédemment sont valables, alors :*

$$\frac{\mathcal{W}_t - \mathbb{E} [\mathcal{W}_t | \mathbb{H}_0]}{\sqrt{\mathbb{V} [\mathcal{W}_t]}} \sim \begin{cases} \mathbb{H}_0 : \mathcal{N}(0, 1) \\ \mathbb{H}_1 : \mathcal{N}\left(\frac{\log(1+\gamma_t)}{\sqrt{\mathbb{V} [\mathcal{W}_t]}}, 1\right) \end{cases} \quad (\text{D.6})$$

Nous présentons à présent les performances du détecteur analysé dans ce chapitre.

Theorem 6 (Les performances du détecteur). *Soit $\xi_t(\alpha)$ une variable réelle telle que : $\mathcal{W}_t \underset{\mathcal{H}_1}{\leq} \log(\xi_t(\alpha))$, alors les probabilités de fausse alarme et de bonne détection ont les formes suivantes :*

$$\begin{cases} \mathbb{P}_{f_a,t} = Q \left(\frac{\log \left(\xi_t(\alpha) \sqrt{\frac{1+2/M}{1+u}} \right)}{\sqrt{\log((1+2/M)(1+u))}} \right) \\ \mathbb{P}_{d,t} = Q \left(\frac{\log \left(\frac{\xi_t(\alpha)}{1+\gamma_t} \sqrt{\frac{1+2/M}{1+u}} \right)}{\sqrt{\log((1+2/M)(1+u))}} \right) \end{cases} \quad (D.7)$$

Le théorème 6 apporte deux contributions importantes : d'une part, il donne une expression explicite des probabilités $\mathbb{P}_{f_a,t}$ et $\mathbb{P}_{d,t}$. Et d'autre part, il permet facilement de calculer la valeur du seuil en fonction de la probabilité de fausse alarme.

$$\log(\xi_t(\alpha)) = Q^{-1}(\mathbb{P}_{f_a}) \sqrt{\mathbb{V}[\mathcal{W}_t]} + \mathbb{E}[\mathcal{W}_t | \mathbb{H}_0]$$

Contrairement au premier modèle proposé par Alexander Sonnenschein et Philip M. Fishman [49], où le modèle d'incertitude suppose connu un intervalle dans lequel la valeur nominale du bruit se trouve, le modèle log-normal permet de calculer *a priori* le seuil à utiliser en fonction des performances désirées. En effet dans le cas du premier modèle du papier [49], repris depuis par [50], le choix des paramètres ne garantit qu'une borne supérieure sur la fausse alarme. Cela mène naturellement à une borne inférieure en terme de pouvoir de détection qui explique la notion de *SNR-wall*.

Le résultat suivant fournit une expression générale du *SNR-wall* en fonction des performances désirées du détecteur et des incertitudes $\{2/M; u\}$.

Theorem 7 (SNR-wall). *Soit \mathcal{W}_t la variable aléatoire définie par l'équation D.4 et soit $\Delta = Q^{-1}(\mathbb{P}_{f_a}) - Q^{-1}(\mathbb{P}_d)$, alors le *SNR-wall* du détecteur d'énergie soumis à une incertitude Log-Normale vérifie :*

$$\gamma_{wall,t} = e^{\Delta \sqrt{\mathbb{V}[\mathcal{W}_t]}} - 1 \quad (D.8)$$

En considérant une incertitude modélisée à l'aide d'une distribution Log-Normale, la valeur du *SNR-wall* dépend des performances désirées en terme de probabilité de fausse alarme et la probabilité de détection. Ainsi, ce résultat peut être interprété en tant que SNR minimum nécessaire afin de garantir les performances désirées. Ce résultat change donc des résultats précédents [50] qui présentent une limite de SNR en deçà de laquelle les détections ne peuvent plus être considérées comme pertinentes. L'impact de l'incertitude du bruit sur les performances du détecteur d'énergie, décrit dans les équations du théorème 6, sont illustrées sur la figure D.3 (figure de gauche).

La figure de gauche représente sept courbes de probabilité de détection pour une fausse alarme de 0.1. La première courbe intitulée "NP-Energy Detector" représente la courbe d'un détecteur d'énergie sans incertitude. La deuxième courbe intitulée "Energy Detector $U = 0$ dB", qui se superpose à la première courbe, illustre les résultats du détecteur modélisé avec une approximation Log-Normale des distributions χ^2 mais avec une incertitude $U = 0$. Les cinq autres courbes, montrent l'impact de l'incertitude sur les performances

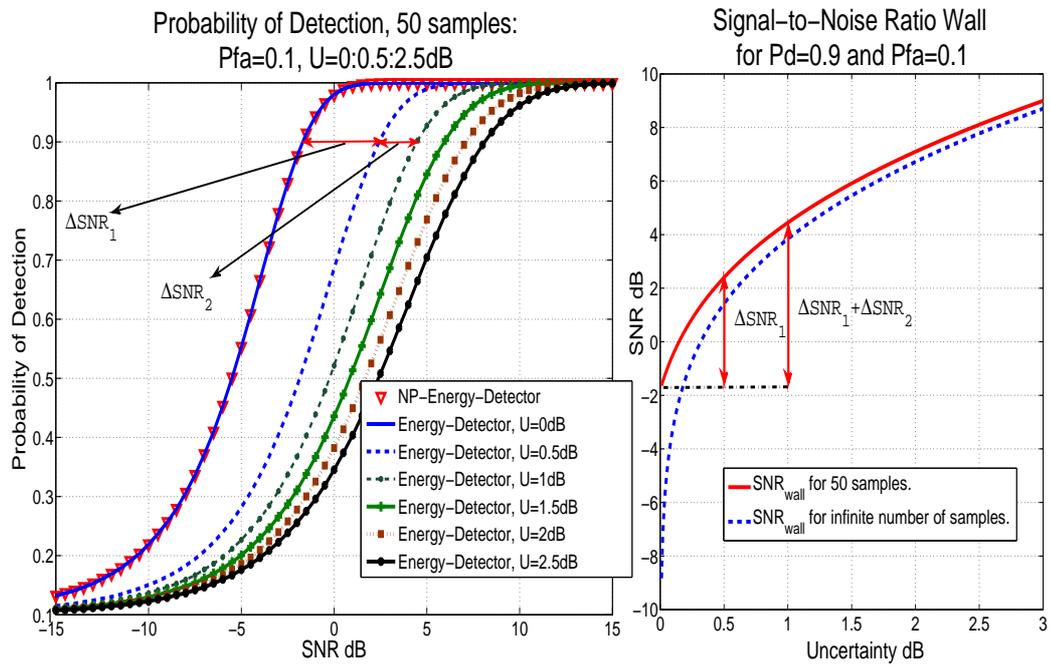


Figure D.3: Probabilité de détection (gauche) et SNR -wall (droite). La figure de gauche montre l'impact de l'incertitude du bruit tel que défini dans ce chapitre sur les performances du détecteur d'énergie. Les courbes de droites montrent que les pertes de performances peuvent être prédites grâce à la nouvelle formule du SNR -wall. Remarquons que l'impact de l'incertitude, dans le sens introduit par [49] et utilisé par [50] n'apparaît pas sur ces courbes. En effet il est en général impossible suivant leur modèle d'imposer une fausse alarme prédéfinie. Par conséquent la comparaison serait biaisée.

du détecteur d'énergie. La courbe de droite montre que le gap qui existe entre les performances du détecteur sans incertitude et celui avec incertitude peut être prédit en fonction du SNR. Ces courbes confirment les théorèmes introduits précédemment et valident l'analyse théorique de ce chapitre. Une analyse plus détaillée est fournie dans la version anglaise de ce manuscrit, au Chapitre 3.

D.5 Conclusion

Cette étude répond simultanément à deux challenges. D'une part nous avons évalué l'approximation de lois χ^2 par une distribution Log-Normal appropriée. Nous avons ainsi montré que celle-ci offre une meilleure approximation que l'approximation Gaussienne. Ensuite cette approximation a été exploitée afin de quantifier la dégradation du pouvoir de détection du détecteur d'énergie lorsque sa connaissance du niveau du bruit devient incertaine.

Appendix E

Apprentissage pour l'Accès Opportuniste au Spectre : Prise en Compte des Erreurs d'Observation

Contents

D.1	Introduction	160
D.2	Détection d'Énergie	161
D.2.1	Modèle, Hypothèses et Notations	161
D.2.2	Évaluation d'un détecteur	161
D.2.3	Détecteur d'Énergie: modèle de Neyman-Pearson	161
D.2.4	Détection d'énergie avec un niveau de bruit incertain	162
D.3	Approximation Log-Normal d'une distributions χ^2	163
D.4	Détecteur d'Énergie et le Modèle Log-Normal de l'Incertitude	165
D.4.1	Approximation de la Statistique d'Énergie et Incertitude sur le Niveau du Bruit	165
D.4.2	Performances et limites du détecteur d'énergie	165
D.5	Conclusion	168

E.1 Introduction

E.1.1 Accès Opportuniste au Spectre

L'accès Opportuniste au Spectre (*en anglais : Opportunistic Spectrum Access-OSA*) est un concept prometteur, suggéré par la communauté radio, pour mieux exploiter les opportunités spectrales aujourd'hui disponibles. En effet, durant le premier siècle d'existence de la radio d'origine humaine, l'allocation statique des bandes de fréquence aux applications et services sans fil (au nombre sans cesse croissant), a mené à une pénurie de la ressource spectrale. Néanmoins, de nombreuses mesures effectuées aux Etats-Unis, d'abord, corroborées ensuite par des études similaires dans le reste du monde, montrent une sous utilisation chronique du spectre [11]. Ces mesures montrent par la même occasion des opportunités de communication substantielles à exploiter. La radio-intelligente (*Cognitive*

Radio -CR) s'est très vite positionnée comme une candidate crédible pour y répondre via l'OSA notamment.

Le concept général de l'accès opportuniste au spectre définit deux classes d'utilisateurs : les utilisateurs primaires (UP) et les utilisateurs secondaires (US). Les UPs ont accès aux ressources spectrales dédiées à leurs services. Ils sont donc prioritaires sur ces bandes de fréquence pour lesquelles ils payent un droit d'exploitation. Les USs, par opposition aux UPs, représentent un groupe d'utilisateurs désireux d'exploiter les opportunités de communication laissées vacantes, à un certain moment dans une certaine zone géographique, par les UPs.

Il est généralement admis que les USs n'ont pas (ou peu) d'information *a priori* sur l'occupation des bandes primaires. De plus, les interférences occasionnées par les USs doivent rester sous un certain seuil toléré par les UPs. Afin de répondre à ces exigences, le concept de la radio intelligente (Cognitive Radio en anglais) a été suggéré [11, 9] à condition que les équipements secondaires soient dotés de capacités cognitives élémentaires, à savoir : observation de l'environnement à travers des capteurs dédiés, analyse des informations collectées, et enfin adaptation du comportement de l'équipement aux fluctuations de l'environnement et aux attentes de l'utilisateur, tout en respectant les contraintes du régulateur.

Il reste néanmoins de nombreux défis à surmonter afin d'exploiter de manière efficace les opportunités présentes dans le spectre⁽¹⁾. Cela implique d'une part, la conception de détecteurs précis et fiables, et d'autre part, l'analyse de mécanismes d'apprentissage et de prise de décision performants. Proposer de tels algorithmes est depuis l'avènement de la radio intelligente au centre de nombreuses recherches [13, 63].

E.1.2 Le Paradigme des Bandits Manchots

Le paradigme des bandits manchots (Multi-Armed Bandit en anglais) a été récemment le centre d'une attention particulière de la part de la communauté radio. Cette thèse se positionne, ainsi, parmi les premiers travaux qui associent ce paradigme à celui de la radio intelligente.

Brièvement, ce paradigme modélise l'agent de prise de décision par un joueur dans un casino. Ce dernier cherche à maximiser les gains cumulés obtenus en tirant le bras de différentes machines à sous (Bandits manchots, Multi-Armed Bandit, en anglais). Ces dernières représentent les ressources à exploiter par l'agent intelligent. Si ce joueur avait une information complète sur les gains moyens de chaque machine à sous, une stratégie optimale serait de jouer en permanence la machine avec le gain moyen le plus élevé. Néanmoins, dans la mesure où le joueur ne dispose d'aucune information sur ce qu'il pourrait gagner en jouant telle ou telle machine, il n'a d'autre choix que de tester les différentes machines afin d'estimer leur gain moyen. La recherche d'un équilibre entre le temps passé à tester les différentes machines afin d'estimer leurs performances respectives, et le temps consacré à la machine qui semble être optimale est ce qui est habituellement appelé *dilemme Exploitation-Exploration*. Si nous imaginons que ces machines à sous représentent des bandes spectrales auxquelles l'utilisateur secondaire cherche à accéder, ce problème de décision et d'apprentissage en radio intelligente s'apparente à un problème des bandits manchots.

⁽¹⁾Ainsi, on cherche à persuader les régulateurs de changer les règles établies depuis 100 ans.

Ainsi, plusieurs algorithmes ont été empruntés au domaine de l'apprentissage machine [101, 102, 103] et suggérés pour répondre à la problématique de l'accès opportuniste au spectre [36, 67, 137, 131]⁽²⁾. Ces algorithmes, néanmoins supposent une observation sans erreur de l'état des bandes de fréquence. Dans ce contexte, l'utilisateur secondaire peut effectivement maximiser ses gains cumulés sans interférer avec les utilisateurs primaires. Nous avons présenté ces travaux dans les papiers [36, 67].

L'objectif de ce chapitre est d'introduire un modèle unifié plus réaliste. Ainsi, nous considérons un modèle des machines à sous dans lequel des erreurs d'observations sont possibles. Le modèle considéré est détaillé dans la section E.2 ainsi que dans nos publications [68, 133, 145]. La section E.3 introduit l'algorithme UCB_1 et montre ses performances dans le cas de l'accès opportuniste au spectre. Ces résultats sont illustrés à l'aide de simulations. Notons que les chapitres 5-6 de la version anglaise décrivent les travaux que nous avons effectués pour étendre cette étude à des contextes toujours plus complexes et plus réalistes. Enfin la section E.4 conclut le chapitre. Pour l'étude détaillée ainsi que les nombreuses extensions réalisées autour de ces travaux, nous invitons les lecteurs à se référer à la version anglaise de ce manuscrit.

E.2 Modélisation de la problématique

E.2.1 Le réseau primaire

Dans le cadre de l'Accès Opportuniste au Spectre [11, 63], nous considérons le cas d'un utilisateur dit "secondaire" qui cherche à exploiter par une bande de fréquence dédiée à un réseau prioritaire dit "réseau primaire". La bande de fréquence d'intérêt est supposée divisée en K sous-bandes indépendantes mais non-identiques.

Soit k l'indice du $k^{\text{ème}}$ canal le plus disponible en probabilité. A chaque fois qu'un canal est sondé, il est observé dans l'un des deux états suivants : {libre, occupé}. Dans le reste du chapitre, nous associons la valeur numérique 0 à un canal occupé, et 1 autrement. L'occupation temporelle d'une sous-bande k est supposée suivre une distribution de Bernoulli inconnue θ_k . De plus les distributions $\Theta = \{\theta_1, \theta_2, \dots, \theta_K\}$ sont supposées stationnaires.

Nous sommes dans le cas particulier d'un réseau primaire synchrone où le temps $t = 0, 1, 2, \dots$, est supposé divisé en paquets de taille fixée. Notons \mathbf{S}_t l'état des canaux à l'itération t : $\mathbf{S}_t = \{S_{1,t}, \dots, S_{K,t}\} \in \{0, 1\}^K$. Pour tout $t \in \mathbb{N}$, la valeur numérique $S_{k,t}$ est supposée être la réalisation aléatoire de la distribution θ_k . De plus, les réalisations $\{S_{k,t}\}_{t \in \mathbb{N}}$ tirées de la distribution θ_k sont supposées indépendantes et identiquement distribuées. La disponibilité moyenne d'un canal est caractérisée par sa probabilité d'être libre. Ainsi, nous définissons la disponibilité μ_k du canal k telle que pour tout t : $\mu_k = \mathbb{P}(S_{k,t} = 1)$, avec $\mu_1 > \mu_2 \geq \dots \geq \mu_k \geq \dots \geq \mu_K$ sans perte de généralité. Le but de l'apprentissage, tel que nous l'entendons dans notre problème, est de permettre à l'agent l'identification la plus rapidement possible du canal avec la disponibilité moyenne maximale, μ^* . Cela afin de pouvoir l'exploiter. L'algorithme choisi pour l'apprentissage devrait ainsi permettre d'apprendre tout en fournissant un service à l'utilisateur. Un service qui devrait s'améliorer au fur et à mesure que la phase d'apprentissage progresse.

⁽²⁾Le papier [131], en 2008, qui semble antérieur à nos travaux [36, 67] n'a été cependant publié que sur le site arxiv. La publication officielle, après un comité de relecture, est apparue quant à elle en février 2011 [144] et présente de nombreuses modifications et corrections.

E.2.2 L'utilisateur secondaire

Nous décrivons dans ce paragraphe le moteur de prise de décision ainsi que les caractéristiques du détecteur de signaux associés à l'utilisateur secondaire.

Nous dénommons "Agent Intelligent" (AI) le moteur de prise de décision de l'équipement de radio intelligente. L'AI peut être vu comme le centre névralgique de l'équipement. A chaque paquet t , il doit choisir un canal à observer. Pour cela, les décisions de l'AI reposent sur les informations passées collectées au fur et à mesure de ses interactions avec l'environnement. Soit i_t le vecteur "information" disponible à l'instant t . Nous supposons que l'AI ne peut observer qu'un canal à la fois à chaque itération t . Ainsi, le choix d'un canal à observer peut être associé à une action $a_t \in \mathcal{A}$ où l'ensemble $\mathcal{A} = \{1, 2, \dots, K\}$ fait référence à l'ensemble de canaux considéré par l'utilisateur secondaire. Par conséquent l'AI peut être considéré en tant que fonction de décision π qui associe pour tout t , une action a_t à l'information i_t : $a_t = \pi(i_t)$

Soit $X_t \in \{0, 1\}$ la réalisation aléatoire calculée à l'issue de l'étape d'observation à l'instant t du canal sélectionné a_t . Dans le cas d'une détection parfaite (sans erreur d'observation), nous aurions : $X_t = S_{a_t, t}$. Dans le contexte considéré, cependant, la valeur de X_t est fonction des caractéristiques opérationnelles du récepteur (COR). Les COR définissent la précision et la fiabilité d'un détecteur via la mesure de deux types d'erreurs : d'une part la détection d'un utilisateur primaire sur la bande sondée alors que la bande est en réalité libre. Cette erreur est communément appelée "fausse alarme". D'autre part, la "détection manquée" revient à considérer une bande libre alors qu'elle est occupée par des utilisateurs primaires à l'instant t . Notons ϵ et δ , respectivement, les probabilités de fausse alarme et de détection manquée caractérisant l'équipement de radio intelligente considéré lors de cette étude :

$$\begin{cases} \epsilon = \mathbb{P}_{fa} = \mathbb{P}(X_t = 0 | S_{a_t, t} = 1) \\ \delta = \mathbb{P}_{md} = \mathbb{P}(X_t = 1 | S_{a_t, t} = 0) \end{cases}$$

Finalement, le résultat de l'étape d'observation peut être associé à la sortie d'une fonction aléatoire $\pi_s(\epsilon, \delta, S_{a_t, t})$ telle que : $X_t = \pi_s(\epsilon, \delta, S_{a_t, t})$. Cependant, nous ne nous intéressons pas aux formes possibles de fonctions de détection et renvoyons le lecteur intéressé à la référence [13].

E.2.3 Stratégie d'accès au canal

En fonction du résultat de l'étape d'observation $X_t \in \{0, 1\}$, l'AI peut choisir d'accéder ou non au canal sélectionné. Soit $\pi_a(X_t) \in \{0, 1\}$ la décision d'accès au canal, telle que, d'une part, 0 désigne un refus d'accès et, d'autre part, 1 désigne une autorisation d'accès au canal. Pour des raisons de simplicité, la stratégie choisie dans cette étude se résume à la règle : "accéder le canal s'il est observé libre". En d'autres termes : $\pi_a(X_t) = X_t$. Cette hypothèse est largement partagée dans la littérature pour des considérations de complexité de réalisation d'une part, et d'efficacité au niveau du réseau d'autre part. Nous supposons dans ces travaux que le détecteur est modélisé de manière à assurer un niveau d'interférence, avec l'utilisateur primaire, inférieur à un certain seuil fixé par les réglementations. Néanmoins, nous ne supposons pas nécessairement connus les paramètres $\{\epsilon, \delta\}$ ⁽³⁾.

⁽³⁾En effet, d'après l'étude réalisée dans le chapitre précédent, les limites rencontrés par les détecteurs actuels ne permettent malheureusement pas de toujours connaître les paramètres $\{\epsilon, \delta\}$. L'algorithme d'apprentissage doit donc pouvoir s'en affranchir.

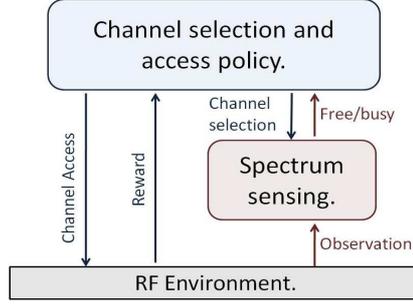


Figure E.1: Représentation de l'interaction d'un agent intelligent avec son environnement RF.

De plus nous supposons qu'il existe un mécanisme permettant à l'utilisateur secondaire d'être informé en cas d'interférence avec l'utilisateur primaire. Dans ce cas, la transmission de l'utilisateur secondaire est assimilée à un échec. Enfin, nous supposons qu'un paquet de taille D_t est envoyé à chaque tentative de transmission (i.e., accès au canal). Ainsi, à la fin de chaque paquet t , l'AI calcule une valeur numérique r_t , habituellement appelée *gain* dans la communauté de l'apprentissage machine, qui quantifie les performances instantanées du moteur de prise de décision de l'utilisateur secondaire.

L'interaction de l'utilisateur secondaire avec son environnement est résumée et illustrée dans la figure E.1.

E.3 Algorithmes et performances

E.3.1 Algorithme de selection du canal

Nous analysons dans ce chapitre l'impact des erreurs⁽⁴⁾ d'observation sur les performances de l'algorithme UCB_1 . Ce dernier avait été précédemment suggéré pour sa simplicité et la garantie de ses propriétés mathématiques de convergence [102][103][67]. Brièvement, cet algorithme repose sur l'affectation d'un indice de qualité à chaque canal en fonction des précédentes observations et tentatives de transmissions. La forme de l'index considéré dans ce papier est la suivante :

$$B_{k,t,T_k(t)} = \bar{X}_{k,T_k(t)} + A_{k,t,T_k(t)}$$

Dans cette expression, d'une part $\bar{X}_{k,T_k(t)} = \frac{\sum_{m=0}^{t-1} r_m \cdot \mathbf{1}_{\{a_m=k\}}}{T_k(t)}$ représente la moyenne empirique des gains obtenus à partir du canal k , après $T_k(t)$ tentatives au bout de t itérations, d'autre part, $A_{k,t,T_k(t)}$ représente un biais ajouté afin d'assurer la convergence de l'algorithme vers le canal optimal. Dans le cadre de cette étude, le biais considéré est le suivant :

$$A_{k,t,T_k(t)} = \sqrt{\frac{\alpha \cdot \ln(t)}{T_k(t)}}$$

⁽⁴⁾Le cas sans erreur n'est pas présenté dans ce résumé. Néanmoins les performances dans le cas sans erreur peuvent se déduire aisément du cas avec erreurs de détection.

où α est un paramètre (réel positif) d'apprentissage. Finalement, l'algorithme π , de sélection des canaux, choisit à chaque itération t le canal avec la plus grande valeur associée à l'indice $B_{k,t,T_k(t)}$, tel que :

$$a_t = \pi(i_t) = \arg \max_k (B_{k,t,T_k(t)})$$

Une version détaillée de l'implémentation de cet algorithme avait déjà été proposée dans notre précédente étude [67] ainsi que dans le chapitre 4 de la version anglaise de ce manuscrit.

E.3.2 Performances

En vue du modèle introduit précédemment, nous considérons un gain de la forme suivante :

$$r_t \triangleq D_t S_{a_t,t} \pi_a(X_t)$$

Néanmoins pour des raisons de simplicité, nous considérons qu'à chaque tentative de transmission, l'utilisateur secondaire transmet $D_t = 1$ bit. Par conséquent, en tenant compte de la forme de la stratégie d'accès, et des simplifications introduites, le gain r_t peut s'exprimer de la manière suivante :

$$r_t = S_{a_t,t} X_t$$

On montre que l'espérance du temps $\mathbb{E}[T_k(t)]$ passé par l'algorithme à sélectionner un canal sous optimal (i.e., $k \in \{2, \dots, K\}$) est borné par une fonction logarithmique du nombre d'itérations (*slot* en anglais) tel que pour $\alpha > 1$ et $\Delta_k = \mu_1 - \mu_k$:

$$\mathbb{E}[T_k(t)] \leq \frac{4\alpha \ln(t)}{((1 - \epsilon)\Delta_k)^2} \quad (\text{E.1})$$

La preuve de ce résultat est une extension des travaux réalisés dans les papiers de la communauté Machine Learning [102, 103]. Les techniques utilisées pour mener la preuve sont aussi similaires. Pour des raisons d'espace, nous ne prouverons pas ce résultat dans ce papier. Il est néanmoins possible d'y accéder dans notre papier [68] ainsi que le chapitre 4.

Ce résultat montre que malgré les erreurs d'observation, il est toujours possible de converger rapidement vers le canal optimal, néanmoins, sans surprise, nous observons une dégradation de la vitesse de convergence qui est directement liée aux performances du détecteur.

Ce résultat a été, de plus, confirmé en simulation, tel que illustré par la figure E.2. Cette figure suppose la disponibilité de dix canaux avec des probabilités de disponibilités μ_k , $k \in \{1, 2, \dots, K\}$ respectivement égaux à [0.9, 0.8, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1]. Ces résultats sont moyennés sur 100 réalisations de l'expérience. On voit notamment que l'algorithme a une première phase d'exploration pendant laquelle aucun canal n'est favorisé. Ensuite, nous observons que la courbe croît rapidement, mettant en avant la capacité de l'algorithme à déterminer le canal optimal quelque soient les erreurs d'observations. Néanmoins, plus le détecteur est précis, meilleure est la phase d'exploitation de l'algorithme. Ceci est illustré par le temps de convergence de l'AI.

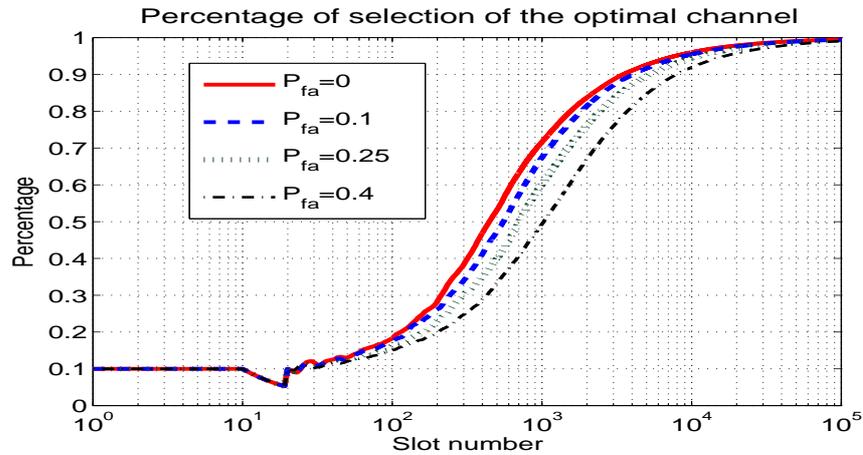


Figure E.2: Pourcentage du temps passé à sélectionner le canal optimal à l'aide de l'algorithme UCB_1 en fonction des erreurs d'observations.

E.4 Conclusion

Nous avons introduit, dans ce papier, un modèle d'accès opportuniste au spectre sous la forme d'un problème de machines à sous. Ce dernier a été complété par un modèle des canaux prenant en compte les erreurs d'observation. En d'autres termes, des erreurs d'observation de l'état des canaux peuvent avoir lieu aléatoirement à chaque mesure. Ensuite, les performances de l'algorithme UCB_1 ont été analysées avant d'être validées par des simulations. Ainsi, ce papier montre que malgré les erreurs d'observations qui peuvent avoir lieu lors de la phase de détection des utilisateurs primaires, l'algorithme UCB_1 reste capable d'apprendre et de converger vers le canal optimal. Ainsi, en fonction de la précision de la détection, la convergence de l'algorithme peut être plus ou moins rapide.

Bien que ces résultats soient encourageants, de nombreux questionnements persistent quant à leur généralisation. En effet dans le cas de réseaux d'utilisateurs secondaires, il est impératif de s'assurer que ces derniers n'interfèrent pas entre eux afin de ne pas ruiner la phase d'apprentissage. Ces points sont notamment discutés dans les chapitres 5 et 6 que nous ne détaillons pas dans cette version résumée.

List of Figures

1	General Synopsis	3
1.1	the global ICT developments during the period 2001-2011 [1].	7
1.2	Major France Telecommunication providers' Wi-Fi access points in Rennes around Saint-Anne metro station. To quickly densify their Wi-Fi network, main operators exploit their subscribers' ADSL box. Thus, usually the Wi-Fi connexion of an ADSL box runs two or three virtual networks that share the same connexion. Usually we find two virtual networks: the first wireless network is dedicated to the subscriber, whereas the second wireless network is managed by the operator. This latter network is shared with other mobile subscribers in the vicinity of the box. Both networks share the same wireless physical card and thus share the same frequency band. Finally, we also noticed, in the case of the operator <i>Free</i> , the existence of a third virtual network dedicated to VoIP.	8
1.3	Estimated economical impact of both licensed and unlicensed radio technologies: results presented by Dave Cleevly during the annual conference of Spectrum management IEEE DySPAN2011. This figure support the idea that more bandwidth should be opened to unlicensed radio exploitation.	10
1.4	Evolution of the bandwidth of the main Radio communications standards (found online: http://3g4g.blogspot.fr/2008_04_01_archive.html) . Similar graphs are provided in [2].	12
1.5	Cognitive radio decision making context: the CR cycle as introduced by Joseph Mitola III [10].	15
1.6	Cognitive radio decision making context [12]. The Cognitive Cycle as introduced by S. Haykin to answer Dynamic Spectrum Access related problems.	16
1.7	Hierarchical and Distributed Cognitive Radio Architecture Management (HDCRAM) [20, 21, 22, 23, 24]. A cognitive cycle management architecture is required to efficiently cycle through the observe, decide and adapt steps. The specificity of the CR context requires the management architecture to be hierarchical and distributed over several processing units. This figure illustrates the basic components and behavior of HDCRAM systems.	18

2.1	Illustration of the basic Cognitive Cycle [38, 39]. As illustrated, an agent, usually referred to as Cognitive Agent (CA) faces an Environment in a broad sense. The CA repeats the Cognitive Cycle where he <i>Observes</i> the environment, <i>Analyzes</i> the collected information and <i>Decides</i> the next action to take. Note that the arrow Action could suggest always an action on the environment. This is possible to evaluate the reaction on the environment to given stimuli. However, the arrow also suggests an action on the CR in order to adapt to the environment	26
2.2	Cognitive radio decision making design space.	29
2.3	Classification of several Dynamic Spectrum Access approaches as suggested in Paper [63]. Three main approaches can be discriminated : Dynamic Exclusive Use Model, Open Sharing Model (Spectrum Commons Model) and Hierarchical Access Model.	33
2.4	Suggested decision making techniques depending on the assumed <i>a priori</i> knowledge.	34
2.5	Multi-criteria optimization problem [32].	36
3.1	Illustration of the energy detection threshold based policy with two outcome classes \mathbb{H}_0 and \mathbb{H}_1 (in this case ‘data’ refers to the estimated power statistic). On the one hand, the first curve on the left refers to the probability density function of the observed energy of noise. On the other hand, the second curve on the right shows the probability density function of the observed energy of the signal.	44
3.2	Approximation Error functions. Both left and right figures plot the functions $\Delta_1(\cdot)$, $\Delta_2(\cdot)$, $\tilde{\Delta}_1(\cdot)$ and $\tilde{\Delta}_2(\cdot)$. However, the left figure shows these functions for a parameter $M = 1000$, while the right figure shows them for $M = 25000$. We observe in this figure that the theoretical approximation introduced in Property 1 seems to converge to the real values as M grows large.	50
3.3	Maximum Absolute Error. In this figure, four curves are represented: two of them, in solid line, illustrate the decreasing rate of the global maximum of the error functions $\Delta_1(\cdot)$ and $\Delta_2(\cdot)$. Whereas, the two other curves, plot the theoretical maximum of the approximations $\tilde{\Delta}_1(\cdot)$ and $\tilde{\Delta}_2(\cdot)$. As we can notice, the theoretical approximations developed in Equations 3.13 and 3.14 describe well the reality in both cases.	51
3.4	Probability of detection (left) and <i>SNR-wall</i> (right). On the left figure, we observe the impact of noise uncertainty, as defined in this chapter, on the detection performances. The right curves shows that the lost of performances can be predicted using the new formula of the <i>SNR-wall</i> . Note that the impact of the “classic” uncertainty[49, 50] does not appear in these curves since it is, generally, impossible to impose a given false alarm, which would make any comparison biased.	55

4.1	Opportunistic Spectrum Access Modeled as a Multi-Armed Bandit Problem. MAB problems represent simple machine learning problems where a gambler have the choice between several machines (or equivalently different arms on a single machine). With no prior knowledge on the machine, the gambler sequentially plays the machines in order to earn information on their expected rewards. Meanwhile, the gambler aims at maximizing his cumulated gains. This last conflict between the time spent on testing the machines and the time spent at playing the machines that seem to be the most profitable is known as Exploration Vs Exploitation dilemma. This figure illustrates side by side the MAB cycle and the CR cycle in OSA contexts. The mean objective is to show that both problems have similar models. In an OSA problem Exploration Vs Exploitation dilemma appears when a secondary user aims at maximizing his cumulated transmitted data parquets while earning more information on the availability of the frequency bands.	65
4.2	Communication Opportunities in Wireless Communication for Cognitive Radio. This figure borrowed from the survey [13] illustrates several possible communication dimensions that could be exploited by a SU within the OSA context. Whether SUs can exploit or not these opportunities depends on their abilities to detect them and to use the available resources while ensuring minimum interference with primary users.	66
4.3	Representation of a CA observing and accessing an RF environment.	70
4.4	A tabular version of a $\pi(i_t)$ policy using a UCB_1 algorithm for computing actions a_t	73
4.5	UCB based policies and dynamic configuration problem: simulation results. Figure on top plots the average cumulated reward as a function of the number of slots for the different UCB based policies. The figures on the bottom represent the number of times every configuration has been selected after a specific number of time slots. From the left to the right, 100 slots, 250 slots, 2500 slots and 5000 slots.	76
4.6	Description of the Dynamic Configuration Adaptation problem.	77
4.7	Slot representation for a radio equipment controlled by a CA. A slot is divided into 4 periods. During the first period , the CA chooses the next configuration. If the new configuration is different from the current one, a reconfiguration is carried out during the second period before communicating. If a reconfiguration is not needed, the CR equipment keeps the current configuration to communicate. At the end of every slot, the CA computes a reward that evaluates its performance during the communication process. It is assumed here that $\tau_1 + \tau_2 + \tau_4$ are small with respect τ_3	77
4.8	Percentage of times a UCB-based policy selects the optimal configuration.	79
4.9	UCB based policies and opportunistic spectrum access problem: simulation results. Figure on top plots the average cumulated reward as a function of the number of slots for the different UCB based policies. The figures on the bottom represent the evolution of the normalized average throughput achieved by these policies.	80

4.10	Percentage of time a UCB-based policy selects the optimal channel in case of perfect channel state information at every iteration t	81
4.11	Presented by J.-Y. Audibert and R. Munos at ICML 2011. Averaged empirical Regret plotted as a function of the exploration parameter α . It considers $N = 5$ machines and a time horizon of 1000 slots. The results suggest that for non infinite horizons, small value of α , that seem to contradict the limits imposed by the theory, can lead to better results than those predicted by the theory.	83
4.12	Percentage of time the UCB_1 -based CA selects the optimal channel under various sensing errors frameworks (over 10 available channels).	86
4.13	UCB_1 algorithm and Opportunistic Spectrum Access problem with sensing errors: regret simulation results.	87
4.14	Simulink model of the OSA framework. At the top, the OFDM transmitter simulates the primary network (many independent transmitters). The cognitive agent refers to the SU's decision making engine. The dashboard presents the main parameters regarding the current simulation.	89
4.15	Validation of the expected theoretical convergence to the optimal throughput (i.e., $(1 - \epsilon)(1 - \mu_{16}) = 0.891$ in this case).	91
4.16	Channel selection proportions: We observe the evolution of the channel selection process conducted by the SU. In this case, we consider 16 channels and 4000 slots. In this specific scenario, we can see that after 4000 iterations, the CA selected the optimal channel almost 65 % of the time. Moreover the 3 best channels are selected 80 % of the time.	91
5.1	Collaboration, Learning and Coordination in the case of Symmetric Networks: averaged regret. The simulation results show that both Hungarian algorithm and Round Robin based coordinators can efficiently learn to allocate the resources among the SUs. All curves are computed with $\alpha = 1.1$. Left Figure shows the impact of collaboration on the learning process in symmetric networks. Right curves compares learning mechanisms with both Hungarian coordination or Round Robbin coordination. We notice that their performance is quite similar.	110
5.2	Percentage of time the Hungarian algorithm based coordinator allocates the different secondary users to their respective optimal sets. The exploration parameter α is chosen equal to 1.1. This value is smaller than the minimum value suggested by the theory. We observe however that the algorithm remains consistent.	111
5.3	The network throughput over simulation time in a scenario with $K=4$ is shown in Figure 5.3(a). The network throughput as a function of the number of SUs (i.e. K) is shown in Figure 5.3(b).	112
6.1	A deterministic 3-armed bandit $MUCB(0.2)$ scenario	124

6.2	Channel selection process over time: a typical run. The dots between the values $[n; n + 1]$ ($Y - axis$) represents the time instants $t \in \mathbb{N}$ where channel n is selected by MUCB policies. This curve illustrates a typical run of MUCB policies. Thus we can see for different exploration coefficients α the selected channels. As predicted by Theorem 5, suboptimal channels are selected regularly on a logarithmic scale depending on their quality.	129
6.3	Average Regret Over 100 experiments: Illustration of Theorem 5. These curves confirm that the regret of MUCB policies grow as a logarithmic function of the number of iterations. Moreover, since $MUCB(1)$ and $MUCB(2)$ seem to respect this trend, these curves suggest that the imposed condition in Theorem 5, $\alpha > 4$, might be improvable.	130
C.1	Evolution des débits des principaux standards de communication sans-fil (Figure trouvé en ligne : http://3g4g.blogspot.fr/2008_04_01_archive.html). Des courbes similaires peuvent être trouvées dans le papier [2].	152
C.2	Cycle cognitif présenter par J.Mitola III [10].	156
D.1	Approximation de la fonction erreur (Résultats et figures issues du papier [143]. Les deux figures de droite et de gauche présentent les fonctions $\Delta_1(\cdot)$, $\Delta_2(\cdot)$, $\tilde{\Delta}_1(\cdot)$ et $\tilde{\Delta}_2(\cdot)$ (C.f. paragraphe 3.3.2). La figure de gauche montre ces fonctions pour un paramètre $M = 1000$, alors que la fonction de droite les montre pour un paramètre $M = 25000$. Nous observons que les approximations théoriques introduites dans la propriété 1, paragraphe 3.3.2- et les équations en haut de la figure ci-dessus- convergent bien vers les vraies valeurs lorsque M grandit.	163
D.2	Erreur Maximale en valeur absolue (Maximum Absolute Error). Sur cette figure, quatre courbes sont affichées: deux en pointillé et deux en trait plein. Les courbes en pointillé illustrent le taux de décroissance de l'amplitude maximale des fonctions $\Delta_1(\cdot)$ et $\Delta_2(\cdot)$. Tandis que, les deux autres courbes, affichent les valeurs théoriques approximées à l'aide des fonctions $\tilde{\Delta}_1(\cdot)$ et $\tilde{\Delta}_2(\cdot)$. Nous pouvons remarquer que les résultats théoriques développés à travers les équations 3.13 et 3.14 décrivent bien la réalité dans les deux cas.	164
D.3	Probabilité de détection (gauche) et $SNR-wall$ (droite). La figure de gauche montre l'impact de l'incertitude du bruit tel que défini dans ce chapitre sur les performances du détecteur d'énergie. Les courbes de droites montrent que les pertes de performances peuvent être prédites grâce à la nouvelle formule du $SNR-wall$. Remarquons que l'impact de l'incertitude, dans le sens introduit par [49] et utilisé par [50] n'apparaît pas sur ces courbes. En effet il est en général impossible suivant leur modèle d'imposer une fausse alarme prédéfinie. Par conséquent la comparaison serait biaisée.	167
E.1	Représentation de l'interaction d'un agent intelligent avec son environnement RF.	173
E.2	Pourcentage du temps passé à sélectionner le canal optimal à l'aide de l'algorithme UCB_1 en fonction des erreurs d'observations.	175

List of Tables

4.1	Summary of the equivalence between MAB paradigm and OSA related problems.	67
-----	---	----

Bibliography

- [1] International Telecommunication Union. <http://www.itu.int/ITU-D/ict/statistics/>.
- [2] G. Fettweis and E. Zimmermann. Ict energy consumption, trends and challenges. *The 11th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, 2008.
- [3] J. Wells. The future of cellular infrastructure. *Online. Open file at the url: <http://www.californiaconsultants.org/download.cfm/attachment/CNSV-0903-Wells.pdf>*.
- [4] V. Chandrasekhar and J. G. Andrews. Femtocell networks: A survey. *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59-67, September 2008.
- [5] M. S. Alouini and A. J. Goldsmith. Area spectral efficiency of cellular mobile radio systems. *IEEE Transactions On Vehicular Technology*, Vol. 48, No. 4, July 1999.
- [6] M. Dohler, R. W. Heath Jr, A. Lozano, B. C. Papadias, and R. A. Valenzuela. Is the phy layer dead? *IEEE Communications Magazine*, 4, 159-165, April 2011.
- [7] SDR Forum. Sdrf cognitive radio definitions. http://data.memberclicks.com/site/sdf/SDRF-06-R-0011-V1_0_0.pdf, November 2007.
- [8] J. Palicot. Radio engineering: From software radio to cognitive radio. *Wiley*, 2011.
- [9] J. Mitola and G.Q. Maguire. Cognitive radio: making software radios more personal. *Personal Communications, IEEE*, 6:13-18, August 1999.
- [10] J. Mitola. Cognitive radio: An integrated agent architecture for software defined radio. *PhD Thesis, Royal Inst. of Technology (KTH)*, 2000.
- [11] Federal Communications Commission. Spectrum policy task force report. http://www.fcc.gov/sptf/files/SEWGFfinalReport_1.pdf, November 2002.
- [12] S. Haykin. Cognitive radio: brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*, 23, no. 2:201-220, Feb 2005.
- [13] T. Yucek and H. Arslan. A survey of spectrum sensing algorithms for cognitive radio applications. *In IEEE Communications Surveys and Tutorials*, 11, no.1, 2009.

- [14] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan. Cooperative spectrum sensing in cognitive radio networks: A survey. *Physical Communication*, vol. 4, no. 1, pp. 40-62, 2011.
- [15] B. Wang, Y. Wu, and K. J. R. Liu. Game theory for cognitive radio networks: an overview. *Computer Networks* 54 (2010) 2537-2561.
- [16] C.J. Rieser. *Biologically Inspired Cognitive Radio Engine Model Utilizing Distributed Genetic Algorithms for Secure and Robust Wireless Communications and Networking*. PhD thesis, Virginia Tech, 2004.
- [17] J. Palicot, C. Moy, and R. Hachemani. Multilayer sensors for the sensorial radio bubble. *Physical Communication*, (2):151-165, May 2009.
- [18] J. Palicot, R. Hachemani, and C. Moy. La bulle sensorielle radio intelligente, in french, ree, no9. October 2007.
- [19] R. Hachemani, J. Palicot, and C. Moy. The "sensorial radio bubble" for cognitive radio terminals. In *Proceeding in URSI, The XXIX General Assembly of the International Union of Radio Science, Chicago (USA)*, August 2008.
- [20] L. Godard. Ph.d. dissertation: Modèle de gestion hiérarchique distribuée pour la reconfiguration et la prise de décision dans les Équipements de radio cognitive. http://hal.archives-ouvertes.fr/docs/00/35/53/52/PDF/these_loig_finale.pdf, 2008.
- [21] J.P. Delahaye, P. Leray, C. Moy, and J. Palicot. Anaging Dynamic Partial Reconfiguration on Heterogeneous SDR Platforms. *SDR Forum Technical Conference'05, Anaheim (USA)*, November 2005.
- [22] L. Godard, C. Moy, and J. Palicot. From a configuration management to a cognitive radio management of sdr systems. *Proceedings of the Second International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrowCom), Mykonos, Greece*, June 2006.
- [23] L. Godard, C. Moy, and J. Palicot. An executable meta-model of a hierarchical and distributed architecture management for the design of cognitive radio equipments. *Special issue on Cognitive Radio*, vol. 64, pp.463-482, number 7-8, August 2009.
- [24] C. Moy. Bio-inspired cognitive phones based on human nervous system. *3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL)*, 2010.
- [25] H. Wang, W. Jouini, L.S. Cardoso, A. Nafkha, J. Palicot, and M. Debbah. Blind bandwidth shape recognition for standard identification using usrp platforms and sdr4all tools. *Sixth Advanced International Conference on Telecommunications (AICT2010) Barcelona, Spain*, May 2010.
- [26] H. Wang, W. Jouini, L.S. Cardoso, A. Nafkha, J. Palicot, and M. Debbah. Blind standard identification with bandwidth shape and gi recognition using usrp platforms and sdr4all tools. *5th International Conference on Cognitive Radio Oriented Wireless Networks and Communications, CrownCom'10, Cannes, France*, June 2010.

- [27] S. Lecomte, W. Jouini, C. Moy, and P. Leray. A systemc radio-in-the-loop modeling for cognitive radio equipments. *Software Defined Radio Forum 2010, Washington DC, USA*, December.
- [28] W. Jouini, M. Lopez-Benitez, A. Nafkha, and J. Pérez-Romero. Joint learning-detection framework: an empirical analysis. *Proceedings of the Joint COST2100 and IC0902 Workshop on Cognitive Radio and Networking*, pages: 1-6, Bologna, Italy, November 2010.
- [29] X. Zhang, W. Jouini, P. Leray, and J. Palicot. Temperature-power consumption relationship and hot-spot migration for fpga-based system. In *Green Computing and Communications (GreenCom), 2010 IEEE/ACM Int'l Conference on Int'l Conference on Cyber, Physical and Social Computing (CPSCom)*, pages 392–397, dec. 2010.
- [30] S.J. Russell and P. Norvig. Definitions of software defined radio (sdr) and cognitive radio system (crs). *Artificial Intelligence: A Modern Approach (2nd ed.)*, Upper Saddle River, New Jersey: Prentice Hall, ISBN 0-13-790395-2, 2003.
- [31] F. K. Jondral. Software-defined radio - basics and evolution to cognitive radio. *EURASIP Journal on Wireless Communications and Networking*, vol. 2005, no. 3, p. 9, 2005.
- [32] T. W. Rondeau, D. Maldonado, D. Scaperoth, and C.W. Bostian. Cognitive radio formulation and implementation. *IEEE Proceedings CROWNCOM, Mykonos, Greece*, 2006.
- [33] T.W. Rondeau. *Application of Artificial Intelligence to Wireless Communications*. PhD thesis, Virginia Tech, 2006.
- [34] FCC. Facilitating opportunities for flexible, efficient and reliable spectrum use employing cognitive radio technologies. *Federal Communications Commission Spectrum Policy Task Force, Docket No. 03-108*, March 2005.
- [35] International Telecommunication Union. Definitions of software defined radio (sdr) and cognitive radio system (crs). *Report ITU-R SM.2152, SM Series, Spectrum management*, September 2009.
- [36] W. Jouini, D. Ernst, C. Moy, and J. Palicot. Multi-armed bandit based policies for cognitive radio's decision making issues. In *Proceedings of the 3rd international conference on Signals, Circuits and Systems (SCS)*, November 2009.
- [37] C. Moy. High-level design approach for the specification of cognitive radio equipments management APIs. *Journal of Network and System Management, Special Issue on Management Functionalities for Cognitive Wireless Networks and Systems*, vol. 18, number 1, March 2010.
- [38] W. Jouini. Seminar on learning for opportunistic spectrum access: a multi-armed bandit framework. <http://www.rennes.supelec.fr/ren/rd/scee/seminaire.html>, March 2011.

- [39] QinetiQ and Ofcom. Cognitive radio technology a study for ofcom, volume 1. <http://stakeholders.ofcom.org.uk/market-data-research/>, February 2007.
- [40] M. Gibson. Tv white space geolocation database. July 2010.
- [41] H.R. Karimi and Ofcom. Geolocation databases for white space devices in the uhf tv bands: Specification of maximum permitted emission levels. *IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, 2011.
- [42] Motorola. Tv white space position paper. fixed tv white space solutions for wireless isp network operators.
- [43] Office of Communication (Ofcom). Implementing geolocation. <http://stakeholders.ofcom.org.uk/binaries/consultations/geolocation/statement/statement>. September 2011.
- [44] M. A. McHenry et al. Spectrum occupancy measurements. Technical report, Shared Spectrum Company, Jan 2004 - Aug 2005. Available at: <http://www.sharedspectrum.com>.
- [45] R. I. C. Chiang, G. B. Rowe, and K. W. Sowerby. A quantitative analysis of spectral occupancy measurements for cognitive radio. In *Proceedings of the IEEE 65th Vehicular Technology Conference (VTC 2007 Spring)*, pages 3016–3020, 2007.
- [46] M. Wellens, J. Wu, and P. Mähönen. Evaluation of spectrum occupancy in indoor and outdoor scenario in the context of cognitive radio. In *Proceedings of the Second International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrowCom 2007)*, pages 1–8, 2007.
- [47] M. H. Islam et al. Spectrum survey in Singapore: Occupancy measurements and analyses. In *Proceedings of the 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom 2008)*, pages 1–7, May 2008.
- [48] M. López-Benítez, F. Casadevall, A. Umbert, J. Pérez-Romero, J. Palicot, C. Moy, and R. Hachemani. Spectral occupation measurements and blind standard recognition sensor for cognitive radio networks. In *Proceedings of the 4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom 2009)*, pages 1–9, 2009.
- [49] A. Sonnenschein and P.M. Fishman. Radiometric detection of spread-spectrum signals in noise of uncertain power. *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, pp. 654–660, July 1992.
- [50] R. Tandra and A. Sahai. SNR walls for signal detection. *IEEE Journal of Selected Topics in Signal Processing*, 2(1):4–17, Feb. 2008.
- [51] W. Jouini. Energy detection limits under log-normal approximated noise uncertainty. *IEEE Signal Processing Letters*, Volume 18, Issue 7, July 2011.

- [52] W. Jouini, C. Moy, and J. Palicot. On decision making for dynamic configuration adaptation problem in cognitive radio equipments: a multi-armed bandit based approach. *6th Karlsruhe Workshop on Software Radios, WSR'10, Karlsruhe, Germany*, March 2010.
- [53] J Mitola. *Cognitive radio architecture-the engineering foundations of radio xml*. Wiley-Blackwell, 2006.
- [54] J Mitola. The future of cognitive radio. <http://www.rennes.supelec.fr/ren/rd/scee/ftp/semin> May 2011.
- [55] N. Colson, A. Kountouris, A. Wautier, and L. Husson. Cognitive decision making process supervising the radio dynamic reconfiguration. In *Proceedings of Cognitive Radio Oriented Wireless Networks and Communications*, page 7, 2008.
- [56] F. Berggren, O. Queseth, J. Zander, B. Asp, C. Jönsson, P. Stenumgaard, N.Z. Kviselius, B. Thorngren, U. Landmark, and J. Wessel. Dynamic spectrum access. http://www.wireless.kth.se/projects/DSA/DSA_report_phase1.pdf, September 2004.
- [57] FCC. Spectrum policy task force, report of the spectrum efficiency working group, November 2002.
- [58] D. P. Reed. How wireless networks scale: the illusion of spectrum scarcity. <http://www.its.bldrdoc.gov/meetings/art/art02/slides02/speakers02.html>, March 2002.
- [59] P. Kolodzy. Spectrum policy, technology leading to new directions? *National Spectrum Managers Association*, May 2002.
- [60] D. P. Reed. Bits aren't bites: Constructing a "communications ether" that can grow and adapt, March 2003.
- [61] FCC. Promoting efficient use of spectrum through elimination of barriers to the development of secondary markets. *Report WT Docket No. 00-230*, October 2003.
- [62] P. Kolodzy. Spectrum policy task force, findings and recommendations. *International Symposium on Advanced Radio Technologies (ISART 2003)*, March 2003.
- [63] Q. Zhao and B. M. Sadler. A survey of dynamic spectrum access: signal processing, networking, and regulatory policy. In *in IEEE Signal Processing Magazine*, pages 79–89, 2007.
- [64] M. M. Buddhikot. Understanding dynamic spectrum access: Models, taxonomy and challenges. *IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, Dublin, April 2007.
- [65] Q. Zhao, L. Tong, and A. Swami. Decentralized cognitive mac for dynamic spectrum access. *IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, 2005.

- [66] K. Liu and Q. Zhao. Channel probing for opportunistic access with multi-channel sensing. *IEEE Asilomar Conference on Signals, Systems, and Computers*, October 2008.
- [67] W. Jouini, D. Ernst, C. Moy, and J. Palicot. Upper confidence bound based decision making strategies and dynamic spectrum access. *Proceedings of the 2010 IEEE International Conference on Communications (ICC)*, May 2010.
- [68] W. Jouini, C. Moy, and J. Palicot. Upper confidence bound algorithm for opportunistic spectrum access with sensing errors. *6th International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications, Osaka, Japan*, June 2011.
- [69] A. Anandkumar, N. Michael, and A. Tang. Opportunistic spectrum access with multiple users: Learning under competition. March 2010.
- [70] K. Liu, Q. Zhao, and B. Krishnamachari. Distributed learning under imperfect sensing in cognitive radio networks. *IEEE Transactions on Signal Processing*, Vol. 58, No. 11, November 2010.
- [71] Y. Gai, B. Krishnamachari, and R. Jain. Learning multiuser channel allocations in cognitive radio networks: A combinatorial multi-armed bandit formulation. *IEEE Symposium on International Dynamic Spectrum Access Networks (DySPAN)*, 2010.
- [72] W. Di, W. Feng, and Y. Shengyao. Cognitive radio decision engine based on priori knowledge. *3rd International Symposium on Parallel Architectures, Algorithms and Programming*, 2011.
- [73] W. Jouini, C. Moy, and J. Palicot. Decision making for cognitive radio equipment: analysis of the first 10 years of exploration. *EURASIP Journal on Wireless Communications and Networking*, 2012(1):26, 2012.
- [74] T. Bray, J. Paoli, C.M. Sperberg-McQueen, E. Maler, F. Yergeau, and J. Cowan. Extensible markup language (xml) 1.1., February 2004.
- [75] DARPA XG Working Group. The XG vision. request for comments. *BBN Technologies, Cambridge MA, USA, Tech. Rep. Version 2.0*, January 2004.
- [76] L. Berlemann, S. Mangold, and B. H. Walke. Policy-based reasoning for spectrum sharing in radio networks. *In Proceedings of IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), Baltimore, MD, USA*, November 2005.
- [77] B. Fette, M. M. Kokar, and M. Cummings. Next-generation design issues in communications. *Portable Design Magazine*, No. 3:20-24, 2008.
- [78] S. Li, M. M. Kokar, and D. Brady. Developing an ontology for the cognitive radio: issues and decisions. *PSDR Forum Technical Conference*, December 2009.
- [79] S. Li, M. M. Kokar, D. Brady, and J. Moskal. Collaborative adaptation of cognitive radio parameters using ontology and policy approach. *In Software Defined Radio Technical Conference, SDR'10. SDRF.*, 2010.

- [80] WIF Forum MLM Working Group. Description of cognitive radio ontology v.1.0, 2010.
- [81] <http://www.ettus.com>.
- [82] E. Blossom. Exploring GNU Radio, November 2004. Available at: <http://www.gnu.org/software/gnuradio/doc/exploring-gnuradio.html>.
- [83] M. Duque-Anton, D. Kunz, and B. Ruber. Channel assignment for cellular radio using simulated annealing. *IEEE Transactions on Vehicular Technology, Volume: 42 Issue: 1*.
- [84] T. Chen, H. Zhang, and Z. Zhou. Swarm intelligence based dynamic control channel assignment in cogmesh. *Proc. the IEEE ICC 2008 (IEEE CoCoNet'08 Workshop), Beijing, China*, May 2008.
- [85] P. Di Lorenzo and S. Barbarossa. Distributed resource allocation in cognitive radio systems based on social foraging swarms. *The 11th IEEE International Workshop on signal processing advances in wireless communications, Marrakech, Morocco*, June 2010.
- [86] B. Atakan and O. B. Akan. Biologically-inspired spectrum sharing in cognitive radio networks. *In Proc. IEEE Wireless Communications and Networking Conference, WCNC, 2007*.
- [87] A. Amanna and J.H. Reed. Survey of cognitive radio architectures. *Proceedings of IEEE SoutheastCon 2010 (SoutheastCon)*, March 2010.
- [88] N. Baldo and M. Zorzi. Learning and adaptation in cognitive radios using neural networks. *5th IEEE Consumer Communications and Networking Conference, CCNC.*, 2008.
- [89] N. Baldo and M. Zorzi. Fuzzy logic for cross-layer optimization in cognitive radio networks. *IEEE Consumer Communications and Networking Conference*, January 2007.
- [90] Charles Clancy, Joe Hecker, and Erich Stuntebeck. Applications of machine learning to cognitive radio networks. *IEEE Wireless Communications Magazine*, 14, 2007.
- [91] N. Kasabov. ECOS : Evolving connectionist systems and the eco learning paradigm. *International Conference on Neural Information Processing, Kitakyushu, Japan*, Oct. 1998.
- [92] N. Kasabov. Evolving connectionist systems. the knowledge engineering approach. *2nd ed. New York : Springer*, 2007.
- [93] T. Weingart, D. Sicker, and D. Grunwald. A statistical method for reconfiguration of cognitive radios. *IEEE Wireless Commun. Mag., vol. 14, no. 4, pp. 34-40*, August 2007.
- [94] H. Urkowitz. Energy detection of unknown deterministic signals. *Proceedings of the IEEE, vol. 55, no. 4, pp. 523-531*, 1967.

- [95] H. V. Poor. An introduction to signal detection and estimation. *New York, NY: Springer-verlag*, 1994.
- [96] Y. Zeng, Y. Liang, A. T. Hoang, and R. Zhang. A review on spectrum sensing for cognitive radio: Challenges and solutions.
- [97] W. Lin and Q. Zhang. A design of energy detector in cognitive radio under noise uncertainty. November 2008.
- [98] M. Abramowitz and I. A. Stegun. Handbook of mathematical functions with formulas, graphs, and mathematical tables. In *New York: Dover Publications*, 1972.
- [99] H. Robbins. Some aspects of the sequential design of experiments. *Bulletin of American Mathematical Society*, 58:527–535, 1952.
- [100] T.L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6:4–22, 1985.
- [101] R. Agrawal. Sample mean based index policies with $O(\log(n))$ regret for the multi-armed bandit problem. *Advances in Applied Probability*, 27:1054–1078, 1995.
- [102] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite time analysis of multi-armed bandit problems. *Machine learning*, 47(2/3):235–256, 2002.
- [103] J.-Y. Audibert, R. Munos, and C. Szepesvári. Tuning bandit algorithms in stochastic environments. In *Proceedings of the 18th international conference on Algorithmic Learning Theory*, 2007.
- [104] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. The nonstochastic multi-armed bandit problem. *SIAM J. Computing*, 32, pages 44–77, 2002.
- [105] C. C. Wang, R. K. Sanjeev, and H. V. Poor. Bandit algorithms for tree search. *IEEE Transactions on Automatic Control*, 50:338–355, 2005.
- [106] P. Auer and R. Ortner. Logarithmic online regret bounds for undiscounted reinforcement learning. *NIPS (pp. 4956)*, year = 2006.
- [107] N. Cesa-Bianchi and G. Lugosi. Prediction, learning, and games. *Cambridge University Press, New York, NY, USA*, 2006.
- [108] P.-A. Coquelin and R. Munos. Bandit algorithms for tree search. In *Uncertainty in Artificial Intelligence*, 2007.
- [109] S. Pandey, D. Agarwal, D. Chakrabarti, and Josifovski V. Bandits for taxonomies: A model-based approach. In *Proceedings of the Seventh SIAM International Conference on Data Mining*, 2007.
- [110] S. Pandey, D. Agarwal, D. Chakrabarti, and Josifovski V. Multi-armed bandit problems with dependent arms. In *ICML '07: Proceedings of the 24th international conference on Machine learning, pages 721-728, New York, NY, USA*, 2007.
- [111] J.F. Hren and R. Munos. Optimistic planning in deterministic systems. *European Workshop on Reinforcement Learning*, pages 151–164, 2008.

- [112] A. Slivkins and E. Upfal. Adapting to a changing environment: the brownian restless bandits. *In COLT, pages 343-354. Omnipress, 2008.*
- [113] A. Garivier and E. Moulines. On upper-confidence bound policies for non-stationary bandit problems. *Arxiv preprint arXiv:0805.3415, 2008.*
- [114] J. D. Abernethy, P. L. Bartlett, A. Rakhlin, and A. Tewari. Optimal strategies and minimax lower bounds for online convex games. *In Servedio and Zhang, 2008.*
- [115] J. Poland. Nonstochastic bandits: Countable decision set, unbounded costs and reactive environments. *Theoretical Computer Science, 397(1-3):77-93, 2008.*
- [116] J. Y. Audibert and S. Bubeck. Minimax policies for adversarial and stochastic bandits. *Proceedings of the 22nd Annual Conference on Learning Theory (COLT 2009) (eds. S. Dasgupta and A. Klivans), pp. 217-226, Omni-press, Eastbourne, United Kingdom, year = 2009.*
- [117] N. Cesa-Bianchi and G. Lugosi. Combinatorial bandits. *In Dasgupta and Klivans, 2009.*
- [118] J. Y. Audibert and S. Bubeck. Regret bounds and minimax policies under partial monitoring. *Journal of Machine Learning Research, 11:2785-2836, Dec.*
- [119] J. Honda and A. Takemura. An asymptotically optimal bandit algorithm for bounded support models. *In Kalai and Mohri, pages 67-79, 2010.*
- [120] J. Honda and A. Takemura. An asymptotically optimal policy for finite support models in the multiarmed bandit problem. *arXiv:0905.2776, 2010.*
- [121] T. Jaksch, R. Ortner, and P. Auer. Near-optimal regret bounds for reinforcement learning. *Journal of Machine Learning Research, 99, 2010.*
- [122] T. Lu, D. Pal, and M. Pal. Contextual multi-armed bandits. *In Yee Whye Teh and Mike Titterton, editors, Proceedings of the 13th international conference on Artificial Intelligence and Statistics, volume 9, pages 485-492, 2010.*
- [123] P. Rusmevichientong and J. N. Tsitsiklis. Linearly parameterized bandits. *Math. Oper. Res., 35:395-411, May, 2010.*
- [124] O.-A. Maillard and R. Munos. Multi-armed bandit problems with dependent arms. *In Proceedings of the 2010 European Conference on Machine Learning and Knowledge Discovery in Databases: Part II, ECML PKDD'10, pages 305-320, Berlin, Heidelberg, 2010.*
- [125] O.-A. Maillard and R. Munos. Adaptive bandits: Towards the best history-dependent strategy. *In Proceedings of the 14th international conference on Artificial Intelligence and Statistics, volume 15 of JMLR W-CP, 2011.*
- [126] O.-A. Maillard, R. Munos, and G. Stoltz. Finite-time analysis of multi-armed bandits problems with kullback-leibler divergences. *In Proceedings of the 24th annual Conference On Learning Theory, COLT '11, 2011.*

- [127] A. Slivkins. Contextual bandits with similarity information. <http://arxiv.org/abs/0907.3986>, 2011.
- [128] S. Filippi, O. Cappé, and A. Garivier. Optimism in reinforcement learning and kullback-leibler divergence. <http://arxiv.org/pdf/1004.5229.pdf>, 2011.
- [129] Gittins J. C. Multi-armed bandit allocation indices. *Wiley-Interscience Series in Systems and Optimization*. John Wiley - Sons Ltd., Chichester, 1989.
- [130] J. Y. Audibert, R. Munos, and C. Szepesvari. Exploration-exploitation trade-off using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410:1876–1902, 2009.
- [131] L. Lai, H.E. Gamal, H.J. Jiang, and V. Poor. Cognitive medium access: Exploration, exploitation and competition. [Online]. Available: <http://arxiv.org/abs/0710.1385>.
- [132] F. Maes, L. Wehenkel, and D. Ernst. Learning to play k-armed bandit problems. In *Proceedings of the 4th International Conference on Agents and Artificial Intelligence (ICAART 2012)*, Vilamoura, Algarve, Portugal, February.
- [133] W. Jouini, R. Bollenbach, M. Guillet, C. Moy, and A. Nafkha. Reinforcement learning application scenario for opportunistic spectrum access. In *Circuits and Systems (MWSCAS), 2011 IEEE 54th International Midwest Symposium on*, pages 1–4, Aug. 2011.
- [134] E. Hossain and V. K. Bhargava. Cognitive wireless communication networks. Springer, 2007.
- [135] P. Steenkiste, D. Sicker, G. Minden, and D. Raychaudhuri. Future directions in cognitive radio network research. *NSF Workshop Report*, March 2007.
- [136] HW Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly - Wiley Online Library*, 1955.
- [137] K. Liu and Q. Zhao. Distributed learning in cognitive radio networks: Multi-armed bandit with distributed multiple players. 2010.
- [138] M. Di Felice, K.R. Chowdhury, and L. Bononi. Learning with the bandit: A cooperative spectrum selection scheme for cognitive radio networks. *on Proceedings of the 2011 IEEE Global Communications Conference (Globecom 2011)*, Houston, TX, USA, December 2011.
- [139] Y. Gai, B. Krishnamachari, and R. Jain. Combinatorial network optimization with unknown variables: Multi-armed bandits with linear rewards. *IEEE Symposium on International Dynamic Spectrum Access Networks (DySPAN)*, 2010.
- [140] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* 58 (301), pages 13–30, March 1963.
- [141] L. P. Kaelbling. *Learning In Embedded Systems*. PhD thesis, Stanford University, Department of Computer Science, 1990.

-
- [142] H. Chernoff. A measure of asymptotic efficiency fo tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, pages 493–507, 1952.
- [143] W. Jouini, D. Le-Guenec, C. Moy, and J. Palicot. Log-normal approximation of chi-square distributions for signal processing. *XXX URSI General Assembly and Scientific Symposium of International Union of Radio Science, Istanbul*, 2011.
- [144] L. Lai, H.E. Gamal, H.J. Jiang, and V. Poor. Cognitive medium access: Exploration, exploitation and competition. *IEEE Transactions on Mobile Computing*, vol. 10, No. 2, Feb 2011.
- [145] W. Jouini, C. Moy, and J. Palicot. Apprentissage pour l'accès opportuniste au spectre: Prise en compte des erreurs d'observation. *XXIII Colloque GRETSI Bordeaux, Septembre*, 2011.