Enabling Energy-Efficient and Backhaul-aware Next Generation Heterogeneous Networks

Athul Prasad

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Electrical Engineering, at a public examination held at the lecture hall S1 of the school on 20 November 2015 at 12 noon.

Aalto University
School of Electrical Engineering
Department of Communications and Networking
Abstract

Heterogeneous networks have been firmly established as the direction in which next-generation cellular networks are evolving. We consider the dense deployment of small cells to provide enhanced capacity, while the macro cells provide wide area coverage. With the development of dual connectivity technology, deploying small cells on dedicated carriers has become an attractive option, with enhanced flexibility for splitting traffic within the network. The power consumption and latency requirements of the backhaul link are also gaining increasing prominence due to these factors. Backhaul link quality itself is expected to play an important role in influencing the deployment costs of next-generation 5G systems.

Energy efficiency as a network design paradigm is also gaining relevance due to the increasing impact cellular networks are having on the global carbon emission footprint. For operators, improving energy efficiency has the added advantage of reducing network operation expenditures. For the end-users, avoiding unnecessary draining of device battery power would improve the user experience.

In this work, we study energy efficient mechanisms for inter-frequency small cell discovery, based on mobility awareness and proximity estimation. Further, we apply generalized small cell discovery concepts in a device-to-device communication environment in order to optimize the energy consumption for device discovery. We also look at energy efficient small cell operations based on traffic characteristics and load constraints-based offloading in relation to the radio access and backhaul power consumption. In addition we study intelligent means of distributing delay-critical functionalities such as Hybrid ARQ, while centralizing the computationally-intense processes in a 5G, cloud-based, centralized radio access network.

Numerical evaluations done using a LTE-Advanced heterogeneous network and analytical settings indicate that significant UE and network power consumption reductions could be achieved with the considered enhancements. Using the optimized small cell operation schemes investigated in this work, reductions in network power consumption and consequent improvements in the overall energy efficiency of the network were observed. The performance of the distributed opportunistic HARQ mechanism for a centralized radio access network is compared to the optimal and static retransmission mechanisms, and the evaluated scheme is shown to perform close to the optimal mechanism, while operating with a non-ideal backhaul link.

Keywords

Small Cell Discovery, Energy Efficiency, LTE-Advanced, Cloud-based Networks, Heterogeneous Networks, Mobility Estimation, 5G, Centralized RAN
Preface

The research work for this doctoral thesis has been carried out at the department of Communication and Networking of Aalto University, Finland, Network Research Division of NEC Laboratories Europe, Germany, and at the Radio Communications Lab of Nokia, Finland.

Firstly, I would like to express my deepest gratitude to Prof. Olav Tirkkonen for his guidance, support, encouragement and infinite patience, without which this work could not have been completed. It has been a great pleasure, honor and privilege, to work with him during the course of my doctoral studies. I would also like to sincerely thank Dr. Andreas Maeder for his continuous support and enlightening discussions during my work at NEC Laboratories, and for graciously agreeing to be one of the instructors for my doctoral thesis. I would like to thank Dr. Maeder and Prof. Tirkkonen for taking the time to provide valuable comments to improve the quality of the thesis, and for helping me revise the article.

I would like to express my sincere gratitude to the pre-examiners Prof. Andrea F. Cattoni, and Prof. Ismail Guvenc, for their invaluable comments, and suggestions for modifying the content of the thesis, which has significantly helped me in revising and improving the quality of this work.

I would also like to express special thanks and gratitude to Dr. Peter Rost for his support and collaboration which helped me in completing the work for my thesis. I would like to thank Dr. Andreas Kunz, Dr. Konstantinos Samdanis, Prof. JaeSeung Song, and Genadi Velev, for the fruitful collaborations we had, which has resulted in many publications and invention disclosures.

I would like to express my gratitude towards Dr. Carl Wijting for his guidance and support during my doctoral studies. I would also like to thank him for the time and effort taken to review my thesis, and for agreeing to be the instructor for this work. I would also like to express
my sincere gratitude towards Petteri Lundén for his patience and continuous support, during the course of my work at Nokia and throughout my doctoral studies, which has led to several joint publications and invention disclosures.

I would also like to thank all my former colleagues at NEC, especially Dr. Bernd Lamparter, Dr. Hans-Joerg Kolbe, Dr. Frank Zdarsky, Prof. Tarik Taleb, Dr. Juergen Quittek, Dr. Heinrich Stuttgen, Chenghock Ng, Hans van der Veen, et al., for their support during my time at the lab. I would like to thank all my colleagues at Nokia, especially Dr. Klaus Doppler, Elena Virtej, Dr. Mikko Uusitalo, Martti Moisio, Dr. Esa Malkamaki, Dr. Kimmo Valkealahti, Dr. Mika Rinne, Karol Schober, Dr. Antti Sorri, Lars Dalsgaard, Dr. Klaus Hugl, Dr. Cassio Ribeiro, Dr. Juha Korhonen, Mauri Honkanen, Osman Yilmaz, et al., for their continuous support and encouragement.

I would like to express my sincere gratitude towards my friends and family for their continuous support, and encouragement during the course of my doctoral studies. During this work, they had continuously motivated me, and showered me with their invaluable warmth and encouragement. I would also like to especially thank my grandmother, who has supported me throughout my life, to whom I would like to dedicate this work.

Helsinki, October 13, 2015,

Athul Prasad
Contents

Preface i

Contents iii

List of Publications v

Author's Contribution vii

List of Abbreviations and Symbols xi

1. Introduction 1

1.1 Motivation ................................ 1
1.2 Problem Definition .......................... 4
1.3 Objective of the Thesis ...................... 6
1.4 Contribution and Structure of the Thesis .... 6
1.5 Summary of the Publications ............... 8

2. Background and State of the Art 11

2.1 Introduction .................................. 11
2.2 Heterogeneous Networks ...................... 12
2.3 RF Fingerprints and Device to Device Communication ... 13
2.3.1 Energy Efficiency in Heterogeneous Networks ... 14
2.3.2 Dual Connectivity .......................... 16
2.4 5G Networks ................................ 17
2.4.1 Cloud-based Centralized RAN Architecture .... 20
2.4.2 Energy Efficiency in 5G .................... 21
2.4.3 HARQ in 5G ................................ 22

3. System and Simulation Model 25

3.1 System Model .................................. 25
3.2 Simulation Setting ............................ 27
### Contents

3.3 Simulator Overview ................................................. 28
3.4 Motivation for Conducting Simulation-based Evaluations .... 30
3.5 Inter-Frequency Small Cell Discovery Modeling .............. 33
3.6 Mobility Models ..................................................... 33
3.7 Power Consumption Models ....................................... 34

4. Optimizing UE Power Consumption in Heterogeneous Small Cell Networks 39
   4.1 Introduction ...................................................... 39
   4.2 Device Speed Estimation Based Discovery .................. 39
   4.3 Received Signal Strength Based Discovery .................. 41
   4.4 Applications in D2D Discovery .............................. 44
   4.5 Performance Evaluations ...................................... 46
   4.6 Conclusion ...................................................... 53

5. Energy-Efficient Small Cell Operation 55
   5.1 Introduction ...................................................... 55
   5.2 Energy-Efficient Small Cell Operation Mechanisms ........ 56
       5.2.1 Backhaul-aware Energy Efficiency Mechanism ........ 56
       5.2.2 Reference Schemes ....................................... 58
   5.3 Simulation Results ............................................. 59
   5.4 Conclusion ...................................................... 67

6. Impact of Backhaul on the Evolution Towards 5G Systems 69
   6.1 Introduction ...................................................... 69
   6.2 Problem Description ........................................... 70
   6.3 Reference Schemes ............................................. 70
   6.4 Opportunistic HARQ for non-Ideal Backhaul ................ 72
   6.5 Simulation Results ............................................. 75
   6.6 Conclusions ...................................................... 76

7. Conclusions and Future Work 77

References ................................................................. 81

Errata ................................................................. 95

Publications ............................................................. 97
This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.


V A. Prasad, K. Samdanis, A. Kunz, and J. Song. Energy Efficient Device Discovery for Social Cloud Applications in 3GPP LTE-Advanced Net-


Author’s Contribution

Publication I: “Energy-Efficient Flexible Inter-Frequency Scanning Mechanism for Enhanced Small Cell Discovery”

The author is the main contributor of this work, and has developed the flexible inter-frequency scanning mechanism, with support from the other authors in the design and implementation of the simulation framework, as well as the analytical evaluation framework.

Publication II: “Energy Efficient Small-Cell Discovery Using Received Signal Strength Based Radio Maps”

The author is the main contributor of this work, and has developed the received signal strength based radio maps for inter-frequency small cell discovery concept, as well as the simulation and analytical evaluation framework, with support from the other authors.

Publication III: “Energy Efficient Inter-Frequency Small Cell Discovery Techniques for LTE-Advanced Heterogeneous Network Deployments”

The author is the main contributor of this work, and has further developed the received signal strength and flexible inter-frequency small cell discovery mechanisms, with support from the other authors. The mechanisms were implemented in a simulation framework, using constraints defined for LTE-Advanced heterogeneous network systems, along with the relaxed inter-frequency scanning mechanism proposed for such systems.
Publication IV: “Enhanced Small Cell Discovery in Heterogeneous Networks Using Optimized RF Fingerprints”

The author is the main contributor of this work, and enhanced the optimized RF fingerprint mechanism for UE-based small cell discovery. The author also implemented the proposed mechanism and the reference schemes in a simulation framework based on LTE-Advanced system settings, with support from the other authors.

Publication V: “Energy Efficient Device Discovery for Social Cloud Applications in 3GPP LTE-Advanced Networks”

The author is the main contributor of this work, and has enhanced the RF fingerprint based localization concept for application in the Device-to-Device discovery process in a social cloud environment. The author also implemented the mechanism along with the reference schemes in a simulation framework, with support from the other authors.

Publication VI: “Energy Efficient Small Cell Activation Mechanism for Heterogeneous Networks”

The author is the main contributor of this work, and has developed the small cell activation mechanism along with the analytical and evaluation framework, with support from the other authors.

Publication VII: “Backhaul-aware Energy Efficient Heterogeneous Networks with Dual Connectivity”

The author is the main contributor of this work, and has enhanced the small cell activation mechanism to include backhaul power consumption awareness. The author also developed and implemented the analytical and simulation framework for evaluations along with the reference schemes, with support from the second author.
Publication VIII: “Opportunistic Hybrid ARQ – Enabler of Centralized-RAN over Non-Ideal Backhaul”

The author has played a major role in developing the basic idea of the article, and was also involved in drafting the article, and developing the simulation framework for the opportunistic Hybrid ARQ concept.
List of Abbreviations and Symbols

Abbreviations

2G  Second Generation
3G  Third Generation
3GPP 3rd Generation Partnership Project
4G  Fourth Generation
5G  Fifth Generation
ACK Acknowledgment
AECID Adaptive Enhanced Cell-ID
AMBR Adaptive Maximum Bit Rate
AP Access Point
ASF Autonomous Search Function
BH Backhaul
BIM Background Inter-frequency Measurement
bpcu bits per channel use
BS Base Station
BW Bandwidth
CAPEX Capital Expenditure
CB Cloud Based
CBR Constant Bit Rate
CDF Cumulative Distribution Function
CN Core Network
CoMP Coordinated Multi-Point transmission and reception
CRC Cyclic Redundancy Check
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS</td>
<td>Compensation</td>
</tr>
<tr>
<td>CSG</td>
<td>Closed Subscriber Group</td>
</tr>
<tr>
<td>CSI</td>
<td>Channel State Information</td>
</tr>
<tr>
<td>D2D</td>
<td>Device-to-Device</td>
</tr>
<tr>
<td>DD</td>
<td>Direct Discovery</td>
</tr>
<tr>
<td>DL</td>
<td>Downlink</td>
</tr>
<tr>
<td>DTX</td>
<td>Discontinuous Transmission</td>
</tr>
<tr>
<td>eICIC</td>
<td>enhanced Inter Cell Interference Coordination</td>
</tr>
<tr>
<td>eNBs</td>
<td>eNodeB</td>
</tr>
<tr>
<td>eNodeB</td>
<td>Evolved Node-B</td>
</tr>
<tr>
<td>ENS</td>
<td>Expression Name System</td>
</tr>
<tr>
<td>EPC</td>
<td>Evolved Packet Core</td>
</tr>
<tr>
<td>ES</td>
<td>Energy Saving</td>
</tr>
<tr>
<td>E-UTRAN</td>
<td>Evolved-UMTS Terrestrial Radio Access Network</td>
</tr>
<tr>
<td>FA</td>
<td>Fiber Access</td>
</tr>
<tr>
<td>FDD</td>
<td>Frequency Division Duplexing</td>
</tr>
<tr>
<td>FEC</td>
<td>Forward Error Correction</td>
</tr>
<tr>
<td>FP</td>
<td>Fingerprint</td>
</tr>
<tr>
<td>GBR</td>
<td>Guaranteed Bit Rate</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>GW</td>
<td>Gateway</td>
</tr>
<tr>
<td>HARQ</td>
<td>Hybrid Automatic Repeat Request</td>
</tr>
<tr>
<td>HCS</td>
<td>Handover Candidate Selection</td>
</tr>
<tr>
<td>HeNB</td>
<td>Home eNodeB</td>
</tr>
<tr>
<td>HetNet</td>
<td>Heterogeneous Network</td>
</tr>
<tr>
<td>HO</td>
<td>Handover</td>
</tr>
<tr>
<td>HSS</td>
<td>Home Subscriber Server</td>
</tr>
<tr>
<td>ICIC</td>
<td>Inter Cell Interference Coordination</td>
</tr>
<tr>
<td>ICT</td>
<td>Information and Communication Technology</td>
</tr>
<tr>
<td>IMS</td>
<td>IP Multimedia Subsystem</td>
</tr>
<tr>
<td>IMT-A</td>
<td>International Mobile Telecommunications - Advanced</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>ISD</td>
<td>Inter-Site Distance</td>
</tr>
</tbody>
</table>
ITU  International Telecommunication Union
i.i.d. independent and identically distributed
LSC  Location Services
LTE  Long Term Evolution
LTE-A Long Term Evolution - Advanced
M2M  Machine-to-Machine
MAC  Medium Access Control
MBR  Maximum Bit Rate
MCS  Modulation and Coding Scheme
MDT  Minimization of Drive Test
MeNB  Master eNB
MIMO Multiple Input Multiple Output
MME  Mobility Management Entity
MS  Mobile Station
MSE  Mobility State Estimation
MTC  Machine Type Communication
NACK  Negative Acknowledgment
NB  Network Based
NLOS  Non-Line Of Sight
OAM  Operations, Administration and Maintenance
OFDMA  Orthogonal Frequency Division Multiple Access
OPEX  Operational Expenditure
PDCP  Packet Data Convergence Protocol
PDSCH  Physical Downlink Shared Channel
PF  Proportional Fair
PHY  Physical Layer
PL  Pathloss
PRB  Physical Resource Block
PUSCH  Physical Uplink Shared Channel
P-GW  Packet Data Network - Gateway
QCI  QoS Class Identifier
QoS  Quality of Service
RB  Resource Block
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>Radio Frequency</td>
</tr>
<tr>
<td>RLC</td>
<td>Radio Link Control</td>
</tr>
<tr>
<td>RLF</td>
<td>Radio Link Failure</td>
</tr>
<tr>
<td>RMS</td>
<td>Root Mean Square</td>
</tr>
<tr>
<td>RN</td>
<td>Relay Nodes</td>
</tr>
<tr>
<td>RR</td>
<td>Round Robin</td>
</tr>
<tr>
<td>RRC</td>
<td>Radio Resource Control</td>
</tr>
<tr>
<td>RRH</td>
<td>Remote Radio Head</td>
</tr>
<tr>
<td>RRM</td>
<td>Radio Resource Management</td>
</tr>
<tr>
<td>RSRP</td>
<td>Reference Signal Received Power</td>
</tr>
<tr>
<td>RSS</td>
<td>Received Signal Strength</td>
</tr>
<tr>
<td>SB</td>
<td>Symmetric Backhaul</td>
</tr>
<tr>
<td>SCANA</td>
<td>Social Cloud Application Network Area</td>
</tr>
<tr>
<td>SCAS</td>
<td>Social Cloud Application Server</td>
</tr>
<tr>
<td>SeNB</td>
<td>Secondary eNB</td>
</tr>
<tr>
<td>SFC</td>
<td>Space Frequency Coding</td>
</tr>
<tr>
<td>SIB</td>
<td>System Information Block</td>
</tr>
<tr>
<td>SINR</td>
<td>Signal to Interference Noise Ratio</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to Noise Ratio</td>
</tr>
<tr>
<td>SON</td>
<td>Self-Organizing Network</td>
</tr>
<tr>
<td>STS</td>
<td>Short Time-of-Stay</td>
</tr>
<tr>
<td>S-GW</td>
<td>Serving Gateway</td>
</tr>
<tr>
<td>TCE</td>
<td>Trace Collection Entity</td>
</tr>
<tr>
<td>TDD</td>
<td>Time Division Duplexing</td>
</tr>
<tr>
<td>ToS</td>
<td>Time-of-Stay</td>
</tr>
<tr>
<td>TTI</td>
<td>Transmit Time Interval</td>
</tr>
<tr>
<td>TTT</td>
<td>Time-to-Trigger</td>
</tr>
<tr>
<td>UDN</td>
<td>Ultra-Dense Network</td>
</tr>
<tr>
<td>UE</td>
<td>User Equipment</td>
</tr>
<tr>
<td>UL</td>
<td>Uplink</td>
</tr>
<tr>
<td>UMTS</td>
<td>Universal Mobile Telecommunications System</td>
</tr>
<tr>
<td>VHD</td>
<td>Vertical Handover Decision</td>
</tr>
<tr>
<td>VPN</td>
<td>Virtual Private Network</td>
</tr>
</tbody>
</table>
WiFi  Wireless Fidelity
WiMAX  Worldwide Interoperability for Microwave Access
WL  Wireless
WLAN  Wireless Local Area Network
Wrd  Wired

Symbols

\( d_{\text{mean}} \)  mean distance traversed by a UE within a small cell
\( E_r \)  Expected rate
\( h_t \)  i.i.d. random channel fading process
\( L_M \)  Distance dependent path-loss from macro BS to mobile terminal
\( L_P \)  Distance dependent path-loss from pico BS to mobile terminal
\( N \)  Blocklength
\( N_A \)  Number of antennas per cell
\( N_{\text{MS}} \)  Number of mobile stations
\( N_P \)  Number of picos
\( N_{\text{RB}} \)  Number of resource blocks
\( N_{\text{UE}} \)  Number of UEs
\( P_0 \)  Power consumption at zero RF output power
\( P_{\text{bh_a}} \)  Active mode backhaul power consumption
\( P_{\text{bh_s}} \)  Sleep mode backhaul power consumption
\( P_C \)  BS Power consumption
\( P_m \)  Maximum RMS transmit power of the BS
\( P_S \)  Sleep mode power consumption
\( \Pr_e \)  Error probability
\( \Pr_{\text{target}} \)  Maximum error probability constraint
\( \Pr_{\text{th}} \)  Error probability threshold
\( r_d \)  Small cell radius
\( R_{\text{eff}} \)  Expected effective rate
\( r_{\text{fp}} \)  Radius of fingerprint matching region
\( R_{\text{Loss}} \)  Offloading loss ratio
\( R_M \)  Distance between macro BS and mobile terminal
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_0$</td>
<td>Cut-off rate</td>
</tr>
<tr>
<td>$R_P$</td>
<td>Distance between pico BS and mobile terminal</td>
</tr>
<tr>
<td>$R_{\text{max}}$</td>
<td>Maximum user data rate</td>
</tr>
<tr>
<td>$R_{\text{min}}$</td>
<td>Minimum user data rate</td>
</tr>
<tr>
<td>$R_{B_s}$</td>
<td>Resource block size</td>
</tr>
<tr>
<td>$S_{\text{eff}}$</td>
<td>Spectral efficiency</td>
</tr>
<tr>
<td>$T$</td>
<td>Number transmissions per codeword</td>
</tr>
<tr>
<td>$T_d$</td>
<td>Small cell switch-off delay</td>
</tr>
<tr>
<td>$T_{\text{Inter}}$</td>
<td>Inter-frequency measurement period</td>
</tr>
<tr>
<td>$T_{\text{Inter,min}}$</td>
<td>Minimum inter-frequency measurement period</td>
</tr>
<tr>
<td>$v$</td>
<td>Mobile terminal speed</td>
</tr>
<tr>
<td>$v_t$</td>
<td>Threshold mobile terminal speed</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Backhaul sleep mode power consumption factor</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Signal-to-noise ratio of one transmission</td>
</tr>
<tr>
<td>$\gamma_{\text{eff}}$</td>
<td>Effective signal-to-noise ratio</td>
</tr>
<tr>
<td>$\gamma_t$</td>
<td>Signal-to-noise ratio of transmission index $t$</td>
</tr>
<tr>
<td>$\delta_{\text{app}}$</td>
<td>D2D application mismatch error</td>
</tr>
<tr>
<td>$\delta_c$</td>
<td>RSRP correction factor</td>
</tr>
<tr>
<td>$\delta_{fp}$</td>
<td>RSRP offset for fingerprint match</td>
</tr>
<tr>
<td>$\delta_{\text{off}}$</td>
<td>Small cell traffic offloading factor</td>
</tr>
<tr>
<td>$\Delta_p$</td>
<td>Slope of the load dependent power consumption</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>Error probability</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Load factor</td>
</tr>
<tr>
<td>$\sigma_x^2$</td>
<td>Variance of Gaussian random variable</td>
</tr>
<tr>
<td>$\psi$</td>
<td>Energy efficiency</td>
</tr>
</tbody>
</table>
1. Introduction

1.1 Motivation

Mobile data rate requirements are increasing at a very fast pace, due to the ubiquitous use of applications that require significantly high data rates [133, 167]. An indication of this trend is shown in Fig. 1.1, with mobile traffic data rates expected to be up to \( \text{25 Exabytes per month} \) [59]. In addition, users are increasingly expecting to have the same browsing experience while using mobile and wired broadband connections [19]. With the current exponential increase in data rate demands expected to continue, operators and vendors have been investigating mechanisms that will improve data rates, satisfy user demand [41, 92], and minimize impacts on the user experience and network signaling load. Heterogeneous networks, with macro cells providing wide area coverage and small cells (such as remote radio heads (RRH), relays, and pico or femto cells) providing coverage and capacity enhancements at relevant hotspot locations are increasingly proving to be an attractive deployment option for network operators [78, 94].

Recent estimates have shown that the contribution of information and communication technology (ICT) to global greenhouse gas emissions is approximately 2% [13, 96]. Based on the survey done in [115], approximately 10-15% of total network operating expenses (OPEX) is due to energy consumption. In developing markets it can be up to 50%. Since this analysis is mainly based on existing networks, it could indicate the impact of macro cell power consumption. It is also assumed that the energy consumption includes backhaul power consumption.

As network density increases because of densely deployed small cells, the network energy consumption also increases, due to the higher num-
Introduction

Figure 1.1. Overall mobile traffic data growth [59].

ber of active nodes in the system [58,135]. Improving the energy efficiency of such networks could involve the efficient use of the energy consumed for network operations, rather than merely reducing the total energy consumption. Optimal use of network resources can be achieved when the traffic demands of the network are taken into account, while optimizing the amount of active nodes in the system [35,56,108,138]. Deactivating cells with limited or no traffic could also reduce the overall interference in the network [7,34]. The anticipated network densification of future wireless networks means that mechanisms for energy-efficient operation of network nodes are an important area of study. The study of backhaul power consumption is also receiving increasing attention, both in academia and the industry [22,101,102]. Energy-efficient techniques that consider both radio access and backhaul link power consumption could be a relevant area of study [140].

With the advent of cloud technologies, it is also proposed to have a cloud-based, centralized radio access network (RAN) architecture in fifth generation (5G) networks [118, 147, 165, 171]. In essence, densely-deployed small cells with basic physical layer capabilities are connected using low-delay, high capacity backhaul with a centralized server using general purpose hardware, that could handle complex physical layer processing and radio resource management (RRM). Various implementation mechanisms are currently being studied [87] for enabling such architectures. Even with such an architecture, having legacy macro cells for ensuring wide area coverage within the cell area could be possible using the dual connectivity mechanism [10,38].

Mobile devices are currently some of the most widely-used electronic de-
Introduction

Figure 1.2. Mobile operator network OPEX distribution in European markets [115].

Vices, with active subscriptions estimated at over 7 billion worldwide [54]. While such devices consume significantly less power than the network infrastructure, due to the economy of scale involved, optimizing device power consumption while ensuring connectivity and avoiding negative impacts on capacity, data traffic, and user experience is still an important consideration. It is essential to identify the functions designed for homogeneous macro-only network deployments, which need to be optimized for the efficient operation in heterogeneous networks.

Since radio access networks are evolving toward higher density in terms of network nodes and UEs, new RAN architectures, etc., the power consumption problem will only become more severe. Investigating methods of improving the overall energy efficiency of such systems is therefore becoming increasingly relevant and important. In this case, 'overall energy efficiency' implies the optimal operation of all the nodes within the access network with no adverse impact on network capacity. Developing such methods, while considering network implementation and deployment constraints, has been the main motivation behind this work.
1.2 Problem Definition

Heterogeneous small cell networks face various challenges that can be broadly classified as self-organizing, backhauling, interference management, and handover or mobility management challenges [92]. The challenges most directly related to power consumption are self-organization, backhauling, and handover or mobility management, as highlighted in Fig. 1.3. Self-organization is relevant because of the power consumption involved in small cell operation, which needs to be optimized. Backhauling involves network and UE power consumption aspects. Network impacts from power consumption involved in the overall backhaul operation need to be coordinated with the network self-organizing functions to enable power savings. UE power consumption is relevant because of the possible small cell optimizations that could be done to reduce the delay constraints on the backhaul, which could consume additional UE battery power. Handover or mobility involves the discovery of small cells deployed in a dedicated frequency band, which could consume considerable UE battery power.

In terms of the nodes involved in a heterogeneous network, it is assumed that the macro cells are deployed after careful network planning operations. It is assumed that wide area coverage is essential and cannot be compromised. The small cells and related backhaul links and the UEs are the nodes which can be operated in an optimized manner to increase power savings. While the contribution due to small cell operation (defined here as the joint operation of the small cell node and its backhaul link) on the operational expenditure of a network operator is evident, the question
arises as to whether the UE operation for handover or mobility management in a network is significant. To highlight this fact, an illustration is made in Fig. 1.4 to indicate the UE battery power consumed for small cell discovery using two different inter-frequency scanning periods. It is assumed that one inter-frequency measurement is made within the scanning period for a time duration of 6 ms, consuming 0.5 W of battery power. Assuming approximately 8 kJ of UE battery is available for cellular radio operation, from the figure we can observe that more than 81% of the UE battery power would be consumed for the small cell discovery operation alone, during the period of one day. The detailed assumptions used for this illustration can be found in Publication III.

There are three key tradeoffs considered in this work. The first tradeoff is between optimizing the UE power consumption for inter-frequency small cell discovery and ensuring that there is a limited impact on the UE throughput by being able to connect to the small cells. Here small cells are assumed to be deployed at capacity hotspots and have low load, thereby offering significantly high throughput for the UE. The second tradeoff is between optimizing the power consumed for small cell operation and ensuring that there is a limited impact on the network capacity. Reducing the small cell operation time would reduce the energy consumed by the network, but would negatively impact the network capacity as well. The third tradeoff is between improving the network spectral efficiency and capacity using a hybrid automatic repeat request (HARQ) mechanism in a network employing centralized RAN architecture with non-ideal backhaul and low-cost RRHs, and optimizing the UE power consumption by avoiding unnecessary uplink retransmissions.

Figure 1.4. UE battery consumption per day due to different inter-frequency scanning periods.
Taking all of the above into account, the key question addressed by this work is: How to reduce the power consumption for the nodes involved in the heterogeneous small cell network, with relaxed operational constraints, and without significant tradeoffs in terms of network capacity?

1.3 Objective of the Thesis

The main objective of the thesis is to propose new mechanisms which could address the key tradeoffs, while staying within the practical constraints involved in real network deployments. The performances of the proposed mechanisms are then evaluated using system-level simulations, emulating real network characteristics. For the first tradeoff, the UE power consumption for discovery is calculated, along with the related impacts on the connected time. Here the connected time quantifies the amount of traffic offloading or capacity enhancements that can be achieved using the enhancements, relative to the reference mechanisms. For the second tradeoff, the network power consumption using the optimized small cell operation mechanism, and impacts on mean throughput relative to different reference mechanisms are evaluated. For the third tradeoff, the achievable uplink throughput using the proposed enhancement in the constrained system is compared with the optimal reference scheme.

1.4 Contribution and Structure of the Thesis

There are three main contributions made in this thesis, related to the considered tradeoffs. First, the mechanisms developed for optimizing the UE power consumption for conducting inter-frequency discovery operation are presented. Then, an energy-based traffic offloading and small cell activation scheme is proposed for optimizing the network power consumption, including small cells and the related backhaul link. Finally, the mechanism for optimizing the small cell HARQ operation without significantly affecting the UE uplink power consumption and throughput is evaluated, while enabling deployments using non-ideal backhaul in a centralized cloud-based RAN.

The inter-frequency discovery mechanisms described in this work can be used by network operators to deploy inter-frequency small cells and configure appropriate measurement gap periods based on the UE speed.
estimation, resulting in minimal power consumption to the UE. The impact to the network in terms of frequent handovers with short connection times can also be avoided, thereby avoiding a higher signaling load towards the core network. Network operators could also reuse the radio environment map built for the macro cells, using the drive test features to estimate the proximity of the UE to small cells before initiating the small cell discovery procedure. These methods could also be combined to filter out connections from fast-moving UEs while saving valuable UE power. The network operators could also use similar techniques for device-to-device (D2D) discovery procedures, if the discovery beacons are sent on a dedicated frequency layer. Since it was shown that overhead for having the fingerprint database on a cloud server is minimal from the UE perspective, third party application providers can also use such mechanisms to provision commercial services using D2D.

Based on the energy-efficient small cell and backhaul operation mechanism, network operators can deploy such cells with high density, by having a coordinated operation of both entities. Since the mechanism for cell activation is distributed at a macro-cellular level, such deployments can operate with minimal operational overhead. Cell activation and deactivation signaling are already standardized as part of Long Term Evolution (LTE)-Advanced networks, so the mechanism proposed here could be deployed on existing networks with minimal overhead. In order to accurately estimate the small cells that need to be activated, an efficient small cell proximity estimation mechanism is required. It is thus assumed that the fingerprint mechanism for small cell discovery could be used as a baseline for the energy efficient small cell operation.

The work in this thesis also elaborates a mechanism for operating low-cost small cells or remote radio heads in which network complexity is transferred to a centralized, virtual base station pool, using a non-ideal, high-latency backhaul link between the RRHs and the centralized entity. Since non-ideal backhaul is more cost-efficient, the overall reduction of capital expenditures (CAPEX) could be one of the main motivating factors behind such network deployments. The mechanism operates with minimal impact on UE uplink transmission power consumption and achievable throughput rates. The main contributions of such a mechanism would be the possibility to ultra-densify the network without significant capital expenditure and to operate the network without negatively impacting the UE battery power consumption.
The thesis is organized as follows. The background of the work and state-of-the-art description based on the literature review is discussed in Chapter 2. The overall system, simulation, mobility and power consumption models used are discussed in Chapter 3. The discovery mechanisms which optimize the UE power consumption in heterogeneous networks are presented in Chapter 4. The techniques for the energy-efficient operation of a heterogeneous small cell network, with detailed performance evaluations are presented in Chapter 5. The impact of backhaul on the evolution towards 5G cloud-based systems, and a mechanism for operating such systems using non-ideal backhaul is presented in Chapter 6. Finally, Chapter 7 concludes the work, and outlines the possible areas of future work.

1.5 Summary of the Publications

The thesis consists of an introduction, followed by eight original publications.

Publication I – Publication IV deals with the first tradeoff of optimizing UE power consumption for inter-frequency small cell discovery. In Publication V, the applicability of such a tradeoff for D2D discovery, which follows similar discovery principles, is investigated. Publication VI – Publication VII deals with the second tradeoff of optimizing small cell power consumption, while ensuring limited impacts on the network capacity. Here the need for jointly optimizing the radio access and backhaul link power consumption is also studied. Publication VIII is related to the third tradeoff between the small cell operation in a constrained environment with non-ideal backhaul and resultant additional UE power consumption due to HARQ retransmissions.

In Publication I, the UE speed and fingerprint-based small cell discovery mechanisms were implemented using the constraints defined in the 3GPP LTE-Advanced heterogeneous network environment. Further, a relaxed measurement scheme, called the background inter-frequency cell search mechanism, was also considered and the performance of each in terms of power consumption and small cell connected time were presented. Optimizations in terms of handover signaling required for each mechanism were also presented. When compared to current cell search mechanisms, the obtained results show that the optimized schemes enable significant power savings, and handover signaling optimizations, all without compro-
mising the small cell connected time.

In Publication II, we consider the use of received signal strength or radio frequency (RF) fingerprint-based small cell proximity detection and discovery for reducing the power consumption associated with inter-frequency cell searches. A dynamic fingerprint-matching region is also presented that takes UE speed into account, excluding fast-moving UEs from initiating small cell search. Simulations conducted in a heterogeneous network environment indicate that significant energy savings can be obtained with some reduction in small cell connected time.

In Publication III, the UE speed and fingerprint based small cell discovery mechanisms were implemented using the constraints defined in the 3GPP LTE-Advanced heterogeneous network environment. Further, a relaxed measurement scheme called the background inter-frequency cell search mechanism was also considered and the performance of each in terms of power consumption and small cell connected time were presented. Optimizations in terms of handover signaling required for each mechanism was also presented. When compared to current cell search mechanisms the obtained results show that the optimized schemes enable significant power savings, and handover signaling optimizations, all without compromising the small cell connected time.

In Publication IV, the cost of sending measurement reports by the UEs for estimating small cell proximity by the network is studied and compared with the mechanism in which optimized fingerprint information is sent directly to the UE and stored on the device. The performance of the optimized fingerprint mechanism is compared with the network-based fingerprint scheme in terms of small cell connected time and power consumption. The simulation results indicate that significant energy savings can be obtained by using the optimized fingerprint mechanism, while further reducing the signaling load related to measurement reporting.

In Publication V, energy-efficient device discovery using an ideal network or cloud-based fingerprint information is evaluated in a D2D environment. The additional energy consumption for having the fingerprint database in the cloud is analyzed using realistic UE power consumption models. Based on the evaluations done, it is shown that the fingerprint mechanism can also be implemented in the cloud without significant additional power consumption, while enabling commercial deployment of D2D communication.

In Publication VI, a fully-distributed small cell activation mechanism
is presented, which takes energy-saving gains based on offloaded traffic from macro to small cells into account when making cell activation decisions. Information elements that need to be communicated over the interface between cells are also discussed. The scheme is evaluated using LTE-Advanced heterogeneous network settings employing different traffic models, and energy saving results are presented.

In Publication VII, the load-based small cell activation mechanism is further extended to take backhaul link power consumption into consideration and is evaluated in an LTE-Advanced heterogeneous network with dual connectivity architecture. In this system, mechanisms for dynamic traffic offloading with control and user plane separation are analyzed under mobility constraints and with indoor and outdoor deployment of small cells. Based on the evaluations done, the backhaul-aware mechanism is shown to provide significant energy saving gains, with acceptable trade-offs in terms of user throughput gains.

In Publication VIII, an opportunistic HARQ mechanism is considered. The mechanism estimates the decoding error probability and enables cloud-based RAN using a non-ideal backhaul. Using the mechanism, RRHs can directly send the feedback for HARQ while still centralizing the decoding process in the cloud by implementing the delay-critical component at the radio access point. The performance of the mechanism is compared with a fixed re-transmission and optimal HARQ and is shown to perform well, especially compared to the optimal scheme.
2. Background and State of the Art

2.1 Introduction

In order to provide significant improvements to the user experience and to enhance the support of a multitude of mobile services and applications, 3GPP introduced LTE as part of release 8 specifications in 2008 [23]. Compared to the previous generation of communication standards, changes were made in the air interface as well as the core network in order to satisfy the challenging requirements defined by 3GPP. Some of the key requirements include: i) 100 Mbps in downlink, 50 Mbps in uplink, ii) higher spectral efficiency and bit rates, iii) lower round-trip time, iv) scalable bandwidth, v) support for time and frequency division duplexing (TDD and FDD), as well as vi) higher cost efficiency [1, 52]. Performance evaluation of the release 8 LTE systems, with features such as multi-antenna transmissions, inter-cell interference coordination and flexible spectrum, indicated that all the requirements set by 3GPP were achieved [16].

As data rate requirements for mobile users continued to increase over time, new requirements for mobile communication systems were specified as part of International Mobile Telecommunications-Advanced (IMT-A) system requirements [158]. Key among them were: extended bandwidth capabilities of up to 40 MHz; higher peak spectral efficiency in uplink and downlink; and lower control and user plane latency. LTE-Advanced systems were specified by 3GPP as part of release 10 standards, in order to satisfy these enhanced requirements, with peak data rate support of up to 1 Gbps [8, 120]. The above-mentioned requirements were achieved with: enhancements in the spectrum supported, using the carrier aggregation feature; enhancements in multi-antenna techniques; and most im-
Background and State of the Art

importantly, by providing improved support for heterogeneous network deployments.

2.2 Heterogeneous Networks

One of the main drivers behind improving data rates for end users has been network densification, which is enabled by heterogeneous network deployments. A heterogeneous network, consisting of a macro cell for wide area coverage and small cells for coverage and capacity improvements, can be broadly classified into two types, based on the frequency reuse mechanism employed: co-channel deployments with macro and small cells deployed in the same frequency layer; or inter-frequency deployments in which small cells are deployed in a dedicated frequency layer. The co-channel deployments are considered to be key enablers for improving spectral efficiency and providing coverage enhancements. But such deployments face significant challenges in interference coordination and mobility, due to co-channel interference between the small and macro cells [91, 93, 123]. Inter-frequency deployment of heterogeneous networks is also an attractive solution, especially for providing capacity enhancements [122]. Several current works look into efficient frequency reuse and the use of the carrier aggregation feature within such deployments [48, 63, 127]. In this work it is assumed that the dedicated frequency carrier allocated for the small cell is fully reused. The possible roles of the UE in such an environment are evaluated in terms of radio link monitoring, radio resource management, channel state information (CSI) feedback, detection of weak cells and interference calculation, and are presented in [40].

While heterogeneous networks enable improvements in user data rates, there are some key technical challenges faced in the operation of such networks. As discussed earlier, these include self-organization, inter-cell interference, handover or mobility performance, and backhauling [92]. Since heterogeneous networks involve random deployment of base stations, network performance depends on the self-organization features of the cells [32, 125]. Self-organization of a heterogeneous network involves the three key processes of self-configuration, self-healing and self-optimization, altogether making it a complex task with a significant number of radio parameters to consider. There has been increased focus from industry and academia on this topic, investigating optimizations related to the various radio parameters involved [43, 51, 168].
The power difference between macro and small cells leads to small cells having a significantly smaller coverage footprint, compared to macro cells. Because of this, UEs tend to connect to macro cells, even when small cells with lower load conditions might be available. To overcome this limitation, the cell range expansion (CRE) feature was standardized in LTE systems, whereby UEs connect to small cells which are weaker by a CRE offset [109,148]. For co-channel deployments, this complicates the issue of inter-cell interference [94,149,156] due to UEs connecting to weaker small cells, which leads to the need for time and frequency domain interference mitigation techniques. But for inter-frequency deployment scenarios, it provides the added opportunity of expanding the small cell coverage footprint without increasing the interference between macro and small cells.

Handover or mobility performance is yet another key issue facing such networks, because of higher base station density, which leads to a higher signaling load towards the core network. Because of this, various solutions, including UE mobility state-based enhancements were studied to overcome such challenges [21, 57, 91]. The mobility performance issues for co-channel deployments could lead to service interruptions caused by radio link failures (RLFs) from handovers that are too late or too early. Thus, mobility performance issues require special handling in such deployment scenarios. A survey of the different vertical handover decision algorithms in 4G heterogeneous networks is presented in [161]. Mobility enhancements for heterogeneous networks with inter-site carrier aggregation are presented in [122], which considers UE autonomous mobility enhancements. Backhaul network design is a relevant and important issue for heterogeneous networks, with the network QoS and user experience intricately linked to the quality of the backhaul [83].

2.3 RF Fingerprints and Device to Device Communication

Received Signal Strength (RSS)-based location estimation in mobile networks has been widely studied in the literature [64, 82, 139] and [146]. The taxonomy of RF fingerprinting mechanisms for location estimation is studied in [80]. But the application of such mechanisms for discovering inter-frequency small cells, and related power consumption aspects have received limited attention. The main motivation behind the application of such mechanisms for small cell discovery was the work done on mobility enhancements for 3GPP LTE-Advanced networks [3]. As part of the work,
energy efficient discovery of small cells deployed on different carriers for offloading purposes using network- and UE-assisted enhancements were investigated.

Device to Device communication for public safety and commercial use is actively being studied in the industry, as part of the 3GPP LTE release 12 standardization efforts. There are several works available in the literature investigating D2D discovery procedures [18,49,75,77]. But limited attention has been given to D2D proximity estimation using network or third-party cloud application server-based approaches. Also, currently available mechanisms do not address the challenge of energy-efficient D2D discovery if the discovery beacons are sent in a different frequency carrier. Such multi-carrier deployments are considered to be one of the main future enhancements for this technology [114].

2.3.1 Energy Efficiency in Heterogeneous Networks

While there are clear advantages to having inter-frequency deployments for capacity enhancements, one basic challenge would be the efficient discovery of small cells that are deployed in a frequency band different from that of the macro cell. The legacy inter-frequency cell search mechanism has been optimized for homogeneous networks, with the UE initiating an event-based cell search based on the serving cell signal strength [145]. In heterogeneous networks, small cells can be present in the proximity of a UE, even while the serving macro cell signal strength is high and so energy-efficient mechanisms that reduce UE battery power consumption for discovering small cells while avoiding unnecessary handover and measurement reporting-related signaling are required. Thus, optimizing the UE power consumption needed for discovering inter-frequency small cells, while minimizing related signaling, was defined as one of the focus areas of this work. In addition, because of the rise of D2D technology [50,128,172], which has a discovery process that can use mechanisms similar to those of small cell discovery, optimizing UE power consumption for D2D discovery with small cell discovery principles was also studied.

The work done in [97] provides a detailed overview of the impact of network upgrades on the overall power consumption and energy efficiency. Based on the analysis done, it was shown that having remote radio heads supporting multiple carriers offers significantly high energy efficiency, while the relative gains from having a dense deployment of small cells are low. However, based on the evaluations done, it was also concluded that
upgrading the existing base stations to the latest LTE-Advanced models and using heterogeneous small cell deployment provides the best trade-off between network performance and power consumption. Also, the importance of considering the backhaul power consumption along with the radio access network operation is elaborated. A detailed analysis of the energy-saving techniques and impacts on the energy consumption of actual network deployments is presented in [126]. From a small cell power consumption perspective, it was shown that for the considered deployment scenario, even with an 8 hour off-time for the small cells, the reduction in energy consumption is only 1.3% of the total network power consumption per year (approximately 2.8 GWh/year). While the work gives a clear indication of the technology potential of small cell deactivation, it considers the activation or deactivation of small cells based on the macro cell load, not on the actual proximity estimation assumptions and energy-based traffic offloading, which is considered in this work.

The work done in [17] evaluates the active and sleep mode power consumption profiles of LTE eNBs, which are used in this work as well. Similar power consumption models have also been evaluated in [36,46,60,131]. Most of the work currently available in the literature focuses on macro cell power consumption optimizations [25,31], with fundamental trade-offs involved in the operation of such networks evaluated in detail in [29]. The work done in [14,103,150] deals with the optimization of small cell power consumption in a heterogeneous network, mainly focusing on optimizing radio access power consumption, while not taking the backhaul link power consumption into account.

In [151,152], the impact of ‘ideal’ fiber optic backhaul links on the overall network power consumption, and the impact of various backhaul types on the operation of heterogeneous networks are also evaluated, mostly with full load assumptions. The mobile backhaul initiative of the broadband forum is investigating various techniques for energy-efficient backhaul architectures to be integrated with radio access networks [22,141]. The studies being done take into account different types of backhaul links. Currently, 3GPP is also working on developing energy-saving enhancements for LTE-Advanced heterogeneous networks [7], where small eNBs are assumed to enter an energy saving (ES) state under certain conditions, while macro cells provide compensation in terms of coverage and capacity. Such a scenario would also comply with the design requirements of the dual connectivity feature.
2.3.2 Dual Connectivity

Currently, the 3GPP has developed a paradigm called dual connectivity that enables simplified mobility within a heterogeneous network environment by optimizing the handover procedure and reducing the signaling load towards the core network [10]. Dual connectivity enables the master eNB (MeNB) to assume more control over RRM and other operational functionalities of the secondary eNB (SeNB) [15,38,68,104,106,166]. With dual connectivity enhancement, the QoS supported for traffic flows depends mainly on the backhaul link quality. Such enhancements have also led to an increasing focus on the importance of the backhaul links for efficient traffic management and QoS support. Considering this, joint optimization of the power consumption of backhaul and radio access networks has been selected as yet another focus area of this work.

The dual connectivity architecture considered with master eNB deployed along with secondary eNBs is as shown in Fig. 2.1. From an architecture and deployment perspective, there are no restrictions on the type of eNBs that could be classified as MeNB or SeNB. The architecture used in this work is based on the agreed assumptions from [10], and the work presented in [131]. Here MeNBs and SeNBs are deployed in different frequency layers F1 and F2 respectively. This maximizes the spectral utilization and enables user throughput enhancements through inter-eNB traffic splitting and with the help of an increased number of buffers and independent RRM operations available within the eNBs.

Dual connectivity architecture could also enable the operators to design the network by providing wide area control and data plane coverage by configuring macro eNBs as MeNBs. They could also help in densifying the network by using low cost remote radio heads, or small cells as SeNBs. Such a deployment alternative is considered to be the basic assumption
in this work, with similar assumptions also studied in [74, 79, 166]. This could provide data plane throughput enhancements, throughout the network or at selected hotspot areas depending on the operator’s deployment strategy, without adding significant control plane overhead. Wide area data plane coverage using MeNBs is essential for improving mobility performance and connectivity throughout the network, while reducing the occurrence of radio link failures and other service interruptions. This could also enable an energy-efficient operation of the network, by enabling the MeNBs to take distributed energy saving decisions.

Such deployments enable operators to have CAPEX reductions by using low-cost small cells that involve minimal infrastructure costs. OPEX reductions are possible due to the low power consumption of such nodes, without the need for artificial cooling, etc. Dense deployments of small cells with limited coverage footprint also implies that significant optimizations are required in the operation of such cells for further reduction in OPEX.

2.4 5G Networks

Cloud-based, centralized RAN architecture is one of the key enhancements currently being studied for future 5G networks. It combines various solution aspects of 4G heterogeneous networks with centralized computing, thereby improving the overall performance of such systems. One of the key focus areas of such an architecture has been the use of ideal, high-capacity and low-latency backhaul [90,121]. But the use of non-ideal backhaul for such architectures would also be an interesting area of study, as it would enable a wider adoption of the technology [28,117,119] by lowering deployment costs.

While capacity enhancement is the main motivation behind the research and development of 5G systems, there are other key drivers as well, including: mobility performance enhancements; latency reductions; improved security and privacy; better reliability; reduced CAPEX and OPEX; and energy efficiency [39]. Some additional requirements, mainly from a physical layer perspective, considered in [99] include: the support of frequency coordination and reuse; support for advanced and pipeline-processing at the receivers; low UE battery power consumption; support for new communication links (beyond eNB-UE link in legacy networks); and self-backhauling.
The perceived increase of connected devices is due to the provisioning of commercial and public safety services using 5G networks, including devices engaged in machine type communications (MTC). Recent advancements in MTC and the internet of things is presented in [70, 134]. Such MTC devices engaging in machine-to-machine (M2M) communication are expected to be one of the main reasons behind the increase in user density in 5G systems [132]. Some of the key requirements, scenarios and performance targets for such a system is also documented in [37, 111], a subset of which is as shown in Fig. 2.2. Currently, massive deployment of MTC devices for enabling smart cities, energy meters, etc., and ultra-reliable MTC for industrial automation is envisioned as a key use case for further studies.

Currently, a thousand-fold capacity increase is one of the main design targets for such a system, with a significant increase in the total amount of connected devices as well as improved individual user experience [69, 88, 116]. The capacity enhancement target is further divided into the three components of spectral efficiency, bandwidth and access point density, each of which is anticipated to provide tenfold gains. Spectral efficiency is expected to be improved by considering advanced transmission technologies, possibly a new physical layer, massive antenna configurations, and better utilization of multi-user and peer-to-peer communication...
mechanisms [112]. For improving system bandwidth, apart from physical layer enhancements, the possibility of utilizing the spectrum available in millimeter and centimeter wave (mmW and cmW) frequency bands is also being studied [153]. Network densification using small cells is yet another key factor expected to provide capacity enhancements for achieving the targets currently set for 5G systems.

One of the main considerations for mobility performance enhancement is seamless mobility within the radio access network, with limited impact on user experience. It is also expected that 5G networks should be able to support high data rates for fast moving users in high-speed trains, with user speeds of up to 500 km/h, compared to 250 km/h in 4G [154]. Mobility support with increased network densification and heterogeneity in terms of multiple radio access technologies, etc., is considered to be a key challenge in 5G as well [11, 111]. Restricting highly mobile users from connecting to 5G mmW cells would be beneficial, considering the challenges involved in handover in terms of aligning transmit and receive beams [11]. Enabling mobility with reliable connectivity, and low UE battery power consumption is considered to be a key challenge in [111], especially taking into account that the small cells in 5G would be deployed in different frequency layers, as compared to the coverage of a wide area cell. This would make the studies for efficient small cell discovery done in this work, which tackle all of these challenges, relevant and important in the context of 5G systems as well.

Latency reductions could be achieved using novel physical layer design, for example with reduced transmission time intervals. This is important for latency-critical industrial automation, automation of vehicles, public safety communication, etc. Key technology enablers for high-capacity, low-latency 5G networks are presented in [100], where an optimized frame structure with time-separated data and control parts are proposed to enable low-latency communication. Enhanced security and privacy would require new mechanisms to be developed at the core network, and should adequately be reflected in the radio access network as well. Some mechanisms for enhancing security, while reducing the latency and improving the reliability of the network are described in [76]. The reduction of network deployment and operation costs is another key factor which would impact the widespread adoption of the system [45]. While the increased CAPEX and OPEX resulting from network densification is identified as a key limitation for 5G networks [42], the work done in [107], based on the
evaluation of the profitability and cost-capacity analysis of 5G mmW cells shows that such cells could provide operators with more than a 50% profit margin.

2.4.1 Cloud-based Centralized RAN Architecture

The main idea behind cloud-based RAN architecture is the deployment of low-cost remote radio heads, with limited radio capabilities, and a centralized, general-purpose processor for handling the radio access and resource management functionalities. The motivation for having such an architecture is to lower the deployment costs (CAPEX) of next-generation wireless access networks while enabling easier implementation of sophisticated features such as Coordinated Multi Point (CoMP) [12, 89, 155]. Currently, the main assumption in cloud-based centralized RAN in terms of backhaul is that the functionalities could be either fully or partially centralized, depending on the backhaul link quality [33]. The possible cloud-based or centralized radio access network architecture considered in this work is as shown in Fig. 2.3. The centralized virtual base station pool could also enable the dynamic sharing of resources between the RRHs as well [142].

A fully centralized architecture would have an ideal, high capacity, low delay backhaul link with the central base station pool handling baseband processing and other higher-layer functionalities. Partially centralized architecture would consist of the baseband processing still being done by
the RRHs, while the rest of the functionalities would be handled at the centralized entity. This partially centralized architecture is used as the baseline assumption in the presented work. The advantage of a fully centralized architecture is that the RRHs could be low-cost radio equipment with limited radio processing capabilities with a high-cost backhaul link. With the partially centralized approach, the RRHs need to have more processing capabilities, increasing the relative deployment costs, with low-cost, non-ideal backhaul links. Here we use the term cloud-based and centralized interchangeably, remaining agnostic regarding the location of the central processor, whether on the cloud server infrastructure or dedicated servers.

2.4.2 Energy Efficiency in 5G

Energy efficiency is a key requirement in 5G, both from the network and UE perspective. With the perceived ultra-dense deployment of different types of cells in 5G, mechanisms for optimizing the power consumption of the network while minimizing the impact on user throughput and QoS would be essential. Such networks also provide the opportunity to provide targeted delivery of data, which could result in efficient use of spectral and energy resources [30, 124, 136]. One such way to target provisioning of capacity to the users could be with the use of the enhanced small cell discovery mechanisms presented in this work, since the ultra-dense small cells deployed in dedicated mmW or cmW frequencies, as compared to the wide area cells, would require energy-efficient discovery [130]. The limited coverage footprint of such cells would necessitate mobility optimizations that take UE power consumption reductions into account, such as the mobility state-based small cell discovery.

Since D2D communication is a key use case in such networks, efficient discovery mechanisms that apply the small cell discovery concepts presented in this work would be relevant as well. The energy-efficient small cell operation mechanisms considered in this work could be a key enabler for the targeted delivery of high capacity to end users, enabling efficient use of energy and spectral resources, with configurable tradeoffs in user QoS as well.

While cloud radio access networks could provide significant improvements in spectral efficiency, they could also be considered to be key enablers for energy-efficient 5G networks [124, 136]. The work done in [124] shows that a significant increase in spectral and energy efficiency can
be achieved using heterogeneous cloud RAN, through higher degrees of freedom in interference control and resource allocation. In [136], the improvement of spectral and energy efficiency is shown using higher levels of flexibility and adaptability, with the help of software functions and algorithms implemented at the cloud server. Such architectures could also enable network controllers to deactivate parts of RAN and backhaul, based on user demand and/or network conditions, in order to reduce network energy consumption. In [137], the energy efficiency benefits of having RAN-as-a-Service concept for 5G with a cloud-based infrastructure and using commodity hardware is evaluated. Based on the performance evaluations done, the proposed paradigm was shown to provide promising energy efficiency gains for 5G RAN.

### 2.4.3 HARQ in 5G

The HARQ procedure has been defined in wireless systems such as LTE [4] to improve physical layer robustness [67,144] and spectral efficiency of the system by enabling the use of a higher spectral efficiency for transmissions. The HARQ ensures the successful reception of packets, and minimizes the error rates in the system. The process is as shown in Fig. 2.4, based on the description in [67]. Upon receiving the packet in the physical uplink/downlink shared channel (PUSCH / PDSCH) the UE or eNB will decode it, and based on the decoding result send an acknowledgment (ACK) or negative ACK (NACK) back to the entity that sent the packet. The ACK/NACK for the packet in frame \( n \) should be sent in frame \( n + 4 \) by the UE or eNB receiving the packet. This would mean that the process-
ing delay requirement for decoding would be approximately 3 ms, which could increase the computational requirements of the RRH in a partially centralized RAN environment.

In this work, we assume that some of the basic functionalities defined in the LTE protocol stack, such as HARQ could be used in 5G as well. The LTE user plane protocol stack is as shown in Fig. 2.5 [84]. Optimizing HARQ based on various constraints has been widely studied in the literature. The work done in [164] considers the rate adaptation design of adaptive modulation and coding (AMC)/HARQ system with Chase-combining, under different block error constraints, while considering the throughput as the performance metric. In [163], various sources of channel quality indicator (CQI) errors at the transmitter in the rate adaptation of AMC/HARQ systems are studied, with CQI errors being modeled based on practical scenarios. In both cases, the evaluations are done with a fixed or maximum number of retransmissions, which is similar to the work considered here. The work in [53] considers the use of an effective SINR parameter to map various combinations of HARQ and AMC into a smaller group of reference states. The work presented here also considers the use of an effective SNR paradigm for estimating whether or not a retransmission is required. In [169], the mapping between SINR and the modulation and coding scheme (MCS) is optimized for maximizing the throughput, considering the HARQ scheme used. While a mapping approach is used in this work as well, the mapping is done between the effective SNR of the UE uplink transmissions and the related block error rates.

The work done in [162] presents a rate compatible structure (irregular...
repeat-accumulate and generalized irregular repeat-accumulate codes), which is considered to be attractive for 5G. Mechanisms for supporting different code rates for incremental redundancy-based HARQ for 5G mmW networks are also elaborated. The work presented in [81] proposes a new air interface for 5G systems, which enables the implementation of a simple HARQ scheme, with reduced complexity and latency requirements. In [144], the use of HARQ as a means for handling the SINR uncertainties in 5G networks is evaluated, and the potential for HARQ to improve the link quality is shown based on simulations.
In this chapter, the overall system model used is discussed, along with the details of the simulation scenario setting and parameters used for evaluation. A generic overview of the simulator is also provided, detailing the process used for scientific validation. The motivation for the use of simulation-based evaluations done in this work is also discussed. The models used in this work for inter-frequency discovery, mobility and power consumption are also presented.

3.1 System Model

The overall system considered in this work is as shown in Fig. 3.1, consisting of three sub-systems: a heterogeneous network sub-system consisting of macro and small cells; a D2D sub-system with macro and D2D UEs with a discovery procedure similar to small cell discovery; and a cloud-RAN sub-system consisting of densely deployed RRHs, with a centralized or cloud-based virtual base station pool. In the heterogeneous network sub-system, small cells were deployed randomly within the coverage area of the macro cells, with minimum inter-cellular distance conditions, in order to have good spatial separation between the cells. For both the heterogeneous network and D2D sub-system, a uniform distribution of users within the scenario is assumed. It is also assumed that users move or remain static within the scenario, discover and connect to the most proximal cell or D2D device with the strongest received reference signal strength, and with an offset in the case of small cells.

The UE is assumed to consume a fixed amount of power for small cell and D2D discovery while searching for the discovery signal sent in a dedicated frequency layer. A backhaul link exists between the macro and small cells, as well as the core network, for data and control plane in-
formation exchange, which is also assumed to consume power, depending on the link type. Macro and small cell operation also consumes power, depending on the load of the cell, which needs to be optimized. The backhaul link is also assumed to be present between the RRHs and the centralized base station pool.

Here, in the context of 4G systems, backhaul is considered to be the link between eNB and the core network, as well as between eNBs in the system. The interface between the eNB and the remote radio heads is often called fronthaul in this context. From a 5G perspective, the terms fronthaul and backhaul are used more ambiguously. Recent studies on 5G backhaul architecture use the term ‘backhaul’ for both the interfaces between eNBs and the core network, and between eNBs and RRH [71]. This convention has been followed in this work as well, as indicated in Fig. 3.1. Backhaul links are assumed to be transport networks that are IP-based, including aggregation points, routers, etc., connecting RAN elements, core network elements and RRHs.

The main focus of the system considered in Publication I to Publication V is on the energy consumed by the UE when discovering small cells deployed in a dedicated frequency band or other devices in its proximity which send D2D discovery signals on dedicated frequency resources. The work done in Publication VI to Publication VII investigates the energy-efficient operation of the heterogeneous small cell network, with the basic assumption that the UEs discover small cells in such a network in an optimal, energy-efficient manner. The work done in Publication VIII mainly focuses on partially decentralizing the HARQ process of a system with a non-ideal backhaul link between RRHs and the cloud BS pool, while
minimizing UE UL transmit power consumption due to unnecessary retransmissions.

In this chapter the simulation setting, small cell discovery, mobility and power consumption models used for evaluations are also discussed. The models discussed here are mainly applicable for the evaluations presented in Publication I to Publication VII.

### 3.2 Simulation Setting

The path loss models, antenna pattern, etc., used in heterogeneous network and D2D sub-systems for evaluation follow the 3GPP case 1 model defined in [2]. The mobility settings have been mainly modeled based on heterogeneous network mobility studies done in 3GPP [3]. Similar models were used for evaluations in several works in literature related to heterogeneous networks [20, 91, 129]. For the simulations with mobility, two-dimensional, spatially-correlated slow fading was also used, in order to more closely model real-world handover and other aspects of mobility-related performance. Thus, fixed slow fading values for each physical location were used in simulations, enabling a more reliable approximation of mobility and handover performance [26, 113]. The distance-dependent path-loss from macro BS to mobile terminal, $L_M$, and from pico cell to the terminal, $L_P$, is given by [2]:

\[
L_M = 128.1 + 37.6 \log_{10}(R_M) \tag{3.1}
\]

\[
L_P = 140.7 + 36.7 \log_{10}(R_P), \tag{3.2}
\]

where $R_M$ and $R_P$ are the distance between the UE and the macro and pico respectively.

The fingerprint database structure considered for small cell discovery is as shown in Table 3.1. The RF fingerprint (FP) information is assumed

<table>
<thead>
<tr>
<th>Fingerprint 1</th>
<th>Fingerprint 2</th>
<th>...</th>
<th>Fingerprint M</th>
</tr>
</thead>
<tbody>
<tr>
<td>MacroBSID(_{11})</td>
<td>MacroBSID(_{21})</td>
<td>...</td>
<td>MacroBSID(_{M1})</td>
</tr>
<tr>
<td>RSRP(_{11})</td>
<td>RSRP(_{21})</td>
<td>...</td>
<td>RSRP(_{M1})</td>
</tr>
<tr>
<td>MacroBSID(_{12})</td>
<td>MacroBSID(_{22})</td>
<td>...</td>
<td>MacroBSID(_{M2})</td>
</tr>
<tr>
<td>RSRP(_{12})</td>
<td>RSRP(_{22})</td>
<td>...</td>
<td>RSRP(_{M2})</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>MacroBSID(_{1N})</td>
<td>MacroBSID(_{2N})</td>
<td>...</td>
<td>MacroBSID(_{MN})</td>
</tr>
<tr>
<td>RSRP(_{1N})</td>
<td>RSRP(_{2N})</td>
<td>...</td>
<td>RSRP(_{MN})</td>
</tr>
</tbody>
</table>
Table 3.2. System Level Simulation Parameters

<table>
<thead>
<tr>
<th>Basic Radio Configuration Parameters [3]</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Macro cell ISD</td>
<td>500 m</td>
</tr>
<tr>
<td>Shadowing Standard Deviation</td>
<td>Macro (Pico) cell 8 (10) dB</td>
</tr>
<tr>
<td>Spectrum Allocation</td>
<td>10 MHz Channels</td>
</tr>
<tr>
<td>Macro (Pico) Max Tx Power [dBm]</td>
<td>46 (30)</td>
</tr>
<tr>
<td>Antenna Gain [dB]</td>
<td>Macro (Pico) 15 (5)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transmit Power Related Parameters [17]</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N, (per cell)</td>
<td>Macro (Pico) 2 (2)</td>
</tr>
<tr>
<td>P₀ [W]</td>
<td>Macro (Pico) 130.0 (6.8)</td>
</tr>
<tr>
<td>Δₚ</td>
<td>Macro (Pico) 4.7 (4.0)</td>
</tr>
<tr>
<td>Pₘ [W]</td>
<td>Macro (Pico) 20.0 (0.5)</td>
</tr>
<tr>
<td>Pₛ [W]</td>
<td>Macro (Pico) 75.0 (4.3)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other Simulation Parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral Efficiency, Sₑff</td>
<td>4.0</td>
</tr>
<tr>
<td>No. of RBs, N_RB</td>
<td>50</td>
</tr>
<tr>
<td>PRB size, RBₜ</td>
<td>180 kHz</td>
</tr>
<tr>
<td>User Placement</td>
<td>Random</td>
</tr>
</tbody>
</table>

to be a combination of the reference signal received power (RSRP) values of N strongest cells, and the corresponding cell IDs. The fingerprint is considered to be a match if the intra-frequency measurements conducted by the UE on the macro cell layer indicate that the macro cell IDs match with the fingerprint entry, and RSRP value differences are within a prescribed limit. This limit is assumed to be a static or dynamic offset, δₕₚ, in the presented work.

While throughput evaluations are not the main focus of this work, the related results were generated using the modified Shannon formula for LTE systems, presented in [98], which has been widely used in the literature [55, 62, 127]. For simplifying the SINR calculations, and to make the interference conditions in the scenarios more challenging, the neighboring cells are always assumed to be fully loaded. Some of the system-level parameters used for simulations are as shown in Table 3.2.

### 3.3 Simulator Overview

A dynamic system-level simulator was used in this work to conduct the evaluations with mobility, with static mode being used for the ones without mobility. A flow-chart showing the overview of the simulator operation...
System and Simulation Model

Create UEs and drop them randomly within the scenario
Move UEs based on mobility model used
Calculate path loss, including slow-fading, etc., between UE and eNB
Execute RRM and mobility functions (handover, scheduling, generating traffic, etc.)
Calculate SINR for each user
Calculate user throughput and other network statistics
Collect statistics

**Figure 3.2.** Overview of the simulator operation.

...is as shown in Fig. 3.2, similar to the one described in [47]. The only difference between static and dynamic mode is that the UEs move within the scenario using different mobility models, described later in this section, and the path-loss values, etc., were calculated after the UE takes each step, defined as the distance moved in 1 ms, depending on the speed of the UE. In the static mode, the UEs are dropped uniformly at random, with the values related to slow-fading, path-loss, etc., remaining static. Publication I – Publication IV uses the dynamic mode, Publication V – Publication VI uses the static mode, and Publication VII uses a combination of both modes. All simulations with conducted with significantly large simulation run-time or number of snapshots, in order to achieve a good averaging of the results.

The main motivation behind using spatially correlated slow-fading was that it was shown to provide more reliable estimates of mobility and handover performance [113]. For throughput evaluations, mainly a full-buffer traffic model was used, where the system would be fully loaded, with data to be sent throughout the simulation duration. Such a traffic model is useful in understanding the potential boundaries of the evaluated system. Constant and maximum bit rate traffic models were used in Publication VII – Publication VI, for load calculations and power consumption evaluations. For the constant bit rate model, based on the downlink SINR experienced by the UE, the scheduler was assumed to do radio resource
reservations for all the users in order to achieve the constant bit rate, $R_{\text{cbr}}$. Any remaining resources were assigned to the other users in a round robin fashion. For the maximum bit rate traffic model, the resources were allocated in a round robin fashion until the users achieved the maximum rate of $R_{\text{max}}$, thereby throttling the rates. If $R_{\text{max}}$ was used along with a minimum bit rate, $R_{\text{min}}$, then a combination of resource reservations until each user achieves $R_{\text{min}}$ was used. The remaining resources were allocated until the maximum bit rate was reached.

A uniform distribution of users within the simulation world was considered for the simulations, in order to achieve a good averaging of results, which can be easily reproduced as well. The mobility models were also chosen with the same goal in mind. Various statistics related to mobility and power consumption were then collected and averaged over the whole simulation to generate the results.

### 3.4 Motivation for Conducting Simulation-based Evaluations

The evaluations presented in this work for the heterogeneous network and D2D sub-systems were mainly done on a dynamic system-level simulator, emulating an LTE-Advanced system. The solutions considered were mainly evaluated using the system-level simulator, to understand the impacts, in a setting that emulates the real-world. This is essential, especially for mobility performance and network energy efficiency evaluations, in order to evaluate whether the enhancements would provide gains when deployed in actual networks.
The evaluations for mechanisms proposed during the course could also be done using a more predictable analytical model. The key difference between such an analytical modeling of the heterogeneous network subsystem and a more complicated system with slow-fading settings is as shown in Fig. 3.3. The figure shows the radio environment map of a heterogeneous network, with each color indicating the strongest cell in each physical region. The radio environment map was created with a resolution of 2 m. From the figure we can observe that while the mobility within an ideal environment without the effects of shadow fading is more predictable and can be evaluated analytically, the scenario becomes quite complex when the effects of shadow fading are added.

When a UE searches for inter-frequency small cells every $T_{\text{Inter}} = 80$ ms, and with UE speed, $v = 3$ km/h, the small cell connection time distribution is as shown in Fig. 3.4. In the purely analytical system, the mean distance $d_{\text{mean}}$ covered by a UE while traversing through a cell of radius $r_d$ is evaluated in Publication I and is given by:

$$d_{\text{mean}} = 2r_d \cdot \frac{2}{\pi} \int_{0}^{\pi/2} \sin \theta \, d\theta = \frac{4r_d}{\pi} \quad (3.3)$$

This mean distance value combined with the knowledge of the UE speed indicates the mean amount of time the UE would be spending in a small cell. But as we can observe from Fig. 3.4, such an analysis would no longer be valid for the more realistic system that was simulated, primarily due to the altering of the cell shape.

A similar evaluation of the radio environment map was also done in Publication II, the effects of which are shown in Fig. 3.5. Here we consider the use of radio fingerprints and fingerprint matching regions along the
boundaries of small cells, with a matching offset of $\delta_{fp} = 1$ dB. The matching offset is the maximum allowable difference between the radio fingerprint information available at the network and the RSRP values of the three strongest macro cells measured by the UE. Here in the case without fading, the path loss, antenna gain and base station feeder losses are considered to enable macro-cell sectorization. The radio environment map also includes a 12 dB cell range expansion offset for small cells. From the figure we can observe that the fingerprint matching region is ideal in the analytical model, with matches occurring only in close proximity of the fingerprint locations. We can also observe that there is no probability of errors in such a scenario.

But in the more realistic case with shadow fading the system becomes much more complicated, with significantly higher false fingerprint matches. The need for understanding the impact of false matches is clearly evident in the realistic cases, as shown in Fig. 3.5(b), generated with the complex simulator. A complex system-level simulator was used for the evaluations because it was considered important to understand the effects and performance of the proposed mechanisms in a realistic network environment. From the figure we can see that there is a significant number of false matching regions, which would consume additional UE battery power for performing inter-frequency measurements.
### 3.5 Inter-Frequency Small Cell Discovery Modeling

For inter-frequency cell discovery, the UEs are assumed to be configured a measurement gap, during which the UE can switch to the new frequency band to search for cells. The measurement gap is assumed to be of 6 ms, in line with the assumptions in LTE-Advanced systems [6], with the measurement gaps configured every 40 or 80 ms, during which time no data is sent in the uplink or downlink directions. Even with relaxed measurement gap periods, the cell discovery is modeled such that the UE makes multiple measurements with a periodicity of 40 or 80 ms before initiating the handover procedure. This is done in order to improve the discovery performance of the UE. In Publication I and Publication II, inter-frequency cell search power consumption is modeled as a function of number of cells detected and based on a minimum RSRP value for cell detection. But in the remaining work, a simpler cell discovery model of fixed power consumption is assumed, based on the model for power consumed when receiving downlink data that is presented in [157]. Similar assumptions have been made for device discovery in Publication V as well, since discovery beacons are assumed to be sent in a dedicated frequency layer.

### 3.6 Mobility Models

The movement models used in the paper are the macro movement and hotspot movement models, as shown in Fig. 3.6. In the macro movement
model, the UE moves in a random fixed direction, at a constant speed until it reaches the end of the simulation world, after which it takes a randomly selected new direction. Such a model enables good averaging of mobility and handover performance with the system. In Publication III and Publication IV, a more realistic model is considered. It combines the macro movement model with stationary periods and a longer time scale for simulations. The stationary periods could indicate the time a user would normally spend at home, the office or at a shopping center, during the course of a day. The state transition probabilities and the time duration spent in each state follow an exponential distribution with mean values as shown in the figure. In Publication I, Publication II, Publication V and, Publication VII a random straight walk model was simulated with users moving in a straight line until the end of the simulation world, and then turning to a random new direction within the simulation world. In Publication VI and Publication VIII mobility was not simulated, since that was not the main focus of the work.

3.7 Power Consumption Models

The used UE power consumption model used is as shown in Fig. 3.7, based on [9, 85, 157]. The total power consumption [W] depends on the radiated transmit power [dBm]. The relationship between the radiated transmit power and the power consumption according to [85] is plotted with dashed green lines in Fig 3.7. In 3GPP, absolute power consumption lev-
Figure 3.8. ScNB transmit power vs. actual power consumption with and without considering backhaul power consumption [Publication VII]. ©2014 Springer Science+Business Media New York

Relative energy consumption per sub-frame were however discussed in [9]. These are plotted in dashed blue in the figure, with the relative energy units shown at the left. According to [157], the power consumed when the UE is receiving is 0.5 W, whereas according to [9], the relative energy consumption per sub-frame is 1, when the UE is receiving. These receive power consumption levels are added to the figure for convenience. Here, these receive power consumption levels are used to measure the power consumed for inter-frequency scanning.

Similar models have been considered for D2D discovery in [128], as well. The receive power consumption [W] has been used in Publication III, and the transmit and receive power models have been used in Publication IV.
and Publication V. The UE transmit power consumption is calculated based on the LTE UL power control models [5, 27], based on ideal path loss estimation assumptions.

The work done in [86] presents the transmit and receive power consumption measurements, based on experiments conducted on LTE smartphones. The measurements show higher power consumption than [85]. For low transmit powers, < 0 dBm, there is an approximately 3 to 3.5 times increase in actual power consumption in the new measurement results. At the maximum transmit power of 23 dBm, there is an approximately 1.33 to 1.67 times increase in the UE battery power consumption. The main relative trend of the transmit power consumption is well captured by the simple models of [85] and [9], used in this work. The ratio of the baseline total transmit power consumption at -20 dBm to the receive power consumption, is almost the same in the measurements of [86], as it is in the models of [9, 85, 157]. Accordingly, if measurements from [86] would be used instead of the simple models of [9, 85, 157], the total absolute power consumption would be roughly 3 times the values presented in Publication III- Publication IV, where the power consumption is dominated by reception. In Publication IV and Publication V, the results would be affected with up to 3.5 times increase in absolute power consumption values.

In Publication I and Publication II, a slightly more complicated model was considered with the energy consumed for the first inter-frequency measurement depending on the number of detected cells. Assuming three cells are detected, this would be about 0.625 W, with subsequent measurements consuming 0.25 W. This model was simplified in the subsequent papers. If a total of six inter-frequency measurements are conducted, on average 0.31 W would be consumed for this operation, with up to 5 times increase in power consumption results compared to the new measurements. An analysis of the actual power consumption, based on the measurements of a smartphone that supports carrier aggregation was also presented in [143].

The small cell transmit vs. actual power consumption is as shown in Fig. 3.8(a). Here, the eNB transmit power is estimated based on the load of the cell, with full power assumed for a cell which is fully loaded, and with the power equally distributed between all physical resource blocks within a subframe. The linear model used for active mode and sleep mode power consumption is based on [17] is used in Publication VI, given by:
\[ P_C = \begin{cases} 
N_A (P_0 + \Delta_p \rho P_m) + P_{bh,a}, & \text{active mode} \\
N_A P_s + P_{bh,s}, & \text{sleep mode}, 
\end{cases} \]  

where \( N_A \) is the number of antennas, \( P_0 \) is the power consumption at zero RF output power, \( \Delta_p \) is the slope of the load-dependent power consumption, \( P_m \) [W] is the maximum RMS transmit power of the base station, and \( P_s \) is the power consumption [W] when the BS is in sleep mode. \( P_{bh,a} \) and \( P_{bh,s} \) are the active and sleep mode power consumption of the backhaul link. The load factor \( \rho \) is the ratio of the number of resources being used \( N_U \), and the total available resources \( N_T \).

The power models used in Publication VI did not consider backhaul power, while those in Publication VII did. The backhaul power consumption for various backhaul link types are presented in Publication VII, based on [72,95,151,152]. The symmetric backhaul (SB) and wired (Wrd) model is based on [151,152], and the fiber access (FA) model on [72]. The wireless backhaul (WL) model used is based on the studies done in [95]. A detailed description of the model and related assumptions is presented in the publications. The power consumption model presented in [44] for future cellular base stations show even lower power consumption of up to 0.2 W for pico cells in sleep mode. This would have significant impact on the results presented in this work, especially on the load-based offloading schemes, since the power savings from small cells in sleep mode would be significantly higher.
4. Optimizing UE Power Consumption in Heterogeneous Small Cell Networks

4.1 Introduction

In this chapter, we mainly deal with the heterogeneous network and D2D sub-systems discussed earlier. The application of device speed and small cell proximity estimation for optimizing the UE power consumption for small cell discovery is studied. The work is then extended to evaluate whether a combination of these mechanisms could enhance the performance in terms of power consumption reduction. The mechanisms were evaluated with the constraint of maximizing small cell discovery, which could result in an increase in the data rates experienced by the UE. Finally, the proximity estimation mechanism is evaluated for D2D device discovery, and the performance evaluation results are presented.

4.2 Device Speed Estimation Based Discovery

The small cell discovery based on device speed estimation considered in this work uses two approaches: the accurate determination of UE speed, assuming a measurement error; or the use of UE-based mobility state estimation currently defined in LTE-Advanced systems. The main idea is to suspend inter-frequency measurements for small cells when the speed estimation indicates that the UE speed has exceeded a certain threshold. For the first approach of speed estimation (assuming a measurement error) we consider the suspension of inter-frequency measurements, when the UE speed $v$ exceeds a threshold speed $v_t$. In Publication I, the method for analytically calculating the threshold velocity is presented, with the small cell connected time results compared for different values of cell search periods. Here for a small cell with radius $R$, we consider the mean
distance that the UE covers within a small cell $d_{\text{mean}}$, assuming uniform distribution of users, to be given by Eq (3.3).

The threshold speed $v_t$ is defined such that above this speed, the amount of time spent by a UE in a small cell would be less than a minimum time-of-stay duration $t_{\text{MTS}}$. Connecting to a small cell for less than the minimum time-of-stay is not beneficial because of relatively low amount of traffic offloaded and the increased handover signaling load involved. Therefore, it is proposed that inter-frequency measurements could be suspended when the UE speed is above this threshold, to save UE battery power for inter-frequency cell search instead. The value of $v_t$ is defined as the ratio of $d_{\text{mean}}$ and $t_{\text{MTS}}$. But here, the minimum periodicity with which the inter-frequency cell search is conducted (which determines the UE power consumption for this activity) and the amount of small cell traffic offloading opportunities that are missed (resulting in lower user throughput values) could be considered to be conflicting metrics. The use of an offloading loss ratio, $R_{\text{Loss}}$, which indicates the target ratio of small cell traffic offloading opportunities used and the total available offloading opportunities, is considered to be a tradeoff value that could be related to the minimum inter-frequency scanning period, $T_{\text{Inter, min}}$ for UE speed $v$ is given as:

$$T_{\text{Inter, min}} = \begin{cases} 
\frac{2 d_{\text{mean}} R_{\text{Loss}}}{v}, & 0 < v < v_t \\
0, & v > v_t 
\end{cases} \quad (4.1)$$

In the work done in Publication III, the problem is generalized by using three quantized levels for UE mobility states instead of actual speed estimations, and a similar mechanism of suspending cell search procedure is considered. Inter-frequency scanning is suspended when the UE is in a medium or high mobility state, as shown in Fig. 4.1.c. From the figure, we can observe that the inter-frequency scanning rate can be controlled, depending on mobility state, thus enabling flexible and configurable mobility performance at various UE speeds. Here mobility states are UE speed estimates quantized into three different levels - normal, medium and high - based on the number of cell selections within a specified time window. In both cases, the speed estimation is done at the UE, with assistance information received from the macro cell. Optionally, the estimation can also be done at the network, utilizing handover history information in combination with time spent in each cell. Thus, the cell search can be autonomously suspended by the UE, depending on the speed or mobility...
state estimation. Alternatively, the macro cell can change the measurement configuration of the UE, based on the estimates received from the UE. Here the mobility state estimation values can be configured in such a way that the UE speeds at which the state transitions occur can be controlled. Based on this, proper values of $T_{\text{Inter, min}}$ could also be chosen.

### 4.3 Received Signal Strength Based Discovery

The RSS-based estimation of small cell proximity uses a combination of RSRP values and corresponding macro cell IDs to determine the approximate location of a UE, relative to a small cell. If the current RF fingerprint of a UE, estimated based on measurements done by the UE, matches a fingerprint in a database, it is assumed that there is a small cell in the proximity, and cell search procedure needs to be initiated. In Publication II, the measurement of the current fingerprint information of a UE with
an \( RSRP_{\text{offset}} \) is compared with the database. Here the offset is selected in a manner that reduces the chances of fast moving UEs having a fingerprint match and initiating the cell search procedure.

Let \( \alpha \) be the path-loss exponent, \( R_{FP} \) be the distance of the fingerprint location from the macro eNB, and \( r_{fp} \) be the radius of the fingerprint matching region that needs to be defined. \( RSRP_{\text{offset}} \) is given by:

\[
RSRP_{\text{Offset}} = 10^\alpha \log_{10} \left( \frac{R_{FP} + r_{fp}}{R_{FP} - r_{fp}} \right) + \delta_c,
\]  

where \( \delta_c \) represents the correction factor that accounts for the effects of slow and fast fading that would be seen in a realistic network deployment scenario. Here, if the search for a fingerprint match is done every \( T_{\text{FPM}} \) seconds, the match region can be designed in such a way, using an appropriate selection of the \( RSRP_{\text{offset}} \) value, that UEs moving with speeds greater than \( \tilde{v} \) can be excluded from finding the small cell, where \( \tilde{v} \) is given by:

\[
\tilde{v} > \frac{2r}{T}
\]  

In Publication II, it is assumed that the fingerprint data is present only for the border regions of the small cells. The work done in Publication III considers a more generic fingerprint mechanism that could be applied in real network deployments, with a fingerprint database containing entries for all physical regions where the UE could connect to a small cell. Further, this mechanism is combined with the mobility state estimation-based cell search procedure, providing further UE power consumption optimization by not even searching for a fingerprint match when the UE is in a medium or high mobility state. A static measurement offset \( \delta_{FP} \) is used in the paper, which also accounts for measurement errors, slow fading.

Figure 4.2. Signaling flow diagram for RF fingerprint-based small cell discovery. ©2013 IEEE
Optimizing UE Power Consumption in Heterogeneous Small Cell Networks

Figure 4.3. Signaling flow diagram for the optimized RF fingerprint-based small cell discovery. ©2013 IEEE

effects, etc. The performance of the received signal strength-based mechanism is compared with the speed estimation mechanism both independently and in a combined manner, along with a relaxed inter-frequency measurement mechanism in which the inter-frequency scanning is done less frequently, without using any additional intelligence.

The location of the fingerprint database is a key aspect in determining the additional signaling load required to implement the fingerprint-based small cell discovery in a real system. In Publication II and Publication III, the mechanism was evaluated while remaining agnostic as to where the database is located and related signaling involved. In Publication IV, we evaluate this aspect, studying the additional costs involved of having the database either at the network or in the UE. If the database is located at the network, there would be additional signaling load, radio resource utilization and power consumption required for sending frequent measurement reports back to the network.

If the fingerprint database is located at the UE itself, such signaling overheads for fingerprint reporting could be avoided. There would be, however, an additional signaling load from sending the fingerprint information to the UE, and for periodic updates depending on the size of the information. The database size would also be limited by the storage capacity of the UE. The signaling flow diagram for the fingerprint-based small cell discovery mechanism is as shown in Fig. 4.2, where both approaches
(fingerprint database located on the network and at the UE) are considered. The signaling flow involved in the optimized fingerprint mechanism presented in Publication IV, is as shown in Fig. 4.3. Here we assume that the inter-frequency measurement gaps are configured only for those UEs that are not carrier aggregation (CA) capable.

4.4 Applications in D2D Discovery

The main consideration of the work so far has been energy-efficient discovery of inter-frequency small cells. In Publication V, we consider the possible power consumption optimizations in a D2D environment, where similar received signal strength-based discovery principles are applied. In a D2D environment, the key difference would be that discovery optimizations would have to be done on the devices sending and searching for discovery beacons, since both operations consume UE battery power. This is different from the small cell discovery procedure, where the power consumed by small cells in periodically sending synchronization signals for UE measurements are not considered.

The performance of having the real-time fingerprint information present either at the network (network-based / NB approach) or at the third party cloud application server (cloud-based / CB approach) is compared. Since D2D devices could be frequently moving around within the network, which would correspondingly require frequent updates of the fingerprint information, the UE-based approach was not considered for such a scenario. The network-based approach has the advantage of relatively simple im-
Optimizing UE Power Consumption in Heterogeneous Small Cell Networks

In the cloud-based approach, the UEs first need to establish a connection with the cloud server before reporting the fingerprint information. There are some advantages to the cloud-based approach, such as application-aware device discovery, where device discovery is initiated only when the devices in proximity are using the same application. Such an advantage would not be possible in the network-based approach, due to the lack of awareness within the 3GPP network of the UE applications initiating the D2D discovery. It is also worth noting that with this approach the additional overhead of generating and maintaining the fingerprint database could be assumed by the owner of the cloud application server, instead of the network operator.

An overview of the social cloud region management, with the CB approach, is as shown in Fig. 4.4. Here the Social Cloud Application Server (SCAS) configures a Social Cloud Application Network Area (SCANA), which is the physical proximity region of a D2D UE sending discovery beacons for the service/application XY. Here, similar to the earlier ap-
Figure 4.6. Small cell connected time normalized to that of $T_{\text{inter}} = 80$ ms vs. UE Speed.

The results presented in this section are based on Publication I – Publication V.

In Publication I, the focus was mainly on finding the tradeoffs between

4.5 Performance Evaluations

The results presented in this section are based on Publication I – Publication V.

In Publication I, the focus was mainly on finding the tradeoffs between
small cell traffic offloading ratios and power savings, using the relaxed inter-frequency cell search period and a speed estimation-based mechanism. The performance of these two indices for various UE speeds and measurement periodicity were investigated. The small cell connected time normalized to that of the $T_{\text{Inter}} = 80$ ms scenario is as shown in Fig. 4.6, with UE speeds varying from 3 - 120 km/h. In the scenario without fading, which represents a theoretical scenario having larger cell sizes, an offloading loss ratio of 20% is observed only at 30 km/h, for $T_{\text{Inter}} = 6$ s, as shown in Fig. 4.6(a). For such a scenario, the mean distance the UE would be traveling within the small cell, as well as the threshold velocity $v_t$, can be calculated analytically. But for the scenario with slow-fading, which represents the realistic network environment, we can see that the offloading loss ratio is much steeper when the inter-frequency measurement periods are relaxed, as observed in Fig. 4.6(b). Here, we can see that the offloading loss ratio target would be exceeded even with a more frequent scanning period of $T_{\text{Inter}} = 800$ ms, with the mean connection times varying from the theoretical values.

For actual deployments, it is proposed that network configurations be done in such a way that the threshold speed information is sent to the UE by the network, enabling the UE to suspend inter-frequency measurements once the threshold speed is exceeded. For slow-moving UEs, the inter-frequency measurements can be relaxed so that the UE can save power, with acceptable tradeoffs in terms of the traffic offloaded to the small cells. The UE inter-frequency measurement power consumption relative to the $T_{\text{Inter}} = 80$ ms scenario can be observed in Fig. 4.7. From
the figure, we can observe that significant power savings can be obtained without significantly compromising small cell connected time by optimizing measurement periodicity, as observed in Fig. 4.6. For example, measurements could be done every 10 s for the scenario with fading, and every 60 s for the ideal case, resulting in only a 20% loss in terms of small cell offloading opportunities.

The values of UE power consumption and small cell connected time normalized to that of $T_{\text{Inter}} = 80$ ms can be seen in Fig. 4.8, which shows the relative gains of the proposed enhancements compared to the reference scheme of scanning with a fixed period of 80 ms. Here, in the case without fading, $\delta_c = 0$ dB, and $\delta_v = 2$ dB, and $\bar{v} = 36$ km/h. From the figures we can observe that while there are significant energy savings resulting from the fingerprint-based small cell discovery mechanism, the filtering of small cell discovery actions at higher speeds is not entirely perfect. This is because of the residual connections occurring when the UE is in the matching region $r$ when the fingerprint matching is done. The perfor-
Optimizing UE Power Consumption in Heterogeneous Small Cell Networks

Figure 4.9. Normalized UE inter-frequency measurement power consumption and small cell connected time vs. UE speed.

mance is slightly worse for the non-ideal case with slow-fading, compared to the ideal case, because of the higher number of fingerprint matching locations within the scenario. In Fig. 4.8(b), the power consumption for a 10 s scanning period was included in order to illustrate the significant amount of power saved, even with such an infrequent cell search period.

From these results we can conclude that while the fingerprint mechanism is important for enabling energy savings for inter-frequency cell search, the mechanism needs to be combined with the device speed estimation mechanism in order to actually enhance performance. This leads us to consider the enhanced mechanisms considered in Publication III, where a combination of the two mechanisms considered so far was evaluated. The hotspot mobility model shown in Fig. 3.6 is more realistic than what has been considered so far (walking in straight lines in random directions). Instead of using dynamic values for $\text{RSRP}_{\text{offset}}$, we consider a static offset $\delta_{fp} = 5 \text{ dB}$ in order to simplify implementation in a real network.

In Fig. 4.9, the normalized values for small cell connected time, and UE power consumed for inter-frequency cell search relative to the $T_{\text{inter}} = 80 \text{ ms}$ case is shown. Here the values using the fingerprint mechanism (1 s and 60 s FP), as well as combining it with the MSE mechanism are also shown (1 s and 60 s FP-MSE). From the figure, we can observe that the enhanced mechanism saves a significant amount of UE battery power as compared to the 40 ms scanning period scheme. Here we can also observe that the FP-MSE scheme provides significant energy savings, even compared to the FP only scheme, irrespective of the period (1 or 60 s), due to
the filtering of fingerprint searches at medium and high mobility states. The cell re-selection parameters for MSE were chosen in such a way that the UE would enter the medium mobility state at around 30 km/h and the high mobility state at 60 km/h, which is reflected in the savings in power consumption as well.

While the results clearly indicate that, as discussed earlier, combining the device speed estimation mechanism with the fingerprint mechanism can provide significant reductions in consumed power due to small cell search at high speeds, they also show that it saves the short connections that occur when the UE is moving at high speeds. This is shown in Fig. 4.10, where a significant drop in handovers can be observed in Fig. 4.10.a. From the connected time distribution shown in Fig. 4.10.b, we can see that the handovers that are filtered out are indeed short time-of-stay connections.
Since throughput values were not computed in this study the impact of the various proposed mechanisms on the UE energy efficiency cannot be accurately predicted. However, based on the normalized values of small cell connected time and UE power consumption values relative to the baseline mechanism of scanning every 40 ms, an ideal estimate of the energy efficiency can be calculated. It is assumed that the normalized maximum power consumption for cell search is 1 W when $T_{\text{Inter}} = 40$ ms, and the corresponding maximum capacity with all possible small cell offloading opportunities utilized to be 1 bit/s. The values for all the other mechanisms having relative power values and capacity values can be calculated from Fig. 4.9. The resultant ratio of the normalized values from the figure is presented in Fig. 4.11. It gives an indication of the estimated energy efficiency values of various schemes, relative to the $T_{\text{Inter}} = 40$ ms case. From the figure, we can see that the FP-MSE schemes perform relatively better than the fingerprint schemes, with the other enhanced schemes providing significant gains as compared to the reference case.

The mechanisms studied so far have been agnostic in terms of the location of fingerprint databases. The work done in Publication IV compares the use of an optimized fingerprint mechanism where the fingerprint entries within a range $\delta_1$ are filtered out and the optimized fingerprint information is sent to the UE, with the fingerprint approach presented in Publication III used as a reference scheme. Here, an offset of $\delta_2$ is used for obtaining a fingerprint match, which is assumed to be configured by the network. From the figure, we can observe that significant energy savings can be obtained by optimizing the fingerprint information and storing it at the UE, thus avoiding the sending of frequent measurement reports. The savings depend on whether the UE is a macro cell edge or mean
Optimizing UE Power Consumption in Heterogeneous Small Cell Networks

Figure 4.13. State diagram and power consumption model [157] for idle UEs sending location information to the network or cloud. ©2014 IEEE

Figure 4.14. Mean power consumed for D2D discovery with different values of $\delta_{app}$.

user, which affects the amount of power consumed while receiving the optimized fingerprint information. Mean users consume even less power than the 60 s FP scheme. Based on performance results, no impact on small cell connected time was observed.

In the studies so far, the UEs were always assumed to be in connected mode and the fingerprint information present on the network. Applying such a discovery procedure to D2D discovery would require possible support from D2D applications where the application mismatch error factor $\delta_{app}$ could add a new dimension. This could also be roughly generalized to the closed subscriber group case in the small cell scenario, with a mismatch indicating the willingness of the operator that deploys such small cells to share connectivity with all UEs. The impact of the fingerprint information for such device discovery scenarios with support from the cloud application server is considered in Publication V, mainly from a UE power consumption perspective for device discovery.
Here, for idle to connected mode state transition, we consider the model shown in Fig. 4.13. The UE remains connected for 1 TTI to send the fingerprint information in the NB approach, whereas in the CB approach it is assumed to remain active for 5 TTIs to send the fingerprint information to the SCAS, to search for a fingerprint match, and to estimate whether the UE is in the SCANA region. In the NB approach, the UE initiates the D2D discovery procedure whenever there is a fingerprint match, whereas in the CB case the discovery procedure is initiated only when the fingerprint matches and the applications for the UEs within the SCANA regions are the same. In the NB case the 3GPP network is unaware of the application used by the UEs, and only the fingerprint information related to all the UEs sending discovery beacons is available. This effect is simulated by having a static application mismatch error value, $\delta_{\text{app}}$ which varies from 0 to 1, with $\delta_{\text{app}} = 0$ indicating that the UEs within the SCANA region are using different applications, and hence need not discover each other.

The performance comparison of both the schemes is as shown in Fig. 4.14. It can be observed that the overhead for having the fingerprint information at the application server is only around 23% when the applications mismatch completely, reducing to 20% when the applications fully match. The relative savings are due to the fact that the UEs need to send fewer fingerprint reports to the SCAS once a matching D2D device is found. The value of $\delta_{\text{app}}$ has no effect on the NB approach, since the fingerprint reporting event would remain constant, irrespective of the applications used by D2D devices within the SCANA. Here the results for the DD approach are not presented, since the resultant trends were similar to the case where the inter-frequency scanning period was varied in the small cell discovery case, as shown in Publication V.

### 4.6 Conclusion

From the studies done in this section so far, the importance of optimizing the inter-frequency small cell discovery procedure is quite evident. It was also observed that combining the device speed estimation and fingerprint mechanisms provide significant UE battery power savings, while effectively filtering out short-time-of-stay connections, which are usually considered detrimental to the operation of heterogeneous small cell networks. Mechanisms for enabling fingerprint information to be located at
the network or at the UE were also considered, and further optimizations of having such information on third party application servers evaluated. Based on the evaluations, it was shown that significant energy savings can be obtained by using the proposed optimization mechanisms, which could eventually lead to an improvement of the UE energy efficiency for inter-frequency small cell or device discovery, as well.

The main consideration of the work here was to optimize the power consumption for UEs for conducting inter-frequency small cell discovery. This would mean that the UEs are not allowed to discover, if other conditions such as the device speed, etc., exceed defined limits. Relevant areas for future work from this perspective would be to investigate the impact on UE transmit power consumption due to the evaluated optimizations. The work done in [110] considers the optimized mechanisms presented here, and proposes an inter-frequency small cell discovery period, based on UE transmit power as the optimization criteria. Since UEs consume higher amount of transmit power to connect to the macro cells which are further away than small cells, a joint optimization of both the metrics would also be an interesting area of further study.
5. Energy-Efficient Small Cell Operation

5.1 Introduction

In this chapter, an energy-efficient small cell activation mechanism is evaluated, which takes into account the energy saving gains obtained based on the amount of traffic load offloaded from the macro cell to small cells. The energy saving gains are achieved while avoiding UE service quality degradation. Offloading traffic to small cells in energy saving mode is proposed only when significant energy saving gains are achievable, thereby minimizing the total energy consumption of the network in the process. The user traffic characteristics and load-based cell activation and offloading is further extended to take into account the backhaul link active and sleep mode power consumption as well. Energy efficiency of the backhaul aware mechanism is evaluated without affecting the service quality experienced by the user. The mechanism is also evaluated in detail, using various user density, mobility, and traffic conditions, with both indoor and outdoor deployment of small cells.

In this work, it is assumed that the MeNBs provide control and data plane coverage throughout the network, with SeNBs providing data plane enhancements at locations where capacity enhancements are required. This enables the macro to control the small cell on/off activity in a dynamic manner through dynamic traffic split within the network, without adding any significant control plane overhead. Data plane coverage by MeNBs is assumed to be essential in providing coverage throughout the network, which is especially required for time-critical services such as public safety, emergency calls, etc. Such an architecture could also enable the dense deployment of small cells, without increasing the load on control plane entities in the core network, thereby also possibly simplifying mo-
Energy-Efficient Small Cell Operation

This chapter is mainly related to the heterogeneous network sub-system.

5.2 Energy-Efficient Small Cell Operation Mechanisms

In this section, we will discuss the energy-efficient small cell operation mechanisms presented in the thesis, as well as the reference schemes used for comparisons.

5.2.1 Backhaul-aware Energy Efficiency Mechanism

The energy efficiency mechanism considered in this work takes both the radio access network load as well as backhaul power consumption into account when taking traffic offloading decisions. In Publication VI, an energy efficiency criteria based on energy offloading – in this case, on the RAN traffic load alone, ignoring the power consumed by the backhaul link – is considered. The work done in Publication VII further extends this concept to take both the RAN and backhaul load into account while activating and deactivating small cells. The main idea of the backhaul-aware criteria is that traffic is only offloaded from a macro to a small cell if it results in energy consumption reduction in the network.

Offloading traffic reduces the load of the macro cell, resulting in reduction of transmit power consumption, which leads to energy savings. If, however, the additional energy consumed by activating a small cell and its backhaul link is more than the energy saved from traffic offloading, it is better not to offload the traffic in order to avoid increasing the power consumption of the network. The offloading decision is also assumed to be taken if it results in service quality improvements for the user, based on network priorities defined by the operator. The backhaul-aware energy efficiency criterion is given by:

$$\rho_{MS} > \frac{(N_{A_m}(P_{0_s} + \Delta p_m P_{m_s} - P_{s_s}) + P_{bh_{a_s}} - P_{bh_{s_s}} - P_{bh_{a_m-off}})}{N_{A_m} \Delta p_m P_{m_m}}$$

(5.1)

where $N_{A_m}$, $\Delta p_m$, $P_{m_m}$ are power consumption values from Eq. 3.4 for the MeNB, and $N_{A_s}$, $\Delta p_s$, $P_{m_s}$ are those of the SeNB. $\rho_{MS}$ is the aggregate load of the users offloaded from MeNB to SeNB, and $\rho_{SM}$ the estimated load due to the offloaded users at the SeNB. $P_{bh_{a_s}}$ and $P_{bh_{s_s}}$ are the active and sleep mode backhaul power consumption of the SeNB backhaul links. Here, $P_{bh_{a_m-off}}$ is the load-dependent part of the MeNB backhaul power.
Energy-Efficient Small Cell Operation

consumption in the ideal backhaul model presented in [152], and is given by:

\[ P_{bh_{a-off}} = (1 - \tau) P_{max_{switch}} \sum_{n=1}^{N_{off}} n_{ports} C_{max_{switch}} R_n, \]  

(5.2)

where \( N_{off} \) is the number of offloaded users, and \( R_n \) is the data rate of each user. Here, \( \rho_{MS} \) and \( \rho_{SM} \) are dependent on the sum rates of the offloaded users, \( \sum_{n=1}^{N_{off}} R_n \), whose values can be calculated as a ratio of the number of PRB utilized and the total amount of PRBs available.

In this way traffic is offloaded from the MeNB to the SeNB if the power consumption reduction due to the offloading action at the MeNB results in a total network power savings. The main challenge in implementing the backhaul-aware mechanism in a real network is the accuracy of the potential load estimation, \( \rho_{MS} \) and \( \rho_{SM} \). For this the SINR of the UE towards the SeNB in an inactive state should be known, since all the other values in the equations are constants, and can be pre-configured at the MeNB. For making a small cell activation decision, the MeNB needs to determine the accurate potential load caused by the offloaded traffic towards the small cell. During the small cell deactivation procedure, the small cell informs the MeNB about the traffic characteristics and load conditions, and the MeNB is assumed to make the deactivation decision.

In both cases, the MeNB is assumed to receive assistance from the latest tools available for network analytics, using dynamic radio environment maps, for making such decisions. While the work done in Publication VII assumes an ideal estimation of this load, a possible signaling flow diagram, which is a modified version of the one presented in Publication VI, indicates how the flow might look with this assistance information. One such standardized feature in LTE-A systems is minimization of drive tests [65, 73], which enables network operators to collect network information such as coverage and capacity maps based on the reports from the UE. Depending on network configurations, such reports maybe periodic or event-based. This enables the operator to build an accurate radio environment map using various data analytics and visualization tools. The use of such information was considered as a possible technique for generating the fingerprint information mentioned in this work as well.

The work done in [151,152] mainly tries to optimize the overall energy consumption of the network with full load assumptions, taking the power consumption for ideal backhaul links into consideration. In a network with realistic traffic and related network load, such optimizations would not be enough. The work done in [14,103,150] optimizes the power con-
Figure 5.1. Signaling Flow Diagram.

...
5.3 Simulation Results

The results presented in this section are taken from Publication VI and Publication VII.

An overview of the load-based offloading mechanism considered in Publication VI and its relative gains for each user density, in terms of pico cell or SeNB power consumption, as compared to the location-based baseline mechanism, is shown in Fig. 5.2. Here the pico power consumption values for the proposed mechanism are normalized to that of the baseline scheme, for each user density. For Scenario-1, $R_{cbr} = 64$ kbps, and data users have full buffer traffic. For Scenario-2, for 80% of the CBR users, $R_{cbr} = 64$ kbps and $R_{cbr} = 512$ kbps for the rest of the CBR users. The MBR data users in this scenario have $R_{\text{max}} = 2$ Mbps, with no minimum
rate constraint, $R_{\text{min}} = 0$ Mbps.

From the results we can observe that the proposed load-based scheme provides significant gains: up to 40% for the full buffer traffic scenario and more than 20% for the MBR scenario. As the density of full buffer or MBR data users increases in the network ($N_{\text{UE}} \geq 40$) the gains tend to be quite similar, since almost the same amount of SeNBs are activated in such scenarios. It can also be observed that the gains are dependent on the traffic type. In Scenario-1, the users not having full buffer traffic have only low-speed voice traffic, which enables a higher amount of SeNBs to remain in an inactive (energy saving) state. For the MBR scenario, however, the other users have a mix of voice and video traffic which consumes a significantly higher amount of resources. The resultant load causes a higher number of SeNBs to be activated, leading to lower gains of 25% at high user density of $N_{\text{UE}} = 60$. The relative energy saving gains increase by up to 8% for the MBR case as the user density increases, compared to the $N_{\text{UE}} = 20$ case. This shows that the mechanism is still effective when the majority of the users have lower traffic rates. Similar trends are observed in the full buffer case, especially because SeNB is activated even if there is a single full-buffer traffic user in its proximity.

In order to give an overview of the potential system energy savings [kWh] that can be obtained using the load-based small cell activation mechanism, the total energy savings over all the eNBs in the considered system is also shown in Fig. 5.3. The energy saving gains are shown for the proposed mechanism, relative to a scenario where no energy saving actions are implemented (All Active) and in comparison with the baseline location-based mechanism. From the figure, we can observe that the gains converge in comparison to the baseline scheme, since at high user densities of full buffer or MBR traffic, all the SeNBs in the proximity of UEs are activated, thereby giving the same performance while using a location or load based scheme. But even in such a scenario, there are significant gains as compared to the case where no energy saving actions are taken. In such cases there might be SeNBs which do not have a proximal UE, thereby consuming energy unnecessarily. These results also highlight the importance of having energy saving mechanisms implemented in heterogeneous networks.

While the basic mechanism considered so far can be applied to any network that has base stations with load-dependent power models, the evaluations still lack the enhancements that could be achieved using dual
connectivity technology. Backhaul links are gaining increasing relevance and importance through the anticipated dense deployment of small cells. Network power consumption optimization must take this into account. The evaluations done in Publication VII address this key consideration while using load- and location-based mechanisms as baseline reference schemes.

In order to present the technology potential of dual connectivity, the throughput distribution results with full buffer traffic and number of picos, $N_F = 10$, are shown in Fig. 5.4, based on results presented in Publication VII. Here we assume that the UEs are able to use all the radio resources allocated to it in a round robin fashion to send data, with the capacity limited based on the Shannon fitting formula presented in [98]. The curves are plotted for user densities of 10 and 50 UEs per MeNB cell, with full SeNB coverage in a network with the dual connectivity feature. From the figure, we can observe that with fixed small cell density there
are significant gains in user throughput with the inter-eNB traffic splitting feature, which is enabled by using dual connectivity. The gains are significantly higher at a low user density of 10 UEs per cell. Gains of more than 20% can be observed for the 5\textsuperscript{th} percentile cell edge users, with significant improvements in throughput for cell center users as well.

As the first case, we consider the ideal fiber optic backhaul presented in [152] to be used for both MeNB and SeNB. Since the sleep mode power values for this type of backhaul were not available, we assumed a scaling factor $\beta$ varying from 0 to 1, indicating the different amounts of power consumed by the backhaul during sleep mode, which would be multiplied with the active mode backhaul power consumption. Thus, $\beta = 0$ indicates that the backhaul link can be fully switched off during sleep mode, and $\beta = 1$ indicates that it cannot be switched-off at all. We also considered two traffic scenarios – static traffic of $R_{cbr} = 128$ kbps, and dynamic traffic with a maximum bit rate of $R_{max} = 512$ kbps, and $R_{min} = 128$ kbps. The results are as shown in Fig. 5.5. Here we consider the location-based small cell activation scheme as the worst possible case, consuming the highest amount of power.

From the figure, we can observe that the user traffic and the backhaul link power consumption during sleep mode have a significant impact on network power consumption. The largest amount of energy savings relative to the location-based scheme (up to 24%) can be observed for the static traffic scenario, when $\beta = 0$ and $N_{\text{UE}} = 30$, since in such scenar-
Energy-Efficient Small Cell Operation

Figure 5.5. System power consumption for symmetric backhaul, normalized to the location-based scheme, with static and dynamic traffic.

ios the location-based scheme tends to activate a significant amount of SeNBs. Such actions could be avoided. There is up to 9% savings relative to the load-based mechanism, highlighting the need for taking the backhaul link power consumption into consideration while optimizing the network power consumption. While the relative gains are lower in the dynamic traffic case, maximum gains of 13% relative to the location-based scheme and up to 11% relative to the load-based schemes are observed, indicating that such mechanisms provide gains at higher user data traffic rates as well. It should also be noted that the gains drop to 6% with $\beta = 0.5$, indicating the significant dependence on backhaul power savings when the radio access network is in sleep mode.

The network power consumption relative to the location-based scheme, with a wired backhaul having different values of $\beta$ and user densities is as shown in Fig. 5.6. The resulting trends remain the same as in the earlier case, with a symmetric backhaul and the absolute values of power savings varying, due to the lower power consumed by such a backhaul. From the
Energy-Efficient Small Cell Operation

Figure 5.6. System power consumption for wired backhaul, normalized to the location-based scheme, with static and dynamic traffic.

In the figure, we can observe that with static traffic and $\beta = 0$ there is a savings of up to 21% compared to the location-based mechanism and 12% compared to the load based mechanism. For $\beta = 1$ the gains (4%) converge with those of the load-based scheme, since the backhaul link power consumption would no longer be a factor for such a scenario. Here it should also be noted that the backhaul link is assumed to enter sleep mode for all the schemes, irrespective of whether or not they were taken into account by the optimization criteria. For $\beta = 0.5$, gains of 13% and 9% are observed, relative to the location- and load-based schemes respectively.

For the dynamic traffic case, the maximum gains observed are reduced to 11% and 5%, due to the relatively lower power savings obtained from the backhaul link nodes entering sleep mode, as well as the higher traffic rates which lead to more cell activations. And at higher user density, all the schemes converge since all the SeNBs are activated.

The normalized power consumption for the fiber-access and wireless backhaul schemes are also shown in Fig. 5.7. Here, it can be observed that the relative gains are lower than for both of the backhaul link types con-
considered previously. Both of these links consume less active mode power than the previous link types. But FA, for example, consumes higher amounts of sleep mode power than SB, $\beta = 0.5$ case, whereas WL consumes more than SB, $\beta = 0.5$ and the same amount as Wrd, $\beta = 0.5$. Thus, the relative gains from these backhaul links entering sleep mode are limited, which also highlights the need for developing backhauling technologies that can coordinate with the radio access network in taking energy saving actions.

From Fig. 5.7(a), we can observe that if we take the fiber-access backhaul link into consideration, maximum power consumption reductions of up to 15% for static traffic and 6% for dynamic traffic can be obtained, compared to the location-based mechanism. The maximum gain compared to the load-based scheme is approximately 4% for both traffic types. It can be noted that the gains would be around 5% for the static traffic case, even at very high user densities of 100 UEs per cell. For the wireless backhaul case shown in Fig. 5.7(b), maximum gains of about 10% and 2% are observed, compared to the location- and load-based schemes. Here the impact of energy savings in sleep mode for the backhaul link is clearly shown in the significantly low gain observed in comparison to the load based mechanism.

Fig. 5.8 shows normalized network power consumption for symmetric backhaul if there is a delay between the MeNB requesting SeNB to switch-off and the time when the SeNB enters the inactive state, when the users are moving at 3 km/h. Here, we also consider a dense deployment of $N_{\text{UE}} = 50$, with users walking in straight lines in random directions. From the figure the impact of mobility on the small cell activation mechanism can be observed, and it can be seen that for the static traffic scenario the proposed load+backhaul-based scheme provides significant gains, even when the switch-off delay is high. While in this work no specific value for deactivation delay was considered in this work, it was shown in [44] that the delay could be up to 1 s. Based on the results presented so far, it can be observed that the small cell operation mechanism could be beneficially deployed in outdoor scenarios, where user mobility is expected, and in indoor scenarios, where users might remain static, while providing significant power consumption reductions compared to the reference schemes.

Normalized mean throughput values for the dynamic traffic scenario are shown in Fig. 5.9. For the static traffic scenario, it is assumed that
the network provides resources for all the users similar to the GBR traffic available in LTE-A systems. For dynamic traffic, the energy-based offloading decisions would have some negative impact on the overall user data rates, since offloading a user to a SeNB would enable more resources to be made available for the remaining users in the MeNB. In order to control this possible deterioration in the QoS, the network could use various traffic offloading criteria, which are not considered in this work. In order to simplify this problem, we consider a traffic offloading factor $\delta_{\text{off}}$ which represents the probability that a user’s traffic is offloaded to the SeNB even when the energy-based offloading condition is not satisfied. Thus, $\delta_{\text{off}} = 0$ would indicate that only the load+backhaul condition is considered, whereas $\delta_{\text{off}} = 1$ would indicate the location-based scheme, where the traffic is offloaded if the UE is in the proximity of an SeNB. The symmetric backhaul case with $\beta = 1$ case is considered here.

From the figure we can observe that the impact on user data traffic rates depends on the user density, with only 10% impact observed when there are 20 UEs per cell. But this can be dynamically controlled by the

Figure 5.7. System power consumption for fiber access and wireless backhaul, normalized to the location based scheme, with static and dynamic traffic.
network by offloading users based on the network tradeoffs policies between network power consumption and user QoS. Here the throughput impacts would depend on the backhaul link type and the user density as well, apart from the traffic type, similar to the energy-saving tradeoffs observed earlier. For example, as the backhaul link power consumption decreases, more users are offloaded, even at medium user densities. This makes no negative impact on user throughput relative to the reference schemes.

5.4 Conclusion

Based on the evaluations done in this section, the relative gains of having energy-efficient small cell operation mechanisms were presented. The importance of having energy-saving schemes, even with relatively sparse de-
ployment of small cells, was shown. Its relevance is expected to continue to grow in the fifth generation networks, with ultra-dense deployment of small cells for delivering high data rates to the end users. The significance of backhaul link power-saving mechanisms, such as sleep modes that are synchronized with the radio access networks, is also clearly evident. While the proposed mechanism was shown to provide significant power-saving gains with limited impacts on user throughput, enhancements which take new network architectures into account, and enable power savings at even shorter time scales need to be investigated. Such work could be especially relevant in the context of 5G networks.
6. Impact of Backhaul on the Evolution Towards 5G Systems

6.1 Introduction

In this chapter, we focus on an efficient operation of the UE in a 5G cloud RAN environment, where the backhaul link between the remote radio heads and the centralized base station pool is non-ideal. The RRH is assumed to have only the basic physical layer capabilities such as channel state estimation, and would be unable to decode packets, in order to make HARQ decisions. Instead, an opportunistic HARQ mechanism is proposed, based on which estimates are made as to whether the UE UL packets should be retransmitted or not, with an appropriate feedback provided to the UE. Thus, the problem is that a HARQ mechanism is required, in the considered non-ideal system, which is as energy-efficient from the UE perspective as the case when the RRH is actually able to decode the packets. Such a system could be built on top of the energy efficient small cell operation mechanism described in Chapter 5, where the backhaul link is assumed to be non-ideal.

The mechanism enables an architectural split in the LTE protocol stack handling uplink HARQ. This helps in reducing the computational requirements at the RRH. Using an opportunistic mechanism, the RRH estimates the probability of successfully decoding the received packet, based on the received signal to noise ratio (SNR), without actually decoding it. Based on the estimate, the RRH sends an ACK/NACK message to the UE, and forwards the received packet (including the HARQ information) to the centralized entity. This enables the implementation of advanced decoding algorithms at the central processor, thereby centralizing processes which require high computational requirements. It also enables reducing the latency requirements on the backhaul interface between RRH and the
centralized base station pool, thereby reducing deployment costs.

Based on performance evaluation, the proposed adaptive scheme is shown to perform close to the optimal HARQ mechanism where accurate estimates are made for retransmissions. It is also shown to perform significantly better than a simple fixed retransmission scheme. This would mean that in such a 5G environment, the investigated scheme would not cause a significant overhead to the UE power consumption, due to unnecessary UL retransmissions. The work presented in this chapter is based on Publication VIII, and is mainly related to the cloud-RAN sub-system. In Publication VIII, the main focus was on enabling uplink HARQ in the constrained environment which gives similar performance as the current state-of-the-art. In this chapter, it is shown that the same methodology could be used to optimize UE uplink power consumption, by minimizing unnecessary HARQ retransmissions in the considered network environment.

6.2 Problem Description

Consider a HARQ system where a maximum of $T$ transmissions per codeword is allowed, with the maximum error probability target of $P_{\text{target}}$ and a given expected effective rate of $R_{\text{eff}}$. The backhaul between the RRH and centralized BS pool where the actual packet decoding is done is non-ideal. The time delay for receiving the feedback result from the centralized BS pool is considered unacceptable, which necessitates local estimation of the packet decoding probability, and sending UL retransmission requests to the UE accordingly. This could lead to a degradation in the UE UL power consumption performance compared to the ideal case, due to the potential unnecessary retransmissions. The main goal is to develop a transmission strategy, that would enable the UE to consume minimum transmit power to support $R_{\text{eff}}$, within $T$ retransmissions. Here the power minimization is with respect to the ideal HARQ mechanism, where retransmissions are sent based on actual decoding of packets.

6.3 Reference Schemes

Two reference mechanisms are described, to compare the performance of the proposed scheme. A simple fixed retransmission strategy is con-
sidered, where retransmissions are sent irrespective of the decoding result. This mechanism represents the worst-case scenario, where the RRH would request the UE for the maximum number of retransmissions. An optimal HARQ scheme is also considered, which represents the ideal case, where it is assumed that retransmissions are sent based on actual decoding results. The main aim is to find the minimum UE UL transmit power for each of these cases, for the given expected effective rate, $R_{\text{eff}}$, the maximum number of retransmissions, $T$, and target outage probability, $P_{\text{target}}$. The minimum required transmit power for the reference schemes could then be compared to the proposed opportunistic scheme, to evaluate the energy efficiency of the mechanism.

An expression which computes the minimum required UE UL transmit power is required to evaluate the schemes. For the $t$:th transmission of a packet over a block-fading channel with SNR $\gamma_t$, let $R_0(\gamma_t)$ be the cut-off rate for energy-constrained Gaussian input signals in an AWGN channel [61, Eq. 7.4.36]:

$$ R_0(\gamma_t) = \left[ 1 + \frac{\gamma_t}{2} - \sqrt{1 + \frac{\gamma_t^2}{4}} \right] \log_2 e + \log_2 \left[ \frac{1}{2} \left( 1 + \sqrt{1 + \frac{\gamma_t^2}{4}} \right) \right] $$

Here $\gamma_t = P_u |h_t|^2$, with channel gain $h_t$ and UE UL transmit power $P_u$ which remains the same all over retransmissions. The cut-off rate can be considered as the lower bound on the capacity of the given channel. It defines the region of rates in which a communication system can operate with an arbitrary probability of error [160]. Thus, cutoff rates provide a capacity bound for finite block length and error probability [170]. While rates higher than cut-off rates are possible, the parameter is used in this work since it is considered valuable for predicting the performance [24, 160].

For $T$ transmission blocks, where the channel during each transmission is modeled using an independent Gaussian channel, we have the following lower bound on the error probability [61, Eq. (7.5.33)]:

$$ \Pr_e(R_{\text{init}}, \gamma(t)) = c_t e^{-NTE_r(R_{\text{init}}, \gamma(t))} $$

$$ E_r(R_{\text{init}}, \gamma(t)) = \frac{1}{T} \sum_{t=0}^{T-1} R_0(\gamma_t) - \frac{R_{\text{init}}}{T} $$

where $N$ is the blocklength, $R_{\text{init}}$ is the initial rate, $E_r$ is the error exponent and $\gamma(t)$ indicates the vector of SNRs of the first $t+1$ transmissions. It is assumed that the constant $c_t = 1$ and is not further detailed here, as it has only minor impact and scales slower than the exponential term.
In order to calculate the minimum UE UL transmit power we need to relate the effective rate with the cut-off rate and the error probability. The effective rate after $T$ transmissions is given by $R_{\text{eff}}(T) = R_{\text{init}}/T$, and the expected effective rate would depend on the expected number of retransmissions. Taking the inverse function of Eq. (6.2) and substituting with eqs. (6.1) and (6.3), we get:

$$R_{\text{init}} = \sum_{t=0}^{T-1} \left\{ \left[ 1 + \frac{\gamma t}{2} - \sqrt{1 + \frac{\gamma t^2}{4}} \right] \log_2 e \right. \left. + \log_2 \left[ \frac{1}{2} \left( 1 + \sqrt{1 + \frac{\gamma t^2}{4}} \right) \right] \right\} + \frac{\ln(Pr_e(R_{\text{init}}, \gamma(t)))}{N} \quad (6.4)$$

For the simple fixed retransmission case, the expected effective rate is the ratio of initial rate and the total number of transmissions that are allowed, $T$. From the above expression, the minimum transmit power for achieving the target effective rate can be calculated for the fixed retransmission case, with $Pr_e(R_{\text{init}}, \gamma(t)) = Pr_{\text{target}}$, the target outage probability. For the optimal HARQ case, the effective rate depends on the the initial rate and the expected number of transmissions [163] to achieve that rate. For this case, the initial rate with the given outage probability constraint $Pr_{\text{target}}$ can be calculated for each transmit power values. Here, the initial rate is chosen based on a target outage probability, rather than the first transmission block error rate (BLER). The expected number of transmissions for the transmit power value can also be calculated based on the initial rate, and the cut-off rate and resultant error probabilities based on eqs. (6.2),(6.3). Based on these, through numerical evaluation, the minimum transmit power at which the ratio of the initial rate and expected number of transmissions equals the target effective rate can be calculated.

### 6.4 Opportunistic HARQ for non-Ideal Backhaul

The opportunistic HARQ mechanism enables the deployment of centralized RAN using non-ideal, high-latency backhaul links, while still minimizing the computational requirements at the RRHs. The possible cost involved here, by estimating the need for retransmissions locally at RRH without packet decoding, would be the higher transmit power consumption from the UE perspective, which needs to be minimized. When HARQ feedback is sent based on estimated error probability rather than actual packet decoding result, there is an additional ambiguity involved that an
ACK response is sent by the RRH but the BS pool is unable to decode the packet. In order to overcome this ambiguity, an outage probability threshold $Pr_{th} < Pr_{target}$ is used in the opportunistic HARQ case.

Another issue is that the RRH cannot estimate the average cut-off rates locally without actually decoding the received packets, since the resultant SNR after $t$ retransmissions is not known to the RRH. If the RRH is able to estimate this independently, the error exponent, and resultant error probability can also be estimated locally at the RRH. Thus, in the constrained environment, the RRH can make the HARQ retransmission decisions based on the expected error probability. In order to achieve this, the concept of effective SNR, which has been used widely in literature [53, 66, 105], is proposed to be used. Here, an estimate of the SNR over all the retransmissions, is used to calculate the cut-off rates in Eq. (6.1), which is then used in Eq. (6.2) to calculate the error probabilities. The effective SNR is computed by taking the maximum of high and low SNR approximations. Thus, the effective SNR is given by:

$$\gamma_{eff} = \max \left( \sum_{t=0}^{T-1} \gamma_t, \left(\frac{\gamma}{4}\right)^T \prod_{t=0}^{T-1} \gamma_t \right)$$  

Here the maximum value is considered in order to get the closed-form lower bound, as both the high and low SNR approximations are lower bounds. The effective SNR estimated could then be used to calculate the average cut-off rates and for calculating the estimated error probability. The estimated error probability can then be compared to the threshold value, $Pr_{th}$, in order to estimate the need for retransmissions. If the estimated error probability is less than $Pr_{th}$, the UE is requested by the RRH to retransmit the UL packet, else the acknowledgment message is sent to the UE. The energy efficiency of this mechanism can be calculated by first evaluating the initial rate as mentioned in the previous section with a given outage probability target, $Pr_{target}$. Then the minimum transmit power at which the estimated number of transmissions that is required to achieve the expected effective rate can be evaluated using the modified error probability threshold for retransmissions, $Pr_{th}$.

For a given initial rate, the opportunistic method could be implemented using a look-up table that maps the effective SNR value with the expected error probability. Based on this mapping, the RRH can determine whether UE UL HARQ retransmissions are required. A simplified flow diagram of how the opportunistic HARQ mechanism could work in practice, is as shown in Fig. 6.1. Here, avoiding the decoding process at the RRH enables
the use of simplified and low-cost hardware. This helps in reducing the CAPEX and OPEX of such deployments, since non-ideal backhaul would also be relatively inexpensive as compared to ideal, low-latency backhaul. While similar concepts such as effective SINR mapping and using BLER constraints to maximize throughput in a system using HARQ have been investigated in the literature, the application of effective SNR to minimize the decoding complexity has not been studied yet.
6.5 Simulation Results

For a fixed expected effective rate, the minimum UE UL transmit power required, $P_u$, with an independent and identical block Rayleigh fading channel, outage error probability target, $P_{\text{target}} = \epsilon = 10^{-4}$, coding over five LTE PRBs, and maximum number of transmissions, $T = \{2, 3, 4\}$, is shown in Fig. 6.2. The probability threshold for retransmissions, used for the opportunistic scheme is given by $P_{\text{th}} = \epsilon/2$.

Here the transmission method is chosen with a target outage probability, $P_{\text{target}}$, and not a target first transmission BLER. This would mean that the first transmission BLER is $T$:th root of $\epsilon$. The BLER target of 10% considered in this work, for $T = 4$, is similar to the value used in [84, 144]. Based on the system-level performance evaluations done in [159], it was shown to provide optimal performance. Using a higher first transmission BLER such as 30% could improve the spectral efficiency of the system, but with a fixed $T$, it would increase the outage probability. The added cost could be an increase in the latency of the system.

The performance of the opportunistic HARQ mechanism with an effective SNR approximation (Opportunistic HARQ, approx.) is compared to the ideal case where the average cut-off rates are known and retransmission decisions are made based on the modified error probability threshold of $P_{\text{th}} = \epsilon/2$, which might not be realistic in actual settings. The mechanism is compared with an optimal HARQ mechanism where feedback is sent based on actual decoding results, which represents the upper bound of the performance that can be achieved. As another reference case, the performance of the static, fixed retransmission mechanism is also presented.

From the figure we can observe that, from an energy efficiency perspective, increasing the value of $T$ in a system with HARQ allowing a maximum of $T$ transmissions with a given expected effective rate $R_{\text{eff}}$, and supported by an imperfect backhaul, $P_u$ needs to be lower. This consequently lowers the energy consumed by the UE to achieve the set expected effective rate. Consider the case where $R_{\text{eff}} = 1$, and $T = 4$. From the figure we can observe that the minimum transmit power $P_u$ required to achieve this for the opportunistic HARQ case is the same as the ideal optimal HARQ scheme.

From the figure, we can also observe that the opportunistic HARQ mechanism performs close to the ideal mechanism, with a slight decrease in
performance for high effective rate targets. In order to achieve the same expected effective rate, the minimum transmit power required for the opportunistic HARQ scheme with effective SNR approximation is about 3.4% higher than the ideal case, for $T = 4$, and approximately 0.5% for $T = 3$ can be observed. Based on these considerations, the opportunistic HARQ mechanism gives almost the same performance as the optimal scheme. Thus, the considered mechanism enables simplified implementation of centralized RAN, with an imperfect backhaul, with no significant impact on the energy consumption of the UE, despite basing HARQ retransmissions on estimated, not decoded block errors.

### 6.6 Conclusions

We consider an opportunistic HARQ mechanism, which provides retransmission feedback to the UE, based on the channel state information, and not on actual decoding of packets. Such an optimization might be relevant for enabling centralized RAN in an environment with a non-ideal backhaul. The main aim is to enable deployments of 5G systems with centralized processors, and RRHs with limited radio capabilities, using low-cost backhaul links with possibly high delay and latency characteristics, without having any negative impacts on UE power consumption. The mechanism is evaluated based on constraints currently defined for 3GPP LTE-Advanced systems, and is compared with optimal HARQ and simple fixed retransmission reference schemes. The core idea is to split the functionalities such that the delay-critical component of sending HARQ retransmission feedback remains at the RRH, while moving the computationally intense part to the centralized and possibly cloud-based virtual base station pool. The mechanism could be considered as a key enabler for low cost and dense deployment of 5G networks that would provision the higher data rate requirements currently envisioned for 5G systems.

In this work, the analytical model of a real system is considered, which gives an indication of the potential gains from using the proposed scheme. Some mechanisms available in such systems, such as uplink transmit power control is not considered in this work, and could be considered as an area for further study. Evaluating the scheme using more realistic system level simulations would also be an interesting area of future work.
7. Conclusions and Future Work

Ultra-dense deployment of small cells for capacity enhancements with macro cells providing coverage and control plane signaling is the direction in which current cellular systems are evolving. In such a paradigm, where small cells are being deployed in a dedicated frequency layer, in contrast to macro cells, energy-efficient and handover signaling optimized mechanisms are required for the overall efficiency of the network. With backhaul and radio access networks becoming increasingly integrated, and due to the increase in backhaul links proportional to the radio access network deployment density, new mechanisms which consider energy-efficient overall network operation are also becoming relevant and important.

The trends in the evolution of next generation 5G systems are also growing in clarity, with system requirements, deployment scenarios, network performance metrics, etc., being defined by various industrial and academic research forums. Current developments indicate that network densification will continue to gain prominence in 5G systems as well, implying the need to further investigate energy-efficient operation mechanisms for the UE and network, which reduce the overall control plane signaling load of the network.

In this work, we first investigate the energy-efficient small cell discovery operation in dense small cell deployment scenarios. Three mechanisms were primarily considered: (i) relaxed measurement gaps, with which the network would configure an infrequent inter-frequency measurement gap for the UE, with the network configuring more frequent measurement gaps upon detection of a small cell (ii) UE mobility state-based measurement gap configuration, where measurement gap configuration depends on the accurate or coarse estimates of the speed of the user, and (iii) small cell proximity estimation based on neighbor cell signal strength measure-
Conclusions and Future Work

ments, called RF fingerprints. A combination of these approaches was also considered, to optimize the signaling load in the network by avoiding handovers when UEs are moving at high speeds.

Evaluations based on LTE-Advanced heterogeneous network settings indicate significant gains in terms of UE power consumption reductions, while optimizing the amount of handovers and related signaling for UEs moving in medium and high mobility states. The small cell connected time performance of the RF fingerprint-based small cell discovery mechanism, used in combination with the UE mobility state-based scheme, was shown to be close to the inter-frequency measurement scheme, with reductions in UE energy consumption. Further optimizations of the proximity estimation scheme based on the fingerprint database location were also investigated, and both network and UE-based approaches were found to be feasible. The small cell proximity estimation mechanism was further generalized to apply to the device-to-device discovery procedure as well, in combination with the relaxed measurement gap mechanism applied for sending discovery beacons in such a scenario. Here, control and user plane location reporting mechanisms were studied. Results indicate that when the fingerprint database for proximity estimation is located on a social cloud server, with application layer location reporting, results in only minimal overhead (in terms of energy consumption and data plane resource utilization).

Energy-efficient small cell operation that takes traffic characteristics and the impact of backhaul power consumption into consideration is also studied. Mechanisms which jointly optimize radio access and backhaul power consumption is presented and evaluated using an LTE-A heterogeneous network setting. Various deployment scenarios, including indoor and outdoor deployment of small cells, dense and sparse deployments, and static and mobile users are also investigated, and related network power consumption values are presented. Evaluations indicate that with the perceived dense deployment of small cells and proportional increase in backhaul link and related power consumption, the joint optimization mechanism can provide significant network power consumption optimizations (up to 20%), with acceptable tradeoffs in terms of the total system capacity. Here the small cell proximity estimation mechanism is considered to be the baseline for the accurate determination of cells to be activated.

Finally, the system considered so far is used as a baseline for 5G central-
ized cloud-based small cell deployments, and further cost optimizations using low cost RRHs deployed using a non-ideal backhaul with a centralized base station pool. Here the key consideration is to keep the delay-critical HARQ component at the RRH, while implementing the computationally intense part at the centralized processor. Simplifying the HARQ procedure by estimating the packet decoding probability, based on channel state feedback, is studied with the constraints of delay requirements currently defined for LTE-Advanced systems. The packet decoding probability is estimated based on mapping SNR with the error probability, using look-up tables. This could be implemented in a real system. Evaluations are done to compare the proposed mechanism with ideal HARQ and static retransmission mechanisms, with the results indicating that the performance is quite similar to that of the optimal scheme. This enables massive and ultra-dense deployment of small cells in a cost-efficient manner with possibly non-ideal backhaul.

Based on the evaluations done on small cell discovery, maintaining the RF fingerprint database has shown significant reductions in energy consumption for the discovery and operation of ultra-dense networks. Further study on small cell discovery procedures could include the implementation of the proposed mechanisms in a real system, making use of network data visualization tools for evaluating the performance of the presented mechanisms. Coordination of such mechanisms with commercial application servers would also be an enabler for provisioning new services, especially using D2D communication procedures. The practical feasibility in terms of deployment and maintenance costs of such new network entities could be an interesting area of future work. Improving the accuracy of the mobility state estimation of UEs, especially at medium mobility state, could also require further study.

Current power consumption models need to be further evaluated with higher system bandwidths to study whether the linear models would still be applicable in such a system. With the development of mmW and cmW small cells, the study of cell discovery and operational power consumption models of such base stations would also be relevant. The joint UE, base station and backhaul power consumption would also be an interesting topic for further study. Currently, high capacity wireless backhaul for small cells is not practical when the backhaul power consumption exceeds that of the radio access. With dense deployment of small cells also envisioned for remote locations where wireless backhaul deployments would
be relevant, energy-efficient wireless backhaul mechanisms in such environments could be a fruitful area of further study.

The practical application of the opportunistic HARQ mechanism and the investigation of realistic gains that can be achieved in terms of RRH computational complexity reduction and impacts on user data rates would be yet another interesting area for future work. The effect of HARQ retransmission prediction errors on the overall system performance could also be further investigated. The impact of mobility on the accuracy of the effective SNR estimates, and the possible related performance impacts could also be an area of further study.
[1] 3GPP. Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN), Sept. 2009. TR 25.913, ver. 9.0.0.


Errata

Publication III

In Fig. 5a, 120 km/h is erroneously printed as 12 km/h.

Publication VIII

In Fig. 3 and related description, $P_{r_{th}}$ is erroneously written as $P_{th}$.

Publication VIII

The effective rate after block $t$ is given by $R_{eff}(t) = \frac{1}{(t+1)}R_{init}$. 