UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Departament de Teoria
del Senyal i Comunicacions

Centre
Tecnològic
de Telecomunicacions
de Catalunya

Ph.D. Thesis

Exploiting cognition for energy and spectrum efficient wireless networks

Author: Agapi Mesodiakaki

Advisors: Luis Alonso, Ph. D.
Associate Professor
Universitat Politècnica de Catalunya (UPC)

Christos Verikoukis, Ph. D.
Senior Researcher
Telecommunications Technological Center
of Catalonia (CTTC)

Department of Signal Theory and Communications
Universitat Politècnica de Catalunya

Barcelona, September 2015

Abstract

Nowadays, there is an increased need for anywhere anytime connectivity, which is expected to increase significantly during the next few years. In order to meet the consequent high data traffic demands, dramatic expansion of network infrastructures as well as fast escalation of energy demands are expected. As a result, it becomes urgent for mobile operators not only to maintain sustainable capacity growth to meet these new demands, but also to limit their electric bill. In parallel, the fact that the spectrum resources are limited has led to another important problem, known as spectrum scarcity, which stresses the need for spectral efficient solutions. The aforementioned goals can be summarized into the joint maximization of energy and spectrum efficiency, which constitutes a fundamental design objective for next generation networks.

To that end, exploiting cognition is expected to play a key role. In general, a *cognitive* network is able to sense its environment and dynamically adapt to it. In particular, cognitive networks, which could be alternatively characterized as context-aware or self-organizing networks (SONs), have the ability to perceive current network conditions, plan, decide, act based on those conditions, learn from the consequences of their actions, all while following end-to-end goals. This loop, the cognition loop, senses the environment, plans actions according to input from sensors and network policies, decides which scenario fits best its end-to-end purpose using a reasoning engine, and finally acts on the chosen scenario. The system learns from the past (situations, plans, decisions, actions) and uses this knowledge to improve the decisions in the future.

Thus, the main objective of this thesis is to propose and evaluate medium access layer algorithms, that will exploit different types of cognition to provide energy and spectrum efficiency enhancement. In particular, two main research directions are followed: i) the first focuses on spectrum-awareness in cognitive radio (CR) networks inspired by the pioneering work conducted by Mitola in 1999, and ii) the second on the context-aware self-adaptation of cellular heterogeneous networks (HetNets) (i.e., SONs). In both contexts, the exploitation of the available information plays a key role on providing sustainable wireless networks. However, given the different needs and features that characterize each of these networks, they are studied separately.

Hence, the first part of this thesis focuses on the coexistence of secondary networks (SNs) that share the same primary user (PU) resources. A coexistence scheme

is proposed as well as a novel energy-efficient contention-aware channel selection algorithm, which: i) exploits cooperative spectrum sensing to detect the free from PU activity licensed channels, ii) for each one of them, it estimates the probability of collision, and iii) selects the less contended (i.e., with the lowest probability of collision) to access first. An analytical model for the throughput and the energy efficiency of the SN under study is provided, which is validated by means of simulation. The proposed channel selection algorithm is shown to outperform its counterparts both in terms of throughput and energy efficiency. The proposed SN coexistence scheme is also shown to achieve throughput and energy efficiency gains, while maintaining fairness among the coexisting SNs.

The second part of the thesis focuses on cognitive HetNets with multi-hop small cell (SC) backhaul (BH) links. The multi-hop BH is a promising solution since i) not all SCs are expected to have a direct connection to the core network and consequently they are likely to forward their traffic to the neighboring SCs to reach it and ii) the expected short length of BH links enables the use of millimeter wave (mmWave) frequencies to provide high capacity BH. In this context, this thesis studies the role of BH aiming to answer to whether or not it could constitute an energy bottleneck for the HetNet. In particular, the BH energy impact is compared to the access network (AN), i.e., the links between the users and their serving cells, under different traffic distribution scenarios and BH technologies.

Moreover, the user association problem is studied aiming at the joint maximization of energy and spectrum efficiency of the network, without compromising the user equipment (UE) quality of service (QoS) requirements. To that end, analytical user association frameworks are provided, which can be used as benchmarks for the performance evaluation of different user association solutions. Given the need for low complexity solutions, we also propose efficient heuristic algorithms which exploit context-aware information (i.e., UE measurements and requirements, the HetNet architecture knowledge and the available spectrum resources of each base station (BS)) to associate the UEs in an energy and spectrum efficient way, while considering both the AN and BH energy consumption. The proposed algorithms as well as the derived optimal solutions are compared with reference approaches for different BH technologies and traffic distribution scenarios. Insights are gained into the energy and spectrum efficiency trade-off. Furthermore, our results indicate that i) mmWave BH is a promising solution to provide low consumption high capacity BH, and ii) that the proposed algorithms can achieve notable performance gains compared to the state-of-the-art, while achieving near-optimal performance.

Resumen

Actualmente, existe una creciente demanda de conectividad en cualquier momento y en cualquier lugar (conocida como anywhere anytime connectivity) que a lo largo de los próximos años se prevé que experimente un significativo aumento. Con el objetivo de hacer frente a la consiguiente alta demanda de tráfico de datos, se anticipa tanto una expansión drástica de la red como un crecimiento del consumo de energía. Como consecuencia, para los operadores de telefonía móvil es urgente no sólo mantener un crecimiento sostenible de la capacidad, sino también limitar su factura eléctrica. En paralelo, el hecho de que los recursos espectrales sean limitados ha conllevado otro problema importante, conocido como escasez espectral, que pone de relieve la necesidad de soluciones eficientes espectralmente. Los objetivos anteriormente mencionados pueden ser resumidos como la necesidad de maximizar conjuntamente la eficiencia espectral y energética, que constituye un objetivo de diseño fundamental para las redes de nueva generación.

Con ese fin, se prevé que la utilización del proceso cognitivo juegue un papel clave. En general, una red cognitiva es capaz de percibir el entorno y adaptarse dinámicamente a él. En concreto, las redes cognitivas, también conocidas como context-aware networks o self-organizing networks (SONs), tienen la capacidad de percibir las condiciones actuales de la red, planificar, decidir, actuar de acuerdo con las condiciones, y aprender del resultado de las acciones, siempre para conseguir los objetivos extremo a extremo. Este ciclo, el ciclo cognitivo, percibe el entorno, planifica las acciones de acuerdo con las señales recibidas de los sensores y las políticas de red, decide qué escenario se adecua mejor a los objetivos extremo a extremo mediante un motor de razonamiento, y finalmente actúa. El sistema aprende del pasado (situaciones, planes, decisiones, acciones) y utiliza este conocimiento para mejorar las decisiones futuras.

Así, el objetivo principal de esta tesis es proponer y evaluar algoritmos de la capa de acceso al medio que exploten distintos tipos de cognición para ofrecer mejoras en la eficiencia espectral y eléctrica. En concreto, se han seguido dos direcciones de investigación principales: i) la primera se centra en redes de radio cognitiva (CR) inspiradas en el trabajo pionero desarrollado por Mitola en 1999, y ii) la segunda se dedica al estudio de la adaptación dinámica de las redes celulares heterogéneas en función del contexto (es decir, SONs). En ambos contextos, la utilización de la información disponible juega un papel crucial en la provisión de redes inalámbricas

sostenibles. A pesar de ello, debido a las diferentes necesidades y características de cada una de las redes, se estudian por separado.

La primera parte de la tesis se centra en la coexistencia de redes secundarias (SNs) que comparten los recursos de los usuarios primarios (PUs). Se propone un esquema que permite dicha coexistencia, así como un nuevo algoritmo de selección de canal energéticamente eficiente basado en la detección de congestión que: i) explota el concepto de cooperative sensing para detectar los canales sin actividad de usuarios primarios, ii) para cada uno de ellos, estima la probabilidad de colisión, y finalmente iii) prioriza el acceso a aquellos canales que tiene menor contención (es decir, que tienen la menor probabilidad de colisión). Se desarrolla un modelo analítico tanto para el throughput como para la eficiencia energética de la red secundaria, que posteriormente se valida mediante simulaciones. El algoritmo de selección de canal propuesto mejora el comportamiento del estado del arte en términos de throughput y de eficiencia energética. Asimismo, el esquema de coexistencia propuesto también ofrece mejoras de throughput y de eficiencia energética y, a su vez, mantiene la equidad entre las redes secundarias que coexisten.

La segunda parte de la tesis se centra en las redes heterogéneas cognitivas (cognitive HetNets) con un backhaul (BH) formado por enlaces multi-hop entre small cells (SC). El multi-hop BH es una solución prometedora debido a que i) se prevé que no todas las SCs tengan conexión directa con la core network y, por consiguiente, deberán encaminar el tráfico a través de las SCs vecinas, y ii) la corta distancia de los enlaces del BH permitirán el uso de las bandas milimétricas (mmWave) para ofrecer un BH de alta capacidad. En este contexto, la tesis tiene el objetivo de responder si el BH puede suponer un cuello de botella para las HetNets. Concretamente, se compara el impacto del BH respecto a la red de acceso (los enlaces entre los usuarios y sus serving cells) en términos de energía para distintas distribuciones de usuarios y tecnologías de BH.

Además, se estudia la user association como un problema de maximización conjunta de la eficiencia espectral y energética de la red, siempre sin comprometer los requisitos de calidad de servicio (QoS) de los usuarios. Con ese fin, se desarrolla un marco analítico para el problema de user association, que supone un benchmark para el análisis del comportamiento de distintas soluciones. Debido a la necesidad de soluciones de baja complejidad, la tesis también propone algoritmos heurísticos eficientes que utilizan la información de contexto disponible (medidas y requerimientos de los de los usuarios, el conocimiento sobre la arquitectura de la HetNet y la disponibilidad de recursos espectrales de cada estación base) para asociar los usuarios de un modo eficiente espectral y energéticamente, siempre teniendo en cuenta el consumo energético del BH y de la red de acceso (AN). Los algoritmos propuestos, así como las soluciones óptimas derivadas, son comparados para distintas distribuciones de usuarios y tecnologías de BH. Ello permite comprender el compromiso existente entre eficiencia espectral y eficiencia energética. Los resultados obtenidos muestran que i) las ondas milimétricas son una solución prometedora para conseguir un BH de bajo consumo de energía y alta capacidad, y ii) que los algoritmos propuestos permiten obtener mejoras notables en comparación con el estado del arte, así como conseguir soluciones cuasi-óptimas.

Στους γονείς μου, Γιάννη και Μαρία,
και στα αδέρφια μου, Χάρη, Στέλλα και Ιωάννα

Acknowledgements

There are so many people I would like to thank for being part of my life these years and for making this an unforgettable experience.

First of all, I would like to thank the Marie Cure project GREENET and all the people involved, for giving me the opportunity to be part of their team. I feel very lucky, since during this program, I came to develop a broad network of contacts (also friends) with whom we shared a lot of nice experiences. I would also like to thank my advisors Dr. Christos Verikoukis and Dr. Luis Alonso for the opportunity they gave me and for their continuous support, motivation and guidance through all the stages of the doctorate program. I feel grateful for the time they devoted to me all these years, and for being always willing to share their knowledge and experience.

Words are not enough to express my gratitude to Dr. Ferran Adelantado for being so kind, supportive and motivating. He is the supervisor I could wish for and I definitely feel lucky for learning from him. His role was determinant both for the completion of this Ph.D. and for making this journey a pleasant and unforgettable experience.

Special thanks go to Dr. Angelos Antonopoulos and Dr. Elli Kartsakli, who were patient with us during our long meetings, and who generously offered their help any time I needed it. Speaking about meetings, I could not forget mentioning Vasilis, Konstantinos and Alexandra, with whom we shared the same anxieties, aspirations and with some even the same offices (Vasili, thank you for breaking the monotony of the working days with your hilarious comments). I wish you all the best and good luck. I would also like to express my appreciation to Melani for her continuous help which definitely made our life easier.

Undoubtedly, "thank you" is not enough to express my gratitude to my friends, the old and the new ones. I will never forget our lunch meetings (with Akis, Kiki, Angelos -again!- and our new member Maria) that were more than a nice break for me, and our endless conversations with my flatmate, Mireia, my favourite Catalan girl, who was always supportive. I have definitely made new good friends during these years and I am grateful for that. Thank you all very much for making my stay in Barcelona a unique experience! I would also like to sincerely thank my old friends Maria (Loumpaki), Maria (Oikonomakou-again!), Ioanna, Christos and Katerina for

standing by my side all the time, being the best friends I could wish for.

Finally, I could not thank enough my family who stood by me all this time, always being supportive and Stamatis for his enormous patience with me!

Thank you, σας ευχαριστώ, gracias, gràcies!

Agapi Mesodiakaki
Barcelona, Spain
September, 2015

Contents

List of Tables

List of Figures

Acronyms

3GPP	3rd Generation Partnership Project
ABS	Almost Blank Subframes
ACK	Acknowledgment
AMC	Adaptive Modulation and Coding
AN	Access Network
BH	Backhaul
BS	Base Station
CAGR	Compound Annual Growth Rate
CAPEX	CAPital EXpenditure
CCC	Common Control Channel
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CFD	Cyclostationary Feature Detection
CH	Cluster Head
CR	Cognitive Radio
CSMA/CA	Carrier Sensing Multiple Access with Collision Avoidance
CTS	Clear to Send
CW	Contention Window
DB	Database
DCF	Distributed Coordination Function
DIFS	DCF Inter Frame Space
DL	Downlink
ED	Energy Detection
EDGE	Enhanced Data rates for GSM Evolution
e-ICIC	Enhanced Inter-Cell Interference Coordination
eNB	eNodeB

EPC	Evolved Packet Core
E-UTRA	Evolved Universal Terrestrial Radio Access
FC	Fusion Center
FD	Feature Detection
GSM	Global System for Mobile communications
GW	Gateway
HeNB	Home eNodeB
HetNet	Heterogeneous Network
IA	Interference-aware
ICIC	Inter-Cell Interference Coordination
IEEE	Institute of Electrical and Electronics Engineers
IIC	Intercell Interference Coordination
ISM	Industrial, Scientific and Medical
LC	Licensed Channel
LOS	Line-Of-Sight
LS	Local Sensing
LTE	Long Term Evolution
LTE-A	Long Term Evolution Advanced
MAC	Medium Access Control Layer
MCS	Modulation and Coding Schemes
MIMO	Multiple-Input Multiple-Output
MME	Mobility Management Entity
MMKP	Multiple-Choice Multidimensional Knapsack Problem
mmWave	Millimeter Wave
NACK	Negative Acknowledgment
NAV	Network Allocation Vector
NGMN	Next Generation Mobile Networks
NLOS	Non-Line-Of-Sight
OFDM	Orthogonal Frequency Division Multiplexing
OPEX	Operational Expenditure
PA	Proposed Algorithm
PCF	Point Coordination Function
PHY	Physical Layer
PRB	Physical Resource Block

PU	Primary User
QoS	Quality of Service
RAN	Radio Access Network
RF	Radio Frequency
RFD	Random Feature Detection
RFS	Request For Sensing
RRM	Radio Resource Management
RSS	Received Signal Strength
RSSI	Received Signal Strength Indicator
RTS	Request to Send
SC	Small Cell
SIFS	Short Inter Frame Space
SINR	Signal-to-Interference-plus-Noise Ratio
SN	Secondary Network
SNR	Signal-to-Noise Ratio
SoA	State-of-the-art
SON	Self-organizing Network
SNR	Signal-to-Noise Ratio
SU	Secondary User
TCO	Total Cost of Ownership
TDM	Time Division Multiplexing
TDMA	Time Division Multiple Access
TTI	Transmission Time Interval
TXOP	Transmission Opportunity
UC	Unlicensed Channel
UE	User Equipment
UL	Uplink
Wi-Fi	Wireless-Fidelity
WLAN	Wireless Local Area Network
WSPRT	Wald's Sequential Probability Ratio Test

Chapter 1

Introduction

1.1 Motivation

During the last years, there is an ever-increasing need for anywhere anytime connectivity, which accounts for the vast use of tablets, laptops with cellular connections, and smart phones along with their data-hungry applications. As a result, the mobile data traffic is growing exponentially. In particular, according to Cisco, the overall mobile data traffic is expected to grow to 15.9 Exabytes per month by 2018, which constitutes a 11-fold increase over 2013 [1]. Thus, network capacity enhancement has become of utmost importance for next generation wireless networks (Fig. 1.1).

One possible way to achieve capacity growth is by assigning more spectrum to the network. However, the spectrum resources are limited and in their majority inefficiently used [2], which is also known as *spectrum scarcity problem*. Therefore, as depicted in Fig. 1.1, an important goal for next generation networks that takes into account both the spectrum scarcity problem and the need for capacity growth, is to achieve high network spectrum efficiency. To that end, cognitive radio (CR) was proposed, which enables the use of licensed bands by unlicensed users (also called secondary users (SUs)) for as long as they remain unused [3–5], thereby achieving high spectrum efficiency.

Nevertheless, the most important way historically to achieve capacity growth is to add more base stations (BSs) to the network. When BSs are added, i) the distance between users and BSs is reduced and thus the signal-to-interference-plus-noise ratio (SINR) increases, while ii) higher frequency reuse can be employed, which result in iii) the increase in the area spectrum efficiency ($bps/Hz/m^2$) (i.e., for given bandwidth more users can be served in a specific area) and iv) the decrease of the energy consumption in the radio access network (RAN). Therefore, adding more BSs is believed to be the only scalable way to meet the current "capacity crunch" [6]. Yet, in many important markets, adding further eNodeBs (eNBs) is not viable due to cost and the lack of available sites and proper connections among

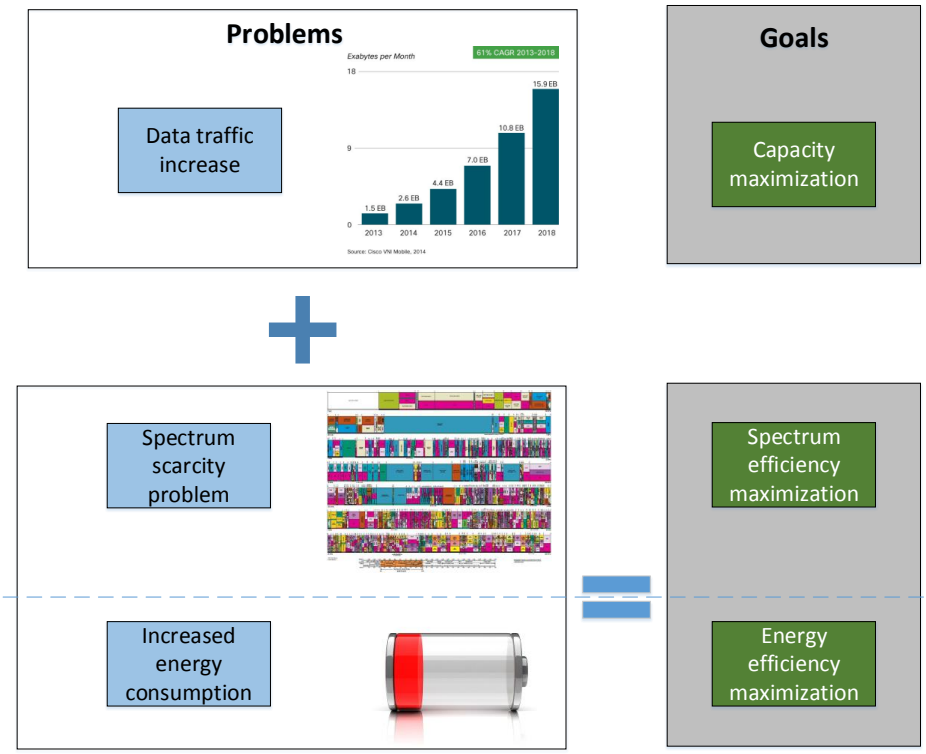


Figure 1.1: Goals of future wireless networks.

them; for example many cities or neighborhood associations are simply not very cooperative about opening up new tower locations. Consequently, future cellular networks are expected to be increasingly organic deployments of BSs[1] (also called small cells (SCs)) of widely varying transmit powers and hence coverage areas, carrier frequencies, backhaul connection types, and communication protocols. A typical smart phone will be capable of communicating via multiple bands over various radio access technologies (RATs), including GSM/EDGE, CDMA, LTE, Wi-Fi, and perhaps others, and will make a network choice based on its needs (e.g., high speed data or voice, or high mobility).

Last but not least, these high capacity future demands are directly connected to increased energy consumption. Hence, it becomes urgent for mobile operators to maintain sustainable capacity growth to meet these new demands and at the same time, to limit the electric bill (i.e., energy consumption). This goal, as depicted in Fig. 1.1, can be summarized into the maximization of the network energy efficiency [7]. Energy efficiency, which is mainly measured in bits/Joule, refers to the total number of successfully transmitted bits divided by the total energy consumption needed. To that end, Next Generation Mobile Networks (NGMN) vision for 5G stresses the need for sustainable 5G services, with network energy efficiency

[1]From now on, in this thesis, the term BS will be used to refer to an eNB and/or a SC.

being a key factor to minimize the total cost of ownership (TCO), as well as the environmental footprint of networks [8]. Specifically, NGMN outlines that 5G should support a 1,000 times traffic increase in the next 10 years (or less) time-frame, with half of the typical consumed energy, compared to today, over the whole network.

Hence, the aforementioned solutions for capacity and spectrum efficiency improvement should also involve low energy consumption. To this direction, cognition is expected to play a key role. In general, a *cognitive* network is able to sense its environment and to plan, decide and act according to it. Cognitive networks could be alternatively characterized as context-aware [9] or self-organizing networks (SONs). In this thesis, these terms will be used interchangeably. Based on the type of acquired information (i.e., context), these networks can be further characterized. For instance, CR networks acquire "spectrum-awareness" through spectrum sensing.

To that end, this thesis provides a contribution to the field of MAC layer algorithm design for wireless networks by proposing and evaluating cognitive mechanisms that enhance both network energy efficiency and spectrum efficiency focusing in two different types of networks, CR networks, and heterogeneous networks (HetNets). The structure of the thesis and the main contributions of this work, which has been conducted under the framework of the Marie Curie project GREENET (PITN-GA-2010-264759), will be discussed in detail in the following section.

1.2 Structure of the Thesis and Contributions

As indicated by its title, this thesis aims to answer to the question: "how to exploit cognition in order to provide energy and spectrum efficient wireless networks?" In this context, two main research directions are followed: i) the first one focuses on spectrum-awareness in CR networks inspired by the pioneering work conducted by Mitola in 1999 [3], and ii) the second on the context-aware self-adaptation of cellular HetNets (i.e., SONs). In both contexts, the exploitation of the available information plays a key role on providing sustainable wireless networks. However, given the different needs and features that characterize each of these networks, their separate study is required. Therefore, as shown also in Fig. 1.2, this thesis is divided into two main parts; the first focuses on CR networks, and the second on cognitive HetNets.

In particular, the remaining part of the thesis consists of four chapters. Chapter 2 comprises two subsections that provide some necessary background information concerning CR networks and cognitive HetNets, respectively (see Fig. 1.2). Specifically, in the CR subsection, the main sensing techniques, the cooperative spectrum sensing procedure as well as the most representative related works are discussed. The CR state-of-the-art mainly includes energy-efficient cooperative spectrum sensing approaches and literature works on secondary networks coexistence mechanisms. In the second subsection, which deals with cognitive HetNets, the main user association technical challenges are described and analyzed together with the expected future HetNet architecture. Finally, being motivated by the new challenges im-

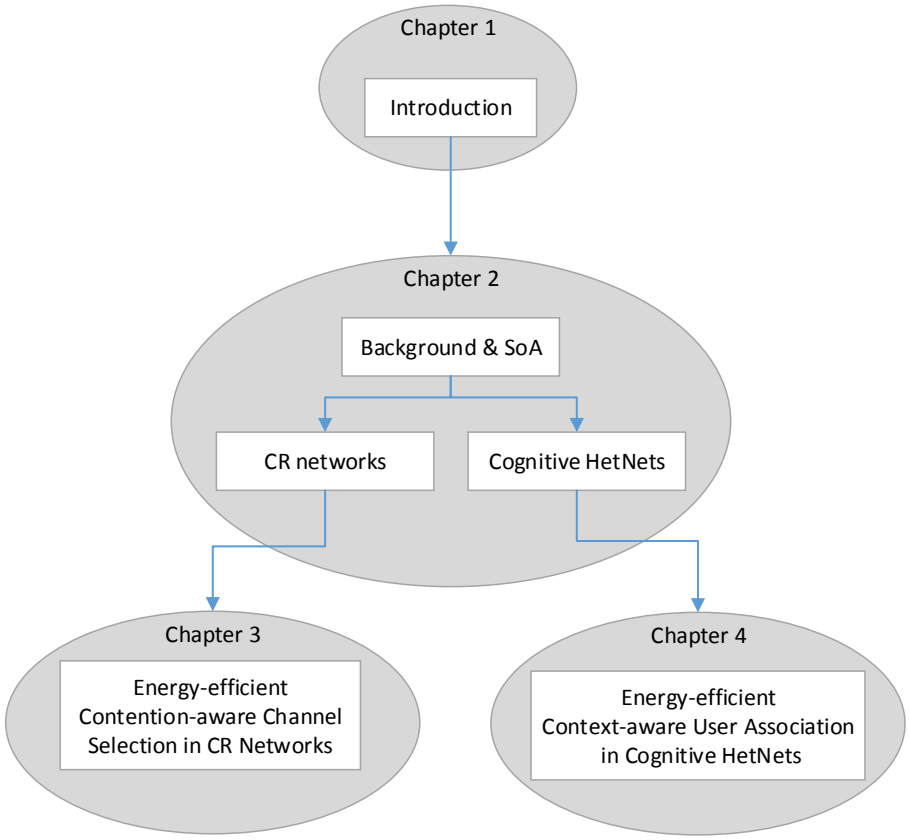


Figure 1.2: Thesis structure.

posed by the backhaul (BH), the related works that have been proposed hitherto on backhaul-aware user association are included.

The innovative contributions of the thesis are organized into two parts. The first part is included in Chapter 3 and is dedicated to a novel energy-efficient contention-aware channel selection algorithm for CR networks. The most important contribution of Chapter 3 is the derivation of a mathematical model for the throughput and energy efficiency of a secondary network (SN), in a scenario where several SNs co-exist and share the same primary user (PU) resources. This model is a solid tool for the evaluation of the proposed energy-efficient contention-aware channel selection algorithm. The second part of the thesis is formed by Chapter 4 and investigates the user association problem in cognitive HetNets. The most important contribution of Chapter 4 is the derivation of upper and lower bounds on the network performance, which can be used as benchmarks for the performance evaluation of different user association solutions. Finally, Chapter 5 discusses the conclusions of the presented work and identifies potential lines for future investigation. In continuation, the main contributions of the thesis will be outlined in more detail.

1.3 Research Contributions

The novel proposals discussed in this thesis have been published in several research contributions. The work presented in the first part of this thesis, concerning the channel selection in cognitive radio networks, has been published in two book chapters, one journal and three international conferences, cited next:

[**BC1**] **A. Mesodiakaki**, F. Adelantado, L. Alonso, and C. Verikoukis, "Chapter 3: Technical Approaches for Improved Spectrum Sharing, Section 3.2.1: A Novel Energy-Efficient Contention-Aware Channel Selection Algorithm for CR Network," *Cognitive Radio Policy and Regulation*, Springer, pp. 124-127, ISBN 978-3-319-04022-6, 2014.

[**BC2**] **A. Mesodiakaki**, F. Adelantado, L. Alonso, and C. Verikoukis, "An Energy-Efficient Contention-Aware Algorithm for Channel Selection in Cognitive Radio Networks," *Resource Management in Mobile Computing Environments, Modeling and Optimization in Science and Technologies*, Springer International Publishing, vol. 3, pp. 419-438, ISBN 978-3-319-06703-2, 2014.

[**J1**] **A. Mesodiakaki**, F. Adelantado, L. Alonso, and C. Verikoukis, "Performance Analysis of a Cognitive Radio Contention-Aware Channel Selection Algorithm," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 5, pp. 1958–1972, May 2015.

[**C1**] **A. Mesodiakaki**, F. Adelantado, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Fairness Evaluation of a Secondary Network Coexistence Scheme," in *Proc. of IEEE 18th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD 2013)*, pp. 180–184, Sep. 2013.

[**C2**] **A. Mesodiakaki**, F. Adelantado, L. Alonso, and C. Verikoukis, "Energy Efficiency Analysis of Secondary Networks in Cognitive Radio Systems," in *Proc. of IEEE International Conference on Communications (ICC 2013)*, pp. 4115–4119, Jun. 2013.

[**C3**] **A. Mesodiakaki**, F. Adelantado, L. Alonso, and C. Verikoukis, "Energy-Efficient Contention-Aware Channel Selection in Cognitive Radio Ad-Hoc Networks," in *Proc. of IEEE 17th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD 2012)*, pp. 46–50, Sep. 2012.

The context-aware user association algorithms, discussed in the second part of this thesis, have been presented in two journals and four international conferences:

[**J2**] **A. Mesodiakaki**, F. Adelantado, L. Alonso, M. D. Renzo, and C. Verikoukis, "Joint Energy and Spectrum Efficient User Association in Millimeter Wave Backhaul Small Cell Networks," *IEEE Transactions on Vehicular Technology*, Jun. 2015 (submitted).

[**J3**] **A. Mesodiakaki**, F. Adelantado, L. Alonso, and C. Verikoukis, "Energy-efficient User Association in Cognitive Heterogeneous Networks," *IEEE Commun. Mag., Energy-Efficient Cognitive Radio Networks*, vol. 52, no. 7, pp. 22–29, Jul. 2014.

[**C4**] **A. Mesodiakaki**, F. Adelantado, L. Alonso, and C. Verikoukis, "Energy and Spectrum Efficient User Association in 5G Heterogeneous Networks," *IEEE International Conference on Communications (ICC 2016)*, Sep. 2015 (to be submitted).

[**C5**] **A. Mesodiakaki**, F. Adelantado, A. Antonopoulos, E. Kartsakli, L. Alonso, and C. Verikoukis, "Energy Impact of Outdoor Small Cell Backhaul in Green Heterogeneous Networks," in *Proc. of IEEE 19th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD 2014)*, pp. 11–15, Dec. 2014.

[**C6**] **A. Mesodiakaki**, F. Adelantado, L. Alonso, and C. Verikoukis, "Joint Uplink and Downlink Cell Selection in Cognitive Small Cell Heterogeneous Networks," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM 2014)*, pp. 2643–2648, Dec. 2014.

[**C7**] **A. Mesodiakaki**, F. Adelantado, L. Alonso, and C. Verikoukis, "Energy-efficient Context-aware User Association for Outdoor Small Cell Heterogeneous Networks," in *Proc. of IEEE International Conference on Communications (ICC 2014)*, pp. 1614–1619, Jun. 2014.

Apart from publications directly related to the thesis contributions, one more research work has been carried out under the framework of Greenet during the elaboration of this thesis. In particular, one conference paper has been produced:

[**C8**] X. Pons, **A. Mesodiakaki**, C. Gruet, L. Navinerb, F. Adelantado, L. Alonso, and C. Verikoukis, "An Energy Efficient Vertical Handover Decision Algorithm," in *Proc. 2nd Workshop on Green Broadband Access, IEEE Global Telecommunications Conference (GLOBECOM 2014)*, Dec. 2014.

Chapter 2

Background and State-of-the-art

2.1 Introduction

As already stated in the previous section, the need for capacity growth has fostered the technical approaches that aim at the joint spectrum efficiency and energy efficiency improvement. To that end, during the last decade, the research works have clearly shown the need to design efficient wireless networks able to adapt to the environment. In this context, the cognition process, and the cognitive or context-aware network behavior, has emerged as one of the cornerstones of future networks. Thus, in this thesis, two of the main research directions to the aforementioned problem are addressed: i) the spectrum-awareness of cognitive radio (CR) networks inspired by Mitola in 1999 [3], and ii) the context-aware self-adaptation of cellular heterogeneous networks (HetNets). Focusing on these two types of networks, the main objective of this thesis is to propose and evaluate medium access layer cognitive algorithms that will enhance both network energy efficiency and spectrum efficiency.

To that end, this chapter begins with an overview of the CR concept and its principles in Section 2.2. Then, in Section 2.2.1, the main sensing techniques are described and analyzed, while in Section 2.2.2, the concept of cooperative spectrum sensing (CSS) and its benefits are presented. In Sections 2.2.3 and 2.2.4, the most representative related works on energy-efficient CSS and on secondary networks (SNs) coexistence are discussed, respectively.

The second part of this chapter deals with cognitive HetNets. In particular, Sections 2.3.1 and 2.3.2, describe and analyze the expected future HetNet architecture as well as the main user association technical challenges. Finally, being motivated by the new challenges imposed by the backhaul (BH), in Section 2.3.3, the related works that have been proposed hitherto on BH-aware user association are also in-

cluded.

2.2 Cognitive Radio Networks

A CR, with its built-in intelligence and cognitive capabilities, can sense the radio spectrum, locate spectrum holes (underutilized or unused portions in the licensed spectrum) and opportunistically access them as long as the licensed users (also called primary users, PUs) do not use the band [3–5].

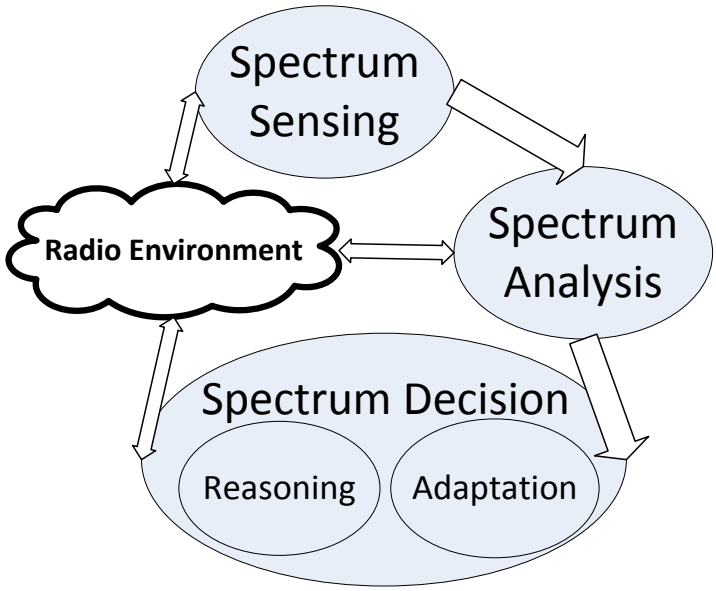


Figure 2.1: Cognitive radio cycle.

A typical duty cycle of a CR, as illustrated in Fig. 2.1, includes detecting spectrum white spaces, selecting the best frequency bands, coordinating spectrum access with other users and vacating the frequency when a PU appears. Through spectrum sensing and analysis, a CR can detect the spectrum white spaces and opportunistically utilize them. When PUs start using the licensed spectrum again, the CRs can detect their activity through sensing and vacate the channels/spectrum, so that no harmful interference is generated due to secondary users' (SUs') transmission.

For the spectrum sensing accuracy, two metrics are traditionally employed, the probability of mis-detection and the probability of false alarm. The first refers to the case where an SU fails to correctly detect that the PU is active on the channel and thus generates unintentional interference to it, whereas the latter to the case where the SU erroneously declares that the spectrum is occupied by a PU, thus decreasing the spectrum opportunities of the SU. Hence, accurate CR spectrum sensing techniques are of high importance to guarantee both PU protection and

high spectrum efficiency.

To that end, there are several spectrum sensing techniques available in the literature, among which the most important are energy detection and feature detection [10, 11].

2.2.1 Spectrum Sensing Techniques

Energy Detection

Energy detection measures the energy received during a finite time interval and compares it to a predetermined threshold [11]. Sensing energy higher than the threshold signifies that a PU is present on the channel. Due to its low computational and implementation complexity, energy detection is the most commonly used technique.

However, it is often accompanied by a number of disadvantages: i) the sensing time taken to achieve a given probability of detection may be high; ii) the detection performance is subject to the noise power uncertainty; iii) energy detection cannot be used to distinguish primary signals from CR user signals. As a result, CR users need to be tightly synchronized and refrained from transmissions during an interval called *quiet period*; iv) energy detection cannot be used to detect spread spectrum signals. Hence, although energy detection is the simplest spectrum sensing technique, its performance in terms of false alarm and mis-detection probabilities may be low. Thus, in order to take advantage of its simplicity, without compromising the accuracy, it has been usually used as the underlying sensing technique of CSS [12]. Thereby, the inaccuracy is overcome by taking advantage of the spatial diversity introduced by the different locations of the SUs.

Feature Detection

On the other hand, feature detection enables the distinction between different types of signals at the expense of higher complexity and longer sensing time. In particular, cyclostationary feature detection (CFD) [11] determines the presence of PU signals by extracting their specific features such as pilot signals, cyclic prefixes, symbol rate, spreading codes or modulation types from its local observation. Therefore, CFD requires prior information about the PU waveforms. However, notice that in CR applications the cyclic frequencies of the PU signals (or at least some of them) are typically known [13], since the waveforms are carefully specified in a standard.

2.2.2 Cooperative Spectrum Sensing (CSS)

In parallel, in order to improve the accuracy of the spectrum sensing technique that is used, CSS is exploited. The CSS benefits from the spatial and multiuser diversity and manages i) to guarantee PU protection by increasing the detection probability (probability that a PU is busy and the channel is sensed as busy) and ii) to utilize the

idle spectrum more efficiently by reducing the false alarm probability (probability that a PU is idle but the channel is sensed as busy) [11, 12, 14]. CSS is usually deployed in the following two successive phases: i) sensing and ii) reporting. In the sensing phase, each CR performs spectrum sensing for a specific amount of time and acquires a decision based on its own observation. In the reporting phase, all local sensing decisions are transmitted to a fusion center (FC), and then, a final decision will be made to indicate the absence (hypothesis H0) or presence (hypothesis H1) of the primary user [15].

- *Sensing phase*: One of the key parameters that strongly affect the sensing performance is the sensing time. A longer sensing time will improve the detection performance. However, if the frame duration is fixed, a longer sensing time will reduce the data transmission time of SUs in the CR network. In [16], the sensing-throughput tradeoff problem has been formulated to find the optimal sensing time that maximizes the SUs' throughput while providing adequate protection to the primary users. In [17], the authors propose a CR system that overcomes the sensing-throughput tradeoff in opportunistic spectrum access CR networks by performing spectrum sensing and data transmission at the same time.
- *Reporting phase*: Typically, the reporting link for each CR uses a control channel to report its sensing result to the FC [18]. As the control channel bandwidth is limited, a summary of the node decisions is reported using one or a few modulated bits [19, 20]. This type of cooperative sensing scheme is called decision fusion [21].
- *Joint sensing and reporting phase optimization*: Recognizing that both the sensing time and cooperative fusion scheme affect the throughput of the SUs, a joint optimization of the sensing time and cooperative fusion scheme has recently been considered in [22]. In [22–24], the k-out-of-N fusion rule has been examined, where all the CRs' reporting links are considered error free. However, in practice, reporting channels are neither ideal nor perfectly known at the FC.

All the aforementioned approaches do not take into account the energy efficiency of the CR network. However, as the CR nodes may be battery-powered wireless devices; their energy-efficiency plays a key role.

2.2.3 Energy-efficient CSS

It is well known that the benefits of CSS come at the cost of control channel overhead and more transmission data, requiring more power consumption and introducing additional transmission delay. To that end, some studies have addressed the problem of power consumption in CSS.

- *Optimal sensing duration*: In [25], the authors study how to choose an optimal sensing duration to strike a balance between energy consumption and system

throughput. They use a comprehensive utility function to formulate the transmission cost in terms of the energy consumption of the sensing process and the transmission process. The maximization of the utility function is obtained with the constraint of providing sufficient protection to PUs.

- *Overhead reduction*: In [26], the authors propose the replacement of observation reports by hard decision reports to reduce the communication overhead. In [27, 28] the authors propose the use of a censorship strategy where only a user that has reliable information can report the sensing result to FC. Another method of network overhead reduction for CSS is to reduce the number of cooperative users, where the performance of sensing can be increased when only a certain number of cognitive radios cooperate (those with the highest signal-to-noise ratio, SNR) [26].
- *Joint optimal sensing duration and overhead reduction*: Accordingly, [29] aims to maximize the energy efficiency of the CR network by optimizing the parameters that affect its average throughput and energy consumption, namely by optimizing the fusion rule threshold, the detector's thresholds at the SUs, the length of the sensing time, and the number of cooperating SUs.

However, all the aforementioned works focus on scenarios where only a single SN is considered. Nevertheless, the scarce transmission opportunities in densely populated areas, in conjunction with the potential high number of SNs in these scenarios, make the SNs' coexistence an ever-hot issue. Hence, as stated in [30], mechanisms for efficient coexistence of more than a single SN are indispensable. The key point of such an efficient coexistence is that the contention of two or more SNs over the same channel is allowed; however, it decisively impacts the achievable throughput and energy efficiency. Therefore, a CR-based medium access control (MAC) layer protocol should i) detect the licensed channels without PU activity and ii) prioritize the access to the channels with low SU contention.

2.2.4 Secondary Networks Coexistence Mechanisms

Although efficient sensing techniques, security or suitable MAC protocols have been extensively addressed by the research community [11], the initial CR technology immaturity and the subsequent lack of real CR applications has hitherto resulted in a slight interest in the coexistence among SNs.

In general, most approaches in the literature decouple the opportunistic spectrum sharing problem into two subproblems: the detection of PUs' activity, and the contention of the SNs. Both the PUs' and the SNs' activity detection can be relied on information provided either by geographical databases or by local sensing procedures. With the use of local sensing for the joint detection of PU and SU activity, however, the coexistence problem is tackled in an holistic manner [31–33].

In parallel, the study of a single licensed channel is not sufficient, especially taking into account the nature of the CR concept. Hence, algorithms designed for multichannel scenarios are needed [33–35].

Nevertheless, the related works proposed hitherto in the literature, either consider single channel models [31, 32], or do not tackle the coexistence problem in a holistic manner (by jointly designing the detection of PU activity and SNs coexistence) [33, 34] or make constraining assumptions (i.e., dedicated common control channel for control information exchange among the SUs, ideal estimation of spectrum availability probability) [33].

To that end, in the first research contribution part of this thesis, the following contributions are provided:

- A novel contention-aware channel selection algorithm is proposed that: i) exploits cooperative spectrum sensing to detect the free from PU activity licensed channels, ii) for each one of them, it estimates the probability of collision, and iii) selects the less contended (i.e., with the lowest probability of collision) to access first.
- An analytical model for the throughput and the energy efficiency of the SN under study is provided, which is validated by means of simulation. Moreover, the impact of the time between two consecutive sensing periods on the aforementioned metrics is studied.
- Finally, the proposed channel selection algorithm is compared with three relevant state-of-the-art algorithms. Simulation results show that the proposed algorithm significantly outperforms its counterparts both in terms of throughput and energy efficiency.

2.3 Cognitive Heterogeneous Networks (HetNets)

The need to exploit cognition in order to provide sustainable networks is not limited to the search of unused spectrum opportunities in ad-hoc networks. It also applies to more sophisticated networks, such as cellular networks. Future cellular networks call for context-aware algorithms, which will exploit their cognition to boost the network efficiency. Thus, the concept of cognitive HetNets is being introduced.

In general, HetNets comprise a conventional cellular network overlaid with a diverse set of lower-power base stations (BSs) such as femtocells, picocells, metrocells and/or microcells, broadly known as small cells (SCs). A SC, according to the Small Cell Forum [36], is a low-power wireless access point that operates in licensed spectrum, is operator-managed and features edge-based intelligence.

SCs are expected to be a key feature of 5G cellular networks as they constitute a viable solution to provide higher end user throughput, as well as expanded indoor and cell edge coverage. Although originally they were envisioned as a means to provide better voice coverage they are now primarily viewed as a cost-effective means of offloading data traffic from the macrocell network. SCs, along with Wi-Fi offloading, are expected to carry over 60% of all global data traffic by 2015 [37]. Moreover, ABI Research forecasts that the outdoor SC units will grow at 53.8%

compound annual growth rate (CAGR) to reach 9.3 million units by 2017 and that the fastest growing SC class is the outdoor femtocells, with output power less than 1 Watt, which will grow at 81.5% CAGR to reach 1.5 million units and $5.4 billion by 2017 [38].

Driven by their attractive features and potential advantages, the development and deployment of SCs have gained tremendous momentum in the wireless industry and research communities in recent years. Small cells have also attracted the attention of standardization bodies, e.g., the 3rd Generation Partnership Project (3GPP), Long Term Evolution (LTE) and LTE-Advanced.

In general, the benefit of adding more BSs is twofold: i) the distance between user equipments (UEs) and BSs is reduced and thus the signal-to-interference-plus-noise ratio (SINR) increases and ii) each UE shares the BS's bandwidth and BH connection with a smaller number of UEs, thus gaining access to a larger portion of resources, which results in extra capacity improvement. However, adding more macrocell BSs (i.e., eNodeBs (eNBs) in the LTE terminology) to the network is not viable in many important markets due to the cost and the lack of available sites and proper connections among them. Consequently, future cellular networks are expected to be dense deployments of BSs of widely varying transmit powers and coverage areas (i.e., eNBs and/or SCs), carrier frequencies, BH connection types and communication protocols. Besides their heterogeneous nature, they will be also characterized by the implementation of cognitive capabilities in some of the involved network entities. In particular, they will be able to perceive current network conditions (i.e., context-awareness), plan, decide and act according to these conditions, the so-called self-organized networks (SONs), learn from the consequences of their actions, while aiming at achieving specific goals. Therefore, in the rest of the thesis, we will refer to these networks as cognitive HetNets.

In cognitive HetNets, UEs will be capable of communicating via multiple bands over various protocols, and thus the user association problem becomes challenging, with future HetNet architecture playing an important role. Thus, in the rest of this section, the future cognitive HetNet architecture will be presented and then the main technical challenges in user association will be discussed and analyzed.

2.3.1 Cognitive HetNet Architecture

LTE-Advanced [39] recently proposed a new HetNet architecture, which is depicted in Fig. 2.2. According to it, a SC can directly connect to core network or through a Home eNodeB (HeNB) gateway (GW). Moreover, the last version of the 3GPP specification supports direct X2-connectivity between HeNBs, independent of whether any of the involved HeNBs is connected to a HeNB GW (Fig. 2.2). Generally, the HeNB GW appears to the mobility management entity (MME) as an eNB and to the HeNB as an MME.

As depicted in Fig. 2.2, SCs will support 3GPP compliant interfaces like S1, X2, Iub, Iuh, etc. However, according to the next generation mobile networks (NGMN) Alliance [40] and Tellabs [41], an optional aggregation gateway can be introduced

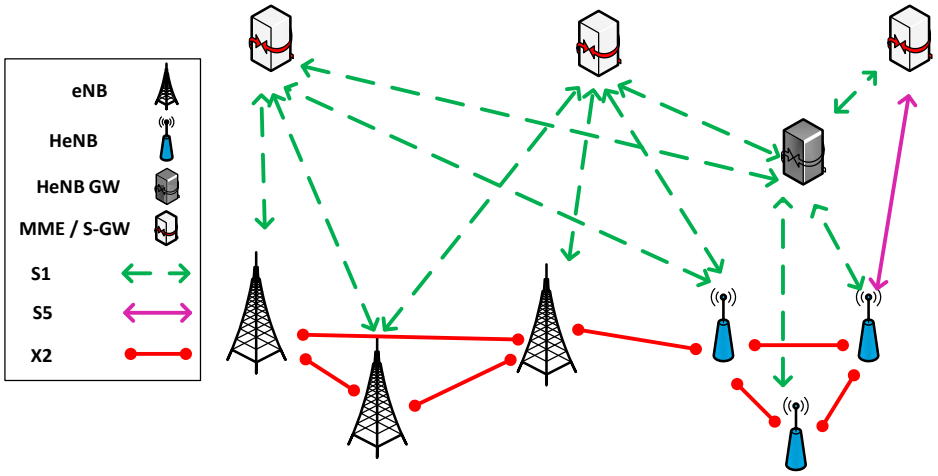


Figure 2.2: Overall E-UTRAN architecture with deployed HeNB GW [39].

to the architecture for better scalability (e.g. reducing the number of logical S1 interfaces to be supported by the evolved packet core, EPC). This optional aggregation gateway can provide functionality on user, control and management plane helping to reduce the signaling load on the core elements (e.g. EPC) as well as to ease operation of SCs. It will be based on the gateway architecture defined by 3GPP for Home NodeB (HNB) and HeNB offering standard interfaces towards the core elements (S1, Iu). Although the Home BS architecture was originally intended to support consumer deployed Home BSs (also known as femtocells), its use is not precluded for operator deployed SC networks.

Depending on the capacity of the aggregation gateway (number of supported SCs) as well as the applied network topology it might be deployed within either the access or aggregation domain. A collocation with a Macro BS (eNB) and supporting from 4 to 12 SC BS (HeNB) might be a reasonable configuration [40].

To that end, assuming that the operator already has a radio network in place, a straightforward option is to connect the SC BS (HeNB) directly to the macrocell site (or any other site offering connectivity to the existing BH network). This option is especially attractive in cases where there is fiber access to the macro site [41]. From topology perspective this would look like a traditional hub-and-spoke, with SCs as spokes and the macro BS site as hub.

Alternatively, e.g. in case of a greenfield deployment or when other transport services are more applicable from cost or availability perspective, the SC BSs can be connected to any other transport network offering suitable BH services. The connectivity can be established either directly or via an aggregation site.

Optionally the hub point offering the connectivity to the cell sites could also be a dedicated aggregation site as shown in the lower half of Fig. 2.3. In this case the node at the aggregation site has the connectivity to the existing BH network.

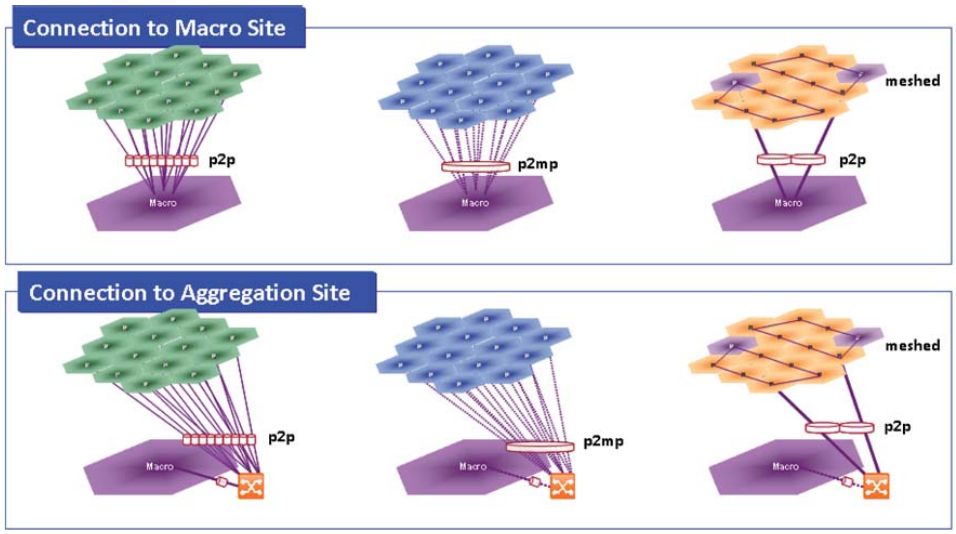


Figure 2.3: Basic connectivity options when using an existing macro cell layer [40].

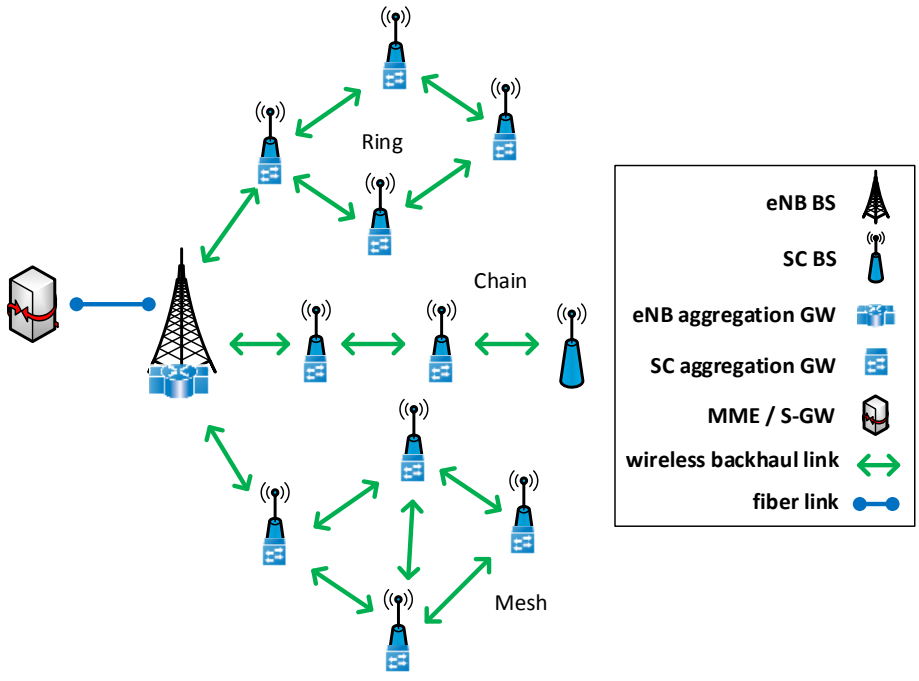


Figure 2.4: Network topologies chart [41].

The connectivity between the SCs and the hub point (being either macro cell or aggregation site) could be based on point-to-point or point-to-multipoint topologies (independently whether wired or wireless connectivity is used). As a further option, instead of connecting every single SC BS to the macro site, chain, tree or mesh topologies can be used between the SC sites themselves for providing further connectivity [41] (Fig. 2.4).

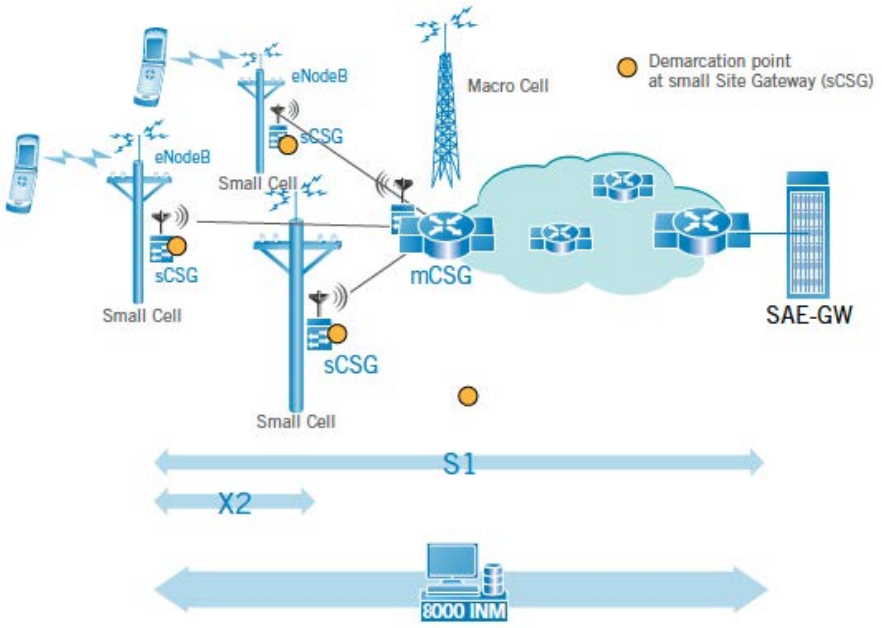


Figure 2.5: Small cell aggregation with demarcation at small cell site [41].

In such topologies (chain, ring and meshed), where one SC acts as an aggregation point for other SC sites, some additional packet aggregation capabilities are required. The connectivity between the macro site and the SC site is no longer a simple point-to-point connection, but requires transit through other SC site GWs (Fig. 2.5).

A topology with alternative traffic paths and the need for some form of forwarding intelligence increases the importance and value of a packet switch at the SC sites. A SC site GW, which is purpose-built for SC sites, can be utilized to provide both the packet aggregation and an extension of the demarcation point (that is, the handover point between the radio and the BH domain, as close to the cell site devices as possible [41]) down to the SC site. This approach also extends all the management capabilities and visibility of the S1 interface down to the SC site. An aggregation device must also address the physical design and cost structure requirements for SC site BH. The SC site GW must provide the same level of functionality as the macrocell site GWs, i.e. quality of service (QoS), connectivity and synchronization capabilities.

The demarcation functionality becomes critical in leased line scenarios where the connectivity between the SC site and the aggregation location is provided by a separate service provider, for example a wholesale provider. The mobile network operator must be able to control the traffic management and QoS functions, as well as, monitor the network and traffic status in its own domain. Therefore, a cell site GW device is required at the SC site. Additionally, connectivity over broadband aggregation networks with varying quality may need to be monitored more closely using a demarcation device at the SC site. A cost and size optimized, feature rich SC site GW will be able to provide the required functionality while strengthening the operator business case.

2.3.2 User Association Technical Challenges

The main objective of user association in future cognitive HetNets is to achieve network capacity enhancement, while satisfying the UE QoS requirements. To that end, three main challenges[1] can be identified: interference management, spectrum efficiency maximization and energy efficiency maximization.

Interference Management

As future HetNets become denser and spectrum resources are limited, high spectrum reuse is necessary to achieve higher spectrum efficiency. Under spectrum reuse, interference management becomes a major concern. User association in future HetNets imposes additional challenges on the interference management, since a UE may communicate with a different BS than the one from which it receives the strongest signal. Thus, in this case, the UE experiences strong downlink interference, while causing high uplink interference to UEs in its proximity. Furthermore, the complex BH architecture should be taken into account to efficiently avoid/mitigate the generated interference.

Several interference management techniques have been proposed so far. Although fractional frequency reuse has been traditionally used, it presents low efficiency, while its complexity increases when applied in dense HetNets consisting of many tiers. Thus, in releases 8/9 of LTE, inter-cell interference coordination (ICIC) was introduced, where the inter-cell interference is controlled by radio resource management (RRM) methods based on the cell spectrum usage and traffic load [39]. In the enhanced ICIC (eICIC), introduced in release 10 of LTE-A, the transmissions of multiple cells are coordinated not only in the frequency domain but also in the time domain using power control schemes. Furthermore, aiming at improving the interference management, eICIC introduced the almost blank subframes (ABS), during which the eNB remains silent (i.e., only transmitting control information at very low power), thus enabling the UE communication with their associated SCs without interference caused by the eNB [39].

[1] Resource allocation fairness is another main HetNet challenge. Yet, it has not been included, since it is a cross-cutting objective that encompasses the three main challenges discussed in this section and all the radio resource management functions, particularly scheduling.

In general, in future cognitive HetNets the user association and the interference management constitute two interdependent problems and therefore, should be jointly studied and designed.

Spectrum Efficiency Maximization

Due to the previously explained spectrum scarcity problem, the spectrum efficiency maximization (i.e., maximization of the achievable data for a given spectrum band) is another challenge mobile operators have to meet. This metric is directly connected to the achieved SINR. Specifically in LTE-A networks, the spectrum efficiency of a UE is a scalar step function of its SINR, with each step corresponding to the use of a specific MCS and thus to a specific achievable rate [42].

Network densification is expected to improve the overall spectrum efficiency, since the distance between UEs and BSs decreases, and thus higher SINR is achieved. In LTE, further spectrum efficiency enhancement is provided by Multiple-Input Multiple Output (MIMO) technology and carrier aggregation. MIMO, which uses multiple antennas at both the transmitter and receiver, offers significant increases in data throughput and link range without additional bandwidth or increased transmit power. On the other hand, through carrier aggregation higher data rates are achieved, since the overall used bandwidth increases.

In future cognitive HetNets, a UE may not be always connected to the BS from which it receives the highest SINR. Hence, spectrum efficiency becomes even more challenging, stressing the need for spectrum-aware user association strategies.

Energy Efficiency Maximization

Maximizing the network energy efficiency may be defined as maximizing the successfully sent data while minimizing the total energy consumption. In case of specific UE requirements, this can be expressed as satisfying the UE traffic demands, while minimizing the total energy consumption.

The total energy consumption is the sum of the energy consumed in the access network, i.e., between the UE and the BS (Uu interface) and in the BH links i.e., between BSs and/or the core network. The BH energy consumption definitely impacts the overall energy efficiency, especially when considering scenarios as the ones previously described with BH links, and thus cannot be neglected.

Given the common assumption that the total transmit power of a BS is equally distributed among its subcarriers [43], the energy consumed in the access network is a function of the number of physical resource block (PRBs) pairs needed to serve the UE traffic. The more the PRBs allocated to the UE, the higher the access network energy consumption to serve the UE traffic. Also note that as eNBs have a much higher total transmit power than SCs, the power allocated to an PRB, and thus the access network energy consumption, is higher when a UE is associated with an eNB than with a SC.

Then, the BH link energy consumption is a scalar function of the aggregated throughput that passes through the link (i.e., the sum of the total throughput of all UEs associated with the SCs that backhaul their traffic through this link). Specifically, given that adaptive MCS is used, the energy consumption in a BH link can be expressed as a function of the aggregated throughput and thus of the required SINR to achieve this throughput. This function depends on the employed BH technology and the distance between the transmitter and the receiver of the BH link. Depending on the UE association, the BH energy consumption may vary significantly. The closer the serving BS to the core network (i.e., the less the number of hops until the data packets reach the core network), the less the BH energy consumption. In addition, further energy efficiency gains can be achieved when BH load balancing is performed, since the energy consumption of a BH link does not increase linearly with its traffic load [42].

Overall, the network energy efficiency is highly dependent on the UE association decision, with BH energy consumption having a significant impact. As a result, user association algorithms taking into account both the access network and BH energy consumption should be designed.

2.3.3 Backhaul-aware User Association Algorithms

In topologies as the ones previously described, where one SC backhauls its traffic to the neighboring SC that acts as an aggregation point, the extension and configuration of the BH involves a multidimensional trade-off between the reduction of the capital expenditure (CAPEX), the reduction of the energy consumption, with the consequent reduction of operational expenditure (OPEX), and the maximization of the access network capacity. It is precisely in this context that, for a BH configuration, which is often determined by external constraints, the user association strategy should exploit context-awareness to deal with the aforementioned complexity. Thus, although backhauling has, thus far, been a largely overlooked issue, future architectures call for BH-aware user association strategies. At the same time, the need to design a framework for the efficient performance evaluation of the existing user associations solutions has ultimately become very important.

However, the related works proposed heitherto in the literature, either consider only the AN [39, 44–50], thus totally overlooking the BH capacity constraints and energy impact, or/and do not take into account the energy consumption and hence, their energy efficiency cannot be guaranteed [43, 51–55].

Thus, the second research contribution part of this thesis focuses on next generation cognitive HetNets. In particular, the user association problem is studied in scenarios where several SCs backhaul their traffic to the neighboring cells until they reach the core network. Analytical frameworks for optimal user association are derived that aim at the joint energy efficiency and spectral efficiency maximization, without compromising the UE QoS. Using the derived frameworks as benchmarks, the performance of existing user association algorithms is evaluated and the high energy efficiency and spectral efficiency potential of a low-complexity heuristic al-

gorithm is shown that exploits the available context-aware information to achieve a good trade-off between energy efficiency and spectral efficiency.

2.4 Conclusions

This chapter has provided some background information that is relevant to the contributions of this thesis, which will be thoroughly presented in the following chapters. At the first part of the chapter, an overview of the CR concept and its principle has been presented and the main sensing techniques have been described and analyzed. The concept of cooperative spectrum sensing and its benefits have been discussed next together with the most representative related works on energy-efficient cooperative spectrum sensing and on SNs' coexistence.

In the second part of this chapter, the expected future HetNet architecture as well as the main user association technical challenges have been described and analyzed. Finally, the related works that have been proposed hitherto on BH-aware user association have been summarized.

Chapter 3

Energy-efficient Contention-aware Channel Selection in Cognitive Radio Networks

3.1 Introduction

As already mentioned, cognitive radio (CR) has received much attention as a possible solution to the spectrum scarcity problem, since it enables the use of licensed channels by unlicensed users (also called secondary users (SUs)) for as long as they remain unused [3–5]. Although efficient sensing techniques, security or suitable MAC protocols have been extensively addressed by the research community [11], the initial CR technology immaturity and the subsequent lack of real CR applications has hitherto resulted in a slight interest in the coexistence among secondary networks (SNs).

The opportunistic spectrum sharing, on which SNs' operation is based, relies upon two main premises: the protection of the primary users' (PUs') transmissions and the maximization of the spectrum usage. The former is achieved by applying effective sensing techniques (cooperative or not) [5,14,29]. Therefore, most proposals on CR networks aim at exploring the radio environment and detect transmission opportunities in licensed channels. The ability to identify such opportunities, and the accuracy with which they are detected, are essential to efficiently exploit them. In this context, proposals on suitable sensing and access mechanisms have been stated [56–58].

The maximization of the spectrum usage, though, can only be met by implementing efficient coexistence mechanisms among SNs, particularly in congested en-

vironments. The new challenges posed by SNs' coexistence are a consequence of the scarce transmission opportunities in densely populated areas, in conjunction with the potential high number of SNs in these scenarios. Hence, as stated in [30], mechanisms for efficient coexistence of more than a single SN are indispensable.

The key point of such an efficient coexistence is that the contention of two or more SNs over the same channel is allowed, but it impacts decisively on the achievable throughput and energy efficiency. Therefore, a CR-based MAC protocol should i) detect the licensed channels without PU activity, and ii) prioritize the access to the channels with low SU contention.

To that end, in this chapter, which includes the first research contribution part of this thesis, the following contributions are provided:

- A SN coexistence scheme is proposed as well as a novel contention-aware channel selection algorithm that: i) exploits cooperative spectrum sensing to detect the free from PU activity licensed channels, ii) for each one of them, it estimates the probability of collision, and iii) selects the less contended (i.e., with the lowest probability of collision) to access first. It is worth noting that this metric can be applied to various traffic patterns.

- An analytical model for the throughput and the energy efficiency of the SN under study is provided, which is validated by means of simulation. Furthermore, the effect of the time between two consecutive sensing periods on the aforementioned metrics is studied and analyzed.

- The proposed channel selection algorithm is compared with three relevant state-of-the-art algorithms. Simulation results show that the proposed algorithm significantly outperforms its counterparts both in terms of throughput and energy efficiency.

- Finally, given the importance of a fair SN coexistence scheme, in the end of this chapter, the proposed SN coexistence scheme is also compared with other state-of-the-art approaches and it is shown that it can achieve throughput and energy efficiency gains, while maintaining fairness among the coexisting SNs.

The rest of the chapter is organized as follows: In Sections 3.2, 3.3 and 3.4, the related work, the system model and the proposed channel selection algorithm are respectively described. In Sections 3.5 and 3.6, the throughput and energy efficiency analysis are presented, respectively. Section 3.7 validates the model accuracy by comparing it with the results obtained by means of simulation and evaluates i) the performance of the proposed channel selection algorithm and ii) the performance of the proposed SN coexistence scheme compared to other relevant state-of-the-art algorithms. Finally, concluding remarks are given in Section 3.8.

3.2 Related work

Most approaches in the literature decouple the opportunistic spectrum sharing problem into two subproblems: the detection of PUs' activity, and the contention of the SNs. Both the PUs' and the SNs' activity detection can be relied on information provided either by geographical databases (DBs) or by local sensing (LS) procedures.

The use of databases to detect the PU and SU activity presents less flexibility, while it requires the deployment of SNs' infrastructure and signaling between SNs and geographical databases [59, 60].

On the other hand, by using local sensing for the joint detection of PU and SU activity, the coexistence problem is tackled in an holistic manner [31–33]. In [31], a set of known SNs accesses the channel in a TDMA fashion. However, the proposed algorithm is designed for a single licensed channel, and the throughput analysis exposes details on the channel access but it does not gain insight in the recovery procedure when PUs resume their activity. Additionally, it requires synchronization between the SNs. In [32], the authors address the coexistence problem between SNs, although the proposal is not designed for a multichannel scenario. Furthermore, SUs require two transceivers to operate (one devoted to data and another to sensing).

Being the closest to this work, [33–35] focus on multichannel scenarios. In [33], the authors propose a MAC protocol for opportunistic spectrum access that uses two channel selection methods, a uniform and a spectrum opportunity-based. According to the first, each SU chooses a channel randomly, whereas the latter takes into account the different spectrum availability probabilities in the channels. However, the authors assume that each SU can correctly estimate the spectrum availability probability (i.e., the number of active secondary flows). Moreover, they consider a dedicated common control channel for control information exchange among the SUs. In [34], the authors propose two algorithms to rank the channels according to their interference severity in terms of strength and activity. Equivalently, in [35], a new carrier sense multiple access (CSMA) protocol that uses a distributed channel selection scheme is proposed, according to which, the transmitter selects an appropriate channel for transmission based on its interference power measurements in the channels. Nevertheless, unlike this work, [34,35] do not tackle the coexistence problem in a holistic manner, by jointly designing the detection of PU activity and SNs coexistence, since the original problem is decoupled into a multichannel access problem without PUs. For the reader's convenience, the aforementioned differences between this work and the state-of-the-art are summarized in Table 3.1.

In this context, another fundamental objective is to guarantee fairness among the coexisting SNs. In general, achieving fairness among the SUs that share the same PU spectrum is a research topic that has received a lot of attention. In [61], a fair opportunistic spectrum access scheme is proposed that, based on a fast catch up strategy, manages to reduce the amount of time after which all SUs have equal access rights to the available licensed channels. In [62], a Homo-Egualis based learning model was proposed to achieve fairness among dissimilar SUs, while in [63] the

Table 3.1: Energy-efficient Contention-aware Channel Selection in Cognitive Radio Networks: Related Work Table

	[59, 60]	[31]	[32]	[33]	[34, 35]	This work
PU activity detection	DB	LS	LS	LS	✗	LS
SNs' activity detection	DB	LS	LS	LS	LS	LS
High flexibility	✗	✓	✓	✓	✓	✓
Need for SNs infrastructure deployment	✓	✗	✗	✗	✗	✗
Need for SN-Database signaling	✓	✗	✗	✗	✗	✗
PU recovery procedure	✗	✗	✓	✓	✗	✓
Number of required transceivers	1	1	2	1	1	1
SNs synchronization need	✗	✓	✗	✗	✗	✗
Dedicated common control channel	✗	✗	✗	✓	✗	✗
Multi-channel design	✗	✗	✗	✓	✓	✓
Spectrum availability calculation	✗	✗	✗	✗	✓	✓

authors proposed heuristic channel allocation algorithms based on multi-channel contention graphs and linear programming aiming at achieving a good trade-off between throughput and fairness, while ensuring interference-free transmissions. In [64], the authors derive the optimal access probabilities for two independent SUs focusing on achieving a good trade-off between spectrum efficiency and fairness. However, all these proposals assume that a licensed channel that is occupied by an SU can not be accessed by another SU. In particular, the licensed channel appears as being busy to the SU and thus it is avoided. Hence, most coexistence schemes in the literature totally overlook the case where several SNs coexist and share the same PU resources.

To overcome the aforementioned problem, in [32], the authors propose FMAC, a MAC protocol, that utilizes a three-state sensing model. Specifically, FMAC uses a spectrum sensing algorithm [65] to distinguish whether a busy channel is occupied by a PU or by an SU and, in the latter case, gives the option to the SU to share the channel with the SUs of other SNs that are currently using it. Nevertheless, in [32] a simple system model consisting only of one licensed channel is considered, while more importantly, the scheme employs a constant back-off window. As a result, unlike the proposed coexistence scheme, it shows low adaptability to any changes in the number of contending SUs in a licensed channel.

3.3 System Model

In the system model under study, M licensed channels are considered that are allocated to PUs and can be opportunistically accessed by SUs, as long as they remain unused. The PU activity is assumed to follow an exponential on-off traffic model, with the mean durations of on and off periods denoted by T_{on} and T_{off}, respectively. While being idle[1] the licensed channels are further characterized by

[1] Although an idle channel implies the absence of any type of activity, in this chapter, a channel will be characterized as busy or idle only based on the PU activity. Thus, an idle channel may still be occupied by SNs.

their SU contention level, i.e., the probability of collision among the SUs that operate on them. In particular, for the licensed channel k, the collision probability among the SUs is denoted by p_{Ck}.

A highly congested unlicensed channel (e.g., belonging to the industrial, scientific and medical band) is also considered, which is operated by users with and without cognitive capabilities[2]. Among the users operating on the unlicensed channel, we will focus on N users with cognitive capabilities that, due to the high contention in the unlicensed channel and given that there is information to be transferred among them, they set up an ad hoc SN to exploit the spectrum opportunities in licensed channels. These users will be referred as SUs, whereas the number of users that operate on the unlicensed channel but do not belong to the SN is denoted by N_{unlic}.

The SN intends to exploit exclusively the idle licensed channels. However, there are two situations where the operation on the unlicensed channel is inevitable: i) during the initial set up of the SN, the exchange of control information is carried out on the unlicensed channel, and ii) when all the available licensed channels are/become busy. Although it will be detailed in Section 3.4, it is worth noting that no dedicated common control channel is used, since the licensed channels are shared for both control and data transmissions, and the unlicensed channel is only used as a common control channel in the two situations stated above.

The SN under study consists of a cluster head[3] [5, 11, 12, 14], whose role may be assigned to the SUs in a round robin way. The sequential assignment of the cluster head role among the SUs improves the algorithm performance in two ways: i) it achieves energy consumption fairness among the SUs [66], and ii) it limits the negative effect of a selfish cluster head, since this is restricted to the time it takes up this role. Furthermore, the SUs of the SN under study are assumed to be adequately close to each other to be exposed to the same channel activity. However, note that their reported sensing results may differ due to false alarm and mis-detection probability.

All SUs that are considered in this system model are equipped with a half duplex transceiver. Thus, even if they are capable of operating over multiple channels, including the licensed channels, they can either transmit or receive over a single channel at any given time. Obviously, the use of a single transceiver is less energy-consuming and costly compared with the use of multiple transceivers and is already considered in some CR devices and prototypes [67].

The SUs' transmissions both in the unlicensed channel and licensed channels use the CSMA/CA access method [68]. To that end, a node wishing to transmit data has to first listen to the channel for a predetermined amount of time (t_{DIFS}) to determine whether or not another node is transmitting. If no other node transmits,

[2]The term *cognitive capabilities* is defined as the set of features that confer on users the ability to tune and sense different channels, and transmit over them if they are detected idle.

[3]It is assumed that the clustering of the SUs is done beforehand. Nevertheless, it is worth mentioning that this takes place in the unlicensed channel, where all users are able to communicate with each other without the presence of PUs.

the node is permitted to begin the transmission process. Otherwise, the node defers its transmission for a random period of time (t_{BO}, back-off time). It is worth noting, however, that the PUs may use their own access method, when accessing the licensed channels (e.g., SC-FDMA in the uplink and OFDMA in the downlink for LTE access).

Although the set of licensed channels sensed by the users of the SN is higher than one, all the users operate on the same single channel by employing CSMA/CA (in fact, several channels are used, but in a sequential manner, since they have to be vacated when they become busy). Thereby, two objectives are achieved: i) the connectivity between all SUs of the SN is guaranteed, and ii) collisions are avoided (or at least minimized) due to the use of CSMA/CA.

3.4 Secondary Network Coexistence Scheme and Algorithm Description

The SN is assumed to be initially located in a highly congested unlicensed channel (shared with other N_{unlic} users). There, the cluster head initiates a sensing procedure aiming at finding new spectrum opportunities in licensed channels for the SN to exploit. Upon the sensing procedure completion, the sensing information is exchanged over the unlicensed channel (Section 3.4.1), and a list containing the licensed channels detected idle is constructed. Then, there are two possible cases:

i) *All the licensed channels have been sensed busy:* The list is empty; the SN stays in the unlicensed channel and another sensing procedure is initiated.

ii) *There is at least one licensed channel sensed idle:* The list is not empty; the SN hops to the first channel of the list and operates there, as described in Section 3.4.2.

The protocol flowchart is depicted in Fig. 3.1 and it is elaborated in the following.

3.4.1 Operation on the unlicensed channel

The operation on the unlicensed channel includes only sensing-related control information exchange. During this procedure the unlicensed channel is used as a common control channel. A sensing procedure can be divided into three periods (t_{ph1}, t_{ph2}, t_{ph3}), as depicted in Fig. 3.2.

Time period t_{ph1}

During this period, in order to limit the experienced delay, only the cluster head contends with the other N_{unlic} users to gain access to the unlicensed channel to

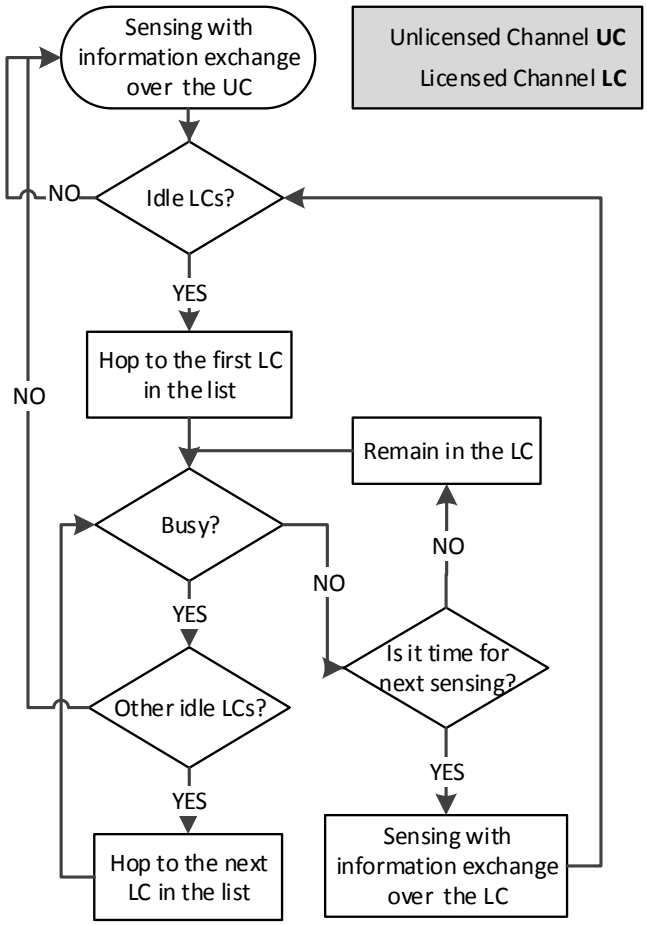


Figure 3.1: Protocol flowchart.

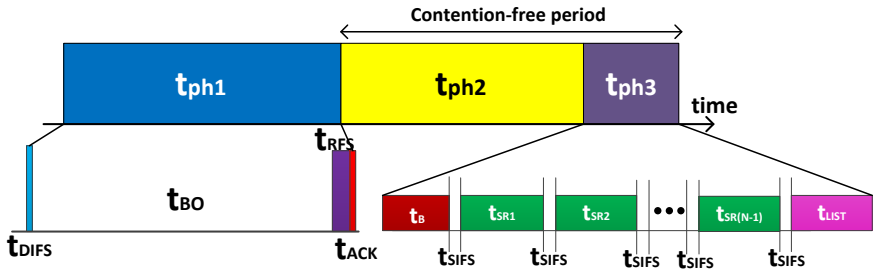


Figure 3.2: Time periods of a sensing procedure.

carry out the whole process (i.e., the rest $N-1$ users of the SN remain idle). Thus, the cluster head first listens to the channel for a predetermined amount of time (t_{DIFS}) to determine whether or not another node is transmitting. If no other node

transmits, it broadcasts a request for sensing (RFS) packet. Otherwise, it defers its transmission for a random period of time (i.e., back-off time (t_{BO})). The RFS packet defines: i) which licensed channels will be sensed by each SU, ii) the order in which the SUs will report their sensing results to the cluster head, and iii) how often the sensing procedure will be triggered. To than end, the following must be noted:

i) All SUs sense the same number of channels in each sensing period, thereby guaranteeing energy consumption fairness. This number is decided during the initial setup, whereas the particular sensing channel assignment is decided and informed by the cluster head with every RFS packet. For this assignment, the cluster head may use information collected in previous periods (while exploiting learning and/or predictive mechanisms) or apply any of the sensing channel assignment algorithms available in the literature [66, 69, 70]. Notice that the use of the round robin algorithm would correspond to the lower performance bound, while any other algorithm, being based on additional information, could improve the SN performance, since the number of channels correctly detected idle would be higher.

ii) Each SU is assigned a network id in the setup process (0 to N-1). Consequently, the cluster head does not need to include in the RFS packet the reporting order of the SUs during t_{ph3} nor the identity of the next cluster head. In particular, before the beginning of every sensing process, the next cluster head id is computed as id_{next}=$(id_{current} + 1)modN$, whereas the reporting order is $(id_{current} + 1)modN$ to $(id_{current} - 1)modN$. For instance, if $N = 4$ and the current cluster head has an id equal to 2, the next cluster head will have $id = 3$, and the reporting order will be {3,0,1}. This strategy does not induce any additional overhead.

iii) As the licensed channels state varies along time, the sensing procedure should be repeated periodically to update the channel information. The parameter T_S is defined as the time elapsed between the completion of a sensing procedure and the triggering of the next (by broadcasting a new RFS packet). This value is tightly coupled with the PU activity, namely for fast changing PU activity, a low T_S should be chosen to keep the information for every channel updated.

To guarantee the successful transmission of RFS, the algorithm in [71] is applied. Accordingly, one node (i.e., the SU scheduled to send its sensing results first) acts as a leader for the purpose of sending feedback to the cluster head. On erroneous RFS reception, the leader does not send an ACK, prompting a retransmission. On erroneous RFS reception at receivers other than the leader (i.e., at the rest $N - 2$ users of the SN), the protocol allows negative ACKs from them to collide with the ACK from the leader, thus prompting the cluster head to retransmit the packet.

Time period t_{ph2}

Upon the RFS successful reception, t_{ph2} begins, with each SU sensing the channels that were assigned to it. In CR networks, as the channels are licensed, it is important to sense a set of channels to have alternatives to hop to in case SUs have to vacate the channel. It is also worth pointing out that when a licensed channel is sensed by more than one SU, cooperative spectrum sensing is applied. In this work, the OR fusion rule[4] is used, which presents low mis-detection and high false alarm probability. In the OR rule, which is the most conservative fusion rule, when at least one of the cooperating SUs senses the licensed channel as busy, the final decision declares a PU is present. Although the application of other fusion rules could achieve better trade-off between false alarm and mis-detection probabilities, the OR rule minimizes the probability of interfering with the PUs, which is the reason why it is selected in this approach.

During sensing (t_{ph2}), cyclostationary feature detection is used [11], which enables the SU that senses the licensed channel to distinguish between PUs' and SUs' signals, at the expense of higher complexity and longer sensing time. Since this technique determines the presence of PU signals by extracting their specific features (e.g., pilot signals, cyclic prefixes), it requires prior information about the PU waveforms. However, notice that this is typically known for most standard technologies that operate on licensed channels [13]. Moreover, in coexisting scenarios its use is fundamental, since a simpler technique, unable to distinguish between PUs' and SUs' signals (i.e., energy detection), would result in very low spectrum efficiency, as all the idle channels being used by other SNs, would be considered busy and thus would be avoided.

Time period t_{ph3}

After the sensing has finished, all SUs of the SN hop back to the unlicensed channel to report their sensing results. Given the importance of exchanging them as soon as possible, the reservation of the unlicensed channel for the constant and known period of t_{ph2}+t_{ph3} is considered, as long as its duration is lower than the maximum tolerable delay[5].

Hence, the cluster head broadcasts a beacon frame (of duration t_B) asking for the sensing results of the rest of the SUs, as depicted in Fig. 3.2. Subsequently, each SU waits t_{SIFS} and sends its sensing results (t_{SR}) to the cluster head in the previously defined order. Thereafter, the cluster head constructs and broadcasts the list (t_{LIST}) and the contention-free period ends.

[4]In cooperative spectrum sensing, the SUs report their sensing results to a central entity (to the cluster head in this case), which process them and makes a final decision according to a predefined rule, also called fusion rule.

[5]This channel reservation is compatible with existing standards, such as the transmission opportunity (TXOP) in 802.11 [72].

3.4.2 Operation on the licensed channels

When the SN under study hops to a licensed channel, operates there using CSMA/CA. Thus, all SUs that have a packet to send (belonging to the SN under study or/and to the other coexisting SNs) contend to gain access to the licensed channel. Hence, the operation time of the SN under study on the licensed channel consists of successful transmission, collision and idle slots[6]. This normal CSMA/CA operation on the licensed channel is interrupted in the following cases:

i) *The PU of the licensed channel remains idle and it is time to initiate the next sensing procedure (i.e., T_S has elapsed).* In this case, only the cluster head contends with the other coexisting SNs to gain access to the licensed channel and trigger a new sensing procedure (consisting again of t_{ph1}, t_{ph2} and t_{ph3}, as described in the unlicensed channel operation).

ii) *The PU becomes busy earlier than T_S.* In this case, the SUs have to leave the licensed channel immediately in order not to interfere with the PU. The time that the SN requires to detect the PU activity and react accordingly by hopping to the next licensed channel in the list is denoted by t_r. In case all the channels of the list have been visited and have become busy before T_S, the SN hops to the unlicensed channel to trigger a new sensing procedure.

3.4.3 Channel selection algorithm

As cyclostationary feature detection is able to discern between PU and SU activity, after the sensing procedure completion, the cluster head constructs a list containing the licensed channels where no PU activity has been detected (idle channels). These channels may have other SNs operating on them and thus may be characterized by the probability of collision among the SUs.

The main goal of the algorithm is to achieve throughput and energy efficiency improvement by reducing the time spent in highly contended licensed channels. Therefore, the channels in the list are sorted in ascending order by the estimation of their probability of collision among the SUs, p_C (i.e., the channel with the lowest p_C takes the first place and, thus, higher priority). Notice that as the licensed channels are classified and accessed based on their activity, the SNs (and so the SUs belonging to them) are distributed among the licensed channels, thereby achieving: load balancing over the channels, connectivity for each particular SN, reduction of the coordination signaling burden, and minimization of the need for a dedicated common control channel.

A SU that senses a licensed channel can efficiently estimate p_C by simply monitoring the channel activity. Specifically, it is able to understand the collisions and the

[6] Please note that although the term *slot* is used, the SUs' access is not slotted since they use CSMA/CA. Yet, we will refer to a slot, as defined in [68], to determine the duration of a successful transmission (successful transmission slot), of a collision (collision slot) or of an idle period (empty slot).

successful transmissions of the other SNs by listening to their packet exchange [73]. Thus, p_C can be measured by counting the number of slots that a successful transmission occurs (C_{succ}), as well as the number of slots that a collision of the other SNs occurs (C_{coll}), as in each of these slots a potential packet transmission of the SN under study would have failed [74]. Thus, p_C may be expressed as

$$p_C = \frac{C_{succ} + C_{coll}}{B} = 1 - \frac{C_{idle}}{C_{succ} + C_{coll} + C_{idle}} \tag{3.1}$$

where B is the total number of observed slots that also includes the number of idle slots (C_{idle}).

The estimation accuracy is highly dependent on the observation time (B) (e.g., for constant SU activity, the longer the observation time, the more accurate the p_C). Therefore, the algorithm should be robust enough to overcome situations of overestimation and/or underestimation of p_C. To that end, in [75] it is shown that the correct construction of the list is slightly impacted by p_C estimation inaccuracy, since it depends more significantly on the comparison between the estimated values and not on the estimated values themselves.

3.5 Throughput Analysis

The throughput of the SN under study may be expressed as

$$S = \frac{\mathbf{E}[D]}{\mathbf{E}[T_U] + \mathbf{E}[T_L]} \tag{3.2}$$

where $\mathbf{E}[D]$ is the expected number of useful bits (i.e., payload) sent by the SN in a representative time period, T_p, defined as the sum of the time spent in the unlicensed (T_U) and licensed channels (T_L), until the SN hops back to the unlicensed. Notice that T_p is a random variable, since it depends on the contention in the unlicensed channel and on the PU activity in the licensed channels (the SN hops back to the unlicensed channel when all the licensed channels previously detected idle become busy). In particular, T_U and T_L may be defined as

- T_U: Time spent by the SN in the unlicensed channel, until there is at least one licensed channel sensed idle. Then, the SN hops to the first channel on the list and T_L begins.
- T_L: Time spent by the SN in licensed channels, until the moment that there is no licensed channel available and the SN has to hop again to the unlicensed channel for the recovery process. T_L may consist of a number of complete periods (T_{CP}) and an incomplete period (T_{INP}). During a complete period (T_{CP}), the SN operates on licensed channels for T_S and a sensing procedure takes place, while in an incomplete period (T_{INP}), the SN operates on licensed channels for less than T_S and there is no other available licensed channel in the list (i.e., all the licensed channels become busy before T_S). After an incomplete

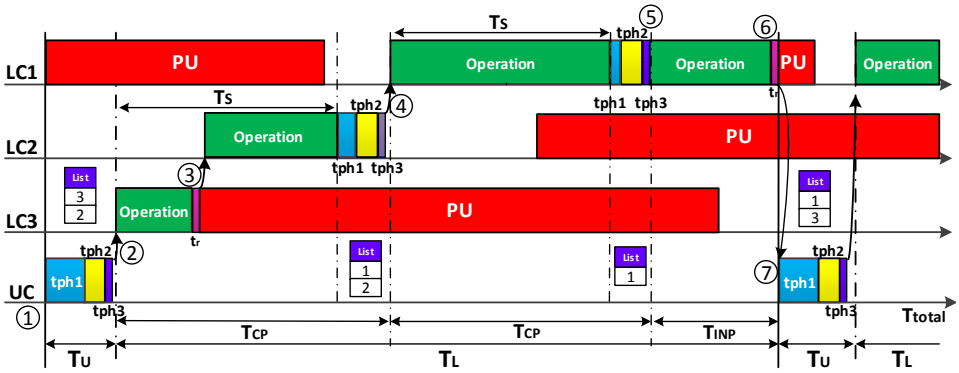


Figure 3.3: Example of the operation of the secondary network under study.

period, the SN hops back to the unlicensed channel, where it will remain for T_U. Thus, the expected value of T_p can be expressed as the sum of the expected values of T_U and T_L.

For a better understanding of T_p and consequently of T_U and T_L, an operation example is given in Fig. 3.3. According to it, 1) the SN under study while operating on the unlicensed channel (UC), it initiates a sensing procedure (consisting of t_{ph1}, t_{ph2} and t_{ph3}) to find new spectrum opportunities in licensed channels 1, 2 and 3 (LC1, LC2 and LC3), while T_U begins. 2) In the end of the sensing procedure, a list is constructed, which contains only the sensed as idle licensed channels categorized by their probability of collision among SUs. In this example, LC3 was estimated to be less contended than LC2, whereas LC1 was sensed as busy. Therefore, the SN under study hops to LC3 (T_U ends and T_L begins) and starts operating there. However, at a point, 3) the PU resumes its transmission and thus after t_r, the SN under study hops to the next channel in the list, i.e., LC2. There it keeps operating until it is time for the next sensing procedure to take place. Thereafter, an updated list is constructed which contains LC1 and LC2. Thereby, 4) a complete period (T_{CP}) ends with the SN hopping to LC1, which was estimated to be less contended than LC2. Subsequently, the SN operates on LC1 until it is time for the next sensing procedure. Given that at this stage the list only contains LC1, 5) another complete period ends (T_{CP}) and the SN keeps operating on the same channel (i.e., LC1). At a point, 6) the PU resumes its transmission and thus, given that there is no other channel in the list, the SN hops back to the unlicensed channel. This period was an incomplete period (T_{INP}). In the unlicensed channel, 7) a new T_U period begins with the SN initiating a new sensing procedure, and the process continues in a similar way.

Calculation of E$[D]$

Prior to further calculations, two key points must be clearly stated. First, the amount of data transmitted over the licensed channels is tightly coupled with the

$$f_{\tau_k}(\tau) = \begin{cases} P(\mathcal{O}_k(0) = idle)(1 - e^{-\frac{S_{k-1}}{T_{on}}} - e^{-\frac{S_{k-1}}{T_{off}}}) + e^{-\frac{S_{k-1}}{T_{on}}} & \text{if } \tau = 0 \\ \frac{1}{T_{off}} e^{-\frac{\tau}{T_{off}}} \left((e^{-\frac{S_{k-1}}{T_{off}}} + e^{-\frac{S_{k-1}}{T_{on}}} - 1)\ P(\mathcal{O}_k(0) = idle) + (1 - e^{-\frac{S_{k-1}}{T_{on}}}) \right) & \text{otherwise} \end{cases} \tag{3.3}$$

number of available channels and with the time that these channels remain idle. The more the available channels, the more the transmission opportunities for the SN. Second, the SN aims to operate exclusively on licensed channels (after the initial setup on the unlicensed channel) or at least for as long as possible. This mainly depends on T_S and the PUs' (in)activity period. If T_S is longer compared to the inactivity periods, the probability that all the available channels become busy between two consecutive sensing procedures increases. On the contrary, if T_S is considerably shorter, unnecessary sensing procedures are triggered, thereby reducing the SN's effective transmission opportunities.

In this context, the set of ordered channels detected idle after a cooperative sensing procedure of the SN is denoted by $\mathcal{B}$[7]. After the sensing procedure, all SUs operate on the first channel in $\mathcal{B}$ for as long as it remains idle. Then, when the channel turns into the busy state, all the users hop to the second channel in $\mathcal{B}$. This operation is repeated until there are no channels available.

Lemma 1. *Given a set of channels in $\mathcal{B}$ with activity and inactivity periods independently and exponentially distributed with T_{on} and T_{off} mean values respectively, the time elapsed between the beginning and the end (due to PU activity resumption) of the SN operation on the kth channel in $\mathcal{B}$ is denoted by τ_k. The probability density function (pdf) of τ_k can be written as in* (3.3)*, where $\mathcal{O}_k(t) \in \{idle, busy\}$ is the actual state of the kth channel at time t and S_{k-1} denotes the total time spent in the previous $k-1$ licensed channels,*

$$S_{k-1} = \sum_{j=1}^{k-1} \tau_j + (j-1)\delta \tag{3.4}$$

where $\delta = t_r + t_{sw}$ is the time required to detect the change in licensed channel activity (t_r) and switch to the following channel (t_{sw}).

Proof. See Appendix 3.1. □

As already mentioned, the SN exchanges data packets only in the licensed channels. Thus, given the set $\mathcal{B}$ of licensed channels the expected payload sent by the

[7]The set $\mathcal{B}$ contains all the licensed channels sensed with no PU activity, ordered by the other SNs' activity; the lower the other SNs' activity in the channel, the lower the channel ordinal in $\mathcal{B}$, and the higher the probability of being visited by the SN.

SN, $\mathbf{E}[D]$, equals to

$$\mathbf{E}[D] = \sum_{k \in \mathcal{B}} \mathbf{E}[N_{pack_k}]\mathbf{E}[P] \tag{3.5}$$

where $\mathbf{E}[N_{pack_k}]$ denotes the expected number of successfully transmitted packets (or equivalently of successful transmission slots, since a successful transmission slot corresponds to the successful transmission of one packet) by the SN under study in the kth licensed channel during $T_p = T_U + T_L$ and $\mathbf{E}[P]$ the average packet payload size. Then, $\mathbf{E}[N_{pack_k}]$ may be expressed as

$$\mathbf{E}[N_{pack_k}] = \frac{\mathbf{E}[T_k]}{\mathbf{E}[T_{slot_k}]} P_{s_k} \tag{3.6}$$

where $\mathbf{E}[T_k]$ is the expected operation time (i.e., successful transmission, collision and idle slots) on the kth channel, P_{s_k} [8] the probability of having a successful transmission by the SN in the kth channel and $\mathbf{E}[T_{slot_k}]$ denotes the average slot duration in the kth channel.

As previously expounded, a period can be defined as the time between the completion of two consecutive sensing procedures. Henceforth, the periods during which at least one licensed channel remains idle are called complete periods, whereas the periods during which all idle channels change their state are denoted as incomplete periods.

Lemma 2. *Given a set of channels $\mathcal{B}$, ordered according to the sensed contention level, the expected operation time on the kth channel (T_k), is given by*

$$\boldsymbol{E}[T_k] = \sum_{x=0}^{\infty} P(X = x)((x-1)\boldsymbol{E}[T_k]_{CP} + \boldsymbol{E}[T_k]_{INP}) \tag{3.7}$$

where X is the number of successive periods (i.e., $X-1$ complete and one incomplete periods) operating exclusively on licensed channels, and $\boldsymbol{E}[T_k]_{CP}$ and $\boldsymbol{E}[T_k]_{INP}$ denote the expected operation time of the SN on the kth channel in a complete and an incomplete period, respectively.

Proof. See Appendix 3.2. □

Regarding the average slot duration in the kth channel, $\mathbf{E}[T_{slot_k}]$, it can be easily derived as

$$\mathbf{E}[T_{slot_k}] = P_{i_k}\sigma + (P_{s_k} + P'_{s_k})T_{rs} + (P_{c_k} + P'_{c_k})T_c \tag{3.8}$$

where P_{i_k} is the probability of having an idle slot in the kth channel, P_{c_k} the one of a collision slot of the SN under study, and P'_{s_k} and P'_{c_k} the probability of a successful

[8] Please note that the proposed model can be applied to any traffic pattern of contending users in the unlicensed channel and licensed channels. Closed form expressions for the traffic-dependent parameters (e.g., probability of a successful transmission, collision and idle slot) can be found in [68, 76] for saturated and non-saturated conditions, respectively. For the reader's convenience, the ones for saturated conditions are derived in Appendix 3.3.

$$\begin{aligned}\mathbf{E}[T_{U_{un}}] &= \frac{P_i\sigma + P_s'T_{rs} + (P_c + P_c')T_c}{P_s} + M_{us}(t_{sr} + t_{sw}) + t_{sw} + t_{RFS} + \\ &+ t_{SIFS}(N+1) + t_{ACK} + 2t_{DIFS} + t_B + t_{SR}(N-1) + t_{LIST}\end{aligned} \tag{3.11}$$

transmission and collision slot, respectively, of the other SNs in the kth channel. The parameter σ denotes the empty slot duration, while T_{rs} and T_c the durations of a successful transmission and collision slot. Further details on the calculation of these parameters are included in Appendix 3.3.

Calculation of $\mathbf{E}[T_U]$

The time spent in the unlicensed channels is devoted to sense the licensed channels and share the information on their availability. The SUs will not be able to operate on licensed channels if no available channels have been detected. Hence, the procedure consisting of gaining access to the unlicensed channel, sensing licensed channels and exchanging the information, will be repeated until there is at least one licensed channel detected as idle.

Lemma 3. *The expected time spent in the unlicensed channel is given by*

$$\boldsymbol{E}[T_U] = \boldsymbol{E}[T_{U_{un}}](\frac{1}{P_{U_s}} - 1) + \boldsymbol{E}[T_{U_s}] \tag{3.9}$$

where $P_{U_s} = 1 - \prod_{n=1}^{M}(1 - P_{s_{idle_n}})$ *is the probability that there is at least one licensed channel sensed as idle (*M *is the total number of licensed channels that are sensed, with* $|\mathcal{B}| \leq M$*), and* $T_{U_{un}}$ *and* T_{U_s} *are the time spent when there is not any channel sensed as idle and when there is at least one channel sensed as idle, respectively. The expected value of* T_{U_s} *is given by*

$$\boldsymbol{E}[T_{U_s}] = \boldsymbol{E}[T_{U_{un}}] + t_{sw} \tag{3.10}$$

Then, the expected value of $T_{U_{un}}$ *is given by* (3.11)*, where* M_{us} *is the number of licensed channels to be sensed by each SU;* t_{sn} *the time to sense a licensed channel;* t_{sw} *the time required to switch between two channels;* σ*,* T_{rs} *and* T_c *the durations of an idle, a successful transmission and a collision slot;* t_{RFS}*,* t_{ACK}*,* t_B*,* t_{SR} *and* t_{LIST} *time required to transmit an RFS, ACK, beacon, report, and list packet, respectively;* P_i *the probability of having an idle slot in the unlicensed channel;* P_c *the probability of a collision slot of the SN under study in the unlicensed channel; and* P_s' *and* P_c' *are the probabilities of a successful transmission and collision slot of the other SNs in the unlicensed channel, respectively.*

Proof. See Appendix 3.4. □

$$\begin{aligned}\mathbf{E}[E_{U_{un}}] &= \frac{N\boldsymbol{P}_{idle}(P_i\sigma + P_s'T_{rs} + P_c'T_c) + P_c(t_{data}(\boldsymbol{P}_{T_x} + (N-1)\boldsymbol{P}_{R_x}) + N\boldsymbol{P}_{idle}t_{DIFS})}{P_s} + \\ &+ N(M_{us}(\boldsymbol{P}_{sn}t_{sn} + \boldsymbol{P}_{sw}t_{sw}) + \boldsymbol{P}_{sw}t_{sw})(\boldsymbol{P}_{T_x} + (N-1)\boldsymbol{P}_{R_x})(t_{RFS} + t_B + t_{LIST}) + \\ &+ (\boldsymbol{P}_{T_x} + \boldsymbol{P}_{R_x} + (N-2)\boldsymbol{P}_{idle})(t_{ACK} + (N-1)t_{SR}) + N\boldsymbol{P}_{idle}((N+1)t_{SIFS} + 2t_{DIFS})\end{aligned} \tag{3.15}$$

Calculation of $\mathbf{E}[T_L]$

The time spent in licensed channels can be divided into two parts: the effective time devoted to data transmission, and the time devoted to sensing, to detect the resumption of the PUs activity and to switch to an alternative channel. Therefore,

$$\mathbf{E}[T_L] = \mathbf{E}[T_{rsn}] + \sum_{k\in\mathcal{B}} \mathbf{E}[T_k] \tag{3.12}$$

where $\mathbf{E}[T_{rsn}]$ is the expected time spent in both the reaction periods and the sensing procedures during T_p and $\mathbf{E}[T_k]$ is the operation time on the kth channel calculated in (3.29). For the sake of clarity, the calculation of $\mathbf{E}[T_{rsn}]$ is detailed in Appendix 3.5.

3.6 Energy Efficiency Analysis

The energy efficiency of the SN under study can be expressed as

$$Q = \frac{\mathbf{E}[D]}{\mathbf{E}[E_U] + \mathbf{E}[E_L]} \tag{3.13}$$

where $\mathbf{E}[D]$ has been derived in (3.5) and $\mathbf{E}[E_U]$ and $\mathbf{E}[E_L]$ are the expected energy consumptions in the unlicensed channel and licensed channels, respectively, during T_p.

Lemma 4. *The expected energy consumed in the unlicensed channel is given by*

$$\boldsymbol{E}[E_U] = \boldsymbol{E}[E_{U_{un}}](\frac{1}{P_{U_s}} - 1) + \boldsymbol{E}[E_{U_s}] \tag{3.14}$$

where $E_{U_{un}}$ and E_{U_s} are the energy consumptions when there is not any channel sensed as idle and when there is at least one channel sensed as idle, respectively. The expected value of E_{U_s} is given by

$$\boldsymbol{E}[E_{U_s}] = \boldsymbol{E}[E_{U_{un}}] + N\boldsymbol{P}_{sw}t_{sw} \tag{3.16}$$

whereas the one of $E_{U_{un}}$ is given by (3.15)*, where $\boldsymbol{P}_{T_x}$, $\boldsymbol{P}_{R_x}$ and $\boldsymbol{P}_{idle}$ are the transmission, reception and idle power, while $\boldsymbol{P}_{sn}$ and $\boldsymbol{P}_{sw}$*[9] *denote the sensing*

[9]The rest of the power consumptions (e.g., by the cluster head to construct the list) can be

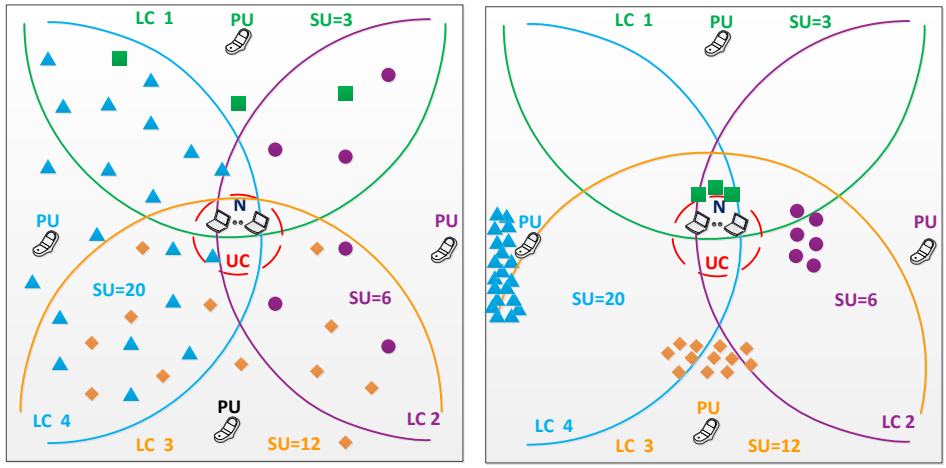


Figure 3.4: Simulation scenarios: (a) Scenario 1, where the SUs of each licensed channel are uniformly distributed around the SN under study and (b) Scenario 2, where the more the SUs on a licensed channel, the further they are located from the SN under study.

power and the power to switch to another channel, respectively.

Proof. See Appendix 3.6. □

Then, the expected energy consumed in the licensed channels can be expressed as

$$\mathbf{E}[E_L] = \mathbf{E}[E_{rsn}] + \sum_{k \in \mathcal{B}} \mathbf{E}[E_{cont_k}] \tag{3.17}$$

where $\mathbf{E}[E_{rsn}]$ is the expected energy consumed by the SN in both the reaction periods and the sensing procedures during T_p and $\mathbf{E}[E_{cont_k}]$ during the contention time in the kth channel. For the sake of clarity, the calculations of $\mathbf{E}[E_{rsn}]$ and $\mathbf{E}[E_{cont_k}]$ are detailed in Appendix 3.7.

3.7 Analysis and Simulation Results

3.7.1 Simulation scenarios

In the extensive simulations executed in MATLAB$^{\text{TM}}$, a SN of N SUs and a set of M= 6 licensed channels are considered, while N_{unlic} = 50. In order to focus

considered negligible.

Table 3.2: Energy-efficient Contention-aware Channel Selection in Cognitive Radio Networks: Simulation Values

Parameter	Value	Parameter	Value
$P_{fa_{kj}}$ / $P_{md_{kj}}$	0.01 / 0.1	m / CW_{min}	6 / 16
T_{on}, T_{off}	1 s	t_r, t_{sn}	1 ms
$TXOP$	6.016 ms	t_{sw}	9 ms
σ	9 μs	t_{SIFS} / t_{DIFS}	10 / 28 μs
PLCP & PHY header	20 μs	MAC Header	34 bytes
l_{ACK}, l_B	14 bytes	Payload	1000 bytes
l_{RFS}	20 bytes	l_{SR}, l_{LIST}	16 bytes
Control T_x Rate	6 Mbps	Data T_x Rate	24 Mbps
$\boldsymbol{P}_{R_x}$, $\boldsymbol{P}_{idle}$, $\boldsymbol{P}_{sn}$	1340 mW	$\boldsymbol{P}_{T_x}$, $\boldsymbol{P}_{sw}$	1900 mW

on the performance assessment of the channel selection algorithms, ideal channel conditions are assumed (i.e., no fading), while all users have the same probability of false alarm and mis-detection for all channels equal to 0.01 and 0.1, respectively. Without loss of generality, all users both in the unlicensed channel and licensed channels are assumed to be in saturated conditions (i.e., always having a packet to transmit). Hence, due to the same traffic conditions, only the number of SUs in each licensed channel is sufficient to define its contention level. Thus, $N_{SU_{lic}}$ is defined as the maximum number of SUs of other SNs that operate on a licensed channel (this also corresponds to the maximum probability of collision). The MAC layer parameters (e.g., m, CW_{min}, Data T_x Rate) have been selected according to IEEE 802.11g Standard [72], while all simulation parameters are summarized in Table 3.2.

Regarding the system topology, the SUs of the SN under study are assumed to be located in the center of a 100m $\times$ 100m square region, while the following scenarios are considered:

- Scenario 1: In this general scenario, the SUs of each licensed channel are uniformly distributed around the SN. An example of four licensed channels is given in Fig. 3.4(a).
- Scenario 2: In this scenario, the more the SUs on a licensed channel, the further they are located from the SN under study (i.e., the SUs of high contended channels are located further, while the SUs of low contended channels are located closer to the SN under study). An example is depicted in Fig. 3.4(b), where the SUs of licensed channel 1 (LC 1), which is the least contended, are located the closest to the SN, while the users of LC 4, which is the most contended, are located the furthest compared to the SUs operating on LC 2 and LC 3. The purpose of this scenario is to show the dependency of the applied channel selection algorithm on the SUs' topology in each licensed channel. Still, notice that such a scenario could correspond to a heterogeneous network scenario with hotspot traffic at these locations (e.g., in shopping malls).

3.7.2 State-of-the-art Algorithms

Being the closest to this work, the following three channel selection algorithms, previously discussed in Section 3.1, will be adapted to the considered scenario, to fairly compare them with the proposed approach. To that end, in all cases, the same SN coexistence scheme is applied, as described in Section 3.4, but with a different channel selection algorithm each time.

Feature detection (FD) algorithm

This algorithm refers to the case where only cyclostationary feature detection is used without any extra estimation technique for the number of contending SUs. In this case, the algorithm is able of distinguishing between channels with PU activity that are avoided and thus are not included in the list, channels with SU activity and channels with no activity at all. For a fair comparison the channels with no activity (neither PU nor SU) will be preferable and, thus, will take the first place in the list. Then, the rest of the channels will be positioned in a random order as in [33, 35].

Interference-aware (IA) algorithm

This algorithm proposes the classification of the licensed channels according to their interference level and the selection of the channel with the least interference [34,35]. In the considered simulations, the interference is measured by $I = \sum_{s=1}^{N_k} G_s P_{trsm} + \sigma_k^2$, where N_k is the number of SUs on the licensed channel k, σ_k^2 the variance of the additive white Gaussian noise and G_s the channel gain between the SN under study and the user s. Then, $G_s = \frac{l_s}{{d_s}^n}$, where l_s is a random variable representing the log-normal shadowing, d_s is the distance between the SN under study and the user s and n denotes the path loss exponent. In the considered simulations, l_s has mean 0 and variance 8 (dB), $\sigma_c^2 = 1\ mW$ and $n = 3$.

Energy detection (ED) Algorithm

Although a comparison between feature detection and energy detection is out of the scope of this chapter but can be found in [13,77], it could provide us with interesting insights that justify the motivation of this work. As previously mentioned, ED, unlike FD, is unable to distinguish if a licensed channel is occupied by a PU or by a SN. For a realistic and fair comparison, in the executed simulations, the parameter values of [13] will be used.

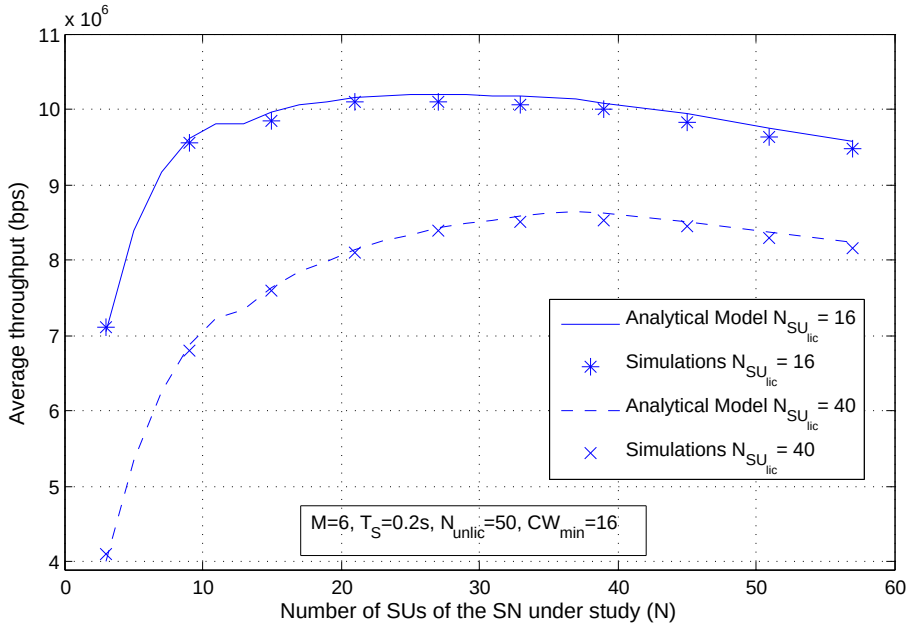


Figure 3.5: Throughput versus N for different maximum numbers of competing SUs in any licensed channel, $N_{SU_{lic}}$.

3.7.3 Model validation

In Fig. 3.5 and 3.6, the throughput and the energy efficiency of the SN versus its number of SUs (N) are depicted analytically and verified by simulations for $N_{SU_{lic}} = 16$ and $N_{SU_{lic}} = 40$. As it can be noticed, as N increases, the throughput is also increased until an upper bound is reached, due to the saturation of the licensed channel (i.e., saturation throughput) and then it decreases. Notice that this decrease is smooth, due to the trade-off between the detected transmission opportunities and the collision probability. In particular, as N increases, the detection accuracy of the idle channels increases, but so does the collision probability. On the other hand, the energy efficiency of the SN is decreased with the increase of its SUs' number, as the energy consumption increases in a greater extent than the successful bits transmitted by the SN. It is worth noting that, due to the high number of SUs belonging to the SN under study, the energy consumption during sensing as well as during the contention slots (idle, collision and successful slots) increases significantly. Furthermore, the less the contention in the licensed channels (i.e., the lower the $N_{SU_{lic}}$), the higher the throughput and the energy efficiency of the SN under study, as it experiences less collisions.

In Fig. 3.7 and 3.8, the effect of the parameter T_S (i.e., time between two consecutive sensing periods) on the throughput and the energy efficiency of the SN under study is studied for three different values of T_{on} and T_{off}. The analytical results

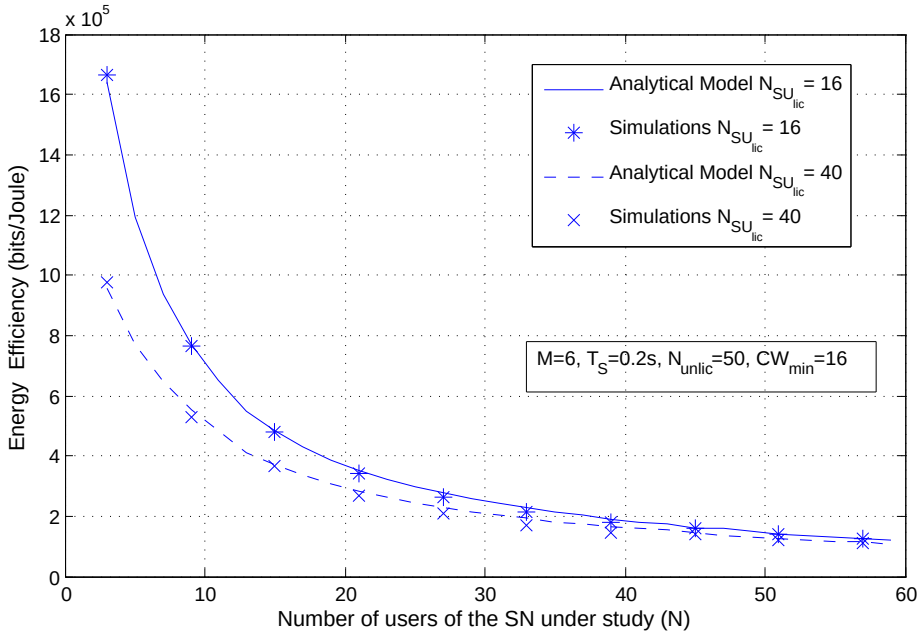


Figure 3.6: Energy efficiency versus N for different maximum numbers of competing SUs in any licensed channel, $N_{SU_{lic}}$.

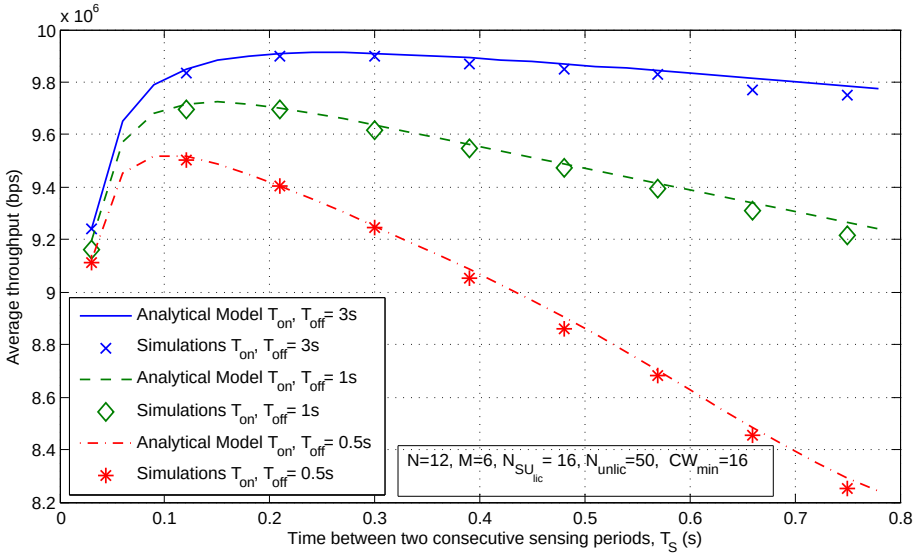


Figure 3.7: Throughput versus time between two consecutive sensing periods, T_S.

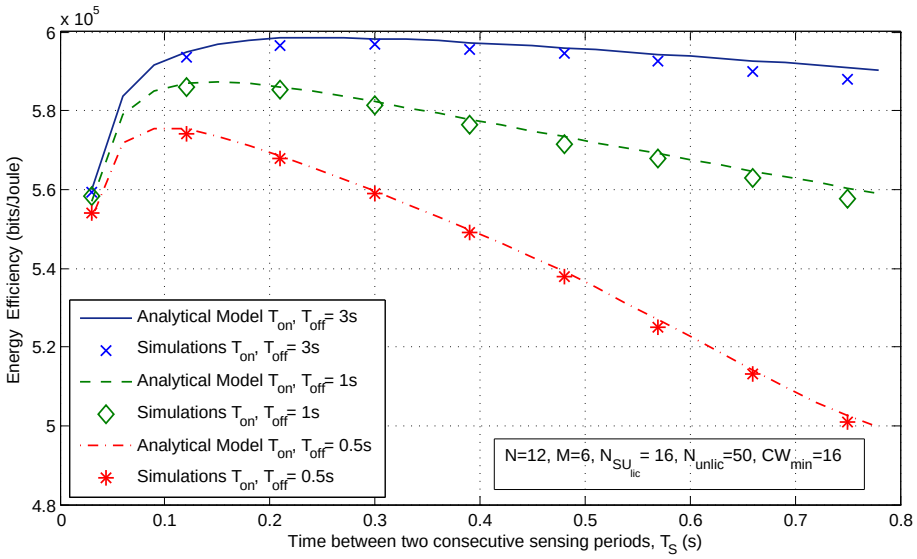


Figure 3.8: Energy efficiency versus time between two consecutive sensing periods, T_S.

are also presented both for throughput and energy efficiency and they are in a good agreement with the simulations. As it can be noticed, there is a maximum throughput and energy efficiency value achieved for each one of the curves. This maximum corresponds to the optimal value of T_S. For lower values than the optimal, there is a lot of time and energy spent in unnecessary frequent sensing procedures and, thus, less time available for data transmission, whereas for higher ones the list is not updated frequently and, thus, the SN has to switch licensed channels to avoid interfering with the PU. Moreover, as it can be observed, this optimal value depends on the PU activity pattern. Hence, for slowly-changing PU activity (i.e., high values of T_{on}, T_{off}), there is no need of frequent sensing procedures and, thus, the optimal value of T_S increases, to appropriately adapt to the PU activity. In addition, notice that for a fix value of T_S, slowly-changing PU activity (i.e., high values of T_{on}, T_{off}) results in higher throughput and energy efficiency, since the SN operates for a longer amount of time on the channel without PU transmission resumption. On the contrary, for quickly-changing PU activity, the SN has to switch among licensed channels frequently in order not to interfere with the PU, resulting in less time devoted to data transmission and more energy consumption.

3.7.4 Performance evaluation

In Fig. 3.9 and 3.10, the comparison of the proposed algorithm (PA) with the aforementioned state-of-the-art algorithms is given. In particular, the throughput and energy efficiency of the SN under study of all algorithms are respectively depicted with $CW_{min} = 16$, versus the maximum number of SUs in a licensed channel

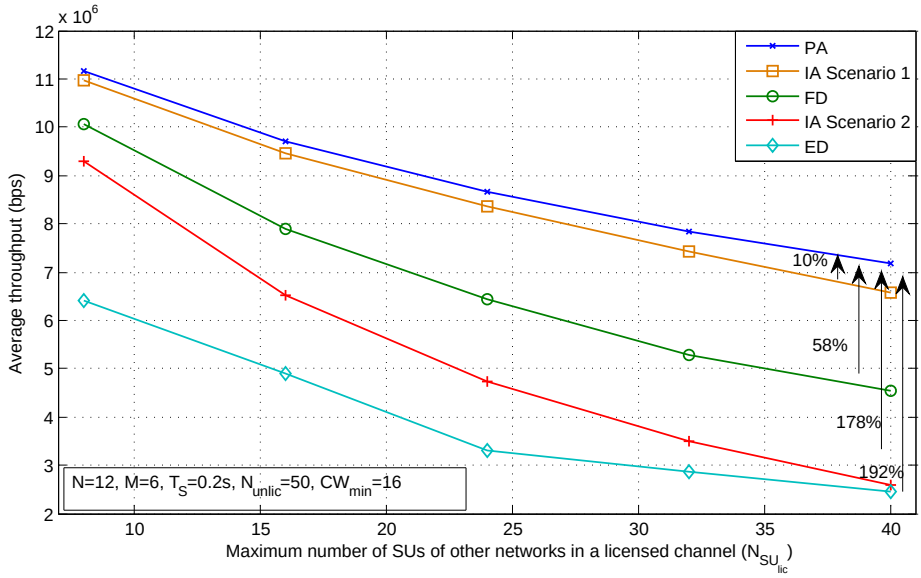


Figure 3.9: Throughput versus maximum number of competing SUs in any licensed channel, $N_{SU_{lic}}$, with $CW_{min} = 16$.

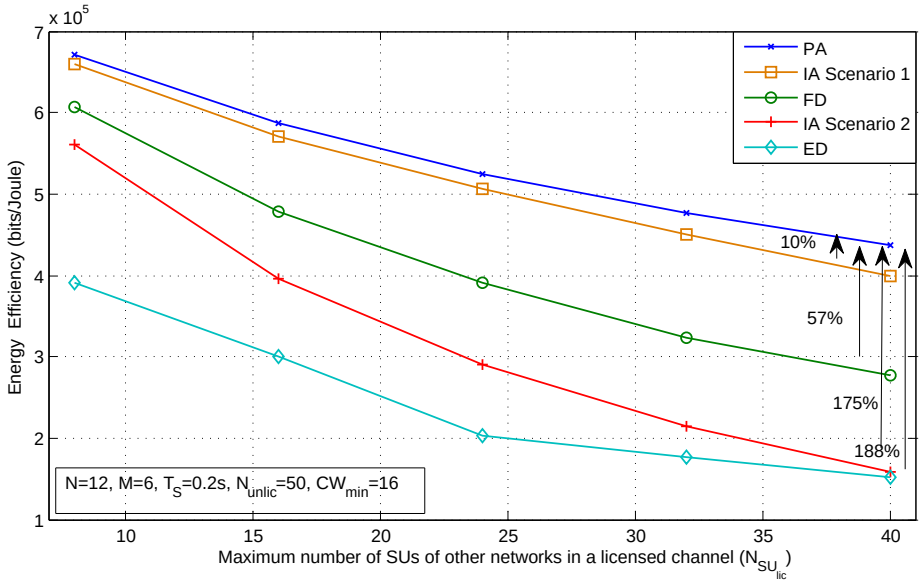


Figure 3.10: Energy efficiency versus maximum number of competing SUs in any licensed channel, $N_{SU_{lic}}$, with $CW_{min} = 16$.

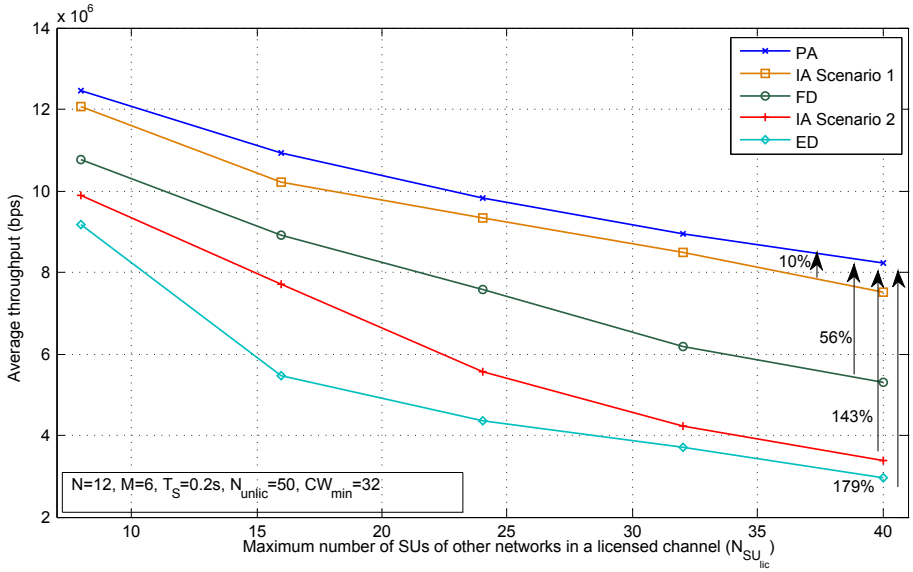


Figure 3.11: Throughput versus maximum number of competing SUs in any licensed channel, $N_{SU_{lic}}$, with $CW_{min} = 32$.

($N_{SU_{lic}}$). Although it will be detailed later, notice that only the IA algorithm performance was found to be dependent on the system topology, and therefore, is depicted separately for Scenario 1 and 2, unlike the rest of the algorithms. In addition, it can be observed that, as the parameter $N_{SU_{lic}}$ increases, both the throughput and energy efficiency of the SN decreases for all algorithms, by virtue of the fact that the contention in the licensed channels increases and so does the energy consumption.

In comparison with FD, PA shows better performance in both throughput and energy efficiency. This stems from the fact that in FD, the SN randomly chooses an idle licensed channel for transmission, thus having higher probability to spend more time in highly contended licensed channels compared to PA. Furthermore, as the parameter $N_{SU_{lic}}$ increases, the relative gain of PA in both throughput and energy efficiency increases due to its contention-awareness and for $N_{SU_{lic}} = 40$ it can present up to 58% improvement in throughput and 57% in energy efficiency.

As far as the IA algorithm is concerned, PA can present up to 178% improvement in throughput and 175% in energy efficiency in cases such as the Scenario 2. In that case, the channels are classified in the opposite order than in PA, namely the channel with the highest contention will present the least interference and thus, it will be the most preferable by the IA algorithm. Therefore, this scenario defines the maximum gain that can be achieved compared to IA. However, even in cases where the users are uniformly distributed (i.e., in Scenario 1), PA can still present a 10% improvement, as its performance does not depend on the topology and thus, it achieves higher accuracy in detecting the low contended channels.

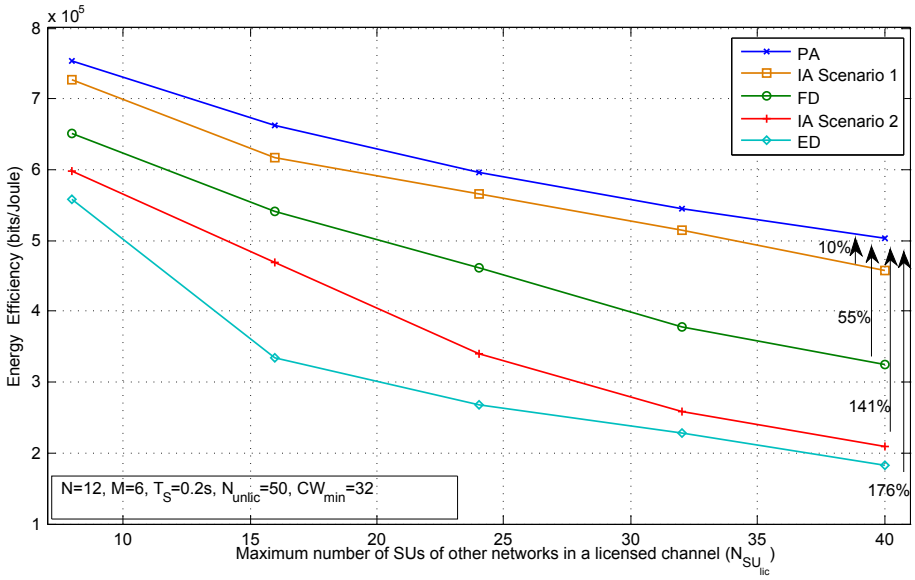


Figure 3.12: Energy efficiency versus maximum number of competing SUs in any licensed channel, $N_{SU_{lic}}$, with $CW_{min} = 32$.

Finally, PA significantly outperforms the ED Algorithm, as the spectrum opportunities that are exploited by the ED algorithm are much less than those of PA. This stems from the fact that, when ED is applied, the channels occupied by SNs are sensed as busy and thus are avoided. The consequent considerably less spectrum efficiency results in a significant degradation of throughput and energy efficiency, as the SN mostly remains in the highly congested unlicensed channel. The gain under high contention in the licensed channels can reach up to 192% in throughput and 188% in energy efficiency.

Further experiments were conducted with $CW_{min} = 32$ aiming at studying the impact of the minimum back-off window value on the algorithms' performance, as depicted in Fig. 3.11 and 3.12. To that end, it can be observed that a higher minimum back-off window value (i.e., with $CW_{min} = 32$) results in higher throughput and energy efficiency for all algorithms. This is due to the fact that in this case the SUs have to defer their transmissions for a longer time, when another node transmits and, thus, the collisions are avoided more efficiently. However, this highly depends on the number of contending users; the more the users, the higher the probability of collision and, thus, the higher the minimum back-off window value should be. In addition, it can be noticed that PA significantly outperforms the reference algorithms for both considered values of minimum back-off window, with the highest performance gains being achieved for $CW_{min} = 16$. This stems from the fact that the rest of the algorithms spend more time in highly contended channels, where a very low back off window value has a severe impact on the network performance due to the increased number of collisions among the SUs.

3.7.5 Fairness evaluation of the coexistence scheme

Having shown the outperformance of the proposed channel selection algorithm compared to the state-of-the-art and given that in such coexistence scenarios, it is fundamental to guarantee fairness among the coexisting SNs, in this section, we evaluate the performance of the proposed SN coexistence scheme, previously described in Section 3.4, when the proposed channel selection algorithm of Section 4.20 is applied. In particular, the proposed SN coexistence scheme of Section 3.4 is compared with the reference approach (FMAC) [32] by means of simulations and it is shown that it can achieve significant throughput and energy efficiency gains, while maintaining or even achieving better fairness among the coexisting SNs. Furthermore, the impact of different minimum back-off window values and different PU activity patterns on the performance of both considered coexistence schemes is studied.

In the considered results, the same simulation parameters of Section 3.7.4 are employed, while the proposed SN coexistence scheme is denoted by SNCS, and the reference scheme by FMAC [32]. In order to calculate the fairness among the coexisting SUs that contend for the same PU resources, the well known Jain's fairness index is employed [78], that is given by

$$J(x_1, x_2...x_n) = \frac{(\sum_{i=1}^{n} x_i)^2}{n \sum_{i=1}^{n} {x_i}^2} \tag{3.18}$$

where n is the number of contending SUs and x_i denotes the number of transmission opportunities of the SU i. An SU is considered to have a transmission opportunity every time it transmits a packet on the channel independently of whether a successful transmission or a collision occurs.

In Fig. 3.13 and 3.14, the average throughput and energy efficiency of the SN under study are respectively depicted for both SNCS and FMAC versus the maximum number of SUs of other networks in a licensed channel, $N_{SU_{lic}}$, for different minimum back-off window values, CW_{min}. As it can be observed, the throughput and the energy efficiency of the SN under study are decreased as the contention in the licensed channels increases, due to the increased number of collisions among the SUs. However, notice that the SNCS achieves better performance than FMAC for all the considered values of CW_{min}. This is due to the fact that, in FMAC, the SUs use a constant back-off window every time a collision takes place, while the SNCS employs an exponential one and thus it manages the collisions more efficiently. Therefore, the maximum performance gain of SNCS is achieved for the lowest minimum back-off window value ($CW_{min} = 16$) and for high contention in the licensed channels ($N_{SU_{lic}} = 40$).

Moreover, as far as the fairness of the considered approaches is concerned, the average Jain's index of all the SUs, that contend to gain access to a licensed channel, is depicted for both approaches in Fig. 3.15, for different minimum back-off window values, CW_{min}. As we can notice, SNCS can achieve up to 25% better fairness than FMAC ($CW_{min} = 16$) for high contention in the licensed channels ($N_{SU_{lic}} = 40$).

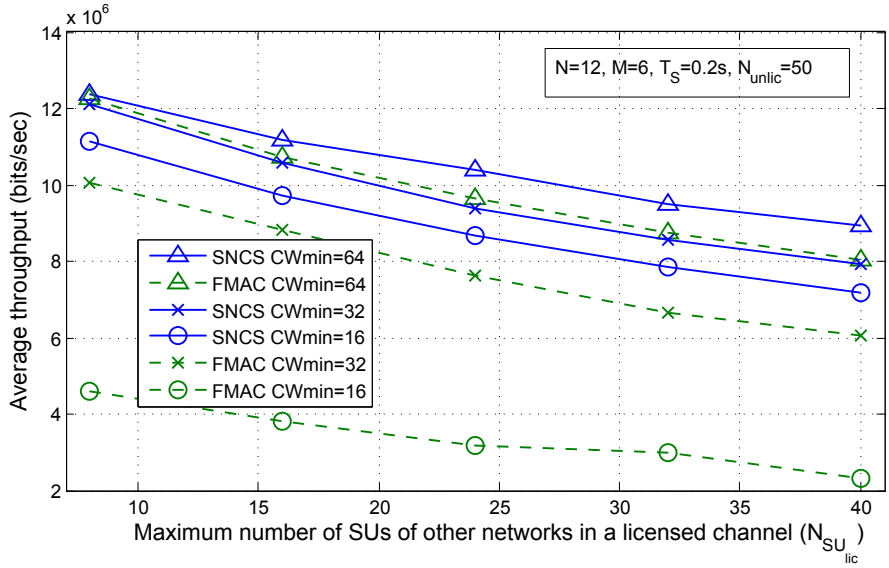


Figure 3.13: Average throughput of the SN under study versus the maximum number of SUs of other networks in a licensed channel, $N_{SU_{lic}}$, for different minimum back-off window values, CW_{min}, when the proposed channel selection algorithm (PA) is applied.

This stems from the fact that for short contention periods, i.e., in the case that the PU resumes its activity in the licensed channel shortly after the SN under study has hopped to it, the SNCS achieves much better fairness among the SUs than in FMAC, as an SU that is involved in a collision defers its transmission for a longer time, and thus the transmissions opportunities are more equally distributed among the contending SUs.

To that end, in Fig. 3.16, 3.17 and 3.18, the performance of the considered coexistence schemes is studied for different PU activity patterns. Specifically, two different values of T_{on} and T_{off} are considered, that correspond to quickly (T_{on}=T_{off}=0.5s) and slowly changing PU activity (T_{on}=T_{off}=2s), respectively. Moreover, notice that the T_S value is adapted according to the considered PU pattern. In particular, for quickly changing PU activity a low value of T_S is employed to repeat the sensing procedure more frequently to keep the information for every licensed channel updated, while for slowly changing PU activity, this value is chosen to be higher to let more time available for data transmissions to the SUs.

As it can be observed in Fig. 3.16 and 3.17, for slowly changing PU activity, the SN under study achieves higher throughput and energy efficiency, as there is more time devoted to transmissions and less to frequent unnecessary sensing procedures. In addition, the SNCS achieves better performance than FMAC for both the considered PU traffic patterns. Please note that CW_{min} = 64 has been selected for

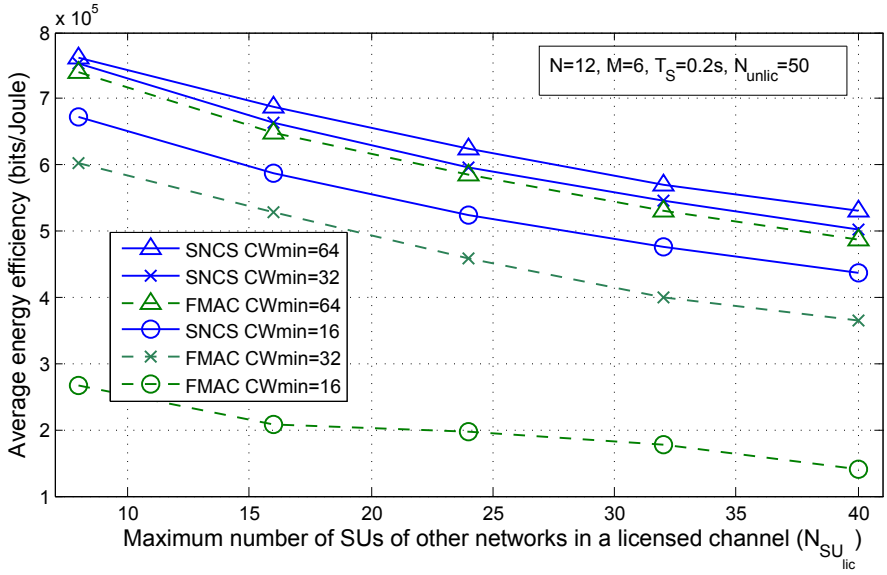


Figure 3.14: Average energy efficiency of the SN under study versus the maximum number of SUs of other networks in a licensed channel, $N_{SU_{lic}}$, for different minimum back-off window values, CW_{min}, when PA is applied.

both approaches, to show the minimum gains that can be achieved in comparison to FMAC.

The average Jain's index is depicted in Fig. 3.18 for different PU activity patterns. As we can notice, for slowly changing PU activity, both approaches achieve slightly better fairness among the contending SUs than in the quickly changing PU activity case, as there is more time devoted to transmissions. Moreover, for low contention in the licensed channels, FMAC achieves better performance in terms of fairness. However, according to the previously analyzed reasoning, for higher contention there is a cross point where the SNCS starts achieving better fairness. The SNCS maximum performance gain in this case is again achieved for high contention in the licensed channels ($N_{SU_{lic}} = 40$).

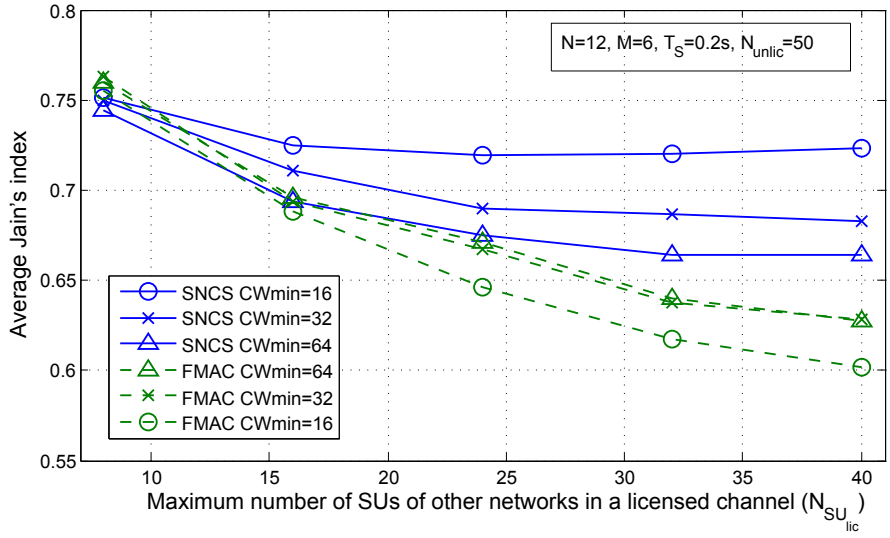


Figure 3.15: Average Jain's index versus the maximum number of SUs of other networks in a licensed channel, $N_{SU_{lic}}$, for different minimum back-off window values, CW_{min}, when PA is applied.

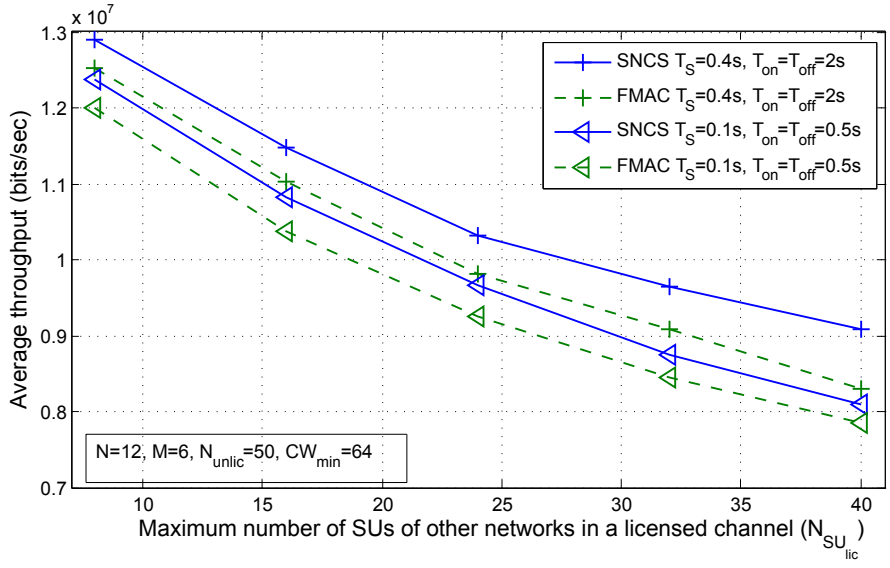


Figure 3.16: Average throughput of the SN under study versus the maximum number of SUs of other networks in a licensed channel, $N_{SU_{lic}}$, for different PU activity patterns, when the proposed channel selection algorithm (PA) is applied.

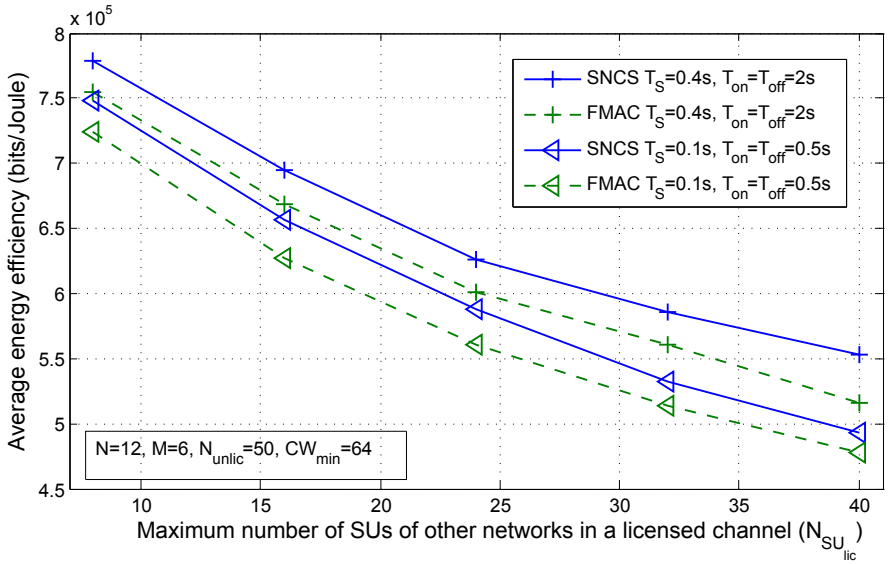


Figure 3.17: Average energy efficiency of the SN under study versus the maximum number of SUs of other networks in a licensed channel, $N_{SU_{lic}}$, for different PU activity patterns, when PA is applied.

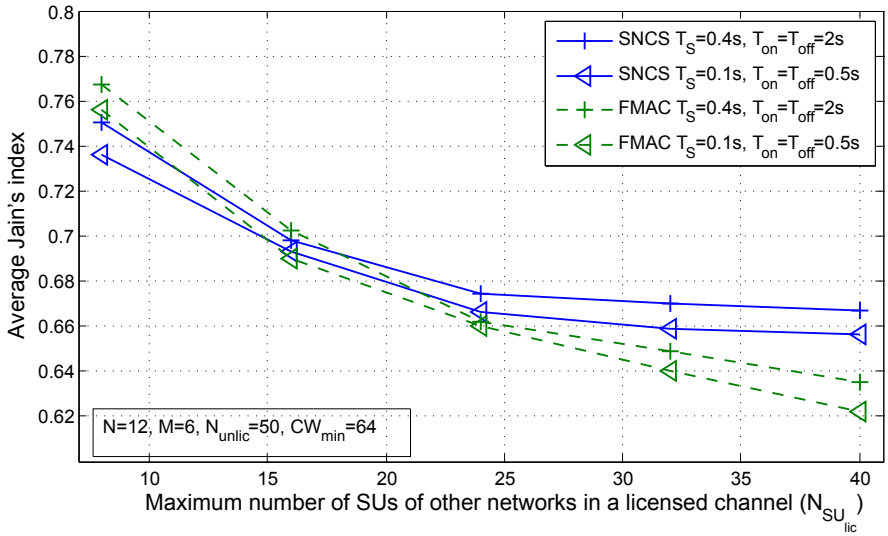


Figure 3.18: Average Jain's index versus the maximum number of SUs of other networks in a licensed channel, $N_{SU_{lic}}$, for different PU activity patterns, when PA is applied.

3.8 Conclusions

In this chapter, a SN coexistence scheme was proposed as well as a novel contention-aware channel selection algorithm that aims at improving the throughput and energy efficiency of a SN, that coexists with other SNs that use the same PU resources. Analytical expressions for the throughput and the energy efficiency of the SN under study have been derived and verified through extensive simulations. Moreover, it has been proved that there is an optimal value for maximum performance for the time between two consecutive sensing periods, which is highly dependent on the PU activity pattern. The proposed algorithm has been compared to three reference algorithms and it has been shown that it significantly outperforms its counterparts both in terms of throughput and energy efficiency.

Moreover, in the same scenario and given that the proposed channel selection algorithm (PA) is employed, the performance of the proposed SN coexistence scheme was evaluated in terms of throughput, energy efficiency and fairness, by comparing it to a reference approach. By means of simulation, the effect of different minimum back-off window values on the performance of the coexistence scheme was studied. Furthermore, the impact of different PU activity patterns on the considered coexistence schemes was analyzed. It was shown that the considered coexistence scheme can achieve throughput and energy efficiency gains, while maintaining fairness among the coexisting SNs in comparison to the reference approach. In particular, the maximum gain is achieved for the lowest minimum back-off window value ($CW_{min} = 16$) and for high contention in the licensed channels ($N_{SU_{lic}} = 40$).

Appendix 3.1 Proof of Lemma 1

The (in)activity periods duration in licensed channels are modeled as exponential independent and identically distributed (i.i.d.) random variables. Accordingly, the pdf of the (in)activity period duration of the kth channel can be expressed as $f_{t_k}(t_k) = \frac{1}{A} e^{-\frac{t_k}{A}}$, where A is equal to the mean value of the inactivity or activity period duration, T_{off} and T_{on}, respectively. Hence, if we define $\mathcal{O}_k(t) \in \{idle, busy\}$ as the state of the kth channel at time t, the time during which the SN operates on the kth channel (τ_k) may be written as

$$\tau_k = \begin{cases} 0, \text{if } \mathcal{O}_k(S_{k-1}) = busy \\ t_k - S_{k-1}, \text{otherwise} \end{cases} \tag{3.19}$$

where S_{k-1} is the total time spent in the previously visited channels as well as the required time to detect the PUs' activity resumption and the consequent channel switching time, as shown in (3.4), and t_k is the idle state duration given that $\mathcal{O}_k(S_{k-1}) = idle$. By definition the pdf is equal to the derivative of the cumulative distribution function, i.e., $f_{\tau_k}(\tau) = \frac{\partial}{\partial \tau} F_{\tau_k}(\tau)$. Taking into account that the channels are divided into actual idle channels detected idle and busy channels erroneously detected idle, the cumulative distribution function is depicted in (3.20). The first

$$F_{\tau_k}(\tau) = P(\tau_k \leq \tau) = \\ = P(\tau_k \leq \tau | \mathcal{O}_k(0) = idle) P(\mathcal{O}_k(0) = idle) + P(\tau_k \leq \tau | \mathcal{O}_k(0) = busy) P(\mathcal{O}_k(0) = busy) \quad (3.20)$$

$$P(\tau_k \leq \tau | \mathcal{O}_k(0) = idle) = \\ = P(\tau_k \leq \tau | \mathcal{O}_k(0) = idle \cap \mathcal{O}_k(S_{k-1}) = busy) P(\mathcal{O}_k(S_{k-1}) = busy | \mathcal{O}_k(0) = idle) + \\ + P(\tau_k \leq \tau | \mathcal{O}_k(0) = idle \cap \mathcal{O}_k(S_{k-1}) = idle) P(\mathcal{O}_k(S_{k-1}) = idle | \mathcal{O}_k(0) = idle) \quad (3.21)$$

$$P(\tau_k \leq \tau | \mathcal{O}_k(0) = busy) = \\ P(\tau_k \leq \tau | \mathcal{O}_k(0) = busy \cap \mathcal{O}_k(S_{k-1}) = busy) P(\mathcal{O}_k(S_{k-1}) = busy | \mathcal{O}_k(0) = busy) + \\ + P(\tau_k \leq \tau | (\mathcal{O}_k(0) = busy \cap \mathcal{O}_k(S_{k-1}) = idle) P(\mathcal{O}_k(S_{k-1}) = idle | \mathcal{O}_k(0) = busy) \quad (3.22)$$

part of the equation models the actual idle channels, whereas the second part models the busy channels. In the case of an idle channel, it will be available for as long as it remains in the idle state. Therefore, the first part can also be expressed as in (3.21). With regard to the second part of (3.20), it models a busy channel erroneously detected idle which can be expressed as in (3.22). In such a case, the channel will only be available for the SN if the state changes before being visited by the SN and for as long as it remains in the new state (i.e., idle state). In (3.20) the probability that the kth visited licensed channel is idle at $t = 0$ is calculated as

$$P(\mathcal{O}_k(0) = idle) = \frac{P_{idle}(1 - P_{fa_k})}{P_{idle}(1 - P_{fa_k}) + P_{md_k}(1 - P_{idle})} \quad (3.23)$$

where P_{idle} is the probability of a channel being idle, and P_{fa_k} and P_{md_k} the false alarm and mis-detection probability, respectively, resulted from the cooperative sensing on the kth channel. Equivalently, $P(\mathcal{O}_k(0) = busy) = 1 - P(\mathcal{O}_k(0) = idle)$. After some algebra, the pdf of τ_k is given by (3.3).

Appendix 3.2 Proof of Lemma 2

The parameter X is defined as the number of successive periods (i.e., $X-1$ complete periods and one incomplete period) operating exclusively on licensed channels. Thus, the probability of having x successive periods may be expressed as $P(X = x) = P(T = T_{CP})^{x-1} P(T < T_{CP})$, where T is the time that the SN operates on licensed channels, with a maximum duration of T_{CP} ($0 \leq T \leq T_{CP}$) and $P(T = T_{CP})$, $P(T < T_{CP})$ the probabilities of having a complete and an incomplete period, respectively, with $P(T = T_{CP}) = 1 - P(T < T_{CP})$ and $P(T < T_{CP})$ is given by

$$P(T < T_{CP}) = \sum_{n=1}^{M} P(n_i = n) P(T < T_{CP} | n_i = n) \quad (3.24)$$

$$P(n_i = n) = \begin{cases} \prod_{k=1}^{M}(1 - P_{s_{idle\,k}}), n = 0 \\ \frac{1}{n}\sum_{k=1}^{n}(-1)^{k-1}P(n_i = n-k)\sum_{j=1}^{M}(\frac{P_{s_{idle\,j}}}{1-P_{s_{idle\,j}}})^k, n > 0 \end{cases} \quad (3.25)$$

$$P_{md_k} = 1 - \sum_{i=1}^{l_k} P(l_d = i) = 1 - \sum_{i=1}^{l_k}\left(\frac{1}{i}\sum_{m=1}^{i}(-1)^{m-1}P(l_d = i-m)\sum_{j=1}^{l_k}\left(\frac{1-P_{md_{k\,j}}}{P_{md_{k\,j}}}\right)^m\right) \quad (3.26)$$

$$P_{fa_k} = \sum_{i=1}^{l_k} P(l_{fa} = i) = \sum_{i=1}^{l_k}\left(\frac{1}{i}\sum_{m=1}^{i}(-1)^{m-1}P(l_{fa} = i-m)\sum_{j=1}^{l_k}\left(\frac{P_{fa_{k\,j}}}{1-P_{fa_{k\,j}}}\right)^m\right) \quad (3.27)$$

where M denotes the total number of licensed channels that are sensed, and n_i only the number of channels detected idle (i.e., n_i=$|\mathcal{B}|$). According to the algorithm description, the random variable n_i follows a Poisson binomial distribution and, thus, its probability mass function is given by (3.25), where $P_{s_{idle\,k}} = (1 - P_{fa_k})P_{idle} + P_{md_k}(1 - P_{idle})$ is the probability that the kth licensed channel is sensed idle, with P_{idle} the probability of a licensed channel being idle, and P_{md_k}, P_{fa_k} the total probabilities of mis-detection and false alarm of the kth channel, respectively. When the OR fusion rule is used to combine the individual sensing reports, P_{md_k} and P_{fa_k} are given by (3.26) and (3.27), respectively, where l_k denotes the number of cooperating SUs that sense the kth channel, l_d and l_{fa} are random variables that represent the number of users (with a maximum of l_k) that correctly detect the PU activity in the kth channel or that cause a false alarm, respectively and $P_{md_{k\,j}}, P_{fa_{k\,j}}$ denote the probability of mis-detection and false alarm of user j in the kth channel [5]. Notice that all these parameters depend on the applied sensing channel assignment algorithm, while similar to n_i, l_d and l_{fa} follow a Poisson binomial distribution. Even though a wide range of hard decision fusion rules for cooperative sensing have been proposed in the literature, here the OR rule is considered. As previously mentioned, the OR rule is the most conservative fusion rule, and consequently its application diminishes the mis-detection probability while increases the false alarm probability. Despite the fact that some other proposals in the literature could achieve better trade-off between false alarm and mis-detection probabilities, the OR rule minimizes the probability of interfering the incumbents of the primary channel. This is the reason why this fusion rule has been chosen. Yet, the analysis presented in this work holds regardless of the applied fusion rule. The selection of an alternative fusion rule would solely result in different expressions in (3.26) and (3.27).

Given that there are n licensed channels sensed as idle, an incomplete period takes place when all the n channels become busy before T_S. In other words, when the SN operates for less than $T_S - S_{n-1}$ on the nth licensed channel (i.e., the last channel of the list), with S_{n-1} denoting the total time spent in the previous $n-1$ channels. Thus, $P(T < T_{CP}|n_i = n)$ is given by (3.28), where f_{τ_n} is the pdf of τ_n

$$P(T < T_{CP}|n_i = n) = P(\tau_n \leq T_S - S_{n-1}) = \int_0^{T_S} f_{\tau_1}(\tau_1) \cdots \int_0^{T_S - S_{n-2}} f_{\tau_{n-1}}(\tau_{n-1}) \int_0^{T_S - S_{n-1}} f_{\tau_n}(\tau_n) d\tau_n d\tau_{n-1} \cdots d\tau_1 \quad (3.28)$$

$$\mathbf{E}[\tau_k|n_i = n] = \int_0^{T_S} f_{\tau_1}(\tau_1) \cdots \int_0^{T_S - S_{k-2}} f_{\tau_{k-1}}(\tau_{k-1}) \int_{T_S - S_{k-1}}^{\infty} (T_S - S_{k-1}) f_{\tau_k}(\tau_k) d\tau_k d\tau_{k-1} \cdots d\tau_1 + \cdots + \int_0^{T_S} f_{\tau_1}(\tau_1) \cdots \int_0^{T_S - S_{k-1}} \tau_i f_{\tau_k}(\tau_{k-1}) \cdots \int_{T_S - S_{n-1}}^{\infty} f_{\tau_n}(\tau_n) d\tau_n d\tau_{n-1} \cdots d\tau_1 \quad (3.31)$$

$$\mathbf{E}[\tau'_k|n_i = n] = \int_0^{T_S} f_{\tau_1}(\tau_1) \cdots \int_0^{T_S - S_{k-1}} \tau_k f_{\tau_k}(\tau_{k-1}) \cdots \int_0^{T_S - S_{n-1}} f_{\tau_n}(\tau_n) d\tau_n d\tau_{n-1} \cdots d\tau_1 \quad (3.32)$$

(i.e., operation time on the nth visited channel). The expected operation time on the kth channel, when having x successive periods, equals

$$\mathbf{E}[T_k] = \sum_{x=0}^{\infty} P(X = x)((x-1)\mathbf{E}[T_k]_{CP} + \mathbf{E}[T_k]_{INP}) \quad (3.29)$$

Then, $\mathbf{E}[T_k]_{CP}$ can be expressed as

$$\mathbf{E}[T_k]_{CP} = \sum_{n=1}^{M} P(n_i = n)\mathbf{E}[\tau_k|n_i = n] \quad (3.30)$$

where $\mathbf{E}[\tau_k|n_i = n]$ is the expected operation time on the kth visited channel, when having a complete period and there are n licensed channels sensed idle, which can be expressed as in (3.31). The expression of $\mathbf{E}[T_k]_{INP}$ is analogous to (3.30). However, as the distribution of τ_k differs for incomplete periods, for $\mathbf{E}[T_k]_{INP}$, $\mathbf{E}[\tau_k|n_i = n]$ should be replaced by $\mathbf{E}[\tau'_k|n_i = n]$, which is given by (3.32).

$$\bar{X}_k = \frac{1}{P_{c_k}}[Np(1-p)^{N-1}\sum_{i=1}^{\bar{N}_k}\binom{\bar{N}_k}{i}p^i(1-p)^{\bar{N}_k-i} + \sum_{j=2}^{N} j\binom{N}{j}p^j(1-p)^{N-j}\sum_{l=0}^{\bar{N}_k}\binom{\bar{N}_k}{l}p^l(1-p)^{\bar{N}_k-l}] \quad (3.34)$$

Appendix 3.3 Calculation of traffic-dependent parameters for saturated conditions

As proved in [68], the traffic-dependent parameters under saturated conditions are given by $P_{s_k} = Np(1-p)^{\bar{N}_k+N-1}$ and $P'_{s_k} = \bar{N}_k p(1-p)^{\bar{N}_k+N-1}$, where N is the number of users of the SN, $\bar{N}_k$ the average number of SUs that belong to other SNs and operate on the kth channel and p the probability that an SU transmits in a randomly chosen slot time. The probability of an idle slot is given by $P_{i_k} = 1 - P_{tr_k}$, where $P_{tr_k} = 1-(1-p)^{\bar{N}_k+N}$ denotes the probability that there is at least one transmission in the considered slot time. Then, the probabilities of collision are calculated by subtracting from the total collision probability derived in [68], the probability of a collision in which only the $\bar{N}_k$ SUs are involved for the calculation of P_{c_k}, and the probability of a collision in which only the N SUs are involved for the calculation of P'_{c_k}.

$$P_{c_k} = P_{c_{k\,tot}} - (1-p)^N \sum_{j=2}^{\bar{N}_k}\binom{\bar{N}_k}{j}p^j(1-p)^{\bar{N}_k-j} \quad (3.33)$$

Note that all probabilities that concern the unlicensed channel operation are given by the equivalent ones for the licensed channels by substituting $\bar{N}_k = N_{unlic}$ and $N = 1$. Finally, $\bar{Y}_k$ is 1 for centralized networks or different (e.g., equal to $\bar{X}_k$) for ad hoc networks, while $\bar{X}_k$ is given by (3.34).

Appendix 3.4 Proof of Lemma 3

As the process in the unlicensed channel will be repeated until at least one channel is sensed idle, the expected time spent in the unlicensed channel can be expressed as

$$\begin{aligned}\mathbf{E}[T_U] &= \sum_{i=0}^{\infty}(1-P_{U_s})^i P_{U_s}(i\mathbf{E}[T_{U_{un}}] + \mathbf{E}[T_{U_s}]) \\ &= \mathbf{E}[T_{U_{un}}](\frac{1}{P_{U_s}} - 1) + \mathbf{E}[T_{U_s}] \quad (3.35)\end{aligned}$$

where $P_{U_s} = 1 - \prod_{n=1}^{M}(1 - P_{s_{idle_n}})$ is the probability that there is at least one licensed channel sensed idle, and $T_{U_{un}}$, T_{U_s} denote the time spent when there is no channel sensed idle and when there is at least one channel sensed idle, respectively. Regarding

$\mathbf{E}[T_{U_{un}}]$, it may be claimed, by inspecting Fig. 3.2, that $\mathbf{E}[T_{U_{un}}] = \mathbf{E}[t_{ph1}] + t_{ph2} + t_{ph3}$, where $\mathbf{E}[t_{ph1}]$ is the expected duration of ph1 and t_{ph2}, t_{ph3} the ph2, ph3 durations in the unlicensed channel. It is worth noting that, the difference between $\mathbf{E}[T_{U_{un}}]$ and $\mathbf{E}[T_{U_s}]$ lies in the fact that the latter includes the additional time to switch to a licensed channel at the end of the process and thus $\mathbf{E}[T_{U_s}] = \mathbf{E}[T_{U_{un}}] + t_{sw}$.

The expected duration of t_{ph1} can be split up into the contention time to gain access to the unlicensed channel, T_{cont}, and the constant time T_{tr_1} that includes the time required to transmit the RFS and ACK packets, and the DIFS and SIFS waiting times. Therefore, $\mathbf{E}[t_{ph1}] = \mathbf{E}[T_{cont}] + T_{tr_1}$, where $T_{tr_1} = t_{RFS} + t_{SIFS} + t_{ACK} + t_{DIFS}$. Then, the contention time can in turn be expressed as $\mathbf{E}[T_{cont}] = \mathbf{E}[N_{unsuc}]\mathbf{E}[T_{unsuc}]$, where $\mathbf{E}[N_{unsuc}]$ denotes the expected number of unsuccessful slots until the cluster head gains access to the unlicensed channel and $\mathbf{E}[T_{unsuc}]$ the expected duration of an unsuccessful slot. Thus, $\mathbf{E}[N_{unsuc}]$ is equal to

$$\mathbf{E}[N_{unsuc}] = \sum_{i=0}^{\infty} i(1-P_s)^i P_s = \frac{1}{P_s} - 1 \tag{3.36}$$

where P_s is the probability of having a successful transmission by the cluster head in the unlicensed channel. Then, $\mathbf{E}[T_{unsuc}]$ is given by

$$\mathbf{E}[T_{unsuc}] = \frac{P_i\sigma + P'_s T_{rs} + (P_c + P'_c)T_c}{1 - P_s} \tag{3.37}$$

where P_i is the probability of having an idle slot in the unlicensed channel; P_c the probability of a collision slot of the SN under study in the unlicensed channel; P'_s and P'_c the probabilities of a successful transmission and collision slot, respectively, of the other SNs in the unlicensed channel; and σ, T_{rs} and T_c are the durations of an idle, a successful transmission and a collision slot. Finally, $t_{ph2} = M_{us}t_{sn} + (M_{us}+1)t_{sw}$ and $t_{ph3} = t_B + t_{SR}(N-1) + t_{LIST} + t_{SIFS}N + t_{DIFS}$, where M_{us} is the number of licensed channels sensed by each SU, t_{sn} the time to sense a channel, and t_B, t_{SR}, t_{LIST} the time required for beacon, report and list packets transmission, respectively.

Appendix 3.5 Calculation of $\mathbf{E}[T_{rsn}]$

Similarly to the calculation of the time devoted to data transmissions, the time devoted to detect the resumption of the PUs' activity and switch to another channel may be written as

$$\mathbf{E}[T_{rsn}] = \sum_{x=0}^{\infty} P(X = x)((x-1)\mathbf{E}[T_{rsn}]_{CP} + \mathbf{E}[T_{rsn}]_{INP}) \tag{3.38}$$

where $\mathbf{E}[T_{rsn}]_{CP}$, $\mathbf{E}[T_{rsn}]_{INP}$ denote the expected time spent in the reaction periods and the sensing procedures in a complete and an incomplete period, respectively.

$$P_k = \int_0^{T_S} f_{\tau_1}(\tau_1) \cdots \int_0^{T_S - S_{k-2}} f_{\tau_{k-1}}(\tau_{k-1}) \int_{T_S - S_{k-1}}^{\infty} f_{\tau_k}(\tau_k) d\tau_k d\tau_{k-1} \cdots d\tau_1 \tag{3.40}$$

$$\mathbf{E}[E_{unsuc}] = \frac{N\boldsymbol{P}_{idle}(P_i\sigma + P_s'T_{rs} + P_c'T_c) + P_c(t_{data}(\boldsymbol{P}_{T_x} + (N-1)\boldsymbol{P}_{R_x}) + N\boldsymbol{P}_{idle}t_{DIFS})}{1 - P_s} \tag{3.42}$$

Then, $\mathbf{E}[T_{rsn}]_{CP}$ can be expressed as

$$\mathbf{E}[T_{rsn}]_{CP} = \sum_{k \in \mathcal{B}} P_k \mathbf{E}[T_{sn_k}] + \sum_{k \in \mathcal{B}} P_k(k-1)\delta + t_{sw} \tag{3.39}$$

where P_k is the probability of having a sensing period in the kth channel, namely the probability of having a complete period in the kth channel, M is the number of sensed channels, n_i is the number of channels sensed idle, and $\mathbf{E}[T_{sn_k}]$ the expected duration of the sensing procedure in the kth licensed channel. P_k is given by (3.40) and $\mathbf{E}[T_{sn_k}] = \mathbf{E}[t_{ph1_k}] + t_{ph2} + t_{ph3}$, where $\mathbf{E}[t_{ph1_k}]$ is calculated as $\mathbf{E}[t_{ph1}]$ but taking into account the respective probabilities of the kth channel. Finally, $\mathbf{E}[T_{rsn}]_{INP} = \sum_{n=1}^{M} P(n_i = n) n\delta$.

Appendix 3.6 Proof of Lemma 4

Following the rationale of (3.35), $\mathbf{E}[E_U]$ can be expressed as

$$\mathbf{E}[E_U] = E_{U_{un}}\left(\frac{1}{P_{U_s}} - 1\right) + E_{U_s} \tag{3.41}$$

where $E_{U_s} = \mathbf{E}[E_{ph1}] + E_{ph2} + E_{ph3} + N\boldsymbol{P}_{sw}t_{sw}$ is the energy consumption in the unlicensed channel when there is at least one licensed channel sensed idle and $E_{U_{un}} = \mathbf{E}[E_{ph1}] + E_{ph2} + E_{ph3}$ otherwise. $\boldsymbol{P}_{sw}$ is the power to switch channel and $\mathbf{E}[E_{ph1}]$, E_{ph2}, E_{ph3} the expected energy consumed in t_{ph1}, t_{ph2} and t_{ph3}, respectively. Then, $\mathbf{E}[E_{ph1}] = \mathbf{E}[N_{unsuc}]\mathbf{E}[E_{unsuc}] + E_{tr_1}$, where $\mathbf{E}[N_{unsuc}]$ calculated in (3.36) and $\mathbf{E}[E_{unsuc}]$ is the expected energy consumed in an unsuccessful slot, while E_{tr_1} the one during T_{tr_1}. $\mathbf{E}[E_{unsuc}]$ is given by (3.42), where $\boldsymbol{P}_{T_x}$, $\boldsymbol{P}_{R_x}$ and $\boldsymbol{P}_{idle}$ are the transmission, reception and idle powers. Then, $E_{tr_1} = t_{RFS}(\boldsymbol{P}_{T_x} + (N-1)\boldsymbol{P}_{R_x}) + t_{ACK}(\boldsymbol{P}_{T_x} + \boldsymbol{P}_{R_x} + (N-2)\boldsymbol{P}_{idle}) + N\boldsymbol{P}_{idle}(t_{DIFS} + t_{SIFS})$. Finally, the energy consumptions during the time periods t_{ph2} and t_{ph3} are equal to $E_{ph2} = N(M_{us}\boldsymbol{P}_{sn}t_{sn} + (M_{us}+1)\boldsymbol{P}_{sw}t_{sw})$, with $\boldsymbol{P}_{sn}$ denoting the sensing power,

and $E_{ph3} = (\boldsymbol{P}_{T_x} + (N-1)\boldsymbol{P}_{R_x})(t_B + t_{LIST}) + t_{SR}(N-1)(\boldsymbol{P}_{T_x} + \boldsymbol{P}_{R_x} + (N-2)\boldsymbol{P}_{idle}) + N\boldsymbol{P}_{idle}(t_{DIFS} + t_{SIFS}N)$.

Appendix 3.7 Calculation of $\mathbf{E}[E_{rsn}]$ and $\mathbf{E}[E_{cont_k}]$

The expected energy consumption in the reaction periods and sensing procedures equals to

$$\mathbf{E}[E_{rsn}] = \sum_{x=0}^{\infty} P(X=x)((x-1)\mathbf{E}[E_{rsn}]_{CP} + \mathbf{E}[E_{rsn}]_{INP}) \tag{3.43}$$

where $\mathbf{E}[E_{rsn}]_{CP}$, $\mathbf{E}[E_{rsn}]_{INP}$ are the expected energy consumptions both in the reaction periods and the sensing procedures in a complete and an incomplete period, respectively, and equal to

$$\mathbf{E}[E_{rsn}]_{CP} = \sum_{k\in\mathcal{B}} P_k \mathbf{E}[E_{sn_k}] + \sum_{k\in\mathcal{B}} P_k(k-1)E_\delta + E_{sw} \tag{3.44}$$

$$\mathbf{E}[E_{rsn}]_{INP} = \sum_{n=1}^{M} P(n_i = n) n E_\delta \tag{3.45}$$

where $E_\delta = N(\boldsymbol{P}_{R_x} t_r + \boldsymbol{P}_{sw} t_{sw})$ and $\mathbf{E}[E_{sn_k}] = \mathbf{E}[E_{ph1_k}] + E_{ph2} + E_{ph3_k}$ are the energy consumptions during δ and the sensing procedures in the licensed channels, respectively. $\mathbf{E}[E_{ph1_k}]$ is calculated as $\mathbf{E}[E_{ph1}]$ but taking into account the respective probabilities of the kth channel. Then, $\mathbf{E}[E_{cont_k}]$ may be expressed as $\mathbf{E}[E_{cont_k}] = \mathbf{E}[E_{s_k}] + \mathbf{E}[E_{c_k}] + \mathbf{E}[E_{i_k}]$, where $\mathbf{E}[E_{s_k}] = \beta P_{s_k} E_s$, $\mathbf{E}[E_{c_k}] = \beta P_{c_k} E_c$ and $\mathbf{E}[E_{i_k}] = \beta N \boldsymbol{P}_{idle}(P_{i_k}\sigma + P'_{s_k} T_{rs} + P'_{c_k} T_c)$ are the expected energies consumed in the successful transmissions, collisions or idle slots of the SN on the kth channel, with $\beta = \mathbf{E}[T_k]/\mathbf{E}[T_{slot}]$ and E_s, E_c the energy consumed in a successful transmission, collision slot of the SN, respectively, and are given by $E_s = (t_{data} + t_{ACK})(\boldsymbol{P}_{T_x} + \boldsymbol{P}_{R_x} + (N-2)\boldsymbol{P}_{idle}) + N\boldsymbol{P}_{idle}(t_{DIFS} + t_{SIFS})$ and $E_c = t_{data}(\bar{X_k}\boldsymbol{P}_{T_x} + \bar{Y_k}\boldsymbol{P}_{R_x} + (N - \bar{X_k} - \bar{Y_k})\boldsymbol{P}_{idle}) + N\boldsymbol{P}_{idle} t_{DIFS}$, where $\bar{X_k}$, $\bar{Y_k}$ are the average numbers of SUs of the SN that are involved in a collision in transmission and reception mode, respectively.

Chapter 4

Energy-efficient Context-aware User Association in Cognitive Heterogeneous Networks

4.1 Introduction

This chapter includes the second research contribution part of this thesis which focuses on the proposal of energy-efficient user association algorithms and analytical frameworks for cognitive heterogeneous networks (HetNets).

As discussed in Chapter 2, the mobile data traffic is expected to grow significantly during the next few years, which results in an urgent need for mobile operators to maintain capacity growth. Serving more traffic leads to increased energy consumption, and therefore, how to minimize the energy consumption becomes also important. In parallel, the spectrum scarcity problem stresses the need for spectral efficient solutions. The aforementioned goals can be summarized into the joint maximization of spectrum and energy efficiency, which constitutes a fundamental design objective for next generation cellular networks.

To that end, the dense deployment of small cells (SCs), overlaying the existing macrocell networks, is a promising solution. The deployment of SCs reduces the distance between user equipments (UEs) and base stations (BSs)[1] and, consequently, i) the area spectral efficiency (bps/Hz/m^2) increases, and ii) the energy consumption in the access network (AN), i.e., the links between the UEs and their serving BSs, decreases. Hence, dense deployment of SCs is expected during the next years, with

[1]In this thesis, we will use the term BS to refer to a macrocell BS and/or a SC BS (i.e., an eNodeB (eNB) and/or a Home eNB in LTE-Advanced, respectively).

SC radius being eventually of the order of 50 meters [36].

However, the dense deployment of SCs also poses new challenges. Due to the high number of deployed SCs, the direct connection of all SCs to the core network becomes complicated. Fiber connections, which have been traditionally considered as the best backhaul (BH) solution, are prohibitive in this case due to their high deployment cost [79]. A promising solution lies in exploiting the existing connection between the macrocell site and the core network (most of the times it is a fiber connection), and to provide core network connectivity to SCs through the macrocell site [40, 41, 80]. Still, in order to connect the SCs to the macrocell site (thus providing them core network connectivity), new cost-efficient wireless BH solutions are required.

In addition, this wireless BH is expected to provide high-capacity services from the SCs to the core network, in order to meet the expected traffic demands of the order of Gbps [36]. Therefore, a promising solution that could provide wireless BH connectivity between the SCs and the core network lies in using millimeter wave (mmWave) frequencies, because of the large amount of available bandwidth at these frequencies, which results in high capacity connections [79]. It has been shown, however, that mmWave frequencies are capable of providing good coverage performance only if the transmission distance is shorter than 200 meters [36, 40, 81]. Otherwise, links may not be established. Since the macrocell radius is even in dense deployments of the order of 500 meters, this implies that a multi-hop architecture is needed, in order to allow each of the SCs to reach the macrocell site (i.e., macrocell aggregation gateway) [39–41].

In this context, user association becomes challenging due to the multi-hop BH architecture [82] and therefore new optimal solutions need to be developed aiming at the joint energy and spectrum efficiency maximization of the network. To that end, in this chapter, the following contributions are provided:

- The role of BH in future outdoor HetNets is studied aiming to answer to whether or not it could constitute an energy bottleneck for the HetNet. In particular, the BH energy impact is compared to the access network (AN), i.e., the links between the users and their serving cells, under different traffic distribution scenarios and BH technologies.

- The user association problem is studied aiming at the joint maximization of energy efficiency and spectral efficiency of the network, without compromising the user equipment (UE) quality of service (QoS) requirements. In this framework, the following additional contributions are provided.

  - A heuristic algorithm is proposed that exploits context-aware information (i.e., UE measurements and requirements, the HetNet architecture knowledge and the available spectrum resources of each base station (BS)) to associate the UEs in an energy-efficient way, while considering both the AN and BH energy consumption. The performance of the proposed algorithm is studied under two study-case scenarios and it is

proved that it achieves significantly higher energy efficiency than the reference algorithms, while maintaining high spectral efficiency.

- Moreover, the joint uplink (UL) and downlink (DL) user association problem is studied. To that end, the previously proposed context-aware algorithm is adapted accordingly to associate the UEs in an energy-efficient way, while considering the energy consumption both in UL and DL. The proposed joint UL-DL algorithm performance is evaluated and it is shown to achieve significantly higher energy efficiency than the reference approaches, while maintaining high spectral efficiency and low UE power consumption.
- The aforementioned problem is formulated as a generalized assignment problem, which is shown to be NP-hard. Therefore, by relaxing the capacity constraints, an upper bound on the network performance is derived which can be used as a benchmark for the performance evaluation of user association algorithms. An enhanced heuristic user association algorithm is also proposed, which considers the total power consumption (AN and BH) for the traffic of a UE to be served. Notice that this association metric relaxes the assumption of homogeneous BH links made in our previous proposal, by considering the actual power consumption of each BH link and not just the number of hops. Finally, the derived bound and the proposed enhanced algorithm are compared with existing user association solutions in scenarios that employ mmWave BH links. In the provided results, the proposed algorithm is shown to achieve notable performance gains compared to its counterparts, while achieving near-optimal performance.
- Aiming at the analytical study of the trade-off between energy and spectrum efficiency, the aforementioned problem is formulated as an ε-constraint problem. The optimal Pareto front solutions of the problem are analytically derived for different BH technologies and insights are gained into the energy and spectrum efficiency trade-off. The proposed optimal solutions, despite their high complexity, can be also used as a benchmark for the performance evaluation of practical user association algorithms. Furthermore, an adaptive heuristic algorithm is proposed, which is compared with reference solutions under different traffic scenarios and BH technologies. The provided results motivate the use of mmWave BH, while the proposed algorithm is shown to achieve near-optimal performance. In particular, the proposed algorithm was shown to be able to select any point of the Pareto front, by accordingly tuning one parameter, and thus, to achieve a *good* trade-off between the aforementioned metrics.

The rest of the chapter is organized as follows: In Section 4.2, the system setup is described and useful parameters are derived. Moreover, the user association algorithms that will be used as reference approaches throughout this chapter are presented and analyzed. In Section 4.3, the BH energy impact is studied in outdoor

SC scenarios with multi-hop BH architecture. The AN and BH power consumption are compared and light is shed to whether or not BH could become an energy bottleneck for the network under different traffic distribution scenarios and BH technologies. In Section 4.4, the UE association problem is studied and a cognitive BH-aware heuristic algorithm is proposed that exploits the available context-aware information to associate the UEs in an energy-efficient way. In the same setup, the joint UL-DL user association problem is also studied with the proposed algorithm being accordingly adapted to take both UL-DL network conditions into account. In Section 4.5, analytical frameworks are proposed that can be used as benchmark solutions for the performance evaluation of user association algorithms. In particular, Section 4.5.2 focuses on the generalized assignment problem formulation, which aims at energy efficiency maximization given that the maximum spectral efficiency is achieved. The trade-off between energy efficiency and spectrum efficiency is then analytically studied in Section 4.5.3, where an ε-constraint problem formulation is followed. Finally, concluding remarks are given in Section 4.6.

4.2 System Setup

In this section, the system model that will be employed throughout this chapter is presented. Without loss of generality and in accordance with the scenarios proposed by 3GPP [84], an eNB sector is considered, which is overlaid with multiple SCs. In particular, a set of BSs is considered, denoted by $\mathcal{C}$, which includes one eNB (j=0) and $C-1$ SCs (j=1...C-1), with C representing the cardinality of the set $\mathcal{C}$. The SCs are divided in N_{cl} clusters (k=1...N_{cl}), as depicted in Fig. 4.1, with SC_k denoting the number of SCs in cluster k [84]. The downlink is studied, unless otherwise stated, and the following assumptions are made:

- Each SC is connected to the core network through the eNB aggregation gateway either directly or through one or more SC aggregation gateways [39–41, 80, 82].
- There is a fiber link from the eNB site to the core network and a set of out-of-band wireless BH links $\mathcal{L}$=$\{\mathcal{L}_1, \mathcal{L}_2, ...\mathcal{L}_l, ...\mathcal{L}_{C-1}\}$. Each wireless BH link l is represented by a set $\mathcal{L}_l$ that includes all cells j that backhaul their traffic through it (i.e., $\forall j \in \mathcal{L}_l$).
- The total transmission power of each BS is assumed to be equally distributed among its spectrum resources, i.e., among its subcarriers in case of LTE [39]. This constitutes a common literature assumption [43], which is also used in practical LTE implementations [85].
- Slow fading channels are considered due to shadowing.
- A set of $\mathcal{N}$ UEs (i=1,...,N) is considered with strict guaranteed bit rate (GBR) QoS requirements, denoted as r_i, based on their service/application [86].
- Each UE can be associated only with one BS at a time.

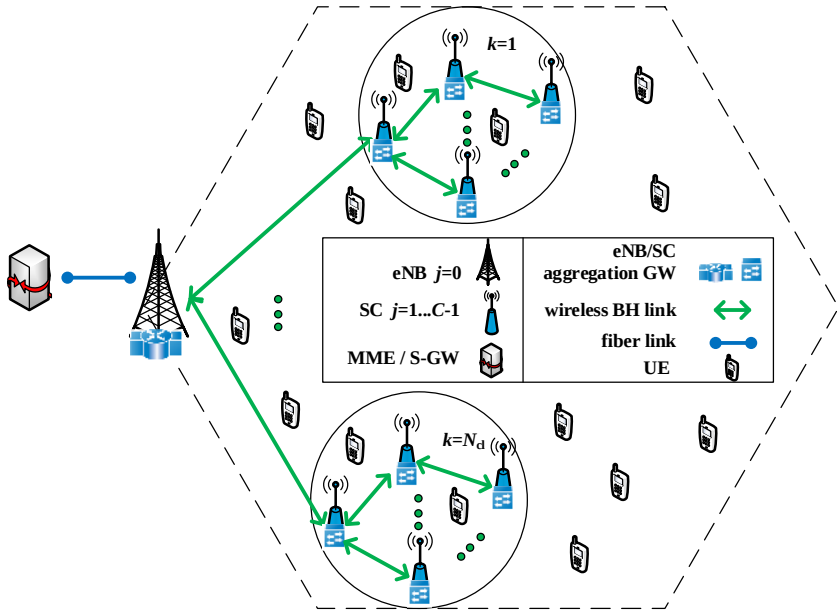


Figure 4.1: System model.

- There is a maximum number of spectrum resource units available to each BS j, i.e., physical resource blocks (PRBs)[2], denoted by $c_{j_{max}}$.

For reader's convenience, in the following a table with the main notation used throughout the chapter is included. Moreover, the most important parameters involved in the total network energy efficiency calculation are derived. In particular, the signal-to-interference-plus-noise ratio (SINR) calculation is given in Section 4.2.1, while the power consumption models for both the AN and the BH are provided in Section 4.2.2.

4.2.1 SINR Calculation

The signal-to-noise ratio (SNR) received by UE i from BS j is given by [87]

$$\begin{aligned} SNR_{ij\,(dB)} = P_{j_c\,(dBm)} + G_{T_{x\,j}\,(dBi)} - L_{cb_j\,(dB)} \\ -L_{p_{ij}\,(dB)} - L_{f_{ij}\,(dB)} - N_{th\,(dBm)} - NF_{(dB)} \end{aligned} \tag{4.1}$$

with $P_{j_c} = 10log_{10}(P_{j_{max}}/(N_{ant_j} c_{j_{max}}))$ standing for the power allocated by BS j to a PRB, where $P_{j_{max}}$ is its maximum transmission power in mW, N_{ant_j} is the number of antennas of BS j and $c_{j_{max}}$ is the maximum number of PRBs allocated to it. The parameter $G_{T_{x_j}}$ represents the antenna gain of BS j and L_{cb_j} is the cable

[2]Please note that 1 PRB is equal to 12 subcarriers in the frequency domain and 0.5 ms in the time domain [39].

Table 4.1: Context-aware User Association in Cognitive HetNets: Notation Table

Parameter	Meaning
$\mathcal{C}$	set of BSs
$\mathcal{N}$	set of UEs
$\mathcal{L}$	set of BH links
$\mathcal{L}_l$	BH link l / set of SCs whose BH traffic passes through BH link l
C	total number of BSs
N	total number of UEs
N_{cl}	number of clusters in the eNB sector area
SC_k	number of SCs in cluster k
j	index used for BSs, with $j = 0$ for eNB and $j \neq 0$ for SCs
i	index used for UEs
l	index used for BH links
k	index used for clusters
w	index used to define the type of link, i.e., UL or DL
r_i	throughput QoS requirement of UE i
f_x	frequency used by x
B_x	bandwidth allocated to x
$c_{j_{max}}$	maximum number of PRBs available at BS j
$P_{j_{max}}$	maximum transmission power of BS j
P_{j_c}	power allocated to a PRB by BS j
$L_{p_{ij}}$	path loss between UE i and BS j
$L_{p_{\mathcal{L}_l}}$	path loss at distance $d_{\mathcal{L}_l}$, equal to the length of the BH link $\mathcal{L}_l$
L_{p_o}	path loss at distance equal to 1 m
L_{cb_j}	cable loss between RF connector and antenna of BS j
h_x	height of the antenna of x
C_H	antenna height correction factor
NF	noise figure
N_{th}	thermal noise
$G_{T_{xj}}$, $G_{T_{x\mathcal{L}_l}}$	transmitter antenna gain of BS j, of the BH link $\mathcal{L}_l$
$G_{R_{xj}}$, $G_{R_{x\mathcal{L}_l}}$	receiver antenna gain of BS j, of the BH link $\mathcal{L}_l$
N_{ant_j}	number of antennas of BS j
R	eNB radius
r	hotspot radius
b	bandwidth of a PRB
$f(SINR_{ij})$	spectral efficiency, as a function f(·) of $SINR_{ij}$
$\lceil \cdot \rceil$	ceiling function operator
$P_{AN_{ij}}$	power consumption of the AN link between UE i and BS j
$P_{BH_{\mathcal{L}_l}}$	power consumption of the BH link $\mathcal{L}_l$
$P_{BH_{max}}$	maximum transmit power of a BH link transmitter
I_{ij}	total interference experienced by UE i, when associated with SC j
N_{total}	total noise power
$SINR^{trg}_{\mathcal{L}_l}$	(minimum) target SINR needed for the aggregate traffic of BH link $\mathcal{L}_l$ to be successfully transmitted, assuming that link adaptation is employed
a_{ij}	association vector, equal to 1 when the UE i is associated with BS j and 0 otherwise
$\alpha_{\mathcal{L}_l}$	link budget parameter of the BH link $\mathcal{L}_l$, derived by subtracting from the total losses the transmitter and receiver antenna gains of the BH link $\mathcal{L}_l$
λ	operating wavelength of the signal
IL	implementation loss
t_{TTI}	transmission time interval, system observation time (i.e., one subframe time)
N_{hops}	number of hops (i.e., number of BH links) until the traffic of the SC reaches the eNB site
p_{ij}	total profit when UE i is associated with BS j
$s_{\mathcal{L}_l j}$	BH architecture parameter, equal to 1 if the traffic of BS j passes through the BH link $\mathcal{L}_l$ and 0 otherwise
β_1, β_2, β_3	normalized weighting coefficients
$\mathcal{F}$	Pareto front / set of Pareto front points
$Candidates_i$	candidate cells for a UE i , i.e., subset of cells selected for the association of UE i
$P_{tot_{ij}}$	total power consumption required for the traffic of UE i to be served, when the UE is associated with BS j

loss between the radio frequency (RF) connector and the antenna. The path loss between UE i and BS j is denoted by $L_{p_{ij}}$, while $L_{f_{ij}}$ represents the losses due to shadowing. Finally, N_{th} stands for the thermal noise and NF is the noise figure. Then, the SINR received by UE i from BS j is expressed as

$$SINR_{ij_{(dB)}} = SNR_{ij_{(dB)}} - 10log_{10}\left(\frac{I_{ij(mW)}}{N_{total(mW)}} + 1\right) \tag{4.2}$$

where I_{ij} is the total interference experienced by UE i, when associated with BS j, which strongly depends on the applied frequency allocation scheme. The proposed works in this thesis take as an input the SINR of UEs, and thus, they can be applied regardless of the employed channel allocation scheme. This is due to the fact that a different channel allocation scheme would solely impact the generated interference (I_{ij}) and consequently the received SINR of UEs. Still, although it is out of the scope of this thesis, the combination of the proposed works with a sophisticated channel allocation solution could further improve the system performance. Finally, the parameter $N_{total} = 10^{(N_{th\,(dBm)} + NF_{(dB)})/10}$ denotes the total noise power in mW experienced by UE i.

4.2.2 Power Consumption Models

The total power consumption of the network can be divided into the one consumed in the AN links and the one consumed in the BH links. Since this chapter is focused on user association, only the transmit power consumption of BSs and BH links is taken into account, as the fixed power consumption part is independent of the user association decision[3]. To that end, the employed power consumption models are presented in the following.

Access link

The power consumption of the AN link between BS j and UE i is given by

$$\begin{aligned} P_{AN_{ij}(W)} &= P_{j_c(W)} c_{ij} \\ &= \left(\frac{P_{j_{max}}}{N_{ant_j} c_{j_{max}}}\right) \left\lceil \frac{r_i}{f(SINR_{ij})} \right\rceil \\ &= \left(\frac{P_{j_{max}}}{N_{ant_j} c_{j_{max}}}\right) \left\lceil \frac{r_i}{b \log_2 (1 + SINR_{ij})} \right\rceil \end{aligned} \tag{4.3}$$

where P_{j_c} is the power allocated by BS j to a PRB and c_{ij} represents the number

[3]Please note that the inclusion of the fixed power component (i.e., power supply and cooling, baseband unit, etc. [88]) would impact all the algorithms equally by shifting all the curves in the same way.

of PRBs needed for the association of UE i with BS j. As already mentioned, N_{ant_j} is the number of antennas of BS j, r_i is for the throughput demand of UE i, while $f(SINR_{ij})$ represents the spectral efficiency and $\lceil \cdot \rceil$ is the ceiling function operator. In general, the spectral efficiency can be expressed as a function f($\cdot$) of the effective $SINR_{ij}$ [89], which maps the SINR received by UE i from BS j to a corresponding rate in bps. Although in LTE, f($\cdot$) is a scalar function with each step corresponding to a specific modulation and coding scheme (MCS) [42], for analytical reasons, the function depicted in (4.3) is traditionally used, which is derived by Shannon's theorem [90] and represents the maximum rate that can be achieved with $SINR_{ij}$ and bandwidth of a PRB equal to b.

Backhaul link

The power consumed in a wireless BH link $\mathcal{L}_l$ can be given by [91]

$$P_{BH_{\mathcal{L}_l\,(dBm)}} = SINR^{trg}_{\mathcal{L}_{l\,(dB)}} + \alpha_{\mathcal{L}_l} \tag{4.4}$$

where $SINR^{trg}_{\mathcal{L}_l}$ corresponds to the (minimum) target SINR that is needed so that the aggregate traffic of BH link $\mathcal{L}_l$ is successfully transmitted, assuming that link adaptation is employed [87]. $SINR^{trg}_{\mathcal{L}_l}$ may be given by [90]

$$SINR^{trg}_{\mathcal{L}_{l\,(dB)}} = 10log_{10}\left(2^{\frac{\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{L}_l} r_i a_{ij}}{B_{\mathcal{L}_l}}} - 1\right) \tag{4.5}$$

where $B_{\mathcal{L}_l}$ denotes the bandwidth of the BH link $\mathcal{L}_l$ and $\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{L}_l} r_i a_{ij}$ is the aggregate traffic that passes through the BH link in bps, with a_{ij} denoting the association vector (equal to 1 when the UE i is associated with BS j and 0 otherwise).

The parameter $\alpha_{\mathcal{L}_l}$ of (4.4) depends on the link budget equation of the BH link $\mathcal{L}_l$. This parameter is derived by subtracting from the total losses (e.g., path loss) the transmitter and receiver antenna gains of the BH link. For instance, in the case of a mmWave BH link, the parameter $\alpha_{\mathcal{L}_l}$ can be given by [92]

$$\begin{aligned}\alpha_{\mathcal{L}_l} = {} & L_{p_{o\,(dB)}} + L_{p_{\mathcal{L}_{l\,(dB)}}} + IL_{(dB)} + N_{th\,(dBm)} \\ & - NF_{(dB)} - G_{T_x\mathcal{L}_{l\,(dBi)}} - G_{R_x\mathcal{L}_{l\,(dBi)}}\end{aligned} \tag{4.6}$$

where L_{p_o} denotes the path loss at distance equal to 1 m and $L_{p_{\mathcal{L}_l}} = 20log_{10}(4\pi\frac{d_{\mathcal{L}_l}}{\lambda})$ is the path loss at distance $d_{\mathcal{L}_l}$ equal to the length of the link (i.e., distance between transmitter and receiver). Finally, λ is the operating wavelength of the signal (e.g., for frequency equal to 60 GHz, $\lambda = 0.005$ m) and IL stands for the implementation loss that may account for e.g., distortion, intermodulation and/or phase noise.

Energy efficiency

The network energy efficiency can be expressed as the total number of successfully transmitted bits by all UEs divided by the total energy consumption (i.e., the sum of the energy consumed in the AN and BH links). Equivalently, it can be expressed as the total throughput of all UEs in bps ($\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{C}} r_i a_{ij}$), divided by the total power consumption in Watts, i.e., the sum of the AN power consumption ($\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{C}} P_{AN_{ij}} a_{ij}$) and the BH power consumption ($\sum_{\mathcal{L}_l\in\mathcal{L}} P_{BH_{\mathcal{L}_l}}$), and thus can be formulated as

$$EE_{(bits/Joule)} = \frac{\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{C}} r_i a_{ij\,(bps)}}{\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{C}} P_{AN_{ij}} a_{ij\,(W)} + \sum_{\mathcal{L}_l\in\mathcal{L}} P_{BH_{\mathcal{L}_l\,(W)}}} \tag{4.7}$$

4.2.3 State-of-the-art User Association Algorithms

During the last few years, the UE association problem has received a lot of research attention. To that end, the most important user association algorithms may be summarized in the following ones:

Reference signal received power (RSRP)

In LTE-Advanced (LTE-A), the user association is based on the reference signal received power (RSRP) and/or reference signal received quality (RSRQ). The first measures the average received power over the resource elements that carry cell-specific reference signals within certain frequency bandwidth, while the latter measures the portion of pure reference signal power over the total power received by the UE [39]. Although these criteria maximize the SINR of UEs [44], simulations and field trials have shown that they do not increase the overall throughput as much as hoped, because many SCs typically have few active UEs [45]. This stems from the fact that the maximum transmit power of a SC is much lower with respect to the eNB, resulting to a smaller downlink SC coverage. However, this is not the case for the uplink, where the UE transmission range is the same for all UEs. The algorithm presents high downlink spectrum efficiency, as the UE gets associated with the BS from which it receives the strongest signal. However, its energy efficiency cannot be guaranteed, since it only considers the radio AN.

Range expansion (RE)

RSRP presents relatively poor overall throughput performance, since most UEs get connected to the eNB, whereas many SCs have few active UEs [45]. To that end, range expansion (RE) was proposed, which adds a *bias* (in dB) to the RSRP in the case the signal comes from a SC, thus actively pushing UEs onto SCs [43, 45, 83, 93–99]. Despite a potentially significant SINR hit for that UE, this has

the potential for a win-win situation, because the UE gains access to a much larger portion of resources, while the eNB reclaims the ones that would have been allocated to it. RE presents lower downlink spectrum efficiency than RSRP, especially in cases an aggressive bias is used, as a UE is connected to a BS that does not provide the highest SINR. However, at the same time, this also results in RE mitigating cross-tier interference in the uplink. For maximum downlink performance, RE should be combined with sophisticated interference mitigation techniques. To that end, enhanced inter-cell interference coordination (eICIC) in release 10 of LTE-A introduced the almost blank subframes (ABS), during which the eNB remains silent (i.e., only transmitting control information at very low power), thus enabling full SC range expansion, since the UE communication with their associated SCs does not experience interference caused by the eNB [39]. Finally, equivalently to RSRP, the high energy efficiency of RE cannot be guaranteed.

Minimum path loss (MPL)

A UE is connected to the BS from which it has minimum path loss (MPL), independently of its received power [43]. In ideal environments with no fading, the algorithm would connect the UE to its closest BS. In hotspot traffic distribution scenarios, MPL achieves maximum traffic offloading to SCs, since most UEs will be associated with SCs being in their proximity. Moreover, it achieves much lower spectrum efficiency than the maximum, as SINR is not taken into account in the user association decision. Although MPL minimizes the UE power consumption in uplink, it fails to guarantee high network energy efficiency, due to high BH energy consumption.

4.3 Energy Impact of Outdoor Small Cell Backhaul in Green Heterogeneous Networks

4.3.1 Introduction and Related Work

In [100, 101], Tombaz *et al.* stress the need for jointly considering the BH and AN energy consumption when studying the energy efficiency of future HetNets. In [100], the impact of indoor femtocell deployment on the energy efficiency of wireless networks is investigated, taking into account the BH power consumption. In [101], the BH impact on the total network power consumption is studied under different BH architectures and capacity requirements. It is shown that BH can potentially become an issue in dense SC HetNets, since it can amount to up to 50% of the total network power consumption. However, the BH architecture scenarios and technologies of the aforementioned works concerned exclusively indoor scenarios. Nevertheless, further study is needed for outdoor scenarios, since the outdoor environmental characteristics impose new challenges (e.g., dependency on weather conditions, larger distances, many interference sources), and thus different types of BH solutions may be required (e.g., home digital subscriber (DSL) lines are not available).

Thus, in this section, the BH energy impact is studied in outdoor scenarios where several SCs backhaul their traffic to the neighboring SCs until they reach the core network. The relation between AN and BH power consumption is investigated and light is shed to whether or not BH could become an energy bottleneck for the network. Finally, to gain valuable insights on the topic, extensive simulation results under different traffic distribution scenarios and BH technologies are provided.

4.3.2 Wireless Backhaul Solutions

In this section, the different BH solutions that can be employed for backhauling future outdoor HetNets are analyzed. This thesis is focused on wireless BH solutions, since wired solutions are more difficult to implement under the considered scenarios (e.g., although fiber can provide very high capacity, its deployment cost is high and it can take many years to deploy). In particular, in accordance with Small Cell Forum (see Fig. 5-2 in [36]), the following main BH solutions are considered: mmWave, microwave, and sub-6GHz band.

Millimeter wave (mmWave)

The mmWave frequencies (60, 70-80 GHz, also known as the V- and E-band, respectively) [92] is expected to be one of the main BH solutions especially for scenarios as the ones previously described with multiple hops. Due to the high path loss at these frequencies, high gain antennas with narrow beams are required. Therefore, mmWave frequencies enable only short range point-to-point (PTP) line-of-sight (LOS) radio links and the connection to the local aggregation gateway would most probably require a number of hops. On the other hand, the high path loss and narrow beams reduce the risk of interfering with other mmWave radio links.

At the same time, the huge amount of available bandwidth at high frequencies can provide significant capacity enhancement. Lower frequency technologies rely on complex RF techniques to deliver higher capacities, such as multi-path propagation and channel aggregation below 6 GHz, or spatial multiplexing at higher frequencies up to 42 GHz. In contrast, mmWave technologies rely on generous availability of wideband RF-channels to deliver Gbps of throughput using simple single-channel configurations, which gives mmWave a potential for cost-per-bit advantage for high capacity BH [102].

Microwave

Microwave (6-60 GHz) is widely used, mostly because of its low deployment cost. Microwave technologies support link throughputs of multiple Mbps. By exploiting the Shannon capacity formula, the link capacity (C) of the microwave communication channel can be given by

$$C_{(Mbps)} = B_{BH\,(MHz)} \;\; SE_{(bps/Hz)} \tag{4.8}$$

where B_{BH} stands for the channel bandwidth and SE for the spectral efficiency. Microwave channel sizes are typically multiples of 7 MHz, while modern-day microwave equipment can transmit over multiple adjacent channels of up to 56 MHz. With modern transmission methods, the capacity of a microwave radio link can be boosted up to 1 Gbps [36]. Depending on the frequency and link budget, microwave frequencies allow a coverage range from a few hundreds of meters to a few kilometers. Similarly with mmWave frequencies, the microwave frequencies require narrow beams to cope with the path loss. However, the beams are wider than mmWave beams, requiring less accurate antenna alignment. While the mmWave frequencies are suitable for densely placed BSs, microwave frequencies are more suitable for long distance rooftop-to-rooftop connections due to less available spectrum and lower path loss.

Sub-6GHz band

Sub-6GHz spectrum is able to operate in harsh mobile non-line-of-sight (NLOS) propagation environments. Therefore, it is considered suitable as a BH solution for small cells deployed in locations where high capacity NLOS connectivity is required. The available spectrum at this band is generally ranging from 5 to 20 MHz, while sub-6GHz frequencies can achieve NLOS coverage of some kilometers.

4.3.3 Simulation Results

In the extensive simulations in MATLAB®, the system setup, described in Section 4.2, was employed. Two clusters were considered ($N_{cl} = 2$), and 5 SCs in total (i.e., $SC_1 = 3$ and $SC_2 = 2$), as depicted in Fig 4.2. The employed BH technologies were:

i) LOS mmWave links (f_{BH1}= 60 GHz band) of B_{BH1}= 100 MHz channel bandwidth [92],

ii) LOS microwave links (f_{BH2}= 28 GHz) of B_{BH2}= 28 MHz [103],

iii) sub-6GHz (f_{BH3}= 3 GHz) of B_{BH3}= 10 MHz [104].

For a fair comparison, the path loss models of the provided references were used, while the antenna gains were selected equal to 20, 15 and 10 dBi, respectively.

The proposed work, as already explained, is independent of the employed channel allocation scheme. Therefore, for the sake of simplicity and without loss of generality, inter-sector interference is assumed to be mitigated through some form of fractional frequency reuse scheme or sophisticated frequency allocation [83]. Moreover, the channels allocated to the eNB are assumed to be orthogonal to the channels allocated to SCs, while SCs that are adequately far from each other may reuse the same bands.

In each realization, N UEs of different GBR requirements were considered. In particular, 60% of UEs demand 512 kbps, 30% 728 kbps and 10% 1024 kbps [86].

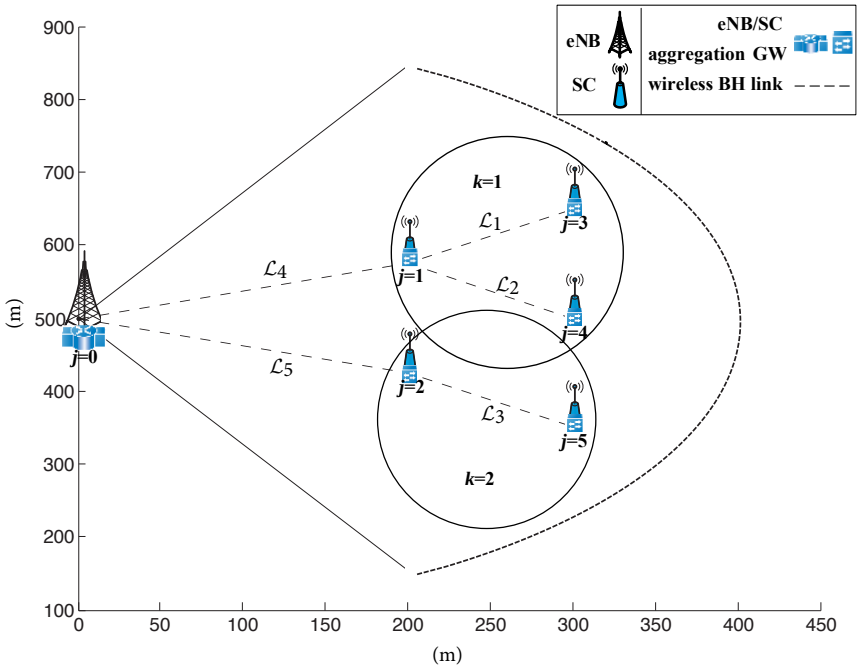


Figure 4.2: Simulation scenario with BH architecture 1.

Moreover, in accordance with the previous sections, the following two UE traffic distribution scenarios were employed: In each realization (1000 in total), N UEs of different throughput requirements were considered. Specifically, 10% of UEs demand 512 kbps, 10% 728 kbps and 80% 1024 kbps [86]. Moreover, two UE traffic distribution scenarios were considered:

- Uniform: the UEs are uniformly distributed in the sector area of radius R.
- Hotspot: the UEs form hotspots. In particular, 2/3 of the total traffic is generated in a radius r= 70 m from SC $j = 3$ and SC $j = 4$ according to 3GPP [84] and the rest is uniformly distributed in the sector area. Notice that this scenario is more realistic, as in future HetNets, UEs are expected to be very close to SCs and to generate bursty hotspot traffic [36].

All simulation parameters are summarized in Table 4.2, where f_{AN} denotes the frequency used in the AN. Moreover, h_{eNB} and h_{SC} denote the BS antenna height of the eNB ($j = 0$) and the SCs ($j \neq 0$), respectively. Furthermore, h_m denotes the mobile antenna height, while C_H stands for the antenna height correction factor and d for the distance between the BS and the UE. The slow fading is modeled by a log-normal random variable with zero mean and deviation 8 or 10 dB in the case the signal is transmitted by an eNB or a SC, respectively. Finally, in this section, the RSRP user association criterion is applied for all BH technologies and UE traffic

Table 4.2: BH Energy Impact: Simulation Values

Parameter	Value
f_{AN}	2.0 GHz
B_{eNB}, B_{SC}	10 MHz
$c_{eNB_{max}}, c_{SC_{max}}$	50
$P_{eNB_{max}}$	46 dBm
$P_{SC_{max}}$	30 dBm
$P_{BH_{max}}$	46 dBm
$L_{p_{eNB}}$	69.55+26.16 $\log f_{AN}$-13.82 $\log h_{eNB}$-C_H+(44.9- 6.55 $\log h_{eNB}$) $\log d$, d in km
$L_{p_{SC}}$	69.55+26.16 $\log f_{AN}$-13.82 $\log h_{SC}$+(44.9- 6.55 $\log h_{SC}$) $\log d$, d in km
C_H	0.8+ (1.1 $\log f_{AN}$- 0.7) h_m -1.56 $\log f_{AN}$
$L_{cb_{eNB}}$	2 dB
h_m	1.5 m
h_{eNB}	25 m
h_{SC}	2.5 m
L_{p_o}	57.5 dB
NF	9 dB
N_{th}	-174 dBm/Hz
$G_{T_{xeNB}}$	14 dBi
$G_{T_{xSC}}$	5 dBi
IL	9 dB
N_{eNB}, N_{SC}	2
R	450 m
r	70 m

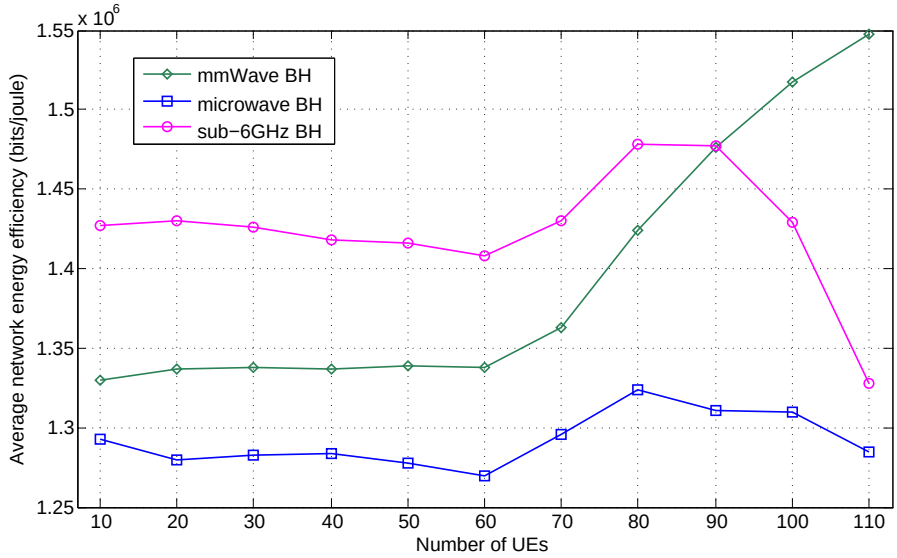


Figure 4.3: Average energy efficiency under different BH technologies for uniform traffic.

distribution scenarios.

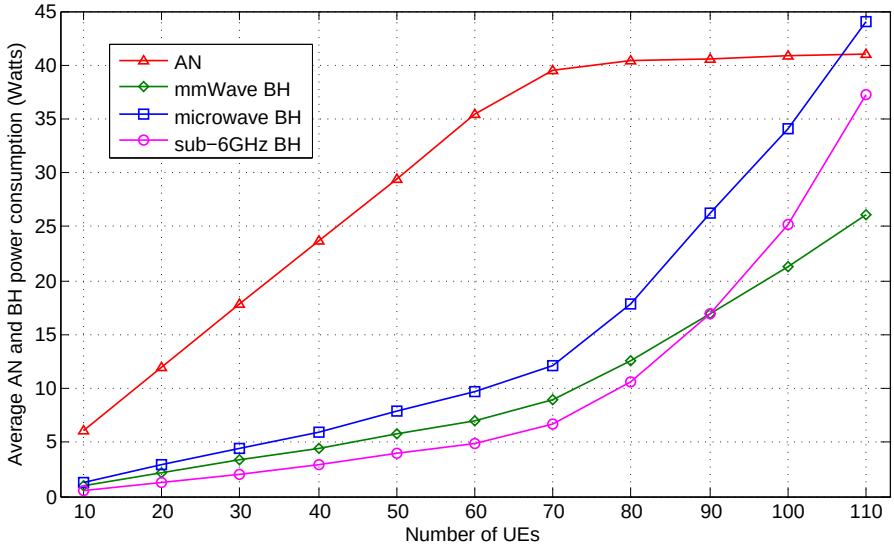


Figure 4.4: Average AN and BH power consumption under different BH technologies for uniform traffic.

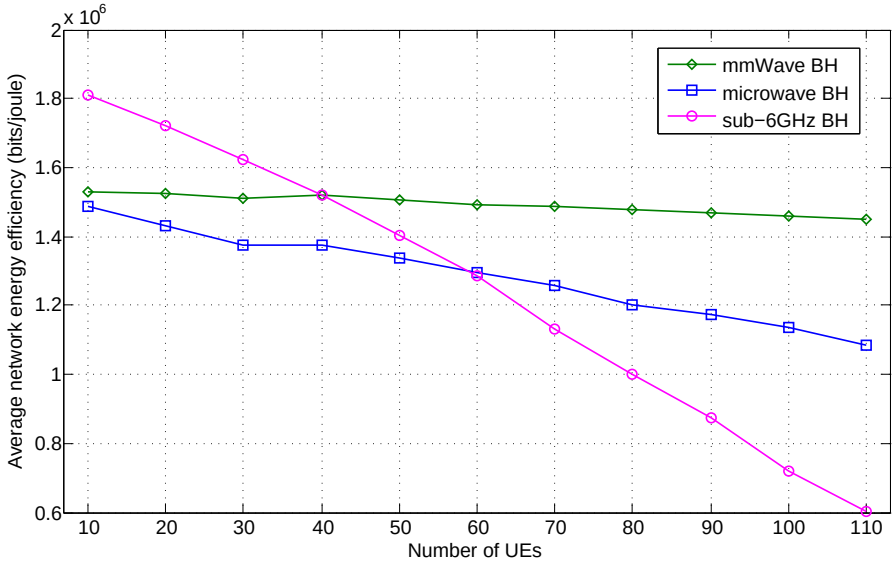


Figure 4.5: Average energy efficiency under different BH technologies for hotspot traffic.

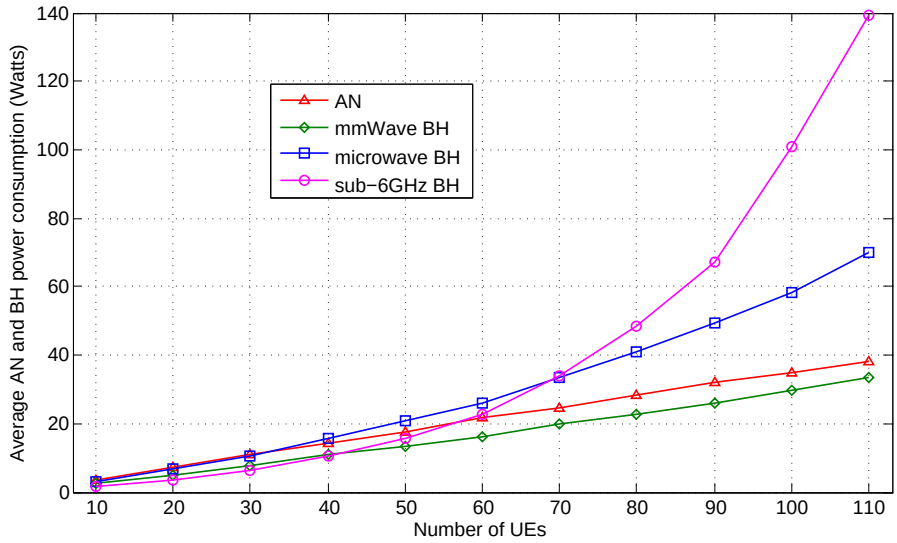


Figure 4.6: Average AN and BH power consumption under different BH technologies for hotspot traffic.

The range of the parameter N has been appropriately selected to avoid system overloading, and thus all UE throughput demands are satisfied (i.e., the UE QoS is guaranteed) in all experiments. Therefore, for a given number of UEs, the differences among the total network energy efficiency of each BH technology only depend on the BH energy consumption.

To that end, in Fig. 4.3 the average network energy efficiency is depicted for different BH technologies under uniform traffic. As it can be observed, for low number of UEs (i.e., low total traffic), the sub-6GHz BH solution presents the highest energy efficiency. This stems from the fact that sub-6GHz experiences the lowest path loss compared to the other technologies and thus for the same aggregate traffic to be sent, it involves lower power consumption (see also Fig. 4.4, where the BH power consumption compared to the AN is depicted for every BH technology under uniform traffic). The connection between path loss and power consumption is the following: for given aggregate rate demands, lower path loss leads to higher SINR, which enables the use of a higher order MCS, thus resulting in higher spectral efficiency (without any power consumption increase). On the contrary, higher path loss, leads to lower SINR, and therefore an increase in the transmitted power is necessary to increase the SINR at the receiver. Thereby, the spectral efficiency is improved at the expense of higher power consumption.

In Fig. 4.3, as the number of UEs increases, the total traffic increases, given that each UE has specific throughput demands. Consequently, the BH traffic also increases and thus the available bandwidth of each BH technology becomes very important. In particular, for given BH channel bandwidth and path loss, if the ag-

gregate traffic is very high and the available bandwidth is not sufficient, an increase in the transmitted power is necessary to increase the SINR at the receiver. Thereby, the spectral efficiency is again improved at the expense of lower energy efficiency. Hence, due to the high bandwidth availability of 60 GHz, mmWave shows the best performance for very high traffic (i.e., 90 UEs), since it is able to send high amount of data, for a given SINR and MCS, without increasing the transmitted power. On the contrary, the available bandwidth of sub-6GHz is very limited. Therefore, for high traffic (i.e., 90 UEs), a significant increase in the transmitted power of the BH links is required (see crosspoint between sub-6GHz and mmWave at 90 UEs in Fig. 4.4), so that the SINR at the receiver increases. Thereby, higher order MCS can be used, which results in higher spectral efficiency. However, as previously mentioned, due to the much higher transmit power consumption, this also results in lower energy efficiency (see Fig. 4.3).

Furthermore, as it is shown in Fig. 4.4, the AN is much more significant than the BH power consumption when the UEs are uniformly distributed in the sector area. This is due to the fact that i) in the uniform scenario more UEs receive stronger signal from the eNB and thus get connected to it and ii) the AN power consumption in the case a UE is connected to the eNB is much higher than when connected to a SC. Moreover, for uniform traffic the AN power consumption increases initially at a high rate, as the number of UEs increases, since more UEs get connected to the eNB. However, for very high traffic (more than 70 UEs) there are no available resources left in the eNB and thus more UEs get connected to SCs, which results in a smoother increase in the AN power consumption.

Equivalently, in Fig. 4.5, the average network energy efficiency is depicted for different BH technologies under hotspot traffic. In this case, notice that the crosspoint where the performance of sub-6GHz gets worse than mmWave occurs earlier (i.e., at 40 UEs) than in the uniform scenario. This is due to the fact that, in the hotspot scenario, most UEs get associated with SCs, and hence the BH traffic increases significantly, thus making the available bandwidth of each technology very important (even for low total traffic). Therefore (due to its very limited bandwidth), sub-6GHz shows the worst performance for high hotspot traffic, i.e., worse than mmWave for traffic higher than 40 UEs and worse than microwave for traffic higher than 60 UEs.

In order to gain further insights on whether or not BH could constitute an energy bottleneck, in Fig. 4.6, the BH power consumption compared to the AN is depicted for every BH technology under hotspot traffic. As it can be observed, in this case, the BH power consumption becomes higher than the AN. This stems from the fact that under hotspot traffic, more UEs receive higher SINR from SCs and thus get associated with them, thus leading to higher BH power consumption and lower AN power consumption (it is reminded that the AN power consumption in the case a UE is connected to the eNB is much higher than when connected to a SC).

Focusing on high traffic scenarios (i.e., 110 UEs), Fig. 4.7 depicts the total power consumption (the sum of AN and BH power consumption) for uniform and hotspot traffic, when different BH technologies are employed. It can be seen that the BH power consumption can reach up to 38%, 51% and 47% of the total power consump-

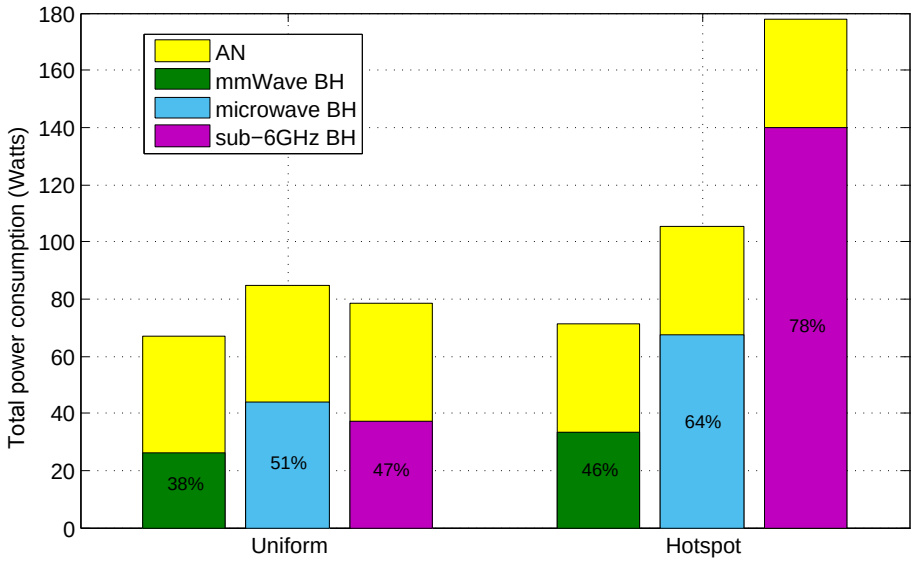


Figure 4.7: AN and BH power consumption under different BH technologies for high uniform and hotspot UE traffic (N=110).

tion for uniform traffic and 46%, 64% and 78% of the total power consumption for hotspot traffic, when mmWave, microwave and sub-6GHz are employed, respectively. In general, the following conclusions can be derived:

- The BH power consumption constitutes a significant part of the total power consumption, which becomes more important as the number of UEs increases and the traffic becomes hotspot.

  ⇒ Given that future HetNets are expected to deal with very high and mainly hotspot traffic [36], backhaul-aware algorithm design becomes essential for next generation wireless networks.

- mmWave seems the best choice to avoid a potential energy bottleneck mainly due to its very high bandwidth availability. However, given that it requires LOS, a mixture of the studied BH technologies is anticipated to be used in future HetNets.

4.4 Energy-efficient Context-aware User Association in Cognitive HetNets

4.4.1 Introduction and Related Work

All the user association approaches, previously described in Section 4.2.3, consider only the AN, thus totally overlooking any BH issues. Still, the BH, as already discussed in Section 4.3, is about to become the most challenging part for future SC networks, especially when considering scenarios as the ones previously described with wireless BH links. In such cases, considering only the best-received signal strength or/and the load of each cell when examining the user association problem is not sufficient. To that end, in [51], the authors model a backhaul-aware BS assignment problem as an optimization problem using a utility-based framework, imposing constraints on both radio and BH resources. The main idea behind their algorithm is to distribute traffic among BSs according to a load balancing strategy, considering both radio and BH load status. However, their algorithm, unlike the work presented in this thesis, reduces the BH congestion at the expense of lower spectral efficiency, since some UEs may be assigned to BSs not being their "best" radio choice. Moreover, it considers a simple system model consisting only of eNBs, whereas future cellular networks are about to be dense HetNet deployments of more sophisticated architecture, as previously described, thus posing new challenges in the user association problem. Finally, although the network energy efficiency is expected to play a key role, the algorithm ignores the energy consumption impact.

Thus, in this section, the UE association problem is studied in multi-hop BH scenarios. To that end, the aforementioned problem is formulated as an optimization problem, which is shown to be NP-hard. Therefore, a cognitive heuristic algorithm is proposed that exploits the available context-aware information to associate the UEs in an energy-efficient way. In particular, i) it maximizes the spectral efficiency by considering only a subset of cells as candidates for the association of the UE, i.e., the cells that satisfy the UE QoS requirements with the fewest PRBs needed, and ii) it maximizes the network energy efficiency, while exploiting the HetNet architecture cognition, by favoring the candidate cell with the fewest hops to reach the core network, or in case there are more candidate cells with the same number of hops, it favors the one with the least loaded BH route in order to achieve load balancing.

4.4.2 Problem Formulation

As previously mentioned, the network energy efficiency is selected as the global objective to be maximized, under the condition that the UE QoS requirements are satisfied. As already described in (4.7), the energy efficiency can be expressed as the total number of successfully transmitted bits by all UEs, $\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{C}} r_i a_{ij} t_{TTI}$, divided by the total energy consumption, $(\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{C}} P_{AN_{ij}} a_{ij} + \sum_{\mathcal{L}_l\in\mathcal{L}} P_{BH_{\mathcal{L}_l}}) t_{TTI}$, i.e., the sum of the energy consumed in the AN and in the BH links. Hence, the

aforementioned problem can be formulated as

$$\max_{a_{ij}} \quad \frac{\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{C}} r_i a_{ij} t_{TTI}}{(\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{C}} P_{AN_{ij}} a_{ij} + \sum_{\mathcal{L}_l\in\mathcal{L}} P_{BH_{\mathcal{L}_l}})t_{TTI}}$$

$$\text{s.t.} \quad \begin{aligned} &a)\ a_{ij} \in \{0,1\},\ \forall i \in \mathcal{N},\ \forall j \in \mathcal{C} \\ &b)\ \sum_{j\in\mathcal{C}} a_{ij} = 1, \forall i \in \mathcal{N} \\ &c)\ \sum_{i\in\mathcal{N}} a_{ij} c_{ij} \leq c_{j_{max}},\ \forall j \in \mathcal{C} \end{aligned} \tag{4.9}$$

where t_{TTI} is the system observation time (i.e., one subframe time). Notice that the energy efficiency is independent of the observation time for steady state behavior. Then, the parameter a_{ij} is the association vector that equals to 1 when user i is associated with BS j and 0 otherwise (4.9a). Each UE can be associated only with one BS at a time (4.9b). Furthermore, the total number of PRBs used by BS j cannot exceed the maximum number of the PRBs allocated to it (4.9c).

As it can be easily understood, the aforementioned problem is a non-convex problem that cannot be solved in polynomial time (i.e., it is NP-hard). Therefore, in the next section we propose and analyze a novel context-aware heuristic algorithm that associates the UEs aiming at maximizing the network energy efficiency, while having low computational complexity.

4.4.3 Energy-efficient Context-aware Algorithm (EE)

All the aforementioned approaches consider only the AN, thus totally overlooking any BH issues. Still, the BH is expected to play a key role in next generation HetNets, especially when considering scenarios as the ones previously described with wireless BH links [36]. In such scenarios, the BH energy consumption can be responsible for a high percentage of the total energy consumption and consequently cannot be ignored.

Thus, an energy-efficient context-aware algorithm (EE) is proposed, summarized in Fig. 4.8, that takes as input the available context-aware information (i.e., the UEs' measurements and requirements, the HetNet architecture knowledge and the available spectrum resources of each BS) to associate the UEs aiming at maximizing the network energy efficiency. Note that this context-aware information can be readily available to all nodes in a LTE-A network, while introducing low overhead (e.g., by transmitting it during the almost blank subframes (ABS) or through the X2 interface between BSs (eNBs and/or SCs) [39]).

At the same time, EE achieves high spectrum efficiency, since only the BSs that require the minimum number of PRBs to satisfy the UE requirements (i.e., the BSs with the best received SINR) are included in the UE candidate cell set. To ensure

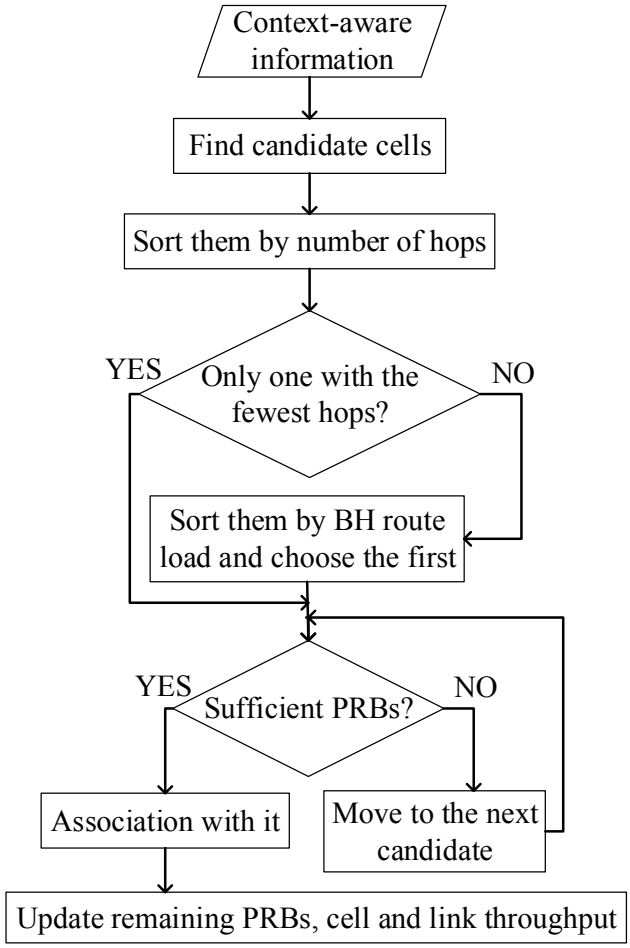


Figure 4.8: Energy-efficient context-aware algorithm (EE) flowchart.

that all the UEs will be associated, EE sorts the UEs by their number of candidates and starts with the ones with the fewest candidates. Then, to maximize the network energy efficiency, EE sorts the candidate cells by the number of hops until their traffic reaches the core network and it associates the UE to the candidate with the fewest hops, as long as it has sufficient PRBs to serve it. Otherwise, it moves to the next candidate. Every time a UE is associated with a BS, the algorithm updates the remaining PRBs of the BS, its cell throughput and the throughput that passes through the BH links that are used until its traffic reaches the core network. In the case there are more candidates with the same number of hops, EE associates the UE to the one with the least loaded BH route, as long as it has sufficient PRBs, to achieve load balancing at the BH links.

In terms of scalability, EE may be executed in each eNB sector at a specific time

Algorithm 1 Energy-efficient context-aware user association algorithm

Input: $\mathcal{N}, \mathcal{C}, SINR_{ij}, r_i, c_{j_{max}}, \mathcal{L}_l, \mathcal{L}$

Calculate c_{ij} from (4.3)
$Candidates_i \leftarrow j : min(c_{ij})$
Sort all UEs i by $Candidates_i$ size in ascending order
Calculate N_{hops} $\forall j \in Candidates_i$
Sort $Candidates_i$ by N_{hops} in ascending order
if $\exists$ only one candidate having the least N_{hops} **then**
    Choose this one
else
    Choose the one with the least loaded BH route
end if
if the chosen BS has sufficient PRBs **then**
    Associate the UE to it
    Update remaining PRBs, cell throughput and the throughput that passes through the link $\mathcal{L}_l \ni j$
else
    Move to the next candidate and repeat the process
end if

interval based on the dynamics of the UE traffic distribution, so that the system performance is optimized. For the new UEs that appear in the meantime, EE is executed as before given the associations of the rest of the UEs. However, in that case, the context-aware information includes the remaining PRBs and the throughput of each BS/cell, and the throughput that passes through each BH link, given the traffic of the already associated UEs. The proposed algorithm is summarized in Algorithm 1.

Simulation Results

In the extensive simulations, executed in MATLAB®, the same simulation scenario and parameters (unless otherwise stated), described in Section 4.3.3, were considered. However, in order to study the impact of the BH architecture on the algorithm performance, apart from the BH architecture 1 scenario, which is depicted in Fig. 4.2, one more BH architecture (BH architecture 2) was considered, which is depicted in Fig. 4.9. Notice that although the same BS deployment is considered in both scenarios, the BH architecture varies significantly (i.e., SCs $j = 3$ and $j = 5$ are connected through SC $j = 4$ in the BH architecture 2 scenario). In general, this may happen due to deployment impairments (e.g., buildings), that impose restrictions on the actual BH link configurations.

The BH network consisted of line-of-sight (LOS) mmWave links (f_{BH}= 60 GHz band) of B_{BH}= 50 MHz channel bandwidth [92]. Moreover, the noise figure, NF, was equal to 6 dB and the mobile antenna height, h_m, was selected equal to 1.7 m. The transmitter/receiver antenna gain of a BH link is equal to $G_{T_{x\mathcal{L}_l}} = G_{R_{x\mathcal{L}_l}} = 15$

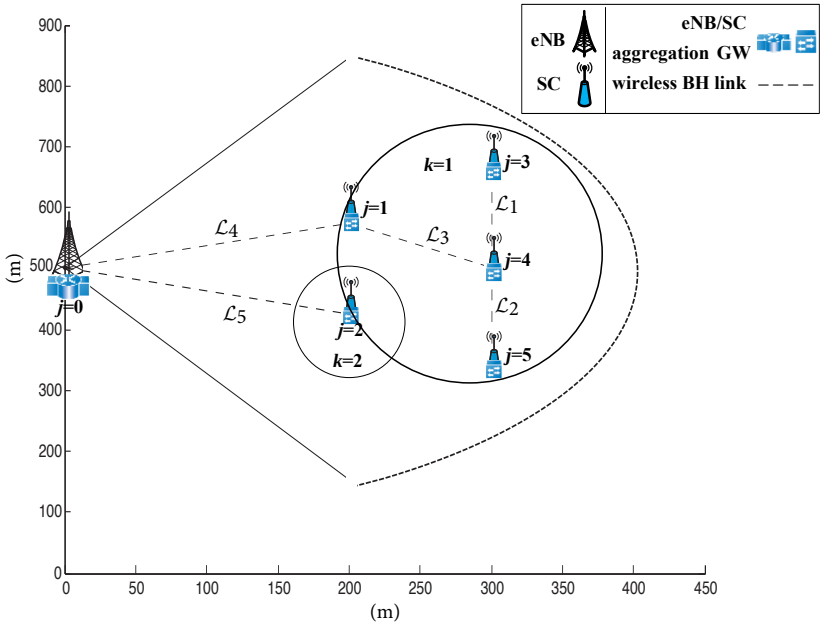


Figure 4.9: Simulation scenario with BH architecture 2.

dBi, whereas IL =7 dB.

In each realization, N fixed UEs of different throughput requirements were considered. Specifically, 80% of UEs demand 500 kbps, 10% 700 kbps and 10% 1 Mbps [86]. In accordance with section 4.3.3, two UE traffic distribution scenarios were considered:

- Uniform: the UEs are uniformly distributed in the sector area of radius R.
- Hotspot: the UEs form hotspots. In particular, 40% of the UEs are considered to be uniformly distributed in a radius $r = 60$ m from SC $j = 3$ and 40% in a radius $r = 60$ m from SC $j = 4$ according to 3GPP [84] and the rest is uniformly distributed in the sector area. It is reminded that this scenario is more realistic, as in future HetNets, UEs are expected to be very close to SCs and to generate bursty hotspot traffic [36].

The slow fading is modeled by a log-normal random variable with zero mean and deviation 8 dB. Furthermore, the association of the received SNR to the achievable spectral efficiency in bps/Hz is given by Table A.2 in [42], while the mmWave link budget equation is given by (4.6).

The reference algorithms, described in section 4.2.3, are considered, which are summarized in the following:

- RSRP, where a UE connects to the BS from which it receives the strongest

RS [39, 44]

$$j^* = argmax_{j \in \mathcal{C}}(RSRP_{ij}) \tag{4.10}$$

- Range expansion (RE), where a $bias = 6$ dB is added to the RSRP if the signal comes from a SC [45, 83, 93–99]

$$j^* = argmax_{j \in \mathcal{C}}(RSRP_{ij} + bias_j) \tag{4.11}$$

- Minimum path loss (MPL), where a UE connects to the BS from which it has the minimum path loss ($L_{ij} = L_{p_{ij}} + L_{f_{ij}}$) [43], independently of its received power

$$j^* = argmin_{j \in \mathcal{C}}(L_{ij}) \tag{4.12}$$

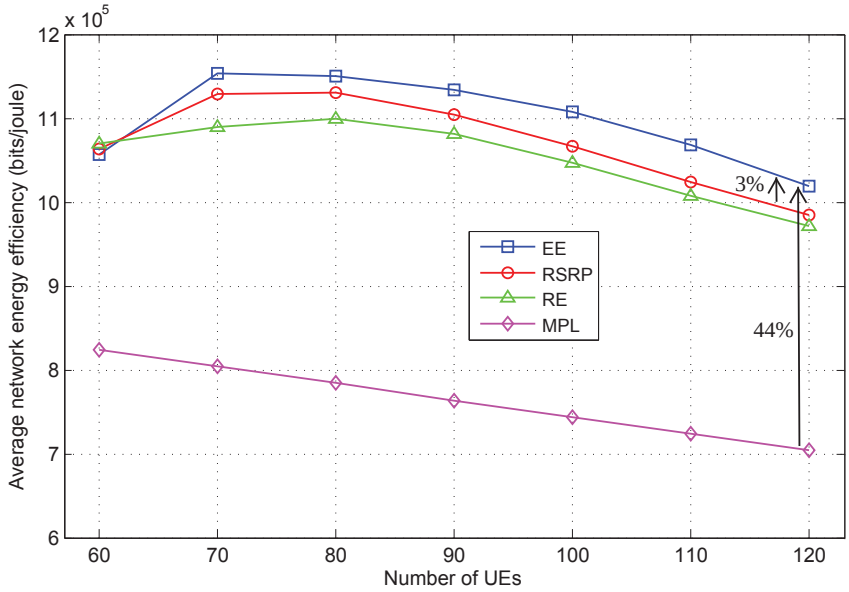
Maximizing the network energy efficiency becomes more challenging, as the number of UEs increases due to the capacity constraints. Therefore, in our experiments, we consider highly loaded scenarios consisting of more than 60 UEs.

In terms of throughput, all algorithms achieve the same performance in both scenarios (e.g., for N=100 UEs, the total network throughput is 57 Mbps), as all algorithms ensure that UEs' demands are satisfied.

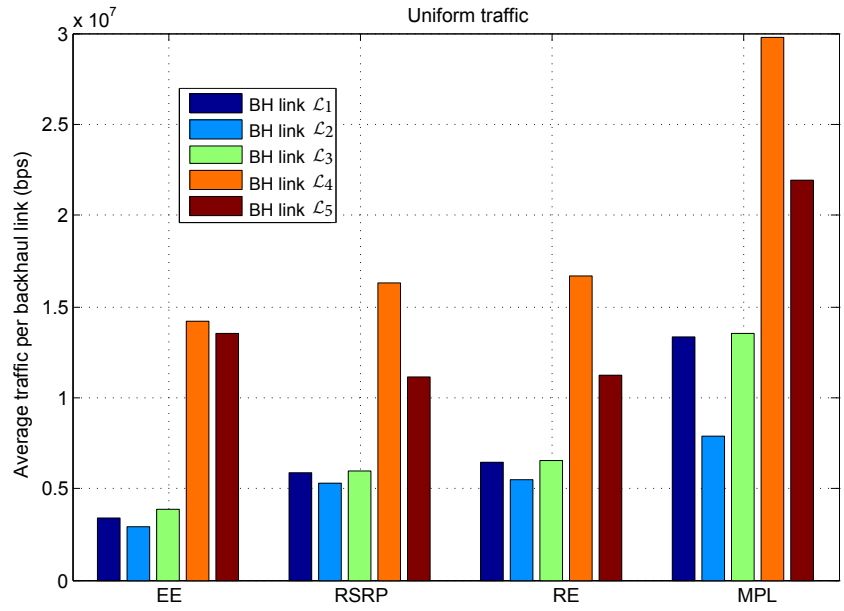
BH architecture 1

As far as the network energy efficiency of the uniform scenario is concerned, EE achieves slightly better performance than RSRP for values higher than N=70 UEs, as depicted in Fig. 4.10(a). This is due to the fact that EE gives priority to the candidate cell with the fewest hops and thus most UEs get connected to the eNB. In this case, the additional energy consumption experienced by EE in the Uu interface with respect to RSRP, is not compensated by the reduction of the BH energy consumption. However, for higher values, there are no available PRBs in the eNB at some point, and thus UEs are associated with the candidate SC with the fewest hops, which results in lower BH energy consumption compared to RSRP. This can be also noticed in Fig. 4.10(b), where the average traffic of each BH link of Fig. 4.2 is depicted, for N=100 UEs. According to it, EE achieves better load balancing among the BH links that are the same number of hops away from the core network (e.g., among BH links $\mathcal{L}_4$ and $\mathcal{L}_5$) than the other algorithms. Notice that BH load balancing achieves better utilization of BH resources, decreases the possibility that a BH link becomes the bottleneck and offers higher security in the case of a BH link failure. Then, regarding RE, it achieves lower energy efficiency, since there are more UEs associated with SCs, resulting in higher BH energy consumption. Finally, MPL can be considered as an aggressive RE algorithm, where it is likely for the UEs to be closer to a SC and thus to be associated to it. Although MPL achieves maximum traffic offloading to SCs, it has very poor energy efficiency performance, as the BH traffic (see Fig. 4.10(b)) and thus the BH energy consumption is much higher than the other approaches.

In Fig. 4.11(a), the average network energy efficiency of all algorithms is depicted, when hotspot traffic is considered. In this case, EE achieves gains up to

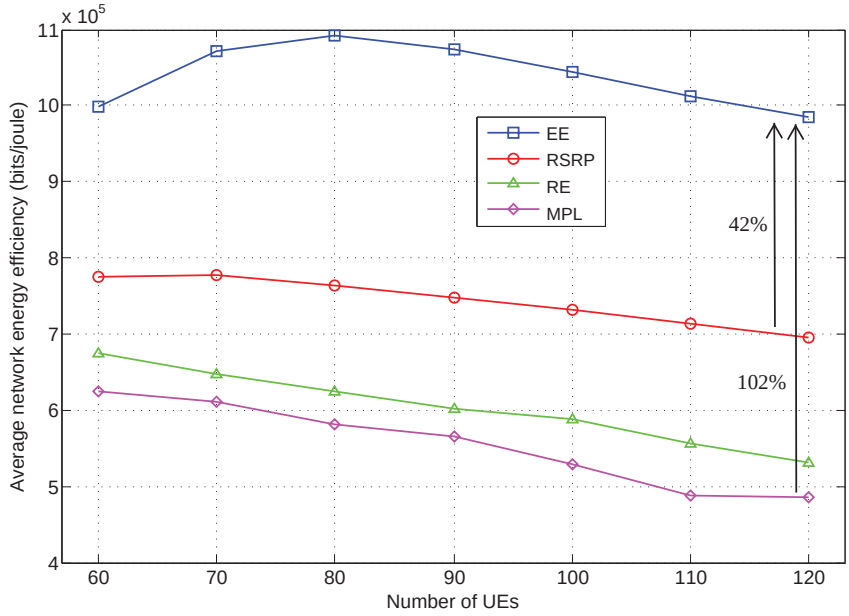


(a) Average network energy efficiency for different number of users (N), when the UEs are uniformly distributed in the sector area.

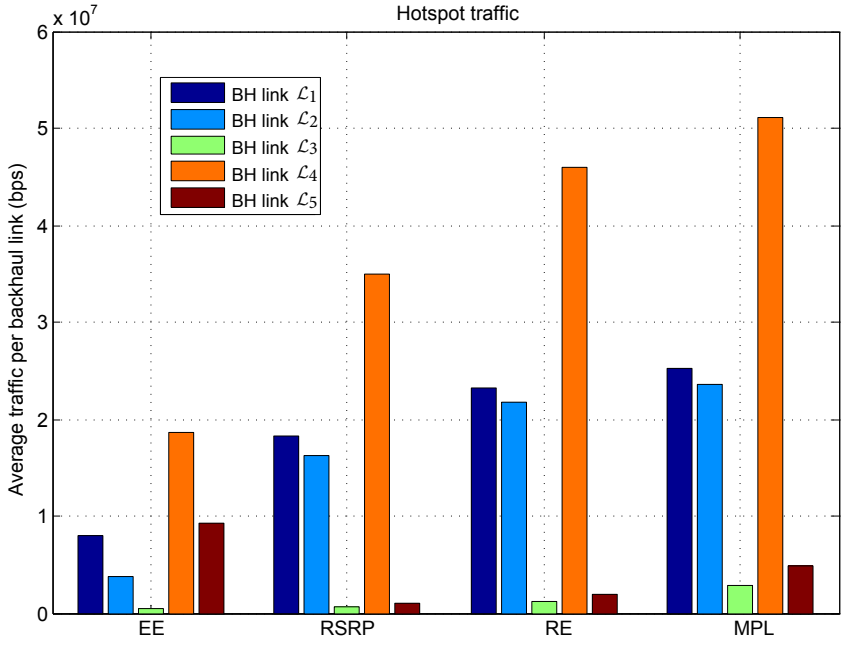


(b) Average traffic of each backhaul link for uniform traffic and N=100 UEs.

Figure 4.10: Performance evaluation results for the simulation scenario with BH architecture 1, depicted in Fig. 4.2, and uniform traffic.



(a) Average network energy efficiency for different number of users (N), for hotspot traffic.



(b) Average traffic of each backhaul link for hotspot traffic and N=100 UEs.

Figure 4.11: Performance evaluation results for the simulation scenario with BH architecture 1, depicted in Fig. 4.2, and hotspot traffic.

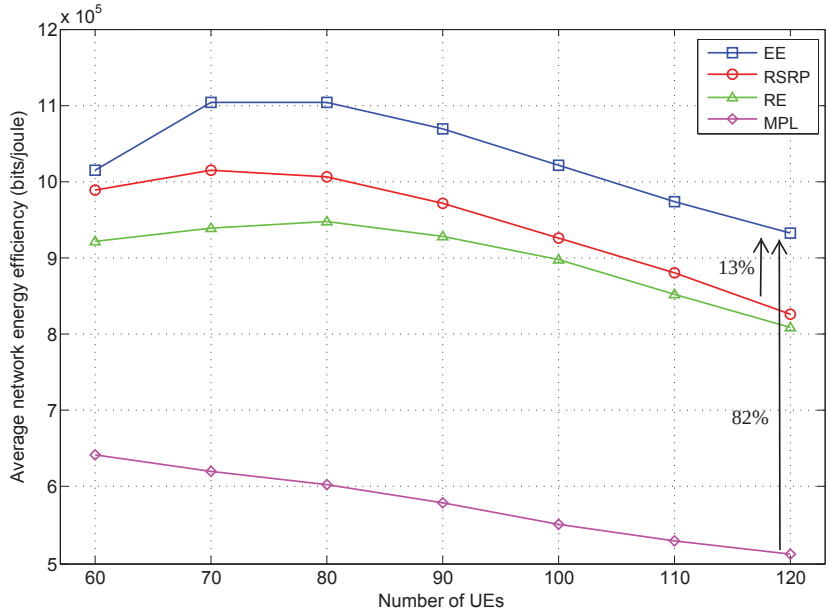
100%, since it manages the BH traffic so that lower energy consumption is generated. Specifically, in comparison to RSRP, EE achieves up to 40% energy efficiency gain. This stems from the fact that the UEs located around SC $j = 3$ or $j = 4$, that have as candidate cells the SCs $j = 1$ or $j = 2$, will be associated with SC $j = 3$ or $j = 4$ when RSRP is applied, given that they receive higher SINR, whereas to SC 1 or 2 with EE, given that less hops are required to reach the core network. As a result, EE generates globally less traffic for the BH links and thus the BH energy consumption becomes much lower. This is better explained in Fig. 4.11(b), where the average traffic of each BH link is depicted for scenario 2, when N=100 UEs. As it is shown, EE achieves again much better BH load balancing (e.g., the UEs around SC $j = 4$ that have as a candidate cell the SC $j = 5$, will be connected to the SC that has the least loaded BH route). Then, as far as RE is concerned, it has poor energy efficiency performance, since there are more UEs associated to SCs than in RSRP. As a result, higher BH traffic is generated and consequently higher BH energy consumption. Finally, MPL presents the lowest energy efficiency, since most UEs are associated with the closest SCs, resulting in the most highly loaded BH links (see Fig. 4.11(b)) and consequently in the highest BH energy consumption.

BH architecture 2

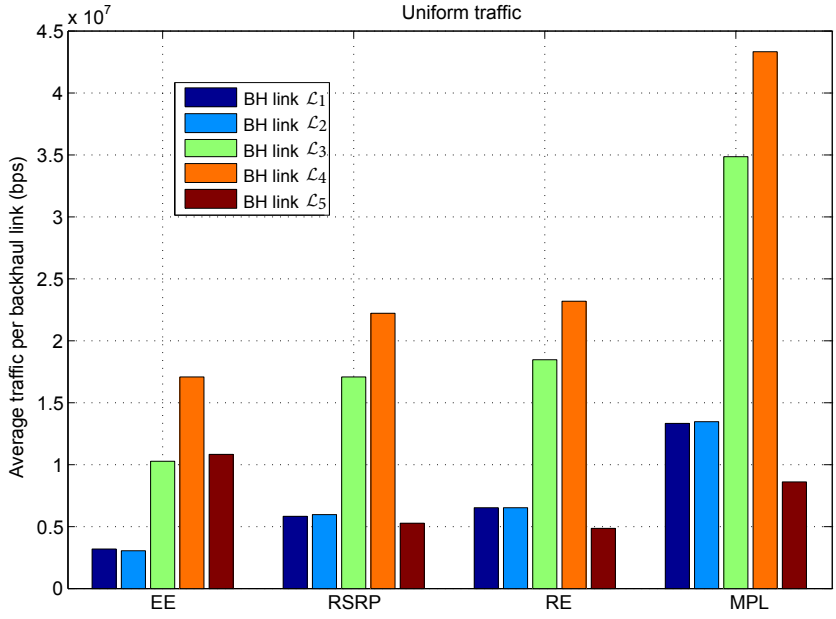
When the BH architecture scenario of Fig. 4.9 is applied, the total network energy efficiency, under the same user distribution scenarios (Fig. 4.12(a) and 4.13(a)), is lower than in the previous BH architecture scenario. This is due to the fact that the number of hops has a strong impact on the BH energy consumption (the SC $j = 3$ and SC $j = 5$ are now three hops away from the core network). Thus, apart from the need for a careful design of the BH architecture, backhaul-aware user association strategies should be developed, with the aim of achieving high network energy efficiency.

In Fig. 4.12(b) and 4.13(b), the average traffic of each BH link for N=100 UEs for the BH architecture of Fig. 4.9 is depicted, when uniform and hotspot are applied, respectively. In comparison to Fig. 4.10(b) and 4.11(b), it can be observed that the BH traffic distribution may differ significantly, when the BH architecture changes. This fact justifies the need for backhaul-aware user association strategies, especially in dense deployment scenarios. To that end, unlike the rest of the algorithms, EE takes into account both the AN and BH information in a dynamic way to achieve energy efficiency gains. For instance, notice that in the previous BH architecture, load balancing was applied between BH links $\mathcal{L}_1$, $\mathcal{L}_2$ and $\mathcal{L}_3$, whereas in this architecture, between BH links $\mathcal{L}_1$ and $\mathcal{L}_2$, since they are the same number of hops away from the core network. It is also worth noting that due to its backhaul-awareness, EE could also deal with BH link failures.

In Fig. 4.14, the average network spectrum efficiency is depicted for all algorithms. This metric was shown to be independent of the applied BH architecture scenario. However, it highly depends on UEs' SINR and consequently on UEs' distribution. As it is shown in Fig. 4.14, EE achieves equally high spectrum efficiency to RSRP and RE for both user distribution scenarios, since it achieves the same

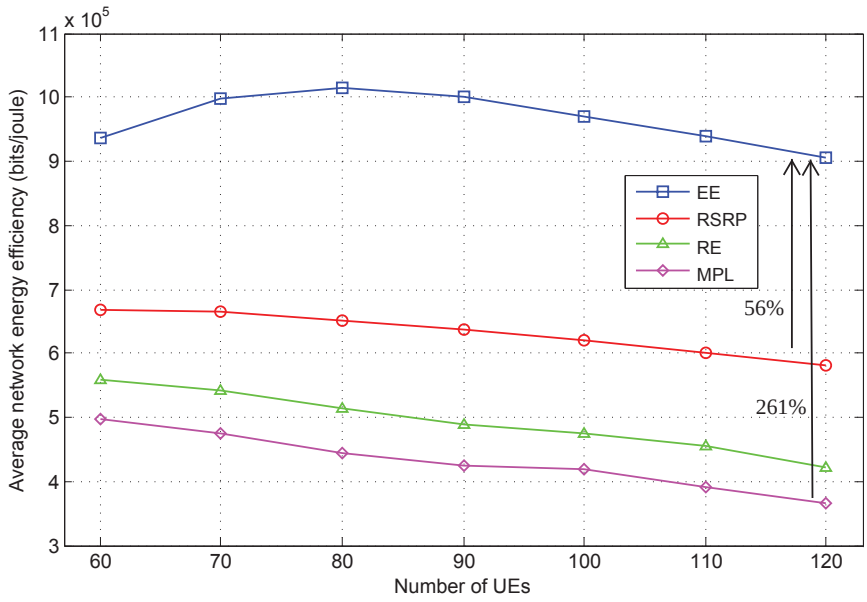


(a) Average network energy efficiency for different number of users (N), when the UEs are uniformly distributed in the sector area.

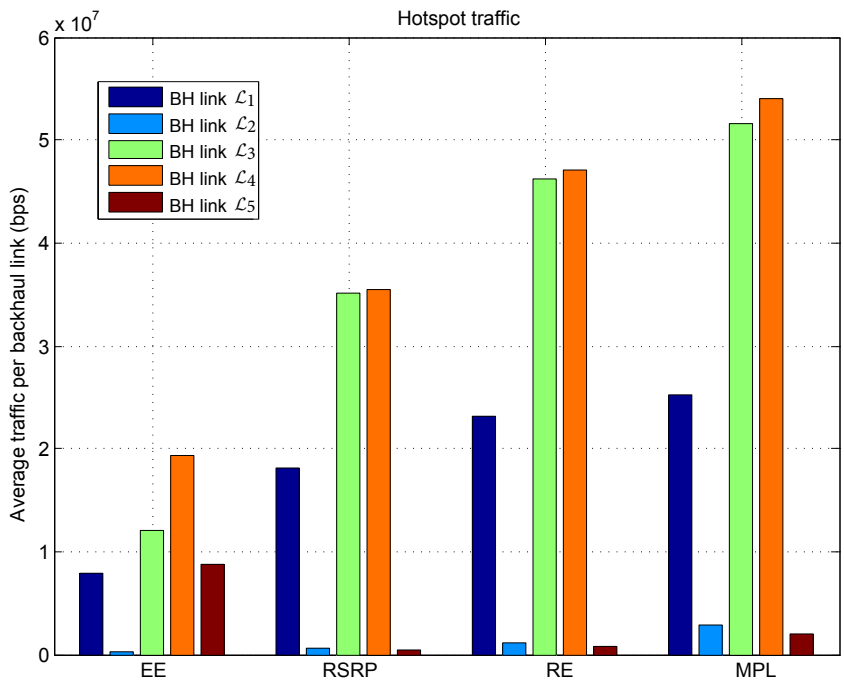


(b) Average traffic of each backhaul link for uniform traffic and N=100 UEs.

Figure 4.12: Performance evaluation results for the simulation scenario with BH architecture 2, depicted in Fig. 4.9, and uniform traffic.



(a) Average network energy efficiency for different number of users (N), for hotspot traffic.



(b) Average traffic of each backhaul link for hotspot traffic when N=100 UEs.

Figure 4.13: Performance evaluation results for the simulation scenario with BH architecture 2, depicted in Fig. 4.9, and hotspot traffic.

throughput with the same amount of total PRBs used. In particular, the high spectrum efficiency of EE, stems from basing the user association not only on the BH configuration, but also on the link quality (i.e., EE considers as candidate cells for a UE only the cells with the highest SINR, that satisfy the UE requirements). MPL, unlike the rest of the algorithms, presents much lower spectrum efficiency, as it associates the UEs independently of their SINR. Thus, a UE may be associated with a SC, having the minimum path loss, although its received SINR is lower, and therefore more PRBs will be needed to achieve the same throughput. Moreover, notice that MPL achieves higher spectrum efficiency in the hotspot scenario. This is due to the fact that UEs that form a hotspot around a SC will have both low path loss and high SINR received from it, and thus will need a lower number of PRBs to satisfy their requirements.

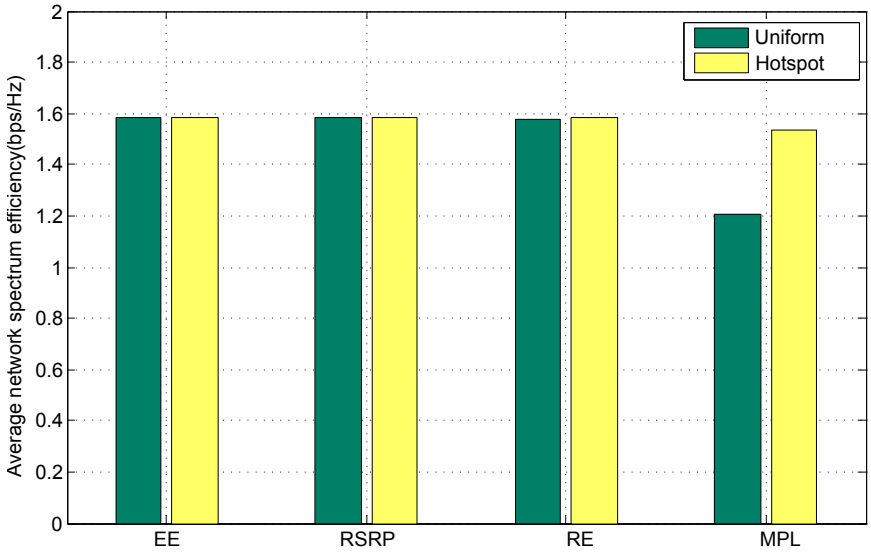


Figure 4.14: Average network spectrum efficiency for both UE traffic distribution scenarios.

4.4.4 Joint Uplink and Downlink User Association in Cognitive HetNets

Introduction and Related Work

In parallel, considering a user association criterion that takes into account only the UL or downlink (DL) [39, 43, 44, 51, 83, 96] is not sufficient. Although the channel conditions may be similar for UL and DL (i.e., UL-DL duality), the traffic load may vary significantly. This is mainly due to increase of asymmetric traffic applications that can be biased toward the DL (e.g., Internet access) or the UL (e.g., uploading a video) [105]. Furthermore, associating a UE with the BS that provides the highest DL SINR, may require higher UE trasmission power in UL. Thus, next generation user association algorithms should jointly optimize UL and DL. To that

end, the authors of [106] propose an association algorithm that provides the best service to UEs on UL and DL, while minimizing the interference. Specifically, the UEs are associated with the BS with minimum path loss in UL and the strongest received power in DL. Nevertheless, the association decision is based only on the AN conditions and the algorithm performance in terms of energy consumption is not studied.

Hence, in this section, the joint UL and DL user association problem is studied aiming at maximizing the total network energy efficiency, without compromising the UE QoS, in cognitive HetNets. The problem is formulated as an optimization problem, which is shown to be NP-hard. Therefore, the heuristic algorithm previously proposed has been accordingly adapted to associate the UEs in an energy-efficient way taking into account both UL and DL. The proposed algorithm first selects a subset of cells as candidates separately for UL and DL based on the radio access conditions and the UE requirements, and then selects the best one among them based on the BH conditions and the available spectrum resources of each BS. In the case the best cell for UL is different from DL, UL/DL split is applied [86] to maximize the network performance. Finally, the performance of the proposed algorithm is evaluated and it is shown that it significantly outperforms the reference algorithms, while maintaining high spectral efficiency and low UE power consumption.

In this section, the system model described in Section 4.2 is applied, with the following additional assumptions.

- In the UL, open loop power control is used [107].
- Each UE can be associated with at most two different BSs at a time, one for UL and one for DL [86].
- There is a maximum number of physical resource blocks (PRBs) available to each BS j in UL and DL, denoted by $c_{j_{max}}^{UL}$ and $c_{j_{max}}^{DL}$, respectively.

For reader's convenience, in the following the calculation of the UL power consumption model for the AN is provided.

Given that open loop power control is used, the UE i transmission power to BS j is given by [108]

$$P_{AN_{ij\,(dBm)}}^{UL} = min(P_{UE_{max}}, P_0 + \gamma L_{p_{ij}}^{DL} + 10log_{10}B) \tag{4.13}$$

where $P_{UE_{max}}$ is the UE maximum transmission power in dBm, P_0 is the target received power in dBm, appropriately chosen so that the BS is able to demodulate and decode the transmission information with a given reliability, $L_{p_{ij}}^{DL}$ is the DL path-loss estimated by the UEs in dB, which is used to compensate the UL path-loss. The parameter γ is the compensation coefficient ($0 \leq \gamma \leq 1$), while B indicates the instantaneous bandwidth measured in number of PRBs.

Algorithm 2 Find candidate cells

Input: $\mathcal{N}$, $\mathcal{C}$, $P_{UE_{max}}$, γ, P_0, $L_{p_{ij}}^{DL}$, B, P_{thres}, $SINR_{ij}$, r_i^{DL}
Output: $Candidates_i^{UL}$, $Candidates_i^{DL}$
1: **for all** $i \in \mathcal{N}$, $j \in \mathcal{C}$ **do**
2: Calculate $P_{AN_{ij}}^{UL}$
3: $Candidates_i^{UL} \leftarrow j : P_{AN_{ij}}^{UL} \leq (min\ (P_{AN_{ij}}^{UL}) + P_{thres})$
4: Calculate $c_{ij}^{DL} = \lceil \frac{r_i^{DL}}{f(SINR_{ij})} \rceil$
5: $Candidates_i^{DL} \leftarrow j : min(c_{ij}^{DL})$
6: **end for**

Problem Formulation

The problem under study aims at the maximization of the sum of UL and DL network energy efficiency, while satisfying the UEs' GBR throughput demands (i.e., guaranteeing the UE QoS). The UL and DL network energy efficiency can be expressed as the total number of successfully transmitted bits by all UEs in both UL and DL divided by the total energy consumption (i.e., the sum of the energy consumed in the AN and in the BH links) in both UL and DL. Thus, the aforementioned problem can be formulated as in (4.14), where the parameter a_{ij}^w equals

$$\begin{aligned} \max_{a_{ij}^w} \quad & \frac{\sum_{w=UL,DL} \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{C}} r_i^w a_{ij}^w}{\sum_{w=UL,DL} (\sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{C}} P_{AN_{ij}}^w a_{ij}^w + \sum_{\mathcal{L}_l\in\mathcal{C}} P_{\mathcal{L}_l}^w)} \\ \text{s.t.} \quad & a)\ a_{ij}^w \in \{0,1\},\ \forall i \in \mathcal{N},\ \forall j \in \mathcal{C} \\ & b)\ \sum_{j\in\mathcal{C}} a_{ij}^w = 1, \forall i \in \mathcal{N} \\ & c)\ \sum_{i\in\mathcal{N}} a_{ij}^w c_{ij}^w \leq c_{j_{max}}^w,\ \forall j \in \mathcal{C} \end{aligned} \tag{4.14}$$

to 1 when the UE i is associated with BS j and 0 otherwise (4.14a). Each UE can be associated only with one BS at a time in UL or DL (4.14b). However, as previously mentioned, the UL association may differ from the DL. Furthermore, the total number of PRB pairs used by BS j, denoted by c_{ij}^w, cannot exceed the maximum number that is allocated to it (4.14c).

The aforementioned problem is a non-convex problem that cannot be solved in polynomial time (i.e., it is NP-hard). Therefore, in the next section a novel context-aware heuristic algorithm is proposed that associates the UEs aiming at maximizing jointly the UL and DL network energy efficiency, while inducing low computational complexity.

Proposed Joint Energy-efficient Context-aware Algorithm (Joint EE)

Algorithm 3 Joint UL-DL energy-efficient user association

Input: $\mathcal{N}$, $\mathcal{C}$, c_{ij}^{w}, $c_{j_{max}}^{w}$, $\mathcal{L}_l$, $\mathcal{L}$, $Candidates_i^w$

1: Sort all UEs i by $Candidates_i^w$ size in ascending order
2: Calculate N_{hops} $\forall j \in Candidates_i^w$
3: Sort $Candidates_i^w$ by N_{hops} in ascending order
4: **if** $\exists$ only one candidate having the fewest N_{hops} **then**
5: Choose this one
6: **else**
7: Choose the one with the least loaded BH route
8: **end if**
9: **if** the chosen BS has sufficient PRBs **then**
10: Associate the UE to it
11: Update remaining PRBs, cell throughput and the throughput that passes through the link $\mathcal{L}_l \ni j$
12: **else**
13: Move to the next candidate and repeat the process
14: **end if**

The proposed joint UL-DL energy-efficient user association algorithm (Joint EE) consists of two sequential stages. In the first stage, Algorithm 2 is applied, to define the candidate cells (separately for UL and DL) based on the radio access conditions (i.e., UE measurements) and the UE requirements. In the second stage, Algorithm 3 is used, that selects the best cell among the candidates (separately for UL and DL) based on the BH conditions (i.e., BH network architecture, BH traffic) and the available spectrum resources of each BS. It is reminded that all the aforementioned context-aware information can be readily available to all nodes in a LTE-A network, while introducing low overhead (e.g., by transmitting it during the ABS or through the X2 interface between BSs (eNBs and/or SCs) [39]). In the case the best cell is the same for UL and DL, the UE is associated with one BS. Otherwise, the UE is associated with two BSs, one for UL and one for DL, using UL/DL split, as previously mentioned.

According to Algorithm 2, Joint EE chooses as candidate cells: i) for the UL, the cells that require lower UE power consumption than an upper threshold value (i.e., $min\ (P_{AN_{ij}}^{UL})+P_{thres}$), to minimize the UE power consumption (line 3), and ii) for the DL, the cells that satisfy the UE requirements with the minimum number of finite PRBs pairs needed, to achieve high spectral efficiency (line 5).

After the selection of candidate cells in UL and DL, Algorithm 3 chooses the best candidate cell to associate the UEs in UL and DL in an energy efficient way, without compromising the UE QoS. To ensure that all UEs will be associated, Joint EE sorts the UEs by their number of candidates and starts with the UEs with the fewest candidates (line 3). A cell cannot be considered as candidate if the received SINR by the UE is so low that the spectral efficiency equals to zero and thus infinite PRBs are needed to meet the UE requirements. Then, to maximize the network energy efficiency, Joint EE sorts the candidate cells by the number of hops (i.e., number of BH links) until their traffic reaches the eNB site, denoted by

Table 4.3: Joint UL-DL User Association: Additional Simulation Values

Parameter	Value
B^w_{eNB}, B^w_{SC}	10 MHz
$c^w_{eNB_{max}}$, $c^w_{SC_{max}}$	50
$P_{UE_{max}}$	23 dBm
NF_{UL}	7 dB
$G_{T_{xeNB}}$	17 dBi
$G_{T_{x\mathcal{L}_l}}, G_{R_{x\mathcal{L}_l}}$	15 dBi
IL	7 dB
P_{thres}	100 mW
$SINR^{UL}_{target}$	7 dB
R	500 m

N_{hops}, (line 3) and it associates the UE to the candidate with the fewest hops[4], as long as it has sufficient PRBs to serve it (lines 5, 9). Otherwise, it moves to the next candidate (line 13). Every time a UE is associated with a BS j in link $w = UL, DL$, the algorithm updates the remaining PRBs of j, its cell throughput and the throughput that passes through the BH link $\mathcal{L}_l \ni j$ (line 11). In case there are more candidates with the same number of hops, Joint EE associates the UE to the one with the least loaded BH path, as long as it has sufficient PRBs, to achieve BH load balancing (lines 7, 9). Load balancing achieves further energy efficiency improvement, since the energy consumption of a BH link does not increase linearly to its traffic load.

The scalability of Joint EE depends on the number N of UEs in the eNB sector, the number C of BSs involved and the time interval at which the algorithm is executed. This time-interval should be based on the dynamics of the UE traffic distribution, so that the system performance is optimized. For the new UEs that appear in the meantime, Joint EE is executed as before given the associations of the rest. However, in that case, the context-aware information includes the remaining PRBs and the throughput of each BS/cell, and the throughput that passes through each BH link, given the traffic of the already associated UEs.

Simulation Results

In the extensive simulations executed in MATLAB®, the same simulation setup and parameters of section 4.3.3 were considered (unless else stated), with $N_{cl} = 2$ clusters. Each cluster consists of 4 SCs (SC_1=SC_2=4) according to 3GPP specifications [84]. The BH links are LOS mmWave links (60 GHz band) with B_{BH}=50 MHz channel bandwidth [92].

However, unlike the static BH architecture scenarios considered up until now (i.e., BH architecture 1 and 2), from this section on, a more general and dynamic BH architecture will be adopted, which complies with Scenario 1 of 3GPP specifications [84]. According to it, the SC clusters are uniformly distributed within the eNB sector, and the SCs of each cluster are uniformly dropped within the cluster area.

[4]Due to the considered architecture, if the eNB is included in the candidate cells, it will be preferred, since no hops are required to reach the eNB site.

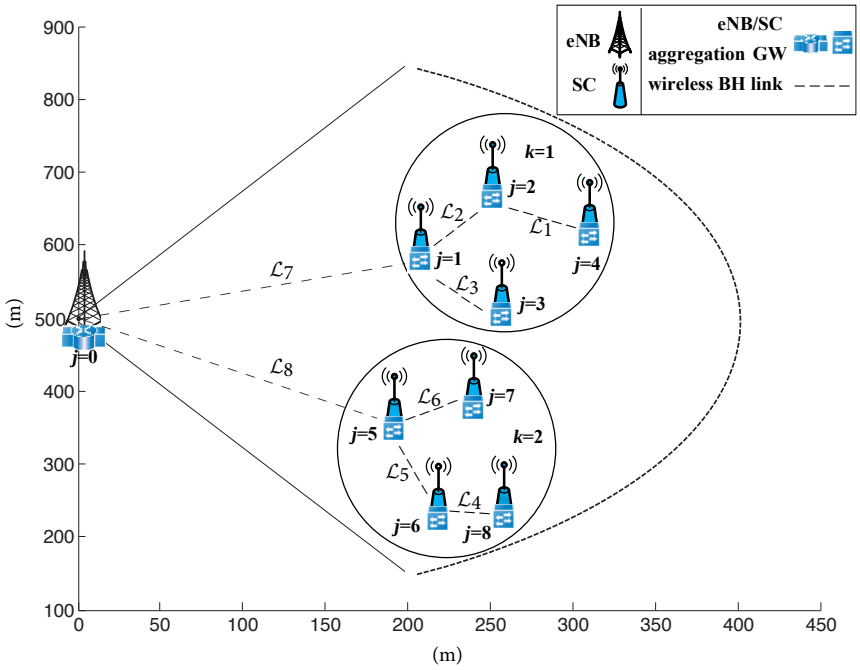


Figure 4.15: Example of simulation scenario with general BH architecture [84], where e.g., $\mathcal{L}_7 = \{1, 2, 3, 4\}$, $\mathcal{L}_1 = \{4\}$, $\mathcal{L}_2 = \{2, 4\}$, and $\mathcal{L}_3 = \{3\}$.

The minimum distance between two SCs is 20 m and between the eNB and a SC cluster center center is 105 m. Moreover, the minimum distance of a UE from the eNB is 35 m and from a SC is 5 m. Furthermore, in each cluster, one SC (the one being the closest to the eNB) is considered to be one hop away from the eNB site and thus plays the role of the aggregator of the cluster traffic, two SCs (the ones being the closest to the aggregator) are considered to be two hops away from the eNB site and the last SC to be three hops away and connected to the closest two-hop-away SC of the cluster. In order to ease understanding, an example of the considered architecture is given in Fig. 4.15.

The proposed work, as already explained, is independent of the employed channel allocation scheme. Therefore, for the sake of simplicity and without loss of generality, inter-sector interference is assumed to be mitigated through some form of fractional frequency reuse scheme or sophisticated frequency allocation [83]. Moreover, the channels allocated to the eNB are assumed to be orthogonal to the channels allocated to SCs. However, SCs belonging to different clusters reuse the same bands, thus interfering to each other.

Moreover, in each realization, N UEs are considered with different GBR requirements in UL and DL. Specifically, 20% of UEs demand 1024 kbps in the DL and 256 kbps in the UL, 20% 728 kbps in the DL and 1024 kbps in the UL and 60% have symmetric requirements for UL and DL equal to 512 kbps in each link [86]. In the UL, we set γ=1, thus considering *full path-loss compensation* for all algorithms.

In accordance with Scenario 1 of [84], the following hotspot traffic distribution scenario is considered:

- Hotspot: the UEs form hotspots. In particular, 2/3 of UEs are randomly and uniformly dropped within the clusters (in a radius $r = 70$ m from cluster center) and 1/3 UEs randomly and uniformly dropped throughout the eNB sector area [84].

The additional (or modified compared to section 4.3.3) simulation parameters are summarized in Table 4.3, where the subscript $x = \{eNB, SC\}$ refers to the eNB or to a SC, respectively and B_x^w is the bandwidth allocated to x in link w and $c_{x_{max}}^w$ is the maximum number of PRBs allocated to x in link w.

As reference algorithms we consider the following ones:

- RSRP, where UE i is associated for both UL and DL with BS j^* from which it receives the strongest RS [39, 44]

$$j^* = \text{argmax}_{j \in \mathcal{C}} RSRP_{ij} \tag{4.15}$$

- Range expansion (RE), where $bias = 13$ dB is added to the RSRP if the signal comes from a SC [45, 83, 93–99]

$$j^* = \text{argmax}_{j \in \mathcal{C}} (RSRP_{ij} + bias_j) \tag{4.16}$$

- Minimum path loss (MPL), where UE i is associated for both UL and DL with BS j^* from which it has the minimum estimated path loss, $L_{p_{ij}}^{DL}$, independently of its received power [43]

$$j^* = \text{argmin}_{j \in \mathcal{C}} L_{p_{ij}}^{DL} \tag{4.17}$$

- Joint UL-DL [106] (Joint), where UE i is associated with BS j^* that provides the minimum path loss in the UL and the strongest received power in the DL

$$\begin{aligned} UL: j^* &= \text{argmin}_{j \in \mathcal{C}} L_{p_{ij}}^{DL} \\ DL: j^* &= \text{argmax}_{j \in \mathcal{C}} RSRP_{ij} \end{aligned} \tag{4.18}$$

The number of UEs, N, has been appropriately selected to avoid system overloading. As a result, all algorithms satisfy the specific UE throughput demands (i.e., they guarantee the UE QoS), and therefore they achieve the same total throughput. Consequently, the total network energy efficiency will depend only on the total power consumption in UL and DL.

To that end, in Fig. 4.16, the total UL-DL network energy efficiency is depicted for all algorithms. As it can be noticed, Joint EE significantly outperforms the rest of the algorithms under low load conditions. In particular, it achieves up to 45% higher total UL-DL energy efficiency than RSRP. This is due to the fact that Joint

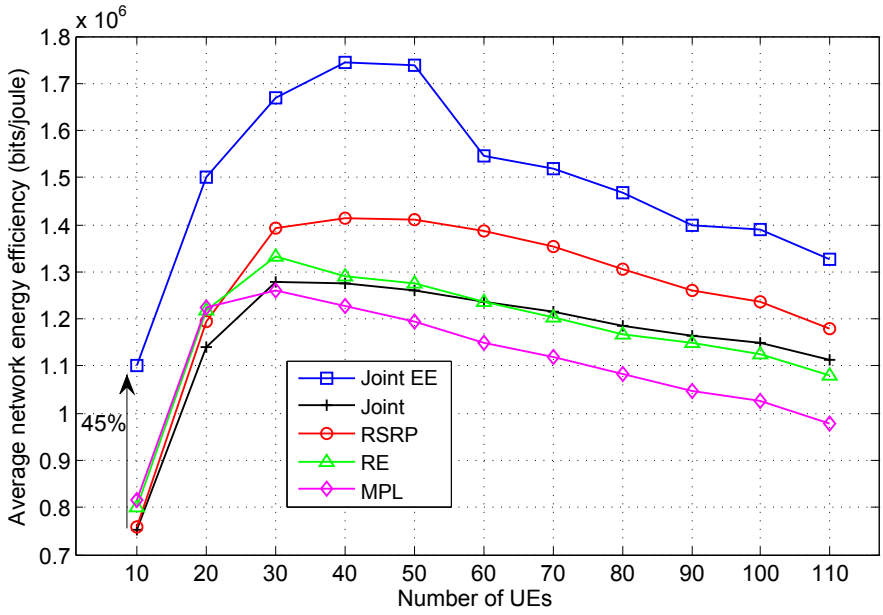


Figure 4.16: Average total UL-DL energy efficiency for different N values and hotspot traffic.

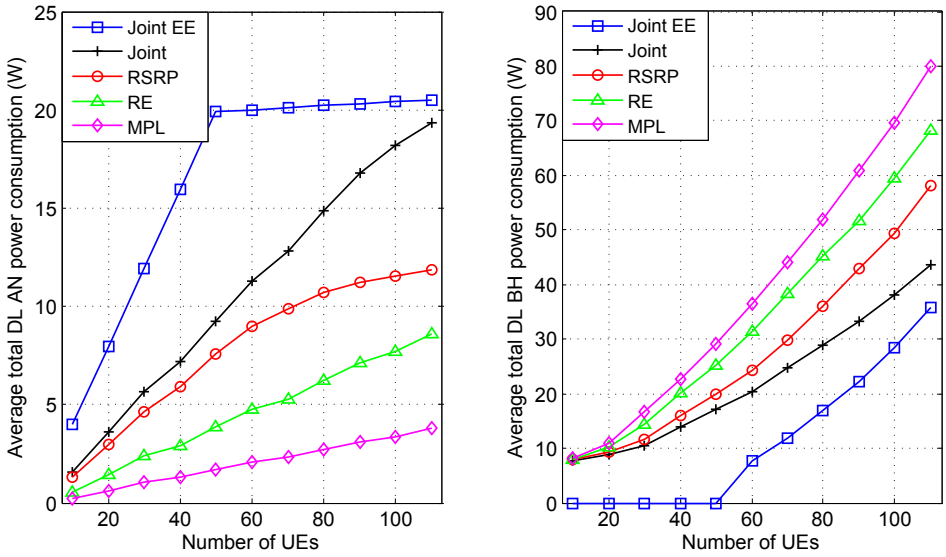


Figure 4.17: Average DL AN and BH power consumption for different N values and hotspot traffic.

EE gives priority to the candidate cell with the fewest hops and thus, under the considered architecture scenario, most UEs get connected to the eNB[5]. As a result,

[5]It is worth noting that in the considered architecture scenario, the UE connection with the

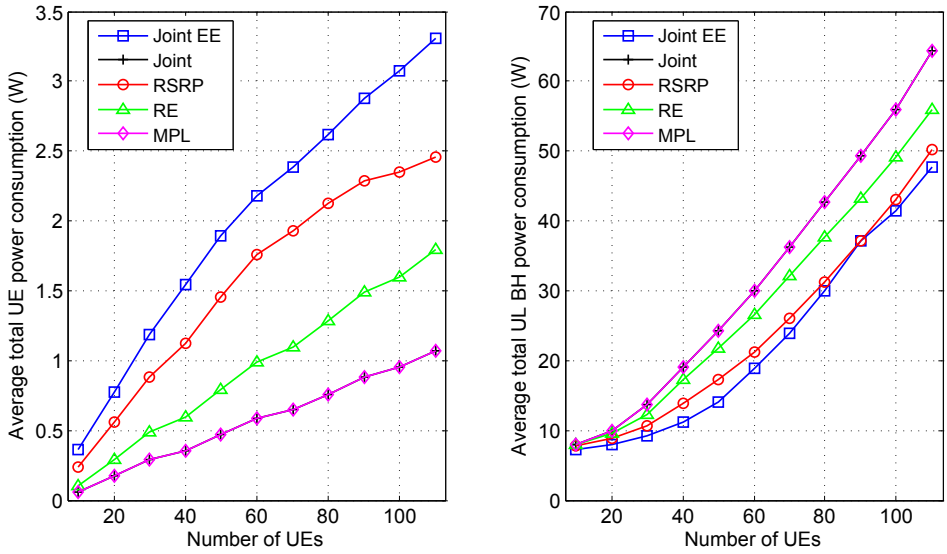


Figure 4.18: Average UL AN and BH power consumption for different N values and hotspot traffic.

for low values of N, although there is higher DL energy consumption in the AN compared to RSRP (the power per subcarrier is much higher for the eNB than for a SC), the BH energy consumption is close to zero (see Fig. 4.17) and thus the total DL power consumption is lower than RSRP. Notice that the BH energy consumption strongly depends on the BH architecture (e.g., considering two SCs of three hops in each cluster, would result in higher BH energy consumption, and thus higher energy efficiency gains for our proposal).

In UL, due to the hotspot traffic distribution and the applied threshold in the UE power consumption, the BH links cannot be switched off completely (see Fig. 4.18). Thus, as the number of UEs increases, the Joint EE gain compared to the reference approaches decreases. However, it still remains positive, since, when there are no available PRBs in the eNB, the UEs are associated with the candidate SC being the fewest hops away from the eNB site, thus minimizing the BH energy consumption. As a conclusion, Joint EE generates globally less traffic for the BH links and thus the BH energy consumption becomes much lower.

Then, as for RE, it has poor total energy efficiency performance, as there are more UEs associated to the SCs than in RSRP, thus presenting higher BH energy consumption in both UL and DL (see Fig. 4.17, 4.18). Moreover, MPL achieves the lowest total energy efficiency, since most UEs are associated with the closest SCs, resulting in the most highly loaded BH links and therefore in the highest BH energy consumption, as depicted in Fig. 4.17, 4.18. Finally, Joint [106] presents

eNB is likely to be preferred, since it involves no wireless BH power consumption. Still, given the capacity constraints of the eNB as well as the very high amount of traffic that is expected to be generated by next generation networks, SC offloading is also expected to play an important role. Therefore, efficient context-aware algorithms that will associate UEs so that the overall network performance is maximized are essential.

Table 4.4: Joint UL-DL User Association: Average Total UL-DL Network Spectral Efficiency

User association algorithm	bps/Hz
Joint EE	3.19
Joint	3.18
RSRP	3.02
RE	2.99
MPL	2.83

similar performance in UL with MPL and in DL with RSRP. However, note that Joint [106] performs better than these algorithms as the traffic load increases, since it exploits the possibility of connecting a UE to two different BSs in UL and DL, thus maximizing the total UL-DL spectral efficiency.

In Table 4.6, the average total UL-DL spectral efficiency of the network is presented for all algorithms. As it is shown, the considered algorithms that jointly study UL-DL (i.e., Joint EE, Joint [106]) achieve the highest spectral efficiency, since the UEs are always connected to the BSs that require the fewest PRBs for their requirements to be fulfilled for both UL and DL. On the contrary, RSRP and RE achieve slightly lower spectral efficiency, as they associate the UEs only based on the DL conditions, which, however, may differ significantly from the UL. MPL, unlike the rest of the algorithms, presents much lower total spectral efficiency, since it associates the UEs independently of their SINR. Hence, it is very likely that a UE is associated to a BS with low received SINR, thus requiring more PRBs to achieve the same throughput.

4.5 Optimal User Association Framework in Cognitive HetNets

4.5.1 Introduction and Related Work

As already discussed, the user association problem has received a lot of research attention, since it impacts both the network and UE performance. In this context, apart from efficient and low complexity user association algorithms, special research focus has been given to the design of analytical frameworks for optimal user association, which could be used to evaluate the performance of existing user association solutions.

To that end, in [50], an energy efficient user association problem is studied from a population game perspective. The game payoff function considers both the energy cost for using a BS and the user selfish performance objective. In [49], the authors propose a low-complexity distributed algorithm that converges to a near-optimal solution and they show that a per-tier biasing loses little, if the bias values are chosen carefully. In [48], the joint user association and resource allocation problem is studied. The authors aim to find the optimal association so that the total resources

required to satisfy the given UE traffic demands are minimized. Focusing also on the joint spectrum allocation and user association problem, in [47], a proportionally fair utility function based on the coverage rate is defined. The authors associate the UEs with BSs based on the biased downlink received power, while stochastic geometry is used to model the placement of BSs. In [46], the authors formulate two different user association problems. The first one is based on a sum utility of long term rate maximization with rate quality of service (QoS) constraints, and the second on minimizing a global outage probability with outage QoS constraints.

Taking into account the BH, in [55], a new theoretical framework is introduced to model the downlink user association problem, while upper bounds are derived for the achievable sum rate and minimum rate using convex optimization. In [51], the authors model a BH-aware BS assignment problem as a multiple-choice multidimensional Knapsack problem (MMKP). In the considered utility-based framework, they impose constraints on both radio and BH resources. The main idea behind their algorithm is to distribute traffic among BSs according to a load balancing strategy, considering both AN and BH load status. Yet, the proposed algorithm, reduces the BH congestion at the expense of lower spectral efficiency, since some UEs may be assigned to non-optimal BSs in terms of RSRP. In [53], a load-balancing based mobile association framework is proposed under both full frequency reuse and partial frequency reuse, and pseudo-optimal solutions are derived using gradient descent method. In [52], a new theoretical framework is introduced to model the downlink user association problem, while upper bounds are derived for the achievable sum rate and minimum rate using convex optimization. In [43], a joint user association and resource allocation optimization problem is proposed, which is shown to be NP-hard. Therefore, the authors developed techniques to obtain upper bounds on the system performance. In [54], the joint problem of downlink user association and wireless backhaul bandwidth allocation is studied in two-tier cellular HetNets. According to the considered architecture, the SCs are connected through wireless BH with the macro BS. The problem is formulated as a sum logarithmic user rate maximization problem, and wireless BH constraints are also considered.

However, unlike the work presented in this thesis, the aforementioned approaches either consider only the AN [39, 44–50], thus totally overlooking the BH capacity constraints and BH energy impact, or/and do not take into account the energy consumption and hence, their energy efficiency cannot be guaranteed [43, 51–55].

4.5.2 Energy and Spectrum Efficiency Maximization

In this section, the user association problem is formulated as a generalized assignment problem (GAP) [109], which considers both the AN and BH energy consumption. The objective of the GAP problem is the maximization of the energy and spectrum efficiency of all UEs, without compromising their QoS (i.e., the specific UE throughput demands are satisfied). Since the considered problem is NP-hard, we relax the capacity constraints to derive an upper bound, which can be used as a benchmark for the performance evaluation of user association algorithms. The energy-efficient context-aware algorithm, proposed in Section 4.4.3, is enhanced to

take into account the sum of the AN and BH power consumption for the traffic of a UE to be served, thereby relaxing the assumption of homogeneous BH links of Section 4.4.3. Finally, the considered bound and the proposed algorithm are compared with existing user association solutions in scenarios with mmWave BH links. The provided results show that there is still room for energy efficiency improvement for the existing user association solutions, which justifies the motivation of this work. On the other hand, the proposed algorithm is shown to significantly outperform its counterparts, while achieving near-optimal performance.

Problem Formulation and Optimal association

As already mentioned, the problem under study aims at the joint maximization of the energy and spectrum efficiency of all UEs, without compromising their QoS (i.e., the specific UE throughput demands are satisfied). The energy efficiency can be expressed as throughput divided by the total power consumption (i.e., both in the AN and BH links). Maximizing the total network energy efficiency could imply favoring some UEs at the expense of others. Therefore, in this section, the maximization of the energy and spectrum efficiency of every UE is taken into account. This goal can reflect an operator's objective, i.e., to provide the same service (throughput) with the minimum power consumption and PRBs, thereby achieving fairness among the UEs. The aforementioned problem is a GAP [109], which can be formulated as in (4.19). The problem refers to how to optimally assign each UE (it is reminded that

$$
\begin{aligned}
\max_{a_{ij}} \quad & \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{C}} a_{ij}p_{ij} \\
\text{where} \quad & p_{ij} = r_i - \beta_1 c_{ij} P_{j_c} - \\
& \beta_2 \sum_{\mathcal{L}_l\in\mathcal{L}} s_{\mathcal{L}_l j}\left(2^{\frac{r_i}{B_{\mathcal{L}_l}}} - 1\right) 10^{\frac{\alpha_{\mathcal{L}_l}-30}{10}} - \beta_3 c_{ij} \\
\text{s.t.} \quad & a)\ a_{ij} \in \{0,1\},\ \forall i \in \mathcal{N},\ \forall j \in \mathcal{C} \\
& b)\ \sum_{j\in\mathcal{C}} a_{ij} = 1, \forall i \in \mathcal{N} \\
& c)\ \sum_{i\in\mathcal{N}} a_{ij} c_{ij} \le c_{j_{max}},\ \forall j \in \mathcal{C} \\
& d)\ s_{\mathcal{L}_l j} \in \{0,1\},\ \forall \mathcal{L}_l \in \mathcal{L},\ \forall j \in \mathcal{C}
\end{aligned}
\tag{4.19}
$$

a_{ij} is equal to 1 when the UE i is associated with BS j and 0 otherwise (4.19a)) to exactly one BS (4.19b), so as to maximize the total profit, denoted by p_{ij}, without reserving from any BS more spectrum resources than its capacity, $c_{j_{max}}$ (4.19c). The parameter $s_{\mathcal{L}_l j}$ is 1 if the traffic of the BS j passes through the BH link $\mathcal{L}_l$ and

0 otherwise (4.19d).

In order to maximize the energy and spectral efficiency of all UEs, the profit function, p_{ij}, has been appropriately selected to consider the UE QoS demands, r_i, the AN and BH power consumption (both calculated in W) as well as the number of PRBs needed so that the UE traffic is served by BS j. Notice that p_{ij} is expressed as the weighted sum of our objective functions. Thereby, the multi-objective optimization problem can be cast as a single-objective one, with normalized weighting coefficients β_1, β_2 and β_3, as defined in [110].

Since the considered problem is NP hard [109], we derive the Ross and Soland upper bound [109], by relaxing the capacity constraints to

$$c_{ij}a_{ij} \leq c_{_{j_{max}}} \forall i \in \mathcal{N},\ \forall j \in \mathcal{C} \tag{4.20}$$

The optimal solution $\hat{a}$ to the resulting problem is obtained by determining for each $i \in \mathcal{N}$ the BS j, that maximizes p_{ij}, while having sufficient PRBs to serve the UE traffic, i.e.,

$$j(i) = \arg\max\{p_{ij} : j \in \mathcal{C}, c_{ij} \leq c_{_{j_{max}}}\} \tag{4.21}$$

Setting $\hat{a}_{j(i),i}$=1 and $\hat{a}_{ji} = 0$ for all $j \in \mathcal{C} \setminus \{j(i)\}$, the resulting upper bound equals to

$$U_0 = \sum_{i \in \mathcal{N}} p_{j(i),i} \tag{4.22}$$

This bound is not yet feasible, since the capacity constraints are not fulfilled. Therefore, we define the following parameters

$$\mathcal{N}_j = \{i \in \mathcal{N} : \hat{a}_{ji} = 1\}, j \in \mathcal{C} \tag{4.23}$$

$$d_j = \sum_{i \in \mathcal{N}_j} c_{ij} - c_{_{j_{max}}}, j \in \mathcal{C} \tag{4.24}$$

$$\mathcal{C}' = \{j \in \mathcal{C} : d_j > 0\}, \quad \mathcal{N}' = \cup_{j \in \mathcal{C}'} \mathcal{N}_j \tag{4.25}$$

where $\mathcal{N}_j$ is the set of UEs associated with BS j, d_j is the number of extra PRBs needed for which the relaxed capacity constraint of BS j is violated, C' is the set of those cells whose relaxed constraint (4.19c) is violated and $\mathcal{N}'$ is their set of UEs. Given a set $\mathcal{S}$ of numbers, we denote with $max_2\mathcal{S}$ the second maximum value in $\mathcal{S}$ and with $\arg\max_2 \mathcal{S}$ the corresponding index. Thus, (4.26) gives the minimum penalty incurred if the UE i, currently assigned to a BS in $\mathcal{C}'$, is reassigned.

$$q_i = p_{j(i),i} - max_2\{p_{ij} : j \in \mathcal{C}, c_{ij} \leq c_{_{j_{max}}}\}, i \in \mathcal{N}' \tag{4.26}$$

Hence, for each $j \in \mathcal{C}'$, a lower bound on the loss of profit to be paid to satisfy the constraint (4.19c) is given by the solution of the 0-1 single Knapsack problem,

Algorithm 4 Proposed enhanced energy-efficient user association algorithm

Input: $\mathcal{N}, \mathcal{C}, SINR_{ij}, r_i, b, c_{j_{max}}, \mathcal{L}_l, \mathcal{L}$

Calculate c_{ij} as in (4.3)
$Candidates_i \leftarrow j \ : \ min(c_{ij})$
Sort all UEs i by $Candidates_i$ size in ascending order
Calculate $P_{AN_{ij}} + P_{BH_{ij}}$ as in Section 4.2.2 $\forall j \in Candidates_i$
Sort $Candidates_i$ by $P_{AN_{ij}} + P_{BH_{ij}}$ in ascending order
Choose the candidate with the minimum $P_{AN_{ij}} + P_{BH_{ij}}$
if the chosen BS has sufficient spectrum resources **then**
    Associate the UE to it
    Update remaining spectrum resources
else
    Move to the next candidate and repeat the process
end if

$KP_j^1(j \in \mathcal{C}')$, defined as

$$\begin{aligned} v_j = \min \quad & \sum_{i \in \mathcal{N}_j} q_i y_{ij} \\ \text{s.t.} \quad & a)\ y_{ij} \in \{0,1\},\ i \in \mathcal{N}_j \\ & b)\ \sum_{i \in \mathcal{N}_j} y_{ij} c_{ij} \geq d_j, \end{aligned} \tag{4.27}$$

where $y_{ij} = 1$ if and only if UE i is removed from BS j. The resulting bound is then equal to

$$U_1 = U_0 - \sum_{j \in \mathcal{C}'} v_j \tag{4.28}$$

Proposed Enhanced Energy-efficient Context-aware Algorithm (e-EE)

The proposed enhanced energy-efficient (e-EE) algorithm, summarized in Algorithm 4, takes as input the available context-aware information, i.e., the UEs' measurements (SINR) and QoS rate requirements (r_i), the HetNet architecture knowledge, i.e., the BH link path that follows the traffic of each SC ($s_{\mathcal{L}_l j}$) and the available spectrum resources of each BS ($c_{j_{max}}$), to associate the UEs in an energy efficient way.

As already mentioned in previous sections, this context-aware information, which can be divided into information being reported by the network and information being reported by the UEs, can be easily available to all nodes in a LTE-Advance network (i.e., eNBs and/or SCs) [39, 111]. In particular, the information being reported by the network does not impose additional constraints, since the standard defines the X2 logical interface to allow the exchange of information among BSs (eNBs and/or SCs) [39]. Moreover, the information about the network architecture (s_{L_l}) requires very limited or nil update due to its static nature. Hence, the

only additional information to be exchanged is the current traffic of each BH link. Regarding the information being reported by the UEs, it is worth pointing out that each UE is required to measure the SINR received from the neighboring BSs. For such a purpose, Release 8 has already defined the radio resource management (RRM) measurement set, i.e., the set of BSs from which a UE measures and reports parameters, such as RSRP or reference signal received quality (RSRQ). Later on, in order to support coordinated multi-point (CoMP), Release 10 defined a subset of the RRM measurement, namely CoMP measurement set, to allow the UEs to measure and report short-term channel state information [111]. Thereby, the aforementioned mechanisms guarantee the availability of the required information.

As shown in Algorithm 4, e-EE considers as candidate cells for a UE the cells that satisfy its requirements with the minimum needed spectrum resources (c_{ij}) (line 2), thus achieving high spectral efficiency. In addition, by considering only a subset of cells, the computational burden induced in the system is reduced. To ensure that all the UEs will be associated, e-EE first sorts the UEs by their number of candidates and starts with the UEs with the fewest candidates (line 3). Note that a cell j cannot be included in the candidates set of a UE i if $SINR_{ij}$ is too low and, consequently, $c_{ij} > c_{j_{max}}$.

At a second step, in order to maximize the network energy efficiency, e-EE sorts the candidate cells by the total power consumption, denoted by $P_{AN_{ij}} + P_{BH_{ij}}$, needed in each case for the traffic of the UE i to be served (line 5). Then, e-EE associates the UE to the candidate cell which involves the minimum power consumption, as long as it has sufficient spectrum resources to serve it (lines 7). Otherwise, it moves to the next candidate (line 11). Every time a UE is associated with a BS j, the algorithm updates the remaining spectrum resources of j.

The proposed algorithm (e-EE) may be executed in each eNB sector at a specific time interval based on the dynamics of the UE traffic, so that the system performance is optimized. For the new UEs that appear in the meantime, e-EE may be executed as before given the associations of the rest, thereby, achieving high scalability.

Simulation Results

For the extensive simulations executed in MATLAB®, the same simulation scenario and parameters as in section 4.4.4 (unless otherwise stated), are considered. In each realization, we also consider N UEs of different GBR requirements. Specifically, 60% of UEs demand 512 kbps, 20% 728 kbps and 20% 1024 kbps [86].

In accordance with section 4.4.4 and Scenario 1 of [84], hotspot user distribution is assumed as described below:

- Hotspot: the UEs form hotspots. In particular, 2/3 of UEs are randomly and uniformly dropped within the clusters (in a radius $r = 70$ m from cluster center) and 1/3 UEs randomly and uniformly dropped throughout the eNB sector area [84].

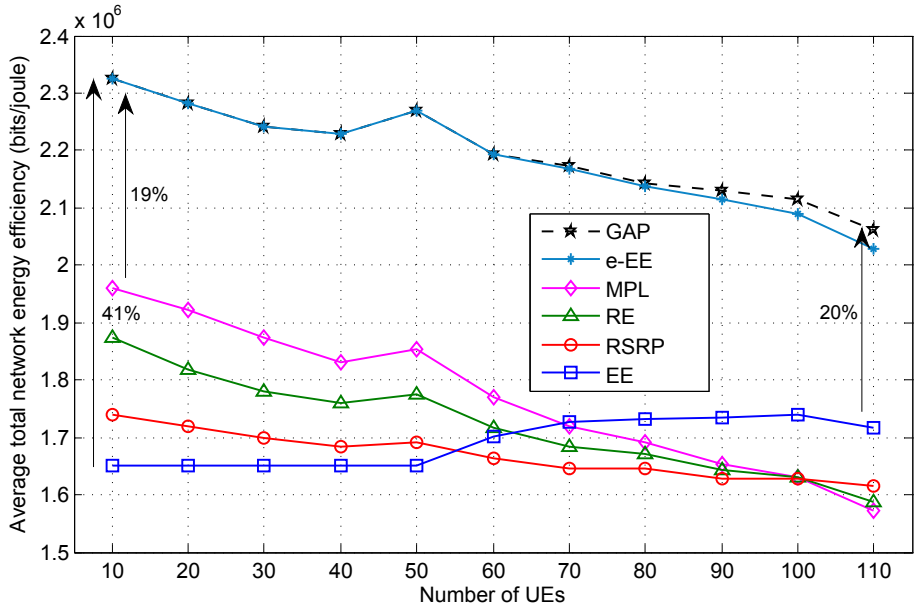


Figure 4.19: Average total network energy efficiency for different N values and hotspot traffic.

For the performance evaluation of reference algorithms we consider the following ones:

- GAP: the user association solution of the upper bound, that was derived in Section 4.5.2.
- e-EE: the user association algorithm proposed in Section 4.5.2.
- EE: the user association algorithm proposed in Section 4.4.3.
- RSRP: a UE connects to the BS from which it receives the strongest signal [39, 44].
- Range expansion (RE): a $bias = 6$ dB is added to the RSRP if the signal comes from a SC [45, 83, 93–99].
- Minimum path loss (MPL), where a UE connects to the BS from which it has the minimum path loss [43], independently of its received power.

The range of the number of UEs, N, has been appropriately selected to avoid system overloading. As a result, all algorithms satisfy the specific UE throughput demands (i.e., they guarantee the UE QoS), and therefore they achieve the same total throughput. Consequently, the total network energy efficiency will depend only on the total power consumption.

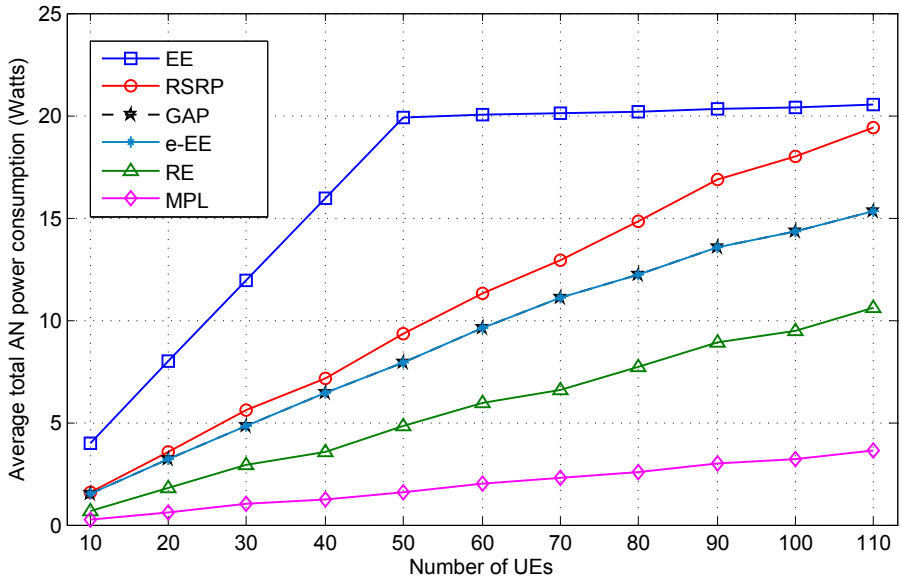


Figure 4.20: Average total access network power consumption for different N and hotspot traffic.

In Fig. 4.19, the average network energy efficiency is depicted for all algorithms versus the number of UEs, N. As it can be observed, e-EE significantly outperforms its counterparts, since it achieves near-optimal performance. Notice that EE achieves worse performance, since it gives priority to the candidate cell with the fewest hops and thus most of the UEs get connected to the eNB. As a result, for low values of N, although the BH energy consumption is equal to zero, there is high energy consumption in the AN (the power per subcarrier is much higher for the eNB than for a SC). On the contrary, e-EE takes into account the possibility of having heterogeneous BH links and adapts the user association decision accordingly. Therefore, it presents less dependency on the employed scenario. Then, regarding RE, it achieves lower energy efficiency, as there are more UEs associated with SCs, resulting in higher BH energy consumption. Finally, MPL can be considered as an aggressive RE, since UEs get associated to the closest BS, and so hotspot traffic is mainly offloaded to SCs. Although MPL achieves the maximum offloading, it has very low energy efficiency, as the BH traffic and thus the BH energy consumption is much higher than the other approaches. To gain further insights, we depict in Fig. 4.20 and 4.21 the total AN and BH power consumption, versus the number of UEs, N, respectively. As it can be noticed, the AN power consumption increases as N increases for all algorithms. In particular for EE, this happens at a high rate, since more UEs get connected to the eNB. However, for very high traffic there are no resources left in the eNB and thus more UEs get connected to SCs, which results in a smoother increase in the AN power consumption. As depicted in Fig. 4.21, the BH energy consumption also increases as the number of UEs increases, since more traffic is generated in the BH links and thus more energy consumption. It is also

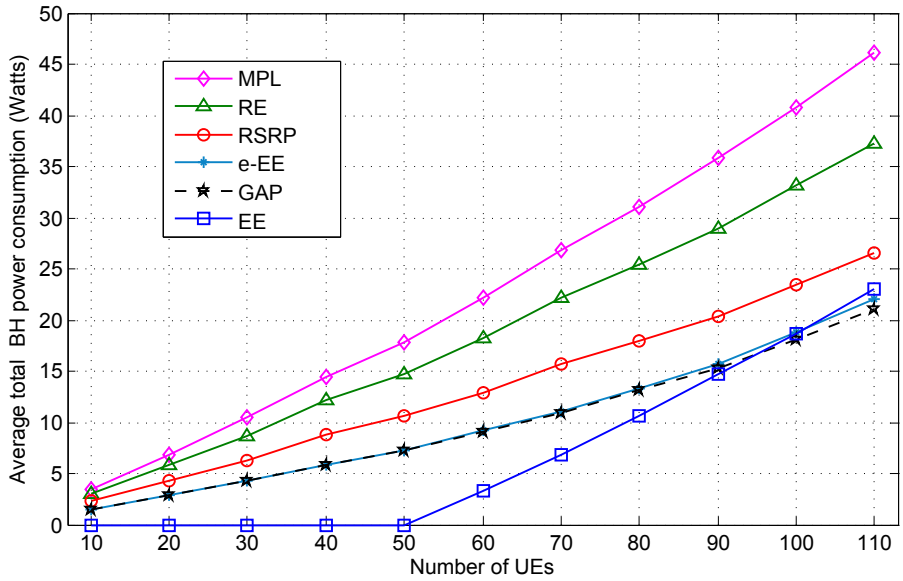


Figure 4.21: Average total backhaul power consumption for different N values and hotspot traffic.

Table 4.5: GAP User Association: Average Network Spectral Efficiency

User association algorithm	**bps/Hz**
GAP, e-EE, EE	1.827
RSRP	1.824
RE	1.816
MPL	1.717

worth noting that the optimal association achieves a good balance between AN and BH power consumption and so does the proposed algorithm.

In Table 4.5, the average network spectral efficiency is presented for all algorithms. As it is shown, the considered algorithms that aim at the maximization of the spectral efficiency (i.e., GAP, e-EE, EE) achieve the highest spectral efficiency, since the UEs are connected to the BSs that require the minimum spectrum resources for their requirements to be fulfilled. On the contrary, RSRP and RE achieve slightly lower spectral efficiency, as the UEs, under high traffic load conditions, may be connected to BSs that require more spectrum resources. MPL, unlike the rest of the algorithms, presents much lower spectral efficiency, since it associates the UEs independently of their SINR. Hence, it is very likely that a UE is associated to a BS with low SINR, thus requiring more spectrum resources to achieve the same throughput.

4.5.3 Spectrum Efficiency and Energy Efficiency trade-off

In this section, an ε-constraint problem formulation is proposed to study the trade-off between energy and spectrum efficiency. By solving the ε-constraint problem for all different ε values, the Pareto front solutions of the problem are analytically derived. To that end, the previously proposed e-EE is modified accordingly so that a *good* trade-off between energy and spectrum efficiency is achieved. In particular, this novel adaptive algorithm, denoted as Ae-EE, aims at the maximization of the energy efficiency given a specific spectral efficiency target. The performance of the proposed Ae-EE algorithm is compared with existing user association solutions (e.g., RSRP, RE, MPL) as well as with the derived optimal solutions under different spectral efficiency targets, traffic distribution scenarios and BH technologies. The provided results motivate the use of mmWave frequencies to provide high capacity BH, while the proposed algorithm is shown to achieve notable performance gains.

Problem formulation

The problem under study aims at the joint maximization of the energy and spectrum efficiency of the network, without compromising the UE QoS (i.e., the specific UE throughput demands). The energy efficiency is expressed as the total number of successfully transmitted bits divided by the total energy consumption or equivalently as the total throughput of the network divided by the total power consumption (i.e., the sum of the power consumed in the AN and in the BH links). Under the condition that the specific UE throughput demands are satisfied, the network energy efficiency maximization is equivalent to the minimization of the total power consumption, while the spectral efficiency maximization is equivalent to the minimization of the PRBs needed.

$$
\begin{aligned}
\underset{a_{ij}}{argmin} \quad & f_1(a_{ij}) = \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{C}} \overbrace{a_{ij}c_{ij}P_{j_c}}^{P_{AN_{ij}(W)}} + \sum_{\mathcal{L}_l\in\mathcal{L}} \overbrace{\left(2^{\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{C}} a_{ij}s_{\mathcal{L}lj}\frac{r_i}{B_{\mathcal{L}_l}}}-1\right)10^{\frac{\alpha_{\mathcal{L}_l}-30}{10}}}^{P_{BH_{\mathcal{L}_l}(W)}} \\
\text{s.t.} \quad & a)\ a_{ij}\in\{0,1\},\ \forall i\in\mathcal{N},\ \forall j\in\mathcal{C} \\
& b)\ \sum_{j\in\mathcal{C}} a_{ij}=1, \forall i\in\mathcal{N} \\
& c)\ \sum_{i\in\mathcal{N}} a_{ij}c_{ij}\le c_{j_{max}},\ \forall j\in\mathcal{C} \\
& d) P_{BH_{\mathcal{L}_l}} \le P_{BH_{max}}\ \forall\mathcal{L}_l\in\mathcal{L} \\
& e)\ s_{\mathcal{L}lj}\in\{0,1\},\ \forall\mathcal{L}_l\in\mathcal{L},\ \forall j\in\mathcal{C} \\
& f) f_2(a_{ij}) = \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{C}} a_{ij}c_{ij} \le \varepsilon
\end{aligned}
\tag{4.29}
$$

The aforementioned problem is therefore a multi-objective problem, which can be formulated with the use of the ε-constraint method [112]. According to it, one of the objectives is included in the utility function to be optimized (i.e., minimization of the total power consumption of the network), while the others (i.e., minimization of the total number of required PRBs) are converted into constraints by setting an upper bound to them. The problem formulation is depicted in (4.29), where the first term of the objective function represents the total AN power consumption, while the second the total BH power consumption. It is reminded that a_{ij} denotes the association vector that is equal to 1 when the UE i is associated with BS j and 0 otherwise (4.29a). Each UE can be associated only with one BS at a time (4.29b). Furthermore, the total number of resources used by BS j, denoted by c_{ij}, cannot exceed the maximum number of PRBs that is allocated to it (4.29c). The power consumption of the BH link $\mathcal{L}_l$ cannot exceed a maximum value, denoted by $P_{BH_{max}}$ (4.29d). The parameter $s_{\mathcal{L}_l j}$ is 1 if the traffic of the BS j passes through the BH link $\mathcal{L}_l$ and 0 otherwise (4.29e). Finally, constraint (4.29f) refers to the total number of PRBs used and hence to the spectrum efficiency of the network.

Theorem 1. *The solution of the ε-constraint problem in* (4.29) *is weakly Pareto optimal.*

Proof. Let $a^{\star}_{ij}$ be a solution of the ε-constraint problem. Let us assume that $a^{\star}_{ij}$ is not weakly Pareto optimal. In this case there exists some other a_{ij} such that $f_k(a_{ij}) < f_k(a^{\star}_{ij})$ for k=1,2. This means that $f_2(a_{ij}) < f_2(a^{\star}_{ij}) \leq \varepsilon$. Hence, a_{ij} is feasible with respect to the ε-constraint problem. While in addition $f_1(a_{ij}) < f_1(a^{\star}_{ij})$, we have a contradiction to the assumption that $a^{\star}_{ij}$ is a solution of the ε-constraint problem. Thus, $a^{\star}_{ij}$ [6] has to be weakly Pareto optimal. □

Although, according to Theorem 1, every solution of the ε-constraint problem is weakly Pareto optimal, there is no Pareto optimal solution, since there is no solution that optimizes both objectives simultaneously. Therefore, it is reasonable to search for a *good* trade-off between the two objectives instead. To that end, the increase of ε leads to a relaxation of the spectral efficiency constraint (i.e., f_2) and consequently to a more energy efficient solution. On the contrary, the decrease of ε improves the spectral efficiency of the solution by degrading its energy efficiency. The set of solutions for the subproblems resulting from the variation of ε define the Pareto front, hereafter denoted by $\mathcal{F}$.

In practice, because of the high number of subproblems and the difficulty to establish an efficient variation scheme for the ε-vector, this approach has mostly been integrated within heuristic and interactive schemes. However, due to the nature of the problem (4.29), it is possible to derive the exact Pareto front with the use of an iterative algorithm [113]. The idea is to construct a sequence of ε-constraint problems based on a progressive reduction of ε.

Let $\vec{\phi}^I = (\phi^I_1, \phi^I_2)$ be the ideal point, where $\phi^I_1 = \min(f_1)$ and $\phi^I_2 = \min(f_2)$ stand for the minimum value of f_1 and f_2, respectively. Equivalently, let $\vec{\phi}^N =$

[6] Please note that, in the rest of the section, a_{ij} is omitted.

Algorithm 5 Exact Pareto front calculation of problem (4.29)

1: Calculate the ideal and nadir points, $\vec{\phi}^I$ and $\vec{\phi}^N$.
2: Add $\vec{f}_1^* = (\phi_1^I, \phi_2^N)$ to $\mathcal{F}$.
3: Set $m = 2$.
4: Set $\varepsilon^m = \phi_2^N - \Delta$, with $\Delta = 1$.
5: **while** $\varepsilon^m \geq \phi_2^I$ **do**
6: Solve problem (4.29) and add the optimal solution value $\vec{f}_m^* = \big((f_m^*)_1, (f_m^*)_2\big)$ to $\mathcal{F}$.
7: Set $\varepsilon^{m+1} = (f_m^*)_2 - \Delta$.
8: Set $m = m + 1$.
9: **end while**
10: Remove dominated points if required.

(ϕ_1^N, ϕ_2^N) be the nadir point, with ϕ_1^N and ϕ_2^N being the minimum values of f_1 and f_2, when $f_2 = \phi_2^I$ and $f_1 = \phi_1^I$, respectively, i.e., $\phi_1^N = \min\{f_1{:}f_2{=}\phi_2^I\}$ and $\phi_2^N = \min\{f_2{:}f_1{=}\phi_1^I\}$. Thus, (ϕ_1^I, ϕ_2^N) is the solution of the Pareto front that minimizes the total power consumption (i.e., f_1) without spectral efficiency constraints. On the contrary, (ϕ_1^N, ϕ_2^I) is the Pareto front solution that minimizes the total number of PRBs used (i.e., f_2).

Lemma 5. *Both (ϕ_1^I, ϕ_2^N) and (ϕ_1^N, ϕ_2^I) belong to $\mathcal{F}$, i.e., $(\phi_1^I, \phi_2^N) \in \mathcal{F}$ and $(\phi_1^N, \phi_2^I) \in \mathcal{F}$.*

Proof. Let us assume that $(\phi_1^I, \phi_2^N) \notin \mathcal{F}$. Then, $\exists\ \vec{f}{=}(f_1, f_2) \in \Phi$: $(f_1, f_2) \succ (\phi_1^I, \phi_2^N)$, where Φ denotes the *objective space* and the expression $\vec{f}{=}(f_1, f_2) \succ (\phi_1^I, \phi_2^N)$ denotes that (f_1, f_2) dominates (ϕ_1^I, ϕ_2^N). In general, we say that $\vec{f}{=}(f_1, f_2)$ dominates $\vec{f}'{=}(f_1', f_2')$, with $\vec{f}, \vec{f}' \in \Phi$ if and only if (iff) $f_1 \leq f_1'$ and $f_2 \leq f_2'$, where at least one inequality is strict. Thus, $\vec{f}{=}(f_1, f_2) \succ (\phi_1^I, \phi_2^N)$ is true when a) $f_1 < \phi_1^I$ and $f_2 < \phi_2^N$ or b) $f_1 < \phi_1^I$ and $f_2 = \phi_2^N$ or c) $f_1 = \phi_1^I$ and $f_2 < \phi_2^N$. Since a) and b) contradict the definition of an ideal point and since c) contradicts the definition of a nadir point, then $(\phi_1^I, \phi_2^N) \in \mathcal{F}$. The proof of $(\phi_1^N, \phi_2^I) \in \mathcal{F}$ is analogous.

□

Lemma 6. *For each $(f_1, f_2) \in \Phi$, if $(f_1, f_2) \in \mathcal{F}$, then $\phi_1^I \leq f_1 \leq \phi_1^N$ and $\phi_2^I \leq f_2 \leq \phi_2^N$.*

Proof. As proved in Lemma 5, $(\phi_1^I, \phi_2^N) \in \mathcal{F}$, and thus it is non-dominated. Since $\phi_1^I = \min(f_1)$, $f_1 \geq \phi_1^I$, $\forall\ (f_1, f_2) \in \mathcal{F}$. Moreover, if $f_2 > \phi_2^N$, $(\phi_1^I, \phi_2^N) \succ (f_1, f_2)$ and $(f_1, f_2) \notin \mathcal{F}$. Hence, $f_1 \geq \phi_1^I$ and $f_2 \leq \phi_2^N$ $\forall\ (f_1, f_2) \in \mathcal{F}$. The proof for $\phi_2^I \leq f_2 \leq \phi_2^N$ is analogous.

□

According to Lemma 5 and Lemma 6, Algorithm 5 generates the exact Pareto front of the multi-objective optimization problem described in (4.29).

Theorem 2. *Algorithm 5 generates one feasible solution for each point of the Pareto front.*

Proof. Let us denote the sequence of solutions of Algorithm 5 by $\{\vec{f}_1^*, \ldots, \vec{f}_m^*, \ldots, \vec{f}_M^*\}$, where, e.g., $\vec{f}_m^* = \big((f_m^*)_1, (f_m^*)_2\big)$, with 1, 2 denoting the first and the second objective, respectively. We have to prove that if $\vec{f} \in \Phi \setminus \{ \vec{f}_1^*, \ldots, \vec{f}_m^*, \ldots, \vec{f}_M^*\}$, then $\vec{f} \notin \mathcal{F}$. Let us assume that there is a solution $\vec{f'} = (f_1', f_2') \in \Phi \setminus \{ \vec{f}_1^*, \ldots, \vec{f}_m^*, \ldots, \vec{f}_M^*\}$ such that $\vec{f'} \in \mathcal{F}$. By Lemma 6, for the first objective we have $\phi_1^I \leq f_1' \leq \phi_1^N$. Thus, either a) $f_1' = (f_m^*)_1$ for a given $m = 1 \ldots M$ or b) $(f_{m-1}^*)_1 < f_1' < (f_m^*)_1$ and $(f_{m-1}^*)_2 < f_2' \leq (f_m^*)_2$ for a given $m = 1 \ldots M$. In the first case (i.e., case a), f_2' must be lower than $(f_m^*)_2$ for $\vec{f'}$ to be efficient. However, since $\Delta = 1$ and the second objective is integer by definition, $\exists\ \varepsilon^{m'}$ that will eventually reach a value for which the optimum of the corresponding ε-constraint problem is $\vec{f'}$ for $m+1 \leq m' \leq M$, that is $\vec{f'} \in \{\vec{f}_{m+1}^*, \ldots, \vec{f}_M^*\}$, which contradicts the hypothesis. Regarding the second case (i.e., case b), f_2' must be such that $(f_{m-1}^*)_2 < f_2' \leq (f_m^*)_2$, which is impossible since $\vec{f}_m^*$ is the optimal value of problem (4.29), with $\varepsilon^m = \varepsilon^{m-1} - \Delta$, $\Delta = 1$, and the second objective is integer. □

It is worth noting that some dominated solutions may be generated by the sequence of subproblems that are derived according to Theorem 2. However, since all dominated points can be identified, one can simply exclude the non-efficient solutions to obtain the exact Pareto front. Furthermore, although Algorithm 5 limits the number of subproblems, a subproblem may be very hard to solve. This stems from the fact that an exhaustive search would require the examination of C^N possible solutions, which results in prohibitive complexity ($\mathcal{O}(n^n)$), as the number of BSs, C, and the number of UEs, N, increase. Therefore, alternative algorithms, available in the literature, should be used, able to come up with very close to the optimal solutions with acceptable computational complexity [112]. In this work, a meta-heuristic method [115] was applied, which has been shown to lead to high-quality solutions (the average gap is less than 1% with respect to best-known solutions) in almost real time. The applied method uses biased randomization together with an iterated local search meta-heuristic algorithm. Although the meta-heuristic algorithm involves lower complexity than $\mathcal{O}(n^n)$[7], it still requires a high number of iterations (50000 in our case). Therefore, there is need for low-complexity algorithms, able to achieve solutions close to the Pareto front.

Proposed adaptive e-EE Algorithm (Ae-EE)

In accordance with the previously proposed algorithms, Ae-EE algorithm takes into account the available context-aware information, i.e., the UEs' measurements (signal-to-interference-plus-noise ratio, SINR) and rate requirements (r_i), the HetNet architecture knowledge ($s_{\mathcal{L}_{lJ}}$) and the available spectrum resources of each BS ($c_{j_{max}}$) to efficiently associate the UEs. As already explained in Section 4.5.2, this context-aware information can be readily available to all nodes in a LTE-A network (i.e., eNBs and/or SCs).

[7]Meta-heuristics have no predefined end, and thus big O notation cannot be used to describe their complexity. Yet, they can be compared empirically (through number of objective function evaluations/iterations).

Algorithm 6 Proposed adaptive energy-efficient user association algorithm

Input: $\mathcal{N}, \mathcal{C}, SINR_{ij}, r_i, b, c_{j_{max}}, \mathcal{L}_l, \mathcal{L}$

1: Calculate c_{ij} as in (4.3)
2: $Candidates_i \leftarrow j \ : \ c_{ij} \leq min(c_{ij_{min}} + c_{thres}, c_{j_{max}})$
3: Sort all UEs i by $Candidates_i$ size in ascending order
4: Calculate $P_{tot_{ij}} = P_{AN_{ij}} + P_{BH_{ij}}$ as in Section 4.2.2 $\forall j \in Candidates_i$
5: Sort the UEs with the same $Candidates_i$ size by the difference in $P_{tot_{ij}}$ between the candidate cells in descending order
6: Sort $Candidates_i$ by $P_{tot_{ij}}$ in ascending order
7: Choose the candidate with the minimum $P_{tot_{ij}}$
8: **if** the chosen BS has sufficient spectrum resources **then**
9: Associate the UE to it
10: Update remaining spectrum resources
11: **else**
12: Move to the next candidate and repeat the process
13: **end if**

The proposed algorithm, which is summarized in Algorithm 6, aims at the maximization of the energy efficiency given a specific spectral efficiency target. From this point on, we will refer to this algorithm as adaptive e-EE (Ae-EE) user association algorithm.

As shown in Algorithm 6, Ae-EE considers as candidate cells for a UE i the set of cells, denoted by $Candidates_i$, that satisfy its rate demands, while requiring less spectrum resources (c_{ij}) than a target $c_{thres} = \delta \ c_{ij_{min}}$ (line 2). In practice, this threshold corresponds to a specific SINR threshold, which defines the subset of cells that will be selected, i.e., the ones with SINR higher than the threshold. The spectral efficiency target is defined by the tuning parameter $\delta > 0$, which controls the deviation in the number of needed PRBs from the association that requires the minimum number of PRBs. For instance, selecting $\delta = 0$, and thus, $c_{thres} = 0$, would result in the maximum spectral efficiency, while setting $\delta > 0$ would decrease the spectral efficiency accordingly in favor of higher energy efficiency. By considering only a subset of cells, the computational burden induced in the system is reduced. It is also worth noting that a BS j cannot be included in the set of candidates of a UE i, if $SINR_{ij}$ is too low and, consequently, $c_{ij} > c_{j_{max}}$. To ensure that all the UEs will be associated, Ae-EE sorts the UEs by their number of candidates and starts with the UEs with the fewest candidates (line 3).

In order to maximize the network energy efficiency, Ae-EE calculates for each UE i and candidate cell j the total power consumption needed for the traffic of the UE i to be served, denoted by $P_{tot_{ij}} = P_{AN_{ij}} + P_{BH_{ij}}$, (line 4). Ae-EE then sorts the UEs with the same $Candidates_i$ size by the difference in $P_{tot_{ij}}$ between the candidate cells in descending order, i.e., starting with the UE with the maximum difference between the first and the second candidate (line 5). Thereafter, Ae-EE sorts the candidate cells of each UE i by $P_{tot_{ij}}$ in ascending order (line 6) and associates the UE to the candidate cell, which involves the minimum power consumption, as long as it has sufficient spectrum resources to serve it (lines 8). Otherwise, it moves to the next candidate (line 12). Every time a UE is associated with a BS j, the algorithm

updates the remaining spectrum resources of j. Contrary to the algorithm providing the exact Pareto front solutions, previously presented, the proposed algorithm is much less complex, i.e., $\mathcal{O}(n \log n)$ [114].

The proposed algorithm may be executed in each eNB sector at a specific time interval based on the dynamics of the UE traffic, so that the system performance is optimized. If a new UE becomes active in the meantime (i.e., after the last execution of the algorithm and before the next one), its association can be decided by Ae-EE given the associations of the rest of the UEs. In particular, Algorithm 6 is applied, excluding lines 3 and 5. Thereby, the proposed algorithm can provide high network scalability.

Results

Extensive simulations were executed in MATLAB®, under the same simulation scenario and parameters with the ones described in Section 4.4.4 (unless else stated). In this section, however, in order to gain further insights into the benefits of mmWave, three different BH technologies were considered:

i) LOS mmWave links (f_{BH1}= 60 GHz band) of B_{BH1}= 200 MHz channel bandwidth [92],

ii) LOS microwave[8] links (f_{BH2}= 28 GHz) of B_{BH2}= 28 MHz [103],

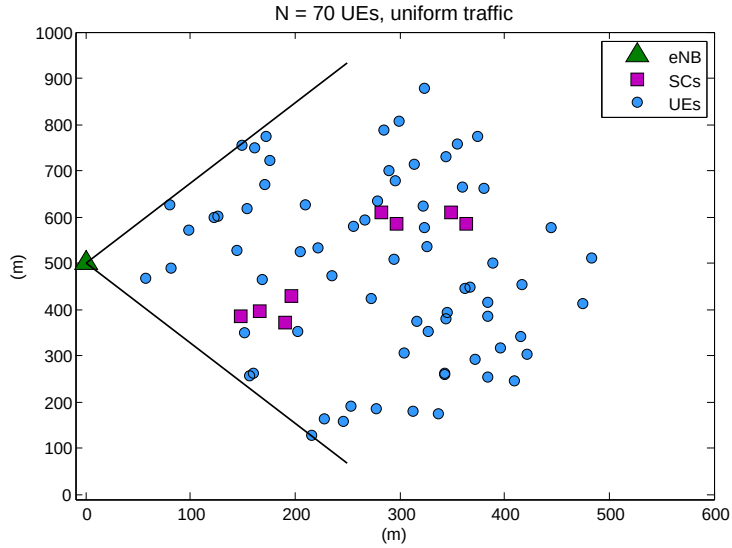
iii) sub-6GHz (f_{BH3}= 3 GHz) of B_{BH3}= 10 MHz [104].

For a fair comparison, the path loss models of the provided references were used, while the antenna gains were selected equal to 28, 15 and 10 dBi, respectively. In each realization (1000 in total), N UEs of different GBR requirements were considered. In particular, 60% of UEs demand 512 kbps, 30% 728 kbps and 10% 1024 kbps [86]. Moreover, in accordance with the previous sections, the following two UE traffic distribution scenarios were employed:

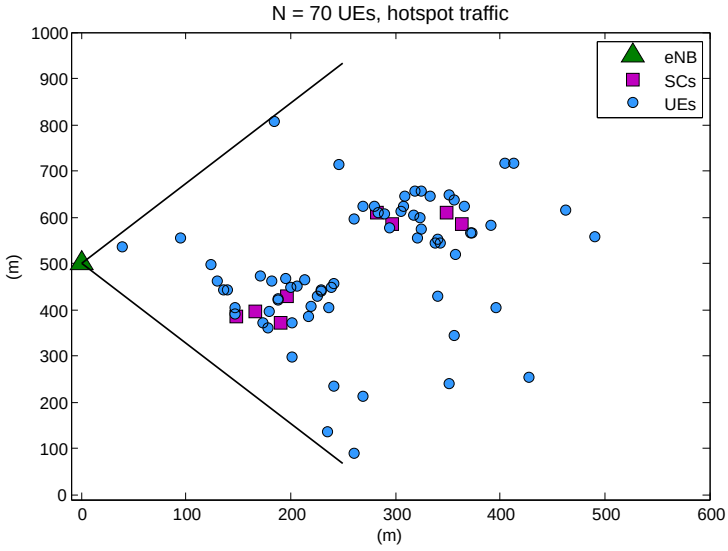
- Uniform: the UEs are uniformly distributed in the sector area of radius R.
- Hotspot: the UEs form hotspots. In particular, 2/3 of UEs are randomly and uniformly dropped within the clusters (in a radius r = 70 m from cluster center) and 1/3 UEs randomly and uniformly dropped throughout the eNB sector area [84].

As already justified in sections 4.3.3 and 4.4.4, for the sake of simplicity and without loss of generality, inter-sector interference is assumed to be mitigated through some form of fractional frequency reuse scheme or sophisticated frequency allocation [83]. Moreover, the channels allocated to the eNB are assumed to be orthogonal

[8]In [103], 28 GHz is considered as mmWave. Still, as already commented, in this thesis, we adopt the SC Forum categorization (Fig. 5-2 in [36]).



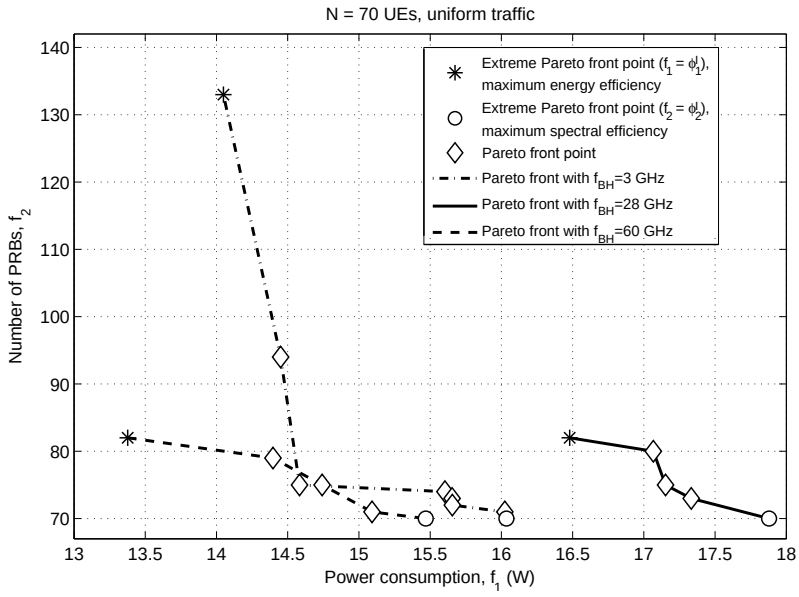
(a)



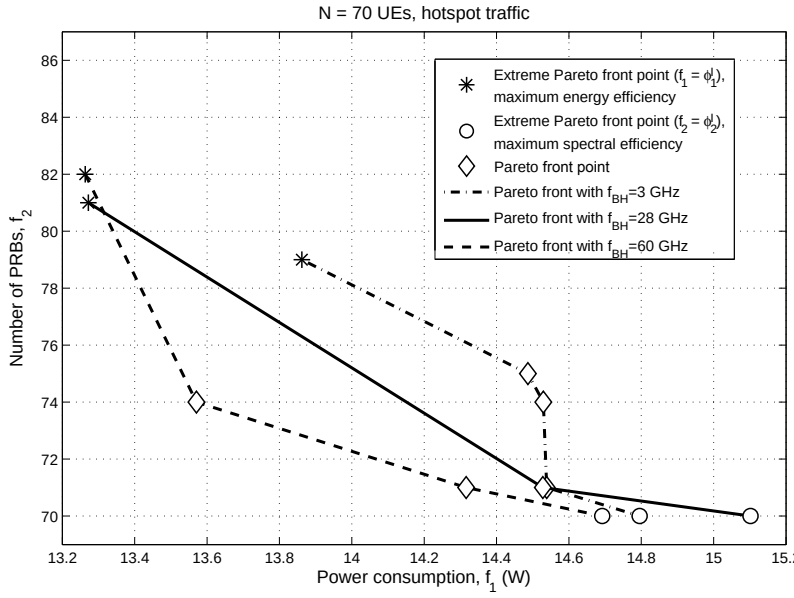
(b)

Figure 4.22: Snapshots of a) uniform and b) hotspot traffic distribution scenarios with N=70 UEs.

to the channels allocated to SCs. However, SCs belonging to different clusters reuse the same bands, thus interfering to each other. The slow fading is modeled by a log-normal random variable with zero mean and deviation 8 dB for the eNB and 10 dB for the SC signal.



(a)



(b)

Figure 4.23: Pareto front: number of PRBs vs. total power consumption for N=70 UEs with a) uniform and b) hotspot traffic for different BH technologies with frequency equal to 3 GHz, 28 GHz and 60 GHz, respectively.

1) Pareto front solutions Two different simulation scenarios are considered, as depicted in Fig. 4.22 a) and b). In the first scenario, the UEs are uniformly

distributed, while in the second they form hotspots. To that end, in Fig. 4.23 a) and b) the exact Pareto front points of the ε-constraint problem in (4.29) are depicted for the considered BH technologies.

As already mentioned, the number of PRBs and the power consumption are two metrics that can not be minimized at the same time, and thus a *good* trade-off between them has to be found. Hence, each Pareto front point corresponds to a dominant solution of the ε-constraint problem for a different ε value, as described in Theorem 2. In general, in multi-objective optimization, none of the Pareto front solutions is better than the others. However, depending on the preference for each of the conflicting objectives, a Pareto front solution may be more preferable than another. For instance, in (4.29), the preference for one of the objectives (f_1 or f_2) may vary based on the network state. In particular, in scenarios where spectral efficiency becomes important, e.g., under highly loaded scenarios, the operators may select a point near the right extreme Pareto front to maximize the spectral efficiency ($f_2 = \phi_2^I$). On the other hand, when the spectrum resources do not limit the system (except for (4.29c)), the operators could select a point of operation near the left Pareto front solution ($f_1 = \phi_1^I$), thus minimizing the network energy consumption.

In the considered example, for all BH technologies when $f_1 = \phi_1^I$ (maximum energy efficiency and satisfied rate QoS requirements), most UEs are associated to SCs to minimize the AN power consumption (the AN power consumption is much higher[9] when a UE is associated to the eNB than to a SC). Moreover, the UE association with the SC that involves the minimum BH power consumption (e.g, the one with the fewest hops or shortest BH links) is favored. Therefore, when $f_1 = \phi_1^I$ (maximum energy efficiency), the number of required PRBs is higher in the uniform scenario (than in the hotspot), since the UEs are located further from the SC cluster centers. On the contrary, when $f_2 = \phi_2^I$ (maximum spectral efficiency and satisfied rate QoS requirements), more UEs are associated to the eNB (in the uniform scenario) to reduce the required PRBs at the expense of higher AN energy consumption. Thus, the AN power consumption increase is higher in the uniform scenario for all BH technologies, as more UEs are associated to the eNB.

Regarding the rest of the Pareto front points, we notice that in the uniform scenario, for the same power consumption as in the hotspot, more PRBs are required for all BH technologies. This stems from the fact that the UEs located in a hotspot mostly get associated with SCs both to use fewer PRBs, and to have much less AN power consumption. On the contrary, when the UEs are uniformly distributed, they are located further from the SC cluster centers and thus to achieve network power consumption decrease, a proportional increase in the required PRBs is needed. This results in the Pareto front curve being steeper for the hotspot scenario, i.e., the hotspot Pareto front points provide better trade-offs between the two objectives than the uniform ones.

Among the different BH technologies, mmWave presents the best performance,

[9] It is reminded that the total transmit power of a BS is equally distributed among its PRBs [43]. Therefore, given that the maximum transmit power of an eNB is much higher than that of a SC, the allocated power at a given bandwidth (e.g., at a PRB) is also much higher.

since its Pareto front is shifted on the left. This implies that mmWave can provide better trade-offs than the rest of the BH technologies. It is worth noting that although mmWave experiences the highest path loss, it is able to send high amount of data without increasing the transmitted power due to its high bandwidth availability[10]. On the contrary, the available bandwidth of sub-6GHz is very limited. Therefore, in hotspot scenarios where higher BH traffic is generated, a significant increase in the transmitted power of the BH links is required, so that the SINR at the receiver increases. Thereby, higher order MCS can be used, which result in higher spectral efficiency. However, due to the much higher power consumption at the transmitter, this also results in lower energy efficiency. Finally, 28 GHz presents in both considered scenarios worse performance than 60 GHz, since it experiences lower path loss, with, however, much less bandwidth available. Compared to 3 GHz, 28 GHz experiences higher path loss. Nevertheless, in the hotspot scenario, where more BH traffic is generated, it achieves better performance due to higher bandwidth availability.

2) Performance Evaluation In this section, the performance of the proposed algorithm is evaluated and compared with both the state-of-the-art and the optimal (yet complex) solutions described in Algorithm 5 for all BH technologies. In particular, the algorithms that will be compared in this section are summarized in the following.

- ε-constraint: the two extreme Pareto front solutions of the ε-constraint problem described in Section 4.5.2. In particular, we refer with ε-constraint EE to the extreme Pareto front solution that maximizes the energy efficiency and with ε-constraint SE to the Pareto front solution that maximizes the spectral efficiency.
- Ae-EE: the proposed adaptive energy-efficient algorithm, described in Section 4.5.3, with c_{thres}=0, 1, 2.
- EE: the user association algorithm proposed in 4.4.3.
- RSRP: a UE connects to the BS from which it receives the strongest signal [39, 44].
- Range expansion (RE): a $bias = 13$ dB is added to the RSRP if the signal comes from a SC [45, 83, 93–99].
- Minimum path loss (MPL), where a UE connects to the BS from which it has the minimum path loss ($L_{ij} = L_{p_{ij}} + L_{f_{ij}}$) [43], independently of its received power.

[10] As discussed in section 4.3.3, the path loss of the BH technology as well as its available bandwidth play an important role in the power consumption of the BH link. For given aggregate rate demands, the path loss of the technology impacts the SINR at the receiver. A higher SINR enables the use of a higher order MCS, thus resulting in higher spectral efficiency. At the same time, for higher available bandwidth, more bits can be sent with a given SINR and MCS. In the case the bandwidth is not sufficient or the received SINR is very low, an increase in the transmitted power is necessary to increase the SINR at the receiver. Thereby, the spectral efficiency is improved at the expense of lower energy efficiency.

The range of the number of UEs, N, has been appropriately selected to avoid system overloading. As a result, all algorithms satisfy the specific UE throughput demands (i.e., they guarantee the UE QoS), and therefore they achieve the same total network throughput (e.g., for $N = 90$ UEs, the total throughput is 56.52 Mbps for all algorithms). Consequently, the total network energy efficiency will depend exclusively on the total power consumption.

Uniform traffic distribution

In Fig. 4.24, the average network energy efficiency is depicted for all algorithms and BH technologies versus the number of UEs, N, with uniform traffic distribution. To gain further insights into the energy efficiency, in Fig. 4.25 and 4.26, the total AN and BH power consumption are depicted, respectively, versus the number of UEs, N, for all BH technologies. In general, it can be noticed that mmWave achieves much higher energy efficiency than the rest of the BH technologies (i.e., 2 times more than 3 GHz and 3.5 times more than 28 GHz) for all algorithms and N values. This is due to its high bandwidth availability which, as already explained, results in much lower BH power consumption (of the order of mW, as depicted in Fig. 4.26).

As far as the different user association algorithms are concerned, it is reminded that ε-constraint EE shows the maximum energy efficiency that can be achieved independently of the spectral efficiency, while ε-constraint SE corresponds to the maximum energy efficiency given that the spectral efficiency is maximized. These solutions achieve better performance than the other approaches (up to 13, 7.5, and 28 times higher energy efficiency for 3, 28 and 60 GHz, respectively). However, unlike the rest of the algorithms, they present very high complexity, which increases with an increasing number of UEs and BSs, as discussed earlier in the problem formulation of this section.

In Fig. 4.24, it can be also noticed that for all BH technologies the proposed Ae-EE algorithm significantly outperforms the state-of-the-art algorithms, while achieving similar performance to the ε-constraint solutions with, however, much less complexity (i.e., $\mathcal{O}(n \log n)$ [114]). Nevertheless, the selection of the parameter value c_{thres} is important. Ae-EE with $c_{thres} = 0$ achieves equal performance to the ε-constraint SE, while as c_{thres} increases, the performance of the algorithm in terms of energy efficiency is improved at the expense of lower spectral efficiency.

Regarding the rest of the algorithms, as depicted in Fig. 4.24, they achieve much lower performance for all BH technologies. In particular, EE achieves low performance, since it gives priority to the candidate cell with the fewest hops to reach the core network and thus most of the UEs get connected to the eNB. As a result, for low values of N, although the BH energy consumption is equal to zero, there is high energy consumption in the AN (we remind that the power per subcarrier is much higher for the eNB than for a SC). On the contrary, Ae-EE takes into account the possibility of having heterogeneous BH links and adapts the user association decision accordingly. Therefore, it presents less dependency on

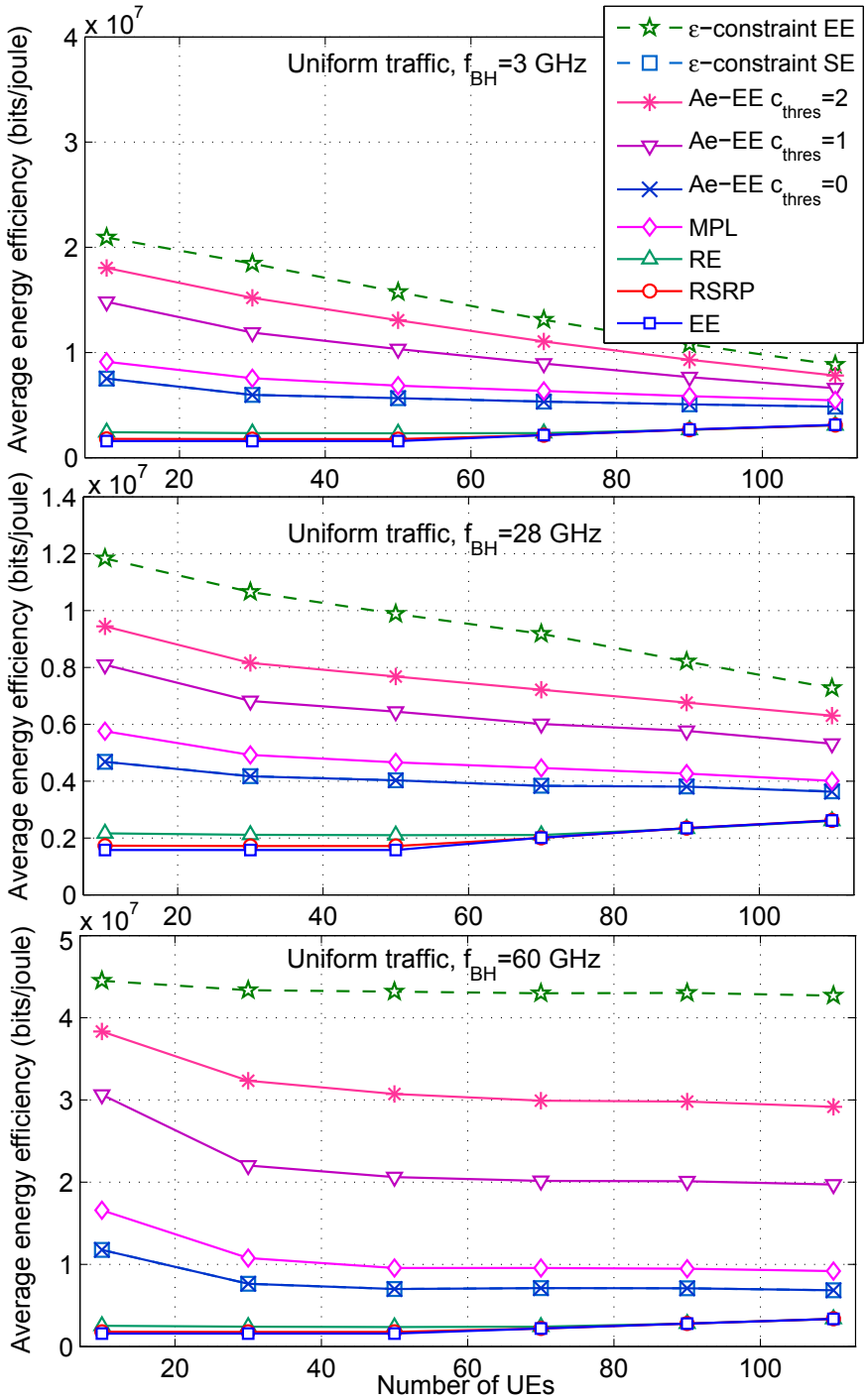


Figure 4.24: Average total network energy efficiency for different N values and BH technologies with frequency equal to 3 GHz, 28 GHz and 60 GHz, respectively, when the UEs are uniformly distributed.

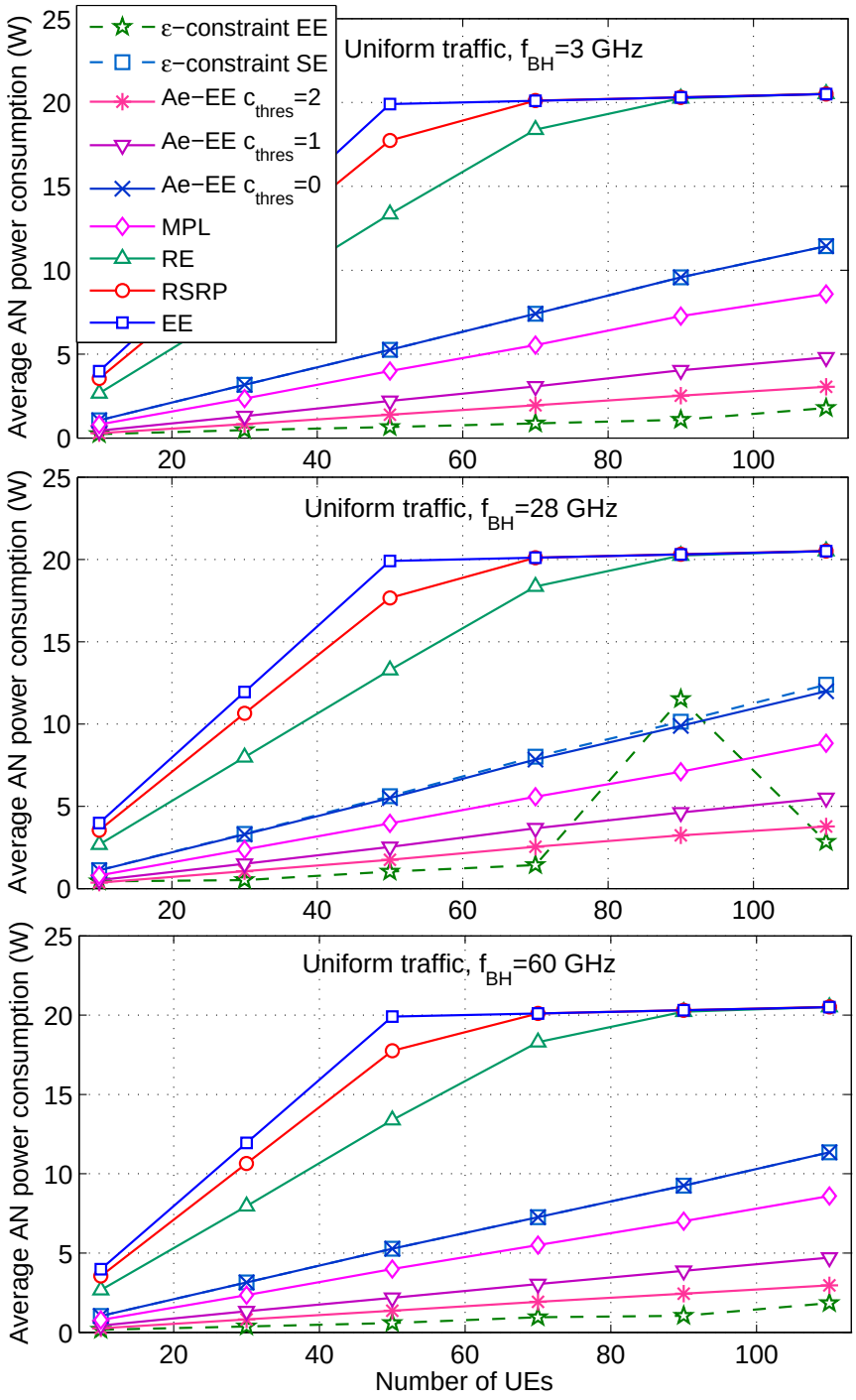


Figure 4.25: Average total access network power consumption for different N values and BH technologies with uniform traffic.

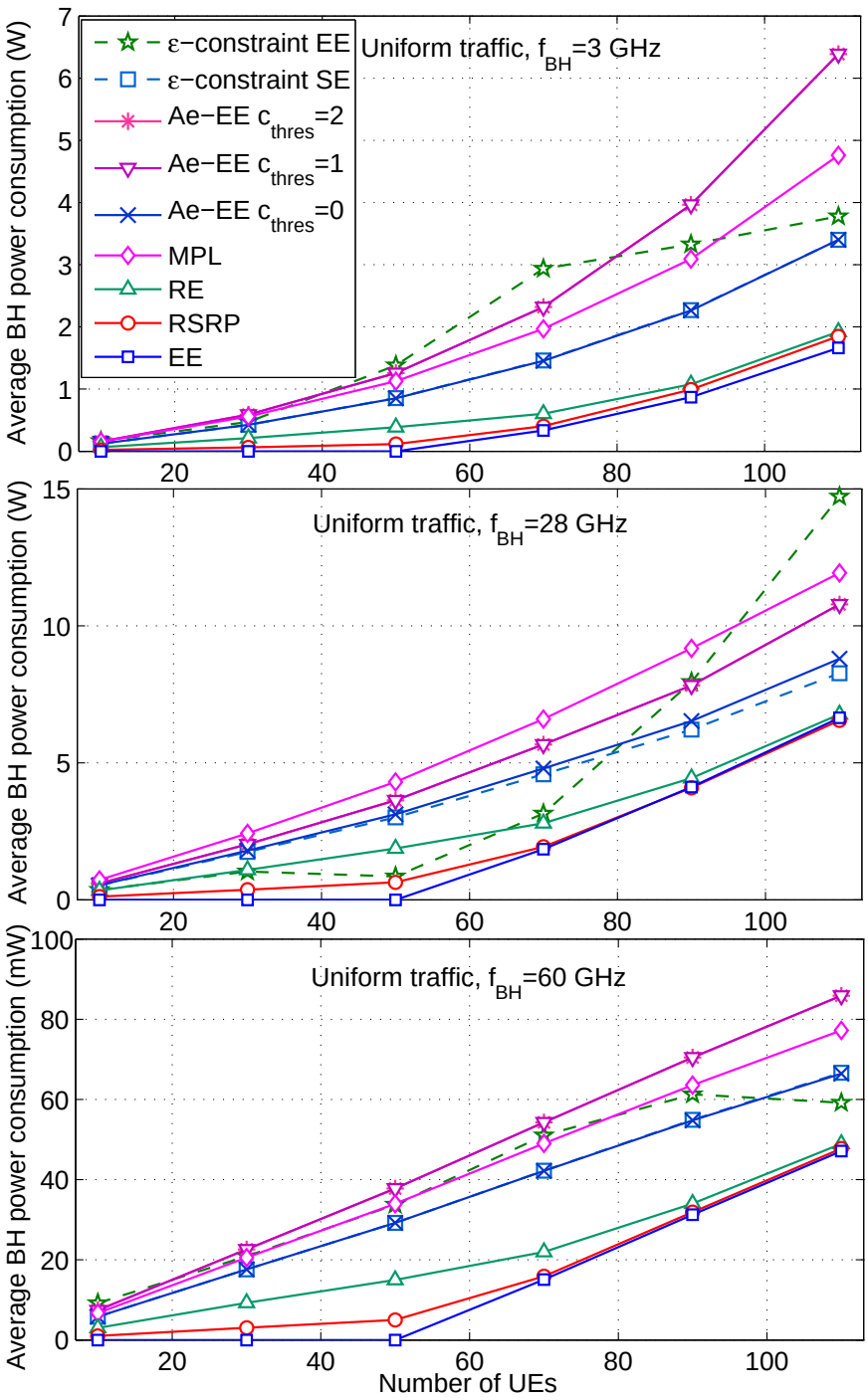


Figure 4.26: Average total backhaul power consumption for different N values and BH technologies with uniform traffic.

the employed scenario. Then, regarding RE, it achieves lower energy efficiency, since there are more UEs associated with SCs, resulting in higher BH energy consumption. Finally, MPL can be considered as an aggressive RE, as UEs get associated to the closest BS, and so traffic is mainly offloaded to SCs. Although MPL achieves the maximum offloading, it has very low energy efficiency, since the BH traffic and thus the BH energy consumption is much higher than the other approaches (see Fig. 4.26).

Moreover, as it can be observed in Fig. 4.25, the AN power consumption increases as N increases for all algorithms (except for ε-constraint EE) and BH technologies. In particular, for EE, the AN power consumption increases initially at a high rate, since more UEs get connected to the eNB. However, for very high traffic there are no resources left in the eNB, and thus more UEs get associated with SCs, which results in a smoother AN power consumption increase. As depicted in Fig. 4.26, the BH energy consumption also increases for all algorithms, as the number of UEs increases, since more traffic is generated in the BH links and thus more energy consumption.

In general, ε-constraint EE favors the combination of user associations that minimizes the total power consumption at a specific instant, and thus, it presents a different behavior than the rest of the algorithms. For instance, in the case of 28 GHz, the power consumption in BH is almost as high as in the AN. Hence, for low traffic, ε-constraint EE favors the association of most UEs with SCs and especially with the SC cluster located closer to the core network to minimize both the AN and BH power consumption. However, for higher traffic load (90 UEs), the BH aggregated traffic increases a lot (it is reminded that the power consumption of a BH link increases in an exponential way with the amount of traffic that passes through the link) and therefore the association of UEs with the eNB is preferable (both in terms of energy consumption and number of PRBs). This is also due to the fact that the association with the eNB at this point gives the possibility of totally switching off one or even both SC clusters, thus resulting in higher energy efficiency gain. To that end, it is worth noting that the most energy-consuming links in the considered model are the links that are one hop away from the core network, which not only aggregate all the traffic of the cluster but also may be much longer than the rest of the BH links. Therefore, the complete switch off of a cluster corresponds to the highest energy efficiency gain. Nevertheless, for even higher traffic (110 UEs), given the capacity constraints of the eNB, it is not possible to associate the majority of UEs to the eNB and thus to switch off the SC clusters. Consequently, in order to avoid having a significant BH power consumption increase, only the UEs with very high QoS demands get associated to the eNB. As a result, the AN power consumption decreases significantly, whereas the BH power consumption increases. Finally, as depicted in Fig. 4.25 and 4.26, both ε-constraint SE and the proposed algorithm achieve a good balance between AN and BH power consumption for all BH technologies.

Hotspot traffic distribution

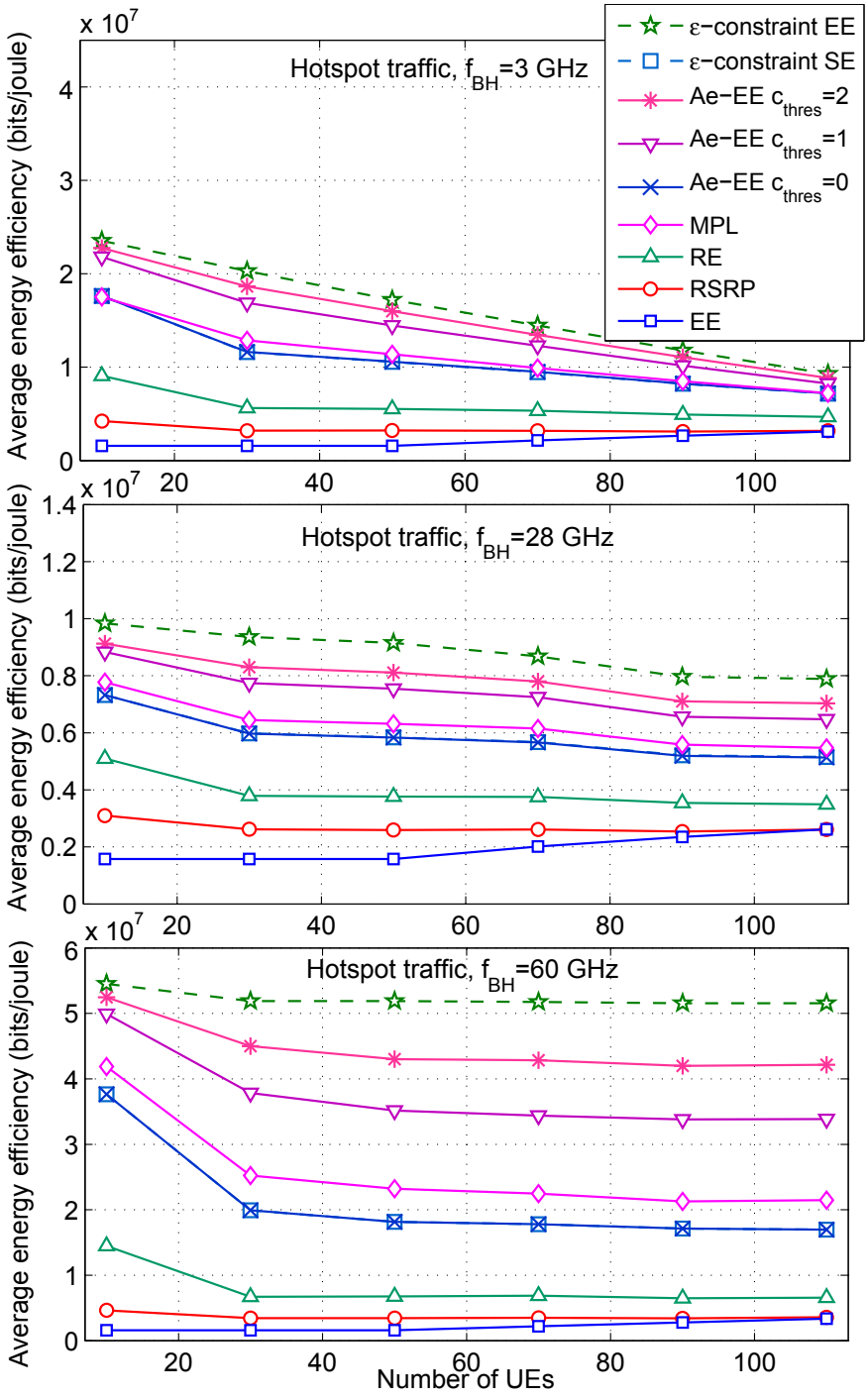


Figure 4.27: Average total network energy efficiency for different N values and BH technologies with frequency equal to 3 GHz, 28 GHz and 60 GHz, respectively, when the UEs form hotspots.

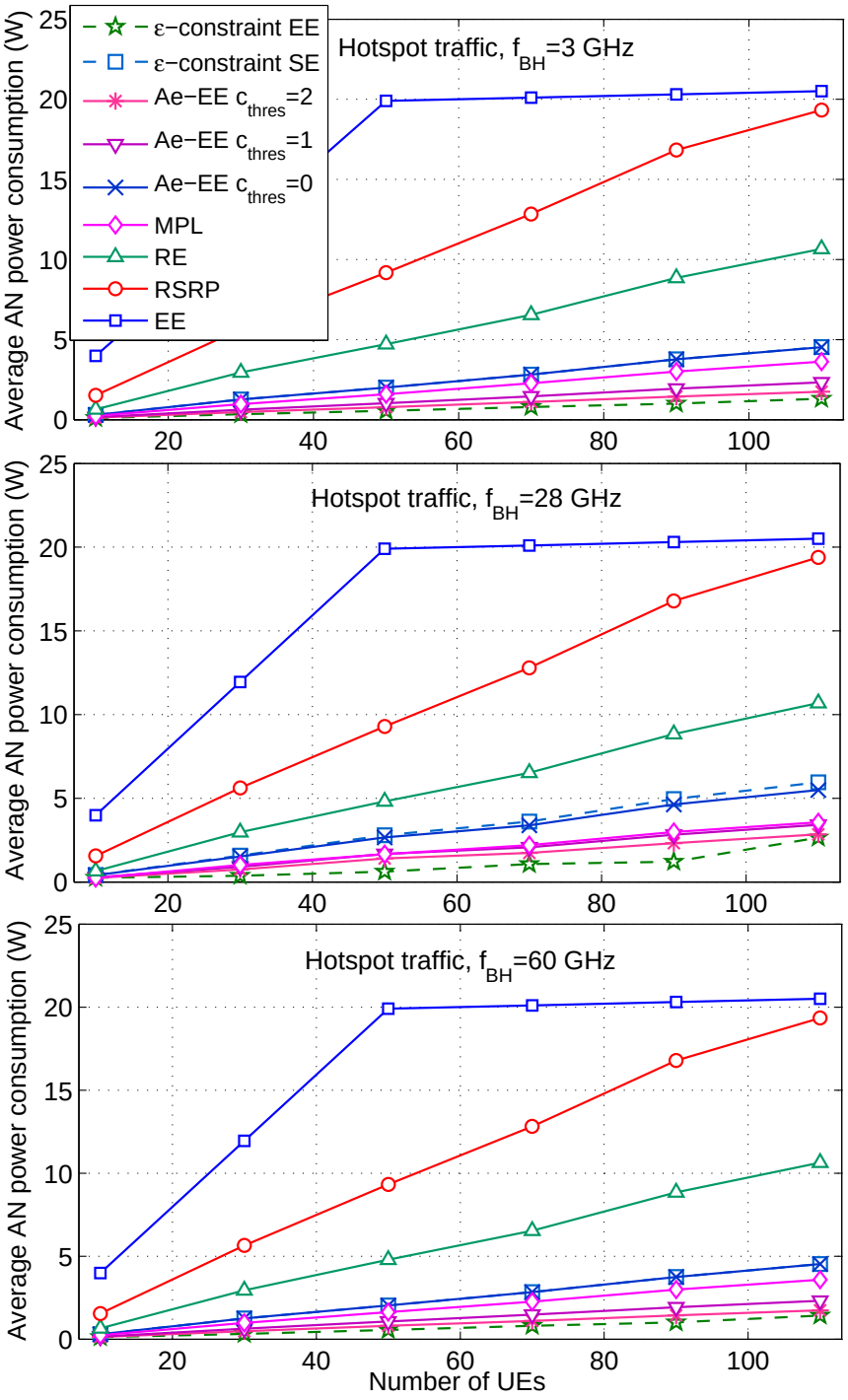


Figure 4.28: Average total access network power consumption for different N values and BH technologies with hotspot traffic.

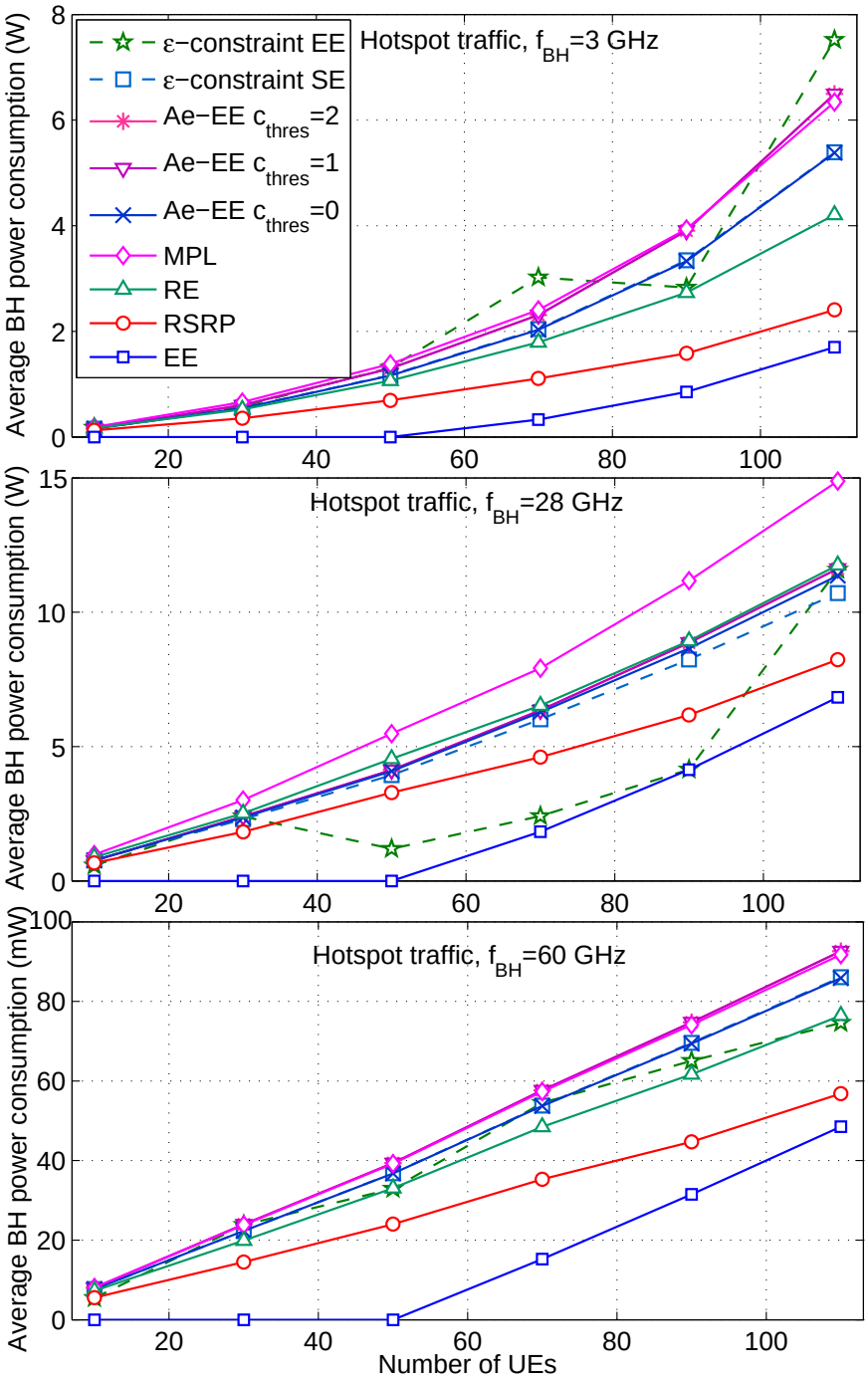


Figure 4.29: Average total backhaul power consumption for different N values and BH technologies with hotspot traffic.

Table 4.6: ε-constraint User Association: Average Network Spectral Efficiency

	Uniform (bps/Hz)			Hotspot (bps/Hz)		
User association algorithm	**3 GHz**	**28 GHz**	**60 GHz**	**3 GHz**	**28 GHz**	**60 GHz**
ε-constraint SE, Ae-EE $c_{thres} = 0$, EE, RSRP	1.74	1.74	1.74	1.74	1.74	1.74
RE	1.73	1.73	1.73	1.74	1.74	1.74
MPL	1.49	1.49	1.49	1.65	1.65	1.65
Ae-EE $c_{thres} = 1$	1.49	1.49	1.51	1.66	1.62	1.67
Ae-EE $c_{thres} = 2$	1.39	1.30	1.41	1.61	1.53	1.62
ε-constraint EE	1.22	0.95	1.27	1.53	1.18	1.56

Accordingly, in Fig. 4.27, the average network energy efficiency of all algorithms is depicted in a hotspot scenario for all different BH technologies. In this case, the proposed algorithm can achieve up to 14.5, 6, and 33 times higher energy efficiency than the rest of the algorithms for 3, 28 and 60 GHz, respectively. To gain again further insights, the AN and BH power consumption of the hotspot scenario are depicted in Fig. 4.28 and 4.29, respectively. Compared to the uniform traffic scenario, it can be noticed that the BH power consumption increases for most algorithms, since more UEs are closer to SCs and thus get associated with them. For the same reason, the AN power consumption decreases for the majority of the algorithms, as depicted in Fig. 4.28.

Moreover, we can notice that ε-constraint EE has a different behavior than in the uniform scenario. This stems from the fact that in the hotspot scenario more BH traffic is in general generated and thus the BH energy consumption is higher. Therefore, for a specific traffic, which differs for different technologies (N=90 UEs for 3 GHz and N=50 UEs for 28 GHz), ε-constraint EE favors the association of a portion of UEs with the eNB in order to switch off completely one of the clusters. Thereby, a significant increase in the BH power consumption is avoided.

Regarding the rest of the algorithms, it is shown that ε-constraint SE achieves again a good balance between AN and BH power consumption and so does the proposed algorithm. The EE algorithm has still very high AN power consumption. This stems from the fact that it generates globally less traffic for the BH links and thus the BH energy consumption becomes much lower at the expense of higher AN power consumption. Then, regarding RE, it has poor energy efficiency performance, as there are more UEs associated to the SCs than in RSRP and thus it presents higher BH energy consumption. Finally, MPL presents the lowest energy efficiency, as most UEs are associated with the closest SCs, resulting in the most highly loaded BH links (see Fig. 4.29) and thus in the highest BH energy consumption for all BH technologies.

In Table 4.6, the average network spectral efficiency is presented for all algorithms and BH technologies. As it can be observed, the considered algorithms that aim at the maximization of the spectral efficiency (i.e., ε-constraint SE, Ae-EE with $c_{thres} = 0$, EE, RSRP) achieve the highest spectral efficiency for all BH technologies, since the UEs are connected to the BSs that require the minimum spectrum resources for their QoS requirements to be fulfilled. On the contrary, RE achieves slightly lower spectral efficiency in the case of uniform traffic, as the UEs, due to biasing, may not be connected to their best choice in terms of spectral efficiency. It is also worth pointing out that the RE spectral efficiency is higher in the hotspot sce-

nario, since, in that case, both biasing and maximum SINR favor the association of UEs with SCs. MPL, unlike the rest of the algorithms, presents much lower spectral efficiency, since it associates the UEs independently of their SINR. Hence, it is very likely that a UE is associated to a BS with low SINR, thus requiring more spectrum resources to achieve the same throughput. This holds also for Ae-EE $c_{thres} = 1, 2$ since energy efficiency is increased at the expense of lower spectral efficiency, which becomes even lower in ε-constraint EE, where energy efficiency is maximized. Furthermore, it is worth noticing that the spectrum efficiency of RSRP, RE and MPL is independent of the employed BH technology, since these user association algorithms take into account only the AN conditions. However, for the proposed solutions, it can be observed that 28 GHz results in lower spectral efficiency than the other BH technologies, since it involves higher BH power consumption and thus less UEs get connected to their closest SCs.

4.6 Conclusions

In the first part of this chapter, the BH energy impact in future green HetNets was studied under different UE traffic scenarios and BH technologies. To that end, some interesting conclusions were drawn and are summarized next:

- It was shown that the BH constitutes a significant part of the total power consumption, which becomes more important as the number of UEs (i.e., total traffic) increases. The BH energy impact was also shown to be dependent on the UE traffic distribution. Specifically, for hotspot traffic the BH energy impact becomes more significant, since more UEs get connected to SCs (i.e., the BH traffic increases). At the same time, the fact that future HetNets are expected to deal with very high amount of hotspot traffic, predicates the need for backhaul-aware algorithm design.

- Finally, it was shown that mmWave, which is considered an appropriate BH solution in terms of capacity due to its high bandwidth availability, is also the best solution in terms of energy efficiency. However, given that it requires LOS, a mixture of the studied BH technologies is anticipated for future SC HetNets.

Moreover, the user association problem was studied aiming at the joint maximization of energy efficiency and spectral efficiency of the network, without compromising the UE QoS requirements. In this context, from the results presented in this chapter, the following conclusions can be derived.

- An energy-efficient user association algorithm was proposed, which exploits the available context-aware information to assign the UEs to the candidate cell with the least number of hops in order to minimize the BH energy consumption or in case there are more candidate cells with the same number of

hops, to the candidate with the least loaded BH route to achieve load balancing. The performance of the proposed algorithm was evaluated under two simulation scenarios and it was shown that it can achieve up to two and a half times more network energy efficiency than the reference algorithms, without compromising the UE QoS requirements.

- In the same setup, the joint UL-DL user association problem was studied, aiming at the total UL-DL energy efficiency maximization, while maintaining high spectral efficiency and low UE power consumption. Given that the considered optimization problem was shown to be NP-hard, the previously proposed energy-efficient context-aware algorithm was adapted accordingly to minimize the BH energy consumption both in UL and DL. The algorithm performance was evaluated by means of simulation and it was shown that it can achieve up to 45% energy efficiency gain.

- The DL user association problem was formulated as a generalized assignment problem, which was shown to be NP-hard. Therefore, an upper bound on the network performance was derived. Millimeter wave BH links were employed and it was shown that there is still room for energy efficiency improvement, which predicates the need for new energy-efficient BH-aware user association solutions. To that end, the previously proposed energy-efficient context-aware algorithm was enhanced to take into account the possibility of having heterogeneous BH links, by considering the actual power consumption of each BH link and not just the number of hops. Thereby, the proposed algorithm was shown to present less dependency on the employed scenario. In particular, in the simulation results, e-EE was shown to significantly outperform its counterparts, while achieving near-optimal performance.

- Another DL user association formulation was proposed (i.e., as an ε-constraint problem), aiming at the analytical study of the trade-off between energy efficiency and spectral efficiency, while the exact Pareto front points of the problem were derived for different BH technologies. Thereby, a framework was developed, which can be used as a benchmark for the performance evaluation of user association algorithms. Moreover, the enhanced low-complexity energy-efficient algorithm, which was previously proposed, was modified accordingly to be able to achieve performance between to the two extreme Pareto front solutions. In particular, the proposed adaptive algorithm was shown to be able to select any point of the Pareto front, by tuning the parameter c_{thres} accordingly, and thus, to achieve a *good* trade-off between the aforementioned metrics. The proposed algorithm was also compared with existing user association solutions under different BH technologies. The provided results indicated that i) the proposed adaptive algorithm (Ae-EE) is able to achieve notable energy and spectrum efficiency gains and that ii) mmWave is a promising solution to provide high capacity and low energy consumption multi-hop BH.

Chapter 5

Conclusions and Future Work

This chapter completes the dissertation by summarizing its main contributions, while also providing some potential research lines for future investigation. In particular, Section 5.1 contains the most significant concluding remarks of each chapter, while Section 5.2 outlines the open research issues related to our contributions.

5.1 Conclusions

In order to meet the ever-increasing high data traffic demands, dramatic expansion of network infrastructures as well as fast escalation of energy demands are expected. As a result, it becomes urgent for mobile operators not only to maintain sustainable capacity growth, but also to limit their electric bill. In parallel, the spectrum scarcity problem stresses the need for spectral efficient solutions. The aforementioned goals can be summarized into the joint maximization of energy and spectrum efficiency. This joint goal constitutes a fundamental design objective for next generation networks and therefore was the main goal of this thesis.

To that end, exploiting cognition was shown to play a key role. In particular, this thesis proposed and evaluated medium access layer algorithms, that exploit different types of cognition to provide energy and spectrum efficiency enhancement. In particular, two main research directions were followed: i) the first has focused on spectrum-awareness in cognitive radio (CR) networks, where the use of licensed bands by unlicensed users (also called secondary users (SUs)) is enabled for as long as they remain unused, thereby achieving high spectrum efficiency, and ii) the second on the context-aware self-adaptation of cellular HetNets (i.e., SONs). In both contexts, the exploitation of the available information was shown to play a key role on providing sustainable wireless networks. However, given the different needs and features that characterize each of these networks, they were studied separately.

Regarding the first research direction, it was emphasized that the opportunistic spectrum sharing, on which secondary networks' (SNs)' operation is based, relies upon two main premises: the protection of the primary users' (PUs') transmissions and the maximization of the spectrum usage. The former is achieved by applying effective sensing techniques (cooperative or not). The maximization of the spectrum usage, though, can only be met by implementing efficient coexistence mechanisms among SNs, particularly in congested environments. Hence, mechanisms for efficient coexistence of more than a single SN are indispensable. The key point of such an efficient coexistence is that the contention of two or more SNs over the same channel is allowed, but it impacts decisively on the achievable throughput and energy efficiency.

Therefore, a CR-based MAC protocol should i) detect the licensed channels without PU activity, and ii) prioritize the access to the channels with low SU contention. To that end, in the first research contribution part of this thesis, the following contributions were provided:

- A SN coexistence scheme was proposed as well as a novel contention-aware channel selection algorithm, which: i) exploits cooperative spectrum sensing to detect the free from PU activity licensed channels, ii) for each one of them, it estimates the probability of collision, and iii) selects the less contended (i.e., with the lowest probability of collision) to access first.
- An analytical model for the throughput and the energy efficiency of the SN under study has been provided, which has been validated by means of simulation. Furthermore, the effect of the time between two consecutive sensing periods on the aforementioned metrics have been studied. It has been proved that there is an optimal value for maximum performance for the time between two consecutive sensing periods, which is highly dependent on the PU activity pattern.
- The proposed channel selection algorithm has been compared with three relevant state-of-the-art algorithms. Simulation results have shown that the proposed algorithm significantly outperforms its counterparts both in terms of throughput and energy efficiency.
- Finally, after having shown the notable gains of the proposed channel selection algorithm compared to the state-of-the-art and given that in such coexistence scenarios, it is fundamental to guarantee fairness among the coexisting SNs, the proposed SN coexistence scheme was evaluated, when the proposed channel selection algorithm is applied. In particular, the proposed SN coexistence scheme was compared with the reference approach (FMAC) by means of simulations and it was shown that it can achieve throughput and energy efficiency gains, while maintaining or even achieving better fairness among the coexisting SNs. Furthermore, the impact of different minimum back-off window values and different PU activity patterns on the performance of both considered coexistence schemes was studied. The maximum gain is achieved for the lowest minimum back-off window value and for high contention in the licensed channels.

The second research direction has focused on the proposal of energy-efficient user association algorithms and analytical frameworks for cognitive heterogeneous networks (HetNets). This is a timely topic since dense deployment of small cells (SCs), overlaying the existing macrocell networks, is expected during the next few years. This is due to the fact that the deployment of SCs reduces the distance between user equipments (UEs) and base stations (BSs) and, consequently, i) the area spectral efficiency (bps/Hz/m^2) increases, and ii) the energy consumption in the access network (AN), i.e., the links between the UEs and their serving BSs, decreases.

However, the dense deployment of SCs also poses new challenges. Due to the high number of deployed SCs, the direct connection of all SCs to the core network becomes complicated. Fiber connections, which have been traditionally considered as the best backhaul (BH) solution, are prohibitive in this case due to their high deployment cost. A promising solution lies in exploiting the existing connection between the macrocell site and the core network (most of the times it is a fiber connection), and to provide core network connectivity to SCs through the macrocell site. Still, in order to connect the SCs to the macrocell site (thus providing them core network connectivity), new cheap wireless BH solutions are required.

In addition, this wireless BH is expected to provide high-capacity services from the SCs to the core network, in order to meet the expected traffic demands of the order of Gpbs. Therefore, a promising solution that could provide wireless BH connectivity between the SCs and the core network lies in using millimeter wave (mmWave) frequencies, because of the large amount of available bandwidth at these frequencies, which results in high capacity connections. It has been shown, however, that mmWave frequencies are capable of providing good coverage performance only if the transmission distance is shorter than 200 meters. Otherwise, links may not be established. Since the macro cell radius is even in dense deployments of the order of 500 meters, this implies that a multi-hop architecture is needed, in order to allow each of the SCs to reach the macrocell site (i.e., macrocell aggregation gateway).

In this context, user association becomes challenging due to the multi-hop BH architecture and therefore new optimal solutions need to be developed aiming at the joint energy and spectrum efficiency maximization of the network. To that end, in the second research part of this thesis, the following contributions were provided:

- The role of BH in future outdoor HetNets was studied aiming to answer to whether or not it could constitute an energy bottleneck for the HetNet. In particular, the BH energy impact was compared to the access network (AN), i.e., the links between the users and their serving cells, under different traffic distribution scenarios and BH technologies. It was shown that the BH constitutes a significant part of the total power consumption, which becomes more important as the number of UEs (i.e., total traffic) increases. The BH energy impact was also shown to be dependent on the UE traffic distribution. Specifically, for hotspot traffic the BH energy impact becomes more significant, since more UEs get connected to SCs (i.e., the BH traffic increases). At the same time, the fact that future HetNets are expected to deal with very high amount

of hotspot traffic, predicates the need for backhaul-aware algorithm design. Finally, it was shown that mmWave, which is considered an appropriate BH solution in terms of capacity due to its high bandwidth availability, is also the best solution in terms of energy efficiency. However, given that it requires LOS, a mixture of the studied BH technologies is anticipated for future SC HetNets.

- The user association problem is studied aiming at the joint maximization of energy efficiency and spectral efficiency of the network, without compromising the user equipment (UE) quality of service (QoS) requirements. In this framework, the following additional contributions were provided:
  - An energy-efficient user association algorithm was proposed, which exploits the available context-aware information to assign the UEs to the candidate cell with the least number of hops in order to minimize the BH energy consumption or in case there are more candidate cells with the same number of hops, to the candidate with the less loaded BH route to achieve load balancing. The performance of the proposed algorithm was evaluated under two simulation scenarios and it was shown that it can achieve up to two and a half times more network energy efficiency than the reference algorithms.
  - In the same setup, the joint UL-DL user association problem was studied, aiming at the total UL-DL energy efficiency maximization, while maintaining high spectral efficiency and low UE power consumption. Given that the considered optimization problem was shown to be NP-hard, the previously proposed energy-efficient context-aware algorithm was adapted accordingly to minimize the BH energy consumption both in UL and DL. The algorithm performance was evaluated by means of simulation and it was shown that it can achieve up to 45% energy efficiency gain.
  - The DL user association problem was formulated as a generalized assignment problem, which was shown to be NP-hard. Therefore, an upper bound on the network performance was derived. Millimeter wave BH links were employed and it was shown that there is still room for energy efficiency improvement, which predicates the need for new energy-efficient BH-aware user association solutions. To that end, the previously proposed energy-efficient context-aware algorithm was enhanced to take into account the possibility of having heterogeneous BH links. Thereby, the proposed algorithm was shown to present less dependency on the employed scenario. In particular, in the simulation results, e-EE was shown to significantly outperform its counterparts, while achieving near-optimal performance.
  - Another formulation was also proposed (i.e., as an ε-constraint problem), aiming at the analytical study of the trade-off between energy efficiency and spectral efficiency, while the exact Pareto front points of the problem were derived for different BH technologies. Thereby, a framework was developed, which can be used as a benchmark for the performance

evaluation of user association algorithms. Moreover, the enhanced low-complexity energy-efficient algorithm, which was previously proposed, was modified accordingly to be able to achieve performance between to the two extreme Pareto front solutions. In particular, the proposed adaptive algorithm was shown to be able to select any point of the Pareto front, by tuning the parameter c_{thres} accordingly, and thus, to achieve a *good* trade-off between the aforementioned metrics. The proposed algorithm was also compared with existing user association solutions under different BH technologies. The provided results indicated that i) the proposed adaptive algorithm (Ae-EE) is able to achieve notable energy and spectrum efficiency gains and that ii) mmWave is a promising solution to provide high capacity and low energy consumption multi-hop BH.

5.2 Future Work

The research contributions presented in this work has opened several new lines for future investigation. The main goals for future work with respect to the first part of the thesis on SN coexistence schemes for CR networks can be summarized as follows:

- So far, the performance of the contention-aware channel selection algorithm has been evaluated through an analytical framework and with the help of extensive simulations. An important step forward would be the hardware implementation of the algorithm on a testbed. This would provide further insight on the SNs' operation, it would permit the practical selection and fine tuning of several parameter values and would, without doubt, open the road to many interesting experiments.
- As CR networks are wireless in nature, they face all common security threats found in the traditional wireless networks. Nevertheless, they also face new security threats and challenges (i.e., PU emulation attack (PUEA), spectrum sensing data falsification (SSDF) attack, common control channel threats and vulnerabilities) that have arisen due to their unique cognitive characteristics. Therefore, another interesting line of investigation involves the enhancement of the proposed channel selection algorithm in terms of security. In this context, the joint study of security and energy efficiency needs to be considered, since most of the times providing security to the network comes at the cost of higher energy consumption (longer sensing time, additional overhead, processing burden, etc.). It is worth noting that in scenarios where various SNs coexist and share the same PU resources, guarantying network security becomes even more challenging as there is a need of more advanced sensing techniques that require further sensing overhead and thus energy consumption. Therefore, algorithms that will provide the CR network with the highest achievable security at a plethora of attacking threats while aiming at maximizing its energy efficiency in scenarios where various SNs share the same PU resources need still to be developed.

- The presented results have shown that the proposed channel selection algorithm can improve the total system throughput and energy efficiency but there is still a margin for performance improvement. Future work can be focused on the joint study of spectrum sensing and spectrum access. Although spectrum sensing optimization has already attracted a lot of research attention (e.g., by improving the accuracy of the applied spectrum sensing technique, and/or by applying/proposing more sophisticated fusion rules during cooperative spectrum sensing), their joint optimization with the applied spectrum access method could result in significant overall performance gains.

There are also several open issues regarding the second part of this thesis, focused on energy-efficient context-aware user association in cognitive HetNets:

- It would be interesting to extend the work presented in this thesis by taking also into account the fixed part of the power consumption of the BSs and BH links. To that end, efficient switching off algorithms could be proposed that would aim at the joint energy and spectrum efficiency maximization. In this context, it would be interesting to consider the possibility of switching off BH links and provide alternative BH traffic routes. These type of algorithms could be used also for self-healing, i.e., in case a BH link fails.

- The joint analysis of user association with resource allocation is another interesting field to be studied. By jointly optimizing the user association and the frequency allocation scheme higher performance gains can be achieved. As a result, different frameworks will be needed, which could exploit e-ICIC to provide higher performance. In this context, different scheduling policies (e.g., giving priority to UEs with delay sensitive type of traffic) could also be employed for QoS provisioning and fairness among UEs.

- The analysis provided in this thesis has considered a multi-hop BH architecture and GBR UE traffic demands. Nevertheless, in a realistic scenario, the BH architecture and the UE traffic may differ in different locations or they may change over time. In order to fully exploit the potential of the proposed schemes, it would be interesting to evaluate their performance under different BH architectures and UE traffic pattern models. To that end, it would be interesting to use stochastic geometry tools to derive parameters of the network such as coverage probability and throughput by modeling the location of the UEs and BSs as Poisson point processes.

- The mmWave frequencies, as it was also shown in this thesis, are expected to be a key technology for next generation networks. Although in this work mmWave has been used as an out-of-band BH technology, it would be interesting to study its performance when mmWave is employed also for the AN. In this case, however, there are new challenges to be addressed, e.g., frequent handovers, due to short link length that is required. To that end, new and realistic channel models for mmWave propagation are still to be developed.

- Finally, similar to the contention-aware channel selection algorithm case, the implementation of the proposed user association techniques in real hardware

would overcome the inherent limitations of theoretical analysis and system-level simulations. After detailed analytical and simulation studies, we are convinced that the implementation of the algorithms in real testbeds would further highlight the need for designing novel, adaptive context and BH-aware user association algorithms for next generation networks that will aim at the joint energy and spectrum efficiency maximization of the network.

Concluding, this thesis has advanced the state of the art first by presenting a novel contention-aware channel selection algorithm for CR networks, and second, by proposing efficient context-aware user association algorithms for cognitive HetNets. The two parts of the thesis have provided valuable lessons on the MAC layer design. Even though they have been treated independently throughout this dissertation, it is possible to envision a system where both parts are combined. This joint scenario could consist of a HetNet, where UEs or even SCs could use the licensed bands as long as they are unused, thereby resulting in further spectral efficiency gains. The road ahead lies open for further research following the new lines of investigation that have been identified.

Bibliography

[1] "Cisco visual networking index: global mobile data traffic forecast update," tech. rep., 2013-2018, Feb. 2014.

[2] "Spectrum policy," tech. rep., Task Force Report, ET Docket 02-155, Nov. 2002.

[3] J. Mitola and G. J. Maguire, "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, pp. 13–18, Aug. 1999.

[4] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 201–220, Feb. 2005.

[5] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/ dynamic spectrum access/cognitive radio wireless networks: a survey," *Computer Networks Journal (Elsevier)*, vol. 50, pp. 2127–2159, Sep. 2006.

[6] J. G. Andrews, "Seven ways that HetNets are a cellular paradigm shift," *IEEE Communications Magazine*, vol. 51, pp. 136–144, Mar. 2013.

[7] C. Yan, S. Zhang, S. Xu, and G. Y. Li, "Fundamental trade-offs on green wireless networks," *IEEE Communications Magazine*, vol. 49, pp. 30–37, Jun. 2011.

[8] "5G White Paper - Executive Version," tech. rep., NGMN Alliance, Dec. 2014.

[9] P. Makris, D. Skoutas, and C. Skianis, "A survey on context-aware mobile and wireless networking: on networking and computing environments' integration," *IEEE Communication Surveys &Tutorials*, vol. 15, pp. 362–386, First 2013.

[10] T. Yucek and H. Arsian, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Communication Surveys &Tutorials*, vol. 11, pp. 116–130, First 2009.

[11] B. Wang and K. Liu, "Advances in cognitive radio networks: a survey," *IEEE Journal on Selected Topics in Signal Processing*, vol. 5, pp. 5–23, Feb. 2011.

[12] F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: a survey," *Physical Communications*, vol. 4, pp. 40–62, Mar. 2011.

[13] J. Lunden, V. Koivunen, A. Huttunen, and H. V. Poor, "Collaborative cyclostationary spectrum sensing for cognitive radio systems," *IEEE Transactions on Signal Processing*, vol. 57, pp. 4182–4195, Nov. 2009.

[14] F. Adelantado, A. Juan, and C. Verikoukis, "Adaptive sensing user selection mechanism in cognitive wireless networks," *IEEE Communication Letters*, vol. 14, pp. 800–802, Sep. 2010.

[15] Y. Zheng, X. Xie, and L. Yang, "Cooperative spectrum sensing based on SNR comparison in fusion center for cognitive radio," in *Proc. of International Conference on Advanced Computer Control*, pp. 212–216, Jan. 2009.

[16] Y.-C. Liangen, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 1326–1337, Apr. 2008.

[17] S. Stotas and A. Nallanathan, "Optimal sensing time and power allocation in multiband cognitive radio networks," *IEEE Transactions on Communications*, vol. 59, pp. 226–235, Jan. 2011.

[18] D. Cabric, S. M. Mishra, and R. W. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," in *Proc. of 38th Asilomar Conference on Signals, Systems and Computers*, pp. 772–776, Nov. 2004.

[19] P. Kaligineedi and V. K. Bhargava, "Sensor allocation and quantization schemes for multiband cognitive radio cooperative sensing system," *IEEE Transactions on Wireless Communications*, vol. 10, pp. 284–293, Jan. 2011.

[20] S. Appadwedula, V. Veeravalli, and D. Jones, "Decentralized detection with censoring sensors," *IEEE Transactions on Signal Processing*, vol. 56, pp. 1362–1373, Apr. 2008.

[21] Z. Quan, S. Cui, and A. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE Journal on Selected Topics in Signal Processing*, vol. 2, pp. 28–40, Feb. 2008.

[22] E. Peh, Y.-C. Liang, Y. L. Guan, and Y. Zeng, "Cooperative spectrum sensing in cognitive radio networks with weighted decision fusion schemes," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3838–3847, Dec. 2010.

[23] W. Han, J. Li, Z. Tian, and Y. Zhang, "Efficient cooperative spectrum sensing with minimum overhead in cognitive radio," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3006–3011, Oct. 2010.

[24] W. Zhang, R. Mallik, and K. Letaief, "Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 8, pp. 5761–5766, Dec. 2009.

[25] X. Feng, X. Gan, and X. Wang, "Energy-constrained cooperative spectrum sensing in cognitive radio networks," in *Proc. of IEEE Global Telecommunications Conference (GLOBECOM 2011)*, pp. 1–5, Dec. 2011.

[26] E. Peh and L. Ying-Chang, "Optimization for cooperative sensing in cognitive radio networks," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC 2007)*, pp. 27–32, Mar. 2007.

[27] S. Chunhua, W. Zhang, and K. Letaief, "Cooperative spectrum sensing for cognitive radios under bandwidth constraints," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC 2007)*, pp. 1–5, Mar. 2007.

[28] Z. Xiangwei *et al.*, "Bandwidth efficient combination for cooperative spectrum sensing in cognitive radio networks," in *Proc. of IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 3126–3129, Mar. 2010.

[29] E. C. Y. Peh, Y. C. Liang, Y. L. Guan, and Y. Pei, "Energy-efficient cooperative spectrum sensing in cognitive radio networks," in *Proc. of IEEE Global Telecommunications Conference (GLOBECOM 2011)*, pp. 1–5, Dec. 2011.

[30] B. Gao, J. M. Park, Y. Yang, and S. Roy, "A taxonomy of coexistence mechanisms for heterogeneous cognitive radio networks operating in TV white spaces," *IEEE Wireless Communications Magazine*, vol. 19, pp. 41–48, Aug. 2012.

[31] D. T. C. Wong and F. Chin, "Sensing-saturated throughput performance in multiple cognitive CSMA/CA networks," pp. 1–5, May 2010.

[32] Y. Zhao, M. Song, and C. Xin, "FMAC: A fair MAC protocol for coexisting cognitive radio networks," in *Proc. of the 32nd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2013)*, pp. 1474–1482, Apr. 2013.

[33] L. Le and E. Hossain, "OSA-MAC: a MAC protocol for opportunistic spectrum access in cognitive radio networks," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC 2008)*, pp. 1426–1430, Apr. 2008.

[34] M. Hossian, A. Mahmood, and R. Jantti, "Channel ranking algorithms for cognitive coexistence of IEEE 802.15.4," in *Proc. of IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2009)*, pp. 112–116, Sep. 2009.

[35] N. Jain and S. Das, "A multichannel CSMA MAC protocol with receiver-based channel selection for multihop wireless networks," in *Proc. of 10th International Conference on Computer Communications and Networks (IC3N)*, pp. 432–439, Oct. 2001.

[36] "Backhaul technologies for small cells: use cases, requirements and solutions," tech. rep., Small Cell Forum, 049.04.02, Feb. 2013.

[37] "Wifi and femtocell integration strategies 2011-2015," tech. rep., Juniper research white paper, Mar. 2011.

[38] "More outdoor small cells will ship in 2012 than macrocells," tech. rep., ABI research, 2012.

[39] "EUTRA & EUTRAN overall description," tech. rep., 3GPP TS 36.300, Mar. 2011.

[40] "Small cell backhaul requirements," tech. rep., NGMN Alliance, Jun. 2012.

[41] "Small cells address the growing demand for data," tech. rep., Tellabs, Sep. 2012.

[42] "Radio frequency (RF) system scenarios," tech. rep., 3GPP TS 36.942, Sep. 2012.

[43] D. Fooladivanda and C. Rosenberg, "Joint resource allocation and user association for heterogeneous wireless cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 248–257, Oct. 2012.

[44] H. Dhillon, R. Ganti, F. Baccelli, and J. Andrews, "Modeling and analysis of K-Tier downlink heterogeneous cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, pp. 550–560, Apr. 2012.

[45] "DL pico/macro HetNet performance: cell selection," tech. rep., Alcatel-Lucent, R1-100945, 3GPP RAN1 Meeting 60, Feb. 2010.

[46] H. Boostanimehr and V. K. Bhargava, "Unified and distributed QoS-driven cell association algorithms in heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 14, pp. 1650–1662, Mar. 2015.

[47] Y. Lin, W. Bao, W. Yu, and B. Liang, "Unified and distributed qos-driven cell association algorithms in heterogeneous networks," *IEEE Journal on Selected Areas in Communications, Recent Advances in Heterogeneous Cellular Networks*, 2015.

[48] C. Liu, P. Whiting, and S. V. Hanly, "Joint resource allocation and user association in downlink three-tier heterogeneous networks," in *Proc. of IEEE Global Telecommunications Conference (GLOBECOM 2014)*, pp. 4232–4238, Dec. 2014.

[49] Q. Ye *et al.*, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 2706–2716, Jun. 2013.

[50] S. Moon, Y. Yi, and H. Kim, "Energy-efficient user association in cellular networks: a population game approach," in *Proc. of 11th International Symposium on Modeling & Optimization in Mobile, Ad Hoc & Wireless Networks (WiOpt)*, pp. 388–395, May 2013.

[51] H. Galeana-Zapien and R. Ferrus, "Design and evaluation of a backhaul-aware base station assignment algorithm for OFDMA-based cellular networks," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3226–3237, Oct. 2010.

[52] S. Corroy, L. Falconetti, and R. Mathar, "Cell association in small heterogeneous networks: downlink sum rate and min rate maximization," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC 2012)*, pp. 888–892, Apr. 2012.

[53] Q. Li, R. Q. Hu, G. Wu, and Y. Qian, "On the optimal mobile association in heterogeneous wireless relay networks," in *Proc. of the 31st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2012)*, pp. 1359–1367, Mar. 2012.

[54] N. Wang, E. Hossain, and V. K. Bhargava, "Joint downlink cell association and bandwidth allocation for wireless backhauling in two-tier HetNets with large-scale antenna arrays," *IEEE Transactions on Wireless Communications*, to appear in 2015.

[55] S. Corroy, L. Falconetti, and R. Mathar, "Cell association in small heterogeneous networks: downlink sum rate and min rate maximization," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC 2012)*, pp. 888–892, Apr. 2012.

[56] R. Misra and A. P. Kannu, "Optimal sensing-order in cognitive radio networks with cooperative centralized sensing," in *Proc. of IEEE International Conference on Communications (ICC 2010)*, pp. 566–1570, May 2012.

[57] H. Jiang, L. Lai, R. Fan, and H. V. Poor, "Optimal selection of channel sensing order in cognitive radio," *IEEE Transactions on Wireless Communications*, vol. 8, pp. 297–307, Jan. 2009.

[58] Z. Khan, J. J. Lehtomaki, L. A. DaSilva, and M. Latva-Aho, "Autonomous sensing order selection strategies exploiting channel access information," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 274–288, Feb. 2013.

[59] C. Sun *et al.*, "Optimizing the coexistence performance of secondary-user networks under primary-user constraints for dynamic spectrum access," *IEEE Transactions on Vehicular Technology*, vol. 61, pp. 3665–3676, Oct. 2012.

[60] K. Ishizu, H. Murakami, and H. Harada, "TV white space database for coexistence of primary-secondary and secondary-secondary systems in mesh networking," in *Proc. of 15th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pp. 118–122, Sep. 2012.

[61] Z. Ma, Z. Cao, and W. Chen, "A fair opportunistic spectrum access (FOSA) scheme in distributed cognitive radio networks," in *Proc. of IEEE International Conference on Communications (ICC 2008)*, pp. 4054–4058, May 2008.

[62] N. Shankar, *Cognitive radio, software defined radio, and adaptive wireless systems*. Springer Netherlands, Apr. 2007.

[63] J. Tang, S. Misra, and G. Xue, "Joint spectrum allocation and scheduling for fair spectrum sharing in cognitive radio wireless networks," *Computer Networks*, vol. 52, no. 11, pp. 2148–2158, 2008.

[64] B. Wang, Z. Ji, and K. Liu, "Primary-prioritized Markov approach for dynamic spectrum access," in *Proc. of the 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2007)*, pp. 507–515, Apr. 2007.

[65] Y. Zhao, M. Song, C. Xin, and M. Wadhwa, "Spectrum sensing based on three-state model to accomplish all-level fairness for co-existing multiple cognitive radio networks," in *Proc. of the 31st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2012)*, pp. 1782–1790, Mar. 2012.

[66] D. Nam and H. Min, "An energy-efficient clustering using a round-robin method in a wireless sensor network," in *Proc. of 5th ACIS International Conference on Software Engineering Research, Management Applications, (SERA)*, pp. 54–60, Aug. 2007.

[67] H. Harada and I. Tinnirello, "A small-size software defined cognitive radio prototype," in *Proc. of IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2008)*, pp. 1–5, Sep. 2008.

[68] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 535–547, March 2000.

[69] L. T. Tan and L. Le, "Channel assignment with access contention resolution for cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 61, pp. 2808–2823, Jul. 2012.

[70] W. Wang, B. Kasiri, J. Cai, and A. Alfa, "Channel assignment schemes for cooperative spectrum sensing in multi-channel cognitive radio networks," *Wireless Communications Mobile Computing*, Oct. 2013.

[71] J. Kuri and S. K. Kasera, "Reliable multicast in multi-access wireless LANs," in *Proc. of the 18th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 1999)*, pp. 760–767, Mar. 1999.

[72] "IEEE standard for information technology – Telecommunications and information exchange between systems – Local and metropolitan area networks – Specific requirements – Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications," *IEEE Std 802.11-2012*, 2012.

[73] S. Rayanchu *et al.*, "Diagnosing wireless packet losses in 802.11: separating collision from weak signal," in *Proc. of the 27th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2008)*, pp. 1409–1417, Apr. 2008.

[74] L. Zhao, J. Zhang, and H. Zhang, "Using incompletely cooperative game theory in wireless mesh networks," *IEEE Network*, vol. 22, pp. 39–44, Jan./Feb. 2008.

[75] A. Mesodiakaki, F. Adelantado, L. Alonso, and C. Verikoukis, "Energy-efficient contention-aware channel selection in cognitive radio ad-hoc networks," in *Proc. IEEE 17th International Workshop onComputer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, pp. 46–50, Sep. 2012.

[76] D. Malone, K. Duffy, and D. Leith, "Modeling the 802.11 distributed coordination function in non-saturated heterogeneous conditions," *IEEE/ACM Transactions on Networking*, vol. 15, pp. 159–172, Feb. 2007.

[77] H. Kim and K. G. Shin, "In-band spectrum sensing in IEEE 802.22 WRANs for incumbent protection," *IEEE Transactions on Mobile Computing*, vol. 9, pp. 1766–1779, Dec. 2010.

[78] R. Jain, D.-M. Chiu, and W. R. Hawe, *A quantitative measure of fairness and discrimination for resource allocation in shared computer system.* Eastern Research Laboratory, Digital Equipment Corporation Hudson, MA, 1984.

[79] L. Wei, R. Hu, Y. Qian, and G. Wu, "Key elements to enable millimeter wave communications for 5G wireless systems," *IEEE Wireless Communications*, vol. 21, pp. 136–143, Jan. 2015.

[80] "Release three: urban foundations," tech. rep., Small Cell Forum, 103.03.01, Feb. 2014.

[81] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[82] A. Mesodiakaki, F. Adelantado, L. Alonso, and C. Verikoukis, "Energy-efficient user association in cognitive heterogeneous networks," *IEEE Communications Magazine*, vol. 52, pp. 22–29, Jul. 2014.

[83] D. Lopez-Perez, C. Xiaoli, and I. Guvenc, "On the expanded region of picocells in heterogeneous networks," *IEEE Journal on Selected Topics in Signal Processing*, vol. 6, pp. 281–294, Jun. 2012.

[84] "Small cell enhancements for E-UTRA and E-UTRAN - Physical layer aspects," tech. rep., 3GPP TS 36.872, Aug. 2013.

[85] "Evolved universal terrestrial radio access (E-UTRA), physical layer procedures," tech. rep., 3GPP TS 36.213, Dec. 2013.

[86] "Study on small cell enhancements for E-UTRA and E-UTRAN; Higher layer aspects," tech. rep., 3GPP TS 36.842, Dec. 2013.

[87] H. Holma and A. Toskala, *LTE for UMTS: OFDMA and SC-FDMA based radio access.* John Wiley & Sons, 2009.

[88] G. Auer *et al.*, "How much energy is needed to run a wireless network?," *IEEE Wireless Communications Magazine*, vol. 18, pp. 40–49, Oct. 2011.

[89] K. Brueninghaus *et al.*, "Link performance models for system level simulations for broadband radio access systems," in *Proc. of IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2005)*, pp. 2306–2311, Sep. 2005.

[90] A. J. Goldsmith, *Wireless communications*. Cambridge University Press, 2005.

[91] J. Zyren and A. Petrick, "Tutorial on basic link budget analysis," Jun. 1998.

[92] K.-C. Huang and Z. Wang, *Millimeter wave communication systems*. Wiley, 2011.

[93] "DL pico/macro HetNet performance: cell selection," tech. rep., Alcatel Lucent, R1-100945, 3GPP RAN1 Meeting 60, Feb. 2010.

[94] "Performance of eICIC with control channel coverage limitation," tech. rep., NTT DOCOMO, R1-103264, 3GPP Std., May 2010.

[95] H.-S. Jo, Y. Sang, P. Xia, and J. Andrews, "Outage probability for heterogeneous cellular networks with biased cell association," in *Proc. of IEEE Global Telecommunications Conference (GLOBECOM 2011)*, pp. 1–5, Dec. 2010.

[96] "Evaluation of rel-8/9 techniques and range expansion for macro and outdoor hotzone," tech. rep., Huawei, R1-100945, 3GPP RAN1 Meeting 61, May 2010.

[97] A. Damnjanovic *et al.*, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Communications*, vol. 18, pp. 10–21, Jun. 2011.

[98] "LTE advanced: heterogeneous networks," tech. rep., Qualcomm Incorporated, Feb. 2010.

[99] C. S. Chiu and C. C. Huang, "An interference coordination scheme for picocell range expansion in heterogeneous networks," in *Proc. of IEEE 75th Vehicular Technology Conference (VTC 2012 Spring)*, pp. 1–6, May 2012.

[100] S. Tombaz, Z. Zheng, and J. Zander, "Energy efficiency assessment of wireless access networks utilizing indoor base stations," in *Proc. of IEEE 24th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2013)*, pp. 3105–3110, Sep. 2013.

[101] S. Tombaz *et al.*, "Is backhaul becoming a bottleneck for green wireless access networks?," in *Proc. of IEEE International Conference on Communications (ICC 2014)*, pp. 4029–4035, Jun. 2014.

[102] A. Ghosh *et al.*, "Millimeter-wave enhanced local area systems: a high-data-rate approach for future wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, pp. 1152–1163, Jun. 2014.

[103] Y. Azar *et al.*, "28 GHz propagation measurements for outdoor cellular communications using steerable beam antennas in New York City," in *Proc. of IEEE International Conference on Communications (ICC 2013)*, pp. 5143–5147, Jun. 2013.

[104] V. Abhayawardhana *et al.*, "Comparison of empirical propagation path loss models for fixed wireless access systems," in *Proc. of IEEE 61st Vehicular Technology Conference (VTC 2005 Spring)*, pp. 73–77, May 2005.

[105] C. S. Chiu and C. C. Huang, "Asymmetric uplink-downlink assignment for energy-efficient mobile communication systems," in *Proc. of IEEE 75th Vehicular Technology Conference (VTC 2012 Spring)*, pp. 1–5, May 2012.

[106] R. Q. Hu *et al.*, "Mobile association in a heterogeneous network," in *Proc. of IEEE International Conference on Communications (ICC 2010)*, pp. 1–6, May 2012.

[107] "Physical layer procedures (FDD)," tech. rep., 3GPP TS 25.214, Sep. 2013.

[108] E. Dahlman, S. Parkvall, and J. Skold, *4G LTE/LTE-Advanced for mobile broadband*. Academic Press, 2011.

[109] S. Martello and P. Toth, *Chapter 7: Generalized assignment problem, Knapsack problems: algorithms and computer implementations*. John Wiley & Sons, 1990.

[110] O. J. Grodzevich and O. Romanko, "Normalization and other topics in multi-objective optimization," in *Proc. of the Fields-MITACS Industrial Problems Workshop (FM-IPSW)*, pp. 89–101, Aug. 2006.

[111] "Further advancements for evolved universal terrestrial radio access (E-UTRA) physical layer aspects," tech. rep., 3GPP TR 36.814, Mar. 2010.

[112] K.-M. Miettinen, *Nonlinear multi-objective optimization*. Kluwer Academic Publishers, Boston, 1999.

[113] J.-F. Berube, M. Gendreau, and J. Y. Potvin, "An exact ε-constraint method for bi-objective combinatorial optimization problems: application to the traveling salesman problem with profits," *European Journal on Operational Research, Elsevier*, vol. 194, no. 1, pp. 39–50, 2009.

[114] T.-H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*. The MIT Press, 3rd ed., 2009.

[115] A.-A. Juan, I. Pascual, D. Guimarans, and B. Barrios, "Combining biased randomization with iterated local search for solving the multidepot vehicle routing problem," *International Transactions in Operational Research*, vol. 194, no. 1, pp. 1–21, 2014.