UNIVERSITAT DE
BARCELONA

Energy Efficient Radio Resource Management for Heterogeneous Networks

Joaquim Manuel Camões Sobral de Bastos

ADVERTIMENT. La consulta d’aquesta tesi queda condicionada a l’acceptació de les següents condicions d'ús: La difusió d’aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) i a través del Dipòsit Digital de la UB (**diposit.ub.edu**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d’investigació i docència. No s’autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d’un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s’autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) y a través del Repositorio Digital de la UB (**diposit.ub.edu**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you’re accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service and by the UB Digital Repository (**diposit.ub.edu**) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it’s obliged to indicate the name of the author.

Doctoral thesis

Energy Efficient Radio Resource Management for Heterogeneous Networks

Joaquim Manuel Camões Sobral de Bastos

UNIVERSITAT DE BARCELONA

Energy Efficient Radio Resource Management for Heterogeneous Networks

Programa de doctorat en Enginyeria i Ciències Aplicades

Autor: Joaquim Manuel Camões Sobral de Bastos

Directors: Dr. Christos Verikoukis
Dr. Jonathan Rodriguez

Tutor: Dr. Manuel Puig Vidal

UNIVERSITAT DE BARCELONA

RESUMEN

La continua y rápida evolución de los sistemas de comunicación actuales presenta diversas líneas de actuación, con aspectos muy diversos como el despliegue e implementación eficiente de redes celulares. En este ámbito, el mayor desafío se presenta en cómo aprovechar los recursos disponibles, principalmente en cuanto al espectro RF, para satisfacer los requisitos asociados a los servicios o aplicaciones específicas ofrecidas al usuario final.

La eficiencia energética también se ha convertido en un aspecto importante, dentro de los mencionados requisitos, debido al impacto que tiene para los operadores con respecto al coste operacional de las redes, y para el usuario final, debido al efecto en la autonomía de los dispositivos móviles. Por lo tanto, aparte de la disponibilidad del servicio, el cual se ve afectado por la cobertura, capacidad, ancho de banda y latencia, el coste energético en bits-per-joule es otro factor importante a tener en cuenta al optimizar las redes de comunicaciones, que tiene como objetivo un incremento en la eficiencia desde una perspectiva más completa.

Esta tesis aborda las crecientes preocupaciones con respecto a la eficiencia energética en la industria de las TIC, centrándose específicamente en las redes inalámbricas celulares. Además, teniendo en cuenta el panorama actual y los desarrollos en este campo, es decir, en relación con los sistemas y redes de comunicación 5G emergentes, el concepto de red heterogénea (HetNet) también es parte del núcleo del trabajo que se ha llevado a cabo y se presenta en esta tesis.

Este trabajo toma en consideración los dos extremos del sistema de redes considerado. Los primeros son los usuarios finales nómadas que deberían poder moverse libremente en dicho sistema, siempre con un servicio continuo, ininterrumpido y proporcionado de manera fluida, y el segundo es la red central administrada por el operador de telecomunicaciones. Los principales beneficios de lograr una mayor eficiencia energética son, para el primero, la autonomía extendida de los dispositivos móviles,

que se refleja en la movilidad extendida de sus usuarios, y para el segundo, costos operativos más bajos o estables, que se reflejan también en la huella global de las TIC en el medio ambiente.

Para lograr esto, el trabajo realizado se centró en abordar dos desafíos clave en HetNets, específicamente, optimizar la eficiencia energética asociada a los traspasos verticales (VHO) y los mecanismos y enfoques de gestión de recursos de radio (RRM) para permitir una comunicación eficiente desde un punto de vista energético.

El trabajo aborda la eficiencia energética en dos perspectivas diferentes, cada una utilizando una solución diferente que explota los distintos mecanismos disponibles especificados de las redes inalámbricas celulares. Uno considera que los dispositivos móviles integran múltiples interfaces RAT e involucran una solución basada en VHO que explota el estándar IEEE 802.21, y el otro apunta a las redes de acceso radio y núcleo de red, explotando una solución basada en RRM.

Teniendo en cuenta el escenario HetNet considerado en esta tesis, así como los problemas inherentes de ICI, para tener una gestión de interferencia eficiente dentro de dicha red se necesita una coordinación efectiva entre los diferentes tipos de nodos (alta y baja potencia). El desafío en este escenario es lograr dicha coordinación de la manera más energéticamente eficiente, a través de la gestión adecuada de los recursos físicos disponibles, es decir, PRBs.

En el primer capítulo de esta tesis, junto con los conceptos más relevantes y los desafíos existentes, que proporcionan una motivación relevante para el trabajo realizado, se ha realizado una descripción general de los desarrollos pasados y actuales, que culminan en las redes 5G emergentes. Además, se llevó a cabo una investigación extendida sobre temas relevantes, es decir, VHO / MIH, DRA, gestión de interferencias y eficiencia energética, el cual se consolida en el segundo capítulo de esta tesis.

La primera solución de eficiencia energética propuesta aprovecha las características proporcionadas por IEEE 802.21 (MIH / MIIS) y VHO para obtener un ahorro de energía en los dispositivos móviles modernos multi-RAT, y puede eventualmente reducir el consumo de energía en sus interfaces RAT en aproximadamente un 30% en promedio, con una sencilla implementación, de acuerdo con los resultados de la simulación. Sin embargo, el esquema propuesto actualmente también conlleva una pérdida de paquetes adicional vinculada a la velocidad de los UE, que es causada principalmente por la latencia en la re-asociación de la red.

En la segunda solución, para el sistema OFDMA HetNet de enlace descendente, el algoritmo de optimización de eficiencia energética propuesto para la asignación de recursos de radio, teniendo en cuenta un dado requisito de velocidad de datos, asociado a QoS, presentó una convergencia rápida, lo cual es clave en el diseño de sistemas EE HetNet. El algoritmo considera no solo la potencia radiada, sino también los dos tipos de potencia del circuito. Los resultados de la simulación pueden aprovecharse para diseñar redes de consumo de energía óptimas basadas en el método HetNet orientado a QoS con una potencia total fija.

Como líneas de investigación futuras en el alcance de las soluciones RRM de eficiencia energética, se pueden llevar a cabo varios enfoques y / o extensiones de las soluciones EE propuestas. Por ejemplo, el desarrollo e integración de un algoritmo de conmutación de interfaz RAT más complejo que proporcione a la solución EE basada en MIH / MIIIS propuesta más inteligencia, es decir, capacidad para mitigar la pérdida de paquetes adicional detectada asociada a la velocidad del UE y la eventual re-asociación de red. Además, con respecto a la solución de optimización EE RRM, se podrían abordar otros enfoques de restricción de QoS junto con la comparación de otros métodos relevantes.

Por otro lado, también sería importante tener en cuenta el cambio de paradigma que surge con la aparición de nuevos estándares, concretamente la versión 15 de 3GPP, que coloca los peldaños para los sistemas y redes 5G. En esta perspectiva, en primera instancia, es clave ir más allá de las suposiciones hechas en este trabajo, al considerar la arquitectura C-RAN en el desafío abordado para lograr una EE más alta, lo cual es unos de los objetivos principales al rediseñar los sistemas de comunicación móvil existentes y futuros, concretamente en cuanto a sostenibilidad medioambiental.

C-RAN tiene el potencial de reducir la interferencia multinivel de una manera más fluida, es decir, cuando se asocia a SDN / NFV, virtualizando el procesamiento de banda base. En dicha implementación, los recursos de radio pueden verse como un "grupo de recursos" para ser compartidos y utilizados por la red, de una manera cooperativa centrada en la red del usuario, en lugar de en el paradigma clásico no cooperativo implementado típicamente en las redes actuales. Como C-RAN tiene como objetivo aprovechar todo el procesamiento de banda base para todos los usuarios dentro de una unidad común, proporciona al operador un control completo sobre la red y la capacidad de coordinar las transmisiones de señal, lo que proporciona un paso significativo hacia la mitigación de la interferencia en la red, asociada con técnicas relevantes como eICIC y CoMP.

En un ámbito más amplio, el trabajo realizado y presentado en esta tesis también podría ampliarse teniendo en cuenta el concepto emergente de células móviles pequeñas, donde el dispositivo móvil adopta el papel de punto de acceso o cabezal de radio remoto. Esto cambia el paradigma de las redes móviles heredadas de centrarse únicamente en la red a centrarse en el dispositivo, donde los dispositivos móviles podrían verse como un conjunto de recursos de red adicionales que el operador utilizará para extender la cobertura de la red bajo demanda. Dicho nuevo paradigma desencadena nuevos desafíos de investigación, además de los centrados en EE, también con respecto a incentivos para la cooperación de los propietarios de dispositivos móviles. Así, estos también podrán desempeñar un papel clave con respecto a la autonomía de los dispositivos, que está directamente asociada a la movilidad involucrada, y el respectivo RRM virtual requerido para extenderla tanto como sea posible en beneficio de los utilizadores.

ABSTRACT

In the continuous evolution of mobile communication systems, there are many issues and aspects to be addressed, as well as in their effective implementation and deployment on cellular networks. In a general way, the most outstanding challenge is how to take the most advantage from the available resources, namely RF spectrum, to satisfy the communication requirements associated to a specific service or application offered to an end user.

Energy efficiency has also become one important aspect associated to such requirements, since it has a significant practical impact on operators, namely regarding their networks operational expenditure, as well as on UEs, concerning their battery duration and associated autonomy. Therefore, on top of service availability, involving network coverage and capacity, with appropriate throughput and latency, the energy cost per bit, or bits-per-Joule, is also an important factor to account for in such optimized networks, driven towards increased all-round efficiency.

This thesis addresses and targets contributing to the growing concerns regarding energy efficiency in the ICT industry, specifically focusing on cellular wireless networks. Moreover, taking into account the current panorama and developments in this field, namely concerning the emerging 5G communication systems and networks, the heterogeneous network (HetNet) concept is also at the core of the work that has been carried out and is presented in this thesis.

The work takes into consideration the two ends of the considered system of networks. The first are the nomad end users that should be able to roam freely in such system, always with continuous service provided in a seamless way, and the second is the core network managed by the respective telecommunications operator. The main benefits of achieving higher energy efficiency are, for the former, extended UE autonomy, which reflects in their users' extended mobility, and for the latter, lower or stable operational costs, reflecting as well in the ICT global footprint on the environment.

To achieve this, the work carried out focused in addressing two key challenges in HetNets in general, specifically towards optimizing the energy efficiency associated to vertical handovers (VHO) and to radio resource management (RRM) mechanisms and approaches to allow energy efficient communication in a HetNet.

The work addresses energy efficiency in two different perspectives, each using a different solution exploiting distinct available mechanisms specified for cellular wireless networks. One considers UEs integrating multiple RAT interfaces, and involve a VHO-based solution exploiting IEEE 802.21 standard, and the other targets both core and radio access networks, exploiting a RRM-based solution.

Taking into account the HetNet scenario, which is considered in this thesis, as well as the inherent ICI issues, in order to have efficient interference management inside such network, effective coordination amongst the different kinds of nodes (high and low power) is needed. The challenge here is also to achieve such coordination in the most energy efficient way, through the appropriate and efficient management of the available physical resources, i.e. PRBs.

An overview of the past and ongoing developments, culminating in the emerging 5G networks, has been made and summarized in the first chapter of this thesis, together with the most relevant concepts and existing challenges, which provide relevant motivation for the work carried out. In addition, an extended investigation was carried out on relevant topics, i.e. VHO/MIH, DRA, Interference Management and Energy Efficiency, which is consolidated in the second chapter of this thesis.

The first proposed energy efficient solution exploiting features provided by IEEE 802.21 (MIH/MIIS) and VHO, targeting energy saving at modern multi-RAT mobile UEs, can eventually reduce the energy consumption at their RAT interfaces by roughly 30% on average, with a straightforward implementation, according to the attained simulation results. However, the currently proposed scheme presents a trade-off in terms of added packet loss linked to UEs velocity, which is mostly caused by delayed network re-association.

In the second solution, for downlink OFDMA HetNet system, the proposed energy efficiency optimization algorithm for radio resource allocation, taking a given data rate requirement into account, associated to QoS, presented fast convergence, which is key in the design of real EE HetNet systems. The algorithm considers not only the radiated power, but also both types of the circuit power. Simulation results can eventually be exploited towards designing optimal energy consumption networks based on QoS-oriented HetNet method, while total power is fixed.

As future work in the scope of the energy efficient RRM solutions for HetNets and investigation lines of research, several subsequent approaches and/or extensions of the proposed EE solutions can eventually be carried out. For instance, by developing and integrating a more complex RAT interface switching algorithm providing the proposed MIH/MIIIS-based EE solution with further intelligence, namely towards mitigating the detected additional packet loss associated to the UEs velocity and eventual network re-association. Also, regarding the EE RRM optimization solution, other QoS constraint approaches could be addressed along with the comparison of other relevant methods, taking into account the considered scenario.

On the other hand, it would also be important to take into account the paradigm shift that arise with the emergence of new standards, namely 3GPP Release 15, placing the stepping stones for 5G systems and networks. In this perspective, at a first instance, it is key to go beyond the assumptions made in this work, by considering C-RAN architecture in the addressed challenge for achieving higher EE, which is a prime concern when redesigning existing and future mobile communication systems, namely in sustainable and environmental perspectives.

C-RAN has the potential to reduce multi-tier interference in a more seamless way, namely when associated to SDN/NFV, by virtualising baseband processing. In such implementation, radio resources can be seen as being in a “pool of resources” to be shared and used by the network, in a cooperative user-network centric way, rather than in the classical non-cooperative paradigm typically implemented in today’s networks. As C-RAN aims to harness all the baseband processing for all users within a common unit, it provides the operator complete control over the network and the ability to coordinate signal transmissions, which provides a significant step towards mitigating interference in the network, associated with relevant techniques, e.g. eICIC and CoMP.

On a broader scope, the work carried out and presented in this thesis could also be extended taking into account the emerging concept of mobile small cells, where the mobile UE adopts the role of access point or remote radio head. This shifts the paradigm of legacy mobile networks from purely being network centric towards being device/UE centric, where mobile devices could then be seen as a pool of additional network resources to be used by the operator to extend network coverage on demand. Such new paradigm triggers new research challenges, besides de ones focused on EE, also in terms of incentives for promoting UEs owners cooperation. The latter will thus also play a key role, namely with respect to the devices autonomy, which is directly associated to the involved mobility, and the respective virtual RRM required to extend it as much as possible, ultimately benefiting the end users.

ACKNOWLEDGMENTS

To my family, colleagues, friends and supervisors for the endless support and patience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

#G	1: First, 2: Second, 3: Third, 4: Fourth, or 5: Fifth Generation
3D	Three-Dimensional
3GPP	3rd Generation Partnership Project
3GPP2	3rd Generation Partnership Project 2
5G-NR	5G New Radio
ABC	Always Best Connected
AMPS	Advanced Mobile Phone System
AP	Access Point
ARQ	Automated Repeat Request
AWGN	Additive White Gaussian Noise
BBU	Baseband Unit
BER	Bit Error Rate
BH	Backhaul
BLER	Block Error Rate
BS	Base Station
CA	Carrier Aggregation
CAPEX	Capital Expenditure
CBR	Constant Bit Rate
CDMA	Code Division Multiple Access
CDMA2000	*3GPP2 Family of mobile technology standards*
CI	Maximum Carrier-to-Interference Ratio
CN	Core Network
CoMP	Coordinated Multipoint
CoS	Class of Service
CPRI	Common Public Radio Interface
CPS	Cyber-Physical System
CR	Cognitive Radio
C-RAN	Cloud (or Centralized) Radio Access Network
CRC	Cyclic Redundancy Check
CRE	Cell Range Extension
CS	Command Service
CSI	Channel State Information
CSIT	Channel State Information at the Transmitter
CQI	Channel Quality Indicator
DAS	Distributed Antenna System
dB	Decibel
D2D	Device-to-Device

DCD	Download Context Descriptor
DL	Downlink
DeNB	Donor eNB
DoS	Denial-of-Service
DRA	Dynamic Resource Allocation
E-UTRAN	Evolved UMTS Terrestrial Radio Access Network
EDGE	Enhanced Data rates for GSM Evolution
EE	Energy Efficiency
eICIC	Enhanced ICIC
eNB	Evolved Node B
EPC	Evolved Packet Core
ES	Event Service
ETSI	European Telecommunications Standards Institute
EV-DO	Evolution-Data Optimized
FBMC	Filter Bank MultiCarrier
FDMA	Frequency-Division Multiple Access
FEC	Forward Error Correction
FeICIC	Further Enhanced ICIC
FFR	Fractional Frequency Reuse
FH	Fronthaul
FR	Full Reuse
FTP	File Transfer Protocol
GBR	Guaranteed Bit Rate
GERAN	GSM EDGE Radio Access Network
GMSK	Gaussian Minimum-Shift Keying
GPRS	General Packet Radio Service
GPS	Global Positioning System
GSM	Global System for Mobile Communications (*Groupe Spécial Mobile*)
HARQ	Hybrid Automated Repeat Request
HeNB	Home eNB
HetNet	Heterogeneous Network
HO	Handover
HSDPA	High Speed Downlink Packet Access
HSPA	High Speed Packet Access
ICI	Inter-Cell Interference
ICIC	Inter-Cell Interference Coordination
ICT	Information and Communication Technology
IE	Information Element
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IoT	Internet of Things

IP	Internet Protocol
IS	Information Service
IS-95	Interim Standard 95 (*also known as cdmaOne*)
IT	Information Technology
KKT	Karush–Kuhn–Tucker
LA	Link Adaptation
LAN	Local Area Network
LGD	Link Going Down
LLC	Logical Link Control
LTE	Long Term Evolution
M2M	Machine-to-Machine
MAC	Medium Access Control
MBS	Macro Base Station
MCS	Modulation and Coding Scheme
MIH	Media Independent Handover
MIHF	Media Independent Handover Function
MIIS	Media Independent Information Service
MIMO	Multiple Input Multiple Output
MIP	Mobile Internet Protocol
MMS	Multimedia Messaging Service
mmWave	Millimetre Wave
MN	Mobile Node
MPEG	Moving Picture Experts Group
MSC	Mobile Switching Centre
MT	Mobile Terminal
MU-MIMO	Multi-User MIMO
ND	Neighbour Discovery
NFV	Network Function Virtualization
NIST	National Institute of Standards and Technology
NMT	Nordic Mobile Telephone
ns-2	Network Simulator *ns-2* (*discrete event network simulator software*)
OFDMA	Orthogonal Frequency-Division Multiple Access
OPEX	Operational Expenditure
OSTBC	Orthogonal Space-Time Block Coding
PBR	Peak Bit Rate
PF	Proportional Fair
PFR	Partial Frequency Reuse
PHY	Physical layer (OSI model)
PLUL	Path Loss for the Uplink
pN	pico Node (low power node)
PoA	Point of Attachment

PR	Partial Reuse
PRB	Physical Resource Block
PSK	Phase Shift Keying
QoE	Quality of Experience
QoS	Quality of Service
RA	Resource Allocation
RAd	Router Advertisement
RAN	Radio Access Network
RAT	Radio Access Technology
RDF	Resource Description Framework
RF	Radio Frequency
RM	Resource Management
RN	Relay Node
RNC	Radio Network Controller
RR	Round Robin
RRH	Remote Radio Head
RRM	Radio Resource Management
RS	Router Solicitation
RSSI	Received Signal Strength Indicator
SAE	System Architecture Evolution
SAP	Service Access Point
SC-FDMA	Single-carrier FDMA
SDN	Software-Defined Networking
SE	Spectral Efficiency
SFR	Soft Frequency Reuse
SINR	Signal-to-interference-plus-noise ratio
SMS	Short Message Service
SNR	Signal-to-Noise Ratio
SSDL	Strongest received Signal in Downlink
TACS	Total Access Communication System
TCP	Transmission Control Protocol
TDMA	Time Division Multiple Access
TTI	Transmission Time Interval
UC	Use Case
UCD	Upload Context Descriptor
UDP	User Datagram Protocol
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
UL	Uplink
URLLC	Ultra-Reliable Low-Latency Communications
UTRAN	UMTS Terrestrial Radio Access Network

VHD	VHO Decision
VHO	Vertical Handover
VM	Virtual Machine
VoIP	Voice over IP
WAP	Wireless Application Protocol
WCDMA	Wideband Code Division Multiple Access
WiMAX	Worldwide Interoperability for Microwave Access
Wi-Fi	Wireless Fidelity (*trademark associated to IEEE 802.11 standards*)
WLAN	Wireless Local Area Network
WWW	World Wide Web
X2AP	X2 Application Protocol

1. INTRODUCTION

The generally adopted vision for ultimate wireless ubiquity is based on all-IP networks, where users can globally roam freely by using distinct points of attachment in the operators' networks, always with continuous service and in the most seamless way, without any glitch being perceived whatsoever by the itinerant users [1][2]. The ability to handover the user's connection between different heterogeneous networks (HetNet) is denominated as Vertical Handover [3].

In such truly mobile environment, a key factor in the nomadic perspective is undoubtedly the energy efficiency of the established wireless communication, which is responsible for draining the battery of the user's mobile terminal(s), which evidently have limited capacity and operational duration [4].

This work focus on radio resource management (RRM) mechanisms and approaches to allow energy efficient communication in a HetNet [5].

1.1. Towards 5G Networks

The first cellular wireless communication technology that effectively caused a true worldwide impact emerged in the beginning of the nineties under the denomination of GSM. Nonetheless, its analogue-based precursors AMPS and TACS deployed in the eighties, considered to represent the first generation in such kind of technology (1G), also had an important role by showcasing a glimpse of what could be possible, and which would change the existing personal and business communication paradigm.

The second generation (2G) of cellular wireless systems, which emerged one decade after 1G, differentiated from such systems, namely for being based on digital technology, which allowed not only better call quality, but also text and image messaging (SMS/MMS). Later, with the deployment of GPRS, and posteriorly EDGE, general data exchange started to be possible through packets transmission, and the Internet started to be accessible via truly mobile devices by using the Wireless Application Protocol (WAP), even though at low data rates, as can be seen in Table 1, due to the communication channel narrow band. The evolution of the 2G systems has been possible namely through the exploitation of higher-order digital modulation schemes, e.g. 8-PSK in EDGE, while GMSK in GSM/GPRS. Nonetheless, other techniques have also contributed towards that evolution, such as rate adaptation algorithms, which adapt the modulation and coding scheme (MCS) according to the radio channel quality, having thus a direct impact on bit rate and robustness of the data transmission.

Roughly another decade after, third generation (3G) systems started to appear, although still inheriting a few characteristics from the previous generation, such as a core network architecture based on dual switching, i.e. circuit-switching for voice communication, and packet-switching for data transmission.

Table 1: Generations of cellular wireless communication technologies.

	1G	**2G**	**3G**	**4G**	**5G**
Early Deployment	1981	1991	2001	2010	2020
Technology & Standards	*Analogue* AMPS, TACS, NMT	*Digital* **GSM**, IS-95, GPRS, EDGE	*Digital* **UMTS**, EV-DO, HSPA, HSPA+	*Digital* LTE, WiMAX, **LTE-Advanced**	*Digital* FBMC **5G-NR**
Core Network	Circuit-switching	Circuit & Packet-switch.	Circuit & Packet-switch.	All-IP (**SAE**) Packet network	All-IP Packet-switching
Multiplexing	FDMA	**TDMA/FDMA**, CDMA	CDMA, **WCDMA**	**OFDMA (DL)** **SC-FDMA (UL)**	**OFDMA (DL)** **SC-FDMA (UL)**
Channel Bandwidth	30 kHz	**200** - 1250 kHz	1.25 - **5** MHz	1.25 - **20** MHz	FR1 10 - 100 MHz FR2 50 - 400 MHz
Latency	N/A	0.5 - 1 s	50 - **400** ms	**10** - 50 ms	**1** - 4 ms
Throughput	*(2.4 kbps)*	**9.6/14.4** - 236 kbps	**0.4** - 21/…/84 Mbps	0.1 - **1.2** Gbps	DL: 20 Gbps UL: 10 Gbps
Major added Services	Voice, (*data through modems)*	SMS, MMS, WAP/basic Internet data	Internet, Video calls, Mobile TV	Gamming, 3D Video, Cloud computing	Tactile Internet, WWWW, URLLC, Smart verticals

The increasing popularization of the Internet fuelled the huge 3G advances in the first decade of the 2000’s, not only because of the World Wide Web (WWW), where users’ experience benefits from higher downlink (DL) rates, but also pushed by the emerging social networks, which made users increasingly more eager and demanding also for higher uplink (UL) data rates.

Several techniques have been investigated and applied to 3G communication systems in order to be possible to achieve such an amazing progression of the base UMTS standard and the involved WCDMA multiplexing technology. For instance, UMTS implementing HSPA+ can deliver a throughput gain of more than 200 times the original 384 kbps of UMTS/UTRAN specified in 3GPP Release '99. HSPA+ currently allows up to 84 Mbps by using carrier aggregation (CA), of 4 carriers in this case, together with multiple input multiple output (MIMO) technology, which depends on multiple transmitting and receiving antennas, as well as on the associated signal processing.

In order to advance in the quest for higher spectral efficiency, but also for more efficient usage of the existing radio resources, a new multiplexing technique, denominated Orthogonal Frequency-Division Multiple Access (OFDMA), has been introduced in fourth generation (4G) systems. Many of the techniques already exploited in 3G are also used in 4G, such as MIMO, although benefiting from a possible higher number of antennas and evolved schemes, e.g. MIMO 4x4 and MU-MIMO.

However, another key evolution, particularly in the Long Term Evolution (LTE) 4G system is the migration from a dual switching core network to an all-IP packet network architecture, denominated as System Architecture Evolution (SAE). The most striking characteristic of this new architecture evolved from GSM’s GPRS core network, is its reduced complexity when compared against 2G and 3G systems. SAE was designed targeting to also address another key communication aspect, besides throughput, that not only enhances user experience but also allows for new applications, e.g. enhanced (real-time) gamming and cloud computing, to be possible through 4G cellular wireless communication systems. That communication aspect, or metric, is known as latency.

The SAE architecture, besides being less complex by involving fewer entities or subsystems, allowing lower latency and higher throughput, and by providing any kind of communication through data packets transmission, just like Internet, it also allows the coexistence and mobility between multiple different radio access networks (RANs). The support provided by SAE, not only covers LTE's E-UTRAN, but also 3GPP legacy systems, such as GSM/GPRS/EDGE's GERAN or UMTS's UTRAN. Moreover, SAE's support even encompasses non-3GPP systems, such as 3GPP2's CDMA2000 and IEEE's Wi-Fi (802.11) and WiMAX (802.16).

Although several advances continue to be achieved presently on current 4G technology, and some are still being implemented on 3G systems, the attention is now driven towards the next generation, 5G [6], specified in 3GPP Release 15 [7], which should be deployed in the early 2020's. The 5G systems, based on 5G-NR (New Radio), are expected to focus on enhanced mobile broadband, namely including high reliability and low latency. This should enable a vast new set of applications that are not fully possible with the current 4G systems. Examples of such kind of demanding applications can be found throughout most "smart vertical" domains, e.g. autonomous driving (Smart Mobility), critical remote telemedicine (Smart Health), or seamless remote working and real-time collaboration, as well as augmented reality (Smart Society/Cities).

The emerging 5G architecture should inherit most of the aspects already existing in SAE, namely based on low complexity and reduced number of entities involved in the actual communications in order to enable even lower latency and higher throughput. Also, the architecture should provide seamless interoperability between heterogeneous networks (HetNets), eventually even towards a truly converged wired and wireless communication network using the same infrastructure.

Symmetrical and asymmetrical transmission of data on extreme mobile broadband should be possible in any place, at any time, by any device connected to a 5G network, thus being possible to achieve perceived ubiquitous mobility by any user, or device, on such network. The envisaged characteristics of 5G systems are illustrated in Figure 1.

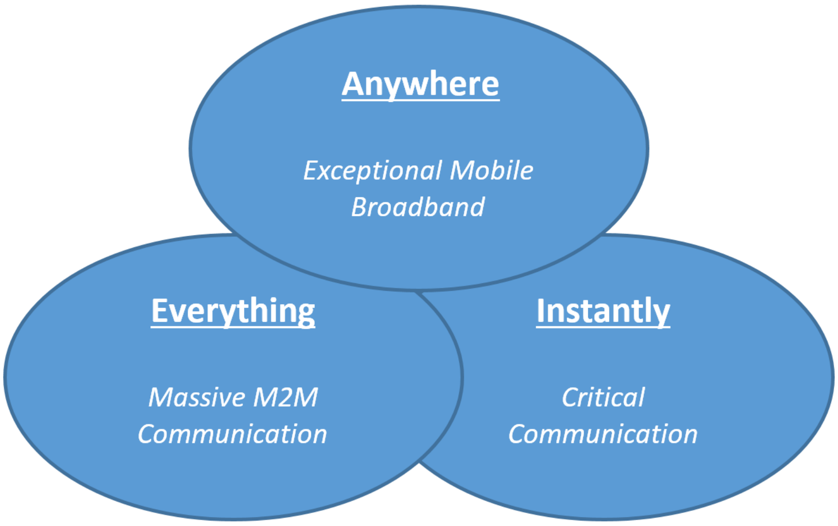

Figure 1: Fundamental characteristics of 5G systems.

Mobile broadband will continue to improve in 5G systems, being expected throughput peak rates above 10 Gbps and at least typical rates of 100 Mbps on most common conditions, accessible pretty much anywhere. This will allow to satisfy the expected increase of overall traffic by 10.000 times, which should be demanded by users and connected machines/things alike, by early 2020's.

Machine-to-machine (M2M) and Device-to-Device (D2D) communication will coexist with communications among users and with data servers, in the same 5G networks. It is expected that most existing users' devices, appliances and objects will also be connected to the network, which should represent an increase of up to 100 times more connected devices, really paving the way for the "everything connected" motto and Internet of Things (IoT) paradigm, contributing to an expected huge data traffic growth.

Also, some of such connected devices and machines will require very low communication latency, of 1 ms or below, which is also the case for some users' mobile applications. Therefore, the aspect of enabling (near) instantaneous actions and reactions, in the so called Tactile Internet, together with the assurance of ultra-reliable low-latency communications (URLLC), are fundamental characteristics for 5G systems to support critical communication. This would ultimately enable the new envisaged applications, such as full support for autonomous driving, as well as many other applications associated to “smart vertical” domains, as already introduced above.

Moreover, many of the envisaged connected devices and objects will not be connected to a continuous power source, but will rely on embedded batteries with limited capacity and duration lifetime. Therefore, in 5G systems it is also further relevant to consider energy efficiency (EE) techniques and solutions that can be applied in order to allow maximum energy saving on the connected objects, providing them with as much power autonomy as possible. On the other hand, EE is also critical in the overall 5G networks, where it is important to have low energy consumption by the involved entities and subsystems composing the network, allowing the operator to have lower network running costs, also known as operational expenditure (OPEX).

Apart from all these mentioned features envisaged in 5G communication systems to enable a set of broad and rich high-quality personalised services to the end users, as individuals and their devices and objects, or corporations with their servers, machines and further connected resources, security and privacy are also fundamental. The end users’ trust while depositing their data, more or less critical and/or private, to be transmitted and used on the 5G networks to be implemented in the future, is key for the technology to fully thrive. Moreover, multiple future envisaged applications and scenarios will only effectively work once the end users share potentially confidential information and/or disclose privacy threatening data, namely about themselves, which will be uploaded, possibly processed and eventually downloaded through 5G systems. Therefore, appropriate security schemes and privacy preserving mechanisms must be considered and effectively implemented in the future systems, enabling preventive measures against cybercrime, e.g. denial-of-service (DoS), tampering and eavesdropping attacks. Nonetheless, cautious assessment should be made when designing and implementing such schemes and mechanisms providing security and privacy, not to induce unbearable increase of latency times, which would on the other hand incapacitate the same envisaged innovative applications to function properly as designed. Thus, a trade-off between the required communication latency and the level of security to be used in the actual communications should be necessary to be considered and set prior to the effective establishment and transmission of data.

From the recent developments of 3G and 4G systems towards the next generation of mobile communication systems, the following subsections introduce with additional detail some of the key concepts and technologies that are included or being considered for such systems, as well as some others emerging for the envisaged 5G systems.

1.1.1. Small cells

An effective planning of the cellular network is key to address the continuously increasing demand for mobile broadband required by new intensive applications and services, as well as data hungry users, depending on the limited available spectrum and radio resources. Many solutions have been devised and deployed to help the network operators, already since the early cellular wireless technologies, to overcome such network capacity challenge, being the simplest one the increase of network capacity through additional radio spectrum. As this solution relies on the rather limited radio spectrum resources, other solutions had also to be investigated and implemented during the early 2G and 3G development periods.

Some of such solutions comprised the implementation of more efficient modulation and coding schemes (MCS), and/or were based on multiple antenna techniques, i.e. MIMO. Although these solutions were important, and continue to be, they alone are insufficient in scenarios of crowded urban areas, as well as at the edges of cells, where the performance of some services can become rather diminished.

To mitigate such performance hampering, research as proven and standardized the concept of small cells together with their tight integration with the associated macro networks in order to spread the traffic load. At the same time, overall network performance can also be maintained, as well as quality of service (QoS), while reusing the valuable spectrum more efficiently. This concept is tightly associated to HetNet.

1.1.2. Heterogeneous network

The evolution of modern mobile communication systems towards 5G is envisaged to accomplish an effective convergence of services, which are already also provided through legacy mobile networking standards. Also, in order to be possible to achieve high throughput connectivity “anywhere”, it is necessary “to bring” the user and connected devices closer to the network. This is possible by increasing the density of the network, i.e. deploying a larger number of base stations (BS), known as Node B in UMTS’s UTRAN, and Evolved Node B (eNB) in LTE’s E-UTRAN. The densification of the network, as well as the HetNet concept itself can thus be seen and classified in two different perspectives. Figure 2 illustrates a multi-tier HetNet, using a single radio access technology (RAT), composed of macrocells, microcells/picocells and femtocells. Such cell sizes depend not only on the BS irradiated power but also on the respective antennas’ position, and number when MIMO is considered, together with the location environment. Such location can be either indoor or outdoor, which can be affected by RF obstacles, e.g. walls, and/or by the rural topography and natural RF propagation barriers, as well as buildings and urban city density, among other aspects.

The densification of cellular networks through the deployment of what is called small cells, also known as microcells, picocells or femtocells, relying on the same RAT, is one of the promising solutions to meet the required capacity increase for satisfying the end users high demand in terms of data throughput, no matter where they are located.

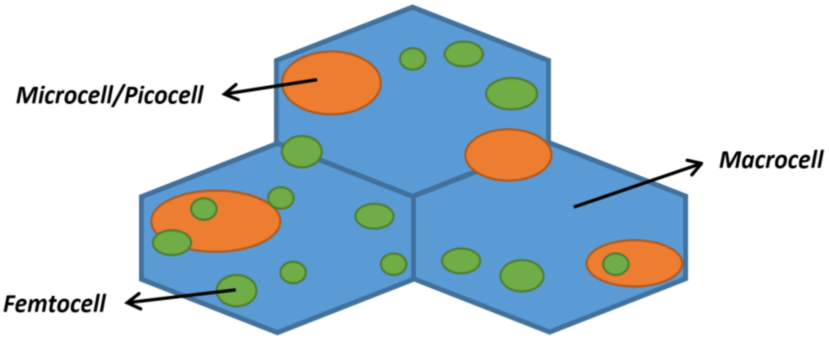

Figure 2: Multi-tier HetNet.

The HetNet, which appears to be less organized spatially when compared to conventional cellular planning, as it is overlaid with such small cells of the same technology, actually allows to achieve better EE on the network side, and on the user equipment (UE) side as well. Furthermore, this innovative solution can even help to significantly enhance the area spectral efficiency of the network.

The other kind of HetNet, known as multi-RAT HetNet, is also based on a denser network, but in this case encompassing multiple RATs by overlaying a conventional cell with small cells generated and operating under diverse technologies. This requires the coexistence and interoperability between different wireless communication systems, which can be cellular or not, and from different families of standards, e.g. 3GPP's LTE, and UMTS/HSPA, and IEEE's WiMAX and Wi-Fi. A multi-RAT HetNet is represented in Figure 3, where such a network comprises three different RATs, which should operate seamlessly and efficiently, providing benefits to the users and to the operator as well.

The cellular networks are thus being subject nowadays to a new paradigm, which involves enhanced optimisation, not only at their deployment, but also throughout their existence. There are now new concepts and elements in the infrastructure being massively deployed, making future 5G cellular networks significantly heterogeneous.

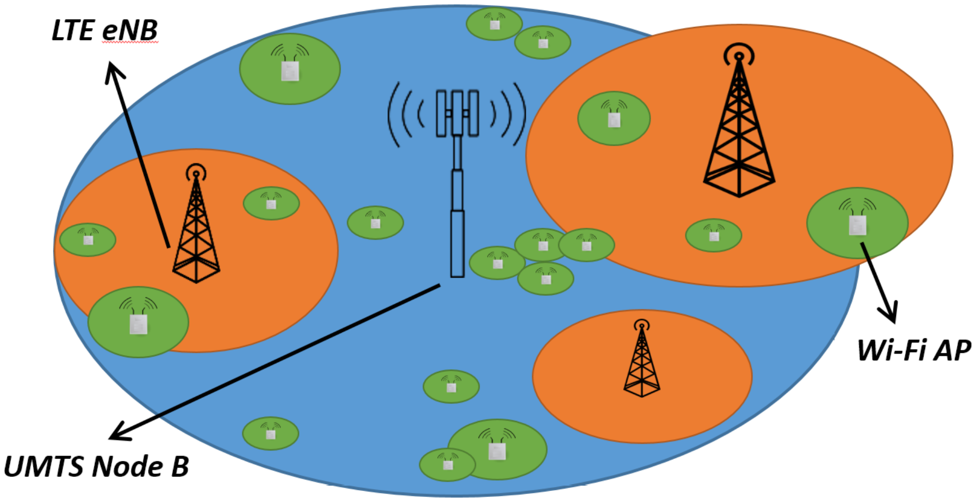

Figure 3: Multi-RAT HetNet.

In such future networks, small cells and associated innovative techniques developed to assist in the required densification, or even “hyper-densification” of the cellular networks, will play a fundamental role. The most promising solution to achieve this is not simply by adding more BSs, which would be unpractical and expensive namely in urban centres, but by deploying low-power BSs, as introduced in the next subsection.

1.1.3. Home eNB, Remote radio head and Relay node

Cost effective alternatives to deploying additional BSs, e.g. in particular geographical areas of a cellular network, have been researched and standardized in 3GPP Release 9 and Release 10. Such solutions include the low-power home eNB (HeNB), the relay node (RN) and the remote radio head (RRH), which provide the cellular network with higher capacity and/or better performance at the (macro) cells edge, in a HetNet. For that, it is important determining the best location where such new elements should be placed geographically in the existing network, not only to better serve the end users, but also to allow energy savings to the operator. Any of these solutions, responsible for small cells, target to increase capacity in hotspots where users’ demand is higher, either because of larger concentration of users and/or due to further demanding services and applications being used in that area. Also, they are used to assure network coverage in areas that are not appropriately served by the typical macro cells, including at their edge, and not only outdoors, but also indoors. Moreover, they also improve the performance of the network, as well as QoS, namely by offloading traffic from the conventional (macro) cells of the HetNet.

An illustration of these solutions is presented in Figure 4, where a 4G LTE network is considered, including the Evolved Packet Core (EPC) that aggregates and controls all the introduced 4G base stations. Essentially, HeNBs are introduced to provide indoor coverage, e.g. at office premises, through femto cells, and mainly act as low-power eNBs [8]. Nevertheless, and although HeNBs are also connected to the EPC, as they are privately owned, they are deployed without tight coordination with the network. It means there is a risk of interference between such cells and the surrounding network.

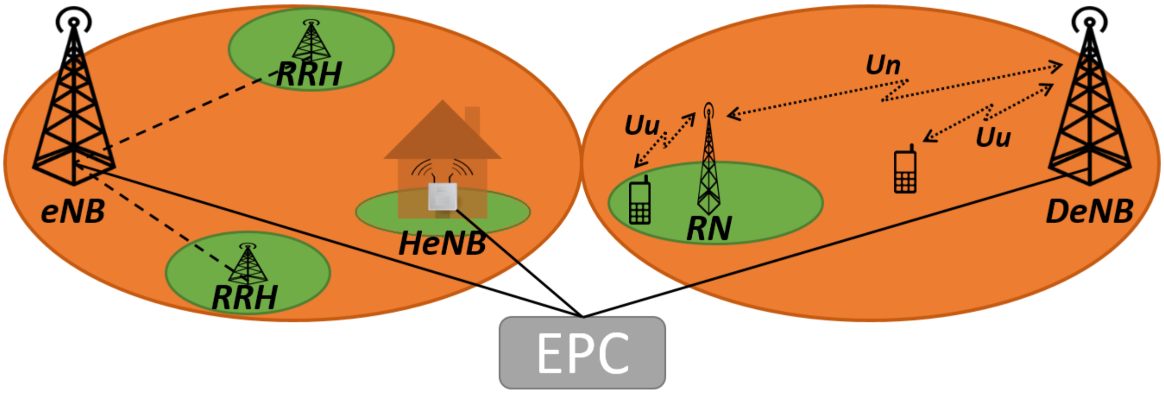

Figure 4: HeNB, RN and RRH small cells.

Small cells can also be provided through the use of RRHs, which in this case entirely rely on the eNB’s BBU (Baseband Unit) to which they are connected through a Common Public Radio Interface (CPRI), usually via fibre, as they merely act as “dumb” distributed antennas. The intelligence of this solution lies at the aggregator eNB, which is physically connected to the EPC.

Finally, the RNs are responsible for small cells within a macro (donor) cell served by what is denominated as donor eNB (DeNB), which connects and feeds its RNs through the *Un* radio interface, based on the LTE *Uu* interface that is used to connect with UEs. The RN essentially acts as an eNB from the UE perspective, and the DeNB, to which the RN is associated, sees it as an UE. Nonetheless, when the frequency allocated to *Uu* interface is the same as of *Un* for the RN, there is the possibility of self-interference in the RN small cell. As previously indicated, such interference can somehow similarly occur in the small cells solution involving HeNBs deployment, in this case in the surrounding areas of the HeNBs’ small cells.

It is clear that exploiting the presented small cell solutions is not always straightforward, and building a HetNet with different cell sizes contributes significantly to raise the complexity of network planning. This is particularly the case since the higher risks of interference, as already mentioned, effectively can take place and cause serious service degradation when such solutions, which are supposed to provide the exact opposite, are implementing without a concerted interference management approach.

1.1.4. Centralized-RAN and Software-Defined Networking

The solutions offered by 4G systems are significantly more advanced than in the earlier mobile systems generations. However, such solutions can be considered still rather rigid, to some extent, as there is not much flexibility provided by the few offered degrees of freedom for optimizing network efficiency in terms of infrastructure deployment and operational cost. Moreover, these systems provide limited support for future emerging use cases, such as D2D, M2M and Broadcasting events. For instance, in conventional distributed RAN in 4G cellular systems, all the eNB radio components are co-located at each respective cell site, namely the RRH, BBU and respective router to interconnect the eNB with other eNBs in the same region, through *X2* interface, and to the Mobile Switching Centre (MSC) via Ethernet. This typical distributed setup is illustrated in Figure 5.

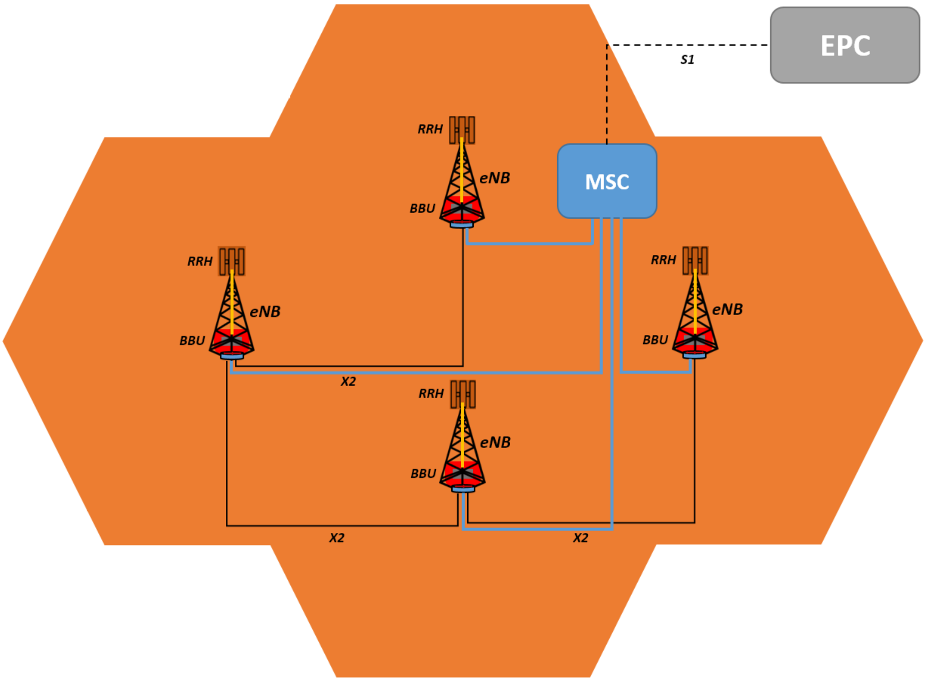

Figure 5: Typical distributed RAN in 4G cellular systems.

In order to allow addressing many emerging and future use cases and applications, 5G systems rely on virtualizing the network, supporting Cloud Radio Access Network (C-RAN) services and network sharing in an attempt to reduce capital and operational expenditure (CAPEX and OPEX) [9]. In addition, 5G systems enhanced architecture targets to support many rich services, such as IoT (involving massive amounts of communications), ultra-reliable and low latency services, as well as enhanced broadband connectivity.

The Cloud RAN, also called Centralized RAN, is a new implementation of the previous generation eNB radio, allowing to increase system efficiency and significantly reduce costs [10]. In the C-RAN implementation, multiple eNBs' BBUs, and associated routers, are placed together at the same respective MSC, which is typically also known as Central Office, BBU Hotel or Super Macro Site. The connectivity between each aggregated BBU and its respective RRH is still assured through CPRI, in this case over longer distances using optical fibre. A representation of C-RAN implementation is shown in Figure 6.

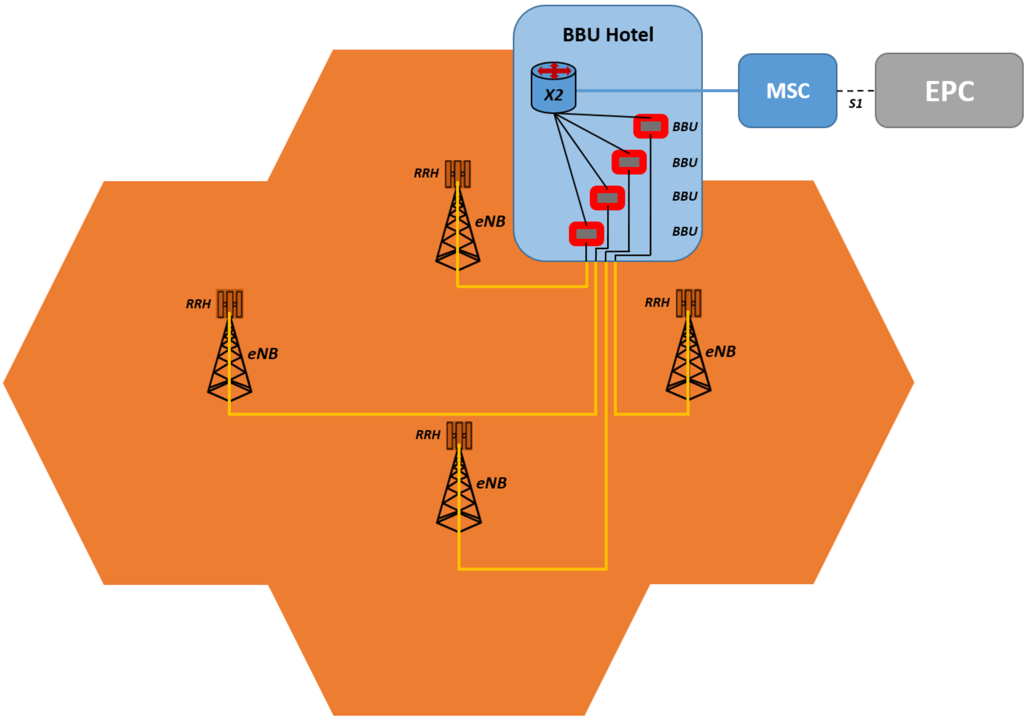

Figure 6: C-RAN implementation in 5G cellular systems.

The potential straightforward reduction of costs associated to the network infrastructure and its deployment (CAPEX) is rather obvious, namely by aggregating the eNBs' BBUs of a specific area or region in one single place (BBU Hotel) where all coordinated interactions also take place, via *X2* interfacing. Such costs reduction originate namely from the space savings at eNB sites, as well as the lower generation of heat, which thus implies requiring less cooling and lower power, all this while allowing easier access for testing, maintenance and repair. Essentially, the BBU Hotel centralized site functions as a data or processing centre, where individual BBUs can be stacked together without direct linkage, to allocate resources to UEs, dynamically, based on network needs. Moreover, communication between BBUs has high bandwidth and presents low latency, contributing to higher network performance, which allows offering new envisioned services and applications.

Evidently, this C-RAN centralised architecture simplifies networks, namely as deployment and scaling can be achieved faster, benefitting network owners and operators. The increased performance allowed by the co-located BBUs and respective *X2* interfaces, which cost-effectively interconnect the various BBUs locally through one single intermediate or large-scale router that also provides aggregation and transport to the MSC via Carrier Ethernet backhaul (BH) connection.

Not only the C-RAN topology offers out-of-the-box increased performance, enhanced flexibility and reduced cost in mobile fronthaul (FH) networks, but also its centralised aspect allows implementing the radio access network in a completely different way, which potentiates even further network performance, efficiency and costs reduction. Such alternative implementation exploits the concept of virtualization, which in this particular case would imply to virtualize the functionality of the BBU, i.e. to implement it in software and keep it running in a server, replacing the actual BBU hardware.

Software-Defined Networking (SDN) and Network Function Virtualization (NFV) are two complementary concepts, where essentially the latter can provide the network architecture conceptual support to the former, using IT virtualization technologies to virtualize into functional software blocks the standardised network node functions. Such functions can be, respectively, implemented using SDN technology to manage the NFV infrastructure, by building data networking equipment and software, separating and abstracting the elements of those systems by decoupling the control and data planes from each other, such that the former can reside centrally while the forwarding components remain distributed [11].

In practical terms, in a SDN/NFV C-RAN implementation, once BBUs are centralised, a server can easily replace the respective purpose-built hardware by virtualizing the BBU functionality, resulting in a virtual BBU. This virtual component can, of course, be easily replicated to build the pool of BBUs existing in a typical BBU Hotel, through multiple independent virtualizations of that network component. Such approach results into greater flexibility, higher performance and even lower cost, as the BBUs functions, including the respective *X2* interfacing, are provided by software processes and threads running on the same physical machine, or server, inside different Virtual Machines (VMs), respectively. A representation of such server-based C-RAN implementation, which in this context can truly be denominated as Cloud-RAN, is illustrated in Figure 7.

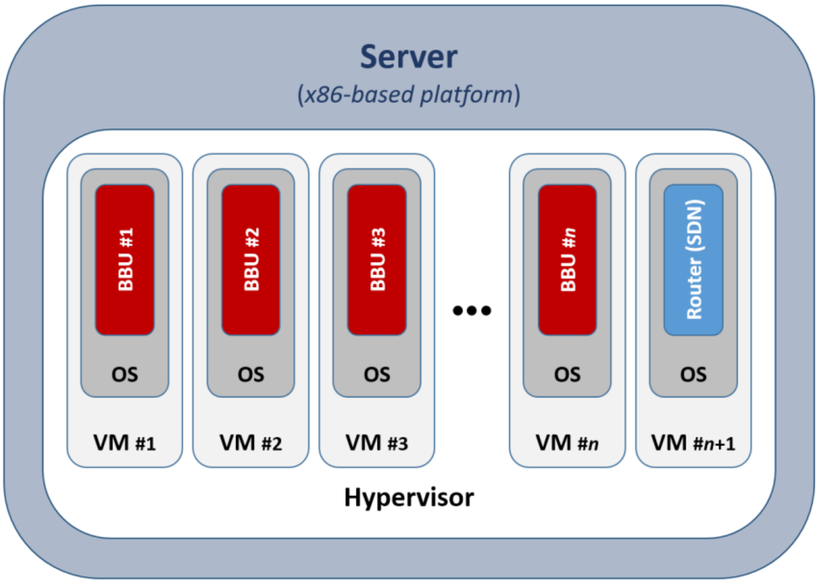

Figure 7: Server-based, virtualised, Cloud-RAN implementation.

This type of cloud computing environment, which provides a platform for managing computing resources (networking, processing, and storage) as a service through a cloud-based approach, is essentially built on open hardware and interface boards that handle fibre links and interconnections, dynamically, within the C-RAN server platform. Moreover, besides allowing the significant reduction of required purpose-built hardware, i.e. physical BBUs in this case, such cloud computing open platforms and real-time virtualization allows the allocation of shared resources between virtual BBUs, dynamically, which contributes to achieve higher spectrum efficiency and capacity. Such environment also expedites supporting further or new communication standards.

C-RANs are significant in the continuous evolution of wireless technologies, namely 5G and IoT systems, where seamless scaling is key, e.g. many pico-cells served by the many eNBs, respectively [12]. With easier deployment and scaling capability, the transition from LTE/4G to 5G networks will rely heavily on C-RAN development. It also provides a cost-effective, manageable solution for supporting more users. Such solution, hardware-wise, is mostly reflected by deploying more RRUs, associated to the respective virtual BBUs, which nevertheless depend on resources availability at the hosting server. Together with the lower cost, higher performance and efficiency derive naturally as well from the lower complexity involved in this reference implementation.

1.1.5. Challenges in 5G systems and networks

The technologies briefly introduced in the previous subsection, together with others, e.g. millimetre wave (mmWave) [13] and massive MIMO, are key in 5G systems towards overcoming the challenging demands that such emerging systems and networks are set to achieve, namely associated to the following representative use cases (UCs):

- **Massive IoT communications**, outdoor and indoor, as IoT seamless integration at early stage in 5G network architecture design is considered fundamental, namely considering two main classes of devices, responsible for different traffic patterns and requirements. The “high-end IoT” devices, which 5G embedded modem cost is not relevant due to the total cost of the object being made connected by this technology, e.g. cars, where also high bandwidth and low latency are required due to the associated core applications, e.g. autonomous driving. On the other hand, “low-end IoT” devices typically present sporadic and delay-tolerant traffic mostly composed of short packets, e.g. smart meters. Both types of devices will need support, taking into account that the latter, due to their typical low-cost and being battery operated could be almost any object, which would potentiate amounts of devices several orders higher when compared against the former type, albeit more demanding individually;

- **Broadband wireless**, both outdoor and indoor, is also a key goal, eventually addressed by massive deployment of small cells to provide uniform broadband experience to the users demanding high data rate and low latency for provisioning of applications, e.g. high-resolution video streaming, video-calling, gamming and cloud services. Seamless transitions between outdoor and indoor environments are also envisaged, exploiting efficiently the associated or co-existent backhaul technologies, typically heterogeneous, e.g. xDSL and fibre, used to transport data to/from the small cells. No single technology will alone accommodate the current and future increase in data traffic, but many solutions, working in concert, can and should be considered in 5G systems and networks. The challenge is how to best integrate such multiple technologies within the operators’ networks, and find the right balance to maximize their cumulative benefits, while leveraging existing assets to facilitate the evolution of their networks. In order to meet the 5G requirements characterizing this UC, are required new Medium Access Control (MAC)/RRM mechanisms for an efficient usage of the available spectrum and RATs, e.g. LTE small cells and Wi-Fi hotspots, co-locating the respective eNB and AP equipment, sharing site-lease agreements and backhaul. However, such extreme resource reuse may create high level of interference and require flexible solutions of coordination to improve the overall network performance.

- **Ultra-Reliable Low-Latency Communications (URLLC)**, outdoor and indoor, is crucial for enabling some emerging and envisaged applications through 5G systems, e.g. eHealth and M2M, for which methods are required to enable high degrees of network availability, with assurance of performance metrics, such as latency. Scalable solutions for networks supporting services with extreme requirements on availability and reliability are thus necessary. The reliability and latency of past communication systems were considered having the human user in mind, but for emerging and future wireless systems it is envisioned to have new M2M and IoT functionalities depending on communication with real-time constraints, enabling new safety-critical Cyber-Physical System (CPS) applications. Such emerging applications, e.g. in the automotive, industrial, health or military domains, require much higher reliability and lower latency than the ones provided by current communication systems.
- **High-speed mobility** is a challenging UC for 5G systems, even when considering vehicles moving at high speed in highways. GSMA in a recent article [14] indicates that automated and connected driving will be a pillar of Europe's industrial renaissance. The vehicle of the future is part of a connected world where superfast digital networks give access to communication, higher safety, improved environmental standards, entertainment, knowledge and personal contacts, to anyone, anywhere and at any time. Positive societal benefits are anticipated, with researchers pointing to: new jobs across the automotive value chain; increased road safety and lower fatalities; increased fuel-efficiency and lower environmental impact; reduction of traffic congestion and higher comfort standards for users [14].

To these introduced use cases, and associated challenges, such as negligible latency, high reliability and availability, adequate security and privacy must also be assured, both at network and information levels, for which specific or dedicated mechanisms and schemes should be put into place ensuring end-to-end security in communications.

1.2. Challenges and Thesis Objectives

There are many issues and aspects to be addressed in the continuous evolution of mobile communication systems, as well as in their effective implementation and deployment on cellular networks. In a general way, the most striking challenge is how to take the most advantage from the available resources, namely RF spectrum, to satisfy the communication requirements associated to a specific service or application offered to an end user.

Energy efficiency has also become one important aspect associated to such requirements, since it has a significant practical impact on operators, namely regarding their networks OPEX, as well as on UEs, concerning their battery duration and associated autonomy. Therefore, on top of service availability, involving network coverage and capacity, with appropriate throughput and latency, the energy cost per bit is also an important factor to account for in such optimized networks, driven towards increased all-round efficiency.

This thesis addresses two main challenges in HetNets in general specifically towards optimizing the energy efficiency associated to:

1. Vertical Handover (VHO);
2. Radio Resource Management (RRM).

1.2.1. Energy efficient VHO

Nowadays, it is quite common for different wireless technologies to be available simultaneously in a growing number of places. This poses several new requirements, challenges, and opportunities. In a multi-RAT HetNet, where network coverage and service to end users can be possible by using such distinct technologies and interoperable systems, the concept of vertical handover is fundamental, namely with respect to achieving seamless operation.

Briefly, the functionality associated to the VHO is essentially the same as in a regular HO when a UE, typically moving away from the network cell that is currently serving it, continues to get service from a new cell and responsible BS, in a seamless way. The main difference, though, is that this handover process occurs between cells supported by distinct technologies, i.e. RAT and RAN.

With the increasing service demand from end users, whose mobile terminals (MTs), or UE, are nowadays typically most likely equipped with multiple distinct wireless interfaces, it is possible to establish a new paradigm regarding network access selection polices. Such polices, together with appropriate management mechanisms, can be defined taking into consideration several communication aspects, such as throughput, QoS and EE, as well as by associating priorities according to network high-level management directives and to the services being provided to the UE.

Taking into account the mobility aspect associated to the majority of users, always carrying at least one UE, combined with their increasing dependency on such equipment, it is clear the importance of EE on the UE side towards longer battery duration and consequent UE autonomy.

Therefore, one of the key challenges to address here is to make the UEs consume the least energy as possible from their batteries by making them operate on the RAT and RAN that is the least demanding for them energy wise, while entirely complying with the respective service requirements. For that, it is necessary to develop VHO algorithms taking several contextual information into account.

Such information is continuously monitored and gathered by the involved networks, in order to provide an optimal energy-based decision, leading to the respective VHO trigger and subsequent RAT/RAN migration. It is also important to refer that such trigger and effective handover can depend not only on the UE dynamic mobility, but also on circumstantial changes of physical and/or network properties.

1.2.2. Energy efficient RRM

Information and communication technology (ICT), including the equipment in the respective infrastructures and their operation, is one of the major increasing sources contributing to global warming, as the involved energy consumption continuously increased in a relatively short time period, namely in the past two decades. Evidently, it becomes essential to assure high EE in such systems operation, which requires particular attention in their design process.

Overall system EE and specific energy saving techniques are thus currently at the core of the design, implementation, deployment and operation of cellular wireless networks, not only to reduce their CO_2 footprint, but also to cut network running costs, i.e. OPEX, to their operators. The latter, look thus to increase their networks bits-per-Joule metric in order to get the most network performance out of the energy investment, which reflects a significant part of the operational costs.

In reaction to such global trends, 3GPP standards currently consider new energy efficient approaches and techniques in the design and specification of 4G and beyond mobile networks. These EE schemes are reflected across the entire protocol stack, from physical layer to network layer, coherently with new networking topologies and deployment strategies.

Summarizing, instead of spectral efficiency (SE), which once was the key performance metric driving the development of cellular wireless systems, now EE joins SE into an engineering trade-off, between both metrics, leading such systems developments.

Taking into account the HetNet scenario, which is considered in this thesis, as well as the inherent ICI issues, in order to have efficient interference management inside such network, effective coordination amongst the different kinds of nodes (high and low power) is needed. The challenge here is also to achieve such coordination in the most energy efficient way, through the appropriate and efficient management of the available physical resources, i.e. PRBs. The result should be optimized methods and algorithms for efficient radio resource allocation, while maximizing achievable EE.

1.2.3. General objectives

The work carried out towards this thesis addresses energy efficiency in two different perspectives, each using a different solution exploiting distinct available mechanisms specified for cellular wireless networks:

a) UEs integrating multiple RAT interfaces, through VHO;
b) Core and RAN network, through RRM.

Therefore, on one hand, it is important to understand and consider UE design in a “smart green” way, i.e. that some “intelligence” can be embedded in the UE regarding the way that it senses and reacts to its surrounding environment in order to extend its battery duration by becoming more energy conscious. An energy efficient UE or MT means that its user is able to enjoy attractive, but inevitably power hungry broadband applications, in real-time, which will start and continue to emerge. Hence, the research carried out for this thesis, concerning the UE and VHO mechanisms, covers the following aspects:

a1) Survey of VHO algorithms;
a2) Investigation of new VHO decision steps and algorithms to promote EE;
a3) Consider further decision parameters to make VHO algorithms context aware;
a4) Implement EE context aware VHO algorithms;
a5) Integrate implemented VHO algorithms using MIIS in IEEE 802.21 architecture;
a6) Test and assess performance of the most promising VHO algorithms.

The main factor driving VHO decisions will be the minimization of energy consumption at the UE side, based on available context information, with an insight into parameters that can be used as decision input, and also different VHO triggers. That knowledge will serve as a base for the proposed development, adaptation or extension, of energy efficient policies and novel VHO algorithms towards EE.

The implementation of the most promising VHO decision algorithms is carried out taking into account the context information that is used for triggering the VHO decision, which can be stored in the Media Independent Information Service (MIIS) server in the existing specified IEEE 802.21 architecture [15]. This work will go beyond state of the art also by identifying new parameters that have the potential to provide high impact for energy efficient VHOs, and how it is possible to provision for these parameters in the current IEEE 802.21 architecture.

On the other hand, EE is also fundamental on the network side, not only to comply with the necessary “green” initiatives imposing a reduction of ICT systems impact, from a CO_2 footprint perspective, but also providing higher bits-per-Joule performance metric in the operator’s network, thus reducing its operational cost. Thus, it is important, first of all, to understand and consider such network design paradigm, including the involved entities and dynamic processes, following the global priorities on energy management, by exploiting the EE mechanisms considered in the evolving 3GPP standards. For that, the research carried out for this thesis, concerning the operator’s network and RRM mechanisms, covers the following aspects:

b1) Survey on dynamic resource allocation (DRA) algorithms;
b2) Study HetNet scenarios and associated constraints;
b3) Investigation on interference management techniques;
b4) Investigation of new RRM schemes to promote EE;
b5) Implementation of EE RRM solutions integrated in a HetNet scenario;
b6) Test and assessment of performance of the most promising RRM algorithms.

The considered reference scenarios are thus based on HetNet, in this case multi-tier, implementing 4G technology, i.e. LTE, comprising a combination of larger (macro) cells and small cells supported by low-power BSs. HetNet is considered towards gaining more SE and enhanced area of coverage, albeit introducing the inherent interference issues, i.e. ICI, which are investigated. By positioning low-power nodes associated with macro cells, small cells are created and typically used to increase the bandwidth or the area of coverage, e.g. through RNs, where desired signals from macro BS don´t reach appropriately, or to increase capacity of the network in places where high density of data usage is required. The strategic placement of such RNs is also here investigated.

In order to have efficient interference management inside a HetNet, it is key to develop and implement the necessary coordination amongst the different kinds of BSs. The coordination schemes between the macro base station (MBS), e.g. eNB, and small cells' low-power BSs, e.g. RNs, are studied in order to allow appropriate implementation and integration in the developed EE solutions.

The ultimate focus is on the maximization of EE in the HetNet, which requires optimizing the radio resources through RRM schemes implementing developed algorithms that also fulfil the minimum requirements for the respective service or application. The performance of the optimizing RRM methods developed to achieve maximum EE is finally assessed through simulation results.

1.3. Scientific Contribution

The work developed towards this thesis can be seen in the two distinct perspectives presented in the previous subsection. The first, referring to EE on the UE side, contributing to increased autonomy of MTs by exploiting VHOs, while the second focus on network EE, providing lower running costs to the operator and higher bits-per-Joule through optimized RRM schemes, both considering HetNets.

The following subsections provide a short introduction to each of the two main innovations achieved in the work that culminated in this thesis, one mostly based on a conceptual EE solution for the UE, and the other addressing EE on the network side, using a specific analytical approach. Additionally, the scientific publications that have derived from this work are also listed here, respectively, for reference.

1.3.1. Energy efficiency at the UE side

The current availability of mobile devices integrating multiple wireless radio interfaces, as well as other features such as low-power GPS modules, presents a good opportunity to capitalise on all such functionalities, with the support from IEEE 802.21 Media Independent Handover (MIH) specification. This IEEE standard provides the means to exploit available dynamic context information related to a UE, through the specified MIIS, allowing to intelligently control the radio network scanning mechanism in such modern mobile handsets.

In this thesis, a conceptually straightforward, but innovative, geo-referenced based RAT selection approach taking advantage of an IEEE 802.21 MIH/MIIS software implementation, is proposed targeting energy efficiency at the modern UE, extending its mobility further, exploiting the VHO mechanism in a multi-RAT HetNet. Numerical results have been attained through computer simulations using an *ns-2* based platform. Not only this approach leads to significant energy saving at the UE, but also it is considered a practical design, which is technology agnostic by nature, and eventually can be easily extended and implemented in real-life UE applications.

Publications:

- J. Bastos, M. Albano, H. Marques, J. Ribeiro, J. Rodriguez, and C. Verikoukis, "**Smart Interface Switching for Energy Efficient Vertical Handovers in ns-2**", IET Communications, Energy Aware Wireless Network Protocols, vol. 6, no. 14, pp. 2228-2238, DOI: 10.1049/iet-com.2011.0811, September 2012.
- J. Bastos, J. Rodriguez, and C. Verikoukis, "**Mobile Terminal Interfaces Management for Energy Efficiency**", 17th Int'l Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD 2012), Barcelona, Spain, 17-19 September 2012.

- J. Bastos, M. Albano, J. Rodriguez, and C. Verikoukis, "**Location Assisted Energy Efficiency for Multi-Interfaced Mobile Terminals**", 2nd IEEE Workshop on Convergence among Heterogeneous Wireless Systems in Future Internet (CONWIRE 2012), integrated in IEEE ICC 2012, Ottawa, Canada, 10-15 June 2012.

1.3.2. Energy efficiency at the network side

Energy efficiency has currently turn into one of the major challenges facing HetNets in today's wireless communication domain. In this thesis, an EE optimization is proposed for downlink OFDMA system in HetNet taking into account realistic network power consumption model (circuit power), together with other realistic constraints.

Many distinct approaches have been proposed to address the EE aspect in HetNets, taking advantage of various concepts coming for multiple domains, which have also been briefly tackled within the work carried out in the scope of this thesis, as it is briefly introduced in section 4.1.1.

In this work, it is explored how to use convex optimization theory to optimize the data rate in a DL multiuser HetNet. Given the optimization problem is non-convex in nature, we reconstruct the optimization problem as a convex one and devise a practical novel resource assignment algorithm for maximizing achievable EE, with quick convergence. Using these results, it is possible to design optimal energy consumption networks, by maximizing EE, based on QoS-oriented method for HetNet, while fixing total power.

Publications:

- J. Bastos, K. M. S. Huq, S. Mumtaz, J. Rodriguez, C. Verikoukis, "**Energy efficiency optimization for downlink OFDMA system in heterogeneous network with QoS constraints**", Int'l Journal of Communication Systems, Energy Efficient Wireless Communication Networks with QoS, DOI: 10.1002/dac.2969, 25 Jan. 2017.
- A. Antonopoulos, J. Bastos, C. Verikoukis, "**Analogue Network Coding-Aided Game Theoretic Medium Access Control Protocol for Energy-Efficient Data Dissemination**", IET Science, Measurement & Technology, DOI: 10.1049/iet-smt.2013.0192, July 2014.

- D. Yang, J. Bastos, C. Verikoukis, J. Rodriguez, "**Location-aided Round Robin Scheduling for Fractional Frequency Reused LTE-A Relay Network**", 17th Int'l Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD 2012), Barcelona, Spain, 17-19 September 2012.

1.4. Methodology and Thesis Organization

The general methodology carried out throughout the work leading to this thesis, and the presented novel approaches, followed six essential steps, which were applied to both main scientific contributions:

1. Literature survey on the respective state of the art;
2. Definition of appropriate realistic scenarios for evaluating EE solutions;
3. Modelling of scenarios, including technology specifications, using software;
4. Development of proposed solutions and their implementation in software;
5. Validation of simulation models and performance evaluation of solutions;
6. Analysis of simulation results and derivation of conclusions, leading towards possible iterations on proposed solutions.

After the broad introduction on the evolution of mobile communication systems, made in Chapter 1 of this thesis, the following Chapter 2 contains a further detailed revision of the state of the art more specific to each of the two main topics, respectively, which are key to support the EE solutions and insights presented in the subsequent chapters. Chapter 3 includes the outcomes from the work associated to the VHO-based solution towards EE in the UE, while Chapter 4 presents the most relevant work related to the RRM-based solution towards EE at the network side. Finally, Chapter 5 consolidates the conclusions achieved throughout this thesis work, together with some suggestions on relevant follow-up work and broader guidelines on investigation at longer term.

2. RESOURCE MANAGEMENT IN HETNETS

2.1. Vertical Handovers

Wireless RF-based communication is currently a fundamental technology that is used on many of our interactions within society, and on which people rely even without being aware of the involved mechanisms, in a rather transparent way. Different wireless communication systems have been designed, deployed and grown according to each's specific goals and major associated services or applications, e.g. users mobility, range, throughput, full duplex voice/video, internet browsing, etc. The deployment of such different systems, each focused on its specific scopes, can easily overlap in terms of coverage areas, while the deployment of identical systems can actually be planned in order to offer complementary coverage, as in a cellular network.

A communication network mechanism, known as handover (HO), has been considered since the early cellular systems. This mechanism allows transferring an ongoing voice call, video or data session, from the established channel connecting the UE to the network, to another channel that eventually is being used by a different BS in the same cellular network. A similar mechanism, known as vertical handover (VHO), can also allow transferring such ongoing communications, between different communication systems in a heterogeneous network, in a seamless way.

2.1.1. Handover mechanism

A handover typically occurs between two distinct cells, and respective BSs, when the UE is moving out of the coverage area of the cell where it is currently connected and is heading towards the coverage area of another cell. The latter cell is usually denominated as target cell, and the former as serving cell. Such kind of handover involving different cells, known as inter-cell handover, aims to allow maintaining an ongoing call or session established with the UE, while the user moves within the network coverage region. However, intra-cell handovers are also possible, where the channel being used in the ongoing call or session is replaced by the network, at the BS, during such handover, without involving any other cells or respective BSs. Such channel substitution is usually triggered by excessive fading or interference on the current channel, towards a better channel, i.e. with less fading and/or interference.

A cellular communication network that has been designed and deployed to allow mobility to UEs and respective users, as well as the appropriate QoS, should implement the HO mechanism and associated processes from top to bottom. Summarising, such mechanism should typically redirect an ongoing call or session in progress from the current serving cell and respective BS to the BS responsible for the target cell, which becomes the new access for the network to maintain such ongoing call or session. The main reasons for triggering an HO can be summarised as follows:

a) UE moving from one covered area (source cell) to another (target cell), without risking to drop the call or session;
b) Channel being used becomes excessively degraded and QoS cannot be met, or even lead to call or session termination, and thus an ongoing call or session gets allocated another channel, which can belong to the same cell or to a neighbour;
c) Source cell capacity is reached, or nearly reached, and the network implements a policy to transfer ongoing calls or sessions, from UEs located in coverage areas overlapping with other cells, to one of such cells and respective BS. This allows to keep some capacity in the serving cell for UEs that can only be served by it;
d) Depending on other network implemented policies, and also on the network RAT itself, several other handover situations are considered. Such as, to allocate ongoing UE communications to larger cells when the UE travels faster, and vice-versa. This allows to free some capacity in the respective serving cells, namely to balance the network capacity in the respective region, while also reducing possible interference to other UEs or cells.

One key aspect in the progression of the handover mechanism, associated to the UE's continuous monitoring of the different available channels, both from its serving cell and neighbour ones, is the signal level threshold that actually triggers the handover itself. Effectively, once the received signal power at the UE, coming from the serving BS, drops below a predetermined threshold, when compared against the received signal power from a neighbour BS, the HO is performed in order to avoid an eventual premature termination of a UE ongoing call or session. This process, involving received power measurements at the UE while it moves away from its serving BS and respective cell, is illustrated in Figure 8, where it is also represented the HO threshold margin, also known as HO hysteresis margin. This margin is necessary in order to prevent borderline situations where the UE can eventually be subjected to unnecessary consecutive handovers, back and forth between two cells, in what is commonly denominated as ping-pong HO effect. Once the received signal power, measured at the UE, coming from its serving BS is below the measured received signal power from the target BS by the predetermined HO threshold margin amount, the handover is executed and the target BS becomes the UE's new serving BS, and respective cell. This transition is illustrated in Figure 8 through the green coloured received signal power curves, respectively indicating which BS is serving the UE as it moves away from BS$_1$ towards BS$_2$.

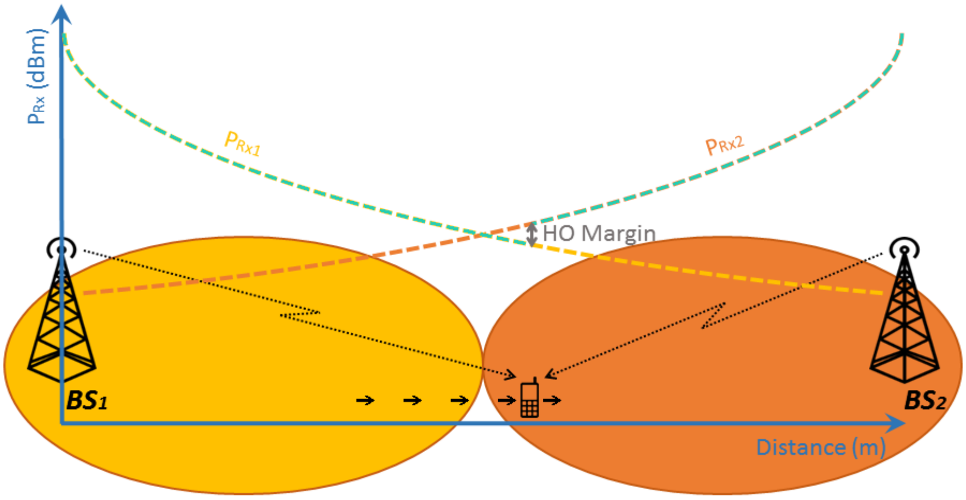

Figure 8: Typical handover including HO threshold margin.

In typical cellular networks where the handover mechanism is implemented, each cell and respective BS holds a list of all possible target BSs in its neighbourhood, which could take in charge ongoing calls or sessions from UEs currently being served by it, upon HO. Such list, specific to each BS and respective cell, is denominated as neighbour list. Each list is produced by dedicated algorithms, which can take as input numerical estimations of RF wave propagation in the region covered by those BSs, using simulation tools, being many times complemented by field measurements.

As previously mentioned, some characteristics of the signal being transmitted through the channel used for an ongoing call or session, are required to be monitored in order to be possible to consider the need of an opportune handover, avoiding an unexpected termination. Upon appropriate monitoring, once predetermined limits are reached, a handover can be typically requested by the UE or by its serving BS, according to measurements being made at any of the two, on the downlink or on the uplink. Additionally, the UE and the neighbour BSs also perform similar signal measurements between themselves, using specific channels, in order to be possible to select the most appropriate potential target BSs out of all the neighbours, thus allowing to rank them. Some of the factors, and associated measurements, decisive for a UE or BS to request a HO are received signal power, as already mentioned, typically provided by the received signal strength indicator (RSSI), and signal-to-noise ratio (SNR). Nevertheless, depending on the cellular network RAT, other criteria can also be considered, such as bit error rate (BER) and block error rate (BLER), or even the estimated distance between UE and BS taking into account the RF signal propagation delay. Moreover, In some cellular networks, depending on its RAT, it is possible that a HO is triggered to a target BS that is not in the neighbour list, whose BSs are more frequently checked namely in terms of measurements, since other nearby BSs are also checked, but less frequently.

Furthermore, as mentioned above, a handover can be carried out following one of three main straightforwardly denominated HO techniques: Network controlled HO; UE assisted HO; and UE controlled HO.

2.1.2. Hard and Soft handovers

As introduced in the previous subsection, the handover mechanism counts with an appropriately predefined HO threshold, among other key factors, which are decisive to trigger the effective handover. However, such effective transition of a UE's ongoing call or session from one BS to another, and respective cells, can actually be performed in two rather distinctive ways, depending on the cellular network implementation, and respective RAT. The most obvious approach, which essentially fits with the handover description presented in the previous subsection, in known as hard handover. In such handover, the channel being used for the UE ongoing call or session at the serving cell is firstly released, and only afterwards the ongoing communication is resumed on a channel belonging to the target cell and its BS. This channel transfer is, of course, made promptly in order to be as imperceptible as possible to the UE user, even though a communication brake is effectively made with the serving BS, which is immediately followed by its reestablishment through the target BS. This type of handover is also known as “break-before-make”.

On the other hand, the soft handover approach is considered to be more sophisticated, allowing further seamless transition of the UE ongoing call or session, although also involving further complexity in its implementation. In this type of HO, the channel being used by a UE and its ongoing call or session, through the serving BS, is not released immediately after the UE establishes connection with the target BS, through one of its available channels. Since the serving BS channel is released only after a certain period during which the UE is connected to both serving and target BSs simultaneously, this kind of handover is also known as “make-before-break”. Such period can be considered effectively as a state of the ongoing call or session since its duration can be significant, and the handover transition itself typically doesn't occur as a mere short event during the ongoing communication, as it actually happens during a hard HO. Moreover, the soft HO mechanism allows the possibility for the UE to establish connections with more than one BS while continuing with a communication using a channel from its serving BS.

The set of BSs and respective cells to which the UE is connected, receiving their signals during a soft HO, is denominated as the active set. The cell which BS produces the stronger signal received by the UE is denominated as the primary cell, providing the UE with the primary channel, while the other cells BSs in the active set provide what are denominated as handover channels.

Once a signal from an another BS, responsible for an additional cell, is received by the UE above a predetermined threshold (Th_{Add}), that cell and respective BS is added to the existing active set as long as it remains above that threshold for a certain amount of time ($Add_{TimeDelay}$) in order to prevent ping-pong effect. On the other hand, if a cell supported by its BS is currently in the active set and the signal received by the UE gets below another threshold (Th_{Drop}), that BS and cell will be dropped from the active set after a specified period of time ($Drop_{TimeDelay}$). Typically these thresholds are imposed on the measurements of the cells' pilot channel SINR (Ec/Io) at the UE. This process is illustrated here below in Figure 9.

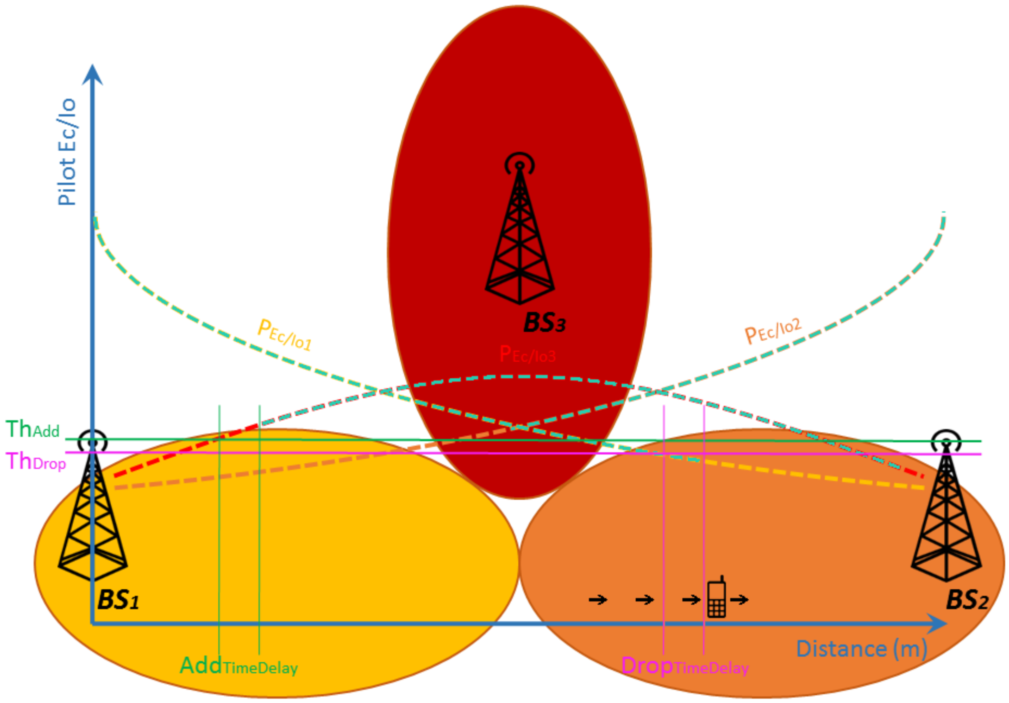

Figure 9: Soft handover including HO thresholds.

It is clear that soft handovers are only possible through specific advanced processes implemented on both UEs and cellular network, on several entities. Taking the latter aspect into account, due to the specificity in terms of the involved network entities, an additional kind of handover, the softer HO, is a variation of the soft HO mechanism, in which the BSs involved in the handover belong to the same sectorized cell site.

In the soft HO, being a call or session that is in that state, the signal of the best channel being used in the HO can be taken for carrying on with the ongoing communication, while in a softer HO it is possible to combine the signals in all such channels being used, producing a better signal. Such signal gain is known as macrodiversity gain, and it is possible to attain due to the uncorrelated nature of fast fading, which is diverse on the different involved cells, as well as the slow fading variation, to some extent. For such handling and combination of signals received in parallel, the UE typically uses what is known as a rake receiver. Moreover, in the softer HO, such signal combination is typically done on both downlink and uplink channels.

Although both types of hard and soft handovers could theoretically be implemented on any kind of cellular network and respective RAT, due to some aspects of the soft HO specific processes and respective implementation costs, on both UE and/or network, it has only been implemented in CDMA-based systems. Actually, the soft HO mechanism is important in CDMA cellular networks, namely because such networks typically are affected by interference caused by what is known as near-far effect. Such interference issue can be solved by the network, by requesting the soft HO of the UE ongoing call or session to a neighbouring BS in order to mitigate the interference, even if the UE connection to its serving BS is flawless.

Summarizing, due to its nature, the soft HO requires more resources, on top of the higher involved complexity, since during its processes it seizes more than one channel for a certain amount of time, while only a single channel is used in a hard HO at any time, for any UE ongoing call or session. The latter also requires less processing or handling from the cellular network, and the HO event can be practically unperceivable by the UE user. Moreover, the required hardware and equipment for supporting hard HO, namely at the UE, can be simpler, not needing to be able to receive two or more signals from different channels, and can thus be cheaper. However, it is possible that in the hard HO the respective ongoing call or session might be briefly interrupted or even terminated unexpectedly if some fault occurs during the execution of the involved processes. To prevent that, typically the hard HO mechanism also implements procedures for re-connecting to the serving BS if the handover to the target BS fails.

On the other hand, since the main principle of the soft HO mechanism is that the UE only releases its connection to the serving BS once it has established an appropriate connection to a target BS, handovers are usually successful and unexpected termination of the ongoing call or session are not common. Moreover, since the UE is connected simultaneously to more than one BS, through the respective allocated channels, an unexpected termination of the ongoing call or session could only happen if all channels would suffer from fading or interference at the same time, which is highly unlikely. In this way, and as most handovers occur due to cell's coverage limitations, the multiple connectivity inherent to soft HO assures higher reliability of the established UE's ongoing call or session, making one channel shortcomings not so critical in such situations. Again, of course such advantages are only possible through higher complexity, both in UE hardware and cellular network resources, where one single call or session is supported by several channels belonging to the different involved cells and respective BSs. This resource inefficiency limits the number of free channels in such cells and can eventually hinder the cellular network capacity.

A trade-off in the cellular network HO mechanism setup is evidently necessary in order to balance the advantages and the associated costs brought by the soft HO added complexity. Such approach can be done, namely by adjusting soft HO duration and associated thresholds, as well as time delays, in order to limit the areas where soft HO can occur. In this way it is possible to achieve further robust and reliable UE calls or sessions, while ensuring that the impact on network capacity is a small as possible.

2.1.3. Systems coexistence and IEEE 802.21 MIH architecture

The HO mechanism is also employed in further complex networks, involving different coexisting RATs, i.e. heterogeneous networks (HetNets), in order to allow the possibility to enhance UE connectivity, but also to optimise energy efficiency and communication costs. Such solution involves implementing the VHO mechanism in the respective HetNet, which should follow the IEEE 802.21 Media Independent Handover (MIH) architecture [15]. This standard provides support to allow seamless handovers between different kinds of communication networks, both wired and wireless, through specific VHO algorithms, without considering HO between similar networks, also known as horizontal or intra-technology HO. Therefore, it has been necessary to re-think the existing access selection polices and also the appropriate management mechanisms, including HO, namely towards addressing throughput, QoS, energy efficiency, and spectrum efficiency, among other aspects.

The emergence of this standard was preceded by extensive research towards effective interoperability between wireless local area networks (WLAN) and 3GPP cellular networks [16]. Taking into account such previous standardisation efforts, both from ETSI and 3GPP, and considering the point of integration regarding system architecture as main differentiator, all solutions for managing VHOs between WLAN and 3GPP networks can be classified into the following three categories.

Loose coupling is the broadly used solution among most proposed architectures. In this kind of approach the different technologies are interconnected by using Mobile IP (MIP), which has been standardised by the Internet Engineering Task Force (IETF). Such protocol, allowing mobile UEs to transit from one network to another while keeping the same IP address, is used here as the basis for such loose coupled integration and envisaged mobility between systems, in an independent manner. This approach allows a network infrastructure with advanced HO management, including HO initiation and execution, where both UE-assisted and UE-controlled HOs are possible, which is rather adequate for service continuity. However, since the interconnected networks are considered as independent concerning the handling of data traffic, this solution doesn’t offer tight performance assurance. Essentially, in loose coupling the focus to improve the offered call or session quality is done by addressing the various aspects of the HO mechanism, such as acceleration of Mobile IP procedures, advanced HO initiation algorithms and policy-based frameworks.

Tight coupling interworking solutions are closer to the cellular network operators’ concept, where a WLAN is considered as another access network to the cellular core network, and both data and signalling traffic are exchanged through the cellular network. This approach should be considered when a WLAN is attached to a 3GPP core network component, and affecting its functionality. Here, network-controlled and UE-assisted HOs are possible, which comply with the respective HO types existing in 3GPP networks.

Very tight coupling solutions essentially focus on interworking at the UTRAN level, namely by incorporating the functionality of the radio network controller (RNC) or lower level UMTS entities into WLAN components. Such functionality involves thus controlling the APs that are connected to the WLAN, including radio resource management. As in tight coupling solutions, soft HO decisions can be taken and executed. The implementation of very tight coupling solutions demand for combined WLAN and 3GPP/UMTS capabilities, both at the UEs and in specific UMTS nodes for supporting WLAN functionality seamlessly, and is therefore considered rather complex.

The IEEE 802.21 standard allows the possibility to take advantage of the different wireless technologies that are common to find simultaneously available in a growing number of places, following a loose coupling approach. Such concurrent variety of communication accesses poses several new requirements, challenges, and opportunities that most of the currently deployed architectures are not suitably ready to deal with. To allow this, the standard provides link-layer intelligence, as well as further information related to the network, to higher layers in order to optimize HO between networks based on distinct RATs.

The base and mechanisms of this standard were also motivated by the increasing service demands from end users, whose mobile terminals, or UE, are nowadays in most cases equipped with multiple distinct wireless interfaces, for several RATs, respectively. To take advantage of such RAT variety, both in terms of covered areas and respective interfaces at the UE, the latter needs to monitor continuously the channels available over different RATs, on the various networks, as in the previously introduced HO mechanisms. Such UE monitoring, not only can allow the anticipation of a necessary HO between network cells supported by the same technology or belonging to different RATs, but also can be used by the network to decide through which RAT a call or session is established for that UE. This allows, e.g., selecting the most energy efficient RAT while maintaining appropriate QoS. Otherwise, handovers can occur for mobile UEs when channel conditions change due to UE movement, while for stationary UEs handovers might occur when the network conditions change, eventually justifying such a switch of the UE towards a further appropriate network. In order to enhance the experience of both mobile and stationary UE users, and ensure service continuity through seamless HO between heterogeneous networks, as facilitated by the IEEE 802.21 standard, the respective architecture should be implemented.

Essentially, the standard provides what can be called as link-layer intelligence and related network information to higher layers so it is possible to optimize handovers involving different RATs in a heterogeneous network. Such kind of RATs, or media types, include the ones specified by 3GPP, 3GPP2, and by IEEE 802 family of standards, both wired and wireless. A key aspect, is that it supports using information available at the UE, also denoted as mobile node (MN), together with information existing within the network infrastructure, in a the cooperative way. The MN is also assumed to be is a good position to detect the available RATs or networks. Moreover, the network infrastructure appropriately stores overall network information, e.g. neighbourhood cell lists, MNs location, and higher layer service availability. It is specified that decisions about connectivity, to one or another BS and/or via which RAT, can be made by both the MN and the network. Typically, the network points of attachment (PoA), e.g. BSs and APs, can be multi-modal by supporting different radio standards and supporting simultaneous connections on more than one RF interface, as similarly happens in most currently available UEs, as previously mentioned.

For the considered heterogeneous networks, the IEEE 802.21 standard offers the following functionalities, which are provided through its specific components.

Service continuity is provided during a MN handover and the respective transitions that occur between the different link-layer technologies. This is supported by a framework, which includes media independent handover (MIH) reference models for the various link-layer technologies, and relies on a mobility management protocol stack within the network entities that support the HO.

VHO distributed support is provided by a set of HO-enabling functions within the protocol stacks of the network entities, hosting a new component known as the MIH Function (MIHF).

MIHF services are provided to the MIH users (layer 3 and above) through a MIH service access point, denominated as MIH_SAP, defined together with its associated primitives. The media independent services provided by the MIHF are the following:

1) **Event** service (ES) detects changes in link-layer properties and triggers appropriate events both on local and remote interfaces;

2) **Command** service (CS) provides a set of commands for MIH users to control link properties that are relevant to HOs and switching between links if required;

3) **Information** service (IS) provides information about the networks based on different technologies, as well as their services, enabling more effective HO decisions, to be made across such diverse networks and respective technologies.

The new link-layer service access points (SAPs) and associated primitives for each link-layer technology allow the MIHF not only to collect link information, but also to control link behaviour during VHOs.

The place where the MIHF resides within the protocol stack of a multi-interface MN or of a network element is illustrated in Figure 10. The MIHF can provide services to the MIH users, e.g. upper layers of mobility protocols, through a single media independent interface (MIH_SAP), while it can get services from the networks' lower layers through various specific media dependent interfaces (MIH_LINK_SAPs).

For the implementation of the standard and the envisaged effective operation of the considered networks, or HetNet, besides the need for the UE or MN to support multiple link-layer technologies and respective interfaces, it should also implement the MIHF. As a logical component, the MIHF must evidently also be implemented in the networks that will support VHO among them. Moreover, the MIHF, independently of where it is implemented, should receive and transmit information about the configuration and status of access networks that are within the MN's reach. Furthermore, such information is typically available at different layers of the protocol stack implemented at MNs or at the network's elements, and depending where it originated or is available, it is obtained by the local MIHF using distinct ways, as further detailed in Figure 11.

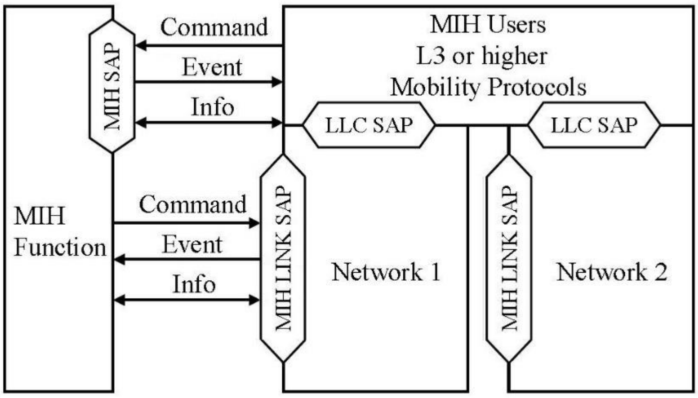

Figure 10: MIH services and interfacing.

Basically, when the MIHF on a local network element needs information about configuration or status of remote access networks, it can obtain such information by exchanging MIH messages with the peer MIHF instance at the remote network element. On the other hand, if the MIHF of a MN or of a network element needs information existing at lower layers of the protocol stack within the same entity where it resides, the MIHF can obtain such information locally by using service primitives of the SAPs interfacing it with those layers.

The IEEE 802.21 standard defines SAPs and primitives that allow to provide generic link-layer intelligence independent of the specifics of MNs or RANs, essentially through a generic interface between the link-layer users in the mobility management protocol stack and the existing media-specific link layers. Therefore, the considered media-specific technologies, e.g. specified by 3GPP and IEEE 802 standards, must enhance their media-specific SAPs and primitives to satisfy this standard generic abstractions.

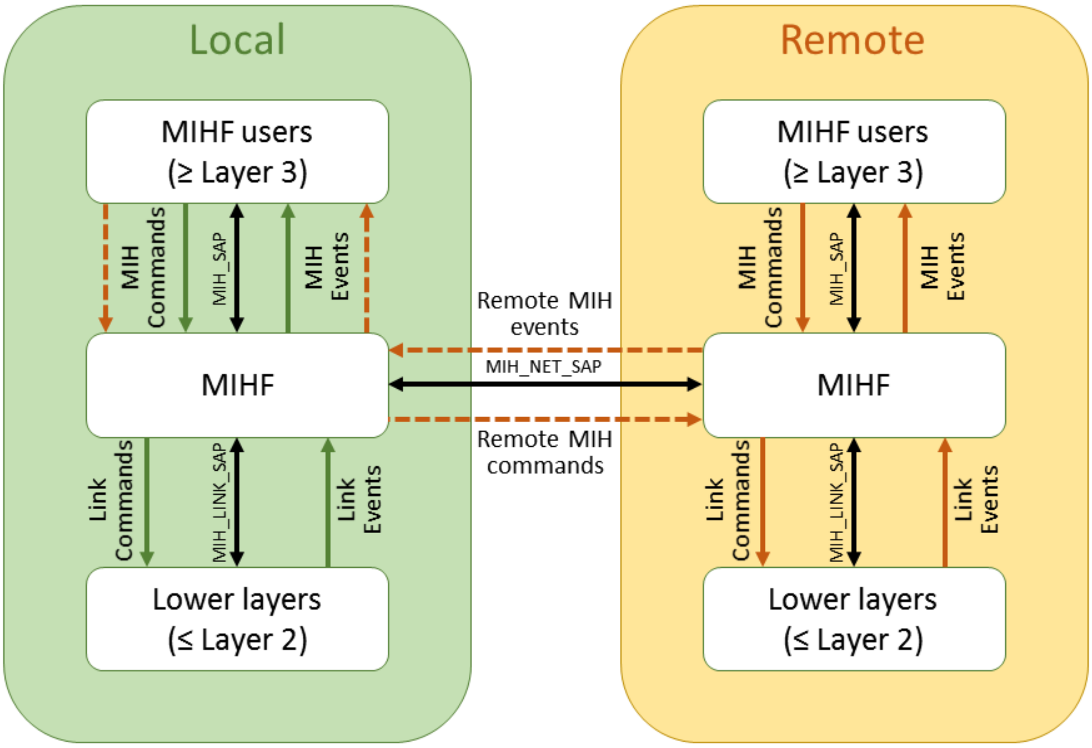

Figure 11: Services flow at entities, locally and remotely.

2.2. Dynamic Resource Allocation

The seamless integration and interoperability between different RATs is at the core of HetNets, namely towards increasing the system performance and energy efficiency both at the network and the users side. In order to achieve that, as previously mentioned, the deployment of low power BSs, providing microcells/picocells and femtocells, inside a macrocell supported by a typical BS allows exploiting lower power shorter radio links, leading to enhanced EE in the network. Furthermore, it also allows traffic load balancing to different BSs, and RATs, which nevertheless implies better resource allocation and utilization.

In RRM, besides the eventual load-balancing between macro and femto/small tiers, and also between distinct RATs, there is another common approach, which is typically complementary, denominated as Dynamic Resource Allocation (DRA). This procedure, continuously performed on the network side, targets to allocate resources to mobile terminals in a further optimized way, dynamically according to the time-varying environment, while taking into consideration the required QoS in question.

Current and future cellular wireless networks are, and will be, highly influenced by how efficient the exploitation of the available radio resources is, e.g. bandwidth (frequency bands or channels), time slots, and power. Moreover, the limited available radio spectrum represents a high cost to cellular/wireless operators. It is, thus, key to investigate and deploy techniques and mechanisms that allow maximizing the network spectral efficiency, while providing the required QoS, as expected by the end users. Such networks should therefore adopt efficient strategies for radio resource management taking into account the diverse packet-based services that are provided, or supported, to the end users.

DRA targets adapting the transmission capability of the network’s existing transport channels according to the time-varying traffic conditions in order to accommodate several users’ communications in a shared common channel. Additionally, cross-layer information provided by the physical layer (PHY) of the communication system, but also by higher layers, e.g. Application layer, can be exploited towards achieving further SE by providing efficient mapping of data packets onto the available radio resources at the Medium Access Control (MAC) sublayer. This, results in the optimization of network performance, while satisfying the applications required QoS.

2.2.1. DRA general architecture

In a cellular wireless network system, the DRA entity offers the respective operators the flexibility to provide broadband services, and associated traffic, with high spectral efficiency by managing radio resources in an optimal fashion, following specific policies. The DRA is typically part of the MAC sublayer in a cellular wireless network, allowing the possibility of also achieving significant energy savings, namely by applying specific scheduling policies, and transmission protocols, in order to appropriately serve the users, and their UEs, according to the services in question. Typically, the DRA architecture includes the Scheduler, the Link Adaptation (LA), the Resource Allocation (RA) and Hybrid Automated Repeat Request (HARQ) blocks, as illustrated in Figure 12.

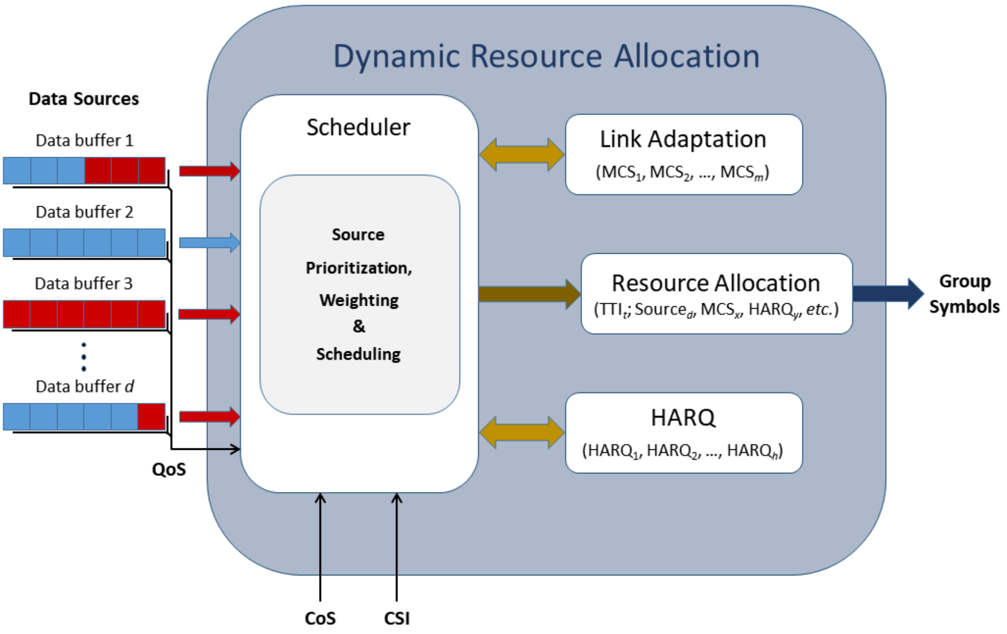

Figure 12: DRA typical architecture.

Essentially, the DRA implements at its core a Scheduler, which performs the prioritization, weighting and scheduling of the arriving data packet streams to be sent to the multiple users, and distinct services, respectively. In order to perform such operations the Scheduler counts with the information of the respective buffers utilization and QoS, as well as of the Class of Services (CoS) involved. The latter are associated to parameters that are set at the beginning of the communication sessions, which remain unchanged throughout the duration of the whole sessions, respectively. These parameters, which are typically defined in standard service applications for each CoS, include the Peak Bit Rate (PBR), the Guaranteed Bit Rate (GBR) and the tolerated Bit Error Rate (BER), jointly with the maximum allowed application delay (T_{CoS}).

To achieve its main purpose, the Scheduler runs a scheduling algorithm, such as Round Robin (RR), Maximum C/I (CI) and Proportional Fair (PF), as introduced here in the next subsection.

As a channel adaptive scheduling module, the DRA not only takes into account QoS aspects, as previously mentioned, but it considers them jointly with Channel State Information (CSI) for the optimized resources allocation to better serve users, respectively. This means that, the assigning of priorities to transmit the respective data to the users, is carried out together with selecting the best location for transmitting users' data, e.g. in terms of subcarriers in an OFDMA system, by allocating the most suitable resources to them, according to each own CSI.

The DRA selects the most spectrally efficient implemented Modulation and Coding Scheme (MCS), involving iteratively the LA and Scheduler functional blocks, that maximize the maximum bit rate while keeping the predicted Block Error Rate (BLER) below a predefined threshold. It is worth mentioning that such MCS selection can also take into consideration other aspects, as, for instance, more complex MCS typically involve more energy per bit to be transmitted, than simply considering transmission with multiple parallel code channels.

As previously introduced, in order to be possible to perform adaptive modulation and coding, CSI is necessary, eventually through Channel Quality Indicator (CQI) messages reported by each UE to the BS, upon local measurements, respectively, through a feedback channel or UL transmissions. This cross-layer architecture benefits not only from fast and efficient channel estimation and feedback, but also from control channels for the fast exchange of signalling messages for the HARQ block. Such swiftness is important for the latter, allowing UEs detecting a corrupted message upon reception, to request a new message to the BS, receiving it promptly and eventually combining the respective messages locally. For that, in Hybrid ARQ, the data is encoded, at the BS before transmission, with a forward error correction (FEC) code, which is iteratively selected in the DRA to correct an expected subset of errors that might occur. Typically, such selection depends on signal quality, e.g. simple HARQ FEC code is selected when poor signal conditions exist and quality is low, and standard ARQ is chosen when the signal quality is above a certain threshold. Basic ARQ doesn’t involve any FEC code, and redundant bits are simply added to the data to be transmitted, employing an error-detecting code, e.g. a cyclic redundancy check (CRC).

The RA essentially consolidates, after DRA iterations, the specific resources, e.g. PRBs to be allocated, arranging them for transmission of the data packets to the UEs, respectively. These are mapped and updated at every frame by the BS, i.e. employing specific MCS and HARQ, among others, including, e.g. the best carrier gains associated to each UE, in each transmitter group block, in an OFDMA system.

2.2.2. Scheduling for capacity-QoS optimization in OFDMA

As previously introduced, the DRA’s ultimate goal is to maximize the cellular system capacity, while assuring the respective QoS requirements are met, respectively, for the services being provided to the UEs served by the BS in question. Essentially, the BS needs to distribute its transmission time, typically split in multiple fractions denominated as transmission time intervals (TTI), and schedule it to appropriately serve the different UEs according to predefined policies.

Possible approaches to address this scheduling problem are to use complex cost functions, or also sets of QoS restrictions to balance network system capacity and users' satisfaction. There should be, thus, a certain fairness being maintained in the scheduling time to serve each of the UEs in question, while using the available network system resources in the an optimal way. The three most common ways to schedule UEs in a BS DL transmission period are introduced as follows.

Round Robin (RR) is the simplest of all scheduling approaches, where priority is straightforwardly given to each UE in respective sequential order, without taking into account any associated QoS requirements. This basic time division multiplexing approach, where the delay between successive transmissions to the same UE is constant and identical for any UE being served by the BS in question, can actually be seen as non-prioritized scheduling. Also, taking into account that no CSI is used in this scheduling method, it is clear that system capacity is far from the channel limits and valuable resources are most likely wasted. Nonetheless, for the exact same exposed aspects, this scheduling approach is very simple and easy to implement, and assures all UEs required to be served actually get served in an perfectly equal way.

Maximum C/I (CI) scheduler, based on Maximum Carrier-to-Interference ratio, targets potentiating system throughput maximization by allocating resources to the UE that has the highest channel gain. For that, the UE providing the BS with the better (highest) CQI get its data opportunistically scheduled just before each frame *n* is to be transmitted, respectively, according to the following rule:

$$k^{CI}(n) = \arg \underset{i\in\{1,\ldots,K\}}{max} CQI_i(n)\,, \quad n = 0, 1, 2, \ldots \tag{2.1}$$

In the above expression, *K* stands for the number of UEs being served by the BS, from which the i^{th} UE with the highest CQI(*n*) is selected for respective resources allocation. Although this scheduling approach is very effective in resources allocation towards capacity maximization, eventually offering maximum system capacity, it totally lacks fair distribution of the available resources to be shared by the UEs. The latter can be particularly harsh for UEs at the cell border, typically with lower CQIs, which may suffer significantly also due to this approach's total disregard of the QoS requirements associated to the services being provided to the UEs.

Proportional Fair (PF) scheduler was proposed in [17] as an approach to mitigate the inadequacies of the CI scheduler. Essentially, this approach targets to balance between achieving the maximum system throughput, while providing at least minimal service to every UE being served by the BS in question. For that, just before each frame *n* is to be transmitted, respectively, the scheduler selects for transmission the data to be sent to the UE with the highest ratio of estimated maximum data rate $R_i(n)$ to current average throughput $T_i(n)$, according to the rule:

$$k^{PF}(n) = \arg \max_{i \in \{1,\dots,K\}} \frac{R_i(n)}{T_i(n)}, \quad n = 0, 1, 2, \dots \tag{2.2}$$

It is clear that, according to the above expression, if a UE is not being served, its average throughput tends to zero, and when that starts to happen its priority increases dramatically, eventually to infinity, placing its data, to be transmitted by the BS, in front of the queue for DL transmission. Assuming the radio channels from BS to the UEs are symmetrical, independent and identically distributed, the PF scheduler can be considered rather fair in terms of resource allocation, while being effective in terms of maximizing the system capacity.

As the PF scheduler targets to balance between QoS requirements and respective priorities, as well as total system throughput, this approach is typically the one preferred in commercial cellular networks by the operators. Nonetheless, there are also some disadvantages in the PF scheduling approach, and there are scenarios where CI, or even RR scheduling, can actually be preferred over it.

For instance, non-opportunistic scheduling, such as PF, reduces system throughput in the cell being served by the BS in question as saturation is reached at lower loads when comparing against CI. The latter can actually be the best option when most of the cell traffic is UDP-based, such as VoIP or video streaming, which relies on fixed bandwidth and thus fairness is not necessary. On the other hand, when considering mostly TCP-based traffic, such as FTP or web browsing, which essentially can be considered as being elastic since the more bandwidth is available the more it is used, if CI is implemented there could be occasions when a TCP connection with a UE having a high CQI may starve the other cell's UEs, namely at its edge. In this scenario, PF scheduling would surely prevent such harmful situations to happen.

The bottom line is that opportunistic CI scheduling pays off, since the system essentially works with larger frames. Nonetheless, small adjustments can be useful, namely by considering QoS requirements even if at the expense of some system throughput. One possible straightforward solution could be to prioritize traffic that is less tolerant to latency, such as VoIP connections, by placing them ahead of other more latency tolerant traffic, eventually using CI scheduling. Another approach could be to implement an urgent queue for such less latency tolerant traffic, where UEs' traffic near the respective deadline get into that queue in order to be served with higher priority, while all other UEs get served using CI, and even UEs whose VoIP packets still have distant deadlines are also served opportunistically.

2.3. Interference Management

The deployment of typical HetNets and respective operation, also involving small cells as envisaged in 5G, and introduced earlier in section 1.1, can be rather challenging. In addition, considering cells of different sizes can increase the complexity of planning the respective network, namely due to potentially higher risks of RF interference, which could jeopardize the intended benefits when considering such approach in the first place, and could eventually lead to service degradation. Thus, it is crucial to consider some form(s) of interference management, which can be addressed in different ways, from thorough pre-deployment planning, to sophisticated operational approaches.

2.3.1. HetNet planning

In a cellular network using frequency reuse of one, like 3GPP LTE, and considering a HetNet with different cell sizes, the UEs typically establish connection to the BS, responsible for a cell, that provides them with the strongest received signal in DL (SSDL). Therefore, the limiting edge between two cells is found where SSDL is equal on those cells. The same normally coincides with the place where equal path loss for the UL (PLUL) occurs in both cells, in a homogeneous network. On the other hand, in a HetNet where BSs of diverse power coexist, e.g. higher power BSs responsible for macro cells and lower power BSs sustaining small cells, the places of similar SSDL are not always necessarily the same as that of identical PLUL. This can be easily settled as a downlink cell border defined by SSDL observed from the macro cell and the small cell, which are identical at a location that typically is closer to the lower power BS, while an uplink cell border defined by PLUL that is equal for both macro and small cells at a farther distance from that same BS. This derives from the fact that UE’s transmission power is identical whether it is connected to the macro cell’s BS or to the small cell’s BS.

In order to be possible for small cells to actually serve enough UEs, which is key in HetNet planning, a possible solution is to increase their coverage area. This can be done by adding an offset to the SSDL of small cells, thus slightly fostering small cells selection by increasing their range, which would attract more UEs to them, increasing HetNet efficiency. This is called Cell Range Extension (CRE), and is illustrated in Figure 13.

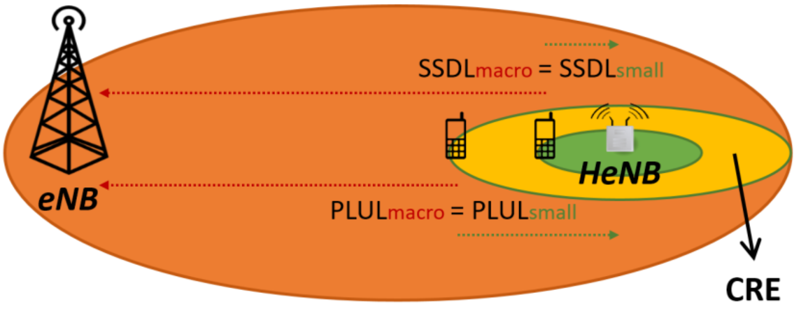

Figure 13: HetNet Downlink and Uplink cell edges and CRE.

Essentially, when a UE is in the CRE region, depicted in yellow in Figure 13, it is not necessarily served by the BS with the strongest DL received power. However, implementing CRE in a HetNet, besides increasing network efficiency by allowing offloading macro cell's traffic to the small cell, it contributes to reduced UL interference in the system. This occurs since a UE served by a macro cell, when locating close to the small cell border, not implementing CRE, would be transmitting at a higher power, which possibly would interfere with the UL of UEs served by the small cell. When implementing CRE, that same UE is more likely to be served by the closer small cell, thus involving lower path loss, which would allow the UE to transmit at lower power and eventually cause much less interference.

The range of the small cells is extended by implementing CRE through a selection offset when the UEs are in idle mode, or through a handover (HO) threshold adjustment in the measurement configuration, when the UEs are in connected mode, in favour of the small cells, benefiting HetNet efficiency. However, any UEs in the CRE area, which are thus being served by the low power BS responsible for that small cell, will most likely experience significant DL inter-cell interference caused by the higher power BS serving the surrounding macro cell. This DL inter-cell interference occurs, as already mentioned, since the macro cell's SSDL is higher at the location where those UEs are receiving transmissions from the small cell that presents lower SSDL at the UEs. Therefore, since the UEs' DL received power from the macro cell is higher than the small cell, the signal-to-interference-plus-noise ratio (SINR) at the UEs being served in the CRE area by the low power BS, responsible for the small cell, is below 0 dB. A direct consequence of such inter-cell interference is the potential impact on such UEs' reception, particularly of the DL control channels, which can be problematic.

3GPP has standardized several solutions that can be implemented in LTE networks to mitigate inter-cell interference issues involving HetNets and small cells. Typically, CRE is jointly considered and implemented together with such interference management schemes, namely involving coordination, to mitigate these interference problems.

2.3.2. Inter-cell interference solutions

In parallel to the CRE approach specification, several interference coordination solutions have also been standardized throughout several 3GPP Releases, starting in Release 8, where inter-cell interference coordination (ICIC) was specified.

As already introduced, LTE implements unitary frequency reuse factor for maximizing the efficiency associated to the allocated RF spectrum, which implies that, considering different cells, any transmissions using the same time-frequency resource will cause interference at the edges of contiguous cells. Therefore, in such network, a UE located close to the border of the cell that is hosting it is prone to interference caused by a BS responsible for a neighbouring cell, on top of the desired DL signals transmitted by the BS that is serving it. Additionally, any UE in the same circumstances can also cause UL interference to a BS of a neighbouring cell.

ICIC has been specified to mitigate such inter-cell interference (ICI) through radio resource management (RRM) techniques, by taking into account several aspects and metrics concerning the network, such as resource usage status and traffic load situation, from multiple cells. Essentially, ICIC is a multicell RRM function, which addresses ICI on traffic channels, by managing radio resources, namely radio resource blocks, in a way that the involved cells' BSs coordinate the usage of frequency domain resources. The general ICIC scheme and related processes are illustrated in Figure 14.

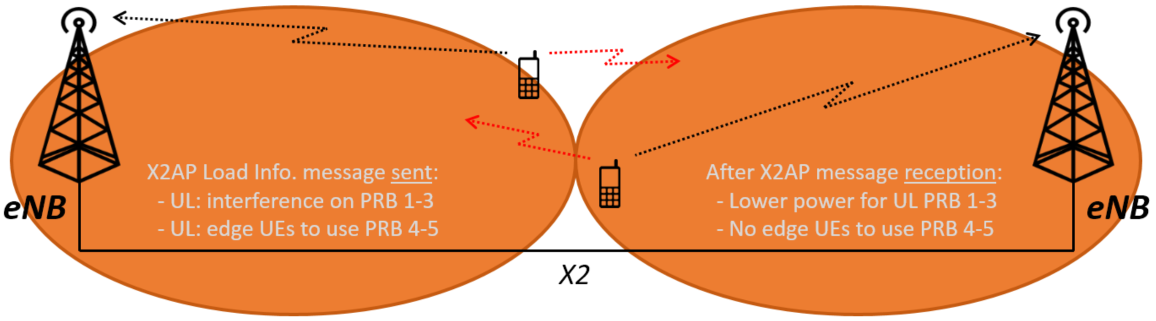

Figure 14: Radio resource management associated to ICIC.

It is possible to observe In Figure 14 an example of a potential UL interference scenario involving UEs at the cells edge.

The required radio resource management to implement ICIC is achieved by using the existing *X2* interfaces that interconnect neighbouring eNBs in the network, to pass ICIC signalling messages for the coordination between such eNBs responsible for the cells in question. The *X2* Application Protocol (X2AP), among other functionalities, provides the necessary means for implementing ICIC, namely by using the specified “Load Information” message, through which an eNB can inform neighbouring eNBs about relevant network status and metrics. Such information contains the indication about the UL interference level per physical resource block (PRB), as well as UL PRBs that are allocated to UEs at cell edge, which are thus prone to UL interference. Also, there is an indication if the sending eNB DL transmit power is above or below a set threshold value. All this information can be used by an eNB receiving such X2AP “Load Information” messages to optimize the scheduling of radio resources to be allocated to the UEs that are at the cell edge. This kind of ICIC solution is typically classified as a coordinated distributed approach as it takes advantage of the signalling (coordinating) messages exchanged between neighbouring eNBs via the *X2* interface.

2.3.3. Fractional Frequency Reuse

Another ICIC approach, Fractional Frequency Reuse (FFR) doesn’t require exploiting any sort of cooperation between cells or signalling messages with indicators on network resources, exchanged between neighbouring eNBs. FFR is an ICIC technique that essentially targets to improve spectral efficiency of LTE networks by exploiting frequency planning, or reusage, as previously introduced in GSM networks in a simpler fashion, to increase network capacity while minimizing ICI [18][19].

In LTE downlink OFDMA scheme, the available bandwidth is partitioned into many narrow subcarriers, which are orthogonal between each other and can thus accommodate the transmission of distinct data streams. Twelve of these subcarriers, spanning 180 kHz in the frequency domain, form one PRB, which in the time domain is composed of seven OFDM symbols, equivalent to one time slot of 0.5 ms. PRBs and respective power are allocated periodically by DRA schedulers at every TTI (1 ms).

Considering the division of network cells into inner and outer regions, respectively, served UEs would typically be subject to distinct experiences, i.e. the ones in the latter region would be closer to neighbouring cells and further distant from their serving eNB, thus, being more likely subject to ICI at cell-edge. In FFR, where such a static division of each cell is considered, the main goal is to safeguard from typical ICI issues the PRBs to be allocated for such UEs at cell-edge zone. To address this, a set of protected PRBs is created, which can only be allocated to UEs being served in that problematic area.

FFR distributes the existing PRBs, and respective RF spectrum, into two distinct sets, or sub-bands, to be allocated for UEs being served by a cell, depending if they are located close to the centre of the cell or near its edge. The set to be allocated for the UEs at cell centre is called Full Reuse (FR) since the allocated spectrum follows frequency reuse-1 model in the surrounding cells, while the set for the UEs at cell-edge is named Partial Reuse (PR) since its PRBs usage in the neighbouring cells follows frequency reuse-3 model, as illustrated in Figure 15. The available RF sub-band for the latter UEs is therefore partitioned across neighbouring cells in order to mitigate ICI, while a common sub-band is allocated to any other UEs inside any cells’ centre.

While this static ICIC technique effectively allows reducing ICI inflicted on UEs at cell-edge, FFR doesn’t adapt dynamically the distribution of resources (PRBs) between the two cell zones taking into account the demand from UEs, respectively. Moreover, FFR is also dependent on a reasonable classification of the UEs, being served by a cell, regarding their precise location within it, in order to be possible allocating them the most appropriate set of PRBs, i.e. FR or PR. The availability of all UEs positioning information is thus at the core of this ICIC technique, which should eventually imply some additional exchange of signalling messages.

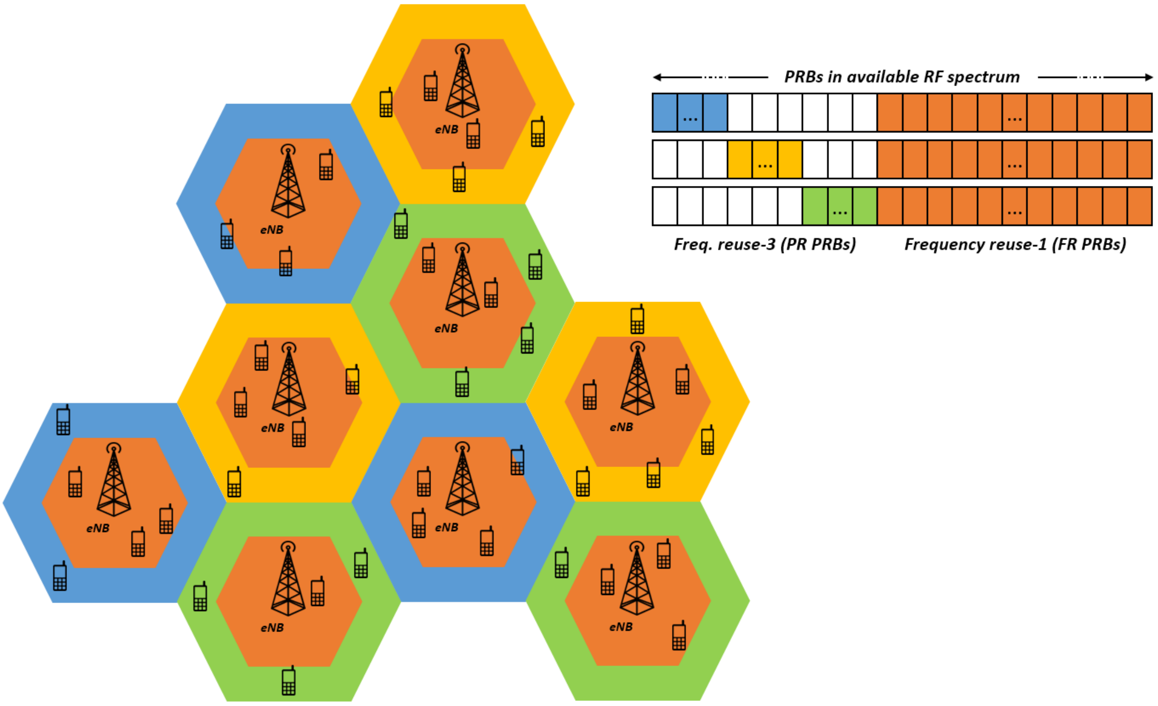

Figure 15: FFR dual frequency reuse approach.

This FFR ICIC approach can be deployed strictly, sometimes also under the name of Partial Frequency Reuse (PFR), but there is also a variation to it denominated Soft Frequency Reuse (SFR), which addresses its inefficiency regarding bandwidth utilization [20][21]. As opposed to what is typically called strict FFR, in SFR, each cell can take advantage of the total available bandwidth and respective associated PRBs. In order for this to be feasible, while strict FFR only sets restrictions on PRBs allocation for the UEs in each cell, SFR also performs that RRM procedure, but jointly with power allocation regarding the used PRBs.

Essentially, an SFR deployment considers the same frequency reuse-3 model on cell-edge, employing the same cell-edge bandwidth partitioning strategy, as in strict FFR. The key difference is that in SFR the UEs closer to the cell centre can be allocated PRBs belonging to the sub-bands reserved for UEs at cell-edge in other cells. Since the former UEs share bandwidth with neighbouring cells, the respective transmission power is typically lower than the power used for the latter UEs, as also represented in Figure 16.

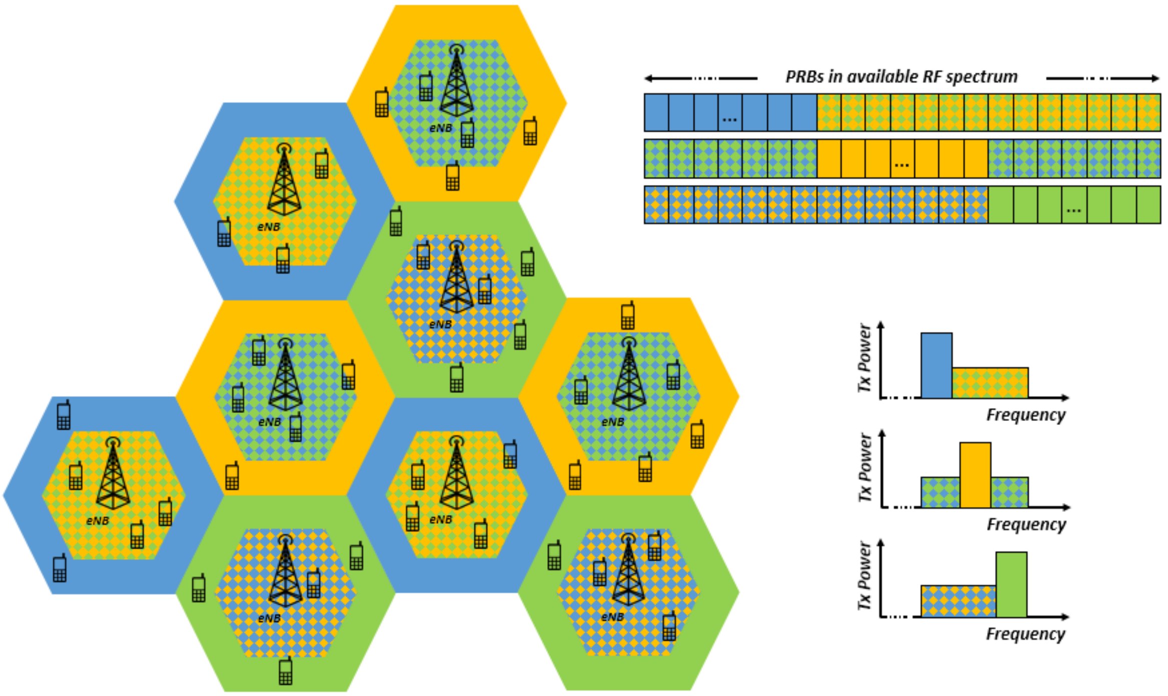

Figure 16: SFR approach, including PRBs power allocation.

While SFR is more bandwidth efficient than strict FFR, all UEs, either at cell edge or closer to its centre, can be afflicted from more interference as all RF spectrum can be used simultaneously, even with the considered restrictions put in place.

As previously introduced, the presented ICIC approaches are independently used in each cell without requiring any cooperation between neighbour eNBs. Therefore, they can be considered as autonomous distributed static techniques, as opposed of being coordinated distributed or centralized ICIC approaches, as is the case of the one presented in the following subsection.

2.3.4. Coordinated Multipoint

Coordinated Multipoint (CoMP) has been specified in 3GPP Release 11 in order to better support HetNets, namely to improve QoS for UEs on the edge of a macro cell and inside small cells or hotspots. In typical homogenous networks, UEs that are located at a cell edge undergo through lower DL signal strength in addition to interference caused by the neighbour cell(s). The CoMP scheme, considered as a centralized ICIC approach, allows using neighbouring cell(s)' eNB(s) to also transmit the same signal as the eNB serving a UE at its cell edge, enhancing QoS all around the edge of that cell, and ultimately on edges of all other cells in the network.

Moreover, implementing CoMP in a network can allow that a UE is able to benefit from the best DL and the best UL resources. That can be achieved by using a range of techniques, which enable the dynamic coordination of transmission and reception via multiple BSs located apart geographically. This allows achieving overall enhanced quality for the UE/user, while also improving the utilisation and efficiency of the network. Essentially, CoMP ensures that what usually is ICI, particularly at cells edge where DL and UL performance usually deteriorates, becomes in fact useful signal.

Several transmission or reception nodes, such as eNBs, RRHs, RNs and HeNB, are coordinated in CoMP to serve a UE through DL reception from multiple nodes and/or through UL transmission to multiple nodes. This is achieved by transmitting DL data in the same PRBs, at the same time, by more than one transmitting node, as what is called *Joint Transmission*. Another way is to receive DL data coming from one transmitting node in one subframe and coming from another transmitting node in the next subframe. This technique is known as *Dynamic Point Selection*. Figure 17 illustrates both of these CoMP schemes.

In a HetNet implementing CoMP, several macro cells and small cells can then participate jointly in DL and UL data transmission to and from a UE, respectively. The UE even has the possibility to benefit from the best UL via the small cell and the best DL through the macro cell, if it is inside the CRE zone, which would nevertheless require synchronization between the respective BSs.

The concepts, schemes and techniques introduced in the previous subsections, as well as their evolution, among others, such as advanced MIMO schemes, are considered fundamental towards the deployment of future sophisticated cellular networks.

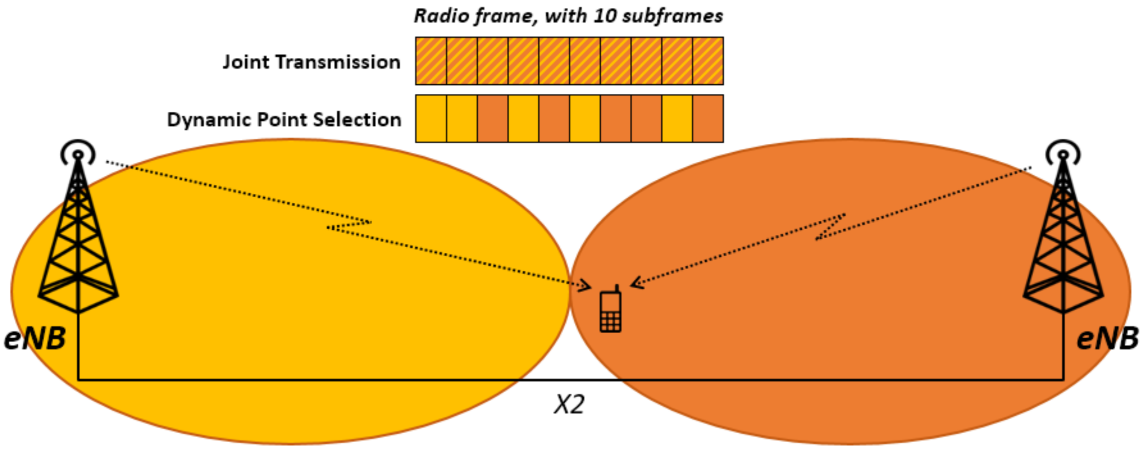

Figure 17: Multipoint transmission in CoMP.

These approaches will not only participate in the EE challenge, but will also enable significantly enhanced user experience, ultimately rewarding the respective service providers, and thus benefiting both end users and operators. It is therefore important that any cellular network infrastructure and UEs are designed in a holistic and concerted approach also towards further energy efficient processes involving its multiple entities.

ICIC continued to be further developed, namely towards better supporting HetNets, having resulted in Enhanced ICIC (eICIC), which provides interference control also for DL control channels, and not just data channels. Further enhanced ICIC (FeICIC) is also specified in 3GPP Release 11, where the time domain was also considered in the ICIC scheme. Both of these schemes also take advantage of information passed through the *X2* interface, using X2AP messages between eNBs, together with other techniques. Furthermore, in emerging 4G and 5G networks, additional benefits can be achieved by adopting C-RAN implementation, eventually together with virtualisation and cloud computing using SDN/NFV technology and architecture, as introduced in section 1.1.4.

Efficient ICIC techniques can improve not only cellular networks' spectral efficiency but also their energy efficiency, which is key for the network operators to further capitalize on their exploitable spectrum share, as well as by helping to reduce the respective network operational costs.

2.4. Energy Efficiency

Any process used for communication between an end user device and a supporting infrastructure, or another user device, requires an amount of energy for its accomplishment. Such required energy results from its spending in several different ways, both at the device(s) belonging to the end user(s) (UEs) and in the network equipment composing the operator's infrastructure.

The focus of the work considered in this thesis lays mostly on the required energy to be diffused in the transmission of data in a cellular network, both by the user's UE and by the involved base station(s) in the operator's network, since those typically represent the biggest part in the communication budget.

2.4.1. Why energy efficiency

The "always connected" paradigm, which started to be massively adopted by modern society in the past decades, progressed in an Always Best Connected (ABC) fashion, where any telecommunication service or application should be available to an end user through the devices and network technologies that best suit his/her needs at that time. This trend of seamless full connectivity among people and devices, any place and at any time, also towards accessing knowledge and culture, through multiple services and applications, continues to evolve, even though embracing necessary trade-offs.

It is well known that ICT, from which mobile communication systems represent a significant share, is responsible for a growing share of energy consumption worldwide, inflicting respective non-negligible impact on our planet's environment [22][23][24]. Such increase derives directly also from the non-stopping growth in data traffic, not only originated by the multiplication in the number of always connected devices, which nowadays typically sum up to "many" per individual user, but also due to further sophisticated services requiring the exchange of higher amounts of data, and faster. This escalation in the amount of energy expected to be used to provide further demanding mobile communications services, keeping up with modern trends and end users behaviour and expectations, requires appropriate solutions.

Such solutions must allow operators to accommodate the additional traffic volume, if possible, without increasing energy costs, which would contribute to keep modern services also environmentally sustainable. Moreover, the network operational costs (OPEX), supported by an operator, also have an impact on the price for the service(s) eventually charged to their clients, as users of their network.

The quest for increasing energy efficiency (EE) in mobile data communication should of course take into account the current consolidated trend of differentiated communication networks, i.e. HetNets, by exploring their diversity and eventual distinct associated energy spending. Thus, HetNets should not only be considered as a possible solution towards enhancing coverage and capacity, but also EE. In this respect, context awareness approaches can be considered, where the network(s) and UEs could take advantage upon specific knowledge on current communication scenario characteristics and associated conditions, e.g. accurate positioning and dynamics of UEs, towards enhancing the overall HetNet EE and performance.

Apart from the possible contention or reduction of costs, obviously appreciated by network operators, while contributing at the same time towards providing more sustainable and cleaner telecommunication services, EE is also quite valued on the UEs side, namely by their users, although mostly for rather distinct reasons.

As mobile devices, UEs are naturally designed targeting to be small and light in order to be possible to be carried effortlessly, namely in the case of wearable devices. Of course, such typical dimensions follow trends, which always consolidate towards satisfying the end users demands and expectations, namely regarding the more frequently used services and applications, as illustrated in Figure 18 in a caricaturized way. Such necessary limitation on UE’s dimensions imposes as well an obvious restriction to the maximum amount of energy that can be stored in the UE battery, independently of the currently existing technologies. Therefore, UEs autonomy is always bounded by such limited energy storage capacity.

There are of course several ways to address such autonomy and EE issues in the UEs, namely by limiting its available features according to the user needs. However, most of the communication aspects, e.g. transmit power and MCS, which would eventually contribute to a higher EE at the UEs, are typically set by the network that provides services to their users, according to the dynamics of the communication scenario, e.g. channel conditions. Such kind of EE approaches, which can benefit both UEs and network regarding the necessary energy to be spent for transmitting a certain amount of data, can include the most appropriate selection of RAT and/or MCS, energy-wise, as also pointed out in previous sections. For instance, under a further strict EE policy, the DRA at a BS can instruct for the selection of less complex MCS if that would involve spending less energy per bit to be transmitted, which would certainly contribute to a potentially higher UE autonomy, assuming QoS is also met.

It is thus important to address this delicate balance between offering multiple services to end users, with distinct QoS requirements, without having to sacrifice UEs autonomy or ergonomics, by considering further EE oriented solutions, while also not discarding updating the network infrastructure, e.g. with multiple antenna technology, i.e. MIMO.

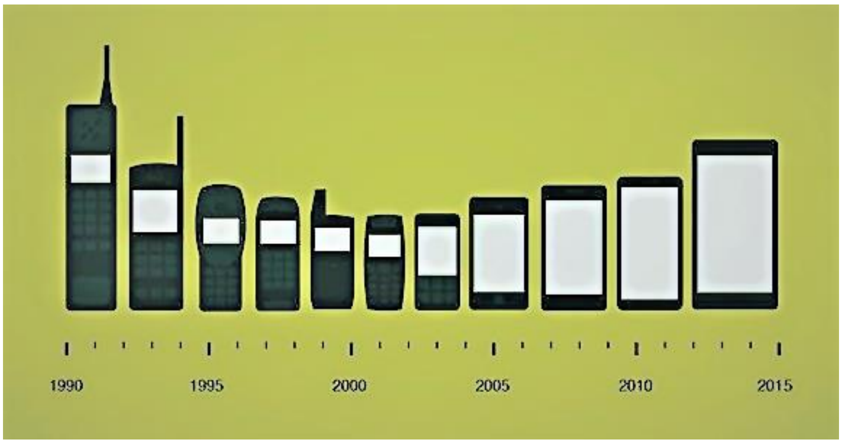

Figure 18: Evolution of mobile phone dimensions.

2.4.2. Energy efficiency metrics

In order to be possible to assess the required energy for providing modern high-speed services ubiquitously, allowing to evaluate and compare the potential EE solutions towards achieving further sustainable wireless broadband communications, it is key to consider established references and appropriate metrics for such comparisons. Those EE metrics can also be used in dynamic resource allocation decisions towards contributing to meet objectives regarding the respective energy necessary for transmitting a certain piece of information, or the associated normalized cost per transmitted bit.

In order to anticipate an effective EE metric, it is necessary to take into account the point-to-point transmission considering additive white Gaussian noise (AWGN) channels, in Shannon's capacity formula

$$R = B \log_2 \left(1 + \frac{P}{BN_0}\right), \tag{2.3}$$

where transmission rate R is achievable, considering transmit power P and system bandwidth B, while N_0 is AWGN's power spectral density (i.e., -174 dBm/Hz) [25][26].

The normalized energy cost to accomplish a task, in this case data transmission, as the ratio between the amount of transmitted data D while dispending energy E, per unit of time, is known by energy efficiency, which can be expressed as follows:

$$\eta_E = \frac{D}{E} = \frac{R}{P} = \frac{B \log_2\left(1+\frac{P}{BN_0}\right)}{P} \tag{2.4}$$

This expression provides a consensual EE metric, in bits/Joule, which in practical terms indicates how many bits of data can be transmitted, point-to-point, in a communication system with a particular configuration, i.e. B and P, while spending a unitary amount of energy. However, since this metric only applies when modelling point-to-point transmissions, it is not at all adequate when considering modern cellular systems. Thus, EE metrics for OFDMA-based communication systems are considered hereafter.

OFDM is currently accepted as one of the major technologies for delivering sustainable high performance in modern cellular networks. This happens, namely because it is possible to have a very flexible framework for RRM, where a DRA entity can allocate to each of the served UEs different portions of radio resources, or PRBs, both in frequency and time domains, if necessary, eventually with distinct associated transmit power as well, as considered here in previous sections [27]. The two main scenarios to be considered in OFDMA-based systems, and respective consolidated EE metrics, can be summarized in the following paragraphs.

Single cell scenario, as illustrated in Figure 19, where an OFDMA-based network BS provides service(s) to multiple UEs, at DL, serves here essentially to establish a sustainable reference without having to consider inter-cell interference, which ultimately typically occurs in effective deployed cellular wireless systems. The following will thus allow consolidating metrics on a link level perspective before addressing the respective full system level.

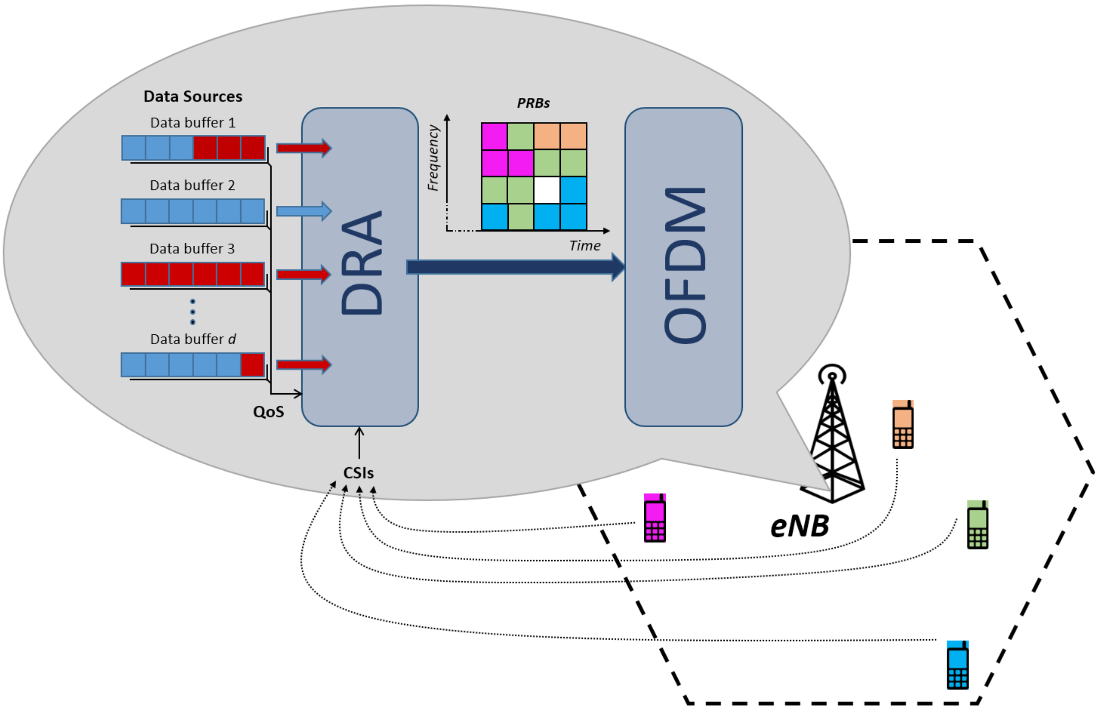

Figure 19: Typical DL transmission in OFDMA single cell, including DRA.

In the DL of a modern OFDMA-based cellular system, such as 3GPP LTE, the base station (eNB in LTE networks) placed at the centre of a cell, as illustrated in Figure 19 in the middle of a representative hexagonal coverage area, will serve multiple UEs by transmitting data utilizing a number of orthogonal subcarriers. Following 3GPP LTE standard and relevant literature, it is assumed that one subcarrier can only be assigned to no more than one UE in order to avoid interference among UEs, being the subcarrier frequency spacing sufficiently wide and inter-subcarrier interference can be ignored [28][29].

The eNB is assumed to have a single antenna, as well as all UEs, and each subcarrier associated to a particular UE is considered to be under frequency-selective fading, e.g. following ITU Pedestrian-B model. Channel information (CSI) from each UE is sent over a feedback channel to the eNB, which is thus assumed to have perfect channel knowledge for each UE at any time. Using such information, the eNB allocates a set of subcarriers to each UE and sets a number of bits in each subcarrier.

Considering that the total bandwidth B is uniformly divided into S subcarriers, having each, thus, a bandwidth $W = B/S$, using Shannon's formula it is possible to determine the maximum achievable data rate for UE u on subcarrier s, in the considered OFDMA-based system, as

$$r_{u,s} = W \log_2\left(1 + \frac{p_{u,s} \cdot g_{u,s}}{W N_0}\right), \quad u = 1, \dots, U; \;\; s = 1, \dots, S \qquad (2.5)$$

where $p_{u,s}$ and $g_{u,s}$ are, respectively, the transmit power and the channel power gain associated to the DL transmission from eNB to UE u, on subcarrier s, while N_0 is AWGN's power spectral density. Taking this into account, the system total DL throughput R, and the total transmit power P, for a single cell OFDMA system can be expressed as follows:

$$R = \sum_{u=1}^{U} \sum_{s=1}^{S} r_{u,s} \qquad (2.6)$$

$$P = \sum_{u=1}^{U} \sum_{s=1}^{S} p_{u,s} \qquad (2.7)$$

The energy efficiency, as the ratio between the amount of transmitted bits per energy unit, as expressed in Eq. 2.4, at the transmitter side, also taking into account the energy consumed in the electronic circuits of the latter, with power P_c, for transmission, always considering DL in OFDMA-based network, can be expressed as:

$$\eta_E = \frac{D}{E} = \frac{R}{P+P_c} = \frac{\sum_{u=1}^{U}\sum_{s=1}^{S} r_{u,s}}{\sum_{u=1}^{U}\sum_{s=1}^{S} p_{u,s}+P_c} \tag{2.8}$$

Multicell scenario, as effectively is the case in modern OFDMA-based networks, as represented in Figure 20, where interference from surrounding cells (ICI) and co-channel interference can have significant impact in DL, and EE mostly depends on accurate CSI at the transmitter (CSIT) and feedback on power consumption [30].

Similarly to the single cell scenario, the system total bandwidth *B* is uniformly divided into *S* subcarriers, or subchannels, with bandwidth *W* = *B*/*S* each, and all subcarriers have reuse factor *C* among the neighbouring cells. Each cell results from an eNB located at its centre, which serves multiple UEs scattered randomly throughout its coverage area, and has other cells surrounding it in multiple tiers.

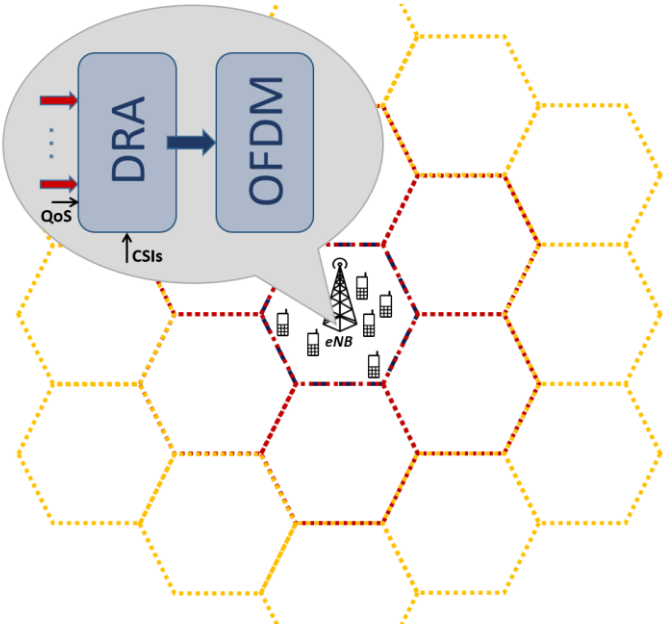

Figure 20: Typical DL transmission in OFDMA multicell scenario.

The typical propagation channel is frequency selective and slow time-varying, allowing that each UE being served in the cell is able to exchange CSI, respectively, with the eNB via a dedicated feedback channel.

The key difference when considering a multicell system, against a single cell scenario, is of course the interference that takes place in the former. To assess such interference it is possible to take into account one or more tiers of neighbouring cells where the same frequency sub-bands are reused according to the implemented frequency reuse model. For instance, it is possible to consider, as a starting point, a two-tier system consisting of six cells surrounding a central one, which is exposed to all transmitted signals, acting the surrounding six cells tier as interferers. In a three-tier system, with 19 cells in total, as represented in Figure 20, the outer twelve cells (tier three), are also considered to cause interference in the central cell (tier one).

Interference on the signal transmitted by the central cell eNB, inflicted by the surrounding eNBs, also known as co-channel interference, behaves typically as a flat fading signal. Therefore, defining signal-to-noise ratio (SNR), and signal-to-interference-plus-noise ratio (SINR), respectively as

$$SNR = \frac{p_{u,s} \cdot g_{u,s}}{WN_0} \quad ; \quad SINR = \frac{p_{u,s} \cdot g_{u,s}}{I + WN_0} \tag{2.9}$$

it is possible to replace SNR in Eq. 2.5 by SINR.

Taking into account the implemented subcarriers reuse factor, *C*, among cells, as indicated previously, as well as the considered number of UEs, *U*, and available subcarriers, *S*, in the DL, due to co-channel interference from surrounding cells, the SINR at the receiver of an UE *u*, on subcarrier *s*, from eNB *c*, can be expressed as

$$SINR = \frac{p_{u,s}^{c} \cdot g_{u,s}^{c}}{I + WN_0} \tag{2.10}$$

where $p^c_{u,s}$ and $g^c_{u,s}$ are, respectively, the transmit power and the channel power gain associated to DL transmission from eNB *c* to UE *u*, on subcarrier *s*, while N_0 is AWGN's power spectral density. The interfering signal power caused by the surrounding eNBs is designated by *I* in the above expression, which can be represented as:

$$I = \sum_{i=1, i \neq c}^{C} p_{u,s}^{i} \tag{2.11}$$

Expression 2.10 can thus be rewritten as:

$$SINR = \frac{p_{u,s}^{c} \cdot g_{u,s}^{c}}{\sum_{i=1, i \neq c}^{C} p_{u,s}^{i} + W N_0} \tag{2.12}$$

The maximum achievable data rate for UE *u* on subcarrier *s* from eNB *c*, in the considered multicell OFDMA-based system, can be expressed as follows:

$$r_{u,s}^{c} = W \log_2 \left(1 + \frac{p_{u,s}^{c} \cdot g_{u,s}^{c}}{\sum_{i=1, i \neq c}^{C} p_{u,s}^{i} + W N_0} \right) \tag{2.13}$$

It is possible, according to the previously introduced EE definitions, represented in Eqs. 2.6, 2.7 and 2.8, to finally define an adequate EE metric for modern multicell OFDMA-based systems, as follows, applicable to any considered eNB *c* in the system in question:

$$\eta_E = \frac{\sum_{u=1}^{U} \sum_{s=1}^{S} r_{u,s}^{c}}{\sum_{u=1}^{U} \sum_{s=1}^{S} p_{u,s}^{c} + P_C} \tag{2.14}$$

It is clear that the energy efficiency of a typical multicell OFDMA-based mobile system will suffer from the inherent interference eventually caused by eNBs neighbours to the cell where a UE is being served, as can be easily observed in above expressions, where ICI is appropriately taken into account. However, such EE limitation is the result of the trade-off towards attaining higher spectral efficiency (SE), which is possible by the typically imposed low frequency reuse factor, leading to ICI [28][31].

2.4.3. Energy efficient vertical handovers

As introduced here in sections 1.2.1 and 2.1, the vertical handover process is naturally context-based, and is generally considered to have three main stages, as illustrated in Figure 21. The first one, denominated *Network discovery and context gathering*, is immediately followed by the *VHO decision* stage, and the *Handover execution*, which concludes the VHO process.

In the initial stage, the *Network discovery* provides a candidate set of the available access networks, which are accessible to the UE being subject to the triggered VHO, and will be considered in the *VHO decision* stage [32]. In this same debuting stage, several context parameters under monitoring, e.g. UE's RSSI, SNR, BER or even geographical location, are collected and can trigger the subsequent *VHO decision* phase.

The second stage will use the gathered context information towards choosing a new RAN, more convenient essentially for the UE, based on the set of input parameters. Any VHO decision is based on pre-established directives towards the maximisation or minimisation of some particular parameter(s) regarding radio access, such as throughput, QoS and transmit power, or also eventually targeting the minimisation of costs to the user and/or to the operator [33]. Pre-determined user policies, such as restricted networks, cost threshold and user strategy (e.g. preference of energy efficiency over price), and operator policies, e.g. restricted networks, inter-operator HO limitations and load balancing, are also considered in the *VHO decision* phase [3].

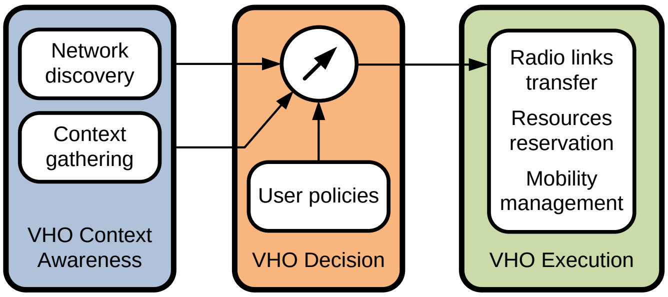

Figure 21: Main stages of VHO process.

In the last stage of the VHO process, the *Handover execution*, the UE connection and respective ongoing service should be transferred seamlessly to the new access network conveniently assigned, shifting the UE's point of attachment (PoA). This operation involves radio links transfer, resources reservation at the new access network, and also mobility management tasks, such as the maintenance of IP network addressing, between the two access networks, to provide the necessary respective service continuity [34][35].

Summarizing, distinct policies can thus be considered in the VHO process [33]. Those are essentially divided between schemes that are focused either on the UE or on the network, or eventually on both. These mechanisms can be layer-2 based or cross-layer, and target the VHO execution optimisation, but also the process to extract the parameters that actually trigger the VHO decision. The objective is to find the best cells, or RAT, which will provide connection to the UE while minimising energy consumption by unnecessary activities, such as scanning. The variety of input parameters reflects the differences between available heterogeneous accesses, and also user preferences, which are defined as context of the considered environment. Those parameters are used as the base for various analytical tools that are applied in obtaining VHO decisions.

The VHO decision (VHD) algorithms can be classified into different groups, differentiating the approaches in terms of the VHO problem analysis [32][35]: Objective function based (the most popular); Fuzzy logic based; Policy-driven; Pattern recognition based; and Markov Decision Process based. Furthermore, VHO decision making is considered to be a highly complex problem, as it can be optimised towards multiple distinct criteria, such as energy saving, financial cost reduction, QoS maximisation, or load balancing at the network side [33].

Energy efficient VHO policies focus on identifying the UE connection that has the highest EE, but also focus on limiting the energy consumption caused by the moving UE in the considered heterogeneous environment. Some of these mechanisms are only based on the UE, creating policies to set the scanning, while other schemes require the

UE to cooperate with the network, providing the latter with the necessary information in order to optimise its actions. It is necessary to establish threshold EE figures for the diverse RATs and respective RANs coexisting in the HetNet scenario where UEs are served, assuming they implement such RATs. For that, it is essential to assess each RAN/RAT under similar EE metrics in order to understand each's limitations. For instance, until when it is beneficial to serve a certain UE through a specific RAN and respective RAT, and not through another available RAN that could seamlessly accommodate such UE after respective VHO, if well justified and opportunely triggered. An illustration of a generic VHD algorithm is shown in Figure 22, where key parameters measurements are taken every T_{meas}, considering N coexisting RANs.

It is possible to find in the literature algorithms that target improving EE in various ways and considering different technologies. In [36], authors present several HO algorithms designed for small cells, but only the algorithm in [37] focus on energy, targeting to exploit the potential energy saving allowed by the low-power small cells operation. This algorithm considers power consumption, but not EE in LTE networks. In [38], a VHD algorithm considers Wi-Fi, HSDPA and WiMAX networks, and transmission power, bandwidth and the financial cost to make the decision are taken into account, while applying a scheme based on a cost function.

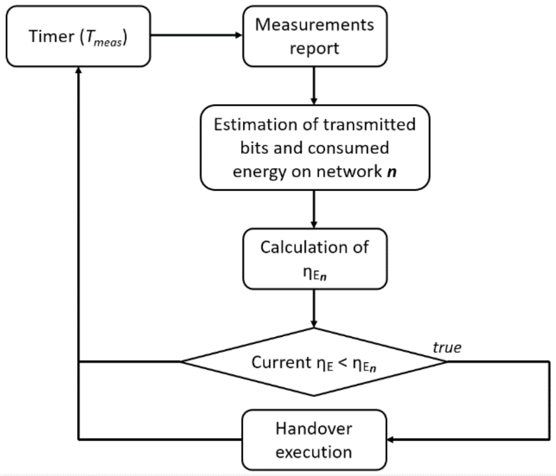

Figure 22: Generic VHD algorithm.

In [39], HO execution time is considered as one of the main aspects to make the right decision, since that is quite importance, namely for real-time services, but the different modulations used in the RATs are not taken into account. Two mechanisms that estimate the UEs' speed are presented in [40], where this aspect is exploited to avoid unnecessary HOs and the subsequent unnecessary energy consumption due to UE's velocity. Improving the scanning procedure in WLAN networks is considered in [41] as another way to improve EE even though it is not directly associated to the VHD, but since it immediately precedes the VHD procedure that can lead to EE improvement. Different scanning and VHD procedures are presented in [42], for every single RAN, taking into account their own characteristics, such as coverage and bit rate. WiMAX and WLAN are considered in [43], where a broadcasted message is used in WLAN to know the load of the RANs and improve the scanning procedure.

Although EE has been extensively applied to several VHO scenarios, as previously summarized here above, not much consideration has been given to practical energy efficient VHO based on available standards. It has been identified that the network scanning mechanisms in mobile UEs is one of the top energy spenders in such devices, and any intelligence introduced to the way the scanning process is managed could certainly lead to energy gains. On the other hand, previous works mainly focus on VHO algorithms striving for EE, although these have to coexist with HO execution protocols that trigger the HO process. Therefore, it is key that any concerted approach towards reducing HOs must explore the available standards, and consider ways on how to exploit available parametric information in order to trigger VHOs at the right occasion.

In this thesis, both the IEEE MIH 802.21 standard and the available context information are exploited to intelligently control the network scanning mechanism in mobile UEs. This approach will not only lead to significant energy saving, but is considered a practical design with real-life application, while being technology agnostic in nature. This will ultimately allow such saved energy to be used for other relevant features in the mobile UE, not necessarily related to communication, in advanced services and applications.

2.4.4. Energy efficient resource management

Mobile data traffic has grown exponentially in the past decades, right from the introduction of data-driven services for the mass market, over 2G networks, e.g. exploiting GPRS technology. Such demand for mobile data traffic is not slowing down and anywhere-anytime connectivity is practically taken for granted by the end users. Mobile network operators need thus to continue striving to achieve sustainable capacity improvement of their networks, as introduced here in sections 1.2.2, 2.2 and 2.3 as well, if possible without increasing operational costs, and preferably even reducing them.

In modern cellular networks, as introduced in 3GPP's latest revisions, the achievable network capacity is highly dependent on the channel quality. For instance, by using link adaptation (LA), the modulation and coding scheme (MCS) can be adapted according to the SINR, and thus considering a higher existing SINR, a higher order MCS can be used, resulting in higher achievable capacity, which can also lead to higher EE of the system. In order to guarantee high SINR, one typical approach in nowadays networks is to assure that the transmitter is closer to the receiver, i.e. by deploying small cells, and at the same time make sure to mitigate as much as possible the subsequent received interference added by the significant densification of the cellular network.

Such cellular densification, by increasing the number of BSs in the network and decreasing each one's coverage area, apart from the necessary investment in the network infrastructure, presents some straightforward benefits. First, the distance between BS and served UEs is reduced, leading to higher SINR, but also eventually allowing lower transmit power depending on the technology. Second, each UE being served shares the BS bandwidth, and respective backhaul connection, with a smaller number of UEs, gaining therefore access to a larger portion of resources, which results in additional capacity improvement. Of course, adding more macrocell BSs to the network, e.g. LTE eNBs, is not always viable due to the cost, or capital expenditure (CAPEX), or even the lack of available sites and the required connections among them.

Moreover, in a HetNet scenario, where BSs can have different transmit power and coverage areas, carrier frequencies, BH connection types, as well as various communication protocols, it should also be possible to exploit cognitive capabilities in some of the involved network entities. For instance, the knowledge on current network conditions, should allow for better planning, deciding and acting according to such conditions, always targeting to achieve the envisaged network performance and QoS.

Typically, there are three levels where resources can, and should, be managed, i.e. at link level, in system level and at the network level, all of which taking into account the appropriate dynamics at each's scope within the cellular environment. Thus, optimization of cellular networks can be done on those different levels and involving a large parameter space, e.g. from static ones during the network planning process, up to highly dynamic optimization at the frame level by exploiting advanced resource scheduling algorithms. An illustration of the schedulable physical resource blocks (PRBs) in LTE's frame structure is presented in Figure 23.

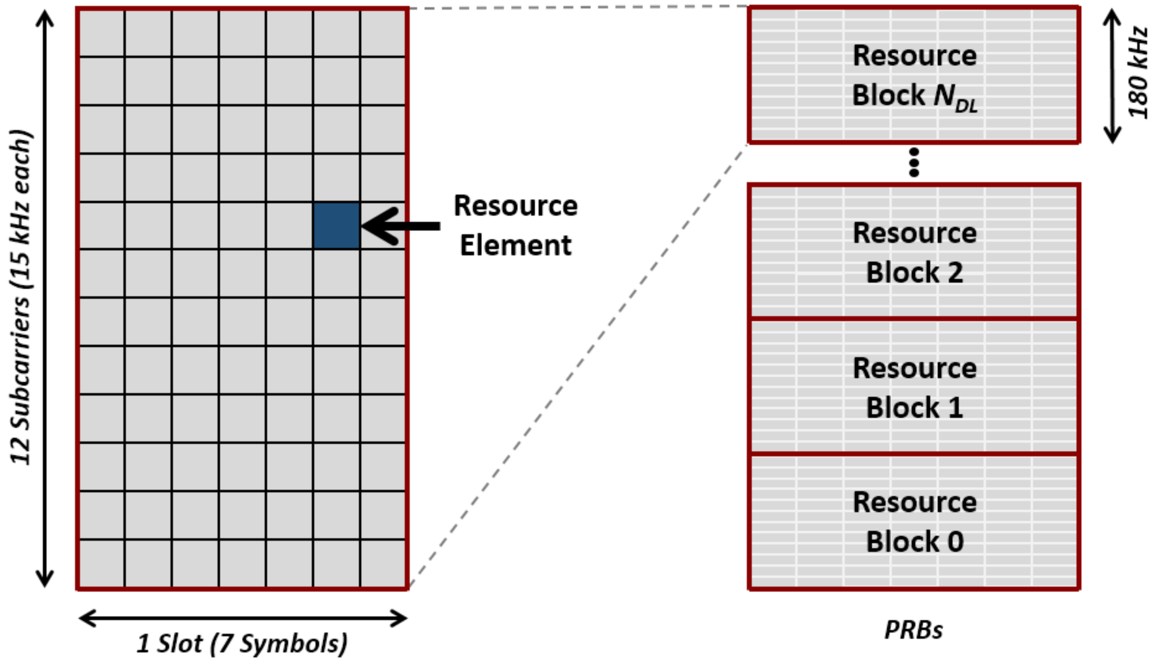

Figure 23: PRBs in LTE's frame structure.

In modern wireless access networks, the resources optimization can be done individually at each of these levels, in separated ways, or in common approaches instead, typically in cross-layer optimization, which can also address energy efficiency. For example, at the lower level, link adaptation is performed to adapt the transmission characteristics more suitably according to the channel conditions. Such adaptation, in early mobile systems generations, typically involved power control, while in further evolved systems usually deals with optimized MCS selection, e.g. also towards controlling the level of interference. Essentially, the channel quality is measured and a block of data is transmitted using the most suitable transmission mode that will both maximize the amount of data transmission, while minimizing potential errors that would originate subsequent retransmissions. The latter, from an EE perspective would also directly represent energy saving since no additional energy would have to be spent in powering such necessary retransmissions, leading to a decrease in system EE.

As introduced in section 2.2, there are several RRM approaches involving MAC layer protocols and dynamic resource allocation algorithms, such as scheduling and link adaptation, which can be used to improve EE in a modern cellular system. Cross-layer approaches are thus key for EE design in mobile cellular systems. For instance, in [44] particular focus is given on a system-based approach towards EE optimal transmission, and resource management (RM) is addressed considering not only time and frequency domains, but also involving the spatial domain.

Cross-layer solutions exploit the interactions between the different layers towards improving EE, but also in adapting conveniently to the service(s) in question, traffic and to the environment dynamics. Moreover, DRA policies are considered and extended to be EE-oriented as well. Essentially, EE resource management solutions should, on one hand, target to reduce the number of wasted transmissions corrupted by interference, i.e. by other BSs in the DL, or by other UEs in the UL, while on the other hand, should adopt efficient scheduling algorithms exploiting the variations across users’ needs to maximize overall EE on the network and UEs.

Summarizing, in practical terms, energy efficiency involving RRM and the MAC layer, is possible by exploiting the following approaches [44]:

Shutdown UE components when inactive, by enabling inactive periods through scheduling shutdown intervals according to buffer status, traffic requirements, and channel states, thus saving energy in such UE(s).

Assure both individual QoS and network fairness by reducing the number of wasted transmissions corrupted by interference, i.e. by other BSs (DL), or by other UEs (UL), while adopting efficient scheduling algorithms exploiting the variations across users' needs to maximize overall EE on the network and UEs.

Power management at the MAC layer to reduce the energy consumption at standby by developing tight coordination between UEs such that each one can wake up precisely when it needs to transmit or receive data, respectively.

In addition, cooperation-based approaches, based on coalitional game theory, using such tool to target increasing the overall energy saving, and EE, namely of the network and not only of UEs. Game theory is used to fairly distribute gains among "players", e.g. entities or components, hence motivating them to cooperate. A game allows also identifying selfish players and free riders, towards ultimately excluding them from the established cooperative process. This makes such an approach interesting for implementation is effective operational systems.

Existing technologies addressing energy efficient RM for cooperative and competitive HetNet systems considering various network goals and constraints are still not mature, and some key challenges still remain under investigation. The EE optimization for downlink OFDMA system in HetNet is one of such cases. In [45], the performance of typical HetNet regarding EE is introduced, although only for picocell. In addition, HetNet energy optimization considering multiple radio access technology is presented in [46], although multi-tier EE optimization is not considered. In [47], the design of EE mobile networks, by employing of BS sleep mode strategies, is investigated, as well as small cells, and the trade-off aspects associated with such approaches. Furthermore, optimization problems towards the minimization of energy consumption are formulated, and also is also determined the optimal operating frequency of the macro BSs. Energy efficient approaches to operate dense networks based on small cells, by applying concepts from cognitive radio (CR), are addressed in [48]. An energy optimized VHO mechanism for HetNet is proposed in [49], considering static circuit and transmit power. However, there are few studies analysing EE optimization considering also the dynamic aspect of circuits' power consumption in HetNet.

In current developments and standardisation towards cognitive 5G networks, high focus is being given to efficient and powerful RRM strategies based on sensed radio parameters and/or on accurately estimated ones [50]. Such strategies should exploit spectrum sensing, interference avoidance and efficient power allocation, as well as appropriate channel selection and bandwidth requirement. This will allow to increase network capacity, while satisfying QoS requirements, through SE maximization, advanced interference management, and EE maximization.

Dynamic RRM algorithms could be executed in 5G networks to optimise the use of radio resources, ultimately in real-time, meeting the required QoS. Such advanced RRM strategies can be implemented in centralised or distributed architectures. The latter is under standardisation in IEEE 802.16h, which proposes a coexistence protocol to enable all RRM related functions, such as detecting neighbourhood topology, registering to the defined database and negotiating for sharing RF spectrum. Joint dynamic RRM algorithm(s) can be deployed over the existing networks, including the cognitive mobile computing network, significantly improving network performance and end users' QoE.

3. ENERGY EFFICIENT VHO FOR HETNETS

Following the challenges and objectives of the work carried out towards this thesis, as summarized and introduced in sections 1.2 and 2.1, while further debated in subsection 2.4.3, this chapter focus on a specific EE solution in the scope of VHO for HetNets, exploiting context information, proposed in the ambit of this thesis [51][52][53].

The growing number of coexisting heterogeneous wireless access networks, paving the way for effective HetNet systems, together with the increasing service demands from the end users, require re-thinking of current access selection polices and appropriate management mechanisms, namely concerning EE. The IEEE 802.21 standard introduces link layer intelligence, as well as related network information to be shared with upper layers in order to optimise VHOs between networks of different technologies, e.g. 3GPP UMTS, IEEE Wi-Fi and WiMAX.

With the massification of mobile terminals (MT), or UEs, with multiple wireless interfaces, it becomes important to manage those RF interfaces efficiently, not only to appropriately provide the requested service(s) to the end user, but also to achieve that with higher EE. The latter contributes directly towards allowing higher users mobility by extending their UEs autonomy, or also for such spared energy to be used for other relevant MT features not directly related to communication, such as enhanced multimedia rendering in advanced mobile applications.

In the proposed solution, the implementation of the IEEE 802.21 standard is fundamental, particularly regarding the necessary signalling for pertinent HOs to occur, in the specific considered case, between IEEE WiMAX and Wi-Fi networks. An implementation of the considered scenario was made using *ns-2* network simulation framework, including the proposed EE scheme, which allowed, respectively, carrying out relevant tests and validating results.

3.1. Introduction

In response to the increasing demands from mobile data end users, always eager for more bandwidth and further mobility, an inevitable solution was introduced to allow users' mobile devices to use any compliant wireless network in their range, independently of its technology, in order to have proper coverage or better service. The IEEE 802.21 standard [15] represents a key enabler in order to allow handovers and interoperability between heterogeneous network types, including both IEEE 802 and non-802 networks, as introduced in subsection 2.1.3. This standard enables implementing the vision for ultimate wireless ubiquity, where users with their mobile terminals can globally roam freely by using distinct points of attachment, always with continuous service, in a seamless way. In this envisioned fully mobile environment, another key factor from the nomad user perspective is the EE of the established wireless connection, which typically is one of the most responsible functionalities for draining the MT's battery, with limited capacity.

Implementing the VHO concept and mechanisms in actual operational networks requires applying a significant amount of effort in different layers of the communication systems. To start with, in order for a VHO decision to happen, it is necessary to gather and evaluate diverse scenario/environmental context information, which must be continuously monitored by the involved networks, and only afterwards, that may eventually lead to a VHO execution [32]. When a VHO decision occurs, based on a set of input parameters, a new RAN is chosen targeting to maximise or minimise specific aspects of the radio access, such as improving the associated EE, as introduced in subsection 2.4.3.

The solution proposed in this work represents an attempt to address some aspects concerning EE in a HetNet scenario, which has been modelled in a simulation platform based on the *ns-2* network simulator [54], allowing to simulate multiple VHOs under an IEEE 802.21 implementation. The 802.21 functionality was incorporated in *ns-2* through add-on modules based in the 802.21 (draft 3) standard, developed by the National Institute of Standards and Technology (NIST).

In the following section, the implementation of IEEE 802.21 in *ns-2* using NIST's add-on modules is presented, as well as its key features. Section 3.3 introduces the overall description of the required signalling involved in VHOs between IEEE WiMAX and Wi-Fi networks, including the respective messages sequence diagram, as implemented in *ns-2*. A short revision of the IEEE 802.21 Media Independent Information Service (MIIS) is presented in section 3.4, together with the proposed approach for EE taking advantage from the functionalities offered by that service. Section 3.5 summarizes the scenario considered in this work for evaluating the proposed approach, and presents some numerical results attained after simulations performed using the *ns-2* simulation tool. The chapter ends with some conclusions and insights on future work.

3.2. VHO and IEEE 802.21 in *ns-2*

Taking into account the VHO concept and associated mechanisms introduced in subsections 1.2.1, 2.1 and 2.4.3, in this work, most of the development, implementation and testing was carried out exploiting the *ns-2* simulation framework. This framework was chosen as it provides support for heterogeneous networks and VHO scenarios by using NIST add-on modules developed for *ns-2* version 2.29 [55].

The modules contain numerous additional and adapted files over the standard release of *ns-2* in order to provide support for mobility scenarios [56][57][58][59][60][61]. The following subsections provide practical insights on IEEE 802.21 support in *ns-2*, which are particularly important in the carried out work.

3.2.1. IEEE 802.21 support in *ns-2*

The IEEE 802.21 NIST add-on module contains an implementation of the Media Independent Handover Function (MIHF) based on the IEEE 802.21 (draft 3) specification. An overview of the MIHF interactivity is presented in subsection 2.1.3, and illustrated in Figure 10 and Figure 11.

The MIHF and MIH User are implemented in *ns-2* as Agents. An Agent is a class defined in *ns-2* and extended in NIST's add-on modules. Such Agent class allows communication between lower and higher layers (e.g. MAC and MIH Users, respectively), providing the mapping between the media independent interface service access point (MIH_SAP) and the media dependent interface (MIH_LINK_SAP), including media specific primitives. In this way, the MIHF can send layer-3 packets to remote MIHF and MIH User can register with the MIHF to receive events from local and remote interfaces. The MIHF is also responsible for getting the list and status of local interfaces and control their behaviour. Essentially, MIH Users make use of the functionalities provided by the MIHF, typically by sending commands to the MIHF and receiving in turn events or messages from it in order to optimise the HO process.

3.2.2. IEEE 802.21 supported technologies in *ns-2*

The *ns-2* simulation framework supports the following technologies in IEEE 802.21 scenarios: WiMAX (IEEE 802.16), WLAN/Wi-Fi (802.11), Ethernet (802.3), and UMTS (3GPP). For the developed work presented here, only WiMAX and Wi-Fi components were required. The WiMAX module employed in this work was entirely developed by NIST [57]. However, one limitation of the considered WiMAX model is that, when using WiMAX cells, at the beginning of a simulation there must be at least one UE, or mobile node (MN) in the standard's nomenclature, in range within each cell in order for the WiMAX BS to function properly. This limitation was taken into account during the scenario setup prior to any relevant simulation. The NIST Wi-Fi *ns-2* module includes several changes in order to support basic IEEE 802.21 functionality [61].

As introduced in subsection 2.1.3, IEEE 802.21 standard defines three types of services, which are provided by the MIHF, i.e. Event, Command and Information services (ES, CS and IS). The latter, IS, is not supported in the considered version of IEEE 802.21 in *ns-2*.

Some relevant supported primitives of such services are summarized in Table 2.

Table 2: Supported primitives in NIST's WiMAX *ns-2* module.

MIH Commands	MIH Events
- Link Event Subscribe	- Link Up
- Link Event Unsubscribe	- Link Down
- Link Configure Threshold	- Link Going Down
- Link Get Parameters	- Link Detected
- MIH Get Status	- Link Parameters Report
- MIH Link Scan	- Link Handover Imminent
	- Link Handover Complete

Another supported set of functionalities is subnet discovery and change of address, which is key in HO execution. In the considered version of *ns-2*, it makes use of Router Advertisement (RAd), Router Solicitation (RS) and Neighbour Discovery (ND) messages. RAd messages are broadcasted periodically by APs or BSs to inform the MNs about the network prefix. Essentially, in *ns-2*, each Wi-Fi AP or WiMAX BS is on a different subnet, also known in *ns-2* as domain or cluster, and therefore will require a layer-3 HO, being the prefix the address of the AP or BS sending the RAd. When a MN is initiated or enters a cell, it uses RS messages to discover APs or BSs, which continues to happen as well after a HO, to discover new ones.

Also, for the identification of power boundaries in Wi-Fi and WiMAX cells, to be used in *ns-2* simulations, three variables associated to the respective RF interfaces are considered, and described as follows [61]:

CSTresh: is the minimum power level to sense wireless packets and switch the interface's MAC from idle to busy state;

RXTresh: is the minimum power level to receive wireless packets without error;

prlimit: implicitly defines the minimum power level that an interface senses before triggering a Link Going Down (LGD) event. This variable assumes a value which is always equal or superior to one, and is multiplied with *RXTresh* variable (*prlimit***RXTresh*) to actually define such minimum power level. The higher the *prlimit* coefficient, the sooner a LGD event is generated.

Finally, the considered version of *ns-2* supports energy modelling, to some extent, as it happens in the native version, which was enhanced in further released versions [62]. The *ns-2* energy model is based on an *EnergyModel* class, which provides mechanisms to describe the energy consumption of the RF interfaces in a MN, while in the different states in which they can be. Essentially, a MN is considered to have an energy budget, or an initial amount of stored energy (e.g. in a battery), which will be consumed in the network operations it performs with each of its wireless/RF interfaces. Each MN's interface is considered to be in one of the following states: *Transmit*, *Receive*, *Idle*, or *Sleep*. An interface can actually also be set to an *Off* state, but this implies a significantly longer transition time when such interface is requested to operate.

The *Transmit* and *Receive* states are self-explanatory. The *Idle* state stands for an interface that is powered-on and ready to receive messages from the physical layer, while the *Sleep* state represents an energy saving mode in which the interface is unreachable, but with the straightforward benefit of significantly involving lower energy consumption. An illustration of the interfaces states is presented in Figure 24.

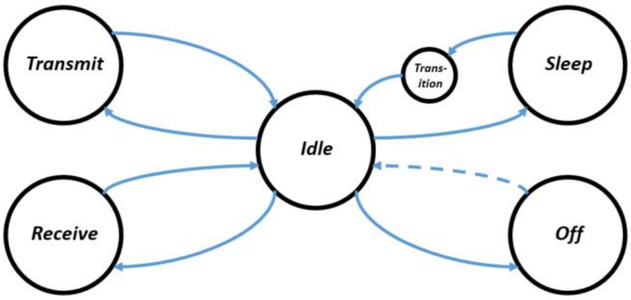

Figure 24: States of RF interfaces in *ns-2*.

This model also considers a *Transition* state, which occurs when the interface transitions from *Sleep* state to *Idle*. Each state is described also by its own power consumption, which mostly depends on the PHY layer.

In practical terms, when the MAC layer performs a send operation, the interface initially in *Idle* state is transitioned to the *Transmit* state. The time that is necessary to complete such operation is computed by *ns-2*, which multiplies it by the predefined interface power consumption (e.g. from vendors' data sheets) in order to determine how much energy is spent in such send operation. Afterwards, that estimated spent energy is subtracted to the MN's initial energy. This same process of energy expenditure estimation, and respective subtraction, is performed for every operation occurring in the MN's interfaces. Moreover, when a *Transition* state occurs, the transition power, set in this work to be 200 mW for any interface, is multiplied by the transition time, considered to be 5 ms, and such established energy spent in switching from *Sleep* to *Idle* is also subtracted from MN's energy budget [63]. Table 3 presents typical values of power consumption for different RF interfaces, and their distinct states, which are commonly integrated in modern MTs [64][65][66]. These values tend to become lower with the continuous technology evolution towards further EE components and systems.

Table 3: Power consumption for different RF interfaces.

Power (mW)	WiMAX	Wi-Fi (802.11b)	UMTS
Transmit	532	890	1100
Receive	510	690	1100
Idle	80	256	555

3.3. VHO between WiMAX and Wi-Fi

In order to be possible for clearly understanding what is exactly involved when a UE transitions between two RANs, in a VHO, this subsection presents a further detailed description of the respective sequential events occurring throughout such process. The description considers in this specific case, when a UE, or MN, being served in a WiMAX cell, migrates into a Wi-Fi hotspot, as well as, after continuing to move along its path, it reconnects back again to the WiMAX BS in the earlier RAN.

3.3.1. VHO triggering process

In Figure 25 it is possible to observe the power received at each of the two RF interfaces in the MN as it moves closer to the Wi-Fi AP, as well as when it moves away from it afterwards. Essentially, the MN was served by the WiMAX cell where it stood still for the first 10 seconds, after which it started moving towards the BS, and along its path, it detected a Wi-Fi network.

Once the Wi-Fi minimum received power is reached, just before *t* = 22 seconds, a VHO is triggered and executed since the considered HO algorithm is set to assume Wi-Fi to be a better network when compared against WiMAX. The MN operates in the Wi-Fi network for roughly 7 s, until the received power from the AP is considered to be insufficient to operate and remain connected to the Wi-Fi AP and respective network.

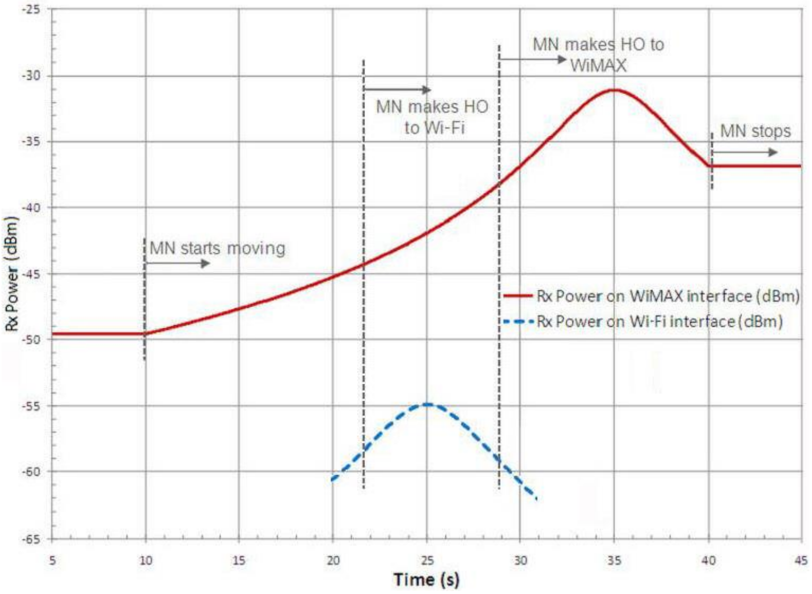

Figure 25: Received power at MN’s interfaces (WiMAX and Wi-Fi).

Since the signal from the earlier WiMAX BS is still available and powerful enough for the MN to operate, a new VHO is triggered and executed, this time connecting the MN back to the WiMAX RAN. The MN is as close to the WiMAX BS as it will ever get at instant *t* = 35 s, when the received power at MN's respective RF interface is at its highest value, around -31 dBm. The MN continued to follow its path until it stops at *t* = 40 s, continuing to be served by the WiMAX BS and respective network. It is worth highlighting that, in an energy wise perspective, in this particular case the MN's Wi-Fi interface is powered for more than 35 seconds, or 78% of the considered time interval (45 seconds), without being actually used although still consuming a significant amount of energy, even though operating at its *idle* state.

3.3.2. VHO execution and exchanged messages

A detailed representation of the sequence of events occurring during the considered VHOs is represented in Figure 26, already also reflecting the IEEE 802.21 NIST implementation in *ns-2*. The detailed description of the represented events and exchanged messages between the involved entities is as follows:

1. MN's MIH User sends a *MIH Get Status* request to MN's MIHF, which replies with a *MIH Get Status* response, informing that interfaces type 19 (Wi-Fi) and type 27 (WiMAX) are available, supporting MIH commands and events;
2. MN's MIH User sends a *MIH Event Subscribe* request to MN's MIHF for subscribing events on both MN's interfaces, and receives back confirmation;
3. MN's WiMAX interface receives a Download Context Descriptor (DCD) and an Upload Context Descriptor (UCD) from the BS and triggers a *Link Detected* event towards MN's MIHF, which forwards it to the MN's MIH User;
4. MN's MIH User receives the *Link Detected* and commands (via MN's MIHF) its WiMAX interface to connect to BS, since that is its only interface detecting a PoA;
5. MN's WiMAX interface connects to the BS and triggers a *Link Up* event towards MN's MIHF, which forwards it to the MN's MIH User, initiating an *RS* command from MN's ND/MIPv6 agent (via MN's WiMAX interface);

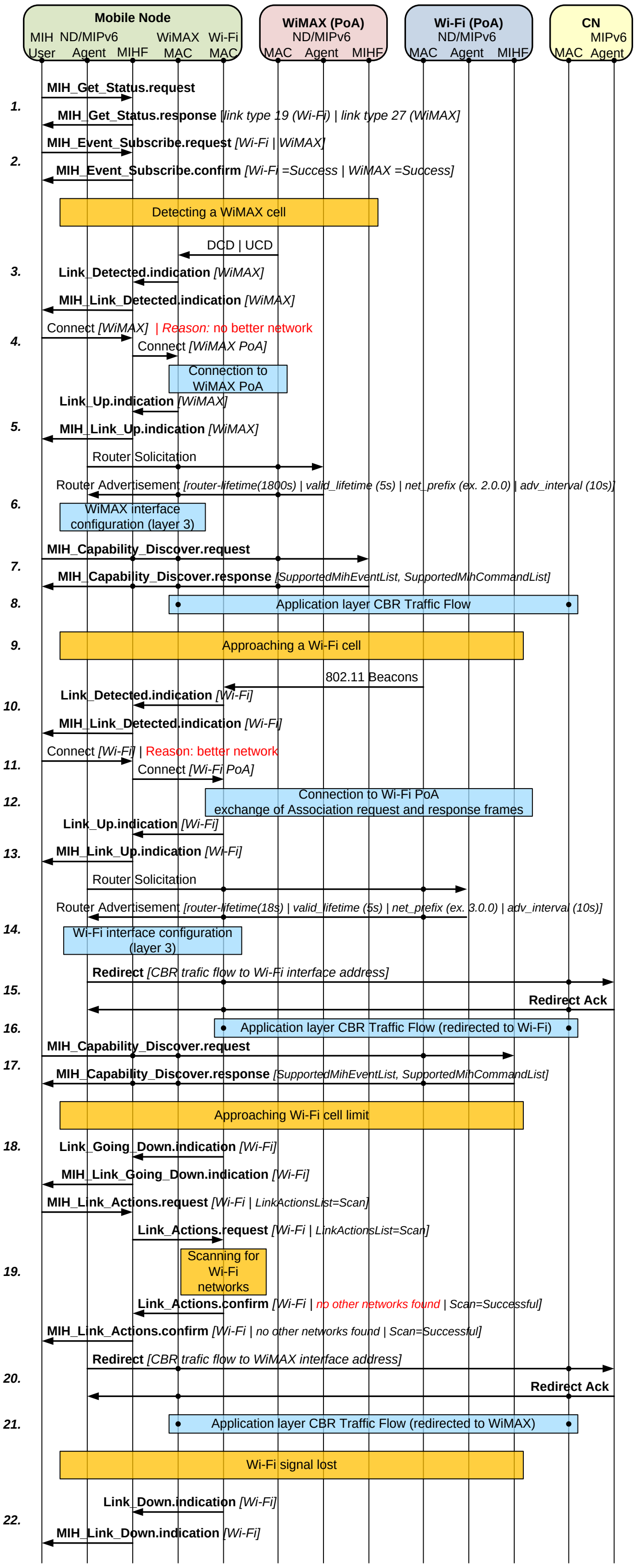

Figure 26: Sequence of events in WiMAX–Wi-Fi–WiMAX VHO, in *ns-2*.

6. WiMAX BS receives the *RS*, detects that the MN is a new neighbour, and sends it an *RA* including the necessary information for the MN's WiMAX interface to connect, using that info to reconfigure its interface address according to the received prefix (e.g. 2.0.1);
7. MN's MIH User is notified and commands the WiMAX interface to send (via MIHF) an MIH *Capability Discover* request to the BS, which sends back an MIH *Capability Discover* response including its MIHF identification, for the MN's MIH User knowledge of its PoA;
8. At t = 5 secs., the Core Network (CN) starts to send constant bit rate (CBR) traffic to the MN through WiMAX, which is received by the MN's WiMAX interface;
9. At t = 10 secs., the MN starts moving towards the BS, with a Wi-Fi cell in its path;
10. At t ≈ 22 secs., the MN's Wi-Fi interface detects 802.11 beacons and triggers a *Link Detected* event towards MN's MIHF, which forwards it to the MN's MIH User, when the received power of the beacon frames is above threshold value;
11. MN's MIH User receives the *Link Detected* and commands (via MN's MIHF) its Wi-Fi interface to connect to the AP, since that is considered to be a better PoA;
12. MN's Wi-Fi interface and the AP exchange *Association Request* and *Response* frames in order to connect the MN to the AP and its Wi-Fi cell;
13. MN's Wi-Fi interface receives the *Association Response*, and triggers a *Link Up* event towards MN's MIHF, which forwards it to the MN's MIH User, initiating an *RS* command from MN's ND/MIPv6 agent (via MN's Wi-Fi interface);
14. AP receives the *RS*, detects that the MN is a new neighbour, and sends it an *RA* including necessary information for the MN's Wi-Fi interface to connect, using it to reconfigure its interface address according to the received prefix (e.g. 3.0.1);
15. MN's MIH User is notified, and MN's MIPv6 Agent commands the Wi-Fi interface to send a *Redirect* message to the CN in order to inform it of MN's new location, to which the CN's MIPv6 Agent replies back by sending an *Ack* message to the MN's Wi-Fi interface, which then notifies MN's MIH User (this behaviour is

considered a *make-before-break* HO, meaning that the MN will use both interfaces at the same time in order to perform a seamless HO);

16. MN's MIH User has the confirmation that the CN has been notified of the MN's new address and redirects the reception of CBR traffic from the WiMAX interface (2.0.1) to Wi-Fi interface (3.0.1), and traffic starts then being received through the link Wi-Fi interface – AP;
17. MN's MIH User commands the Wi-Fi interface to send (via MIHF) a MIH *Capability Discover* request to the AP, which sends back a MIH *Capability Discover* response including its MIHF identification, for the MN's MIH User knowledge of its new PoA;
18. At t ≈ 28 secs., the MN approaches the Wi-Fi cell boundary and Wi-Fi interface triggers a *Link Going Down* event (based on received power of beacon frames);
19. MN's MIH User commands the Wi-Fi interface to execute a *Link Scan* in order to search for other nearby Wi-Fi networks, for which the MN's Wi-Fi interface sends a *Probe Request* in each defined IEEE 802.11 channel frequencies in order to execute the scan (a *Probe Response* is received only in channel 2, where the MN currently is, and based on that the MN's MIH User is notified that there is no other available Wi-Fi network);
20. Because of MN's speed, the probability of the Wi-Fi link to go down increases, when it reaches a specified value, and also since the MN's WiMAX interface is still active, MN's MIPv6 Agent commands the WiMAX interface to send *Redirect* message to CN in order to inform it of the MN's new location (2.0.1), back to WiMAX network, to which the CN sends back a *Ack* message through its MIPv6 Agent, being the MN's MIH User notified;
21. At t ≈ 29 secs., MN's MIH Agent has the confirmation that the CN has been notified of the MN's new address and redirects the reception of CBR traffic from the Wi-Fi interface (3.0.1) to the WiMAX interface (2.0.1), and traffic starts then again being received through the link WiMAX interface – BS;
22. MN's Wi-Fi interface triggers a *Link Down* event and disconnects from Wi-Fi AP.

3.4. MIIS and Energy Efficiency

The Media Independent Information Service (MIIS) specified in IEEE 802.21 standard [15], provides a framework by which an MIHF implemented in a MN or in a network, can discover and obtain network information within a geographical area to assist in MN's network selection and HOs. This service can acquire and provide a global view of all the networks of different types that are relevant to a MN in the region where it is currently located, facilitating roaming across these networks through seamless HOs.

3.4.1. MIIS functionalities

The MIIS considers various Information Elements (IEs), which it supports, and that provide essential information to network selection algorithms towards making intelligent HO decisions. Depending on the type of mobility of MNs, the support for different types of IE is required for performing HOs. MIIS provides the capability for obtaining information associated to lower layers, such as neighbour maps and other link-layer parameters, as well as information about available higher layer services, such as Internet connectivity. The most relevant available IEs, in the scope of the presented work, are presented in Table 4.

This service provides a generic mechanism to allow an operator's network and a mobile user's UE to exchange information regarding different HO candidate access networks. The HO candidate information includes many aspects on different RATs, such as IEEE 802 networks (e.g. Wi-Fi, WiMAX, UWB), 3GPP networks (e.g. UMTS, LTE) and 3GPP2 networks (e.g. CDMA2000).

The MIIS also allows this collective information to be accessed from any single RAN. For instance, by using a Wi-Fi RAN the MN's Wi-Fi interface can get information, not only about all other 802.x based networks in a specific geographical region, but also about 3GPP or 3GPP2 networks in that region as well. Thus, it allows a MN to use its currently active RAN connection to inquire about other available access networks in a region.

Table 4: Relevant available IEs in MIIS.

Name	Description
- IE_NETWORK_TYPE	- Link types of the RANs available in a given region
- IE_COST	- Indication of cost for service or network usage
- IE_NETWORK_QOS	- QoS characteristics of the link layer
- IE_NETWORK_DATA_RATE	- The max. data rate value supported by RAN's link layer
- IE_POA_LINK_ADDR	- Link-layer address of PoA.
- IE_POA_LOCATION	- Georeferenced location (e.g. coordinates, cell ID) of PoA
- IE_POA_IP_ADDR	- IP Address of PoA

The MIIS enables such functionality across all available access networks by providing a uniform way to retrieve heterogeneous network information in any geographical region. Thus, this service allows the MNs and network entities to discover relevant information that can adequately influence the selection of the most appropriate networks in HO processes.

Such relevant information is intended to be mainly used by a policy engine entity, as included in Figure 21, which can make efficient HO decisions based on such MIIS output. Some of the key motivations for an effective MIIS implementation are as follows:

a) Provide information about the available access networks in a geographical area;
b) Provide static link-layer information parameters (e.g. security and QoS support), assisting MN's in access network selection;
c) Provide information about capabilities of different PoAs in neighbour reports to help configure MN's RF interfaces for connecting to available or selected RANs;
d) Provide an indication of higher layer services supported by different access networks and CNs (e.g. public or enterprise) that can be relevant in HO decisions.

The IEEE 802.21 Information Service is supported by every MIHF that implements the MIIS to support flexible and efficient information queries. The MIIS defines two methods for representing IEs, which are binary representation and Resource Description Framework (RDF) representation [67]. It also defines two query methods [15]. A RDF representation schema [68] is used in the IEEE 802.21 Information Service to define the structure of each IE, as well as the relationship among the existing IEs.

3.4.2. Proposed EE solution exploiting MIIS

In the energy saving approach proposed in the scope of this work, the MN exploits the MIIS functionality through its currently active interface to decide which of its RF interfaces it should keep powered-on or off, without having to establish a specific network connection for the sole purpose of retrieving information on available RANs. In this way, it is possible for the mobile UE to save valuable energy by deactivating its RF interfaces whenever those are foreseen to be useless based on BSs and APs geolocation information provided by the MIIS, and according to MN’s current location. The information on existing RANs could be obtained by the MN using any RAT, whether by means of request/response signalling, or by means of information that is specifically or implicitly broadcasted over those cellular networks. Instead, that information could also be gathered and maintained in an internal database at the MN.

In a mobile UE with multiple RF interfaces, the fact that all RAT interfaces could be consuming power all the time represents a clear waste of energy, reflected in precious UE battery draining, even if those interfaces are in idle state since the UE is still powering such interfaces that are kept up waiting to receive data.

What is proposed in the scope of this work is to reduce to a minimum such energy depletion at UE’s RF interfaces, by selectively turning them off or putting them into a condition of very low energy consumption, e.g. sleep state. This selective deactivation and activation of the UE’s RAT interfaces is done here according to context-aware information made available by the network, through MIIS functionalities.

The required information is essentially the georeferenced position of RANs’ BSs and APs. This information allows the UE to decide, according to its own estimated position (e.g. from inbuilt GPS and/or provided by the network) and a selection algorithm, which of its RAT interfaces should be active, or not, potentially allowing VHOs if effectively justified. Significant amounts of energy could thus be saved through the implementation of this solution, allowing extra mobility to the users of such UEs.

Basic primitives were implemented in the scope of this work, specifically for the proposed energy saving approach, namely for requesting geolocation information of BSs and APs to the core network, in an *ns-2* based simulation platform implementing IEEE 802.21 with NIST's add-on modules. Also, was developed a module to drive the state of an RF interface, in order to be possible to switch it on and off based on context information regarding the geolocation of the BSs, APs, and UEs, as proposed.

Taking into consideration the portrayed VHO between WiMAX and Wi-Fi in section 3.3, and the respective sequence of events represented in Figure 26, the necessary implemented additional or adapted exchanges to support the proposed EE approach taking advantage of MIIS functionality are represented in Figure 27.

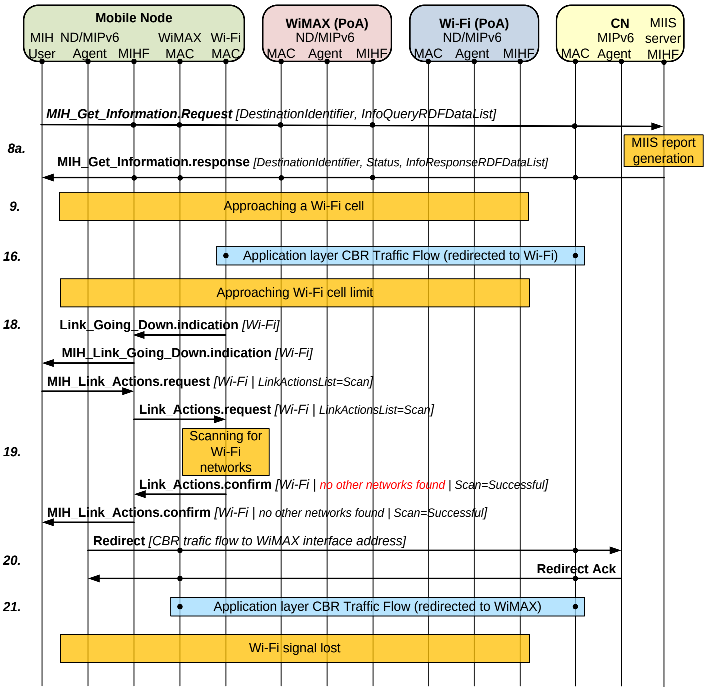

Figure 27: MIIS adapted exchanges in WiMAX–Wi-Fi–WiMAX VHO, in *ns-2*.

In this proposed approach targeting energy saving at the UE, taking Figure 26 as reference, after achieving step 8, once the MN is connected to the WiMAX network, since no other network is detected, and traffic flow is established, the MN commands its Wi-Fi RF interface to go into sleep state. Once traffic flow is established between MN and the WiMAX network (step 8), upon MN's MIH User request the MIIS client within the MN's MIHF communicates with the WiMAX BS's MIHF by sending a *MIH Get Information* request, which is carried over the appropriate transport (layer-2 or layer-3). The WiMAX BS's MIHF forwards this request to the CN's MIHF, where the MIIS server resides, which after generating a report based on the received query, returns the requested information, via WiMAX's MIHF, to the MN's MIH User through the appropriate *MIH Get Information* response frame. This sequence of events is represented in Figure 27 as step 8a.

The implemented query requests the MIIS server for a list of available networks and respective PoAs in the surroundings of the WiMAX BS where the MN is connected, more specifically such PoA's link types (*IE_NETWORK_TYPE*), PoAs' link-layer addresses (*IE_POA_LINK_ADDR*), and PoAs' geographical coordinates (*IE_POA_LOCATION*). With that received information, the mobile UE can determine according to its own estimated geolocation (e.g. provided by GPS and/or the network), speed and direction, when it should power down or wake up any of its RF interfaces.

In the considered WiMAX–Wi-Fi–WiMAX VHO case, such above mentioned assessment happens at $t \approx 20$ s, when the UE determines that, according to its own position mapped onto the georeferenced information provided by the MIIS on existing PoAs and the estimated power received from them, it should wake up its Wi-Fi interface. This happens just in time to proceed with steps 9-18. After traffic flow is successfully redirected to Wi-Fi network (step 16), the UE can put its WiMAX interface to sleep, for approximately 5 seconds, after which it is waken up in order to assure that WiMAX connectivity is re-established (steps 20-21). At t ≈ 30 s, the UE can put its Wi-Fi interface back to sleep state, until further notice.

After receiving the Wi-Fi *Link Going Down* indication (step 18), although the UE kept in its internal database the list of nearby available networks and respective PoAs supplied by the MIIS server upon request, the Wi-Fi *Link Scan* procedure (step 19) is nonetheless commanded by the MN for searching other nearby Wi-Fi networks. This could prevent damage from a MIIS server incomplete database, and could also eventually allow MNs to help maintain the MIIS server database up-to-date, by using additional appropriate information exchange. However, such additional aspect was not considered to be in the scope of this work and therefore was not implemented. It is also worth highlighting that the periodicity of information exchange between MN and MIIS server should be enough to assure full knowledge of all the available networks in the MN's path, and it should depend mainly on the MN's velocity.

In the proposed EE approach, it was assumed a Wi-Fi AP coverage range of 100 m. The following UE's RAT interfaces deactivation and (re)activation rules were implemented:

a) Wi-Fi deactivation when UE is more than 20 m away from any AP coverage area;
b) Wi-Fi (re)activation when UE is less than 20 m away from any AP coverage area;
c) WiMAX deactivation when UE is less than 80 m away from any AP;
d) WiMAX (re)activation when UE is more than 80 m away from any AP.

A UE's RAT (re)activation occurs just before such interface is considered to be useful according to the virtual map of PoAs, and corresponding estimated received powers, as stored in the respective UE's internal database, built on the provided MIIS information. The considered distances threshold values can be adjusted for a more, or less, conservative approach towards addressing potential VHOs.

3.5. Simulation Results

The proposed approach for taking advantage of MIIS, targeting EE at the UEs by selectively switching their RAT interfaces on and off, was implemented on an *ns-2* based simulation platform including NIST's IEEE 802.21 modules, previously evaluated in [69].

The MIIS primitives presented in subsection 3.4.2 were implemented over the MIHF functionality in the considered *ns-2* simulation platform. The main objective of this work was to evaluate the proposed energy saving approach through a set of pondered simulations, considering a mobile wireless HetNet scenario potentiating VHOs, using the prepared *ns-2* platform.

3.5.1. Simulation system model and setup

The system model represented in Figure 28 was implemented in the customised *ns-2* simulator. Essentially, it consists in a network topology composed by a single WiMAX BS, three Wi-Fi APs and 25 UEs placed randomly, following a uniform distribution, inside the WiMAX cell. The WiMAX BS has a coverage range of 500 meters, and each AP provides coverage in 100 meters radius.

The WiMAX BS and each Wi-Fi AP have dedicated (wired) links to the CN through a router, and any of those links supports 1 Gbps bit rate. All mobile UEs, or MNs, have WiMAX and Wi-Fi RAT interfaces, as well as an inbuilt GPS module.

The MNs start communicating inside the WiMAX cell, where they initially remain static for a short period. After that, all MNs start moving with constant speed in a random,

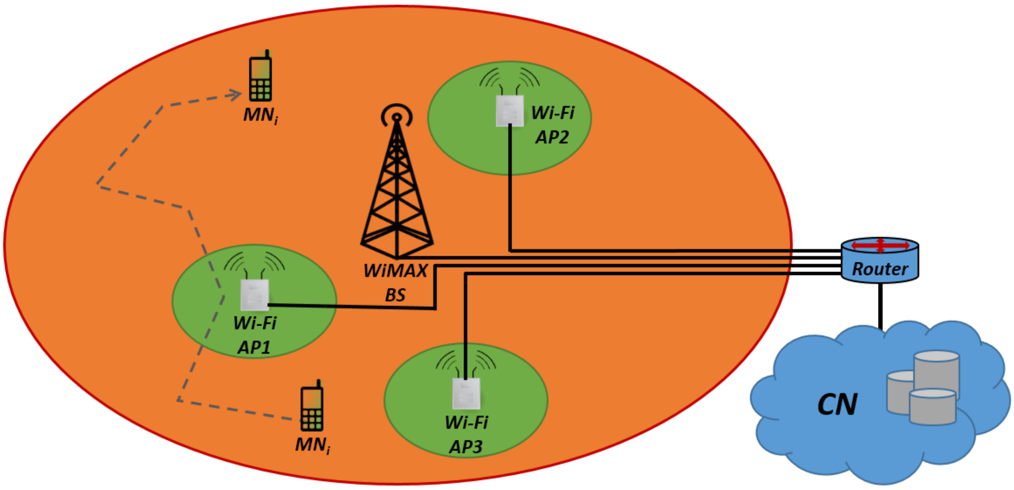

Figure 28: Simulation scenario for evaluating EE solution exploiting MIIS.

but steady, direction, which changes randomly after each simulated minute, following the random waypoint model [70], as illustrated with a dashed arrow in Figure 28, portraying the movement of one MN. Thus, considering this setup, each simulated MN should most likely enter some Wi-Fi AP(s) coverage area(s) throughout its path during the 10 minutes considered in each simulation.

The most relevant parameters considered in this simulation setup are summarised in Table 5. The two-ray ground propagation model, although quite simplistic taking into account the dynamic scenario, was nonetheless considered for the simulations since it was assumed that typical channel phenomena are sufficiently compensated by the VHO algorithm hysteresis and respective threshold.

Table 5: Simulation parameters.

Parameter	Value
- WiMAX cell radius	- 500 m
- WiMAX modulation	- 16QAM (10 Mbps)
- WiMAX BS transmit power	- 15 W (42 dBm) @ 3.5 GHz
- WiMAX *RXThresh*	- 1.215 nW (-59 dBm)
- WiMAX *CSThresh*	- 80% of RXThresh
- Wi-Fi coverage range	- 100 m
- Wi-Fi standard	- IEEE 802.11b (11 Mbps)
- Wi-Fi AP transmit power	- 100 mW (20 dBm) @ 2.417 GHz
- Wi-Fi *RXThresh*	- 0,989 nW (-60 dBm)
- Wi-Fi *CSThresh*	- 90% of RXThresh
- Wi-Fi *prlimit*	- 1.2
- Propagation channel model	- Two-ray ground
- Traffic bit rate	- 50 kbps (CBR)
- Traffic type	- UDP, full-duplex
- MNs velocity	- 1, 2, 4, 8 m/s
- MNs interfaces power consumption	- (*see Table 3*)
- GPS chipset power consumption	- 45 mW [71]

In every simulation run, it was considered a period of 600 seconds, during which MNs move with a specific constant speed. The considered traffic was UDP, full duplex, where each MN receives and transmits at a certain bit rate. The MNs velocities considered in different simulation runs, as well as the considered traffic bit rate, are also indicated in Table 5. In the proposed EE solution, the GPS chipset, as a MN component, is always active, thus consuming energy constantly at the indicated power consumption, in order that the MN's position estimation is updated at every elapsed second for subsequent decision on changing the states of its RAT interfaces, respectively. In addition, such GPS energy expenditure is continuously subtracted to the MN's stored energy. Moreover, power control was not considered in any of the two RATs integrated in each MN.

3.5.2. Numerical results

The first set of results, after 10 independent runs of a complete simulation considering a period of 600 seconds, indicates that during such period the total energy consumed by a MN implementing the proposed EE approach is significantly lower than if all RAT interfaces in the MN were consuming energy all the time, which is normally the case. Such outcome derives from the fact that the GPS chipset typically consumes much less energy than any of the RAT interfaces, which commonly need to be at least in idle state.

These results are presented in Figure 29, showing the average and standard deviation of the total energy consumed per MN, considering distinct MNs velocities. The red bars represent the results considering the typical implementation where both RAT interfaces are always active in the MN, and the green ones represent the results obtained with the proposed EE solution. The presented results include the 95% confidence intervals superposed on the respective chart bars, which are minimal in the traditional solution. In the latter, such consistent total consumed energy, independently of the considered MNs velocity, or path with eventual VHOs is clear, and occurs since transmit power control was not implemented in this work, and also obviously because MNs keep both interfaces at least in idle state, and are all transmitting or receiving at the same bit rate.

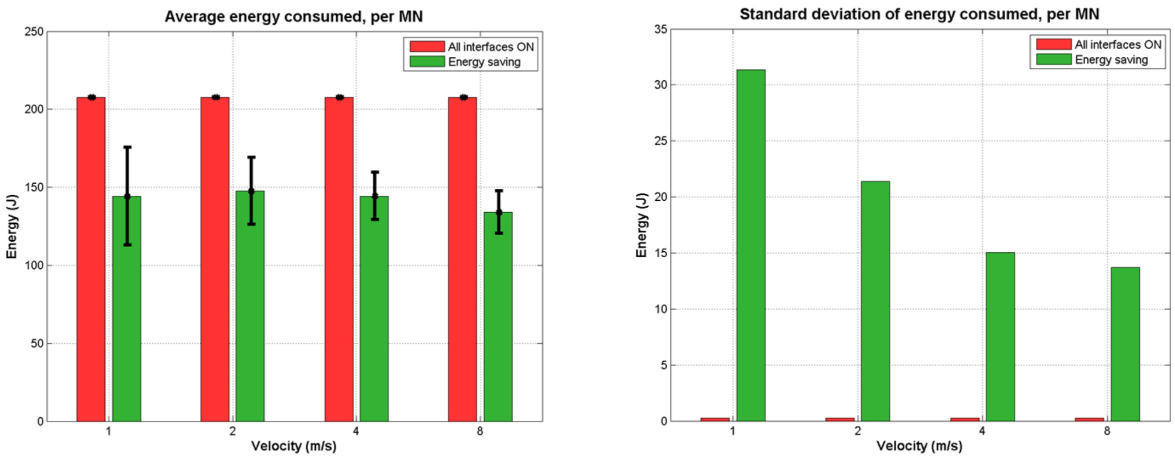

Figure 29: Avg. and Std. dev. of consumed energy per MN, for different velocities.

On the other hand, the MNs using the proposed energy saving approach present diverse total energy consumption, which reflects each one's path with associated VHOs, and consequent deactivation of the alternative RAT interface until it is foreseen to be needed. Such diverse results are visible particularly in the representation of the standard deviation, in Figure 29 on the right side.

The reactivation of the Wi-Fi interface or the WiMAX interface, to idle state, is done according to the rules presented at the end of subsection 3.4.2, before the one that is deactivated is foreseen to be useful, according to the virtual map of PoAs stored in each MN internal database, built after the exchanged information with the MIIS server.

Although it is not shown in the simulations results, it is clear that turning off the Wi-Fi interface is much more beneficial for energy saving, than powering off the WiMAX interface, since this interface consumes less energy in idle state, as shown in Table 3.

Once MNs are moving faster, the probability that a MN enters a different RAN coverage area is higher and also is the possibility of occurrence of respective VHOs. It is noticeable that the proposed EE approach appears on average to be slightly further efficient when MNs are moving faster since they tend to be subject to more VHOs, and therefore would, on average, benefit further from the proposed EE solution. It is also visible that the benefits of this solution are more consistent throughout the different MNs as their velocity increases, since any MN would be involved in a similar amount of VHOs.

The average number of executed VHOs swapping a MN from one RAN to the other, during the considered simulations, together with the average amount of "smart" RAT interface activation/deactivation at the MN, either Wi-Fi or WiMAX, in presented in Figure 30. The results in this chart confirm what is expected in the considered scenario when MNs move faster, which on average effectively lead to a higher occurrence of VHOs. In addition, it is visible that, when using the proposed EE solution, more RAT interfaces switching (*on* and *off*) happens when MNs move faster, always targeting to optimise energy spending at the MNs' RF interfaces.

The average total time spent by the simulated MNs on each RAN is presented in Figure 31, together with the average total accumulated time during which each of the RAT interfaces was activated on the respective MN. The total time that an MN is connected to a Wi-Fi AP decreases with its velocity in the considered scenario setup, as expected, since the smaller AP coverage area can be crossed in a shorter period of time. In Figure 31, it is also visible the reduction in the accumulated time that the Wi-Fi interface is activated as the MN moves faster.

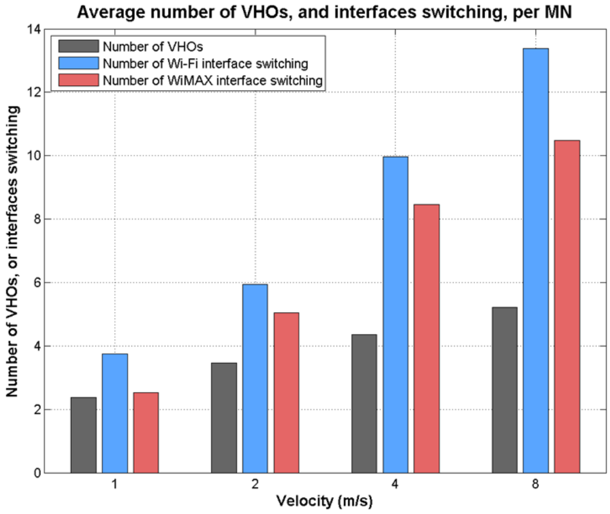

Figure 30: Avg. number of VHOs and RF interfaces switching, for different velocities.

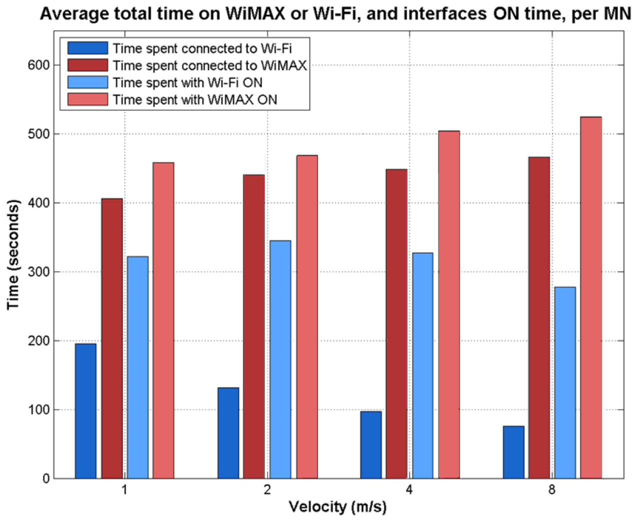

Figure 31: Avg. total time in each RAN, and total interfaces active time, by velocity.

This latter fact justifies the higher total energy saving when MNs move faster, as shown in Figure 29, as the considered Wi-Fi interface consumes significantly more energy than the WiMAX interface, namely when comparing at idle state.

Finally, from the carried out simulations, the total packet loss occurred in each simulation run, concerning each involved MN, was also accounted for, and the respective average and standard deviation results are presented in Figure 32 for each of the considered MNs velocities.

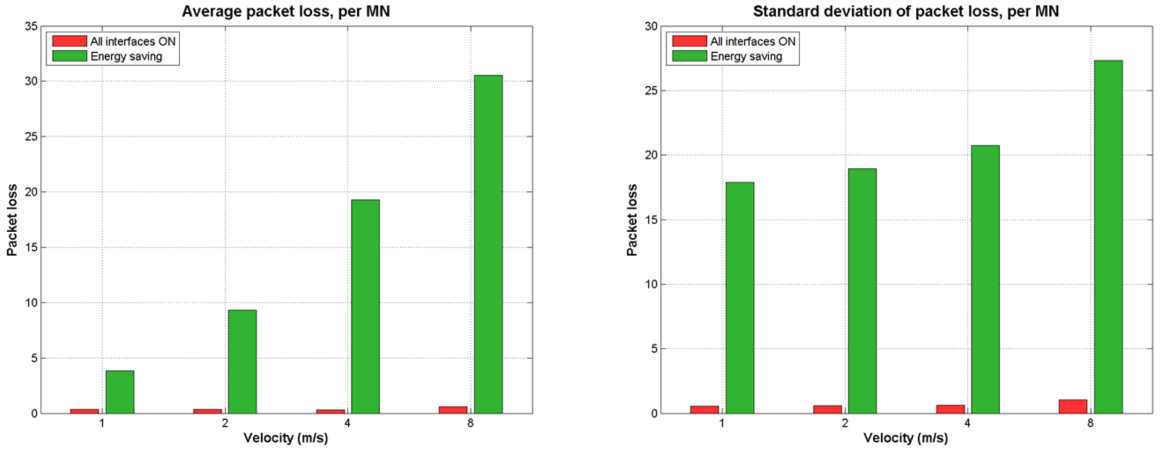

Figure 32: Avg. and Std. dev. of packet loss, per MN, for different velocities.

The average total number of packets transmitted and received in a simulation run, by each MN, is approximately 49550.

In its current implementation, the smart interface switching EE solution sacrifices reliability of packet delivery in exchange for energy saving, since more packets are lost with the proposed EE solution, also increasing with MNs velocity. This packet loss was identified to occur namely at the reactivation of deactivated interfaces, when the network association of the interface is delayed by channel congestion, in particular causing some packets to be lost when the WiMAX is not available and the Wi-Fi access point has bad channel conditions due to the distance to the MN.

Taking specific applications into consideration, this trade-off occurring with the proposed EE solution could eventually be tolerated in VoIP calls while the MN is moving at lower speeds, but eventually there would be degradation of video calls and streaming (MPEG4), which would become significantly damaged when increasing MN's speed.

Nevertheless, the mitigation of this shortcoming should be possible by implementing a more complex EE interface switching algorithm, ultimately also taking into account the speed of the MNs to enlarge the regions where both wireless interfaces are active.

3.6. Conclusion

The typical availability of multiple RF/RAT interfaces integrated in mobile UEs nowadays, as well as other features such as low power GPS chipsets, presents a good opportunity to capitalise on all those provided functionalities, with the support of MIH/MIIS specified in IEEE 802.21 standard, already well established. Assuming a proper implementation of this standard in a HetNet, the georeferenced-based solution is proposed targeting EE at the modern mobile UE, allowing to extend its autonomy and the user's mobility further.

The results attained through simulations using an *ns-2* based platform show that the proposed EE solution can reduce UE's energy spending on RAT interfaces by approximately 30%, on average, with a fairly simple implementation. Obviously, this energy saving figure can be higher when a larger number of different RAT interfaces are integrated in the UE. On the other hand, the currently proposed approach, as is, presents a trade-off in terms of additional packet loss, which is aggravated with UEs velocity, being mostly caused by delayed network re-association.

As future work, the identified added packet loss trade-off, associated to the currently implemented EE solution exploiting IEEE 802.21 MIIS, should be ultimately mitigated with the integration of a more complex RAT interface activation/deactivation algorithm providing the proposed EE solution with further intelligence.

In addition, it would be interesting to evaluate further the extra signalling overhead coming from the required information exchange with a MIIS server, as another eventual trade-off for the desired energy efficiency. Moreover, a further realistic and detailed study concerning the whole architecture required to support the proposed EE solution, as well as other HetNet solutions further exploiting IEEE 802.21 MIH/MIHF/MIIS, would surely provide valuable insights, ultimately instigating further interesting possibilities.

Furthermore, taking into account the emergence of new standards, e.g. 3GPP Release 15, which paves the way to 5G systems, another interesting path steering investigation, would be the adaptation and implementation of an equivalent solution exploiting 3GPP architectures together with their entities and functionalities. Such approach would be complementary to the IEEE-based solution presented in this thesis, here considering IEEE’s information service specified in IEEE 802.21 standard, and also IEEE’s wireless communication systems, i.e. WiMAX and Wi-Fi. On the other hand, it could also represent a straightforward alternative, as IEEE MIH/MIIS has not been widely adopted, as once intended, and 3GPP will have its own approach to provide context information for achieving multiple purposes, including RRM optimization. Such effort, together with the carried out work presented in this thesis, could be used to provide technical recommendations to 3GPP on how to better cater for mobility in 5G HetNets.

4. ENERGY EFFICIENT RRM FOR HETNETS

This chapter is dedicated to the work carried out in the scope of this thesis, proposing an EE approach in radio resource management for HetNets, as introduced in sections 1.2, 2.2 and 2.3, towards addressing the challenges presented in subsection 2.4.4. The proposed solution deals with an optimization problem taking into account specific QoS requirements for the downlink of an OFDMA system in a HetNet [72].

Many distinct approaches have been proposed to address the EE aspect in HetNets, taking advantage of various concepts coming for multiple domains, which have also been briefly tackled within the work carried out in the scope of this thesis [73][74].

The considered optimization approach takes into account a realistic network power consumption model, in this case also considering circuit power, in line with what is introduced in subsections 2.4.2 and 2.4.4. The proposed solution addresses EE maximization using convex optimization theory where the primary optimization criterion is data rate in a DL multiuser HetNet. Such criterion is considered as a QoS requirement while maximizing EE in this constrained based optimization problem.

Since the optimization problem is non-convex in nature, it has been reconstructed as a convex one and a novel straightforward efficient resource allocation algorithm was devised for maximizing achievable EE, with quick convergence. Such reconstruction was carried out transforming the considered problem into a convex optimization problem by redefining the constraint using cubic inequality, which results in an efficient iterative resource allocation algorithm. In each iteration, the transformed problem is solved by using dual decomposition with a projected gradient method.

The achieved numerical results and analytical insights presented in the following sections clearly exhibit the potential of the developed scheme for the considered challenging wireless HetNet system.

4.1. Introduction

Information and communication technology (ICT), through the equipment, devices and overall supporting infrastructure being continuously deployed, represents a significant source contributing to the increase of global emissions, due to the fast growth of the subsequent energy being consumed to satisfy ICT end users demand [24]. It has thus become evident that achieving higher energy efficiency is nowadays a prime concern when redesigning existing and future mobile communication systems.

As debated throughout this thesis, energy saving is predominantly positioning itself at the forefront of communication system design, as operators look to increase EE, i.e. bits per Joule, and decrease their network OPEX, which is also fully in line with global recommendations towards a sustainable economy with regard to environment [75].

In reaction towards these global trends, the 3GPP standards have been considering new energy efficient approaches into the design of 4G and beyond mobile networks, across the entire protocol stack, from physical layer to networking, in line with new networking topologies and deployment strategies. It can be said that a paradigm change has been taking place, where once spectral efficiency (SE) was the ruling performance metric of choice, now EE and the trade-off between both metrics are taking over in system design guidelines.

As briefly introduced in subsection 1.1.2, one of today’s reference scenarios is the heterogeneous network (HetNet) [5]. This system of networks comprises a blend of large cells, i.e. macro-cells, and small cells supported by low-power nodes, some of which may be designed and deployed with limited access, and some even without wired backhaul. HetNet is a promising approach for gaining more SE and enhanced area of coverage in future cellular wireless systems.

By appropriately deploying low-power nodes associated with macro-cells, small cells are created and usually utilized to broaden the area of coverage, or the bandwidth, e.g.

through relays nodes, where desired signals from a macro station only barely reach, or to increase the network capacity in places where high density of data usage is required. In order to have efficient interference management inside a HetNet, a clear point of coordination between different kinds of nodes is essential. A typical example of that is the coordination between the macro base station (MBS), e.g. an eNB, and small cell BSs, e.g. low-power nodes such as relay transmitters.

4.1.1. Previous work

Many scenarios involving wireless communication networks have been considered while investigating EE, in relevant research literature. The leading work where OFDM system has been applied for EE optimization using circuit power consumption is [76]. In [77], an energy efficient transmission rate and transmit power adaptation scheme is proposed for the orthogonal space-time block coding (OSTBC) MIMO system where channel state information at the transmitter (CSIT) is not perfect. EE is analysed considering a single cell OFDMA network, in [28]. The study in [78] also targeted a scenario similar to the latter, and applied particle swarm optimization techniques. In addition, the resource allocation is formulated in [79] to maximize EE for secure OFDMA systems as a mixed non-convex and combinatorial optimization problem with negligible circuit power taking into consideration single cell scenario. In [80], EE in OFDM distributed antenna systems (DAS), in single cell scenario is also analysed. Comparison of EE in different types of MIMO systems is also presented in [81].

Current technologies that address energy efficient RM for cooperative and competitive HetNet systems under a variety of network performance targets and constraints are not yet fully developed [82]. Although much work on EE network deployment strategies has been done, current outcomes are still quite basic and some important challenges remain to be investigated [83], e.g. EE optimization in HetNet for downlink OFDMA systems. In [45], the performance of conventional heterogeneous networks in terms of EE is first introduced, although only picocell is considered. Also, a multi-RAT HetNet energy optimization is presented in [46], but multi-tier EE optimization is still unavailable. The design of EE cellular networks by employing of BS sleep mode strategies, as well as small cells, and the trade-off issues associated with these techniques, is investigated in [47]. Moreover, optimization problems in the form of power consumption minimization are formulated and the optimal operating frequency of the MBS is also determined. The work in [48] looks into EE ways of operating dense small cell networks by applying concepts from cognitive radio (CR). A power-optimized vertical handover mechanism, considering static circuit and transmit power, is proposed for HetNet in [49]. Nevertheless, there are only few studies analysing EE optimization considering both types of circuit power (static and dynamic).

4.1.2. Optimization problem and EE approach

The work carried out in the scope of this thesis provides more insights into the optimal EE for the DL of OFDMA-based cellular system in a multi-tier HetNet system, under necessary QoS requirements, i.e. radio resource constraints. The maximization of EE as a convex optimization problem for HetNet is addressed. The consumption of both transmit and circuit power (static and dynamic) was considered in the design of an optimal EE system. The objective of the optimization problem is to maximize EE while fulfilling minimum QoS requirements, in this case bit rate.

To solve such problem, the optimization problem has been remodelled as a convex one and then a novel rate optimization algorithm is proposed to achieve maximum EE, with a good convergence time. Therefore, first the optimization problem is formulated where constraints are modelled as a cubic inequality, and a novel resource allocation algorithm is proposed to achieve maximum EE for a given SE, which includes both transmit and circuit power for HetNet scenario. An algorithm has been developed using gradient method by constructing Lagrangian function and checking that solution with Karush–Kuhn–Tucker (KKT) to meet the optimality of the solution.

The considered system model is presented in section 4.2, and an optimization problem to maximize EE is formulated in section 4.3. A novel optimizing rate allocation algorithm to achieve maximum EE in HetNet is provided in section 4.4. The attained numerical results are presented and discussed in section 4.5. Some open problems are identified in section 4.6, and future work is proposed as well, respectively.

4.2. System model

HetNets are networks composed by multiple and distinct nodes with different transmit power associated to the implemented RAT and technology being considered. For example, a HetNet may comprise macro cells, small cells and relays. In the considered scenario, apart from a conventional HetNet, it is also proposed a HetNet with interference coordination between different cells, with a system architecture based on LTE technology, as illustrated in Figure 33.

One HetNet cell site has one eNB (which is the main MBS) and several picocells served by low power pico nodes (pN) and relays nodes (RN), linked via optical fibre and microwave backhauls, respectively. It is assumed that the HetNet can be associated with other HetNets to coordinate inter-cell interference (ICI), and thus no ICI is present. For this work, it was considered an adaptation of the coordination techniques from [84].

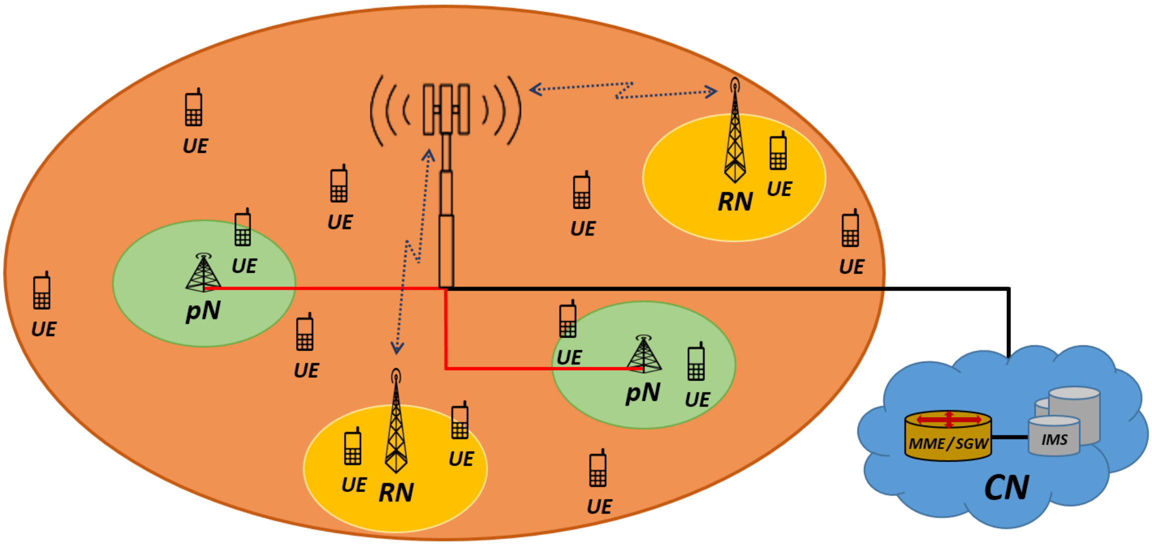

Figure 33: System model considering a HetNet system.

Therefore, in this work it was necessary to address only intra-cell interference in each HetNet, caused by the unavoidable cell overlap. Each macro, pico and relay is designed considering a different frequency range, as illustrated in Figure 33 using different colours. Essentially, each picocell's pN, is affected by interference from other pNs within the same HetNet cell site. This same consideration, in this work, also applies to RNs, respectively.

The impact of the intra-cell interference for HetNet in this scenario is negligible since the pNs and RNs are appropriately deployed with enough distance among them, respectively, and the MBS manages this interference through power control, and the backhaul between the eNB and pico or relay links [85][86]. Thus, it is possible to consider there is no significant amount of interference between the low power nodes inside a HetNet cell site to degrade the system efficiency in a harsh way.

We take into account a HetNet cell site with downlink OFDMA network containing U active users. The aggregate bandwidth B is evenly divided into M physical resource blocks (PRB), where each PRB has a bandwidth of $W = B / M$. The sets of active users and all PRBs, are defined as follows, respectively:

$$\mathcal{U} = \{1, 2, \ldots, U\}\ ; \quad \mathcal{M} = \{1, 2, \ldots M\} \tag{4.1}$$

The transmit power for user u on PRB m is expressed as $p_u^m > 0$ and obviously there is only one $m \in \mathcal{M}$ for each $u \in \mathcal{U}$. Taking these into account, it is possible to define the maximum achievable rate according to Shannon theorem, for user u on PRB m, as

$$\overline{d_u^m} = W \log_2 \left(1 + \frac{p_u^m . g_u^m}{I + W N_0}\right) \tag{4.2}$$

where $g_u^m = |h_u^m|^2$ is the channel power gain of user u on PRB m, while h_u^m is the corresponding frequency response. Moreover, N_0 is the single-sided noise power spectral density, and I is the received interference coming from the interfering transmitter(s) inside the HetNet between similar types of small base stations, or low power nodes, i.e. pNs and RNs, respectively, which is here assumed negligible.

The total transmit power, *P*, and system throughput, *D*, are defined as follows, also in line with what is presented in subsection 2.4.2:

$$P = \sum_{u=1}^{U} \sum_{m=1}^{M} p_u^m \; ; \quad D = \sum_{u=1}^{U} \sum_{m=1}^{M} d_u^m \tag{4.3}$$

The transmission power also includes the reciprocal of the drain efficiency of the power amplifier, which is here denoted as α. Transmission power is thus represented as αp_u^m. Apart from the transmission power, circuit power is considered here as well. Circuit power, $P_{circuit}$, can be separated into two parts [28]. One is static, or fixed, and the other is dynamic depending on different parameters, and can be defined as follows:

$$P_{circuit} = P_{static} + \delta D \tag{4.4}$$

The fixed circuit power while transmitting is denoted by P_{static}, and δ is an invariant expressing dynamic power consumption per unit data rate.

4.3. Problem Formulation for Optimal EE

An energy efficiency (EE) metric, as introduced in subsection 2.4.2, namely in expression 2.4, can be defined for the downlink of OFDMA HetNet as follows:

$$EE = \frac{D}{\alpha P + P_{circuit}} \tag{4.5}$$

EE is denoted as transmitted bits per unit energy consumption, i.e. bits/Joule, where the energy consumption includes transmission, or radiation energy consumption (αP multiplies over transmission time) and circuit energy consumption ($P_{circuit}$ multiplies over transmission time) of the transmitter.

Accordingly, the optimization problem can be formulated as follows:

$$\max_{d} EE = \frac{\sum_{u=1}^{U} \sum_{m=1}^{M} d_u^m}{\alpha_1 P_{mc} + \alpha_2 \sum_{n=1}^{N} P_{pc_n} + \alpha_3 \sum_{l=1}^{L} P_{rl_l} + P_{static} + \delta \sum_{u=1}^{U} \sum_{m=1}^{M} d_u^m} \tag{4.6}$$

This expression can be extended, and subject to the conditions as defined below:

$$\max_{d} EE = \frac{\sum_{u=1}^{U}\sum_{m=1}^{M} d_u^m}{\alpha_1\left[\sum_{u=1}^{U}\sum_{m=1}^{M} p_u^m\right]_{mc} + \alpha_2 \sum_{n=1}^{N}\left[\sum_{u=1}^{U}\sum_{m=1}^{M} p_u^m\right]_{pc_n} + \alpha_3 \sum_{l=1}^{L}\left[\sum_{u=1}^{U}\sum_{m=1}^{M} p_u^m\right]_{rl_l} + P_{static} + \delta \sum_{u=1}^{U}\sum_{m=1}^{M} d_u^m}$$

$$1)\ \widehat{d_u^m} \leq d_u^m \leq \overline{d_u^m}\ or\ d_u^m = 0,$$

$$2)\ d_u^m \geq 0, \qquad (4.7)$$

where $\widehat{d_u^m}$ denotes the minimum rate requirement for user *u* on PRB *m*, while P_{mc}, P_{pc}, and P_{rl} represent the transmit power from macro BS, pico and relay node stations, respectively. The coefficients α_1, α_2, and α_3 denote the power consumption that calibrates with average transmitted power due to RF amplifier and feeder losses. The static circuit power, for the different types of transmitter, considering *N* pico cell nodes and *L* relays, can also be expressed as:

$$P_{static} = P_{static_{mc}} + \sum_{n=1}^{N} P_{static_{pc_n}} + \sum_{l=1}^{L} P_{static_{rl_l}} \qquad (4.8)$$

The proposed optimization problem, where the optimization variable is the rate vector **d**, is non-convex by nature. To solve the problem using convex optimization method it is necessary to remodel the constraints of the optimization problem. For that, constraint *1)* is redefined using the approach as in [87]. The reformulation of constraint *1)* is performed based on non-negativity of constraint *2)*, ensuring convexity of the problem, as given in the following expression:

$$(d_u^m).\left(d_u^m - \widehat{d_u^m}\right).\left(\overline{d_u^m} - d_u^m\right) \geq 0, \qquad (4.9)$$

Four conditions are subsequently derived from this expression, being necessary to find the 1st and 3rd conditions that satisfy the imposed constraints' non-negativity and convexity. Those four conditions are the following:

1) $d_u^m = 0$

2) $d_u^m \in \left(0, \widehat{d_u^m}\right)$

3) $d_u^m \in \left[\,\widehat{d_u^m}, \overline{d_u^m}\,\right]$

4) $d_u^m \in \left(\overline{d_u^m}, +\infty\right)$ (4.10)

The solution space defined by the constraints is convex, and the objective function of the proposed optimization problem is concave, thus the formulated problem has a solution, and that means it is a convex optimization problem [88].

4.4. Optimization for Maximum EE

In order to solve the defined problem, an algorithm has been developed using gradient method by constructing Lagrangian' function and satisfying the solution with Karush-Kuhn-Tucker (KKT). The Lagrangian of our problem is formulated with Lagrange multiplier λ, as

$$L(\mathbf{d},\lambda) = EE + \sum_{u=1}^{U} \sum_{m=1}^{M} \lambda_u^m \left[(d_u^m) . \left(d_u^m - \widehat{d_u^m} \right) . \left(\overline{d_u^m} - d_u^m \right) \right], \tag{4.11}$$

with the reciprocal Lagrange dual function

$$\mathrm{g}(\lambda) = \max_{d} L(\mathbf{d},\lambda). \tag{4.12}$$

Subsequently the dual problem is expressed as follows:

$$\min_{\lambda} \mathrm{g}(\lambda)\,;\ \lambda \geq 0 \tag{4.13}$$

The objective functions of the problem formulated in expressions 4.11 and 4.13 are thus differentiable, with respect to the primal variable **d** and dual variable λ. Therefore, these two problems can be solved by the gradient projected method [88], defined as

$$d_u^m(i+1) = \left[d_u^m(i) + \theta \frac{\partial L(\mathbf{d},\lambda)}{\partial d_u^m} \right]^+, \tag{4.14}$$

$$\lambda_u^m(i+1) = \left[\lambda_u^m(i) - \psi \frac{\partial L(\mathbf{d},\lambda)}{\partial \lambda_u^m} \right]^+, \tag{4.15}$$

where *i* denotes the iteration index, while θ and ψ are positive step sizes, and $[.]^+$ is a projection onto the set of $\mathbb{R}^+$. By applying the KKT condition

$$\frac{\partial L(\mathbf{d},\lambda)}{\partial \mathbf{d}} = 0, \tag{4.16}$$

it is possible to obtain a solution. The proposed approach can be summarized in Algorithm 1, as introduced in [89] and presented here below.

Algorithm 1: QoS (data rate) constrained optimization algorithm to maximize EE

```
1) Step 1: Initialization
   Set $d_u^m(0)$ and $\lambda_u^m(0)$ as non-negative value for all users u and PRBs m.
2) Step 2: Optimization
   Apply the gradient method:
   - update $d_u^m(i+1)$ according to Eq. 4.14
   - update $\lambda_u^m(i+1)$ according to Eq. 4.15
3) Step 3: Iteration
   Iterate until the implementation converges to the optimality (or the number of
   iterations is reached) and the algorithm stops, otherwise keep returning to step 2.
```

4.5. Simulation Results

In this section are presented the numerical results attained after simulation considering the system model defined in section 4.2, as illustrated in Figure 33, for a specific setup to demonstrate and discuss the effectiveness of the proposed EE solution. For this, a system level simulation tool was used to implement the proposed iterative optimization algorithm and to evaluate its convergence behaviour. Frequency reuse-1 model is used and all simulations are done for the DL case.

4.5.1. Simulation setup

An LTE network was considered, with a central cell as reference in a hexagonal deployment, surrounded by six cells in the first tier and by other twelve cells in the second tier. Wrap-around was considered to avoid border effects. The UEs are deployed independently with uniform distribution throughout the cell. Monte Carlo simulation is used with full-queue traffic model for all the users, which means they always have information ready to be transmitted.

The implementation in the simulation tool, with one macro cell and respective MBS (*eNB*), several relays (*RL*) and other low power nodes serving pico cells (*PC*), is shown in

Figure 34 for the main central cell. Each circle represents the coverage area of the respective transmitter, and the active UEs are marked as blue asterisks. The most important parameters of the simulated system are set according to the LTE standard [90], and are summarized in Table 6.

4.5.2. Numerical results and analysis

The convergence behaviour of the proposed optimization algorithm, according to different step sizes is presented in Figure 35. It is visible that using the proposed algorithm the system EE approaches the optimal value after some iterations. The algorithm should converge quickly enough to have a realistic implementation in a system level basis. Therefore, the comparison between different types of step size number is important. The graph also shows the convergence behaviour of EE with two types of static step size, fractional (0.1) and integer (1), and dynamic decreasing step size 1/(1+0.01.*i*) [87].

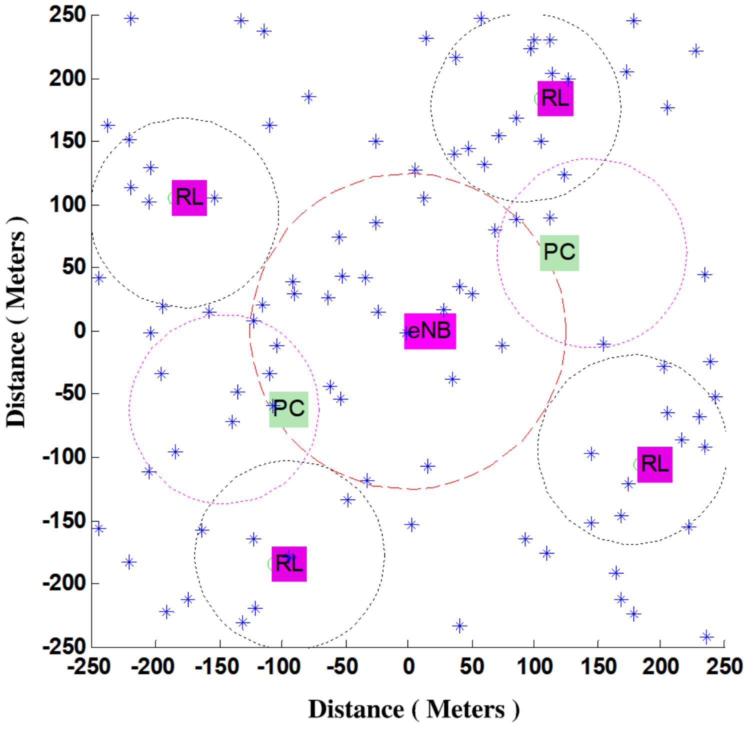

Figure 34: Deployment of one reference HetNet main cell site.

Table 6: Simulation parameters.

Parameter	Value
- Deployment	UEs randomly deployed, uniformly, at each cell
- Cellular layout	Hexagonal, 19 cell sites, 3 cell sectors per cell-site
- Inter-HetNet distance	500 m
- Macro transmit power (P_{mc})	20 W
- Pico transmit power (P_{pc})	2 W
- Relay transmit power (P_{rl})	1 W
- Static circuit power (P_{static})	Macro: 15 W; Pico: 10 W; Relay: 5 W
- Power consumption coeff. α	Macro (α_1): 3.8; Pico (α_2): 5.5; Relay (α_3): 5.5 [91]
- Dynamic power cons. coeff. δ	33.0103 dBm/Mbps [28]
- Carrier Frequency	2.6 GHz
- Bandwidth	10 MHz
- Number of PRB	50
- Noise density	-174 dBm/Hz
- Path-loss model	
▪ Macro	PL (dB) = $40.\log_{10}(d) - 11.02$
▪ Pico; Relay	PL (dB) = $22.\log_{10}(d) + 34.02$
- Log-normal shadowing	Macro: 4 dB; Pico: 6 dB; Relay: 6 dB
- Noise figure	7 dB
- Antenna gain	Macro: 14 dBi; Pico: 5 dBi; Relay: 5 dBi
- Maximum num. of UEs per cell	40

It can be observed that the convergence behaviour with a dynamic step size $1/(1+0.01.i)$ is smoother but converges more slowly than with integer static step size. The convergence behaviour with fractional static step size 0.1 not only demonstrates the slowest convergence but also presents the least smoothness. A considerable gain in employing the dynamic step size is that it allows to converge rapidly first with bigger step sizes and then fine-tuning itself at later stage with smaller step sizes. In conclusion, the static step size appears to be more suitable and can converge more rapidly, although a dynamic step size can be preferable due to its slow-transform rate profile, which is key for the smoothness of system quality. Otherwise, a sudden change of access data rate will often result in undesirable quality fluctuation.

Figure 36 shows the static power (P_{static}) impact in the proposed optimization algorithm.

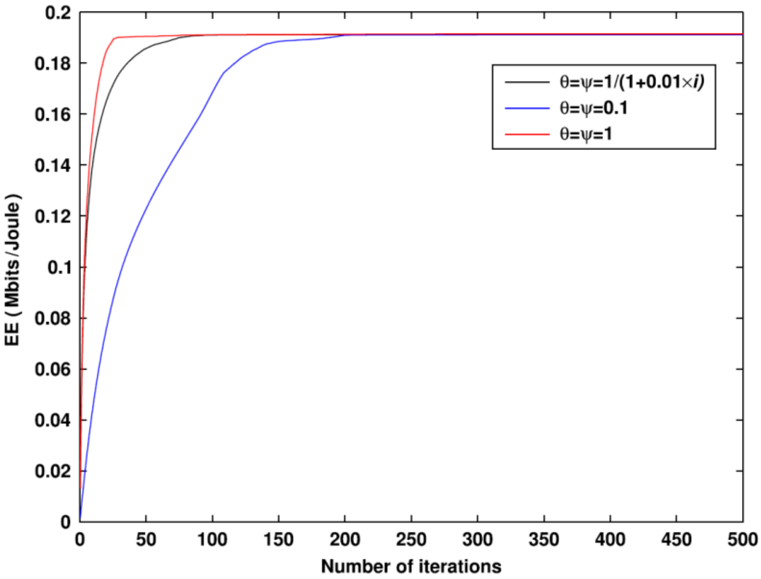

Figure 35: Impact of the iteration step size.

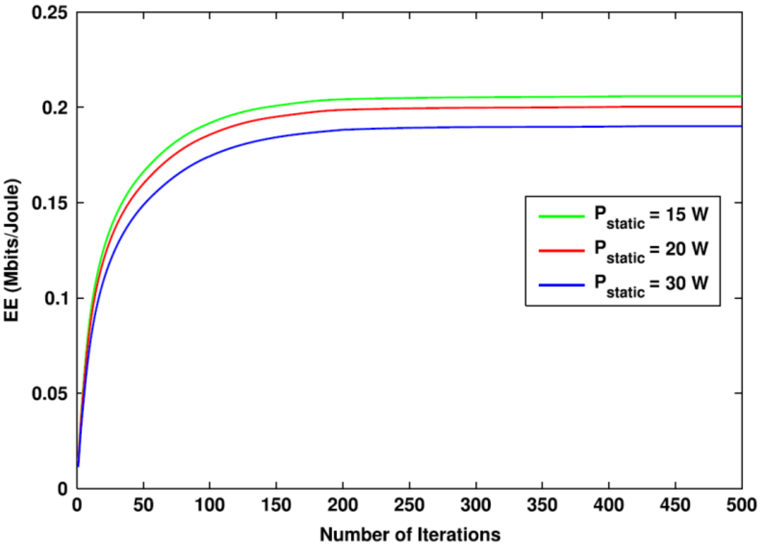

Figure 36: Comparison of convergence of static circuit power.

The curves show that the EE is improved if we consider low static power factor P_{static}, while also converging faster when more static circuit power is considered. For this graph, we assume total static circuit power of one HetNet cell site, i.e. $P_{static} = P_{static_{mc}} + P_{static_{pc}} + P_{static_{rl}}$. The value of P_{static} makes an impact on the system EE, i.e. higher P_{static} leads to lower system EE. On the other hand, although for higher values of P_{static} further iterations are necessary for convergence, the optimization is smoother and, thus, abrupt changes would make less impact on the system.

The effect of having different number of active UEs in the system, and their respective demands in terms of PRBs, using the proposed optimization algorithm and considering 3 PRBs per UE, is shown in Figure 37. Increasing the quantity of UEs leads to higher system EE as the overall system data rate increases as well. In this chart, the curves show that the smaller the quantity of UEs the faster the proposed algorithm converges, as the more UEs the more iterations are required for convergence. In addition, thus, if offloading is necessary in the HetNet, then faster algorithm convergence occurs for lower QoS requirements, while convergence becomes slower otherwise.

For some specified power consumption value, we can deduce that the system capacity and EE is proportionately related.

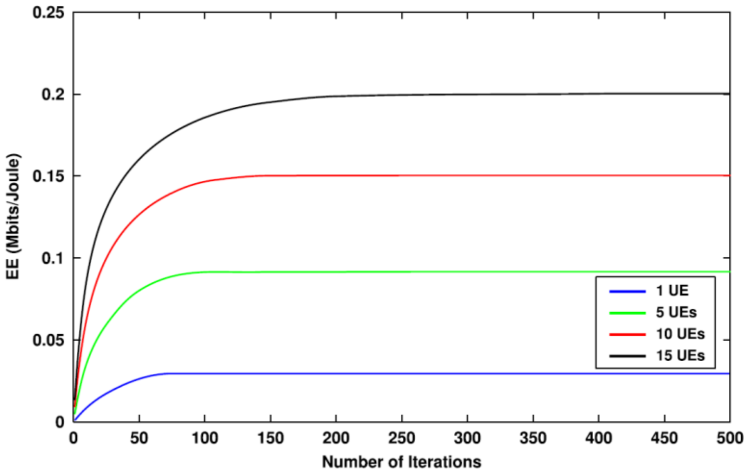

Figure 37: Comparison of system EE convergence for different quantities of UEs.

It means that when system capacity is higher, the system EE should also be higher, and vice versa. This feature allows less power requirement to satisfy the demand for a defined capacity requirement. Hence, this optimization method is practical for assessing how energy efficient a wireless system is.

To help clarifying this aspect, Figure 38 presents the system EE considering a single UE with variable consumed power for different amounts of PRBs demand. It is clear that maximum system EE is achieved when allocating all possible PRBs to the UE and adjusting the power accordingly. The curves represented in this chart also show the average cell EE as a function of SE, according to Eq. 4.7. It is visible that EE increases with SE until some point and then decreases. To interpret this result, it is important to recall that the EE is a monotonic function until circuit power is considered. The curves tend to be quasi-concave on SE, which justifies the basic criterion of SE–EE curve depicted in [28]. According to [92], a function $f : Q \rightarrow \mathbb{R}$ is quasi-concave on Q if and only if for all $x, y \in Q$ and for all $\beta \in (0, 1)$ it is the case that

$$f[\beta x + (1-\beta)y] \geq \min\{f(x), f(y)\}, \tag{4.17}$$

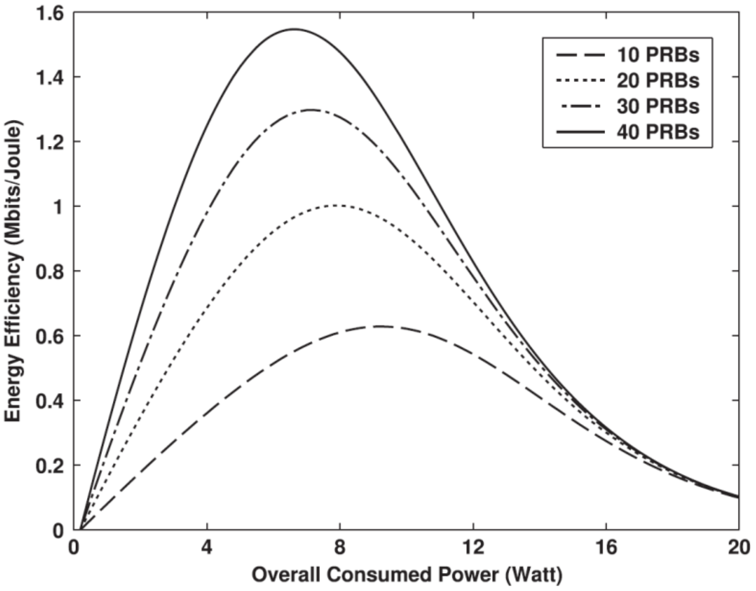

Figure 38: Relationship of EE, overall consumed power, and PRBs allocation.

which justifies the quasi-concavity present in the curves of Figure 38. Furthermore, those curves also demonstrate the optimum envelop of the overall SE-EE region, which provides a complete perspective on the system SE-EE trade-off. The optimum EE emphasizes the presence of a saturation point, beyond which the EE can no longer grow regardless of how many additional resources are utilized.

With such result, on one hand, it is possible to plan a HetNet system with optimal energy consumption while the system capacity is not limited. On the other hand, it is possible performing optimization to achieve maximum EE while fulfilling a given SE (QoS) requirement. Therefore, to achieve further EE gain, the network operators should set up operational SE of the system close to the respective peak value of EE.

The three-dimensional (3D) representation in Figure 39 shows the complete picture achieved through the carried out numerical simulations, where the SE and EE relation, as well as bandwidth, can be visualized in a further extensive way, after Figure 38 where bandwidth, SE and EE are considered separately, and do not provide the best picture. The presented results are obtained with a specific assignment for 10 PRBs for different amounts of bandwidth allocations and SE.

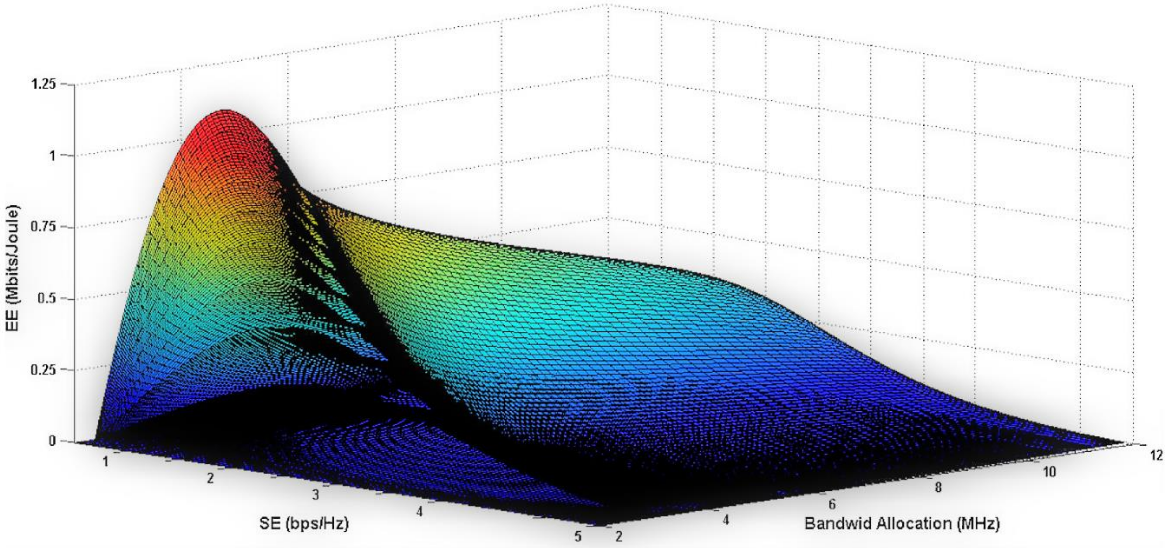

Figure 39: EE vs. SE and bandwidth allocations for single UE scenario.

Two major features are visible in Figure 39. First, if SE is increasing, the optimal solution of EE always appears inside the feasible domain as the global optimality, but after reaching the optimality, it decreases with the SE. Second, the overall curve surface is quasi-concave, as well is the curve line connected by all the boundary optimal points. These two features can be easily extended to the curves presented in Figure 38, since the considered optimality is actually searched among many similar curves with various PRB assignments.

Finally, the proposed algorithm, for which globally optimal PRB assignment policies were developed to maximize the overall network EE, was analysed considering different quantities of UEs, in terms of the transmitted frame, while transmit power is constant. Figure 40 presents the relationship between system EE and the number of transmitted frame. Using the proposed optimization algorithm, the system EE increases with the amount of UEs although the transmit power is constant (e.g. 20 W in this case). The more frames are transmitted, the higher the probability of attaining further EE, since with the increase of frame transmission better utilization of the resource allocation is ensured, thus eventually enhancing throughput.

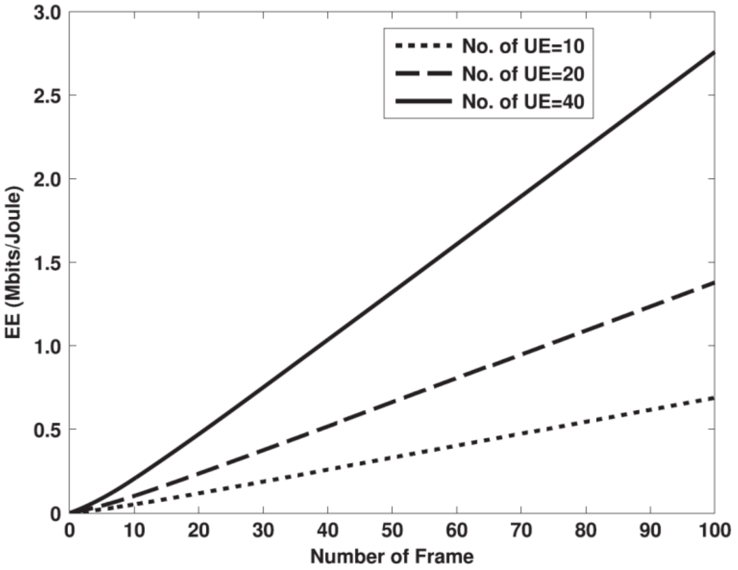

Figure 40: EE vs. the number of frame (transmit power is constant).

4.6. Conclusion

This chapter introduces a concrete approach towards increasing the energy efficiency in HetNet RRM. The investigated EE optimization in downlink OFDMA system for HetNet, using convex optimization theory, is presented and an optimization problem is formulated to maximize EE, where a constraint is formulated as a cubic inequality for a given data rate requirement. The formulated problem not only takes into consideration the radiated power, but also both types of the circuit power, i.e. static and dynamic, without which the HetNet system SE-EE trade-off analysis would be incomplete.

The innovative optimizing method, developed to achieve maximum system EE, for radio resource allocation was implemented in a system level simulation tool, which provided interesting numerical results. The results show that the proposed optimization algorithm converges rapidly, which is key in the design of practical EE HetNet systems. Such results can be exploited towards designing optimal energy consumption networks (maximizing EE) based on QoS-oriented HetNet method, while the total power is fixed.

As future research work, taking into account particularly the considered scenario and proposed solution, different QoS constraint approaches could be addressed along with the comparison of other relevant methods.

In addition, also targeting to provide some technical recommendations to 3GPP towards higher EE in 5G HetNets, taking into account the approach proposed in this work, further research could be made considering C-RAN architecture. Such investigation could, on one hand, taking into account the subsequent potentially enhanced mitigation, or coordination, of interference provided by C-RAN, target optimised fronthaul design, which is key in emerging HetNets [93]. On the other hand, considering C-RAN together with SDN/NFV, thus in a virtualised implementation, it would be interesting to consider EE optimization taking into account both centralisation and virtualisation aspects, namely considering energy consumption by the circuitry needed for the C-RAN.

5. CONCLUSION AND FUTURE WORK

This thesis addresses and targets contributing to the growing concerns regarding energy efficiency in the ICT industry, specifically focusing on cellular wireless networks. Taking into account the current panorama and developments in this field, namely concerning the emerging 5G communication systems and networks, the HetNet concept is at the core of the work that has been carried out and presented in this thesis.

The work takes into consideration the two ends of the considered system of networks. The first are the nomad end users that should be able to roam freely in such system, always with continuous service provided in a seamless way, and the second is the core network managed by the respective telecommunications operator. The main benefits of achieving higher energy efficiency are, for the former, extended UE autonomy, which reflects in their users' further mobility, and for the latter, lower or stable operational costs, reflecting as well in the ICT global footprint on the environment.

An overview of the past and ongoing developments, culminating in the emerging 5G networks, has been made and summarized in the first chapter of this thesis, together with the most relevant concepts and existing challenges, which provide relevant motivation for the work carried out. A key overall challenge is how to take the most advantage from the available resources, such as RF spectrum, to meet the communication requirements for providing a specific service or application, in a further energy efficient way. The solutions that have been considered in this thesis work, for addressing such challenge, exploit radio resource management mechanisms and schemes to achieve further energy efficient HetNets.

Two main approaches were considered in the investigation and development of such EE solutions for HetNets, which also address the EE ambition at each of the two ends of the considered HetNet, i.e. at the UEs and at the CN, respectively. In one approach, the concepts, processes and dynamics associated to VHO, and the respective IEEE 802.21 supporting standard functionalities, are exploited through a straightforward, but innovative, conceptual solution to achieve significant energy saving at multi-RAT UEs. The other approach exploits HetNet mechanisms for radio resource allocation optimizing the energy consumption at the network side, namely at the BS, following a specific analytical method. For that, an extended investigation was carried out on the relevant topics, i.e. VHO/MIH, DRA, Interference Management and Energy Efficiency, which is consolidated in the second chapter of this thesis. The remaining chapters include two specific solutions, addressing each of the two considered approaches targeting to enhance EE at each of the two ends of a HetNet, respectively, UEs and CN.

The proposed EE solution exploiting features provided by IEEE 802.21 (MIH/MIIS) and VHO, targeting energy saving at modern multi-RAT mobile UEs, can eventually reduce the energy consumption at their RAT interfaces by roughly 30% on average, with a straightforward implementation, according to the attained simulation results. However, the currently proposed scheme presents a trade-off in terms of added packet loss linked to UEs velocity, which is mostly caused by delayed network re-association.

For downlink OFDMA HetNet system, the proposed EE optimization algorithm for radio resource allocation, taking a given data rate requirement into account, associated to QoS, presented fast convergence, which is key in the design of real EE HetNet systems. The algorithm considers not only the radiated power, but also both types of the circuit power. Simulation results can eventually be exploited towards designing optimal energy consumption networks based on QoS-oriented HetNet method, while total power is fixed.

As future work in the scope of the EE RRM solutions for HetNets and investigation lines of research, several subsequent approaches and/or extensions of the proposed EE solutions can eventually be carried out. For instance, by developing and integrating a more complex RAT interface switching algorithm providing the proposed MIH/MIIIS-based EE solution with further intelligence, namely towards mitigating the detected additional packet loss associated to the UEs velocity and eventual network re-association. Also, regarding the EE RRM optimization solution, other QoS constraint approaches could be addressed along with the comparison of other relevant methods, taking into account the considered scenario.

Moreover, on one hand, it is also worth pursuing evaluating the extra signalling overhead coming from the required information exchange with a MIIS server, in the proposed EE VHO/MIH-based solution, as well as performing further realistic and detailed studies concerning the whole HetNet architecture required to support the proposed EE solutions.

On the other hand, it would also be important to take into account the paradigm shift that arise with the emergence of new standards, namely 3GPP Release 15, placing the stepping stones for 5G systems and networks. In this perspective, at a first instance, it is key to go beyond the assumptions made in this work, by considering C-RAN architecture in the addressed challenge for achieving higher EE, which is a prime concern when redesigning existing and future mobile communication systems, namely in sustainable and environmental perspectives.

C-RAN has the potential to reduce multi-tier interference in a more seamless way, namely when associated to SDN/NFV, by virtualising baseband processing. In such implementation, radio resources can be seen as being in a “pool of resources” to be shared and used by the network, in a cooperative user-network centric way, rather than in the classical non-cooperative paradigm typically implemented in today’s networks. As C-RAN aims to harness all the baseband processing for all users within a common unit, it provides the operator complete control over the network and the ability to coordinate signal transmissions, which provides a significant step towards mitigating interference in the network, associated with relevant techniques, e.g. eICIC and CoMP. This is, in fact, part of the 3GPP technology roadmap that still has several research challenges to be solved, e.g. in terms of fronthaul deployment strategy. The developments carried out in this work could thus be adapted or extended under the light of such new paradigm, and could eventually be used for contributing to recommendations to 3GPP for designing future mobile communication systems.

On a broader scope, the work carried out and presented in this thesis could also be extended taking into account the emerging concept of mobile small cells, where the mobile UE adopts the role of access point or remote radio head. This shifts the paradigm of legacy mobile networks from purely being network centric towards being device/UE centric, where mobile devices could then be seen as a pool of additional network resources to be used by the operator to extend network coverage on demand. Such new paradigm triggers new research challenges, besides de ones focused on EE, also in terms of incentives for promoting UEs owners cooperation. The latter will thus also play a key role, namely with respect to the devices autonomy, which is directly associated to the involved mobility, and the respective virtual RRM required to extend it as much as possible, ultimately benefiting the end users.

REFERENCES

[1] I. F. Akyildiz, S. Mohanty and J. Xie, “A Ubiquitous Mobile Communication Architecture for Next-Generation Heterogeneous Wireless Systems,” IEEE Communications Magazine, vol. 43, no. 6, pp. 29–36, June 2005.

[2] A. Sanmateu et al., “Seamless Mobility across IP Networks Using Mobile IP,” International Journal on Computers and Telecommunications Networking, vol. 40, pp. 181–190, September 2002.

[3] J. McNair and F. Zhu, “Vertical handoffs in fourth-generation multinetwork environments,” IEEE Wireless Communications vol. 11, no. 3, pp. 8–15, June 2004.

[4] C. Anderson, D. Freeman, I. James, A. Johnston and S. Ljung, “Mobile Media and Applications, From concept to Cash: Successful Service Creation and Launch”, Wiley, November 2005.

[5] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, D. Malladi, “A survey on 3GPP heterogeneous networks,” IEEE Wireless Communications, vol. 18, no. 3, pp. 10–21, June 2011.

[6] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, “What Will 5G Be?,” IEEE Journal on Selected Areas in Communications, vol. 32, no. 6, pp. 1065–1082, June 2014.

[7] 3GPP, "Technical Specification Group Services and System Aspects; Release 15 Description; Summary of Rel-15 Work Items (Release 15). Technical Report", TR 21.915 V1.1.0, March 2019.

[8] V. Chandrasekhar, J. Andrews, and A. Gatherer, “Femtocell networks: a survey,” IEEE Communications Magazine, vol. 46, no. 9, pp. 59–67, September 2008.

[9] China Mobile Research Institute “C-RAN: The Road Towards Green RAN,” Technical Report, April 2011.

[10] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, “Cloud RAN for Mobile Networks – a Technology Overview,” IEEE Communications Surveys Tutorials, vol. 17, no. 1, pp. 405–426, September 2014.

[11] V. Jungnickel, K. Habel, M. Parker, S. Walker, C. Bock, J. Ferrer Riera, V. Marques, and D. Levi, “Software-defined open architecture for front- and backhaul in 5G mobile networks,” in 2014 16th International Conference on Transparent Optical Networks (ICTON), 2014, pp. 1–4.

[12] R. Wang, H. Hu, and X. Yang, “Potentials and Challenges of C-RAN Supporting Multi-RATs towards 5G Mobile Networks,” IEEE Access, vol. 2, no. 1, pp. 1187–1195, October 2014.

[13] R. J. Weiler, M. Peter, W. Keusgen, E. Calvanese-Strinati, A. De Domenico, I. Filippini, A. Capone, I. Siaud, A.-M. Ulmer-Moll, A. Maltsev, T. Haustein, and K. Sakaguchi, “Enabling 5G backhaul and access with millimeter-waves,” in 2014 European Conference on Networks and Communications (EuCNC), pp. 1–5, June 2014.

[14] GSMA, “Powerful Digital Highways for Europe’s Connected and Automated Driving”, September 2015, URL: https://www.gsma.com/newsroom/press-release/powerful-digital-highways-for-europes-connected-and-automated-driving/ (*accessed September 2019*)

[15] IEEE P802.21-2008, “Standard for Local and Metropolitan Area Networks: Media Independent Handover Services,” IEEE, January 2009.

[16] G. Lampropoulos, N. Passas, L. Merakos, and A. Kaloxylos, “Handover management architectures in integrated WLAN/cellular networks,” IEEE Communications Surveys Tutorials, vol. 7, no. 4, pp. 30–44, Fourth Quarter 2005.

[17] A. Jalali et al., “Data throughput of CDMA-HDR—a high efficiency-high data rate personal communication wireless system,” IEEE VTC, vol. 3, pp. 1854–1857, May 2000.

[18] N. Hassan and M. Assaad, “Optimal Fractional Frequency Reuse (FFR) and Resource Allocation in Multiuser OFDMA System,” in 2009 International Conference on Information and Communication Technologies, Karachi, pp. 88–92, August 2009.

[19] N. Himayat, S. Talwar, A. Rao, and R. Soni, “Interference management for 4G cellular standards [WIMAX/LTE update],” IEEE Communications Magazine, vol. 48, no. 8, pp. 86–92, August 2010.

[20] J. Li, N. Shroff, and E. Chong, “A reduced-power channel reuse scheme for wireless packet cellular networks,” IEEE/ACM Transactions on Networking, vol. 7, no. 6, pp. 818–832, December 1999.

[21] Huawei, “R1-050507: Soft frequency reuse scheme for UTRAN LTE,” 3GPP TSG RAN WG1 Meeting #41, May 2005.

[22] L. Belkhir, A. Elmeligi, “Assessing ICT global emissions footprint: Trends to 2040 & recommendations,” Journal of Cleaner Production, vol. 177, pp. 448–463, March 2018.

[23] J. Malmodin and D. Lundén, "The energy and carbon footprint of the global ICT and E&M sectors 2010 - 2015", in proc. 5th International Conference on Information and Communication Technology for Sustainability (ICT4S2018), EPiC Series in Computing, vol. 52, pp 187–208, May 2018.

[24] G. Fettweis and E. Zimmermann, “ICT energy consumption – trends and challenges”, in proc. 11th International Symposium on Wireless Personal Multimedia Communications (WPMC '08), September 2008.

[25] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, “Fundamental trade-offs on green wireless networks,” IEEE Communications Magazine, vol. 49, no. 6, pp. 30–37, June 2011.

[26] C. E. Shannon, “A mathematical theory of communication”, The Bell System Technical Journal, vol. 27, no.3; no. 4, pp. 379–423; 623–666, July; October 1948.

[27] G. Li, Z. Xu, C. Xiong, C. Yang, S. Zhang, Y. Chen, and S. Xu, “Energy-efficient wireless communications: Tutorial, survey, and open issues,” IEEE Wireless Communications, vol. 18, no. 6, pp. 28–35, December 2011.

[28] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, “Energy- and spectral-efficiency tradeoff in downlink OFDMA networks,” IEEE Transactions on Wireless Communications, vol. 10, no. 11, pp. 3874–3886, November 2011.

[29] I. F. Akyildiz, D. M. Gutierrez-Estevez, and E. C. Reyes, “The evolution to 4G cellular systems: LTE-Advanced”, Physical Communication, Elsevier, vol. 3, no. 4, pp. 217–244, December 2010.

[30] G. Miao, N. Himayat, G. Li, A. Koc, and S. Talwar, “Interference-aware energy efficient power optimization”, in proc. of IEEE International Conference on Communications 2009 (ICC’09), pp. 1–5, June 2009.

[31] S. Mumtaz, J. Rodriguez, "Green Communication in 4G Wireless Systems", River Publishers, March 2013.

[32] X. Yan, Y. A. Şekercioğlu and S. Narayanan, “A survey of vertical handover decision algorithms in Fourth Generation heterogeneous wireless networks,” Computer Networks, Elsevier, vol. 54, no. 11, pp. 1848–1863, August 2010.

[33] M. Kassar, B. Kervella and G. Pujolle, “An overview of vertical handover decision strategies in heterogeneous wireless networks,” Computer Communications, Elsevier, vol. 31, no. 10, pp. 2607–2620, June 2008.

[34] M. R. Kibria and A. Jamalipour, “On Designing Issues of the Next Generation Mobile Network,” IEEE Network Magazine, vol. 21, no. 1, pp. 6–13, February 2007.

[35] N. Nasser, A. Hasswa and H. Hassanein, “Handoffs in fourth generation heterogeneous networks,” IEEE Communications Magazine, vol. 44, no. 10, pp. 96–103, October 2006.

[36] D. Xenakis et al., “Mobility Management for Femtocells in LTE-Advanced: Key Aspects and Survey of Handover Decision Algorithms,” in IEEE Communications Surveys & Tutorials, vol. 16, no. 1, pp. 64–91, First Quarter 2014.

[37] D. Xenakis et al., “An energy-centric handover decision algorithm for the integrated LTE macrocell-femtocell network,” Computer Communications, Elsevier, vol. 35, no. 14, pp. 1684–1694, August 2012.

[38] H. Liu, C. Maciocco, V. Kesavan, A. L. Y. Low "Energy efficient network selection and seamless handovers in mixed networks," in proc. IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks & Workshops, June 2009.

[39] I. Kanno, K. Yamazaki, Y. Ikeda, H. Ishikawa,"Adaptive energy centric radio access selection for vertical handover in heterogeneous networks," in proc. IEEE Wireless Communication and Networking Conference, April 2010.

[40] V. Rakovic, V. Atanasovski, L. Gavrilovska, "Velocity aware vertical handovers," in proc. IEEE International Symposium on Applied Sciences in Biomedical and Communication Technologies, November 2009.

[41] J.-T. Park, J.-W. Nah, S. Wang, S.-M. Chun, "Context-aware mobility management with energy efficiency for multimedia streaming service in wireless LAN," in proc. IEEE Consumer Communications and Networking Conference, January 2009.

[42] W.-H. Yang, Y.-C. Wang, Y.-C. Tseng, B.-S. P. Lin, "An energy-efficient handover scheme with geographic mobility awareness in WiMAX-WiFi integrated networks," in proc. IEEE Wireless Communications and Networking Conference, April 2009.

[43] Y. Choi and S. Choi, "Service Charge and Energy-Aware Vertical Handoff in Integrated IEEE 802.16 e/802.11 Networks," in proc. IEEE International Conference on Computer Communications (INFOCOM 2007), May 2007.

[44] G. Miao, N. Himayat, G. Li, A. Swami, "Cross-layer optimization for energy-efficient wireless communications: A survey", Wireless Communications and Mobile Computing, Hindawi – John Wiley & Sons, vol. 9, no. 4, pp. 529–542, April 2009.

[45] W. Wang, G. Shen, "Energy efficiency of heterogeneous cellular network", IEEE 72nd Vehicular Technology Conference - Fall (VTC 2010-Fall), September 2010.

[46] N. Bayer, D. Sivchenko, H. J. Einsiedler, A. Roos, A. Uzun, S. Gondor, A. Kupper, "Energy optimisation in heterogeneous multi-RAT networks", 15th International Conference on Intelligence in Next Generation Networks, pp. 139–144, October 2011.

[47] Y. S. Soh, T. Q. S. Quek, M. Kountouris, H. Shin, "Energy efficient heterogeneous cellular networks", IEEE Journal on Selected Areas in Communications, vol. 31, no. 5, pp. 840–850, May 2013.

[48] M. Wildemeersch, T. Q. S. Quek, C. H. Slump, A. Rabbachin, "Cognitive small cell networks: energy efficiency and tradeoffs," IEEE Transactions on Communications, vol. 61, no. 9, pp. 4016–4029, September 2013.

[49] Y. Song, P. Y. Kong, Y. Han, "Power-optimized vertical handover scheme for heterogeneous wireless networks", IEEE Communications Letters 2014, vol. 18, no. 2, pp. 277–280, February 2014.

[50] Jonathan Rodriguez, "Fundamentals of 5G Mobile Networks", John Wiley & Sons Ltd., May 2015

[51] J. Bastos, M. Albano, H. Marques, J. Ribeiro, J. Rodriguez, and C. Verikoukis, “Smart Interface Switching for Energy Efficient Vertical Handovers in ns-2”, IET Communications, Energy Aware Wireless Network Protocols, vol. 6, no. 14, pp. 2228–2238, September 2012.

[52] J. Bastos, J. Rodriguez, and C. Verikoukis, “Mobile Terminal Interfaces Management for Energy Efficiency”, 17th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks, September 2012.

[53] J. Bastos, M. Albano, J. Rodriguez, and C. Verikoukis, “Location Assisted Energy Efficiency for Multi-Interfaced Mobile Terminals”, 2nd IEEE Workshop on Convergence among Heterogeneous Wireless Systems in Future Internet, in association with IEEE ICC 2012, June 2012.

[54] ns-2 network simulator, URL: https://www.isi.edu/nsnam/ns/ (*accessed September 2019*)

[55] National Institute of Standards and Technology (NIST), ns-2 add-on IEEE 802.21 (draft 3) modules support, URL: https://www.nist.gov/services-resources/software/simulation-models-ns-2 (*accessed September 2019*)

[56] National Institute of Standards and Technology (NIST), “The Network Simulator ns-2 NIST add-on – IEEE 802.21 model (based on IEEE P802.21/D03.00)”, NIST, January 2007.

[57] National Institute of Standards and Technology (NIST), “The Network Simulator ns-2 NIST add-on – IEEE 802.16 model (MAC+PHY), NIST, January 2009.

[58] IEEE 802.16 WG, IEEE Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Fixed Broadband Wireless Access Systems, IEEE Std. 802.16-2004, October 2004.

[59] IEEE 802.16 WG, Amendment to IEEE Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Fixed Broadband Wireless Access Systems – Physical and Medium Access Control Layer for Combined Fixed and Mobile Operation in Licensed Bands, IEEE Std. 802.16e-2005, December 2005.

[60] National Institute of Standards and Technology (NIST), “The Network Simulator ns-2 – NIST add-on – Neighbor discovery”, NIST, January 2007.

[61] National Institute of Standards and Technology (NIST), “The Network Simulator ns-2 – NIST add-on – MAC 802.11”, NIST, January 2007.

[62] The VINT Project, “The ns Manual (formerly ns Notes and Documentation)", November 2011, URL: http://www.isi.edu/nsnam/ns/doc/ns_doc.pdf (*accessed September 2019*)

[63] V. Kakadia and W. Ye, “Energy model update in ns-2”, URL: http://www.isi.edu/ilense/software/smac/ns2_energy.html (*accessed September 2019*)

[64] F. Beaudoin, et. al., “A Fully Integrated Tri-Band, MIMO Transceiver RFIC for 802.16e”, in proc. IEEE Radio Frequency Integrated Circuits Symposium, pp. 113–116, June 2008.

[65] Y. Agarwal, R. Chandra, A. Wolman, P. Bahl, K. Chin and R. Gupta, “Wireless wakeups revisited: Energy management for VoIP over Wi-Fi smartphones”, in proc. ACM/USENIX Mobile Systems, Applications, and Services (MobiSys 2007), pp. 179–191, June 2007.

[66] G. Lampropoulos, A. Kaloxylos, N. Passas, L. Merakos, “A power consumption analysis of tight-coupled WLAN/UMTS networks”, in proc. IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications, September 2007.

[67] G. Klyne and J. Carroll, “Resource Description Framework (RDF): Concepts and Abstract Syntax”, World Wide Web Consortium, Recommendation, February 2004.

[68] D. Brickley and R. Guha, “RDF Vocabulary Description Language 1.0: RDF Schema”, World Wide Web Consortium, Recommendation, February 2004.

[69] H. Marques, J. Ribeiro, P. Marques, J. Rodriguez, “Simulation of 802.21 Handovers Using ns-2”, Journal of Computer Systems, Networks, and Communications, Hindawi, vol. 2010, Article ID 794749, June 2010.

[70] T. Camp, J. Boleng, and V. Davies, “A survey of mobility models for ad hoc network research”, Wireless Communications and Mobile Computing, Wiley, vol. 2, no. 5, pp. 483–502, September 2002.

[71] Qualcomm Atheros Inc., AR1520 (GPS chip) specifications, January 2010, URL: https://www.ath-drivers.eu/qualcomm-atheros-datasheets-for-AR1520.html

[72] J. Bastos, K. M. S. Huq, S. Mumtaz, J. Rodriguez, C. Verikoukis, “Energy efficiency optimization for downlink OFDMA system in heterogeneous network with QoS constraints”, International Journal of Communication Systems, Energy Efficient Wireless Communication Networks with QoS, January 2017.

[73] A. Antonopoulos, J. Bastos, C. Verikoukis, “Analogue Network Coding-Aided Game Theoretic Medium Access Control Protocol for Energy-Efficient Data Dissemination”, IET Science, Measurement & Technology, July 2014.

[74] D. Yang, J. Bastos, C. Verikoukis, J. Rodriguez, “Location-aided Round Robin Scheduling for Fractional Frequency Reused LTE-A Relay Network”, 17th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks, September 2012.

[75] M. De Sanctis, E. Cianca, V. Joshi, “Energy efficient wireless networks towards green communications”, Wireless Personal Communications, Springer, vol. 59, no. 3, pp. 537–552, August 2011.

[76] G. Miao, N. Himayat, Y. Li, D. Bormann, “Energy efficient design in wireless OFDMA”, in proc. IEEE International Conference on Communications 2008, May 2008.

[77] L. Chen, Y. Yang, X. Chen, G. Wei, “Energy-efficient link adaptation on Rayleigh fading channel for OSTBC MIMO system with imperfect CSIT,” IEEE Transactions on Vehicular Technology, vol. 62, no. 4, pp. 1577–1585, January 2013.

[78] Z. Li, H. Jiang, Z. Pan, N. Liu, X. You, “Energy spectral efficiency tradeoff in downlink OFDMA network,” International Journal of Communication Systems, Wiley, January 2014.

[79] D. W. K. Ng, E. S. Lo, R. Schober, “Energy-efficient resource allocation for secure OFDMA systems,” IEEE Transactions on Vehicular Technology, vol. 61, no. 6, pp. 2572–2585, July 2012.

[80] C. He, B. Sheng, X. You, “Energy and spectral efficiency trade-off for the downlink OFDM-distributed antenna systems,” International Journal of Communication Systems, Wiley, vol. 27, no. 11, pp. 3339–3351, November 2014.

[81] C. He, B. Sheng, P. Zhu, D. Wang, X. You, “Energy efficiency comparison between distributed and co-located MIMO systems,” International Journal of Communication Systems, Wiley, vol. 27, no. 1, pp. 81–94, January 2014.

[82] Z. Hasan, H. Boostanimehr, V. K. Bhargava, “Green cellular networks: A survey, some research issues and challenges,” IEEE Communications Surveys Tutorials, vol. 13, no. 4, pp. 524–540, November 2011.

[83] D. Feng, C. Jiang, G. Lim, L. J. Cimini, G. Feng, G. Y. Li, “A survey of energy-efficient wireless communications,” IEEE Communications Surveys & Tutorials, vol. 15, no. 1, pp. 167–178, First Quarter 2013.

[84] K. M. S. Huq, S. Mumtaz, J. Rodriguez, C. Verikoukis, "Investigation on Energy Efficiency in HetNet CoMP Architecture", in proc. 2014 IEEE International Conference on Communications, pp. 1112–1117, June 2014.

[85] K. M. S. Huq, S. Mumtaz, M. Alam, J. Rodriguez, R. L. Aguiar, "Frequency Allocation for HetNet CoMP: Energy Efficiency Analysis", International Symposium on Wireless Communication Systems, August 2013.

[86] J. Kim, D. Lee, W. Sung, "Interference coordination of heterogeneous LTE systems using remote radio heads," EURASIP Journal on Advances in Signal Processing, pp. 1–8, December 2013.

[87] C. Li, H. Xiong, J. Zou, Z. He, "Joint source and flow optimization for scalable video multirate multicast over hybrid Wired/Wireless coded networks," IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, no. 5, pp. 550–564, May 2011.

[88] S. Boyd, L. Vandenberghe, “Convex Optimization”, Cambridge University Press, 2004.

[89] K. M. S. Huq, S. Mumtaz, J. Bachmatiuk, J. Rodriguez, X. Wang, R. Aguiar, “Green HetNet CoMP: Energy Efficiency Analysis and Optimization,” IEEE Transactions on Vehicular Technology, vol. 64, no. 10, pp. 4670–4683, October 2015.

[90] 3GPP, “Evolved universal terrestrial radio access (E-UTRA); further advancements for E-UTRA physical layer aspects (release 9). Technical Report”, TR 36.814, V9.0.0, March 2010.

[91] A. J. Fehske, F. Richter, G. P. Fettweis, “Energy Efficiency Improvements through Micro Sites in Cellular Mobile Radio Networks”, in proc. 2009 IEEE Globecom Workshops, December 2009.

[92] R. K. Sundaram, "A First Course in Optimization Theory", Cambridge University Press, 1996.

[93] M. Paolini, “Fronthaul or backhaul for micro cells? A TCO comparison between two approaches to managing traffic from macro and micro cells in HetNets”, Technical report, Senza Fili Consulting, 2014, URL: https://senzafili.com/fronthaul-or-backhaul-for-micro-cells-form/