



Departament de Teoria
del Senyal i Comunicacions



UNIVERSITAT POLITÈCNICA DE CATALUNYA



Self-organized Femtocells: a Time Difference Learning Approach

PhD Thesis Dissertation

by

Ana María Galindo Serrano

Submitted to the Universitat Politècnica de Catalunya (UPC)
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Barcelona, May 2012

Supervised by: Dr. Lorenza Giupponi
Tutor: Prof. Ana Isabel Pérez-Neira

PhD program on Signal Theory and Communications

A ti madre,

Abstract

The use model of mobile networks has drastically changed in recent years. Next generation devices and new applications have made the availability of high quality wireless data everywhere a necessity for mobile users. Thus, cellular networks, which were initially designed for voice and low data rate transmissions, must be highly improved in terms of coverage and capacity. Networks that include smart entities and functionalities, and that allow to fulfil all the mobile networks' new requirements are called heterogeneous networks. The gains introduced by these networks are basically due to the reduction of the distance between transmitters and receivers, which increases the network capacity per unit area, the ubiquitous coverage and the spectral efficiency. One key component in heterogeneous networks is femtocells. Femtocells are low range, low power mobile base stations deployed by the end consumers, which underlay the macrocell system and provide a solution to the problem of indoor coverage for mobile communications. Femtocells drop off the macrocell load and, therefore, macrocells can devote their resources exclusively to outdoor and mobile communications. Furthermore, the energy consumption decreases significantly because femtocells have very low transmission powers and are active only when needed. Femtocells can reuse the radio spectrum and, thereby, they allow increasing the spectral efficiency. Moreover, under appropriate algorithms for interference control, they give a viable alternative to the problem of spectrum static allocation.

In the case of femtocells reusing the spectrum, it must be guaranteed that the interference they generate does not affect the performance of the underlying macrocellular system. To this end, we propose to model the femtocell network as a decentralized system and to introduce a learning algorithm in each femto node, providing self-organization capabilities, that perfectly fit with the femtocells deployment pattern. We thus introduce a multiagent learning algorithm that performs the radio resource management in the femtocell system, so that femtocells control the interference they generate at macrocell in a decentralized and uncoordinated manner. Femtocells, then, are able to maintain their interference under a desirable threshold as a function of the environmental situation they perceive.

In distributed systems, learning can be a long process. For this reason, we introduce a new cooperative method, known as docitive algorithm, where agents exchange information that allow them to accelerate their learning process and increase its precision. We also present a learning technique based on Fuzzy Inference Systems, which allows femtocells to represent the environment they perceive and the actions they can perform in a continuous way. In this way, agents have more accurate behaviors and better adapt to the environmental conditions. Besides, we extend the learning method to partial observable environments in order to provide a 3GPP standard compliant solution, which does not rely on the existence of an X2 interface between macro and femto nodes. This means that agents do not have a complete representation of the surrounding environment and do not receive any feedback from the macro network. Finally,

since the proposed solutions are software-based self-optimization algorithms to be embedded in integrated circuits, we present a study regarding their implementation requirements in terms of computation and memory demands, in order to determine if they fit in state of the art communication processors.

Resumen

El modelo de uso de las redes móviles ha cambiado drásticamente en los últimos años. La aparición y rápida adopción de dispositivos de última generación y, con ellos, nuevas y múltiples aplicaciones, ha convertido en una necesidad para los usuarios móviles la disponibilidad de servicios de datos inalámbricos de alta calidad en todo momento y lugar. Por tanto, las redes móviles, originalmente diseñadas para la transmisión de voz y bajas tasas de datos, han de ser mejoradas en términos de cobertura y capacidad. Las redes que incluyen nuevas e inteligentes entidades y funcionalidades, las cuales permiten cumplir con los requisitos anteriormente mencionados, son denominadas redes heterogéneas. Las ganancias introducidas por estas redes son debidas a la reducción de la distancia entre el transmisor y el receptor, aumentando así la capacidad de la red por unidad de área, la cobertura total y la eficiencia espectral. Un componente clave en las redes heterogéneas son las femtoceldas, las cuales son estaciones base de comunicaciones móviles de bajo rango de cobertura y potencia de transmisión, instaladas por los usuarios. Éstas estaciones base se ubican de forma subyacente respecto a las macroceldas y proporcionan una solución para los problemas de cobertura de servicios móviles en interiores. Las femtoceldas disminuyen la carga de servicios que deben proporcionar las macroceldas, por tanto éstas pueden dedicar sus recursos exclusivamente a las comunicaciones exteriores. Por otra parte, el consumo de energía total del sistema disminuye notablemente ya que la potencia de transmisión de las femtoceldas es muy baja y sólo están activas cuando son necesarias. Las femtoceldas pueden reutilizar el espectro radioeléctrico, por lo tanto, permiten aumentar la eficiencia espectral y, con el uso de algoritmos apropiados de control de interferencia, brindan una alternativa viable al problema de la asignación estática del espectro.

En el caso particular en el cual las femtoceldas reutilizan el espectro, se debe garantizar que la interferencia generada por ellas no afecte el rendimiento del sistema macrocelular subyacente. Para ello, se propone modelar la red de femtoceldas como un sistema descentralizado e introducir un algoritmo de aprendizaje en cada femtocelda, brindándoles, de esta forma, capacidad de auto organización, lo cual encaja perfectamente con el modelo de despliegue de las femtoceldas. Con este fin, se introduce un algoritmo de aprendizaje multiagente para realizar la gestión de recursos radio, de modo que las femtoceldas controlen la interferencia que generan a los usuarios asociados a las macroceldas de manera descentralizada y sin coordinación. Las femtoceldas, por tanto, son capaces de mantener su interferencia bajo un umbral en función de la situación del entorno que perciben.

El aprendizaje en sistemas distribuidos lleva tiempo, razón por la cual se introduce un nuevo método de cooperación, conocido como algoritmo “docitive”. Éste algoritmo contempla el intercambio de información entre agentes, lo que les permite acelerar el proceso de aprendizaje y aumentar su precisión. También se presenta una técnica de aprendizaje basado en sistemas de inferencia difusos, el cual permite representar el entorno percibido por los agentes y las acciones

que estos pueden realizar, de forma continua. De este modo, los agentes tienen comportamientos más precisos y una mayor capacidad adaptativa. Además, se amplía el método de aprendizaje para entornos con observaciones parciales, con el fin de proporcionar una solución compatible con los estándares 3GPP, que no dependa de la existencia de una interfaz X2 entre macroceldas y femtoceldas. Por último, ya que las soluciones propuestas son algoritmos de auto optimización para ser incorporados en circuitos integrados, se presenta un estudio con respecto a sus requisitos en cuanto a exigencias de cálculo y de memoria, a fin de determinar si se ajustan a los procesadores de comunicación actuales.

Agradecimientos

Durante estos años en los que he realizado esta tesis, indudablemente he crecido como persona y profesional, lo cual ha sido posible gracias al buen ambiente de nuestro CTTC y las personas que lo componen. Quisiera agradecer especialmente a todos esos amigos/compis que han contribuido a que nuestro lugar de trabajo tenga ese toque divertido y amigable que ha hecho menos duro este largo proceso. En especial agradezco a todos mis compañeros de despacho (tanto a los del 108 en su momento, como a todos los que han pasado y ahora están en el 109), el mejor despacho nunca visto!

Vivir en Barcelona ha sido una experiencia maravillosa, llena de enriquecedores momentos. Aquí tengo una gran familia, mis amigos, que me han apoyado en todo momento, dándome ánimos cuando creía no poder más, dándome alegrías y risas, compañía y cariño. Sobre todo, gracias por la tranquilidad que me ha dado el saber que siempre puedo contar con ustedes y por aguantar estoicamente mis malos momentos.

Madre, gracias por estar ahí siempre para mí, a pesar de no entender a qué me dedico y los muchos kilómetros que nos separan. Nunca olvidaré tu famosa frase: “si por lo menos yo supiera lo que haces para poder ayudarte!”. No ha sido necesario, tu amor incondicional ha sido el mejor apoyo que se puede pedir. Tita Leli, esta tesis te la dedico con mucho amor, siempre estuviste ahí para mí.

Lorenza, sin ti la realización de esta tesis no habría sido posible. Cuando empezamos hicimos un trato de esfuerzo y aprendizaje conjunto, creo que el balance general ha sido positivo, tanto en el aspecto profesional como en el aspecto personal. Muchas gracias por tu paciencia, tiempo y dedicación, especialmente en estos últimos meses en los que has tenido que sacrificar parte de tu tiempo con Leo para revisar la tesis.

To all the people with whom I have worked during these years, thank you very much to all of you!. Mischa, I learned very much about the research world thanks to you, it was really interesting. Eitan, you showed me a different way to see the problems and you gave me the opportunity to meet amazing people and places, Thank you!. Gunther, thank you for your help with the standards. Marc, gracias por tus explicaciones y ayuda sobre la implementación de algoritmos en procesadores. Polin, ha sido, es y espero que siga siendo un placer trabajar contigo, creo que ha sido una experiencia muy positiva e instructiva para ambos. David, qué sería de mí sin ti? gracias por tu ayuda y paciencia infinita. Jaime, desentrañar contigo los misterios de LTE ha sido muy productivo y al mismo tiempo divertido. Andrea, estos años de trabajo y aventuras compartidos contigo definitivamente me han cambiado, gracias por todo, especialmente por tu paciencia.

Finalmente me gustaría agradecer a David, Paolo, Pol, Jaime, Mischa, Jessica, Chantal,

Toni y Lorenza, claro, por su ayuda con la revisión de la tesis. Sus consejos y correcciones han contribuido enormemente a que el trabajo que presento aquí sea más claro, organizado y de mejor calidad.

Ana M.

Mayo 2012.

List of Publications

Journal articles

- A. Galindo-Serrano and L. Giupponi, “Managing Femto to Macro Interference without X2 Interface Support Through POMDP”, submitted to *Mobile Networks and Applications (MONET) Journal. Special Issue on Cooperative and Networked Femtocells*.
- A. Galindo-Serrano and L. Giupponi, “Designing an Online Learning Method for Interference Control”, submitted to *Dynamic Games and Applications Journal*.
- A. Galindo-Serrano and L. Giupponi, “Q-learning Algorithms for Interference Management in Femtocell Networks”, submitted to *EURASIP Journal on Wireless Communications and Networking*.
- L. Giupponi, A. Galindo-Serrano, P. Blasco and M. Dohler, “Docitive Networks - An Emerging Paradigm for Dynamic Spectrum Management”, *IEEE Wireless Communications Magazine*, vol. 17, no. 4, pp. 47–54, Aug. 2010.
- L. Giupponi, A. Galindo-Serrano and M. Dohler, “From Cognition To Docition: The Teaching Radio Paradigm For Distributed & Autonomous Deployments”, *Computer Communications*, Elsevier, Aug. 2010.
- A. Galindo-Serrano and L. Giupponi, “Distributed Q-learning for Aggregated Interference Control in Cognitive Radio Networks”, *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1823–1834, May 2010.
- M. Dohler, L. Giupponi, A. Galindo-Serrano and P. Blasco, “Docitive Networks: A Novel Framework Beyond Cognition”, *IEEE Communications Society, Multimedia Communications TC, E-Letter*, January 2010.

Conference articles

- A. Galindo-Serrano, E. Altman and L. Giupponi, “Equilibrium Selection in Interference Management Non-Cooperative Games in Femtocell Networks”, submitted to *6th International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS 2012)*.
- A. Galindo-Serrano and L. Giupponi, “Managing Femto-to-Macro Interference without X2 Interface Support”, submitted to *23rd Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2012)*.
- A. Galindo-Serrano, L. Giupponi and M. Majoral, “On Implementation Requirements and Performances of Q-Learning for Self-Organized Femtocells”, in *Proceedings of the IEEE Global Communications Conference (IEEE GLOBECOM 2011), second Workshop on Femto cell Networks (FEMnet)*, 5-9 December, 2011, Houston, USA.
- A. Galindo-Serrano, L. Giupponi, “Femtocell Systems with Self Organization Capabilities”, in *Proceedings of International Conference on NETWORK Games, CONTROL and OPTimization (NetGCooP 2011)*, 12-14 October, 2011, Paris, France.
- M. Simsek, A. Czylik, A. Galindo-Serrano, L. Giupponi, “Improved Decentralized Q-learning Algorithm for Interference Reduction in LTE-femtocells”, in *Proceedings of IEEE Wireless Advanced 2011*, 20-22 June 2011, London, UK.
- A. Galindo-Serrano, L. Giupponi, G. Auer, “Distributed Femto-to-Macro Interference Management in Multiuser OFDMA Networks”, in *Proceedings of IEEE 73rd Vehicular Technology Conference (VTC2011-Spring), Workshop on Broadband Femtocell Technologies*, 15-18 May, 2011, Budapest, Hungary.
- A. Galindo-Serrano, L. Giupponi, “Downlink Femto-to-Macro Interference Management based on Fuzzy Q-Learning”, in *Proceedings of the Third IEEE International workshop on Indoor and Outdoor Femto Cells (IOFC’ 2011)*, May 13, 2011, Princeton, USA. **BEST PAPER AWARD**
- A. Galindo-Serrano, L. Giupponi and M. Dohler, “Cognition and Docition in OFDMA-Based Femtocell Networks”, in *Proceedings of the Global Telecommunications Conference, 2010 (IEEE GLOBECOM 2010)*, Dec. 2010, Miami, USA.
- P. Blasco, L. Giupponi, A. Galindo-Serrano and M. Dohler, “Energy Benefits of Cooperative Docitive over Cognitive Networks”, in *Proceedings of the 3rd European Wireless Technology Conference 2010 in the European Microwave Week*, Sept 26 - Oct. 1, 2010, Paris, France.
- A. Galindo-Serrano, L. Giupponi and M. Dohler, “BeFEMTO’s Self-Organized and Docitive Femtocells”, in *Proceedings of Future Network and MobileSummit 2010 Conference*, 16-18 June 2010, Florence, Italy.
- A. Galindo-Serrano, L. Giupponi, P. Blasco and M. Dohler, “Learning from Experts in Cognitive Radio Networks: The Docitive Paradigm”, in *Proceedings of the 5th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM 2010)*, 9-11 June 2010, Cannes, France.

- P. Blasco, L. Giupponi, A. Galindo-Serrano, M. Dohler, “Aggressive Joint Access & Backhaul Design For Distributed-Cognition 1Gbps/km² System Architecture”, in *Proceedings of 8th International Conference on Wired/Wireless Internet Communications (WWIC 2010)*, 1-3 June, 2010, Lulea, Sweden.
- A. Galindo-Serrano and L. Giupponi, “Distributed Q-learning for Interference Control in OFDMA-based Femtocell Networks”, in *Proceedings of IEEE 71st Vehicular Technology Conference (VTC2010-Spring)*, 16-19 May 2010, Taipei, Taiwan.
- A. Galindo-Serrano and L. Giupponi, “Decentralized Q-learning for Aggregated Interference Control in Completely and Partially Observable Cognitive Radio Networks”, in *Proceedings of IEEE Consumer Communications & Networking Conference (IEEE CCNC 2010)*, 9-12 January 2010, Las Vegas, USA. **BEST CONFERENCE PAPER AWARD.**
- A. Galindo-Serrano, L. Giupponi, “Aggregated Interference Control for Cognitive Radio Networks based on Multi-agent Learning”, in *Proceedings of the 4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM 2009)*, 22-24 June 2009, Hannover, Germany.

Events

- Co-chair of the “Second Workshop on Cooperative Heterogeneous Networks (coHetNet)”, to be held in conjunction with ICCCN 2012 July 30-August 2, 2012 Munich, Germany.
- “Self-organization in distributed systems: A practical application in femtocell networks”, Invited talk at 5th International Workshop on Femtocells (and Hetnets) 2012, King’s college London, 13th-14th of February, 2012.
- “Cognitive & Docitive Femtocell Networks”, Invited talk at the BeFEMTO Femtocell Winter School, 6-10 February 2012, Barcelona, Spain.

Other contributions

- Contribution to “D2.3: The BeFEMTO System Concept and its Performance,” EU FP7-ICT BeFEMTO project, June 2012.
- Contribution to “D4.4: Integrated SON techniques for femtocells radio access,” EU FP7-ICT BeFEMTO project, May 2012.
- Contribution to “Small Cell Deployments: Recent Advances and Research Challenges (A 2012 Update from the 5th Intl’ Workshop on Femtocells)”, Small Cell Forum white paper, to appear.

Contents

1	Introduction	1
1.1	Current situation overview	1
1.2	Problem statement	5
1.3	Objectives	8
1.4	State of the art	10
1.5	Outline of the thesis	14
	Bibliography	20
2	Scenarios and Simulation Parameters	27
2.1	3GPP scenarios	27
2.1.1	3GPP suburban modeling	28
2.1.2	3GPP dense-urban modeling	28
2.2	HeNB networks deployment models proposed by the BeFEMTO project	30
2.3	System model and simulation scenarios	31
2.3.1	System model	31
2.3.2	Single-cell scenario	33
2.3.3	Multicell scenario	33
2.3.4	Simulation parameters	34
2.4	Functional architecture	35
2.5	Conclusions	39
	Bibliography	41
3	Multiagent time difference methods: study and design for interference control	43
3.1	ML overview	45
3.1.1	Learning in single-agent systems	46

3.1.2	Learning for multiagent systems	49
3.2	TD learning methods	51
3.2.1	Study of TD Q-learning and Sarsa methods	53
3.3	Q-learning and Sarsa for interference control	55
3.3.1	Q-learning and Sarsa comparison	57
3.3.2	Parameters selection for Q-learning algorithm	59
3.4	Conclusions	63
	Bibliography	65
4	Multiagent Q-learning for interference control	69
4.1	Learning algorithm details	70
4.1.1	Learning design: case study 1	70
4.1.2	Learning design: case study 2	72
4.2	Simulation results	73
4.2.1	Single-cell scenario results	73
4.2.2	Multicell scenario results	74
4.3	Multiuser scheduling support	77
4.3.1	Practical implementation in 3GPP LTE	79
4.4	Conclusions	82
	Bibliography	83
5	Docition: a cooperative approach	85
5.1	Docitive cycle	86
5.2	Emerging docitive algorithms	88
5.3	Learning and teaching techniques	90
5.3.1	Simulation results	90
5.4	Conclusions	94
	Bibliography	95
6	Interference management based on Fuzzy Q-learning	97
6.1	Fuzzy Inference Systems	99
6.2	Fuzzy Q-learning	101
6.2.1	FQL-based interference management	103

6.3	Simulation results	104
6.3.1	FQL-PMFB results for the single-cell scenario	105
6.3.2	FQL-PMFB results for the multicell scenario	108
6.4	Conclusions	111
	Bibliography	113
7	Interference management without X2 interface support based on Partial Observable Markov Decision Process	115
7.1	Proposed learning methodology for partially observable environments	116
7.2	Spatial characterization of interference in femtocell networks	118
7.3	Q-learning in partially observable environments	121
7.4	Simulation results	123
7.4.1	Single-cell scenario results for POMDP	124
7.4.2	Multicell scenario results for POMDP	126
7.5	Conclusion	130
	Bibliography	131
8	Memory and computational requirements for a practical implementation	133
8.1	Integrated circuit architectures	133
8.2	Practical implementation in state of the art processors	134
8.2.1	Memory requirements of expert knowledge	136
8.2.2	Computational requirements	137
8.2.3	Comparison of neural networks and lookup table representation mechanisms for Q-learning	139
8.3	Conclusions	141
	Bibliography	142
9	Conclusions and Future Work	143
9.1	Summary of results	143
9.2	Future work	145
	Bibliography	149
A	Notation	151
B	Acronyms and Definitions	155

List of Figures

1.1	General interpretation of the proposed learning-based solution.	9
1.2	Thesis main contributions.	15
2.1	Suburban modeling.	28
2.2	3GPP Dual Stripe femtocell block model.	29
2.3	5 × 5 apartment grid	30
2.4	Single-cell system proposed layout.	33
2.5	Multicell system proposed layout.	34
2.6	LTE system architecture including enhancement for learning support.	37
3.1	Learner-environment interaction.	47
3.2	Taxonomy of the formulation of the given problem.	52
3.3	Q-table representation.	53
3.4	Policy diagram for Q-learning and Sarsa.	54
3.5	Average cost for Q-learning and Sarsa methods during the learning iterations. . .	57
3.6	Macrocell and average femtocell capacity for Q-learning and Sarsa methods during the learning iterations.	58
3.7	Probability of being below the capacity and above the power thresholds for Q-learning and Sarsa learning approaches.	59
3.8	CDF of the probability of being below the capacity threshold for different α and γ	60
3.9	CDF of the probability of being above the power threshold for different α and γ	61
3.10	Probability of being below the capacity threshold as a function of the learning iterations for different α and γ	61
3.11	Probability of being above the power threshold as a function of the learning iterations for different α and γ values.	62
3.12	Probability of being below the capacity threshold as a function of the learning iterations for different action selection policies.	64

3.13	Probability of being above the power threshold as a function of the learning iterations for different action selection policies.	64
4.1	Q-table for task r of agent f , for case study 2.	73
4.2	Probability of being above the power threshold as a function of the learning iterations for a single-cell scenario with different femtocell densities.	74
4.3	Convergence of SINR at macrouser to three desired values (i.e. 17, 20 and 23 dB).	75
4.4	Macrocell, average femtocell and total system capacity as a function of the femtocell occupation ratio for the single-cell scenario.	75
4.5	Macrocell, average femtocell and total system capacity as a function of the maximum total transmission power for a femtocell occupation ratio of $p_{oc}=20\%$	76
4.6	Probability of being below the capacity threshold as a function of the learning iterations for the multicell scenario.	76
4.7	Femtocells average capacity over the learning iterations for the multicell scenario.	77
4.8	Macrocell and femtocell average capacity versus the femtocell occupation ratio p_{oc} for the multicell scenario.	77
4.9	Transfer learning scheme at macrouser u_m	78
4.10	Message flow structure for Q-learning in LTE networks.	81
4.11	Average probability of not to fulfil the thresholds in time.	81
5.1	Docitive cycle which extends the cognitive cycle by cooperative teaching.	86
5.2	Docitive unit structure: main functionalities.	88
5.3	Taxonomy of docitive algorithms with different degrees of docition.	90
5.4	Macrocell capacity as a function of femtocell density.	92
5.5	CCDF of the average SINR at macrouser for a femtocell occupation ratio $p_{oc}=50\%$	93
5.6	Probability of being above the power threshold as a function of the learning iterations for different docitive cases.	93
6.1	System performance for Q-learning with different amount of available actions.	98
6.2	Fuzzy Inference System scheme.	100
6.3	FIS structure for the FQL-PMFB algorithm.	103
6.4	Membership functions of the input linguistic variables for FQL-PMFB.	104
6.5	Macrocell capacity as a function of the femtocell occupation ratio in single-cell scenario.	107
6.6	Average femtocell system capacity as a function of the femtocell occupation ratio in single-cell scenario.	108

6.7	Total system capacity as a function of the femtocell occupation ratio in single-cell scenario.	108
6.8	Macrocell capacity as a function of the femtocell occupation ratio in multicell scenario.	109
6.9	Average femtocell capacity as a function of the femtocell occupation ratio for multicell scenario.	110
6.10	Probability of being above the power threshold as a function of learning iterations when applying FQL.	111
6.11	Probability of being below the capacity threshold as a function of learning iterations when applying FQL	111
7.1	Messages exchanged when a macrouser attempts to access a femtocell.	118
7.2	Generic variogram parameters.	120
7.3	Basic structure of POMDP technique.	122
7.4	Variogram fit with spherical model for the single-cell scenario with $p_{oc} = 60\%$. . .	124
7.5	CCDF of the error of the SINR estimated by femto BS in the single-cell scenario.	125
7.6	Probability of being below the capacity threshold as a function of the learning iterations in the single-cell scenario.	125
7.7	Macrocell and femtocell average capacity as a function of the femtocell occupation ratio for single-cell scenario.	126
7.8	Variogram fit with different models for the multicell scenario.	126
7.9	CDF of the error for the SINR estimated by femtocells in the multicell scenario.	127
7.10	Probability of being above the power threshold as a function of the learning iterations in multicell scenario.	128
7.11	Probability of being below the capacity threshold as a function of the learning iterations in multicell scenario.	128
7.12	Macrocell capacity as a function of the femtocell occupation ratio for multicell scenario.	129
7.13	Femtocell average capacity as a function of the femtocell occupation ratio for multicell scenario.	130
8.1	Neural network scheme.	139

List of Tables

2.1	3GPP path loss models for urban deployment	35
2.2	Simulation parameters	36
6.1	FQL-PMFB simulation parameters	105
8.1	Operations and their computational requirements	136
8.2	Computational requirement for Q-learning	137
8.3	Computational requirement for FQL	138
8.4	Computational requirement for POMDP	138
8.5	Computational requirement for neural network	140
8.6	Comparison between lookup table and neural network representation mechanisms	141

Chapter 1

Introduction

1.1 Current situation overview

In the last four years, global mobile data traffic has increased more than 130% per year and it is predicted to continue augmenting in the coming years, i.e. from 2012 to 2016, at a compound annual growth rate of 78% [1]. This significant growth in the mobile data traffic comes as a clear result of the proliferation of data-oriented devices, i.e. smartphones, tablets, laptops with mobile broadband, etc., and, with them, the emergence and availability of abundant services and applications.

High quality and fast mobile data everywhere is becoming a necessity for many network users. In order for operators and service providers to supply the increasingly required traffic, they need to plan and deploy highly efficient future networks. The concept of future networks involves the integration of smart, flexible, scalable, robust and environment aware functionalities able to improve the spectral efficiency per unit area. New challenges that need to be addressed by future networks include timely data retrieval, automatic managements of systems and elements, mobility, reliability, automatic identification of new network entities, scalability, efficient energy management and security [2]. To cope with the mentioned issues, the 3rd Generation Partnership Project (3GPP) standardization body has developed the Evolved Packet System (EPS) which consists of the Long Term Evolution (LTE) wireless mobile broadband technology for radio access and the System Architecture Evolution (SAE) for the non-radio aspects, as part of release 8 [3]. Currently, 3GPP is addressing the LTE-Advanced (LTE-A), whose standards are defined in 3GPP release 10 [4].

It has been observed [5] that since 1957 there has been a million fold increase in wireless capacity. If broken down into constituent components, a 25-times improvement is due to a wider spectrum, a 5-times gain to chopping the spectrum into smaller slices, a 5-times enhance is due to advances in modulation and coding schemes and a 1600-times increase is due to reduced cell

sizes and transmit distances that allow efficient spatial reuse of spectrum. Currently, wireless cellular systems with one Base Station (BS) can achieve a performance close to the optimal, characterized by information theoretic capacity limits. Further gains, therefore, depend on the development of advanced network topology and transmission techniques. LTE-A offers high spectral efficiency, low latency and high peak data rates, which are achieved mainly thanks to concepts such as higher order Multiple-Input Multiple-Output (MIMO) techniques, carrier aggregation, heterogeneous networks and self-organization strategies. The work presented in this thesis focuses on these last two concepts.

Heterogeneous networks integrate new techniques and low-power smart nodes (small cells), i.e. femtocell and picocells, into traditional cells, i.e. macrocells, microcells, metrocells and repeaters [6]. This topology decreases the distance between nodes and users, and provides greater capacity at a lower cost per bit. Furthermore, with the introduction of low power nodes, coverage holes in the macrocell system can be eliminated, indoor and cell-edge coverage enhanced, the network capacity in hot spots can be improved and broadband mobility services added [7]. A large amount of traffic can be offloaded from macrocells to small cells, which do not introduce high network overhead and may highly reduce the global network energy consumption. Small cells can be either deployed by operators, i.e. picocells, or users, i.e. femtocells, and may potentially share the same spectrum with traditional cells [8].

Despite all these positive aspects, the insertion of new underlying cells brings complex challenges in terms of, e.g., interference, handover, backhauling and self-organization.

- **Interference:** Traditionally, interference in cellular networks is mitigated via frequency reuse schemes. Using these schemes reduce the spectral efficiency per area unit and involves a careful planning. Thus, frequency reuse schemes are not suitable in heterogeneous networks. The trend is then to develop specialized and competent intercell interference coordination techniques that allow multiple cells to coexist while working on the same frequency band [8].
- **Handover:** The co-existence of multiple cells highly increases the number of vertical and horizontal handovers in the cellular networks, which comes at the expense of system overhead. Also, the probability of handover failures and unnecessary handovers in scenarios consisting of numerous cells increases, hence causing degradation of services, reduction in throughput and an increment in blocking probability and packet loss. New techniques able to consider future trends and provide efficient handover phases processes—i.e. network discovery, handover decision, and handover execution—need to be provided [9].
- **Backhauling:** The introduction of small cells in the cellular network context requires high performance and flexible backhaul solutions that must be cost-efficient, easy to deploy and able to provide uniform end-to-end performance. Technologies defined for backhaul include

millimeter-wave, microwave, optical fiber, category 5/6 Local Area Network (LAN) copper cable and Digital Subscriber Line (DSL) technologies. For outdoor small cells, Line of Sight (LOS) millimeter-wave and microwave or fiber is recommended. On the other hand, for indoor small cells, it is recommended to reuse existing copper and fiber infrastructure and to take advantage of the users DSL connection. In any case, the selection of the backhaul technology in every scenario implies careful analysis in order to guarantee a better relation between cost and Quality of Service (QoS) [10].

- **Self-organization:** Coexistence and management tasks of various types of nodes in the mobile network require self-responsive and intelligent forms of organization, in such a way that entities have the ability to understand and react to the environment in an autonomous manner. Furthermore, in cellular networks, the growing number of different types of nodes implies a notable increase in the number of network parameters with complex interdependencies that have to be considered and optimized. This can be achieved by introducing the concept of self-organization, deeply studied and applied in multiple branches of science and technology. The concept of self-organization was first introduced by Ashby in [11] and recently defined as the global order emerging from local interactions, where systems aim at finding a structure with function through coordination [12]. Then, components of self-organized systems must be able to evolve behaviors without planning actions, changing its structure and function as a result of the sum of all the interactions of its components and the environment as a whole. Currently, self-organizing capabilities contemplated in wireless systems can be classified in self-configuration, self-optimization and self-healing. The introduction of these capabilities would allow operators to reduce Operational Expenditure (OPEX) and Capital Expenditures (CAPEX) and to enhance scalability, robustness, performance and QoS of the network.

In the context of new tendencies in the use of mobile networks we are analyzing, it is worth mentioning that it is estimated that in mobile networks two thirds of calls and over 90% of data services occur indoors. Hence, it is important for cellular operators to provide good indoor coverage not only for voice services, but also for high speed data services. Some surveys, however, show that 45% of households and 30% of businesses experience poor indoor coverage problems [13]. Improving indoor coverage and service quality will generate more revenue for operators and enhance subscribers loyalty. These improvements are proposed to be achieved with the introduction of the innovative concept of femtocell technology.

Femtocells are short-range, low-power, low-cost cellular BSs designed to serve very small areas, such as a home or an office environment, providing radio coverage of a certain cellular network standard, e.g., Universal Mobile Telecommunications System (UMTS), Worldwide Interoperability for Microwave Access (WiMAX), LTE. They are connected to the service provider via broadband connection, e.g., DSL or optical fiber. Due to these characteristics, they can be

deployed far more densely than macrocells, so that the femtocell spectrum can be reused more efficiently than in only-macro networks. Femtocells enable reduced distance between the transmitter and the receiver, and reduced transmit power while maintaining good indoor coverage, since penetration losses and outdoor propagation attenuation insulate the femtocell from surrounding femtocell and macrocell transmissions. Besides, as femtocells serve around 1 to 5 users, they can devote a larger portion of their resources to fewer users compared to large coverage macrocells. Indoor subscribers are served through the user-installed femtocell, providing high data rates and reliable traffic, while the operator reduces traffic on the macrocell, thus focusing only on outdoor and mobile users. It is also worth mentioning that femto BSs only need to be switched on when the users are at home, or at work, so that their use is greener than macrocells, and that they can provide significant power savings to the User Equipment (UE). In fact, the Path Loss (PL) to the femto BS is much smaller than the one to the macro BS, and so is the required transmitted power from the UE to the femto node. This would increase the battery life of UEs, which is one of the biggest bottlenecks for providing high speed data services in mobile networks.

From the operators' perspective, since a large amount of traffic can be offloaded from macrocells, macrocell sites can be reduced, which would result in important CAPEX savings in the radio access network and in the backhauling. This will also lead to associated savings on the OPEX. Preliminary studies have shown that the 60000 USD/year maintenance of a macrocell reduces to 10000 USD/year for an equivalent capacity femto network [5] and that self-optimizing femtocells enable further reductions in OPEX. As for the business case, many issues are still open, since even though femtocells offer savings in site lease, backhaul and electricity costs for operators, they incur strategic investments due to the competition with ubiquitous Wi-Fi. Therefore, operators will have to decide to aggressively price femtocells, despite tight budgets and high manufacturing costs. In February 2009 the Small Cell Forum (formerly Femto Forum) [14]—a non-profit membership organization in charge of representing the operators point of view with the goal of marketing and promoting femtocell solutions as well as input in standardization activities—published the research conducted by a US-based wireless telecommunications consultancy (Signals Research Group—SRG), which used data that had been provided by a group of mobile operators and vendors. They found that femtocells can generate attractive returns for operators by significantly increasing the expected lifetime value of a subscriber across a range of user scenarios. Since then, numerous reports on business cases have been published in [14].

From the subscribers' perspective, femtocells will offer users a single billing account for land line phones, broadband connections and mobile phones, besides improved voice and data services and reduced dropped calls. Femtocells can act as the focal point to connect all domestic devices to a home server and act as the gateway for all domestic devices to the Internet. Numerous applications regarding localization, security, health and information mobility, to name just a

few, are being proposed to be carried by femtocells [15].

Despite these announced economic and technical benefits provided by this new technology, the deployment of femtocells will also cause some problems to operators. To achieve the expected spectral efficiency, macro and femto layers should operate in the same frequency band, so that interference becomes more random and harder to control, and the Radio Resource Management (RRM) task is much more complicated than in traditional cellular networks.

Current state of the art femtocell products incorporate all the functionalities of a typical BS, thus, they only require a data connection like, e.g., the DSL, through which they are connected to the mobile operator's core network. The capabilities include a maximum transmission power up-to 20 dBm; UMTS/High Speed Packet Access (HSPA) indoor coverage; and 4–8 simultaneous voice calls or data sessions. Very basic plug and play functionalities are implemented, including auto-configuration and activation.

This thesis deals with the introduction of femtocells in heterogeneous networks. In more detail, some of the above mentioned challenges are considered and combined, resulting in an intercell interference coordination approach based on self-organization techniques, as detailed in the following section.

1.2 Problem statement

Femtocell technology faces several issues that may be common to all heterogeneous networks components or particular of this technology. These issues need to be solved urgently, before femtocells expected mid-term massive deployment. Some of these issues are summarized below.

1. *Interference management and coexistence with heterogeneous networks:* Femtocells can share their operating frequency band with the existing macro network, or can operate in a dedicated frequency band. The interference management is more challenging in case of co-channel operation, but this option is more rewarding for the operator due to the increased spectral efficiency. The 3GPP LTE-A standard ensures intracell orthogonality among macrocellular users and mitigate intercell interference through fractional frequency reuse. However, since femtocells will be placed by end consumers, their number and position will be unknown to the network operator, so that the interference cannot be handled by means of a centralized frequency planning, which generates a distributed interference management problem.
2. *Self-organization:* Considering that the home will be the basic unit at which femtocells will be deployed, self-organization is essential to the femtocell mass deployment and management, for two main reasons. First, it will be unfeasible to foresee a centralized node

managing radio resources of femtocells, due to their huge number and unknown position, so that femto nodes need to have the capability of making all RRM decisions. Second, self-organization will reduce signaling burden on the backhaul, resulting in improved capacity.

3. *User's access privileges, open versus closed access:* Femtocells can be deployed in a Closed Subscriber Group (CSG) fashion, which implies that only registered users may establish connection with them. Alternatively, femtocells could be characterized by open access, so that they can serve all the operator's subscribers. The closed access option is more challenging from the interference point of view, since a macro user may be located in the coverage area of a femtocell, thus receiving interference from it, while having forbidden access to it. On the other hand, security issues related with the open access scheme and the backhaul, which may result in the bottleneck of the traffic served by the femtocell, may preclude this option. Hybrid privileges may represent the optimum solution, but further research is needed in this field.
4. *Handover:* From the timing perspective, the handover from macrocell to femtocell is not an issue, since the user has plenty of time to make the handover, when the QoS perceived by the macrocell is acceptable. On the other hand, handover from femtocell to macrocell has to be very quick. In case of open access femtocells, other issues come up. In current Second-Generation (2G) and Third-Generation (3G) systems, mobiles use neighbor lists (broadcasted by the current cell) to learn where to search for potential handover cells. Such protocols do not scale to the large number of femtocells that overlay the macrocell, motivating the proposal of novel handover algorithms for heterogeneous networks that take into account the presence of femtocells.
5. *QoS provided by the backhaul:* Backhaul dimensioning and corresponding resource allocation are important aspects of femtocell network design, since femtocells are expected to be deployed at massive scale and a shared backhaul may easily become the traffic bottleneck. Towards this objective, reducing the signaling load is also a challenge. Further research on joint backhaul and radio access design is required.
6. *Timing and Synchronization:* Femtocells will require synchronization to align received signals, minimize multi-access interference, ensure tolerable carrier offset, and correct handover of users from and to the macro network. On the one hand, the intercarrier interference arising from a carrier offset causes loss of subcarrier orthogonality. In Time Division Duplex (TDD) systems, femtocells will require an accurate reference for coordinating the absolute phases to forward and reverse link transmissions, and bounding the timing drift. Both these issues apply to macro BSs as well, but the low cost burden and the difficulty of synchronizing over backhaul make the synchronization an important issue for femtocells.
7. *Low cost:* Despite the savings foreseen thanks to the adoption of femtocells, the operators

will have to face the competition with the ubiquitous Wi-Fi. This is why the cost issue will be a central factor for the actual deployment of femtocells.

8. *Low power consumption:* Keeping the output power of femto BSs low is desirable from interference management and electromagnetic pollution perspectives. However, maintaining both coverage and capacity with low output power is a challenge.

The main scope of this thesis is to investigate the autonomous RRM coordination of femtocells, from the point of view of the aggregated interference they may generate at macrocell users. The interpretation we give to autonomous decisions relies on the theory of self-organization. More precisely, we implement self-organization through Machine Learning (ML) techniques, specifically following a Reinforcement Learning (RL) formulation, since these algorithms allow to learn online from environmental sensed information and based on this to take actions. Environmental information includes any kind of information related with the surrounding scenario in which the femtocell is operating. The concept of self-organization based on learning techniques is then applied in such a form that each femtocell is considered as a component of a complex system able to adjust to the interference circumstances of the surrounding environment. Femtocells are therefore modeled as a decentralized system, lacking of a central authority and working in a coordinated fashion to find a stable and reliable behavior function. However, in some cases, some cooperative techniques are required in order to successfully and efficiently accomplish the learning processes. Also, efficiency in learning approaches is highly related with the possibility of including expert knowledge in the learning algorithms design, in order to improve the accuracy in the representation of the environment and the potential actions the femtocell may take, i.e. a detailed interpretation in the environment and actions by the learning entity brings interesting gains in terms of precision. In addition, sometimes it is not feasible for the learning entities to have all the information required for the environmental representation, therefore the available information for them is another topic to be considered. This thesis covers and presents solutions regarding these issues.

The work presented in this thesis focuses on femtocells with closed access, i.e. they can be accessed only by registered users in their CSGs. Also, femtocells work in co-channel operation with the macrocells and both systems have Orthogonal Frequency Division Multiple Access (OFDMA), as in 3GPP LTE-A. In order to find out the impact of femtocell deployment on the macrocell layer, and of femtocells among each other, we use a system level simulation approach, including accurate radio propagation models [13, 14]. This thesis was developed under the framework of the Integrated Project (IP) Broadband Evolved Femto Networks (Be-FEMTO) [16], therefore, references regarding the functional architecture, use cases, femtocells requirements, etc. proposed in the project and approved by many industrial partners such as Sagemcom, NEC, Telefonica, docomo, Qualcomm, mimoOn, etc, will be found throughout the text.

1.3 Objectives

Self-organization techniques based on ML approaches are introduced to perform the RRM procedures for coexistence of macro and femto networks. In particular, we propose to map the femtocells onto a multiagent system [17], where each femto BS is an intelligent and autonomous agent that learns [18] by directly interacting with the environment and by properly utilizing the past experience, as it is presented in Figure 1.1. Multiagent systems are characterized by the following: i) the intelligent decisions are made by multiple and uncoordinated nodes; ii) the nodes partially observe the overall scenario; and iii) their input to the intelligent decisions process are different from node to node since they come from spatially distributed sources of information. The reason for proposing a multiagent system is found in the impossibility for the femto network to be managed by means of a centralized node, due to the number of femtocells and the lack of information to the network operator regarding their location.

The environment in which the multiagent system is operating is dynamic due to the characteristics of the mobile wireless scenario e.g., existence of lognormal shadowing, fading, mobility of user, etc, and to the cross dependencies of actions made by the multiple agents. We model the natural evolution of the environment through states and the multiagent system, through a stochastic game. A stochastic game is the extension of Markov Decision Processes (MDPs), which are the natural model of a single agent scenario, to multiple agents [19].

Real scenarios formed by simultaneously performing multiple agents commonly present highly dynamic and unstable behaviors. This occurs because the policy learnt by an agent at a given moment may not be valid anymore when the environment switches to a new state due to potential actions performed in parallel by other agents in the system. This characteristic does not allow to define a probabilistic state transition model, therefore, we propose to solve the stochastic game through Time Difference (TD) RL algorithms, presented in Chapter 3. RL paradigm is based on learning from interactions with the environment through actions, where knowledge is built based on the observed consequences when a given action is executed. In this context, we focus on the paradigm of independent learning [20] where each femtocell (agent) learns independently a power allocation strategy for interference avoidance. The interactive learning dynamics will be evaluated in terms of system performances and speed of convergence.

We solve the interference management problem from femto to macro systems in Chapter 4 through Q-learning, which is a form of TD RL. When implementing multiuser scheduling in the macro network, femto nodes need to be able to react to instantaneous changes of macro user allocation per frequency band. This problem can be solved at femtocell level by exploiting the knowledge of the frequency distribution planning of the macrocell, which can be obtained via the X2 interface and by reusing the acquired knowledge for the different frequency bands instead of learning from scratch. This is also covered in Chapter 4.

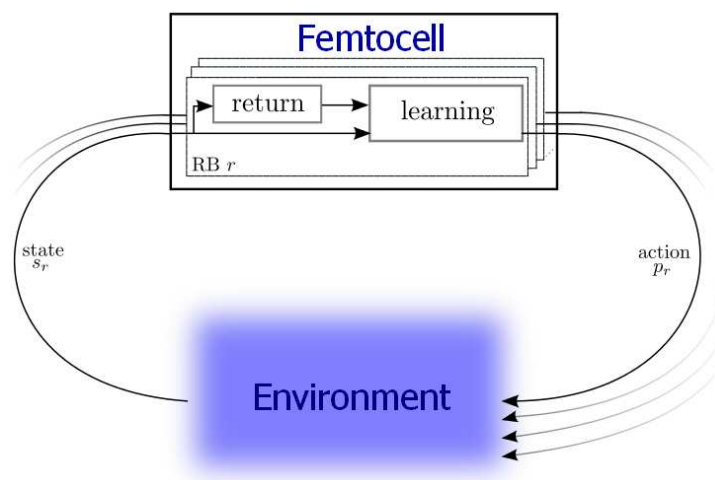


Figure 1.1: General interpretation of the proposed learning-based solution.

To speed up the self-organization and learning process, which is a critical issue of all learning algorithms, in Chapter 5 we propose an unprecedented cooperative paradigm, according to which more expert agents facilitate expert knowledge dissemination, helping other agents to learn, thus mimicking our society-driven pupil-teacher paradigm. We refer to this paradigm as *docitive radio* [21], from Latin *docere* (to teach), and *cognoscere* (to know), and to the femtocells implementing it as “*docitive femtocells*”. Different levels of cooperation among femto nodes [22, 23] are proposed, taking into account the signaling overhead that this would imply over the backhaul.

In RL algorithms, the state of the surrounding environment and available actions to the agents are commonly represented by discrete sets. The TD algorithms are based on quantifying, by means of the Q-function, the quality of an action in a certain state. Therefore, to be able to learn from the past, the Q-values have to be stored in a representation mechanism. The lookup table is the most commonly used and the most direct method when memory requirements are not a problem. However, when the number of state-action pairs is large or the input variables are continuous, the memory requirements may become unfeasible, so that there is a need for a more compact representation mechanism, such as a neural network. On the other hand, the neural network representation mechanism may result in intractable computational complexity for detailed or continuous state or action representations. In addition, the definition of state and action spaces highly depends on the scenario and may significantly affect the performances. To solve these problems and build a system capable of working independently of the scenario and the system designer criterion, in line with Self-organized Networks (SON) requirements, in Chapter 6 we introduce a Fuzzy Q-learning (FQL) scheme, which combines the advantages of Fuzzy Inference System (FIS) and multiagent RL. In particular, the FIS allows to generalize the state space and to generate continuous actions [24]. Furthermore, it reduces the learning period

since previous expert knowledge can be naturally embedded into the fuzzy rules, which results in more adaptable and accurate learning algorithms.

In order for femtocells to have a picture of the surrounding environment they need to receive some information regarding the performance perceived by the macrouser. This information is proposed to be conveyed through a X2 interface between femtocells and macrocells. The existence of the mentioned interface, however, has not been considered yet in last standardization release 11 [25], therefore, femtocells will be supposed to perform under a partial information scheme. In this case the learning processes rely on a set of state beliefs, locally built and maintained by the femtocells, who take actions based on them. The theoretical framework considered in this case is that of Partially Observable Markov Decision Process (POMDP) and it is presented in Chapter 7.

Computational requirements have also to be discuss to evaluate the feasibility of the proposed schemes. This is studied in Chapter 8.

In this thesis we focus on stand-alone femtocells for residential sector, connected to the operator's network via a traditional wired backhaul and on networked femtocells for large indoor spaces and dense urban deployments. Stand-alone femtocells are not supposed to exchange information to carry out a certain task, but they may interact with the macro BS by means of the X2 interface, as considered in BeFEMTO functional architecture, with respect to the interference perceived by the macrouser. On the other hand, networked femtocells are stand-alone femtocells interconnected through a X2 interface, introduced in 3GPP release 10 for some particular cases, i.e. mobility enhancement between femtocells when the target cell is an open access femtocell or for closed/hybrid access femtocells with the same CSG ID. Then, femtocells with X2 interface among them may exchange information for e.g., location information, performance estimation, etc.

1.4 State of the art

The work presented in this thesis embraces multiple areas of interest. In what follows, a summary of the state of the art in literature and with respect to standardization, self-organization and RRM in femtocell systems, is presented.

- **Standardization efforts:** Driven by the increasing interest in femtocells and to guarantee products' interoperability, the Small Cell Forum [14] has emerged and closely collaborates with 3GPP, 3GPP2, Next Generation Mobile Networks (NGMN), as well as with the broadband forum. Its current focus is on the development and adoption of small cells for the provision of high-quality 2G, 3G and Fourth-Generation (4G) coverage and services within residential, enterprise, public and rural access markets.

Femtocells, or Home eNodeBs (HeNBs), following the 3GPP nomenclature, have been included in standards since release 8, which covers basic HeNB architectures with focus on how H(e)NB are accommodated into the operator core network. Release 8 introduced the access control mechanisms for HeNBs: open, closed and hybrid and the concept of CSG in TS 25.467 (UTRAN architecture for 3G Home Node B). Advances in standardization are reflected in release 9, where the HeNB architecture and support are included in TS 36.300-870 (The UTRAN Overall description). TR 23.830 release 9 includes roaming support for access control, handover from macro to femtocells, IP Multimedia Subsystem (IMS) enabled HeNB, etc. Release 10 TS 36.300 introduces the X2 interface between HeNBs and provides HeNB mobility enhancements including intra-CSG/inter-CSG in a HeNB Gateway (HeNB GW). Release 11 presents HeNB security features for UE mobility scenarios in TS 33.320.

Other significant technical specifications and reports are: TS 22.220 (Service requirements for H(e)NB), TR 23.830 (Architecture aspects of H(e)NB), TR 23.832 (IMS aspects of architecture for HeNB), TS 25.467 (Universal Terrestrial Radio Access Network (UTRAN) architecture for 3G HeNB), TS 25.367 (Mobility procedures for HeNB), TS 25.469 (UTRAN Iuh interface HeNB Application Part signaling), TR 25.820 (3G HeNB study item), TR 25.967 (Frequency Division Duplex (FDD) HeNB RF requirements), TS 32.581-2-3 (HeNB Operation, Administration and Maintenance (OA&M) concepts and requirements), TS 32.583, TR 32.821 (Study of Self-organizing networks related OA&M interfaces for HeNB), TR 33.820 (Security of H(e)NB) and TR 36.921 (FDD HeNB radio frequency requirements analysis).

Parallel activities in NGMN [26] and 3GPP [27] on self-configuring and SON have been achieved for some time now. Release 8 includes SON functionalities regarding initial equipment installation and integration, i.e. Automatic Neighbor Relation (ANR), self-configuration of the Evolved NodeB (eNB), the Mobility Management Entity (MME) and automatic Physical Cell Identity (PCI) configuration. The SON functions developed in release 9 are designed to optimize deployed LTE networks. This includes requirements, goals and parameters of SON use cases, i.e. interference reduction, automatic configuration, mobility robustness and load balancing optimization, inter-cell interference coordination, etc., which have been captured in TR 36.902 [28]. Release 10 introduces SON functions to enhance interoperability between small cells and macrocells and includes the recommendations provided by NGMN. Summarizing, new functionalities such as coverage and capacity optimization, enhanced inter-cell interference coordination, cell outage detection and compensation, self-healing functions, minimization of drive testing and energy savings, are introduced. Release 11 SON functions are related to the automated management of heterogeneous networks. It includes mobility robustness optimization enhancements and inter-radio access technology handover decision optimization [29].

- **Literature:** From the scientific point of view, the number of peer-reviewed publications dedicated to femtocells is already quite high. Among the first publications about femtocells are [13, 30–51]. These contributions have targeted solutions for the most pertinent problems reported by product developments and standardization activities: market impact and business model [30, 31], evaluation of performance through system level simulations [32, 33], performance of WiMAX-based femtocells [34–36], self-optimization and organization [37–39], interference avoidance [40–47, 52] and access control [13, 42, 48–51]. More specifically, [32] presents a method for power control for pilot and data that ensures a constant femtocell radius in the downlink, and its theoretical performance is evaluated through system level simulations. Reference [38] presents two interference mitigation strategies, based on open-loop and closed-loop control, which adjust the maximum transmit power of femtocell users to suppress the cross-tier interference at the macrocell BS. Those strategies are evaluated through simulation results. Reference [39] presents two dynamic frequency selection algorithms that permit the smart selection of an operating band, among the available ones, to the generic femtocell. In [45], the authors propose a dynamic resource partitioning to mitigate the downlink femto to macro interference. In this case femtocells are denied access to downlink resources assigned to closer macro users. On the other hand, in [46], the authors propose an opportunistic channel scheduling scheme that determines optimal channel and power allocation to femtocell users in order to manage the uplink interference from macrocell users to femtocells.

Open versus closed access policies have been investigated in [53], the effect of the number of femtocells on the capacity of a macrocell have been described in [54], and examinations of interference impacts involving various combinations of femtocells and macrocells can be encountered in [55]. The conclusion from these studies is that a closed co-channel femtocell can lead to significant interference problems for all parties, particularly if the femtocell does not adaptively change its transmit power in order to minimize its interference on existing networks [55]. Tradeoffs associated with different levels of open access are investigated in [49], where the level of open access is adaptively controlled as a function of factors including the instantaneous load on the femtocell. In fact, while a closed femtocell can be problematic, in an environment with high density of mobile stations a completely open femtocell can also suffer problems, because it will potentially force the sharing of limited femtocell wireless bandwidth and internet backhaul capacity among a significant number of mobile stations.

Downlink power control in femtocells working in CSG mode and deployed in co-channel operation is a key area of research. Recent contributions in this field include a power control scheme based on macrouser Reference Signal Received Power (RSRP) reports, presented in [56] and in [57]. The authors improve the power control proposed by 3GPP in [58], which is based on femtocells Signal to Interference Noise Ratio (SINR) measurements, by

introducing a macrouser proximity control. In [59], the authors introduce a game theory-based universal power allocation algorithm to be executed simultaneously at each BS to satisfy user demands. A general analysis regarding the development of femtocells up to date and an evolution forecast for this technology in coming years is presented in [60].

A central point is the interference management in OFDMA femtocells, which is why the Small Cell Forum [14] presented a study regarding this [61] and concluded that interference between femtocells and between macro and femtocells remains the most detrimental performance factor. This important issue is also given significant emphasis by 3GPP [62]. The requirements exposed by the Small Cell Forum found their input into 3GPP 3G HeNB and LTE HeNB activities. Due to the importance of this issue, many contributions can be found. To give some examples, in [44], the authors introduce an uplink capacity analysis and interference avoidance strategy based on feasible combinations of average number of active macrocell UEs and femtocell BSs per cell-site. Reference [63] proposes a decentralized interference control based on potential games. Some guidelines on spectrum allocation and interference mitigation based on self-configuration and self-optimization techniques are presented in [41]. A distributed and dynamic carrier assignment method for downlink interference avoidance is proposed in [52]. In [64] a downlink interference management in OFDMA networks, based on distributed gradient descent method is presented and in [65] the authors present an algorithm based on game theory to mitigate femto-to-macrocell cross-tier interference.

In literature, multiple examples of self-organized techniques can be found. In [66], the authors propose an adaptive frequency reuse and deployment for OFDMA cellular networks based on a self-organizing framework. Biologically inspired mutually coupled oscillator techniques have been applied for automatic synchronization in sensor networks [67]. Also, transmission power selection in distributed sensor networks through consensus average methods have been proposed in [68]. More in particular, multiple references can be found in literature regarding self-organizing femtocells, to name a few, in [69] the authors propose two approaches for inter-cell interference control based on messages exchanged by the femtocells or measurement reports coming from the users. In [70] the authors propose solutions to automatically tune parameters such as radio spectrum, pilot power, resource blocks, and access control mechanisms for optimal performance for enterprise femtocells. Reference [71] summarizes the BeFEMTO project proposed interference mitigation algorithms for macro-femtocell coexistence, which are based on distributed learning algorithms.

- **European projects:** Among the projects related to the topic, we highlight the Integrated Project (IP) BeFEMTO (Broadband Evolved Femto Networks) [16], and the Specific targeted Research Project (STREP) FREEDOM (Femtocell-based Network Enhancement by Interference Management and Coordination of Information for Seamless Connectivity) [72],

led by SAGEM Communications SAS and Universitat Politècnica de Catalunya (UPC), respectively, and funded in the framework of the 4th call of the 7th Framework Program of European Union. The activities planned in both these projects, aim to provide a new vision of a femto-based broadband network, giving solutions to the major open issues of the femtocell technology, promising new advances beyond the state of the art and significant impact on standardization bodies. We also emphasize the work carried out in the context of the Celtic initiatives HOMESNET (Home Base Station: An Emerging Network Paradigm) and Winner+ (Wireless World Initiative New Radio +), where femtocells are studied in the context of spectrum sharing concepts between operators, with macrocells as the primary system and femtocells as the secondary one. ARTIST4G (Advanced Radio Interface Technologies for 4G Systems) is a FP7 project building upon the 3GPP LTE-A standard [73]. The main project objective is to improve the ubiquitous user experience of cellular mobile radio systems by satisfying the requirements of high spectral efficiency and user data rate across the whole coverage area, fairness between users, low cost per information bit, and low latency. Regarding self-organization applied to wireless networks, the SOCRATES (Self-Optimisation and self-ConfiguRATion in wirelEss networkS) European funded FP7 project stands out [74]. This project is aimed at the development of self-organization methods to enhance the operations of wireless access networks, by integrating network planning, configuration and optimization into a single, mostly automated process requiring minimal manual intervention.

1.5 Outline of the thesis

This section gives a brief overview of the contents of the following chapters, which are summarized in Figure 1.2.

Chapter 2

This chapter presents the scenarios considered to validate the proposed solution following the 3GPP recommendations and the scenarios considered in the BeFEMTO project. It also summarizes the system model, the simulation parameters for the macrocell and femtocell systems and the enhanced system functional architecture, based on the 3GPP release 10 EPS recommendations.

Chapter 3

This chapter summarizes the main concepts related to multiagent learning systems based on RL. Then, it presents a study to select one TD learning method proposed to be implemented

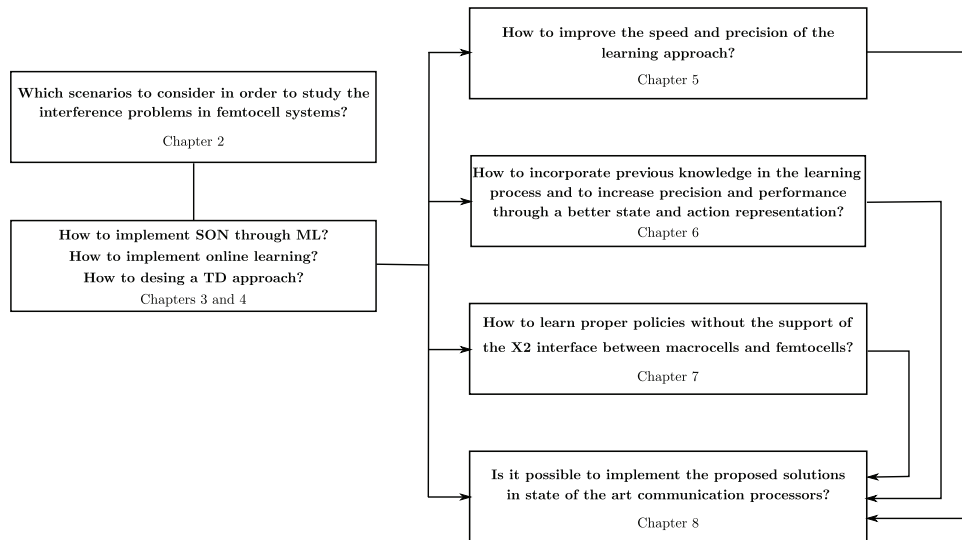


Figure 1.2: Thesis main contributions.

in the femtocell BSs to perform the interference control. Once the learning approach is selected, a study regarding the design of the learning process is presented. This study consists of the selection of three important parameters, i.e. learning rate, discount factor and action selection policy.

Results related to this chapter have been submitted for possible publication to:

- A. Galindo-Serrano and L. Giupponi, “Designing an Online Learning Method for Interference Control”, submitted to *Dynamic Games and Applications Journal*.
- M. Simsek, A. Czylik, A. Galindo-Serrano, L. Giupponi, “Improved Decentralized Q-learning Algorithm for Interference Reduction in LTE-femtocells”, in *Proceedings of IEEE Wireless Advanced 2011*, 20-22 June 2011, London, UK.

Chapter 4

This chapter presents the different state and cost representations used by the learning approaches proposed in this thesis. Then, simulation results are presented for different scenarios, in terms of algorithm ability to fulfil the constraints over the time and macrocell and femtocell systems performance. Furthermore, in order to represent the state of the environment, femtocells require some information from the macrocell system. Hence, a study regarding the implementation of the proposed approach in 3GPP systems is carried out and a strategy to handle the multiuser scheduling, based on transfer learning, is introduced.

The work of this chapter has been published in the following papers:

- A. Galindo-Serrano, L. Giupponi, G. Auer, “Distributed Femto-to-Macro Interference Management in Multiuser OFDMA Networks”, in *Proceedings of IEEE 73rd Vehicular Technology Conference (VTC2011-Spring), Workshop on Broadband Femtocell Technologies*, 15-18 May, 2011, Budapest, Hungary.
- A. Galindo-Serrano and L. Giupponi, “Distributed Q-learning for Interference Control in OFDMA-based Femtocell Networks”, in *Proceedings of IEEE 71st Vehicular Technology Conference (VTC2010-Spring)*, 16-19 May 2010, Taipei, Taiwan.
- A. Galindo-Serrano and L. Giupponi, “Distributed Q-learning for Aggregated Interference Control in Cognitive Radio Networks”, *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1823–1834, May 2010.
- A. Galindo-Serrano, L. Giupponi, “Aggregated Interference Control for Cognitive Radio Networks based on Multi-Agent Learning”, in *Proceedings of the 4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM 2009)*, 22-24 June 2009, Hannover, Germany.

Chapter 5

This chapter introduces a novel cooperative technique among learning entities in order to facilitate them to deal with common problems in decentralized multiagent systems. The proposed technique is known as docition. Docition allows agents to speed up the learning process and to create rules for unseen situations based on expert knowledge exchange among learners. This chapter focuses on introducing the concept of docition and its working taxonomy. The contents of this chapter are the results of a joint work with Pol Blasco and it covers the initial stages of the research regarding the docitive approach. Further and deeper results will be part of Pol Blasco Ph.D. thesis.

The following articles deal with the contents of this chapter:

- A. Galindo-Serrano, L. Giupponi and M. Dohler, “Cognition and Docition in OFDMA-Based Femtocell Networks”, in *Proceedings of the Global Telecommunications Conference, 2010 (IEEE GLOBECOM'10)*, Dec. 2010, Miami, USA.
- P. Blasco, L. Giupponi, A. Galindo-Serrano and M. Dohler, “Energy Benefits of Cooperative Docitive over Cognitive Networks”, in *Proceedings of the 3rd European Wireless Technology Conference 2010 in the European Microwave Week*, Sept 26 - Oct. 1, Paris, France.
- A. Galindo-Serrano, L. Giupponi and M. Dohler, “BeFEMTO’s Self-Organized and Docitive Femtocells”, in *Proceedings of Future Network and MobileSummit 2010 Conference*,

16-18 June 2010, Florence, Italy.

- A. Galindo-Serrano, L. Giupponi, P. Blasco and M. Dohler, “Learning from Experts in Cognitive Radio Networks: The Docitive Paradigm”, in *Proceedings of the 5th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM 2010)*, 9-11 June 2010, Cannes, France.
- L. Giupponi, A. Galindo-Serrano, P. Blasco and M. Dohler, “Docitive Networks—An Emerging Paradigm for Dynamic Spectrum Management”, *IEEE Wireless Communications Magazine*, vol. 17, no. 4, pp. 47–54, Aug. 2010.
- L. Giupponi, A. Galindo-Serrano and M. Dohler, “From Cognition To Docition: The Teaching Radio Paradigm For Distributed & Autonomous Deployments”, *Computer Communications*, Elsevier, Aug. 2010.
- M. Dohler, L. Giupponi, A. Galindo-Serrano and P. Blasco, “Docitive Networks: A Novel Framework Beyond Cognition”, *IEEE Communications Society, Multimedia Communications TC, E-Letter*, January 2010.

Chapter 6

In this chapter, a novel solution to the basic discrete state and action representation in RL methods is proposed. The combination of FIS and RL, called FQL, allows a more compact and effective expertness representation mechanism, the avoidance of subjectivity in the algorithm design, the independence of the scenario and the designer criterion, in line with SON requirements, and the possibility of speeding up the learning process by incorporating offline expert knowledge are some of the advantages introduced by this solution with respect to traditional Q-learning.

Obtained results have been published in:

- A. Galindo-Serrano and L. Giupponi “Q-learning Algorithms for Interference Management in Femtocell Networks”, submitted to *EURASIP Journal on Wireless Communications and Networking*.
- A. Galindo-Serrano, L. Giupponi, M. Majoral, “On Implementation Requirements and Performances of Q-Learning for Self-Organized Femtocells”, in *Proceedings of the IEEE Global Communications Conference (IEEE Globecom 2011), second Workshop on Femtocell Networks (FEMnet)*, 5-9 December, Houston, USA.
- A. Galindo-Serrano, L. Giupponi, “Downlink Femto-to-Macro Interference Management based on Fuzzy Q-Learning”, in *Proceedings of the Third IEEE International workshop*

on *Indoor and Outdoor Femto Cells (IOFC'11)*, May 13, 2011, Princeton, USA. **BEST PAPER AWARD.**

Chapter 7

The main limitation of the approach presented in previous chapters is the assumption of the existence of an X2' interface between macrocells and femtocells, through which macrocells communicate to the femtocell the degree of interference at macrousers. X2' interface has not been standardized in last 3GPP release 11. This chapter presents a completely autonomous solution, where a femtocell network does not require any information from the macrocell network. The proposed solution relies on the theory of POMDP, which furthermore avoids the signaling burden on the backhaul network introduced by the signaling overhead over the X2' interface required for the Q-learning implementation.

Results regarding contents of this chapter appear in:

- A. Galindo-Serrano and L. Giupponi, “Managing Femto to Macro Interference without X2 Interface Support Through POMDP”, submitted to *Mobile Networks and Applications (MONET) Journal. Special Issue on Cooperative and Networked Femtocells*.
- A. Galindo-Serrano and L. Giupponi “Managing Femto-to-Macro Interference without X2 Interface Support”, submitted to *23rd Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2012)*.
- A. Galindo-Serrano and L. Giupponi, “Distributed Q-learning for Aggregated Interference Control in Cognitive Radio Networks”, *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1823–1834, May 2010.
- A. Galindo-Serrano and L. Giupponi, “Decentralized Q-learning for Aggregated Interference Control in Completely and Partially Observable Cognitive Radio Networks”, in *Proceedings of IEEE Consumer Communications & Networking Conference (IEEE CCNC 2010)*, 9-12 January 2010, Las Vegas, USA. **BEST CONFERENCE PAPER AWARD.**

Chapter 8

Since proposed learning approaches are supposed to be embedded in femtocell BSs, this chapter presents a study regarding the possibility to incorporate them in state of the art communication processors. To this end, an analysis regarding the learning approaches memory and computational requirements is performed.

Results presented in this chapter appear in the following publications:

- A. Galindo-Serrano and L. Giupponi, “Managing Femto to Macro Interference without X2 Interface Support Through POMDP”, submitted to *Mobile Networks and Applications (MONET) Journal. Special Issue on Cooperative and Networked Femtocells*.
- A. Galindo-Serrano and L. Giupponi “Q-learning Algorithms for Interference Management in Femtocell Networks”, submitted to *EURASIP Journal on Wireless Communications and Networking*.
- A. Galindo-Serrano, L. Giupponi, M. Majoral, “On Implementation Requirements and Performances of Q-Learning for Self-Organized Femtocells”, in *Proceedings of the IEEE Global Communications Conference (IEEE Globecom 2011), second Workshop on Femtocell Networks (FEMnet)*, 5-9 December, Houston, Texas, USA.
- A. Galindo-Serrano, L. Giupponi, G. Auer, “Distributed Femto-to-Macro Interference Management in Multiuser OFDMA Networks”, in *Proceedings of IEEE 73rd Vehicular Technology Conference (VTC2011-Spring), Workshop on Broadband Femtocell Technologies*, 15-18 May, 2011, Budapest, Hungary.
- A. Galindo-Serrano and L. Giupponi, “Distributed Q-learning for Aggregated Interference Control in Cognitive Radio Networks”, *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1823–1834, May 2010.

Chapter 9

This chapter summarizes the main results of the thesis and possible future lines of research directly related to this work. Some preliminary results on a work carried out with Dr. Eitan Altman, and described in this section, have been submitted to the *6th International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS 2012)*.

Bibliography

- [1] Cisco, “Cisco visual networking index: Global mobile data traffic forecast update, 2011-2016,” Cisco, White paper, February 2012, available online (29 pages). [Online]. Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html
- [2] Net!Works, “Future networks and management,” Net!Works European Technology Platform, White paper, June 2011, available online (14 pages). [Online]. Available: http://www.networks-etp.eu/fileadmin/user_upload/Publications/Position_White_Papers/White_Paper_Future_Network_Management.pdf
- [3] “3GPP TS 36.201 v8.3.0 evolved universal terrestrial radio access (E-UTRA); LTE physical layer - general description (release 8),” 3GPP organization, Tech. Rep., March 2009.
- [4] 3GPP organization. [Online]. Available: <http://www.3gpp.org/LTE-Advanced>
- [5] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, “Femtocell networks: A survey,” *IEEE Communication Magazine*, vol. 46, no. 9, pp. 59–67, Sept. 2008.
- [6] Qualcomm, “LTE advanced: Heterogeneous networks,” Qualcomm Incorporated, White paper, January 2011, available online (15 pages). [Online]. Available: <http://www.qualcomm.com/media/documents/lte-advanced-heterogeneous-networks-0>
- [7] 4G Americas, “Optimizing the mobile application ecosystem,” 4G Americas, White paper, April 2011, available online (19 pages). [Online]. Available: <http://lteworld.org/whitepaper/optimizing-mobile-application-ecosystem>
- [8] D. López-Pérez, I. Güvenç, G. de la Roche, M. Kountouris, T. Q. S. Quek, and J. Zhang, “Enhanced inter-cell interference coordination challenges in heterogeneous networks,” *CoRR*, vol. abs/1112.1597, 2011.
- [9] A. A. I. Hassane, R. Li, and F. Zeng, “Handover necessity estimation for 4G heterogeneous networks,” *International Journal of Information Sciences and Techniques (IJIST)*, vol. 2, no. 1, January.
- [10] Ericsson, “It all comes back to backhaul,” Ericsson, White paper, February 2012, available online (11 pages). [Online]. Available: <http://www.ericsson.com/res/docs/whitepapers/WP-Heterogeneous-Networks-Backhaul.pdf>
- [11] W. R. Ashby, “Principles of the self-organizing dynamic system,” *The Journal of General Psychology*, vol. 37, no. 2, pp. 125–128, 1947.

-
- [12] F. Heylighen, *Self-organization in communicating groups: the emergence of coordination, shared references and collective intelligence*. Language and Complexity (Barcelona University Press), 2011.
- [13] J. Zhang and G. de la Roche, *Femtocells: Technologies and Deployment*. Wiley, Jan. 2010.
- [14] Small Cell Forum. [Online]. Available: <http://www.smallcellforum.org/>
- [15] “D2.1: Description of baseline reference systems, use cases, requirements, evaluation and impact on business model,” EU FP7-ICT BeFEMTO project, Dec. 2010.
- [16] The BeFEMTO website. [Online]. Available: <http://www.ict-befemto.eu/>
- [17] K. P. Sycara, “Multiagent systems,” *AI Magazine*, vol. 19, no. 2, pp. 79–92, 1998.
- [18] M. E. Harmon and S. S. Harmon, “Reinforcement learning: A tutorial,” 2000. [Online]. Available: <http://www.nbu.bg/cogs/events/2000/Readings/Petrov/rltutorial.pdf>
- [19] G. Chalkiadakis, “Multiagent reinforcement learning: Stochastic games with multiple learning players,” University of Toronto, Tech. Rep., 2003.
- [20] P. Hoen and K. Tuyls, “Analyzing multi-agent reinforcement learning using evolutionary dynamics,” in *In Proc. of the 15th European Conference on Machine Learning (ECML)*, June 2004.
- [21] M. Dohler, L. Giupponi, A. Galindo-Serrano, and P. Blasco, “Docitive networks: A novel framework beyond cognition,” *IEEE Communications Society, Multimedia Communications TC, E-Letter*, Jan. 2010.
- [22] M. Tan, *Multi-Agent Reinforcement Learning: Independent vs. Cooperative Agents*. In M. N. Huhns and M. P. Singh, editors. Morgan Kaufmann, San Francisco, CA, USA., 1993, ch. 26, pp. 451–480.
- [23] L. Panait and S. Luke, “Cooperative multi-agent learning: The state of the art,” *Autonomous Agents and Multi-Agent Systems*, vol. 3, no. 11, pp. 383–434, Nov. 2005.
- [24] Y.-H. Chen, C.-J. Chang, and C. Y. Huang, “Fuzzy Q-learning admission control for WCDMA/WLAN heterogeneous networks with multimedia traffic,” *IEEE Transactions on Mobile Computing*, vol. 8, pp. 1469–1479, 2009.
- [25] 3GPP, “Evolved Universal Terrestrial Radio Access (E-UTRA); Overall description; Stage 2 (Release 11),” 3GPP TS 36.300 V11.1.0 (2012-03), March 2012.
- [26] NGNM Alliance, “NGMN recommendation on SON and O&M requirements,” Next Generation Mobile Networks, White paper, December 2008, available online (53 pages).

- [Online]. Available: http://www.ngmn.org/uploads/media/NGMN_Recommendation_on_SON_and_O_M_Requirements.pdf
- [27] “Telecommunication management; self-organizing networks (SON); concepts and requirements,” 3GPP organization, TS TS 32.500.
- [28] “Evolved universal terrestrial radio access network (E-UTRAN); self-configuring and self-optimizing network (SON) use cases and solutions (release 9),” 3GPP organization, Tech. Rep., June 2010.
- [29] 4G Americas, “Self-optimizing networks - the benefits of SON in LTE,” 4G Americas, White paper, July 2011, available online (69 pages). [Online]. Available: <http://www.4gamericas.org/documents/Self-Optimizing%20Networks-Benefits%20of%20SON%20in%20LTE-July%202011.pdf>
- [30] S. Ortiz, “The wireless industry begins to embrace femtocells,” *Computer*, vol. 41, pp. 14–17, July 2008.
- [31] C. Edwards, “The future is femto,” *Engineering & Technology*, vol. 3, no. 15, pp. 70–73, Sept. 6 2008.
- [32] H. Claussen, “Performance of macro- and co-channel femtocells in a hierarchical cell structure,” in *In Proc. of the 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, IEEE PIMRC 2007*, 3-7 Sept. 2007.
- [33] L. T. W. Ho and H. Claussen, “Effects of user-deployed, co-channel femtocells on the call drop probability in a residential scenario,” in *In Proc. of the 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, IEEE PIMRC 2007*, 3-7 Sept. 2007.
- [34] S.-P. Yeh, S. Talwar, S.-C. Lee, and H. Kim, “WiMAX femtocells: a perspective on network architecture, capacity, and coverage,” *IEEE Communication Magazine*, vol. 46, no. 10, pp. 58–65, Oct. 2008.
- [35] D.-Y. Kwak, J.-S. Lee, Y. Oh, and S.-C. Lee, “Development of WiBro (mobile WiMAX) femtocell and related technical issues,” in *Global Telecommunications Conference, 2008. IEEE GLOBECOM'08*, Dec. 2008, pp. 5638–5642.
- [36] A. Valcarce, G. de la Roche, A. Jüttner, D. López-Pérez, and J. Zhang, “Applying FDTD to the coverage prediction of WiMAX femtocells,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Feb. 2009.
- [37] H. Claussen, L. T. W. Ho, and L. G. Samuel, “Self-optimization of coverage for femtocell deployments,” *Wireless Telecommunication Symposium. WTS 2008*, pp. 278–285, 24-26 April 2008.

- [38] H.-S. Jo, J.-G. Yook, C. Mun, and J. Moon, "A self-organized uplink power control for cross-tier interference management in femtocell networks," in *IEEE Military Communications Conference, IEEE MILCOM 2008*, 17-19 Nov. 2008.
- [39] F. Mazzenga, M. Petracca, R. Pomposini, F. Vatalaro, and R. Giuliano, "Algorithms for dynamic frequency selection for femto-cells of different operators," in *The 21th Personal, Indoor and Mobile Radio Communications Symposium 2010 (PIMRC'10)*, Sept. 2010, pp. 1548–1553.
- [40] I. Guvenc, M.-R. Jeong, F. Watanabe, and H. Inamura, "A hybrid frequency assignment for femtocells and coverage area analysis for co-channel operation," *IEEE Communications Letters*, vol. 12, no. 12, pp. 880–882, Dec. 2008.
- [41] D. López-Pérez, A. Valcarce, G. de la Roche, and J. Zhang, "OFDMA femtocells: A roadmap on interference avoidance," *IEEE Communications Magazine*, vol. 47, no. 9, pp. 41–48, Sept. 2009.
- [42] D. López-Pérez, G. de la Roche, A. Valcarce, A. Jüttner, and J. Zhang, "Interference avoidance and dynamic frequency planning for WiMAX femtocells networks," in *Communication Systems, 2008. ICCS 2008. 11th IEEE Singapore International Conference on*, 19-21 Nov. 2008, pp. 1579–1584.
- [43] N. Arulselvan, V. Ramachandran, S. Kalyanasundaram, and G. Han, "Distributed power control mechanisms for HSDPA femtocells," in *Proceedings of the 69th IEEE Vehicular Technology Conference, VTC Spring 2009*, April 2009.
- [44] V. Chandrasekhar and J. G. Andrews, "Uplink capacity and interference avoidance for two-tier femtocell networks," *IEEE Trans. Wireless Communications*, vol. 8, no. 7, pp. 3498–3509, July 2009.
- [45] Z. Bharucha, A. Saul, G. Auer, and H. Haas, "Dynamic resource partitioning for downlink femto-to-macro-cell interference avoidance," *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, pp. 14–17, 2010.
- [46] Y.-Y. Li and E. S. Sousa, "Cognitive uplink interference management in 4G cellular femto-cells," in *The 21th Personal, Indoor and Mobile Radio Communications Symposium 2010 (PIMRC'10)*, Sept. 2010, pp. 1565–1569.
- [47] M. Husso, J. Hämäläinen, R. Jäntti, J. Li, E. Mutafungwa, Risto Wichman, Z. Zheng, and A. M. Wyglinski, "Interference mitigation by practical transmit beamforming methods in closed femtocells," *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, 2010.

- [48] D. López-Pérez, G. de la Roche, A. Valcarce, A. Jüttner, and J. Zhang, "Access methods to WiMAX femtocells: A downlink system-level case study," in *Communication Systems, 2008. ICCS 2008. 11th IEEE Singapore International Conference on*, 19-21 Nov. 2008, pp. 1657–1662.
- [49] D. Choi, P. Monajemi, S. Kang, and J. Villasenor, "Dealing with loud neighbors: the benefits and tradeoffs of adaptive femtocell access," in *Global Telecommunications Conference, 2008. IEEE GLOBECOM'08*, Dec. 2008.
- [50] S. B. Kang, Y. M. Seo, Y. K. Lee, M. Z. Chowdhury, W. S. Ko, M. N. Irlam, S. W. Choi, and Y. M. Jang, "Soft QoS-based CAC scheme for WCDMA femtocell networks," *Advanced Communication Technology, 2008. ICACT 2008. 10th International Conference on*, vol. 1, pp. 409–412, 17-20 Feb. 2008.
- [51] D. Choi, P. Monajemi, S. Kang, and J. Villasenor, "Implementation of network listen modem for WCDMA femtocell," in *Cognitive Radio and Software Defined Radios: Technologies and Techniques, 2008 IET Seminar on*, Sept. 2008.
- [52] S. Uygungelen, Z. Bharucha, and G. Auer, "Decentralized interference coordination via autonomous component carrier assignment," in *Workshops Proceedings of the Global Communications Conference, GLOBECOM 2011, 5-9 December 2011, Houston, Texas, USA*, 2011, pp. 219–224.
- [53] Nortel, Vodafone, "Open and closed access for Home NodeBs," 3GPP, Athens, Greece, 3GPP document reference R4-071231, Aug. 2007, 3GPP TSG-RAN WG4 Meeting 44.
- [54] Nokia Siemens Networks, "Initial home NodeB coexistence simulation results," 3GPP, Orlando, USA, 3GPP document reference R4-070902, Jun. 2007, 3GPP TSG-RAN WG4 Meeting 43bis.
- [55] Orange, "Home BTS consideration and deployment scenarios for UMTS," 3GPP, Kobe, Japan, 3GPP document reference R4-070825, May 2007, 3GPP TSG RAN WG4 Meeting 43.
- [56] Z. Wang, W. Xiong, C. Dong, J. Wang, and S. Li, "A novel downlink power control scheme in LTE heterogeneous network," in *The 9th Annual IEEE Consumer Communications and Networking Conference - Wireless Consumer Communication and Networking*, January 2012, pp. 241–245.
- [57] T. Yang and L. Zhang, "Approaches to enhancing autonomous power control at femto under co-channel deployment of macrocell and femtocell," in *IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2011, Toronto, ON, Canada, September 11-14, 2011*, 2011, pp. 71–75.

- [58] “3GPP TR 36.921 evolved universal terrestrial radio access (E-UTRA); FDD home eNode B (HeNB) radio frequency (RF) requirements analysis,” 3GPP, Tech. Rep., March 2010.
- [59] S. Ramanath, V. Kavitha, and M. Debbah, “Satisfying demands in a multicellular network: A universal power allocation algorithm,” in *9th International Symposium on Modeling and Optimization in Mobile, Ad-Hoc and Wireless Networks (WiOpt 2011), May 9-13, 2011, Princeton, NJ, USA*, 2011, pp. 175–182.
- [60] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, “Femtocells: Past, present, and future,” *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 497–508, 2012.
- [61] FemtoForum, “Interference management in OFDMA femtocells.” [Online]. Available: <http://www.femtoforum.org>
- [62] “Home Node B (HNB) radio frequency (RF) requirements (FDD) (release 10),” 3GPP organization, Technical Report TR 25.967, April 2011.
- [63] L. Giupponi and C. Ibars, “Distributed interference control in OFDMA-based femtocells,” in *Proceedings of the IEEE 21st International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2010, 26-29 September 2010, Istanbul, Turkey*, pp. 1201–1206.
- [64] R. Combes, Z. A. M. Haddad, and E. Altman, “Self-optimizing strategies for interference coordination in OFDMA networks,” in *Proc. of the 2011 IEEE International Conference on Communications Workshops (ICC)*, Kyoto, Japan, 5-9 June 2011.
- [65] M. Bennis and S. M. Perlaza, “Decentralized cross-tier interference mitigation in cognitive femtocell networks,” in *Proceedings of IEEE International Conference on Communications, ICC 2011, Kyoto, Japan, 5-9 June, 2011*.
- [66] A. Imran, M. Imran, and R. Tafazolli, “A novel self organizing framework for adaptive frequency reuse and deployment in future cellular networks,” in *Personal Indoor and Mobile Radio Communications (PIMRC), 2010 IEEE 21st International Symposium on*, September 2010, pp. 2354–2359.
- [67] S. Barbarossa and F. Celano, “Self-organizing sensor networks designed as a population of mutually coupled oscillators,” 2005, pp. 475–479.
- [68] S. Barbarossa, G. Scutari, and A. Swami, “Achieving consensus in self-organizing wireless sensor networks: The impact of network topology on energy consumption,” in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 2, April 2007, pp. II-841—II-844.

-
- [69] D. López-Pérez, Á. Ladányi, A. Jüttner, and J. Zhang, “OFDMA femtocells: A self-organizing approach for frequency assignment,” in *Personal, Indoor and Mobile Radio Communications, 2009 IEEE 20th International Symposium on*, sept. 2009, pp. 2202–2207.
- [70] G. de la Roche, A. Ladányi, D. López-Pérez, C.-C. Chong, and J. Zhang, “Self-organization for LTE enterprise femtocells,” in *GLOBECOM Workshops (GC Wkshps), 2010 IEEE*, Dec. 2010, pp. 674–678.
- [71] M. Bennis, L. Giupponi, E. Diaz, M. Lalam, M. Maqbool, E. Strinati, A. De Domenico, and M. Latva-aho, “Interference management in self-organized femtocell networks: The BE-FEMTO approach,” in *Wireless Communication, Vehicular Technology, Information Theory and Aerospace Electronic Systems Technology (Wireless VITAE), 2011 2nd International Conference on*, 28 2011-march 3 2011.
- [72] FREEDOM project. [Online]. Available: <http://www.ict-freedom.eu>
- [73] ARTIST4G project. [Online]. Available: <https://ict-artist4g.eu/>
- [74] SOCRATES project. [Online]. Available: <http://fp7-socrates.org/>

Chapter 2

Scenarios and Simulation Parameters

As introduced in Chapter 1, the aim of this thesis is to give a solution for the interference problem in macrocell and femtocell coexisting systems, being compliant with the 3GPP recommendations and following the framework of the BeFEMTO European project [1], under which this thesis has been carried out.

In this chapter, the scenarios considered to validate the proposed solution are presented. First, the urban and suburban deployment models proposed by 3GPP for the study of interference between LTE-based macrocell and femtocells [2] are presented in Section 2.1. Second, the scenarios considered in the BeFEMTO project, from the femtocell business model point of view, considering the future trends of femtocell networks deployment [3], are introduced in Section 2.2. Third, Section 2.3 summarizes the system model, the two scenarios defined for simulations, i.e. single-cell and multicell scenarios, following the 3GPP recommendations, and the simulation parameters for the macrocell and femtocell systems. The single-cell scenario is a simpler scenario where the proposed solutions are proven to correctly perform, while the multicell scenario is an extension of the first one to a more complex and realistic deployment. Finally, the enhanced system functional architecture, based on 3GPP release 10 [4] EPS recommendations, including the support to the solution proposed in this thesis, is presented in Section 2.4.

2.1 3GPP scenarios

This section briefly describes the scenarios proposed by 3GPP technical report [2] for the study of interference between LTE-based macrocell and femtocell systems. In what follows, the 3GPP nomenclature is adopted, then macrocells are referred to as eNBs, femtocells as HeNBs and macro and femto users as UEs. The 3GPP technical report analyzes a suburban model and two dense-urban models, as presented below.

2.1.1 3GPP suburban modeling

A suburban model is considered when HeNBs are assumed to be deployed inside houses located within the eNB coverage area. Specifically, in this model houses are represented as a 2 dimensional squared 12×12 m shapes, as presented in Figure 2.1. Houses are deployed following a random uniform distribution within the eNB coverage area, subject to minimum separation of 35 m to the eNB and assuming a non-overlapping constraint. The HeNBs are installed randomly within houses and it is assumed that they are always “active”, i.e. there is at least one active call. The density of HeNBs per eNB is variable.

With respect to the UEs distribution, in each house the HeNB UEs are randomly dropped within a specified distance from the center of the house and with a minimum separation from the HeNB of 20 cm. HeNB UEs have a 10% probability of being outdoor around the house. All UEs associated to the eNB are assumed to be randomly dropped within the eNB coverage area and therefore they also may be inside a house containing a HeNB.

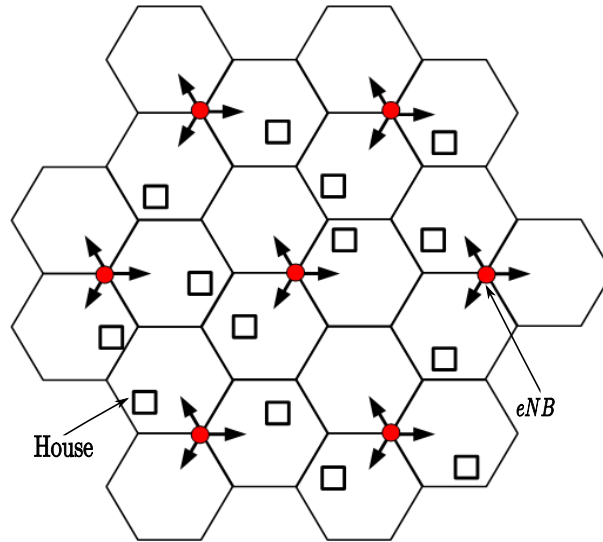


Figure 2.1: Suburban modeling.

2.1.2 3GPP dense-urban modeling

In [2] two urban models are proposed by 3GPP, i.e. Dual Stripe and 5×5 Grid models. They are described in the following subsections.

Dual Stripe model

A Dual Stripe model is defined as a dense HeNB deployment model and it considers HeNBs to be installed inside the apartments forming the block of apartments. As it is shown in Figure 2.2,

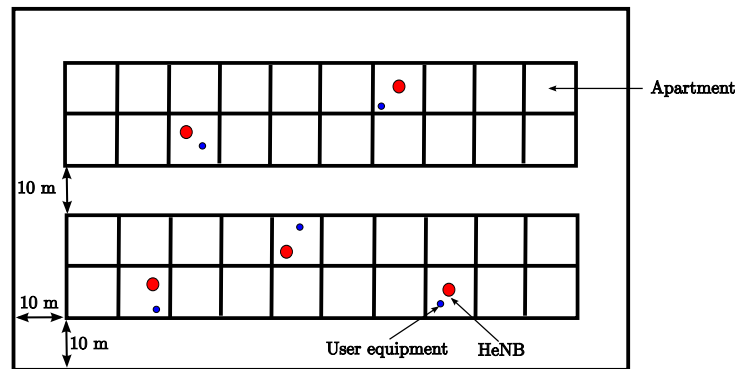


Figure 2.2: 3GPP Dual Stripe femtocell block model.

each block of apartments has 2 stripes, separated by a 10 m wide street. Each stripe has 2 rows of $A = 10$ apartments with squared form and size of 10×10 m. Around the 2 stripes there is also a 10 m wide street, so that, the total block size is $10(A + 2) \times 70$ m. The streets around the blocks of apartments and between the two stripes are necessary to make sure that the HeNBs from different femtocell blocks are not too close to each other. In each eNB sector, one or several femtocell blocks are randomly dropped assuming that the femtocell blocks are not overlapping with each other. For each femtocell block, the number of floors fl , is chosen randomly between 1 and 10.

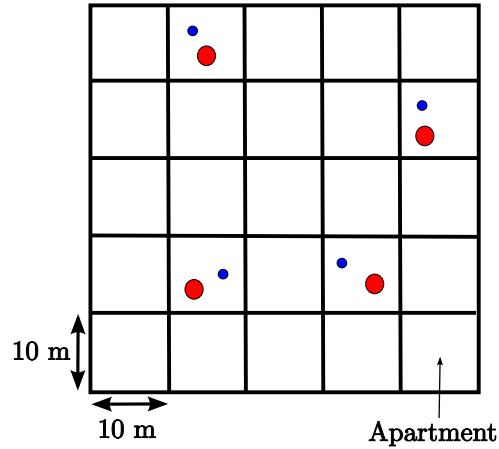
An occupation ratio, p_{oc} , which can vary from 0 to 1, is introduced and determines whether inside an apartment there is a HeNB or not. Also, each HeNB has a random activation parameter which determines the moment in which it is switched on. In Figure 2.2, active HeNBs are represented by red dots, and HeNB UEs are represented by blue dots.

HeNBs are randomly located in the apartments and it is assumed that there is only one UE per HeNB, which is dropped randomly in the active HeNB coverage area with a minimum separation from the HeNB of 20 cm. eNB UEs are dropped uniformly and randomly throughout the indoors/outdoors eNB coverage area with a minimum separation of 35 m to the eNB BSs.

5 × 5 Grid model

This model is a simpler alternative for HeNB urban deployment. HeNBs are located in apartments distributed in 5×5 grids of 25 apartments, as shown in Figure 2.3, with fl number of floors. The apartments' size is $10 \text{ m} \times 10 \text{ m}$ and are placed next to each other.

The HeNB and its associated UEs are dropped randomly and uniformly inside the apartment area. Similar to the Dual Stripe model, an occupation ratio and an activation parameter are defined. They respectively determine the presence or not of a HeNB inside an apartment and the probability whether the HeNB is active or not.

Figure 2.3: 5×5 apartment grid

2.2 HeNB networks deployment models proposed by the BeFEMTO project

The strategic scenarios for HeNB systems are studied by 3GPP, the Small Cells Forum [5] and by several European projects such as FREEDOM [6] and BeFEMTO. As already mentioned, this thesis has been developed under the framework of the BeFEMTO project, where different residential, enterprise and mobile scenario models are studied. In order to introduce the HeNB scenarios considered in the work presented in this thesis, in what follows we summarize the BeFEMTO's vision of broadband evolved HeNBs, which includes four major themes:

- *Indoor Standalone HeNBs*: This strategic scenario refers to HeNBs installed at homes located in urban and suburban areas. Standalone HeNBs work without exchanging any information with other neighboring HeNBs. HeNBs installed at homes will guarantee to the users a broadband service, similar to the one available in urban areas, thanks to their fast backhaul connection through a fix line such as DSL or optical fiber. New interesting services regarding localization, health services, advanced learning and security will be available for HeNBs owners as well as the advantage of having a reduced radiation exposure. HeNBs installed at homes, using advanced Radio Frequency (RF), interference control and RRM procedures and co-existing with eNBs are the key enablers in the achievement of a significant enhancement in the spectral efficiency [7].
- *Indoor Networked HeNBs*: Networked HeNBs are formed by networking standalone HeNBs through a X2 interface, as introduced in 3GPP TS 36.300, release 10. Each HeNB ought to be connected to at least one other HeNB. This strategic scenario is considered for large confined (indoor) spaces such as enterprises, shopping malls, airports, stadiums and dense urban HeNB deployments, enabling localized and high quality coverage and ded-

icated capacity. The Networked HeNBs are also enablers for local communications with the advantages of resource efficient communication, low latency, controlled security, and support of a high level of QoS, particularly for deployment in large indoor environments with planning restrictions.

- *Outdoor Fixed Relay HeNB*: This scenario is considered for mainly enhancing cell-edge capacity and reliable coverage. BeFEMTO proposes to use and adapt the fixed HeNB technology enhanced with self-optimization operation for fixed relay HeNB.
- *Outdoor Mobile HeNB*: BeFEMTO aims at developing a viable mobile HeNBs technology to provide large coverage and availability in environments such as public transport (buses, trains, etc). The Mobile HeNB theme enables the transport industry to offer a better user experience while on the move, as well as provisioning of local and special services.

The work presented in this thesis focuses on the Indoor Standalone HeNBs and the Indoor Networked HeNBs deployment models.

2.3 System model and simulation scenarios

In this section, first, the system model is introduced and second, the two scenarios considered to validate the algorithms proposed in this thesis are presented. A dense-urban HeNB deployment model is considered, since it is more complex in terms of interference management, where HeNBs are located following the Dual Stripe model presented in Subsection 2.1.2. First, a so called *single-cell* scenario is presented. It consists of a single eNB and a block of apartments where the HeNBs are located. Second, an extension of this scenario to a more complex one, consisting of multiple eNBs and blocks of apartments within the eNBs coverage area is presented. This is referred to as *multicell* scenario.

2.3.1 System model

In general, we consider $|\mathcal{M}| = M$ eNBs providing service to its U_m associated UEs and coexisting with N HeNBs. Each HeNB provides service to its U_f associated UEs. An OFDMA downlink is considered, where the system bandwidth BW is divided into R Resource Blocks (RBs). A RB represents one basic time-frequency unit formed by 12 subcarriers of width $\Delta f = 15 \text{ kHz}$, thus being 180 kHz wide in the frequency domain, for a duration of one slot, so it is 0.5 ms long in the time domain [8].

We assume that both eNBs and HeNBs operate in the same frequency band and have the same amount R of available RBs, which allows to increase the spectral efficiency per area through spatial frequency reuse. The work presented in this thesis focuses only on the downlink operation.

We denote by $\mathbf{p}^{f,F} = (p_1^{f,F}, \dots, p_R^{f,F})$ and $\mathbf{p}^{m,M} = (p_1^{m,M}, \dots, p_R^{m,M})$ the transmission power vector of HeNB f and eNB m with $p_r^{f,F}$ and $p_r^{m,M}$ denoting the downlink transmission power of eNB and HeNB in RB r , respectively. We assume that the R RBs in both systems are defined according to a proportional scheduling policy. The maximum transmission power for eNB and HeNB BSs are P_{\max}^M and P_{\max}^F , respectively, where $\sum_{r=1}^R p_r^{m,M} \leq P_{\max}^M$ and $\sum_{r=1}^R p_r^{f,F} \leq P_{\max}^F$.

The SINR at UE $u^m \in U_m$ allocated in RB r of macrocell m is:

$$SINR_r^m = \frac{p_r^{m,M} h_{mm,r}^{MM}}{\sum_{k=1, k \neq m}^M p_r^{k,M} h_{km,r}^{MM} + \sum_{f=1}^N p_r^{f,F} h_{fm,r}^{FM} + \sigma^2} \quad (2.1)$$

with $m = 1, \dots, M$. Here, $h_{mm,r}^{MM}$ indicates the link gain between the transmitting macro BS m and its UE u^m ; $h_{km,r}^{MM}$ denotes the link gain between the transmitting macro BS k and UE u^m in eNB m ; $h_{fm,r}^{FM}$ indicates the link gain between the transmitting HeNB f and UE u^m of eNB m ; finally, σ^2 is the noise power.

The capacity of eNB m is:

$$C^{m,M} = \sum_{r=1}^R \frac{BW}{R} \log_2 (1 + SINR_r^m) \quad (2.2)$$

with $m = 1, \dots, M$.

The SINR at UE $u^f \in U_f$ allocated in RB r of HeNB f is:

$$SINR_r^f = \frac{p_r^{f,F} h_{ff,r}^{FF}}{\sum_{m=1}^M p_r^{m,M} h_{mf,r}^{MF} + \sum_{k=1, k \neq f}^N p_r^{k,F} h_{kf,r}^{FF} + \sigma^2} \quad (2.3)$$

with $f = 1, \dots, N$. Here, $h_{ff,r}^{FF}$ denotes the link gain between the transmitting HeNB f and its UE u^f ; $h_{mf,r}^{MF}$ indicates the link gain between the eNB m and UE u^f in HeNB f and $h_{kf,r}^{FF}$ denotes the link gain between the transmitting HeNB k and UE u^f of HeNB f .

The capacity of HeNB f is:

$$C^{f,F} = \sum_{r=1}^R \frac{BW}{R} \log_2 (1 + SINR_r^f) \quad (2.4)$$

with $f = 1, \dots, N$.

Finally, the total system capacity is given by:

$$C_{Total} = \sum_{m=1}^M C^{m,M} + \sum_{f=1}^N C^{f,F} \quad (2.5)$$

2.3.2 Single-cell scenario

The so called single-cell scenario is deployed in an urban area and is modeled as shown in Figure 2.4. We consider $M = 1$ eNBs with radius $D = 500$ m and $F = 1$ blocks of apartments with $fl = 1$ floors. Each HeNB provides service to its $U_f = 1$ associated HeNB UEs, which is randomly located inside the HeNB area with a minimum separation from the HeNB of 20 cm. eNB UEs $U_m = 1$ is located outdoor and also randomly between the 2 stripes of apartments in order to consider a difficult situation in terms of interference management. We consider that HeNB UEs are always indoors.

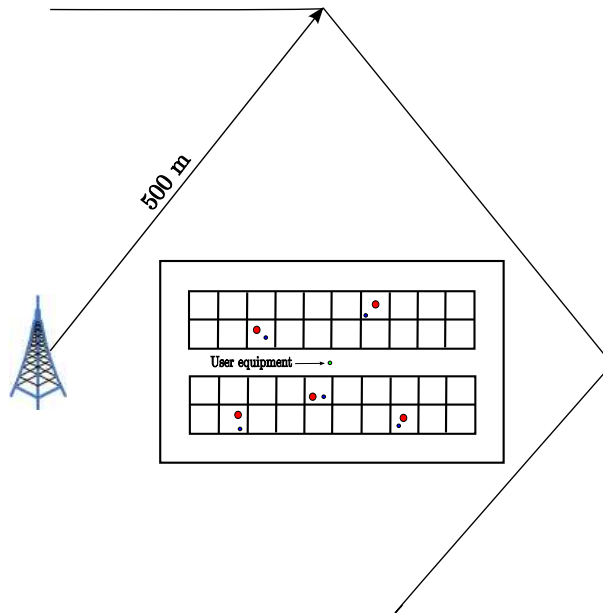


Figure 2.4: Single-cell system proposed layout.

2.3.3 Multicell scenario

The simulation area consists of a 2 tier hexagonal cell distribution, which translates into $M = 19$ eNBs with radius $D = 500$ m and $F = 10$ blocks of apartments, three of them are located in the F1 sector of the central eNB coverage area, where statistics are analyzed, and the rest are randomly located inside the other eNBs coverage area. The eNBs are placed at the junction of 3 hexagonal cells. Each cell is considered as a sector and therefore, a eNB serves 3 sectors. This scenario is deployed based on the frequency reuse scheme $1 \times 3 \times 3$ as shown in Figure 2.5. The channel is divided into 3 segments: $F1$, $F2$, $F3$, and each segment is assigned to each sector. This scheme is very simple from the operator's point of view and it mitigates intercell interference by reducing the probability of slot collision by a factor of 3. However, the sector capacity is also

reduced by a factor of 3 [9]. For each sector, the azimuth antenna pattern is modeled as [2]:

$$Az(\theta) = -\min \left[12 \left(\frac{\theta}{\theta_{3dB}} \right), Az_m \right] \quad (2.6)$$

where $\theta_{3dB} = 70^\circ$ is the angle from central lobe at which the gain reduces to half the maximum value and $Az_m = 20$ dB is the maximum possible attenuation due to sectorization.

UEs are located outdoors and randomly within the hexagonal system. Statistics for eNB system are taken from UEs served by sector $F1$ of central eNB.

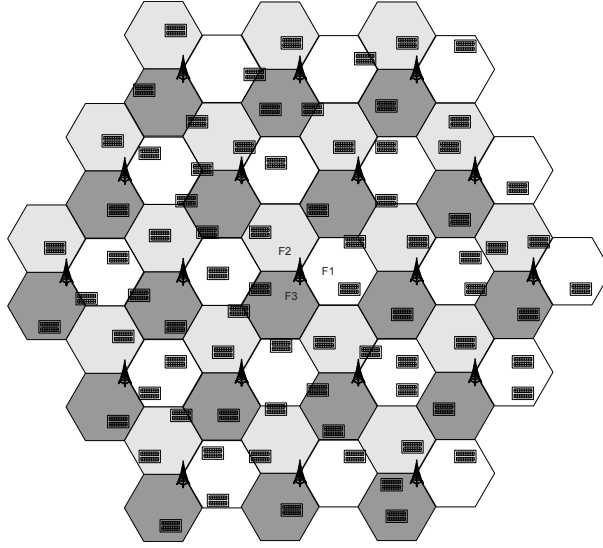


Figure 2.5: Multicell system proposed layout.

2.3.4 Simulation parameters

We consider the eNB and HeNB systems operating at 1850 MHz and to be based on LTE. Therefore, the frequency band is divided into RBs. For simulations, we consider an amount of RBs $R = 6$, which corresponds to the minimum LTE implementation and corresponds to a channel bandwidth of $BW = 1.4$ MHz. The antenna patterns for eNB, HeNB and eNB/HeNB UEs are omnidirectional, with 18 dBi, 0 dBi and 0 dBi antenna gains, respectively. The shadowing standard deviation is 8 dB and 4 dB, for eNB and HeNB systems, respectively. The eNB and HeNB noise figures are 5 dB and 8 dB, respectively. The transmission power of the eNB is 46 dBm, whereas the HeNB adjusts its power over all its RBs through the learning scheme to a total maximum value of $P_{\max}^F = 10$ dBm.

The considered PL models are for urban scenarios and are summarized in Table 2.1. Here, d and d_{indoor} are the total and indoor distances between the eNB/HeNB and the eNB/HeNB UE, respectively. The factor $0.7d_{\text{indoor}}$ takes into account the penetration losses due to the walls inside the apartments. $WP_{\text{out}} = 15$ dB and $WP_{\text{in}} = 5$ dB are the penetration losses of the building

external walls and of the walls separating the apartments, respectively. Finally, w_p represents the number of walls separating apartments.

Table 2.1: 3GPP path loss models for urban deployment.

eNB to eNB/HeNB UEs	outdoors	$PL(\text{dB}) = 15.3 + 37.6 \log_{10} d$
	indoors	$PL(\text{dB}) = 15.3 + 37.6 \log_{10} d + WP_{\text{out}}$
HeNB to eNB/HeNB UEs	UE in the same apartment stripe	$PL(\text{dB}) = 38.46 + 20 \log_{10} d + 0.7d_{\text{indoor}} + w_p WP_{\text{in}}$
	UE outside the apartment stripe	$PL(\text{dB}) = \max(15.3 + 37.6 \log_{10} d, 38.46 + 20 \log_{10} d) + 0.7d_{\text{indoor}} + w_p WP_{\text{in}} + WP_{\text{out}}$
	UE inside a dif- ferent apartment stripe	$PL(\text{dB}) = \max(15.3 + 37.6 \log_{10} d, 38.46 + 20 \log_{10} d) + 0.7d_{\text{indoor}} + w_p WP_{\text{in}} + 2 WP_{\text{out}}$

We also consider the 3GPP recommendation of frequency-selective fading model specified in [10] (urban macro settings) for eNB to UE propagation, and a spectral block fading model with coherence bandwidth 750 kHz for indoor propagation. Relevant simulation parameters for eNBs and HeNBs are summarized in Table 2.2 [2].

2.4 Functional architecture

In this section we present the EPS functional architecture proposed by 3GPP's release 10 [4]. This architecture is extended so that our proposed solution, based on learning, can be supported. To this end, one new architecture component is introduced, i.e. the Local Femtocell GateWay (LFGW) and a new interface between HeNBs and eNBs is added, i.e. the X2'.

Figure 2.6 shows in black the most important architecture components and the relationships among them, as standardized by 3GPP, and in orange the newly introduced entity and interface, as contemplated in the BeFEMTO project deliverable D 2.2 [11]. In what follows, we summarize the architectural components main functionalities.

- Packet Data Network Gateway (P-GW): The P-GW is the node that provides logical connectivity from the UE to external packet data networks by being the point of exit and entry of traffic for the UE. A UE may have simultaneous connectivity with more than one P-GW for accessing multiple packet data networks. The P-GW is responsible for anchoring the user plane mobility within the LTE/Evolved Packet Core (EPC) network as well as for inter-Radio Access Technology (RAT) handovers between 3GPP and non-3GPP technologies such as WiMAX and 3GPP2. The P-GW supports the establishment of data bearers between the Serving Gateway (S-GW) and itself and is responsible for Internet

Table 2.2: Simulation parameters

Parameter	Value
Carrier frequency	1.85 GHz
System bandwidth BW	1,4 MHz
Subframe time duration	1 ms
Number of subcarriers per RB, N_{sc}	12
Number of RBs, R	6
Thermal noise, N_0	-174 dBm/Hz
UE antenna gain	0 dBi
Outside wall penetration loss, WP_{out}	15 dB
Indoor wall penetration loss, WP_{in}	5 dB
eNB	
Inter-site distance D	500 m
Sectors per eNB	3
BS Tx power per sector	46 dBm
Elevation BS antenna gain	14 dBi
Azimuth antenna element gain	$-\min \left[12 \left(\frac{\theta}{\theta_{3dB}} \right)^2, A_m \right]$ [dB] where $A_m = 20$ and $\theta_{3dB} = 70^\circ$
Avg. UE per sector	1
Channel coherence bandwidth	750 kHz
User mobility	0 ~ 50 km/h
OFDMA scheduling	Proportional fair
Traffic model	Full buffer
Inter-cell interference modeling	Cells always active
HeNB	
Avg. dual stripe apart- ment blocks per eNB sector	1
UEs per active HeNB	2
Max. HeNB Tx power, P_{max}^F	10 dBm
HeNB antenna gain	0 dBi

Protocol (IP) address allocation for the UE, Differentiated Services Code Point (DSCP) marking of packets, traffic filtering using traffic flow templates and rate enforcement.

- Serving Gateway (S-GW): The S-GW handles the user data plane functionality and is involved in the routing and forwarding of data packets from the EPC to the P-GW via the S5 interface. The S-GW is connected to the eNB via an S1-U interface which provides user plane tunneling and inter-eNB handovers (in coordination with the MME). The S-GW also performs mobility anchoring for intra-3GPP mobility. Unlike with a P-GW, a UE is associated to only one S-GW at any point in time.
- Mobility Management Entity (MME): The MME is a node in the EPC that handles mobility related signalling functionality. It is the key control-node for the LTE access-network.

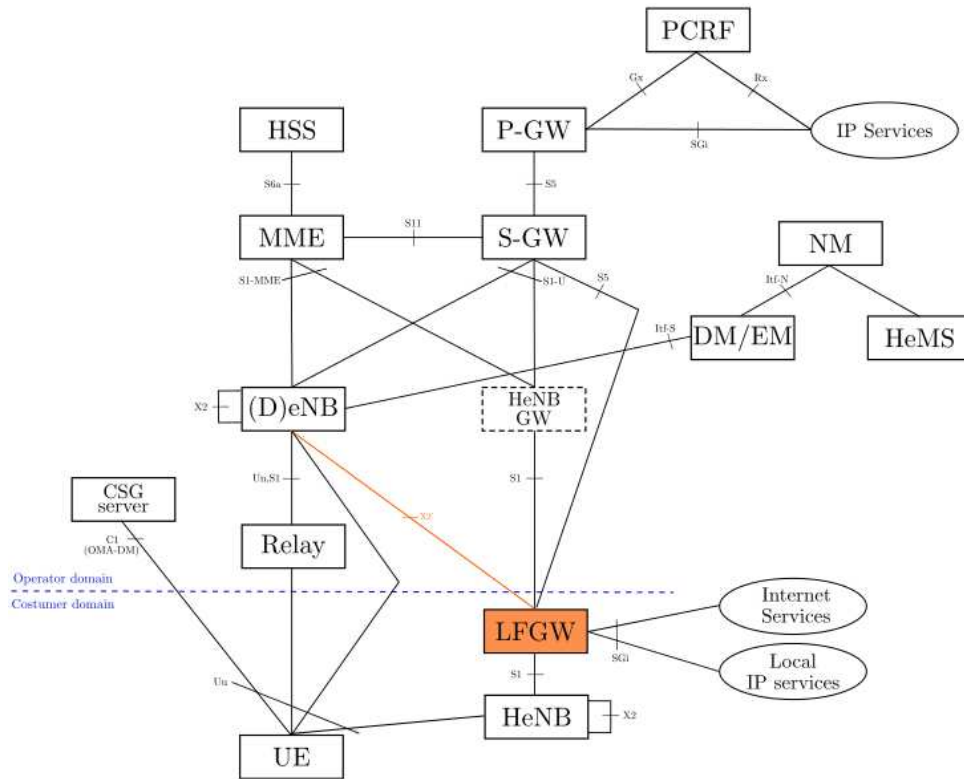


Figure 2.6: LTE system architecture including enhancement for learning support.

Specifically, the MME tracks and maintains the current location of UEs allowing the MME to easily page a mobile node. The MME is also responsible for managing UE identities and controls security both between the UE and the eNB (Access Stratum (AS) security) and between UE and MME (Non-Access Stratum (NAS) security). It is also responsible for handling mobility related signalling between UE and MME (NAS signalling).

- (Donor) evolved NodeB ((D)eNB): The Evolved Universal Terrestrial Radio Access Network (E-UTRAN) architecture consists of multiple eNBs and RRM is the main functionality of each eNB. This functionality includes radio bearer control, radio admission control and scheduling and radio resource allocation for both the uplink and downlink. The eNB is also responsible for the transfer of paging messages to the UEs and header compression and encryption of the user data. eNBs are interconnected by the X2 interface and connected to the MME and the S-GW by the S1-MME and the S1-U interface, respectively. An eNB is called a DeNB if it controls one or more relays.
- HeNB GW: The HeNB GW is an optional gateway through which HeNBs access the core network using the S1 interface. The HeNB GW may also be used only for the S1-MME interface. In this case, the S1-U interface is directly between the HeNB and the S-GW.
- HeNB: A HeNB is a customer-premises equipment that connects a 3GPP UE over the

E-UTRAN wireless air interface (Uu interface) and to an operator's network using a broadband IP backhaul. Similar to the eNBs, RRM is a main functionality of a HeNB.

- **User Equipment (UE):** A UE is a device that connects wirelessly over the E-UTRAN wireless air interface (Uu interface) to a cell of a (D)eNB or a HeNB.
- **Relay:** A relay is a node that is wirelessly connected to a Donor evolved NodeB (DeNB) and relays traffic from UEs to and from that DeNB. The relay can be a cell on its own as is the case in LTE release 10 or a node that only supports a reduced protocol stack of an eNB.
- **Policy and Charging Rule Function (PCRF):** The PCRF functionalities include policy control decisions and flow-based charging control. The PCRF is the main QoS control entity in the network and is responsible for building the policy rules that will apply to user services and passing these rules to the P-GW.
- **Home Subscriber Server (HSS):** The HSS is a user database that stores subscription-related information to support other call control and session management entities. It also stores user identification, numbering and service profiles. It is mainly involved in user authentication and authorization.
- **Closed Subscriber Group Server (CSG Server):** The CSG server hosts functions used by the subscriber to manage membership to different CSGs. For example, the CSG server includes the UE CSG provisioning functions which manage the allowed CSG list and the operator CSG list stored on the UE.
- **Network Management (NM):** The NM is the main controlling entity of the OA&M part and is responsible for managing the network, mainly as supported by the Element Managements (EMs) but it may also involve direct access to the network elements. All communication with the network is based on open and well-standardized interfaces supporting management of multi-vendor and multi-technology network elements.
- **Domain Management / Element Management (DM/EM):** The Domain Management (DM) provides element management functions and domain management functions for a sub-network. Inter-working DMs provide multi-vendor and multi-technology network management functions. The EM provides a package of end-user functions for management of a set of closely related types of network elements. These functions can be divided into two main categories: EM Functions and Sub-Network Management Functions.
- **HeNB Management System (HMS):** The HMS assumes either the role of an initial HMS (optional) or of a serving HMS. The initial HMS may be used to perform identity and location verification of a HeNB and assigns the appropriate serving HMS, security gateway

and HeNB GW or MME to the HeNB. The serving HMS supports identity verification of the HeNB and may also support HeNB GW or MME discovery. Typically, the serving HMS is located inside the operator's secure network domain and the address of the serving HMS is provided to the HeNB via the initial HMS.

In what follows, the new LFGW functional entity main functionalities are presented. Also, the added logical interface is introduced.

- **Local Femtocell GateWay (LFGW)**: The LFGW is a functional entity deployed within a HeNB network, specifically within a customer premises. Similar to a HeNB GW, it can serve as a concentrator for S1 interfaces, and can also serve as a local mobility anchor, a local mobility control entity and central local breakout point for Local IP Access (LIPA) and Selected IP Traffic Offload (SIPTO). It further supports local routing and load balancing and may act as a HeNB controller for centralized radio resource and interference management support.
- **X2'**: The X2' interface logically connects eNB and HeNB with each other, analogous to the X2 interface between eNBs. It is introduced to support interference management and mobility optimization between eNBs and HeNBs. Some issues regarding the X2' interface require further study, i.e. 1) the interface should be direct or via the LFGW, 2) the LFGW would just concentrate the connections and provide some isolation between domains or also translate to the S1 interface towards the HeNBs and, 3) the functionality of the X2' and the X2 interfaces should be identical or the X2' interface is enhanced (e.g., for Over-The-Air operation) or reduced.

In femtocell systems, the RRM tasks could be implemented in decentralized and centralized forms. Decentralized solutions are those cases where RRM algorithms are implemented at femto node level. On the other hand, centralized RRM could be implemented at LFGW level. This appears as a good solution since the LFGW has a wider picture of the situation in the system, in comparison to the femto node. Nevertheless, in femtocell systems, decentralized RRM algorithms appear as the most desirable solutions since they allow femtocells to take decisions in a completely autonomous way.

2.5 Conclusions

This chapter gives a general overview of the suburban and dense-deployed scenarios considered by 3GPP standardization body to study the interference in HeNB/eNB heterogeneous systems. Also, it includes a brief summary of HeNBs use cases considered by the BeFEMTO project giving

an overview of future business models for HeNB systems. Based on this, two scenarios have been presented, the single-cell and the multicell scenarios which are proposed to be deployed following the Indoor Standalone HeNBs and the Indoor Networked HeNBs models. Also, the simulation parameters considered for the validation of the solutions proposed in this thesis have been introduced. Finally, the EPS functional architecture proposed by 3GPP in release 10 and enhanced with the LFGW functional entity and the X2' entity to support the interference control learning algorithm proposed in this thesis, have been presented.

Bibliography

- [1] The BeFEMTO website. [Online]. Available: <http://www.ict-befemto.eu/>
- [2] 3GPP, “3GPP R4-092042 TSG RAN WG4 (Radio) Meeting 51: Simulation assumptions and parameters for FDD HeNB RF requirements,” Tech. Rep., 4-8 May 2009.
- [3] “D2.1: Description of baseline reference systems, use cases, requirements, evaluation and impact on business model,” EU FP7-ICT BeFEMTO project, Dec. 2010.
- [4] 3GPP, “3GPP TS 25.401 V 10.2.0 , UTRAN overall description (release 10),” Tech. Rep., June 2011.
- [5] Small Cell Forum. [Online]. Available: <http://www.smallcellforum.org/>
- [6] FREEDOM project. [Online]. Available: <http://www.ict-freedom.eu>
- [7] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, “Femtocell networks: A survey,” *IEEE Communication Magazine*, vol. 46, no. 9, pp. 59–67, Sept. 2008.
- [8] S. Sesia, I. Toufik, and M. Baker, *LTE, The UMTS Long Term Evolution: From Theory to Practice*. Wiley Publishing, 2009.
- [9] D. López-Pérez, A. Jüttner, and J. Zhang, “Dynamic frequency planning versus frequency reuse schemes in OFDMA networks,” in *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*, 26-29 April 2009.
- [10] J. Salo, G. Del Galdo, J. Salmi, P. Kysti, M. Milojevic, D. Laselva, and C. Schneider, “MATLAB implementation of the 3GPP Spatial Channel Model (3GPP TR 25.996),” Online, Jan. 2005, <http://www.tkk.fi/Units/Radio/scm/>.
- [11] “D2.2: The BeFEMTO system architecture,” EU FP7-ICT BeFEMTO project, Dec. 2011.

Chapter 3

Multiagent time difference methods: study and design for interference control

Femtocell networks have to be considered as decentralized systems due to their deployment model, i.e. femtocell BSs are installed by the end consumer and therefore their location, amount and impact, in terms of interference to the macrocell system, are unknown to the operator [1]. As a result, femtocells are required to be able to evolve self-constructed coherent behaviors in accordance with the environment and as autonomously as possible. To this end, self-organization has recently been associated with capabilities that femtocells have to include [2]. Self-organization is defined as the ability of entities to spontaneously arrange given parameters following some constraints and without any human intervention. To do this, entities have to represent somehow the environment where they perform and the gathered information has to be interpreted for them to correctly react [3]. Learning algorithms appear as a logical interpretation of self-organization since, through them, the environmental sensed information can be translated into actions. ML is the branch of Artificial Intelligence (AI) discipline concerning the design of algorithms able to evolve behaviors based on empirical information, as sample data from sensors or past experience from databases [4].

In terms of learning, decentralized systems are interpreted as multiagent systems. Multiagent systems are a reasonable form to solve complex, large and unpredictable problems since they offer modularity given by the implementation of different agents in the system. Each of these agents is specialized at solving a specific problem, having accurate and local reactions. Multiagent systems can be *deliberative*, when a model can be formalized for each agent behavior in terms of beliefs, desires and goals, or *reactive*, when agents cannot have an environment representation and act using stimulus-response type of behavior [5]. When interdependent and dynamic problems are considered (i.e. when the structure of the system dynamically changes),

as it is the case of femtocell systems, reactive agents are the more adequate solution. As a form to implement reactive agents, ML introduces the concept of RL, which works based on learning from interactions with the environment, and on the observed consequences when a given action is executed. To solve RL problems there are three fundamental classes of methods, i.e. dynamic programming, Monte Carlo methods and TD learning.

Dynamic programming methods are mathematically developed, but they require a complete and exact model of the dynamics of the environment to be analytically solved. On the other hand, Monte Carlo methods do not require a complete model, but depend upon a model of sample transitions which is built based on sequences of states, actions and returns from online or simulated interactions with the environment. Monte Carlo methods are suitable for episodic tasks and the gathered experience is updated at the end of each episode. Finally, TD learning processes do not require any environmental dynamic model and knowledge is updated in each learning step. They are more complex to analyze than Monte Carlo methods [6], but require less memory and computation. We focus on the study of TD learning methods since in the problem to face in this thesis an incremental learning method able to adapt to the environment online and without environmental models is required. In fact, it would be highly complicated to construct a comprehensive model of a realistic wireless context.

In literature, RL is often proposed to solve problems related to RRM in centralized architectures, e.g., dynamic channel allocation [7], multirate transmission control [8], call admission control [9]. Recently, the solution of RRM problems by means of decentralized approaches based on multiagent RL has been receiving a growing interest in our community. In [10], the authors present a channel and power algorithm selection, based on the construction of behavioral rules. A frequency resource selection approach, based on Multi-Armed Bandit (MAB) formulation is introduced in [11]. Reference [12] presents a spectrum sensing optimization for cognitive radio systems based on a two stages RL approach. In [13], the authors propose a solution for femtocell interference mitigation based on stochastic approximation. Power allocation strategies following a heterogeneous and delayed learning are presented in [14].

This chapter presents a study to select one TD learning method to control the aggregated interference generated from the femtocells to the macrocells. It also includes an analysis regarding the selection of three important parameters in the learning method design, i.e. learning rate, discount factor and tradeoff between exploration and exploitation. The chapter is structured as follows. First, a brief introduction to learning in single and multiagent systems concept is given in Section 3.1. Then, TD algorithms are studied in Section 3.2. Finally, an analysis of the TD parameters for the interference control problem is presented in Section 3.3.

3.1 ML overview

ML is related to the theory of unsupervised learning, supervised learning and RL, where intelligence is centralized in a single agent or distributed across a multiagent system.

- *Supervised learning*: is the most common and successful ML technique so forth. The task of the learner is to predict the value of the outcome for any valid input object after having seen a number of training examples. The training examples are pairs of input objects, usually vectors, and desired outputs. If desired outputs are continuous, the learning problem is a regression problem. On the other hand, if the desired outputs are discrete values, it is a classification problem. Once the training process has finished, the learning approach predicts the value of the function for any unseen valid input object [4]. This method is called “supervised” learning because of the presence of the variable outcome to guide the learning process.
- *Unsupervised learning*: these methods objective is to learn to represent statistical structures of unlabeled input patterns. The inputs to the system are usually assumed to be independent samples of an underlying unknown probability distribution. The learned statistical structures are then used by the system to reconstruct patterns from noisy input data. This form of learning is called “unsupervised” learning because of the lack of explicit target outputs, as required in supervised learning, or environmental-based reward, as in RL, to evaluate a potential solution [15]. Unsupervised learning may be divided into two types of problems, data clustering and feature extraction. Data clustering, aims to unravel the structure of the provided data set. Feature extraction, on the other hand, often seeks to reduce the dimensionality of the data so as to provide a more compact representation of the data set.
- *Reinforcement Learning*: The ability of learning new behaviors online and automatically adapting to the temporal dynamics of the system is commonly associated with RL [16]. At each time step, the agent perceives the state of the environment and takes an action to transit in a new state. A scalar cost is received, which evaluates the quality of the selected action and its impact in the agents’ environment [17]. RL is applied for constructing autonomous systems that improve themselves with experience.

Supervised and unsupervised methods are not suitable for interactive problems where agents must learn from their own experience, which is the context where the problem we aim to solve falls. The majority of studies that can be encountered in RRM literature, applying RL techniques, are formulated for centralized settings where all decisions are taken by a single entity e.g., Radio Network Controller (RNC) in UTRAN systems. Femtocell systems cannot be formulated through

a centralized learning process due to their deployment model, scalability and signaling overhead constraints. We therefore focus on decentralized learning processes based on RL.

The topic of learning in distributed systems has actually been studied in game theory since 1951 when Brown proposed the fictitious play algorithm [18]. The underlying assumption of fictitious play is that an agent assumes that its opponents sample the actions from some fixed distribution, i.e. at each time step opponents use a stationary mixed strategy. Then, each agent in the game estimates its opponents strategies by keeping a score of the appearance frequencies of the different actions. This means that, each agent needs to know the strategies followed by the other players in the game.

In ML, the literature of single agent learning is extremely rich, while it is only in recent years that attention has been focused on distributed learning aspects, in the context of multiagent learning. It has been yielding some enticing results, being arguably a truly interdisciplinary area and the most significant interaction point between computer science and game theory communities. The theoretical framework to formulate RL problems can be found in MDPs for the single agent system, and in stochastic games, for a multiagent system. In what follows, we give a brief introduction of learning in single and multiagent systems.

3.1.1 Learning in single-agent systems

A MDP provides a mathematical framework for modeling decision-making processes in situations where outcomes are partly random and partly under the control of the decision maker. A MDP is a discrete time stochastic optimal control problem. Here, operators take the form of actions, i.e. inputs to a dynamic system, which probabilistically determine successor states. A MDP is defined in terms of a discrete-time stochastic dynamic system with finite state set $\mathcal{S} = \{s_1, \dots, s_k\}$. Time is represented by a sequence of time steps, $t = 0, 1, \dots, \infty$. At each time step, a controller observes the system's current state and selects an action, which is executed by being applied as input to the system. Let us assume that s is the observed state, and that the action is selected from a finite set of admissible actions $\mathcal{A} = \{a_1, \dots, a_l\}$. When the controller executes action $a \in \mathcal{A}$, the system state at the next step changes from s to v , with a state transition probability $P_{s,v}$. We further assume that the application of action a in state s incurs an immediate cost $c(s, a)$. When necessary, we refer to states, actions, and immediate costs by the time steps at which they occur, by using s_t , a_t and c_t , where $a_t \in \mathcal{A}$, $s_t \in \mathcal{S}$ and $c_t = c(s_t, a_t)$ are, respectively, the state, action and cost at time step t . A graphic representation of the learner-environment interaction is shown in Figure 3.1, here by the use of c_{t+1} and s_{t+1} we aim to emphasize that, as consequence of the performed action at time t , a_t , in the next time step, the agent receives a cost c_{t+1} and finds itself in a new state $v = s_{t+1}$. To sum up, a MDP consists of:

- a set of states \mathcal{S} .

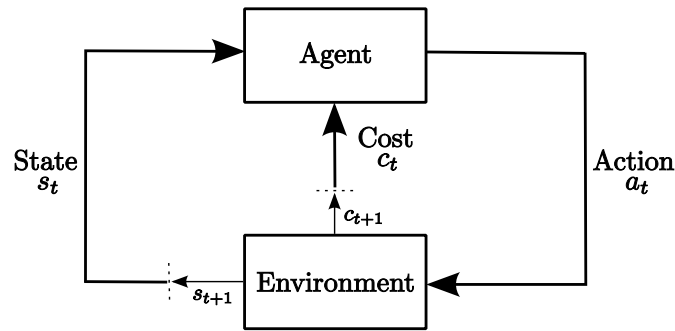


Figure 3.1: Learner-environment interaction.

- a set of actions \mathcal{A} .
- a cost function $C : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$.
- a state transition function $P : \mathcal{S} \times \mathcal{A} \rightarrow \Pi(\mathcal{S})$, where a member of $\Pi(\mathcal{S})$ is a probability distribution over the set \mathcal{S} (i.e. it maps states to probabilities).

The state transition function probabilistically specifies the next state of the environment as a function of its current state and the agent's action. The cost function specifies expected instantaneous cost as a function of current state and action. The model is a *Markov* model if the state transitions are independent of any previous environment states or agent actions. The objective of the MDP is to find a policy that minimizes the cost of each state s_t . As a result, the aim is to find an optimal policy for the infinite-horizon discounted model, relying on the result that, in this case, there exists an optimal deterministic stationary policy [16].

RL problems model the world using MDP formulism. In the literature, three ways have been identified to solve RL problems. The first one consists of the knowledge of the state transition probability function from state s to state v , $P_{s,v}(a)$, and is based on dynamic programming. The second and third forms to solve RL problems, on the other hand, do not rely on this previous knowledge and are based on Monte Carlo and TD methods. As a result, Monte Carlo and TD are primarily concerned with how an agent ought to take actions in an environment so as to minimize the notion of long-term cost, that is, so as to obtain the optimal policy, when the state transition probabilities are not known in advance. When state transition probability is not known, but a sample transition model of states, actions and costs can be built, Monte Carlo methods can be applied to solve the MDP problem. On the other hand, if the only way to collect information about the environment is to interact with it, TD methods have to be applied. TD methods combine elements of dynamic programming and Monte Carlo ideas, they learn directly from experience which is a characteristic of Monte Carlo methods and they gradually update prior estimate values, which is common of dynamic programming. TD methods allow an online learning which is crucial for long-term/continuous applications.

RL algorithms are based on the computation of *value functions*, i.e. the state-value function, $V(s)$, or the state-action value function, $Q(s, a)$, which measure how good, based on the future expected cost, is for an agent to be in a given state or to execute an action in a given state, respectively. The expected costs for the agent in the future are given by the actions it will take and therefore, the value functions depend on the policies being followed. The state-value of state s is defined as the expected infinite discounted sum of costs that the agent gains if it starts in state s and then executes the complete decision policy π ,

$$V^\pi(s) = \mathbb{E}_\pi \left\{ \sum_{t=0}^{\infty} \gamma^t c_t \mid s_t = s \right\} \quad (3.1)$$

where $0 \leq \gamma < 1$ is a discount factor which determines how much expected future costs affect decisions made now.

Similarly, the Q-value $Q(s, a)$ represents the expected decreased cost for executing action a at state s and then following policy π thereafter.

$$Q^\pi(s, a) = \mathbb{E}_\pi \left\{ \sum_{t=0}^{\infty} \gamma^t c_t \mid s_t = s, a_t = a \right\} \quad (3.2)$$

Solving a RL problem means to find the best return in the long term. This is defined as finding an optimal policy, which is the one giving minimum expected return. We define the optimal value of state s as:

$$V^*(s) = \min_{\pi} V^\pi(s) \quad (3.3)$$

According to the principle of Bellman's optimality [16], the optimal value function is unique and can be defined as the solution to the equation:

$$V^*(s) = \min_a \left(C(s, a) + \gamma \sum_{v \in \mathcal{S}} P_{s,v}(a) V^*(v) \right) \quad (3.4)$$

which asserts that the value of state s is the expected cost $C(s, a) = \mathbb{E}\{c(s, a)\}$, plus the expected discounted value of the next state, v , using the best available action. Given the optimal value function, we can specify the optimal policy as:

$$\pi^*(s) = \arg \min_a \left(C(s, a) + \gamma \sum_{v \in \mathcal{S}} P_{s,v}(a) V^*(v) \right) \quad (3.5)$$

Applying the Bellman's criterion in the action-value function, first we have to find an intermediate minimum of $Q(s, a)$, denoted by $Q^*(s, a)$, where the intermediate evaluation function for every possible next state-action pair (v, a') is minimized, and the optimal action is performed with respect to each next state v . $Q^*(s, a)$ is:

$$Q^*(s, a) = C(s, a) + \gamma \sum_{v \in \mathcal{S}} P_{s,v}(a) \min_{a' \in \mathcal{A}} Q^*(v, a') \quad (3.6)$$

Then, we can determine the optimal action a^* with respect to the current state s . In other words, we can determine π^* . Therefore, $Q^*(s, a^*)$ is minimum, and can be expressed as:

$$Q^*(s, a^*) = \min_{a \in \mathcal{A}} Q^*(s, a) \quad (3.7)$$

3.1.2 Learning for multiagent systems

The characteristics of the distributed learning systems, as mentioned in the introduction, are as follows: i) the intelligent decisions are made by multiple intelligent and uncoordinated nodes; ii) the nodes partially observe the overall scenario; and iii) their inputs to the intelligent decision process are different from node to node since they come from spatially distributed sources of information. These characteristics can be easily mapped onto a multiagent system, where each node is an independent intelligent agent. The theoretical framework is found in stochastic games [19] described by the five-tuple $\{\mathcal{N}; \mathcal{S}; \mathcal{A}; P; C\}$. Here, $|\mathcal{N}| = N$ is the set of agents, indexed 1, 2, . . . , N ; $\mathcal{S} = \{s_1, s_2, \dots, s_k\}$ is the set of possible states, or equivalently, a set of N -agent stage games; \mathcal{A} is the joint action space defined by the product set $\mathcal{A}^1 \times \mathcal{A}^2 \times \dots \times \mathcal{A}^N$, where $\mathcal{A}^f = \{a_1^f, a_2^f, \dots, a_l^f\}$ is the set of actions (or pure strategies) available to the f -th agent; P is a probabilistic transition function defining the probability of migrating from one state to another provided the execution of a certain joint action or, equivalently, it specifies the probability of the next stage game to be played based on the game just played and the actions taken in it; $C = \{c^1 \times c^2 \times \dots \times c^N\}$, where c^f is the immediate cost of the f -th agent in a certain stage of the game, which is a function of the joint actions of all N nodes [20].

In multiagent systems, the distributed decisions made by the multiple nodes strongly interact among each other. These kind of problems are usually modeled as non-cooperative games. The simplest and most common interpretation of a non-cooperative game is that there is a single interaction among players (“one-shot”), after which the payoffs are decided and the game ends. However, many, if not all strategic endeavors occur over time, and in a state dependant manner. That is, the games, and so the environment in which the nodes make decisions progress over time, passing through an infinite number of states, and the current game is decided based on the history of the interactions. Stochastic games form a natural model for such interactions [21]. A stochastic game is played over a state space, and is played in rounds. In each round, each player chooses an available action simultaneously with and independently from all other players, and the game moves to a new state under a possible probabilistic transition relation based on the current state and the joint actions [19]. We distinguish in this context two different forms of learning. On the one hand, the agent can learn the opponent’s strategies, so that it can then devise a best response. Alternatively, the agent can learn a strategy of his own that does well against the opponents, without explicitly learning the opponent’s strategies. The first approach is sometimes referred to as model-based learning, and it requires at least some partial information of the other players strategies. The second approach is referred to as model-free learning, and it

does not necessarily require to learn a model of the strategies played by the other players.

We will discuss in the following a very partial sample of multiagent learning techniques, which we consider representative for the aim of this taxonomy:

- *Model-based approaches:* This approach, generally adopted in game theory literature, is based on building some model of the other agents strategies, following which, the node can compute and play the best response strategy. This model is then updated based on the observations of their actions. As a result, these approaches require knowledge or observability of the other agents' strategies, which may pose severe limits from the feasibility point of view in terms of information availability and signalling overhead. The best known instance of this scheme is fictitious play [18], which is a static game that simply counts the plays of the other agents in the past. Different variations of the original schemes exist, for example those considering that the agent does not play the exact best response, but assigns a probability of playing each action. Other algorithms in literature that can be classified into this group are the Metastrategy [22] and the Hyper-Q algorithms [23]. An example of a stochastic game approach in this category is the Non-stationary Converging Policies (NSCP) game [24].
- *Model-free approaches:* A completely different approach, commonly considered by the AI literature, is the model-free approach, also known as TD learning, which avoids building explicit models of other agents' strategies. Instead, over time, each agent learns how properly the various available actions work in the different states. TD methods typically keep memory of the appropriateness of playing each action in a given state by means of some representation mechanism, e.g., lookup tables, neural networks, etc. This approach follows the general framework of RL and has its roots in the Bellman equations [16]. TD methods can be roughly classified into two groups, i.e. on-policy and off-policy methods. On-policy methods learn the value of the policy that is used to make decisions and off-policy methods can learn about policies other than that currently followed by the agent [6].

We focus on model-free learning techniques, which avoid building explicit models of other agents' strategies, opposite to model-based approaches, where nodes compute and play the best response following a previously constructed model of the other agents' strategies. In model-free algorithms, each agent learns over time how properly the various available actions work in the different states. Among the different TD methods, we study the off-policy algorithm, Q-learning, and the on-policy method, Sarsa. Both approaches are proven to converge to an optimal policy in single agent systems, as long as the learning period is long enough, i.e. after a training period, the optimal solution for the given situation can be chosen in one-shot, which allows to have highly responsive approaches after a short period of learning [25].

Q-learning and Sarsa algorithms, typically used in centralized settings, can be extended to the multiagent stochastic game by having each agent simply ignore the other agents and pretend that the environment is stationary. Even if this approach has been shown to correctly behave in many applications, it is not characterized by a strict proof of convergence, since it ignores the multiagent nature of the environment and the Q-values are updated without regard for the actions selected by the other agents. A first step in addressing this problem is to define the Q-values as a function of all the agents' actions; however, the rule to update the Q-values is not easy to define in this more complicated case. Littman in [26] suggests the minimax Q-learning algorithm for a two-player zero-sum repeated game. Later works, such as the joint action learners [27] and the Friend or Foe Q-algorithm [28], propose other update rules for the particular case of common payoff games. More recent efforts can be found for the more general case of general-sum games [29], however, the problem still remains not efficiently solved. Other algorithms falling in this category are the correlated equilibrium Q-learning [30], the asymmetric Q-learning [31], the regret minimization approaches [32, 33], etc.

3.2 TD learning methods

TD learning, formally defined by R.S. Sutton in [34], is a prediction method based on the future values of a given signal. The name TD comes from the use of the differences in predictions over successive time steps to drive the learning process [6]. Agents implementing TD methods are naturally implemented in an online fashion, i.e. agents learn from every transition without considering the subsequent actions. Therefore, after a short period of training, agents implementing TD methods rapidly improve their behavior, improvements that continue with time and, in our particular case, translate in decreasing damage to macrocell users from the earliest moments of the learning process.

As presented in Figure 3.2, which summarizes what has been discussed in previous sections, we consider the femtocell system as a distributed system given its deployment model. Distributed systems are commonly modeled by means of stochastic games, where players select their actions independently from the other agents in the system. Since we aim to perform an online learning in such a way that agents can adapt to the environmental changes automatically, we formulate the problem through RL. In the problem we are considering, building explicit models of other agents' strategies is highly complex. The solution is then found through model-free techniques, also known as TD learning methods, which construct the knowledge based on experience.

One of the intrinsic challenges of TD algorithms, and of the RL methods in general, is the tradeoff between exploration and exploitation. Exploration is necessary to guarantee learning across all available actions, and therefore that the best ones are selected at the end of the learning process. Exploitation refers to the use by the agent of the knowledge already acquired to obtain

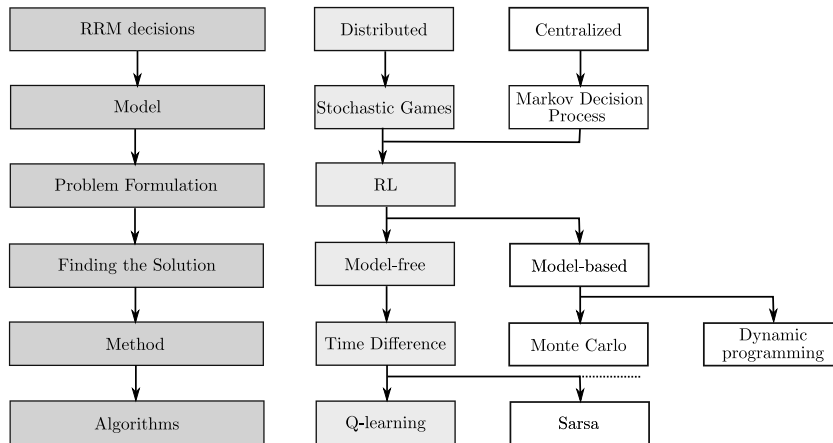


Figure 3.2: Taxonomy of the formulation of the given problem.

reward. To guarantee optimal behaviors all actions have to be infinitely often selected.

A policy maps states to actions, hence, it determines the actions the agent will follow depending on the given situation. Policies can be classified into two groups, 1) the behavior policy, which determines the conduct of the agent, i.e. the actual action selected by the agent in the current state and 2) the estimation policy, which determines the policy evaluated, or the action in the next state used for the evaluation of behaviour policy. In RL there are two methods to ensure a sufficient exploration, the on-policy and off-policy methods, which basically differ on the form they select the estimation policy.

- *On-policy methods*: These methods evaluate or improve the policy, π , used to perform the decisions. In other words, they estimate the value of a policy while using it for control. This means that, the policy followed by the agent to select its behavior in a given state (behavior policy) is the same used to select the action (estimation policy) based on which it evaluates the behavior followed.
- *Off-policy methods*: These methods do distinguish between behavior and estimation policies. Therefore, the policy to generate behavior, π is unrelated to the policy evaluated [6]. The policy evaluated in off-policy methods is the one corresponding to the best action in the next state, π^* , given the current agent experience.

In TD learning, agents attempt to select actions that minimize the discounted costs they receive over the future, which is why the discount rate, γ , is introduced in the state value function, equation (3.1) and in the state-action value function, equation (3.2). The new information in the state or state-action value update, is weighted through the learning rate, α . The correct selection of these parameters highly influences the performance of the learning process. Another important characteristic of TD methods is the action selection policy, which gives the agent the

criterion to follow when selecting an action. The criterion can be to perform exploration or to exploit the already acquired knowledge. Exploration have to be included in the action selection policies in order to achieve good behaviors based on explicit trial-and-error processes.

3.2.1 Study of TD Q-learning and Sarsa methods

Q-learning and Sarsa are based on the estimation of state-action value function, $Q(s, a)$, therefore they consider transitions form state-action pair to state-action pair. Learning is performed by iteratively updating the Q-values, which represent the expert knowledge of the agent, and have to be stored in a representation mechanism [6]. The most intuitive and common representation mechanism is the lookup table, so that, in our case, the TD methods represent their Q-values in a Q-table, whose dimension depends on the size of the state and action sets, $k \times l$, as depicted in Figure 3.3.

	a_1	...	a_l
s_1	$Q(s_1, a_1)$		$Q(s_1, a_l)$
s_2	$Q(s_2, a_1)$		$Q(s_2, a_l)$
\vdots			
s_k	$Q(s_k, a_1)$		$Q(s_k, a_l)$

Figure 3.3: Q-table representation.

The Q-learning and Sarsa approaches are based on an iterative algorithm which consists in estimating the Q-values on the run. The estimation for agent f is performed as follows: (1) all Q-values are initialized to an arbitrary number H ; (2) the agent measures the current state of the environment; (3) then, it selects the action a following the action selection policy; (4) the agent executes the selected action, which causes a transition to a new state; (5) the environment sends a feedback c to the agent; (6) the agent updates $Q(s, a)$; (7) the cycle is repeated from step (2). The processes from step (2) to (6) form a learning iteration. In what follows we present in more details both Q-learning and Sarsa algorithms.

Q-learning procedure

Q-learning algorithm was first introduced in 1989 by Watkins in his Ph.D., thesis [35] and the proof of convergence of this algorithm was presented later by Watkins and Dayan in [25]. The Q-learning process tries to find $Q^*(s, a)$ in a recursive manner using available information (s, a, v, c) , where s and v are the states at time t and $t + 1$, respectively; and a and c are the action taken at time t and the immediate cost due to executing a at s , respectively. Q-learning algorithm is an off-policy algorithm, it estimates π^* while following π , as it is shown in the right side of Figure 3.4. This means that the behavior of the agent is determined by the action selection policy followed by it, which is represented by policy π , while the Q-value update process is performed based on the minimum Q-value in the next state, independently of the policy being followed [6]. Then, the minimum Q-value is the optimal policy π^* , as expressed in equation (3.9). Algorithm 1 presents the Q-learning procedure in a formal form. The Q-learning rule to update the Q-values:

$$Q(s, a) \leftarrow Q(s, a) + \Delta Q(s, a) \quad (3.8)$$

where $\Delta Q(s, a)$ is:

$$\Delta Q(s, a) = \alpha [c + \gamma \min_a Q(v, a) - Q(s, a)] \quad (3.9)$$

where α is the learning rate, which weights the importance given to the information observed after executing action a .

The advantage of off-policy methods is that they do not include the cost of exploration in the Q-value update. This characteristic makes this form of learning consistent with the principle of knowledge exploitation, i.e. after the learning process ends the policy found by the algorithm is applied without including exploration. We can conclude that agents applying off-policy learning exploit the acquired knowledge in a very effective way since the beginning of the learning process.

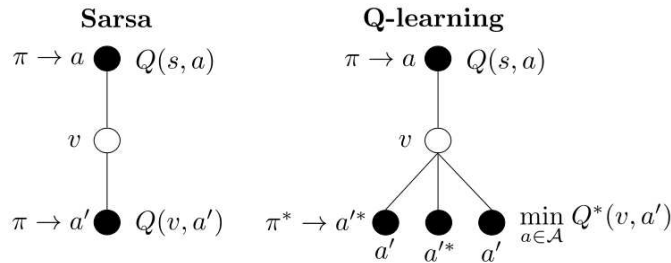


Figure 3.4: Policy diagram for Q-learning and Sarsa.

Sarsa procedure

Sarsa learning algorithm follows the same logic as the Q-learning process. It tries to find $Q^*(s, a)$ in a recursive manner using available information (s, a, c, v, a') , where a' is the action taken at time $t + 1$, following the action selection policy, π . Sarsa is an on-policy learning method, meaning

Algorithm 1 Q-learning

Initialize:**for** each $s \in \mathcal{S}$, $a \in \mathcal{A}$ **do** initialize $Q(s, a) \leftarrow H$ **end for**evaluate the starting state s **Learning:****loop** select the action a following the action selection policy execute a receive an immediate cost c observe the next state v select the action a' corresponding to the minimum Q-value in v

update the table entry as follows:

$$Q(s, a) \leftarrow Q(s, a) + \alpha[c + \gamma \min_a Q(v, a') - Q(s, a)]$$

 update $s = v$ **end loop**

that it estimates π while following π , as shown in the left side of Figure 3.4. Therefore, it evaluates or improves the followed policy in every Q-value update, and the update process is strictly based on experience [6]. This can be seen in the Sarsa updating rule (3.10), where the action for the next state, a' , and, therefore, its associated state-action value, $Q(v, a')$, is determined by π , the action selection policy the agent follows. Algorithm 2 presents the pseudocode of Sarsa procedure. The Sarsa learning rule to update the Q-values differs from equation (3.9) in the computation of $\Delta Q(s, a)$ as:

$$\Delta Q(s, a) = \alpha[c + \gamma Q(v, a') - Q(s, a)] \quad (3.10)$$

The advantage of an on-policy algorithm is that it optimizes the same policy that it follows, so that, the policy followed by the agent will be more effective. This form of knowledge update gives agents applying Sarsa a safer learning behavior. However, after the learning period ends, exploration is removed from the action selection policy, this means that the agent will follow π^* . In on-policy learning, this does not correspond with the policy optimized during the learning process, which always follows the action selection policy π . This makes it inconsistent with the knowledge exploitation principle followed after the learning period ends.

3.3 Q-learning and Sarsa for interference control

In this section, a comparison in terms of convergence and system performance, between the Sarsa and the Q-learning approaches, is presented for the interference control between femtocells and macrocells. To this end, we first briefly introduce the states, actions and cost used in the learning algorithm. This case study is presented more in detail in next chapter, Section 4.1,

Algorithm 2 Sarsa learning**Initialize:****for** each $s \in \mathcal{S}$, $a \in \mathcal{A}$ **do** initialize $Q(s, a) \leftarrow H$ **end for**evaluate the starting state s **Learning:****loop** select the action a following the action selection policy execute a receive an immediate cost c observe the next state v select the action a' following the action selection policy

update the table entry as follows:

$$Q(s, a) \leftarrow Q(s, a) + \alpha[c + \gamma Q(v, a') - Q(s, a)]$$

 update $s = v$ **end loop**

together with other case studies used across this thesis. Also, results obtained for the learning algorithm configuration in order to select important learning parameters, such as learning rate, discount factor and action selection policy, are presented.

We define system state, actions, associated cost and next state for agent f and RB r as follows:

State: The system state for agent f and RB r is defined as:

$$\tilde{s}_r^f = \{Pow^f, \bar{C}_r^m, \bar{C}_r^f\} \quad (3.11)$$

where Pow^f indicates the femtocell total transmission power over all RBs. \bar{C}_r^m indicates whether the capacity of macrouser u^m in RB r of macrocell m most affected by the activity of femtocell f , is above or below the macrocell capacity threshold C_{\min}^M . Finally, \bar{C}_r^f is the femtocell f capacity indicator.

Actions: The set of possible actions are the l power levels that femtocell f can assign to RB r .

Cost: The cost c assesses the immediate return incurred due to the assignment of action a at state s . The considered cost function is:

$$c = \begin{cases} K & Pow^f > P_{\max}^F \text{ or } C_r^m < C_{\min}^M \\ K \exp^{-C_r^f} & \text{otherwise} \end{cases} \quad (3.12)$$

where K is a constant value. The rationale behind this cost function is that the total transmission power of each femtocell does not exceed the allowed P_{\max}^F , the capacity of the macrocell does not fall below a target C_{\min}^M , and the capacity of the femtocells is maximized.

Next State: The state transition from s to v in RB r is determined by the learning-based power allocation.

3.3.1 Q-learning and Sarsa comparison

In what follows we present some simulation results in order to compare Q-learning and Sarsa learning methods. The following results have been obtained for the multicell scenario, presented in Section 2.3.3, with an occupation ratio of $p_{oc} = 45\%$ and a required C_{min}^M per RB of 1.2 Mbit/s. In the following results, exploration has been removed after the 80% of the learning iterations, which corresponds to iteration 220000.

First, a comparison between average cost values obtained by Q-learning and Sarsa algorithms is presented in Figure 3.5. Here, it is shown how both proposed approaches reduce their received cost value during the learning period, as it is required by their cost functions. Another important aspect to highlight in this figure is that, after a first period of learning (e.g., 100000 iterations), Q-learning approach does not decrease any longer its average cost value and incurs in some oscillations due to the exploration process. On the other hand, Sarsa approach smoothly decreases the average cost because its learning process is more conservative. After the exploration process is removed, both learning processes rapidly decrease their cost values. However, it can be observed that Q-learning average cost decreases faster than Sarsa's, since the received cost value, after power and macrocell capacity constraints are met, depends on the achieved capacity by the femtocell, then, when Q-learning is applied, agents' learned optimal policies correspond to higher transmission power levels, than when Sarsa is applied. This is because cost function used in the learning systems are the same but not the learnt policies. So, the policies learnt by Q-learning are able not only to guarantee the constraints, but also to achieve higher femtocell systems capacities.

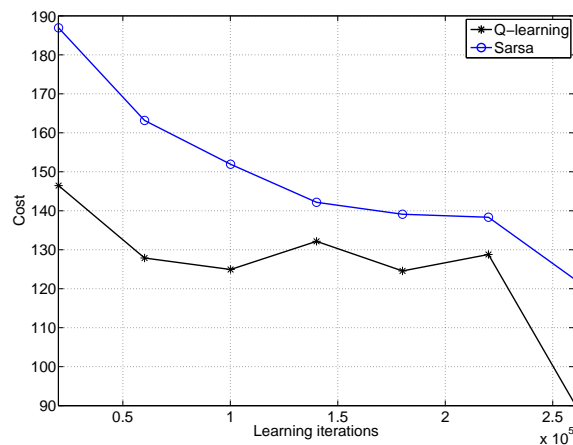


Figure 3.5: Average cost for Q-learning and Sarsa methods during the learning iterations.

Figure 3.6 depicts the macrocell and the average femtocell systems capacity during the learning period. In this plot, it is clearly reflected that Sarsa algorithm offers a safer learning process, which translates in a more conservative action selection in order to guarantee less failures, i.e. not to fulfil the constraints imposed by the cost function, in the long term. As it can be observed, the different learning methods highly impact the system performance. At the end of the learning iterations, when Sarsa algorithm is applied, the macrocell system has a capacity 0.45 Mbit/s higher than the Q-learning case. Here, it is worth noting that even though when applying Q-learning, macrocell system performance is lower than that obtained by Sarsa, Q-learning approach still fulfils the required macrocell performance constraints introduced in the learning system design. On the other hand, Q-learning allows femtocells to have an average capacity 2.5 Mbit/s higher than the Sarsa method, which translates in a higher system capacity.

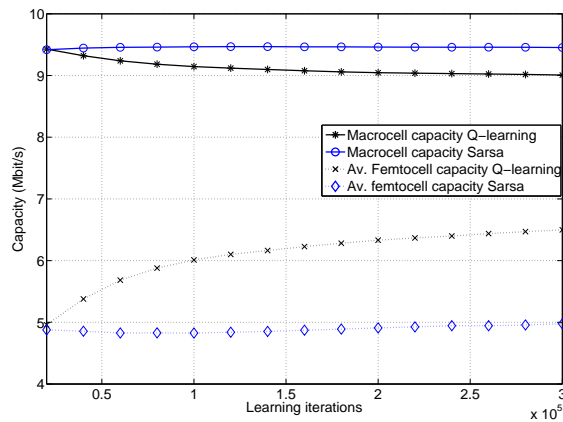


Figure 3.6: Macrocell and average femtocell capacity for Q-learning and Sarsa methods during the learning iterations.

In the following results, and across the rest of this thesis, we will make use of the following definitions:

Definition 1 The probability of being below the capacity threshold *is the average probability over the R RBs that the macrocell capacity is below C_{\min}^M .*

Definition 2 The probability of being above the power threshold *is the average probability that the femtocells total transmission power, Pow^f , is above P_{\max}^F .*

Figure 3.7 shows the average probability of being below the capacity threshold and the probability of being above the total power threshold for Q-learning and Sarsa. As it was expected, both probabilities decrease faster for the Q-learning approach due to its off-policy nature, which exploits the acquired knowledge in a more active form. On the other hand, in the long-term, Sarsa approach has an average better performance than Q-learning due to its on-policy methodology,

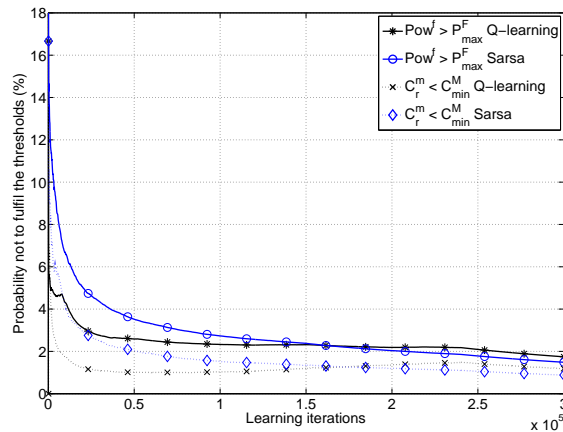


Figure 3.7: Probability of being below the capacity and above the power thresholds for Q-learning and Sarsa learning approaches.

which consists in less risky decisions, i.e. it selects actions which will guarantee to fulfil the constraints. More in details, Sarsa has a lower probability than Q-learning of not to fulfil the thresholds (i.e. in 0.26% and 0.3%, respectively), for both capacity and power, at the end of the learning process. On the other hand, after 50000 learning iterations, Q-learning approach presents a stable behavior and has a probability of not to fulfil the threshold of 1% less than Sarsa, for both, the macrocell capacity and femtocells total transmission power. Sarsa reaches this probability after 150000 learning iterations.

Based on the previous results, Q-learning algorithm is selected because it allows to guarantee i) a macrocell system capacity above the C_{\min}^M threshold and ii) a femtocell total transmission power below the P_{\max}^F threshold since early stages of the learning process. This will guarantee less harmful interference at macrousers and therefore a better coexistence of the underlay systems. Furthermore, even if the macrocell capacity achieved by the Q-learning process is lower than the one provided by the Sarsa method, it highly increases the femtocell systems performance, which implies a remarkable gain in terms of total system capacity.

3.3.2 Parameters selection for Q-learning algorithm

The main advantage of Q-learning algorithm is that it allows agents in the system to perform an online and off-policy learning (i.e. agents learn through real time interactions with the environment) based on optimal policy exploitation. Therefore, after a short period of training, agents rapidly adapt their transmission powers not to damage macrocell users. In what follows, we perform an analysis regarding three key design aspects in TD methods, the discount factor, the learning rate selection and the exploitation/exploration calibration.

Discount factor and learning rate selection

In this section we present some simulations in order to select the discount factor, γ , and the learning rate, α , of the proposed decentralized Q-learning algorithm. As it was explained before, RL algorithms objective is to minimize the expected discount cost they receive in an infinite time horizon. To this end, the discount factor is introduced to weight the importance of future returns in the current learning process. This means that costs received t steps ahead in the future are weighted less than those received in the present, by a factor of γ^t . If $\gamma = 0$, the agent is said to be “myopic”. This means that it will focus on minimizing immediate cost and in the update of the knowledge, i.e. in the Q-value update, it will only consider the current received cost. When γ is close to 1, the agent seeks for long-term low costs.

The learning rate α , is an important characteristic of the TD methods. This rate determines the weight given to the newly gathered information for state s , after executing action a , i.e. the next state transition consequence, and how it will modify the one currently stored in $Q(s, a)$. If $\alpha = 0$, the agent will not learn anything, i.e. there will not be modification in the already stored Q-value, while $\alpha = 1$ will make the agent only consider the new information.

Figures 3.8 and 3.9 show the Cumulative Distribution Function (CDF) of the probability of being below the capacity and above the power thresholds, respectively, for different value combinations of γ and α . We present results in two plots. In the left-hand side subplot, we fix the discount factor at $\gamma = 0.9$ and we plot the curves for $\alpha = \{0.1, 0.3, 0.5, 0.7, 0.9\}$. In the right-hand side subplot we do the opposite, we fix the learning rate at $\alpha = 0.5$ and we plot the CDF for $\gamma = \{0.1, 0.3, 0.5, 0.7, 0.9\}$. As it can be observed, the combination $\gamma = 0.9$ and $\alpha = 0.5$ gives the lower probability of being above the total femto transmission power and below the macrocell capacity conditions during the learning period.

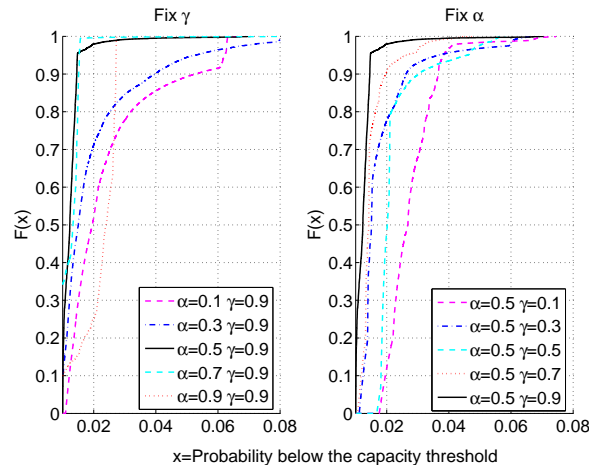


Figure 3.8: CDF of the probability of being below the capacity threshold for different α and γ .

Figures 3.10 and 3.11 show the probability of being below the capacity and above the power

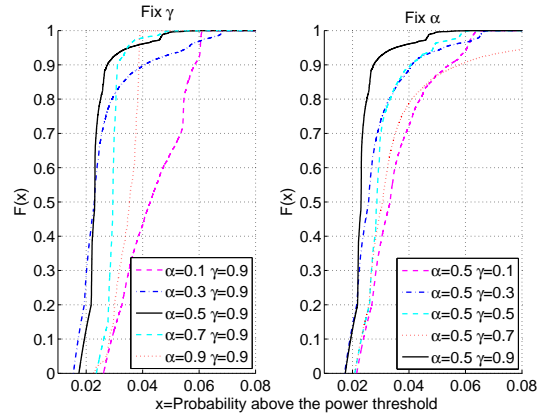


Figure 3.9: CDF of the probability of being above the power threshold for different α and γ .

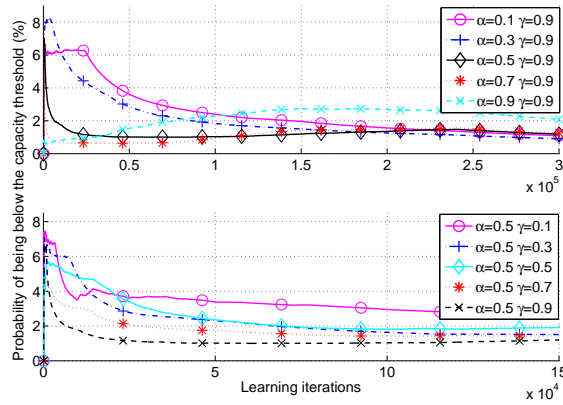


Figure 3.10: Probability of being below the capacity threshold as a function of the learning iterations for different α and γ .

thresholds as a function of the learning iterations, respectively. Again, we divide the results in two subplots, the first one contains the combinations $\gamma = 0.9$ and $\alpha = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and the second one shows the plots for combinations $\alpha = 0.5$ and $\gamma = \{0.1, 0.3, 0.5, 0.7, 0.9\}$. From these figures it can be deduced that when $\gamma = 0.9$, learning algorithms with low α , i.e. 0.1 and 0.3, require a longer learning period with respect to those learning processes with higher α and therefore the performance of the macrocell system will be jeopardized during longer periods. On the other hand, for those learning processes with high α , i.e. 0.7 and 0.9, the learning is faster but more inaccurate in the long-term since they do not consider enough the past experience, which makes the learning being instable with the time. For the case of fix $\alpha = 0.5$, when the discount factor is low, convergence is slower than for high discount values, since in the knowledge construction, the long-term learning is considered in lower degree. To sum up, $\gamma = 0.9$ and $\alpha = 0.5$ are a good combination for the case study, since they provide a good tradeoff between convergence to stable behaviors from early stages of the learning process and good performance in the long-term.

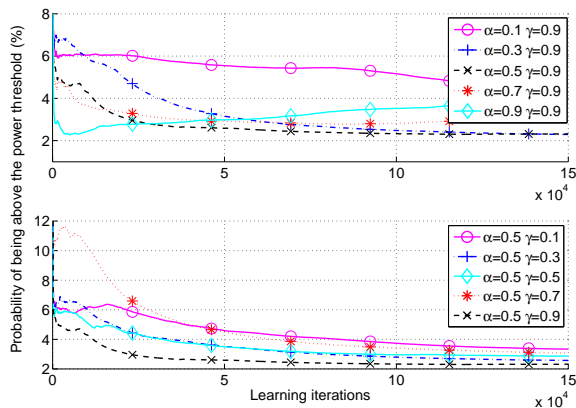


Figure 3.11: Probability of being above the power threshold as a function of the learning iterations for different α and γ values.

Action selection policy

In each iteration, the agent chooses one of the available actions in the row of the Q-table for the given state, following an action selection policy. An online learning algorithm must effectively balance the exploration and exploitation in the action selection policy due to its trial and error nature, where selected actions are evaluated through training information. The exploration is the search for better policies and it has to be included in order to guarantee that, at the end of the learning process, the agent knows which are the best and the worst available actions. On the other hand, exploitation refers to the use of the current acquired knowledge in the given state to select the best policy to minimize the received cost [6]. We compare four methodologies; three of them are based on the ε -greedy action selection policy and the last one is a softmax action selection policy.

- *ε -greedy constant:* In general, the ε -greedy action selection policy chooses the action a associated with the minimum Q-value with probability $1 - \varepsilon$, and with probability ε the action is selected randomly. In particular, the proposed ε -greedy constant policy selects actions randomly, with a probability $\varepsilon=6\%$, during the first 80% of the learning iterations. During the final 20% of the learning iterations, the algorithm only exploits the acquired knowledge.
- *ε -greedy two steps:* This action selection policy randomly chooses actions during the first 40% of the learning iterations with a probability $\varepsilon=6\%$. Then, the probability ε is decreased to 3% during the following 40% of the learning iterations. During the final 20% of the learning iterations, the algorithm exploits the acquired knowledge.
- *Time-based:* This action selection policy starts with a probability $\varepsilon=6\%$ of visiting random states. This probability uniformly decreases with time as a function of the learning

iterations.

- *Softmax*: The uniformly random action selection can be a drawback of the previous defined methods since the worst possible action is as likely as one of the best ones. A solution to this problem is to assign a weight in the form of selection probability to each one of the available actions, according to their Q-value. We apply the Boltzmann method, according to which, actions are selected randomly with regard to their associated probabilities, meaning that worst actions are unlikely to be chosen. The probability to choose an action is given by:

$$pbb(s, a_i) = \frac{e^{-Q(s, a_i)/\tau}}{\sum_{a \in \mathcal{A}} e^{-Q(s, a_i)/\tau}} \quad (3.13)$$

where τ is a positive parameter called temperature. High τ cause actions to have nearly the same probabilities to be chosen, while low τ values cause a big difference in selection probabilities depending on the actions associated Q-values [6].

Figures 3.12 and 3.13 show the probability of being below the capacity and above the power thresholds, respectively, for the four studied action selection policies. The ε -greedy two steps action selection policy better performs than the other three studied policies, since after a period (i.e. the first 40% of the learning iterations), decreasing the probability of visiting random states allows the agents to better exploit the acquired knowledge, which is more appropriate for that stage of the learning process. The time-based action selection policy presents a good behavior after the first 30% of learning iterations, but this is not the case at the beginning of the learning process, therefore we do not consider it as a suitable action selection policy because it results in higher damages to the macrouersers performance than the ε -greedy methods. Finally, the softmax policy becomes highly greedy after a point, which makes it not appropriate in the given case study due to its dependence on the Q-values, which in our case may significantly vary.

3.4 Conclusions

Due to the decentralized deployment nature of femtocell systems, in this chapter we have proposed to model them as a self-organized system following the concept of multiagent systems. To achieve the desired self-organization, we rely on the theory of model-free learning. We studied two TD learning approaches, one on-policy method, the Sarsa learning and one off-policy method, the Q-learning. We have concluded that, given the interference control problem we are dealing with, the most appropriate learning method to apply is the Q-learning approach as it allows to cause less damage to macrouersers performance since early stages of the learning process and generates higher system performance results. Furthermore, in this chapter an analysis regarding the selection of three important RL parameters, i.e. discount factor, γ , learning rate, α ,

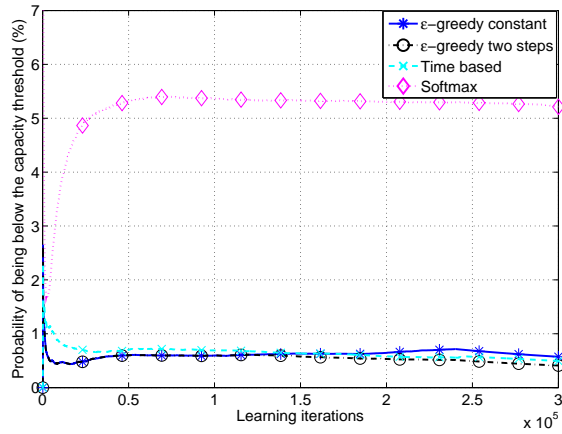


Figure 3.12: Probability of being below the capacity threshold as a function of the learning iterations for different action selection policies.

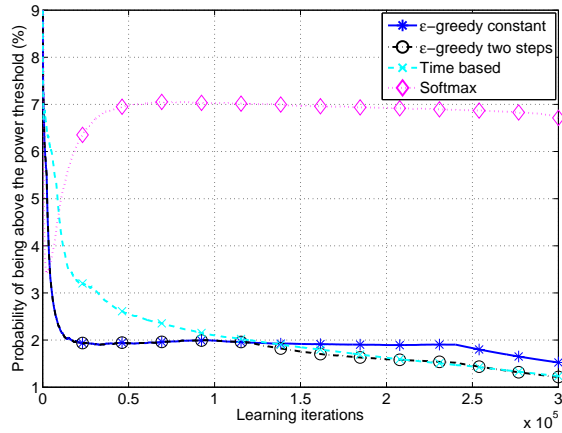


Figure 3.13: Probability of being above the power threshold as a function of the learning iterations for different action selection policies.

and action selection policy, has been presented. The results of the study leads to the following design choices, we propose to apply the Q-learning approach with $\gamma = 0.9$, $\alpha = 0.5$ and the ϵ -greedy two-steps action selection policy.

Bibliography

- [1] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: A survey," *IEEE Communication Magazine*, vol. 46, no. 9, pp. 59–67, Sept. 2008.
- [2] The BeFEMTO website. [Online]. Available: <http://www.ict-befemto.eu/>
- [3] C. Prehofer and C. Bettstetter, "Self-organization in communication networks: principles and design paradigms," *IEEE Communications Magazine*, vol. 43, no. 7, pp. 78–85, July 2005.
- [4] P. Clark, *Machine Learning: Techniques and Recent Developments*, A. R. Mirzai. London, UK: Chapman and Hall ed. Artificial intelligence: concepts and applications in engineering, 1990.
- [5] K. P. Sycara, "Multiagent systems," *AI Magazine*, vol. 19, no. 2, pp. 79–92, 1998.
- [6] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. The MIT Press, 1998.
- [7] J. Nie and S. Haykin, "A Q-learning-based dynamic channel assignment technique for mobile communication systems," *In IEEE Transactions on Vehicular Technology*, vol. 48, no. 5, pp. 1676–1687, Sept 1999.
- [8] Y.-S. Chen, C.-J. Chang, and F.-C. Ren, "Q-learning-based multirate transmission control scheme for RRM in multimedia WCDMA systems," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 1, Jan. 2004.
- [9] E.-S. El-Alfy, Y.-D. Yao, and H. Heffes, "Autonomous call admission control with prioritized handoff in cellular networks," in *In Proc. of IEEE International Conference on Communications, IEEE ICC 2001*, 11-14 June 2001.
- [10] S. M. Perlaza, S. Lasaulce, H. Tembine, and M. Debbah, "Learning to use the spectrum in self-configuring heterogeneous networks," in *4th International ICST Workshop on Game Theory in Communications Networks (GAMECOMM), Paris, France*, May 2011.
- [11] A. Feki and V. Capdevielle, "Autonomous resource allocation for dense LTE networks: A multi armed bandit formulation," in *22nd Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'11)*, 2011, pp. 66–70.
- [12] T. Jiang, D. Grace, and Y. Liu, "Two stage reinforcement learning based cognitive radio with exploration control," *IET Communications (COM-2009-0803.R2)*, vol. 5, pp. 644–651, October 2010.

- [13] M. Bennis and S. M. Perlaza, "Decentralized cross-tier interference mitigation in cognitive femtocell networks," in *Proceedings of IEEE International Conference on Communications, ICC 2011, Kyoto, Japan, 5-9 June, 2011*.
- [14] H. Tembine, A. Kobbane, and M. E. Koutbi, "Robust power allocation games under channel uncertainty and time delays," in *Wireless Days (WD), 2010 IFIP*, October 2010.
- [15] P. Dayan, *Unsupervised learning*. The MIT Encyclopedia of the Cognitive Sciences, 1999.
- [16] R. Bellman, *Dynamic Programming*. Princeton, NJ: Princeton Univ. Press, 1957.
- [17] M. E. Harmon and S. S. Harmon, "Reinforcement learning: A tutorial," 2000. [Online]. Available: <http://www.nbu.bg/cogs/events/2000/Readings/Petrov/rltutorial.pdf>
- [18] G. W. Brown, *Iterative solution of games by fictitious play*, in: *Activity Analysis of Production and Allocation*. New York: John Wiley and Sons, 1951, ch. 24, pp. 374–376.
- [19] D. Fudenberg and D. K. Levine, *The Theory of Learning in Games*. The MIT Press, June 1998, vol. 1.
- [20] P. Hoen and K. Tuyls, "Analyzing multi-agent reinforcement learning using evolutionary dynamics," in *In Proc. of the 15th European Conference on Machine Learning (ECML)*, June 2004.
- [21] E. Altman and G. Koole, "Stochastic scheduling games with markov decision arrival processes," *Journal Computers and Mathematics with Appl*, vol. 26, no. 6, pp. 141–148, 1993.
- [22] R. Powers and Y. Shoham, "New criteria and a new algorithm for learning in multi-agent systems," in *In Proc. of Advances in Neural Information Processing Systems (NISP2004)*, 13-18 Dec. 2004, pp. 1089–1096.
- [23] G. Tesauro, "Extending Q-learning to general adaptive multi-agent systems," in *In Proc. of Advances in Neural Information Processing Systems (NISP2003)*, 8-13 Dec. 2003, pp. 1089–1096.
- [24] M. Weinberg and J. S. Rosenschein, "Best response multi-agent learning in non stationary environments," in *In Proc. of 3rd International Joint Conference on Autonomous Agents and Multi agent systems (AAMAS 2004)*, 19-23 Aug. 2004, pp. 506–513.
- [25] C. J. Watkins and P. Dayan, "Technical note: Q-learning," *Machine Learning*, vol. 8, pp. 279–292, 1992.
- [26] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *In Proc. of the 11th International Conference on machine Learning*, 1994, pp. 157–163.

-
- [27] C. Claus and C. Boutilier, “The dynamics of reinforcement learning in cooperative multi-agent systems,” in *In Proc. of the 15th national Conference on Artificial Intelligence*, 1998, pp. 746–752.
- [28] M. L. Littman, “Friend-or-foe Q-learning in general-sum games,” in *In Proc. of the 18th International Conference on machine Learning*, 2001, pp. 322–328.
- [29] J. Hu and M. P. Wellman, “Nash Q-learning for general-sum stochastic games,” *Journal on Machine Learning Research*, vol. 4, pp. 1039–1069, 2003.
- [30] A. Greenwald and K. Hall, “Correlated Q-learning,” in *In Proc. of 20th International Conference on Machine Learning (ICML 2003)*, 21-24 Aug. 2003, pp. 242–249.
- [31] V. Könönen, “Asymmetric multiagent reinforcement learning,” in *In Proc. of IEEE/WIC International Conference on Intelligent Agent Technology (IAT 2003)*, 13-17 Oct. 2003, pp. 336–342.
- [32] A. Greenwald and A. Jafari, “A class of no-regret algorithms and game-theoretic equilibria,” in *In Proc. of 2003 Computational Learning Theory Conference*, 2003, pp. 1–11.
- [33] Y. Freund and R. Schapire, “A decision-theoretic generalization of online learning and an application to boosting,” in *In Proc. of 2nd European Computational Learning Theory Conference*, 1995, pp. 23–37.
- [34] R. S. Sutton, “Learning to predict by the methods of temporal differences,” in *MACHINE LEARNING*, 3. Kluwer Academic Publishers, 1988, pp. 9–44.
- [35] C. J. Watkins, “Learning from delayed rewards,” Ph.D. dissertation, Cambridge University, 1989.

Chapter 4

Multiagent Q-learning for interference control

As discussed in the previous chapter, femtocells are proposed to be modeled as agents with self-organization capabilities such that, the femtocell system can be considered as a decentralized multiagent system. Self-organization is applied through Q-learning, which is gaining in popularity in our field as an adequate and reliable learning approach. In [1] the authors proposed a solution for dynamic channel assignment in mobile communication systems. Reference [2] presents a Q-learning-based multirate transmission control for RRM to enhance spectrum utilization while meeting the QoS requirements of heterogeneous services. A distributed channel and power allocation is proposed in [3] to deal with co-channel interference heterogeneous networks. Power assignment policies in cognitive wireless mesh networks considering the secondary user SINR requirements and the energy efficiency are learnt based on Q-learning in [4]. Reference [5] presents a link adaptation solution through Adaptive Modulation and Coding (AMC) based on Q-learning for Orthogonal Frequency Division Multiplexing (OFDM) wireless systems.

By applying the Q-learning algorithm proposed in this thesis, femtocells can autonomously select their transmission power per RB, as a function of the current state of the environment and following the objectives dictated by the cost function, which in our case would be directly related with the aggregated interference generated by the multiple femtocells at macrousers. Compact state representation is one key issue in learning algorithms. States have to be as compact and precise as possible in terms of crucial information for the agent. The same occurs with the cost function, it has to precisely map how good is for one agent (and for the system, when referred to multiagent structures) to execute one action in a given state [6]. Cost defines the agent goals according to the states the agent can perceive and the actions it can execute.

In order for femtocells to have an idea of their impact to the macrocell system performance, the RB r proposed state representation, contains an indicator that gives a notion about the amount of interference perceived by the macrouser allocated in RB r . This indicator is assumed

to be conveyed through the X2' interface (see Section 2.4) between the macrocells and the femtocells [7]. A study regarding the introduction of the proposed scheme into 3GPP systems is carried out in Section 4.3. Furthermore, a strategy to handle the multiuser scheduling, based on the so called transfer learning [8], is also proposed. In LTE systems, the user allocation may change in multiples of a Scheduling Block (SB) duration, i.e. every 1 ms [9], which consists of two consecutive RBs, whereas the latency induced by the X2 interface is on average 10 ms. Consequently, the femto nodes can only react to changes of the macrouser resource allocation with a significant delay. The proposed transfer learning mechanism would allow macrouser to suffer less damage from femtocell nodes since agents will be able to carry out the learning process continuously while a macrouser has a session open, regardless of the scheduling process in the macrocell.

Section 4.1 presents the different state and cost representations used across this thesis to then continue in Section 4.2 with interesting simulation results for the single-cell and the multicell scenarios. Section 4.3 introduces the proposed solution to exchange the required information between macrocells and femtocells in 3GPP systems to perform the learning approach and to handle the multiuser scheduling.

4.1 Learning algorithm details

In this section we first define the states, actions and cost function, introduced in Section 3.1.1, which define the proposed learning algorithm. We present two case studies, starting from a simpler case and then increasing the complexity in the state representation and in the cost function objectives. The state and cost, used in case study 1, have been designed in order to highlight the ability of the learning approach to converge to desired SINR values. On the other hand, case study 2 selected state representation and cost function, look for an accurate control in the macrocell and femtocell systems performance.

4.1.1 Learning design: case study 1

In our system the multiple agents with learning capabilities are the femtocell BSs, so that for each RB they are in charge of identifying the current environment state, select the action based on the action selection policy and execute it. In the following, for each agent $f = 1, 2, \dots, N$ and RBs $r = 1, 2, \dots, R$ we define system state, action, associated cost and next state.

State: The system state for agent f and RB r is defined as:

$$\tilde{s}_r^f = \{\bar{I}_r^m, P_{ow}^f\} \quad (4.1)$$

Notice that, to fit the state \tilde{s}_r^f into a Q-table, the entries of (4.1) need to be quantized to a

k -dimensional state space \mathcal{S} . This is done by using multiple ranges of values for each component of the state, in such a way that each combination will represent a state $s \in \mathcal{S}$.

- **Macrocell interference indicator:** $\bar{I}_r^m \in I$ represents a binary indicator to specify whether the femtocell system is generating aggregated interference above or below the macrouser required threshold. This measure is based on the SINR value computed at the macrouser allocated at RB r . The set of possible values is based on:

$$\bar{I}_r^m = \begin{cases} 1 & \text{if } SINR_r^m < SINR_{Th}^M, \\ 0 & \text{otherwise} \end{cases} \quad (4.2)$$

where $SINR_r^m$ is the instantaneous SINR at the macrouser u^m allocated in RB r and $SINR_{Th}^M$ represents the minimum value of SINR that can be perceived by macrousers.

- **Femtocell total transmission power indicator:** One of the requirements of femtocells is to transmit at low-power levels, for this reason we include in the state definition the \bar{Pow}^f indicator, in order to guarantee that the femtocell total transmission power over all RBs, Pow^f , is below the threshold P_{\max}^F . It is given by:

$$Pow^f = \sum_{r=1}^R p_r^f \quad (4.3)$$

As a result, \bar{Pow}^f is a binary indicator defined as follows:

$$\bar{Pow}^f = \begin{cases} 1 & \text{if } Pow^f < P_{\max}^F, \\ 0 & \text{otherwise} \end{cases} \quad (4.4)$$

Therefore, in this case study the state of the environment is characterized by $k = 4$ possible situations.

Actions: The set of possible actions are the l power levels $\mathcal{A} = \{p_1, p_2, \dots, p_l\}$ that femtocell can assign to RB r .

Cost: The cost c assesses the immediate return incurred due to the assignment of action a at state s . The considered cost function is:

$$c = \begin{cases} K & \text{if } Pow^f > P_{\max}^F, \\ (SINR_r^m - SINR_{Th}^M)^2 & \text{otherwise} \end{cases} \quad (4.5)$$

where K is a constant value. The rationale behind this cost function is that the Q-learning aims to minimize it, so that: i) the total transmission power of each femtocell does not exceed the allowed P_{\max}^F , and ii) the SINR at the macrouser is below the selected threshold $SINR_{Th}^M$.

Next State: The state transition from s to v in RB r is determined by the learning-based power allocation.

4.1.2 Learning design: case study 2

Differently from case study 1, here we consider a more complex state and cost definition since we furthermore consider a femtocell system performance indicator. The actions and next state have the same definition as the ones given in the case study 1. We define, the state and cost, with respect to each task r of femtocell f as follows.

State: The state of RB r for femtocell f is defined as:

$$\tilde{s}_r^f = \{Pow^f, \bar{C}_r^m, \bar{C}_r^f, \} \quad (4.6)$$

- **Macrocell capacity \bar{C}_r^m indicator:** We consider a macrocell capacity indicator since another main requirement for femtocells is not to jeopardize the macrocell performance. We define \bar{C}_r^m as a binary indicator to determine whether the capacity of macrouser u^m in RB r of macrocell m most affected by the activity of femtocell f , is above or below the macrocell capacity threshold C_{\min}^M , which is the minimum capacity per RB that the macrocell has to fulfil.

$$\bar{C}_r^m = \begin{cases} 1 & \text{if } C_r^m \geq C_{\min}^M, \\ 0 & \text{otherwise} \end{cases} \quad (4.7)$$

- **Femtocell capacity \bar{C}_r^f indicator:** In our design, we aim to maintain the above mentioned constraints, but we also want to maximize the capacity reached by the femtocells, so as to control the femto-to-femto interference. This is why, the third component of the state vector is a femtocell capacity \bar{C}_r^f indicator. We normalize C_r^f with respect to C_{\min}^F , so that $C_r^f/C_{\min}^F = \hat{C}_r^f$. C_{\min}^F is the minimum capacity to guarantee an acceptable service to the femto-users. We divide the possible values into four intervals given by:

$$\bar{C}_r^f = \begin{cases} 3 & \text{if } 0 \leq \hat{C}_r^f < 0.25, \\ 2 & \text{if } 0.25 \leq \hat{C}_r^f < 0.5, \\ 1 & \text{if } 0.5 \leq \hat{C}_r^f < 0.75, \\ 0 & \text{otherwise} \end{cases} \quad (4.8)$$

In this case, the state of the environment is then characterized by $k = 16$ possible states.

Cost: The cost c assesses the immediate return incurred due to the assignment of a certain action in a given state. The considered cost function is:

$$c = \begin{cases} K & Pow^f > P_{\max}^F \text{ or } C_r^m < C_{\min}^M, \\ K \exp^{-C_r^f} & \text{otherwise} \end{cases} \quad (4.9)$$

The rationale behind this cost function is that the total transmission power of each femtocell does not exceed the allowed P_{\max}^F , the capacity of the macrocell does not fall below a target C_{\min}^M , and the capacity of the femtocells is maximized.

In order to make more clear how states are quantized to a k -dimensional space for agents to represent the knowledge they acquire on the run, Figure 4.1 represents the Q-table for case study 2.

	$a_1 = p_1$...	$a_l = p_l$
$\left\{ \begin{array}{l} \bar{Pow}^f = 1 \\ \bar{C}_r^m = 0 \\ \bar{C}_r^f = 3 \end{array} \right\} = s_1$	$Q(s_1, a_1)$		$Q(s_1, a_l)$
$\left\{ \begin{array}{l} \bar{Pow}^f = 1 \\ \bar{C}_r^m = 0 \\ \bar{C}_r^f = 2 \end{array} \right\} = s_2$	$Q(s_2, a_1)$		$Q(s_2, a_l)$
\vdots			
$\left\{ \begin{array}{l} \bar{Pow}^f = 0 \\ \bar{C}_r^m = 1 \\ \bar{C}_r^f = 0 \end{array} \right\} = s_k$	$Q(s_k, a_1)$		$Q(s_k, a_l)$

Figure 4.1: Q-table for task r of agent f , for case study 2.

Finally, mentioning that we assume that the C_r^m , $SINR_r^m$ and C_r^f can be computed by the macrocell and the femtocell, respectively, based on the Reference Signal Received Quality (RSRQ) reported by the UEs allocated in RB r . The RSRQ is the quantification of the user received signal considering both, signal strength and interference [10].

4.2 Simulation results

This section presents some exciting results obtained when applying Q-learning approach in the single-cell scenario, introduced in Section 2.3.2 and in the multicell scenario, summarized in Section 2.3.3. Simulations are run for a C_{\min}^M per RB of 1.2 Mbit/s.

4.2.1 Single-cell scenario results

Results presented in this section correspond to a learning process modeled following the case study 1 and with an ε -greedy constant action selection policy. First of all, it has to be noted that the decentralized Q-learning algorithm, as any other learning scheme, needs a learning phase to learn the optimal decision policies. However, once completed the learning process and acquired the optimal policy, the multiagent system takes only one iteration to reach the optimal power allocation configuration, when starting at any initial state $s \in \mathcal{S}$.

Figure 4.2 shows the probability of being above the power threshold as a function of the learning iterations. It can be observed that the probability of Pow^f above P_{\max}^F decreases with iterations reaching very low values, for low and high density of femtocells, and that when the

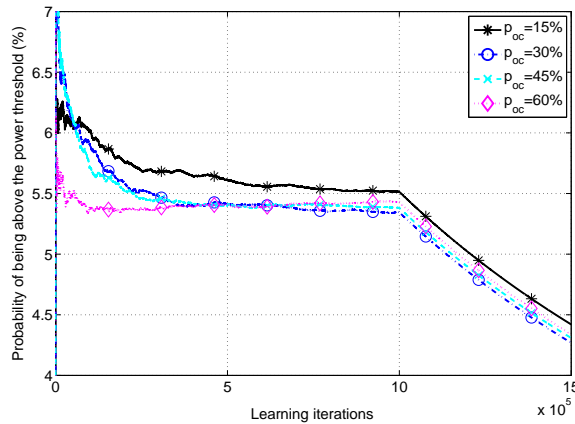


Figure 4.2: Probability of being above the power threshold as a function of the learning iterations for a single-cell scenario with different femtocell densities.

probability of visiting random states ε is set to zero (at iteration 1000000), it decreases faster.

Figure 4.3 shows the convergence curves for different values of desired SINR at the macrouser allocated in RB r , for a femtocell occupation ratio of $p_{oc}=40\%$. It can be observed how the Q-learning is able to maintain at different desired values (e.g., 17, 20, 23 dB) the SINR at the macrouser. Similar results can be obtained for different values of occupation ratio ranging from 10 to 50%.

As for the system level performances, Figure 4.4 shows macro, average femto and total system capacity. As it was expected, the introduction of femtocells increases the total system capacity in the scenario. On the other hand, the macrocell capacity remains constant since the SINR at the macrousers is kept at a fixed target value by the learning scheme. As a result, the perception of quality of service of macrousers is not jeopardized due to the presence of the femto network. In addition, the average femto capacity remains constant since the total transmitted power of each femtocell decreases with the density of femtocells.

Finally, Figure 4.5 shows macro, average femto and total system capacity as a function of the maximum total transmission power for a femtocell occupation ratio of $p_{oc}=20\%$. It can be observed that the total system capacity and average femto capacity grow with P_{max}^F . On the other hand, macrocell capacity again remains constant independently of the maximum transmission power of femtocells thanks to the SINR control at macrousers performed by means of the Q-learning algorithm.

4.2.2 Multicell scenario results

In this section, the presented results have been obtained for the case study 2 when applied in the multicell scenario. We evaluate the system performance in terms of femtocell total transmission

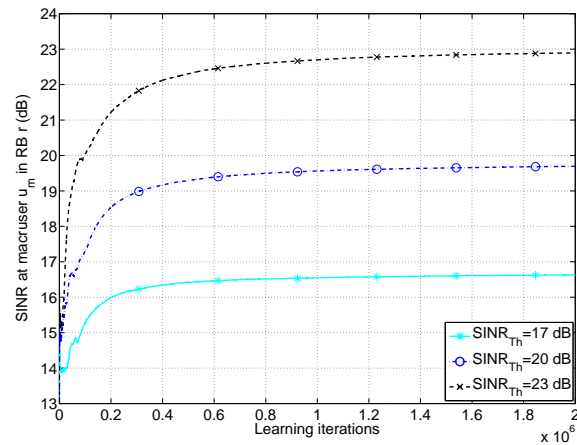


Figure 4.3: Convergence of SINR at macrouser to three desired values (i.e. 17, 20 and 23 dB).

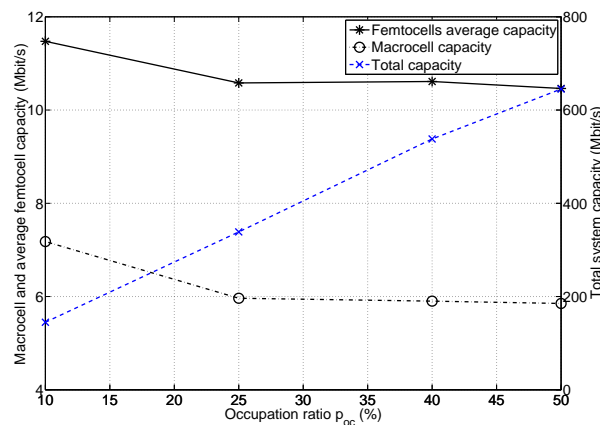


Figure 4.4: Macrocell, average femtocell and total system capacity as a function of the femtocell occupation ratio for the single-cell scenario.

power, average femtocell and macrocell systems capacity. Figure 4.6 depicts the probability of being below the capacity threshold as a function of the learning iterations. It can be observed that the proposed Q-learning approach is able to decrease the probability in all occupation ratio cases. In particular, for an occupation ratio of $p_{oc}=60\%$, which represents the worst case scenario presented, our learning scheme ensures that the capacity of the macrouser is above the threshold in more than 97% of all cases. Secondly, Q-learning is able to decrease the average probability that the femtocell total transmission power exceeds P_{max}^F . For instance, in case $p_{oc}=60\%$, Q-learning is able to reduce this probability to 2%.

Figure 4.7 shows the femtocells average capacity over the Q-learning iterations. It can be observed that the proposed algorithm is able to increase the femtocell capacity over time. Naturally, as the femtocell occupation ratio p_{oc} increases, the transmission power decreases in order not to cause excessive aggregated interference to the macrouser, which explains why the femto

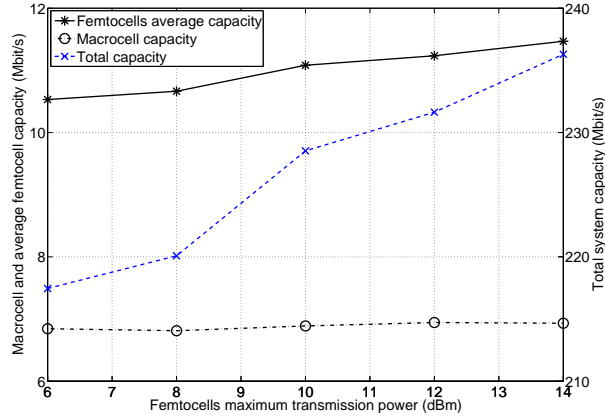


Figure 4.5: Macrocell, average femtocell and total system capacity as a function of the maximum total transmission power for a femtocell occupation ratio of $p_{oc}=20\%$.

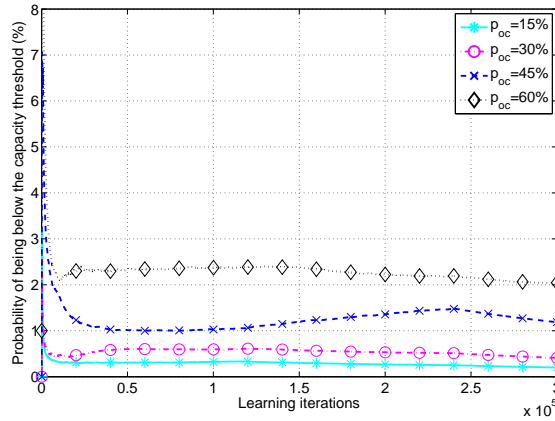


Figure 4.6: Probability of being below the capacity threshold as a function of the learning iterations for the multicell scenario.

capacity decreases as p_{oc} increases.

We compare system performance results obtained by the Q-learning with a benchmark algorithm known as Smart Power Control (SPC), which is based on interference measurements and which was proposed by 3GPP in [11]. In this algorithm, the femtocell BS adjusts its RBs transmission power based on the total received interference at the femtocell BS, according to:

$$p_r^f = \max(\min(\eta \cdot (E_c + 10 \log(R \times N_{sc})) + \beta, P_{\max}^F), P_{\min}^F) \quad (4.10)$$

where $\eta = 0.35$ is a linear scalar that allows altering the slope of power control mapping curve and $\beta = 0.8$ is a parameter expressed in dB, both of which are femtocell configuration parameters. Furthermore, N_{sc} is the number of sub-carriers, P_{\min}^F is the minimum femtocell transmit power, and E_c is the reference signal received power per resource element allocated to the femto node.

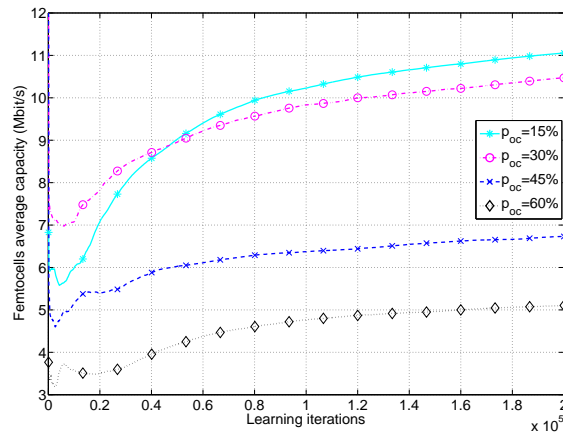


Figure 4.7: Femtocells average capacity over the learning iterations for the multicell scenario.

Figure 4.8 presents the macrocell and femtocell average capacities for Q-learning and SPC (4.10). The Q-learning algorithm outperforms the benchmark algorithm for both macro and femtocell average capacity, since Q-learning contemplates in its cost function the maximization of the femtocell capacity. This ensures that the capacity of the femtocell is the maximum possible, provided that a macrocell capacity per RB of at least C_{\min}^M is maintained.

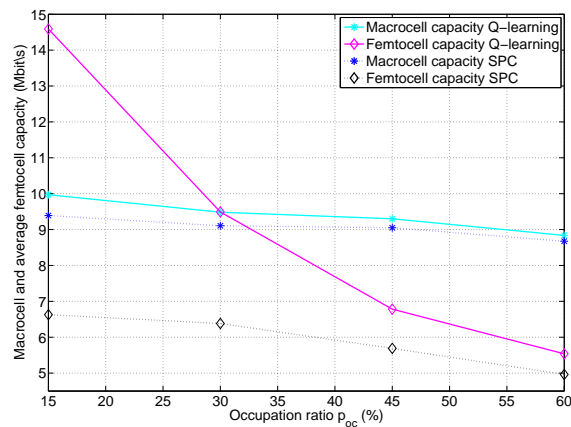


Figure 4.8: Macrocell and femtocell average capacity versus the femtocell occupation ratio p_{oc} for the multicell scenario.

4.3 Multiuser scheduling support

When the resource allocation of macrousers is updated by the macrocell, the perception of the state of the environment may change on a time frame basis of 1 ms. Due to unavoidable latencies incurred by signaling updates related to the state, given in equations (4.1) and (4.6) the femtocell cannot react quickly enough not to generate interference to vulnerable macrousers. Suppose that

macrocell m reschedules macrouser u^m from RB r to r' ; however, the task that an interfering femto node has learnt for RB r' until this instant may no longer be useful, since u^m now confronts the femto with a different perception of the state $s_{r',t}$, in terms of perceived interference at macrocell user u^m , reflected by the SINR indicator $SINR_r^m$ and the macro capacity $C_{r',t}^m$ of RB r' in (4.1) and (4.6), respectively. Provided the femto node is informed *a priori* about the resource assignment of macrouser u^m , the femtocell can proactively avoid interference on u^m , since the proper policy for protecting macrouser u^m was already learnt by the femto node in task r , as illustrated in Figure 4.9.

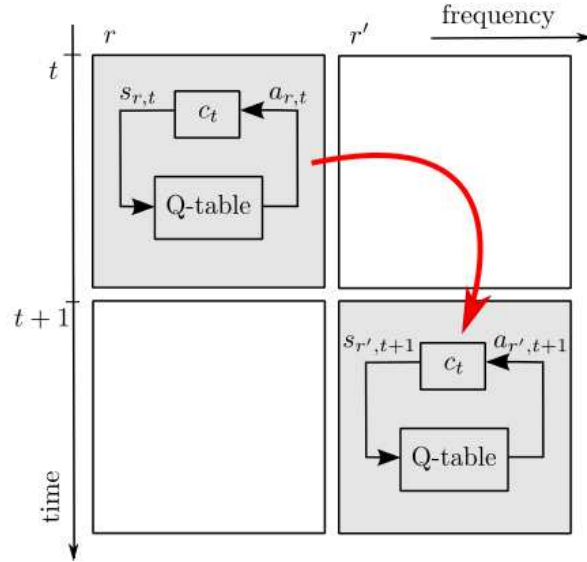


Figure 4.9: Transfer learning scheme at macrouser u_m .

To take advantage of this expert information already available at the femtocell, we propose that the future scheduling policies of the macrocell, with respect to the macrouser trapped in the coverage area of the femtocell, is communicated via X2' interface to the femtocell. In this way, the femtocell can reuse the expert knowledge that has already learnt in task r for task r' . The ML paradigm which allows to transfer this task related expert knowledge is referred to in literature as *transfer learning* [8].

The core idea of transfer learning is that experience gained in learning to perform one task can help improve learning performance in a related, but different task. Examples from the real world are numerous, for example, we may find that learning to recognize apples might help recognizing pears. The two tasks involved in the transfer learning process are the source and target tasks, which may be characterized by similar or different state spaces \mathcal{S} , action sets \mathcal{A} and cost functions \mathcal{C} . The information that can be exchanged range from information about the expected outcome when performing an action in a particular state (e.g., a state-action Q-value, a policy, a full task model) to a general heuristic that attempts to guide the learning (e.g., a subset of the full set of actions, rules or advices).

Many methods of transfer information between tasks, characterized by different state and action spaces, require an inter-task mapping to transfer effectively, in the sense that it is necessary to know how the two tasks are related. Some transfer learning algorithms allow the agent to learn multiple source tasks and learn from them all. More sophisticated algorithms build a library of seen tasks and use only the more relevant for transfer. Different types of knowledge may transfer better or worse depending on task similarity. For instance, particular information may transfer across closely related tasks, while high level concepts may transfer across pairs of less similar tasks.

In our case, the different tasks, referring to the different RBs, are characterized by the same state and action spaces and cost functions, and the information that is exchanged is the full task model, i.e. the full Q-table, as is shown in Figure 4.9. Since transfer learning is a femtocell internal process, it does not incur in delays.

4.3.1 Practical implementation in 3GPP LTE

In order to facilitate distributed Q-learning, the macrocell should report a feedback of 1 bit per RB to the offending femto node. The r -th entry of the corresponding bitmap termed Downlink High Interference Indicator (DL-HII) of dimension R is defined by:

$$\text{DL-HII}_r = \begin{cases} 1 & \text{SINR}_r^m \geq \text{SINR}_{Th}^M \\ 0 & \text{SINR}_r^m < \text{SINR}_{Th}^M \end{cases} \quad (4.11)$$

or

$$\text{DL-HII}_r = \begin{cases} 1 & C_r^m \geq C_{\min}^M \\ 0 & C_r^m < C_{\min}^M \end{cases} \quad (4.12)$$

depending on the used case study, which indicates whether SINR_r^m or C_r^m are above or below SINR_{Th}^M or C_{\min}^M thresholds, respectively. The DL-HII bitmap is to be exchanged between victim macrousers associated macrocell and interfering femtocells. The 3GPP LTE network architecture connects neighboring BSs via the X2 interface [12, 13], which conveys control information related to handover and interference coordination.

The DL-HII bitmap is equivalent to that of the Relative Narrowband Transmit Power (RNTP) indicator standardized in LTE [14]. The RNTP indicator contains R bits; each bit corresponds to one RB in the frequency domain, indicating to neighboring BSs whether the sending BS intends to transmit the associated RB with high power. The value of the threshold and the time period for which the indicator is valid are configurable parameters. As the LTE standard does *not* specify the action of a BS upon receiving a RNTP bitmap, it is possible to convey the DL-HII messages over the X2' interface.

Unfortunately, the X2 interface induces significant delays of up to $\Delta_{\max}=20$ ms, with an average of $\bar{\Delta}=10$ ms. This means that any change in the resource allocation of a victim macrouser u^m

is observed with an average delay of $10T$, where T denotes the subframe duration, which coincides to the length of one RB.

We therefore propose to augment the DL-HII bitmap by the resource allocation vector of macrocell m , denoted by $\mathbf{u}^m = [u_I^m, \dots, u_R^m]^T$, where entry u_r^m accounts for the macrouser scheduled at RB r . As typically only a small subset of the macrousers served by BS m are within the coverage area of a particular femtocell, it is sufficient to only send a RNTP-like bitmap to femtocell f , termed User Resource Block Allocation (URBA) bitmap. The URBA bitmap contains the anticipated RBs to be scheduled at time instant $t + \Delta_{\max}T$ to that macrouser u^m who is trapped within the coverage area of femtocell f .

In summary, Q-learning is integrated to the LTE network architecture by the following procedure:

1. Macrouser u^m determines the cell-ID of surrounding femto BSs, by reading the corresponding Physical Broadcast Channel (PBCH). LTE system key information consists on the Master Information Block (MIB) which is broadcast on the PBCH, and a number of System Information Blocks (SIBs), which are sent on the Physical Downlink Shared Channel (PDSCH) through the Radio Resource Control (RRC) messages. Macrouser u_m reads the MIB and extract the cell-ID information contained in the *SIB1*.
2. The cell-IDs of the surrounding femto BSs are reported to the serving macrocell.
3. The macrocell sends a DL-HII bitmap via the X2' interface to its surrounding femtocells, containing information about which RBs are subject to high interference. RNTP is part of the X2 load indication procedure [15]. Since the DL-HII would be equivalent to the RNTP, it is assumed that this indicator would be part of the load information messages of the X2 interface.
4. The macrocell sends a URBA bitmap, with RBs that are scheduled for trapped macrouser u^m at future time instants $t + \Delta_{\max}T$.

The message flow structure of given procedure would be as Figure 4.10 shows.

Figure 4.11 represents the average probability that the macrocell capacity is below the threshold C_{\min}^M over the R RBs and the average probability of the femtocells total transmission power above the threshold P_{\max}^F . Considering the presented 3GPP practical implementation, one learning iteration is performed every 10 ms due to the average delay of the X2 interface. Results show that after 5×10^4 iterations, the learning algorithm has learnt proper decision policies since its behavior becomes stable, despite the exploration process is still active, the multiple agents simultaneously learning and the dynamism of the system. Translating this into time, the learning process takes 500 s to achieve a stable behavior starting from a situation in which it has no

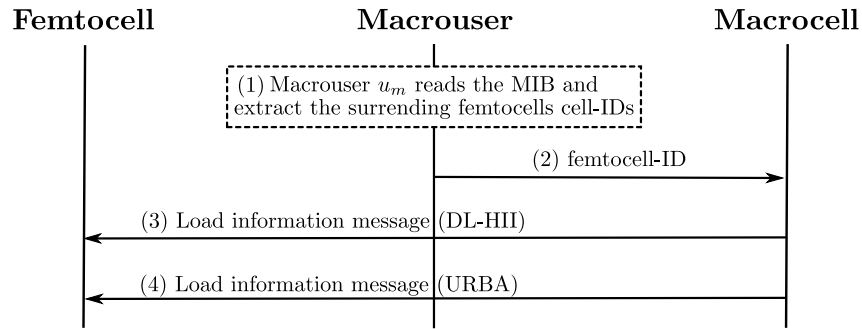


Figure 4.10: Message flow structure for Q-learning in LTE networks.

knowledge, i.e. the Q-table is initialized at H values, which is a very good convergence time considering the complexity of the system we are dealing with and the fact that the validity of the acquired knowledge does not expire with time.

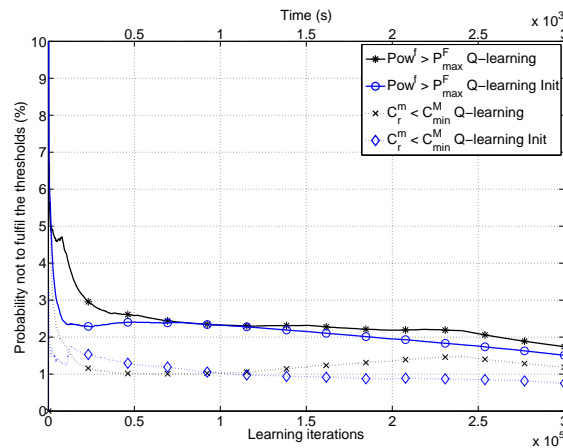


Figure 4.11: Average probability of not to fulfil the thresholds in time.

As it has been observed across the results presented in this chapter, the beginning of the learning process is always unstable due to the lack of information of the agent regarding how good or bad the multiple available actions are. This is because when the Q-table is initialized, all the Q-values, i.e. all the actions, have the same value and therefore the same probability to be chosen. One possible improvement is based on the algorithm designer intervention in the Q-table initialization. Results presented in Figure 4.11 correspond to a smart initialization, where Q-values corresponding to those states with macrocell user capacity below the threshold and actions with low transmission power were initialized to lower Q-values than the rest of Q-values. This intuitive initialization procedure allows to decrease the learning period and to have a more stable behavior at the beginning of the learning process. We will propose an effective way to implement expert knowledge in Chapter 6 through fuzzy inference rules.

In any case, this solution still implies the human intervention and, as we may see in Fig-

ure 4.11, it does not solve all the problems of the Q-learning approach, in terms of stability and speed of learning. That is the reason why we study and implement other more complex, smart and automatic solutions based on a cooperative technique, which we call Docition and which are presented in Chapter 5.

4.4 Conclusions

This chapter has presented the design of the proposed decentralized Q-learning algorithm based on the theory of multiagent learning to deal with the problem of interference generated by multiple femtocells at macrouers. We have shown that the multiagent system is able to automatically learn a policy to maintain the interference at the macrouers under a desired value. Simulation results have shown that constraints in the cost function can be fulfilled in both, single-cell and multicell scenarios, by introducing learning capabilities. We have proposed that the macrocell conveys through the X2' interface to the femtocell network a URBA bitmap containing information about macrouers scheduling in future instants. Taking advantage of this information, femtocells can perform transfer learning among internal tasks in advance, ensuring that an excessive interference is not generated at macrouers. The training periods from situations of total lack of knowledge have been shown to be in the order of 500 s. This performance results can be further improved through cooperative techniques which are proposed in next Chapter 5 or through the introduction of expert knowledge, as we will propose in Chapter 6. Another drawback of the solution presented is the assumption of the existence of an X2' interface between macrocell and its underlying femtocells. To solve this problem, we will propose the use of learning in partial observable environments in Chapter 7.

Bibliography

- [1] J. Nie and S. Haykin, “A Q-learning-based dynamic channel assignment technique for mobile communication systems,” *In IEEE Transactions on Vehicular Technology*, vol. 48, no. 5, pp. 1676–1687, Sept 1999.
- [2] Y.-S. Chen, C.-J. Chang, and F.-C. Ren, “Q-learning-based multirate transmission control scheme for RRM in multimedia WCDMA systems,” *IEEE Transactions on Vehicular Technology*, vol. 53, no. 1, Jan. 2004.
- [3] M. Bennis and D. Niyato, “A Q-learning based approach to interference avoidance in self-organized femtocell networks,” in *IEEE Globecom 2010 Workshop on Femtocell Networks*, December 2010, pp. 706–710.
- [4] X. Chen, Z. Zhao, and H. Zhang, “Power allocation for cognitive wireless mesh networks by applying multi-agent Q-learning approach,” *CoRR*, vol. abs/1102.5400, 2011.
- [5] J. P. Leite, P. H. P. de Carvalho, and R. D. Vieira, “A flexible framework based on reinforcement learning for adaptive modulation and coding in OFDM wireless systems,” in *2012 IEEE Wireless Communications and Networking Conference: PHY and Fundamentals*, April 2012, pp. 819–824.
- [6] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. The MIT Press, 1998.
- [7] “D2.2: The BeFEMTO system architecture,” EU FP7-ICT BeFEMTO project, Dec. 2011.
- [8] M. E. Taylor and P. Stone, “Transfer learning for reinforcement learning domains: A survey,” *Journal of Machine Learning Research*, vol. 10, pp. 1633–1685, July 2009.
- [9] “Evolved universal terrestrial radio access (E-UTRA); physical channels and modulation (release 8),” 3GPP organization, TS 3G TS36.211, Nov. 2007.
- [10] S. Sesia, I. Toufik, and M. Baker, *LTE, The UMTS Long Term Evolution: From Theory to Practice*. Wiley Publishing, 2009.
- [11] “3GPP TR 36.921 evolved universal terrestrial radio access (E-UTRA); FDD home eNode B (HeNB) radio frequency (RF) requirements analysis,” 3GPP, Tech. Rep., March 2010.
- [12] 3GPP, “X2 General Aspects and Principles (Release 8),” 3GPP TS 36.420 V8.0.0 (2007-12), Dec. 2007.
- [13] —, “X2 Application Protocol (X2AP) (Release 8),” 3GPP TS 36.423 V8.2.0 (2008-06), June 2008.

- [14] —, “Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures (Release 8),” 3GPP TS 36.213 V 8.8.0 (2009-09), Sep. 2009.
- [15] “Evolved universal terrestrial radio access network (EUTRAN); X2 application protocol (X2AP) (release 8),” 3GPP organization, TS TS 36.423, June 2008.

Chapter 5

Docition: a cooperative approach

In decentralized multiagent systems, the environment perceived by a given agent is no longer stationary, since it consists of other nodes who are similarly adapting. This may generate oscillating behaviors that not always reach an equilibrium and that are not yet fully understood, even by ML experts. The dynamics of learning may thus be long and complex in terms of required operations and memory, with complexity increasing with an increasing observation space. A possible solution to mitigate this problem, to speed up the learning process and to create rules for unseen situations, is to facilitate expert knowledge exchange among learners [1, 2].

A major disadvantage of learning solutions is that no practically viable solution is available as of today for wireless settings. This is because cognitive entities with learning capabilities need to be trained at start-up as well as during run-time to adapt to the wireless system dynamics, all of which consumes considerable time and energy. Furthermore, precision and time of convergence of learning algorithms may be poor, which prevents effective decision taking on a real time basis, something often overlooked, but eventually preventing real-world deployment.

To overcome above shortcomings, the novel concept referred to as *docitive radio* was introduced in [3, 4]. Whilst the emphasis in cognitive radios is to learn (“cognoscere” in Latin), the focus of docitive radios is on teaching (“docere” in Latin). It capitalizes on the fact that some nodes have naturally acquired a more pertinent knowledge for solving a specific system problem and are thus able to teach other, less able, nodes on how to cope under the same or similar situations. The potential impact of docitive networks into next generation high capacity wireless networks is ensured by means of latest European Telecommunications Standards Institute Broadband Radio Access Networks (ETSI BRAN) standardisation activities [5].

To apply docition, intelligent systems have to be able to represent their expertise and measure it. Then, they must establish a relation pattern with the other agents in the system to decide which of them are potential entities to cooperate with. Finally, docitive agents have to decide the degree of cooperation and the moment or moments to execute the cooperative process.

The aim of the work presented in this chapter is to outline a working taxonomy which shall aid future research endeavors of our community. To this end, we position the concept of docition and introduce a viable working taxonomy. More in particular, we apply the docitive approach in femtocell systems. Docitive femtocells are not (only) supposed to teach end-results, but rather elements of the methods of getting there. This concept perfectly fits a femtocell network scenario, where a femtocell is active only when the users are at home. When a femto BS is switched on, instead of starting a very energy expensive context awareness phase to sense the spectrum and learn the proper RRM policy, it can take advantage of the decision policies learnt by the neighbor femtocells, which have been active during a longer time. This novel paradigm for femtocells will be shown to capitalize on the advantages but, most importantly, to mitigate major parts of the drawbacks of purely self-organized and cognitive schemes, thus increasing their precision and accuracy and speeding up the learning process.

5.1 Docitive cycle

The general idea of docition is to extend the cognitive radio concept, introduced by Mitola in [6], to a system able, not only to intelligently interact with the environment, but also with collective principles, i.e. knowledge acquired by the system components individually becomes collective expertise. Logically, from a global point of view, higher and faster levels of expertise can be reached by parallel learning approaches. This easily translates into better individual behaviors when entities in the system can take advantage of the global knowledge. Of course, the utilization of the mentioned global expertise comes together with some constraints, in terms of inherent system characteristics, scalability, required signaling exchange, etc.

In what follows we present the high-level cognitive radios operational cycle, which consists in acquisition, decision and actuation processes. We extended this cycle by the introduction of the docitive functionalities, given by the docitive entity, as shown in Figure 5.1.

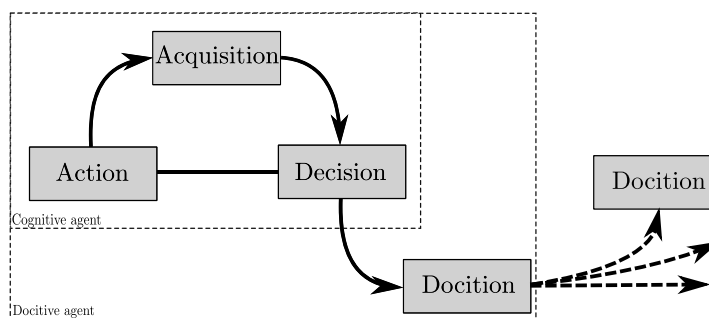


Figure 5.1: Docitive cycle which extends the cognitive cycle by cooperative teaching.

- *Acquisition.* The acquisition unit provides quintessential information of the surrounding

environment, such as spectrum occupancy or interference temperature. This data can be obtained by means of numerous methods, such as sensing performed by the node itself and/or in conjunction with spatially adjacent cooperating nodes; docitive information from neighboring nodes; databases; etc. Once the cognitive agent has the information describing the current situation of the surrounding environment it has to interpret it and send it to the next unit, where decisions are made.

- *Intelligent Decision.* The core of a cognitive radio is without doubt the environmental-state dependent intelligent decision engine, which typically learns from past experiences gathered from e.g., the dynamics of interference or statistics of spectral occupancy. Based on some intelligent algorithms, it then draws decisions on choice of band and resource block, transmission power, etc.
- *Action.* With the decision taken, an important aspect of the cognitive radio is to ensure that the intelligent decisions are being carried out, which is typically handled by a suitably reconfigurable software defined radio, some policy enforcement protocols, among others.
- *Docition.* The docition unit has two main tasks, the knowledge relation and the knowledge dissemination and propagation among agents. Those tasks have to be realized under the non-trivial aim of improving the own or other agent's learning process and performance. As shown in Figure 5.1, the docition unit is linked to the intelligent decision unit and communicates with the other agents' docition units. By these means, each docition unit builds its relationships with the other agents in the system and decides on the key docitive parameters.

Going more deeply into the docitive concept, there are four main aspects to be considered, i.e. 1) docitive mode: should the agent teach or learn from other agents?; 2) nodes relation: what are the entities in the system with whom the agent can interact?; 3) knowledge to share: what to teach or to learn?; 4) moment of docition: when to perform the docition process?. Figure 5.2 shows the structure of docition unit. The decision evaluation unit is in charge to assess the choices made in the decision unit in terms of how well the agent is performing with respect to the goals it pursues. Based on the resulting information, the mode selector unit determines if the docitive agent should perform as a teacher or as a learner. It is important to notice that the docitive agent may be an expert in some tasks but it can be completely inexpert in others. Therefore, the mode selector unit should keep a track per task reflecting the mode, i.e. teacher or learner, in which the docitive agent should perform. The negotiation unit is in charge of evaluating other potential docitive partners in the system, to decide the information to be exchanged and the moment to share it, as well as to evaluate the tradeoff between performing docition or cognition. This is the core unit in the docitive agent since it is in charge of the key decisions in the cooperative process. Finally, the docitive executor unit enforces the decisions

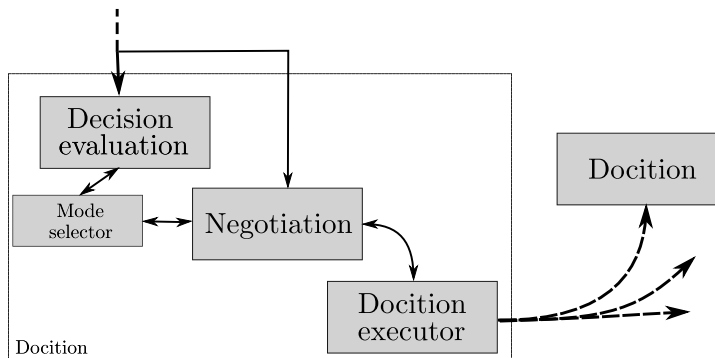


Figure 5.2: Docitive unit structure: main functionalities.

made by the negotiation unit considering the constraints in the network data transport. This information is available to the negotiation unit and has to be considered by it in the decision of performing docition or cognition in a given moment.

5.2 Emerging docitive algorithms

The state of human cognition heavily depends on the teachers encountered during one's life, who generally impact learning space, learning speed and teaching abilities. The henceforth introduced concept of docition is inspired by the Problem Based Learning (PBL) concept. It has been advocated by the pedagogy expert psychologists Lev Vygotsky, John Dewey, Jean Piaget, among others, with the prime aim that teachers are encouraged to be coaches and not information givers. PBL makes pupils work as a team using critical thinking to synthesize and apply knowledge; they apprehend through dialogue, questioning, reciprocal teaching, and mentoring.

Mimicking above well-functioning society-driven teacher-pupil paradigm, we capitalize on the advantages of PBL by encouraging radios to teach other radios with the aim to significantly improve performance of current (cognitive) systems. Said teaching process requires the exchange of information in a cooperative fashion, the rate of which however is negligible when compared to the data volumes handled by the system as a whole. Therefore, even if docitive systems require some (often sporadic) exchange of low-rate information, they are asymptotically distributed and generally autonomous.

Whilst applicable to other learning techniques, we focus here on the decentralized implementation of Q-learning presented in Section 3.2.1. The main challenge in decentralized learning systems, is how to ensure that individual decisions of the nodes result in jointly optimal decisions for the group, considering that the standard convergence proof for Q-learning does not hold in this case as the transition model depends on the unknown policy of the other learning nodes. Whilst one could treat the distributed network as a centralized one, it brings along many problems [4].

We thus propose a distributed approach where nodes share potentially differing amounts of intelligence acquired on the run. This is expected to sharpen and speed up the learning process. Depending on the degree of docition among nodes, the following cases can be distinguished:

- *Startup Docition.* Docitive radios teach their policies to any newcomers joining the network. In this case, again, each node learns independently; however, when a new node joins the network, instead of learning from scratch how to act in the surrounding environment, it learns the policies already acquired by more expert neighbors. Gains are due to a high correlation in the environments of adjacent expert and newcomer nodes. Policies are shared by exchanging Q-tables.
- *IQ-Driven Docition.* Docitive radios periodically share part of their policies with less expert nodes, based on the degree and reliability of their expert knowledge. Policies are shared by exchanging (a weighted version) of the entire Q-table or rows thereof, corresponding to states that have been previously visited.
- *Performance-Driven Docition.* Docitive radios share part or the entirety of their policies with less expert nodes, based on their ability to meet prior set performance targets. Example targets are maximum created interference, achieved capacity, etc.
- *Perfect Docition.* The multi-user system can be regarded as an intelligent system in which each joint action is represented as a single action. The optimal Q-values for the joint actions can be learnt using standard centralized Q-learning. In order to apply this approach, a central controller, implemented e.g., in the LFGW, should model the MDP and communicate to each node its individual actions. Alternatively, all nodes should model the complete MDP separately and select their individual actions; whilst no communication is needed here, they all have to observe the joint actions and individual rewards. Due to an exponential growth of the states, this approach is typically not feasible.

The degree of cooperation, and thus the overhead, augments with an increasing degree of docition. The optimum operating point hence depends on the system architecture, performance requirements, etc. A summary of the taxonomy introduced is shown in Figure 5.3.

A major factor influencing the degree, intensity and direction of docition is clearly the quantification of the level of expertness of nodes. The optimum operating point depends on the system architecture, performance requirements, etc. However, as already stated, the overhead due to the exchange of docitive information is asymptotically negligible when compared with actual data volumes transported through the network.

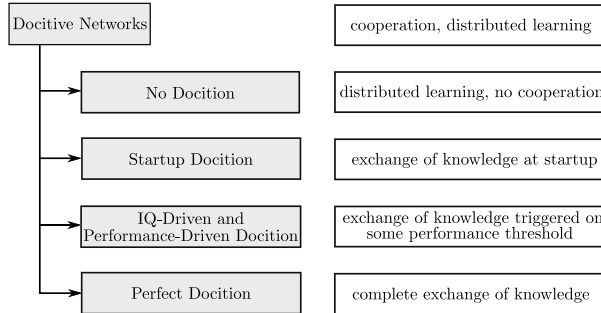


Figure 5.3: Taxonomy of docitive algorithms with different degrees of docition.

5.3 Learning and teaching techniques

In our scenario, a femto BS which has been switched on, could take advantage of the exchange of information and expert knowledge from other femtocell in the neighborhood [1], the so-called docitive femtocells. The agents select the most appropriate femto BS from which to learn, based on the level of expertness and the similar impact that their actions may have on the environment, which is captured by a gradient, ∇_f^r , which captures the similar impact that femtocells actions may have on the environment. Notice that, in terms of signaling overhead, the gradient is only a numerical value to exchange sporadically among femto BSs. This gradient, for femtocell f and RB r , is defined as:

$$\nabla_f^r = \frac{SINR_t^r - SINR_{t-1}^r}{a_t^r - a_{t-1}^r}, \quad (5.1)$$

where a_t^r and a_{t-1}^r represent the actions taken for RB r at time t and $t-1$, respectively, and $SINR_t^r$ and $SINR_{t-1}^r$, represent the SINR at the macrouser in RB r at time t and $t-1$, respectively. The rationale behind the definition of this gradient is that nodes should learn from nodes in similar situations, e.g., a femtocell which is located close to a macrouser should learn the policies acquired by a femtocell operating under similar conditions. Depending on the degree of docition among nodes, we consider two cases, the startup docition and the IQ-Driven docition.

The degree of cooperation, and thus the overhead, augments with an increasing degree of docition. The optimum operating point hence depends on the system architecture, performance requirements, etc.

5.3.1 Simulation results

Results presented in this chapter were modeled for the single-cell scenario, presented in Section 2.3.2 and case study 1, presented in Section 4.1.1. The Q-learning scheme, following the docitive paradigm, has been compared to two reference algorithms:

- *Distance-Based Non-Cognitive*. The rationale behind this reference algorithm is that fem-

tocell f selects the transmission power of RB r based on its distance from the macrouser using that RB. The set of possible values of power to assign is the same as for the Q-learning, as defined in Section 4.1.1. Notice that this reference algorithm is only proposed as a non-cognitive benchmark for comparison purposes, and for its implementation we make the hypothesis that the femto network has at least some approximate knowledge of the position of the macrouser, which is a quite difficult hypothesis in a realistic cellular network.

- *Iterative Water-Filling (ITW)*. It is a non-cooperative game where agents are selfish and compete against each other by choosing their transmit power to maximize their own capacity, subject to a total power constraint, such that:

$$\begin{aligned} \max_{p_r^{f,F}} \sum_{r=1}^R \log \left(1 + \frac{p_r^{f,F} h_{f,f,r}^{FF}}{\sum_{k=1, k \neq f}^N p_r^{k,F} h_{k,f,r}^{FF} + \sigma^2} \right) \\ \text{s.t.} \quad \sum_{r=1}^R p_r^{f,F} \leq P_{\max}^F, \quad p_r^{f,F} \geq 0 \end{aligned} \quad (5.2)$$

The solutions to (5.2) are given by the ITW power allocation solutions [7]:

$$p_r^{f,F} = \max \left(\frac{1}{\lambda^{f,F}} - \frac{\sum_{k=1, k \neq f}^N p_r^{k,F} h_{k,f,r}^{FF} + \sigma^2}{h_{f,f,r}^{FF}}, 0 \right) \quad (5.3)$$

where $\lambda^{f,F}$ is the Lagrangian multiplier chosen to satisfy the power constraint.

To evaluate the proposed dicitive approaches we divide the femtocells in the scenario into two groups. In the first group, we have the dicitive or teaching entities and in the second group we have the learning entities, which start their learning process 500000 learning iterations later than the dicitive ones. The given results were obtained for the second group of femtocells in order to show the improvement achieved when agents take advantage of acquired knowledge of other comparable entities in the system. In the startup case, learning entities update their Q-tables at the beginning of the learning process based on the selected dicitive entity policies. In the IQ-Driven case, learning entities update every 10000 learning iterations the Q-values of those rows with lower knowledge, i.e. higher Q-values.

Figure 5.4 shows the macrocell capacity as a function of the femtocell density. It can be observed that learning techniques do not jeopardize the macrocell capacity, maintaining it at a desired level independently of the number of femtocells. On the other hand, with the distance-based reference algorithm, the macrocell capacity decreases when the number of femtocells increases, since the reference algorithm does not adaptively consider the aggregated interference coming from the multiple femtocells in the power allocation process. Furthermore, the ITW algorithm dramatically reduces the macrocell capacity due to its selfish power allocation policy. Finally,

with respect to the implementation, it is worth mentioning that the Q-learning approaches only need feedback from the macro network about the SINR at the macrousers. However, the non-cognitive distance-based approach relies on stronger hypotheses, such as the positions of the macrousers.

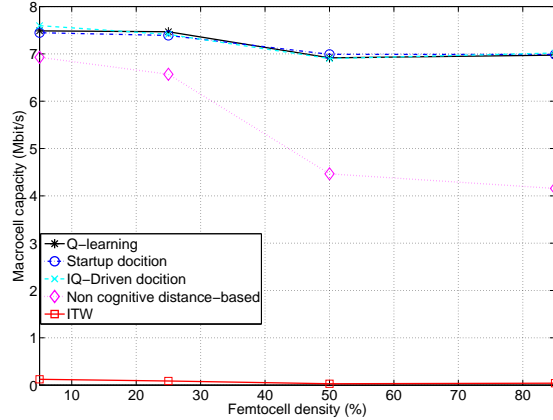


Figure 5.4: Macrocell capacity as a function of femtocell density.

As for the performance of docition, Figure 5.5 shows performances in terms of precision, i.e. oscillations around the target SINR. We assumed a 50% femtocell occupation ratio, composed of the probability that a femtocell is present and that it is switched on. In particular, it represents the Complementary Cumulative Distribution Function (CCDF) of the variance of the average SINR at the control point with respect to the set target of $SINR_{Th} = 20$ dB. It can be observed that due to the distribution of intelligence among interactive learners the paradigm of docition stabilizes the oscillations by reducing the variance of the SINR with respect to the specified target. More precisely, at a target outage of 1%, we observe that the IQ-Driven docition outperforms the startup docition by a factor of two, and the Q-learning algorithm by about an order of magnitude.

Figure 5.6 shows the average probability that the total power at femtocells is higher than P_{\max}^F as a function of the learning time when docition is applied. It can be observed that the docitive approaches better satisfy the constraint in terms of total transmission power since the early stages of the learning process. More in particular, for the startup docition case, after docition, the femtocell continues with its learning process adapting the knowledge to its own situation. On the other hand, for the IQ-Driven docition case, since the learner agent periodically updates the knowledge corresponding to states of the environment where it performs poorly, the agent presents a very accurate behavior during all the learning process. This accurate behavior is achieved thanks to the continuous adjustment in the agent policy.

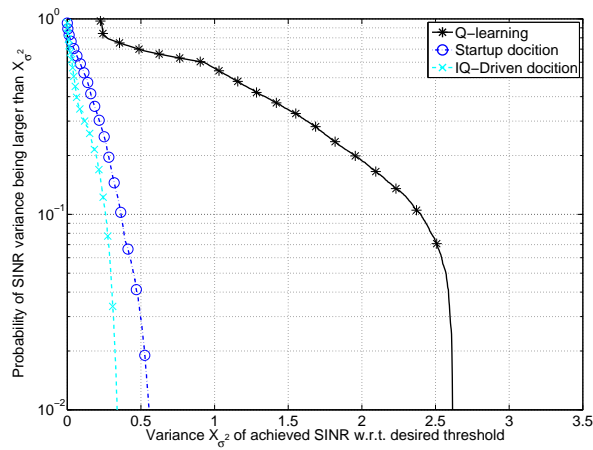


Figure 5.5: CCDF of the average SINR at macrouser for a femtocell occupation ratio $p_{oc}=50\%$.

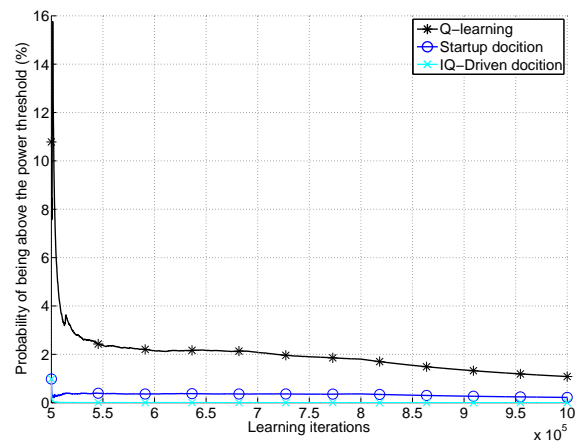


Figure 5.6: Probability of being above the power threshold as a function of the learning iterations for different docitive cases.

Challenges in docition

Although docition presents interesting and enticing advantages since it promises to have a more efficient use of the collective knowledge in a system, there are some important and difficult challenges and questions to consider. In what follows, the main ones are summarized:

- *Information Theory.* One of the core problems is how to quantify the degree of intelligence of a cognitive algorithm. With this information at hand, intelligence gradients can be established where docition should primarily happen along the strongest gradient. This would also allow one to quantify the tradeoff between providing docitive information versus the cost to deliver it via the wireless interface. Some other pertinent questions are how much information should be taught?, can it be encoded such that learning radios with

differing degrees of intelligence can profit from a single multicast transmission?, how much feedback is needed?, how often should be taught?, etc.

- *PHY/MAC Layers.* A pertinent question is which of the states should be learned individually, and which are advantageously taught? Another open issue is how much rate/energy should go into docition versus cognition?
- *Docitive System Design.* At system level, numerous questions remain open, such as what is the optimal ratio of docitive versus cognitive entities?, what is the optimal docition schedule?, should every cognitive entity also be a docitive one?, what is the docition overhead versus the cognitive gains?, etc.

5.4 Conclusions

The main drawback of online learning approaches is the length of their learning process. As a result, we have focused on the novel paradigm of docition, with which a femto BS can learn the interference control policy already acquired by a neighboring femtocell which has been active during a longer time, thus saving significant energy during the startup and learning process. Notably, we have shown in a 3GPP compliant scenario that, with respect to Q-learning, docition applied at startup as well as continuously on the run yields significant gains in terms of convergence speed and precision. Also, we have highlighted different types of docition as well as different quantifications of expertness.

Finally recall that, whilst we applied the docitive paradigm to the model-free distributed Q-learning algorithm, it is equally applicable to concepts which involve decision making based on cooperative spectrum sensing, distributed consensus building, etc. We believe that we just touched the tip of an iceberg as these preliminary investigations have shown that docitive networks are a true facilitator for utmost efficient utilization of scarce resources and thus an enabler for emerging as well as unprecedented wireless applications. Further research and results related to the docitive theory will be presented by Pol Blasco in his thesis.

Bibliography

- [1] M. Tan, *Multi-Agent Reinforcement Learning: Independent vs. Cooperative Agents*. In M. N. Huhns and M. P. Singh, editors. Morgan Kaufmann, San Francisco, CA, USA., 1993, ch. 26, pp. 451–480.
- [2] M. N. Ahmadabadi and M. Asadpour, “Expertness based cooperative Q-learning,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 32, no. 1, pp. 66–76, Feb. 2002.
- [3] M. Dohler, L. Giupponi, A. Galindo-Serrano, and P. Blasco, “Docitive networks: A novel framework beyond cognition,” *IEEE Communications Society, Multimedia Communications TC, E-Letter*, Jan. 2010.
- [4] L. Giupponi, A. Galindo-Serrano, P. Blasco, and M. Dohler, “Docitive networks - an emerging paradigm for dynamic spectrum management,” *IEEE Wireless Comm. Magazine*, vol. 17, no. 4, pp. 47–54, Aug. 2010.
- [5] Alvarion, CTTC, Polska Telefonia Cyfrowa, Siklu, and Thales, “Very high capacity density BWA networks; system architecture, economic model and technical requirements,” ETSI TC BRAN, TR 101 534, Feb. 2012.
- [6] J. M. III and J. Gerald Q. Maguire, “Cognitive radio: making software radios more personal,” *IEEE [see also IEEE Wireless Communications] Personal Communications*, vol. 6, no. 4, pp. 13–18, Aug. 1999.
- [7] G. Scutari, D. P. Palomar, and S. Barbarossa, “Asynchronous iterative waterfilling for gaussian frequency-selective interference channels,” *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 2868–2878, July 2008.

Chapter 6

Interference management based on Fuzzy Q-learning

As studied in Chapter 3, the Q-learning approach builds incrementally a Q-value per each state and action pair. Q-values are estimated through the Q-function, given in equation 3.8, which attempts to estimate the discounted future costs of executing an action in the agent's current perceived state. The knowledge of the agent is then represented by the Q-values, which are usually stored in a Q-table. Consequently, states characterizing the environmental situation and the available actions have to be represented by discrete values and therefore, the use of thresholds is mandatory. This entails an important intervention of the learning system designer selecting the mentioned thresholds for the state representation, and setting the amount and the values of the available actions.

When designing the learning system, the selection of the amount of states representing the environment and the actions available in each state, plays an important role in the agent behavior. The size of those sets directly affects the system adaptability and therefore its performance. Besides, it is directly related with the feasibility in the knowledge representation, i.e. when the number of state-action pairs is large or the input variables are continuous, the memory requirement to store the Q-table may become impracticable, as well as the required learning time.

More in particular, our experience in working with Q-learning for interference management in dynamic scenarios is that, the performance of the system and the convergence capabilities strongly depend on the granularity in the state and action spaces definition. The use of discrete state and action sets may result in an inexact state representation and/or the selection of a not accurate enough action in a given situation. Figure 6.1 represents the system performance in terms of average macrocell capacity for the multicell scenario, presented in Section 2.3.3 and case study 2, introduced in Section 4.1.2. As it can be observed, when applying a Q-learning approach with different quantities of actions, i.e. 40, 50 and 60, the learning processes with higher amount of actions better perform than those with less available actions. This is because

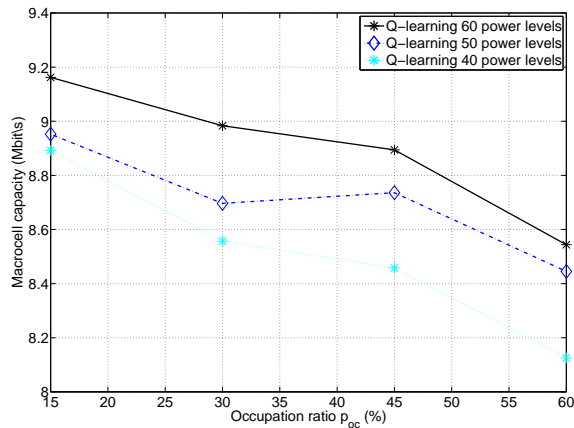


Figure 6.1: System performance for Q-learning with different amount of available actions.

a learning algorithm with less actions has to choose policies which may be far from the optimal solution, resulting in poor performance or even precluding the agent to be able to fulfil the learning requirements. Summarizing, by cutting down 17% of available actions, the femtocell system performance, in terms of capacity, may experience a reduction of up to 10%.

A solution to the previously presented weak points, is to use a form of continuous state and action representation without the requirement of near infinite Q-tables. This would allow to build a system capable of working independently from the scenario and designer criterion, which would be in line with the self-organized requirements of future networks. To this end, we propose to improve the Q-learning algorithm, presented in Chapter 3, by the introduction of FIS, in order to represent state and action spaces continuously. This approach is called Fuzzy Q-learning (FQL). Fuzzy logic, introduced by Lotfi Zadeh [1], is a way to map a fuzzy input space to a crisp output space by means of membership functions. The combination of FIS and RL was introduced by Berenji in [2] and then extended by Glorennec and Jouffe in [3] and [4]. They also presented an interesting case study, where applied FQL for navigation system in autonomous robots [5].

Additionally to the benefits already mentioned, FQL offers other interesting advantages such as: (1) a more compact and effective expertness representation mechanism and (2) the possibility of speeding up the learning process by incorporating offline expert knowledge in the inference rules.

In the field of telecommunications, FQL is lately arising as an attractive approach, mainly because of its properties of continuous state and action representation and its ability to deal with uncertainty and imprecision when learning by reinforcement methods is used. Specifically, in wireless systems we can find FQL applied as a situation-aware approach to manage the data access in multi-cell Wideband Code Division Multiple Access (WCDMA) systems [6] and for admission control in WCDMA/Wireless Local Area Network (WLAN) heterogeneous networks [7].

In [8], the authors use FQL to select the modulation and coding scheme in High Speed Downlink packet Access (HSDPA) systems. In [9], the authors propose a Nash-Stackelberg FQL approach based on which mobile users individually select the best available system to connect in a heterogeneous network. An Hybrid Automatic Repeat Request (HARQ) for MIMO configuration and Modulation and Coding Scheme (MCS) selection based on FQL for HSPA evolution systems is proposed in [10]. Finally, [11] and [12] propose two distributed and autonomous solution, based on fuzzy RL techniques, for coverage and capacity self-optimization through BSs' downtilt angle adjustment in LTE networks.

In what follows, first we give a brief overview of the FIS in Section 6.1. Second, Section 6.2 presents a formal definition of the FQL concept. Third, Section 6.3 introduces the proposed FQL algorithm with different degrees of complexity. Finally, some enticing results for the single-cell and multicell scenarios presented in Chapter 2, are given in Section 6.3.2.

6.1 Fuzzy Inference Systems

Fuzzy inference, is the process of formulating the mapping from a given input to an output using fuzzy logic. The mapping provides a basis from which decisions can be made, or patterns discerned. The parallel nature of the rules is one of the more important aspects of fuzzy logic systems. Instead of sharp switching between modes based on breakpoints, logic flows smoothly from different regions of behavior depending on the dominant rule.

The purpose of fuzzy systems is to perform as control systems considering that many times real problems cannot be efficiently expressed through mathematical models. So, fuzzy set theory models the vagueness that exists in real world problems. According to this theory, when \mathcal{X} is a fuzzy set and x is a relevant object, the proposition “ x is a member of \mathcal{X} ” is not necessarily true or false, but it may be true or false only to some degree, the degree to which x is actually a member of \mathcal{X} [1, 2].

In fuzzy logic, each object can be labeled by a linguistic term, where a linguistic term is a word as “small”, “medium”, “large”, etc. so that, x is defined as a linguistic variable. Each linguistic variable is associated with a term set $T(x)$, which is the set of names of linguistic values of x . Each element in $T(x)$ is a fuzzy set. In our work, we refer to the *Takagi-Sugeno* FIS [13], which is given by generic rules:

$$\mathcal{R}_i : \quad \text{If } x_1 \text{ is } \mathcal{X}_1^i \text{ and } \dots \text{ and If } x_z \text{ is } \mathcal{X}_z^i, \text{ Then } O = o_i(\tilde{s})$$

\tilde{s} is formed by L linguistic variables or input values characterized by z linguistic values, then $\tilde{s} = \{x_1, \dots, x_z\}$ is an input and \mathcal{X}_h^i is a fuzzy set in the domain of x_h , for $h=1, \dots, z$ and $i=1, \dots, n$. We have denoted by z the dimension of the input space and by n the number of rules.

The output function $o_i(\tilde{s})$ is a polynomial function of \tilde{s} . In our work we use the *0-Takagi-Sugeno* FIS, which means that the output polynomial is a constant.

A fuzzy inference process consists of three parts: fuzzification of the input variables, computation of truth values and defuzzification, as presented in Figure 6.2. The first step takes the fuzzy inputs \tilde{s} and determines to which degree they belong to each of the appropriate fuzzy sets via membership functions. A fuzzy set \mathcal{X}_h^i is characterized by a membership function $\mu_{\mathcal{X}_h^i}(x_h)$ that associates each point in \mathcal{X}_h^i with a real number in the interval $[0, 1]$. This number represents the grade of membership of x_h to \mathcal{X}_h^i [1].

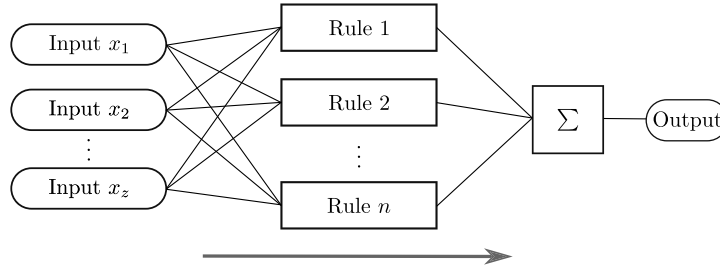


Figure 6.2: Fuzzy Inference System scheme.

After the inputs are fuzzified, for each rule we know to which degree each part of the inputs is satisfied. If the inputs of a given rule have more than one part, the fuzzy operator is applied to obtain one number that represents the result of the inputs for that rule. This number is known as the truth value of the considered rule. The inputs to the fuzzy operator consist of two or more membership values from fuzzified input variables. The output is a single truth value given by:

$$w_i(\tilde{s}) = \prod_{h=1}^z \mu_{\mathcal{X}_h^i}(x_h) \quad (6.1)$$

In general, each rule is also described by a membership function which gives the degree of reliability of the rule. We consider the reliability to be a constant, equal for all the rules. Since decisions are based on the testing of all the rules in a FIS, the rules must be combined in some manner in order to make a decision. The input for the defuzzification process is a fuzzy set and the output is a single number given by:

$$O(\tilde{s}) = \frac{\sum_i w_i(\tilde{s}) \times o_i}{\sum_i w_i(\tilde{s})} \quad i = 1, \dots, n \quad (6.2)$$

In the following section we present the FQL which is used to approximate the action value function in RL problems through Takagi-Sugeno FIS [13].

6.2 Fuzzy Q-learning

Let us consider an input state vector \tilde{s} , represented by L fuzzy linguistic variables. For each RB r we denote $\bar{\mathcal{S}} = \{\bar{s}_1, \dots, \bar{s}_n\}$ the set of fuzzy state vectors of L linguistic variables. For state \bar{s}_i , we denote $\mathcal{A} = \{a_1, \dots, a_l\}$ the set of possible actions. The rule representation of FQL for state \bar{s}_i is:

$$\begin{aligned} \text{If } \tilde{s} \text{ is } \bar{s}_i, \text{ Then } a_1 \text{ with } q(\bar{s}_i, a_1) \\ \dots \\ \text{or } a_j \text{ with } q(\bar{s}_i, a_j) \\ \dots \\ \text{or } a_l \text{ with } q(\bar{s}_i, a_l) \end{aligned}$$

where a_j is the j -th action candidate which is possible to choose for state \bar{s}_i , and $q(\bar{s}_i, a_j)$ is the fuzzy Q-value for each state-action pair (\bar{s}_i, a_j) . The number of state-action pairs for each state \bar{s}_i equals the number of the elements in the action set, i.e. each antecedent has l possible consequences. As one associates actions in every state in Q-learning, one associates several competing solutions in every rule in FQL. As a result, every fuzzy rule needs to choose an action a_j from the action candidate set \mathcal{A} by an action selection policy. A fuzzy Q-value which is incrementally updated is associated to each conclusion. The result of FQL is the output of the defuzzification process. The first output is the inferred action after defuzzifying the n rules and is given by:

$$a = \frac{\sum_{i=1}^n w_i \times \hat{a}}{\sum_{i=1}^n w_i} \quad (6.3)$$

where w_i represents the truth value (i.e. the fuzzy-AND operator) of the rule representation of FQL for \bar{s}_i , and \hat{a} is the action selected for state \bar{s}_i , after applying ε -greedy two-steps action selection policy, presented in Section 3.3.2. The second output represents the Q-value for the state-action pair (\tilde{s}, a) , and is given by:

$$Q(\tilde{s}, a) = \frac{\sum_{i=1}^n w_i \times q(\bar{s}_i, \hat{a})}{\sum_{i=1}^n w_i} \quad (6.4)$$

Q-values have to be updated after the action selection process. Since there is a fuzzy Q-value per each state-action pair, in each iteration, n fuzzy Q-values have to be updated based on:

$$q(\bar{s}_i, \hat{a}) = q(\bar{s}_i, \hat{a}) + \alpha \Delta q(\bar{s}_i, \hat{a}) \quad (6.5)$$

where

$$\Delta q(\bar{s}_i, \hat{a}) = [c(\tilde{s}, a) + \gamma(Q(\tilde{v}, a') - Q(\tilde{s}, a))] \times \frac{w_i}{\sum_{i=1}^n w_i} \quad (6.6)$$

and c represents the cost obtained applying action a in state vector \tilde{s} and $Q(\tilde{v}, a')$ is the next-state optimal Q-value defined as:

$$Q(\tilde{v}, a') = \frac{\sum_{i=1}^n w_i \times q(\bar{v}_i, a^*)}{\sum_{i=1}^n w_i} \quad (6.7)$$

and

$$\mathbf{a}^* = \arg \min(q(\bar{v}_i, \mathbf{a}_j^*)) \quad j = 1, \dots, l \quad (6.8)$$

is the optimal action for the next state \bar{v}_i , after the execution of action $\hat{\mathbf{a}}$ in the fuzzy state \bar{s}_i .

FQL is performed through a four layer structure FIS. The functionalities of each layer are the following:

- *Layer 1*: This layer has as input L linguistic variables each of which is defined by the term set $T(x)$. Therefore, considering the number of fuzzy sets it will have z term nodes. Every node is defined by a membership function with a bell shape form, so that the output $O_{1,h}$ for a generic component x of \tilde{s} is given by:

$$O_{1,h} = \exp \frac{-(x-e^h)^2}{(\rho^h)^2} \quad h = 1, \dots, z \quad (6.9)$$

where e^h and ρ^h are the mean and the variance of the bell shape function associated to node h , respectively.

- *Layer 2*: is the rule nodes layer. It is composed by n nodes and each of them gives as output the truth value of the i -th fuzzy rule. Each node in layer 2 has L input values, one from one linguistic variable of each of the L components of the input state vector, so that the i -th rule node is represented by the fuzzy state vector \bar{s}_i . The layer 2 output $O_{2,i}$ is the product of L membership values corresponding to the inputs. Truth values are represented as:

$$O_{2,i} = \prod_{h=1}^L O_{1,h} \quad i = 1, \dots, n \quad (6.10)$$

- *Layer 3*: In this layer each node is an action-select node. Here the set of possible actions for each layer 3 node are l power levels. In this layer the amount of nodes is n and they select the action $\hat{\mathbf{a}}$ based on the ε -greedy policy explained in Section 3.3.2 and the $q(\bar{s}_i, \hat{\mathbf{a}})$ values are initialized based on expert knowledge. The node i generates two normalized outputs, which are computed as:

$$O_{3,i}^A = \frac{O_{2,i} \times \hat{\mathbf{a}}}{\sum_{d=1}^n O_{2,d}} \quad i = 1, \dots, n \quad (6.11)$$

$$O_{3,i}^Q = \frac{O_{2,i} \times q(\bar{s}_i, \hat{\mathbf{a}})}{\sum_{d=1}^n O_{2,d}} \quad i = 1, \dots, n \quad (6.12)$$

- *Layer 4*: This layer has two output nodes, action node O_4^A and Q-value node O_4^Q , which represent the defuzzification method. The final outputs are given by:

$$O_4^A = \sum_{i=1}^n O_{3,i}^A \quad (6.13)$$

$$O_4^Q = \sum_{i=1}^n O_{3,i}^Q \quad (6.14)$$

6.2.1 FQL-based interference management

In this section, the FQL algorithm proposed to perform the aggregated interference control from femtocells to macrocells, is presented. This solution is given for case study 2, introduced in Section 4.1.2. Differently from the Q-learning case, in FQL, the components of the state vector are the actual values of the input variables, which eliminates the subjectivity of the state vector definition of Q-learning.

FQL power, macro and femto capacity-based (FQL-PMFB)

Figure 6.3 shows the FQL structure for the FQL power, macro and femto capacity-based (FQL-PMFB) algorithm as a four layer FIS. The functionalities of each layer are the following:

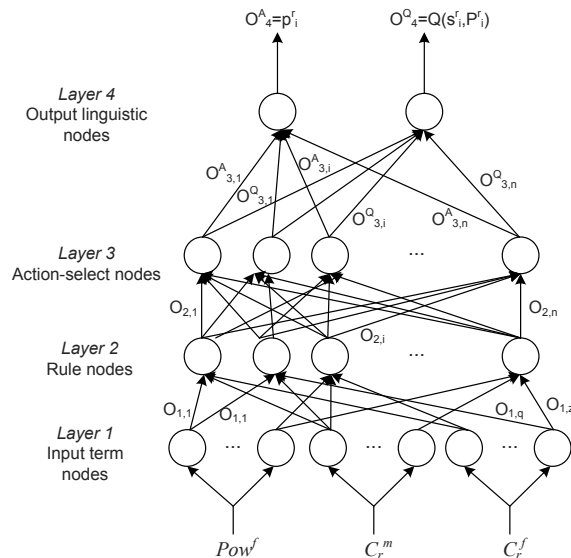


Figure 6.3: FIS structure for the FQL-PMFB algorithm.

- Layer 1:** This layer has as input $L = 3$ linguistic variables, i.e. the femtocell total transmission power Pow^f , the macrocell capacity at RB r , C_r^m and the femtocell capacity at RB r , C_r^f . The input linguistic variables are defined by the term sets: $T(Pow^f) = \{\text{Very Low (VL)}, \text{Low (L)}, \text{Medium (M)}, \text{High (H)}, \text{Very High (VH)}\}$, $T(C_r^m) = \{\text{L}, \text{Medium Low (ML)}, \text{Medium High (MH)}, \text{H}\}$ and $T(C_r^f) = \{\text{L}, \text{ML}, \text{MH}, \text{H}\}$. Therefore, considering the number of fuzzy sets in the three term sets in layer 1 we have $z = |T(Pow^f)| + |T(C_r^m)| + |T(C_r^f)| = 13$ term nodes. Every node is defined by a membership function with a bell shape form (Figure 6.4), so that the output $O_{1,h}$ for a generic component x of \tilde{s} is given by equation 6.9.

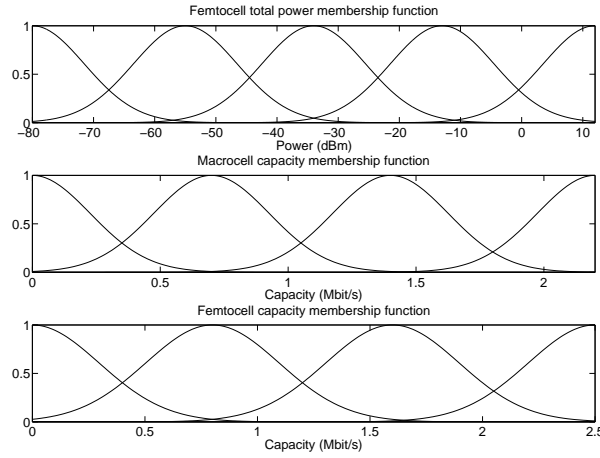


Figure 6.4: Membership functions of the input linguistic variables for FQL-PMFB.

The membership function definitions for the proposed FQL are based on expert knowledge. In particular, we choose five fuzzy sets for Pow^f term set because the range of values to be considered is large and, to keep the total femto transmission power below a maximum value, is a system requirement. For C_r^m and C_r^f fuzzy terms we consider four fuzzy sets per each one, because both indicators are very important from the system performance point of view.

- *Layer 2*: it is composed by $n = |T(Pow^f)| \times |T(C_r^m)| \times |T(C_r^f)| = 80$ nodes. Each node in layer 2 has three input values, one from one linguistic variable of each of the three components of the input state vector.
- *Layer 3*: Here the set of possible actions for each layer 3 node are $l = 60$ power levels. Those power levels range from -80 dBm to 10 dBm Effective Radiated Power (ERP).

Different mean and variance values, and the amount of fuzzy sets for the three term sets description, have been tested through simulations. The actual values used in simulation presented in this chapter are summarized in Table 6.1. Notice that, subjectivity in the term set definitions is absorbed by the learning process in layer 3, due to the inherent adaptive capability of FQL algorithms. The impact of the particular membership functions shapes (i.e. mean and variance) is not significant, since the learning process allows self-adaptation of the FIS to the environment.

6.3 Simulation results

This section is divided into two parts. In the first one, simulation results for FQL-PMFB obtained for the single-cell scenario, introduced in Section 2.3.2, are presented. The performance

Table 6.1: FQL-PMFB simulation parameters

Parameter	Value
Total transmission power variance ρ^h $h = (0, \dots, 4)$	12
Macrocell capacity variance ρ^h $h = (5, \dots, 8)$	0.32
Femtocell capacity variance ρ^h $h = (9, \dots, 12)$	0.42
Total transmission power mean values e^h	-80, -55, -34, -13, 12
Macrocell capacity mean values e^h	0, 0.7, 1.4, 2.2
Femtocell capacity mean values e^h	0, 0.8, 1.7, 2.5

of FQL-PMFB is compared with less complex FQL algorithms in order to represent the proposed approach adaptability. Then, in the second part, the FQL-PMFB is compared with the Q-learning approach proposed in Chapter 3, in the context of multicell scenario. Results presented have been obtained for case study 2, introduced in Section 4.1.2. In both cases C_{\min}^M was set at 1.2 Mbit/s.

6.3.1 FQL-PMFB results for the single-cell scenario

In this section we first present the reference algorithms we use to compare the proposed learning method and then we show some simulation results.

Reference algorithms

The proposed algorithms are compared to two reference FQL algorithms with lower grade of complexity than the proposed FQL-PMFB and a SPC based on interference measurements proposed in [14] and presented in Section 4.2.2.

- **FQL Power-Based (FQL-PB):** The algorithm's objective is to maintain only the total femto BS transmission power below a given threshold. The vector state is represented as:

$$\tilde{s} = \{Pow^f\}$$

And the cost equation is the following:

$$c = \begin{cases} K & \text{if } Pow^f > P_{\max}^F, \\ 0 & \text{otherwise} \end{cases}$$

The rational behind this cost function is that the total transmission power of each femtocell does not exceed the allowed P_{\max}^F .

Since the structure of this algorithm is very similar to the FQL-PMFB we only highlight the different parameters.

- *Layer 1*: The term set of the input linguistic variables is defined by the following fuzzy set: $T(Pow^f) = \{\text{VL}, \text{L}, \text{M}, \text{H}, \text{VH}\}$. The layer 1 of the fuzzy system is composed by $z = |T(Pow^f)| = 5$ term nodes.
 - *Layer 2*: it is composed by $n = |T(Pow^f)| = 5$ nodes.
- **FQL power and macrocell capacity-based (FQL-PMB)**: This algorithm objective is to maintain the total femto BS transmission power below the threshold and at the same time maximize the macrocell capacity. The input state variable is the following:

$$\tilde{s} = \{Pow^f, C_r^m\}$$

and the cost equation:

$$c = \begin{cases} K & \text{if } Pow^f > P_{\max}^F \text{ or } C_r^m < C_{\min}^M, \\ 0 & \text{otherwise} \end{cases}$$

The rationale behind this cost function is that, besides the total transmission power control of the femtocell, it guarantees that the macrocell capacity is above a desired threshold.

The layered structure is as for FQL-PMFB, but:

- *Layer 1*: The linguistic variables of the system are defined by the following fuzzy sets: $T(Pow^f) = \{\text{VL}, \text{L}, \text{M}, \text{H}, \text{VH}\}$, $T(C_r^m) = \{\text{L}, \text{ML}, \text{MH}, \text{H}\}$, so that layer 1 is composed by $z = |T(Pow^f)| + |T(C_r^m)| = 9$ term nodes, each one representing a fuzzy term of an input linguistic variable.
- *Layer 2*: it is composed by $n = |T(Pow^f)| \times |T(C_r^m)| = 20$ nodes.

We evaluate the fuzzy algorithms behavior in terms of macrocell, average femtocell, total system capacity and convergence speed. Figure 6.5 depicts the behavior in terms of macrocell capacity. Analyzing the fuzzy algorithms behavior, we can see that the FQL-PMB and FQL-PMFB algorithms, differently from the FQL-PB algorithm, are required in their cost equations to maintain the macrocell capacity above the C_{\min}^M , and so they behave, contrarily to FQL-PB, which is not required to maintain this target and therefore the agents' adopted actions do not contemplate the macrocell system performance. In addition, it is worth mentioning that the FQL-PMFB obtains lower values of macrocell capacity than FQL-PMB, since due to its cost function definition, it also aims at maximizing the femtocells capacity. Femtocells applying FQL, to fulfil the macrocell performance requirements, can transmit at higher power levels which increases the femto capacity but decreases the macro one. Finally, the FQL-PMB and FQL-PMFB better perform than the benchmark algorithm, since when increasing the occupation ratio of femtocells, it does not adaptively operate to maintain the interference below a threshold.

Figure 6.6 shows the system behavior in terms of average femtocell capacity. It can be observed that average femtocell capacity decreases with the femtocell occupation ratio due to

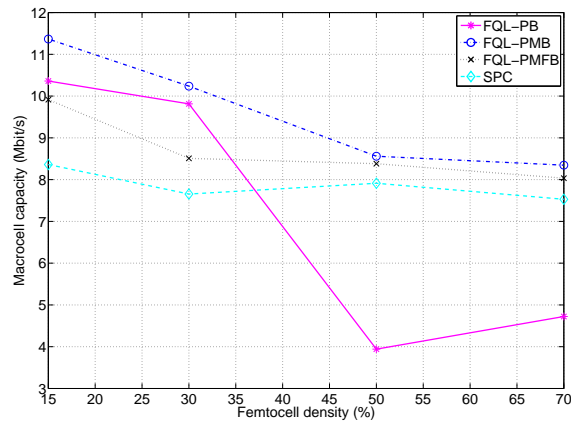


Figure 6.5: Macrocell capacity as a function of the femtocell occupation ratio in single-cell scenario.

the increase of the femto-to-femto interference in all cases. Since FQL-PMFB include in its cost equation (6.3.1) the maximization of the femtocell capacity, it is able to increase the average femtocell capacity up to more than 1 Mbit/s with respect to FQL-PMB and 2 Mbit/s with respect to FQL-PB. As mentioned before, the FQL-PMFB is able to choose more precise actions, which allow femtocells to reach higher performances maintaining the macrocell system requirements. In addition, similarly to the macrocell capacity case, the FQL-PMB and FQL-PMFB algorithms outperform the SPC algorithm.

Figure 6.7 represents the system behavior in terms of total system capacity. As it was expected FQL-PMFB has the higher system capacity, outperforming the FQL-PMB and the SPC algorithm of up to 20 Mbit/s and the FQL-PB of up to 40 Mbit/s. Total system capacity increases with the femtocell occupation ratio since there are more nodes in the network. On the other hand, femtocells capacity decreases as a function of femto nodes occupation ratio due to the interference between them.

Finally, to assess the speed of convergence, we compare the iterations before convergence required by the FQL algorithm and by the Q-learning, under the same conditions, in a scenario with 70% of femtocell occupation ratio. We define convergence as the iteration point where the learning process presents a stable behavior, i.e. when analyzing the probability of being below the capacity threshold, oscillations are not longer present. The FQL increases the speed of convergence up to 25% with respect to the Q-learning algorithm, since it achieves a stable behavior with 25% less learning iterations. The reduction of the FQL required iterations to achieve a stable behavior is due to the previous knowledge that can be embedded in the rules.

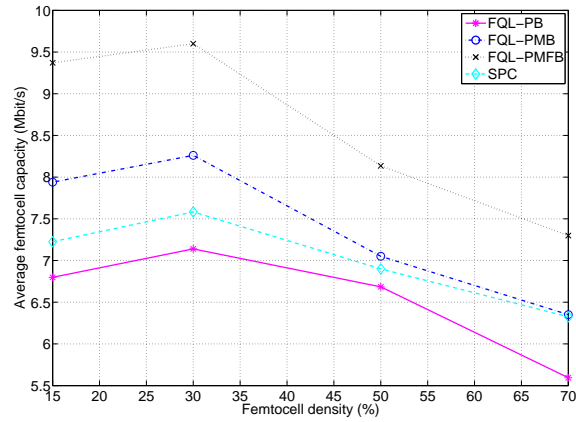


Figure 6.6: Average femtocell system capacity as a function of the femtocell occupation ratio in single-cell scenario.

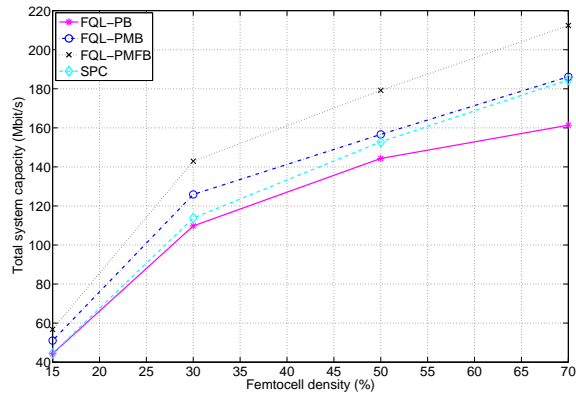


Figure 6.7: Total system capacity as a function of the femtocell occupation ratio in single-cell scenario.

6.3.2 FQL-PMFB results for the multicell scenario

In what follows we present results in terms of system performances and convergence capabilities for multicell scenario.

System performance

In the following figures, we compare the Q-learning and the FQL-PMFB results to the SPC and ITW algorithm [15], introduced in Section 5.3.1.

Figure 6.8 depicts the behavior in terms of macrocell capacity as a function of the femtocell occupation ratio, p_{oc} . The FQL-PMFB algorithm better performs with respect to the Q-learning algorithm due to the fact that it is able to find more accurate actions in each state of the envi-

ronment. This improvement in the system behavior comes associated with the continuous state and action representation allowed by the FQL-PMFB algorithm, which is the main advantage with respect to the Q-learning approach. Finally, FQL-PMFB and Q-learning better perform than the benchmark algorithms, i.e. the SPC and the ITW, since when increasing the occupation ratio of femtocells, they are not able to adaptively operate to maintain the interference below a threshold. These algorithms are not able to react to the dynamics of the environment when users are moving around, femtos are switched on and off, etc. and it is not capable of keeping memory of previous experience.

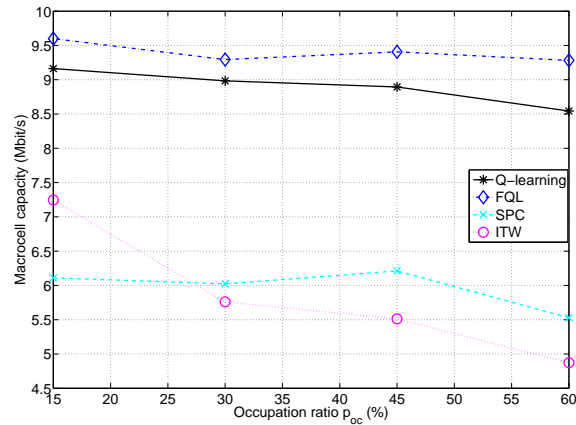


Figure 6.8: Macrocell capacity as a function of the femtocell occupation ratio in multicell scenario.

Figure 6.9 shows the system behavior in terms of average femtocell capacity as a function of the occupation ratio p_{oc} . It can be observed that average femtocell capacity decreases with the femtocell occupation ratio due to the increment of the femto-to-femto interference. Both learning methods try to maximize the femtocell capacity as it is guided by the cost equation. However, FQL-PMFB is able to better perform with respect to the Q-learning and the SPC algorithms, thanks to a better adaptation capability. It is worth mentioning that for FQL-PMFB the same results have been obtained with different membership functions with respect to those shown in Figure 6.9. As a result, the performances do not significantly depend on the selected membership function, since the overlaid Q-learning is able to adapt the fuzzy system to the environment needs. On the other hand, the ITW better perform than the FQL-PMFB, as it was expected, since it performs an optimal power allocation without considering the macrocell system performance.

Convergence capabilities

We now discuss the convergence capabilities of the proposed approaches. We compare Q-learning, to FQL-PMFB and we also study the impact of initializing the inference rules of the FIS in order

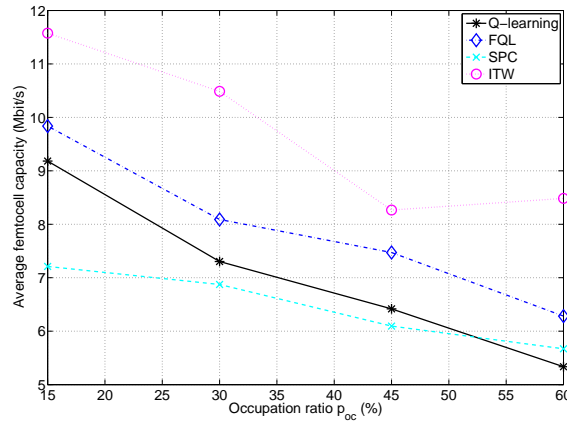


Figure 6.9: Average femtocell capacity as a function of the femtocell occupation ratio for multicell scenario.

to incorporate in the learning scheme offline expert knowledge. To do this, we implement the Init-FQL, which consists of an expert initialization of some of the Q-values in layer 3 nodes. Specifically, the action selection model chooses the action corresponding to the lowest Q-value, therefore, we propose to initialize the Q-values corresponding to critical states and appropriate actions at a lower initial value than the rest of actions. In this way, the agent is expected to find more adequate solutions since the beginning of the learning process, which results in a faster learning period and a lower interference at macrousers. For instance, Q-values corresponding to states with “L” macrocell capacity and “L” power levels can be initialized at lower Q-values than other states. The rationale behind this is that if the macrocell capacity indicator is low, this means that femtocell’s actions may jeopardize the macrouser performance and consequently the transmission power has to be decreased.

Figure 6.10 shows the probability of being above the power threshold as a function of the learning iterations for a femtocell occupation ratio of 45%. In particular, we compute the average required iterations for the three learning systems to reach a probability lower than a benchmark fixed at 2%. Results show that FQL-PMFB needs 57% less iterations than the Q-learning algorithm to reach the target, and the Init-FQL needs 95% less iterations than FQL-PMFB.

Finally, Figure 6.11 represents the probability of being below the capacity threshold as a function of the learning iterations, for a scenario with a femtocell occupation ratio of 60%. As it can be observed, in terms of interference, the FQL-PMFB algorithm is able to better adapt its actions since the beginning of iterations, which results in lower interference at the macro system.

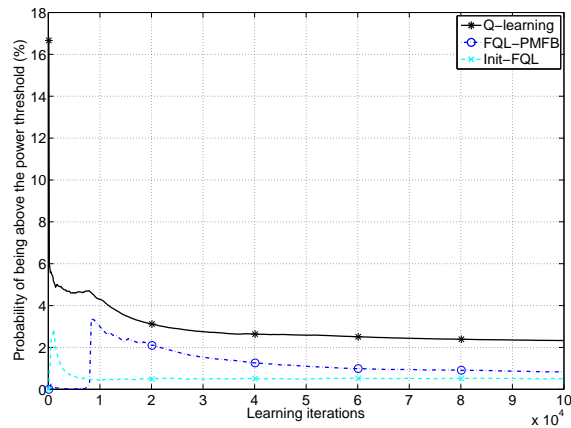


Figure 6.10: Probability of being above the power threshold as a function of learning iterations when applying FQL.

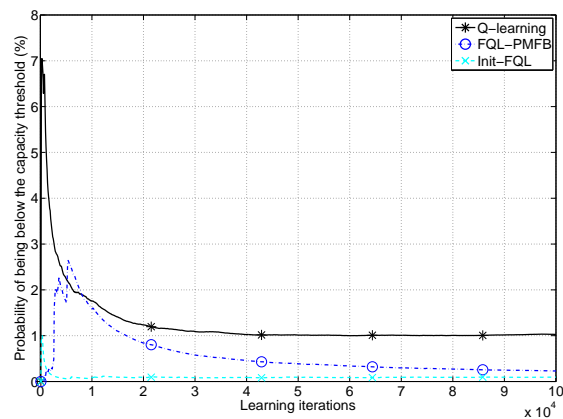


Figure 6.11: Probability of being below the capacity threshold as a function of learning iterations when applying FQL

6.4 Conclusions

In this chapter we have presented a decentralized FQL approach for interference management in a macro-femto network to foster coexistence between macro and femto systems in the same band. We have shown that with respect to distributed Q-learning, FQL allows to operate with detailed and even continuous representation of state and action spaces, reducing the algorithm complexity. In addition, since previous expert knowledge can be naturally embedded in the fuzzy rules, the learning period can be significantly reduced.

First, we have compared the FQL for three different environment perception and objectives in order to show that this system is able to fulfil different requirements from both macrocell and femtocell systems. These results have been obtained for the single-cell scenario. Then, we

have applied the proposed FQL to a more complex scenario, the multicell scenario. Here, we have demonstrated that the proposed scheme outperforms the Q-learning solution and a SPC heuristic approach and the classical ITW algorithm. Finally, very interesting results appear when an expert initialization of the Q-values, in the layer 3 of the fuzzy structure, is performed. With this simple process important gains can be achieved in terms of precision and speed of convergence.

Bibliography

- [1] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.
- [2] H. R. Berenji, "Fuzzy Q-learning: a new approach for fuzzy dynamic programming," in *IEEE World Congress on Computational Intelligence., Proceedings of the Third IEEE Conference on Fuzzy Systems, 1994*, vol. 1, June 1994, pp. 486–491.
- [3] P. Y. Glorennec and L. Jouffe, "Fuzzy Q-learning," in *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems, 1997*, vol. 2, July 1997, pp. 659–662.
- [4] L. Jouffe, "Fuzzy inference system learning by reinforcement methods," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 28, no. 3, pp. 338–355, Aug. 1998.
- [5] P. Y. Glorennec and L. Jouffe, "A reinforcement learning method for an autonomous robot," in *Proceedings of Fourth European Congress on Intelligent Techniques and Soft Computing (EUFIT96)*, Sept. 1996.
- [6] Y.-S. Chen, C.-J. Chang, and F.-C. Ren, "Situation-aware data access manager using fuzzy Q-learning technique for multi-cell WCDMA systems," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 2539–2547, Sept. 2006.
- [7] Y.-H. Chen, C.-J. Chang, and C. Y. Huang, "Fuzzy Q-learning admission control for WCDMA/WLAN heterogeneous networks with multimedia traffic," *IEEE Transactions on Mobile Computing*, vol. 8, pp. 1469–1479, 2009.
- [8] C.-Y. Huang, W.-C. Chung, C.-J. Chang, and F.-C. Ren, "Fuzzy Q-learning-based hybrid ARQ for high speed downlink packet access," in *2009 IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall)*, Sept. 2009.
- [9] M. Haddad, Z. Altman, S. E. Elayoubi, and E. Altman, "A Nash-Stackelberg fuzzy Q-learning decision approach in heterogeneous cognitive networks," in *2010 IEEE Global Telecommunications Conference (GLOBECOM 2010)*, Dec. 2010.
- [10] W.-C. Chung, Y.-Y. Chen, and C.-J. Chang, "HARQ control scheme by fuzzy Q-learning for HSPA+," in *IEEE 73rd Vehicular Technology Conference (VTC Spring), 2011*, May 2011.
- [11] R. Razavi, S. Klein, and H. Claussen, "Self-optimization of capacity and coverage in LTE networks using a fuzzy reinforcement learning approach," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2010 IEEE 21st International Symposium on*, Sept. 2010, pp. 1865–1870.

-
- [12] M. N. ul Islam and A. Mitschele-Thiel, “Reinforcement learning strategies for self-organized coverage and capacity optimization,” in *2012 IEEE Wireless Communications and Networking Conference: PHY and Fundamentals*, April 2012, pp. 2845–2850.
- [13] T. Takagi and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control,” *IEEE transactions on systems, man, and cybernetics*, vol. 15, no. 1, pp. 116–132, 1985.
- [14] “3GPP TR 36.921 evolved universal terrestrial radio access (E-UTRA); FDD home eNode B (HeNB) radio frequency (RF) requirements analysis,” 3GPP, Tech. Rep., March 2010.
- [15] D. López-Pérez, A. Jüttner, and J. Zhang, “Dynamic frequency planning versus frequency reuse schemes in OFDMA networks,” in *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*, 26-29 April 2009.

Chapter 7

Interference management without X2 interface support based on Partial Observable Markov Decision Process

Results presented in former chapters showed that decentralized Q-learning is able to learn a policy to maintain the aggregated interference generated by the femtocells at macrousers under a given threshold. To do this, femtocells need feedback from the macro network about its performance and the aggregated impact they are having on it. The 3GPP LTE network architecture connects neighboring macrocells via the X2 interface [1, 2]. In the solutions given until now across this thesis, we assumed the existence of an X2' interface between macrocells and femtocells, through which femtocells receive a bitmap feedback from near macrocells, equivalent to the RNTP indicator standardized in LTE [3], about the interference perceived by the macrousers. A similar assumption about the existence of an interface between macrocells and femtocells, is also the basis of several schemes proposed in literature, e.g., [4–6].

The main limit of the approach presented in preceding chapters, is the assumption of the existence of the mentioned X2' interface, which has not been yet standardized in release 11 [7]. As a consequence of the lack of direct communication between macrocells and femtocells, the interference management task becomes even more challenging, since the femto network has to completely autonomously make decisions without any feedback about the impact it is having on the victim macrousers. To solve this problem, we rely on the theory of Partially Observable Markov Decision Process (POMDP) [8], a suitable tool for decision making in scenarios with some degree of uncertainty, which has been applied to solve several problems in wireless systems e.g., power control in wireless sensor networks [9], decentralized cognitive radio spectrum access [10], spectrum sensing and access in cognitive radio [11], cognitive radio handoff based on partially observable channel state information [12], network routing [13], coding rate [14], etc.

POMDP works by constructing a set of beliefs about the current state of the environment based on empirical observations. In our particular case, the beliefs depend on the service perception that femtocells estimate at the macrouser receivers. We propose that femtocells, based on the SINR measured at their receivers, build a belief set through spatial interpolation techniques, such as ordinary Kriging [15]. In particular, these beliefs are built by first, estimating the position of potential victim macrousers and then, by interpolating the SINR in those locations. The POMDP learning process is then executed based on this estimated information.

The scenario that we consider is that of networked femtocells, proposed in the framework of ICT BeFEMTO project [16, 17] and presented in Section 2.2. In particular, we focus on networked femtocell systems for residential and corporate scenarios, where femtocells are able to exchange signaling information and interact among each other.

The advantage of the proposed solution is twofold. On the one hand, it allows femtocells to work in a completely autonomous fashion, which responds to the increasing need of self-organization of the overall network and to the possibility that the X2' interface will never be standardized. On the other hand, it avoids the signaling burden on the backhaul network introduced by the signaling overhead over the X2' interface required for the Q-learning implementation. We compare the performances of both cases of femtocells with complete and partial observation of the environment. Results show that, after a training phase, the proposed solution for partially observable scenarios is able to solve the aggregated interference management problem generated by the femto network, while continuously and autonomously adjusting to the high dynamics of the surrounding environment and without introducing signaling overhead in the system.

In what follows, first Section 7.1 presents the proposed methodology for partially observable environments. Then, Section 7.2 introduces the spatial interpolation method used to construct the required POMDP's observations. Afterwards, the Q-learning for partially observable environments is presented in Section 7.3. Finally, Section 7.4 summarizes relevant simulation results for both cases of complete and partial information applied in both, the single-cell and the multicell scenarios.

7.1 Proposed learning methodology for partially observable environments

In this chapter we propose an aggregated femto to macro interference management where femtocells autonomously operate, without receiving any feedback from the macro network through the X2' interface, about the interference they are causing to macrousers in downlink operation. We propose that each femtocell first estimates the position of potential victim macrousers,

and then estimates the aggregated interference those users may be receiving through a spatial characterization based on local femtocells measurements. These measurements can be shared among femtocells since we assume they are networked and can exchange signaling information. In what follows, we propose a methodology consisting of four steps. Notice that, the information gathered in Steps 1 and 2 of the proposed methodology is based on references that will be provided, whereas in this chapter we will focus on Steps 3 and 4, which are explained in detail in Sections 7.2 and 7.3, respectively.

Step 1. Femtocells location determination: We assume femtocells to be able to evaluate their own position through femtocell positioning techniques e.g., as those proposed in [18] by ICT BeFEMTO project. In femtocells, all location determination algorithms are included in the HMS [19].

Step 2. Estimation of macrousers position: Based on a Motorola's proposal discussed in a patent filed in February 2011 [20], we assume that every time a macrouser is in the coverage area of a femtocell, it attempts an access to it, which may be either accepted, in case the macrouser belongs to the closed subscriber group of the femtocell, or rejected in case the macrouser does not belong to it. If the access is rejected, the femtocell is aware of the presence of a potential victim macrouser. We also suppose that the macrouser reports about the RB in which it is operating. When the macrouser is in the coverage area of at least three femtocells, and has consequently attempted three accesses, we assume that by some positioning technique e.g., by triangulation, the networked femtocells are able to jointly estimate the position of the macrouser. Notice that, differently from what happens in traditional cellular networks, the handover macro to femto in [20], is not started by the network, but a direct signaling message (supposed not to be power controlled) is sent from the macrouser to the femtocell. This is shown in Figure 7.1, extracted from [20], which represents the exchanged messages between a macrouser and a femtocell when the macrouser attempts to access the femtocell.

Step 3. SINR estimation at macrouser: Once the femtocells have detected the presence of a victim macrouser in the coverage area of the femtocell network, and have estimated their position, we propose that based on the SINR they measure, they perform a spatial interpolation to approximate the SINR at the victim macrouser position. We perform the spatial interpolation through the ordinary Kriging interpolator algorithm [15], which is explained in detail in Section 7.2. The estimated SINR at victim macrousers will be required in Step 4 as input to the learning process, as a macrocell system performance indicator.

Notice that the SINR interpolation could be performed in a centralized form at LFGW

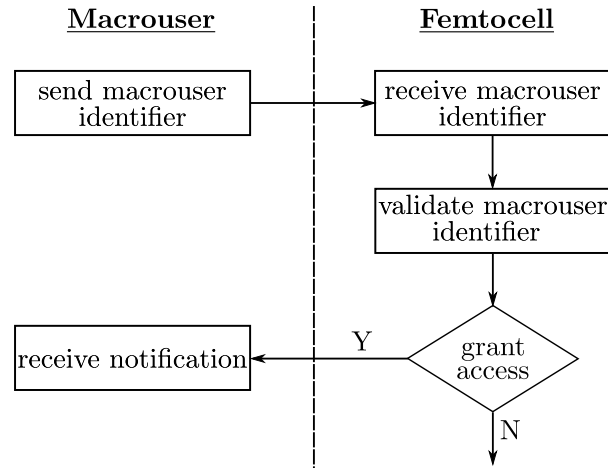


Figure 7.1: Messages exchanged when a macrouser attempts to access a femtocell.

level. In particular, the LFGW would perform the SINR approximation based on the SINR reported by the femtocells through the S1 interface. Then, the LFGW would communicate the approximated SINR value back to the femtocells, also through the S1 interface. This centralized alternative would diminish the computational requirements since parallel interpolation processes at the multiple femto nodes would be avoided. Furthermore, information exchange would be reduced, instead of SINR measurements exchange among femtocells, they would have to report only once their measurements and receive back the estimated value by the LFGW.

Step 4. Learning in partially observable environments: With the approximated SINR estimated in Step 3, agents implemented in the femtocells have enough information to learn decision policies through the theory of POMDP. The estimated SINR is the POMDP's required observation, which provides information about the state of the environment. Further details of the proposed learning method are presented in Section 7.3.

7.2 Spatial characterization of interference in femtocell networks

In this section, we present the probabilistic analysis and modeling of the interference perceived at macrousers. This interference modeling is based on the measurements of the SINR gathered by those femtocells which the user attempts an access to, as defined in Step 3 in Section 7.1. Femtocells measurements are treated as a realization of a random field, assuming the SINR as a stationary stochastic process [21]. Once measurements have been gathered, the random field fitting process is performed in two steps:

1. A structure analysis of the spatial continuity properties is performed, which consists of measuring the variability of the measured femtocells SINR, through a variogram model. Then, one of the available variogram models is selected and properly fitted.
2. The value at an unmeasured location is estimated through an interpolation process using variogram properties of neighboring data.

We aim to estimate the aggregated interference generated at macrouser receiver $\hat{Z}(x^*)$, whose location is estimated in position x^* . We consider as input the SINR $Z(x_1), \dots, Z(x_n)$ perceived at n known locations x_1, \dots, x_n , which are the locations of the n closest femtocells receivers. To evaluate the spatial behavior of the SINR over an area, a variogram analysis among the n known locations is required. The experimental variogram data are computed as the expected squared increment of the values between locations x_i and x_j , characterized as a sample of a random field, such as:

$$\gamma(x_i, x_j) = \mathbf{E}[(Z(x_i) - Z(x_j))^2]$$

More in particular, for the case of a stationary field, it is possible to use an empirical variogram based on sample measurements at the n different locations $Z(x_1), \dots, Z(x_n)$. First, the distance is divided in a set of lags with separation h , then distances between measurement sites $\|x_i - x_j\|, \forall i, j$, with similar separation, are grouped into bins $N(h_i)$ centered in h_i . Finally, the empirical variogram is obtained through:

$$\gamma(h_i) = \frac{1}{N(h_i)} \sum_{i=1}^{N(h_i)} (Z(x_i) - Z(x_j))^2$$

The empirical variogram cannot be used directly for the interpolation process because not all the distances are present in the sample data, so that a variogram model is required.

Variogram models are mathematical functions that describe the degree of spatial dependence of a spatial random field. There are infinitely possible variogram models, and the more commonly used are linear, exponential, gaussian and spherical. Variogram models are characterized by sill ρ_0^2 , range a_0 and nugget c_0 parameters, as shown in Figure 7.2, whose values are computed through the fitting process.

1. The sill, ρ_0^2 , is the variogram model upper bound. It is equal to the total variance of the data set.
2. The range, a_0 , is the distance h where the fitted variogram model becomes constant with respect to the lag distance. For the particular case of exponential and gaussian models (see details of the models in the following) where the variogram model increases asymptotically toward its sill value, the term practical range is also used, and it is chosen so that the value

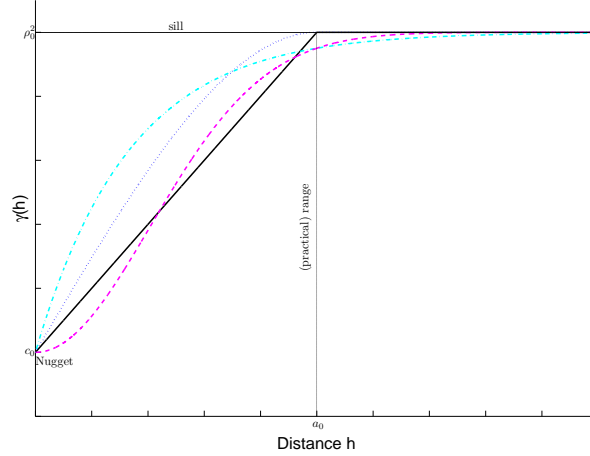


Figure 7.2: Generic variogram parameters.

of the resulting exponential or gaussian function evaluated at the practical range lag is 95% of the sill value.

3. The nugget, c_0 , is the value at which the variogram model intercepts the y-axis.

The following equations characterize the typical examples of variogram models and they are represented in Figure 7.2.

Linear model:

$$\gamma(h) = \begin{cases} \frac{\rho_o^2}{a_o} |h|, & 0 < |h| \leq a_o, \\ \rho_o^2, & a_o < |h|, \end{cases}$$

Spherical model:

$$\gamma(h) = \begin{cases} \rho_o^2 \left[1.5 \frac{|h|}{a_o} - 0.5 \left(\frac{|h|}{a_o} \right)^3 \right], & 0 < |h| \leq a_o, \\ \rho_o^2, & a_o < |h|, \end{cases}$$

Exponential model:

$$\gamma(h) = \begin{cases} 0, & h = 0, \\ \rho_o^2 \left[1 - \exp\left(-\frac{3|h|}{a_o}\right) \right], & h \neq 0, \end{cases}$$

Gaussian model:

$$\gamma(h) = \begin{cases} 0, & h = 0, \\ \rho_o^2 \left[1 - \exp\left(-\frac{|h|}{a_o}\right)^2 \right], & h \neq 0, \end{cases}$$

The experimental variogram data are fitted to the variogram model by matching the shape of the curve of the experimental variogram data $\gamma(x_i, x_j)$, with the shape of the mathematical

function $\gamma(h)$ by least-square regression. The variogram model parameters are then accordingly determined. The most appropriate variogram model is chosen by computing the error between experimental variogram data and the corresponding variogram model values in those points.

Once the variogram model is selected, the interpolation procedure can be performed. Interpolation allows the estimation of a variable at an unmeasured location based on the observed values at surrounding locations. We apply the ordinary Kriging [15] interpolation technique, whose main advantages are that it compensates the effects of data clustering (i.e. data within a cluster have less weight than isolated data points) and it provides estimation of errors, which allows to find the best unbiased estimation of the random field values between the measurement points. Assuming a stationary field, where the mean expected value $\mathbf{E}[Z(x^*)] = \mu$ is unknown but constant, and the variogram is known, the ordinary Kriging estimation is given by a linear combination such as:

$$\hat{Z}(x^*) = \sum_{j=1}^n \lambda_j Z(x_j)$$

where λ_j is the weight given to the observed value $Z(x_j)$. Weights should be chosen such that the variance $\sigma^2(x^*) = \text{var}(\hat{Z}(x^*) - Z(x^*))$ of the prediction error $\hat{Z}(x^*) - Z(x^*)$ is minimized, subject to the unbiased condition $\mathbf{E}[\hat{Z}(x^*) - Z(x^*)] = 0$. The weights are computed based on the ordinary Kriging equation system:

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_q \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma(x_1, x_1) & \cdots & \gamma(x_1, x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_q, x_1) & \cdots & \gamma(x_q, x_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma(x_1, x^*) \\ \vdots \\ \gamma(x_q, x^*) \\ 1 \end{pmatrix}$$

where the additional parameter μ is a Lagrange multiplier used in the minimization of the variance $\sigma^2(x^*)$ to fulfil the unbiasedness condition $\sum_{j=1}^n \lambda_j = 1$.

7.3 Q-learning in partially observable environments

In many real world problems, it is not possible for the agent to have perfect and complete perception of the state of the environment. As a result, it makes sense to consider situations in which the agents make observations of the state of the environment, which may be noisy, or in general do not provide a complete picture of the state of the scenario. This is exactly the situation of a femtocell network autonomously making decisions without the assistance of the macro network through the X2' interface. The femtocells measure the SINR, and through this, they estimate the SINR at a given position where it has been estimated that a victim macrouser is located. This information is only a partial and noisy representation of the reality, and is in general affected by error. The resulting formal model to operate in this kind of environments is called POMDP [8][22]. A POMDP is based on a State Estimator (SE), which computes the agent's *belief state* b , as a function of the old belief state, the last action and the current

observation the agent makes of the environment o , as it is shown in Figure 7.3. In this context, a belief state is a probability distribution over states of the environment, indicating the likelihood that the environment is actually in each of those states given the agent's past experience. The SE can be constructed straightforwardly using the estimated world model and Bayes' rule [22]. So a POMDP consists of:

- a set of agents N .
- a set of states \mathcal{S} .
- a set of actions \mathcal{A} .
- a cost function $\mathcal{C} : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$.
- a state transition function $P : \mathcal{S} \times \mathcal{A} \rightarrow \Pi(\mathcal{S})$.
- a set of observations Ω .

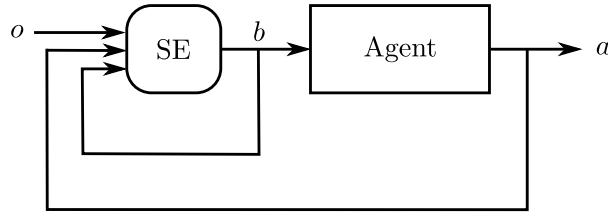


Figure 7.3: Basic structure of POMDP technique.

Similarly to the case of complete information, an optimal policy cannot be found. To implement a solution, we rely on the results in [8], where mechanisms to find reasonably good suboptimal policies are studied. In partially observable environments, the goal is to find a policy for selecting actions that minimize an infinite-horizon, discounted optimality criterion, based on the information available to the agents. The way to proceed is to maintain a probability distribution over the states of the underlying environment. We call this distribution *belief state* $\mathcal{B} = \{b(1), b(2), \dots, b(k)\}$, and we use the notation $b(s)$ to indicate the agent's belief that it is in state s .

Based on the belief state, we can find an approximation of the Q-function, Q_b , as follows:

$$Q_b(b) = \sum_s b(s)Q(s, a) \quad (7.1)$$

and we can use Q_b as a basis for action selection in the environment. Once the action a is selected and executed, the Q-learning update rule has to be generalized, so that the Q-value $Q(s, a)$, corresponding to state s and action a , is updated according to a weight, which is the belief that the agent is actually occupying the state s [8]:

$$Q(s, a) \leftarrow Q(s, a) + \Delta Q_b(s, a) \quad (7.2)$$

where $\Delta Q_b(s, a)$ is:

$$\Delta Q_b(s, a) = \alpha b(s)[c + \gamma \min_a Q(b', a) - Q(s, a)]$$

where b' is the resulting belief state, after the execution of action a .

Now, to define the POMDP system, it is necessary to identify the system state, belief state, action, associated cost and the observations.

- **State:** The environment state is characterized as defined in equation (4.6). Given the lack of communication between macrocells and femtocells, the uncertainty in the state definition is only related with the \hat{C}_r^m indicator. The other state components, i.e. the femtocell total transmission power and its corresponding capacity are already known by the femtocell.
- **Action:** The set of possible actions are the l power levels. Here, the action selection procedure is performed based on the Q_b obtained from equation (7.2).
- **Belief state:** For each learning process, the femtocells have to build the belief state $\mathcal{B} = \{b(H), b(L)\}$, defined by two components, i.e. $b(H)$ and $b(L)$, which represent the belief of the femtocell that the capacity at the macrouser is above or below the threshold, respectively. These beliefs are the result of the interpolation process.
- **Cost:** The cost equation is the same as the one defined for completely observable environments in equation (4.9), but replacing \hat{C}_r^m by \tilde{C}_r^m . In particular, \tilde{C}_r^m is the capacity of macrocell m in RB r estimated by the interpolation method defined in Step 3 of the proposed learning methodology.
- **Observations:** The set of observations Ω is characterized by all the estimations that the femtocell has to compute, as described in Steps 2 and 3 of the proposed methodology, which consist of positioning of the macrouser, and of the estimation of the aggregated interference at the macrouser, based on the aggregated interference received by the networked femtocells.

7.4 Simulation results

In this section we present the simulation results obtained for the POMDP proposed approach. Again, we divide the results in two parts, the first one presents the outputs obtained for single-cell scenario, presented in Section 2.3.2 and the second presents the outputs obtained when applying POMDP in the multicell scenario, introduced in Section 2.3.3, both, for case study 2 and required C_{\min}^M per RB of 1.2 Mbit/s. More in particular, we present some significant simulation results regarding the implementation of ordinary Kriging spatial interpolator, the learning algorithms behavior, and the femto and macro systems performance. We will refer to

the Q-learning algorithm for completely observable environments, presented in Section 3.2.1 as Q-learning and to the Q-learning algorithm for partially observable environments as POMDP.

7.4.1 Single-cell scenario results for POMDP

In this section we have considered the spherical variogram model. Figure 7.4 represents the spherical model fitting curve for the sample variogram obtained for an occupation ratio of $p_{oc} = 60\%$. As it can be observed, the fitting curve reveals that the sample variogram can be correlated depending on the distance.

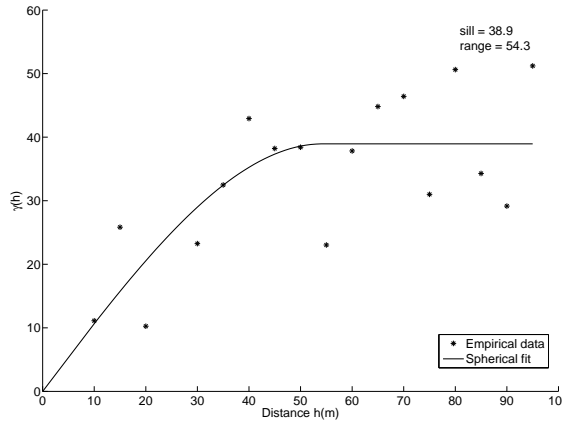


Figure 7.4: Variogram fit with spherical model for the single-cell scenario with $p_{oc} = 60\%$.

In order to test the ordinary Kriging estimator performance, we compute an Interpolation Error (IE) in such a way that the error counter is incremented if the capacity estimated at femtocells is above the C_{min}^M threshold and the actual capacity perceived by macrouser is below it and viceversa. Figure 7.5 represents the CCDF which indicates how often the IE is above a particular level. As it can be observed, the probability to obtain an IE higher than 6% is less than 8%. Therefore we can affirm that, despite the complexity of the proposed methodology, errors resulting from the interpolation process can be tolerated. It is also worth mentioning that we are assuming hard constraints to measure the IE. This means that interpolated values that are very close to the actual values may still be counted as IEs if the actual and the interpolated values do not fall into the same side of the C_{min}^M threshold.

We now discuss the convergence capabilities of the proposed approach. We compare the Q-learning process for both cases of partially and completely observable environments. Figure 7.6 represents the probability of being below the capacity threshold as a function of the learning iterations, for a scenario with a femtocell occupation ratio of $p_{oc} = 45\%$. As it can be observed, in terms of interference, the femtocell applying the learning process with partially observable environment generates a more unstable performance at macrouser compared to Q-learning with

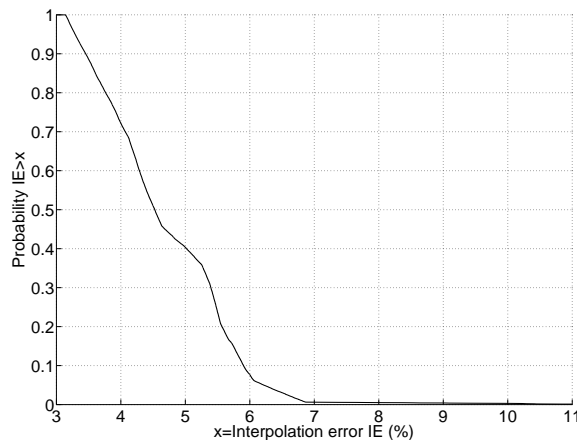


Figure 7.5: CCDF of the error of the SINR estimated by femto BS in the single-cell scenario.

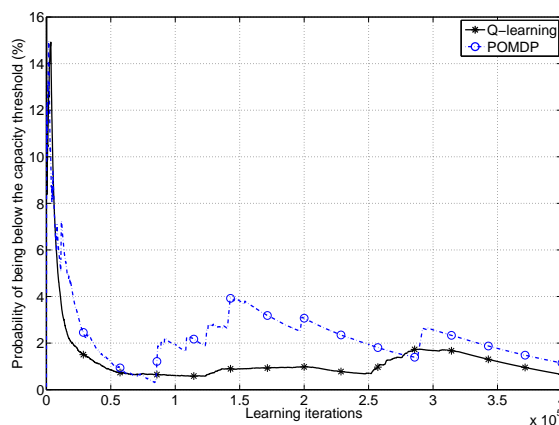


Figure 7.6: Probability of being below the capacity threshold as a function of the learning iterations in the single-cell scenario.

complete information. This instability is associated to the inherent errors in the femtocells observations based on which they perform the action selection process.

Figure 7.7 represents the macrocell and femtocell average capacities for partially observable and completely observable environments. As can it be observed, both learning systems keep the macrocell capacity above the C_{\min}^M threshold. As a result, it is demonstrated that through the proposed approach it is reasonable to operate without the support of X2' interface in partially observable environments. However, for the partially observable case, power levels are selected in a more conservative way with respect to the completely observable case, as a result of the errors in the observations. This results in higher macro capacities and lower femto capacities for the partially observable case with respect to the completely observable case. In this sense, Q-learning with partial information achieves a poorer load balancing in the network, which is the price to pay for the unavailability of the X2' interface.

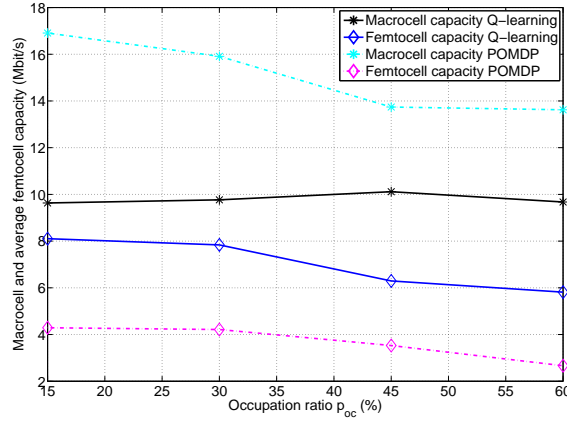


Figure 7.7: Macrocell and femtocell average capacity as a function of the femtocell occupation ratio for single-cell scenario.

7.4.2 Multicell scenario results for POMDP

Here, in each learning iteration the most appropriate variogram model is selected based on the computation of the least squared error, according to which, the variogram parameters, the sill and the range, are determined. Figure 7.8 represents the linear, spherical, exponential and gaussian variogram model fitting curves for the sample variogram obtained for an occupation ratio of $p_{oc} = 45\%$. As it can be observed, the linear and gaussian variogram models present a better adjustment to the sample variogram. Based on empirical results, we observe that, across simulations, these two models have been the most widely used to fit the scenario sample data.

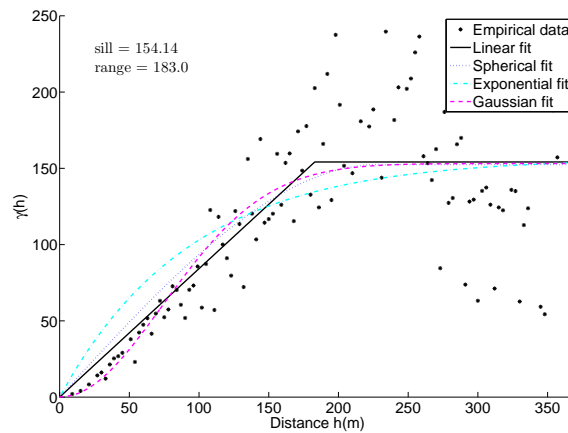


Figure 7.8: Variogram fit with different models for the multicell scenario.

In order to test the ordinary Kriging estimator performance, we represent the CDF of the IE in Figure 7.9 for different occupation ratios. As it can be observed, the IE is lower than 4% with a probability higher than or equal to 0.95, for all the studied cases. In particular, it is worth

mentioning that the error decreases with the occupation ratio due to the availability of more SINR samples to interpolate, and since it is more likely that the SINR measurements based on which interpolation is performed are taken by femtocells closer to the macrouser. Therefore, it is more likely that femtocells measured values are similar to the actual one experienced at the macrouser.

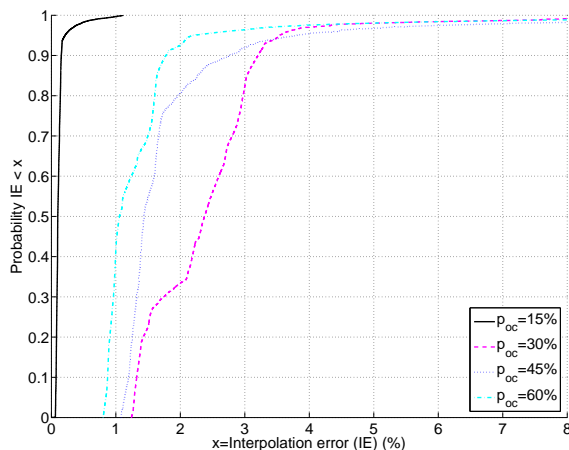


Figure 7.9: CDF of the error for the SINR estimated by femtocells in the multicell scenario.

We now discuss the convergence capabilities of the proposed approaches, results are represented in Figures 7.10 and 7.11. We observe that in all cases the learning process is characterized first, by a training phase where the agent explores and learns proper decision policies and then, by an exploitation phase, where these policies are enforced and the algorithm's behavior is stable. In particular, independently of the dynamism of the scenario, the system's objective, defined in the cost function (4.9), are achieved. We compare the Q-learning process for both cases of partially and completely observable environments in terms of capacity and total transmission power, for a scenario with a femtocell occupation ratio of $p_{oc} = 45\%$. Figure 7.10 shows the probability of being below the power threshold (see Definition 2 in Chapter 3), which is the first constraint to be met in cost equation (4.9), as a function of the learning iterations. Here, as it was expected, Q-learning presents a faster convergence behavior than POMDP learning algorithm, which besides presents more fluctuations.

Figure 7.11 represents the probability of being below the capacity threshold (see Definition 1 in Chapter 3), which is the second constraint to be met in cost equation (4.9), as a function of the learning iterations. POMDP performances are poorer than Q-learning's, especially during the training phase, in the first part of the learning process. The reason is that, by receiving an explicit feedback from the macro network, Q-learning can learn more quickly the appropriate policies for decisions. In addition, POMDP performance is affected by the inherent errors in the femtocells observations, based on which it performs the action selection process. Despite this longer training phase, we observe that in the exploitation phase POMDP as well correctly learns

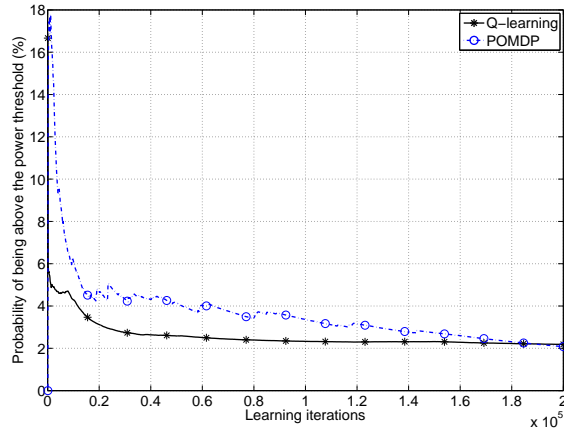


Figure 7.10: Probability of being above the power threshold as a function of the learning iterations in multicell scenario.

how to properly act in the proposed dynamic environment.

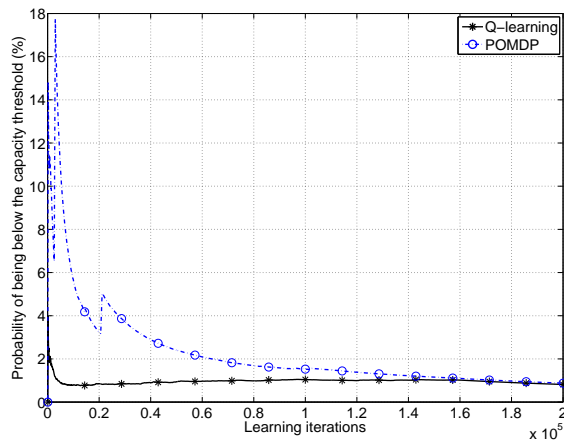


Figure 7.11: Probability of being below the capacity threshold as a function of the learning iterations in multicell scenario.

Translating the learning iterations, presented in Figures 7.10 and 7.11, into time and considering that after 100000 learning iterations, the POMDP algorithm presents a more stable behavior, we can affirm that the learning process takes 100 s. This computation is based on the fact that the POMDP algorithm does not require feedback from the macrocell system, then, the learning process can be performed every 1 ms, i.e. LTE scheduling time basis. In next chapter, an analysis regarding the feasibility of this proposal, is presented.

We now discuss the systems performance in terms of femto and macrousers achieved capacity. Figure 7.12 presents the macrocell capacity for partially and completely observable environments, as well as for the SPC algorithm, presented in Section 4.2.2. As it can be observed,

both learning systems keep the macrocell capacity above the C_{\min}^M threshold. As a result, it is demonstrated that through the proposed approach it is possible to operate without the support of X2' interface in partially observable environments, in compliance with 3GPP specifications. Notice that POMDP selects an average lower power than Q-learning since, having to operate in an uncertain and dynamic scenario, it makes more conservative decisions than Q-learning. This behavior was also observed in the single-cell scenario. This results in higher macro capacity for POMDP than Q-learning. Both learning systems, however, better behave than SPC, since this approach, differently from the learning schemes proposed, is not able to timely react to the high variability that we have introduced in the simulated scenario. This remarkable capacity of the learning schemes to react to the continuous changes of the surrounding environment, where appropriate decisions policies have been learnt, is the most important distinguishing feature with respect to non cognitive approaches.

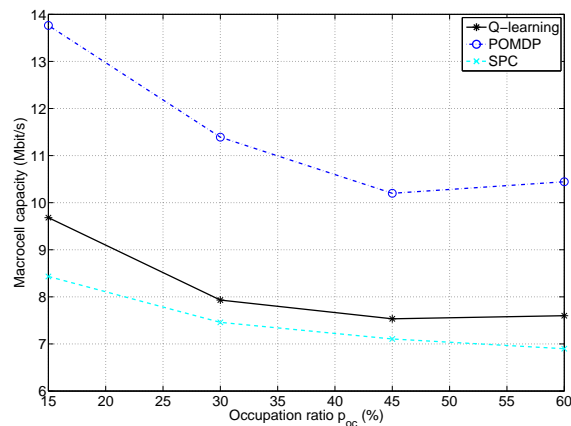


Figure 7.12: Macrocell capacity as a function of the femtocell occupation ratio for multicell scenario.

Finally, Figure 7.13 presents the femtocell average capacities for the three algorithms. As it can be observed, Q-learning offers higher femtocells average capacities than the POMDP and SPC. The reason is the same as discussed for macrocell capacity, that is, POMDPs' transmission power levels selected by femtocells are more conservative with respect to the Q-learning case, as a response to the errors in the observations. In particular, we observe through simulations that power levels selected by Q-learning and POMDP have average differences of up to 5 dB, which explains the resulting systems performance. As already mentioned, this results in higher macro capacities and lower femto capacities for the partially observable case with respect to the completely observable case. In this sense, POMDP achieves a poorer load balancing in the network, which is the price to pay for the unavailability of the X2' interface.

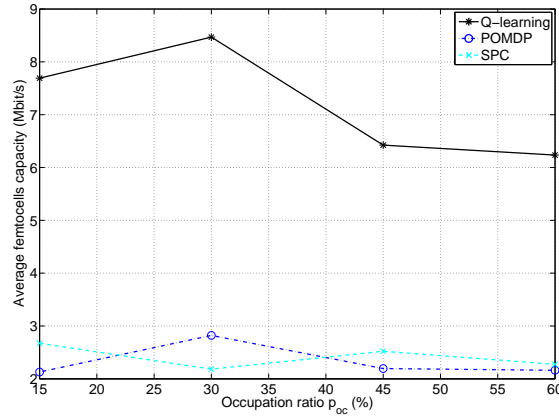


Figure 7.13: Femtocell average capacity as a function of the femtocell occupation ratio for multicell scenario.

7.5 Conclusion

In this chapter we have presented an autonomous learning algorithm for a heterogeneous network scenario, consisting of macrocells and femtocells working in co-channel operation, to solve the aggregated interference generated by the multiple femtocells at macrocell receivers. Differently from the work presented in previous chapters and from several works available in literature, we consider 3GPP release 11 compliant architectural hypothesis, according to which the X2' interface from macrocells to femtocells is not available. The lack of feedback, which could be gathered through this X2' interface, makes the interference management problem even more challenging, since femtocells have to make decisions in a completely autonomous fashion.

The resulting theoretical framework to model the interference management problem is a stochastic game with partial information, where the agents' decisions affect the perception of the environment of neighbor nodes, and where the environment is also affected by the typical dynamics of a wireless scenario. This game is proposed to be solved by means of the theory of POMDP, which works by constructing beliefs about the state of the environments. The belief set is built through the femtocells SINR measurements and by spatially interpolating them through the ordinary Kriging technique. Extensive simulation results have shown that POMDP algorithm is able to learn a sub-optimal solution, which guarantees to maintain the macrocell system performance above a desired threshold, allowing an autonomous femtocells deployment and avoiding the introduction of signaling overhead in the system when considering both, the single-cell and the multicell scenario.

Bibliography

- [1] 3GPP, “X2 Application Protocol (X2AP) (Release 8),” 3GPP TS 36.423 V8.2.0 (2008-06), June 2008.
- [2] —, “X2 General Aspects and Principles (Release 8),” 3GPP TS 36.420 V8.0.0 (2007-12), Dec. 2007.
- [3] —, “Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures (Release 8),” 3GPP TS 36.213 V 8.8.0 (2009-09), Sep. 2009.
- [4] S. Oh, H. Kim, B. Ryu, and N. Park, “Inbound mobility management on LTE-Advanced femtocell topology using X2 interface,” in *Proc. of 2011 20th International Conference on Computer Communications and Networks (ICCCN)*, Maui, Hawaii, July 31 2011-Aug. 4 2011.
- [5] W. Liu, C. Hu, D. Wei, M. Peng, and W. Wang, “An overload indicator & high interference indicator hybrid scheme for inter-cell interference coordination in LTE system,” in *Proc. of the 2010 3rd IEEE International Conference on Broadband Network and Multimedia Technology (IC-BNMT)*, Beijing, China, 26-28 Oct. 2010, pp. 514–518.
- [6] R. Combes, Z. A. M. Haddad, and E. Altman, “Self-optimizing strategies for interference coordination in OFDMA networks,” in *Proc. of the 2011 IEEE International Conference on Communications Workshops (ICC)*, Kyoto, Japan, 5-9 June 2011.
- [7] 3GPP, “Evolved Universal Terrestrial Radio Access (E-UTRA); Overall description; Stage 2 (Release 11),” 3GPP TS 36.300 V11.1.0 (2012-03), March 2012.
- [8] M. L. Littman, A. R. Cassandra, and L. P. Kaelbling, “Learning policies for partially observable environments: Scaling up,” *Proceedings of the Twelfth International Conference on Machine Learning*, pp. 362–370, 1995.
- [9] A. Udenze and K. McDonald-Maier, “Partially observable markov decision process for transmitter power control in wireless sensor networks,” in *Proc. of the Bio-inspired Learning and Intelligent Systems for Security, 2008. BLISS '08. ECSIS Symposium on*, Edinburgh, 4-6 Aug. 2008, pp. 101–106.
- [10] Q. Zhao, L. Tong, A. Swami, and Y. Chen, “Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework,” *IEEE Journal on Selected Areas in Communications*, vol. 25, pp. 589–600, 2007.
- [11] J. Unnikrishnan and V. V. Veeravalli, “Dynamic spectrum access with learning for cognitive radio,” in *Proc. of the Signals, Systems and Computers, 2008 42nd Asilomar Conference on*, Pacific Grove, CA, 26-29 Oct. 2008, pp. 103–107.

-
- [12] R.-T. Ma, Y.-P. Hsu, and K.-T. Feng, "A POMDP-based spectrum handoff protocol for partially observable cognitive radio networks," in *Proc. of the 2009 IEEE conference on Wireless Communications & Networking Conference (WCNC'09)*. Budapest, Hungary: IEEE Press, 2009, pp. 1331–1336.
- [13] B. Rathnasabapathy and P. Gmytrasiewicz, "Formalizing multi-agent POMDPs in the context of network routing," in *Proc. of the 36th Hawaii International Conference on System Sciences (HICSS03)*, vol. 9. IEEE Computer Society, 2003.
- [14] A. K. Karmokar, D. V. Djonin, and V. K. Bhargava, "POMDP-based coding rate adaptation for type-I hybrid ARQ systems over fading channels with memory," *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, pp. 3512–3523, 2006.
- [15] G. Bohling, "Kriging," *C&PE 940*, October 2005.
- [16] The BeFEMTO website. [Online]. Available: <http://www.ict-befemto.eu/>
- [17] "D2.1: Description of baseline reference systems, use cases, requirements, evaluation and impact on business model," EU FP7-ICT BeFEMTO project, Dec. 2010.
- [18] "D4.1: Preliminary SON enabling & multi-cell RRM techniques for networked femtocells," EU FP7-ICT BeFEMTO project, Dec. 2010.
- [19] 3GPP, "3GPP TS 25.467 V8.1.0., UTRAN architecture for 3G home nodeB (stage 2)," Tech. Rep., March 2009.
- [20] Y. Cai, X. C. P. Ding, X. Jin, and R. P. Moorut, "Management interference from femtocells," U.S. Patent 12/536,125, Feb. 11, 2010.
- [21] S. Firouzabadi, M. Levorato, D. O'Neill, and A. J. Goldsmith, "Learning interference strategies in cognitive ARQ networks," in *Proc. of the Global Communications Conference (GLOBECOM 2010)*, Miami, USA, 6-10 Dec. 2010.
- [22] K. P. Murphy, "A survey of POMDP solution techniques," *Technical Report, UC Berkeley*, 2000. [Online]. Available: <http://www.cs.ubc.ca/murphyk/papers.html>.

Chapter 8

Memory and computational requirements for a practical implementation

When new algorithms are proposed to be included in communication systems, it has to be considered that they must be embedded in Integrated Circuits (ICs). ICs have limits in terms of amount of processing capacity in a given amount of time as well as in the information they can store. Despite the fact that these constraints could preclude the introduction of a given algorithm in new communication systems, literature regarding this kind of analysis is almost nonexistent, to the best of the author's knowledge. In this chapter, in order to test if the proposed learning approaches can be incorporated in state of the art communication ICs, a study regarding memory and computational requirements is performed.

In this chapter, first an introduction to the different IC architectures is presented in Section 8.1. Then, Section 8.2 introduces the Digital Signal Processors (DSPs) main characteristics. This section also summarizes the results, in terms of memory and computational requirements, obtained for Q-learning, presented in Chapter 3, FQL, introduced in Chapter 6, and POMDP, described in Chapter 7. Furthermore, a comparison between the lookup table and the neural networks representation mechanisms is presented.

8.1 Integrated circuit architectures

Currently, the majority of communication algorithms are implemented in one or several of the following IC architectures:

- Digital Signal Processors (DSPs): are specialized microprocessors designed to perform signal processing algorithms in real-time. DSPs can be potentially reprogrammed by the

customer and they can perform mathematical operations faster and with less energy consumption than general purpose microprocessors, since they often include several independent execution units that are capable of operating in parallel. DSPs also have hardware extensions to implement specific computationally intensive algorithms, which allow them to accelerate efficiently complex processes [1, 2].

- **Field Programmable Gate Arrays (FPGAs):** are ICs designed to be configured by the customer or designer after manufacturing. They are formed by numerous logic elements such as registers, logic gates, multipliers, memories, addresses, etc. and reconfigurable interconnections. The FPGA is programmable using a hardware description language such as VHSIC Hardware Description Language (VHDL). This hardware description language specifies how the internal elements of the FPGA shall be connected in order to perform the required signal processing algorithms [3]. FPGAs are more powerful than DSPs because many signal processing blocks can perform in parallel. On the other hand, their energy consumption at chip level is higher, their hardware structure is more complex, and therefore more expensive, and they are more difficult to program.
- **Application Specific Integrated Circuits (ASICs):** are ICs customized for a particular use, such as signal processing applications, rather than intended for general-purpose use. ASICs are very expensive to design and they are applicable if thousands or millions of devices need to be manufactured. Therefore, the design steps to produce ASICs for a specific application follow a standard product design.
- **General purpose microprocessors:** Most of the communication ICs also contain general purpose microprocessors. They are used as control processors and to implement the functionalities in upper communication layers.

The main characteristics to consider when deciding which IC architecture to use, are the processing requirements of the algorithm, which are measured in Million Instructions per Second (MIPS) and the complexity of the mathematical operations. For the implementation of the proposed learning solutions contemplated in this thesis, i.e. Q-learning, FQL and POMDP, we choose DSPs since they offer enough processing capabilities. Furthermore, there are abundant and different DSPs available in the market, allowing to choose from a wide range of processors depending on the features required by the communication system to be implemented.

8.2 Practical implementation in state of the art processors

It is assumed that femtocell BSs functionalities are implemented in DSPs. The study presented in this section focuses on the memory and computational requirements of Q-learning, FQL and POMDP when considering DSPs.

In the past, DSPs used to be simple, with a small group of specific assembly instructions to perform the mathematical operations. Many applications were programmed in assembly language and each instruction required a fixed number of DSP cycles or clock cycles. As a result, it was fairly easy to predict how many cycles a particular algorithm would take. Nowadays, most of the DSPs are very sophisticated, except for those that are targeted for very low power consumption applications. State of the art DSPs have many execution units in parallel and can perform a variable number of instructions per second, depending on the ordering of the assembly instructions in the algorithm. They also perform pipelining, i.e. partial overlapping of instructions in the time domain. Because of this complexity, in nearly all the cases, assembly language is not a suitable program language for DSPs, instead they are programmed in C (which is more comfortable for the developer) and the compiler takes care of all the optimization and instruction packing and reordering.

More in particular, besides common mathematical operations such as addition, multiplication, division, etc. memory accesses, i.e. memory reading and writing, also consume DSP cycles. The required DSP cycles for a memory access depend on the type of memory being accessed, i.e. internal Random Access Memory (RAM), external RAM, Synchronous Dynamic Random Access Memory (SDRAM), etc. Basic operations also take a different number of cycles depending on where the data are stored, for instance, adding two numbers stored in DSP registers can take one clock cycle, while adding two numbers stored in the DSP memory can take one clock cycle to read each number, one to perform the addition and one to store the result.

Due to all these characteristics, the prediction of the operations or DSP cycles required by a particular algorithm is a very hard task. In what follows we present a theoretical estimation of the operational requirements for the mathematical operations required in the learning approaches, assuming that every basic DSP instruction takes one DSP cycle, except for division, whose requirements are determined from processors datasheet. Before doing so, it is necessary to establish which DSP is more appropriate for the considered case.

We assume that femto BSs functionalities are implemented in a DSP C64x of Texas Instruments. In particular, we assume the use of a TMS320C6416 Fixed-Point DSP, commonly used for communication applications [4]. TMS320C6416 is the DSP with higher performance in the TMS320C6x series. It has the same DSP core that is embedded in Systems on a Chip (SoCs) designed for femtocell and LTE applications, such as the TCI6489 [5].

The computational analysis presented in this section does not take into account the compiler optimizations and the ability of the TMS320C64x DSPs to execute various instructions per clock cycle. Therefore, this analysis provides an upper bound for the computational resources that are needed by the algorithms.

In DSPs, addition, multiplication and comparison operations between two numbers as well

as reading or writing the memory, require one DSP cycle. For the proposed FQL algorithm, more complex operations are required, i.e. exponentials and divisions. Divisions in a TMS320C6x DSP need between 18 and 42 operations [6], so that, we assume the worst case of 42 operations. Also, we consider that exponential functions, are solved through the piecewise linear approximation [7], which results in 11 operations per exponential computation. Table 8.1 summarizes the required computational expenses of the operations performed by the learning algorithms.

Table 8.1: Operations and their computational requirements

Operation	Required Instructions
Memory access	1
Comparison	1
Sum	1
Multiplication	1
Storage	1
Division	42
Exponential	11

In what follows the analysis in terms of memory and computational requirements for a practical implementation of the Q-learning, FQL and POMDP is presented.

8.2.1 Memory requirements of expert knowledge

In learning algorithms, the expert knowledge can be represented in different ways. The selected representation mechanism is directly related with memory requirements of the learning method. We assume femtocells implementing LTE standard with 20 MHz bandwidth channel, which corresponds to 100 RBs [8]. We also consider a latency of 10 ms over the X2 interface [9, 10] for the cases of perfect information algorithms, i.e. Q-learning and FQL.

- Q-learning:** Since in Q-learning there is a Q-value $Q(s, a)$ per each state-action pair, the memory requirements for this kind of systems are given by the size of the state and action spaces. We define the set of possible actions as l power levels that the femtocell can assign to RB r and the set of states as k possible states. In the lookup table, each Q-value is uploaded in 1 B. The considered set of actions is formed by $l = 60$ power levels. The amount of states the agent can perceive depends on the considered state representation. Therefore, the total memory requirement for a femtocell implementing Q-learning following the case study 1, with $k = 4$, is $(k \times l) \cdot 100 = 24$ kB; and for case study 2, with $k = 16$, gives $(k \times l) \cdot 100 = 96$ kB.
- FQL:** In FQL the knowledge is stored in layer 3, each node in this layer has a Q-value per

each action, therefore this occupies $n \times l$ B. So, for the implementation of FQL the total memory requirement for case study 2 is $(n \times l) \cdot 100 = 480$ kB.

- **POMDP:** For the case of partial information, the memory requirements are the same as for the Q-learning approach since the knowledge representation used by POMDP is the same as the one used by Q-learning.

The TMS320C6416 processor only has cache memory, hence the learning algorithms would use the 1280 MB addressable external SDRAM. The remaining memory would be occupied by the program code and other LTE functionalities that may be implemented in the same DSP.

8.2.2 Computational requirements

The computational requirements of the learning processes are given by the operations they have to execute in order to fulfil the representation of the acquired knowledge in a learning iteration.

- **Q-learning:** The computational cost for Q-learning is given by the Q-value estimation through equation (3.8) in Section 3.2, which is summarized in Table 8.2. In particular, the total number of operations required per RB is 246. Since a learning iteration is performed every 10 ms, the average latency of the X2 interface, the total amount operations is then 2.46 MIPS.

Table 8.2: Computational requirement for Q-learning

Operations	Required Instructions
Identification of current and next state in the Q-table	2
Memory access	$2 \times l$
Comparison	$2 \times (l-1)$
Sum	3
Multiplication	2
Storage	1

- **FQL:** For FQL, the computational requirements are given by the processes performed in each node of the four layers of the FIS. Here, some exponential operations are required in layer 1 to compute the membership values, equation (6.9), and some divisions are required to compute the outputs of layer 3 based on equations (6.11) and (6.12), as presented in Section 6.2. The computational operations of each layer are summarized in Table 8.3. The amount of operations per RB learning task is 195053. Therefore, the total operations required are 1950.53 MIPS.

Table 8.3: Computational requirement for FQL

Layer	Nodes	Required Instructions
Layer 1	z	$11 \times z$
Layer 2	n	$3 \times n$
Layer 3	n	$2 \times (1 + 42 + (n - 1)) \times n$
Layer 4	2	$3 \times (n - 1)$
Memory access		$2 \times (l \times n)$
Comparison		$2 \times (l - 1) \times n$
Q-value location		$2 \times n$
Storage		n

- POMDP:** The computational cost in POMDP algorithm is determined by the computation of the Q-value and by the operations required in the spatial interpolation process. We consider two possible belief states, so that, two Q-values have to be updated according to weights b in each learning iteration. Similarly to what described in Table 8.2 for Q-learning, 2428 operations are required for executing the POMDP according to equation (7.2) given in Section 7.3. The computational complexity of Kriging spatial interpolation to estimate the macrocell capacity, \tilde{C}_r^m , depends on many aspects such as the variogram fitting, the number of femtocells providing measurements, etc. We estimate a number of operations around 10000. Since this information is local, the learning process is not subject to any time constraint related to interfaces among entities and can be performed e.g., every 1 ms, which is the LTE scheduling period. In this case, the total amount of operations is around 1242.8 MIPS.

Table 8.4: Computational requirement for POMDP

Operations	Required Instructions
Q-value computation	2428
\tilde{C}_r^m computation	~ 10000

The TMS320C6416 processor has a maximum capacity of 8000 MIPS [4], so that Q-learning, FQL and POMDP algorithms can be implemented in the IC together with other LTE algorithms. It is worth mentioning that both, FQL and POMDP are computationally more expensive than Q-learning, which relies on feedback received from the macro network. For the case of FQL this is the cost of having faster learning processes and continues state and action representation. For the case of POMDP, this is the price to pay for the non availability of the X2' interface. The advantage of POMDP is, on the other hand, the 3GPP standard compliance.

8.2.3 Comparison of neural networks and lookup table representation mechanisms for Q-learning

Q-learning is based on quantifying, by means of the Q-function, the quality of an action in a certain state. The learned Q-values have to be stored in a representation mechanism. In this thesis we have been considering the lookup table as representation mechanism however, when the number of state-action pairs is large, the lookup table becomes unfeasible, so that there is the need for a more compact representation mechanism. A proposal has been described in Chapter 6, based on combining fuzzy logic and Q-learning. Another option is to use neural networks to estimate the Q-values. In this section we focus on the neural network option and we compare it, in terms of memory and computational requirements, with the lookup table representation mechanism for case study 2.

The neural network representation structures are compact and scalable [11]. We consider a neural network with three layers, input, hidden, and output layers, as it is shown in Figure 8.1. The N_i nodes at input layer represent the state-action space, therefore, there is an input node per each state indicator possible value, and per each action, resulting in $N_i = 2 + 2 + 4 + l = 68$. The amount of nodes in the hidden layer, N_h , is approximately the square root of the number of nodes in the input layer [11], then $N_h = 8$. The output layer consists of N_o nodes computing the approximated Q-value for a certain state-action pair, thus $N_o = 1$. All nodes in one layer are connected with all nodes in the next layer, so that the neural network is fully interconnected. Each connection between a couple of nodes is associated with a weight, which represents the synaptic strength of the connection, so that it determines the impact that each node has on the decision making process. The ability of the neural network to correctly approximate the Q-values, lies on the proper selection of the weights, which have to be trained by means of a learning process. For this purpose, it is considered the error back propagation [11].

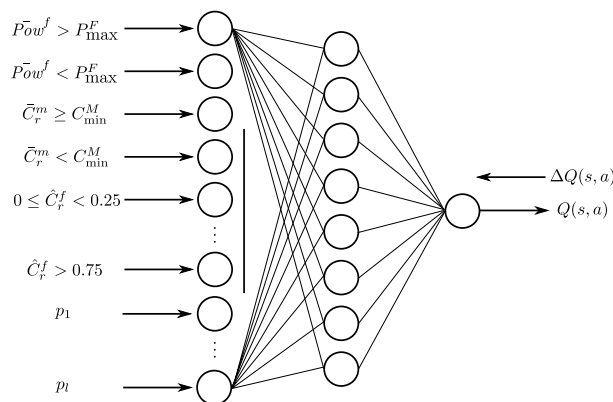


Figure 8.1: Neural network scheme.

The nodes in the input layer are passive. Instead, nodes in hidden and output layers are active. This means that they process data through their activation function. We propose the use

of a linear activation function for the output node, to produce Q-values of arbitrary magnitude. In turn, for hidden nodes, we consider a sigmoid function, $\text{sig}(x)$, defined as follows:

$$\text{sig}(x) = \frac{1}{1 + e^{-x}} \quad (8.1)$$

which requires 54 operations, considering the operation requirements summarized in Table 8.1. The error value propagated in the back direction is $\Delta Q(s, a)$, defined in equation (3.9).

Computational requirements in each layer of the neural network are summarized in Table 8.5.

Table 8.5: Computational requirement for neural network

Layer	Nodes	Required Instructions
Layer 1	N_i	N_i
Layer 2	N_h	$(54 + 2 \times N_i) \times N_h$
Layer 3	N_o	$2 + 2 \times N_h$
Backpropagation		$N_h(1 + 2 \times N_i) + 71 \times (N_i \times N_h + N_h \times N_o)$

The computation of each Q-value in case of neural network, requires approximately $N_i + (54 + 2 \times N_i) \times N_h + 2 + 2 \times N_h = 1606$ operations. In addition, the neural network has to account for the cost of the back propagation to update the weights. In the back propagation method signals are propagated from output to inputs one after the other and the connection weight used are equal to the ones used during the computation of the output value. Only the direction of data flow is changed. Then, when the error signal for each neuron is computed, the weights coefficients of each neuron input node may be modified. The new weight of the connections is the derivative of the neuron activation function. Derivative of activation function used in the hidden nodes requires 71 operation according to Table 8.1.

Back propagation results in $N_h(1 + 2 \times N_i) + 71 \times N_i + N_h = 5992$ operations, for every Q-value update. This amount of operations accounts for two processes, the first one correspond to the propagation of the error signal $\Delta Q(s, a)$ back to all neurons, $N_h(1 + 2 \times N_i)$, and the second process corresponds to the modification of the weights coefficients of each neuron connection, $71 \times N_i + N_h$. Then, the approximate total number of operations to update a Q-value when using neural networks is 7598 operations. We conclude that the advantages of the neural network representation mechanism are the limited memory requirements and scalability, whereas the main drawback is the increased number of computational operations, with respect to those needed by the lookup table representation mechanism. In any case, the total amount of operations required by the neural network is about 75.98 MIPS, therefore, it can be implemented in the TMS320C6416 processor.

Table 8.6 presents a comparison regarding the memory requirements and computational complexity of both the lookup table and neural network representation methods, considering

that a learning iteration is performed every 10 ms. In particular, the memory required by the neural network to store the knowledge is determined by the number of weights and consequently, by its number of connections $N_i \times N_h + N_h \times N_o = 552$. The memory requirements of the lookup table are 40 kB higher than those of the neural network.

Table 8.6: Comparison between lookup table and neural network representation mechanisms

Method	Memory (kB)	MIPS
Lookup table	96	2.46
Neural network	55.2	75.98

8.3 Conclusions

This chapter presents a study regarding the requirements of the proposed Q-learning, FQL and POMDP approaches, when they are embedded in ICs. The analysis is performed assuming that femtocell BSs are implemented in TMS320C6416 processors. Evaluation is performed in terms of memory and computational requirements and results show that state of the art processors can support the proposed learning approaches.

Bibliography

- [1] J. Eyre and J. Bier, "The evolution of DSP processors," *IEEE Signal Processing Magazine*, vol. 17, pp. 43–51, March 2000.
- [2] S. W. Smith, *Digital Signal Processing: A Practical Guide for Engineers and Scientists*. Newnes, 2002.
- [3] The "FPGA place-and-route challenge". [Online]. Available: <http://www.eecg.toronto.edu/~vaughn/challenge/challenge.html>
- [4] Tms320c6416 fixed-point digital signal processor. [Online]. Available: <http://focus.ti.com/lit/ds/symlink/tms320c6416t.pdf>
- [5] Ref. tci6489. [Online]. Available: <http://www.ti.com/litv/pdf/sprt522a>
- [6] Y.-T. Cheng, "TMS320C6000 integer division," Texas Instruments Application Report, Tech. Rep. SPRA707, October 2000.
- [7] M. Bajger and A. Omondi, "Implementations of square-root and exponential functions for large FPGAs," in *Asia-Pacific Computer Systems Architecture Conference*, 2006, pp. 6–23.
- [8] S. Sesia, I. Toufik, and M. Baker, *LTE, The UMTS Long Term Evolution: From Theory to Practice*. Wiley Publishing, 2009.
- [9] 3GPP, "X2 Application Protocol (X2AP) (Release 8)," 3GPP TS 36.423 V8.2.0 (2008-06), June 2008.
- [10] ———, "X2 General Aspects and Principles (Release 8)," 3GPP TS 36.420 V8.0.0 (2007-12), Dec. 2007.
- [11] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New York: Macmillan, 1994.

Chapter 9

Conclusions and Future Work

This thesis has proposed modeling femtocells as decentralized agents with learning capabilities. The objective is to control the aggregated interference that multiple femtocells working simultaneously can generate at macrocell users. The use of a decentralized model allows us to divide the computational load between the multiple femto nodes, introducing scalability and robustness to the network. The introduction of learning capabilities in the femtocells enables a highly adaptive and autonomous behavior, which perfectly fits with the paradigm of heterogeneous networks and current self-organized trends in wireless communications. The proposed learning approach is based on the RL theoretical framework, which allows agents to construct their knowledge based on online interactions, without requiring a transition model between states. The technical chapters of this thesis have hence been focused on the design of learning approaches with complete and partial information, discrete and continuous state and action spaces representation and cooperative capabilities.

9.1 Summary of results

Currently, interference management in femtocell systems is one of the key open issues that hampers the massive deployment of this technology. That is the reason why the 3GPP standardization body has proposed dense-deployed scenarios models to study the interference in HeNB/eNB heterogeneous systems. Based on these recommendations, two scenario models, the single-cell and the multicell scenarios, presented in Chapter 2, have been designed to test the proposed solutions. Femtocells are installed by end users, then, a centralized RRM is not feasible due to scalability issues and the macrocell lack of information regarding its underlying femtocells. For that reason, recently, attention to solve the RRM in femtocell systems has been focused on autonomous and self-organized solutions. In this thesis, self-organization of femtocells is accomplished following model-free TD learning methods to control the interference that multiple femtocells simultaneously transmitting can cause at macrocell users.

Among the multiple TD learning methods, Chapter 3 presents a comparison between an on-policy method, called Sarsa, and an off-policy method, the Q-learning approach. We compared both algorithms in terms of cost value, system performance and their capacity to fulfil the established constraints, as a function of the learning iterations, in order to select the most suitable one, given the proposed interference problem. Based on the obtained results, we selected the off-policy method, i.e. the Q-learning approach. This decision has been driven by the better performance given by the Q-learning algorithm since the very beginning of the learning process, which allows guaranteeing less damage to macrousers due to the presence of the femtocells. This chapter also deals with the design details of the learning algorithm, resulting in the selection of a discount factor, $\gamma = 0.9$, a learning rate, $\alpha = 0.5$, and the ε -greedy two-steps action selection policy.

Chapter 4 shows that macrocell user capacity and total transmission power constraints can be fulfilled in both single-cell and multicell scenarios with the proposed learning approach. To this end, some information regarding the macrocell user capacity must be conveyed from the macrocell to its underlying femtocells. This information is proposed to be sent through a X2' interface between macrocells and femtocells, as contemplated in the functional architecture, presented in Section 2.4 and also supported by the BeFEMTO consortium [1]. This chapter presents a solution for the inclusion of the learning approach required information in 3GPP systems related to the interference perceived by the macrocell users and the macrousers scheduling in the future. Relying on this information, femtocells can then perform a transfer learning among internal tasks, carrying out a macrouser oriented learning process, which will guarantee to cause less damage to the macrouser performance, given that state transitions would be smoother while the macrouser session remains open. Hence, the interference at macrousers can be better controlled, mostly for the case when agents are starting their learning process.

Results discussed in Chapters 3 and 4 show the drawbacks of decentralized online learning algorithms: the length of the training process, oscillatory behaviors, just to name a few. As a solution to this problem, the paradigm of docition has been introduced in Chapter 5. Following this cooperative technique, it has been shown that femtocells can learn the interference control policy already acquired by neighboring femtocells, which have been active during a longer time and have already learnt proper decision policies. This translates in less interference at macrousers and less energy consumption due to the decrease of required learning iterations. In any case, further research is needed in this interesting field since there are multiple open issues that still require to be fulfilled. For instance, the quantification of the degree of intelligence of one agent, what information to teach, when to perform the docition, etc.

Another aspect which has been considered in this thesis is the use of a discrete state and action representation. The correct selection of these sets highly influences the learning process performance and speed of convergence. Also, when states and actions need to have a detailed

representation, the use of simple knowledge representation (i.e. lookup tables), is more complicated due to the memory requirements and the increment in the learning search space. Tackling these aspects, Chapter 6 combines Fuzzy logic with RL, resulting in the FQL approach, which allows to operate with detailed and even continuous representations of state and action spaces, reducing the algorithm complexity. In addition, since previous expert knowledge can be naturally embedded in the fuzzy rules, the learning period can be significantly reduced. Results show that the proposed scheme outperforms the Q-learning and a heuristic approach introduced by 3GPP and that, with a simple expert initialization of the Q-values, important gains can be achieved in terms of precision and speed of convergence.

For the state representation in the learning processes defined in Chapters 4, 5 and 6, existence of a X2' interface between femtocells and macrocells is assumed. The standardization of the mentioned interface has not been contemplated in last 3GPP releases, so that this assumption is not yet standard compliant. In order to provide a solution which could be implemented in current 3GPP systems, Chapter 7 deals with the problem of making decisions without feedback from the macrocells to the femtocells about the interference perceived by the macrouser. Then, the interference management problem is modeled by means of the theory of POMDP, which works by constructing beliefs about the state of the environment. The belief set is built through the networked femtocells SINR measurements and by spatial interpolating them through the ordinary Kriging technique. Results show that the POMDP algorithm is able to learn a sub-optimal solution, which guarantees to maintain the macrocell system performance above a desired threshold, allowing a completely autonomous femtocell system deployment and avoiding the introduction of signaling overhead.

Finally, Chapter 8 presents a study about the memory and computational requirements of proposed solutions and concludes that state of the art processors can support the introduced learning approaches.

9.2 Future work

The work presented in this thesis left multiple investigation lines open for future work. In what follows we summarize the most important ones.

- The solutions presented in this thesis have been stated from a decentralized point of view, given the important advantages of this form of modeling. Nevertheless, it lacks a comparison with a centralized solution in order to have a quantitative measure of the gains introduced by the decentralization in the RRM procedures. Centralized solutions commonly have better performance given their complete knowledge about the system. However, they are difficult to implement, require large signaling and can incur in important

delays. On the other hand, decentralized systems are fast and require little signaling, but can result in sub-optimal behaviors, given their partial knowledge about the system and usually show a slow convergence time. Therefore, comparisons have to be measured as tradeoffs. In femtocell systems, a centralized solution could be implemented at the LFGW, presented in Section 2.4 and introduced in the BeFEMTO deliverable D 2.2 [2].

The problem we have considered in this thesis can be formulated as a centralized problem as follows:

$$\begin{aligned} & \underset{\{p_r^{f,F}\}}{\text{maximize}} && \sum_{r=1}^R \frac{BW}{R} \log_2 \left(1 + SINR_r^f \right) \\ & \text{s.t.} && Pow^f \leq P_{\max}^F \\ & && \frac{BW}{R} \log_2 \left(1 + SINR_r^m \right) \geq C_{\min}^M, \quad r = 1, \dots, R \end{aligned}$$

which is a non-convex problem. An optimal solution to this problem, to the best of the authors knowledge, would require an exhaustive search over the feasible set of transmit powers, which entails high complexity. On the other hand, to solve this problem through learning techniques is not feasible due to scalability issues given by the potential amount of femtocells. Further work to find a feasible solution to this problem remains as future work. We then propose, as a possible starting point, to investigate the possibility of performing the RRM at the LFGW following the solution given in [3] for large scale systems, which combines model predictive control, multiagent systems and RL.

- Another important aspect, when dealing with decentralized systems, is the convergence to an equilibrium point. Some work has been developed in the context of a collaboration with Dr. Eitan Altam and has been recently submitted to the *6th International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS 2012)*. It has not been included in this thesis since further results still have to be obtained. We can summarize, however, the main idea of the contribution.

We propose to model the interference problem as a non-cooperative n -person game, where players are the multiple femtocells facing an optimization problem with restrictions. Depending on the relationship between the set of strategies and the utility function of each agent and other players' strategies, n -person games can be classified in three models: orthogonal, coupled and generalized. We focus on coupled constraint games since our players aim to maximize their individual utility, subject to a common global constraint, the total interference at macrocell users. We follow the Generalized Nash Equilibrium Problem (GNEP) natural extension of the standard Nash equilibrium concept proposed by Nash for players sharing common resources or limitations [4]. GNEP is currently widely used in multiple and different fields, since it perfectly describes competition situations in dis-

tributed decision making systems. Some routing games with capacity constraints have this structure, where the set S_i of available strategies to player i are those for which the sum of flows in each link cannot exceed the link's capacity. It is an extension of the constrained satisfaction games recently introduced by Perlaza et al [5]. GNEP solution in n -person non-cooperative games with common constraints gives infinitely many solutions. In order to select a unique equilibrium point among the solutions of the game, we propose to follow the approach proposed by Rosen in [6], which is known as normalized equilibrium, since it has suitable properties for decentralized scenarios.

We consider N non-cooperative players, i.e. the N femtocells in the system, where player $f = 1, \dots, N$ controls the variable $p_r^f \in \mathbb{R}^{n_f}$. Let p_r be the N -dimensional vector of all players strategies with dimension $n = \sum_{f=1}^N n_f$ and p_r^{-f} the $N - 1$ vector formed by all players' strategies but f . Let $S \subset \mathbb{R}^n$ be a compact convex set and R is a convex compact set of constraints.

Each player in the game has a utility function, $U^f : \mathbb{R}^{n_f} \rightarrow \mathbb{R}$. An equilibrium in this game consists of a vector $p_r^* \in R$ such that for each player f , $U^f(p_r^f)$ attains its maximum over all p_r^f for which $(p_r^f, p_r^*[-f]) \in R$. Here, (p_r^f, p_r^{-f*}) is the policy obtained from p_r^* by the strategies of all players, except for that player f who uses p_r^f instead of p_r^* . The maximization problem is then given by:

$$\text{maximize } U^f(p_r^f, p_r^{-f}) \quad \text{subject to } p_r^f \in \mathbf{p}^f(p_r^{-f}) \quad (9.1)$$

We consider the setting in which the achievable utility of all femtocells is given by the convex region ν defined by the set of constraints:

$$\begin{aligned} \sum_{f=1}^N \hat{h}_r^f p_r^f &\leq I_{Th} \\ 0 &\leq p_r^f \leq P_{\max}^f \end{aligned}$$

where I_{Th} is an interference constraint at macrouser u^m . Every player f maximizes its own utility U^f , which is assumed to be a strictly concave increasing function of its strategy vector \mathbf{p}^f . We assume that the utility of a player depends only on its own strategy. The interference constraint I_{Th} at macrousers is a common constraint that all player strategies are required to satisfy. Therefore, this places this game in the category of *coupled constraints* defined by Rosen [6]. In games with coupled constraints the choice of strategies of a player depends on the strategies chosen by other players.

- In this thesis it has been proven that the correct selection of some tuneable parameters in the learning algorithm i.e. learning rate and learning period, as well as a smart Q-table initialization, can bring important gains in the learning process in terms of speed of convergence and accuracy, from the beginning of the learning process. For the learning algorithm tuneable parameters, it could be interesting to test state-action pair driven learning rate

(α) as presented in [7]. State-action pair driven learning rate consists in diminish every time the action is selected the α in the knowledge update of ΔQ . Some work regarding this issue can be found in [8], where the *asynchronous Q-learning* is proposed. The duration of the learning period is related to one of the main challenges in RL approaches, that is the tradeoff between exploration and exploitation. Therefore, the selection of the learning period is a key factor in learning processes. In our work, we assumed a fixed learning period, after which the exploration is eliminated. The introduction of state driven exploration could guarantee a better adaptation when large state problems are formulated. We plan to improve our algorithm at this point following the work presented in [9].

On the other hand, for the lookup table initialization, in Section 4.3.1 we presented a very simple initialization method, which we called Init Q-learning, and we showed the previously mentioned advantages. Furthermore, in this framework, some cooperative work has been done with Meryem Simsek [10], from Universität Duisburg-Essen. In this paper we have proposed that every time a new state is visited, the corresponding Q-values of the row in the Q-table representing the given state are initialized as a function of received cost after the execution of the selected action. We consider that further research on this aspect is required since smart initialization procedures of the Q-table can bring notable improvements regarding the main drawbacks in distributed learning algorithms.

- We propose to introduce the given solutions in a LTE simulator in order to test the presented self-organization techniques in a more realistic environment. To do this we are considering the ns-3 simulator developed as part of the LTE-EPC Network Simulator (LENA) project, realized by Ubiquisys and CTTC, which is completely 3GPP standard compliant. The following step would be to introduce the proposed algorithms in a networking test bed such as the one developed at CTTC in the framework of BeFEMTO project.
- Another interesting point to be fulfilled is the combination of FQL and POMDP techniques, which would allow to achieve a completely autonomous learning algorithm based on POMDP, with better performances in terms of speed of convergence and accuracy, introduced by FIS. To this end, we propose to follow the solution presented in [11], where the authors introduce the fuzzy multiagent POMDP algorithm, which successfully combines FQL and POMDP in such a way that the FIS based RL controller approximates optimal policies for multiagent POMDPs, modeled as a sequence of Bayesian games.

Bibliography

- [1] “D2.1: Description of baseline reference systems, use cases, requirements, evaluation and impact on business model,” EU FP7-ICT BeFEMTO project, Dec. 2010.
- [2] “D2.2: The BeFEMTO system architecture,” EU FP7-ICT BeFEMTO project, Dec. 2011.
- [3] V. Javalera, B. Morcego, and V. Puig, “Distributed MPC for large scale systems using agent-based reinforcement learning,” in *12th IFAC Symposium on Large Scale Systems: Theory and Applications (2010)*, 2010.
- [4] F. Facchinei and C. Kanzow, “Generalized Nash equilibrium problems,” *Annals OR*, vol. 175, no. 1, pp. 177–211, 2010.
- [5] S. M. Perlaza, H. Tembine, S. Lasaulce, and M. Debbah, “Satisfaction equilibrium: A general framework for QoS provisioning in self-configuring networks,” in *GLOBECOM*, 2010.
- [6] J. Rosen, “Existence and uniqueness of equilibrium points for concave N-person games,” *Econometrica*, vol. 33, no. 3, pp. 520–534, 1965.
- [7] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. The MIT Press, 1998.
- [8] E. Even-dar and Y. Mansour, “Learning rates for Q-learning,” in *Journal of Machine Learning Research*, vol. 5, 2003, pp. 1–25.
- [9] Y. Achbany, F. Fouss, L. Yen, A. Pirotte, and M. Saerens, “Managing the exploration/exploitation trade-off in reinforcement learning,” Information System Unit, Universite catholique de louvain, Belgium, Tech. Report, 2005, available online (20 pages). [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.137.9971>
- [10] M. Simsek, A. Czylik, A. Galindo-Serrano, and L. Giupponi, “Improved decentralized Q-learning algorithm for interference reduction in LTE-femtocells,” in *Conference on Wireless Advanced, 20-22 June 2011, London, UK*, 2011.
- [11] R. Sharma and M. T. J. Spaan, “A Bayesian game based adaptive fuzzy controller for multi-agent POMDPs,” in *FUZZ-IEEE 2010, IEEE International Conference on Fuzzy Systems, Barcelona, Spain, 18-23 July, 2010, Proceedings*, 2010.

Appendix A

Notation

α	Learning rate.
β	SPC algorithm parameter expressed in dB.
γ	Discount factor.
$\gamma(h_i)$	Empirical variogram.
$\gamma(x_i, x_j)$	Experimental variogram data of the values between locations x_i and x_j .
Δf	Subcarrier width.
ε	Probability of selecting actions randomly.
η	SPC linear scalar.
θ	Angle from user to the antenna.
θ_{3dB}	Angle from central lobe at which the gain reduces to half the maximum value.
λ_j	Weight given to the observed value in a given location.
π	Followed policy.
π^*	Optimal policy.
ρ_0^2	Variogram model sill.
ρ^h	Variance of the bell shape function associated to FIS Layer 1, node h .
σ^2	Noise power.
τ	Softmax action selection policy temperature.
a	Current state selected action.
a'	Next state selected action.
a_0	Range, distance where the fitted variogram model becomes constant.
\mathcal{A}	Set of actions.
Az	Azimuth antenna pattern.
Az_m	Maximum possible attenuation due to sectorization.
b	Belief state.
BW	Total bandwidth.

c	Cost value.
c_0	Nugget, value at which the variogram model intercepts the y-axis.
C	Cost function.
$C^{f,F}$	Capacity of femtocell f .
$C^{m,M}$	Capacity of macrocell m .
C_{\min}^M	Macrocell minimum capacity per RB.
C_{Total}	Total system capacity.
d	Total distance between BS and UE.
d_{indoor}	Total indoor distance BS and UE.
D	Inter-site distance.
e^h	Mean value of the bell shape function associated to FIS Layer 1, node h .
E_c	Reference signal received power per Resource Element (RE) at the femto node.
f	Femtocell, agent.
fl	Femtocell block number of floors.
F	Blocks of apartments.
$h_{ff,r}^{FF}$	Link gain between transmitting femtocell f and its UE u^f .
$h_{fm,r}^{FM}$	Link gain between transmitting femtocell f and UE u^m of macrocell m .
$h_{mf,r}^{MF}$	Link gain between macrocell m and UE u^f in femtocell f .
$h_{mm,r}^{MM}$	Link gain between transmitting macrocell m and its UE u^m .
\bar{I}_r^m	State indicator for femtocell system aggregated interference.
k	State set size.
K	Constant value in cost function.
l	Number of actions, number of femtocell transmission power levels.
L	Linguistic variables.
m	Macrocell.
\mathcal{M}	Set of macrocells.
M	Number of macrocells.
\mathcal{N}	Set of femtocells, agents.
N	Number of femtocells.
N_o	Thermal noise.
N_{sc}	Number of sub-carriers.
$\mathbf{p}^{f,F}$	Transmission power vector of femtocell f .
$\mathbf{p}^{m,M}$	Transmission power vector of macrocell m .
p_{oc}	Femtocell occupation ratio.
$p_r^{f,F}$	Downlink transmission power of macrocell m in RB r .
$p_r^{m,M}$	Downlink transmission power of femtocell f in RB r .
$P_{s,v}$	State transition probability.

P_{\max}^F	Femtocell maximum total transmission power.
P_{\min}^F	Femtocell minimum total transmission power.
P_{\max}^M	Macrocell maximum total transmission power.
Q	State-action value function.
r	Resource block.
R	Number of RBs.
s	Observed state.
S	Set of states.
$SINR_r^f$	SINR at UE u^f allocated in RB r of femtocell f .
$SINR_r^m$	SINR at UE u^m allocated in RB r of femtocell m .
$SINR_{Th}^M$	macrouser SINR threshold.
t	Time step.
$T(x)$	Term set.
u^f	Femtocell UE.
u^m	Macrocell UE.
U_f	Femtocell associated UEs.
U_m	Macrocell associated UEs.
v	Next state.
V	State value function.
w_i	Single true value.
w_p	Number of walls separating apartments.
WP_{in}	Indoor wall penetration losses.
WP_{out}	Outdoor wall penetration losses.

Appendix B

Acronyms and Definitions

2G	Second-Generation
3G	Third-Generation
4G	Fourth-Generation
3GPP	3rd Generation Partnership Project
AI	Artificial Intelligence
ANR	Automatic Neighbor Relation
AMC	Adaptive Modulation and Coding
AS	Access Stratum
ASIC	Application Specific Integrated Circuit
BS	Base Station
CAPEX	Capital Expenditures
CCDF	Complementary Cumulative Distribution Function
CDF	Cumulative Distribution Function
CSG	Closed Subscriber Group
DeNB	Donor evolved NodeB
DL-HII	Downlink High Interference Indicator
DM	Domain Management
DSCP	Differentiated Services Code Point

DSL	Digital Subscriber Line
DSP	Digital Signal Processor
EM	Element Management
eNB	Evolved NodeB
EPC	Evolved Packet Core
EPS	Evolved Packet System
ERP	Effective Radiated Power
ETSI BRAN	European Telecommunications Standards Institute Broadband Radio Access Networks
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
FDD	Frequency Division Duplex
FIS	Fuzzy Inference System
FPGA	Field Programmable Gate Array
FQL	Fuzzy Q-learning
GNEP	Generalized Nash Equilibrium Problem
HARQ	Hybrid Automatic Repeat Request
HeNB	Home eNodeB
HeNB GW	HeNB Gateway
HMS	HeNB Management System
HSPA	High Speed Packet Access
HSDPA	High Speed Downlink packet Access
HSS	Home Subscriber Server
IC	Integrated Circuit
IMS	IP Multimedia Subsystem
IP	Internet Protocol
ITW	Iterative Water-Filling

LAN	Local Area Network
LFGW	Local Femtocell GateWay
LIPA	Local IP Access
LOS	Line of Sight
LTE	Long Term Evolution
LTE-A	LTE-Advanced
MAB	Multi-Armed Bandit
MDP	Markov Decision Process
MIB	Master Information Block
MIMO	Multiple-Input Multiple-Output
MIPS	Million Instructions per Second
ML	Machine Learning
MME	Mobility Management Entity
MCS	Modulation and Coding Scheme
NAS	Non-Access Stratum
NGMN	Next Generation Mobile Networks
NM	Network Management
NSCP	Non-stationary Converging Policies
OA&M	Operation, Administration and Maintenance
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OPEX	Operational Expenditure
PBL	Problem Based Learning
PBCH	Physical Broadcast Channel
PCI	Physical Cell Identity
PCRF	Policy and Charging Rule Function

PDSCH	Physical Downlink Shared Channel
P-GW	Packet Data Network Gateway
PL	Path Loss
POMDP	Partially Observable Markov Decision Process
QoS	Quality of Service
RAM	Random Access Memory
RAT	Radio Access Technology
RB	Resource Block
RE	Resource Element
RF	Radio Frequency
RL	Reinforcement Learning
RNC	Radio Network Controller
RNTP	Relative Narrowband Transmit Power
RRC	Radio Resource Control
RRM	Radio Resource Management
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
SAE	System Architecture Evolution
SB	Scheduling Block
SDRAM	Synchronous Dynamic Random Access Memory
SE	State Estimator
S-GW	Serving Gateway
SIB	System Information Block
SINR	Signal to Interference Noise Ratio
SIPTO	Selected IP Traffic Offload
SoCs	Systems on a Chip

SON	Self-organized Networks
SPC	Smart Power Control
SRG	Signals Research Group
TD	Time Difference
TDD	Time Division Duplex
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
URBA	User Resource Block Allocation
UTRAN	Universal Terrestrial Radio Access Network
VHDL	VHSIC Hardware Description Language
WCDMA	Wideband Code Division Multiple Access
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network

