Radio Resource Management Strategies based on Hopfield Neural Networks

Núria Garcia Garcia

TESI DOCTORAL UPF / 2009

DIRECTORS DE LA TESI

Dr. Ramon Agustí Comes

Dr. Jordi Pérez Romero

TUTOR

Dr. Miquel Oliver Riera

Departament de Tecnologies de la Informació i de les Comunicacions

UNIVERSITAT POMPEU FABRA

Al Roger, al Joan i a la Núria.

Agraïments

En primer lloc vull fer palès el meu agraïment al Joan Vinyes per haver-me ofert l'oportunitat de fer aquesta tesi.

Vull agrair als meus directors de tesi, Jordi Pérez i Ramon Agustí, el seu mestratge i la paciència que han tingut per a seguir una tasca a la qual no sempre he pogut dedicar tot el temps que requeria.

El meu agraïment també al Miquel Oliver, com a tutor i cap del grup de recerca NETS, per haver-me fet costat en tot allò que he anat necessitant per a poder arribar a escriure aquestes línies. I estenc aquest agraïment a la resta de companys de grup perquè un bon entorn facilita molt la feina.

Agraeixo al Joan Olmos la rutina que em va proporcionar pel càlcul de la resposta freqüencial de canals variant en el temps i al Jordi el seu valuòs ajut en l'estructuració del simulador pel desplegament d'una xarxa mòbil.

Agraeixo al Quim i al Joan Carles els seu enriquidors comentaris.

Un agraïment molt especial al Ramon perquè sempre he tingut el seu suport.

Summary

This doctoral thesis addresses the Radio Resource Management in mobile communications systems: 3G, B3G (Beyond 3G) and 4G. As it is widely known radio resources are scarce, particularly in the cellular mobile communication systems, and therefore it is quite necessary to provide an efficient management of such resources. From a practical point of view, while mobile communications systems are based on the use of standardized radio access technologies, this is not the case of the radio resource management algorithms, which are not standardized, so that it is always possible to investigate new suitable algorithms that improve previous achievements in the framework of the considered technology. It is in this direction toward where we point to with the realization of this thesis by introducing new Radio Resource Management strategies in the framework of the multiple access techniques WCDMA, in the heterogeneous TDMA and WCDMA B3G systems and with the OFDMA multiple access technique for future 4G systems.

This thesis, after identifying some of the more usual elements in Radio Resource Management, introduces a basic description of them as well as the most relevant aspects of the multiple access systems WCDMA and OFDMA. In this sense, the mechanisms for their operation are detailed that afterwards will be used in the definition and specification of the proposed Radio Resource Management algorithms. After a brief introduction to the neural networks, we follow with a more detailed presentation of the particular case of Hopfield Neural Networks, which constitute the thread of the work of this thesis. In particular the equations that characterize these networks as dynamic systems are described and their stability conditions through the Lyapunov stability theorems and the definition of the Energy function are introduced. In this respect, some aspects relating to the robustness of the algorithm, and in particular to its convergence have been also considered.

From the conjunction between the particularities of WCDMA, TDMA and OFDMA access systems and Hopfield Neural Networks a series of novel algorithms were developed and described below briefly.

1. Uplink WCDMA Admission Control through an optimized management of different transmission bit rates assigned to users who share the access and in accordance with different Quality of Service profiles. We have assumed a maximum load restriction of the network, a fair spectrum sharing among users and a maximum power available in the mobile terminals. It is

also assumed a real time service with variable bit rates. The blocking probability is used to display the benefits of the algorithm

2- Management of the bit rates of already admitted users in the uplink of a WCDMA system, in order to guarantee a satisfaction probability, defined as the probability of exceeding a certain bit rate value. It is also assumed a real time service with variables bit rates and the same restrictions as those considered in the Admission Control case.

3- Management of the bit rates of already admitted users in the downlink of a WCDMA system, in order to ensure a maximum delay in delivery of packets. We have assumed fair spectrum sharing restrictions and a maximum power available at the base station. It is assumed an interactive service based on a www traffic model. A reference algorithm is also introduced for comparative purposes. The dropping probability is the parameter used to assess the algorithm behaviour.

4- Joint management of real time and interactive services in the downlink of a WCDMA system. It incorporates part of the algorithms set forth above with the same traffic models and the same restrictions. We have used in this case the satisfaction probability and dropping probability to capture the "throughput" and the packet delay, respectively.

5- Joint Radio Resource Management algorithm for a B3G scenario where a user can be served by more than one access technology. In particular WCDMA and TDMA have been considered. The services are assumed to be interactive and the dropping rate has been used to show the benefits of the algorithm. It is also worth noticing the commitment between the delay introduced to execute a vertical handover and the final dropping rate obtained

6- Algorithms for the Management of the bit rate of users already admitted in a downlink OFDMA system, in order to ensure a maximum delay in the delivery of packets. There have been presented two variants of the algorithm, and it is shown that the dropping probability performs better than that offered by other so far published heuristic algorithms

The thesis also points toward future research lines not explored so far that seek to exploit the basis of the methodology developed here. Initially they consist in multicellular centralized scenarios to proceed later with distributed ones, in particular for the OFDMA systems, which are the basis of the 4G access systems. In this sense it would also include a refinement of the developed methodology that shortens the simulation time and allows for a better precision in the obtained results.

A list of the publications carried out in connection with the results of this thesis has also been included. It consists of international conferences as well as a publication in an indexed magazine.

Resum

Aquesta tesi doctoral s'encara en la Gestió de Recursos de Ràdio en sistemes de comunicacions mòbils: 3G, B3G (Beyond 3G) i 4G. Com és ben conegut, els recursos de ràdio són escassos, especialment en els sistemes de comunicacions mòbils cel·lulars, i per això és molt necessari proporcionar una gestió eficaç de tals recursos. Des d'un punt de vista pràctic, mentre els sistemes de comunicacions mòbils es basen en l'ús de tecnologies d'accés de ràdio estandarditzades, aquest no és el cas dels algoritmes de gestió de recursos radio, que no estan estandarditzats; per tant, sempre és possible investigar-ne de nous que millorin consecucions prèvies en el marc de la tecnologia considerada. És cap aquesta línia on hem adreçat la realització d'aquesta tesi, introduint noves estratègies de Gestió de Recursos de Ràdio en el marc de les tècniques d'accés múltiple WCDMA, de sistemes heterogenis TDMA i WCDMA BG i amb la tècnica d'accés múltiple d'OFDMA per a futurs sistemes 4G.

Aquesta tesi, després d'identificar alguns dels elements més usuals en la Gestió de Recursos Radio, introdueix una descripció bàsica d'aquests així com d'els aspectes més pertinents dels sistemes d'accés múltiple WCDMA i OFDMA. En aquest sentit, s'hi detallen aspectes claus d'aquests sistemes que posteriorment s'utilitzaran en la definició i especificació dels algoritmes de Gestió de Recursos Radio proposats. Després d'una introducció breu a les xarxes neuronals, es fa una presentació més acurada del cas particular de xarxes neuronals de Hopfield, la qual cosa constitueix el fil conductor del treball d'aquesta tesi. En particular es descriuen les equacions que caracteritzen aquestes xarxes com a sistemes dinàmics i s'analitzen les seves condicions d'estabilitat a través dels Teoremes d'estabilitat de Lyapunov i la definició del model de funció Lyapunov (funció energia) que s'introdueix. Tanmateix, es fa referència a alguns aspectes relatius a la robustesa del l'algoritme i, en particular a la seva convergència.

A partir de la conjunció entre les particularitats dels sistemes d'accés WCDMA, TDMA i OFDMA i les xarxes neuronals de Hopfield s'han desenvolupat un sèrie de nous algoritmes que tot seguit s'enumeren i descriuen breument.

1. Control d'admissió en uplink WCDMA a través d'una gestió optimitzada de les diferents velocitats de transmissió assignades als usuaris que comparteixen l'accés i d'acord als seus diferents perfils de Qualitat de Servei (QoS). Hem assumit una restricció de càrrega màxima de la xarxa, un repartiment d'espectre just entre usuaris i un potència màxima disponible als

terminals mòbils. També s'assumeix un servei de temps real amb velocitats de transmissió variables. La probabilitat de blocatge s'utilitza per mostrar els beneficis de l'algoritme

2. Gestió de les velocitats de transmissió d'usuaris ja admesos en l'enllaç ascendent d'un sistema WCDMA, per a garantir una probabilitat de satisfacció , definida com la probabilitat d'excedir un cert valor de velocitat de transmissió. També s'assumeix un servei en temps real amb velocitats de transmissió variables i es consideren les mateixes restriccions que en el cas de control d'admissió.

3. Gestió de les velocitats de transmissió d'usuaris ja admesos en l'enllaç descendent d'un sistema WCDMA, per a assegurar un màxim retard en el lliurament de paquets. Hem assumit restriccions de repartiment just d'espectre i una màxima potència disponible a l'estació base. Se suposa un servei interactiu basat en un model WWW de tràfic. S'ha introduït també un algoritme de referència amb la finalitat de fer comparatives. El paràmetre utilitzat per avaluar el comportament de l'algoritme és la probabilitat de pèrdua de paquets.

4. Gestió conjunta de serveis en temps real i serveis interactius en l'enllaç descendent d'un sistema de WCDMA. Incorpora part dels algoritmes emprats en els casos anteriors amb els mateixos models de tràfic i les mateixes restriccions. En aquest cas s'ha utiitzat la probabilitat de satisfacció i la probabilitat de pèrdua de paquets per a valorar el rendiment i el retard dels paquets, respectivament.

5. Algoritme de Gestió conjunta de recursos Radio per a un escenari B3G on un usuari pot ser servit per més d'una tecnologia d'accés. En particular s'han considerat els accessos WCDMA i TDMA. S'ha suposat que els serveis són interactius i s'ha utilitzat la probabilitat de pèrdua de paquets per a mostrar el bon comportament de l'algoritme. Val la pena notar el compromis entre el retard introduït per executar un Handover vertical i la probabilitat final de pèrdua de paquets.

6. Algoritme per a la gestió de velocitats de transmissió d'usuaris ja admesos en el enllaç descendent d'un sistema OFDMA, per assegurar un màxim retard en el lliurament de paquets. En aquest cas, s'han presentat dues variants de l'algoritme, i es mostra que la probabilitat pèrdua de paquets actua millor que altres algoritmes heurístics publicats fins aleshores.

La tesi també apunta cap a futures línies de recerca avui per avui encara no explorades per a explotar les bases de la metodologia aquí desenvolupada. Inicialment consisteixen en escenaris multicel·lulars centralitzats per a continuar posteriorment escenaris distribuïts, en particular per

als sistemes OFDMA, que són la base dels sistemes d'accés de la quarta generació. En aquest sentit també s'inclouria un refinament de la metodologia desenvolupada per a escurçar els temps de simulació i assolir una millor precisió en els resultats obtinguts

S'ha inclòs també una llista de les publicacions fetes en connexió amb els resultats. Consta de conferències internacionals així com una publicació en una revista indexada.

Preface

Wireless networking has emerged as one of the major areas for research and industrial development in the past couple of decades and it has represented a revolution in the telecommunication world that could already perhaps be comparable to the transition from the analogical to digital technology. In this environment and particularly related with the 3G cellular multimedia mobile has been focused this research work.

The continuous pressure in terms of users capacity demand as shown by the explosive market penetration in the still active 2G (2nd Generation) and 2.5 cellular systems, like GSM/GPRS, it is expected to be continued with the current 3G (3rd Generation) systems and particularly with the Universal Mobile Telecommunication System (UMTS), the dominant one. In that respect, Radio Resource Management (RRM) has been proved to be a very valuable tool for achieving the necessary spectrum economies in front of the users' pressure to get radio access while still guaranteeing the Quality of Service (QoS) agreed. In UMTS, Wide Band Code Division Multiple Access (WCDMA) is the radio access adopted and will be used as a framework where to elaborate the new RRM strategies proposed in this work. In addition to that, Common Radio Resource Management (CRRM) strategies will be also introduced in Beyond 3G systems (B3G) as a guarantee for a specific QoS level. Finally, in 4G (4th Generation)systems, Orthogonal Frequency Division Multiple Access (OFDMA) is the radio access adopted and it it also expected RRM will be used in a similar and even more intensive manner than in 3G systems.

Admission Control is one of the more important RRM functions the mobile operators have to face. The result of executing an Admission Control algorithm has been in the past just to grant or not a requested voice channel. However in 3G systems like UMTS, Multimedia applications can not be managed in the same way. Users will have a multiple options choice in terms of bit rate allocated to each one of the multiple connections they can use in a session. These multiple options will also be available once the users are already admitted. We can then imagine a list of acceptable bit rates a user can apply which the operator has to be able to manage them optimally depending on the QoS required and the system load. Same considerations would aply also in 4G systems.

We have proposed a new bit rate adaptive allocation strategy for both WCDMA and OFDMA that optimises at the same time the number of users getting access to the system in accordance with their contractual user profile and a cost function that penalises the inefficient use of the resources allocation. Then a system model has previously been identified, a simulation

environment has been implemented introducing propagation and mobility models in an isolated cell and a Hopfield Neural Network (HNN) has been used as optimisation tool.

Encouraging results have been obtained showing the gains in terms of Average Bit Rate allocated to each users and the corresponding Blocking probability for real time services. Similar results have also been obtained in terms of the packet delay and the corresponding dropping rate for delay sensitive services. As being this work an initial step in a more long term research, finally a future work is stated as a continuation of the results presented here

Contents

List of figures

List of tables

Acronym List

2D HNN: Two Dimensional Hopfield Neural Networks
3D HNN: Three Dimensional Hopfield Neural Networks
1G: First Generation
2G: Second Generation
3G: Third Generation
4G: Fourth Generation
3GPP: Third Generation Partnership Project

ABR: Average Bit Rate
ADALINE: Adaptative Linear Neuron
AMPS: Advanced Mobile Phone System
ANN: Artificial Neural Network
ART: Adaptive Resonance Theory
ASM: Advanced Spectrum Management
ATM: Asynchronous Transfer Mode

B3G: Beyond 3rd Generation
BB: Base Band
BER: Bit Error Rate
BLER: Block Error Rate
BSC: Base Station Controller

CDF: Cumulative Distribution Function
CDMA: Code Division Multiple Access
CHNN: Continuous Hopfield Neural Network
CIR: Carrier to Interference Ratio
CN: Core Network
CNN: Continuous Neural Network
CRMS: Customer Relationship Management System
CRRM: Common Radio Resource Management
CTM: Continuous Time Model

DHNN: Discrete Hopfield Neural Network
DNN: Discrete Neural Network
DRA: Dynamic Resource Allocation
DSA: Dynamic Spectrum Allocation
DTB: Data Transfer Buffer
DTM: Discrete Time Model
DVB-T: Digital Video Broadcasting-Terrestrial

EDGE: Enhanced Data rates for GSM Evolution
EGPRS: Enhanced GPRS
EUTRA: Evolved UMTS Terrestrial Radio Access

FDMA: Frequency Division Multiple Access
FFT: Fast Fourier Transform
FIFO: First In, First Out
FM: Frequency Modulation
FPGA: Field Programmable Gate Array
FRF: Frecuency Reuse Factors

GERAN: GSM EDGE Radio Access Network
GOBR: Generalized Optimum Bit Rate
GPF: General Proportional Fair
GPRS: General Packet Radio System
GPS: General Processor Sharing
GSM: Global System for Mobile communications

HNN: Hopfield Neural Network
HSDPA: High Speed Downlink Data Access
HSUPA: High Speed Uplink Data Access

IBM: International Business Machines
IEEE: Institute of Electrical and Electronics Engineers
IFFT: Invers Fast Fourier Transform
IMT: International Mobile Telecommunications
IMTSC: IMT Single Carrier
IP: Internet Protocol
ITU: International Telecommunication Union
ITU-R: ITU Radiocommunication

LTE: Long Term Evolution
LVQ: Learning Vector Quantization

MAC: Medium Access Control
MCS: Modulation Coding Scheme
MC-CDMA: Multi-Carrier Code Division Multiple Access
MOSFET: Metal Oxide semiconductor Field Effect Transistor
NP: Nondeterministic Polinomial
NRT: Non Real-Time

OBR: Optimum Bit Rate
OFDMA: Orthogonal Frequency-Division Multiple Access

PBF: Pass Band Filter
PF: Proportional Fair

QAM: Quadrature Amplitude Modulation
QoS: Quality of Service

RAB: Radio Access Bears
RAN: Radio Access Network
RAT: Radio Access Technologies
RB: Resource Block
RBF: Radial Basis Function
RC: Resistance-Capacitance Circuit
RF: Radio Frequency
RNC: Radio Network Controller
RRM: Radio Resource Management
RSS: Reference Scheduling Scheme
RT: Real Time

SB: Spectrum Brokerage
SF: Spreading Factor
SIR: Signal to Interference
SMS: Short Message Service
SOM: Self-Organizing Map
STS: Short-Term Scheduler

TDMA: Time Division Multiple Access
TFCS: Transport Format Combination Set

UE: User Equipment
UMTS: Universal Mobile Telecommunication System
UTRA: Universal Terrestrial Radio Access
UTRA FDD: UMTS Terrestrial Radio Access Frequency Division Duplex
UTRAN: UMTS Terrestrial Radio Access Network

VLSI: Very Large Scale Integration

WCDMA: Wideband Code Division Multiple Access
WLAN: Wireless Local Area Network

WiMAX: Worldwide Interoperability for Microwave Access

Chapter 1

Introduction

1.1 Motivation and problem statement

Wireless networking has emerged as one of the major areas for research and industrial development in the past couple of decades and is expected a continual expansion into new areas. Extension of the cellular voice telephony, wireless hotspot Internet access, wireless home and corporate networking, to ad hoc and mesh networks and wireless personal area networks have deeply impacted our lifestyle. After more than a decade of exponential growth, today's wireless industry is one of the largest industries in the world. At the time of this writing, more than one billion people is subscribed to cellular telephone services worldwide, billions of short messages are weekly exchanged, wireless local area networks are connecting home computers and provide access in corporate and public hotspots and yet we are expecting emerging ad hoc networks to connect a myriad of terminals that range from small sensors to huge appliances in the near future. Mobile Communications have definitively become Wireless Communications. The penetration of the cellular telephone in some countries in Europe has exceeded the population of the country and there are over 250 million subscribers only in China.

Multimedia and Cellular industries marriage is undoubtedly the challenge mobile operators have to face nowadays. After three years of economical turbulences in the technological sector, operators finally decided to strongly deploy 3G Systems. In the short term UMTS (Universal Mobile Telecommunication System), particularly in Europe and Japan as wideband UMTS through HSDPA (High Speed Downlink Data Access) and HSUPA (High Speed Uplink Data Access), appear as a key element for the definitive irruption of the mobile multimedia paradigm. Focusing on the medium and long term, the so called 4G (4 Generation) boosted by the ITU-R (International Telecommunication Union Radiocommunication) as Advanced IMT (International Mobile Telecommunications) 2000 aims at offering wireless mobile systems able to satisfy the current identified needs in terms of both bit rate and network performance. In this respect current developments as WiMAX (Worldwide Interoperability for Microwave Access)

and LTE (Long term Evolution) proposals in the IEEE and 3GPP standardization bodies respectively, are taking the lead to bring us a real 4G.

Inspired and motivated by the extraordinary development the mobile communication has experimented over the near last three decades, our aim is contribute to this field in some aspects related to the radio resource management in both WCDMA (Wideband Code Division Multiple Access) and OFDMA (Orthogonal Frequency Division Multiple Access) systems that we could summarize as the basic pilars in the construction of the 3G and 4G respectively. In particular our research goal has mainly consisted in finding new approaches for the provision of some resources management strategies provided some specified Quality of Service (QoS) attributes restriction are guaranteed.

1.2 Objectives, initial assumptions and approaches considered

The main objective of this thesis work consists in envisaging novel approaches to deal with the Dynamic Resource Allocation (DRA) problem in multi-user mobile communications systems where there are a big number of possible allocated bandwidths to assign per user and where certain requirements in terms of QoS have to be fulfilled. In addition to that, several users' profiles should be managed at the same time.

Two main sub-objectives stemmed from the previous rather general one and have been pursued in this work:

1. To envisage a DRA strategy for a multi-user WCDMA system allowing each multi- rate enabled user to be assigned the most suitable bandwidth (Bit Rate) so as to keep a specified user satisfaction probability measured as temporal percentage of exceeded required troughput. This would apply to both real time /streaming services and non real time services.

2. To envisage a scheduling mechanism for delay sensitive services for multi-user WCDMA and OFDMA packet based systems so as to guarantee a minimum dropping rate of packets exceeding the maximum tolerable delay.

The initial assumptions we have taken to cover the above objectives will consider the key aspects we can find in the deployment of mobile communications that use as access technologies WCDMA and OFDMA. In such scenarios, the key concepts related to power constraints, channel propagation issues and widely accepted traffic models will be entirely captured. However, the link aspects connected to Physical layer impairements like synchronization errors, imperfect channel estimation, etc or those related to the actual deployed systems like multicell interference will not be retained and will be considered out of the scope of this work. Doing this way we only loose some precision in the final figures from an engineering point of view but our understanding is that the proposed schemes still can contribute with the right conclusions highlighting the attained improved performance in comparative terms

The conducting approach through all this work relies in the use of the Hopfield Neural Networks (HNN). To the best of our knowledge it is the first time this approach is used as being part of Radio Resource Management (RRM) schemes aplied to both WCDMA and OFDMA mobile networks. The rationale behind this approach relies in the essentialy combinational nature of the Radio resources allocation problem in the above multi-rate systems. The suitability of the approach follows from the fact that essentialy the HNN is a combinational problem solver.

1.3 Publications

The work presented in this thesis has been published in the following papers (sorted in reverse-chronological order)

Journals

1. Garcia, N.; Perez-Romero, J., R. Agustí *A New OFDMA Scheduler for Delay-Sensitive Traffic Based on Hopfield Neural Networks* EURASIP Journal on Wireless Communications and Networking, vol. 2008, Article ID 817676, 9 pages, 2008. doi:10.1155/2008/817676

In proceedings of peer-reviewed International Conferences

1. Garcia, N.; Perez-Romero, J., R. Agustí *Novel Joint Channel and Queuing-aware OFDMA Scheduler for Delay Sensitive Traffic.*18th IEEE International Symposium on Personal Indoor Mobile Radio Communications (PIMRC) Athens (Greece) .3-7 September 2007 Proceedings PIMRC 2007. IEEE, 2007, p. 1-150.

2. Garcia, N.; Perez-Romero, J., R. Agustí *A new CRRM Scheduling Algorithm for heterogeneous networks using Hopfield Neural Networks* Wireless Personal Multimedia Communications (WPMC). San Diego (USA) 17-20 September 2006 *Proceedings of WPMC*. IEEE, 2006, p. 1-5.

3. Garcia, N.; Perez-Romero, J. *A DRA scheme based on Hopfield neural networks methodology* The 13th IEEE Mediterranean Electrotechnical Conference (MELECON) Benalmadena (Spain) 16-19 May 2006 Proceedings Electrotechnical Conference, 2006 p. 583-586

4. Garcia, N.; Perez-Romero, J., R. Agustí *A Novel Scheduling Algorithm for Delay-Oriented Services based on Hopfield Neural Networks Methodology* IEEE Wireless Communications & Networking Conference (WCNC) Las Vegas (USA) 3-6 April 2006 *Proceedings*. WCNC'06 E-ISBN: 1-4244-0270-0 ISSN: 1525-3511 ISBN: 1-4244-0269-7

5. Garcia, N.; R. Agustí, Perez-Romero, J., *A User-Centric approach for Dynamic Resource Allocation in CDMA systems based on Hopfield Neural Networks* The 14th IST Mobil & Wireless Communication Summit. Dresden (Germany) June 2005 Proceedings. Editorial Information Society, 2005, p. 1-5

These Publications are also related with contributions to following research projects:

1. Network of Excellence in Wireless Communications (NEWCOM). Period: 2004 / 2006

Technical Documents

1. 507325 NEWCOM DR7.4 Final Report on the activities carried out in the European Network of Execellence in Wireless COMmunications-Dept. 7 . 28th February 2007

1.4 Summary of the thesis work

Chapter 1, this thesis introduction, presents the motivation and main contributions from the author.

Chapter 2, provides as related work an introduction to the Radio Resource Management (RRM) concepts that are a basic part of the operation of cellular networks, after having introduced a rather generic approach to the Mobile Communications world. In particular all these concepts are mainly described as being part of the Radio Acces Technologies (RAT) we are going to use

in a great extent in this thesis: Wide Code Division Multiple Access (WCDMA) and Orthogonal Frequency Division Multiple Access (OFDMA) based RAT. Furthermore, in addition of depicting a global picture of the above RATs, some key RRM basic elements conected with these RATs will be also overviewed since thay are going to be used later in this thesis. In particular, the concepts and performance indicators of Load factor and breathing effect in uplink WCDMA and Power allocation in downlink WCDMA have been illustrated

Chapter 3 is an overview of the Artificial Neural Networks (ANN) leading finally to the Hopfield Neural Networks (HNN) basics that constitute the core RRM element to be extensively exploited in this thesis. Continuous and Discrete HHN have been introduced, focusing mainly in the differential equations capturing the HNN dynamic sytem features as well as the stability conditions stated by the Theorem of Lyapunov. In particular the concepts of Energy function, equilibrium states and global stability of the continuous Hopfield Neural Network are introduced in depth as they will be exploited later in the different RRM mechanisms has been proposed and assessed in this thesis. Finally a two dimensional HNN model for Radio Resource Management is described and their working hipotesis settled.

Chapter 4 introduces novel approaches for considering the Dynamic Resource Allocation (DRA) in a multi-rate WCDMA wireless network. Three novel dynamic resource allocation algorithms are proposed that make use of the HNN methodology. It provides a fast way of finding the optimum resource allocation that minimises a given energy function reflecting specific service and system constraints. So, it introduces a novel HNN algorithm for managing different service profiles for a user centric approach in the uplink of a WCDMA system supporting non real-time (NRT) communication, in which each user has a range of bit rates that can be allocated depending on its service profile, defined by a minimum satisfaction bit rate. Significant and promising results are shown for a single isolated cell WCDMA scenario including realistic propagation and mobility models. Furthermore, novel RRM strategies HNN-based are introduced to schedule the downlink transmissions in a WCDMA scenario with delay-oriented services. Finally, RRM applied to both schedule the downlink transmissions in a WCDMA scenario with delay-oriented services and real time services are also devised. Algorithms are evaluated by means of simulations and compared with a reference scheme, revealing its ability to adapt to the specific services and traffic conditions.

Chapter 5 proposes a novel Common Radio Resource Management (CRRM) algorithm for heterogeneous scenarios making use of the HNN methodology reflecting specific service and system constraints. The proposed algorithm is applied in a heterogeneous wireless scenario with

WCDMA and Time Division Multiple Access (TDMA) radio access networks to schedule the downlink transmissions of a delay-constrained service. The algorithm is assessed using the same approach already stated in Chapter 4.

Chapter 6 introduces a novel joint channel and queuing-aware OFDMA scheduler for delay sensitive traffic based on a Hopfield Neural Network scheme. It allows providing an optimum OFDMA performance by solving a complex combinational problem. The algorithm is based on distributing the available subcarriers among the users depending, on the one hand, on the time left for the transmission of the different packets in due time, so that packet droppings are avoided. On the other hand, it also accounts for the available channel capacity in each subcarrier depending on the channel status reported by the different users. The different requirements are captured in the form of an energy function that is minimised by the algorithm. In that respect the chapter illustrates two different algorithms coming from two settings of this energy function. The algorithms have been evaluated for delay sensitive traffic and have been compared against other state-of-the-art algorithms existing in the literature, exhibiting a better behaviour in terms of packet dropping probability.

Chapter 7 presents the general conclusions of the thesis and suggests possible future research lines. In particular an extension from the unicellular case to multicellular scenarios and ulterior distributed new Radio Resource Management strategies is first envisaged. Furthermore new analytical approaches aimed to improve the optimization procedures for the resolution of the HNN-based dynamical system as well as to shorten and improve their numerical resolution are also devised.

Annex A summarises reference scenarios for evaluation of common radio resource management strategies considered in this work.

Annex B presents the relationship between each energy function considered in the HNN-based algorithms and its associated interconnection matrix and input bias currents.

Chapter 2

Related work

2.1 Mobile Communications

The Mobile Communications concept is as old as the radio itself. However the scarcity of the available radio spectrum resources prevented the masses from its use. The cellular mobile communications, as new concept born in the Bell Labs in the sixties decade in last XX century, allowed millions of people being able to use the spectrum resources allocated to the mobile communications service breaking down the spectrum barrier. That come to reality in what was the first commercial mobile systems Advanced Mobile Phone System (AMPS) appeared in Chicago in the last seventies. This concept is based in a systematic and intelligent reuse of the spectrum available through thousands of cells deployed in a territory, being each cell covered by a Base Station equipped with the transmission/reception radio equipment [1].

The AMPS was the start of the so named First Generation of Mobile Communications (1G) characterized by its analogical principles imposed by the FM modulation and the Frecquency Division Multiple Access (FDMA) Technique. However technological limitations due to the large size and power consumption of the terminals restricted his use in practice to vehicles. Moreover the analogical basis of this Generation prevented it from its expansion due to its inherent technical limitations.

Global System for Mobile communication (GSM) was the mobile communication system that actually made the sector increase in a dramatic way. Political factors derived from the sector liberalization and the European Commission implication as well as the adoption of a digital technology as a basis for the GSM design fuelled an unforeseen GSM success, expected still to be continued for the next 5-10 years. GSM was built on four pillars:

- Digital Technology
- International roaming

- Security: Authentication and Confidentiality
- Digital data up to 9.6 Kb/s and SMS

GSM was initiated in its design phase at the beginning of 80's and is the paradigm of the Second Generation of Mobile Communications (2G). Then Internet was at that time still a researching concept and was not incorporated in GSM. Later on, General Packet Radio System (GPRS) as a GSM extension, incorporated the packet data oriented protocols used in Internet so as GPRS is able to interconnect with the Internet world. This GSM/GPRS evolution is included in the so called 2.5 Generation (2.5 G) and holds still the limited bit transmission rate imposed by the GSM basis. Then GPRS represents just a small step ahead on the 2G.

Universal Mobile Telecommunication System (UMTS) is a new mobile communications system trying to solve the problems raised by GSM in terms of technology, services and capacity. It is a European system fuelled by the European Commission from the II Framework Program in the early 90's. UMTS initiates the Third Generation of Mobile Communications (3G). Internet native concepts and Multimedia are relevant aspects of UMTS in addition to the capacity improvement in terms of radio deployment. The wideband capacity enabled by UMTS can provide actual multimedia features by the first time in the mobile arena. The new radio access technology based in Wideband Code division Multiple Access (WCDMA) is another key point able to introduce the required flexibility elements in the new multimedia dominated era, e.g changing easily a user specific rate in a certain point of the time. Then, allowing to different users sharing different multimedia sessions among them in a dynamical and transparent way, keeping still under control the required Quality of Service (QoS). This flexibility requires however the development of new concepts of Radio Resource Management (RRM) to operate the network in an efficient way.

During the evolution from 2G to 3G, a range of wireless systems, like International Mobile Telecommunications 2000 (IMT-2000), Bluetooth, Wireless Local Area Network (WLAN), and HiperLAN, have been developed. All these systems were designed independently, targeting different service types, data rates, and users. As all these systems have their own merits and shortcomings, there is no single system that is good enough to replace all the other technologies. In particular, IEEE standardisation bodies succeeded with the 802.xx sytems, in particular WLAN and Worlwide Interoperability for Microwave Access (WiMAx).

As a possible new vision trying to better exploit the above heterogeneous scenario full of different systems, the so called 3.5G has appeared to operate in a coordinate manner the different systems thet would be presented to the user as a unique system. This heterogenous system would exploit the truncking gain in terms of system capacity as well as the enrichment with the different system features each isolated system presents in terms of 3.5G overall system performance. This is however a slow process that might be only partially accomplished due to the different integration issues it raises.

Currently, and following a similar evolutive path that follows the 3G, it is being fostered by the International Telecommunications Union Radio ITU-R the Advanced IMT 2000 systems or Fourh Generation of Mobile Communications (4G). 4G networks are all-IP based heterogeneous networks planned to be deployed from 2015 that finally allow users to use any system at any time and anywhere at a low transmission cost. Among the key features 4G will bring it is worth mentioning peak user bit rates from 100Mb/s to 1Gb/s and end to end packet delays below 10 ms, being the Ortogonal Frequency-Division Multiple Access (OFDMA) the multiple access technique selected.

Figure 2-1 shows a picture of the mentioned Mobile Generations. Basically a trade-off between the mobile speed (and the coverage area associated), and the Data transfer Rate can be noticed.

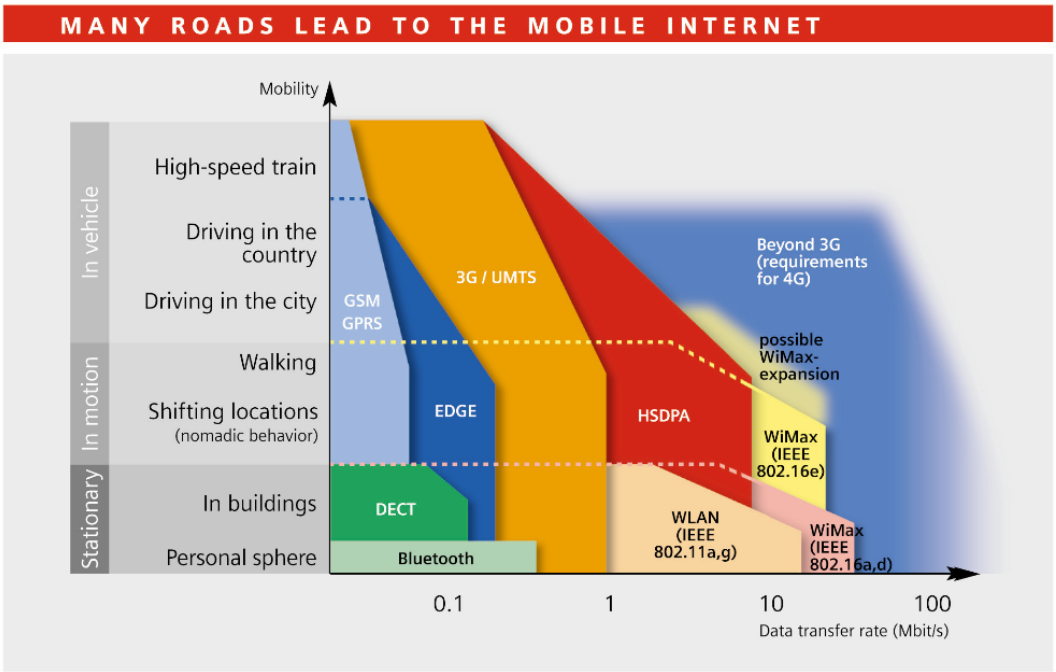

Figure 2-1 Evolution Mobile Communication Systems

2.2 Radio Resource Management concept

The mobile networks starting in the 3G are characterized by a broad range of services expected to be supported. In particular the 3G networks services are divided into four Service classes: conversational (e.g. voice), streaming (e.g. video), interactive (e.g. www browsing) and background (e.g. e-mail). Even more services can be foreseen in the currently planned 4G. The provision of such mobile multimedia services has to be provided under QoS guarantees, which would not be possible without a proper utilization of the air interface resources by means of suitable RRM strategies. These RRM functions have a key role in the development of the modern mobile networks since they dictate how to exploit the always scarce radio resources in the mobile network, and still ensuring the target QoS values, the planned coverage area and a high system capacity [2].

Such strategies should deal with the specific peculiarities of the radio access technology that in the UMTS Terrestrial Radio Access Frequency Division Duplex (UTRA FDD) mode of UMTS is based on WCDMA. One of the peculiarities of these access schemes is that it lacks from a constant value for the maximum available capacity, since it is tightly coupled to the amount of interference in the air interface. Therefore, RRM functions become crucial to manage this interference depending on the provided services. Moreover, RRM strategies are not subject of standardisation, so that they can provide a differentiation aspect among manufacturers and operators. Additionally, RRM functions can be implemented in many different ways, this having an impact on the overall system efficiency and on the operator infrastructure cost, so that definitively RRM strategies will play an important role in such scenarios.

2.2.1 Quality of Service and RRM

The Quality of Service concept can be understood in many different forms and levels. Although QoS is inherently subjective, the realization of QoS in a communication network needs to be associated with some QoS parameters that characterize a satisfactory user’s perceived service. QoS figures may range from parameters at a system or global network level to parameters on a per connection basis.

QoS provisioning in cellular mobile systems is particularly complex given the large number of effects that tend to degrade the communication links. Fading in propagation conditions, interference, channel distortion, etc. are only some examples.

2.2.1.1 WCDMA Basics

In WCDMA users transmit at the same time and frequency and by means of different spreading sequences, none perfectly orthogonal in most of the cases. Consequently, there is a natural coupling among the different users that makes the performance of a given connection much more dependant on the behavior of the rest of users sharing the radio interface compared to other multiple access techniques (i.e. FDMA or TDMA). Additionally, the number of tunable parameters in a WCDMA network is significantly higher than that of a 2G TDMA-based one, not to mention those from 1G FDMA.

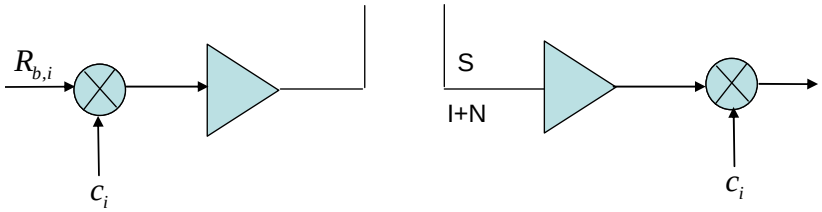

Figure 2-2 Scheme of a link used in WCDMA

One key performance figure at connection level that is strongly related to QoS is the Bit Error Ratio (BER). The calculation of this figure usually requires of involved physical layer simulations. In the Figure 2-2 a schematic of a link used in WCDMA is shown. An i-th user has data to be transmitted at the user rate $R_{b,i}$ or R_b in general terms. This data sequence is multiplied by a so named Pseudo-random code sequences, which determines the receiver to which this data train is headed to. The Pseudo-random code sequence rate R_c is much higher than the so called Base Band (BB) bit rate R_b according to the Spreading Factor, SF, defined as [3]

$$SF = \frac{R_c}{R_b} = \frac{T_b}{T_c} >> 1, \tag{2.1}$$

where T_b and T_c are the bit and chip duration respectively. In practice we can set the offered spectrum bandwidth W, as $W = 1/T_c = R_c$, named chip rate. As a consequence of that the BB spectral occupancy is spread out to W, filling up the entire spectrum available.

After up-conversion of this base band signal to the Radio Frequency (RF) format, this signal is mixed up in the air with other similar ones. The radio bandwidth W available for the radio

access channel then is shared via different Pseudo-random code sequences by all the users. That is, a WCDMA access encompasses many links like the shown in Figure 2-2, each one including a differente uncorrelated Pseudo-random code c_i, so that the receiver can exploit this uncorrelated property to separate the different user's transmissions conveyed under the same WCDMA bandwidth W.

At the receiver, ideally only the desired signal should be retained after down conversion, multiplication and base band filtering of the received compound signal at the receiver antenna output. The rest of users are separated thanks to the Pseudo-random code sequences cross-correlation properties. However a complete removal of these non desired signals is not possible and there is always left a residual signal called multiuser interference at the receiver output. The Signal to Interference (2.2) plus thermal noise ratio at the receiver output is called SIR_{out}. It can be shown that its value can be obtained from the Signal to Interference plus thermal noise ratio at the receiver input SIR_{in} as

$$SIR_{out} = \frac{S}{I+N} SF \approx SIR_{in} \cdot SF \,, \tag{2.2}$$

where S is the desired signal power input, I is the addition of all non desired users input powers and N is the generated power thermal noise: $N = K \cdot T \cdot F \cdot W$, K is the Boltzman constant, $T = 290$ Kelvin degrees and F is the noise factor of the receiver. N is usually much lower than I.

A typical QoS parameter is the E_b / N_0 value where E_b is the desired bit energy given by $E_b = S \cdot T_b$ and N_0 is the equivalent noise spectral density power given by

$$N_0 = \left(I + N\right) / W = \left(I + N\right) T_c \,. \tag{2.3}$$

The Bit Error Rate in the user data train delivered to the destination user can be calculated as

$$BER = f\left(\frac{E_b}{N_0}\right). \tag{2.4}$$

that depends on the coding and modulation used as well as the channel characteristics. Consequently, E_b / N_0, a key parameter in WCDMA RRM , will be used later in this work. In

addition to that, other system level QoS values would be the end to end delay, the dropping probability, etc, that are strongly related to the actual E_b / N_0 value.

2.2.1.2 OFDMA Basics

Figure 2-3 shows the basics of an OFDM Transmitter scheme. The complex serial user's transmited symbols, S_i of duration T_s, are parallelized in blocks of N symbols $(S_0, S_1, \ldots, S_{N-1})$ through the serial parallel converter (S/P), then they are QAM modulated, at $1/T$ symbols per second, by N subcarriers $(F_0, F_1, \ldots, F_{N-1})$ separated each other $1/T$, being $T = NT_s$, and finally pass band filtered (PBF). This scheme is easily implemented in practice by means of IFFT chips. The N modulated subcarrier is then orthogonal and hence easily separated in the receiver. The rationale behind this scheme relies in its capability to transmit without Intersymbol Interference wide band signals through dispersive channels like those appearing in wide band mobile communications. Certainly the user bit rate as seen by OFDM subcarrier system is about N times (depending on the bit rate of each QAM modulator) lower than the actual rate as transmitted by the user. This scheme is therefore especially convenient when high bit rates are required as in 4G systems.

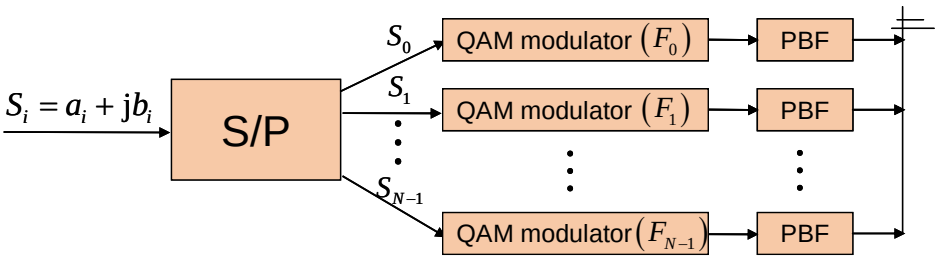

Figure 2-3 Scheme of a transmiter used in OFDMA with subcarriers $F_0, \ldots, F_{N-1}$

OFDMA turns out when a Multiple Access is to be provided taking OFDM as a basis. Then a 2D frequency-time scheme is arranged as shown in Figure 2-4. By defining the unit of resource as that formed by one subcarrier and one time slot, then, the multiple access is easily accomplished by assigning each user with a number of subcarriers and time slots. It is denoted in the above Figure 2-4 by $TX\,\#1$, $TX\,\#2$, $TX\,\#3$, $TX\,\#4$, $TX\,\#5$ and $TX\,\#6$. The great amount of possible assignments makes the OFDMA system extraordinary flexible and at the same time very difficult of managing in an optimal way. Answering to the question: how to assign resources to users?, is then a key issue in terms of RRM and one of the main objectives to be covered in this work.

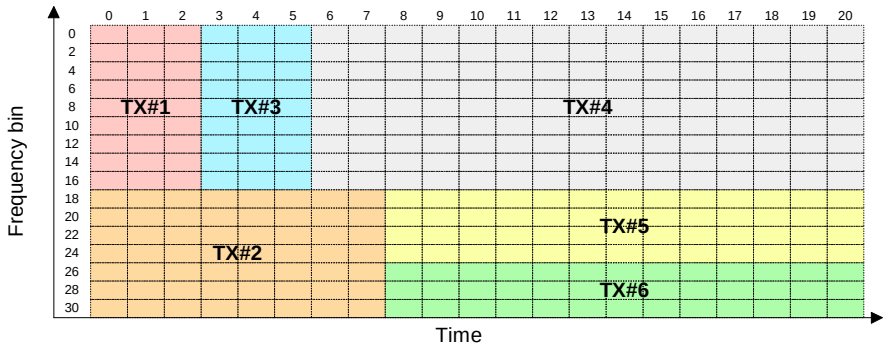

Figure 2-4 OFDMA scheme with several users accessing a common channel

2.2.2 RRM Functions

The problem faced by a network operator is to offer a system where the network usage is maximized for a given set of QoS requirements. In the traditional approach to solve this problem two aspects can be clearly distinguished: the network planning (e.g. the design of the fixed network infrastructure in terms of number of cell sites, cell site location, number and architecture of concentration nodes, etc.) and the radio resource allocation (e.g. for a given network deployment, the way in which radio resources are dynamically managed in order to meet the instantaneous demand of the users moving around the network).

In the framework of 2G mobile systems (i.e. GSM), the network planning is key. Second Generation wireless systems focused their effort on providing mobile voice applications to the end user with an acceptable quality. The QoS for voice service is mainly controlled through a suitable frequency assignment among cell sites in order to provide a sufficient Carrier to Interference Ratio (CIR). On the other hand, the call blocking probability is the other fundamental QoS parameter and it is controlled through providing in a first step several frequencies to a given cell site and in a second step by adding new sites. For a given network configuration, there is an almost constant value for the maximum capacity because radio resource allocation actions in the short term scale have a limited impact. Additionally, radio resource allocation in the short term (e. g. in the order of tenths/hundreds of milliseconds) has not much to do in a scenario where the supported service (e. g. voice) requires for a channel with constant quality and tight delay constraints.

The evolution of the end user needs towards multimedia applications has pushed the wireless community to conceive the so called 3G systems (such as UMTS or CDMA 2000), where a very large amount of both circuit switched services and packet switched services for voice and data will be provided. In WCDMA based systems, unlike the FDMA and TDMA ones, there is not a

constant value for the maximum available capacity, since it is tightly coupled to the amount of interference in the air interface. In WCDMA we have the so called soft capacity as a specific system identity parameter. Another relevant point relies to the multi-service scenario where constant delay requirement for specific data packet services can be defined. This fact, consequently, opens the ability to exploit RRM functions so as for instance to grant a lot of resources to a service in case when its deadline time is about to expire, instead of doing that in a constant pace. In terms of radio network planning, a range of new planning challenges specific to WCDMA arise: soft-handover overhead, cell dominance and isolation, etc. In short RRM in 3G systems is a very involved process not comparable to what occurred in 1G and 2G.

In any wireless system there is clearly a very scarce resource, which is the available bandwidth. 3G systems intend to provide several classes of services at different bit rates while assuring a QoS for each one. In order to harmonise these two contradictory points (scarce bandwidth and a stringent QoS) Radio Resource Management is the functionality located in both User Equipment (UE) and Radio Network Controller (RNC) inside UTRAN (UMTS Terrestrial Radio Access Network) in charge to perform a proper management of the available radio resources. RRM contains the algorithms suitable to enable the radio path to fulfil the QoS criteria required by the services transmitted through this radio segment part.

In particular, Release 99 of UMTS from 3GPP [4] is mainly focused on an efficient provision of circuit switched services. Packet switched services are still considered as complementary, and, moreover, the optimisation of capacity in a multi-service environment is not envisaged as foreground. However, Release 4 and 5 of UMTS in 3GPP and future evolutions in 3G systems are focusing in an all-IP end to end architecture, and consequently the packet switched services in the air interface will gain momentum and the optimisation of capacity in a multi-service environment is going to be faced definitely. Then, addressing the RRM topic to guarantee a given QoS in a packet driven environment such as the above mentioned in the framework of the 3G systems and in particular in the WCDMA scenario is retained as a key research issue in such all-IP scenarios.

Then, RRM strategies are a key in the current 3G and forthcoming 4G communicaction mobile systems. The Radio Resource Management is carried out through the execution of specific algorithms. These algorithms are left to the operator or system vendor to define and consequently are not going to be object of standardization. In short, they relate the end user requirements in terms of QoS, with the network resources necessary to accomplish this final objective while at the same time guaranteeing the QoS of the rest of users already admitted in the radio system [5].

In a first general approach, Figure 2-5 illustrates the functional actions these RRM algorithms have to carry out. First, the Admission Control is in charge of accepting the entrance of the user in the network in the light of the required QoS. Later on the RR Management allocates the necessary radio resources: Power, Bandwidth and Codes to obtain the required QoS. That is accomplished in different time scales, from short term up to long term. Congestion Control mechanisms should be devised to face situations in which the system has reached a congestion status. That is, the RRM cannot guarantee anymore to the admitted users the initial QoS contracted due to the evolution of system dynamics (mobility aspects, increase in interference, etc.) and a new setting of the assigned QoS should be undertaken. This fact has to be taken into account in the admission phase negotiation. Both Admission and Congestion control rely widely in the measurements carried out mainly in terms of load and interference.

Measurements are always the input to all the RRM functions. The amount of measurements obtained is a trade-off between the final performance and the signalling load. The greater the number of measurements we have, the better the performance attained in the RRM is and the higher the signalling load needed to transport the measurements from the measuring points to the decision points in the network. Handover Management is another key RRM function that allows the service continuity when the user is moving from one cell to another. In this process the old accees point (Base Station), which the user was attached to, is replaced by the new one. As a final point of a RRM process a set of Radio Access Bearers (RAB) in the radio access part (Base Stations to terminal link) are set up.

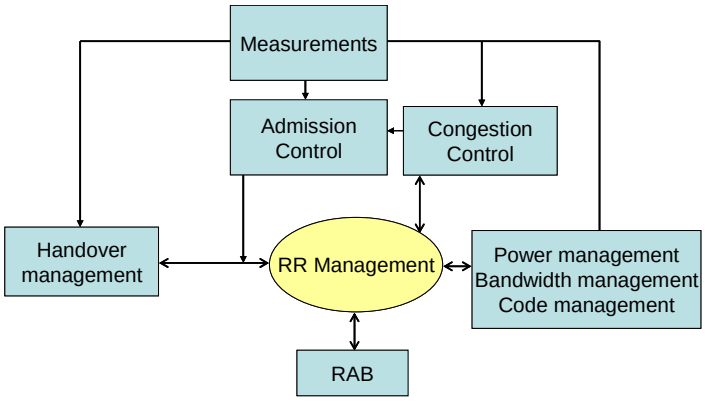

Figure 2-5 RRM Functions

In this work we will focus on the Admission Control algorithms and Management of Transmission parameters applied to a multirate enabled user terminal. As we will see later the problem could be stated as not only whether to admit or not a user requesting a service, but also as how to select a proper bit rate from a set of possible values as a granted bandwidth. That is

the key point in both WCDMA and OFDMA environments and constitutes a key contribution of this work. For this purpose the assignments have to be done to get maximum system efficiency assuming that the system restrictions and the fairness among users are respected under any situation resulting from the specific services to be provided.

A summary of the main RRM functions follows as:

2.2.2.1 Admision Control

Admission Control functions in the radio network controller must control requests for setup/reconfiguration of radio resources taking into account the following issues:

- How to rearrange already assigned resources to handle a request to setup/reconfiguration.
- Interworking with congestion control functions
- Different Admission Control behaviour depending on QoS requirements, user profile and load state.

Admission Control will be the RRM function to be covered in this work under different scenarios in both WCDMA and OFDMA systems. It is particularly to underline here the main role the scheduling functions play inside the Admission Control. In that respect, the schedulers are the entities enforced by the Admission Control to manage the order the packets coming from the different users are to be delivered to the radio channel.

2.2.2.2 Congestion Control

Congestion control mechanisms should be devised to face situations in which the system has reached a congestion status and therefore the QoS guarantees are at risk due to the evolution of system dynamics (mobility aspects, increase in interference, etc.). To this end, it takes care of the following issues:

- Design of criteria to decide that the system has reached the congestion status depending on different parameters such as performance and interference measurements.
- Development of proper algorithms to counteract congestion by suitably combining different mechanisms such as limiting TFCS (Transport Format Combination Set), inter-frequency handovers, etc.

2.2.2.3 Management of Transmission parameters

These mechanisms are devoted to decide the suitable radio transmission parameters for each connection. It includes:

- Transport Formats selection.
- Outer loop power control (E_b / N_0 target definition)
- Random access parameters management

2.2.2.4 Code management

This mechanism is responsible for the management of the limited number of codes available in downlink WCDMA

2.2.2.5 Handover management

Handover is the mechanism leading to keep the continuity in the user communication of mobile crossing different cells in a cellular system. In general terms the Handover Process consists of the three main phases: Measurements, Decision and Execution. The Decision Phase, which considers the Algorithm parameters and Handover criteria for soft handover, is that addressed initially in RRM.

2.2.3 Current and future trends

Nowadays, the Beyond 3G paradigm is already with us. The perspective of Beyond 3G systems, among others, is that of heterogeneous networks, where the multiplicity of access technologies as well as the diversity of terminals with reconfigurability capabilities will be key in order to allow users on the move to enjoy seamless wireless services irrespective of geographical location, speed of movement and time of the day. In addition to the need for a proper interworking among RATs, a new dimension into the radio resource management problem is introduced. That is, instead of performing the management of the radio resources independently for each RAT, some form of overall and global management of the pool of radio resources can be envisaged. Common Radio Resource Management (CRRM) is the envisaged process to manage dynamically the allocation and de-allocation of radio resources (e.g. time slots, codes, frequency carriers, etc.) within a single or between different radio access systems for the fixed spectrum bands allocated to each of these systems. With CRRM a more efficient usage of the radio resources will follow.

CRRM is aimed to jointly manage the specific RRM entities located on a particular RAT so as to obtain efficiency gains in the use of radio resources. When a new user is willing to gain a

network access to get some service from a network operator, CRRM will set the more convenient RAT satisfying all the user requirements stated in the user profile. The different available RATs will be hidden to the user that will only be concerned with perceived QoS and cost. The CRRM continues working along the session life of the service, and not just in the admission phase, so that it can be moved from one RAT to another according to CRRM decisions.

The Common Radio Resource Management server (CRRM Server) is proposed as a new logical node in UTRAN and GERAN. The purpose of the CRRM Server is to collect resource management functionality that can be related to multi-system functionality, so that this can be processed in the same algorithm (i.e. handover, load control, QoS control, etc) (Figure 2-6)

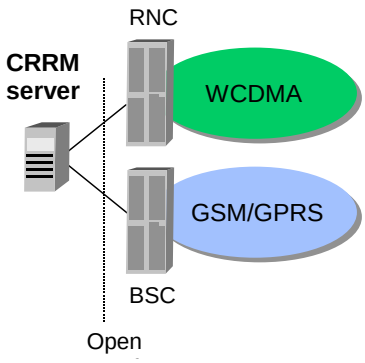

Figure 2-6 CRRM concept

The CRRM Server shall provide RRM Decision Support and ask for Reporting Information to the logical entities (i.e. RNC, BSC and other CRRM servers) that are connected to it, so that they can proceed being aware about what is happening in other Radio Access Technologies operating in the same coverage area. For this purpose, Cell Measurement Gathering and Prioritisation of the list of candidate cells of a UE for a specific operation (handover, network controlled cell reselection) are identified. Open interfaces are likely to be standardised.

In addition to the above, the traditional concept of static allocation of licensed spectrum resources to networks operators in wireless communications seems not to be most suitable approach in Beyond 3G scenario, characterized by changing traffic along time and space, changing availability of RATs, etc. In order to overcome these constraints and to achieve a better utilization of the scarce spectrum, Advanced Spectrum Management (ASM) techniques are envisaged: ASM enables to manage dynamically (allocation, de-allocation, sharing of spectrum blocks) within a single or between different radio access systems so that spectrum bands allocated to each of the systems are not fixed. On one hand, the Spectrum Brokerage (SB)

approach considers spectrum to be a tradable economic good similar to stocks or real estate, etc. Spectrum users place purchase orders providing information like maximum price and validity or expiry date of the order, not dissimilar to the procedures in a stock market. On the other hand, the actual management must consider also the technical issues like interference avoidance between spectrums not spatially or temporally separated, management of intersystem interference, definition of guard bands, etc.

With the arrival of 4G, new RRM strategies will appear, ASM above mentioned and the Spectrum Brokerage will be one of them and probably fully exploited by the use of the OFDMA strategies that allow in a natural and simple way moving pieces of spectrum. Certainly at this end new allocation rules will be put in place by the national regulatory agencies. Focusing actually in what is the scope of this thesis work, we intend to contribute in the developing of proper scheduling strategies leading to the satisfactory share of large amount of radio resources among a big variety of users requirements, what will be for sure a prime issue in the management of such 4G networks.

2.3 Basic RRM concepts for WCDMA

Research in multi-rate wideband code-division multiple access (WCDMA) networks for future wireless communication systems has recently attracted a lot of attention in the light of the new high bit rate data services that will be provided through 3G and 4G systems, including video, multimedia, circuit and packet mode data, together with the traditional speech services and low bit rate data. In such a scenario, it becomes crucial a proper management of the available radio resources in order to benefit from the traffic multiplexing to maximise the channel utilisation while at the same time keeping the specific service requirements. Specifically, in a WCDMA system, the amount of radio resources allocated to each user is controlled by regulating the transmission powers and the spreading gains (equivalently the transmission rates).

In the following the relevant RRM related concepts we have used in our work are described. We have been concerned with both WCDMA uplink and downlink. In particular in the following the two key RRM concepts consisting of the load factor as for the uplink and the Transmitted Power in relation to the downlink are introduced.

2.3.1 Uplink characterization: single cell case

Within a WCDMA cell, all users share the common bandwidth, W, and each new connection increases the interference level of other connections, affecting their quality expressed in terms of certain E_b / N_0. For n users transmitting simultaneously at a given isolated cell and in a given frame, the following inequality must be satisfied for each user [5]:

$$\frac{P_i \cdot \left(W / R_{b,i}\right)}{P_N + \left[P_R - P_i\right]} \geq \left(\frac{E_b}{N_0}\right)_i \qquad i = 1, \ldots, n \ , \tag{2.5}$$

$$P_R = \sum_{i=1}^{n} P_i \ , \tag{2.6}$$

where P_i is the i-th user received power at the base station, $R_{b,i}$ is the i-th user bit rate, P_N is the thermal noise power and $\left(E_b / N_0\right)_i$ stands for the i-th user requirement. P_R is the total received own-cell power at the base station. Implicitly in the above inequalities a certain received power level is assumed in each case:

$$P_i \geq \frac{P_N + P_R}{\dfrac{(W / R_{b,i})}{\left(\dfrac{E_b}{N_0}\right)_i} + 1} \qquad i = 1, \ldots, n \ . \tag{2.7}$$

Adding all n inequalities, it holds:

$$\sum_{i=1}^{n} P_i = P_R \geq \sum_{i=1}^{n} \frac{P_N + P_R}{\dfrac{(W / R_{b,i})}{\left(\dfrac{E_b}{N_0}\right)_i} + 1} \ . \tag{2.8}$$

Then, P_N can be isolated in Equation (2.8), assuming equality, as:

$$P_R = \frac{P_N \cdot \sum_{i=1}^{n} \frac{1}{\frac{(W/R_{b,i})}{\left(\frac{E_b}{N_0}\right)_i}+1}}{1-\sum_{i=1}^{n} \frac{1}{\frac{(W/R_{b,i})}{\left(\frac{E_b}{N_0}\right)_i}+1}}. \quad (2.9)$$

Substituting P_R in Equation (2.7) gives:

$$P_i = \frac{P_N}{\left[\frac{(W/R_{b,i})}{\left(\frac{E_b}{N_0}\right)_i}+1\right] \times \left[1-\sum_{i=1}^{n} \frac{1}{\frac{(W/R_{b,i})}{\left(\frac{E_b}{N_0}\right)_i}+1}\right]}. \quad (2.10)$$

Claiming in (2.10) for the inherent positivity of P_R (i.e. $P_R > 0$) and/or P_N (i.e. $P_R > 0$) leads to:

$$\sum_{i=1}^{n} \frac{1}{\frac{(W/R_{b,i})_i}{\left(\frac{E_b}{N_o}\right)_i}+1} < 1. \quad (2.11)$$

If all users in the cell are of the same type, the above condition clearly indicates that the uplink WCDMA capacity is limited to:

$$n < 1 + \frac{(W/R_{b,i})}{\left(\frac{E_b}{N_0}\right)_i}. \quad (2.12)$$

Thus, the need of controlling the number of transmissions on the air interface is clearly stressed by Equation (2.12). If Equation (2.12) does not hold, it means that it is not possible for all the n simultaneous users to achieve their respective E_b / N_0 targets simultaneously. A key point in

the above formula is the existing dependency between the number of transmissions and the bandwidth allocated to each user when transmitting. In particular this point will be relevant in our work.

2.3.1.1 Camping, active and simultaneous users

From expressions (2.6) and (2.7) readily follows that the number of simultaneous users transmitting, n, plays a key role in the radio interface management since it mainly defines the interference level and must be controlled in order to keep the required transmitted power levels bounded. Nevertheless, the number of simultaneous users transmitting exhibits complex dependences. In order to explore these dependences, let consider an area where a cellular system is deployed. Let consider M subscribers in the area generating calls (or sessions) at an average rate per user of λ calls/s, transmission rate of $R_{b,i}$ and aggregate rate of $M \cdot \lambda$ calls/s. Let also assume that the average call duration is $1/\mu$ s and that N is the number of active users in a session call. Then, we can represent the situation by a continuous time birth-death Markov chain, the states being the number of users N in a call. The birth coefficient and death coefficient in the N -th state are given by:

$$\lambda_N = (M - N)\lambda, \tag{2.13}$$

$$\mu_N = N\mu. \tag{2.14}$$

If the traffic source in a call is characterized by an activity factor α, representing the percentage of time that the source is generating traffic, then the number of simultaneous users occupying the radio interface provided that N users are in a call can be represented as:

$$f_n(n) = \binom{N}{n} \alpha^n (1-\alpha)^{N-n} \tag{2.15}$$

Although the methods of analysis are complex and require several hypotheses in order to become tractable, the above equations are useful in the senses that do provide insight into the basic behavior. In particular, three different concepts and a framework for their relationships have arisen: M, the number of users camping on a cell, N, the number of active users in a session and n, the number of users transmitting in a given moment. In this work we will assume that the activity factor is set to 1, so as only continuous transmission within a session will occur. Then $n = N$.

M depends on the population density in the area of the cell as well as on the service penetration. The long term variation of M indicates that this parameter is mainly controlled through a suitable radio network planning, so that the network deployed (i.e. the number of cells and their configuration) is such that the resulting M camping on a cell's planned coverage area is manageable with the amount of radio resources devoted to that cell.

N depends on one hand on M and on the other hand on the call birth/death process related to the user's behavior with respect to service usage. For a given M, the independence among the different call generation processes of the different subscribers makes N to be variable. The Admission Control will be in charge of controlling N, the number of users accepted on the cell for a connection. The main performance figure related to admission is the blocking probability (i.e. the probability that a user makes a call attempt and is not allowed to establish the connection).

Admission Control would allow absorbing deviations on N occurring either by deviations on M with respect to the expected level or by deviations on the user's call generation process from the expected behavior (e.g. more calls attempts than expected, calls longer than expected, etc.). In such situations, the resulting blocking probability would be higher than planned.

For a given N, the number of simultaneous users depends on their activity factor and on the traffic multiplexing process, so that n results to be variable. In case of deviations on n, the congestion control mechanism needs to be activated in order to preserve the performance of the different connections. The variations on n would occur on the short term, in the order of frames (i.e. in the order of tenths of milliseconds for UTRA), so that the congestion control mechanism will be coupled at some extent with the bit rate control mechanism (notice that preventing a user's transmission is equivalent to set its transmission rate to 0 bit/s).

All the above issues in turn are impacted by the user bit rate variations that unlike the more oriented voice systems are going to have a key role in the multimedia 3G systems scenarios.

2.3.1.2 Capacity and coverage; Breathing effect

Capacity and coverage are closely related in WCDMA networks, and therefore both must be considered simultaneously. The coverage problem is directly related to the power availability, so that the power demands deriving from the system load level should be in accordance with the

planned coverage. So, it must be satisfied that the required transmitted power will be lower than the maximum power, $P_{T\max}$, allowed and high enough to be able to get the required $\left(E_b / N_0\right)$ target even at the cell edge. Then the user-i th transmited power follows from Equation (2.10) as

$$P_{T,i} = \frac{L_{p,i} P_N}{\left[\frac{(W / R_{b,i})}{\left(\frac{E_b}{N_o}\right)_i} + 1\right] \times \left[1 - \sum_{i=1}^{n} \frac{1}{\frac{(W / R_{b,i})}{\left(\frac{E_b}{N_o}\right)_i} + 1}\right]} \qquad i = 1, \ldots, n, \tag{2.16}$$

where $L_{p,i}$ stands for the propagation losses appearing in the link base station–user i-th.

The i-th user is said to be in outage if the requires transmitted power given by Equation (2.16) is higher than $P_{T\max}$, which is equivalent to say that the i-th user observes at the cell site an E_b / N_0 below its target value. If this situation persists, the i-th user is said to be out of coverage and experiences no service availability. It is worthy to be noticed that increasing the number of simultaneous users transmitting, n, or the user bit rate, results in a decrease of the coverage area. That is a very important issue in the development of RRM strategies and also known as breathing effect. That is, all happens as the coverage area increases or decreases (it breaths) depending on the interference load.

2.3.1.3 Load factor

A relevant parameter that will be used somehow in the RRM algorithms is the load factor that will be described in the following [6]

The uplink $\left(E_b / N_0\right)$ value can be formulated as

$$\left(\frac{E_b}{N_0}\right)_i = \frac{W}{R_i} \frac{P_i}{I_{tot} - P_i}, \tag{2.17}$$

where P_i is the power received by the i-th user, I_{tot} is the total power received at the Base station and R_i is the rate allocated to the i-th user. W is the bandwidth available.

Then, from Equation (2.17) it follows

$$P_i = \frac{1}{1+\frac{W}{\left(\frac{E_b}{N_0}\right)_i R_i}} I_{tot} . \tag{2.18}$$

Defining

$$P_i = L_i I_{tot} \tag{2.19}$$

it results

$$L_i = \frac{1}{1+\frac{W}{\left(\frac{E_b}{N_0}\right)_i R_i}} . \tag{2.20}$$

The noise raise originated by the presence of interference is defined as

$$\Delta N = \frac{I_{tot} / W}{P_N / W} . \tag{2.21}$$

Let consider

$$I_{tot} - P_N = \sum_{i=1}^{N} \nu_i P_i = I_{tot} \sum_{i=1}^{N} \nu_i L_i , \tag{2.22}$$

where N is the number of users in the cell and ν_j is the activity factor of the i-th user.

From Equations (2.21) and (2.22), it follows

$$\Delta N = \frac{1}{1-\sum_{i=1}^{N} \nu_i L_i} = \frac{1}{1-\eta_{UL}}, \tag{2.23}$$

where P_N is the the thermal noise power and η_{UL} is the so called load factor.

Finally, from Equations (2.20) and (2.23) it follows

$$\eta_{UL} = \sum_{i=1}^{N} \nu_i L_i = \sum_{i=1}^{N} \nu_i \frac{1}{1+\frac{W}{\left(\frac{E_b}{N_0}\right)_i R_i}}. \tag{2.24}$$

The load factor plays a very important role in the planning process of WCDMA. Due to the statistical nature of this factor originated mainly by the uncertainties in the random nature of the sources signals, captured in the above expression through the activity factor. In practice there is a maximum value assigned to the load factor, η_{max}, that will be used as a data in our work.

2.3.2 Downlink characterization: single cell case

2.3.2.1 Power Allocation

Within a WCDMA cell, all users share the common bandwidth and each new connection increases the interference level of other connections, affecting their quality expressed in terms of a certain E_b / N_0. For n users receiving simultaneously from a given cell, the following inequality for the i-th user must be satisfied [7]:

$$\frac{\frac{P_{T_i}}{L_p(d_i)} \times \frac{W}{R_{b,i}}}{P_N + \rho \times \left[\frac{P_T - P_{T_i}}{L_p(d_i)}\right]} \geq \left(\frac{E_b}{N_o}\right)_i, \tag{2.25}$$

$$P_T = P_p + \sum_{i=1}^{n} P_{T_i}, \tag{2.26}$$

P_T being the base station transmitted power, P_{T_i} being the power devoted to the i-th user, $L_{p,i}$ being the path loss at distance d_i (including shadowing), $R_{b,i}$ the i-th user transmission rate, W the bandwidth, P_p the power devoted to common control channels and P_N the background noise. ρ is the orthogonality factor since some orthogonality is lost due to multipath. Additionally, physical limitations into the power levels are given by the maximum base station transmitted power, $P_{T_{\max}}$. Then P_{T_i} can be isolated in Equation (2.25), assuming equality, and substituting in (2.26), it can be obtained that the total transmitted power to satisfy all the users' demands should be:

$$P_T = \frac{P_p + \sum_{i=1}^{n} \frac{P_N}{\frac{W / R_{b,i}}{\left(\frac{E_b}{N_o}\right)_i} + \rho} L_p(d_i)}{1 - \sum_{i=1}^{n} \frac{\rho}{\frac{W / R_{b,i}}{\left(\frac{E_b}{N_o}\right)_i} + \rho}} \leq P_{T_{\max}} \tag{2.27}$$

and the power devoted to the i-th user, P_{T_i}, is given by:

$$P_{Ti} = L_p(d_i) \frac{P_N + \rho \times \frac{P_{T_{\max}}}{L_p(d_i)}}{\frac{W / R_{b,i}}{\left(\frac{E_b}{N_o}\right)_i} + \rho}. \tag{2.28}$$

Besides, it is reasonable to put some limits on the maximum power to be devoted to a single connection, otherwise high demanding users could retrieve from service to a number of users sharing the same cell downlink power level. Then, the following restriction will be observed along the connection dynamics:

$$P_{Ti} \leq P_{c,\max}. \tag{2.29}$$

In our work, some assumptions will be carried out in order to retain the main results and make the problem manageable in a first approach. That is, it will be assumed that

$$P_{Ti} \geq L_p(d_i)\frac{P_N}{\frac{W/R_{b,i}}{\left(\frac{E_b}{N_o}\right)_i}} \tag{2.30}$$

where it has been assumed that the orthogonally factor is near to zero.

The load factor and transmitted power are two key parameters to be considered when dealing with WCDMA Admission Control in uplink and downlink respectively and will be retained later in this work when considering WCDMA Admission Control in packet based services.

2.4 Basic RRM concepts for OFDMA

Orthogonal Frequency Division Multiple Access has been selected as the proper technology for achieving flexible radio interfaces in WiMAX and LTE as key RAT in next 4G of Mobile Communications. As already mentioned, the radio resources in OFDMA RATs are divided in both time and frequency building a time-frequency grid, [8], where the minimum radio resource that can be assigned to a user is usually named as a Resource Block (RB). In frequency, the whole available bandwidth is divided into groups of adjacent subcarriers or chunks whereas in time it is divided into frames.

Furthermore, a flexible chunk (or groups of a set of contiguous subcarriers) allocation mechanism that adapts to users' requirements is also desirable. Conventional frequency reuse schemes (e.g., reuse 1 where all chunks are assigned to all cells or reuse 3 where all chunks are distributed regularly between groups of 3 cells) are static since the allocation of frequency resources to cells is fixed and regular and cannot be changed on-line. Dynamic Spectrum Allocation (DSA) algorithms that consider the heterogeneous spatial traffic distributions to decide dynamically a proper chunk-to-cell assignment are also currently under study and constitute one of the new RRM key issues in OFDMA and strongly impacts its Admission Control mechanisms at the long term. On the short term, the OFDMA RAT, intracell interference is usually avoided by means of scheduling schemes that only allow that a Resource Block can be assigned to a single user within a given cell at a time. On the other hand, intercell interference cannot be easily avoided with scheduling due to traffic variations among cells and the lack of coordination between the schedulers in them. Hence, in order to combat intercell interference different Frequency Reuse Factors (FRF) are deployed off-line in the network within the planning phase, distributing the total bandwidth of the system among different cells.

The simplest FRF scheme is the total reuse of the system bandwidth among cells (i.e., all the chunks are available at any cell) denoted in this work as FRF1 ($FRF = 1$).This scheme maximizes network capacity but at the cost of higher BER and lower throughput especially for the users at the edge of the cell that experiment high interference because all neighboring cells are reusing the same frequency. One possibility to mitigate this interference for the users at the edge of the cell is increasing the reuse factor. Applying a FRFM ($M \geq 1$) implies that the total bandwidth is divided into M equal sub-bands, and distributed among groups of M contiguous cells (or clusters). This scheme reduces the intercell interference at the cost of reducing the cell capacity as well. Typically, M takes a value equal to 3 ($FRF = 3$). There are also other reuse schemes with FRF between 1 and 3 that have been recently proposed (see Figure 2-7, extracted from [9]). They are based on the division of the frequency band and the users between central and edge sets.

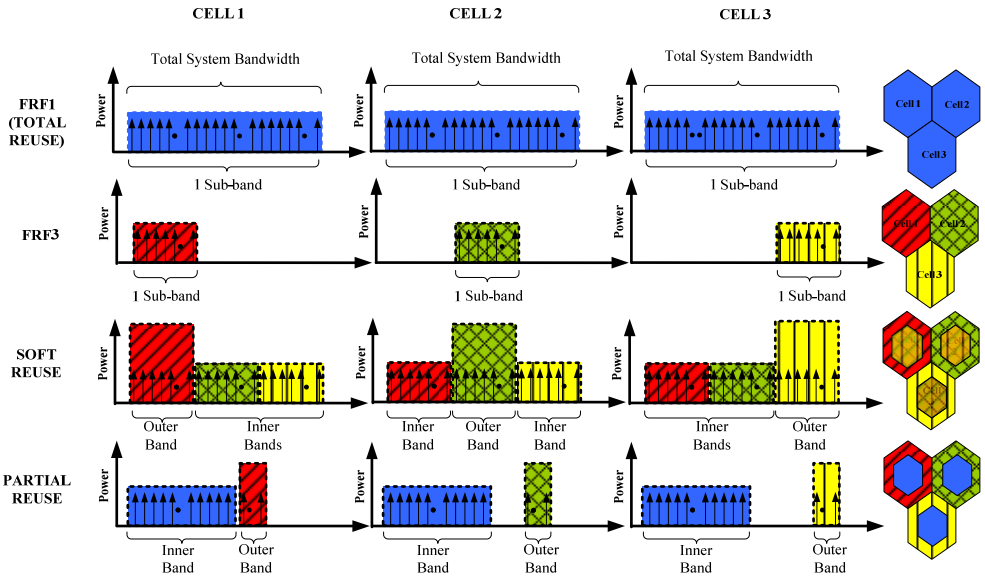

Figure 2-7 Reuse scheme based on the frequency band and the users between central and edge sets.

Actually the OFDMA scheduling framework is divided into two functional decision blocks. One provides the dynamic chunk-to-cell allocation in the system in the medium-long term. The other one is called Short-Term Scheduler (STS) and tries to assign the allocated chunks or subcarriers at the station level to the different users trying to keep at the same time the QoS service requirements and the optimum explotation of the OFDMA system in terms of spectral efficiency, fairness, maximum throughput, etc.

Different policies can be followed by the above STS. For example, the users can be scheduled in function of the channel quality, the buffer delay, the throughput, the buffer occupancy, the

service, etc. Thus, the STS is traffic and channel aware in its decisions, and tries (frame by frame) to exploit the multiuser diversity by making use of the available resources at each cell. Several scheduling schemes could be developed and constitute a hot research topic nowadays. Round Robin is often employed, but its performance is poor since it assigns the channel cyclically to users and then is not channel aware in its decisions. Proportional Fair (PF) and Generalized Proportional Fair (GPF) have been proposed to provide a trade-off between fairness and system throughput and to exploit multiuser diversity in time. GPF is a generalized version of PF where weights are introduced to change the trade-off between fairness and throughput. Then, not only multiuser diversity is exploited in time but also in frequency. The allocation of OFDMA resource blocks to users when delay sensitive services are to be satisfied is a particular open problem this thesis work is going to face and contribute. In this work, and as a first step, we will assume just one cell with a given number of already allocated subcarriers and we will be interested in identifiying new short term schedulers to map resources to users in due time.

Chapter 3

An introduction to Hopfield Neural Networks

3.1 Introduction

This chapter is an approach to Artificial Neural Networks (ANN) based on references [10] and [11]. ANNs are systems that consist of many interconnected processing units or nodes working together imitating the human nervous system. In the first following sub-sections we expose basic general concepts. The last sub-sections contain more specific results to be applied to our work.

The chapter is organized as follows. Section 3.2 is an overview of ANN. It describes the origins, basic neural networks concepts and some historical notes. There are many types of ANNs and they are applied to many research areas. The section contains also a brief description of the most significative architectures and summary of their applications. The section concludes introducing the mathematical model of a node. Section 3.3 is focused on one nonlinear dynamical system, the so-called Hopfield Neural Network (HNN), which is the model used in this work. It describes the mathematical model of that dynamical system. Some issues of stability that arise in the model will be also introduced to show its robustness. This section summarizes also the most relevant results from classical papers to analyze how HNNs work when they are applied to optimization problems.

HNN can be modelled by either numerical simulations or hardware implementations. Since early stages of a HNN model development require a loop of analysis-changes to obtain a well-working and prototype model, numerical simulations are the most common used techniques in front hardware implementations. Section 3.4 contains details for numerical simulation which is the one used in our work. This section includes also some references for hardware implementations, but this technique is beyond the scope of our work.

Section 3.5, the last one, describes the particular notation of HNN models we have used in this thesis. Because of this notation they are often referred to as 2D HNN. Futhermore, this section exposes and analyzes which of the main characteristics described in previous sections have been adapted and it will be the basis of the 2D HNN that we have developped to optimize the problems presented in our work.

3.2 An overview of Neural Networks

3.2.1 Origin and neural networks concept

The origins of neural networks are based on efforts to model information processing in biological systems as the human brain. Human brain contains many billions of very special kind of cells or neurons organized into a very complicated intercommunicating network. It is estimated to contain approximately 10^{11} neurons, each connected, on average, to 10^{4} others. Using these connections neurons can pass electrical signals between each other. The fastest neuron switching times are known to be on the order of 10^{-3} seconds, quite slow compared to computer switching speeds of 10^{-10} seconds. Yet humans are able to make surprisingly complex decisions, surprisingly quickly. It is this simultaneous cooperative behavior of many very simple processing units which is at the root of the enormous sophistication and computational power of the brain. This observation has led many to speculate that the information-processing abilities of biological neural systems must follow from highly parallel processes operating on representations that are distributed over many neurons.

An *Artificial Neural Network* (ANN), often just called a *Neural Network* (NN), is an interconnected group of artificial neurons, simple processing elements called *artificial neurons*, *nodes* or *units*. The term *node* will be preferably used in this work. Nodes are wired together in a complex communication network, akin to the vast network of neurons in the human brain. The units are connected by connection strengths known as *synaptic weights* and operate only on their local data and on the inputs they receive via the connections. An *activation rule* acts on the set of input signals at a unit to produce a new output signal. Thus the internal characteristics of the network are modified, sometimes by the network itself, until a satisfactory performance level is attained. Therefore neural networks are adaptive systems that change its structure based on external or internal information that flows through the network.

Optionally, there is a *training process* and a *learning process* that specify how signals adjust the weights for a given input/output pair. Investigation of learning and training approaches forms an important facet of work on neural networks.

A neural network is usually implemented by using electronic components or is simulated in software on a digital computer. Neural Networks are also referred to in literature as *neurocomputers*, *connective networks*, *parallel distributed processors*, etc.

The power and usefulness of ANNs have been demonstrated to a wide variety of different areas including speech synthesis, diagnostic problems, medicine, business and finance, robotic control, signal processing, computer vision, optimisation, complex mapping and many other problems that fall under the category of pattern recognition. An extensive list of applications can be found in [12].

3.2.2 Historical introduction

This sub-section is a summary of the most important events in the history of neural networks. The most relevant contributions to its developpement are referenced for further information to the reader.

The pioneer work in neural network was a paper written by the neurophysiologist Warren McCulloch and the mathematician Walter Pitss in 1943, [13]. This paper describes how neurons in the brain might work modelling a simple neural network that uses electrical circuits. They assumed that each neuron was binary and had a finite threshold, that each synapse was either excitatory or inhibitory and caused a finite delay (of 1 cycle), and that artificial networks could be constructed with multiple synapses between any pair of nodes. In order to show that any logical expression is computable, all that is necessary is to build the functions AND, OR and NOT and, in order to build larger functions, one needs only to glue these primitive together. Such artificial networks are kwown as McCulloch-Pitts netwoks.

In 1949, an important contribution on learning rules of the brain came from the Donald Hebb's book, *The Organization of Behavior*, in which pointed out two fonamental concepts for the future developpement of neural networks: memory is stored in connections and learning takes place by synaptic modification. Therefore, the connectivity of the brain is continually changing as an organism learns differing functional tasks, and neural assemblies are created by such

changes. He was able to piece together a theory known as Hebbian theory and the models which follow this theory are said to exhibit Hebbian learning..

The first computer simulation to test a well-formulated artificial neural network theory was made in IBM 701 and 704, in 1956, by Rochester, Holland, Haibt and Duda from the IBM research laboratories [14]. But the technology available at that time did not allow them studying complexes models.

In the 1950-60s other interesting contributions were done by Ashby's book, *Design for a Brain: The Origin of Adaptative Behavior*; Minsky's thesis, *Theory of Neural-Analog Reinforcement Systems and Its Application to the Brain-Model Problem*; Taylor's works on *associative memory* [15] and Steinbuch, with the introduction of learning matrix [16]. John Von Neumann with his unfinished manuscript, published after his death as a book, *The computer and the Brain*, extended the McCulloch-Pitts model. He thought of imitating simplistic neuron functions by using telegraph relays or vacuum tubes. This led to the invention of the Von Neumann machine. It was also in this period that Bernard Widrow and Marcian Hoff of Stanford University published a paper, [17], in which they proposed their ADALINE circuits (ADAptative LINEear neuron). In a McCulloch-Pitts network they used a new algorithm to recognize binary patterns and predict the next bit in a flowing stream of bits. This new approach was very usefull in the learning concept process. Widrow-Hoff network was the first ANN applied to a real world problem, using an adaptive filter that eliminates echoes on phone lines (the system is as ancient as air traffic control systems, it is still in commercial use). A more extensive list of works with a brief description can be found in [10].

In 1958 Frank Rosenblatt , [18], initiated a new phase in neural network research introducing the basic concept of a *perceptron* which computes a single *output* from multiple real-valued *inputs* by forming a linear combination according to its input *weights* and then possibly putting the output through some nonlinear activation function. The original Perceptron received two inputs, and gave a single output. In this model neurons could have continuous signals in inputs and outputs in front of the binary values that need to have McCulloch-Pitss. Rosemblatt demonstrated that the perceptron could learn any classification or relation between input-output in a limited number of steps if hardware system was able to produce this classification.

Although the perceptron model worked well for simple problems, Marvin Minsky and Symour Papert, published in 1969 the book *Perceptrons, An Introduction to Computational Geometry*,

in which they provided mathematical proofs of the limitations of the perceptron and pointed out its weakness in computation. In particular, it is incapable of solving the classic exclusive-or (XOR) logic problem. This problem was later resolved by the so-called hidden neurons. Minsky and Papert showed other weaknesses in the neural network. Especially the popular training of the networks was a problem. If the number of inputs increased, the time to train the network would increase exponentially, and thereby set a limit of the efficiency of the network. This work jointly with technological limitations for experimentation on computers stopped the wave of research of the neural network. Such drawbacks led to the temporary decline of the field of neural networks.

In the 1980s major contributions to the theory and design of neural networks were made on several fronts. The development of more-powerful networks, better training algorithms, and improved hardware has all contributed to resurgence of interest in neural networks.

In 1982 John Hopfield published a scientific paper, [19], which attracted much interest. He stated that the approach to ANN should not be to purely imitate the human brain but instead to use its concepts to build machines that could solve dynamic problems. He showed what such networks were capable of and how they would work. It was his articulate, likeable character and his vast knowledge of mathematical analysis that convinced scientists and researchers at the National Academy of Sciences to renew interest into the research of artificial neural networks. His ideas gave birth to a new class of neural networks that over time became known as the *Hopfield Neural Networks* (HNN) and was the starting point of the new era of neural networks, which continues today. Hopfield used the idea of an energy function to formulate a new way of understanding the computation performed by recurrent networks with symmetric synaptic connections between nodes and for understanding their dynamic behavior. Hopfield networks always evolve in the direction that leads to lowest network energy. This implies that if a combinatorial optimization problem can be formulated as minimizing this energy, the network can be used to find the optimal (or suboptimal) solution by letting the the network evolve freely. These HNN will be described in section 3.3 as they are an important part of our research. Such systems could be implemented in hardware by combining standard components such as capacitors and resistors. Since then, new versions of the Hopfield network have been developed. We remark the extension made in 1983 by G.E. Hinto and T.J. Sejnowski, [20], in which they added probabilistic theory to a Hopfield network. Each node is a stochastic unit that generates an output according to the Boltzmann statistical distribution. These networks are named Boltzmann machines.

In 1986, Rumelhart and McClelland published the two volume book *Parallel Distributed Processing: Explorations in the Microstructures of Cognition* which still is regarded as a bible for cognitive scientists. They established a novel algorithm for learning role based on back-propagation learning and has been a major influence in the use of back-propagation learning, which has emerged as the most popular learning algorithm for the training of multilayer perceptrons.

Neural network paradigms in recent years include: the Adaptive Resonance Theory (ART) model, first introduced by Carpenter and Grossberg (1983, [21]) and becoming the advanced ART II and ART III network models; the Learning Vector Quantization (LVQ), invented by Teuvo Kohonen; the Self-Organizing Map (SOM) model introduced by Kohonen (1989, [22]) , the Fukushima's model (1988, [23]), the Radial Basis Function (RBF) network (1988, [24]) and others. Figure 3-1 shows the evolution of the most popular neural networks. The field has generated interest from researchers in such diverse areas as engineering, computer science, psychology, neuroscience, physics, and mathematics.

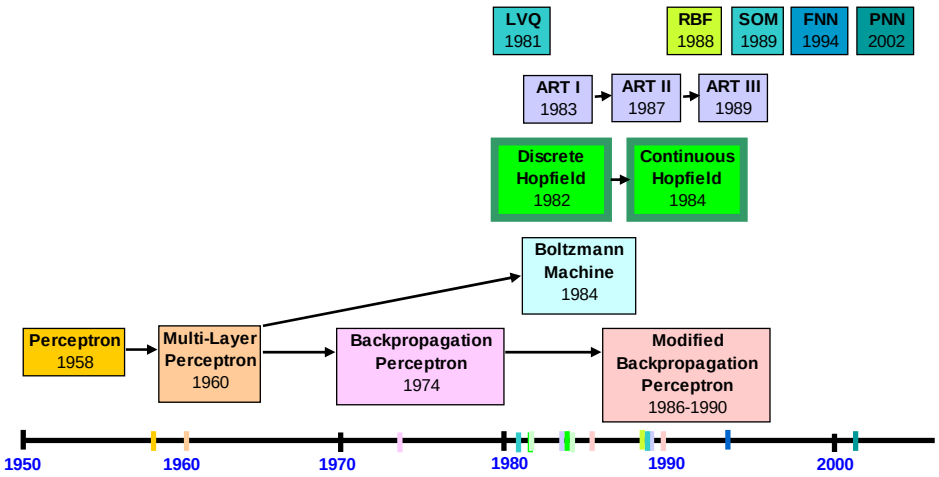

Figure 3-1 Evolution of the most popular Artificial Neural Networks

With the dawn of the 1990's and the technological era, many advances into the research and development of artificial neural networks are occurring all over the world. Nature itself is living proof that neural networks work. The challenge today lies in finding ways to electronically implement the principles of neural network technology. Electronics companies are working on three types of neuro-chips namely, digital, analog and optical. With the prospect that these chips may be implemented in neural network design, the future of neural network technology looks very promising [25]

3.2.3 Structure of neural networks

The behaviour of a neural network depends largely on the interaction between the nodes, the nature of its connections, the network size (number of nodes) and the algorithm running on them. There are many many kinds of NNs by now and there is not only one factor to classify them. In the following we will describe the major factors to be considered, not claiming to be complete.

Nodes in networks are grouped in layers. We can distinguish single-layer and multi-layer architectures. In a single-layer organisation all nodes are connected to one another, in multi-layer networks, nodes are often numbered by layer, instead of following a global numbering and the connections are between nodes from different layers. Multi-layer networks can have three types of node layers. Figure 3-2 shows the most common type of artificial neural network consisting of three groups, or layers, of units: an *input layer* a *hidden layer* and an *output layer*.

An input layer, in which the activity of nodes represents the raw information that is fed into the network. The number of nodes in this layer depends on the number of possible inputs we have. Hidden layers, in which the function of the nodes is to process information and to communicate between others layers. The number of hidden layers and how many nodes in each of them are characteristics of networks. The third type is the output layer, in which the behaviour of the output units depends on the activity of the hidden units and the weights between the hidden and output units. The number of nodes in this layer depends on the number of desired outputs. Rosenblatt's original perceptron model is a multi-layer network containing only the input and the output layer. Hopfield network is a single-layer network, it consists of only one layer whose neurons are fully connected with each other.

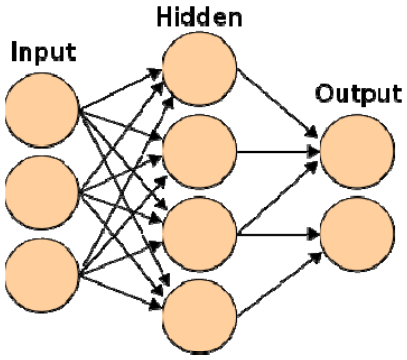

Figure 3-2 An Artificial Neural Network.

A characteristic that distinguishes multi-layer networks is the pattern of weighted connections between the units from different layers and the propagation of data. As for this pattern of connections, the main distinctions that can be made are the following:

Feedforward neural networks, in wich the information moves in only one direction, forward, from the input nodes, through the hidden nodes and to the output nodes. There are no cycles nor loops in the network. Classical examples of feedforward neural networks are the Perceptron and the Adaline.

Feedback neural networks, in which the output of one layer can be routed to a previous layer introducing loops in the network. Their state is changing continuously until they reach an equilibrium point. Feedback architectures are very powerful and can get extremely complicated. Sometimes in literature they are also referred to as interactive or recurrent, although the latter term is often used to denote feedback connections in single-layer organisations or internal state space feedback as it is described in the paragraph below.

Recurrent neural networks have been developed to deal with the time varying or time-lagged patterns and are usable for the dynamic problems. They are different from above structures in the sense that they not only operate on an input space but also on an internal state space, i.e. a context layer added to the structure which retains information between observations of what has been already processed by the network. Figure 3-3 shows the flow through recurrent neural network architecture. Inputs are received from an external source, passed to a hidden layer, and then on to the output layer. The signal from the output layer is passed to the context layer that acts as a new input layer to the hidden layer on the next pass. The network's response can be stable (successive iterations produce smaller and smaller differences between values in the output layer and values in the context layer until the outputs become constant) or unstable (the network's outputs never cease changing). This stability issue proved a problem for early researchers, but Cohen and Grossberg [26] devised a theorem showing that at least a subset of recurrent networks were guaranteed to produce outputs with stable states. Much of the early work on recurrent networks was pioneered by John Hopfield.

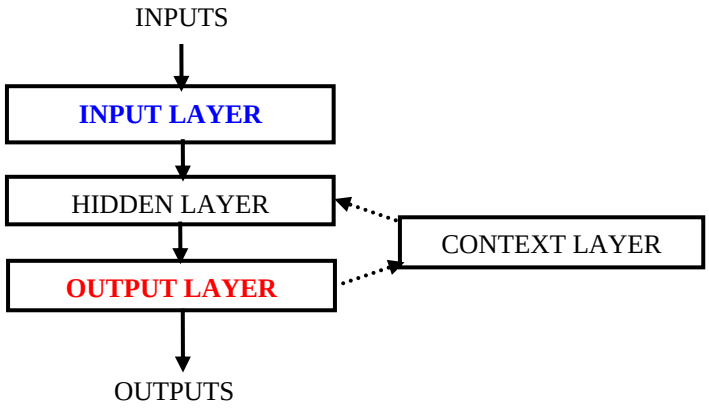

Figure 3-3 Recurrent neural network architecture.

One of the most important aspects of neural networks is the learning process. It is said that a node is *learning* when its weight is updated. The process through which neural networks learn is named the *training* process. Training a net can be seen as a search for a goal point in gigantic n-dimensional space of points defined by the weight vectors. The point is the set of weights which minimizes some error function. *Learning algorithms* are the rules for training neural networks. They define how connection weights must be adjusted in order to fit a series of inputs to another series of known outputs. There are many algorithms for training neural networks; most of them can be viewed as a straightforward application of optimization theory and statistical estimation.

In terms of their learning features neural networks can be classified as neural networks with *supervised* learning algorithms, neural networks with *unsupervised* learning algorithms and neural networks with *reinforcement learning*. Usually any given type of network architecture can be employed in these algorithms. This classification is not unique and different research groups make different classifications.

In supervised learning, both the inputs and the outputs are provided. The network then processes the inputs and compares its resulting outputs against the desired outputs. An output error is then calculated, causing the system to adjust the weights by some algorithm. The greater the computed error value is, the more the weight values will be changed. This process occurs over and over as the weights are continuously tweaked. The goal of this is to minimize the output error with respect to the given inputs. Supervised learning is fairly common in classification problems and digit recognition.

In unsupervised learning, the network is provided with inputs but not with desired outputs. The system itself has to adapt its parameters autonomously. It can't be determined what the result of the learning process will look like. During the learning process, the units (weight values) of such a neural net are arranged inside a certain range, depending on given input values. The goal is to group similar units close together in certain areas of the value range. This effect can be used efficiently for pattern classification purposes.

The third type, called the reinforcement learning, bridges a gap between supervised and unsupervised categories. In reinforcement learning, the learner does not explicitly know the input-output instances, but it receives some form of feedback from its environment. The feedback signals help the learner to decide whether its action on the environment is rewarding or punishable. The learner thus adapts its parameters based on the states (rewarding / punishable) of its actions.

Application areas of ANNs depend on their structures. Some of them are more appropriate for a specific application than others. Figure 3-4 shows a list of relevant application areas for some of the ANN we have mentioned.

	PERCEPTRON	ART	LVQ	RBF	SOM	FNN	PNN	BOLTZMANN	HOPFIELD
Autoassociation									
Pattern recognition									
Image processing									
Decision making									
Data compression									
Filtering									
Optimization									

Figure 3-4. Relevant application areas for some neural networks

3.2.4 Mathematical representation of a node.

The mathematical model of a node used in this work is described in [10]. Figure 3-5 shows the notation used in this work for modeling a node k, denoted as u_k. Let $x_1, x_2, \ldots, x_S$ be S incoming connections in it. When a pulse x_j comes from a connection, it is first multiplied by a number, w_{kj}, called the *weight* or *strength* of the connection, which assigns a certain importance to that connection, and then all the overall results are accumulated. The accumulated value passes through an *activation function* which emits a pulse when a certain value is reached. The output of the activation function stage, v_k, is in turn connected to the inputs of several

other nodes, which forms a complete network. The neural model can also include an externally applied *bias*, denoted I_k, that has the effect of increasing or lowering the net input of the activation function.

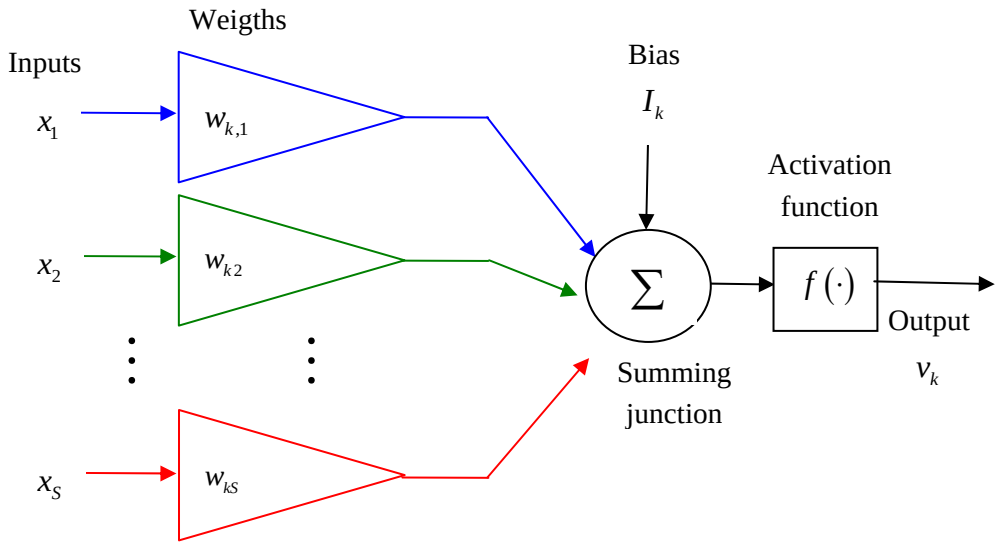

Figure 3-5 Model of a node

In mathematical terms a node k can be described by the following pair of equations:

$$y_k = \sum_{j=1}^{S} w_{kj} x_j \tag{3.1}$$

and

$$v_k = f\left(y_k + I_k\right). \tag{3.2}$$

By denoting

$$u_k = y_k + I_k \tag{3.3}$$

we can add a new input x_0 with associated weigh $w_{k0} = I_k$ and rewrite (3.1) and (3.2) respectively as

$$\begin{aligned} u_k &= \sum_{j=0}^{S} w_{kj} x_j \\ x_0 &= 1 \end{aligned} \tag{3.4}$$

and

$$v_k = f(u_k). \tag{3.5}$$

The activation function denoted by $f(u)$, acts as a squashing function, such that the output of a neuron in a neural network is kept between certain values (usually 0 and 1, or -1 and 1) and is usually defined by one of the following mathematical models:

1. *Threshold Function* or *Heaviside Functions* (Figure 3-6).

Defined by

$$f(u) = \begin{cases} 1 & \text{if} \quad u \geq 0 \\ 0 & \text{if} \quad u < 0 \end{cases} \tag{3.6}$$

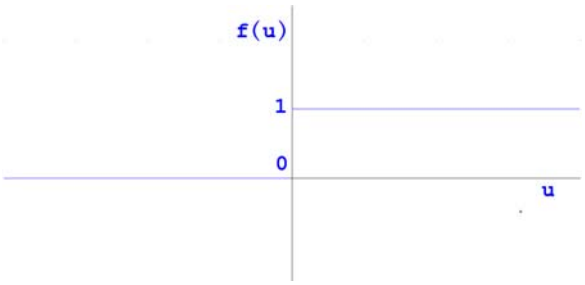

Figure 3-6 Heaviside function

2. *Piecewise-Linear Function*

Defined by

$$f(u) = \begin{cases} 1 & \text{if} \quad u \geq 0 \\ u & \text{if} \quad 0 < u < 1 \\ 0 & \text{if} \quad u \leq 1 \end{cases} \tag{3.7}$$

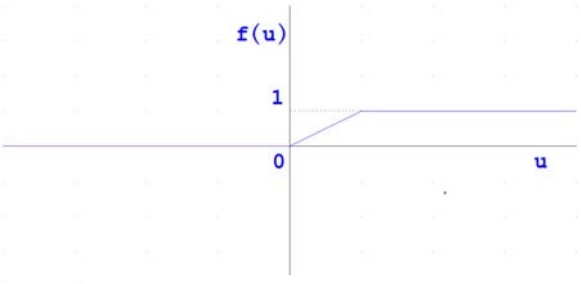

Figure 3- 7 Piecewise Linear function

3. *Sigmoid Function* or *Logistic Function*

Defined by

$$f(u) = \frac{1}{1 + \exp(-\alpha u)} \tag{3.8}$$

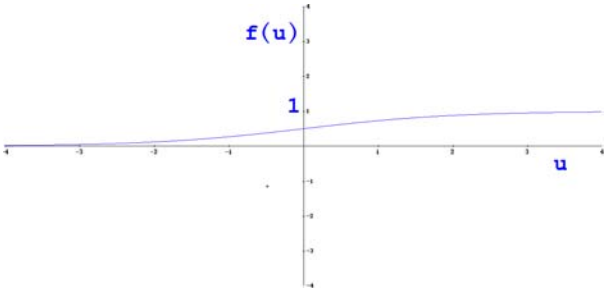

Figure 3- 8 Sigmoid function

4. *Hyperbolic Tangent Function*

Defined by

$$f(u) = \tanh(\alpha u) \qquad (3.9)$$

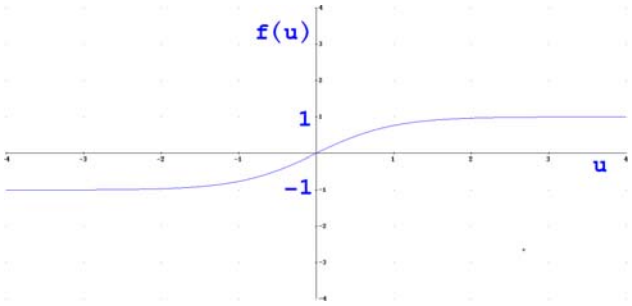

Figure 3- 9 Hyperbolic Tangent function

The sigmoid and the tangent hyperbolic functions are both strictly increasing functions and differentiable functions. The value α is the *slope parameter*.

In case of activation function with a discrete number of range values we will have the *Discrete Neural Networks* (DNN). Definitions (3.6) and (3.7) can be examples of activation funtions for this model. In case of a continuous activation function the networks will be named *Continuous Neural Networks* (CNN). Definitions (3.8) and (3.9) are examples of activation functions for a Continuous model

3.3 Hopfield Neural Networks

3.3.1 Structure of the Hopfield Neural Network

A *Hopfield Neural Network* (HNN) is one layer recurrent neural network with no hidden units in which all connections are bi-directional and symmetric; therefore, all neurons are both input and output nodes. Invented by John Hopfield in 1982, [19], his idea was to get a convergence of weights to find the minimum value for a *energy function* that has a scalar value associated with each state of the network. This value is referred to as the *energy of the network*. This type of network guarantees that its dynamics will converge to states which are local minima in the energy function. Thus, if a state is a local minimum in the energy function it is a stable state for the network. Such system could be implemented in hardware by combining standard components such as capacitors and resistors. Hopfield proposed his neural network circuit model in [27]

3.3.2 Mathematical model of the Continuous Hopfield Neural Network

3.3.2.1 Additive mathematical model of a neuron

A model of a node in HNN can be presented by an electrical circuit as shown in Figure 3-10. In physical terms, the synaptic weights $w_{k1}, w_{k2}, \ldots, w_{kS}$ are represented by *conductances*, (i.e. inverse of resistances), where S is the number of inputs. A HNN, as it has been said, is a recurrent network, in which the output of the nodes become the new inputs during the convergence process; therefore, the inputs are represented now by the same variables, $v_1(t), v_2(t), \ldots, v_S(t)$, as the outputs and are represented by *potentials* (i.e., voltages). These inputs are applied to a summing *junction* characterized by a low input resistance, unity current gain, and high output resistance; that is, it acts as a summing node for incoming currents. Afterwards, an external current source b_k, can be applied.

In neural network literature, the neuronal model described is commonly referred to as the additive model.

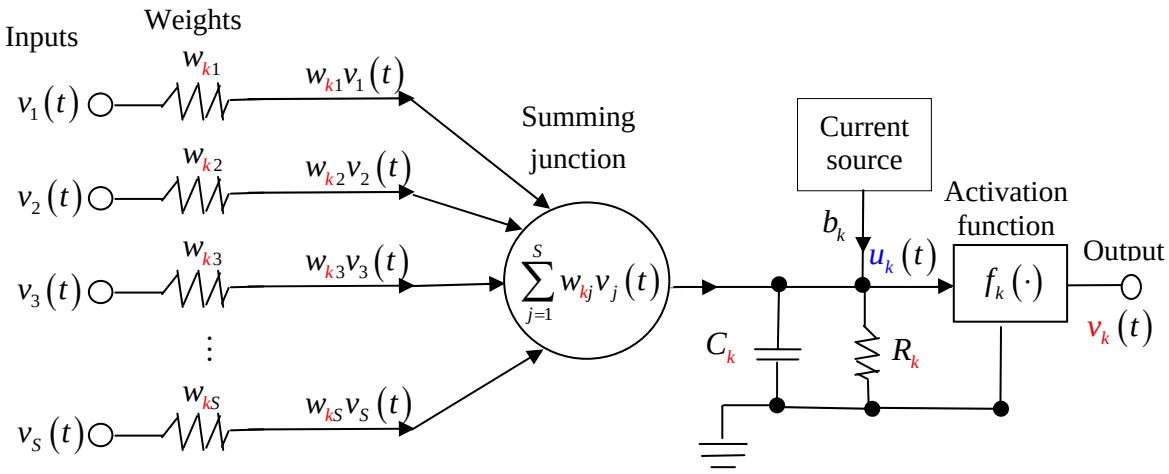

Figure 3-10 Continuous Additive model of node v_k

Let $u_j(t)$ denote the induced local field at the input of the nonlinear activation function $f_k(\cdot)$.

We may express the current fed into the resistance-capacitance (RC) circuit as follows:

$$\sum_{j=1}^{S} w_{kj} v_j(t) + b_k \tag{3.10}$$

and the total current flowing away from the input node of the nonlinear element as

$$\frac{u_k(t)}{R_k} + C_k \frac{du_k(t)}{dt}, \tag{3.11}$$

where the first term is due to leakage resistance R_k and the second term is due to leakage capacitance C_k. From Kirchoff's current law, we know that the total current flowing toward any node of an electrical circuit is zero. By applying Kirchoff's current law to the input node of the non linear activation function $f_k(\cdot)$ in Figure 3-10, we get

$$\frac{u_k(t)}{R_k} + C_k \frac{du_k(t)}{dt} = \sum_{j=1}^{S} w_{kj} v_j(t) + b_k . \tag{3.12}$$

Finally, we may determine the output of node k by using the nonlinear relation

$$v_k(t) = f_k(u_k(t)) \tag{3.13}$$

3.3.2.2 Mathematical model for a recurrent network of S interconnected nodes

Let us consider a *recurrent network* consisting of an interconnection of S nodes, each one of which is assumed to have the same mathematical model described by (3.12) and (3.13). Then, ignoring internode propagation time delays, we may define the dynamics of the network by the following *system of coupled first-order differential equations:*

$$\frac{u_k(t)}{R_k} + C_k \frac{du_k(t)}{dt} = \sum_{j=1}^{S} w_{kj} v_j(t) + b_k \;;\; k = 1, 2, \ldots, S . \tag{3.14}$$

The activation function $f_k(u_k(t))$ relating to the output of node k is assumed to be a continuous and differentiable function of its induced local field $u_k(t)$. Therefore, we have a Continuous Hopfield Neural Network (CHNN). A commonly used activation function is a

sigmoid function given by equation (3.15) and Figure 3.8. It does not matter the election of the activation function while it satisfies good differentiable and continuous properties.

$$f_k\left(u_k\left(t\right)\right)=\frac{1}{1+\exp\left(-\alpha_k u_k\left(t\right)\right)}\ ;\ k=1,2,\ldots,S\ . \tag{3.15}$$

Dividing by C_k in both sides of (3.14) we have,

$$\frac{du_k\left(t\right)}{dt}=-\frac{u_k\left(t\right)}{R_k C_k}+\sum_{j=1}^{S}\frac{w_{kj}}{C_k}v_j\left(t\right)+\frac{b_k}{C_k};\ \ k=1,2,\ldots,S\ . \tag{3.16}$$

We can rewrite equations (3.16), incorporating the circuit characteristics to constants. We define the *time constant,* $\tau_k=R_kC_k$, new weights, T_{kj} and new bias, I_k respectively as follows:

$$\tau_k=R_kC_k,\ T_{kj}=\frac{w_{kj}}{C_k}\ \text{and}\ \ I_k=\frac{b_k}{C_k}, \tag{3.17}$$

and we assume the same time constant for all nodes,

$$\tau=\tau_k\ ;\ \ k=1,2,\ldots,S\ . \tag{3.18}$$

Using (3.13), (3.17) and (3.18) in equations (3.16) the next ones follow

$$\frac{du_k\left(t\right)}{dt}=-\frac{1}{\tau}u_k\left(t\right)+\sum_{j=1}^{S}T_{kj}\,f_j\left(u_j\left(t\right)\right)+I_k;\ \ k=1,2,\ldots,S\ . \tag{3.19}$$

Defining the following vectors and matrix

$$\begin{aligned}&U\left(t\right)=\left[u_k\left(t\right)\right]_{k=1,\ldots,S};V\left(t\right)=\left[v_k\left(t\right)=\ f_k\left(u_k\left(t\right)\right)\right]_{k=1,\ldots,S},\\&\text{Interconnection matrix}: T=\left[T_{k,j}\right]_{\substack{k=1,\ldots,S\\ j=1,\ldots,S}},\\&\text{Bias current vector: } I=\left[I_k\right]_{k=1,\ldots,S}.\end{aligned} \tag{3.20}$$

Equation (3.19) can be written in the compact form

$$\frac{dU(t)}{dt} = -\frac{1}{\tau}U(t) + TV(t) + I\,. \tag{3.21}$$

3.3.3 Dynamical model of the Continuous Hopfield Neural Network

3.3.3.1 Some theoretical concepts in dynamical systems

Properties of Hopfield Neural Networks can be studied in terms of dynamical system models. For this, we briefly describe in this sub-section the following theorical concepts that we will use in next sections to study HNN convergence properties: ***dynamical system, state-space model, autonomous dynamical system, equilibrium state, stability definitons*** and the ***Lyapunov function*** and ***Lyapunov's theorems*** More details of these concepts can be found in classical references in dynamical sistems theory as [28] and [29].

A ***dynamical system*** consists of an abstract *state space*, whose coordinates describe the state at any instant and a *dynamical rule* that specifies the immediate future of all state variables, given only the present values of those same state variables. For example the state of a pendulum is its angle and angular velocity, and the evolution rule is Newton's equation. In HNN the state is the set of output value nodes $\left(v_1 = f_1\left(u_1(t)\right), v_2 = f_2\left(u_2(t)\right), \ldots, v_S = f_S\left(u_S(t)\right)\right)$ and equations (3.19) can be regarded as its dynamical rule. A formal definition describes a dynamical system as a tuple (R, S, T) where S is a state space, T is a set of times and R is a rule for evolution, $R : S \times T \to S$, that gives the future state to a state $s \in S$.

A ***state-space*** model is a mathematical model for describing the dynamics of a system such that, according to the model, the values of a set of state variables, at any particular instant of time, are supposed to contain sufficient information to predict the future evolution of the system. The ordered collection of states that follow from using the evolution rules is named the *trajectory* or *forward orbit* of the state space. The state space model (also known as the "time-domain approach") provides a convenient and compact way to model and analyze systems with multiple inputs and outputs. It may be modeled by a system of first-order differential equations as follows,

$$\frac{dx_j(t)}{dt} = F_j\left(x_j(t)\right); \quad j = 1, 2, \ldots, S\,, \tag{3.22}$$

where the function $F_j(\cdot)$ is, in general, a nonlinear function of its argument.

Using vector notation we can put the above equations as follows,

$$\frac{d\boldsymbol{x}(t)}{dt} = F\left(\boldsymbol{x}(t)\right), \tag{3.23}$$

where $\boldsymbol{x}(t) = \left[x_1(t), x_2(t), \ldots, x_S(t)\right]^T \in R^S$ is the vector of the state variables and $F : R^S \to R^S$ a vectorial function.

A nonlinear dynamical system described by equation like (3.23) where the vector function $F\left(\boldsymbol{x}(t)\right)$ does not depend explicitly on time t is said to be an ***autonomous dynamical system***.

Equilibrium state definition. A constant vector $\bar{\boldsymbol{x}}$ is said to be an equilibrium (stationary) state of a dynamical system described by (3.23) if the following condition is satisfied:

$$F\left(\bar{\boldsymbol{x}}\right) = \bar{0} \tag{3.24}$$

Notice that if a constant vector $\bar{\boldsymbol{x}}$ is an equilibrium state of a dynamical system, the velocity vector $\frac{d\boldsymbol{x}}{dt}$ vanishes at the equilibrium state $\bar{\boldsymbol{x}}$, and therefore the constant function $\boldsymbol{x}(t) = \bar{\boldsymbol{x}}$ is a solution of equation (3.23) and consequently a state included in the trajectory of the system

Stability definition 1. The *equilibrium state* $\bar{\boldsymbol{x}}$ is said to be *uniformly stable* if for any given positive ε there exists a positive δ such that, if $\boldsymbol{x}(0)$ is the initial state, the condition $\left\|\boldsymbol{x}(0) - \bar{\boldsymbol{x}}\right\| < \delta$ implies that $\left\|\boldsymbol{x}(t) - \bar{\boldsymbol{x}}\right\| < \varepsilon$ for all $t > 0$

This definition means that a trajectory of the system can be made to stay within a small neighborhood of the equilibrium state $\bar{\boldsymbol{x}}$ if the initial state $\boldsymbol{x}(0)$ is close to $\bar{\boldsymbol{x}}$

Stability definition 2. The *equilibrium state* $\bar{\boldsymbol{x}}$ is said to be *convergent* if there exists a positive δ such that, if $\boldsymbol{x}(0)$ is the initial state, the condition $\left\|\boldsymbol{x}(0)-\bar{\boldsymbol{x}}\right\|<\delta$ implies that $\boldsymbol{x}(t)\to\bar{\boldsymbol{x}}$ as $t\to\infty$.

This definition means that if the initial state $\boldsymbol{x}(0)$ of a trajectory is close enough to the equilibrium state $\bar{\boldsymbol{x}}$, then the trajectory described by the state vector $\boldsymbol{x}(t)$ will approach $\bar{\boldsymbol{x}}$ as time t approaches infinity.

Stability definition 3. The *equilibrium state* $\bar{\boldsymbol{x}}$ is said to be *asymptotically stable* if it is both stable and convergent.

Stability definition 4. The *equilibrium state* $\bar{\boldsymbol{x}}$ is said to be *globally asymptotically stable* if it is asymptotically stable and all trajectories of the system converge to $\bar{\boldsymbol{x}}$ as time approaches infinity.

Definition 4 implies that the system cannot have other equilibrium states, and it requires that every trajectory of the system remains bounded for all time $t>0$. Therefore, global asymptotic stability implies that the system will ultimately settle down to a steady state for any choice of initial conditions.

Lyapunov function definition. A *Lyapunov function* for the equilibrium state $\bar{\boldsymbol{x}}$ is a scalar function, $V(\boldsymbol{x}):U\subset R^S\to R$, defined on a neighbourhood U of $\bar{\boldsymbol{x}}$, that satisfies the following requirements:

1. $V(\boldsymbol{x})$ locally positive-definite, that is, on a neighbourhood U of $\bar{\boldsymbol{x}}$, $V(\bar{\boldsymbol{x}})=0$ and $V(\boldsymbol{x})>0$, $\forall\boldsymbol{x}\in U$, $\boldsymbol{x}\neq\bar{\boldsymbol{x}}$

2. $V(\boldsymbol{x})$ is continuous.

3. $V(\boldsymbol{x})$ has continuous first-order partial derivatives with respect to the elements of the state vector $\boldsymbol{x}$.

Remark. It can be noted that if $V(\boldsymbol{x}): U \subset R^S \rightarrow R$ satisfies 2. and 3. and $V(\boldsymbol{x}) > V(\bar{\boldsymbol{x}})$, but $V(\bar{\boldsymbol{x}}) = V_{\bar{x}} \neq 0$ then, for $\forall \boldsymbol{x} \in U$, $\boldsymbol{x} \neq \bar{\boldsymbol{x}}$, it can be taken $\tilde{V}(\boldsymbol{x}) = V(\boldsymbol{x}) - V_{\bar{x}}$ as a valid Lyapunov function now satisfaying $\tilde{V}(\bar{\boldsymbol{x}}) = V(\bar{\boldsymbol{x}}) - V_{\bar{x}} = 0$ and $\tilde{V}(\boldsymbol{x}) = V(\boldsymbol{x}) - V_{\bar{x}} > 0$. Since $\frac{\partial \tilde{V}(\boldsymbol{x})}{\partial x} = \frac{\left(\partial V(\boldsymbol{x}) - V_{\bar{x}}\right)}{\partial x} = \frac{\partial V(\boldsymbol{x})}{\partial x}$ all the below results related with minimum properties will remain valid without paying attention that $V(\bar{\boldsymbol{x}}) = 0$ will be satisfied.

Lyapunov functions are of interest in mathematics, especially in stability theory. Named after Aleksander Mikhailovich Lyapunov, they are functions which prove the stability of a certain fixed point in a dynamical system or autonomous differential equation.

Theorem of Lyapunov

Lyapunov's theorem stablishes the conditions for the stability and asymptotic stability of the state-space equation described by an autonomous nonlinear dynamical system with state vector $\boldsymbol{x}(t)$ and equilibrium state $\bar{\boldsymbol{x}}$. They may be stated as follows [29]:

Let $\bar{\boldsymbol{x}} = 0$ *be an equilibrium state for a system described by* $\frac{d\boldsymbol{x}(t)}{dt} = F(\boldsymbol{x}(t))$ *where* $F: U \rightarrow R^n$ *is a locally Lipschitz function and* $U \subset R^n$ *is a small neigborhood around* $\bar{\boldsymbol{x}} = 0$ *. Given a Lyapunov function* $V(\boldsymbol{x})$*, then*

1.If its derivative with respect to time is negative semidefinite for $\boldsymbol{x} \in U \setminus \{\bar{\boldsymbol{x}}\}$*, i.e.* $\frac{dV(\boldsymbol{x})}{dt} \leq 0$*, then* $\bar{\boldsymbol{x}} = 0$ *is a stable equilibrium point.*

2. If its derivative with respect to time is negative definite for $\boldsymbol{x} \in U \setminus \{\bar{\boldsymbol{x}}\}$*, i.e.* $\frac{dV(\boldsymbol{x})}{dt} < 0$*, then* $\bar{\boldsymbol{x}} = 0$ *is an asymptotically stable equilibrium point.*

The important point of this discussion is that Lyapunov theorem is used to make conclusions about trajectories' properties of a system without having to solve the state-space equation of the system. Unfortunately, the theorems give no indications of how to find a Lyapunov function; it is a matter of trial and error in each case. The inability to find a Lyapunov function doesn't mean

that the system is unstable. Functions which might prove the stability of some equilibrium are called Lyapunov-candidate-functions. For dynamical systems (e.g. physical systems), conservation laws can often be used to construct a Lyapunov-candidate-function.

3.3.3.2 Energy function, equilibrium states and global stability of the CHNN

In 3.3.3.1 it has been introduced the basis to study the stability in Continuous Hopfield Neural Networks (CHNN) systems. In this respect the usual way to proceed is the obtention of a Lyapunov function. In HNN terminology they use to be defined as an energy function. Then an important issue in HNN is how to determine the energy function and how to clarify its convergence to equilibrium states. In this sub-section we will describe how the stability of the nonlinear dynamic system that models the Hopfield neural network can be established using an apropiate Lyapunov function.

To study the HNN stability behaviour, whose mathematical model was described by equations (3.19) (or its equivalent (3.21)), the following additional assumptions are introduced:

1. The matrix of synaptic weights is symmetric: $T_{kj} = T_{jk}$, $k, j = 1, 2, \ldots, S$
2. The inverse of the nonlinear activation function $f_k(\cdot)$ exists, so we can write

$$u_j = f_j^{-1}(v_j), \; j = 1, 2, \ldots, S$$

Then, it is shown in [30] that the CHNN described by equations (3.19) has the scalar energy function (Lyapunov function), defined by

$$E = -\frac{1}{2}\sum_{k=1}^{S}\sum_{j=1}^{S} T_{kj} v_k v_j + \frac{1}{\tau}\sum_{j=1}^{S}\int_0^{v_j} f_j^{-1}(v)\,dv - \sum_{j=1}^{S} I_j v_j \tag{3.25}$$

and taking into account a particular activation function ranging from 0 to 1 like atanh (.). For a generic activation function, such that $\int_p^{v_j} f^{-1}(v)\,dv$ satisfies $\lim_{v_j \to p}\int_p^{v_j} f^{-1}(v)\,dv = 0$, $\int_p^{v_j} f^{-1}(v)\,dv > 0$ for $v_j > 0$ and $\lim_{v_j \to 1}\int_p^{v_j} f^{-1}(v)\,dv = 1$, we can express

$$E = -\frac{1}{2}\sum_{k=1}^{S}\sum_{j=1}^{S} T_{kj} v_k v_j + \frac{1}{\tau}\sum_{j=1}^{S}\int_{p}^{v_j} f_j^{-1}(v)\,dv - \sum_{j=1}^{S} I_j v_j \quad , \tag{3.26}$$

being p the abscissa value verifying $f_j^{-1}(p) = 0$. Notice that there is only one value satisfying this condition because of the monotonicity of the function.

The exact form for the nonlinear activation function $f_k(\cdot)$ is not particularly important, and any non linear monotonically increasing function could be used, the only requirement is to adapt the limits of the integral of the second sumand in order to integrate in a domain with positive values for the integrand function. Preference elections are a sigmoid function or the hyperbolic tangent function.

For illustrative purposes the following Proposition is detailed to show that the expression (3.26) is a Lyapunov function. Accordingly, the CHNN described possesses a Lyapunov (energy) function which continually decreases with time. The existence of such an Energy function guarantees that the system converges towards equilibrium states that are local minimum solutions. These equilibrium states are often referred to as *point attractors.* Next sub-section studies where these equilibrium points are located.

Proposition . *If* $\bar{\boldsymbol{x}}$ *is an equilibrium state that satisfies* $E(\bar{\boldsymbol{x}}) = 0$ *and that is a minimum for the energy function defined by equation* (3.26) *and* $f_k^{-1}(\cdot)$ *is monotonically increasing the following results follow,*

i) Function $E(\boldsymbol{x})$ *is bounded and continuous from his definition and* $E(\bar{\boldsymbol{x}}) = 0$

ii) Function $E(\boldsymbol{x})$ *has continuous first-order partial derivatives with respect to the elements of the state vector* $\boldsymbol{x}$

iii) Function E *satisfies* $E(\boldsymbol{x}) > 0, \forall \boldsymbol{x} \in U,\ \boldsymbol{x} \neq \bar{\boldsymbol{x}}$ and being U a neighbourhood of $\boldsymbol{x}$

iv) Function E *is a monotonically decreasing function of time that satisfies* $\frac{dE}{dt} < 0$

Proof:

Statements i), ii) and iii) follow by definition of the point $\overline{\boldsymbol{x}}$ and the energy function.

To establish iv), we proceed to differentiate E with respect to time,

$$\frac{dE}{dt}=\frac{d}{dt}\left(-\frac{1}{2}\sum_{k=1}^{S}\sum_{j=1}^{S}T_{kj}v_k v_j\right)+\frac{1}{\tau}\frac{d}{dt}\left(\sum_{j=1}^{S}\int_{p}^{v_j}f_j^{-1}(v)\,dv\right)-\frac{d}{dt}\left(\sum_{j=1}^{S}I_j v_j\right)=$$

$$=\left(-\frac{1}{2}\sum_{k=1}^{S}\sum_{j=1}^{S}T_{kj}\frac{dv_k}{dt}v_j-\frac{1}{2}\sum_{k=1}^{S}\sum_{j=1}^{S}T_{kj}v_k\frac{dv_j}{dt}\right)+\frac{1}{\tau}\sum_{j=1}^{S}f_j^{-1}\left(v_j\right)\frac{dv_j}{dt}-\sum_{j=1}^{S}I_j\frac{dv_j}{dt}\underset{T_{ji}=T_{ij}}{=}$$

$$=-\sum_{k=1}^{S}\sum_{j=1}^{S}T_{kj}v_k\frac{dv_j}{dt}+\frac{1}{\tau}\sum_{j=1}^{S}f_j^{-1}\left(v_j\right)\frac{dv_j}{dt}-\sum_{j=1}^{S}I_j\frac{dv_j}{dt}\underset{u_j=f_j^{-1}(v_j)}{=}-\sum_{j=1}^{s}\left(\sum_{k=1}^{S}T_{kj}v_k-\frac{1}{\tau}u_j+I_j\right)\frac{dv_j}{dt} \tag{3.27}$$

we recognize equation (3.19) , for neuron j, inside the parentheses on the right-hand side of equation (3.27), hence, we can summarize the result obtained by the following equation

$$\frac{dE}{dt}=-\sum_{j=1}^{s}\left(\frac{du_j}{dt}\right)\frac{dv_j}{dt}. \tag{3.28}$$

In order to analize the sign of the above equation, we proceed to consider the inverse of activation function in equation (3.13), which exists according to the assumption 3, and then after differentiating it we have

$$\frac{du_j}{dt}=\frac{df_j^{-1}\left(v_j\right)}{dt}=\frac{df_j^{-1}\left(v_j\right)}{dv_j}\frac{dv_j}{dt}. \tag{3.29}$$

Using (3.29) in (3.28) we obtain

$$\frac{dE}{dt}=-\sum_{j=1}^{S}\left(\frac{dv_j}{dt}\right)^2\frac{df_j^{-1}\left(v_j\right)}{dv_j}. \tag{3.30}$$

By the hypothesis, $f_j^{-1}(\cdot)$ is monotonically increasing, it follows therefore that

$$\frac{df_j^{-1}(v_j)}{dv_j} \geq 0; \text{ for all } v_j . \tag{3.31}$$

We also note that

$$\left(\frac{dv_j}{dt}\right)^2 \geq 0; \text{ for all } v_j . \tag{3.32}$$

Hence, all the factors that make up the sum on the right-hand side of equation (3.30) are nonnegative and, therefore, we can conclude that the energy function E defined by equation (3.25) satisfies

$$\frac{dE}{dt} \leq 0 . \tag{3.33}$$

In equation (3.30), we may note that all addends have the same sign, therefore $\frac{dE}{dt} = 0$ only if

$$\frac{dv_j}{dt} = 0; \text{ for all } v_j . \tag{3.34}$$

By equation (3.34) we may write

$$\frac{dE}{dt} < 0, \text{ except at a fixed point } \bullet \tag{3.35}$$

3.3.3.3 Relation between the Stable States of the DHNN and CHNN

So far we have been considering CHNN that have a continuous activation function. However, discrete activation functions taking a finite number of output values can be also considered as *Discrete Hopfield Neural Networks* (DHNN). The activation function output in DHNN uses to be binary. The Discrete Hopfield Neural Network model (DHNN) was presented by Hopfield in his paper [19].

In this sub-section a relation between the stable states of DHNN and CHNN will be established. Assuming that a large gain (slope at the origine) in the CHNN activation function is taken, that

is, that negative (positive) small values of the input of the continuos activation function (e.g sigmoid) lead to 0 $(+1)$ output values , it is possible to model the dynamics of the CHNN with the Lyapunov (energy) functions of a DHNN. Then, as in this DHNN model the ideal equilibrium states are located in the vertexes of a hypercube it can be assumed the same behaviour will be present in the CHNN model. That is, in practice the equilibrium states in the CHNN model will tend to be close to the vertexes. Figure 3-11 shows the discrete additive model of a node v_k in a DHNN with S nodes and a binary activation function. In this network an appropriate energy function (Lyapunov function) is the following:

$$E = -\frac{1}{2}\sum_{\substack{k=1 \\ k\neq j}}^{S}\sum_{j=1}^{S} T_{kj} v_k v_j \; - \sum_{k=1}^{S} I_k v_k \tag{3.36}$$

where $v_k = f_k\left(u_k\right)$.

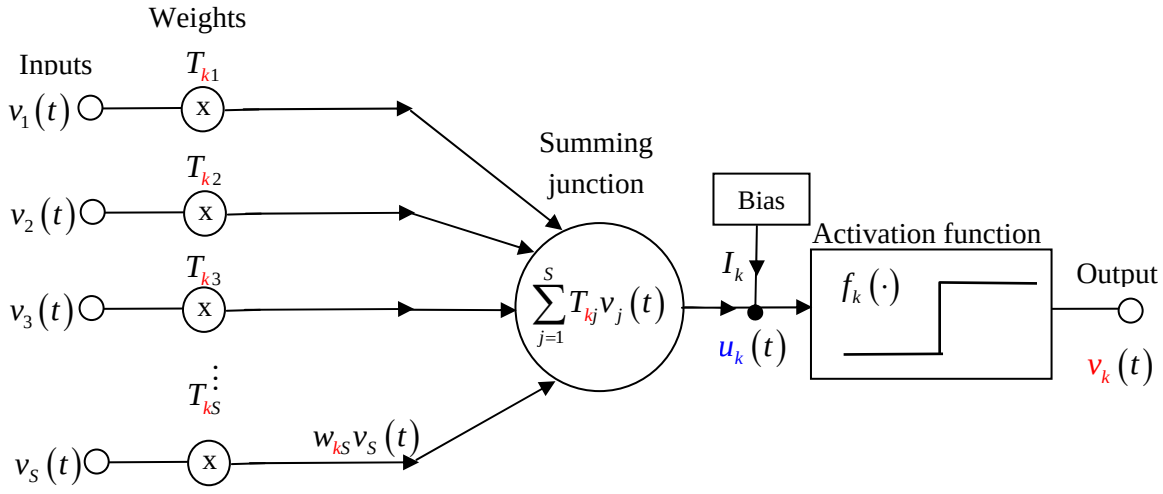

Figure 3-11 Discrete Additive model of node v_k

Discrete Hopfield Neural Network needs that diagonal elements in the matrix equal 0 (i.e. $T_{kk} = 0$). The absence of a connection from each unit to itself is necessary to avoid a permanent feedback of its own state value.

Hopfield showed [30] that this energy always decreased until reaching the equilibrium state at the corners of a S -dimensional hypercube space.

The DHNN features have attracted a great deal of attention in the literature as a *content-addressable memory application*. The ability of a content-addressable memory application is to retrieve a stored pattern, given a reasonable subset of the information content of the pattern. In these applications the desired patterns are stored in the vertices of the mentioned S-dimensional hypercube space. DHNN features have been also exploited in many other fields, as an elegant way of solving combinatorial problems.

We will use the above results to analize where the equilibrium states of a CHNN can be located. It was shown in sub-section 3.3.3.2 that the energy function of a CHNN model can be given by equation:

$$E = -\frac{1}{2}\sum_{k=1}^{S}\sum_{j=1}^{S} T_{jk} v_k v_j + \frac{1}{\tau}\sum_{j=1}^{S}\int_{p}^{v_j} f_j^{-1}(v)\,dv - \sum_{j=1}^{S} I_j v_j \tag{3.37}$$

Let us consider the activation functions

$$v = f_j(u) = \frac{1}{1+\mathrm{e}^{-\alpha_j u}};\quad j = 1,\ldots,S \tag{3.38}$$

its inverse functions satisfies

$$u = f_j^{-1}(v) = -\frac{1}{\alpha_j}\ln\left(\frac{1-v}{v}\right)\quad j = 1,\ldots,S\,. \tag{3.39}$$

Then, defining

$$f^{-1}(v) = -\ln\left(\frac{1-v}{v}\right), \tag{3.40}$$

equations (3.39) can be rewritten as

$$f_j^{-1}(v) = \frac{1}{\alpha_j} f(v)\quad j = 1,\ldots,S\,. \tag{3.41}$$

Using (3.41), (3.37) can be rewritten as

$$E = -\frac{1}{2}\sum_{k=1}^{S}\sum_{j=1}^{S} T_{kj} v_k v_j - \sum_{j=1}^{S} I_j v_j + \frac{1}{\tau}\sum_{j=1}^{S}\frac{1}{\alpha_j}\int_{p}^{v_j} f^{-1}(v)\,dv \tag{3.42}$$

In the RRM scenarios considered in this thesis work the third term on the right tends to zero. Then (3.42) tends to be (3.36) and the equilibrium states of the CHNN system are nearby the vertexes of a hypercube as in DHHN system.

3.3.3.4 Working Hypothesis for Continuous HNN

It will be assumed that Continuous HNN with S *nodes used in this thesis are modeled by the following equations*

$$\begin{aligned} \frac{du_k(t)}{dt} &= -\frac{1}{\tau}u_k(t) + \sum_{j=1}^{S} T_{kj} v_j + I_k \quad ;\ k = 1,2,\ldots,S, \\ v_k &= f_k(u) = \frac{1}{1+e^{-\alpha_k u}} \end{aligned} \tag{3.43}$$

where u_k, v_k *denote inputs and ouputs values of nodes respectively,* $T = \left(T_{kj}\right)_{k,j=1,\ldots,S}$ *is a symmetric interconnection matrix,* $I = \left[I_k\right]_{k=1,\ldots,N}$ *denote a vector of additional conditions (bias current vector),* $\tau = 1$ *and high* α_k*'s values will be considered. Therefore, its Lyapunov (energy) function will be taken*

$$E = -\frac{1}{2}\sum_{k=1}^{S}\sum_{j=1}^{S} T_{kj} v_k v_j \ - \sum_{j=1}^{S} I_j v_j \ . \tag{3.44}$$

Using (3.44)*, equations* (3.43) *can be written as*

$$\frac{du_k(t)}{dt} = -\frac{1}{\tau}u_k(t) - \frac{\partial E}{\partial v_k};\ k = 1,2,\ldots,S \ . \tag{3.45}$$

For this model stable states can be found on the vertices of an appropiate S *-dimensional cube*

Remark. As it was already stated with the DHHN model, it can be said without loss of generality in the search of equilibrium states, that all $T_{kk} = 0$. If not, the square term $\frac{T_{kk} v_k^2}{2}$ in the energy function can be replaced by the linear term $\frac{T_{kk} v_k}{2}$ since the output of the output function tends to be either 0 or 1. Thus, the square term values can be forced to approach to zero by compensatory tems as follows

$$I_k \leftarrow I_k - \frac{T_{kk}}{2} \text{ for } k = 1, \ldots, S . \tag{3.46}$$

3.3.4 Solving optimization problems with Hopfield Neural Networks

Hopfield Neural Networks are proposed as a means of approximately solving combinatorial optimization problems which can somehow be expressed as the minimization of a quadratic objective function over a set of 0-1 variables. In HNN, these problems are converted into a continuous optimization problem that minimizes an energy (Lyapunov) function. The key idea is that this Lyapunov function is usually constructed employing one term associated with the problem's objective function and adding one more term for each constraint. A typical Lyapunov function would take the form

$$E = E^{op} + \mu_1 E_1^{cns} + \mu_2 E_2^{cns} + \ldots, \tag{3.47}$$

where E^{op} is the objective function to be minimized related with the problem, and the E_i^{cns} terms are constraint functions determined for each particularly case. This function when expanded and rearranged properly should be put in the form of equation (3.44).

3.4 Implementation of a Hopfield Neural Network

3.4.1 Numerical simulation of a HNN

Let be a CHNN whose dynamics is described by a set of nonlinear first order differential equations given by . This is a continuous time model (CTM) of the neural network that could be implemented as a Hardware circuit as it is shown in Figure 3-10. However, it is common to

simulate network performance on a digital computer with a discrete time model (DTM), because it is an easier task than to build the hardware equipment.

An Euler discretization of the CTM can be made using the Euler method, i.e., using

$$\frac{du_k(t)}{dt} \approx \frac{u_k(t+h)-u_k(t)}{h} \tag{3.48}$$

equations (3.43) can be discretizated as

$$\begin{aligned} u_k(t+h) &= u_k(t)+h\left(-\frac{1}{\tau}u_k(t)+\sum_{j=1}^{S}T_{kj}v_j+I_k\right);\ k=1,2,\ldots,S. \\ v_j(t) &= \frac{1}{1+\mathrm{e}^{-\alpha_j u_k(t)}} \end{aligned} \tag{3.49}$$

By defining a mesh trajectory times $t_n = t_0 + nh$, for $n = 0,1,2,\ldots$ and for a given initial trajectory started time t_0, Euler's method constructs the following set of values to approximate the trajectory

$$\begin{aligned} & w_k^0 = u_k(t_0). \\ & w_k^{n+1} = w_k^n + h\left(-\frac{1}{\tau}w_k^n+\sum_{j=1}^{S}T_{kj}\left(v_j\right)^n+I_k\right);\ n=1,2,3,\ldots. \\ & \left(v_j\right)^n = \frac{1}{1+\mathrm{e}^{-\alpha_j w_k^n}}. \end{aligned} \tag{3.50}$$

The Euler aproximation depends on the choice of the trajectory time step h; the smaller it is, the more accurate is the approximation but greater the time of simulation.

Note that for large n the S - dimensional point $W^n = \left(w_1^n, w_2^n, \ldots, w_S^n\right)$ should be close enough to the S - dimensional stable state. Then, for a selected norm $\|\ \|$ and a selected ε, the iterative by equations (3.50) is stopped when

$$\left\|W^n - W^{n-1}\right\| < \varepsilon. \tag{3.51}$$

3.4.2 Hardware implementation of a Hopfield Neural Network

A hardware implementation of the Hopfield network according to Figure 3-10 could reach a stable state after several time constants have been transcurred and the permanent state in the analogue circuit has been attained. This analysis assumes that all the components are ideal ones. With real components there remain several problems to overcome as for programmable resistors are not easily implemented in VLSI and amplifiers present propagation delays. So, recently the move has been away from the above circuits towards for example promising FPGA–based and MOSFET-analogue based circuits, which require weights to be stored as voltages [31, 32].

3.5 Hopfield Neural Network model for Radio Resource Management

3.5.1 Introduction

Radio Resource Management (RRM) deals with an efficient distribution of resources and may be formulated as a combinatorial optimization problem subjected to a set of constraints. In some applications, the real-time solution of optimization problems is widely required. However, traditional numerical algorithms may not be efficient since the computing time required for a solution is greatly dependent on the dimension and structure of the problems. Particularly, HNNs are suitable to solve complex combinatorial problems and, if analog hardware is implemented, then they work in real time in a few microseconds.

Sub-section 3.5.2, describes how RRM problems can be formulated as a Hopfield Neural Network. The major problem of the Hopfield model, however, is how to determine the energy fuction. Several energy functions can be proposed for a given problem. Sub-section, 3.5.3 introduces a common mathematical formulation of all the HNN used in our work, named in the related literature, 2D HNN. In this sub-section the most common constraint function summands used in Section 3.5.4 for the general formulation of the 2D HNN energy function are also exposed. In this thesis, novel adaptations of this energy model have been introduced to solve with good results each RRM case studied.

3.5.2 Mathematical Problem formulation

Since our goal is to solve RRM real-time problems with HNN, and such networks use binary outputs to represent problem solutions, this sub-section represents a RRM task with a set of binary variables.

A task RRM problem may be viewed as an optimization problem subjected to a set of constraints. This set of constraints is composed by Quality of Service constraints (delay constraints, target Signal to Noise ratio, …), functional constraints (power, bandwidth, …) and hardware constraints (number of processors, …).

Let us assume that we have a RRM problem consisting in a system with N ordered users and M ordered services to manage. Then, the associated HNN architectural will require $S = N\text{x}M$ nodes and let $U \in N^S$ and $V \in N^S$ be the vectors of the input and output values of the nodes respectively. We group, for convenience, the nodes and its output values into N groups of M nodes. Each group of M nodes will be used to represent an assignment of services to one user connection. Thus, nodes and output value nodes are denoted by two subcripts as follows

$$\begin{aligned} &U_{ij}, i = 1,\ldots,N;\ j = 1,\ldots,M \\ &V_{ij}, i = 1,\ldots,N;\ j = 1,\ldots,M \end{aligned} \tag{3.52}$$

and

$$U = \left[U_{ij}\right]_{\substack{i=1,\ldots,N \\ j=1,\ldots,M}};\ V = \left[V_{ij}\right]_{\substack{i=1,\ldots,N \\ j=1,\ldots,M}}. \tag{3.53}$$

Notice that U_{ij} and V_{ij} designe the components $U_{N(i-1)+j}$ and $V_{N(i-1)+j}$ of vectors U and V respectively.

We identify an assignement of a resource to a user with a output value of a node as follows

Definition. *The j-th resource is assigned to i-th user iff $V_{ij} = 1$.*

By the definition, the set of output value nodes gives the final solution to the RRM problem as follows

$$V_{ij} = \begin{cases} 1, & \text{resource } j \text{ is assigned to user } i \\ 0, & \text{otherwise} \end{cases}; \quad i = 1,\ldots,N \quad j = 1,\ldots,M \,. \qquad (3.54)$$

This notation with two subscripts helps to interpret easily the outputs in terms of users and resources and it is named 2D (dimensional) HNN

As an example, let us consider a HNN architectural with $S = 6\text{x}5 = 30$ nodes corresponding to 6 users and 5 resources. In this HNN, a possible given output value vector notation is written as $V = (V_k)_{k=1,\ldots,30} = (0,1,0,0,0,1,0,0,0,0,0,0,0,0,1,0,0,0,1,0,0,0,0,0,0,0,0,1,0,0)$. In 2D HNN notation this output value vector, V, is grouped in 6 groups of 5 components, as follows: $V_1 = (0,1,0,0,0)$, $V_2 = (1,0,0,0,0)$, $V_3 = (0,0,0,0,1)$, $V_4 = (0,0,0,1,0)$, $V_5 = (0,0,0,0,0)$, $V_6 = (0,0,1,0,0)$. The group V_i shows the assignement for the i-th user. In this 2D HNN notation, the above output should be written as $V_{12} = 1, V_{21} = 1, V_{35} = 1, V_{44} = 1, V_{63} = 1$ and $V_{ij} = 0$ otherwise. That is, user connections number 1, 2, 3, 4 and 6 have the resources 2, 1, 5, 4 and 3 respectively. User connection number 5 has no resources associated. No others assignments have been made (Figure 3-12)

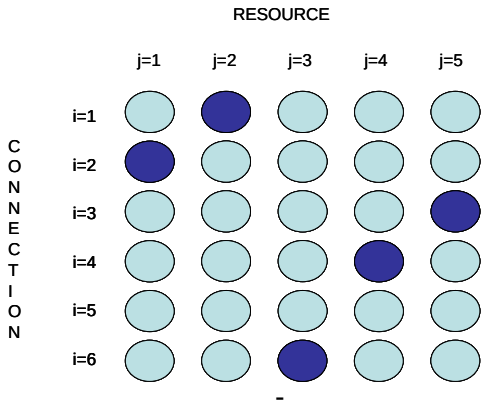

Figure 3-12 An assignment in a 6x5 2D HNN model

A set of constraints between the set of nodes are described, in 2D HNN notation, by the matrix

$$T = \left[T_{kl,ij} \right]_{\substack{i,k=1,\ldots,N \\ j,l=1,\ldots,M}} . \qquad (3.55)$$

Additional conditions that may be required over each node are defined by the vector

$$I = \left[I_{kl} \right]_{\substack{k=1,\ldots,N \\ l=1,\ldots,M}}. \tag{3.56}$$

The additive model of a node described on Figure 3-10 is now shown in Figure 3-13 in for the node U_{kl} in a 2D HNN model

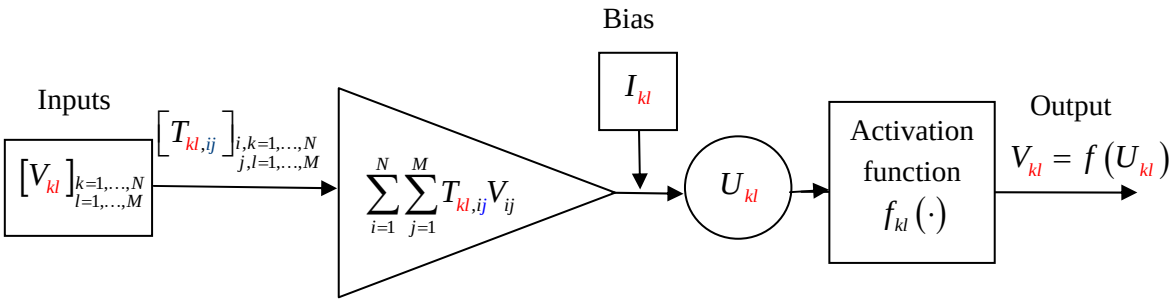

Figure 3-13 Additive model of a node in a 2D HNN model for resources management

Equations for the mathematical model of the CHNN described by equation (3.19), deduced in subsection 3.3.2.2, can now be rewritten as follows,

$$\frac{dU_{kl}(t)}{dt} = -\frac{1}{\tau} U_{kl}(t) + \sum_{i=1}^{N} \sum_{j=1}^{M} T_{kl,ij} f_{ij}\left(U_{ij}(t)\right) + I_{kl};\ k = 1,\ldots,N; l = 1,\ldots,M, \tag{3.57}$$

where we have also assumed, for all nodes, the time constant condition defined by equation (3.17) and the following activation functions,

$$V_{kl} = f_{kl}(U_{kl}) = \frac{1}{1+\exp\left(-\alpha_{kl} U_{kl}\right)}\ ;\ k = 1,2,\ldots,N; l = 1,2,\ldots,M \tag{3.58}$$

and its inverse functions

$$U_{kl} = f_{kl}^{-1}(V_{kl}) = -\frac{1}{\alpha_{kl}} \ln\left(\frac{1-V_{kl}}{V_{kl}}\right);\ k = 1,2,\ldots,N; l = 1,2,\ldots,M. \tag{3.59}$$

By using expression (3.40) equation (3.59) becomes

$$U_{kl} = f_{kl}^{-1}(V_{kl}) = -\frac{1}{\alpha_{kl}} f^{-1}(V_{kl});\ k = 1,2,\ldots,N; l = 1,2,\ldots,M \tag{3.60}$$

By using the vector and matrix notation introduced in (3.53), (3.55) and (3.56), equations (3.57) and (3.58) can be written in the compact form

$$\frac{dU(t)}{dt} = -\frac{1}{\tau}U(t) + TV + I\,. \tag{3.61}$$

In subsection 3.3.3.2 it was established that for a non-decreasing monotonic activation function and a symmetric matrix T the network is asymptotically stable and equations (3.57) and (3.58) have the following Lyapunov (energy) function

$$E = -\frac{1}{2}\sum_{k=1}^{N}\sum_{l=1}^{M}\sum_{i=1}^{N}\sum_{j=1}^{M}T_{ij,kl}V_{ij}V_{kl} + \frac{1}{\tau}\sum_{i=1}^{N}\sum_{j=1}^{M}\int_{p}^{V_{ij}}\frac{1}{\alpha_{ij}}f^{-1}(v)dv - \sum_{i=1}^{N}\sum_{j=1}^{M}I_{ij}V_{ij}\ , \tag{3.62}$$

then, for $\tau = R_{kl}C_{kl};, \alpha_{ij} = \alpha, i = 1,2,\ldots,N;\ j = 1,2,\ldots,M$, the energy function can be rewritten as

$$E = -\frac{1}{2}\sum_{p=1}^{N}\sum_{q=1}^{M}\sum_{i=1}^{N}\sum_{j=1}^{M}T_{ij,kl}V_{ij}V_{pq} + \frac{1}{\tau}\frac{1}{\alpha}\sum_{i=1}^{N}\sum_{j=1}^{M}\int_{p}^{V_{ij}}f^{-1}(v)\,dv - \sum_{i=1}^{N}\sum_{j=1}^{M}I_{ij}V_{ij}\,. \tag{3.63}$$

Or, in a compact form,

$$E = -\frac{1}{2}V^{T}TV + \frac{1}{\tau}\frac{1}{\alpha}\sum_{i=1}^{N}\sum_{j=1}^{M}\int_{p}^{V_{ij}}f^{-1}(v)\,dv - V^{T}I\,. \tag{3.64}$$

However, if the gain α of the activation function times τ is sufficiently high then second term in (3.63) (or in (3.64)) becomes negligible and so the Lyapunov function of the systems becomes equal to the energy function for discrete HNN as it was described in sub-section 3.3.3.3; i.e. the following energy function might be considered

$$E = -\frac{1}{2}V^{T}TV - V^{T}I\,. \tag{3.65}$$

We summarize the above results in a working hypothesis for 2D CHNN

3.5.3 Working Hypothesis for 2D Continuous HNN

It will be assumed that Continuous HNN with S *neurons used in this thesis are modeled by the following equations*

$$\begin{aligned} \frac{dU_{kl}(t)}{dt} &= -\frac{1}{\tau}U_{kl}(t) + \sum_{i=1}^{N}\sum_{j=1}^{M} T_{kl,ij}\, f_{ij}\left(U_{ij}(t)\right) + I_{kl}, \\ V_{kl} &= f_{kl}\left(U_{kl}\right) = \frac{1}{1+\exp\left(-\alpha_{kl}U_{kl}\right)} \;;\; k=1,2,\ldots,N; l=1,2,\ldots,M \end{aligned}, \tag{3.66}$$

where U_{kl}, V_{kl} *denote inputs and ouputs values of nodes respectively,* $T = \left(T_{ij,kl}\right)_{\substack{i,k=1,\ldots,N \\ j,l=1,\ldots,M}}$ *is a symmetric matrix (interconnection matrix),* $I = \left[I_{kl}\right]_{\substack{k=1,\ldots,N \\ l=1,\ldots,M}}$ *denote a vector of additional conditions, named (bias current vector),* $\tau = 1$ *and high* α_{kl}*'s values will be considered. Therefore, its Lyapunov (energy) function will be given by*

$$E = -\frac{1}{2}\sum_{k=1}^{N}\sum_{l=1}^{M}\sum_{i=1}^{N}\sum_{j=1}^{M} T_{ij,kl}V_{ij}V_{kl} - \sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}V_{ij}\,. \tag{3.67}$$

Using (3.66)*, equations* (3.67) *can be written as*

$$\frac{dU_{kl}(t)}{dt} = -\frac{1}{\tau}U_{kl}(t) - \frac{\partial E}{\partial V_{kl}};\; k=1,\ldots,N; l=1,\ldots,M\,. \tag{3.68}$$

For this model stable states can be found on the vertices of an appropiate S *-dimensional cube*

Remark. It can be assumed, without loss of generality in the search of the equilibrium state, that all $T_{kl,kl} = 0$. If not, the square term $\frac{T_{kl,kl}V_{kl}^2}{2}$ in the energy function can be replaced by the linear term $\frac{T_{kl,kl}V_{kl}}{2}$ since the energy at any vertex in the S-dimensional hypercube does not change. Thus, the square term values can be forced to approach to zero by compensatory tems as follows

$$I_{kl} \leftarrow I_{kl} - \frac{T_{kl,kl}}{2};\ k = 1,\ldots,N; l = 1,\ldots,M\ . \tag{3.69}$$

3.5.4 Model of Energy function for RRM problems

In this subsection we describe the model of the energy function that can be used to study optimization problems in RRM. In section 3.3.5 we briefly described how energy functions associated to combinatorial optimization problems could be constructed. These energy functions contained several terms related to the objective function to be minimized and to the constraints of the problem. Energy functions associated to RRM problems follow this pattern, moreover classical terms will be added to ensure the goals of RRM problems.

In order to better illustrate how the above HHN features can be exploited in RRM, we will introduce in the following a problem model. This model will be the basis of the novel energy functions we have designed in our research. Particularly details of each one will be presented in next chapters for each of the RRM problems considered in this thesis.

RRM Problem Model

Let B_T be a total bandwidth to be partitioned among N different users. Let R_{ij}, $j = 1,\ldots,M$ be the bit rates available for the user i and R_i^*, $R_i^* \in \{R_{ij}\}_{j=1,\ldots,M}$, $i = 1,\ldots,N$, the bit rate selected to be assigned to this i-th user. For a given RRM scenario, with a number of constraints the goal is to minimize a cost function:

$$C(B_T, R_{ij}, N, M, V_{ij}) \tag{3.70}$$

where variables V_{ij} are taking values 1 or 0 depending on whether the j-th resource is assigned to the i-th user or not respectively, that is

$$V_{ij} = \begin{cases} 1, & \text{if } R_i^* = R_{ij} \\ 0, & \text{otherwise} \end{cases};\ i = 1,\ldots,N; j = 1,\ldots,M\ . \tag{3.71}$$

Then final values for $\left[V_{ij}\right]_{\substack{i=1,\ldots,N \\ j=1,\ldots,M}}$ *gives the optimal solution to the RRM problem.*

Working Hypotheses for the RRM Problem Model.

i) *We assume that we have a combinational problem, i.e. that the number of users is so large, that it is not possible to provide the required bit rates to all of them. So, the choice of the best solution must be selected among a set of combination candidate solutions.*

ii) *There is, at least, one selection* $R_i^* \in \left\{R_{ij}\right\}_{j=1,\ldots,M}, i=1,\ldots,N$ *satisfying* $\sum_{i=1}^{N} R_i^* = B_T$

Associated HNN and Energy Function for the RRM problem model

The associated HNN schem for the RRM problem model requires $S = N\text{x}M$ *nodes. Their inputs and their outputs are defined by equations* (3.72) *and* (3.73) *respectivally*

$$U_{ij}; i=1,\ldots,N;\ j=1,\ldots,M \tag{3.72}$$

$$V_{ij}; i=1,\ldots,N;\ j=1,\ldots,M\ . \tag{3.73}$$

A Lyapunov (energy) function for this HNN that follows the model of equation (3.67) *can be defined by*

$$E = C(B_T, N, M, V_{ij}) - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} V_{ij}\left(1-V_{ij}\right) + E^{const} \tag{3.74}$$

and so, stable states can be found on the vertices of an appropiate S *-dimensional cube giving a optimal combination of values for* $\left[V_{ij}\right]_{\substack{i=1,\ldots,N \\ j=1,\ldots,M}}$ *that solves tthe RRM problem.*

The first summand in the above equation expresses the criterion to optimize the total bandwith available. This term must be given such that once expanded and rearranged, it could be expressed as in equation (3.67). Due to the working hyphotesis ii), this term has a combinational

selection, V^* represented as a particular set of 1's and 0's of outputs values V_{ij} values satisfying (3.71). Since this term may contain diagonal elements of the T-matrix of the form $\alpha\left(V_{ij}\right)^2$ which are nonzero, we can restore using expression (3.69) in order to have a guarantee that the desired equilibrium solution will be close enough to 0 and 1, that is on the corners of the space.

The second summand is introduced, following the model presented by Hopfield, [27]. Note that this term has minimal value when, for each i, either $V_{ij}=1$ or $V_{ij}=0$. Moreover, the coefficients in this second summand were chosen so as to cancel out the diagonal elements in the first summand. The elimination of diagonal connection strengths will generally lead to stable points only at corners of the space. The second term can also be expanded and rearranged into the form (3.69). This second term is always used in all the energies function associated to HNN formulation.

Last term can be introduced to describe the set of constraints between the set of neurons (time connection, bit rate connection, service available, …) Usually the translation of all these constraints lead to add more than one term. For instance, it could be assumed that one user can not receive more than one resource or that one resource only can be assigned to one user. In the first case it would be necessary to add a term as $-\frac{1}{2}\sum_{j=1}^{M}\left(1-\sum_{i=1}^{N}V_{ij}\right)$. In the second case, it can be added the term $-\frac{1}{2}\sum_{i=1}^{N}\left(1-\sum_{j=1}^{M}V_{ij}\right)$. In chapter 4, 5 and 6 we will use one or the other or both of the two approaches depending on the problem.

Other terms that can be found in energy function formulations are $-\frac{1}{2}\left(\sum_{j=1}^{M}V_{ij}-M\right)^2$, forcing all the resources to be assigned, or $-\frac{1}{2}\left(\sum_{i=1}^{N}V_{ij}-N\right)^2$, forcing all the users to have an assignment. The incorporation of every new term allows us to better model each particular problem restriction. However it is worth to mention that an excessive number of terms can put additional difficulties in the system dynamics and slow down the convergence. Another key issue is to know which weight each term has to be assigned so as to improve the system's convergence. In Chapters 4, 5 and 6 these issues will be considered.

Chapter 4

RRM based on Hopfield Neural Networks for WCDMA

4.1 Introduction

Research in multi-rate wideband code-division multiple access (WCDMA) networks for future wireless communication systems has recently attracted a lot of attention in the light of the new high bit rate data services that will be provided through 3G systems, including video, multimedia, circuit and packet mode data, together with the traditional speech services and low bit rate data. In such scenario, it becomes crucial a proper management of the available radio resources in order to benefit from the traffic multiplexing to maximise the channel utilisation while at the same time keeping the specific service requirements. Specifically, in a WCDMA system, the amount of radio resources allocated to each user is controlled by regulating the transmission powers and the spreading gains (equivalently the transmission rates).

The provision of quality of service (QoS) guarantees in wireless networks has attracted a lot of attention in the last years due to the scarcity of radio resources, time-varying channel conditions, and very diverse QoS requirements of the multimedia users. In that sense, one of the key aspects to achieve an efficient usage of the available radio resources, while at the same time keeping the desired QoS constraints for the different services, is to execute a Dynamic Resource Allocation (DRA) that takes into account both QoS requirements and channel status for each user. The DRA problem supporting packet communication services has been widely treated in the literature. In particular, in [33] and [34] procedures for getting a maximum throughput are proposed in a WCDMA uplink scenario. Similarly, in [35] fairness among users is also introduced while also maximizing the total throughput of the system for a downlink WCDMA system. However, in most of these cases optimal or suboptimal solutions are attained at the expense of a great complexity that could prevent in practice real time operation (i.e. at the frame level scale).

The DRA problem in multi-user and multi-service wireless environments can be considered as a combinational problem where radio resources (e.g. bandwidth) have to be allocated to several users at each frame subject to certain restrictions in terms of Quality of Service (QoS) and of total amount of available resources. In that sense, it is known that a Hopfield Neural Network (HNN) searches for the combinational solution that minimizes an Energy Function under specific constraints. This HNN feature has been already previously introduced in Chapter 3. In addition to that, it is also claimed that real time operation at a very low time scale could be provided either by means of Hardware HNN implementations or in some cases by means of numerical algorithms. Consequently, the use of HNN can be regarded as a firm candidate for implementing the DRA problem in different scenarios whenever a high number of possible combinations of bandwidth allocations to the users exist, thanks to the ability of HNN to find the optimum combination that fits in the specific constraints. In this respect, to cover the DRA issue though from a viewpoint of wireless ATM technology, in [36] a HNN is proposed to manage the multirate and multiservice structure of an ATM network. Nevertheless not much related work has appeared in the literature since then to solve this specific issue.

In the following sections we propose four novel dynamic resource allocation algorithms that make use of the Hopfield Neural Network methodology. In section 4.2, HNN is used to manage the Admission Control Problem in the uplink of a WCDMA system for multirate users modeling different bit rates, Section 4.3 extends this HNN algorithm for managing different service profiles for a user centric approach. Then we focus in the performance assessement of admitted users by introducing a new metric denoted User dissatisfaction probability. Section 4.4 applies HNN to schedule the downlink transmissions in a WCDMA scenario with delay-oriented services. Finally, section 4.5 applies HNN to both schedule the downlink transmissions in a WCDMA scenario with delay-oriented services and real time services.

4.2 Hopfield Neural Networks basis in Admission Control

4.2.1 Introduction

Admission Control is one of the most important RRM functions the mobile operators have to face. The result of executing an Admission Control Algorithm has been in the past just to grant or not a requested voice channel. However in 3G systems like UMTS, Multimedia applications can not be managed in the same way. Users will have multiple choices in terms of the bit rate to be allocated to each one of the multiple connections they can use in a session. We can then

imagine a list of acceptable bit rates a user can apply (e.g Transport formats in UMTS) so that, depending on the signed contract with the network operator and the system load, the operator has to be able to manage them optimally. In addition to that, new users can benefit from a lower blocking probability if the already accepted users tolerate a certain reduction, at the minimum extent, of their allocated bit rates so as to make room for the new user. This is a problem that has not been dealt comprehensibly in a WCDMA scenario. In this section 4.2 we propose a new bit rate adaptive allocation strategy for uplink WCDMA that optimises at the same time the number of users getting access, according to their contract with the operator stated by means of a user profile, and a Cost Function that penalises the unfairness in the resource allocation.

4.2.2 Problem formulation

The proposed strategy starts from the load factor restrictions in uplink WCDMA systems that can be related with the bandwidth allocation as follows:

4.2.2.1 Load constraint

In any WCDMA system, it is required to limit the maximum amount of interference generated by the different transmissions. This can be achieved by limiting the so-called uplink load factor η_{UL} to a maximum value $\eta_{UL\max}$. In the case of a single isolated cell, the load factor is related to the number of instantaneous transmissions *N* by means of the following expression introduced in Section 2.2.4. This expression assumes activity factor equal to 1 as happens for example with streaming services. Then, it follows

$$\eta_{UL} = \sum_{i=1}^{N} \frac{1}{1 + \frac{W}{\left(\frac{E_b}{N_0}\right)_i R_{b,i}}} \leq \eta_{UL\max} \tag{4.1}$$

where *W* is the total transmission bandwidth, $\left(E_b / N_0\right)_i$ is the target for the i-th user and $R_{b,i}$ is the bit rate selected to be allocated to the i-th user. The power control ensures that each user attains the $\left(E_b / N_0\right)_i$ target. N is the number of admitted users in a given cell, $\eta_{UL\max}$ is the maximum planned uplink load factor (that usually takes values around 0.7) and W is the transmission bandwidth. By assuming that, usually, it holds:

$$\frac{W}{\left(\frac{E_b}{N_0}\right)_i R_{b,i}} >> 1 \tag{4.2}$$

and assuming for simplicity that the $\left(E_b / N_0\right)_i$ target is the same for all the users, from the expression (4.1) the sum of allocated bit rates must fulfill the following inequality:

$$\sum_{i=1}^{N} R_{b,i} \leq \frac{W \cdot \eta_{UL\max}}{\left(\frac{E_b}{N_0}\right)} = B_T \,. \tag{4.3}$$

Notice that (4.3) establishes that the maximum bandwidth that can be allocated in the cell is B_T . Actually this condition can be relaxed due the soft capacity featured by WCDMA schemes. Notice also that B_T can be regarded as the equivalent available total bandwidth to be shared among the different users.

4.2.2.2 Bandwidth sharing

In order to make an efficient use of the available bandwidth B_T , the selected $R_{b,i}$ should come out from minimizing the expression

$$\left(B_T - \sum_{i=1}^{N} R_{b,i}\right)^2 \tag{4.4}$$

for $R_{b,i}$ belonging to an allowed set of values.That is, the maximum Bandwidth available B_T should be exploited at the maximum degree among all users by allocating suitable bit rates. Furthermore, to cope with this maximum bandwidth limitation, a new set of variables $\left\{\psi_{i,j}\right\}_{i,j}$ are introduced in the analysis to better manage the user's profile in terms of maximum bit rates contracted and hence allowed. When the j -th bit rate is not allowed for the i -th user (i.e., it is above the maximum permitted) $\psi_{i,j} = 1$ and $\psi_{i,j} = 0$ otherwise. These variables are also used when a user suffers from excessive propagation loss values that prevent a given i -th bit rate from being used because of power constraints in the terminal. That is, in a WCDMA

deployment the greater the user bit rate, the lower the coverage area for a given maximum mobile terminal transmit power as was introduced in Sub-section 2.3.1.2

4.2.2.3 Fairness

In order to be able to guarantee a satisfactory fairness in the allocation process, a fairness condition has to be defined and added to the above requirements. We propose to allocate to the i-th user a bit rate $R_{b,i}$ so as to minimize

$$\sum_{i=1}^{N} \frac{\left(B_{i\max} - R_{b,i}\right)^2}{B_{i\max}{}^2}, \tag{4.5}$$

where $B_{i\max}$ is the maximum bit rate that can be allocated to the i-th user. It is worth mentioning here the Admission Control has to cope with different users profiles (in the sense that they are granted with different maximal transmission bit rates). Then, in the above expression the bit rate allocated relates to the maximum bit rate tolerated.

4.2.2.4 Power constraint

This condition takes into account the fact that the maximum available power in a mobile terminal is limited and, therefore, some bit rates may not be achievable for terminals located at the cell edge.

In particular, the power requirement of the i-th user terminal is given by expression (2.10) as:

$$P_{T_i} = \frac{L_{p,i} P_N}{\frac{W}{\left(\frac{E_b}{N_o}\right)_i R_{b,i}} + 1} \frac{1}{1-\eta_{UL}} \approx \frac{L_{p,i} P_N}{\frac{W}{\left(\frac{E_b}{N_o}\right)_i R_{b,i}}} \frac{1}{1-\eta_{UL}}, \tag{4.6}$$

where $L_{p,i}$ is the path loss corresponding to the i-th user and P_N is the background noise. Consequently, assuming a maximum available power of $P_{T\max}$ the bit rate allocation must take into account the following constraint for the i-th user:

$$P_{T\max} \geq \left(\frac{E_b}{N_o}\right)_i \frac{R_{b,i} L_{p,i} P_N}{W} \frac{1}{1-\eta_{UL}}. \tag{4.7}$$

4.2.3 Optimization process for uplink WCDMA

The Admission Control problem presented above is an RRM optimization process that can be formulated by means of a 2D CHNN introduced in Chapter 3, section 3.5. Let N be the number of users and let R_{ij}, $j = 1,\ldots,M$ be the number of streaming bit rates allowed for user i-th, $i = 1,\ldots,N$ and $R_{b,i}^{j} \in \left\{R_{ij}\right\}_{j=1,\cdots,M}$ the bit rate selected to be assigned to this i-th user. Accordingly, the associated 2D HNN scheme requires *N* times *M* nodes, i.e. $S = N\text{x}M$ nodes. Then, in this 2D HNN node deployment, we identify the output values nodes V_{ij} introduced in Section 3.5.2 with the allocation of bit rate to the user as follows

Definition. *The j-th bit rate is allocated to the i-th user iff $V_{ij} = 1$.*

By the definition, the set of output value nodes gives the following final allocation to the users

$$\begin{gathered} V_{ij} = 1 \Rightarrow R_{b,i}^{j} = R_{ij} \\ V_{ij} = 0 \Rightarrow \text{the allocation of } R_{ij} \text{ is not consider} \\ \text{for the } i\text{-th user; } i = 1,\ldots,N;\ j = 1,\ldots,M \end{gathered}. \tag{4.8}$$

Let us define the following set of cost values

Definition. *We define a set of cost values $\left\{C_{i,j}\right\}_{\substack{i=1,\ldots,N \\ j=1,\ldots,M}}$, where $C_{i,j}$ is associated to the i-th user and for the j-th possible bit rate to be allocated R_{ij}, as follows:*

$$C_{i,j} = \frac{(B_{i,\max} - R_{ij})^2}{\left(B_{i,\max}\right)^2}. \tag{4.9}$$

Then, the cost value in (4.9) is actually set to 0 when $R_{b,i}^{j} = B_{i\max}$, i.e., i-th user is granted with its maximum bit rate. Otherwise the objective is to have a cost as lower as possible.

According [36] a Lyapunov (energy) function for this 2D HNN can be defined by

$$E = \frac{\mu_1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} C_{ij}V_{ij} + \frac{\eta^{\varsigma}\mu_2}{2}\left|1-\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{R_{ij}}{B_T}V_{ij}\right| + \frac{\mu_3}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}\psi_{ij}V_{ij} + \frac{\mu_4}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}V_{ij}(1-V_{ij}) + \frac{\mu_5}{2}\sum_{i=1}^{N}(1-\sum_{j=1}^{M}V_{ij})^2 , \quad (4.10)$$

where

$$\varsigma = u\left(\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{R_{ij}}{B_T}V_{ij} - 1\right) \quad (4.11)$$

and u(.) is the unit step function.

The energy function defined by (4.10) and (4.11) follows the generalized energy function model for RRM problems presented in sub-section 3.5.3. equation (3.74). Notice that, in this case, the term $C(B_T, N, M, V_{ij})$ is given by the first three terms in (4.10), which capture the requirement of the objective function for optimizing the total bandwith available and the constraints to be satisfied. In particular, the first summand incorporates the cost values (4.9) related to the unfairness in the bit rate per user allocation. The second summand penalizes the undesired situations where

$$\sum_{i=1}^{N}\sum_{j=1}^{M} R_{ij}V_{ij} \geq B_T . \quad (4.12)$$

However it is not guaranteed that the aggregation of the allocated bit rates per user is lower or equal to the total Bandwidth B_T available. Fortunately, unlike other 1G or 2G systems, the soft WCDMA nature would tolerate sporadic violations of the above inequality. The third summand simply penalizes non allowed bit rates according to the contractual user profile. That is, for the i-th user, $\psi_{ij} = 0$ if bit rate R_{ij} is allowed and $\psi_{ij} = 1$ otherwise.

The fourth summand, which was presented in sub-section 3.5.3, forces convergence to $V_{ij} \in \{0,1\}$, as ensure rapid convergence to correct and stable states of neurons. The five summand, also mentioned there, forces the physical condition only one R_{ij} is possible for a given i-th user.

The energy function given by (4.10) and (4.11) can be expanded and rearranged into the form (3.67). Then stable states should tend to be found on the vertices of a S - dimensional cube. The Annex B contains the expressions founded for the interconnections matrix $T = \left[T_{kl,ij} \right]_{\substack{i,k=1,\ldots,N \\ j,l=1,\ldots,M}}$ and the bias current vector $I = \left[I_{kl} \right]_{\substack{k=1,\ldots,N \\ l=1,\ldots,M}}$.

As it was stated in Section 3.3.4, the expression 4.10 includes the μ coefficients associated to each term.

The minimisation of the Energy Function (4.10) provides the allocation bit rate of the *i*-th user. Once the final V_{ij} values are achieved as solution of the equations (3.66), the final bit rate assigned to the *i*-th user in one frame is given by the V_{ij} which takes the value 1 for the *j*-th bit rate (i.e. $R_{b,i}^{j}$).

4.2.4 Perfomance evaluation

4.2.4.1 Simulated Scenario

Our WCDMA scenario consists of one isolated circular cell of radius 2 Km. including the propagation model and mobility model (see Annex A). The log-normal shadowing deviation is $10\,\text{dB}$. The noise power is $P_N = -102\,\text{dBm}$. The total transmission bandwidth is 3840 KHz. The available set of bit rates (in Kb/s) belong to the set A = {384, 256, 192, 128, 64 and 32}. Furthermore, the E_b / N_0 target is $5\,\text{dB}$ and $\eta_{UL\max} = 0.7$ which leads to an avalilable bandwith $B_T = 850\,\text{Kb/s}$.

Two kinds of scenarios are considered:

- Homogeneous scenario: All the active users have the same profile, in this case all the users are allowed to use a bit rate in bits/s extracted from the set A

- Heterogeneous scenario: The registered users are assumed to have different profiles. Each profile is defined by the maximum bit rate a user with this profile can transmit. This Maximum bit rate can be any of the bit rates belonging to set A.

In the simulations we have carried out, 6 different profiles associated to the different bit rates of the set A have been assumed. The users have been randomly distributed among the 6 profiles.

Two models to work out have been retained: The ***HNN model*** and the ***Reference model***:

The ***HNN model*** includes the call admission and bit rate adaptive allocations according to the HNN optimisation procedure introduced in sub-section 4.2.3. To this end the numerical iterative solution of (3.45) is obtained following the Euler technique stated in (3.50) with $h = 10^{-4}$ for each k-th node, $k = 1, \ldots, S$ and taking the Euclidian norm and $\varepsilon = 10^{-6}$ in (3.51). The T_{kj} and I_k values are obtained in Annex B. The selected parameters appearing in the formulation of the HNN, , and the energy function considered, (4.10), are $\mu_1 = 1000$, $\mu_2 = 4000$, $\mu_3 = 8000$, $\mu_4 = 800$, $\mu_5 = 6000$, $\tau = 1, \alpha = 1.0$ and $\eta = 10$. These values have been set based on [36]

The ***Reference model*** is used by comparison purposes and it simply does not consider the adaptive allocation of resources. In other words it is a rigid model in which a user is granted with the requested bit rate or otherwise it is blocked.

The user requesting connection is blocked whenever the allowed Maximum bit rate due to the breathing effect, that is including propagation impairments, is below the minimum bit rate possible, 32 Kb/s in our case. A blocking situation would also occur in HNN enabled systems when all the active users were assigned the minimum bit rate of 32 Kb/s and a new user would request a link to get through. For the Reference Model, not including a HNN procedure, in addition to the breathing effect induced blocking, there appears the conventional blocking when the bit rate requested can not be granted because the system is already operating at the full load.

The traffic model assumes Poisson arrivals and exponential distributed service times. Then, the simulated sessions are exponentially distributed with mean duration $T_{m,i}$. The arrival rate is formulated as

$$\lambda_i = \frac{1}{T_i + T_{m,i}}, \tag{4.13}$$

where T_i is the mean interarrival time between users assumed, also exponentially distributed. The resulting traffic load is:

$$\rho = \sum_i \lambda_i T_{m,i} = \sum_i \frac{1}{T_i + T_{m,i}} T_{m,i} \,. \tag{4.14}$$

Actually, rather short sessions have been simulated in order to speed up the simulation effort. So, in our simulations we have set $T_i = 0.3\,\text{s}$ and $T_{m,i} = 0.5\,\text{s}$. The frame duration assumed to update the allocated bit rates is set to $T_a = 0.1\,\text{s}$, so as, at the beginning of each active frame a Hopfield optimisation is carried out, and as a result of that, some selected granted bit rates come out and are kept constant along the frame.

Regarding to the mobility model, the considered mobile speed is 60 Km/h. When a user finishes a session it is moved randomly to another location within the cell just to simulate a different user when it comes into activity again after on average T_i s. later.

4.2.4.2 Simulation results

The calculation of the average bit rate assigned (ABR) with the corresponding blocking probability are the retained key performances indicators pursued as simulation results.

Results using HNN in the Homogenous scenario

Figure 4-1 shows ABR, resulting from the simulations, measured as the fraction (percentage) of bit rate with respect to the maximum allowed bit rate versus the path loss in a homogeneous scenario with a single user profile (i.e. in the case with all the users having the same maximum bit rate of 384 Kb/s) for different simulated number of frames.The results are shown for a number of simulated frames ranging from 5000 to 40000. Notice that the ABR decreases for high path losses revealing the ability of the algorithm to take into account the terminal power constraints due to the cell breathing effect. The convergence is very good in all the significant cases, i.e. for losses greater than 70 dB approximately. Due to the low number of users experimenting losses lower than 75 dB in the cell, the convergence is much slower for these values. In fact less than 2% of users moving around the cell according to the mobility model present losses lower than 70 dB.

The ABR over the considered actual coverage area would be about the 40% of the maximum bit rate demanded of 384 Kb/s. It is worth to notice as for a Path Loss greater than 115 dB approximately the Average Bit Rate decrease quickly, indicating we are moving toward farther locations within the cell and consequently the breathing effect appears.

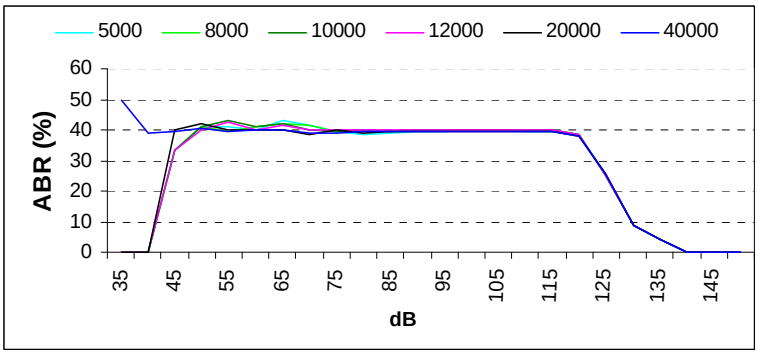

Figure 4-1 Average Bit Rate allocation for a homogeneous scenario with 10 users.

Figure 4-2 and Figure 4-3 show the load impact on the allocated Average Bit Rate. It is shown as the lower the load the higher the Average Bit Rate attained.

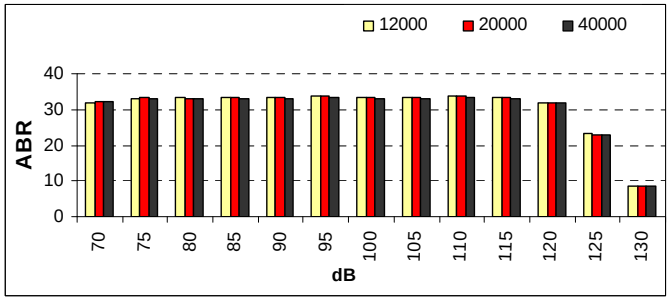

Figure 4-2 Average Bit Rate allocation for a homogeneous scenario with 12 users.

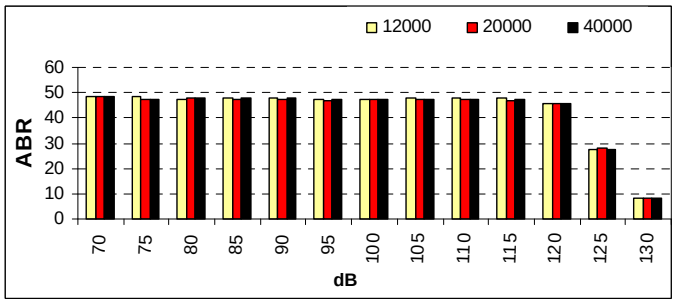

Figure 4-3. Average Bit Rate allocation for a homogeneous scenario with 8 users.

The above figures need to be complemented with the blocking rate. Then Figure 4-4, Figure 4-5 and Figure 4-6 show this blocking rate for the same load figures considered above.

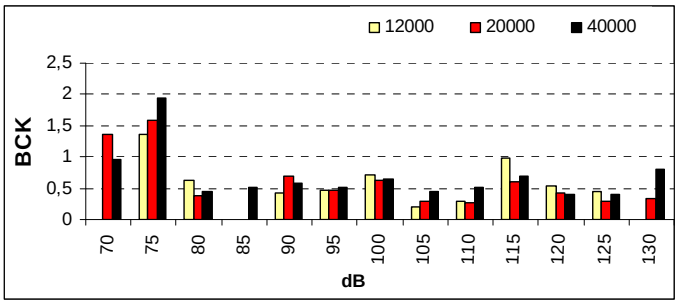

Figure 4-4. Blocking rate for a homogeneous scenario with 8 users.

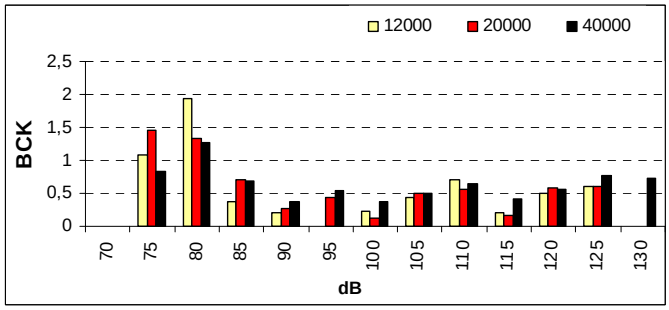

Figure 4-5 Blocking rate for a homogeneous scenario with 10 users.

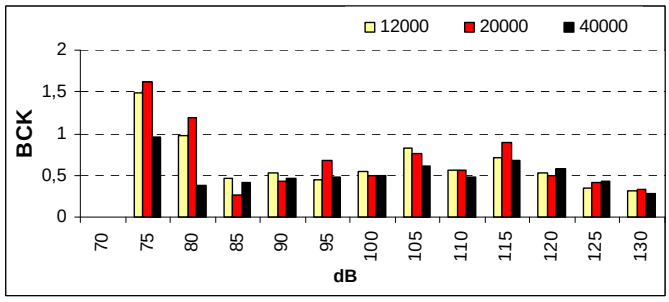

Figure 4-6 Blocking rate for a homogeneous scenario with 12 users.

It can be noticed as the convergence is rather good with the number of frames considered. Blocking percentages lower than 1% would correspond to the desired values in practice. It is worth to note as the blocking rate is not load sensitive. That indicates that the blocking, in practice, depends on the propagation losses rather than on the system load. In other words, increasing the load certainly makes the Average Bit Rate decrease but not to the extent of blocking new users. Due to the fact that the load factor $\eta \le \eta_{\max} = 0.7$, increasing the traffic load in terms of ρ, simply distributes the assigned bit rates but does not change the blocking rate. Certainly for high values of traffic load (not shown in the Figures) the blocking rate will end up increasing as it will not be longer possible to allocate all the users requesting service even at the lower bit rates.

Results using HNN in the Heterogeneous scenario

Figure 4-7, Figure 4-8 and Figure 4-9 show the Average Bit Rates obtained when considering the more realistic heterogeneous scenario. In particular two user`s profiles have been used considering the maximum bit rate allowed in each one 384 Kb/s and 128 Kb/s respectively. The same loads as before have been considered. It is worth to notice as the obtained results greatly improve those obtained in the homogenous scenario. Now it appears as the bit rate allocations can be managed better as much as there are more users satisfied with less bandwidth. This is also a very interesting case that resembles much more a realistic multimedia environment with multiple choices of bit rates for combinations of different traffic loads such as voice, images, data, web browsing, streaming, etc.

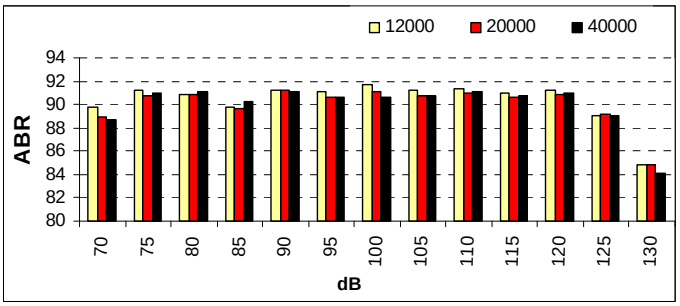

Figure 4-7 Average Bit Rate allocation for a heterogeneous scenario with 8 users.

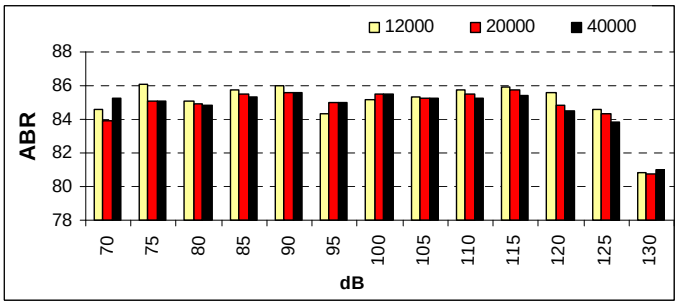

Figure 4-8 Average Bit Rate allocation for a heterogeneous scenario with 10 users.

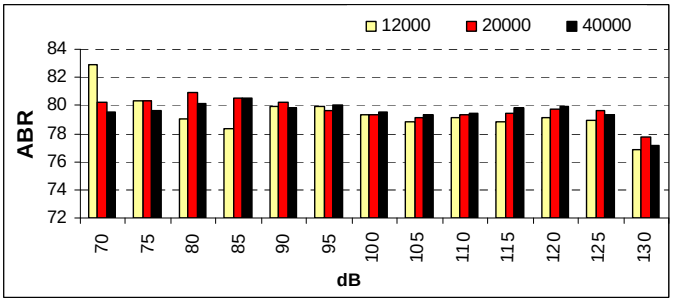

Figure 4-9 Average Bit Rate allocation for a heterogeneous scenario with 12 users.

Figure 4-10 and Figure 4-11 show as the blocking probability is kept at similar margins than obtained with those obtained for the homogenous scenario.

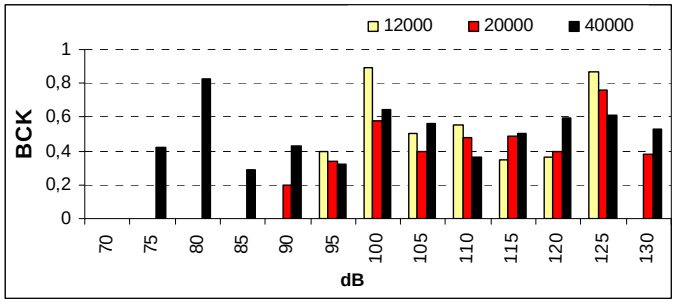

Figure 4-10. Blocking rate for a heterogeneous scenario with 8 users.

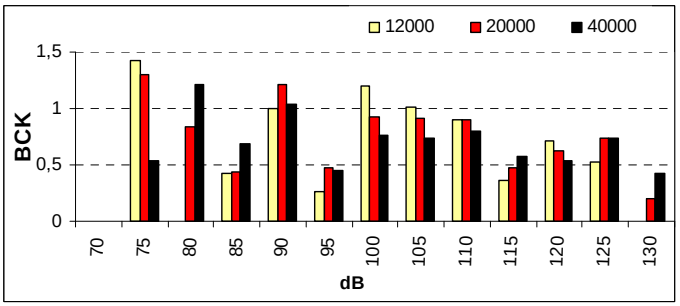

Figure 4-11 Blocking rate for a heterogeneous scenario with 12 users.

Figure 4-12 shows the Average Bit Rate per each user class. Each user class is defined by the maximum allowed bit rate contracted. It can be noticed as users with the lower allowed maximum bit rate get from the network a better Average Bit Rate than those users with higher allowed maximum bit rate. That can be explained taking into account the HNN algorithms always try to make room to a new user. Then, if this is a user with a low allowed maximum bit rate class, always the actual allocated bit rate will be closer to the maximum bit rate allowed than in case of being a high maximum bit rate class of user.

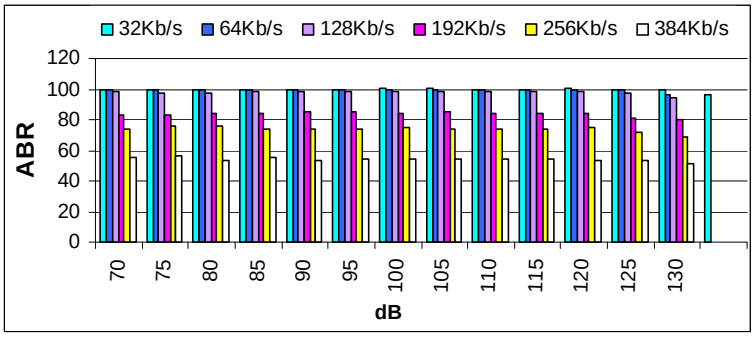

Figure 4-12 Average Bit Rate allocation for classes and a heterogeneous scenario with 10 users.

Results for a Fixed Allocation Strategy

The following results show some obtained results for the same scenario when no HNN strategy is used. That is, we allocate the maximum bit rate allowed in the homogeneous scenario, considered as a fixed bit rate and according to the reference model previously stated in sub-section 4.2.3.1 for comparison reasons. Figure 4-14 and Figure 4-15 show that the blocking probability under this strategy is quite unsuccessful.

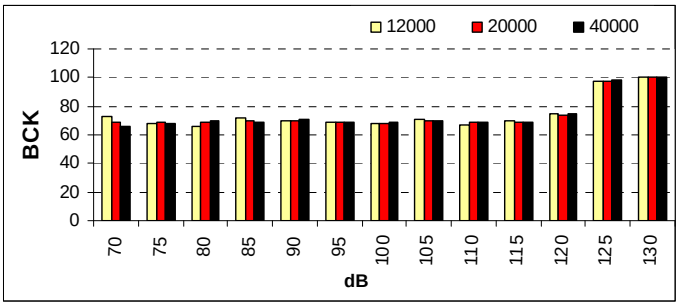

Figure 4-13 No HNN strategy. Blocking rate for a homogeneous scenario with 8 users. and maximum bit rate of 384 Kb/s

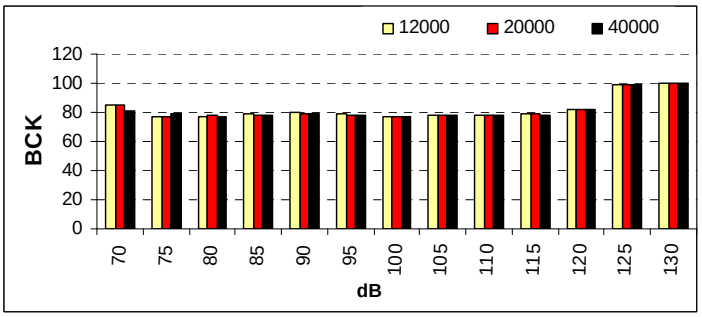

Figure 4-14 No HNN strategy. Blocking rate for a homogeneous scenario with 12 users. and maximum bit rate of 384 Kb/s

It could be argued that a less elaborated allocation algorithm could give similar results than those obtained with the HNN methodology. That is, if the Average Bit Rate is about 40% of the maximum bit rate of 384 Kb/s in the homogenous scenario HNN simulated cases, then by allowing the users to use, let's say, a similar bit rate of about 40% of 384 Kb/s, that is, about 192 Kb/s but being it now a fixed bit rate, perhaps the results could be the same than those obtained with adaptive allocation. Figure 4-15 and Figure 4-16 however show that the blocking rate with this simple strategy increase dramatically, not being acceptable in a real situation.

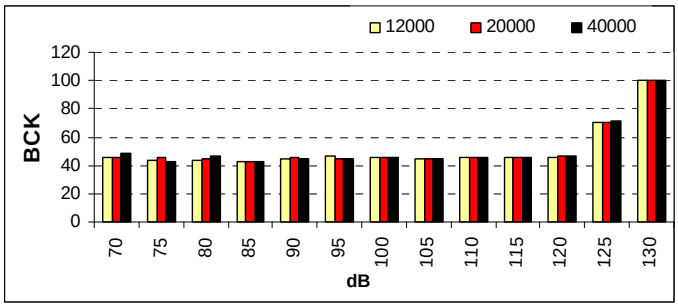

Figure 4-15 No HNN strategy. Blocking rate for a homogeneous scenario with 8 users. and maximum bit rate of 192 Kb/s

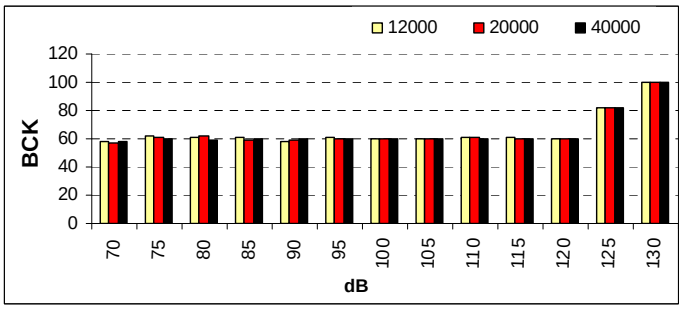

Figure 4-16 No HNN strategy. Blocking rate for a homogeneous scenario with 12 users. and maximum bit rate of 192 Kb/s

Finally, Figure 4-17 summarizes the blocking probability for the above two cases in the reference model.

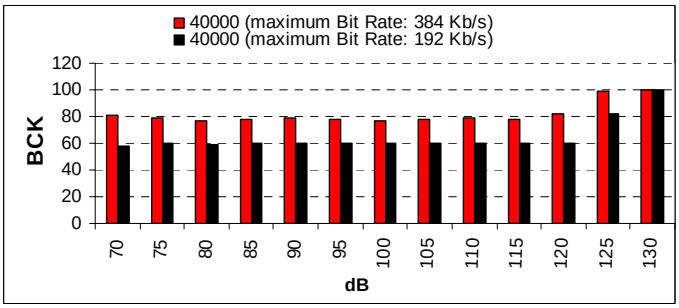

Figure 4-17 No HNN strategy. Comparing blocking rate for a homogeneous scenario with 12 users.

4.2.4.3 Error Analysis

It was already stated in the expression (4.3) that

$$\sum_{i=1}^{N} R_{b,i} \leq \frac{0.7W}{\left(\frac{E_b}{N_0}\right)_i} = B_T \,. \tag{4.15}$$

However, it it was also mentioned that this bound could be relaxed for WCDMA systems due to its soft capacity features. Nevertheless it could be interesting to check the deviations from the above inequality that in all cases should be as small as possible. Figure 4-18, Figures 4-19, Figure 4-20 and Figure 4-21 show the obtained results for a load $\rho = 7.5$. In particular, they show the number of frames out of 40000, in which the above inequality is not verified. It can be noticed in the histogram as deviations are clustered around certain values. That is due to the finite number of combinations of bit rates. The most important feature is however that the maximum deviation around 110 Kb/s is within the 10-15 %, which could be acceptable, and represents hardly the 5% of cell users. This also can be depicted in the Figure 4-18. Certainly lower loads than $\rho = 7.5$, will improve these deviation figures as shown in Figure 4-20 and Figure 4-21

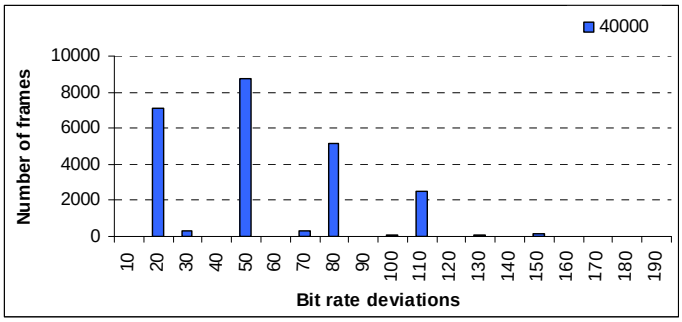

Figure 4-18 Error analysis for a heterogeneous scenario with 12 users.

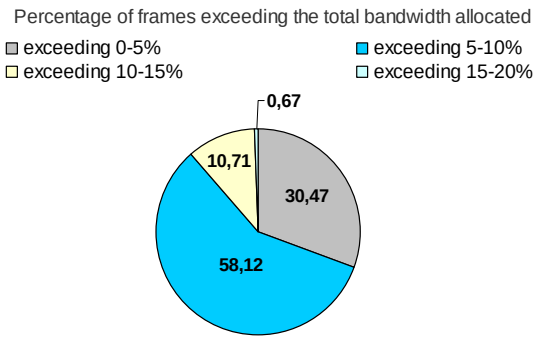

Figure 4-19 Error analysis for a heterogeneous scenario with 12 users.

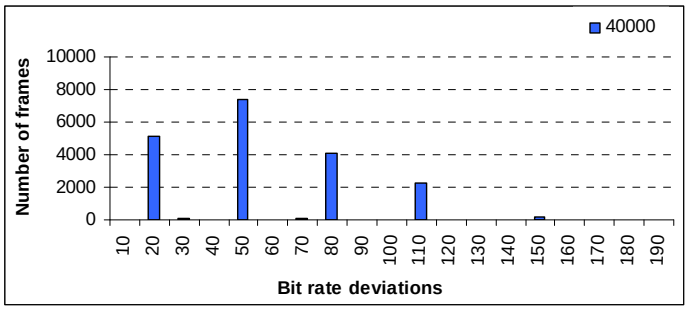

Figure 4-20 Error analysis . for a heterogeneous scenario with 10 users.

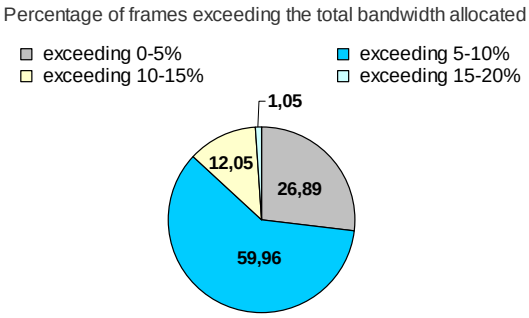

Figure 4-21 Error analysis for a heterogeneous scenario with 10 users.

4.2.4.4 Convergency issues

The Energy function defined by expressions (4.10) and (4.11) is a real valued function of several variables with domain in R^S, i.e for every $(\cdots,V_{ij},\ldots)\in R^S$, $E\left((\cdots,V_{ij},\ldots)\right)\in R$. In the domain definition embraces two regions:

For all $(\cdots,V_{ij},\ldots)\in R^S$ such that $\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{R_{ij}}{B_T}V_{ij}<1$ we have

$$E_1=\frac{\mu_1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}C_{ij}V_{ij}+\frac{\mu_2}{2}\left(1-\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{R_{ij}}{B_T}V_{ij}\right)+$$
$$\frac{\mu_3}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}\psi_{ij}V_{ij}+\frac{\mu_4}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}V_{ij}(1-V_{ij})+\frac{\mu_5}{2}\sum_{i=1}^{N}(1-\sum_{j=1}^{M}V_{ij})^2, \quad (4.16)$$

For all $(\cdots,V_{ij},\ldots)\in R^S$ such that $\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{R_{ij}}{B_T}V_{ij}>1$ we have

$$E_2 = \frac{\mu_1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} C_{ij}V_{ij} + \frac{\eta\mu_2}{2}\left(-1+\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{R_{ij}}{B_T}V_{ij}\right)+ \frac{\mu_3}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}\psi_{ij}V_{ij} + \frac{\mu_4}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}V_{ij}(1-V_{ij}) + \frac{\mu_5}{2}\sum_{i=1}^{N}(1-\sum_{j=1}^{M}V_{ij})^2. \quad (4.17)$$

Then, in each region the energy function keeps its quadratic form, but its derivability is lost in the S-dimensional curves which define the border between the above two regions. As the equation driving the CHNN dynamic system is (see expression (3.68))

$$\frac{dU_{kl}(t)}{dt} = -\frac{1}{\tau}U_{kl}(t) - \frac{\partial E}{\partial V_{kl}}; \quad k=1,\ldots,N; l=1,\ldots,M$$

the energy function associated to the resulting trajectory $U(t)=(\cdots,U_{kl}(t),\ldots)$ would change its value when crossing the region's border. This can affect in both a negative and a positive way the convergence time to a stable solution. The negative impact occurs when the trajectory is close to a stability solution and is forced to change its trend. The positive impact is when the solution we are approaching is not a stable solution and by changing the trajectory we have more chances to approach before the correct solution.

In these circumstances it is really difficult to prove formally convergence, although in practice as the formulation above suggest, the simulation results we have obtained have shown that this convergence is attained either through a smooth convergence process of the HNN Energy function as in Figure 4-22 or by means of a rather involved convergence process as shown in Figure 4-23.

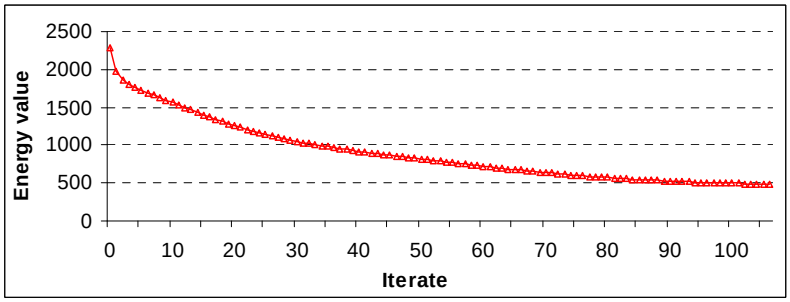

Figure 4-22 Results of convergence in the evolution of the HNN

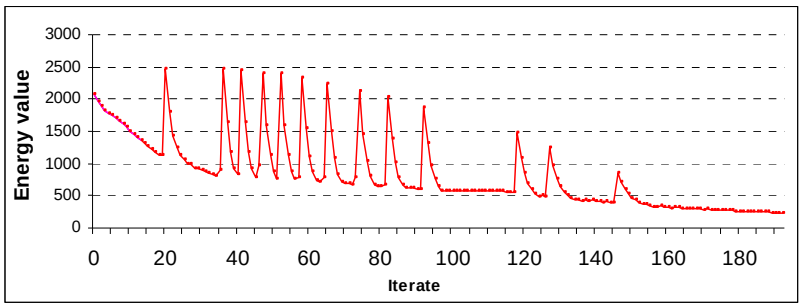

Figure 4-23 Results of convergence in the evolution of the HNN

It can be noticed that the convergence is still attained even in the cases like that illustrated in Figure 4-23 where a trajectory crosses the region's borders, but now the convergence process is disturbed by the presence of sudden peacks that we conjectured are due to the changes of the polarity in the step function appearing in the Energy function definition. When this function becomes positive, then it appears a sudden positive jump in the Energy value or peack that is rapidly corrected. Then the convergence process evolves as usual while the step function in the HNN Energy function formulation is null. This peacks irruption occurs several times until the convergence is finally reached.

4.2.5 Conclusions

Section 4.2 has presented a novel Dynamic Resource Allocation approach for a multi-rate WCDMA wireless network supporting continuous real-time communication. A fair mechanism for the Admission Control has been developed that allocates the available resources according to the specific contracted user profile. For this purpose, a Hopfield Neural Network optimization methodology has been retained with the introduction of a proper cost function characterising the user requirements. It has been shown that the considered approach is able to maintain the blocking probability below its target value while still keeping the best exploitation of the maximum available bandwidth. Result compare favourably with the so-called fixed strategy. It has also been proved that the proposed scheme is robust enough for a soft capacity defined system as WCDMA.

4.3 A User-Centric Approach for Dynamic Resource Allocation in the Uplink

4.3.1 Introduction

In the previous Section we have faced the admission phase and seen as the blocking rate can be dramatically improved when a HNN strategy with variable bit rate users is retained. In this section we will focus on the performace the admitted users can achieve. To this end we have assessed performances different from the blocking rate.

The problem of dynamic resource allocation (DRA) in a multi-service WCDMA wireless network supporting non real-time (NRT) communication services has been widely treated in the literature (see [33], [34] and references therein). However, in most of the cases, a centric-network vision has been the common way to face the DRA, usually with the objective of maximizing the aggregate NRT throughput under different network constraints. As a consequence of that, in many cases, users are treated according to the propagation losses they suffer, so that throughput is maximised by allowing the transmissions of the users with the lowest path loss. Therefore, static and pedestrian users located far from the base station could be discriminated against other closer users.

In this section, a fair user-centric approach is incorporated in the dynamic resource allocation problem, specifically, the actual bit rate allocated per user will be granted according to a set of cost values that incorporates the current achieved bit rate. On the other side the closed loop power control algorithm will provide to each user the required Eb/No target regardless its position in a given instant as in UMTS systems, so the user's performance in term ob bit error rate is guaranteed.

The remainder of the section is organized as follows. In sub-section 4.3.2, we describe the system model and formulate the resource allocation problem by introducing the fairness criterion and some WCDMA wireless network restrictions. In sub-section 4.3.3, we present the 2D HNN optimization procedure used to allocate as most resources as possible but still guaranteeing a certain user satisfaction probability in agreement with the user profile. In sub-section 4.3.4, we present numerical results, and in sub-section 4.3.5 conclusions are summarised.

4.3.2 DRA problem formulation

A WCDMA wireless network is considered where specific bit rates have to be allocated to a set of real time users, already admitted in the network, requiring to transmit continuously and according to different bit rates depending on the radio channel availability. In such user-centric approach with different users' service profiles, performance indicators other than the used as for the Admission Control like the blocking rate as well as traditional ones (e.g. outage probability, BLER, etc.) must be considered to capture the degree of fulfillment of the contracted bit rate. Specifically, the probability that a user is not satisfied with the obtained bit rate will be retained as a mechanism to assess the system performance. Furthermore, different types of users with different rights coexist in the system according to their user profiles. In particular, each user belongs to a user class characterised by the contractual bit rate, which is the maximum bit rate that can be allocated to the user.

It has been assumed that the allocation process is the same one already used in Section 4.2. That means it is carried out periodically in periods of $T_a = 0.1\,\text{s}$, denoted in the following as frames. Let assume that there are N users $i = 1, \ldots, N$ in the cell in the current frame, and that the i-th user has a set of M possible bit rates. The objective of the DRA problem is then to select for each user an optimum allocated bit rate $R_{b,i}$, taking into account a set of constraints as explained in the following.

Being WCDMA an interference-limited technique, the aim of the DRA algorithm is to allocate the highest possible bit rate per user subject to a maximum planned uplink cell load factor, which allows keeping the total interference under certain limits. A fast closed loop power control strategy is assumed like in UMTS aiming to keep the required E_b / N_0 target. The i-th user's satisfaction degree with the experienced service is measured with the non-satisfaction probability

$$P_{ns,i} = \Pr\left(R_{b,i} < \frac{R_{b,i\,\max}}{D_i} \right), \tag{4.18}$$

where $R_{b,i}$ is the bit rate experienced by the i-th user, $R_{b,i\max}$ is the maximum contractual bit rate for the i-th user according to its user class and D_i is a parameter to be selected according to the user contract and represents the fraction of the maximum bit rate that the user still accepts as

satisfactory. A plausible fairness rule will be included in the HNN formulation in order to guarantee a minimum value of the above probability $P_{ns,i}$ according to the agreed user contract.

On the other hand, the WCDMA technique imposes certain restrictions to the above optimization like Load constraints and Power constraints as already formulated in Section 4.2 and summarized in the following as:

4.3.2.1 Load constraint

In any WCDMA system, it is required to limit the maximum amount of interference generated by the different transmissions. This can be achieved by limiting the so-called uplink load factor η_{UL} to a maximum value $\eta_{UL\max}$ .Then, assuming for simplicity that the E_b / N_0 target is the same for all the users, the sum of allocated bit rates must fulfil the inequality already shown in (4.3).

4.3.2.2 Bandwidth sharing

The efficient exploitation of the available bandwidth B_T in terms of the allocated bit rate $R_{b,i}$ to the i-th user can be obtained by minimizing the equation (4.4), which ensures that the maximum aggregated throughput is still attained in this scenario.

4.3.2.3 Power constraint

This condition takes into account the fact that the maximum available power in a mobile terminal is limited. Assuming a maximum available power of $P_{T\max}$ the bit rate allocation must take into account the following constraint for the i-th user already shown in equation

4.3.3 Optimization process for DRA

The DRA problem formulated in the previous sub-section sets the requirement to find the optimal bit rate allocation for the *N* different users given the available bandwidth B_T , the user satisfaction degree and the Load and Power constraints posed by the WCDMA technique. This is an RRM optimization process that can be formulated in the 2D CHNN introduced in the Chapter 3, section 3.5. The same notation as in Section 4.2 will be used. That is, let be N the

number of users and let R_{ij}, $j=1,\ldots,M$ be the number of streaming bit rates allowed per user i-th, $i=1,\ldots,N$ and $R_{b,i}^{j} \in \left\{R_{ij}\right\}_{j=1,\cdots,M}$ the bit rate selected to be assigned to the i-th user. Accordingly, the associated 2D HNN sheme requires $S = N\text{x}M$ nodes. As in Section 4.2 we named V_{ij} .the output value nodes and identify them with the allocation of bit rate to the user as follows

Definition. *The j-th bit rate is allocated to the i-th user iff $V_{ij}=1$.*

The final allocation to the DRA problem is given by (4.8), indicating the allocation of the j-th bit rate, $j=1,\ldots,M$, to the i-th user, $i=1,\ldots,N$.

According to (4.18) the i-th user is not satisfied in the k-th frame whenever the allocated bit rate $R_{b,i}^{(j)}$ is below the satisfaction bound $R_{b,i\max} / D_i$. Then, a fairness rule taking into account the user satisfaction degree is formulated through the following dynamic set of cost values, changing in each frame.

Definition. *We define a set of cost values $\left\{C_{i,j}^{k}\right\}_{\substack{i=1,\ldots,N \\ j=1,\ldots,M}}$, where $C_{i,j}^{k}$ is associated to the i-th user in the k-th frame and for the j-th possible bit rate to be allocated R_{ij}, as follows:*

$$C_{i,j}^{k} = \begin{cases} 0 & \text{if} \quad R_{ij} \geq \dfrac{R_{b,i\max}}{D_i} \\ p_{ns,i}^{*} - \hat{p}_{ns,i}^{k-1} & \text{if } \left(\hat{p}_{ns,i}^{k-1} < p_{ns,i}^{*}\right) \text{ and } R_{ij} < \dfrac{R_{b,i\max}}{D_i} \\ 1 & \text{if } \left(\hat{p}_{ns,i}^{k-1} \geq p_{ns,i}^{*}\right) \text{ and } R_{ij} < \dfrac{R_{b,i\max}}{D_i} \end{cases}, \tag{4.19}$$

where $p_{ns,i}^{*}$ is the target probability that the i-th user is not satisfied. In turn, $\hat{p}_{ns,i}^{k-1}$ is an estimation of the current non-satisfaction probability, measured up to frame $k-1$. It is computed as the ratio of non-satisfied users belonging to the same user class of the i-th user with respect to the total number of users of this class.

Notice that, for bit rates *j* above the satisfaction bit rate bound $R_{b,i\max} / D_i$, the cost function is low (i.e. 0) meaning that these bit rates are suitable for the i -th user. In turn, for lower bit rates, it depends on the non-satisfaction experienced by the user up to frame $k-1$: if the non-satisfaction is above the desired target, the cost is high (i.e. 1), meaning that higher bit rates are desirable, while if the non-satisfaction is below the target, the cost decreases meaning that the user could still accept these low bit rates.

An optimization problem subject to the restrictions as formulated in the 2D CHNN introduced in the Chapter 3, section 3.5 follows. The Lyapunov (Energy) function to be minimized has the same structure that the ones already formulated and described in Section 4.2, (4.10) and (4.11). However, now the cost values incorporate the user centric metric Cost introduced in (4.19)

The minimisation of the Energy Function (4.10) provides the final allocated bit rate of the *i*-th user. Once the final V_{ij} values are achieved as solution of the equations (3.66), the final bit rate assigned to the *i*-th user in one frame. $R_{b,i}^{j}$, is given for the *j*-th bit rate satisfiying V_{ij} =1

4.3.4 Performance evaluation

In order to obtain performance results we consider A = {256, 128, 64, 32 and 16}.and a simulation heterogenous scenario as used in the admission phase in Section 4.2, now we define with two user's profile. Furthermore, the same parameters' values for the HNN model and the energy function (sub-section 4.2.4.1) have been retained. In particular for the heterogeneous scenario Figura 4-24 shows the histogram and the cumulative distribution function (CDF) of the bit rate allocated to Class 1 and 2 users for a moderate-high load of ρ=6.25 (corresponding to $N=10$ users) and differente mix of user profiles .Then it can be noticed in the CDF that the non-satisfaction probability (i.e. the probability of having a bit rate below 128 Kb/s for class 1 users and 64 Kb/s for class 2 users) is below the stated target value of $p_{ns,i}^{*}=20\%$ and the allocated bit rates are most of the time the maximum permitted (i.e. the probability of having a bit rate below 256 Kb/s for class 1 users and 64 Kb/s for class 2 users). can be easily obtained. It is worth noticing as for this traffic load the system can not provide $p_{ns,i}^{*}=20\%$ for a percentage of Clase 1 users as large as 50%, 40% and 40%. On the contrary a 20% of class 1 users and a 80 % of class 2 users could be satsfactory, indicating these figures that too many class 1 users can not be suppported by the system due to their high bit rate requirements.

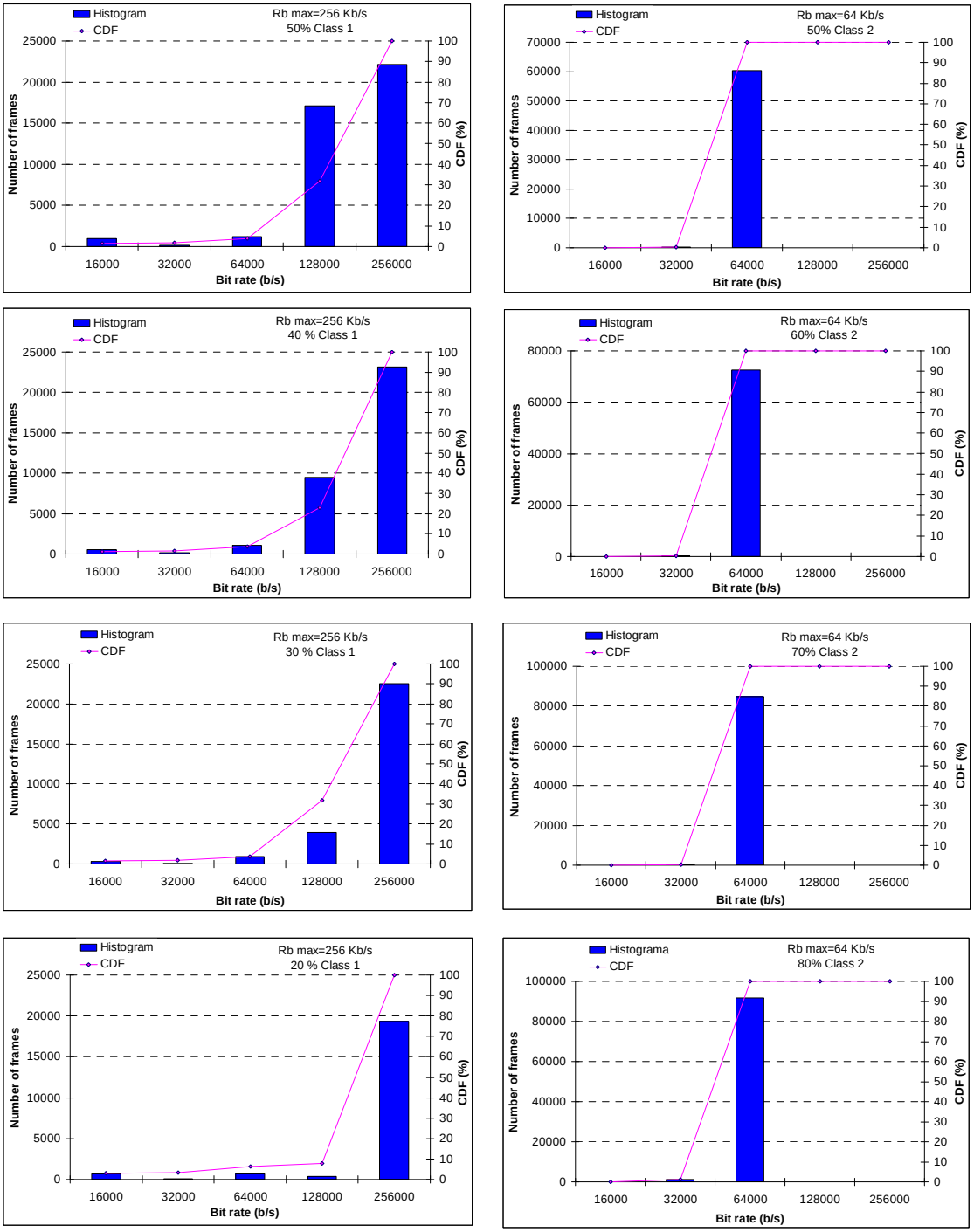

Figure 4-24 Histogram and CDF of the allocated bit rates for different two user classes in a heterogeneous scenario with 10 users.

Figure 4-25 shows the histogram and the cumulative distribution function (CDF) of the bit rate allocated for the same traffic load ρ=6.25 shown in Figure 4-24. Class 1 users keep the maximum bit rate of 256 Kb/s. while Class 2 Users have a maximum bit rate of 128Kb/s instead of 64 Kb/s. Certainly the systems in such circunstances is more capacity demanding. It can be noticed in the CDF that the non-satisfaction probability is below the stated target value of

$p^*_{ns,i} = 20\%$ and the allocated bit rates are most of the time the maximum permitted. Finally, in Figure 4-26 the load has significantly increased up to ρ=8.25 and the maximum permitted bit rates are allocated in a lower number of situations, but the target objective of 20% non-satisfaction probability is still kept. By increasing the traffic load beyond this traffic load the simulations indicated the non-satisfaction probability overcome its allowe limit.

In all the cases, the blocking probability results in values below 0.5%.

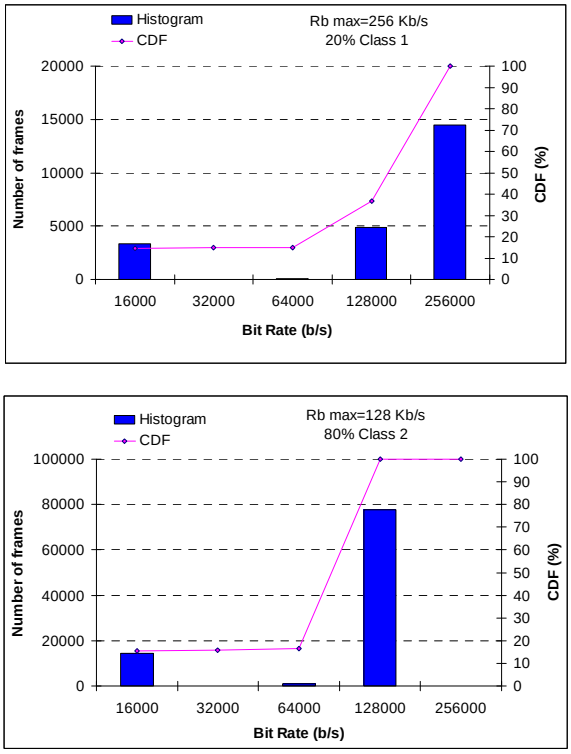

Figure 4-25 Histogram and CDF of the allocated bit rates for the two user classes in a heterogeneous scenario with 10 users.

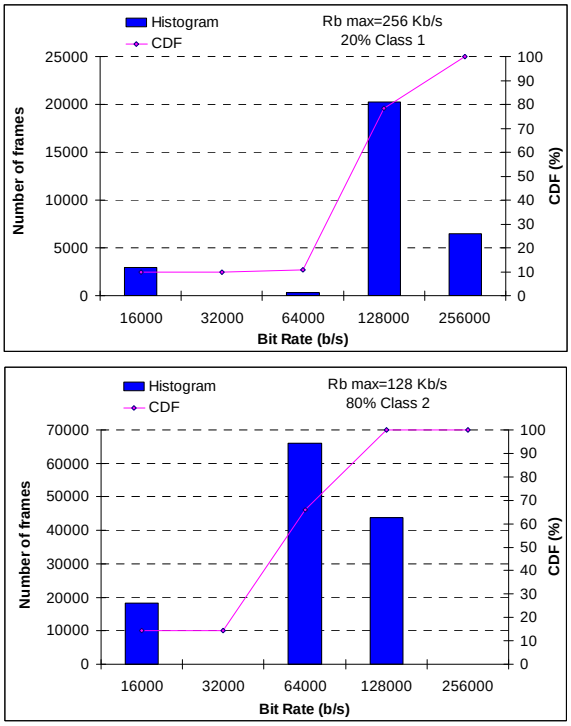

Figure 4-26. Histogram and CDF of the allocated bit rates for the two user classes in a heterogeneous scenario with 14 users.

It can be noticed in the above Figures that a trade-off appears in relation to the dynamic set of cost values (first term in the Energy formulation (4.10)) and the best bandwidth partitioning among users (second term in the above Energy formulation). In other words, the algorithm tends to exploit the maximum available bandwidth (i.e. allocating the maximum permitted bit rates whenever possible), while keeping at the same time the target non-satisfaction probability.

4.3.5 Conclusions

Section 4.3 has presented a novel Dynamic Resource Allocation approach for a multi-rate WCDMA wireless network supporting continuous real-time communication. A user-centric instead of a network-centric approach has been introduced aimed at developing a fair mechanism that allocates the available resources according to the specific contracted user profile. For this purpose, a Hopfield Neural Network optimization methodology has been retained with the introduction of a proper set of cost values changing in each frame and characterising the user requirements. It has been shown that the considered approach is able to maintain the so-called non-satisfaction user probability below its targeted value while still keeping the best exploitation of the maximum available bandwidth. Significant and promising results have been obtained revealing that the proposed strategy performs satisfactorily.

4.4 Delay services in the downlink direction

4.4.1 Introduction

As it has been shown in the Sections 4.2 and 4.3, Dynamic Resource Allocation can be considered as a combinational problem where radio resources (e.g. bandwidth) have to be allocated to several users at each frame subject to certain restrictions in terms of QoS and of total amount of available resources. In that sense, it has been shown as the Hopfield Neural Networks that search for the combinational solution that minimizes an Energy Function under specific constraints can work properly. In turn, it has been shown a good HNN behavior when managing different service profiles in the uplink of a WCDMA system, in which each user has a range of bit rates that can be allocated depending on its service profile, defined by a minimum satisfaction bit rate. Consequently, the use of HNN can be regarded as a firm candidate for implementing the DRA problem in different scenarios whenever a high number of possible combinations of bandwidth allocations to the users exist, thanks to the ability of HNN to find the optimum combination that fits in the specific constraints.

Taking into account the above framework, in this section the basis of the HNN is captured for solving a downlink scheduling problem in a WCDMA packet network where sensitive delay guarantees are to be provided to the users sharing the access. In this case the main overall goal is not necessarily to achieve the maximum throughput, but to deliver each packet without exceeding a specific time deadline. This can be applicable usually to interactive data services but also to real time services, like streaming, where packets lost due to their late arrival affect the reconstruction quality of future elements in the media stream, due to the predictive modules used in the source codecs. A number of transmission scheduling schemes have been recently proposed in the literature, mainly as wireless fair queuing schemes like [37] and [38]. Similarly, in [39] and [7] heuristic scheduling algorithms based on prioritization according to maximum delay bounds and considering different user profiles are presented, although fair scheduling is not specifically addressed. The concept of credit-based scheduling to guarantee bit rate requirements is addressed in [40] and [41], where a practical scheduling algorithm is proposed to guarantee throughput and assess the delay-jitter of streaming video services. Credit-based scheduling schemes are simple to implement but they can not offer tight delay bounds. Dynamic fair scheduling schemes are also proposed in [42] that circumvent this problem by using a Generalized Processor Sharing (GPS) fair scheduling based approach. Also, a solution for the uncontrolled dropping packet rate usually present with these delay-based approaches is covered

in [43] in the context of a Multi-Carrier Code Division Multiple Access (MC-CDMA) scenario relaying on a specific MAC scheme.

Unlike the previous Sections, the HNN methodology is here used for delay-oriented services, which involves the proposal of an adequate scheduling metric and a buffer management strategy able to guarantee the delivery of each packet before its deadline exceeds when several user profiles (i.e. with different delay constraints) appear in a multiuser context. Moreover, unlike most of the previous and current research in wireless scheduling, which focus on the scheduling issue of one single server, our downlink WCDMA HNN-based approach corresponds to a multiple server case, where multiple concurrent packets can be transmitted by the base station simultaneously and where there is not much related work available. The optimum solution is found by solving a nonlinear constrained optimization problem. The HNN here proposed is claimed to do that in a very low time scale of several milliseconds (the typical value for a transmission time interval in a wireless system). The rest of the section is organised as follows. Sub-section 4.4.2 presents the problem formulation and sub-section 4.4.3 provides the proposed HNN-based scheduling approach, which will be compared against a reference scheme presented in sub-section 4.4.4. Results are presented in sub-section 4.4.5 and finally conclusions are summarized in sub-section 4.4.6.

4.4.2 Problem formulation

The considered DRA problem assumes a set of N users, $i = 1,\ldots,N$, with their corresponding queues, located at the base station of the access network, which contains the packets pending to be transmitted in the downlink direction of a CDMA system, as illustrated in Figure 4-27. It is considered that non-shaped traffic is entering the queues so that all the incurred packet delay is introduced at the network level. The scheduling algorithm operates in frames of $T(s)$ and allocates a certain bit rate to each user from a set of M possible bit rates .Multiple transmissions of different users in parallel are allowed by making use of multi-code transmissions. Similarly, variable bit rates are achieved by varying the spreading factor.

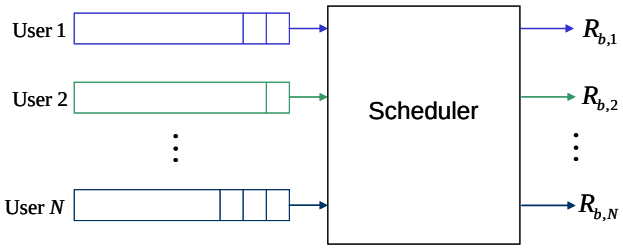

Figure 4-27 Packet scheduler

The proposed strategy starts from the Power available restrictions in downlink WCDMA systems that can be related with bandwidth allocations as follows:

4.4.2.1 Power constraint

In the downlink WCDMA, the total available power at the base station is shared by all the transmissions. Then, as shown in equation (2.28), the power devoted to user *i*-th assuming that the base station is transmitting at maximum power $P_{T\max}$ i.e. assuming that the base station operates at the capacity limit, is bounded by:

$$P_{T_i} = L_{p,i}\frac{P_N + \rho\times\dfrac{P_{T\max}}{L_{p,i}}}{\dfrac{W/R_{b,i}}{\left(\dfrac{E_b}{N_o}\right)_i}+\rho} \approx \frac{\rho\ \ P_{T\max}}{\dfrac{W/R_{b,i}}{\left(\dfrac{E_b}{N_o}\right)_i}+\rho} \approx \frac{\rho\ \ P_{T\max}}{\dfrac{W/R_{b,i}}{\left(\dfrac{E_b}{N_o}\right)_i}}, \tag{4.20}$$

where $L_{p,i}$ is the total path loss including shadowing between the i-th user and the base station, P_N is the background noise, W is the transmission bandwidth after spreading, $R_{b,i}$ is the bit rate of the i-th user, $\left(E_b/N_0\right)_i$ is the target requirement of the i-th user to assure a certain bit error rate and ρ is the orthogonality factor. A downlink power control is considered like in operative WCDMA systems as UMTS, to achieve the target $\left(E_b/N_0\right)$ requirements. The approximations given in equation (4.20) assume, for the sake of simplicity, a single isolated cell, that background noise is negligible with respect to intracell interference and the orthogonality factor negligible in front of the processing gain divided by the E_b/N_0 target. However, the presented formulation could be easily extended to consider the multi-cell case.

On the other hand, assuming that a fraction $(1-\beta)$ of the total transmission power is devoted to downlink common control channels (e.g. pilot channel), the total power devoted to the users is bounded by:

$$\sum_{i=1}^{N} P_{T_i} \approx \sum_{i=1}^{N} \frac{\rho P_{T\max}}{\frac{W / R_{b,i}}{\left(\frac{E_b}{N_o}\right)_i}} \leq \beta P_{T\max} . \tag{4.21}$$

4.4.2.2 Bandwidth Sharing

By assuming the same (E_b / N_0) requirement for all the users, the power constraint can be translated into a bit rate constraint given by:

$$\sum_{i=1}^{N} R_{b,i} \leq \frac{\beta W}{\rho\left(\frac{E_b}{N_o}\right)} = B_T . \tag{4.22}$$

Then, the efficient exploitation of the total available bandwidth B_T in terms of the allocated bit rate $R_{b,i}$ to the i-th user can be obtained by minimizing:

$$\left(B_T - \sum_{i=1}^{N} R_{b,i} \right)^2 . \tag{4.23}$$

4.4.2.3 Queue model

With respect to the queue model, let us assume that at the beginning of the k-th frame the i-th user has L_i packets waiting for transmission, as depicted in Figure 4-28, $l_{i,m}^{k}$ denotes the number of bits of the m-th packet of the i-th user in the k-th frame.

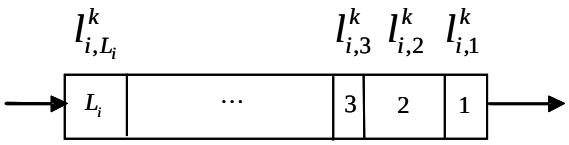

Figure 4-28 Queue of the i-th user

Assuming a FIFO policy for the packets in the queue of each user, the amount of bits that should be transmitted until the complete transmission of the m-th packet of the i-th user is given by:

$$B_{i,m,left}^k = \sum_{n=1}^{m} l_{i,n}^k \,. \tag{4.24}$$

On the other hand, the delay constraint is given by $D_{\max,i}$.measured as the maximum packet delay specified in the contract of each user, and measured in frames. Let $f_{i,m}^k$ be the elapsed time at the beginning of the k-th frame since the arrival of the *m*-th packet in the queue of the i-th user. Then, the maximum time-out left for transmission of this packet is given by:

$$TO_{i,m}^k = D_{\max,i} - f_{i,m}^k \,. \tag{4.25}$$

Consequently, the minimum bit rate required to guarantee the transmission in due time of the m-th packet of the i-th user is given by:

$$v_{i,m}^k = \frac{\sum_{n=1}^{m} l_{i,n}^k}{TO_{i,m}^k} = \frac{B_{i,m,left}^k}{TO_{i,m}^k} \,. \tag{4.26}$$

Definition *-We define the* ***Optimum Bit Rate (OBR)*** *for the i-th user in the k-th frame, as the bit rate $R_{b,i,opt}^k$ that allows transmitting all its packets in due time, i.e:*

$$R_{b,i,opt}^k = \max_{m=1,\cdots,L} \left\{ v_{i,m}^k \right\} \,. \tag{4.27}$$

Notice that a continuous transmission at the OBR would avoid packet losses for this user. However it cannot always be guaranteed for all the users because of the total bandwidth restrictions. It is assumed that packets not served in due time are dropped.

4.4.3 Optimization process for the HNN-based scheduling model

The proposed scheduling algorithm tries to minimise the constraint specified by (4.23) to ensure that the total bandwidth is used efficiently while at the same time it takes into account the time-out requirements of the different packets to ensure an optimal allocation. The bit rate allocation will be executed by means of an optimization process that can be formulated in the 2D CHNN introduced in Chapter 3, section 3.5. The same notation as in above sections will be used That is, let be N the number of users and let R_{ij}, $j=1,\ldots,M$ be the number of bit rates allowed per user i-th, $i=1,\ldots,N$ and $R_{b,i,opt} \in \left\{R_{ij}\right\}_{j=1,\cdots,M}$ the bit rate selected to be assigned to the i-th user. Accordingly, the associated 2D HNN sheme requires $S = N\text{x}M$ nodes. As in Section 4.2 we named V_{ij} the output value nodes and identify them with the allocation of bit rate to the users using equations (4.8) A properly defined Energy Function follows the same structure that the ones already formulated by means of equations (4.10) and (4.11). Then, it includes the designed dynamical set of cost values associated to each bit rate usage and a proper formulation for the downlink network restrictions, and an optimal partitioning of the total available bandwidth B_T, which depends on the specific characteristics of the considered system. It, but now the dynamic set of cost values are defined by (4.28).

Definition. *We define a set of cost values* $\left\{C_{i,j}^k\right\}_{\substack{i=1,\ldots,N \\ j=1,\ldots,M}}$, *where* $C_{i,j}^k$ *is associated to the* i*-th user in the k-th frame and for the* j*-th possible bit rate to be allocated* R_{ij}, *as follows:*

$$C_{i,j}^k = \begin{cases} 1 & \text{if} \quad R_{ij} < R_{b,i,opt}^k(1-a) \\ \dfrac{R_{i,j}\ TO_i^k - B_{i,left}^k(1-a)}{B_{i,left}^k(1+a) - B_{i,left}^k(1-a)} & \text{if} \quad R_{b,i,opt}^k(1-a) \le R_{ij} \le R_{b,i,opt}^k(1+a), \\ 0 & \text{if} \quad R_{b,i,opt}^k(1+a) < R_{ij} \end{cases} \qquad (4.28)$$

where TO_i^k *and* $B_{i,left}^k$ *are the time-out and the number of bits to be transmitted, respectively, corresponding to the most restrictive packet in the queue of the i-th user, i.e. the one having the maximum value of* $v_{i,m}^k$ *as defined in* (4.26). *In turn,* a *is a parameter to be set.*

Notice that, for bit rates R_{ij} that guarantee the transmission of the remaining bits in the contracted time (i.e. above the OBR $R_{b,i,opt}^{k}$), the cost value is low (i.e. 0) meaning that these bit rates are suitable for the *i*-th user. In turn, for lower bit rates the cost function is high (i.e. 1), meaning that these bit rates are not desirable. Finally, for bit rates in the range near to the OBR the cost value is between 0 and 1 depending on the setting of the parameter a, meaning that these bit rates could still be acceptable.

As in the above sections, once the final V_{ij} values are achieved as solution of the equations (3.66), the final bit rate assigned to the *i*-th user in one frame is given by the V_{ij} which takes the value 1 for the *j*-th bit rate (i.e. $R_{b,i,opt} = R_{ij}$).

4.4.4 Reference scheduling scheme

For comparison purposes, a reference scheduling algorithm has been considered. It operates under the same assumptions regarding delay constraints as the proposed HNN-based algorithm but, instead of making the optimisation procedure, it simply allocates to each user its OBR, $R_{b,i,opt}^{k}$, as defined in (4.27). The algorithm operates then in the following steps in the *k*-th frame:

Step 1.- To order the users in increasing value of $R_{b,i,opt}^{k}$.

Step 2.- To allocate sequentially to each user the first available bit rate higher or equal than $R_{b,i,opt}^{k}$ until having exhausted the total bandwidth B_T

Step 3.- Once the bandwidth is exhausted, allocate a bit rate $= 0$ Kb/s (i.e. no transmission) to the remaining users

4.4.5 Performance evaluation

4.4.5.1 Simulated Scenario

The WCDMA wireless network scenario consists of one isolated circular cell with radius 0.5 km. The propagation and mobility models defined in [44] are considered. In particular, the path

loss at distance d is given by $L(\text{dB}) = 128.1 + 37.6 \cdot \log_{10}(d)$, with a log-normal distributed shadowing with deviation 10 dB. Users are randomly distributed. The noise power is $P_N = -102$ dBm. The transmission bandwidth is $W = 3.84$ Mchips/s, the orthogonality factor is $\rho = 0.4$ and $\beta = 0.95$. The maximum power available at the base station is 43 dBm. Furthermore, the E_b / N_0 target is 5 dB which leads to an available bandwidth $B_T = 2.88$ Mb/s The available set of possible bit rates is given by: {384 Kb/s, 256 Kb/s, 128 Kb/s, 64 Kb/s, 32 Kb/s and 16 Kb/s}. The same parameters' values for the HNN model and the energy function (sub-section 4.2.4.1) have been considered. The frame period *T* is set to 10 ms and the parameter *a* in the cost function (4.28) takes the value 0.1. An interactive service has been considered for simulation purposes and a www traffic model has been assumed [44] as representative. Specifically, www sessions are composed by an average of 5 pages, with an average time between pages of 30s. In each page, the average number of packets is 25 with an average time between packets of 0.0277s. The packet length follows a Pareto with cut-off distribution with parameters $\alpha_P = 1.1$, $k = 81.5$ and $m = 6000$ bytes [44]. The average time between www sessions is 0.1s (i.e. it is assumed that a user is continuously generating sessions).

Two user classes, namely Class 1 and Class 2, have been considered, as representative of two different user profiles, with maximum allowed delays of 120 ms and 60ms, respectively.

4.4.5.2 Simulation results

This sub-section presents some illustrative results obtained with the proposed allocation strategy.

Figure 4-29 plots the Cumulative Distribution Function (CDF) of the packet delay for the two classes of users and the two considered scheduling algorithms when there is a total of $N = 300$, 600, 900 and 1200 users and the traffic mix is 60% of class 1 users and 40% of class 2 users. It can be noticed that the HNN approach significantly improves the reference scheduling scheme (RSS), exhibiting delay values much lower than the maximum delay bound for each traffic class. In both schemes, the delay of class 1 users is higher than the delay of class 2 users.

Figure 4-30 and Figure 4-31 show the average packet delay and the packet dropping ratio (i.e. the number of packets that are discarded because they have not been transmitted within the

delay bound with respect to the total number of packets), respectively, as a function of the number of users. It is clearly observed that the HNN achieves a lower average delay and a lower dropping ratio for all the considered traffic loads. Specifically, notice that, for low traffic loads, the dropping ratio is around 1% with HNN while it increases up to 10% with the reference scheduling scheme.

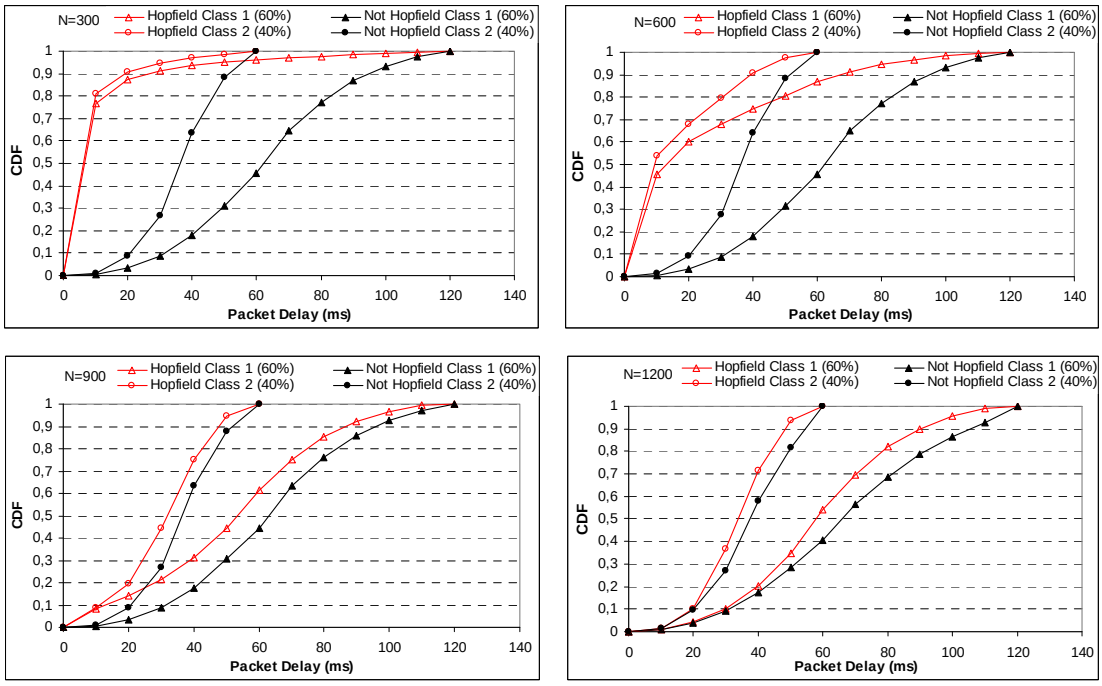

Figure 4-29 Cumulative Distribution Function of the packet delay for a situation with 300, 600 900 and 1200 users (60% class 1 and 40% class 2)

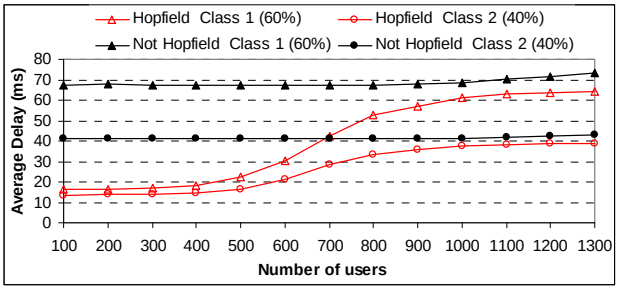

Figure 4-30 Average packet delay for a traffic mix with 60% of class 1 users and 40% of class 2 users

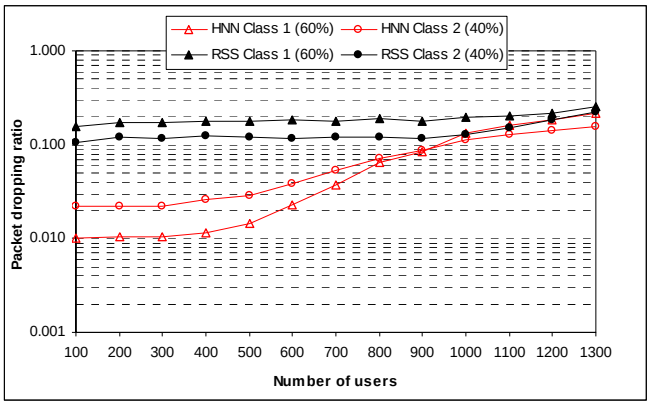

Figure 4-31. Packet dropping ratio for a traffic mix with 60% of class 1 users and 40% of class 2 users

It is worth mentioning that, when the number of users is low, the HNN scheme tends to benefit from the allocation of the highest bit rates in order to achieve lower delays, as it is depicted in Figure 4-32 which exhibits the histogram of the utilisation of the different bit rates. Then, as the number of users increases, the allocated bit rate decreases turning into an increase in the delay and dropping rate. Consequently, the HNN scheme is able to adapt the resource allocation to the specific traffic conditions. In contrast, the reference scheme tends to allocate for a certain user the OBR given by (4.27), which depends on the packet length and on the number of packets in the buffer of this user but not on the number of users in the system. As a result of that, on the one hand the reference scheme does not benefit from the highest bit rates (i.e. those higher than OBR) for low numbers of users (see in Figure 4-33 that the histogram of bit rate utilisation is less sensitive to the number of users than in the HNN case) and, on the other hand, the delay and dropping increase more slowly with the number of users than in the HNN case.

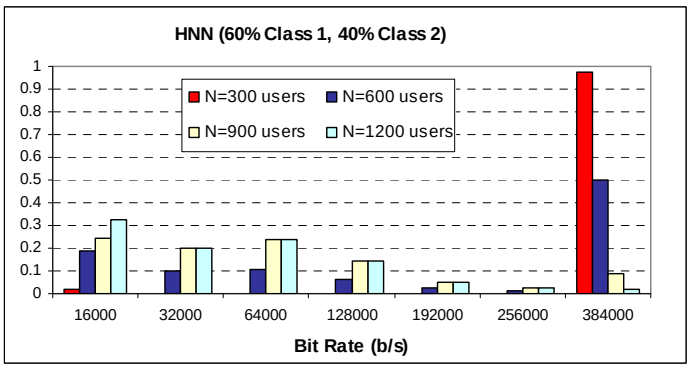

Figure 4-32 Histogram of the utilisation of the different bit rates with HNN for a traffic mix with 60% of class 1 users and 40% of class 2 users

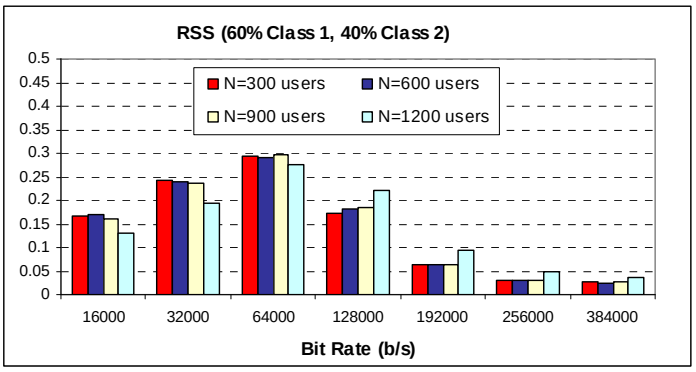

Figure 4-33 Histogram of the utilisation of the different bit rates with the reference scheme for a traffic mix with 60% of class 1 users and 40% of class 2 users

Figure 4-34 shows the dropping ratio for a traffic mix with 80% of class 1 users and 20% of class 2 users. Similar comments to the previous traffic mix apply, reflecting the good behaviour that can be obtained with HNN for different traffic mixes.

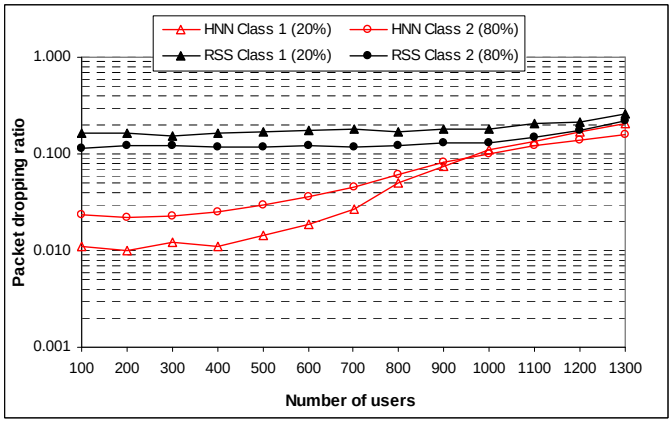

Figure 4-34 Packet dropping ratio for a traffic mix with 80% of class 1 users and 20% of class 2 users

4.4.6 Conclusions

Section 4.4 has presented a scheduling algorithm for delay-oriented services based on the use of Hopfield Neural Networks to decide the optimum instantaneous bit rate that is allocated to each user minimising an energy function that takes into consideration the specific system and service constraints. The algorithm has been evaluated through simulations in the downlink of a WCDMA system and has been compared against a reference scheme that tries to allocate the optimum bit rate to deliver the packets in the specific delay bound. It has been observed that the HNN algorithm provides a better behaviour in terms of both delay and dropping performance, with a higher ability to adapt to the traffic load conditions.

4.5 Multiple service case in the downlink direction

4.5.1 Introduction

In this section the basis of the HNN is captured for a downlink WCDMA system where both real time and delay-sensitive services are to be provided to the users sharing the access. With respect to the real time services, the user requirements are given in terms of a specific bit rate contracted with the mobile operator. In that sense, the main overall goal is that the system can continue providing an optimized throughput, while at the same time delivering to each real time user a bit rate at least equal to the satisfactory bit rate specified in his contract. In turn, with respect to delay-sensitive services, it was described in the above section and the main overall goal is not necessarily to achieve the maximum throughput nor to guarantee a certain bit rate, but to deliver each packet without exceeding a specific time deadline. This can be applicable usually to interactive data services [38]

The rest of the section is organized as follows. Sub-section 4.5.2 presents the DRA problem formulation. Sub-section 4.5.3 presents the proposed HNN model. Sub-section 4.5.4 provides some results obtained through simulation and finally conclusions are summarized in Sub-section 4.5.5.

4.5.2 Problem formulation

The considered DRA problem assumes a set of N_r real time (RT) users and N_d delay sensitive users with their corresponding queues located at the base station of the access network, operating with WCDMA. It is considered that non-shaped traffic is arriving to the queues so that all the incurred packet delay is introduced at the network level. The scheduling algorithm operates in frames of $T(s)$ and allocates a certain bit rate to each user from a set of M possible bit rates. Multiple transmissions of different users in parallel are allowed by making use of multi-code transmissions. Similarly, variable bit rates are achieved by varying the spreading factor. The objective of the DRA problem is then to select for each user an optimum allocated bit rate $R_{b,i}$, taking into account proper formulation for the downlink network restrictions, as explained in the following.

4.5.2.1 Power constraint

In the downlink WCDMA, the total available power at the base station is shared by all the transmissions. Then, the power devoted to the i-th user assuming that the base station is transmitting at maximum power $P_{T\max}$, i.e. assuming that the base station operates at the capacity limit, as it was given in Sectionn 4.4.2 by:

$$P_{T_i} \approx \frac{\rho \quad P_{T,\max}}{\dfrac{W / R_{b,i}}{\left(\dfrac{E_b}{N_o}\right)_i}}. \tag{4.29}$$

4.5.2.2 Bandwidth sharing

To properly formulate the proposed DRA algorithm, first some constraints are introduced in the following. The sum of the powers devoted to the $N = N_r + N_d$ users must be below a fraction β of $P_{T\max}$ (the remaining fraction $1-\beta$ is usually devoted to common control channels). Then, the power constraint (4.29) can be translated into a bit rate constraint given by:

$$\sum_{i=1}^{N} R_{b,i} \le \frac{\beta W}{\rho\left(\dfrac{E_b}{N_o}\right)} = B_T, \tag{4.30}$$

where it has been assumed the same E_b / N_0 for all the users. Then, the efficient exploitation of the total available bandwidth B_T in terms of the allocated bit rate $R_{b,i}$ to the i-th user can be obtained by trying to fulfil (4.30) with equality.

4.5.3 Optimization process

The DRA problem formulated in the previous sub-section sets the requirement to find the optimal bit rate allocation for the *N* different users given the available bandwidth B_T, the user satisfaction degree and the Load and Power constraints posed by the WCDMA technique. This is an RRM optimization process that can be formulated in the 2D CHNN introduced in the

Chapter 3, section 3.5. It will be used the same notation as Section 4.4 , taking into account that now we have different meanings for the bit rates according to whether the corresponding users are real time or delay-sensitive. That is, let $N = N_r + N_d$ be the number of users, N_r RT users and N_d delay sensitive, and let R_{ij}, $j = 1,\ldots,M$ be the number of streaming bit rates allowed per user i-th, $i = 1,\ldots,N$ and $R_{b,i}^{j} \in \left\{R_{ij}\right\}_{j=1,\cdots,M}$ the bit rate selected to be assigned to this i-th user. Accordingly, the associated 2D CHNN scheme requires, $S = N\text{x}M$ nodes. Then, in this 2D CHNN node deployment we named V_{ij} .the output value nodes and identify them with the allocation of bit rate to the users as in Section 4.2, Equation (4.8).

The energy function that we consider is the one given by (4.10) assuming that the dynamic set of cost values $\left\{C_{i,j}^{k}\right\}$ has a double meaning. For the real time users, $\left\{C_{i,j}^{k}\right\}$ is the one defined in section 4.3 by the equation (4.19), which depends on the desired maximum non-satisfaction probability (defined as the probability of receiving a bit rate below the satisfaction bit rate specified in the contract). In turn, for the delay-sensitive users the dynamic set of cost values $\left\{C_{i,j}^{k}\right\}$ associated to the i-th user in the k-th frame and for the j-th possible bit rate to be allocated, $R_{b,i,j}^{k}$, is the one defined in section 4.4 by (4.28).

As in the above sections, once the final V_{ij} values are achieved as solution of the equations (3.66), the final bit rate assigned to the *i*-th user in one frame is given by the V_{ij} which takes the value 1 for the *j*-th bit rate (i.e. $R_{b,i}^{j}$).

4.5.4 Performance evaluation

4.5.4.1 Simulated scenario

The WCDMA wireless network scenario and the data inputs considered are the same that were introduced in Section 4.4.5 and that we summarize in the following. Scenario assumes one isolated circular cell with radius $0.5\,\text{km}$. Users are randomly distributed. The transmission bandwidth is $W = 3.84$ Mchips/s, the orthogonally factor is $\rho = 0.4$ and $\beta = 0.95$ The maximum power available at the base station is 43 dBm. Furthermore, the E_b / N_0 target is

5 dB which leads to an available bandwidth $B_T = 2.88$ Mb/s The available set of possible bit rates, in Kb/s, is given by: {384, 256, 128, 64, 32 and 16}. The selected parameters appearing in the formulation of the HNN are $\mu_1 = 1000$, $\mu_2 = 4000$, $\mu_3 = 8000$, $\mu_4 = 800$, $\mu_5 = 6000$, $\tau = 1, \alpha = 1.0$ and $\eta = 10$. The numerical iterative solution of (3.45) is obtained following the Euler technique stated in (3.50) with $h = 10^{-4}$ for each k-th node, $k = 1, \ldots, S$ and taking the Euclidian norm and $\varepsilon = 10^{-6}$ in (3.51). The frame period T_a is set to 10 ms and the parameter *a* in the cost function (4.28) takes the value 0.1. Also, as in section 4.4.5, an interactive service has been considered for simulation purposes as representative of delay-sensitive users with the www traffic model given in [44]. Similarly, for real time users, sessions are generated according to a Poisson arrival process with exponential call duration. Two delay-sensitive user classes, namely Class 1 and Class 2, have been considered, as representative of two different user profiles, with maximum allowed delays of 120 ms and 60 ms, respectively. In turn, for the real time services, two user profiles are considered: Class 1 users have a maximum bit rate of 256 Kb/s and class 2 users of 64 Kb/s. In turn, the satisfaction bit rate is 192 Kb/s for class 1 and 64 Kb/s for class 2. The maximum non-satisfaction probability for RT users.

4.5.4.2 Simulation results

This sub-section presents some illustrative results with the proposed allocation strategy. Two different case studed are considered.

- **Case study 1:** only www users exist in the scenario. The user profile mix is 60% of class 1 and 40% of class 2.

- **Case study 2:** a variable number of www users coexist with 10 RT users. For both services, the profile mix is 60% class 1 and 40% class 2.

Figure 4-35 plots the CDF of the packet delay in case study 1 for the two classes of users when the total number of www users is 600. The results are compared against a reference scheduling scheme (RSS) in which users are served at the bit rate that allows keeping the delay constraint without taking into consideration the status of the rest of users. It can be noticed that the HNN approach significantly improves RSS, exhibiting delay values much lower than the maximum delay bound for each traffic class.

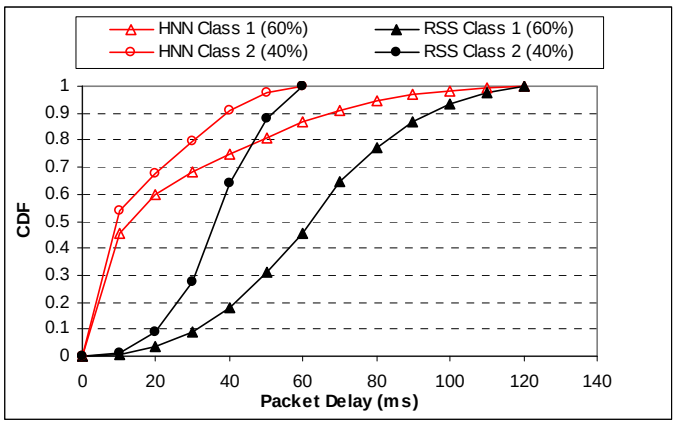

Figure 4-35. Cumulative Distribution Function of the packet delay of www users in case study 1, with 600 www users in the scenario

Figure 4-36 plots the packet dropping ratio (i.e. the ratio of discarded packets due to the expiration of their deadline) for delay-sensitive users in the two case studies. Clearly, class 1 users, having a higher delay bound, exhibit a lower dropping rate than class 2 users. In turn, the inclusion of RT users in the scenario (i.e. case study 2), introduces a certain increase in the dropping ratio of both classes because the algorithm reduces the bit rate allocated to www users. This can be also observed in Figure 4-37, which plots the histogram of the allocated bit rates to the different services in the two considered case studies. Notice that in case 1, most of the www users receive a bit rate equal to 384 Kb/s, while in case 2 the number of allocations of this bit rate are reduced. Similarly, notice in case 2 that the algorithm is able to keep the bit rate of both RT user classes above the satisfaction bit rate.

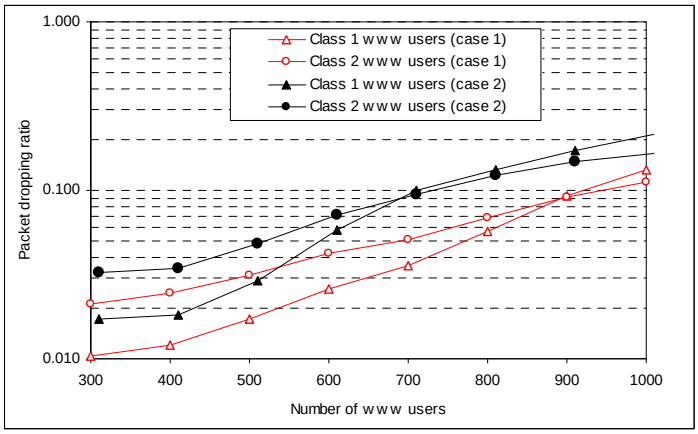

Figure 4-36. Packet dropping ratio for delay-sensitive for the two considered case studies

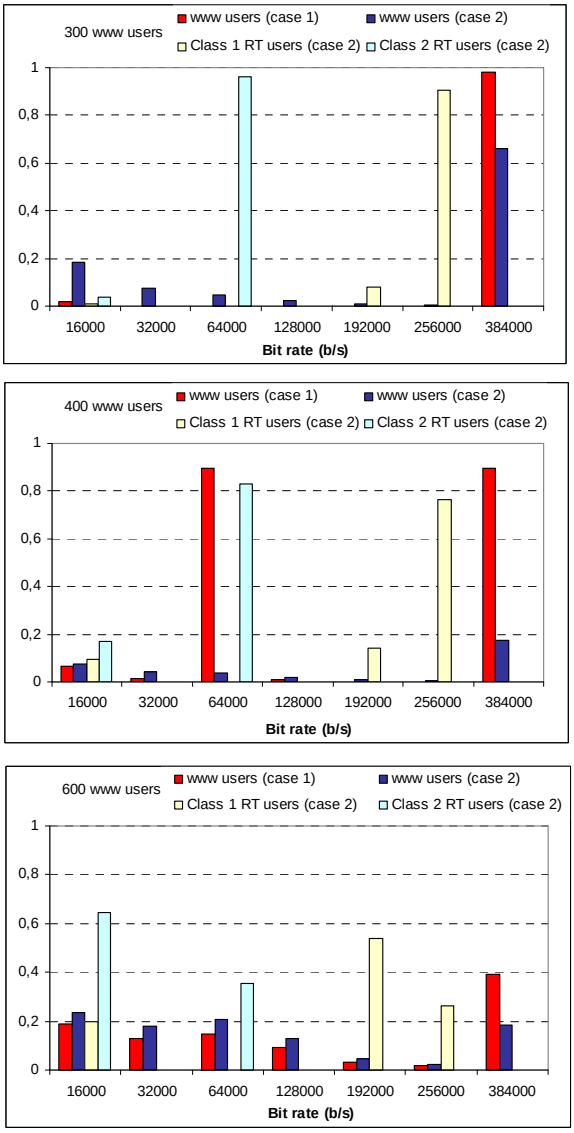

Figure 4-37. Histogram of the utilisation of the different bit rates for www and RT users in the two considered case studies with a total of 300, 400 and 600 www users in the scenario

Regarding the performance of *RT* users, Figure 4-38 plots the satisfaction probability for the two user classes when increasing the number of RT users in the scenario and there are a total of 300 www users. The profile mix is 60% users of class 1 and 40% of class 2. It can be observed that, for the considered load levels, the algorithm is able to keep the satisfaction probability above the limit of 0.8 (i.e. corresponding to the maximum non-satisfaction probability 0.2) for both user classes. Notice also that the satisfaction is higher for class 1 users than for class 2 users because for the latter the satisfaction bit rate is equal to the maximum possible bit rate, so the algorithm has fewer combinations that fulfil the satisfaction rate.

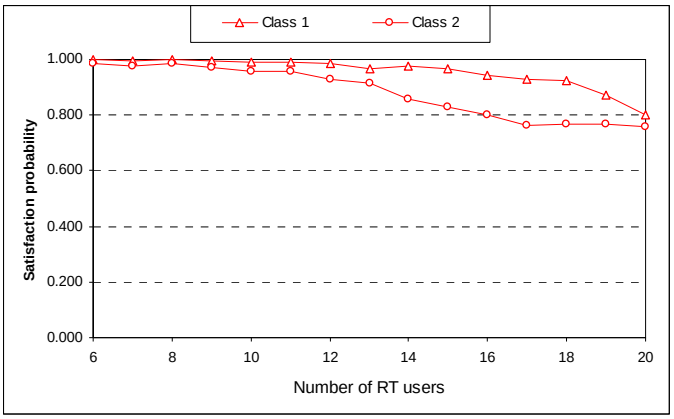

Figure 4-38. Satisfaction probability of RT users as a function of the number of RT users in a scenario with 300 www users

4.5.5 Conclusions

This section 4.5 has exploited the already described DRA strategy based on Hopfield Neural Networks in a multi-service scenario. It has been applied in the downlink of a WCDMA system with both real time and delay-oriented services. The different requirements of each service have been included in the formulation of the cost function. Results have shown that the proposed strategy is able to adapt to the specific traffic conditions and service requirement

Chapter 5

Common Radio Resource Management

5.1 Introduction

The heterogeneous network concept is intended to propose a flexible and open architecture for a large variety of wireless access technologies, applications and services with different QoS demands. A typical heterogeneous scenario is constituted by several Radio Access Technologies (RATs) each having a Radio Access Network (RAN) interfacing a common Core Network (CN). RATs include cellular networks, e.g. UTRAN (UMTS Terrestrial Radio Access Network) and GERAN (GSM/EDGE Radio Access Network), as well as non-cellular systems, e.g. WLAN.

Each radio access network differs from each other by air interface technology, cell-size, access, coverage and ownership. In spite of these basic differences, many services can be carried over any of the deployed RATs. Then, the complementary characteristics that these networks offer make possible to exploit the trunking gain resulting from the joint consideration of the different networks as a whole. Thus interworking among heterogeneous RATs leads to a better overall performance than the accumulated performances of the stand-alone systems. This challenge calls for the introduction of new RRM algorithms operating from a common perspective that take into account the overall amount of resources offered by the available RATs, and therefore are referred to as CRRM (Common Radio Resource Management) algorithms [45] and [46].

In this scenario, whenever there is more than one radio access technology that can provide the requested services, it is needed to decide the RAT that fits better to the user constraints. In that sense, one of the key aspects to achieve an efficient usage of the available radio resources while at the same time keeping the desired QoS constraints for the different services is to execute a Dynamic Resource Allocation (DRA) that takes into account both QoS requirements and channel status for each user. DRA in a heterogeneous scenario can be considered as a combinational problem where radio resources (e.g. bandwidth in each RAT) have to be

allocated to several users at each frame subject to certain restrictions in terms of QoS and of total amount of available resources. In that sense, as it has been already applied in Chapter 4, it is known that a HNN searches for the combinational solution that minimizes an Energy Function under specific constraints. Taking into account the above Chapter 4 framework, in this chapter the basis of the HNN is captured for a downlink scheduling problem in a CRRM scenario. The main overall goal is not to achieve the maximum throughput, but to deliver each packet without exceeding a specific time deadline using the best available RAT. The HNN here proposed is claimed to do that in a very low time scale of several milliseconds (the typical value for a transmission time interval in a wireless system).

The rest of the chapter is organized as follows. Section 5.2 presents a HNN-based CRRM algorithm which includes in the formulation of the energy function some constraints capturing in practice that a user must remain during a certain time in a RAT to avoid ping pong effects (i.e. changing continuously from one RAT to another). Subsection 5.2.2 presents the problem formulation and subsection 5.2.3 provides the proposed HNN-based scheduling approach, which will be compared against a reference scheme presented in subsection 5.2.4. Finally Section 5.3 analyses the performance of the proposed HNN methodology in two different scenarios.

5.2 Problem formulation

The considered DRA problem assumes a set of N users, $i = 1, \ldots, N$, with their corresponding queues, located at the base station of the access network, which contain the packets pending to be transmitted in the downlink direction of a multiple-RAT system, as illustrated in Figure 5-1. It is considered that non-shaped traffic is arriving to the queues so that all the incurred packet delay is introduced at the network level. The scheduling algorithm operates in frames of T (s) and allocates a certain bit rate to each user from a set of M possible bit rates in a given RAT. Multiple transmissions of different users can be allocated in each frame for each RAT multiplexed either on a TDMA or a WCDMA way, depending on the specific characteristics of the RAT. It will be assumed that a user can not be simultaneously allocated in more than one RAT.

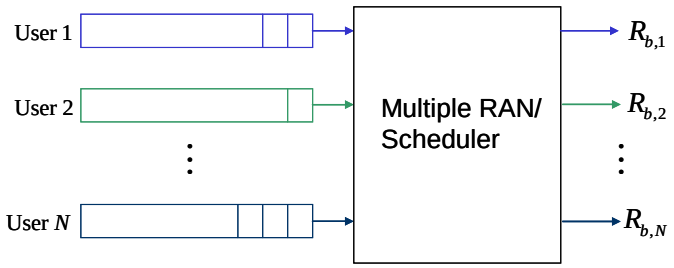

Figure 5-1. Joint Packet Scheduler

Concerning the service type, an interactive www-like service is considered in which the different packets must be transmitted before a specific delay bound. Then, from the point of view of queueing model and service constraints the same conditions and Optimum Bit Rate definition, (4.27), explained in sub-section 4.4.2.3 for delay-sensitive traffic apply.

The set of bit rates allocated at a given instant in a p-th RAT must fulfil the relationship.

$$\sum_{i=1}^{N} R_{b,i,p} \leq B_p , \tag{5.1}$$

where $R_{b,i,p}$ is the optimum allocated bit rate of the i-th user in the p-th RAT. In turn, B_p is the total available bandwidth in the *p*-th RAT. Then, the efficient exploitation of the total available bandwidth B_p in terms of the allocated bit rate to the different users can be obtained by leading (5.1) to equality.

5.1.1 Ping-Pong effect

One problem to face up in wireless networks is the ping pong effect. This is a typical effect in horizontal handover procedures coming up when, due to the pilot signal fluctuations, the user point of attachment is transferred repeatedly from the old base station to the new one and vice versa. Such non desired intermediate handovers cause a signalling and hence a resource waste and must be avoided as much as possible. When CRRM is considered, a Vertical Handover is said to occur when the user is moved from one RAT to another. Then, similar signalling expense could be demanded if the RAT change assignment is executed too often. Furthermore, it could happen that this RAT handover requires a minimum execution time due to implementation constraints. Then, because of these reasons we will define T_{VH} as the minimum

time a user has to be anchored to a given RAT once it is assigned. The obtained final scheduling performances obtained when this time is considered are expected to be degraded with respect to the ideal theoretical case $T_{VH} = 0$ (in which the RAT can be changed from frame to frame) and the purpose here will be to assess this degradation for different T_{VH} values.

5.3 Optimization process

The proposed scheduling algorithm tries to ensure that the total bandwidth is efficiently used while at the same time it takes into account the time-out requirements of the different packets to ensure an optimal allocation for the N different users. It is assumed also that there are a set of RATs $p = 1,\ldots,P$ and let $R_{ij,p}$, $j = 1,\ldots,M$ be the number bit rates allowed per user i-th, $i = 1,\ldots,N$ in the p-th RAT It is assumed also that $R^{j}_{b,i,p} \in \left\{R_{i,j,p}\right\}_{j=1,\cdots,M}$ is the bit rate to be allocated in the *k*-th frame to the *i*-th user, as well as the appropriate *p*-th RAT. As stated in the previous section, the problem consists in selecting, from the predefined set of bit rates $j = 1,\ldots,M$ at the p-th RAN the set of allocated bit rates $\left\{R^{j}_{b,i,p}\right\}_{i=1,\ldots,N}$ A user can only be allocated in one RAT. This bit rate allocation problem will be executed by means of an optimization process that can be formulated in terms of a three-dimensional Continuous Hopfield Neural Network (3D CHNN) with $L = N \times M \times P$ nodes. In such a 3D HNN the notation used extends in a natural way the notation introduced in chapter 3, section 3.5 for a 2D HNN model. Then, in this 3D node deployment each i, j, p-th node has an input value noted as $U_{i,j,p}$ and a output valued noted as $V_{i,j,p}$. We identify this output values nodes with the allocation of bit rate and RAT to the user as follows

Definition. *The j-th bit rate of the p-th RAT is allocated to the i-th user iff* $V_{i,j,p} = 1$.

By the definition, the set of output value nodes gives the following final allocation to the users

$$\begin{aligned} &V_{ij,p} = 1 \;\Rightarrow\; R^{j}_{b,i,p} = R_{i,j,p} \\ &V_{ij,p} = 0 \;\Rightarrow\; \text{the allocation of } R_{i,j,p} \text{ is not consider} \\ &\text{for the } i\text{-th user; } i = 1,\ldots,N;\ j = 1,\ldots,M;\ p = 1,\ldots,P \end{aligned} \tag{5.2}$$

Let us define a dynamic set of cost values associated to each bit rate of each RAT as follows:

Definition. *We define a set of cost values* $\left\{C_{i,j,p}^{k}\right\}_{\substack{i=1,\ldots,N \\ j=1,\ldots,M \\ p=1,\ldots,P}}$ *,where* $C_{i,j,p}^{k}$ *is associated to the* i *-th user in the k-th frame and for the* j *-th possible bit rate* and p -th RAT to *be allocated* $R_{i,j,p}$ *, as follows:*

$$C_{i,j,p}^{k}=\begin{cases} 1 & \text{if} \quad R_{i,j,p} < R_{b,i,p,opt}^{k}(1-a) \\ \dfrac{R_{i,j,p}\ TO_{i}^{k} - B_{i,pleft}^{k}(1-a)}{B_{i,pleft}^{k}(1+a)-B_{i,pleft}^{k}(1-a)} & \text{if} \quad R_{b,i,opt}^{k}(1-a) \le R_{i,j,p} \le R_{b,i,opt}^{k}(1+a), \\ 0 & \text{if} \quad R_{b,i,p,opt}^{k}(1+a) < R_{i,j,p} \end{cases} \quad (5.3)$$

where TO_{i}^{k} *and* $B_{i,pleft}^{k}$ *are the time-out and the number of bits to be transmitted, respectively, corresponding to the most restrictive packet in the queue of the i-th user, i.e. the one having the maximum value of* $v_{i,m}^{k}$ *as defined in (4.26). In turn,* a *is a parameter to be set.*

According to [36] a valid expression for the energy function E follows as:

$$E=\frac{\mu_1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{p=1}^{P}C_{i,j,p}^{k}V_{i,j,p}+\sum_{p=1}^{P}\frac{\eta^{\varsigma_p}\mu_2}{2}\left|1-\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{R_{i,j,p}}{B_p}V_{i,j,p}\right|+\frac{\mu_3}{2}\sum_{p=1}^{P}\sum_{i=1}^{N}\sum_{j=1}^{M}\psi_{i,j,p}V_{i,j,p}+$$
$$+\frac{\mu_4}{2}\sum_{p=1}^{P}\sum_{i=1}^{N}\sum_{j=1}^{M}V_{i,j,p}\left(1-V_{i,j,p}\right)+\frac{\mu_5}{2}\sum_{i=1}^{N}\left(1-\sum_{p=1}^{P}\sum_{j=1}^{M}V_{i,j,p}\right)^{2}. \quad (5.4)$$

The minimisation of this function provides the allocation of the i -th user given by $V_{i,j,p}$, which takes the value 1 for the j -th bit rate and p -th RAT (i.e. $R_{b,i,p}^{j}$) and 0 otherwise. The first summand in (5.4) contains the cost function defined in (5.3) In turn, the second summand is similar to that stated in Section 4. It penalizes the undesired situations in which the total allocated bandwidth in a given RAT is higher than B_p, and, at the same time, it drives expression (5.1) to the equality, so that resources are efficiently used. In this term, the following exponent is defined:

$$\varsigma_p = u\left(\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{R_{i,j,p}}{B_p}V_{i,j,p}-1\right) \quad (5.5)$$

where $u(\ )$ is the unit step function.

The third summand in (5.4) simply penalizes non allowed bit rates because of e.g. the user profile. That is, for the bit rates $R_{i,j,p}$ of the *i*-th user included in its profile, $\Psi_{i,j,p}=0$, while for the bit rates not included in the profile $\Psi_{i,j,p}=1$. Furthermore, the $\Psi_{i,j,q}$ value is used to manage the T_{VH} constraint. Particularly, provided that the i-th user is currently allocated in the p-th RAT, we will set $\Psi_{i,j,p}=1$, for all bit rate, $j=1,\ldots,M$, in any other *r*-th RAT $r \neq p$ during T_{VH} seconds after the initial p-th RAT assignment. In this way, the algorithm will keep the i-th user allocated in the same RAT during T_{VH}.

Again as it was also stated in Chapter 4, the fourth term forces convergence towards $V_{i,j,p} \in \{0,1\}$ and the last term forces the physical condition that only one bit rate $R_{i,j,p}$ and RAT are possible for the i-th connection.

Following the same deployment explained in Chapter 3, the optimization process of the 3D-HNN is carried out on a frame basis and relies again on minimizing the Energy function given by (5.4) through the convergence of the following equations:

$$\frac{dU_{klq}(t)}{dt} = -\frac{1}{\tau}U_{klq}(t) + \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{p=1}^{P} T_{klq,ijp}\, f_{ijp}\left(U_{ijp}(t)\right) + I_{klq};$$

$$V_{klq} = f_{klq}\left(U_{klq}\right) = \frac{1}{1+\exp\left(-\alpha_{klq}U_{klq}\right)}; \qquad (5.6)$$

$$k=1,2,\ldots,N; l=1,2,\ldots,M; p=1,\ldots,P;$$

where $T=\left(T_{ijp,klq}\right)_{\substack{i,k=1,\ldots,N\\ j,l=1,\ldots,M\\ p,q=1,\ldots,P}}$ is a symmetric matrix, $I=\left[\, I_{klq} \,\right]_{\substack{k=1,\ldots,N\\ l=1,\ldots,M\\ q=1,\ldots,P}}$ denote a vector of additional conditions, $\tau=1$ and high α_{klq}'s values will be considered.

The energy function given by (5.4) and (5.5) can be expanded and rearranged into the form given by Equation (5.7). Then stable states should tend to be found on the vertices of a S-dimensional cube.

$$E = -\frac{1}{2}\sum_{k=1}^{N}\sum_{l=1}^{M}\sum_{q=1}^{P}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{q=1}^{P} T_{ijp,klq} V_{ijp} V_{klq} - \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{q=1}^{P} I_{ijp} V_{ijq} \,. \qquad (5.7)$$

Using (5.7), equations (5.6) can be written as

$$\frac{dU_{klq}(t)}{dt} = -\frac{1}{\tau}U_{klq}(t) - \frac{\partial E}{\partial V_{klq}};\ k = 1,\ldots,N; l = 1,\ldots,M; q = 1,\ldots,P\,, \qquad (5.8)$$

which characterizes the 3D-HNN behaviour. The minima of the Energy occur now at the 2^L corners inside the L-dimensional hypercube defined on $V_{i,j,p} \in \{0,1\}$ By solving numerically (5.8) and after reaching a stable state each node is either ON (i.e. 1 if $V_{i,j,p}$ is greater or equal than 0.5) or OFF (i.e. 0, if $V_{i,j,p}$ is lower than 0.5). The numerical iterative solution of (5.8) is obtained following the Euler technique introduced in Chapter 3, subsection 3.4.1, similarly to what has been done in the simulations so far presented.

Annex B contains the expressions obtained for both the interconnection matrix $T = \left[T_{kl,ij}\right]_{\substack{i,k=1,\ldots,N \\ j,l=1,\ldots,M}}$ and the bias current vector $I = \left[I_{kl}\right]_{\substack{k=1,\ldots,N \\ l=1,\ldots,M}}$.

5.4 Reference Scheduling Scheme (RSS)

For comparison purposes, the same reference scheduling algorithm used in Sub-section 4.4.4 has been considered. It operates under the same assumptions regarding to the previous constraint delay and T_{VH} setting but, instead of making the optimisation procedure, it simply allocates to each user its Optimum Bit Rate, OBR, as defined in (4.27) and selects the lowest loaded RAT, so that load balancing considerations are captured. The algorithm operates then in the following steps in the k-th frame:

Step 1.- To order the users in increasing value of $R^k_{b,i,opt}$.

Step 2.- To allocate the RAT having the lower number of connected users at the allocation time and keep it during T_{VH} for a given user.

Step 3.- To allocate sequentially to each user the first available bit rate higher or equal than $R^k_{b,i,opt}$ (i.e. OBR) in the selected RAT or until having exhausted the total bandwidth in this selected RAT.

Step 4.- Once the bandwidth is exhausted in all RAT's, it allocates a bit rate =0 Kb/s (i.e. no transmission) to the remaining users.

5.5 Performance evaluation

5.5.1 Simulated scenario

Two RATs have been selected for illustration purposes in the downlink direction, namely a WCDMA-based and a TDMA-based RAT, representative of e.g. UMTS and GERAN, respectively. The wireless network scenario consists of one isolated circular cell with radius 0.5 km where the two RATs are co-sited. The propagation and mobility models defined in [44] are considered. In particular, the path loss at distance d is given by $L(dB) = 128.1 + 37.6 \cdot \log_{10}(d)$, with a log-normal distributed shadowing with deviation $10\,dB$. Users are randomly distributed. The noise power is $P_N = 102 dBm$

As for the Power constraint and Bandwidth sharing we have considered in the WCDMA scenarios, let us notice now that

a. The power constraint is linked to the type of access. In case of using a WCDMA the power constraint follows that already stated in Section 4.2.4. In case of using TDMA there is just a limited power constraint and it is not shared among the different users. Then, the access is not interference limited due to the own users that share it as it is in WCDMA.

b. The way how the bandwidth is shared among the different users is access dependent. In case of using WCDMA it applies what it was described in sub-section 4.5.2.2. In case of TDMA, each allocated slot contributes with bandwidth (depending on the maximum modulation and coding scheme that can be used), being the sum of the bandwidths for all available slots equal to the total bandwidth available.

The bit rate and power constraints defined for the downlink direction of WCDMA are the same as in sub-section 4.3.5 with the transmission bandwidth $W = 3.84$ Mchips/s, the orthogonality

factor is $\rho = 0.4$ and $(1-\beta)$ is the fraction of total power devoted to common control channels, with $\beta = 0.95$. The maximum power available at the base station is $43\ dBm$. Furthermore, the E_b / N_0 target is $5dB$ which leads to an available bandwidth $B_{CDMA} = 2.88$ Mb/s. The available set of possible bit rates, in Kb/s, is given by $\{384,\ 256,\ 128,\ 64,\ 32 \text{ and } 16\}$.

With respect to the TDMA-based RAT, the GERAN system [47] is taken for illustrative purposes with a total bandwidth of $B_{TDMA} = 1920$ Kb/s corresponding to 4 carriers with modulation coding scheme MCS-9. Assuming a multislot capability of 8 slots (i.e. 8 slots per frame can be allocated to the same user terminal) and MCS-9, the available bit rates for a given mobile are approximated by $60j$ Kb/s where $j = 1,\ldots,8$.

Two user classes, namely Class 1 and Class 2, have been considered, as representative of two different user profiles, with maximum allowed delays of 120 ms and 60ms, respectively. The proposed framework allows them to be jointly managed.

The selected parameters appearing in the formulation of the HNN are $\mu_1 = 1000$, $\mu_2 = 4000$, $\mu_3 = 8000$, $\mu_4 = 800$, $\mu_5 = 6000$, $\tau = 1$, $\alpha = 1.0$ and $\eta = 10$. The frame period is set to 10 ms and the parameter *a* in the cost function (5.3) takes the value 0.1. An interactive service has been considered for simulation purposes and a www traffic model, already introduced in Section 4.4.5 has been assumed as representative.

On the other hand, the iterative numerical solution of (5.8) is obtained following the Euler technique introduced in Chapter 3, sub-section 3.4.1 and is finalized when the iterations n and $n-1$ satisfy $\left\| W^n - W^{n-1} \right\|_2 < \varepsilon$, where $W^n = \left(w_1^n, w_2^n, \ldots, w_L^n \right)$ and $\| \ \|_2$ is the Euclidean Norm. We have set $h = 10^{-4}$ in (3.50) and $\varepsilon = 10^{-5}$. If all these conditions are fulfilled we decide the process converges and the matrix $V_{i,j,p}$ provides us the $R_{b,i,j,p}$ values.

It is worth mentioning that, due to the constraints in the considered DRA problem, the convergence, as usual in the HNN studied so far, follows the pattern shwown in Sub-section 4.2.4.4. Keeping this in mind, and from a practical point of view, a maximum number of iterations $N_{\max} = 1000$ is considered in the iterative solution of (3.50). If this maximum is

reached it is assumed that the procedure has not converged and the $R_{b,i,j,p}$ values are decided to be the same obtained in the last frame. In any case, it should be noted that, under this strategy, only a 0.005% of the frames simulated in the results presented in the Section 5.5 failed to converge.

5.5.2 Simulation results

In this sub-section some illustrative results obtained with the proposed allocation strategy are obtained. Figure 5- 2 shows the average packet loss ratio (i.e. the ratio of packets dropped because of exceeding the deadline) versus the number of users for different T_{VH} settings for class 1 users. It can be observed that the performance degrades when increasing the value of T_{VH} . Figure 5-3 shows the comparison between the proposed HNN strategy and the reference scheme (RSS) in terms of average packet dropping ratio, revealing the improvement attained when a HNN is used. The same trend is observed in Figure 5-4, which compares the CDF of the delay for the two classes of users. Figure 5-5 plots the histogram of bit rate assignment in the WCDMA RATs, with HNN and RSS. Notice that, for a high number of users, RSS assigns the higher bandwidths at the extent of leaving without service (i.e. bit rate $= 0$) most of the users. On the contrary, HNN tries to serve all the users so as all of them receive some no null bit rate, thus offering better performances. Similar results would apply to TDMA.

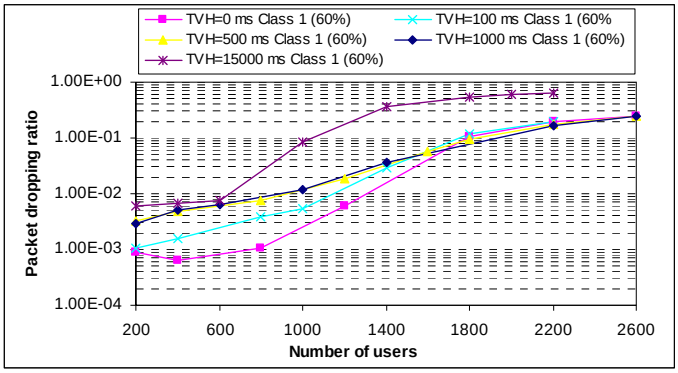

Figure 5- 2 Average packet dropping for different values of T_{VH}

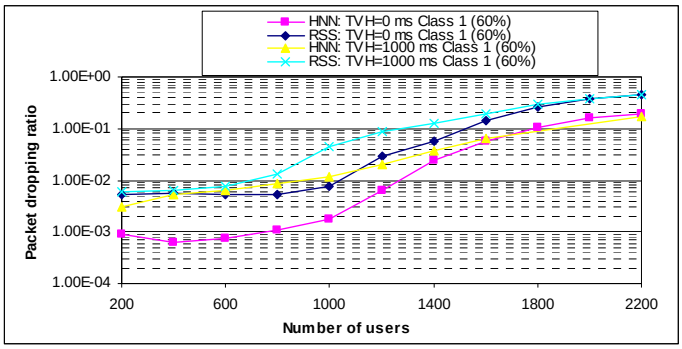

Figure 5-3. Comparison with the Reference Scheme RSS

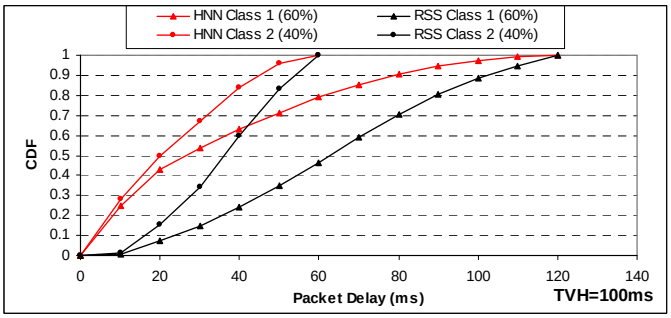

Figure 5-4. Cumulative Distribution Function of the packet delay for a situation with 1200 users (60% class 1 and 40% class 2)

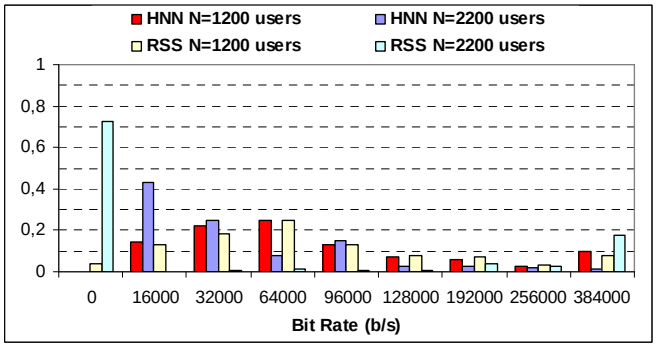

Figure 5-5. Histogram of the utilisation of the different bit rates in WCDMA with HNN and RSS for a traffic mix with 60% of class 1 users and 40% of class 2 users

In the following, it is shown the performance of the proposed HNN methodology in the same scenario S1 with different defined service classes as: class 1 users with a maximum packet delay of 2 s and class 2 users with a maximum packet delay of 1s.

The obtained results follow the same trends that were obtained in the scenario of sub-section 4.3.3. In particular, Figure 5-6 plots the average packet delay for the HNN and the RSS schemes, revealing a lower delay for the HNN case.

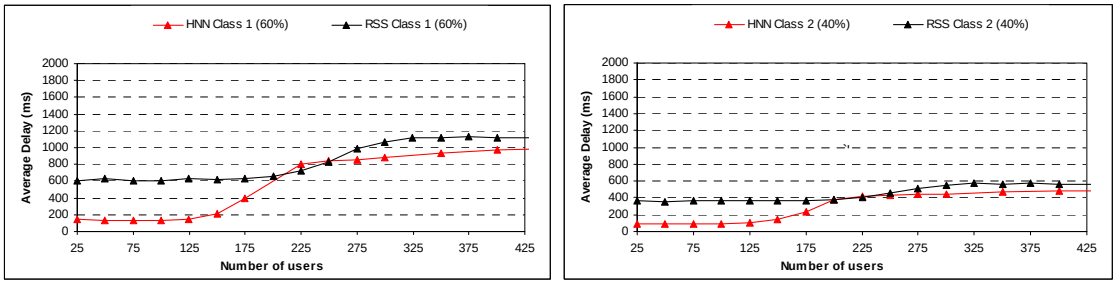

Figure 5-6. Average packet delay for the class 1 and class 2 users .

On the other hand Figure 5-7 plots the bit rate allocated to the different transmissions in WCDMA and in GERAN for the HNN and RSS algorithms and for two different load levels. It can be observed how the HNN is able to make the allocation in the system depending on the overall number of users, so that the higher the number of users the lower the bit rate that is being allocated. As a result of that, with HNN there are very few allocations of 0 Kb/s, so that most of the users transmit to some extent. On the contrary, the results with RSS reveal that the scheme is less sensitive to the increase in the number of users, and in some cases users transmitting with a high bit rate coexist with other users with 0 Kb/s.

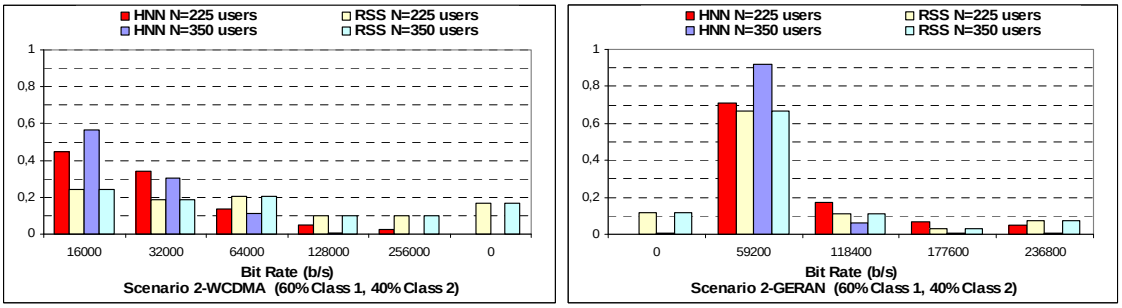

Figure 5-7. Histogram of the bit rates allocated in WCDMA and GERAN.

5.6 Conclusions

In this chapter a novel HNN based scheduling procedure for a constrained delay service has been proposed for the common radio resource management of multiple RATs. It has been shown for a scenario with WCDMA and TDMA RATs that it compares favourably with a reference scheduling procedure. The algorithm allows also taking into consideration timing constraints to avoid continuous changes of RAT for a given user, and it has been shown the impact of this constraint in the obtained performance.

Chapter 6

HNN in OFDMA Systems

6.1 Introduction

Orthogonal Frequency Division Multiple Access (OFDMA) has emerged as one of the most promising schemes for broadband wireless networks. By using multiple parallel low rate subcarriers, OFDMA can offer satisfactory high speed data rate, robust wireless transmission and flexible radio resource management, among other remarkable features, as it is widely documented in the open literature. In fact, current standards like DVB-T, Wireless LAN IEEE.802.11a and fixed-mobile broadband access system IEEE 802.16 i-e, have adopted OFDMA scheme. In addition to that, OFDMA has also been selected as access technology for the future 3G LTE in the Evolved Universal Telecommunication Radio Access (EUTRA) [48], and most of the 4G initiatives are also considering OFDMA as a prime access strategy. As a result of this current trend and from the radio resource allocation point of view, there has recently paid a lot of attention to manage dynamically the inherent flexibility offered by OFDMA in a optimal and still practical way, either in isolated [49] or in multicell OFDMA systems [50]

Concerning packet data transmission most of the subcarrier allocation strategies proposed in OFDMA-based wireless multimedia networks intent somehow to maximize the system throughput or minimize the overall transmit power while achieving the terminal's bit rate requirements [51]. A recent good survey on these topics can be found in [52]. Unfortunately the traffic related queuing impact when considering DRA scheduling schemes is not covered at the same extent. That is particularly relevant for interactive services and in general terms for delay-bounded services, in which packets should be delivered within specified deadlines. In that respect, there has been little work on these relevant performance measures such as the delay bound and the delay violation probability, which are indicative of the worst-case delay behaviour. To the best of our knowledge [53] and [54] are among the first papers to face the constrained delay issue in managing the OFDMA system resources using for this purpose a heuristic approach based on utility and priority functions when assigning resources to users.

OFDMA scheduling should actually include both joint subcarrier and power allocation. This is a rather complex problem, and usually it is simplified by separating these two allocations. Subcarrier allocation provides more gain than power allocation in [55] and in fact it is shown in [56] and [57] that waterfilling allocation only brings marginal performance improvement over fixed power allocation with ACM. Then, in this chapter, we focus on a subcarrier allocation strategy that is aware of the queuing state per each user and that retains a fixed power allocation as well as adaptive QAM (Quadrature Amplitude Modulation).

Subcarrier allocation in OFDMA can be seen as a combinational problem where there are plenty of possible combinations associated to a given user. A user can be granted with many subcarriers at a given point of the time. In turn, each subcarrier provides a given channel capacity depending on the current fading and interference, so that multiuser diversity can be exploited. Also, a given subcarrier can only be assigned to one user. This is a natural choice based on [56] that proves that the optimum OFDMA performance is reached by assigning each subcarrier only to one user in a cell among the many users trying to get access. In queuing aware OFDMA systems like the one considered here, the information about queuing and channel status is exploited to efficiently allocate resources through proper cross-layer designs of the data scheduler. As a matter of fact, like in general OFDMA DRA proposals, heuristic algorithms are usually selected to circumvent the fact that NP-hard algorithms would be necessary to obtain the optimum solutions. This is the case of [58], where a heuristic two-step algorithm is proposed to first allocate a number of subcarriers to each user and then to assign the specific subcarrier to each terminal. Similarly, in [59] an alternative approximate asymptotic mechanism is exploited. It relies on the fact that in a heavy traffic scenario minimizing delay violation is approximately equivalent to minimizing mean waiting time. Similarly, in [60] an allocation strategy depending on the queue size of each terminal relative to the overall data queued at the access point is presented for video streams. It is shown that this allocation achieves significant improvements with respect to static allocation methods, in spite of not including in the allocation neither channel gain information nor any stream specific knowledge. In [61] a cross-layer DRA strategy is presented that combines the channel status information together with the queue status and quality requirements in order to maximize power efficiency and ensure user fairness, using a virtual clock scheduling algorithm and an adaptive subcarrier and power allocation.

In this chapter, and due to the fact that the typical DRA in OFDMA turns out to be actually a combinational problem among all the subcarriers involved, we have devised the collective computation property featured by the Hopfield Neural Networks, which provide an optimal

solution for many combinational problems as a very suitable approach for the problem addressed here. In fact, the HNN approach provides feasible solutions to complex optimization problems, like the Nondeterministic Polinomial time (NP)-hard algorithms mentioned above as it has been introduced in Chapter 3.

6.2 Scheduler for delay sensitive traffic

6.2.1 Introduction

This section proposes a novel HNN-based joint channel and queue aware scheduling strategy for downlink OFDMA systems suitable for delay bounded services. A multiuser scenario with statistically independent fading channels and an isolated cell is considered. Then, the subcarrier allocation is directly related to the remaining time before the agreed bounded delay service per user is violated for each packet as well as to the channel state. The proposed algorithm is compared against other approaches existing in the literature [58] and against a heuristic algorithm, also proposed in this paper as a first simple step in the provision of a joint channel and queue aware strategy, which simply prioritizes the users according to their remaining packet life time and assigns subcarriers until they are exhausted.

The rest of the section is organised as follows. Sub-section 6.2.2 describes the considered system model, including the queueing behavior and the OFDMA considerations. Sub-section 6.2.3 describes the proposed HNN-based algorithm, with two different possibilities depending on the definition of the energy function. The proposed algorithm will be compared against the reference schemes presented in sub-section 6.2.4. Results are given in Sub-section 6.2.5 and finally conclusions are summarised in sub-section 6.2.6.

6.2.2 System model

The considered DRA problem assumes a set of N users, $i = 1, \ldots, N$, with their corresponding queues, located at the base station of the access network, which contains the packets pending to be transmitted in the downlink direction of an OFDMA system, as illustrated in Figure 6-1. It is considered that non-shaped traffic is arriving to the queues so that all the incurred packet delay is introduced at the network level. Also the model allows for differentiating among multiple classes of traffic (services classes) as it will be later discussed. When a packet can not be

delivered within this bounded delay it is dropped and therefore a dropping probability appears as a key performance indicator of the scheduling behaviour.

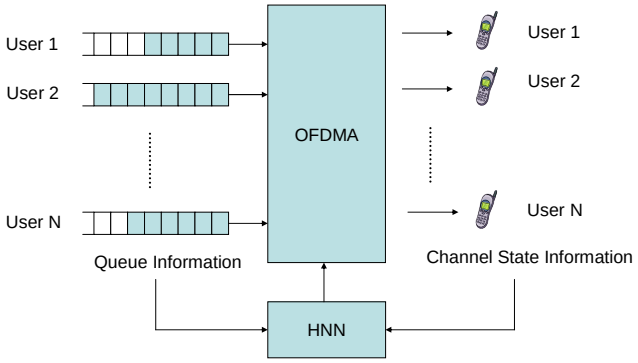

Figure 6-1 System model

The OFDMA system operates in frames of duration T and allocates a certain bit rate to each user by assigning to him a set of subcarriers. Multiple transmissions of different users in parallel are allowed by making use of different subcarrier combinations. A granularity of one subcarrier is considered in the assignment process. Then, a suitable eligible bit rate per user according to the queue status and service class requirements, as well as other terms to include the OFDMA downlink network restrictions has to be obteined These considerations related with queue status and OFDMA model are explained in the following.

6.1.1.1 Queuing Model

With respect to the queue model, we assume the model defined in sub-section 4.4.2.3, equation (4.26) and we introduce a generalisation for the Optimal Bit Rate

Definition *-We define the* ***Generalised Optimum Bit Rate (GOBR)*** *for the* i*-th user in the k-th frame, as the bit rate* $R_{b,i,gopt}(k)$ *that allows transmitting all its packets in due time, i.e:*

$$R_{b,i,gopt}(k) = \max_{m=1,\cdots,L_i} \left\{ v_{i,m}(k) \right\} .(1+\theta), \tag{6.1}$$

where $\theta \ (\geq 0)$ *is a safety empirical factor introduced to face fluctuations in the packet generation of the successive frames.*

The GOBR should be provided to each user by the OFDMA scheduler. To this end it will perform the most suitable aggregation of a given number of subcarriers. Notice that a continuous transmission at the OBR would avoid packet losses for this user. However it cannot always be guaranteed for all the users because of the total bandwidth restrictions. Notice that for $\theta = 0$ we have the definition used in previous chapters.

6.1.1.2 **OFDMA system model**

The system model assumes a total of *S* subcarriers with separation Δf (Hz) to be allocated to the N users. It is assumed that the transmitter knows the channel state at the terminal side, and in particular the receiver signal to noise ratio of the i-th user in the j-th subcarrier $\rho_{ij}(k)$ in the k-th frame. This value should be transmitted regularly by the mobile to the base station via a feedback channel, as illustrated in Figure 6-1 being the elapsed time lower than the channel coherence time. Then the actual capacity $c_{ij}(k)$ of the j-th QAM modulated subcarrier with Gray bit mapping in the k-th frame for the i-th user can be approximated by [62]:

$$\begin{aligned} &c_{ij}(k) = \log_2(1+\beta\rho_{ij}(k)) \quad b/s/Hz; \\ &\beta = -1,6/\ln(5\,BER), \end{aligned}, \tag{6.2}$$

where BER is the target Bit Error Rate. Then, the throughput of the OFDMA system in the *k*-th frame is given by:

$$R(k) = \sum_{i=1}^{N}\sum_{j=1}^{S} \chi_{ij}(k)\, c_{ij}(k)\, \Delta f\ . \tag{6.3}$$

Where $\chi_{ij}(k)$ is set to 1 when the j-th subcarrier is assigned to the i-th user and is set to 0 otherwise. Finally, as not all the $c_{ij}(k)$ values are allowed in a QAM modulation, the value obtained in (6.2) will be rounded to the highest integer lower than or equal to $c_{ij}(k)$ from the set $\{0,1,2,4,6\}$ bits/s/Hz.

The set of subcarriers assigned, in the *k*-th frame, has to satisfy

$$R(k) = \sum_{j=1}^{S} \chi_{ij}(k)\, c_{ij}(k)\, \Delta f \le R_{b,i,gopt}(k);\ i = 1,\ldots,N\ , \tag{6.4}$$

where $R_{b,i,gopt}(k)$ is the bit rate that allows i-th user to transmit all its packets in due time and

$$\sum_{i=1}^{N}\sum_{j=1}^{S}\chi_{ij}(k) \le S\,. \tag{6.5}$$

Then, the efficient exploitation of the total available subcarriers, with respect to their assignment to the differents users, can be obtained by minimizing

$$\left(\sum_{j=1}^{S}\chi_{ij}(k)c_{ij}(k)\Delta f - R_{b,i,gopt}(k)\right)^2;\ i=1,\ldots,N \tag{6.6}$$

and keeping (6.5).

6.2.3 Optimization process for the HNN-based scheduling model

The above DRA problem is an optimation problem that can be formulated by means of a 2D HNN as it is explained in Chapter 3, section 3.5. Assuming N users and S subcarriers the associated 2D HNN sheme requires $L = N \times S$ nodes. As in the above chapters we denote V_{ij}, $i=1,\ldots,N;\ j=1,\ldots,S$, the output value nodes and identify them with the assignment of subcarriers to the users, as follows

Definition. *The j-th subcarrier is assigned to the i-th user iff $V_{ij}=1$.*

Then, we identify the output value node V_{ij} with χ_{ij}. in equation (6.3).

It is worth noticing that, by selecting a suitable expression for the energy function E, an OFDMA queuing aware embedded optimization can be achieved. This optimization process of the HNN is carried out on a frame by frame basis and relies on minimizing the energy function through the convergence of a differential equation solution.

In the following two suitable expressions for the energy function compliant with the definition in (3.67) are introduced that will give rise to two different scheduling HNN based algorithms.

A. HNN1 algorithm

A first expression proposed for the energy E follows as:

$$E = \frac{\mu_1}{2}\sum_{i=1}^{N}\left(1-\frac{\sum_{j=1}^{S} c_{ij}\,V_{ij}\Delta f}{\overline{m_i}}\right)^2 + \frac{\mu_2}{2}\sum_{i=1}^{N}\sum_{j=1}^{S}\psi_{ij}V_{ij} + \frac{\mu_3}{2}\sum_{i=1}^{N}\sum_{j=1}^{S}V_{ij}\left(1-V_{ij}\right) + \frac{\mu_4}{2}\sum_{j=1}^{S}\left(1-\sum_{i=1}^{N}V_{ij}\right)^2 . \tag{6.7}$$

The first term is a cost function intended to be minimized by a proper setting of V_{ij}. It includes the expression

$$\overline{m_i} = \left(\left[\frac{R_{b,i,gopt}}{\Delta f}\right]+1\right)\cdot\Delta f\ , \tag{6.8}$$

where $[\ ]$ denotes the integer part, so that (6.8) is actually a quantification of the OBR value $R_{b,i,gopt}$ in multiples of Δf. The minimum value of the energy E would be achieved for specific combinations of V_{ij} that minimise each summand, by capturing the conditions (6.6), so that each user tends to be allocated with its OBR. Notice that OBR can change at each frame depending on the traffic dynamics and the packets evolution in the queues. Similarly, the channel fading also impacts OBR dynamics as the capacity of the different subcarriers can change on a frame basis.

The second summand in (6.7) simply penalizes the allocated subcarriers with bit rates equal to zero due to a bad channel state. That is, when the *i*-th user is considered with a subcarrier *j*-th in which $c_{ij}=0$, $\psi_{ij}=1$, thus increasing their contribution to the energy function and then bringing this allocation out of the energy minima and leaving the corresponding subcarrier available for other users. Otherwise it is set to $\psi_{ij}=0$.

The third summand of (6.7) was introduced in sub-section 3.5.3 in order to force convergence towards $V_{ij}\in\{0,1\}$ and the fourth term is introduced to reflect the physical OFDMA constraint that a given subcarrier can only be allocated to one user. The relationship between the energy function (6.7), the HNN interconnection matrix $T=\left[T_{ij,pq}\right]$ and the input bias current I_{ij}

values in the general expression of the energy function in (3.67) can be obtained according to the details shown in the Annex B.

B. HNN2 algorithm

A second expression for the energy function that captures additional features concerning both users and channel subcarrier status not considered in algorithm HNN1 is:

$$E=\frac{\mu_1}{2}\sum_{i=1}^{N}\omega_i\left(1-\frac{\sum_{j=1}^{S}c_{ij}V_{ij}\Delta f}{\overline{m_i}}\right)^2+\frac{\mu_3}{2}\sum_{i=1}^{N}\sum_{j=1}^{S}V_{ij}\left(1-V_{ij}\right)+\frac{\mu_4}{2}\sum_{j=1}^{S}\left(1-\sum_{i=1}^{N}V_{ij}\right)^2. \quad (6.9)$$

In this case, the first term has been modified with respect to the HNN1 algorithm with the inclusion of a new coefficient ω_i introduced with a two-fold objective. First, it should favour the users with high OBR that the former formulation of the term could not capture. Second, it should favour the allocation of subcarriers to the users with the best channel capacity thus making a better exploitation of the multiuser diversity. For that purpose, we have defined an order in the set of users , $order_i$, and a weight, ω_i , associated to each user as follows

Definition. *We define* $order_i$ *as the position of the* i*-th user in a list of users ordered in decreasing value of their OBR* $R_{b,i,gopt}$

Definition. *We define a set of weights* $\{\omega_i\}_{i=1,..,N}$ *, where* ω_i *is associated to the* i*-th user in as follows*

$$\omega_i=\left(2-\frac{1}{order_i}\right)\cdot\left(1-\frac{\sum_{j=1}^{S}c_{ij}}{\sum_{i'=1}^{N}\sum_{j=1}^{S}c_{i'j}}\right). \quad (6.10)$$

Notice that, with this definition, users with either a high OBR (i.e. a low value of $order_i$) or a good channel status in the different subcarriers will tend to have smaller values of ω_i and consequently the minima of the energy function will tend to occur in V_{ij} values so that a certain number of subcarriers is allocated to these users.

On the other hand, notice also that the effect of coefficient ω_i already captures to some extent the avoidance to allocate the subcarriers to the users with a bad channel status, which was intended by the second summand in the energy function of the HNN1 algorithm in (6.7), and consequently this summand is not included in the definition of the energy function for the HNN2 algorithm in (6.7) The Appendix B, section B.1 shows the relationship between the energy function in (6.7) and (6.9) and its interconnection matrix and input bias currents

The terms μ_1, μ_2, μ_3 and μ_4 in (6.7) and (6.9) are constants to be set in each model

Then once the final output values nodes, V_{ij}, are achieved as solution of the equations(3.66), the final bit rate assigned to the *i*-th user after the execution of the algorithm in one frame follows as

$$R_{b,i} = \sum_{j=1}^{S} V_{ij}\, c_{ij}\, \Delta f \ . \tag{6.11}$$

6.2.4 Reference scheduling schemes

The proposed HNN-based algorithms described in the previous section have been compared against other approaches. First, a simple heuristic reference scheduling algorithm has been considered, which exploits the OBR concept, but does not take into consideration the optimisation in accordance with the HNN procedure. This algorithm, denoted in the following as RSS1 (Reference Scheduling Scheme 1), simply tries to allocate to each user its optimum bit rate GOBR as defined in (6.1). The algorithm operates then in the following steps in each frame:

Step 1.- To order the users in increasing value of $R_{b,i,gopt}$.

Step 2.- To allocate sequentially to each user *i*-th the necessary number of subcarriers, so that its final scheduled bit rate is higher or equal than $\overline{m_i}$ from (6.8). This allocation is carried out by ordering first all the available subcarriers still pending to be allocated in increasing value of c_{ij}.

Step 3.- Once all the *S* available subcarriers have been allocated, assign a bit rate equal to 0 Kb/s (i.e. no transmission) to the remaining users.

Notice that, as far as the OFDMA capacity can satisfy the required OBR per user and frame, the reference system would lead to a quite satisfactory scheduling approach from a delay point of view.

Furthermore, focusing on the existing approaches in the literature, the proposed algorithm has also been compared against the recent proposal introduced in [58], which will be denoted in the following as RSS2 (Reference Scheduling Scheme 2). This algorithm also focuses on delay sensitive traffic and operates in two different steps:

The first step is the subcarrier allocation algorithm, which determines the number of subcarriers to be allocated to each user. For that purpose, it accounts for the different average channel conditions in all subcarriers as well as for the delay requirements of the different packets in the queue. After an initial computation, the algorithm executes several iterations in order to ensure that the total number of allocated subcarriers equals the number of available subcarriers *S*.

In a second step, the subcarrier assignment algorithm is executed deciding the specific subcarriers allocated to each user. This is done by creating a priority list ordering the different users in accordance with the history of packet droppings experienced by each one, so that users with a higher number of droppings have higher priority. In turn, for users with equal number of droppings, the priority is computed in accordance with the channel quality (i.e. users with better quality have higher priority). Then, according to the priority list, each user selects the best available subcarriers up to the number of subcarriers computed in the first step.

For details of the algorithm the reader is referred to [58]. It is worth mentioning that this algorithm was in turn compared in [58] against other previous references, such as [60], exhibiting a better performance. Consequently, this algorithm has been retained here for comparison purposes as an appropriate reference representative of the state-of-the art in OFDMA dynamic resource allocation algorithms for delay sensitive traffic.

6.2.5 Simulation results

A single cell scenario has been considered to assess the proposed HNN-based DRA strategy for a downlink OFDMA wireless access. We consider $S = 128$ subcarriers and $\Delta f = 15$ kHz. In the simulation, the scheduling algorithm operates in frames of $T = 10$ ms. We will also assume the

coherence time is larger than the frame time so as within a frame it is assumed that the channel impulse response does not vary. In our simulation each user channel suffers from multipath Rayleigh fading with a delay profile characterized by a time variant impulse response following the pedestrian model of [44] with a mobile speed of 5 km/h and an average signal-to-noise ratio equal to 17 dB. We let a target $\text{BER} = 10^{-4}$ and we assume a set of possible transmission bit rates: $15\,m$ Kb/s per subcarrier ($m = \{0,1,2,4,6\}$) by properly adjusting the modulation levels of a 2^m QAM adopted signalling format.

The selected parameters appearing in the formulation of the HNN are $\mu_1 = 4000$, $\mu_2 = 30000$, $\mu_3 = 800$, $\mu_4 = 18000$, τ=1 and α=1.0. Simulations not shown here concerning the variation of these parameters have revealed that they are actually robust values so that changing them to a certain extent (i.e. variations as large as 50% have been tested) does not impact significantly the final results. The only conditions are that these parameters should be positive and satisfy $\mu_3 < \mu_4$, as it is shown in the Appendix B, section B.1.

On the other hand, the iterative numerical solution given by (3.50) is finalized when the iterations n and $n-1$ satisfy $\left\| W^n - W^{n-1} \right\|_2 < \varepsilon$, where $W^n = \left(w_1^n, w_2^n, \ldots, w_S^n \right)$ and $\| \ \|_2$ is the Euclidean Norm. We have set $\Delta = 10^{-4}$ in (6.8) and $\varepsilon = 10^{-5}$. The convergence to a stable value is attained in practice in most of the situations between 1000 and 1500 iterations. As a result of that, a maximum of 2000 iterations has been used to stop the iterative process. If all these conditions are fulfilled we decide the process converges and the values V_{ij} provide us the inputs to calculate the total bit rate allocated to each user in each frame $R_{b,i}$ according to (6.11)

An interactive service following the www traffic model from [44] has been considered as representative of a delay-sensitive service. Specifically, www sessions are composed by an average of 5 pages, with an average time between pages of 30s. In each page, the average number of packets is 25 with an average time between packets of 0.0277s. The packet length follows a Pareto with cut-off distribution with parameters alpha=1.1, minimum packet size 81.5 bytes and maximum packet size 6000 bytes. The average time between www sessions is 0.1s (i.e. it is assumed that a user is continuously generating sessions). Two interactive user classes, namely Class 1 and Class 2, have been included, as representative of two different user profiles, with maximum allowed delays of 120 ms and 60ms, respectively. 60% of the users belong to class 1 and 40% to class 2.

By setting the parameter $\theta > 0$ for the GOBR in (6.1), queues are forced to be emptied faster than for $\theta = 0$, which is particularly true for low loaded systems. However, there is not an optimum θ setting unique for all the loads. Then, from the obtained results $\theta = 0.6$ has been retained as a satisfactory value in all the studied cases. Let us notice that, in general, high values of θ could end up in assigning bandwidth in excess to some users in detriment of others. This is clearly pointed out in the RSS1 scheme where the first ordered users could be provided with an excessive bandwidth (and actually not required), which would prevent the allocation to other users in the ordered list.

Figure 6-2 plots the comparison between the considered strategies in terms of packet dropping probability for class 1 users as a function of the total number of users in the scenario (similar results not shown here would be observed for class 2 users). It can be observed that the worst performance is obtained with the RSS1 scheme, and that the two approaches based on HNN are able to outperform both RSS1 and RSS2 strategies, thanks to the consideration of both queuing time constraints and channel status in the optimisation carried out by Hopfield Neural Networks. Notice that, for low dropping probability values the reduction achieved by HNN-based strategies is in around one order of magnitude with respect to both RSS1 and RSS2. In that respect, notice also that the energy function from HNN2 is able to achieve always a lower dropping probability than the energy function from HNN1. Equivalently, the performance in terms of dropping probability can be translated into a certain system capacity (i.e. maximum number of users that the system can handle for a certain maximum dropping probability of e.g. 1%). Specifically, while RSS2 would exhibit a capacity of around 1200 users, in the case of HNN2 the capacity is increased up to around 1350 users (i.e. a capacity gain of 13%).

With respect to the performance on average terms, Figure 6-3 compares the average packet delay measured for class 1 users with the different approaches. In this case, the comparison reveals that HNN-based approaches achieve an average delay that lies between the RSS1 and RSS2 schemes. However, Figure 6-4, which plots the ratio between standard deviation and average delay for each strategy, indicates that RSS2 is actually the strategy with the highest dispersion in terms of delay, which eventually justifies that, in spite of having a good performance on average terms, the packet dropping ratio is higher than with the HNN-based algorithms. Consequently, whenever delay sensitive traffic is considered, the performance should not be optimised only on average terms but also specific conditions in terms of maximum allowed delays should be considered. Finally, it is worth mentioning that actually both HNN1 and HNN2 provide just an upper bound of the dropping probability due to the above mentioned truncation of the iterative Euler technique and the existence of a minimum in

the energy function. So, even better results could be expected either by exploring other numerical solutions or alternative improved energy function definitions, what is left for future work.

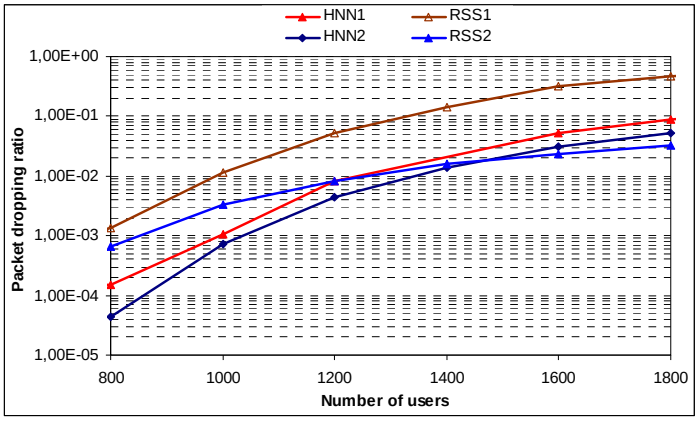

Figure 6-2. Packet dropping ratio for class 1 users as a function of the number of users in the scenario

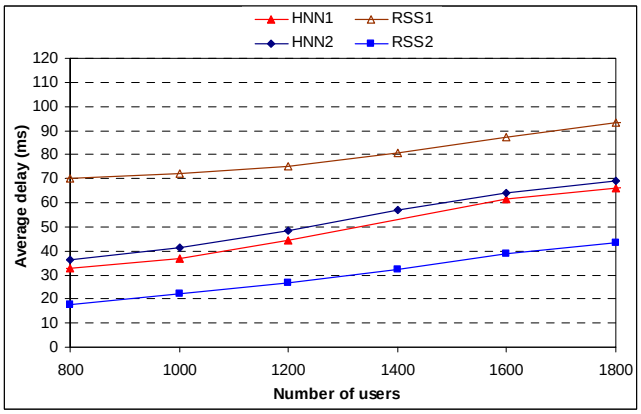

Figure 6-3. Average packet delay for class 1 users as a function of the number of users in the scenario

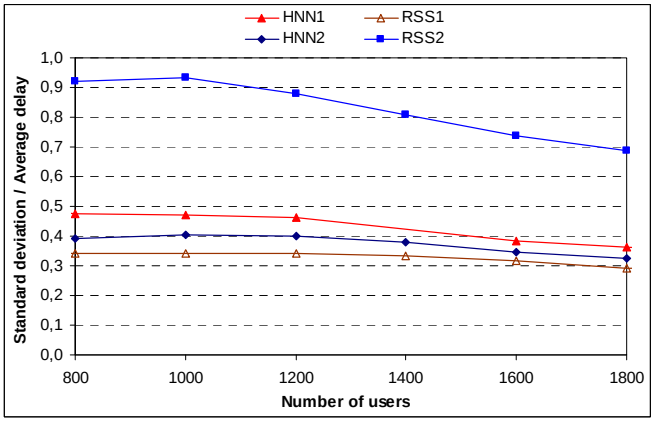

Figure 6-4. Ratio between the standard deviation and the average of the packet delay for class 1 users

As an illustrative result of how the different algorithms operate, Figure 6-5 plots the CDF of the bit rate allocated per user with the different approaches for a situation with 1000 users in the scenario (for illustrative purposes, only the bit rate of class 1 users is presented, but the performance for class 2 users would be similar). Actually only the most relevant part of CDF relative to the highest percentile is stressed in Figure 6-5 to better differentiate the reference and the HNN based scheduler algorithms operation. It can be observed that both HNN-based strategies are able to make allocations of higher bit rates, thanks to the HNN-optimisation accounting for the joint queue and channel status, which ensures that the subcarriers are allocated to the most suitable users. In that respect, the main difference between HNN1 and HNN2 would be for the lowest bit rates (i.e. below 30 Kb/s, not shown in the graph, and where the crossing point between the HNN1 and HNN2 curves occurs), in which HNN1 would exhibit a higher probability than HNN2 of allocating low bit rates.

Finally, Figure 6-6 plots the comparison in terms of the CDF of the total allocated bandwidth obtained with HNN1 with respect to the total requested bandwidth (i.e. the sum of all the OBRs of the different users) for the cases with 1200 users and 1600 users. It can be observed how the total requested bandwidth increases with the number of users but the total allocated bandwidth remains approximately the same, meaning that the system has reached its maximum capacity. However, in spite of the fact that the total requested bandwidth is higher than the total allocated bandwidth, the algorithm carries out a smart allocation that keeps the packet dropping probability at low values, as illustrated in Figure 6-2. Similar results are obtained with HNN2.

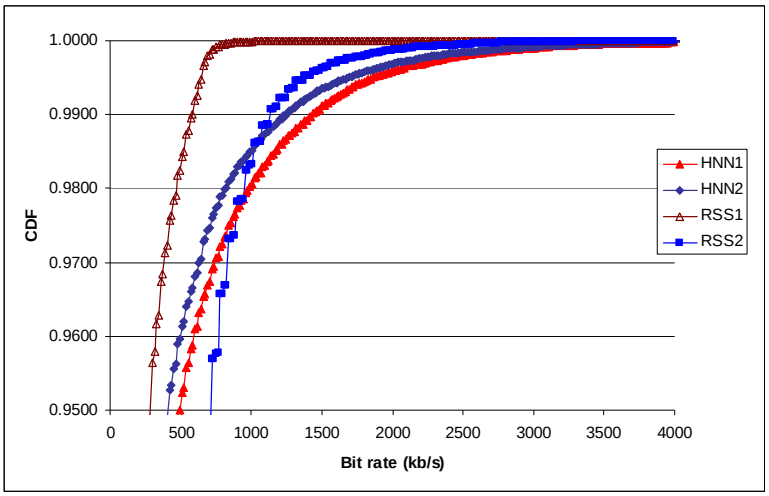

Figure 6-5. CDF of the allocated bit rate for class 1 users for the different strategies with 1000 users in the scenario

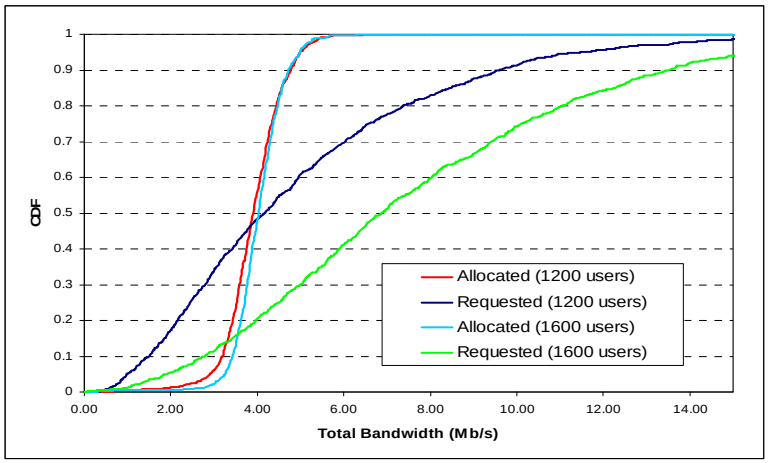

Figure 6-6. Cumulative Distribution Function of the total allocated and requested bandwidth for the HNN1 algorithm with 1200 and 1600 users

6.3 Conclusions

This chapter has presented a novel strategy to carry out the dynamic resource allocation of subcarriers to users in OFDMA systems with delay sensitive service in which packets should be transmitted within a specific maximum delay bound. It is based on Hopfield Neural Network methodology which is a powerful optimisation technique and takes into account both service class constraints in terms of maximum allowed delay as well as channel capacity limitation in each subcarrier. Actually, HNN methodology has been carried out by solving iteratively a numerical differential equation having a hardware implementation in mind, and the different delay requirements are captured in the form of an energy function that is minimised by the algorithm. In that respect, two different energy functions have been analysed by means of simulations and compared against two reference schemes, revealing a better behaviour in terms of packet dropping probability, which eventually turns into system capacity increases. Specifically, capacity gains of around 13% for a maximum dropping probability of 1% have been observed with respect to a representative state-of-the-art algorithm existing in the literature.

Chapter 7

Conclusions and Future work

7.1 Conclusions

This thesis has introduced a framework based in Hopfield Neural Networks suitable for the definition of new RRM algorithms in 3G, B3G and 4G mobile communications. Following the partial detailed conclusions already stated in the different chapters, a summary of them is assembled in the sequel. A mechanism for the Admission Control in uplink WCDMA has been first developed that allocates optimally the available resources according to the specific contracted user profile and provided some restrictions in terms of load, fairness and maximum power available. Then, a new user-centric approach has been presented so as to provide a maximum user satisfaction probability in terms of minimum satisfactory throughput. In both cases real time traffic with variable bit rate has been assumed. As for delay sensitive traffic, a new scheduling algorithm has also been presented for downlink WCDMA and interactive traffic, aiming at guaranteeing the allowed maximum packet delay for a given packet dropping rate provided the base station transmitted power restrictions. In addition to that, new RRM algorithms have also been introduced to manage a mix of real time and interactive services. A B3G scenario as been also considered to manage a proper downlink scheduler for delay sensitive services when two Radio Access Technologies: WCDMA and GERAN were present. In this case the proposed algorithm has considered the timing constraints to avoid ping-pong effects in the vertical Handover and it has been shown the impact of this constraint in the obtained performance. Finally a novel strategy has been presented to carry out the dynamic resource allocation of subcarriers to users in downlink OFDMA systems with delay sensitive service in which packets should be transmitted within a specific maximum delay bound. Hopfield Neural Network methodology has also proved as a powerful optimisation technique and takes into account both service class constraints in terms of maximum allowed delay as well as channel capacity limitation in each subcarrier. In all the cases satisfactory results have been provided which compare favourably with other reference algorithms.

7.2 Future work

7.2.1 New RRM scenarios

The analysis presented in this thesis has revealed the flexibility and the promising behavior of HNN to handle multiple RRM problems in wireless communication scenarios. Based on this analysis, a number of future research lines have been identified to further evaluate the applicability of HNNs in other situations, as summarised in the following:

- Extension to multi-cell scenarios: This work has not considered the effect of the multicell interference arising in a cellular system. In a future work we plan to extend the RRM based in HNN algorithms to such multicell scenario, with a particular focus on the OFDMA schemes that are being widely proposed for the next 4G of Mobile Systems. In a first stage we plan to address a centralized management scenario where the HNN based algorithms have a global picture of all the cells. Each cell is assumed to be allocated with a given number of subcarriers extracted from a frequency planning tool. Then intercell interference between different cells using the same subcarrier could be present and the HNN algorithm should be capable to consider it in a similar way as it considered bad propagation channels for certain subcarriers due to the fading effect in the unicellular case.

- Decentralised HNN operation: One of the current trends in wireless communications is the decentralisation of RRM functions towards end nodes in the radio access network. This is the case of e.g. HSPA, which implements packet scheduling functions in the base station in order to get reduced latencies and to be more efficient from a signalling point of view. In this respect, considering a multi-cell network, we plan to address a decentralised or distributed HNN scheme where the HNN algorithm is run in each cell as in the scenario considered in this thesis but considering now that it has to be aware of the impact of the interference produced by the rest of cells using the same subcarriers. The decentralised approach implies a lower signalling effort than the centralized one as now the subcarrier allocation has not to be distributed to all the cells but each cell decides for its own.

7.2.2 New mathematical approaches for solving HNN-based RRM

We have presented in this thesis how Hopfield Neural Networks can be applied to study Radio Resource Management obtaining satisfactory results in all the studied problems. The obtained results correspond to stable and sub-optimal solutions for the different RRM scenarios analyzed.

Nevertheless, either from the mathematical as well as from the simulation optimization viewpoint there is still room to improve the adopted methodology. In particular as for the mathematical HNN development we have left as future work to explore the stability and convergence of the HNN dynamic differential equations in order to gain a better insight in the existing different suboptimal solutions and be able to reach the optimal one in case. A similar approach is planed in relation to explore the most appropriate numerical analysis to be used to improve the standard Euler method retained in this work to solve the HNN dynamic differential equations. Furthermore, both studies would allow us to use HNN in scenarios where a major number of nodes and more profiles could be considered. In particular, some particular studies to be pursued have been identified in the following.

7.2.2.1 Analysis of Stable states of Hopfield Lyapunov Functions.

In the problems presented in this thesis it has been established a generic energy function to reach globally asymptotically states in a continuous HNN. Depending on its connection matrix (weights) election several energy functions can be proposed following that model. In order to make a choice between them it is convenientto study the conditions under which a solution converges to a vertex, to a point on a surface or to an interior point. Classical approaches that can be found in the literature to study the energy behaviour are:

- Applying an eigenvalue analyses around an equilibrium point. The knowledge of the eigenvalues can be used to gain an understanding of the relationship between the network parameters and the type of solution the network reaches [63, 64].
-
- Studying the convergence region of local minima [65, 66, 67]. An analysis around an equilibrium point clarifies conditions to converge to the surface of a hypercube, especially the condition that a vertex becomes a local minimum of the HNN and the form of the convergence region that can lead to that minimum.
-

7.2.2.2 Numerical simulation of a Hopfield Neural Network

An Euler discretization of the HNN dynamical differential equation can be made using the Euler method, i.e., using

$$\frac{du_k(t)}{dt} \approx \frac{u_k(t+h)-u_k(t)}{h} \tag{7.1}$$

equations (3.43) can be discretizated as

$$u_k(t+h) = u_k(t+h) + h\left(-\frac{1}{\tau}u_k(t) + \sum_{j=1}^{S} T_{kj}v_j + I_k\right), \; k = 1, 2, \ldots, S \,. \qquad (7.2)$$
$$v_j(t) = \frac{1}{1+\mathrm{e}^{-\alpha_j u(t)}}$$

The Euler aproximation depends on the choice of the time step h; the smaller it is, the more accurate is the approximation but greater the time of simulation. So, an adaptatives step h is propose to be implemented to obtain a better performance.

Otherwise, theoretically the error in n-th iterate using the approximation gives by (3.50), i.e. $e^n = \left|u_k(t_n) - w_k^n\right|$, grows slightly as the value of n increases. This controlled error growth is a consequence of the stability of Euler's method, which implies that the errors due to rounding are expected to grow in no worse than a linear manner. When the number of nodes in the HNN increase, it increase also the number of iterations to have good approximations to stable points. Then, other strategies consisting on using more accurate numerical methods to discretize the differential equations, such as Runge-Kutta methods [68] can be used.

Annex A

Reference scenario

The considered scenario for evaluation purposes is a single cell scenario. The cell can be configured either to work only with WCDMA, with WCDMA and GERAN, or with OFDMA. Dynamic simulations are considered, since they are able to model movement, call set-up and release, service demand variations and quality changes in received signal. Service demanding mobiles are generated according to a certain traffic distribution and they move throughout the whole zone. As mobile units wander within the described scenario, they receive data packets, perform power control if needed, undergo inter-system handover and may be dropped if quality falls beyond a limit.

A.1 Mobility model

The pseudo-random mobility model [44] with semi-directed trajectories is considered. In this model the mobiles are characterised by a speed and a direction, and the position is updated every d_c meters following the corresponding direction, where d_c is the decorrelation length typically of 20 m. At each position update the direction is changed with probability 0.2 and in this case the angle for the direction change is selected randomly in the range $[-45, 45]$ degrees with uniform distribution.

The process is illustrated in Figure A-1 At initialisation, the mobiles are scattered in the scenario and the mobile direction is randomly chosen. When reaching the scenario boundary, depending on the mobility type, the users are reflected back with a change in direction by 180 degree.

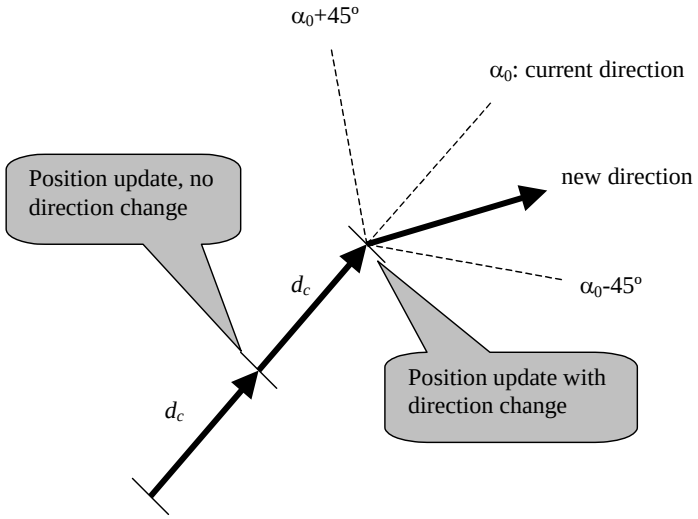

Figure A-1 Mobility model

A.2 Propagation model

Macro-cell environments are characterised by large cells and high-transmitted powers in outdoor urban and sub-urban areas with antennas over the roof top levels. The path loss model used in this case is extracted from [44]

$$L_p\left(dB\right) = 40 \cdot \left(1 - 4 \cdot 10^{-3} \Delta h_b\right) \log_{10} d - 18 \log_{10} \Delta h_b + 21 \log_{10} f + 80, \qquad \text{(A.1)}$$

,

where d is the distance in km between the base station transceiver and the mobile terminal, f is the carrier frequency in MHz and Δh_b is the base station antenna height, measured in m from the average roof top level.

Typical values for these parameters are $\Delta h_b = 15m$ and $f = 2000$ MHz, leading to

$$L_p\left(dB\right) = 128.1 + 37.6 \log_{10} d\,. \qquad \text{(A.2)}$$

The path loss L shall in no circumstances be less than the free space loss given by:

$$L_f\left(dB\right) = 20 \log_{10} d + 20 \log_{10} f + 32.44\,. \qquad \text{(A.3)}$$

In addition to the propagation loss, a lognormal shadowing model is considered, with standard deviation σ. The shadowing in different spatial positions is correlated. In particular, an exponential normalised decorrelation function is considered, for two points separated x m:

$$R(x) = e^{-\frac{x}{d_c}\ln 2}, \qquad \text{(A.4)}$$

where d_c is the decorrelation distance, being $20m$ a typical value.

In the OFDMA analysis an additional fast fading model is considered based on the ITU Pedestrian A [69]

A.3 Interactive traffic model for delay-sensitive services

The web browsing traffic model applies for interactive services in which the end-user requests on-line data from remote equipment. The considered traffic model consists of www browsing sessions, in which the user downloads some web pages, called “packet calls”. Each of these packet calls is formed by different packets representing typically the HTML objects in a www page. Between two consecutive packet calls there is a “reading time”, which models the time that the user would spend reading the content of the downloaded page. The overall www browsing session model is shown graphically inFigure A-1. In the case of File Transfer Protocol (FTP), the same model can be applied simply by considering a single packet call in the session.

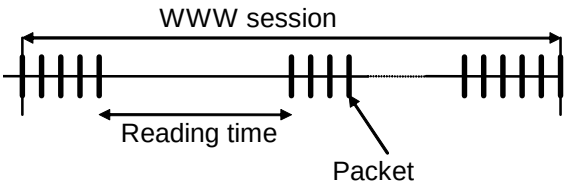

Figure A-2 WWW browsing session model.

The model is based in broad outlines on the one specified in UMTS 30.03 document, but some modifications of the parameters have been done in order to achieve a figure of the total generated traffic per user closer to current traffic measurements, as presented in Table 1.

Parameter	Uplink values	Downlink values
Session arrival process	Users are continuously generating sessions	Users are continuously generating sessions
Number of packet calls (pages) per session	Geometrical random variable with average 5	Geometrical random variable with average 5
Reading time between packet calls. The reading time starts when the last packet of the packet call is completely received by the user, and ends when the user makes a request for the next packet call.	Exponential random variable with average $20s$	Exponential random variable with average $20s$
Number of packets within a packet call.	Geometrical random variable with average 25	Geometrical random variable with average 25
Inter arrival time between packets within a packet call	Exponential random variable with average $0.25s$	Exponential random variable with average $0.25s$
Packet size	Pareto distribution with cut-off and parameters: $k = 65$ bytes, $\alpha = 1.1$, $m = 20000$ bytes (This provides an average packet size equal to 350 bytes)	Pareto distribution with cut-off and parameters: $k = 390$ bytes, $\alpha = 1.1$, $m = 100000$ bytes (This provides an average packet size equal to 2050 bytes and an average web size of about 50kB)
Average workload per user:	2.7 Kb/s	15.6 Kb/s

Table 1. Parameters of the WWW source model.

Annex B

Interconnection matrices and input bias currents for the proposed HNN Models

In a 2D HNN sheme as that described in chapter 3 (sub-section 3.5.3), each input value node $U_{i,j}$ has resistive connections with the other nodes and these connections can be described by the interconnection matrix $T = \left[T_{ij,kl} \right]$. Here, $T_{ij,kl}$ is the interconnection weight from the $ij-$th node to the $kl-$th node. Each node also receives an input bias current I_{ij}, which is the only user-adjustable parameter. The dynamics of the HNN are represented by

$$\frac{dU_{ij}}{dt} = -\frac{1}{\tau} U_{ij} + \sum_{k=1}^{N} \sum_{l=1}^{M} T_{ij,kl} V_{kl} + I_{ij}, \qquad i = 1, \ldots, N; \; j = 1, \ldots, M. \tag{B.1}$$

Hopfield proved that for a symmetric interconnection matrix T and sufficiently high gains of the amplifiers, nodes evolve by gradient descent of the quadratic energy function E,

$$E = -\frac{1}{2} \sum_{k=1}^{N} \sum_{l=1}^{M} \sum_{i=1}^{N} \sum_{j=1}^{M} T_{ij,kl} V_{ij} V_{kl} - \sum_{i=1}^{N} \sum_{j=1}^{M} I_{ij} V_{ij} \; . \tag{B.2}$$

Using (B.2), (B.1) may be rewritten as follows

$$\frac{dU_{ij}}{dt} = -\frac{U_{ij}}{\tau} - \frac{\partial E}{\partial V_{ij}}. \tag{B.3}$$

Then, using (B.1) and (B.3) it follows,

$$\sum_{k=1}^{N} \sum_{l=1}^{M} T_{ij,kl} V_{kl} + I_{ij} = -\frac{\partial E}{\partial V_{ij}}; \quad i = 1, \ldots, N; \; j = 1, \ldots, M \tag{B.4}$$

This annex contains the interconnection matrix and input bias vector for each of the energy functions presented in this thesis.

B.1 HNN WCDMA Model

The HNN WCDMA model and its associated energy were presented in sub-section 4.2.3 and also used in sub-sections 4.3.3, 4.4.3 and 4.4.5. In the first case the cost values do not change throught all the simulation. In the last three cases they change in each frame but the formal general expression is preserved and it is shown in (B.5), where we have omitted the subscrip related to the frame.

$$E = \frac{\mu_1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} C_{i,j}V_{i,j} + \frac{\eta^{\varsigma}\mu_2}{2}\left|1-\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{R_{i,j}}{B_T}V_{i,j}\right| + \frac{\mu_3}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}\psi_{i,j}V_{i,j} + \frac{\mu_4}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}V_{i,j}(1-V_{i,j}) + \frac{\mu_5}{2}\sum_{i=1}^{N}(1-\sum_{j=1}^{M}V_{i,j})^2, \tag{B.5}$$

$$\varsigma = u\left(\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{R_{ij}}{B_T}V_{ij} - 1\right).$$

For the above energy function E and a given V_{i^*,j^*} we have ([36])

$$\frac{\partial E}{\partial V_{i^*,j^*}} = \frac{\mu_1}{2}C_{i^*,j^*} + (-1)^{1-\varsigma}\frac{\eta^{\varsigma}\mu_2}{2}\frac{R_{i^*,j^*}}{B_T} + \frac{\mu_3}{2}\psi_{i^*,j^*} + \frac{\mu_4}{2}\left(1-2V_{i^*,j^*}\right) - \mu_5\left(1-\sum_{j=1}^{M}V_{i^*,j^*}\right). \tag{B.6}$$

Using (B.6) and equating the corresponding coefficients in (B.4), the interconnection matrix and current bias vector in a frame are found to be

$$T_{ij,kl} = \mu_4\delta_{ik}\delta_{jl} - \mu_5\delta_{ik}, \tag{B.7}$$

$$I_{i,j} = -\frac{\mu_1}{2}C_{i,j} - (-1)^{1-\varsigma}\cdot\frac{\eta^{\varsigma}\mu_2}{2}\cdot\frac{R_{i,j}}{B_T} - \frac{\mu_3}{2}\psi_{i,j} - \frac{\mu_4}{2} + \mu_5. \tag{B.8}$$

Then, the interconnection matrix is the following $(N\cdot M)\times(N\cdot M)$-block diagonal matrix

$$T = \begin{pmatrix} A & O & \cdots & O \\ O & A & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ O & \cdots & O & A \end{pmatrix} \tag{B.9}$$

where O and A are the following $M\times M$ -matrices

$$O=\begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 0 \end{pmatrix}; A=\begin{pmatrix} \mu_4-\mu_5 & -\mu_5 & \cdots & \cdots & -\mu_5 \\ -\mu_5 & \mu_4-\mu_5 & -\mu_5 & \ddots & -\mu_5 \\ \vdots & -\mu_5 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -\mu_5 \\ -\mu_5 & \cdots & \cdots & -\mu_5 & \mu_4-\mu_5 \end{pmatrix}. \quad \text{(B.10)}$$

Let C be the $M\times M$ -matrix which all its elements equals to value 1 and let I be the $M\times M$ -identity matrix, then matrix A can be written as

$$A=\mu_4 C-\mu_5 I. \quad \text{(B.11)}$$

Notice as T is a symmetric matrix. The diagonal elements are no zero, then a correction is necessary by changing $I_{kl}\leftarrow I_{kl}-\frac{T_{kl,kl}}{2}$ as it was mentioned in sub-section 3.5.3.

In order to have minimum points of the energy function with respecte the output value nodes, V_{ij}, the energy function must be positive definite; it is necessary that the second derivatives be positive, i.e.

$$\frac{\partial^2 E}{\partial V_{ij}^{\ 2}}>0\Rightarrow \mu_4<\mu_5. \quad \text{(B.12)}$$

B.2 HNN CRRM Model

In this sub-section we present an analysis of the following energy function considered for CRRM problems. Althought the cost values are different in each frame, for simplicity, we will omit the frame's subscript. Then

$$E=\frac{\mu_1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{p=1}^{P}C^k_{i,j,p}V_{i,j,p}+\sum_{p=1}^{P}\frac{\eta^{\varsigma_p}\mu_2}{2}\left|1-\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{R_{i,j,p}}{B_p}V_{i,j,p}\right|+\frac{\mu_3}{2}\sum_{p=1}^{P}\sum_{i=1}^{N}\sum_{j=1}^{M}\psi_{i,j,p}V_{i,j,p}+$$
$$+\frac{\mu_4}{2}\sum_{p=1}^{P}\sum_{i=1}^{N}\sum_{j=1}^{M}V_{i,j,p}\left(1-V_{i,j,p}\right)+\frac{\mu_5}{2}\sum_{i=1}^{N}\left(1-\sum_{p=1}^{P}\sum_{j=1}^{M}V_{i,j,p}\right)^2;$$
$$\varsigma_p=u\left(\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{R_{i,j,p}}{B_p}V_{i,j,p}-1\right). \quad \text{(B.13)}$$

For the above energy function E and a given V_{i^*,j^*,p^*} we have

$$\frac{\partial E}{\partial V_{i^*,j^*,p^*}} = \frac{\mu_1}{2} C_{i^*,j^*,p^*} + (-1)^{1-\varsigma_{p^*}} \frac{\eta^{\varsigma_{p^*}} \mu_2}{2} \frac{R_{i^*,j^*,p^*}}{B_T} + \frac{\mu_3}{2} \psi_{i^*,j^*,p^*} + \frac{\mu_4}{2}\left(1 - 2V_{i^*,j^*,p^*}\right) - \\ -\mu_5 \left(1 - \sum_{p=1}^{P} \sum_{j=1}^{M} V_{i^*,j^*,p^*}\right) \tag{B.14}$$

For CRRM the equivalence to equation (B.4) is

$$\sum_{k=1}^{N} \sum_{l=1}^{M} \sum_{q=1}^{P} T_{ijp,klq} f_{klq}\left(U_{klq}(t)\right) + I_{ijp} = -\frac{\partial E}{\partial V_{ijp}}; i = 1,2,\ldots,N; j = 1,2,\ldots,M; p = 1,\ldots,P \tag{B.15}$$

Using (B.14) and equating the corresponding coefficients in (B.15), the interconnection matrix and current bias vector in a frame are found to be

$$T_{ijp,klq} = \mu_4 \delta_{ik} \delta_{jl} \delta_{pq} - \mu_5 \delta_{ik}, \tag{B.16}$$

$$I_{i,j,p} = -\frac{\mu_1}{2} C_{i,j,p} - (-1)^{1-\varsigma_p} \frac{\eta^{\varsigma_p} \mu_2}{2} \cdot \frac{R_{i,j,p}}{B_T} - \frac{\mu_3}{2} \psi_{i,j,p} - \frac{\mu_4}{2} + \mu_5. \tag{B.17}$$

Then, the interconnection matrix is the following $(N \cdot M \cdot P) \times (N \cdot M \cdot P)$ block diagonal matrix

$$T = \begin{pmatrix} A & O & \cdots & O \\ O & A & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ O & \cdots & O & A \end{pmatrix} \tag{B.18}$$

where O and A are the $(M \cdot P) \times (M \cdot P)$ matrices

$$O = \begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 0 \end{pmatrix}; A = \begin{pmatrix} \mu_4 - \mu_5 & -\mu_5 & \cdots & \cdots & -\mu_5 \\ -\mu_5 & \mu_4 - \mu_5 & -\mu_5 & \ddots & -\mu_5 \\ \vdots & -\mu_5 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -\mu_5 \\ -\mu_5 & \cdots & \cdots & -\mu_5 & \mu_4 - \mu_5 \end{pmatrix}. \tag{B.19}$$

Let C be the $(M \cdot P)\times(M \cdot P)$-matrix which all its elements equals to value 1 and let I be the $(M \cdot P)\times(M \cdot P)$-identity matrix, then matrix A can be written as

$$A = \mu_4 C - \mu_5 I. \tag{B.20}$$

Notice as T is a symmetric matrix. The diagonal elements are no zero, then a correction is necessary by changing $I_{klq} \leftarrow I_{klq} - \frac{T_{klq,klq}}{2}$ as it was mentioned in sub-section 3.5.3.

In order to have minimum points of the energy function with respecte the output value nodes, V_{ijp} , the energy function must be positive definite; it is necessary that the second derivatives be positive, i.e.

$$\frac{\partial^2 E}{\partial V_{ijp}^{\ 2}} > 0 \Rightarrow \mu_4 < \mu_5. \tag{B.21}$$

B.3 HNN OFDMA Model

This sub-section is devoted to study the energy functions presented for 2D HNN OFDMA models, Equations (6.7) and (6.9).

In order to make the derivation valid for both the HNN1 and the HNN2 algorithms, let consider the following common definition of the energy function *E*:

$$E = \frac{\mu_1}{2}\sum_{i=1}^{N}\omega_i\left(1-\frac{\sum_{j=1}^{S}c_{ij}V_{ij}\Delta f}{\overline{m_i}}\right)^2 + \frac{\mu_2}{2}\sum_{i=1}^{N}\sum_{j=1}^{S}\psi_{ij}V_{ij} + \frac{\mu_3}{2}\sum_{i=1}^{N}\sum_{j=1}^{S}V_{ij}\left(1-V_{ij}\right) + \frac{\mu_4}{2}\sum_{j=1}^{S}\left(1-\sum_{i=1}^{N}V_{ij}\right)^2. \tag{B.22}$$

Notice that the energy function of HNN1 in (6.7) is obtained by taking $\omega_i = 1; i = 1,\ldots,N$ in (B.22), while the energy function of HNN2 in (6.9) is obtained by taking μ_2=0.

For the energy function E defined as (B.22) and a given V_{i^*,j^*} we obtain

$$\frac{\partial E}{\partial V_{i^*,j^*}} = \frac{\mu_1}{2}\sum_{i=1}^{N} 2\omega_i \left(1 - \frac{\sum_{j=1}^{S} c_{ij} V_{i,j} \Delta f}{\overline{m_i}}\right)\left(-\sum_{j=1}^{S} \frac{c_{ij}\Delta f}{\overline{m_i}} \frac{\partial V_{i,j}}{\partial V_{i^*,j^*}}\right) + \frac{\mu_2}{2}\sum_{i=1}^{N}\sum_{j=1}^{S} \psi_{ij} \frac{\partial V_{i,j}}{\partial V_{i^*,j^*}} +$$

$$+\frac{\mu_3}{2}\sum_{i=1}^{N}\sum_{j=1}^{S}\left[\frac{\partial V_{ij}}{\partial V_{i^*j^*}}\left(1-V_{i,j}\right) + V_{i,j}\frac{\partial\left(1-V_{i,j}\right)}{\partial V_{i^*,j^*}}\right] + \frac{\mu_4}{2}\sum_{j=1}^{S} 2\left(1-\sum_{i=1}^{N} V_{i,j}\right)\left(-\sum_{i=1}^{N}\frac{\partial V_{i,j}}{\partial V_{i^*,j^*}}\right). \tag{B.23}$$

Furthermore, since $\frac{\partial V_{ij}}{\partial V_{i^*j^*}} = 1$ for $i = i^*, j = j^*$ and $\frac{\partial V_{ij}}{\partial V_{i^*j^*}} = 0$ for $i \neq i^*, j \neq j^*$, equation (B.23) can be expressed as:

$$\frac{\partial E}{\partial V_{i^*,j^*}} = -\mu_1 \frac{c_{i^*,j^*}\Delta f\, \omega_{i^*}}{\overline{m_{i^*}}}\left(1 - \frac{\sum_{j=1}^{S} c_{i^*,j} V_{i^*,j}\Delta f}{\overline{m_{i^*}}}\right) + \frac{\mu_2}{2}\psi_{i^*,j^*} + \frac{\mu_3}{2}\left(1-2V_{i^*,j^*}\right) - \mu_4\left(1-\sum_{i=1}^{N} V_{i,j^*}\right) \tag{B.24}$$

By substituting (B.24) into (B.3) it is obtained

$$\frac{\partial E}{\partial U_{i^*,j^*}} = -\frac{U_{i^*j^*}}{\tau} + \mu_1 \frac{c_{i^*j^*}\Delta f\, \omega_{i^*}}{\overline{m_{i^*}}}\left(1 - \frac{\sum_{j=1}^{S} c_{i^*j} V_{i^*j}\Delta f}{\overline{m_{i^*}}}\right) - \frac{\mu_2}{2}\psi_{i^*j^*} - \frac{\mu_3}{2}\left(1-2V_{i^*j^*}\right) + \mu_4\left(1-\sum_{i=1}^{N} V_{ij^*}\right). \tag{B.25}$$

By identifying the coefficients in (B.25) with the corresponding coefficients in (B.1), it is possible to obtain the interconnection matrix and bias current vector as:

$$T_{ij,kl} = -\mu_1 \frac{c_{ij}c_{il}\left(\Delta f\right)^2 \omega_i}{\left(\overline{m_i}\right)^2}\delta_{ik} + \mu_3\delta_{ik}\delta_{jl} - \mu_4\delta_{jl}, \tag{B.26}$$

$$I_{i,j} = \mu_1 \frac{c_{i,j}\Delta f\, \omega_i}{\overline{m_i}} - \frac{\mu_2}{2}\psi_{i,j} - \frac{\mu_3}{2} + \mu_4, \tag{B.27}$$

where function δ_{pq} is 1 if $p = q$ and 0 otherwise.

Then, the interconnection matrix is the following $\left(N \cdot S\right) \times \left(N \cdot S\right)$-block partitioned matrix

$$T = \begin{pmatrix} A_1 & I_\mu & \cdots & I_\mu \\ I_\mu & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & I_\mu \\ I_\mu & \cdots & I_\mu & A_N \end{pmatrix}, \tag{B.28}$$

where O, I_μ and A_i;i=1,…,N are the following $S \times S$ -matrices

$$O = \begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 0 \end{pmatrix}, I_\mu = \begin{pmatrix} -\mu_4 & 0 & \cdots & \cdots & 0 \\ 0 & -\mu_4 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & -\mu_4 \end{pmatrix}, \tag{B.29}$$

$$A_i = \begin{pmatrix} -\beta_{11}^i \mu_1 + \mu_3 - \mu_4 & -\beta_{12}^i \mu_1 & \cdots & \cdots & -\beta_{1S}^i \mu_1 \\ -\beta_{21}^i \mu_1 & -\beta_{22}^i \mu_1 + \mu_3 - \mu_4 & -\beta_{23}^i \mu_1 & \ddots & -\beta_{2S}^i \mu_1 \\ \vdots & -\beta_{32}^i \mu_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -\beta_{(S-1)S}^i \mu_1 \\ -\beta_{S1}^i \mu_1 & \cdots & \cdots & -\beta_{S(N-1)}^i \mu_1 & -\beta_{SS}^i \mu_1 + \mu_3 - \mu_4 \end{pmatrix}. \tag{B.30}$$

Where, for brevety, we have denoted $\beta_{jl}^i = \dfrac{c_{ij} c_{il} \left(\Delta f\right)^2 \omega_i}{\left(\overline{m_i}\right)^2}$.

In order to have minimum points of the energy function with respect to output voltages V_{ij} of neurons, it is necessary that the second derivatives be positive, or equivalently:

$$\frac{\partial^2 E}{\partial V_{ij}^2} > 0 \Leftrightarrow \mu_1 \frac{c_{ij} c_{iq} \left(\Delta f\right)^2}{\left(\overline{m_i}\right)^2} - \mu_3 + \mu_4 > 0. \tag{B.31}$$

Condition (B.31) is satisfied if we ensure always that $-\mu_3 + \mu_4 > 0$, which yields the following relationship between the parameters of the energy function: $\mu_3 < \mu_4$

References

[1] Carg V.K.; Wilkes, J.E. *Principles and Applications of GSM*, Prentice Hall 1999 .

[2] *Radio resource management strategies 3GPP* TR 25.922 v5.0.0 March 2002..

[3] Sallent, O.; Valenzuela, J.L.; Agusti, R. *Principios de Comunicaciones Móviles*, Edicions UPC, Septiembre 2003.

[4] www.3gpp.org

[5] Pérez-Romero, J.; Sallent,O.; Agustí,R. ; Díaz-Guerra, M., *Radio. Resource Management Strategies in UMTS*, John *Wiley* & Sons,. 2005

[6] H. Holma, A. Toskala (editors), WCDMA for UMTS, John Wiley and Sons, 2000.

[7] O. Sallent, J. Pérez-Romero, F. Casadevall, R. Agustí, *An Emulator Framework for a New Radio Resource Management for QoS guaranteed Services in WCDMA Systems*, IEEE Journal on Selected Areas in Communications, Vol.19, No. 10, October 2001, pp. 1893-1904.

[8] Slawomir, P. *OFDMA for Broadband Wireless Access.* Artech House 2006

[9] Bernardo, F., Agustí, R., Pérez-Romero, J., Sallent, O. "Dynamic Spectrum Assignment in Multicell OFDMA Networks enabling a Secondary Spectrum Management" Special Issue on Cognitive Radio and Advanced Spectrum Management, Wireless communications and Mobile Computing Jounal, 2009.

[10] Haykin, S. *Neural Networks. A comprehensive foundations*. Second Edition Prentice Hall. New Jersey 1999 ISBN 0-13-273350-1

[11] Abe, S.; *Neural Networks and Fuzzy Systems. Theory and Applications* Kluwer academic Publishers USA 1997

[12] http://cse.stanford.edu/class/sophomore-college/projects-00/neural-networks/Applications/index.html (February 2009)

[13] Warren , W.; Walter,W. *A Logical Calculus of Ideas Immanent in Nervous Activity*, Bulletin of Mathematical Biophysics 5:115-133. 1943

[14] Rochester, N., Holland, J.H., Haibt L.H. and Duda, W.L. *Tests on a Cell Assembly Theory of the Action of the Brain Using a Large Digital Computer*, IRE Transaction of Information Theory IT-2 pp: 80-93, 1956

[15] Taylor, W.K. *Electrical Simulation of Some Nervous System. Functional Activities Information Theory*, C Cherry (ed.), 314–328, 1956

[16] Steinbuch, K. . *Die Lernmatrix*, Kybernetik, 1, 1, 36-45, 1961

[17] Widrow, B; Hoff, M; *Adaptive Switching Circuits*, IRE WESCON Convention Record, Part 4, August 1960.

[18] Rosenblatt, F. *The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain* Psychological Review vol. 65, núm. 6, págs. 386-408, 1958

[19] Hopfield, J.J. *Neural Networks and Physical Systems with Emergent Collective Computational Abilities*, Proceedings of the National Academy of Sciences, USA vol. 79 pp:2554-2558, 1982

[20] Hinton, G.; Sejnowski, T. *Optimal perceptual inference* Proc. IEEE Conf. on computer Vision an Pattern Recognition. Pp. 448-453 Washington, 1983

[21] Carpenter, G.A; Grossberg, S.. *Dynamic models of neural systems: Propagated signals, photoreceptor transduction, and circadian rhythms.* J.P.E. Hodgson (Ed.), Oscillations in Mathematical Biology, Springer Series: Lecture Notes in Biology, **51**, 102-196, 1983

[22] Kohonen, T. *Self-Organization and Associative Memory*. Springer-Verlag, Berlin-Heidelberg-New York-Tokyo, 1989.

[23] Fukushima, K.: *Neocognitron: A hierarchical neural network capable of visual pattern recognition*, Neural Networks, **1**[2], pp. 119-130, 1988.

[24] Broomhead, D. S. ; Lowe, D. *Multivariate functional interpolation and adaptive networks* Complex Systems 2, pp. 321 - 355, 1988

[25] http://cse.stanford.edu/classes/sophomore-college/projects-00/neural-networks/Future/

[26] Cohen, M.A.; Grossberg., S. *Absolute stability of global pattern formation and parallel memory storage by competitive neural networks.* IEEE Transactions on Systems, Man, and Cybernetics, 13:815-826, 1983.

[27] Hopfield, J.J.; Tank, D.W. *Simple "Neural" optimization networks: an A/D converter, signal decision circuit, and a linear programming circuit* IEEE Transactions on Circuits and systems. Vol 33, No 5, pp. 533-541, May 1986

[28] Arnold, V. I.; Avez, A. *Ergodic problems of classical mechanics.* Mathematical Physics Monograph Series. A. S. Wightman, Editor, New York 1968

[29] Lefschetz, S.; *Differential equations: geometric theory.* Pure and Applied Mathematics. Vol. VI Interscience Publishers, Inc., New York; Interscience Publishers Ltd., London 1957

[30] Hopfield, J.J. *Neurons with graded response have collective computational properties like those of two-state neuron*, Proceedings of the National Academy of Sciences, USA vol. 81 pp:3088-3092, 1984

[31] Wakamura, M.; Maeda, Y. *FPGA Implementation of Hopfield Neural Network Via Simultaneous Perturbation Rule*, SICE 2003 Annual Conference, Fuki University, Japan 2003

[32] Eberhart, S.P.; Daud, D., Kerns, D. A.; Brown, T. X.; Thakoor, A. P. *Competitive neural architecture for hardware solution to the assignment problem* Neural networks, 4 pp. 431-442, 1991

[33] S.A.Jafar, A. Goldsmith, *Adaptive Multirate CDMA for Uplink Throughput Maximization*, IEEE Transactions on Wireless Communications, Vol. 2, No. 2, March, 2003, pp. 218-228

[34] Seong-Jun Oh, Danlu Zhang, and Kimberly M. Wasserman, *Optimal Resource Allocation in Multiservice CDMA Networks*, IEEE Transactions on Wireless Communications, Vol. 2, No. 4, July 2003

[35] E. Lim, S.Kim *Transmission Rate Scheduling with fairness constraints in Downlink of CDMA Data Networks*, IEEE Transactions on Vehicular Technology, January, 2005

[36] Chang Wook Ahn, and R.S.Ramakrishna, *QoS Provisioning Dynamic Connection-Admission Control for Multimedia Wireless Networks Using a Hopfield Neural Network*, Trans. on Veh. Tech., VOL. 53, NO. 1, Jan 2004

[37] V.Bharghavan, S.Lu, T.Nandagopal, *Fair queueing in wireless networks*, IEEE personal Communications, February 1999

[38] Y.Cao, V.O.K. Li *Scheduling Algorithms in Broadband Wireless Networks*, Proceedings of IEEE, January 2001

[39] I.Akyldiz, A. Levine, I.Joe, *A slotted CDMA Protocol with BER Scheduling for Wireless Multimedia Networks*, IEEE Trans on Networking April 1999

[40] A.C.Kam, T Minn, K_Y S. Siu, *Supporting rate guarantees and fair access for bursty data traffic in W-CDMA* IEEE Journal on Selected Areas in Communications, Vol. 19, No. 11, November 2001

[41] J. Perez-Romero, O. Sallent, R. Agusti *Downlink Packet Scheduling for a Two-Layered Streaming Video Service in UMTS* IST Mobile & Wireless Telecommunications Summit,16-19 June 2002, Thessaloniki, Greece

[42] L. Xu, X. Shen, and Jon W. Mark, Fellow *Dynamic Fair Scheduling With QoS Constraints in Multimedia Wideband CDMA Cellular Networks*, Transactions on Wireless Comm, Jan 2004.

[43] P-Y Kong, K.C Chua, B.Bensau *A Novel Scheduling Scheme to hare Dropping Ratio Guaranteeing a Delay Bound in a MultiCode CDMA network,* IEEE trans on Networking, Dec 2003

[44] UMTS 30.03 v3.2.0 TR 101 112 *Selection procedures for the choice of radio transmission technologies of the UMTS*, ETSI, April, 1998.

[45] 3GPP TR 25.881 v5.0.0 *Improvement of RRM across RNS and RNS/BSS*

[46] A. Tölli, P. Hakalin, H. Holma, *Performance Evaluation of Common Radio Resource Management (CRRM),* ICC Conference, Vol. 5, April, 2002, pp. 3429-3433.

[47] T. Halonen, J. Romero, J. Melero, *GSM,GPRS and EDGE Performance*, John Wiley & Sons, 2002

[48] 3GPP, R1-050779, Texas Instruments, Throughput Evaluations in EUTRA OFDMA Downlink, 2005.

[49] Wong,C. Y.; Cheng, R. S.; Ben Letaief K.; Murch, R. D., *Multiuser OFDM with adaptive subcarrier, bit, and power allocation* , IEEE J. Sel.Areas Commun., vol. 17, pp. 1747-1758, Oct. 1999

[50] Li, G.; Liu, H. *Dowlink RadioResources Allocation for Multicell OFDMA Systems*, IEEE Trans on Comm. December 2006

[51] Kittipiyakul, S.; Javidi, T. *Resource Allocation in OFDMA with Time-Varying Channel and Bursty Arrivals* IEEE Communications Letters, Vol. 9, No. 11, September, 2007, pp. 708-710.

[52] Gross, J.; Bohge, M., *Dynamic Mechanisms in OFDM Wireless Systems: A Survey on Mathematical and System Engineering Contributions*. May 2006, TKN Technical Report Series:TKN-06-001, Editor: A. Wolisz, http://www.tkn.tu-berlin.de/publications/papers/TKN_Report_ 06_001.pdf

[53] Kim, K.; Kang, H.; Providing Quality of Service in Adaptive Resource Allocation for OFDMA Systems, IEEE VTC conference spring, May, 2004.

[54] Pandharipande, A.; Kounrouris', M.; Yang, H.; Park, H. *Subcarrier allocation schemes for multiuser OFDM systems* International Conference on Signal Processing Communications (SPCOM), 2004.

[55] Shen, Z.; Andrews, J. G.; Evans, B. L. *Optimal Power Allocation in Multiuser OFDM Systems*, in Proceedings of the IEEE Global Communications Conference, San Francisco, Vol. 1, pp. 337-341, December 2003

[56] Jang, J.; Lee, K.B. Transmit Power Adaptation for multiuser OFDM systems, IEEE JSAC Feb. 2003

[57] Biglieri, E.; Proakis, J.; Shamai, S. *Fading channels: information theoretic and communications aspects* IEEE Trans. Inf. Theory, vol.44, pp. 2619-2692, Oct. 1998

[58] Khattab, A.K.; Elsayed, K.M.F. *Opportunistic Subcarrier Management for Delay Sensitive Traffic in OFDMA-based Wireless Multimedia Networks*, 2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks, Buffalo-NY, 26-29 June 2006

[59] Song, G.; Cimnini, L.J.; Zheng, H. Joint *Channel-Aware and Queue-Aware Data Scheduling in Multiple Shared Wireless Channels*, IEEE WCNC Conference, 2004

[60] Gross,J.; Klaue, J.; Karl, Wolisz, H. A. *Subcarrier allocation for variable bit rate video streams in wireless OFDM systems*, IEEE VTC Fall conference, Orlando, October, 2003.

[61] Zhang, Y.; Letaief, K. Adaptive resource allocation and scheduling for multiuser packet-based OFDMA networks, IEEE ICC conference, June, 2004.

[62] Chung,S. T.; Goldsmith,A. J. *Degrees of Freedom in Adaptive Modulation: A Unified View*, IEEE Tran. on Comm, VOL. 49, NO. 9, September 2001

[63] Aiyer, S.V.; Niranjan, M.; Fallside, *A Theoretical Investigation into the Performance of the Hopfield Model* IEEE Transactions on Nueral Networks, Vol. 1, No. 2, June 1990

[64] Abe, S. *Global Convergence and Suppression of Spurious States of the Hopfield Neural Networks*. IEEE Transactions on Circuits and Systems I. Vol 40. N. 4 April 1993

[65] Abe, S.; Kawakami, J. *Theories on the Hopfield neural networks with inequality constraints* Proc. IJCNN-WASH-DC-90, vol. 1 pp349-352, Jan, 1990

[66] Abe, S.J. *Theories on the Hopfield neural networks.* Proc. IJCNN Vol. 1 pp. 557-564 1989

[67] Oohori, T.; Yamamoto, H.; Setsu, N.; Watanabe, K. *A solution to combinatorial optimization problems using an accelerated Hopfield Neural Networks.* Electrical engineering in Japan. Vol 115, Issue 3, pp. 76-84 Mar 2007

[68] Lambert, J.D. *Computational Methods in Ordinary Differential Equations*. John Wiley, 1973 London

[69] 3GPP, TR 25.814 v7.1.0, Physical layer *aspects for evolved Universal Terrestrial Radio Access (UTRA),* Release 7, Sep. 2006.